Sparsity-Based Data Reconstruction Models for Biomedical Imaging

Emrah Bostan

Thèse N° 6974 (mai 2016)

Thèse présentée à la faculté des sciences et techniques de l'ingénieur
pour l'obtention du grade de docteur ès sciences
et acceptée sur proposition du jury

Prof. Pierre Vandergheynst, *président*
Prof. Michael Unser, *directeur de thèse*
Prof. Mario Figueiredo, *rapporteur*
Prof. Michael Liebling, *rapporteur*
Dr. Laurent Condat, *rapporteur*

École polytechnique fédérale de Lausanne—2016

Cover design by Annette Unser
Printing and binding by Repro-EPFL
Typeset with LaTeX

Available at **http://bigwww.epfl.ch/**

Abstract

We propose new regularization models to solve inverse problems encountered in biomedical imaging applications. In formulating mathematical schemes, we base our approach on the sparse signal processing principles that have emerged as a central paradigm in the field. We adopt a variational perspective and specify the proposed sparsity-promoting data reconstruction models as energy minimization problems. To design practical algorithms, we develop novel iterative methods to efficiently perform the consequent optimization task.

The thesis is organized in three main parts. In the first part, our main contribution is the introduction of a proper statistically-based discretization paradigm for inverse problems. In particular, our framework considers a continuous-domain stochastic signal model and characterizes a specific class of inference algorithms. We show that derived inference-based methods cover the classical Tikhonov-type techniques as well as a wide range of the sparsity-promoting schemes including the well-known ℓ_1-norm regularization and its nonconvex variants. This provides a unifying stochastic perception of the resolution of inverse problems.

In the second part, we propose a novel phase retrieval algorithm for imaging unstained biological samples that are optically thin. In specific terms, we use the transport-of-intensity equation (TIE) relating the spatial phase map of a field to the derivative of its intensity along the propagation direction. We analyze the implications of using the TIE formalism with finer and coarser approximations of said derivative. Based on this analysis, our contribution is a practical phase reconstruction algorithm that incorporates a sparsity-based regularization. The developed technique operates with a standard bright-field microscope. Experiments on real data illustrate that our phase reconstruction algorithm is viable and can be a low-cost alternative to dedicated phase microscopes.

In the last part, we develop regularization schemes for vector fields, which have an increasing prevalence in medical imaging. In this context, our first contribution is a new regularizer that imposes sparsity on the singular values of the Jacobian of a given vector field. We show that the proposed regularization functional is a valid extension of total variation (TV) regularization to vector-valued functions. We utilize the developed framework for enhancing the streamline visualizations of experimentally acquired 4D flow MRI data. Since vector field regularization requires processing large volumes of multidimensional data, our second contribution is the development of a non-iterative denoising algorithm. In particular, we design model-based tight wavelet frames that are able to remove the spurious divergence content from the vector field, which is of interest in aortic blood flow imaging.

Keywords: Sparsity, linear inverse problems, variational models, iterative algorithms, total variation regularization, convex optimization, phase imaging, phase retrieval, transport-of-intensity, 4D flow MRI, phase-contrast MRI, divergence-free wavelets, vector field regularization.

Résumé

Nous proposons de nouveaux modèles de régularisation pour la résolution de problèmes inverses rencontrés en imagerie biomédicale. Notre formulation mathématique se base sur les principes de parcimonie qui ont récemment émergé comme un paradigme central en traitement du signal. Adoptant une approche variationnelle, nos modèles de reconstruction parcimonieuse s'expriment comme des problèmes de minimisation d'énergie. Les tâches d'optimisation résultantes sont résolues en pratique par des algorithmes basés sur de nouvelles méthodes itératives.

Cette thèse est organisée en trois parties. Dans la première, notre contribution principale est l'introduction d'un formalisme statistique pour la discrétisation des problèmes inverses. En particulier, nous considérons un modèle stochastique continu pour les signaux et caractérisons une classe spécifique d'algorithmes d'inférence. Nous montrons que les méthodes présentées recouvrent les techniques classiques de régularisation Tykhonov ainsi qu'une large classe de schémas de reconstruction parcimonieuse. Cette classe inclue la régularisation en norme ℓ_1 et ses variations non-convexes. Ceci définit un cadre stochastique unifié pour la résolution des problèmes inverses.

Dans la seconde partie, nous proposons un nouvel algorithme d'extraction de phase pour analyser des échantillons biologiques sans coloration et optiquement fins. Plus précisément, nous utilisons l'équation de transport d'intensité (TIE en anglais) qui relie la carte spatiale des phases d'un champ à la dérivée de son intensité le long de la direction de propagation. Nous analysons les implications du formalisme TIE avec des approximations plus fines et plus grossières de cette dérivée. En nous basant sur cette analyse, notre contribution est un algorithme effectif de reconstruction de phase qui identifie des domaines de fréquence adéquats dans les mesures et incorpore la regularisation parcimonieuse. Les techniques développées sont adaptées à des données issues d'un microscope en champ clair standard. Des expériences sur des données réelles illustrent que notre algorithme de reconstruction de phase est viable et peut être une alternative à faible coût aux microscopes à phase.

Dans la dernière partie, nous introduisons des schémas de régularisation pour les champs de vecteurs, qui sont de plus en plus utilisés en imagerie médicale. Dans ce contexte, notre première contribution est un nouveau régularisateur qui induit de la parcimonie sur les valeurs singulières du Jacobien d'un champ vectoriel donné. Nous montrons que la fonctionnelle de régularisation considérée est une extension valide de la régularisation par variation totale (TV en anglais) pour les fonctions à valeurs vectorielles. Nous utilisons cet outil pour améliorer la visualisation de données expérimentales quadridimensionnelles en imagerie à rénonance magnétique (MRI en anglais). De plus, la régularisation de champs de vecteurs requiert l'analyse de grandes quantités de données multidimensionnelles. Notre seconde contribution est donc le développement d'un algorithme non itératif de débruitage adapté pour ce problème. En particulier, nous introduisons des ondelettes redondantes qui sont capables de supprimer la divergence du champ de vecteurs. Ceci est particulièrement intéressant pour l'analyse des flux sanguins aortiques.

Mots clefs : Parcimonie, problèmes inverses linéaires, modèles variationnels, algorithmes itératifs, régularisation par variation totale, optimisation convexe, imagerie de phase, extraction de phase, transport d'intensité, flux IRM à quatre dimensions, imagerie à contraste de phase, ondelettes sans divergence, régularisation de champs de vecteurs.

It is not the brains that matter most, but that which guides them—the character, the heart, generous qualities, progressive ideas.

Fyodor Dostoyevsky

To my family, in recognition of their support.

Acknowledgement

This thesis would not have been written without the support and encouragement of many people. I take this opportunity to express my gratitude to all of them.

First and foremost, I thank my advisor Prof. Michael Unser. I would like to acknowledge his enthusiasm for research and science, which has been a constant source of motivation for me. He has created a unique atmosphere in the group, where I have seen a continual flow of original ideas. In particular, I value his advice on how to properly present scientific opinions. On top of all these, I thank him for his considerate, generous, and respectful attitude.

I express my sincere thanks to the president of the thesis committee, Prof. Pierre Vandergheynst, and the official members, Prof. Mario Figueiredo, Prof. Michael Liebling, and Dr. Laurent Condat, for accepting to review the thesis.

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement nº 267439.

I thank the fellow past and present members of the Biomedical Imaging Group (BIG), Prof. Arash Amini, Anais Badoual, Dr. Jean-Charles Baritaux, Dr. Aurlien Bourquard, Dr. Nicolas Chenouard, Dr. Ning Chu, Dr. Ricard Delgado Gonzalo, Prof. Adrien Depeursinge, Laurene Donati, Julien Fageot, Dr. Denis Fortun, Emmanuel Froustey, Dr. Matthieu Guerquin-Kern, Harshit Gupta, Dr. Ulugbek Kamilov, Dr. Hagai Kirshner, Dr. Stamatis Lefkimmiatis, Dr. Ha Nguyen, Dr. Michael McCann, Junhong Min, Dr. Masih Nilchian, Pedram Pad, Zsuzsanna Puspoki, Dr. Daniel Sage, Daniel Schmitter, Prof. Chandra Sekhar Seelamantula, Dr. Ferreol Soulez, Dr. Martin Storath, Dr. Pouya Tafti, Dr. Philippe Thevenaz, Virginie Uhlmann, Dr. Cedric Vonesch, Dr. John Paul Ward, and Luc Zeng.

I thank my friends, Zafer Dogan, Arda Gunay, Kerem Kapucu, and Eray Molla, who have been absolutely fantastic for as long as I know them.

Finally, I would like to thank my girlfriend, Stéphanie Schneider, my brother, Emir Bostan, my father, Halis Bostan, and my mother, Gungor Bostan, for their unconditional support. This work is humbly dedicated to them.

Contents

Abstract **i**

Résumé **iii**

Acknowledgement **ix**

1 Introduction **1**

1.1 Main Contributions 3
1.2 Organization of the Thesis 5

2 Discrete Linear Inverse Problems **7**

2.1 Overview 7
2.2 Introduction 8
  2.2.1 Forward Problem 8
  2.2.2 Variational Reconstruction 10
  2.2.3 Outline 11
2.3 Linear Reconstruction Methods 11
  2.3.1 Quadratic Regularizations 11
  2.3.2 Bayesian Inference with Gaussian Models 14
  2.3.3 Completing Notes on Linear Reconstructions 18
2.4 Sparsity-Based Reconstruction Methods 20
  2.4.1 Inducing Sparsity via Regularization 21
  2.4.2 Wavelet-Based Regularization 22
  2.4.3 Derivative-Based Regularization 24

2.5 Iterative Optimization Methods 26
    2.5.1 Fast Iterative Shrinkage/Thresholding Algorithm 26
    2.5.2 Alternating Directions Method of Multipliers 28
2.6 Summary 30

3 Sparse Reconstructions: Continuous-Domain Links 33
3.1 Overview 33
3.2 Introduction 34
    3.2.1 Contributions 35
    3.2.2 Outline 36
3.3 Measurement Model 37
    3.3.1 Discretization of the Signal 37
    3.3.2 Discrete Measurement Model 39
3.4 Sparse Stochastic Models 40
    3.4.1 Continuous-Domain Innovation Model 40
    3.4.2 Statistical Distribution of Discrete Signal Model 43
    3.4.3 Illustrative Examples 47
3.5 Bayesian Estimation 48
    3.5.1 MAP Formulation 49
    3.5.2 Potential Functions 50
3.6 Reconstruction Algorithm 51
3.7 Summary 54

4 Sparse Reconstructions: *In-Silico* Considerations 57
4.1 Overview 57
4.2 General Framework 58
4.3 Computational Recipes 59
    4.3.1 Analysis of Circulant Operators 59
    4.3.2 Operator-Specific Regularization 61
    4.3.3 Vectorial Proximal Mappings 62
    4.3.4 Efficient Calculation of Matrix Inverse 64
4.4 Numerical Results 67
    4.4.1 Image Deconvolution 67
    4.4.2 MRI Reconstruction 70
    4.4.3 X-Ray Tomographic Reconstruction 72
4.5 Discussion 75

5 Variational Phase Imaging 77
5.1 Overview . . . 77
5.2 Introduction . . . 78
5.2.1 Contributions . . . 81
5.2.2 Outline . . . 82
5.3 Physical Models . . . 83
5.3.1 Transport-of-Intensity Equation . . . 83
5.3.2 Contrast Transfer Function . . . 86
5.4 Reconstruction Algorithm . . . 89
5.4.1 Spectral Weighting Filters . . . 89
5.4.2 Discrete Formulation . . . 90
5.4.3 Optimization Algorithm . . . 93
5.5 Experiments . . . 96
5.5.1 Synthetic Data . . . 96
5.5.2 Real Data . . . 99
5.5.3 Validation . . . 102
5.6 Summary . . . 104

6 Sparse Reconstruction of Vector Fields 107
6.1 Overview . . . 107
6.2 Introduction . . . 108
6.2.1 Contributions . . . 110
6.2.2 Outline . . . 110
6.3 Flow-Field Regularization . . . 111
6.3.1 Curl- and Divergence-Based Methods . . . 111
6.3.2 Gradient-Based Methods . . . 112
6.4 Jacobian-Based Regularization . . . 113
6.4.1 Connections with the Existing Methods . . . 116
6.4.2 Nuclear Total Variation . . . 117
6.5 Reconstruction Algorithm . . . 119
6.6 Experiments . . . 120
6.6.1 Synthetic Data . . . 121
6.6.2 Real Data . . . 123
6.7 Discussion . . . 123

7 Non-Iterative Model-Based Sparse Flow Reconstruction 131
7.1 Overview . . . 131
7.2 Introduction . . . 132
  7.2.1 Contributions . . . 133
  7.2.2 Outline . . . 133
7.3 Mathematical Framework . . . 134
  7.3.1 Basic Definitions . . . 134
  7.3.2 The Leray Projector . . . 135
7.4 Divergence-Free Wavelets . . . 135
  7.4.1 Decay estimates . . . 137
  7.4.2 Wavelet Implementation Details . . . 138
7.5 Sparse Vector Field Regularization . . . 138
  7.5.1 Variational formulation . . . 138
7.6 Numerical Results . . . 139
  7.6.1 Scale-Adapted Thresholding . . . 139
  7.6.2 Comparison with the Bi-Orthogonal Design . . . 140
7.7 Summary . . . 140

8 Conclusion 147
8.1 Summary of Results . . . 147
8.2 Outlook . . . 149

Bibliography 151

Curriculum Vitæ 169

Chapter 1

Introduction

This thesis is concerned with practical algorithms for *inverse problems* in biomedical imaging. Adopting a suitable language, the concept of such problems is to "invert" the data acquisition process that relates an unknown image to observable measurements [1]. By doing so, one obtains the spatial distribution of physical parameters. Reconstruction of the refractive index distribution inside a cell (the physical parameter) from its diffraction tomograms (the measurements) is a notable example [2]. Since we retrieve information about entities—via solving inverse problems—that we are unable to observe directly, the topic is among the most intensively studied mathematical problems in the field [3].

For a large class of biomedical imaging applications, the data acquisition process is modeled (or well-approximated) by *linear* models [4]. Still, linear inverse problems are not straightforward as they are generally ill-posed. This implies that there may exist several solutions with high data fidelity in the sense that they all would produce hypothetical measurements that are close to the observed ones. It is thus necessary to look for solutions that are not only consistent with the measurements, but also satisfy additional constraints. In mathematical terms, the main idea is to formulate an *energy functional* that measures both the data fidelity and the discrepancy with respect to the desired attributes. The solution is then sought as a minimizer of the energy functional so that preferable solutions are discriminated from spurious ones. The framework is broadly termed *variational image reconstruction.*

The way one measures the data fidelity is dictated by the noise characteristics of the imaging system, leaving not much room for design. Therefore, the quality of the final image reconstruction is determined by the constraints imposed on the solution. Principally, they are introduced in two ways: (a) in the *deterministic* approaches, one regularizes the solution based on physical or geometric features. For instance, it can be preferred that the reconstructed image has bounded variations or that its values are non-negative; (b) in the *statistical* approaches, images are modeled as random quantities. By introducing a prior distribution that describes our "belief" on the solution, one draws statistical inference about the unknown image. A simple example is to prefix spatial regions in which the solution is located with a high probability.

In the traditional deterministic (Tikhonov-type) regularization schemes, we utilize quadratic norms for measuring both the data fidelity and the reconstruction regularity. This provides us with a differentiable optimization problem. The solver is linear and is evaluated efficiently. Remarkably, the use of the linear reconstruction is also justified from a statistical perspective. Under the hypothesis of Gaussianity and stationarity, the quadratic regularization methods are derived as statistical estimators that are based on the Bayesian inference. As we shall detail later, the statistical model establishes that the derived linear scheme is optimal in the mean squared error sense. Moreover, it is the best linear algorithm (in terms of the same error metric), even if the Gaussian assumption is violated [5]. By combining the outcomes of the two schools in a complementary fashion, a firm understanding of the linear techniques has been gained.

In recent years, the research efforts in variational image reconstruction have capitalized on the *sparsity-promoting* algorithms rather than the classical linear ones. The trend is motivated by the observation that naturally occurring signals admit sparse or nearly-sparse representations in some transform domain [6]. The promotion of sparsity is established by using well-chosen non-quadratic regularizers. The practical outcome of such approaches is nonlinear reconstruction methods that demonstratively outperform their linear counterparts [7]. Therefore, nonlinear frameworks are now central to the inverse problems in biomedical imaging.

The present formulations of sparsity-promoting regularization are based on solid principles of functional analysis and approximation theory. They are, however, primarily deterministic. In a purely discrete setting, it is also possible to interpret them in statistical terms as *maximum a posteriori* (MAP) estimators, for instance, by using generalized Gaussian or Laplace distributions as priors [8, 9]. Still, in a certain regard, these models are dependent on the subsequent reconstruction task since they are linked to the choice of a given sparsifying transform. The apparent drawback is that they do not provide further understanding on the true nature of the underlying object. In addition, sparsity-promoting reconstructions necessitate the development of efficient algorithms. The latter is also emphasized by the ever-increasing dimensionality of the measurements. In other words, practical and sparsity-driven image reconstruction methods persistently call for algorithmic solutions with affordable computational complexity.

In this thesis, we attempt to address these problems in a principled manner. To have an expanded perspective over the challenges, we shall be placed at the interplay of the deterministic and the stochastic approaches. Particularly, we specify statistical estimators—based on *continuous-domain stochastic models*—to strengthen the links with the current sparsity-promoting algorithms. Our goal being to solve real world problems, we also allow ourselves to develop deterministic sparsity-promoting methods whose statistical interpretations are provided to the best of our ability. We shall constantly see that the sparsity-based methods ultimately translate into high-dimensional nonlinear optimization problems. Irrespective of the chosen formalism, we thus develop iterative algorithms—based on state-of-the-art tools in convex optimization—that are implemented efficiently.

1.1 Main Contributions

The contributions of this thesis have been grouped under three main headings:

- *Stochastic reinterpretation of sparsity-promoting algorithms.* The recent theory of continuous-domain sparse stochastic processes characterizes an object as the solution of a linear stochastic differential equation [10]. Based on this model, our contribution is the introduction a statistically-based discretization paradigm for linear inverse problems. Once the problem is discretized, we derive a class of MAP estimators. We show that there is a critical outcome of specifying an upfront continuous-domain stochastic model: the class of admissible prior distributions for the discrete problem is restricted and thus cannot be arbitrary. In effect, the said class is necessarily linked to the infinitely divisible prior distributions. The striking aspect is that our estimators still cover both the classical Tikhonov-type methods and the sparsity-promoting schemes that are currently of use. We formulate an algorithm that handles the corresponding optimization problems, which can be nonconvex. We compare the performance of estimators associated with models of increasing sparsity. Particularly, we show that sparsity-promoting algorithms are powerful tools for solving biomedical image reconstruction problems. However, encouraging sparse solutions does not always increase the performance.

- *Digital phase imaging of cells.* Guided by the above-mentioned elements, we present a novel phase retrieval algorithm for imaging optically-thin transparent objects. Our formulation starts with the transport-of-intensity equation (TIE) [11]. The model relates the spatial phase map of a field to its intensity variation along the propagation direction. The latter is measured by using a set of bright-field defocus images. Accordingly, we investigate the effect of the defocus distance on the retrieved phase map. Our contribution is a sparsity-driven weighted phase reconstruction algorithm that nonlinearly combines different ranges of spatial frequencies depending on the defocus value of the measurements. Using both simulated and real microscopic data, we show that the method outperforms commonly used TIE-based phase reconstruction algorithms. One of our important findings is that the developed numerical phase imaging technique is practical and has a lower cost than the hardware-based solutions such as digital holographic microscopy (DHM).

- *Sparsity-based reconstruction of vector fields.* Vector fields bear an increasing appearance (directly or indirectly) in medical imaging applications including 4D flow MRI [12]. We hence develop variational schemes for vector fields by expanding our insights into the image reconstruction. Our contribution within the context is twofold: (a) inspired by the well-known total variation regularization for images, we propose a new regularizer for vector fields. The regularizer, termed *nuclear total variation* (TVN), imposes sparsity on the singular values of the Jacobian of the field. Our simulations reveal that TVN provides improved denoising performance compared to already existing vectorial extension of TV. We also illustrate the applicability of the framework to real data; (b) since vector field reconstruction requires processing large volumes of multidimensional data, we develop a non-iterative denoising algorithm for incompressible flows. Such objects have practical importance in flow MRI. We construct tight wavelet frames based on physical principles.

We show that our construction is much simpler and yields better denoising results than the available approaches.

1.2 Organization of the Thesis

The thesis is organized as follows: In Chapter 2, we provide the links between the stochastic and the deterministic approaches for discrete linear inverse problems. We also review the recently developed convex optimization techniques within a general setting. In Chapter 3, we introduce a class of MAP estimators that combine the continuous-domain stochastic processes with a discretization framework. The key concept that provides links with sparsity-promoting algorithms is the use of infinitely divisible distributions, which we shall explain in detail. In Chapter 4, we apply the developed MAP estimators to deconvolution, MRI, and X-ray tomographic reconstruction problems and compare the performance of estimators associated with models of increasing sparsity. In Chapter 5, we consider the problem of optical phase retrieval and review the TIE approach. Our analysis on the physical model allows us to derive a novel phase reconstruction algorithm whose applicability is illustrated and validated in practical configurations. In Chapter 6, we formulate the TVN regularization. We illustrate that TVN has the ability to enhance real data visualizations. In Chapter 7, we present a denoising technique for incompressible vector fields by using wavelet tight frames that are specifically constructed to be divergence-free.

Chapter 2

Discrete Linear Inverse Problems

2.1 Overview

In this chapter, we introduce the discrete linear inverse problem framework that lies at the core of several practical problems studied in the thesis. Our consideration starts by introducing the general concept of variational image reconstruction where the solution is obtained by minimizing an energy functional. In designing the energies, we first review the classical techniques where the consequent minimization task is linear. During the process, we explain the deterministic and stochastic perspectives and point out the fundamental links between these approaches. We then focus on the recent reconstruction models that are motivated by sparsity precepts. Since such methods are specified by nonlinear minimization problems, we revisit state-of-the-art iterative optimization algorithms that enable efficient resolution. The chapter not only lays the algorithmic foundations for what follows, but also recognizes the need for the contributions presented in the coming chapters.

2.2 Introduction

We have noted in Chapter 1 that the inverse problem formalism is encountered in various biomedical imaging applications [4]. The common conceptualization of an inverse problem is to consider it as the "inverse" of a forward (or direct) problem. Such apprehension (even though it is rather hasty) specifies the starting point of our analysis.

2.2.1 Forward Problem

The forward problem is concerned with the formulation of the data acquisition step that relates our quantity of interest to some measured data.[1] The general mathematical model of the process is given by

$$\mathbf{z} = \mathrm{H}(s), \tag{2.1}$$

where s represents the exact spatial (or spatio-temporal) distribution of the quantities and is an element in the image[2] space $\mathcal{P}$ (i.e., the space that contains the physical parameters we want to reconstruct). Similarly, $\mathbf{z}$ denotes the ideal (noise-free) measurements that lies in the measurement space $\mathcal{M}$. The mathematical description of the acquisition is expressed by the mapping $\mathrm{H} : \mathcal{P} \to \mathcal{M}$. In this setting, the forward problem aims to find an accurate characterization of H (typically in terms of operators) that is, the process of extracting the measurements for a given $s \in \mathcal{P}$. Accordingly, the goal of an ideal inverse problem is to recover s given $\mathbf{z}$ and H. Such schemes are called hybrid systems as one reconstructs a continuous-domain object from measurements that are discrete quantities.

In a more realistic scenario, the formation of the measurements is given by

$$\mathbf{y} = \mathrm{H}(s) + \mathbf{n}, \tag{2.2}$$

where $\mathbf{n}$ models the additive noise and the vector $\mathbf{y} = \mathbf{z} + \mathbf{n} \in \mathbb{R}^M$ denotes the noisy measurements. Practical methods use numeric computations for solving inverse problems and operate with an entirely discrete model, going a step beyond their hybrid analogues (2.1) and (2.2).

---

[1] For example, the quantity of interest can be the refractive index distribution of a cell, whereas the measurements are diffraction tomograms as mentioned in Chapter 1.

[2] Note that we shall use the terms "image" and "signal" interchangeably throughout the thesis. In any case, both are used in a general and unifying sense.

The most basic formulation of a discrete linear inverse problem is

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{2.3}$$

where $\mathbf{s} \in \mathbb{R}^N$ is the discrete representation of s (for instance, vectorized form of its sample values) and the system matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$ is the discrete version of the forward operator.[3] The transition from (2.2) to (2.3) is called the discretization step, which requires a precise understanding of the underlying physics. It will be addressed in Chapter 3 by constructing concrete examples. Meanwhile,—purely for didactic purposes—the discretization is considered to be performed appropriately. In the thesis, we assume that $\mathbf{n}$ is Gaussian.[4]

The goal is then to properly invert (2.3) in order to recover $\mathbf{s}$ from the observed measurements. However, as the measurements contain noise, it is generally not possible to find the exact solution so that one seeks an approximation of $\mathbf{s}$. Still, the task is not trivial. A substantial body of inverse problems are ill-posed,[5] implying that a simple inversion of $\mathbf{H}$ (if it exists) does not usually yield meaningful solutions. Therefore, we need to introduce additional constraints in the problem. In doing so, the main purpose is to state some desired attributes for the solution so that better reconstructions are obtained. This is achieved by formulating a suitable variational problem, from which the sought approximation of $\mathbf{s}$ is computed.

---

[3] We assume that the physical response of the acquisition device is linear. Even in nonlinear cases, it is common practice to linearize the problem, which explains the special emphasis put on linear inverse problems [1].

[4] Gaussian noise is a standard assumption in the field. All of the algorithms that are going to be developed in the thesis can be easily modified to work under Poisson noise assumption [13].

[5] Three main reasons for ill-posedness are as follows:

1. A solution does not exist: As a result of the noise, $\mathbf{y}$ is not an element of the space spanned by the columns of $\mathbf{H}$ (i.e., the measurement domain).

2. Several solutions exist: The nullspace of $\mathbf{H}$ is not empty so that $\mathbf{y}$ cannot be uniquely explained.

3. Solutions are not stable: Small changes in $\mathbf{y}$ result in large changes in the reconstructed image.

2.2.2 Variational Reconstruction

In variational approaches, the solution $\mathbf{s}^\star$ is specified via the following optimization problem:

$$\mathbf{s}^\star = \arg\min_{\mathbf{s}} \ \mathcal{J}_\tau(\mathbf{s};\mathbf{y}), \tag{2.4}$$

where

$$\mathcal{J}_\tau(\mathbf{s};\mathbf{y}) = \mathcal{D}(\mathbf{s};\mathbf{y}) + \tau\mathcal{R}(\mathbf{s}) \tag{2.5}$$

is called the energy (or the cost) functional. In (2.5), the data fidelity term $\mathcal{D}$ measures the consistency of a reconstruction with respect to the measurements whereas the regularization term $\mathcal{R}$ imposes constraints on the solution to favor specific type of solutions. The regularization parameter $\tau > 0$ controls the trade-off between these two terms.

The reconstruction framework in (2.4) bears fundamental importance for solving inverse problems [3] and all of the algorithms developed in the thesis will be in line with this formulation. Considering the central role of (2.5), three aspects need to be addressed attentively:

- An essential step is to specify $\mathcal{R}$ since the characteristics of the final image reconstruction are determined by the chosen regularization strategy. On one hand, the regularizer should provide a precise mathematical description of the prior information. On the other hand, the designed regularization functional should result in tractable algorithms with low computational complexity. For instance, it is highly advantageous to have a convex regularizer.

- We also need to establish how one measures the data fidelity. Similar to the specification of the regularization functional, the term $\mathcal{D}$ should provide us with accurate modeling properties and affordable computational burden.

- Inverse problems (especially in biomedical imaging) usually give rise to large-scale optimization problems. Once the above mentioned ideas are formulated, we must then develop an efficient algorithm for computing the reconstruction. The task should be worked out by taking the underlying structure of the minimization problem into account.

2.2.3 Outline

Our intent in the rest of the chapter is to communicate an understanding of the regularized signal reconstruction concepts. Based on this point, the chapter is organized as follows: In Section 2.3, we first provide an overview of the conventional linear reconstruction methods that are derived from a deterministic perspective. We then explain how variational reconstruction can be recast within a statistical framework. In particular, we bridge the derived statistical reconstruction techniques to the former deterministic approaches by using Gaussian prior models. In Section 2.4, we move to the current nonlinear image reconstructions that are driven by sparsity considerations. Finally, we present the principles of iterative optimization algorithms that are well-suited for nonlinear reconstruction frameworks.

2.3 Linear Reconstruction Methods

In the sequel, we consider the classical reconstruction techniques that are linearly dependent on the measurements. We review the deterministic and stochastic approaches and highlight the functional equivalence between the two methodologies.

2.3.1 Quadratic Regularizations

We begin with the conventional deterministic techniques that involve the use of quadratic norms. We first analyze the case where we do not impose any regularization on the solution by means of a regularization functional.

The most basic reconstruction is given by the least squares (LS) method that is obtained as follows:

$$\begin{aligned} \mathbf{s}^{\star}_{\mathrm{LS}} &= \arg\min_{\mathbf{s}} \ \mathcal{J}_{\mathrm{LS}}(\mathbf{s}) \\ &= \arg\min_{\mathbf{s}} \ \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2, \end{aligned} \tag{2.6}$$

where $\|\cdot\|_2$ is the standard ℓ_2-norm. The LS formulation simply looks for a solution that produces hypothetical measurements that are as close to the observed ones as possible in the Euclidean distance sense. The first step to solve (2.6) is to compute the derivative of the energy with respect to $\mathbf{s}$, which is given by

$$\frac{\partial}{\partial \mathbf{s}} \mathcal{J}_{\mathrm{LS}}(\mathbf{s}) = \mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{s} - \mathbf{y}).$$

Equating this to zero, the solution satisfies that

$$(\mathbf{H}^{\mathrm{T}}\mathbf{H})\mathbf{s}^{\star}_{\mathrm{LS}} = \mathbf{H}^{\mathrm{T}}\mathbf{y}. \tag{2.7}$$

When $\mathbf{H}$ is full column rank, the square matrix $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ is invertible. The LS solution is then given by

$$\begin{aligned}\mathbf{s}^{\star}_{\mathrm{LS}} &= (\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{y}\\ &= \mathbf{s} + \underbrace{(\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{n}}_{\tilde{\mathbf{n}}}.\end{aligned} \tag{2.8}$$

However, $\mathbf{H}$ is badly conditioned in many problems, which means that $\mathbf{H}$ has small singular values [14]. This implies that $\tilde{\mathbf{n}}$ represents nothing but an amplified noise term as a result of the inversion of $\mathbf{H}^{\mathrm{T}}\mathbf{H}$. Specifically, the amplification occurs in the directions that correspond to the small singular vectors of $\mathbf{H}$. As a consequence, the LS reconstruction performs poorly in general [15].

If we consider the situation when $\mathbf{H}$ is not full column rank, the solution of (2.8) is not unique. The LS approach (i.e., the quadratic data fidelity term by itself) is then inadequate and a regularization functional is needed for limiting the noise-driven errors.

Tikhonov Regularization

Tikhonov regularization is the classical approach for solving ill-posed inverse problems. As an addition to the LS energy, it incorporates a quadratic regularization term. In this type of methods, the general form of reconstruction is defined as

$$\mathbf{s}^{\star}_{\mathrm{Tik}} = \arg\min_{\mathbf{s}}\ \mathcal{J}_{\mathrm{Tik}}(\mathbf{s}) \tag{2.9}$$

$$= \arg\min_{\mathbf{s}}\ \frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{s}\|_2^2 + \tau\|\mathbf{R}\mathbf{s}\|_2^2, \tag{2.10}$$

where $\mathbf{R}$ is called the regularization operator that acts linearly on $\mathbf{s}$.

To get an elemental insight into (2.9), we state the optimization problem:

$$\min_{\mathbf{s}}\ \frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{s}\|_2^2 \text{ subject to } \|\mathbf{R}\mathbf{s}\|_2^2 \le \varrho^2, \tag{2.11}$$

where ϱ is a positive constant. This new problem in (2.11) is a constrained version of the LS formulation and its Lagrangian functional is given by

$$L(\mathbf{s}, \tau) = \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2 + \tau\left(\|\mathbf{R}\mathbf{s}\|_2^2 - \varrho^2\right),$$

where τ is the Lagrange multiplier. More importantly, for a suitable value of τ, one can show that the solution of (2.11) is equal to $\mathbf{s}^\star_{\mathrm{Tik}}$ under the mild condition that $\varrho^2 \leq \|\mathbf{s}^\star_{\mathrm{LS}}\|_2^2$. Also, the relation between ϱ and τ is monotonic [15].

We see that the Tikhonov-type regularizations effectively introduce an upper-bound on the ℓ_2-norm of the quantity $\mathbf{R}\mathbf{s}$. We also notice that $\mathbf{s}^\star_{\mathrm{Tik}}$ becomes $\mathbf{s}^\star_{\mathrm{LS}}$ as $\tau \to 0$. In contrast, if $\tau \to \infty$, the minimization of (2.9) amounts to finding the reconstruction for which $\mathbf{R}\mathbf{s}$ has the smallest ℓ_2-norm (i.e., maximum amount of regularity). Since the regularity criteria is communicated in the range of $\mathbf{R}$, it is determined in relation to the chosen prior information:

- A straightforward choice is the identity operator. In this situation, the regularization functional simply introduces a constraint on the energy of the reconstruction. Potential solutions of high energy are discarded.
- An often preferred choice is the first- or second-order derivative operator. Such designs encourage the final reconstruction to have lower variations. The regularizer prefers solutions that have a certain degree of smoothness.

As for the resolution step, we remark that the energy functional is convex and its derivative reads

$$\frac{\partial}{\partial \mathbf{s}}\mathcal{J}_{\mathrm{Tik}}(\mathbf{s}) = \mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{s} - \mathbf{y}) + 2\tau\mathbf{R}^{\mathrm{T}}\mathbf{R}\mathbf{s}.$$

The formal Tikhonov solution is a linear mapping that is written as

$$\mathbf{s}^\star_{\mathrm{Tik}} = (\mathbf{H}^{\mathrm{T}}\mathbf{H} + 2\tau\mathbf{R}^{\mathrm{T}}\mathbf{R})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{y}. \tag{2.12}$$

Under the assumption that the intersection of the nullspaces of $\mathbf{H}$ and $\mathbf{R}$ is trivial (i.e., $\mathrm{Ker}(\mathbf{H}) \cap \mathrm{Ker}(\mathbf{R}) = \{\mathbf{0}\}$), the solution is unique for the system of linear equations given in (2.12), allowing us to appreciate the role of the regularization in comparison to the LS approach.

2.3.2 Bayesian Inference with Gaussian Models

As explained in Section 2.2.1, our primary interest of recovering the unobservable signal $\mathbf{s}$ from the observation vector $\mathbf{y}$ is often ill-posed. To overcome this complication, we now consider statistical inference techniques where one models $\mathbf{s}$ as a random object. The goal of inference is to estimate certain type of realizations of this random object—among infinitely many others—through an assessment of uncertainty. Particularly, to obtain a meaningful realization (that is the reconstruction), the evaluation of uncertainty is based on the measurement model and our prior information about $\mathbf{s}$ encoded by a *prior* probability distribution function (pdf) p_S.

Bayesian estimators specify the solution of an inverse problem through the probability distribution describing $\mathbf{s}$ after all our available information has been incorporated into the problem. Using Bayes' rule, this distribution, called the *posterior*, is given by

$$p_{S|Y}(\mathbf{s}\,|\,\mathbf{y}) \propto p_{Y|S}(\mathbf{y}\,|\,\mathbf{s})p_S(\mathbf{s}), \tag{2.13}$$

where $\propto$ implies equality after normalization, and $p_{Y|S}(\mathbf{y}\,|\,\mathbf{s})$ is the conditional distribution of the measurements given $\mathbf{s}$. Based on the posterior (2.13) that provides a complete statistical characterization of the problem, one can formulate different estimators.

In the sequel, we derive well-established Bayesian estimators that are based on a Gaussian prior model for $\mathbf{s}$. These classical estimators are of interest not only on their own right, but also to interpret and analyze the deterministic regularization methods introduced in Section 2.3.1 from the Bayesian point of view. This will help us understand the implicit assumptions of quadratic regularization methods, which is the main goal of the present section.

To begin, let us first note that an N-dimensional multivariate Gaussian pdf is specified by

$$G(\mathbf{x}\,|\mathbf{m_x}, \mathbf{C_{xx}}) = \frac{1}{\sqrt{(2\pi)^N |\det{(\mathbf{C_{xx}})}\,|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{m_x})^{\mathrm{T}}\mathbf{C}_{\mathbf{xx}}^{-1}(\mathbf{x}-\mathbf{m_x})\right), \tag{2.14}$$

where the parameter $\mathbf{m_x} = \mathbb{E}\{\mathbf{x}\}$ denotes the mean, and the $N \times N$ covariance matrix

$$\mathbf{C_{xx}} = \mathbb{E}\left\{(\mathbf{x}-\mathbf{m_x})(\mathbf{x}-\mathbf{m_x})^{\mathrm{T}}\right\}$$

is assumed to be positive definite.[6] We also note that $\mathbf{x} \sim \mathcal{N}(\mathbf{m_x}, \mathbf{C_{xx}})$ indicates that the random variable $\mathbf{x}$ follows such a multivariate Gaussian distribution.

Considering the prior signal model, we assume that $\mathbf{s} \sim \mathcal{N}(\mathbf{m_s}, \mathbf{C_{ss}})$. The additive Gaussian noise in (2.3) is assumed to be zero mean so that $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{C_{nn}})$. Our final assumption is that $\mathbf{s}$ and $\mathbf{n}$ are mutually independent.

MAP Estimation with Gaussian Priors

The most often used Bayesian approach, called the *maximum a posteriori* (MAP) estimator, is to seek the maximizer of the posterior that is defined as

$$\begin{aligned} \mathbf{s}^\star_{\mathrm{MAP}} &= \arg\max_{\mathbf{s}} \; p_{S|Y}(\mathbf{s} \,|\, \mathbf{y}) \\ &= \arg\max_{\mathbf{s}} \; \log\big(p_{Y|S}(\mathbf{y} \,|\, \mathbf{s}) p_S(\mathbf{s})\big) \\ &= \arg\min_{\mathbf{s}} \; -\log\big(p_{Y|S}(\mathbf{y} \,|\, \mathbf{s})\big) - \log\left(p_S(\mathbf{s})\right), \end{aligned} \tag{2.15}$$

where the second equality follows from the monotonicity of the log function. The remarkable aspect is that the MAP formulation in (2.15) is completely in line with the generic variational reconstruction model in (2.5),[7] explaining the popularity of such estimators.

Next, we derive the conditional distribution $p_{Y|S}(\mathbf{y} \,|\, \mathbf{s})$ based on our Gaussian model. Due to the independence of $\mathbf{s}$ and $\mathbf{n}$, we write that

$$\begin{aligned} p_{Y|S}(\mathbf{y} \,|\, \mathbf{s}) &= p_{Y|S}(\mathbf{Hs} + \mathbf{n} \,|\, \mathbf{s}) = p_N(\mathbf{y} - \mathbf{Hs}) \\ &= \big((2\pi)^N |\det\left(\mathbf{C_{nn}}\right)|\big)^{(-1/2)} \, \mathrm{e}^{-\frac{1}{2}(\mathbf{y}-\mathbf{Hs})^{\mathrm{T}}\mathbf{C_{nn}^{-1}}(\mathbf{y}-\mathbf{Hs})} \end{aligned}$$

Therefore, by ignoring the constants that are irrelevant to the minimization, we

---

[6] This implies that $\mathbf{C_{xx}}$ is invertible. Also, the multivariate Gaussian distribution is said to be "non-degenerate" in this case.

[7] We see that the data fidelity term $\mathcal{D}(\mathbf{s}; \mathbf{y})$ has a direct relationship with the so-called negative log-likelihood function that is given by $-\log\big(p_{Y|S}(\mathbf{y} \,|\, \mathbf{s})\big)$. The same type of equivalence is observed between the regularization functional $\mathcal{R}(\mathbf{s})$ and the prior distribution p_S.

obtain that

$$\begin{aligned}\mathbf{s}^\star_{\mathrm{MAP}} &= \arg\min_{\mathbf{s}} \ -\log\left(p_{Y|S}(\mathbf{y}\,|\,\mathbf{s})\right) - \log\left(p_S(\mathbf{s})\right) \\ &= \arg\min_{\mathbf{s}} \ \frac{1}{2}(\mathbf{y}-\mathbf{H}\mathbf{s})^{\mathrm{T}}\mathbf{C}_{\mathbf{nn}}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{s}) + \frac{1}{2}(\mathbf{s}-\mathbf{m}_{\mathbf{s}})^{\mathrm{T}}\mathbf{C}_{\mathbf{ss}}^{-1}(\mathbf{s}-\mathbf{m}_{\mathbf{s}}) \\ &= \arg\min_{\mathbf{s}} \ \mathcal{J}_{\mathrm{MAP}}(\mathbf{s}). \end{aligned} \tag{2.16}$$

The energy in (2.16) involves two quadratic forms in $\mathbf{s}$. Its derivative can be expressed as

$$\frac{\partial}{\partial \mathbf{s}}\mathcal{J}_{\mathrm{MAP}}(\mathbf{s}) = -\mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathbf{nn}}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{m}_{\mathbf{s}}) + \left(\mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H} + \mathbf{C}_{\mathbf{ss}}^{-1}\right)(\mathbf{s}-\mathbf{m}_{\mathbf{s}}).$$

The general form of the Gaussian MAP estimator is thus written as

$$\mathbf{s}^\star_{\mathrm{MAP}} = \mathbf{m}_{\mathbf{s}} + \left(\mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H} + \mathbf{C}_{\mathbf{ss}}^{-1}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathbf{nn}}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{m}_{\mathbf{s}}). \tag{2.17}$$

Based on (2.17), we can distinguish clear connections with the quadratic reconstruction methods. Consider the case where the noise components are independent and identically distributed (i.i.d.) (i.e., the noise is *white* and $\mathbf{C}_{\mathbf{nn}} = \sigma^2\mathbf{I}$) and the signal is zero mean (i.e., $\mathbf{m}_{\mathbf{s}} = \mathbf{0}$). Then, the MAP solution

$$\mathbf{s}^\star_{\mathrm{MAP}} = \left(\mathbf{H}^{\mathrm{T}}\mathbf{H} + \mathbf{C}_{\mathbf{ss}}^{-1}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{y}, \tag{2.18}$$

is equivalent to the Tikhonov regularization in (2.12) for $\mathbf{R} = \mathbf{C}_{\mathbf{ss}}^{-1/2}$ and $\tau = \sigma^2/2$. It is important to note that $\mathbf{C}_{\mathbf{ss}}^{-1/2}$ is the whitening operator of $\mathbf{s}$, which enriches our understanding of the regularization operator. Furthermore, if our confidence in our prior information decreases, that is to say $\mathbf{C}_{\mathbf{ss}}^{-1} \to 0$, we see that $\mathbf{s}^\star_{\mathrm{MAP}} \to \mathbf{s}^\star_{\mathrm{LS}}$, which is also known as the *maximum likelihood estimation* (MLE). Therefore, the use of the quadratic data fidelity term can be statistically justified if $\mathbf{n}$ is additive white Gaussian noise (AWGN).

MMSE Estimation with Gaussian Priors

We can also draw inference by minimizing an expected risk that is averaged over a set of possible estimations. A standard risk function is the mean squared error (MSE) of an estimator $\hat{\mathbf{s}}$:

$$\mathrm{MSE}(\hat{\mathbf{s}}) = \mathbb{E}\left\{\|\mathbf{s}-\hat{\mathbf{s}}\|_2^2\right\}.$$

The *minimum mean squared error* (MMSE) estimator for $\mathbf{s}$ given $\mathbf{y}$ is the conditional expectation [5]

$$\begin{aligned}\mathbf{s}^{\star}_{\mathrm{MMSE}} &= \arg\min_{\hat{\mathbf{s}}} \ \mathrm{MSE}(\hat{\mathbf{s}}) \\ &= \mathbb{E}\left\{\mathbf{s} \,|\, \mathbf{y}\right\} \\ &= \int_{\mathbb{R}^N} \mathbf{s}\, p_{S|Y}(\mathbf{s} \,|\, \mathbf{y}) \mathrm{d}\mathbf{s}. \end{aligned} \tag{2.19}$$

However, evaluation of (2.19) is very often not tractable as it involves an N-dimensional integration.

The good news is that our present scenario constitutes a rare case where the MMSE estimation has a closed-form expression. We remark that $\mathbf{s}$ and $\mathbf{n}$ are jointly Gaussian due to their independence. Since the observation model is linear, $\mathbf{s}$ and $\mathbf{y}$ are also jointly Gaussian. In this case, a classical result in Bayesian estimation theory [5] establishes that the posterior pdf is also Gaussian with mean

$$\mathbb{E}\{\mathbf{s} \,|\, \mathbf{y}\} = \mathbf{m}_{\mathbf{s}} + \mathbf{C}_{\mathbf{ss}}\mathbf{H}^{\mathrm{T}} \left(\mathbf{H}\mathbf{C}_{\mathbf{ss}}\mathbf{H}^{\mathrm{T}} + \mathbf{C}_{\mathbf{nn}}\right)^{-1} (\mathbf{y} - \mathbf{H}\mathbf{m}_{\mathbf{s}}) \tag{2.20}$$

and covariance matrix

$$\mathbf{C}_{\mathbf{s}\,|\,\mathbf{y}} = \mathbf{C}_{\mathbf{ss}} - \mathbf{C}_{\mathbf{ss}}\mathbf{H}^{\mathrm{T}} \left(\mathbf{H}\mathbf{C}_{\mathbf{ss}}\mathbf{H}^{\mathrm{T}} + \mathbf{C}_{\mathbf{nn}}\right)^{-1} \mathbf{H}\mathbf{C}_{\mathbf{ss}}.$$

We see that (2.20) provides us with the sought Gaussian MMSE estimator, which yields the best possible reconstruction in MSE sense. In particular, under the zero-mean signal assumption, we obtain

$$\mathbf{s}^{\star}_{\mathrm{MMSE}} = \mathbf{C}_{\mathbf{ss}}\mathbf{H}^{\mathrm{T}} \left(\mathbf{H}\mathbf{C}_{\mathbf{ss}}\mathbf{H}^{\mathrm{T}} + \mathbf{C}_{\mathbf{nn}}\right)^{-1} \mathbf{y}, \tag{2.21}$$

which also specifies a linear reconstruction method.

Linear MMSE Estimation

One can also attain the linear estimator (2.21) without using the Gaussian prior. Consider the case where we know only the mean $\mathbf{m}_{\mathbf{s}}$ and the covariance matrix $\mathbf{C}_{\mathbf{ss}}$ of the signal. This implies that only the first- and second-order moments of the joint pdf of $\mathbf{s}$ and $\mathbf{n}$ (which is now arbitrary) are available. We keep the MMSE criterion as our optimality condition and would like to have the best possible estimator that is constrained to be linear. Such an estimator is given by the formulation

$$\mathbf{K}^{\star} = \arg\min_{\mathbf{K}} \; \mathbb{E}\{\|\mathbf{s} - \mathbf{K}\mathbf{y}\|_2^2\},$$

and is known as the *linear* MMSE (LMMSE) or the Wiener solution. Under the zero-mean signal hypothesis, the LMMSE reconstruction is given by

$$\mathbf{s}^{\star}_{\mathrm{LMMSE}} = \underbrace{\mathbf{C_{ss}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{C_{ss}}\mathbf{H}^{\mathrm{T}} + \mathbf{C_{nn}}\right)^{-1}}_{\mathbf{K}^{\star}}\mathbf{y}, \tag{2.22}$$

which implicitly assumes that both $\mathbf{C_{yy}}$ and $(\mathbf{K}^{\star})^{\mathrm{T}}\mathbf{K}^{\star}$ are invertible [5]. We note that the resulting approach provides the MMSE solution among all linear reconstruction schemes—independent of the prior model—and is fully compatible with the Gaussian MMSE solution (2.21).

In estimation theory, it is also well-known that the Gaussian MAP and LMMSE solutions are equivalent [10]. This is established by the identity

$$\mathbf{C_{ss}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{C_{ss}}\mathbf{H}^{\mathrm{T}} + \mathbf{C_{nn}}\right)^{-1} = \left(\mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathbf{nn}}^{-1}\mathbf{H} + \mathbf{C}_{\mathbf{ss}}^{-1}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{C}_{\mathbf{nn}}^{-1}, \tag{2.23}$$

which follows from the pre- and post-multiplication of both sides of (2.23) by $(\mathbf{H}^{\mathrm{T}}\mathbf{C_{nn}}^{-1}\mathbf{H} + \mathbf{C_{ss}}^{-1})$ and $(\mathbf{H}\mathbf{C_{ss}}\mathbf{H}^{\mathrm{T}} + \mathbf{C_{nn}})$, respectively (see [16] for details). Based on (2.23), we simply see that the general form of the Gaussian MAP/MMSE and LMMSE reconstructions can be seen as a weighted variant of Tikhonov-type regularization schemes, where the weights are chosen inversely proportional to the noise power.

2.3.3 Completing Notes on Linear Reconstructions

We have obtained a clear view of the linear reconstruction methods, where there is a reassuring agreement between quadratic energy minimizations and MAP/MMSE solutions under the Gaussian signal assumption. In computational terms, the highly favorable aspect of these methods is that the consequent reconstruction is of the form $\mathbf{A}\mathbf{x} = \mathbf{b}$. This makes the linear schemes particularly interesting within the scope of the thesis since solving linear systems can also take place as a building-block for more sophisticated algorithms. For that reason, we explain a generic resolution strategy.

In a general setting, high-dimensional matrix inversions are computed using an iterative algorithm such as the conjugate gradient (CG) method. CG is an efficient solver for linear systems where $\mathbf{A} \in \mathbb{R}^{N \times N}$ is symmetric and positive definite. The inherent idea of the algorithm is to iteratively minimize the strongly convex energy functional $\mathcal{J}_{\mathrm{CG}}(\mathbf{x}) = (1/2)\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} - \mathbf{x}^{\mathrm{T}}\mathbf{b}$, whose unique minimizer is our formal solution $\mathbf{A}^{-1}\mathbf{b}$.

The starting point is to represent $\mathbf{x}$ in a basis of mutually conjugate vectors (with respect to $\mathbf{A}$).[8] That is to say that

$$\mathbf{x} = \sum_{i}^{N} \alpha_i \mathbf{p}_i,$$

with $\mathbf{p}_i^{\mathrm{T}}\mathbf{A}\mathbf{p}_j = \mathbf{0}$ for $i \neq j$.

At iteration t, the iterate reads $\mathbf{x}_t = \sum_{i \leq t} \alpha_i \mathbf{p}_i$ and the goal is to identify the next conjugate vector $\mathbf{p}_{t+1}$ and the corresponding coefficient α_{t+1}. Note that the residual, that is $\mathbf{r}_t = \mathbf{b} - \mathbf{A}\mathbf{x}_t$, is the negative of the gradient of $\mathcal{J}_{\mathrm{CG}}$ at $\mathbf{x}^t$ (i.e., the direction that the gradient descent method would move towards). The CG algorithm chooses $\mathbf{p}_{t+1} = \mathbf{r}_t - \sum_{i \leq t} \left((\mathbf{p}_i^{\mathrm{T}}\mathbf{A}\mathbf{r}_t)/(\mathbf{p}_i^{\mathrm{T}}\mathbf{A}\mathbf{p}_i)\right) \mathbf{p}_i$, which assures the mutual conjugacy constraint. In this direction, the coefficient $\alpha_{t+1} = \left(\mathbf{p}_{t+1}^{\mathrm{T}}\mathbf{r}_t\right) / \left(\mathbf{p}_{t+1}^{\mathrm{T}}\mathbf{A}\mathbf{p}_{t+1}\right)$ is optimal with respect to the energy $\mathcal{J}_{\mathrm{CG}}$ [17]. An efficient implementation scheme of the presented ideas is described in Algorithm 1.

We see that Algorithm 1 is composed of only one matrix-vector multiplication per iteration, which is highly favorable. For $\mathbf{A}$ being an N-by-N matrix, the method theoretically converges to the exact solution at most after N iterations.[9] We note that the convergence speed of the algorithm is dictated by the condition number of $\mathbf{A}$. A variant of the method, called the preconditioned CG, can be utilized if there is need to improve the convergence speed of the standard algorithm [18].

More favorable scenarios exist for the linear reconstruction methods. In certain cases, the direct inversion of the matrix $\mathbf{A}$ is feasible. This implies that the recon-

---

[8] Since $\mathbf{A}$ is symmetric and positive definite, the expression $\mathbf{p}_i^{\mathrm{T}}\mathbf{A}\mathbf{p}_j$ can be seen as the inner product

$$\langle \mathbf{p}_i, \mathbf{p}_j \rangle_{\mathbf{A}} = \mathbf{p}_i^{\mathrm{T}}\mathbf{A}\mathbf{p}_j.$$

Hence, if two vectors are orthogonal with respect to $\langle \cdot, \cdot \rangle_{\mathbf{A}}$, they are conjugate to each other. In this sense, we realize that the conjugation resembles orthogonality constraint.

[9] In practice, the statement might not be satisfied due to round-off errors. Still, the good news is that the CG iterations always improve the solution (in the sense that the energy value is monotonically decreased).

Algorithm 1: Conjugate gradient method for solving $\mathbf{Ax} = \mathbf{b}$ with $\mathbf{A}$ being symmetric and positive definite matrix

```
input  : Matrix $\mathbf{A}$, vector $\mathbf{b}$, and the initial solution $\mathbf{x}_0$
output: Resulting vector $\mathbf{x}^\star \approx \mathbf{A}^{-1}\mathbf{b}$
$t \leftarrow 0$; $\mathbf{r}_t \leftarrow \mathbf{b} - \mathbf{A}\mathbf{x}_t$; $\mathbf{p}_t \leftarrow \mathbf{r}_t$;
repeat
    $\mathbf{q}_t \leftarrow \mathbf{A}\mathbf{p}_t$;
    $\alpha_t \leftarrow \left(\mathbf{r}_t^{\mathrm{T}}\mathbf{r}_t\right)\left(\mathbf{p}_t^{\mathrm{T}}\mathbf{q}_t\right)^{-1}$;
    $\mathbf{x}_{t+1} \leftarrow \mathbf{x}^t + \alpha^t\mathbf{p}^t$;
    $\mathbf{r}_{t+1} \leftarrow \mathbf{r}^t - \alpha^t\mathbf{q}^t$;
    $\mathbf{p}_{t+1} \leftarrow \mathbf{r}^{t+1} + \left(\mathbf{r}_{t+1}^{\mathrm{T}}\mathbf{r}_{t+1}\right)\left(\mathbf{r}_t^{\mathrm{T}}\mathbf{r}_t\right)^{-1}\mathbf{p}_t$;
    $t \leftarrow t + 1$;
until stopping criteria
return $\mathbf{x}_t$;
```

struction (or some elementary units of more advanced schemes) can be obtained non-iteratively. In Chapter 4, we shall analyze this point in details.

Despite all the computational convenience, linear reconstruction methods have well-known drawbacks. In the case of quadratic regularizers, the choice $\mathbf{R} = \mathbf{I}$ does not introduce any structure on the reconstruction. This becomes a practical issue if the underlying object has components that cannot be measured such as in the case of limited-angle tomography. As for $\mathbf{R}$ being the first-order derivative operator, the regularization term does discourage local derivative values to be large. The downside of such behavior is that the discontinuities (such as edges in images) suffer from being smoothed out, which in general yields sub-optimal reconstructions. The aforementioned shortcomings are eliminated through the incorporation of sparsity-promoting regularization functionals.

2.4 Sparsity-Based Reconstruction Methods

In this section, we examine the use of regularization for incorporating the notion of sparsity in the inverse problem framework. The approach, usually called the *sparsity-promoting* regularization, has emerged as the key element in the development of powerful variational reconstruction algorithms [10]. The practical significance of these approaches is that they provide the state-of-the-art results in various inverse problems in biomedical imaging, obtaining higher quality image reconstructions than their Tikhonov-type counterparts [7, 19, 20, 21, 22].

Primarily inspired by the use of wavelet transforms to concentrate the energy of natural signals on very few wavelet coefficients,[10] sparsity-based regularization functionals encourage solutions that exhibit a sparse structure in some suitably-chosen domain. Basically, the sparsity is achieved through the regularization operator. With this property, **Rs** is expected to be sparse in the sense that it involves many fewer nonzero terms than N.

2.4.1 Inducing Sparsity via Regularization

Consolidation of the sparsity property into variational methods is achieved by specifying non-quadratic regularization functionals. To have a convenient setting to analyze sparsity-promoting frameworks, we consider the generic energy functional of the form

$$\mathcal{J}_{\mathrm{S}}(\mathbf{s}) = \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2 + \tau \underbrace{\|\mathbf{R}\mathbf{s}\|_p^p}_{\mathcal{R}_{\mathrm{S}}(\mathbf{s})}, \tag{2.24}$$

where $p \in [1, 2)$ and

$$\|\mathbf{s}\|_p = \left(\sum_{j=1}^{N} |s_j|^p\right)^{1/p}.$$

Intuitively, p can be seen as a sparsity index. For $p = 2$, one obtains the Tikhonov regularization which do not promote sparsity. On the contrary, the regularizer $\mathcal{R}_{\mathrm{S}}$ encourages more sparsity as p gets smaller.

Among all possible choices for p, we particularly focus on the case $p = 1$. Our fundamental motivation is that it imposes the highest level of sparsity on

---

[10] In an orthonormal basis of P functions $\phi_{\boldsymbol{k}}$ and for $f = \sum_{\boldsymbol{k} \in \Omega \subset \mathbb{Z}^2} c[\boldsymbol{k}]\phi_{\boldsymbol{k}}$,

1. $\exists \Omega_K \subset \mathbb{Z}^2$, $\mathrm{card}(\Omega_K) \ll \mathrm{card}(\Omega) = P$ (sparse support) and
2. $\exists \mathbf{c}$ such that $\|f - \sum_{\boldsymbol{k} \in \Omega_S} c[\boldsymbol{k}]\phi_{\boldsymbol{k}}\|_2 \ll \|f\|_2$ (accurate approximation) [17].

Similar ideas can also be argued as the rapid decay observed for the sorted transform-domain coefficients. For a formal discussion on the topic, we refer the reader to [6].

the solution under the constraint that the energy functional (2.24) stays convex. Even though this property has been long acknowledged in the literature [23, 24, 25], the application of ℓ_1-norms to promote sparse solutions has an uninterrupted popularity. Mainly motivated by the recent developments based on the compressed sensing (CS) [11] framework [27, 28], it currently stands out as the prevailing approach in variational reconstruction methods [29]. Accordingly, we fix the regularization term as

$$\mathcal{R}_{\mathrm{S}}(\mathbf{s}) = \|\mathbf{R}\mathbf{s}\|_1$$

in the rest of the section.

Let us first consider the simplest form of regularization provided by choosing $\mathcal{R}_{\mathrm{S}}(\mathbf{s}) = \|\mathbf{s}\|_1$. This leads to the optimization problem

$$\min_{\mathbf{s}} \ \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2 + \tau\|\mathbf{s}\|_1.$$

To gain more intuition on the implications of using the ℓ_1-norm, we illustrate the geometry of its unit ball in Figure 2.1. As explained in [30], we see that—for a given ℓ_2-norm value—the ℓ_1-norm has a higher cost for vectors containing many small coefficients than the ones that have few large coefficients. In plain terms, the ℓ_1-norm regularization favors the latter case. Thus, replacing an ℓ_2-normed regularizer with its ℓ_1-normed version significantly changes the constraint imposed on the solution (see Section 2.3.1). We note that this empirical idea can be properly stated in mathematical terms (and eventually be extended to a larger class of regularizers) [31]. Next, we point out some common choices for the sparsifying operator $\mathbf{R}$.

2.4.2 Wavelet-Based Regularization

A typical sparsity-driven regularization approach is to use the prior information that the underlying signal $\mathbf{s}$ has a sparse representation in the wavelet domain.

---

[11] The CS theory aims to obtain the "sparsest" signal $\mathbf{s}$ given its underdetermined linear measurements $\mathbf{y} = \mathbf{Q}\mathbf{s}$ via solving the problem

$$\min_{\mathbf{s}} \ \|\mathbf{s}\|_0 \ \text{ subject to } \ \mathbf{y} = \mathbf{Q}\mathbf{s},$$

where $\|\cdot\|_0$ counts the number of non-zero entries in a given vector. In practice, one replaces $\|\cdot\|_0$ with the ℓ_1-norm as the former results in an NP-hard problem. The remarkable aspect of the theory is that the solution recovered under this simplification coincides the correct one when $\mathbf{Q}$ satisfies certain conditions [26].

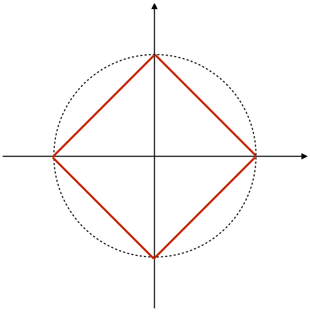


Figure 2.1: Illustration of ℓ_2-norm (dashed) and ℓ_1-norm (solid) unit balls.

In our present formulation, this is simply accomplished by $\mathbf{R} = \mathbf{W}$, where $\mathbf{W}$ represents an orthonormal wavelet transform implying that $\mathbf{W}^{\mathrm{T}}\mathbf{W} = \mathbf{I}$. To have an approximation of the wavelet coefficients of the true signal $\mathbf{w} = \mathbf{W}\mathbf{s}$, we consider the problem

$$\begin{aligned}\mathbf{w}^{\star}_{\mathrm{Wave}} &= \arg\min_{\mathbf{w}} \ \mathcal{J}_{\mathrm{Wave}}(\mathbf{w}) \\ &= \arg\min_{\mathbf{w}} \ \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{W}^{\mathrm{T}}\mathbf{w}\|_2^2 + \tau\|\mathbf{w}\|_1. \end{aligned} \tag{2.25}$$

Once (2.25) is solved, the reconstructed signal is retrieved by the inverse wavelet transform:

$$\mathbf{s}^{\star}_{\mathrm{Wave}} = \mathbf{W}^{\mathrm{T}}\mathbf{w}^{\star}_{\mathrm{Wave}}. \tag{2.26}$$

To boost the performance of wavelet-based regularizations, one can use an augmented version of $\mathbf{W}$ that forms a (normalized) tight frame including the shifted-versions of $\mathbf{W}$. This technique, known as cycle-spinning, aims to account for the lack of shift-invariance of $\mathbf{W}$ and is very often used in practice [32].

Wavelet-based approaches are particularly interesting for deriving simple algorithms for removing noise in signals. For such problems, which means that $\mathbf{H} = \mathbf{I}$, we formulate the denoised signal as

$$\begin{aligned}\mathbf{s}^\star_{\text{Wave}} &= \arg\min_{\mathbf{s}} \ \frac{1}{2}\|\mathbf{y}-\mathbf{s}\|_2^2 + \tau\|\mathbf{W}\mathbf{s}\|_1 \\ &= \mathbf{W}^{\mathrm{T}} \arg\min_{\mathbf{w}} \ \frac{1}{2}\|\tilde{\mathbf{w}}-\mathbf{w}\|_2^2 + \tau\|\mathbf{w}\|_1, \end{aligned} \tag{2.27}$$

where we use the orthonormality of $\mathbf{W}$ for the second equality. It is well-known that the solution of the problem is calculated by soft-thresholding $\tilde{\mathbf{w}}$ that represents the wavelet coefficients of $\mathbf{y}$ [33]. The denoising scheme in (2.27) is implemented efficiently since the forward and inverse wavelet transforms have efficient filter bank implementations [34]. The computational convenience of the algorithm motivates the design of wavelets that sparsely represent certain type of signals. In Chapter 7, we shall develop application-specific wavelets for denoising vector fields with distinct physical features.

2.4.3 Derivative-Based Regularization

As noted in Section 2.3.1, we often utilize the first-order derivative operator with a quadratic regularizer for constraining the solution to exhibit a certain level of smoothness. Alternatively, we now use a regularization term that combines the derivative operator with the ℓ_1-norm. This results in the so-called total variation (TV) regularization [35] that is given by

$$\mathbf{s}^\star_{\text{TV}} = \arg\min_{\mathbf{s}} \ \frac{1}{2}\|\mathbf{y}-\mathbf{H}\mathbf{s}\|_2^2 + \tau\|\mathbf{D}\mathbf{s}\|_1, \tag{2.28}$$

where $\mathbf{D} \in \mathbb{R}^{N\times N}$ is the discrete counterpart of the derivative. By using (2.28), we favor signals whose derivative is sparse, meaning that the reconstructed signal is expected to be piecewise-smooth. Contrary to its Tikhonov-type analogue, TV allows for homogenous regions with sharp boundaries in between. Hence, TV is often called an "edge-preserving" regularization. This aspect makes it yield state-of-the-art reconstruction results in various imaging applications [36].

The definition of TV in the one-dimensional setting is clear. For d-dimensional and scalar-valued signals, the *anisotropic* TV regularization functional is given by

$$\mathrm{TV}_{\text{ani}}(\mathbf{s}) = \sum_{j\in\Omega} \|[\mathbf{D}\mathbf{s}]_j\|_1,$$

where Ω is the index set of pixel locations and $[\mathbf{Ds}]_j \in \mathbb{R}^d$ represents the gradient vector at pixel j. The disadvantage of the anisotropic TV is that piecewise-smoothness is imposed with preference to certain orientations. Therefore, the *isotropic* TV functional

$$\mathrm{TV}_{\mathrm{iso}}(\mathbf{s}) = \sum_{j\in\Omega} \|[\mathbf{Ds}]_j\|_2$$

is the most common choice of TV for such signals.[12] In the thesis, we shall be concerned only with the latter so that we drop the term "isotropic". Still, we notice that the definition of TV functional reveals variations as the inherent dimensionality of the underlying object gets more involved. This point will become clearer when we focus on extending the TV regularization for multi-dimensional vector-valued functions in Chapter 6.

Bayesian Interpretation of ℓ_1-Norm Regularization

A basic statistical perception of the ℓ_1-norm regularizations is obtained by assuming a separable prior distribution for the signal in the range of $\mathbf{R} \in \mathbb{R}^{N\times N}$. In this case, the prior reads that

$$p_S(\mathbf{s}) \propto p_R(\mathbf{Rs}) = \prod_{j=1}^{N} p_{R_j}([\mathbf{Rs}]_j), \tag{2.29}$$

where p_{R_j} is the pdf of the jth entry of $\mathbf{Rs}$. Furthermore, we assume that each p_{R_j} for $j = 1, \dots, N$ is given by a zero-mean Laplace distribution with a scale parameter τ so that

$$p_{R_j}(x) = \frac{\tau}{2}\mathrm{e}^{-\tau|x|}. \tag{2.30}$$

Consequently, one sees that

$$-\log\left(p_S(\mathbf{s})\right) = \tau\|\mathbf{Rs}\|_1,$$

up to a constant that is not important in terms of minimization. Therefore, the considered ℓ_1-norm regularization framework (2.24) coincide with the MAP formulation when we adopt an i.i.d. Laplace prior for the vector $\mathbf{Rs}$ together with the AWGN assumption.

---

[12] Alternative regularizers that are also based on the gradient operator have been discussed in the literature [31, 37, 38]. However, the isotropic TV is the most popular one for solving inverse problems.

2.5 Iterative Optimization Methods

The performance improvement brought by the sparsity-promoting methods over the Tikhonov-type ones comes with an increased computational cost. The reconstruction step is specified as a nonlinear function of the measurements. Another apparent difficulty is the non-differentiability of the ℓ_1-norm, eliminating the option of using standard approaches such as the gradient descent.[13] These challenges have led to the development of several iterative optimization algorithms [39, 40, 41, 42]. In the guidance of these methods, we explain how ℓ_1-norm regularizations are tackled. To that end, we consider the generic reconstruction of the form

$$\mathbf{s}_{\mathrm{S}}^{\star} = \arg\min_{\mathbf{s}} \ \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2 + \tau \mathcal{R}_{\mathrm{S}}(\mathbf{s}), \tag{2.31}$$

where $\mathcal{R}_{\mathrm{S}}(\mathbf{s}) = \|\mathbf{R}\mathbf{s}\|_1$ and it is assumed that $\mathbf{H}$ and $\mathbf{R}$ have distinct nullspaces.

2.5.1 Fast Iterative Shrinkage/Thresholding Algorithm

The basic principle of the fast iterative shrinkage/thresholding algorithm (FISTA) is explained as a "majorize-minimize" (i.e., optimization transfer) approach.[14] In such schemes, a sequence of surrogate functions that upper-bound the original energy is sequentially minimized to attain the sought minimizer. The key goal is to formulate surrogates that are simple to minimize. Seeing the structure of our energy functional in (2.31), it is possible to obtain such a surrogate that provide an upper-bound on the quadratic data term:

$$\begin{aligned}
\mathcal{D}(\mathbf{s};\mathbf{y}) &= \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2 = \frac{1}{2}\|\mathbf{H}(\mathbf{s} - \mathbf{s}^t) + \mathbf{H}\mathbf{s}^t - \mathbf{y}\|_2^2 \\
&= \frac{1}{2}(\mathbf{s} - \mathbf{s}^t)^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H}(\mathbf{s} - \mathbf{s}^t) + (\mathbf{s} - \mathbf{s}^t)^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{s}^t - \mathbf{y}) + \frac{1}{2}\|\mathbf{H}\mathbf{s}^t - \mathbf{y}\|_2^2 \\
&\le \frac{\gamma}{2}(\mathbf{s} - \mathbf{s}^t)^{\mathrm{T}}(\mathbf{s} - \mathbf{s}^t) + (\mathbf{s} - \mathbf{s}^t)^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{s}^t - \mathbf{y}) + \frac{1}{2}\|\mathbf{H}\mathbf{s}^t - \mathbf{y}\|_2^2 \\
&= \frac{\gamma}{2}\left\|\mathbf{s} - \mathbf{s}^t + \frac{1}{\gamma}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{s}^t - \mathbf{y})\right\|_2^2 + O(\mathbf{s}^t, \mathbf{y}),
\end{aligned} \tag{2.32}$$

---

[13] It is possible to replace the ℓ_1-norm with its "smoothed" versions (for example, the Huber functional). However, this decreases the reconstruction performance.

[14] The same framework can also be seen as forward-backward splitting algorithm. See [43] for further details.

where $\gamma \geq \|\mathbf{H}^{\mathrm{T}}\mathbf{H}\|_2$ and the exact form of $O(\mathbf{s}^t, \mathbf{y})$ is omitted since it is not a function of $\mathbf{s}$.

In great generality, the surrogate energy proposes the following iterative strategy

$$\mathbf{s}^{t+1} = \arg\min_{\mathbf{s}} \frac{1}{2}\left\|\mathbf{s} - \left(\mathbf{s}^t - \frac{1}{\gamma}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{s}^t - \mathbf{y})\right)\right\|_2^2 + \frac{\tau}{\gamma}\mathcal{R}_{\mathrm{S}}(\mathbf{s}), \tag{2.33}$$

which is called the iterative shrinkage/thresholding algorithm (ISTA) [44, 45]. Note that (2.33) performs a gradient descent step via the computation of

$$\mathbf{z}^t = \mathbf{s}^t - \frac{1}{\gamma}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{s}^t - \mathbf{y}).$$

This is then followed by the implicit denoising problem

$$\mathbf{s}^{t+1} = \arg\min_{\mathbf{s}} \frac{1}{2}\|\mathbf{s} - \mathbf{z}^t\|_2^2 + \frac{\tau}{\gamma}\mathcal{R}_{\mathrm{S}}(\mathbf{s}),$$

which is elegantly formalized by relating it to the proximal mapping of $\mathcal{R}_{\mathrm{S}}$:

$$\mathbf{s}^{t+1} = \mathrm{prox}_{\mathcal{R}_{\mathrm{S}}}(\mathbf{z}^t; \tau/\gamma) = \arg\min_{\mathbf{s}} \frac{1}{2}\|\mathbf{s} - \mathbf{z}^t\|_2^2 + \frac{\tau}{\gamma}\mathcal{R}_{\mathrm{S}}(\mathbf{s}). \tag{2.34}$$

The convergence of the iterations given in (2.33) to the minimizer is guaranteed since $\mathcal{R}_{\mathrm{S}}$ is convex [45]. Unfortunately, the method shows a poor convergence speed in practice. Hence, FISTA is proposed as an improved version of ISTA [46]. By incorporating what is known as the Nesterov's acceleration scheme [47], the algorithm provably improves the convergence speed compared to ISTA [46]. In the implementation, this is achieved through an "over-shooting" step where one uses the two previous iterates to update the solution. FISTA can be summarized as in Algorithm 2.

An important aspect of ISTA-type schemes is that they do not put any smoothness restriction on $\mathcal{R}_{\mathrm{S}}$ since the evaluation of the proximal mapping is applicable to non-differential functions. The latter implies that the overall computational cost is affected by the calculation of $\mathrm{prox}_{\mathcal{R}_{\mathrm{S}}}$ given in (2.34). Therefore, the algorithm takes full advantage of the specific structure of (2.31) in certain situations:

- The most favorable case is that $\mathbf{R} = \mathbf{I}$. We recall that this occurs for the wavelet-based regularizations given in (2.25). In this setting, the proximal

Algorithm 2: FISTA for solving (2.31)

```
input  : Measurements $\mathbf{y}$, regularization parameter $\tau$, step size $\gamma$,
         and proximal mapping of $\mathcal{R}_{\mathrm{S}}(\mathbf{s})$ proxRegularizer()
output: Reconstructed signal $\mathbf{s}_{\mathrm{S}}^{\star}$
$t \leftarrow 0$; $\mathbf{s}^t \leftarrow \mathbf{0}$; $\mathbf{x}^t \leftarrow \mathbf{0}$; $k^t \leftarrow 0$;
repeat
    $\mathbf{z}^t \leftarrow \mathbf{s}^t - \frac{1}{\gamma}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{s}^t - \mathbf{y})$;
    $\mathbf{x}^{t+1} \leftarrow$ proxRegularizer($\mathbf{z}^t; \tau/\gamma$);
    $k^{t+1} \leftarrow \frac{1+\sqrt{1+4k^t}}{2}$;
    $\mathbf{s}^t \leftarrow \mathbf{x}^{t+1} + \frac{k^t-1}{k^{t+1}}(\mathbf{x}^{t+1} - \mathbf{x}^t)$;
    $t \leftarrow t + 1$;
until stopping criteria
return $\mathbf{s}^t$;
```

mapping is separable and admits the soft-thresholding operator as its closed-form solution. Consequently, FISTA is frequently used in reconstruction methods that impose sparsity in the wavelet-domain [48].

- The case $\mathbf{H} = \mathbf{I}$ (i.e., denoising problems) is also extremely convenient. Formulating an equivalent dual problem of (2.31), the denoised signal is obtained efficiently [49]. We shall develop such an algorithm in Chapter 6 and provide details for the duality-based formulation.

Regarding the general framework of (2.31), it is noteworthy that FISTA involves inner iterations to obtain the proximal mapping of the regularizer $\|\mathbf{R}\mathbf{s}\|_1$. If a closed-form solution is not available, this is approximately computed in practice (we generally perform a small number of inner iterations) and one relies on a slightly modified implementation of Algorithm 2 that ensures monotonic decrease of the energy [49].

2.5.2 Alternating Directions Method of Multipliers

Similar to FISTA, alternating directions method of multipliers (ADMM) decomposes the opimization problem into simpler sub-problems. To do so, the algorithm first transforms the initial unconstrained problem (2.31) into an equivalent constrained problem:

$$\mathbf{s}_{\mathrm{S}}^{\star} = \arg\min_{\mathbf{s}} \ \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2 + \tau\|\mathbf{u}\|_1 \ \text{ subject to } \ \mathbf{u} = \mathbf{R}\mathbf{s}. \tag{2.35}$$

From the constrained form, we introduce the augmented Lagrangian functional

$$L(\mathbf{s}, \mathbf{u}, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2 + \tau\|\mathbf{u}\|_1 - \boldsymbol{\alpha}^{\mathrm{T}}(\mathbf{u} - \mathbf{R}\mathbf{s}) + \frac{\mu}{2}\|\mathbf{u} - \mathbf{R}\mathbf{s}\|_2^2, \tag{2.36}$$

where $\boldsymbol{\alpha}$ is the Lagrange multiplier and μ is called the penalty parameter. The augmented Lagrangian differs from the standard Lagrangian form in that it adds the quadratic penalty term $(\mu/2)\|\mathbf{u} - \mathbf{R}\mathbf{s}\|_2^2$.

As L is a saddle-function, we can solve the saddle-point problem with the method of multipliers (MM) technique [50]. This iterative scheme is expressed as

$$\begin{aligned} (\mathbf{s}^{t+1}, \mathbf{u}^{t+1}) &= \arg\min_{(\mathbf{s}^t, \mathbf{u}^t)} \ L(\mathbf{s}^t, \mathbf{u}^t, \boldsymbol{\alpha}^t) \\ \boldsymbol{\alpha}^{t+1} &= \arg\max_{\boldsymbol{\alpha}^t} \ L(\mathbf{s}^{t+1}, \mathbf{u}^{t+1}, \boldsymbol{\alpha}^t), \end{aligned}$$

where the maximization problem is solved by gradient ascent as $L(\mathbf{s}^{t+1}, \mathbf{u}^{t+1}, \boldsymbol{\alpha}^t)$ is concave over $\boldsymbol{\alpha}^t$. However, the minimization of $L(\mathbf{s}^t, \mathbf{u}^t, \boldsymbol{\alpha}^t)$ over the augmented variable $(\mathbf{s}^{t+1}, \mathbf{u}^{t+1})$ is in general difficult to compute. Therefore, we separate the joint minimization into two individual problems, hence the name "alternating directions". ADMM iterations [51, 52, 53] are then given by

$$\begin{aligned} \mathbf{u}^{t+1} &= \arg\min_{\mathbf{u}^t} \ L(\mathbf{s}^t, \mathbf{u}^t, \boldsymbol{\alpha}^t) \\ \mathbf{s}^{t+1} &= \arg\min_{\mathbf{s}^t} \ L(\mathbf{s}^t, \mathbf{u}^{t+1}, \boldsymbol{\alpha}^t) \\ \boldsymbol{\alpha}^{t+1} &= \boldsymbol{\alpha}^t + \mu(\mathbf{R}\mathbf{s}^{t+1} - \mathbf{u}^{t+1}). \end{aligned}$$

In the current ℓ_1-normed regularization setting, the minimization of $L(\mathbf{s}^t, \mathbf{u}^t, \boldsymbol{\alpha}^t)$ over $\mathbf{u}^t$ develops into the proximal mapping of ℓ_1-norm. As noted previously, this is computed by the soft-thresholding operator. Similarly, we can show that the second step becomes a matrix inversion problem. An implementation of the present ADMM

Algorithm 3: ADMM for solving (2.31)

```
input : Measurements $\mathbf{y}$, regularization parameter $\tau$, penalty parameter $\mu$
output: Reconstructed signal $\mathbf{s}^\star_{\mathrm{S}}$
$t \leftarrow 0$; $\mathbf{s}^t \leftarrow \mathbf{0}$; $\mathbf{u}^t \leftarrow \mathbf{0}$; $\boldsymbol{\alpha}^t \leftarrow \mathbf{0}$; $\mathbf{b} \leftarrow \mathbf{H}^{\mathrm{T}}\mathbf{y}$;
repeat
    $\mathbf{z}^t \leftarrow \mathbf{R}\mathbf{s}^t + \boldsymbol{\alpha}^t/\mu$;
    $\mathbf{s}^{t+1} \leftarrow \mathrm{prox}_{\|\cdot\|_1}(\mathbf{z}^t; \tau/\mu)$;
    $\mathbf{b}^t \leftarrow \mathbf{b} + \mu\mathbf{R}^{\mathrm{T}}\left(\mathbf{u}^{t+1} - \boldsymbol{\alpha}^t/\mu\right)$;
    $\mathbf{s}^{t+1} \leftarrow (\mathbf{H}^{\mathrm{T}}\mathbf{H} + \mu\mathbf{R}^{\mathrm{T}}\mathbf{R})^{-1}\mathbf{b}^t$;
    $\boldsymbol{\alpha}^{t+1} \leftarrow \boldsymbol{\alpha}^t + \mu(\mathbf{R}\mathbf{s}^{t+1} - \mathbf{u}^{t+1})$;
    $t \leftarrow t + 1$;
until stopping criteria
return $\mathbf{s}^t$;
```

is provided in Algorithm 3. This is akin to the linear reconstruction methods explained in Section 2.3.1 and enjoys the same computational benefits.

ADMM-based optimization methods will be particularly interesting within the scope of the thesis as they are highly modular. By suitably modifying the steps of the algorithm, we shall use them to systematically compare a wide range of sparsity-based signal reconstruction algorithms. Moreover, we shall show that both sub-problems have closed-form solutions in certain images applications. This point will enable us to develop a practical phase retrieval algorithm in Chapter 5.

2.6 Summary

We have reviewed the basic principles regarding the resolution of linear inverse problems in this chapter. In particular, we have presented the classical regularization methods from a Bayesian perspective by using Gaussian prior models. Despite the computational simplicity of these approaches, they are now replaced with their sparsity-promoting counterparts that significantly improves the reconstruction quality. We have seen that the use of sparsity-based methods are mainly driven by deterministic considerations. Furthermore, the straightforward statistical justifications in a purely discrete setting is not fully satisfying. We address these issues by applying a proper discretization to a continuous-domain stochastic signal model in Chapter 3. Methodologically, this uncommon approach will enable us to have a rich family of MAP estimators that provide the statistical view of the ℓ_p-type regularizations (for $p = 1, 2$ and $p \to 0$).

Chapter 3

Sparse Reconstructions: Continuous-Domain Links

3.1 Overview

In this chapter,[1] we revisit linear inverse problems. The present approach will noticeably differ from what has been presented in Chapter 2. We take a step further in the problem formulation by establishing a continuous-domain stochastic signal model in the first place. Since the signal model is specified prior to the data acquisition step, the fundamental need is its integration into the inverse problem formalism. Accordingly, we introduce a discretizaton scheme that allows us to derive MAP estimators in a systematic fashion. On a theoretical note, our key finding is that the class of admissible priors for the discretized object is confined to the infinitely divisible distributions. The result—attained as a consequence of our continuous-domain considerations—paves direct and reassuring links with the sparsity-driven schemes. The developed methodology also suggests novel algorithmic solutions for the subsequent signal reconstruction step. This point will allow us to study a variety of reconstruction algorithms within unified settings.

---

[1] The chapter is based on our paper [54].

3.2 Introduction

We consider the problem of reconstructing a signal from its noisy linear measurements. For this purpose, we return to the generic discrete observation model:

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{3.1}$$

where $\mathbf{s} \in \mathbb{R}^N$ is the discrete representation of the unknown object, $\mathbf{y} \in \mathbb{R}^M$ denotes the measurements, and the linear operator $\mathbf{H} \in \mathbb{R}^{M \times N}$, with $M \leq N$, models the physical response of the acquisition device. In (3.1), the vector $\mathbf{n}$ is the measurement noise which, from here on, is assumed to be independent and identically distributed (i.i.d.) Gaussian of variance σ^2.

Throughout this chapter, we shall be concerned with the statistical formulation of the reconstruction task based on the prior knowledge of the distribution of the signal. As mentioned in Chapter 2, when the signal is zero-mean Gaussian with covariance $\mathbf{C}_{\mathbf{ss}}$, the MMSE solution is given by the Wiener filter. It can also be formulated as a variational technique in terms of the quadratic minimization problem

$$\mathbf{s}^\star = \arg\min_{\mathbf{s}} \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2 + \sigma^2 \|\mathbf{C}_{\mathbf{ss}}^{-1/2}\mathbf{s}\|_2^2, \tag{3.2}$$

which provides a clear connection between the Gaussian MMSE/MAP estimation and the Tikhonov-type regularization. We remark that (3.2) effectively imposes a constraint on the ℓ_2-norm of the whitened signal, $\mathbf{C}_{\mathbf{ss}}^{-1/2}\mathbf{s}$, which is i.i.d. Gaussian with unit variance.

Reconstruction algorithms that are derived under the hypothesis of Gaussianity and stationarity have played a central role for solving problems that are in spirit of (3.1). They favor linear filtering type implementations and bring powerful optimality criterion for the assumed stochastic setting. Nonetheless, the trend in variational formulations for signal reconstruction has been to move away from quadratic regularization and to change the second term in (3.2) to promote sparse solutions [55]. Majorly, the change has been triggered by the wavelet transforms. Such tools provide us with sparse (or nearly-sparse) representations of many signals that occur naturally [6]. The promotion of sparsity is achieved by specifying well-chosen non-quadratic regularization functionals . One common choice for the regularization functional is $\mathcal{R}_1(\mathbf{v}) = \|\mathbf{v}\|_1$, where $\mathbf{v}$ represents the wavelet (or a wavelet-like multiscale) coefficients of the signal [56]. Combined with computational techniques like cycle-spinning [57], wavelet-based reconstruction methods can greatly improve the reconstruction quality [48]. Another preferred choice of regularization is $\mathcal{R}_2(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|_1$ where $\mathbf{L}$ is the discrete counterpart of the first-order derivative operator, which is known as the TV regularization [35]. This type of regularization is suitable for reconstructing signals with sparse derivatives [58] and has been used in several imaging applications [59, 19, 21]. Although using the ℓ_1-norm as regularization functional has been around for some time (for instance, see [60, 61]), it is still at the heart of sparse signal reconstruction problems. The practical challenge brought by these algorithms is that the final optimization problem is nonsmooth. Consequently, a significant amount of research is dedicated to the design of efficient optimization methods [62].

The present formulations of sparsity-promoting regularizations are mainly motivated by deterministic arguments. In a solely discrete setting, their use can be statistically justified—by means of MAP estimation—if one considers generalized Gaussian, Laplace, or hyper-Laplace priors [8, 9, 63, 64, 65]. The choice of the prior, however, usually depends on the given sparsifying transform. At times, their use is actuated by the observation that the log-likelihood functionals are compatible with non-quadratic regularizers. One can argue that these points make the process self-coherent but rather unmethodical. Moreover, a purely discrete formulation for (3.1) does not necessarily provide us with further insights on the true nature of the signal.

3.2.1 Contributions

In what follows, we would like to establish ensuring statistical links with the sparsity-promoting reconstruction algorithms. To have a thoroughly generative framework, we specify a continuous-domain stochastic model for the signal that does not depend on the final reconstruction task. For this purpose, we rely on the theory of continuous-domain sparse stochastic processes [66]. In this theory, the stochastic process is defined through an innovation model that can be driven by a *non-Gaussian* excitation. [2] The primary advantage of the continuous-domain model is that it lends itself to an analytical treatment. In particular, it allows for the derivation of the probability density function (pdf) of the signal in any transform domain, which is typically much more difficult in a purely discrete framework.

---

[2] It is noteworthy that the theory includes the stationary Gaussian processes. This point is essential as it provides backward-compatibility.

On the modeling level, it provides us with more accurate characterization of the signal content (and its interpretation) before the acquisition step. From a practical perspective, we shall need to be in line with (3.1) since the discrete formulation is pivotal for implementation. Therefore, we provide a proper discretization method that allows us to derive MAP estimators in a principled manner. Remarkably, the underlying class of models provides us with a strict derivation of the class of admissible regularization functionals, which happen to be confined to two categories: Quadratic or sparse.

The main contributions of this chapter are listed as follows:

- The introduction of continuous-domain stochastic models in the formulation of inverse problems. This leads to the use of non-quadratic reconstruction schemes.

- A general framework for the proper discretization of inverse problems. Combined with the stochastic signal modeling, the scheme provides feasible statistical estimators.

- The characterization of the complete class of admissible potential functions (prior log-likelihoods) and the derivation of the corresponding MAP estimators. The connections between these estimators and the existing deterministic methods are also explained. In particular, our estimators cover the classical Tikhonov-type regularizations and a broad class of sparsity-promoting schemes that are commonly used in the field.

- A generic reconstruction algorithm that is based on variable-splitting and ADMM techniques. The method handles the eventual optimization problem stated by the MAP estimation, which can be nonconvex.

3.2.2 Outline

The chapter is organized as follows: In Section 3.3, we explain the acquisition model and obtain the corresponding representation of the signal $\mathbf{s}$ and the system matrix $\mathbf{H}$. In Section 3.4, we introduce the continuous-domain innovation model that defines a generalized stochastic process. We then statistically specify the discrete-domain counterpart of the innovation model and characterize admissible prior distributions. Based on this characterization, we derive the MAP estimation

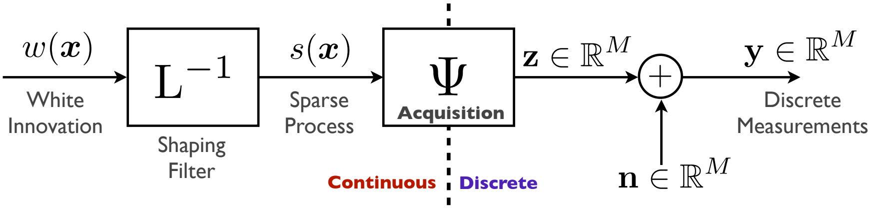


Figure 3.1: General form of the linear, continuous-domain measurement model considered in this chapter. The signal $s(\boldsymbol{x})$ is acquired through linear measurements of the form $z_m = [\Psi s]_m = \langle s, \psi_m \rangle$. The resulting vector $\mathbf{z} \in \mathbb{R}^M$ is corrupted with AWGN. Our goal is to estimate the original signal s from noisy measurements $\mathbf{y}$ by exploiting the knowledge that s is a realization of a sparse stochastic process that satisfies the innovation model $\mathrm{L}s = w$, where w is a non-Gaussian white innovation process.

as an optimization problem in Section 3.5. In Section 3.6, we provide an efficient algorithm to solve the optimization problem for a variety of admissible priors.

3.3 Measurement Model

The foundation of our signal recovery approach is to reconstruct the signal as a d-dimensional function of the space-domain variable $\boldsymbol{x} \in \mathbb{R}^d$ given some noisy measurements $\mathbf{y} \in \mathbb{R}^M$. Since the resulting reconstruction algorithms are implemented numerically, we develop a discretization scheme involving the projection of the solution onto some finite-dimensional reconstruction space. This enables us to obtain a tractable representation of continuously-defined signal reconstruction problem that is in line with (3.1) with minimal loss of information.

3.3.1 Discretization of the Signal

To obtain a clean analytical discretization of the problem, we consider the generalized sampling approach using “shift-invariant” reconstruction spaces [67]. The advantage of such a representation is that it offers the same type of error control as finite-element methods. The approximation error between the original signal and its representation in the reconstruction space can be made arbitrarily small by choosing a sufficiently fine reconstruction grid [67].

The idea is to represent the signal s by projecting it onto a reconstruction space. We define our reconstruction space at resolution T as

$$V_T(\varphi_{\text{int}}) = \left\{ s_T(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} s\,[\boldsymbol{k}]\,\varphi_{\text{int}}\left(\frac{\boldsymbol{x}}{T} - \boldsymbol{k}\right) : s[\boldsymbol{k}] \in \ell_\infty(\mathbb{Z}^d) \right\}, \tag{3.3}$$

where $s[\boldsymbol{k}] = s(\boldsymbol{x})|_{\boldsymbol{x}=T\boldsymbol{k}}$, and φ_{int} is an interpolating basis function positioned on the reconstruction grid $T\mathbb{Z}^d$. We note that the interpolation property is $\varphi_{\text{int}}(\boldsymbol{k}) = \delta[\boldsymbol{k}]$.

For the representation of s in terms of its samples $s[\boldsymbol{k}]$ to be stable and unambiguous, φ_{int} has to be a valid Riesz basis for $V_T(\varphi_{\text{int}})$. Moreover, to guarantee that the approximation error decays as a function of T, the basis function should satisfy the partition of unity property [67]

$$\sum_{\boldsymbol{k}\in\mathbb{Z}^d} \varphi_{\text{int}}(\boldsymbol{x} - \boldsymbol{k}) = 1,\ \forall \boldsymbol{x} \in \mathbb{R}^d. \tag{3.4}$$

The projection of the signal onto the reconstruction space $V_T(\varphi_{\text{int}})$ is then given by

$$P_{V_T} s(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} s(T\boldsymbol{k})\varphi_{\text{int}}\left(\frac{\boldsymbol{x}}{T} - \boldsymbol{k}\right), \tag{3.5}$$

with the idempotent property that $P_{V_T} P_{V_T} s = P_{V_T} s$. To simplify the notation, we shall use a unit sampling $T = 1$ with the implicit assumption that the sampling error is negligible (if the sampling error is large, one can use a finer sampling and rescale the reconstruction grid appropriately).

Thus, the resulting discretization is

$$s_1(\boldsymbol{x}) = P_{V_1} s(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} s[\boldsymbol{k}]\varphi_{\text{int}}(\boldsymbol{x} - \boldsymbol{k}). \tag{3.6}$$

To summarize, $s_1(\boldsymbol{x})$ is the projected version of the original signal $s(\boldsymbol{x})$ and it is uniquely described by the samples $s[\boldsymbol{k}] = s(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{k}}$ for $\boldsymbol{k} \in \mathbb{Z}^d$. The main point is that the reconstructed signal is represented in terms of samples even though the problem is still formulated in the continuous-domain.

3.3.2 Discrete Measurement Model

By using the discretization scheme in (3.6), we are now ready to formally link the continuous model in Figure 3.1 and the corresponding discrete linear inverse problem given in (3.1). Although the signal representation (3.6) is an infinite sum, in practice we restrict ourselves to a subset of N basis functions with $\boldsymbol{k} \in \Omega$, where Ω is a discrete set of integer coordinates in a region-of-interest (ROI). Hence, we rewrite (3.6) as

$$s_1(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \Omega} s[\boldsymbol{k}] \varphi_{\boldsymbol{k}}(\boldsymbol{x}), \tag{3.7}$$

where $\varphi_{\boldsymbol{k}}(\boldsymbol{x})$ corresponds to $\varphi_{\text{int}}(\boldsymbol{x} - \boldsymbol{k})$ up to modifications at the boundaries (periodization or Neumann boundary condition).

Let us first consider a noise-free signal acquisition. The general form of a linear, continuous-domain noise-free measurement system is

$$z_m = \int_{\mathbb{R}^d} s(\boldsymbol{x}) \psi_m(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \ (m = 1, \dots, M) \tag{3.8}$$

where $s(\boldsymbol{x})$ is the original signal, and the measurement function $\psi_m(\boldsymbol{x})$ represents the spatial response of the mth detector which is application dependent as we shall explain in Chapter 4.

By substituting the signal representation (3.6) into (3.8), we discretize the measurement model and write it in matrix-vector form as

$$\mathbf{y} = \mathbf{z} + \mathbf{n} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{3.9}$$

where $\mathbf{y}$ is the M-dimensional measurement vector, $\mathbf{s} = (s[\boldsymbol{k}])_{\boldsymbol{k} \in \Omega}$ is the N-dimensional signal vector, $\mathbf{n}$ is the M-dimensional noise vector, and $\mathbf{H}$ is the $M \times N$ system matrix whose entry $(m, \boldsymbol{k})$ is given by

$$[\mathbf{H}]_{m,\boldsymbol{k}} = \langle \psi_m, \varphi_{\boldsymbol{k}} \rangle = \int_{\mathbb{R}^d} \psi_m(\boldsymbol{x}) \varphi_{\boldsymbol{k}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}. \tag{3.10}$$

We note that this allows us to specify the forward model given in (3.1) which is compatible with the continuous-domain formulation. The solution of this problem yields the representation $s_1(\boldsymbol{x})$ of $s(\boldsymbol{x})$ which is parameterized in terms of the signal samples $\mathbf{s}$. Having the forward model explained, our next aim is to obtain the statistical distribution of $\mathbf{s}$.

3.4 Sparse Stochastic Models

We now proceed by introducing our stochastic framework, which will provide us with a signal prior. For that purpose, we assume that $s(\boldsymbol{x})$ is a realization of a stochastic process that is defined as the solution of a linear stochastic differential equation (SDE) with a driving term that is not necessarily Gaussian. Starting from such a continuous-domain model, we aim at obtaining the statistical distribution of the sampled version of the process (discrete signal) that will be needed to formulate estimators for the reconstruction problem.

3.4.1 Continuous-Domain Innovation Model

As mentioned in Section 3.2, we specify our relevant class of signals as the solution of an SDE in which the process s is assumed to be whitened by a linear operator. This model takes the form

$$\mathrm{L}s = w, \tag{3.11}$$

where w is a continuous-domain white innovation process (the driving term), and L is a (multidimensional) differential operator. The right-hand side of (3.11) represents the unpredictable part of the process, while L is called the whitening operator. Such models are standard in the classical theory of stationary Gaussian processes [68]. The twist here is that the driving term w is not necessarily Gaussian. Moreover, the underlying differential system is potentially unstable to allow for self-similar models.

In the present model, the process s is characterized by the formal solution $s = \mathrm{L}^{-1}w$, where L^{-1} is an appropriate right inverse of L. The operator L^{-1} amounts to some generalized "integration" of the innovation w. The implication is that the correlation structure of the stochastic process s is determined by the shaping operator L^{-1}, whereas its statistical properties and sparsity structure is determined by the driving term w. As an example in the one-dimensional setting, the operator L can be chosen as the first-order continuous-domain derivative operator $\mathrm{L} = \mathrm{D}$. For multidimensional signals, an attractive class of operators is the fractional Laplacian $(-\Delta)^{\frac{\gamma}{2}}$ which is invariant to translation, dilation, and rotation in $\mathbb{R}^d$ [69]. This operator gives rise to "$1/\|\boldsymbol{\omega}\|^{\gamma}$"-type power spectrum and is frequently used to model certain types of images [70, 71, 72]. Such models provide interesting aspects about regularization as we shall detail in Chapter 4.

The mathematical difficulty is that the innovation w cannot be interpreted as an ordinary function because it is highly singular. The proper framework for handling such singular objects is Gelfand and Vilenkin's theory of generalized stochastic processes [73]. In this framework, the stochastic process s is observed by means of scalar-products $\langle s, \varphi \rangle$ with $\varphi \in \mathcal{S}(\mathbb{R}^d)$, where $\mathcal{S}(\mathbb{R}^d)$ denotes the Schwartz class of smooth rapidly decreasing test functions.

A fundamental aspect of the theory is that the driving term w of the innovation model (3.11) is uniquely specified in terms of its Lévy exponent $f(\cdot)$.

Definition 1. *A complex-valued function $f : \mathbb{R} \to \mathbb{C}$ is a valid Lévy exponent iff. it satisfies the three following conditions:*

1. *it is continuous;*
2. *it vanishes at the origin;*
3. *it is conditionally positive-definite of order one in the sense that*

$$\sum_{m=1}^{N} \sum_{n=1}^{N} f(\omega_m - \omega_n) \xi_m \overline{\xi}_n \geq 0$$

under the condition $\sum_{m=1}^{N} \xi_m = 0$ for every possible choice of $\omega_1, \dots, \omega_N \in \mathbb{R}$, $\xi_1, \dots, \xi_N \in \mathbb{C}$, and $N \in \mathbb{N} \setminus \{0\}$.

An important subset of Lévy exponents are the p-admissible ones, which are central to our formulation.

Definition 2. *A Lévy exponent f with derivative f' is called p-admissible if it satisfies the inequality*

$$|f(\omega)| + |\omega||f'(\omega)| \leq C|\omega|^p$$

for some constant $C > 0$ and $0 < p \leq 2$.

A typical example of a p-admissible Lévy exponent is $f(\omega) = -s_0|\omega|^\alpha$ with $s_0 > 0$. The simplest case is $f_{\mathrm{Gauss}}(\omega) = -\frac{1}{2}|\omega|^2$; it will be used to specify Gaussian processes.

Gelfand and Vilenkin have characterized the whole class of continuous-domain white innovation and have shown that they are fully specified by the generic characteristic form

$$\begin{aligned}\widehat{\mathscr{P}}_w(\varphi) &= \mathbb{E}\left\{\mathrm{e}^{\mathrm{j}\langle w,\varphi\rangle}\right\}\\ &= \exp\left(\int_{\mathbb{R}^d} f(\varphi(\boldsymbol{x}))\mathrm{d}\boldsymbol{x}\right),\end{aligned} \tag{3.12}$$

where f is the corresponding Lévy exponent of the innovation process w. The powerful aspect of this characterization is that $\widehat{\mathscr{P}}_w$ is indexed by a test function $\varphi \in \mathcal{S}$ rather than by a scalar (or vector) Fourier variable ω. As such, it constitutes the *infinite-dimensional* generalization of the characteristic function of a conventional random variable.

Recently, Unser et al. characterized the class of stochastic processes that are solutions of (3.11) where L is a linear shift-invariant (LSI) operator and w is a member of the class of so-called Lévy noises [66, Theorem 3].

Theorem 1. *Let w be a Lévy noise as specified by* (3.12) *and* L^{-1*} *be a left inverse of the adjoint operator* L^* *such that either one of the conditions below is met:*

1. L^{-1*} *is a continuous linear map from $\mathcal{S}(\mathbb{R}^d)$ into itself;*
2. *f is p-admissible and L^{-1*} is a continuous linear map from $\mathcal{S}(\mathbb{R}^d)$ into $L_p(\mathbb{R}^d)$; that is,*

$$\|\mathrm{L}^{-1*}\varphi\|_{L_p} < C\|\varphi\|_{L_p}, \quad \forall\varphi \in \mathcal{S}(\mathbb{R}^d)$$

   for some constant C and some $p \geq 1$.

Then, $s = \mathrm{L}^{-1}w$ is a well-defined generalized stochastic process over the space of tempered distributions $\mathcal{S}'(\mathbb{R}^d)$ and is uniquely characterized by its characteristic form

$$\widehat{\mathscr{P}}_s(\varphi) = \mathbb{E}\left\{\mathrm{e}^{\mathrm{j}\langle s,\varphi\rangle}\right\} = \exp\left(\int_{\mathbb{R}^d} f\big(\mathrm{L}^{-1*}\varphi(\boldsymbol{x})\big)\mathrm{d}\boldsymbol{x}\right). \tag{3.13}$$

It is a (weak) solution of the stochastic differential equation $\mathrm{L}s = w$ in the sense that $\langle \mathrm{L}s, \varphi\rangle = \langle w, \varphi\rangle$ for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$.

Before we move on, it is important to emphasize that Lévy exponents are in one-to-one correspondence with the so-called infinitely divisible (i.d.) distributions [74].

Definition 3. *A generic pdf* p_X *is infinitely divisible if, for any positive integer* n*, it can be represented as the* n*-fold convolution* $(p * \cdots * p)$ *where* p *is a valid pdf.*

Theorem 2 (Lévy-Schoenberg). *Let* $\hat{p}_X(\omega) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\omega X}\} = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{j}\omega x} p_X(x)\,\mathrm{d}x$ *be the characteristic function of an infinitely divisible random variable* X*. Then,*

$$f(\omega) = \log \hat{p}_X(\omega)$$

is a Lévy exponent in the sense of Definition 1. Conversely, if $f(\omega)$ *is a valid Lévy exponent, then the inverse Fourier integral*

$$p_X(x) = \int_{\mathbb{R}} \mathrm{e}^{f(\omega)} \mathrm{e}^{-\mathrm{j}\omega x} \frac{\mathrm{d}\omega}{2\pi}$$

yields the pdf of an i.d. random variable.

Another important theoretical result is that it is possible to specify the complete family of i.d. distributions thanks to the celebrated Lévy-Khintchine representation [75] which provides a constructive method for defining Lévy exponents. This tight connection will be essential for our formulation and limits us to a certain family of prior distributions.

3.4.2 Statistical Distribution of Discrete Signal Model

The interest is now to statistically characterize the discretized signal described in Section 3.3.1. To that end, the first step is to formulate a discrete version of the continuous-domain innovation model (3.11). Since, in practical applications, we are only given the samples $(s[\boldsymbol{k}])_{\boldsymbol{k}\in\Omega}$ of the signal, we obtain the discrete-domain innovation model by applying to them the discrete counterpart $\mathrm{L_d}$ of the whitening operator L. The fundamental requirement for our formulation is that the composition of $\mathrm{L_d}$ and L^{-1} results in a stable, shift-invariant operator whose impulse response is well localized [76]

$$\left(\mathrm{L_d}\mathrm{L}^{-1}\delta\right)(\boldsymbol{x}) = \beta_{\mathrm{L}}(\boldsymbol{x}) \in L_1(\mathbb{R}^d). \tag{3.14}$$

The function β_{L} is the generalized B-spline associated with the operator L. Ideally, we would like it to be maximally localized. A necessary requirement for $\mathrm{L_d}$ is that its null space includes the one of L [76].

To give more insight, let us consider $\mathrm{L} = \mathrm{D}$ and $\mathrm{L_d} = \mathrm{D_d}$ (the finite-difference operator associated to D). Then, the associated B-spline is

$$\beta_{\mathrm{D}}(x) = \mathrm{D_d}\mathbb{1}_{+}(x) = \mathbb{1}_{+}(x) - \mathbb{1}_{+}(x-1),$$

where $\mathbb{1}_{+}(x)$ is the unit step (Heaviside) function. Hence, $\beta_{\mathrm{D}}(x) = \mathrm{rect}(x - \frac{1}{2})$ is a causal rectangle function (polynomial B-spline of degree 0). Let us remark that (3.14) bears an important practical consequence that is given by

$$u = \mathrm{L_d}s = \mathrm{L_d}\mathrm{L}^{-1}w = \beta_{\mathrm{L}} * w. \tag{3.15}$$

Since

$$(\beta_{\mathrm{L}} * w)(\boldsymbol{x}) = \langle w, \beta_{\mathrm{L}}^{\vee}(\cdot - \boldsymbol{x})\rangle,$$

where $\beta_{\mathrm{L}}^{\vee}(\boldsymbol{x}) = \beta_{\mathrm{L}}(-\boldsymbol{x})$ is the space-reversed version of β_{L}, it can be inferred from (3.15) that the evaluation of the samples of $\mathrm{L_d}s$ is equivalent to the observation of the innovation through a B-spline window.

From a system-theoretic point of view, $\mathrm{L_d}$ is understood as a finite impulse response (FIR) filter. This impulse response is of the form $\sum_{\boldsymbol{k}\in\Omega} d[\boldsymbol{k}]\delta(\cdot - \boldsymbol{k})$ with some appropriate weights $d[\boldsymbol{k}]$. Therefore, we write the discrete counterpart of the continuous-domain innovation variable as

$$u[\boldsymbol{k}] = \mathrm{L_d}s(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{k}} = \sum_{\boldsymbol{k}'\in\Omega} d[\boldsymbol{k}']s(\boldsymbol{k} - \boldsymbol{k}').$$

This enables us to write in matrix-vector notation the discrete-domain version of the innovation model (3.11) as

$$\mathbf{u} = \mathbf{L}\mathbf{s}, \tag{3.16}$$

where $\mathbf{s} = (s[\boldsymbol{k}])_{\boldsymbol{k}\in\Omega}$ represents the discretization of the stochastic model with $s[\boldsymbol{k}] = s(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{k}}$ for $\boldsymbol{k} \in \Omega$, $\mathbf{L} : \mathbb{R}^N \to \mathbb{R}^N$ is the matrix representation of $\mathrm{L_d}$, and $\mathbf{u} = (u[\boldsymbol{k}])_{\boldsymbol{k}\in\Omega}$ is the discrete innovation vector.

We shall now rely on (3.13) to derive the pdf of the discrete innovation variable, which is one of the key results of this chapter.

Theorem 3. *Let s be a stochastic process whose characteristic form is given by (3.13) where f is a p-admissible Lévy exponent, and $\beta_{\mathrm{L}} = \mathrm{L_d}\mathrm{L}^{-1}\delta \in L_p(\mathbb{R}^d)$ for some $p \in [1,2]$. Then, $u = \mathrm{L_d} s$ is stationary and infinitely divisible. Its first-order pdf is given by*

$$p_U(u) = \int_{\mathbb{R}} \exp\left(f_{\beta_{\mathrm{L}}^{\vee}}(\omega)\right) \mathrm{e}^{\mathrm{j}\omega u} \frac{\mathrm{d}\omega}{2\pi}, \tag{3.17}$$

with Lévy exponent

$$f_{\beta_{\mathrm{L}}^{\vee}}(\omega) = \log \hat{p}_U(\omega) = \int_{\mathbb{R}^d} f\big(\omega\beta_{\mathrm{L}}^{\vee}(\boldsymbol{x})\big)\, \mathrm{d}\boldsymbol{x}, \tag{3.18}$$

which is p-admissible as well.

Proof. Taking (3.15) into account, we derive the characteristic form of u which is given by

$$\begin{aligned}
\widehat{\mathscr{P}}_u(\varphi) &= \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle u,\varphi\rangle}\} = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle \beta_{\mathrm{L}} * w,\varphi\rangle}\} = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle w,\beta_{\mathrm{L}}^{\vee} * \varphi\rangle}\} \\
&= \widehat{\mathscr{P}}_w(\beta_{\mathrm{L}}^{\vee} * \varphi) \\
&= \exp\left(\int_{\mathbb{R}^d} f\big(\beta_{\mathrm{L}}^{\vee} * \varphi(\boldsymbol{x})\big)\, \mathrm{d}\boldsymbol{x}\right).
\end{aligned} \tag{3.19}$$

The fact that u is stationary is equivalent to $\widehat{\mathscr{P}}_u(\varphi) = \widehat{\mathscr{P}}_u\big(\varphi(\cdot - \boldsymbol{x}_0)\big)$ for any $\boldsymbol{x}_0 \in \mathbb{R}^d$, which is established by a simple change of variable in (3.19). We now consider the random variable $U = \langle u, \delta\rangle = \langle w, \beta_{\mathrm{L}}^{\vee}\rangle$. Its characteristic function is obtained as

$$\begin{aligned}
\hat{p}_U(\omega) &= \mathbb{E}\{\mathrm{e}^{\mathrm{j}\omega U}\} = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\langle w,\omega\beta_{\mathrm{L}}^{\vee}\rangle}\} \\
&= \widehat{\mathscr{P}}_w(\omega\beta_{\mathrm{L}}^{\vee}) \\
&= \exp\left(f_{\beta_{\mathrm{L}}^{\vee}}(\omega)\right)
\end{aligned}$$

where the substitution $\varphi = \omega\beta_{\mathrm{L}}^{\vee}$ in $\widehat{\mathscr{P}}_w(\varphi)$ is valid since $\widehat{\mathscr{P}}_w$ is a continuous functional on $L_p(\mathbb{R}^d)$ as a consequence of the p-admissibility condition. To prove that

$f_{\beta_{\mathrm{L}}^{\vee}}(\omega)$ is a p-admissible Lévy exponent, we start by establishing the bound

$$\begin{aligned} C\|\varphi\|_{L_p}^p|\omega|^p &\geq \int_{\mathbb{R}^d} \left|f\left(\omega\beta_{\mathrm{L}}^{\vee}(\boldsymbol{x})\right)\right| \,\mathrm{d}\boldsymbol{x} \\ &\quad + |\omega| \int_{\mathbb{R}^d} \left|f'\left(\omega\beta_{\mathrm{L}}^{\vee}(\boldsymbol{x})\right)\varphi(\boldsymbol{x})\right| \,\mathrm{d}\boldsymbol{x} \\ &\geq \left|f_{\beta_{\mathrm{L}}^{\vee}}(\omega)\right| + |\omega| \left|f'_{\beta_{\mathrm{L}}^{\vee}}(\omega)\right|, \end{aligned} \tag{3.20}$$

which follows from the p-admissibility of f. We are also relying on Lebesgue's dominated convergence theorem to move the derivative with respect to ω inside the integral that defines $f_{\beta_{\mathrm{L}}^{\vee}}(\omega)$. In particular, (3.20) implies that $f_{\beta_{\mathrm{L}}^{\vee}}$ is continuous and vanishes at the origin. The last step is to establish its conditional positive definiteness which is achieved by interchanging the order of summation. We write

$$\begin{aligned} &\sum_{m=1}^{N}\sum_{n=1}^{N} f_{\beta_{\mathrm{L}}^{\vee}}(\omega_m - \omega_n)\xi_m\overline{\xi}_n = \\ &\int_{\mathbb{R}^d} \underbrace{\sum_{m=1}^{N}\sum_{n=1}^{N} f\left(\omega_m\beta_{\mathrm{L}}^{\vee}(\boldsymbol{x}) - \omega_n\beta_{\mathrm{L}}^{\vee}(\boldsymbol{x})\right)\xi_m\overline{\xi}_n}_{\geq 0} \,\mathrm{d}\boldsymbol{x} \geq 0 \end{aligned} \tag{3.21}$$

under the condition $\sum_{m=1}^{N}\xi_m = 0$ for every possible choice of $\omega_1,\dots,\omega_N \in \mathbb{R}$, $\xi_1,\dots,\xi_N \in \mathbb{C}$, and $N \in \mathbb{N}\setminus\{0\}$. $\square$

The direct consequence of Theorem 3 is that the primary statistical features of $\mathbf{u}$ is directly related to the continuous-domain innovation process w via the Lévy exponent. This implies that the sparsity structure (tail behavior of the pdf and/or presence of a mass distribution at the origin) is primarily dependent upon f. The important conceptual aspect, which follows from the Lévy-Schoenberg theorem, is that the class of admissible pdfs is restricted to the family of i.d. laws since $f_{\beta_{\mathrm{L}}^{\vee}}(\omega)$, as given by (3.18), is a valid Lévy exponent. We emphasize that this result is attained by taking advantage of considerations in the continuous-domain.

3.4.3 Illustrative Examples

We now would like to present our formalism by highlighting some specific examples. If we choose $\mathrm{L} = \mathrm{D}$, then the solution of (3.11) with the boundary condition $s(0) = 0$ is given by

$$s(x) = \int_0^x w(x')\mathrm{d}x'$$

and is a Lévy process. It is noteworthy that the Lévy processes—a fundamental and well-studied family of stochastic processes—include Brownian motion and Poisson processes which are commonly used to model physical random phenomena [74].

When Lévy processes are considered, $\beta_{\mathrm{D}}(x) = \mathrm{rect}(x - \frac{1}{2})$ and the discrete innovation vector $\mathbf{u}$ is obtained by

$$\begin{aligned} u[k] &= \langle w, \mathrm{rect}(\cdot + \tfrac{1}{2} - k)\rangle \\ &= s(k) - s(k-1). \end{aligned}$$

Since the B-splines are non-overlapping, we can deduce that the increments $(u[k])_{k\in\Omega}$ are i.i.d., which is the defining property of a Lévy process [74]. This property also qualifies the class of such objects as a convenient benchmarking platform for different denoising algorithms [58].

Evaluating (3.18) together with $f(0) = 0$ (see Definition 1), we obtain

$$f_{\beta_{\mathrm{D}}^{\vee}}(\omega) = \int_{-1}^{0} f(\omega)\mathrm{d}x = f(\omega).$$

In particular, we generate Lévy processes with Laplace-distributed increments by choosing $f(\omega) = \log(\frac{\tau^2}{\tau^2+\omega^2})$ with the scale parameter $\tau > 0$. To see that, we write $\exp(f_{\beta_{\mathrm{D}}^{\vee}}(\omega)) = \hat{p}_U(\omega) = \frac{\tau^2}{\tau^2+\omega^2}$ via (3.18). The inverse Fourier transform of this rational function is known to be

$$p_U(u) = \frac{\tau}{2}\mathrm{e}^{-\tau|u|}.$$

Also, we rely on Theorem 3 in a more general aspect. For instance, a special case of interest is the Gaussian (nonsparse) scenario where $f_{\mathrm{Gauss}}(\omega) = -\frac{1}{2}|\omega|^2$. Therefore, one gets $f_{\beta_{\mathrm{L}}^{\vee}}(\omega) = \log \hat{p}_U(\omega) = -\frac{1}{2}\omega^2\|\beta_{\mathrm{L}}\|_2^2$ from (3.18). Plugging this into (3.17), we deduce that the discrete innovation vector is zero-mean Gaussian with variance $\|\beta_{\mathrm{L}}\|_2^2$ (i.e., $p_U(u) = \mathcal{N}(0, \|\beta_{\mathrm{L}}\|_2^2)$).

Table 3.1: Four members of the family of infinitely divisible distributions and the corresponding potential functions.

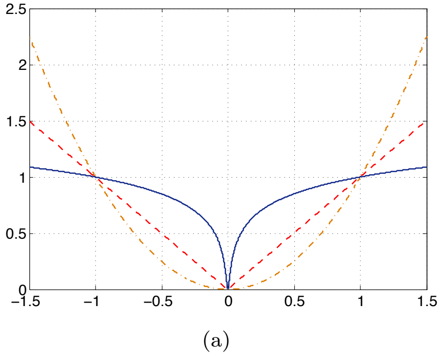
	$p_U(x)$	$\Phi_U(x)$	Property
Gaussian	$\frac{1}{\sigma_0\sqrt{2\pi}}e^{-x^2/2\sigma_0^2}$	$a_1x^2+b_1$	smooth, convex
Laplace	$\frac{\tau}{2}e^{-\tau\lvert x\rvert}$	$a_2\lvert x\rvert+b_2$	nonsmooth, convex
Student's t	$\frac{1}{\epsilon B(r,\frac{1}{2})}\left(\frac{1}{(x/\epsilon)^2+1}\right)^{r+\frac{1}{2}}$	$a_3\log\left(\frac{x^2+\epsilon^2}{\epsilon^2}\right)+b_3$	smooth, nonconvex
Cauchy	$\frac{1}{\pi s_0}\frac{1}{(x/s_0)^2+1}$	$\log(\frac{x^2+s_0^2}{s_0^2})+b_4$	smooth, nonconvex

Additionally, when $f(\omega)=\frac{-|\omega|^\alpha}{2}$ with $\alpha\in[1,2]$, one finds that $f_{\beta_{\mathrm{L}}^\vee}(\omega)=\log\hat{p}_U(\omega)=-\frac{|\omega|^\alpha}{2}\|\beta_{\mathrm{L}}\|_{L_\alpha}^\alpha$. This indicates that $\mathbf{u}$ is a symmetric α-stable ($S\alpha S$) distribution with scale parameter $s_0=\|\beta_{\mathrm{L}}\|_{L_\alpha}^\alpha$. For $\alpha=1$, we have the Cauchy distribution (or Student's t-distribution with $r=1/2$).

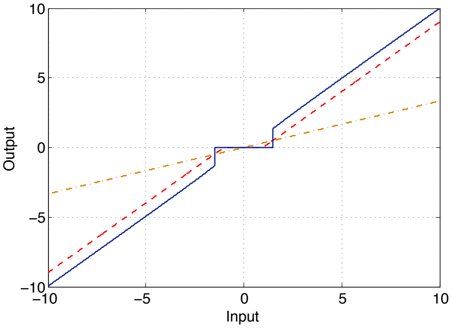
Considering other i.d. laws, the inverse Fourier transformation (3.17) is often harder to compute analytically, but it can still be performed numerically to determine $p_U(u)$ (or its corresponding potential function $\Phi_U=-\log p_U$). In general, p_U will be i.d. and will typically imply heavy tails. Note that heavy-tailed distributions exhibit attractive properties for compressibility [77, 78].

3.5 Bayesian Estimation

We now use the results of Section 3.4 to derive solutions to the reconstruction problem in some well-defined statistical sense. To that end, we concentrate on the MAP solutions that are presently derived under the decoupling assumption that the components of $\mathbf{u}$ are independent and identically distributed (i.i.d.). This assumption is exact when L is a first-order differential operator (such as the derivative) in which case the B-spline is of unit support. For higher-order operators, the decoupling has local dependencies over the support of β_{L} that can be worked out explicitly [79]. However, taking these into account results in more complicated estimation algorithms.



(a)



(b)

Figure 3.2: Potential functions (a) and the corresponding proximity operators (b) of different prior distributions: Gaussian prior (dash-dotted), Laplacian prior (dashed), and Student's t-prior ($\epsilon = 10^{-2}$) (solid). For illustrative purposes, the multiplication factors are set such that $\Phi_U(1) = 1$ for all potential functions.

3.5.1 MAP Formulation

In order to reconstruct the signal, we seek an estimate of $\mathbf{s}$ that maximizes the posterior distribution $p_{S|Y}$ which depends upon the prior distribution p_S, assumed to be proportional to p_U (since $\mathbf{u} = \mathbf{L}\mathbf{s}$). The direct application of Bayes' rule is

$$\begin{aligned} p_{S|Y}(\mathbf{s} \mid \mathbf{y}) &\propto p_N(\mathbf{y} - \mathbf{H}\mathbf{s}) p_U(\mathbf{u}) \\ &\propto \exp\left(-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{s}\|^2}{2\sigma^2}\right) \prod_{\boldsymbol{k}\in\Omega} p_U\big([\mathbf{L}\mathbf{s}]_{\boldsymbol{k}}\big). \end{aligned}$$

Then, we write the MAP estimation for $\mathbf{s}$ as

$$\begin{aligned} \mathbf{s}^\star_{\mathrm{MAP}} &= \arg\max_{\mathbf{s}} p_{S|Y}(\mathbf{s} \mid \mathbf{y}) \\ &= \arg\min_{\mathbf{s}} \left(\tfrac{1}{2}\|\mathbf{H}\mathbf{s} - \mathbf{y}\|_2^2 + \sigma^2 \sum_{\boldsymbol{k}\in\Omega} \Phi_U\big([\mathbf{L}\mathbf{s}]_{\boldsymbol{k}}\big)\right), \end{aligned} \tag{3.22}$$

where $\Phi_U(x) = -\log p_U(x)$ is called the *potential function* corresponding to p_U. Note that (3.22) is compatible with the standard form of the variational reconstruction formulation as explained in Chapter 2. In the next section, we focus on the potential functions.

3.5.2 Potential Functions

Recall that, in the current Bayesian formulation, the potential function

$$\Phi_U(x) = -\log p_U(x)$$

is specified by the Lévy exponent $f_{\beta_{\mathrm{L}}^\vee}$, which is itself in direct relation with the continuous-domain innovation w via (3.18). For illustration purposes, we consider three members of the i.d. family: Gaussian, Laplace, and Student's t (or, equivalently, Cauchy) distributions. We provide the potential functions for these priors in Table 3.1.[3] On one hand, we already know that the Gaussian prior does not correspond to a sparse reconstruction. On the other hand, the Student's t-prior has a slower tail decay and promotes sparser solutions than the Laplace prior. Also, to provide a geometrical intuition of how the Student's t-prior increases the sparsity of the solution, we plot the potential functions for Gaussian, Laplacian, and Student's

---

[3] The exact values of the constants b_1, b_2, b_3, and b_4 and the positive scaling factors a_1, a_2, and a_3 have been omitted since they are irrelevant to the optimization problem.

t estimators in Figure 3.2. By looking at Figure 3.2, we see that the Student's t-estimator penalizes small values more than the Laplacian or Gaussian counterparts do. Conversely, it penalizes the large values less.

Let us point out some connections between the general estimator (3.22) and the standard variational methods. The first quadratic potential function (Gaussian estimator) yields the classical Tikhonov-type regularizer and produces a stabilized linear solution, as explained in Section 3.2. The second potential function (Laplace estimator) provides the ℓ_1-type regularizer. Moreover, the well-known TV regularizer [35] is obtained if the operator L is a first-order derivative operator. Interestingly, the third log-based potential (Student's t-estimator) is linked to the limit case of the ℓ_p relaxation scheme as $p \to 0$ [80]. To see the relation, we note that minimizing $\lim_{p\to 0}\sum_i |x_i|^p$ is equivalent to minimizing $\lim_{p\to 0}\sum_i \frac{|x_i|^p-1}{p}$. Then, it holds that

$$\lim_{p\to 0}\sum_i \frac{|x_i|^p-1}{p} = \sum_i \log|x_i| = \sum_i \tfrac{1}{2}\log|x_i|^2 \leq \frac{1}{2}\sum_i \log(x_i^2+\kappa) \tag{3.23}$$

for any $\kappa \geq 0$. The key observation is that the upper-bounding log-based potential function in (3.23) is interpretable as a Student's t-prior. This kind of regularization has been considered by different authors (see [81, 82, 32] and also [83, 84, 37] where the authors consider a similar log-based potential) to encourage highly sparse reconstructions.

3.6 Reconstruction Algorithm

We have now the necessary elements to derive the general MAP solution of our image reconstruction problem. By using the discrete innovation vector $\mathbf{u}$ as an auxiliary variable, we naturally recast the MAP estimation as the constrained optimization problem

$$\begin{aligned} \mathbf{s}^{\star}_{\mathrm{MAP}} = \quad & \arg\min_{\mathbf{s}\in\mathbb{R}^K} \quad \left(\frac{1}{2}\|\mathbf{Hs}-\mathbf{y}\|_2^2 + \sigma^2 \sum_{\boldsymbol{k}\in\Omega} \Phi_U\left(u[\boldsymbol{k}]\right)\right) \\ & \text{subject to} \quad \mathbf{u} = \mathbf{Ls}. \end{aligned} \tag{3.24}$$

This representation of the solution naturally suggests using the type of splitting-based techniques that have been employed by various authors for solving similar optimization problems [36, 85, 52]. Rather than dealing with a constrained optimization problem directly, we prefer to formulate an equivalent unconstrained problem. To that purpose, we rely on the augmented Lagrangian (AL) method [86] and introduce the corresponding AL functional of (3.24) given by

$$\mathcal{L}_{\mathcal{A}}(\mathbf{s},\mathbf{u},\boldsymbol{\alpha}) \quad = \quad \frac{1}{2}\|\mathbf{H}\mathbf{s}-\mathbf{y}\|_2^2+\sigma^2\sum_{\boldsymbol{k}\in\Omega}\Phi_U\left(u[\boldsymbol{k}]\right)+\boldsymbol{\alpha}^{\mathrm{T}}(\mathbf{L}\mathbf{s}-\mathbf{u})+\frac{\mu}{2}\|\mathbf{L}\mathbf{s}-\mathbf{u}\|_2^2,$$

where $\boldsymbol{\alpha}\in\mathbb{R}^N$ denotes the Lagrange multiplier and $\mu\in\mathbb{R}$ is the penalty parameter. To obtain the solution, we apply ADMM [53] that replaces the joint minimization of the AL functional over $(\mathbf{s},\mathbf{u})$ by partial minimizations of $\mathcal{L}_{\mathcal{A}}$ with respect to each independent variable in turn, while keeping the other variable fixed. These independent minimizations are followed by the update of the Lagrange multiplier. In summary, ADMM results in the following scheme at iteration t:

$$\mathbf{u}^{t+1}\leftarrow\arg\min_{\mathbf{u}}\mathcal{L}_{\mathcal{A}}(\mathbf{s}^t,\mathbf{u},\boldsymbol{\alpha}^t) \tag{3.25a}$$

$$\mathbf{s}^{t+1}\leftarrow\arg\min_{\mathbf{s}}\mathcal{L}_{\mathcal{A}}(\mathbf{s},\mathbf{u}^{t+1},\boldsymbol{\alpha}^t) \tag{3.25b}$$

$$\boldsymbol{\alpha}^{t+1}=\boldsymbol{\alpha}^t+\mu(\mathbf{L}\mathbf{s}^{t+1}-\mathbf{u}^{t+1}). \tag{3.25c}$$

From the Lagrangian duality point of view, (3.25c) can be interpreted as the maximization of the dual functional so that, as the above scheme proceeds, feasibility is imposed [53].

Alternating Minimization Steps

We now explain the alternating minimization steps. Let us focus on the sub–problem (3.25a). First, we observe the relation

$$\arg\min_{\mathbf{u}}\mathcal{L}_{\mathcal{A}}=\arg\min_{\mathbf{u}}\quad\sigma^2\sum_{\boldsymbol{k}\in\Omega}\Phi_U\left(u[\boldsymbol{k}]\right)+\boldsymbol{\alpha}^{\mathrm{T}}(\mathbf{L}\mathbf{s}-\mathbf{u})+\frac{\mu}{2}\|\mathbf{L}\mathbf{s}-\mathbf{u}\|_2^2 \tag{3.26}$$

$$\equiv\arg\min_{\mathbf{u}}\quad\sigma^2\sum_{\boldsymbol{k}\in\Omega}\Phi_U\left(u[\boldsymbol{k}]\right)+\frac{\mu}{2}\left(u[\boldsymbol{k}]-\left([\mathbf{L}\mathbf{s}]_{\boldsymbol{k}}+\frac{\alpha[\boldsymbol{k}]}{\mu}\right)\right)^2, \tag{3.27}$$

where the equivalence is obtained by completing the squares and ignoring the constant values that are irrelevant to minimization purposes.

In effect, we see that the minimization is separable, which implies that (3.25a) reduces to performing K scalar minimizations of the form

$$\min_{u[\boldsymbol{k}]\in\mathbb{R}} \left(\sigma^2\Phi_U(u[\boldsymbol{k}]) + \frac{\mu}{2}\left(u[\boldsymbol{k}] - z[\boldsymbol{k}]\right)^2\right) \text{ , } \forall \boldsymbol{k} \in \Omega, \tag{3.28}$$

where $\mathbf{z} = \mathbf{Ls} + \boldsymbol{\alpha}/\mu$. One sees that (3.28) is nothing but the proximity operator of $\Phi_U(\cdot)$ that is defined below.

Definition 4. *The proximity operator associated to the function $\lambda\Phi_U(\cdot)$ with $\lambda \in \mathbb{R}_+$ is defined as*

$$\operatorname{prox}_{\Phi_U}(y;\lambda) = \arg\min_{x\in\mathbb{R}} \frac{1}{2}(y-x)^2 + \lambda\Phi_U(x). \tag{3.29}$$

At each iteration t, the solution of (3.25a) is obtained by applying $\operatorname{prox}_{\Phi_U}$, with $\lambda = \sigma^2/\mu$, in a *component-wise* fashion to $\mathbf{z}^t = \mathbf{Ls}^t + \boldsymbol{\alpha}^t/\mu$. The closed-form solutions for the proximity operator are well-known for the Gaussian and Laplace priors [56]. For $\boldsymbol{k} \in \Omega$, they are given by

$$\operatorname{prox}_{(\cdot)^2}(z[\boldsymbol{k}];\lambda) = z[\boldsymbol{k}](1+2\lambda)^{-1}, \tag{3.30a}$$

$$\operatorname{prox}_{|\cdot|}(z[\boldsymbol{k}];\lambda) = \max(|z[\boldsymbol{k}]| - \lambda, 0)\operatorname{sgn}(z[\boldsymbol{k}]), \tag{3.30b}$$

respectively. The proximity operator has no closed-form solution for the potential function of Student's t-distribution. However, the global solution of each scalar problem can still be computed. The first-order optimality condition requires that the global minimizer is a stationary point. Hence, the minimizer satisfies

$$\frac{\partial}{\partial u[\boldsymbol{k}]}\left(\sigma^2\Phi_U(u[\boldsymbol{k}]) + \frac{\mu}{2}\left(u[\boldsymbol{k}] - z[\boldsymbol{k}]\right)^2\right) = 0.$$

For $\Phi_U(x) = \log\left((x^2+\epsilon^2)/\epsilon^2\right)$, we have the third-order root finding problem of the form

$$\mu u^3[\boldsymbol{k}] + (z[\boldsymbol{k}] - \mu)u^2[\boldsymbol{k}] + (2\sigma^2 - \epsilon^2)u[\boldsymbol{k}] - \epsilon\mu z[\boldsymbol{k}] = 0.$$

The identification of the global minimizer among the roots can be easily computed and stored in a lookup table (LUT) (see Figure 3.2(b)) for a suitable range of values. This idea suggests a very fast implementation of the proximal step which is applicable to the entire class of potential functions of i.d. distributions.

Lastly, we consider the second minimization problem (3.25b), which amounts to the minimization of a quadratic problem. Again, the minimizer satisfies that

$$\frac{\partial \mathcal{L}_{\mathcal{A}}}{\partial \mathbf{s}} = \mathbf{H}^{\mathrm{T}}\mathbf{H}\mathbf{s} - \mathbf{H}^{\mathrm{T}}\mathbf{y} + \mathbf{L}^{\mathrm{T}}\boldsymbol{\alpha} + \mu\mathbf{L}^{\mathrm{T}}\mathbf{L}\mathbf{s} - \mu\mathbf{L}^{\mathrm{T}}\mathbf{u} = \mathbf{0}. \tag{3.31}$$

At each iteration, the solution of (3.25b) is given by

$$\mathbf{s}^{t+1} = (\mathbf{H}^{\mathrm{T}}\mathbf{H} + \mu\mathbf{L}^{\mathrm{T}}\mathbf{L})^{-1}\left(\mathbf{H}^{\mathrm{T}}\mathbf{y} + \mu\mathbf{L}^{\mathrm{T}}\left(\mathbf{u}^{t+1} - \frac{\boldsymbol{\alpha}^{t}}{\mu}\right)\right). \tag{3.32}$$

Interestingly, this part of the reconstruction algorithm is compatible with the Gaussian MAP/MMSE solution as given in Chapter 2. In a general setting, this problem is solved iteratively using a linear solver such as the conjugate-gradient (CG) method. It is noteworthy that the condition $\mathrm{Ker}(\mathbf{H}) \cap \mathrm{Ker}(\mathbf{L}) = \{\mathbf{0}\}$ should be satisfied to have a well-defined and unique solution of (3.25b). We note that the computational complexity of the shrinkage step is linear with respect to the given data size. Therefore, the dominating part of the computation is the matrix inversion in (3.32). Also in some cases, the direct inversion is possible, which makes the algorithm highly efficient. We shall discuss such aspects in details in Chapter 4.

We conclude this section with some remarks regarding the optimization algorithm. Note that the method remains applicable when $\Phi_U(x)$ is nonconvex, with the following caveat: The convergence is not guaranteed. However, when the ADMM converges and Φ_U is nonconvex, it converges to a local minimum, including the case where the sub-minimization problems are solved exactly [53]. As the potential functions considered in the present context are closed and proper, we stress the fact that if $\Phi_U : \mathbb{R} \to \mathbb{R}_+$ is convex and the unaugmented Lagrangian functional has a saddle point, then the constraint in (3.24) is satisfied and the objective functional reaches the optimal value as $t \to \infty$ [53]. Meanwhile, in the case of nonconvex problems, the algorithm can converge to different local minima depending on the initial point. It is therefore recommended to apply a deterministic continuation method [64, 38] or to consider a reasonable initial solution that can be obtained by solving the problem first with Gaussian or Laplace priors.

3.7 Summary

The purpose of this chapter has been to develop a practical scheme for linear inverse problems by combining a proper discretization method and the theory of continuous-domain sparse stochastic processes. In basic terms, our model is composed of two fundamental concepts: the whitening operator L, which is in connection with the regularization operator, and the Lévy exponent f, which is related to the prior distribution. An important theoretical implication of our formalism is that the potential functions are linked to infinitely divisible distributions. The latter sounds restrictive at first, but it essentially provides us with a sufficiently large class of MAP estimators. Particularly, we have shown that the derived MAP estimators cover the current state-of-the-art methods in the field including TV-type regularizers. Another interesting observation is that we face an optimization problem for MAP estimation that is generally nonconvex, with the exception of the Gaussian and the Laplacian priors. We have proposed a computational solution, based on ADMM, that applies to arbitrary potential functions by suitable adaptation of the proximity operator.

Chapter 4

Sparse Reconstructions: *In-Silico* Considerations

4.1 Overview

In this chapter, [1] we present the MAP reconstruction of biomedical images. In particular, we concentrate on three different imaging modalities and consider the problems of deconvolution, MR image reconstruction from partial Fourier coefficients, and image reconstruction from X-ray tomograms. For each of these problems, we explain how the image acquisition model is constructed—by using the discretization scheme described in the previous chapter—while keeping the computational complexity aspects in mind. The reconstruction is performed via the generic ADMM-based algorithm where we exploit the structure of the optimization problem to have an efficient implementation. In a broader prospect, the goal of this chapter is to gain an understanding of the relation between the regularization and the characteristics of the underlying image. Thus, for a fixed imaging modality, we perform model-based image reconstructions, where we highlight images that suit well to a particular MAP estimator. These examinations will provide us with the right background for choosing (and also designing) suitable regularization frameworks in the coming chapters.

---

[1] The chapter is based on our papers [54, 87].

4.2 General Framework

Let us first explain some points in our reconstruction framework that are common to all different imaging modalities considered in the sequel. To simulate the noisy measurements $\mathbf{y}$, we degrade the noise-free measurements (i.e., the application of the system matrix $\mathbf{H}$ to some noise-free image $\mathbf{s}$) with AWGN. The images are extended using periodic boundary conditions. We analyze three different priors associated with varying levels of sparsity. Based on our developments in Chapter 3, we shall use the following potential functions Φ_U:

- Gaussian prior where $\Phi_{\mathrm{Gauss}}(x) = a_1|x|^2$: We obtain the classical linear reconstruction. In this case, one does not impose sparsity on the solution. The optimization is convex and differentiable.

- Laplace prior where $\Phi_{\mathrm{Laplace}}(x) = a_2|x|$: We obtain the ℓ_1-type reconstruction. The method encourages sparsity in the reconstructions. The optimization is convex but not differentiable.

- Student's t-prior where $\Phi_{\mathrm{Student's}}(x) = a_3\log((x^2+\epsilon^2)/\epsilon^2)$: We obtain the ℓ_p-type reconstruction as $p \to 0$. Within our numerical comparisons, this configuration corresponds to the highest level of sparsity in the reconstructions. The ϵ parameter is set to 10^{-2} based on the geometric arguments illustrated in Figure 3.2(b). The optimization is nonconvex but differentiable.

For the image reconstruction, we use the iterative algorithm introduced in Section 3.6. Working with two-dimensional images, we shall explain how to adapt the method to the chosen settings and the system matrix $\mathbf{H}$ that is to be constructed in a problem-specific way. For the sake of completeness, we also would like to provide insights into how the reconstruction can be carried out numerically. This task is addressed by providing the mathematical description of the operators taking part in the iterations. We also analyze the corresponding image processing operations (such as filtering and pointwise operations) and the complexity issues. This is followed by some numerical experiments where we discuss certain computational aspects in a more detailed manner.

4.3 Computational Recipes

We have seen that the image reconstruction algorithm outlined in Section 3.6 is composed of two essential steps: 1) the application of the proximal mapping associated to a potential function and 2) the matrix inversion whose structure depends on the specific imaging problem at hand. Along these lines, we develop (and refine when necessary) the strategies that make the reconstruction computationally efficient.

To begin with, let Ω define the index set of all sample locations of the image. Without loss of generality, it is assumed to be the square region

$$\Omega = \left\{ \boldsymbol{k} = (k_1, k_2) \in \mathbb{Z}^2 \,\middle|\, 1 \le k_1 \le n, 1 \le k_2 \le n \right\},$$

where $N = n^2$ denotes the total number of sample locations (i.e., $\mathrm{card}(\Omega) = N$). The vector representation of the underlying image is given by

$$\mathbf{s} = \mathrm{vect}\left(s[\boldsymbol{k}]\right)_{\boldsymbol{k}\in\Omega} \in \mathbb{R}^N,$$

where the vectorization is assumed to be lexicographically ordered.

4.3.1 Analysis of Circulant Operators

We now review some basic computational principles regarding the block circulant matrices with circulant entries (BCCB)[2] as they are of special interest in our formulation. In particular, the structure of such operators enables efficient computations. First, we remark that

$$\begin{aligned} \mathbf{g} &= \mathrm{vect}\left(g[\boldsymbol{k}] = (h_t * s)[\boldsymbol{k}]\right)_{\boldsymbol{k}\in\Omega} \\ &= \mathbf{Ts}, \end{aligned}$$

---

[2] A BCCB matrix $\mathbf{T}$ is of the form

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_0 & \mathbf{T}_{M-1} & \dots & \mathbf{T}_1 \\ \mathbf{T}_1 & \mathbf{T}_0 & \dots & \mathbf{T}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{T}_{M-1} & \mathbf{T}_{M-2} & \dots & \mathbf{T}_0 \end{pmatrix}, \tag{4.1}$$

where each block $\mathbf{T}_j$ for $j = 0, 1, \dots, M-1$ is a circulant matrix.

Algorithm 4: Operations with circulant matrix $\mathbf{T}$

```
input  : Dense matrix $\widehat{\mathbf{E}}_\mathbf{T}$ containing the eigenvalues of $\mathbf{T}$ and an image $s$
output: Resulting image $g$
$\hat{s} \leftarrow$ FFT2($s$) (two-dimensional DFT);
if g == Ts
    $\hat{g} \leftarrow \widehat{\mathbf{E}}_\mathbf{T} \circ \hat{s}$ (pointwise multiplication);
else if g == $\mathbf{T}^{-1}$s
    $\hat{g} \leftarrow \widehat{\mathbf{E}}_\mathbf{T}^{-1} \circ \hat{s}$ (pointwise division);
end if
$g \leftarrow$ IFFT2($\hat{g}$) (two-dimensional inverse DFT);
return $g$;
```

where $\mathbf{T} \in \mathbb{R}^{N\times N}$ is a BCCB matrix whose entries can be directly [3] deduced from the ones of the shift-invariant convolution kernel h_t and $*$ denotes the discrete convolution that is circular as a result of our assumption on the boundary conditions.

Theorem 4 (Spectral Decomposition Theorem [14])**.** *If the matrix* $\mathbf{T} \in \mathbb{R}^{N\times N}$ *is BCCB, then there exists a spectral decomposition of the form*

$$\mathbf{T} = \mathbf{F}^{-1}\mathbf{\Lambda}_\mathbf{T}\mathbf{F},$$

where $\mathbf{\Lambda}_\mathbf{T} \in \mathbb{C}^{N\times N}$ *is a diagonal matrix containing the eigenvalues of* $\mathbf{T}$ *and* $\mathbf{F} \in \mathbb{C}^{N\times N}$ *is the unitary two-dimensional discrete Fourier transform (DFT) matrix.*

In the Fourier domain, the output vector is hence computed by using Theorem 4 as $\hat{g}[\boldsymbol{k}] = [\mathbf{\Lambda}_\mathbf{T}]_{\boldsymbol{k},\boldsymbol{k}}\ \hat{s}[\boldsymbol{k}]$, for $\boldsymbol{k} \in \Omega$. Knowing that the first column of $\mathbf{F}$ is a vector of all ones, the eigenvalues of a BCCB matrix is completely characterized by its first column. Let $\mathbf{t}_1$ be the first column of $\mathbf{T}$, and $\mathbf{E}_\mathbf{T}$ be its $(n \times n)$ dense matrix representation in the sense that $\mathbf{t}_1 = \mathrm{vec}(\mathbf{E}_\mathbf{T})$. Then, we have

$$\mathbf{\Lambda}_\mathbf{T} = \mathrm{diag}(\mathrm{vec}(\widehat{\mathbf{E}}_\mathbf{T})),$$

where $\widehat{\mathbf{E}}_\mathbf{T}$ denotes the two-dimensional DFT of $\mathbf{E}_\mathbf{T}$. In terms of image processing, the implementation is carried out via two FFTs and pointwise multiplication. Similarly, since $\mathbf{\Lambda}_\mathbf{T}$ is a diagonal matrix, we can easily invert a BCCB matrix given

---

[3] See [88] for an illustrative example.

that the inverse $\mathbf{T}^{-1}$ exists. In both cases, the computational cost is $\mathcal{O}(N \log n)$ (see Algorithm 4).

4.3.2 Operator-Specific Regularization

Prior to examining the performance of the chosen potential functions, we briefly depict some specific aspects of the regularization. Since our primary goal is to assess the practical influence of sparse reconstructions, we choose the regularization operator as the magnitude of the image gradient.[4] Having fixed the regularization operator, we continue our computational developments by establishing the implementation details with respect to the current settings.

We write the discrete regularization functional as

$$\mathcal{R}(\mathbf{s}) = \tau \sum_{\boldsymbol{k} \in \Omega} \Phi_U \left(\| [\mathbf{L}\mathbf{s}]_{\boldsymbol{k}} \|_2 \right), \tag{4.2}$$

where τ is the regularization parameter that refers to grouping all the multiplicative factors (i.e., $a_i \sigma^2$ for $i = 1, 2, 3$). In (4.2), the discrete gradient operator is denoted by $\mathbf{L} : \mathbb{R}^N \to \mathbb{R}^{N \times 2} : \mathbf{s} \mapsto \big(\mathrm{vect}(\mathbf{D}_1 \mathbf{s})\ \mathrm{vect}(\mathbf{D}_2 \mathbf{s})\big)$, where each partial derivative $\mathbf{D}_j$ for $j = 1, 2$ is discretized by using forward finite differences. Accordingly, the gradient vector at the $\boldsymbol{k}$-th sample position is represented by

$$[\mathbf{L}\mathbf{s}]_{\boldsymbol{k}} = (s[\boldsymbol{k} + \mathbf{e}_1] - s[\boldsymbol{k}], s[\boldsymbol{k} + \mathbf{e}_2] - s[\boldsymbol{k}]), \tag{4.3}$$

where $\mathbf{e}_j$ is the j-th unit vector in $\mathbb{R}^2$. The following proposition provides the implementation of $\mathbf{L}^{\mathrm{T}}$.

Proposition 1. *Let* $\mathbf{L} : \mathbb{R}^N \to \mathbb{R}^{N \times 2}$ *be a linear operator defined as* (4.3). *Then, the adjoint operator* $\mathbf{L}^{\mathrm{T}} : \mathbb{R}^{N \times 2} \to \mathbb{R}^N$ *is given by*

$$[\mathbf{L}^{\mathrm{T}} \mathbf{u}]_{\boldsymbol{k}} = u_1[\boldsymbol{k} - \mathbf{e}_1] - u_1[\boldsymbol{k}] + u_2[\boldsymbol{k} - \mathbf{e}_2] - u_2[\boldsymbol{k}], \tag{4.4}$$

where $\mathbf{u}[\boldsymbol{k}] = (u_1[\boldsymbol{k}], u_2[\boldsymbol{k}])$.

---

[4] We note that this choice is motivated by the most prominent method in the field that is TV regularization.

4.3.3 Vectorial Proximal Mappings

In the present formulation, the potential function acts on the ℓ_2-norm of a vector in $\mathbb{R}^2$. Hence, we need to modify the proximal mappings. To do so, let us first remark that the constrained form of the reconstruction problem is given by

$$\mathbf{s}^{\star}_{\mathrm{MAP}} = \arg\min_{\mathbf{s}\in\mathbb{R}^N} \left(\frac{1}{2}\|\mathbf{Hs}-\mathbf{y}\|_2^2 + \tau \sum_{\boldsymbol{k}\in\Omega} \Phi_U\left(\|\mathbf{u}[\boldsymbol{k}]\|_2\right) \right) \text{ s. t. } \mathbf{Ls} = \mathbf{u} \in \mathbb{R}^{N\times 2}. \quad (4.5)$$

By following the developments given in Section 3.6, we arrive at the vectorial counterpart of (3.25a) that is—at iteration t—of the form

$$\min_{\mathbf{u}[\boldsymbol{k}]\in\mathbb{R}^2} \Big(\tau\Phi_U(\|\mathbf{u}[\boldsymbol{k}]\|_2) + \frac{\mu}{2}\|\mathbf{u}[\boldsymbol{k}] - \mathbf{z}^t[\boldsymbol{k}]\|_2^2 \Big),\ \forall \boldsymbol{k}\in\Omega,$$

where $\mathbf{z}^t = \mathbf{Ls}^t + \boldsymbol{\alpha}^t/\mu$ with $\boldsymbol{\alpha}^t \in \mathbb{R}^{N\times 2}$ denoting the Lagrange multiplier. For the case of Gaussian prior, the modification is trivial and the solution is given by

$$\mathrm{prox}_{\|\|_2^2}(\mathbf{z}^t;\tau/\mu) = \mathbf{z}^t(1+2(\tau/\mu))^{-1}.$$

As for the Laplace case, the vectorial extension of the proximal mapping is less obvious. First, we note that one needs to solve the optimization problem

$$\min_{\mathbf{u}[\boldsymbol{k}]\in\mathbb{R}^2} \Big(\tau\|\mathbf{u}[\boldsymbol{k}]\|_2 + \frac{\mu}{2}\|\mathbf{u}[\boldsymbol{k}] - \mathbf{z}^t[\boldsymbol{k}]\|_2^2 \Big),\ \forall \boldsymbol{k}\in\Omega, \quad (4.6)$$

which is nonsmooth. To address it properly, we shall need to introduce some basic notions regularly used in convex optimization [86, 89, 90, 50].

Definition 5. *Let $f:\mathbb{R}^q \to \mathbb{R}$ be a convex function and $\mathbf{v}$ be a vector in $\mathbb{R}^q$. The subdifferential of f at $\mathbf{v}$ is defined as*

$$\boldsymbol{\partial} f(\mathbf{v}) = \{\mathbf{w}\in\mathbb{R}^q \,|\, f(\mathbf{v}) + \mathbf{w}^{\mathrm{T}}(\mathbf{v}'-\mathbf{v}) \le f(\mathbf{v}')\ \forall \mathbf{v}'\in\mathbb{R}^q\}. \quad (4.7)$$

We note that the elements of the set $\boldsymbol{\partial} f(\mathbf{v})$ are called the subgradients of f at $\mathbf{v}$. The following proposition points the importance of studying the concept of subdifferential.

Proposition 2. *For any convex function $f:\mathbb{R}^q\to\mathbb{R}$, a point $\mathbf{v}\in\mathbb{R}^q$ is a global minimum of f iff. $\mathbf{0}\in\boldsymbol{\partial} f(\mathbf{v})$.*

In effect, if the underlying convex function f is smooth at $\mathbf{v}$, the set $\boldsymbol{\partial} f(\mathbf{v})$ includes only the gradient of f at $\mathbf{v}$ (i.e., $\boldsymbol{\partial} f(\mathbf{v}) = \{\nabla f(\mathbf{v})\}$). Therefore, the provided optimality condition $\mathbf{0} \in \boldsymbol{\partial} f(\mathbf{v})$ is equivalent to the classical first-order optimality condition $\nabla f(\mathbf{v}) = \mathbf{0}$.

Lemma 1. *For $\mathbf{v} \in \mathbb{R}^q$, the subdifferential of the $f(\mathbf{v}) = \|\mathbf{v}\|_2$ is*

$$\boldsymbol{\partial} f(\mathbf{v}) = \begin{cases} \mathbf{v}/\|\mathbf{v}\|_2 & \textit{if } \mathbf{v} \neq 0; \\ \{\mathbf{v}' \mid \|\mathbf{v}'\|_2 \leq 1\} & \textit{otherwise.} \end{cases} \tag{4.8}$$

We are now equipped with the necessary tools to formalize the solution. The following proposition provides a closed-form expression for (4.6).

Proposition 3. *Let $\lambda > 0$ and $\mathbf{v}' \in \mathbb{R}^q$. Then, the minimizer of*

$$\min_{\mathbf{v} \in \mathbb{R}^q} \left(\lambda \|\mathbf{v}\|_2 + \frac{1}{2}\|\mathbf{v} - \mathbf{v}'\|_2^2 \right) \tag{4.9}$$

is given by

$$\operatorname{prox}_{\|\cdot\|_2}(\mathbf{v}'; \lambda) = \max\left(\|\mathbf{v}'\|_2 - \lambda, 0\right) \frac{\mathbf{v}'}{\|\mathbf{v}'\|_2}. \tag{4.10}$$

Proof. [5] Using Proposition 2, we see that the minimizer $\mathbf{v}^\star$ must satisfy $\mathbf{0} \in \boldsymbol{\partial} f(\mathbf{v}^\star)$. Considering this with Lemma 1, it writes for $\mathbf{v}^\star$ that

$$\begin{cases} \lambda \mathbf{v}^\star / \|\mathbf{v}^\star\|_2 + (\mathbf{v}^\star - \mathbf{v}') = \mathbf{0} & \text{if } \mathbf{v} \neq 0; \\ \|\mathbf{v}'\|_2 \leq \lambda & \text{otherwise.} \end{cases} \tag{4.11}$$

For the former case, it holds that $\mathbf{v}' = \mathbf{v}^\star + \lambda \frac{\mathbf{v}^\star}{\|\mathbf{v}^\star\|_2}$ and that $\|\mathbf{v}'\|_2 = \|\mathbf{v}^\star\|_2 + \lambda$. Thus, $(\mathbf{v}^\star)/\|\mathbf{v}^\star\|_2 = (\mathbf{v}')/|\mathbf{v}'\|_2$. Thus, we have

$$\mathbf{v}^\star = \|\mathbf{v}^\star\|_2 \frac{\mathbf{v}^\star}{\|\mathbf{v}^\star\|_2} = \big(\|\mathbf{v}' - \lambda\|_2\big)(\mathbf{v}')/|\mathbf{v}'\|_2.$$

As for the latter case, $\mathbf{v}^\star = \mathbf{0}$ iff. $\|\mathbf{v}'\|_2 \leq \lambda$. Combining both results, we arrive at the desired conclusion. □

---

[5] Similar ideas have been presented in [91, 92].

In the light of Proposition 3, the minimizer $\mathbf{u}^{t+1}$ of (4.6) is specified via the vectorial proximal mapping of the Laplace prior that is given by

$$\mathbf{u}^{t+1}[\boldsymbol{k}] = \operatorname{prox}_{\|\|_2}\big(\mathbf{z}^t[\boldsymbol{k}]; \tau/\mu\big) = \max\left\{ \|\mathbf{z}^t[\boldsymbol{k}]\|_2 - \frac{\tau}{\mu}, 0 \right\} \frac{\mathbf{z}^t[\boldsymbol{k}]}{\|\mathbf{z}^t[\boldsymbol{k}]\|_2}. \tag{4.12}$$

Since the minimization is implemented for each sample location $\boldsymbol{k} \in \Omega$ in parallel, the computational cost is $\mathcal{O}(N)$. Finally, note that the problem (4.3.3) for the Student's t-prior is smooth. This allows us to extend our LUT-based approach for the vector case in a straightforward way. Having explained the vectorial proximal mappings, we focus on the last main computational step of our reconstruction method.

4.3.4 Efficient Calculation of Matrix Inverse

In the sequel, we consider solving (3.32). As discussed earlier in Chapter 3, we need to perform a matrix inversion. At iteration t, we are interested in obtaining the solution that is specified as

$$\mathbf{s}^{t+1} = \underbrace{\left(\mathbf{H}^{\mathrm{T}}\mathbf{H} + \mu\mathbf{L}^{\mathrm{T}}\mathbf{L}\right)}_{\mathbf{A}}{}^{-1} \underbrace{\left(\mathbf{H}^{\mathrm{T}}\mathbf{y} + \mu\mathbf{L}^{\mathrm{T}}\left(\mathbf{u}^{t+1} - \frac{\boldsymbol{\alpha}^t}{\mu}\right)\right)}_{\mathbf{b}^t}. \tag{4.13}$$

Prior to analyzing $\mathbf{A}$, we first note that $\mathbf{L}^{\mathrm{T}}\mathbf{L}$ has a fixed structure.

Proposition 4. *Let* $\mathbf{L} : \mathbb{R}^N \to \mathbb{R}^{N\times 2}$ *be a linear operator defined as* (4.3)*. Then, the linear operator* $\mathbf{L}^{\mathrm{T}}\mathbf{L} : \mathbb{R}^N \to \mathbb{R}^N$ *is given by*

$$[\mathbf{L}^{\mathrm{T}}\mathbf{L}\mathbf{s}]_{\boldsymbol{k}} = 4s[\boldsymbol{k}] - s[\boldsymbol{k} - \mathbf{e}_1] - s[\boldsymbol{k} + \mathbf{e}_1] - s[\boldsymbol{k} - \mathbf{e}_2] - s[\boldsymbol{k} + \mathbf{e}_2]. \tag{4.14}$$

One essentially sees that $\mathbf{L}^{\mathrm{T}}\mathbf{L}$ represents nothing but the circular convolution with a Laplacian kernel (up to a sign change). Hence, it is a BCCB operator. This implies that the inversion of the matrix $\mathbf{A}$ depends on the structure of $\mathbf{H}^{\mathrm{T}}\mathbf{H}$. If $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ is also BCCB, then $\mathbf{A}$ admits the spectral decomposition given in Theorem 4 where the eigenvalues of $\mathbf{A}$ are obtained as the sum of the eigenvalues of $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ and $\mathbf{L}^{\mathrm{T}}\mathbf{L}$. This enables one to obtain $\mathbf{s}^{t+1} = \mathbf{A}^{-1}\mathbf{b}^t$ directly as explained in Algorithm 4. In our numerical experiments, we shall illustrate examples for both cases.

In summary, the main computational complexity of the reconstruction is driven by the structure (and the implementation) of $\mathbf{H}^{\mathrm{T}}\mathbf{H}$. As a concluding remark, we note that the general outline of our image reconstruction framework is described in Algorithm 5.

Algorithm 5: Generic MAP reconstruction of images according to (4.5)

```
input : Noisy measurements $\mathbf{y}$, initial solution $\mathbf{s}^0$,
        regularization parameter $\tau$, penalty parameter $\mu$,
        and implementation of proximalMapping()
output: Reconstructed image $\mathbf{s}_{\mathrm{MAP}}$

if $\mathbf{H}$ is a BCCB operator then
    $\widehat{\mathbf{E}}_{\mathbf{H}} \leftarrow$ computeEigenvalues($\mathbf{H}$);
    $\widehat{\mathbf{E}}_{\mathbf{L}^{\mathrm{T}}\mathbf{L}} \leftarrow$ computeEigenvalues($\mathbf{L}^{\mathrm{T}}\mathbf{L}$);
    $\widehat{\mathbf{E}}_{\mathbf{A}} \leftarrow \widehat{\mathbf{E}}_{\mathbf{H}} \circ \widehat{\mathbf{E}}^*_{\mathbf{H}} + \mu\widehat{\mathbf{E}}_{\mathbf{L}^{\mathrm{T}}\mathbf{L}}$;
else if $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ is a BCCB operator then
    $\widehat{\mathbf{E}}_{\mathbf{H}^{\mathrm{T}}\mathbf{H}} \leftarrow$ computeEigenvalues($\mathbf{H}^{\mathrm{T}}\mathbf{H}$);
    $\widehat{\mathbf{E}}_{\mathbf{L}^{\mathrm{T}}\mathbf{L}} \leftarrow$ computeEigenvalues($\mathbf{L}^{\mathrm{T}}\mathbf{L}$);
    $\widehat{\mathbf{E}}_{\mathbf{A}} \leftarrow \widehat{\mathbf{E}}_{\mathbf{H}^{\mathrm{T}}\mathbf{H}} + \mu\widehat{\mathbf{E}}_{\mathbf{L}^{\mathrm{T}}\mathbf{L}}$;
else
    $A(\cdot) \leftarrow$ constructOperator($\mathbf{H}^{\mathrm{T}}\mathbf{H} + \mu\mathbf{L}^{\mathrm{T}}\mathbf{L}$); (see Note 1)
end if
$t \leftarrow 0$; $\mathbf{u}^t \leftarrow \mathbf{0}$; $\boldsymbol{\alpha}^t \leftarrow \mathbf{0}$; $\mathbf{b} \leftarrow \mathbf{H}^{\mathrm{T}}\mathbf{y}$;
repeat
    $\mathbf{z}^t \leftarrow \mathbf{L}\mathbf{s}^t + \boldsymbol{\alpha}^t/\mu$;
    $\mathbf{u}^{t+1} \leftarrow$ proximalMapping($\mathbf{z}^t; \tau/\mu$);
    $\mathbf{b}^t \leftarrow \mathbf{b} + \mu\mathbf{L}^{\mathrm{T}}\left(\mathbf{u}^{t+1} - \boldsymbol{\alpha}^t/\mu\right)$;
        if $\widehat{\mathbf{E}}_{\mathbf{A}}$ exists then
            $\mathbf{s}^{t+1} \leftarrow \mathbf{A}^{-1}\mathbf{b}^t$; (use Algorithm 4)
        else
            $\mathbf{s}^{t+1} \leftarrow$ conjugateGradient($A(\cdot)$,$\mathbf{b}^t$,$\mathbf{s}^t$); (see Note 2)
        end if
    $\boldsymbol{\alpha}^{t+1} \leftarrow \boldsymbol{\alpha}^t + \mu(\mathbf{L}\mathbf{s}^{t+1} - \mathbf{u}^{t+1})$;
    $t \leftarrow t + 1$;
until stopping criteria
return $\mathbf{s}^t$;
```

Note 1: `constructOperator()` returns a function that implements the action of the argument (i.e. A:$\mathbf{s} \mapsto (\mathbf{H}^{\mathrm{T}}\mathbf{H} + \mu\mathbf{L}^{\mathrm{T}}\mathbf{L})\mathbf{s}$).
Note 2: `conjugateGradient()` implements the conjugate gradient iterations to compute $A^{-1}(\mathbf{b}^t)$ where $\mathbf{s}^t$ is the initial solution.

4.4 Numerical Results

In this section, we apply our MAP reconstruction framework to a set of imaging problems. For each instance of reconstruction, the regularization parameter τ is optimized via an oracle to obtain the highest-possible SNR. The penalty parameter μ is set to 10τ for convex problems and $5 \times 10^3\tau$ for the nonconvex one. These heuristics are found to provide a good convergence speed. The reconstructions are initialized in a systematic fashion: The solution of the Gaussian estimator is used as the initial solution for the Laplace estimator whose result is used as the initial solution for Student's t-estimator.

4.4.1 Image Deconvolution

The first problem we consider is the deconvolution of fluorescence micrographs (i.e., spatial distribution of florescent labels). For simplicity, we consider that the fluorescence emitters are strictly located on a single spatial plane only. This means that the imaging is assumed to be two dimensional. For deconvolution, the measurement function in (3.8) is assumed to be shift-invariant. Therefore, it corresponds to the shifted version of the point-spread function (PSF) of the microscope on the sampling grid.

For $\mathbf{m} = (m_1, m_2)$, we write that

$$\psi_{\mathbf{m}}^{\mathrm{decon}}(\boldsymbol{x}) = \psi^{\mathrm{decon}}(\boldsymbol{x} - \mathbf{m}),$$

where ψ^{decon} represents the PSF. We discretize the model by choosing

$$\varphi_{\mathrm{int}}(\boldsymbol{x}) = \mathrm{sinc}(x_1)\mathrm{sinc}(x_2)$$

with $\varphi_{\boldsymbol{k}}(\boldsymbol{x}) = \varphi_{\mathrm{int}}(\boldsymbol{x} - \boldsymbol{k})$ for $\boldsymbol{k} \in \Omega$. The entries of the resulting system matrix $\mathbf{H}$ are given by

$$\begin{aligned}[\mathbf{H}]_{\mathbf{m},\boldsymbol{k}} &= \langle \psi^{\mathrm{decon}}(\cdot - \mathbf{m}), \mathrm{sinc}(\cdot - \boldsymbol{k}) \rangle \\ &= \left(\psi^{\mathrm{decon}} * \mathrm{sinc}\right)(\mathbf{m} - \boldsymbol{k}). \qquad (4.15)\end{aligned}$$

In effect, (4.15) corresponds to the samples of the band-limited version of the PSF. The current discretization framework implies that $\mathbf{H}$ is BCCB that represents a discrete convolution with the samples of the band-limited PSF. Remark that $\mathbf{H}^{\mathrm{T}}$ is also diagonalized by the discrete Fourier transform. Its eigenvalues are the complex conjugates of $\mathbf{\Lambda_H}$. In the present scenario, we simply write that

$$\begin{aligned}\mathbf{H}^{\mathrm{T}}\mathbf{H} &= \mathbf{F}^{-1}\mathbf{\Lambda}^{*}_{\mathbf{H}}\underbrace{\mathbf{F}\mathbf{F}^{-1}}_{\mathbf{I}}\mathbf{\Lambda_H}\mathbf{F}\\ &= \mathbf{F}^{-1}|\mathbf{\Lambda_H}|^2\mathbf{F},\end{aligned}$$

which means that the eigenvalues of $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ are obtained as the squared modulus of the ones of $\mathbf{H}$. Since $\mathbf{A} = (\mathbf{H}^{\mathrm{T}}\mathbf{H}+\mu\mathbf{L}^{\mathrm{T}}\mathbf{L})$ is BCCB, the direct solution of the matrix inversion in (4.13) is thus obtained at the cost of two FFTs. This makes our ADMM-based reconstruction particularly well-suited for the deconvolution problem.

Reconstruction Results

We perform controlled experiments, where the blurring of the microscope is simulated by a Gaussian PSF kernel of support (9×9) and standard deviation $\sigma_{\mathrm{b}} = 4$, on three microscopic images of size (512×512) that are displayed in Figure 4.1. In Figure 4.1(a), we show stem cells surrounded by numerous goblet cells. In Figure 4.1(b), we illustrate nerve cells growing along fibers, and we show in Figure 4.1(c) bovine pulmonary artery cells.

For deconvolution, the algorithm is run for a maximum of 500 iterations, or until the relative error between the successive iterates is less than 5×10^{-6}. The measurements are degraded with different levels of AWGN. The noise variance σ^2 is determined through blurred SNR (BSNR) that is defined as $\mathrm{BSNR} = \mathrm{var}(\mathbf{Hs})/\sigma^2$.

We conclude from the results of Table 4.1 that the MAP estimator based on a Laplace prior yields the best performance for images having sharp edges with a moderate amount of texture, such as those in Figures 4.1(b)-4.1(c). This confirms the observation that, by promoting solutions with sparse gradient, it is possible to improve the deconvolution performance. However, enforcing sparsity too heavily, as is the case for Student's t-priors, results in a degradation of the deconvolution performance for the biological images considered. Finally, for a heavily textured image like the one found in Figure 4.1(a), the linear image deconvolution algorithm (Gaussian prior based MAP estimator) yields the best performance. We note that the derived algorithms are compatible with the methods commonly used in the field (e.g., Tikhonov regularization [93] and TV regularization [94]).

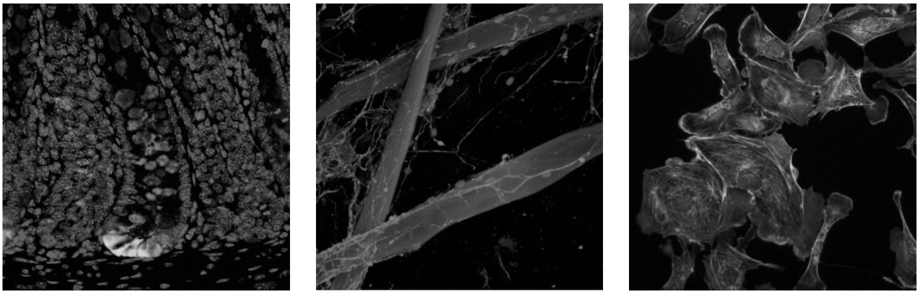

(a) (b) (c)

Figure 4.1: Images used in deconvolution experiments: (a) stem cells surrounded by goblet cells; (b) nerve cells growing around fibers; (c) artery cells.

Table 4.1: Deconvolution performance of MAP estimators based on different prior distributions.

		Estimation Performance (SNR in dB)		
	BSNR (dB)	Gaussian	Laplace	Student's
Stem cells	20	**14.43**	13.76	11.86
Stem cells	30	**15.92**	15.77	13.15
Stem cells	40	**18.11**	**18.11**	13.83
Nerve cells	20	13.86	**15.31**	14.01
Nerve cells	30	15.89	**18.18**	15.81
Nerve cells	40	18.58	**20.57**	16.92
Artery cells	20	14.86	**15.23**	13.48
Artery cells	30	16.59	**17.21**	14.92
Artery cells	40	18.68	**19.61**	15.94

4.4.2 MRI Reconstruction

We now consider the problem of reconstructing MR images (i.e., spatial distribution of spin densities) from undersampled spatial-frequency domain trajectories. The measurement function represents a complex exponential at some fixed frequencies and is given by

$$\psi_m^{\mathrm{mri}}(\boldsymbol{x}) = \mathrm{e}^{-2\pi \mathrm{j}\langle \boldsymbol{\omega}_m, \boldsymbol{x}\rangle},$$

where $\boldsymbol{\omega}_m$ represents the sample point in spatial-frequency space. For convenience, we use the same sinc basis for discretization as in Section 4.4.1. This results in a system matrix with the entries

$$\begin{aligned}[\mathbf{H}]_{m,\mathbf{n}} &= \langle \psi_m^{\mathrm{mri}}, \mathrm{sinc}(\cdot - \mathbf{n})\rangle \\ &= \langle \mathrm{e}^{-2\pi \mathrm{j}\langle \boldsymbol{\omega}_m, \cdot\rangle}, \mathrm{sinc}(\cdot - \mathbf{n})\rangle \\ &= \mathrm{e}^{-2\pi \mathrm{j}\langle \boldsymbol{\omega}_m, \mathbf{n}\rangle}, \end{aligned} \tag{4.16}$$

for $\|\boldsymbol{\omega}_m\|_\infty \leq \frac{1}{2}$. The effect of choosing a sinc function is that the system matrix reduces to the discrete version of complex Fourier exponentials. We assume that ω_m are located on the Cartesian grid. It is noteworthy that the system matrix is not a square matrix but it can be factorized as

$$\mathbf{H} = \mathbf{MF} \in \mathbb{C}^{M \times N}$$

where $\mathbf{M} \in \mathbb{R}^{M \times N}$, with $M < N$, is a binary matrix that is constructed by the suitable subset of rows of the identity matrix. Even though, $\mathbf{H}$ is not BCCB in this case, let us remark that $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ is still BCCB since

$$\mathbf{H}^{\mathrm{T}}\mathbf{H} = \mathbf{F}^{-1}\mathbf{M}^{\mathrm{T}}\mathbf{MF},$$

where the eigenvalues $\boldsymbol{\Lambda}_{\mathbf{H}^{\mathrm{T}}\mathbf{H}} = \mathbf{M}^{\mathrm{T}}\mathbf{M}$ is a diagonal matrix with the entries at each (m, m) being one (remaining diagonal entries are zero). The good news is that the matrix inversion in (4.13) (once again) can be directly performed.

Reconstruction Results

We study the reconstruction of the two MR images of size (256 × 256) illustrated Figure 4.2—a cross-section of a wrist is displayed in the first image, followed by an

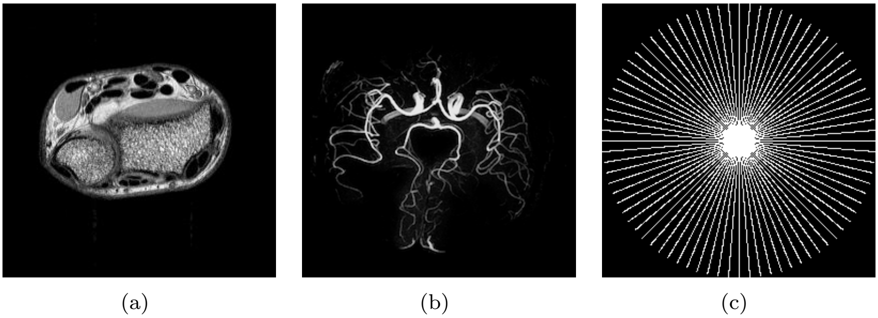


Figure 4.2: Images used in MR reconstruction experiments: (a) cross section of a wrist; (b) angiography image; (c) spatial-frequency space sampling pattern along 40 radial lines.

Table 4.2: MR image reconstruction performance of MAP estimators based on different prior distributions.

	Estimation Performance (SNR in dB)		
	Gaussian	Laplace	Student's
Wrist (20 radial lines)	8.82	**11.8**	5.97
Wrist (40 radial lines)	11.30	**14.69**	13.81
Angiogram (20 radial lines)	4.30	9.01	**9.40**
Angiogram (40 radial lines)	6.31	14.48	**14.97**

MR angiography image—and consider a radial sampling spatial-frequency sampling pattern (see Figure 4.2(c)).

The reconstruction algorithm is run with the stopping criteria set as in Section 4.4.1. We show in Table 4.2 the reconstruction performance of the reconstruction algorithms in question for different number of radial lines.

On one hand, we observe that the MAP estimators based on Laplace priors yield the best solution in the case of the wrist image, which has sharp edges and some amount of texture. Meanwhile, the reconstructions using Student's t-priors are suboptimal because they over-encourage sparsity. This is similar to what has been observed with the microscopic images in Section4.4.1. On the other hand, Student's t-priors are quite suitable for reconstructing the angiogram image, which is mostly composed of piecewise-smooth components. Finally, we observe that the performance of Gaussian estimators is not competitive for the images considered. Our reconstruction algorithms are tightly linked with the deterministic approaches used for MRI reconstruction including TV [95] and its nonconvex counterparts [96], which are in line with Laplace and Student's t- priors, respectively.

4.4.3 X-Ray Tomographic Reconstruction

X-ray computed tomography (CT) aims at reconstructing an object from its projections taken along different directions. The mathematical model of a conventional CT is based on the Radon transform

$$\begin{aligned} g_{\theta_m}(t_m) &= \mathcal{R}_{\theta_m}\{s(\boldsymbol{x})\}(t_m) \\ &= \int_{\mathbb{R}^2} s(\boldsymbol{x})\delta(t_m - \langle \boldsymbol{x}, \boldsymbol{\theta_m}\rangle)\mathrm{d}\boldsymbol{x}\,, \end{aligned}$$

where $s(\boldsymbol{x})$ is the absorption coefficient distribution of the underlying object, t_m is the sampling point and $\boldsymbol{\theta}_m = (\cos(\theta_m), \sin(\theta_m))$ is the angular parameter. Therefore, the measurement function

$$\psi_m^{\mathrm{xray}}(\boldsymbol{x}) = \delta\left(t_m - \langle \boldsymbol{x}, \boldsymbol{\theta}_m\rangle\right)$$

denotes an idealized line in $\mathbb{R}^2$ perpendicular to $\boldsymbol{\theta}_m$.

In our formulation, we represent the absorption distribution in the space spanned by the tensor product of two B-splines

$$s(\mathbf{x}) = \sum_{\mathbf{k}} s[\mathbf{k}]\varphi_{\mathrm{int}}(\boldsymbol{x} - \boldsymbol{k})\,,$$

where

$$\varphi_{\mathrm{int}}(\boldsymbol{x}) = \mathrm{tri}(x_1)\mathrm{tri}(x_2)$$

with $\mathrm{tri}(x) = (1 - |x|)$ denoting the linear B-spline function.

The entries of the system matrix are then determined explicitly using the B-spline calculus described in [97], which leads to

$$\begin{aligned}[\mathbf{H}]_{m,\boldsymbol{k}} &= \langle \delta(t_m - \langle \boldsymbol{x}, \boldsymbol{\theta}_m \rangle), \varphi_{\text{int}}(\boldsymbol{x} - \boldsymbol{k}) \rangle \\ &= \frac{\triangle^2_{|\cos\theta_m|} \triangle^2_{|\sin\theta_m|}}{3!} (t_m - \langle \boldsymbol{k}, \boldsymbol{\theta}_m \rangle)^3_+,\end{aligned}$$

where

$$\triangle_{t_0} f(t) = \frac{f(t) - f(t - t_0)}{h}$$

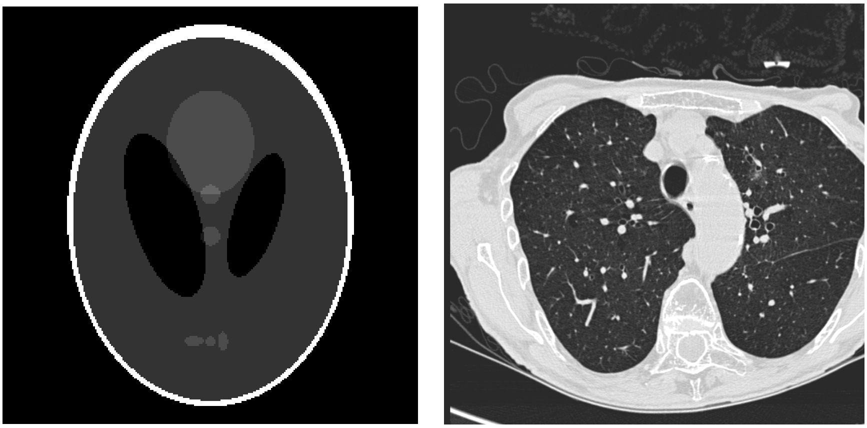
is the finite-difference operator, $\triangle^n_h f(t)$ is its n-fold iteration, and $t_+ = \max(0, t)$. This approach provides an accurate modeling, as demonstrated in [97]. In the current case, $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ is not a BCCB operator so that one has to use an iterative solver to compute $\mathbf{A}^{-1} = (\mathbf{H}^{\mathrm{T}}\mathbf{H} + \mu \mathbf{L}^{\mathrm{T}}\mathbf{L})^{-1}$ in (4.13). Particularly, we use CG method. As noted previously, the algorithm requires one to be able to compute the multiplication of $\mathbf{A}$ with a vector (i.e., the series of image processing operations to compute the output image). Since the operator $\mathbf{L}^{\mathrm{T}}\mathbf{L}$ is performed via FFTs, the remaining task is to implement $\mathbf{H}^{\mathrm{T}}\mathbf{H}$. To do so, one can take advantage of multi-threading and look-up-table techniques as explained in [98].

Reconstruction Results

We consider the two images shown in Figure 4.3. The Shepp-Logan (SL) phantom has size (256 × 256), while the cross section of the lung has size (750 × 750). In the simulations of the forward model, the Radon transform is computed along 180 and 360 directions for the lung image and along 120 and 180 directions for the SL phantom. The measurements are degraded with the Gaussian noise such that the signal-to-noise ratio is 20 dB.

We solve the quadratic minimization problem (3.32) iteratively by using 50 CG iterations. The reconstruction results are reported in Table 4.3.

Before commenting on the results, let us remark that the SL phantom is a piecewise-smooth image with sparse gradient. Therefore, we observe that the imposition of more sparsity brought by Student's t-priors significantly improves the reconstruction quality for this particular image. However, we find that the Gaussian priors for the lung image outperform the other priors. Like the deconvolution and MRI problems, our algorithms are in line with Tikhonov-type [99] and TV [100] reconstructions used for X-ray CT.



(a) (b)

Figure 4.3: Images used in X-ray tomographic reconstruction experiments: (a) the Shepp-Logan (SL) phantom; (b) cross section of the lung.

Table 4.3: Reconstruction results of X-ray computed tomography using different estimators.

	Estimation Performance (SNR in dB)		
	Gaussian	Laplace	Student's
SL Phantom (120 direction)	16.8	17.53	**18.76**
SL Phantom (180 direction)	18.13	18.75	**20.34**
Lung (180 direction)	**22.49**	21.52	21.45
Lung (360 direction)	**24.38**	22.47	22.37

4.5 Discussion

By performing numerical experiments on different types of imaging modalities, we have seen that sparsity-promoting reconstructions are effective methods for solving biomedical imaging problems. Our simulations point the importance of sparse models and demonstrate that ℓ_1-type reconstructions are favorable in general. This case is represented in our family of MAP estimators via Laplace priors. However, encouraging sparser solutions (by using Student's t-priors) improves the reconstruction performance only for a very specific class of images. We have also seen that the classical non-sparse solutions (provided by Gaussian priors) can still yield better reconstructions for images of high texture content, for instance. Therefore, the efficiency of a potential function is primarily dependent upon the type of image being considered. In our model, this is related to the Lévy exponent of the underlying continuous-domain innovation process w which is in direct relationship with the discrete signal prior.

The exact equivalence between our models and the TV regularization is fully justified in the one-dimensional setting as it has been pointed out in Subsection 3.5.2. However, due to images being multi-dimensional signals, the discrete gradient operator $\mathbf{L}$ is a vector-valued mapping which causes mathematical complications to exactly fit it in the theory. As mentioned earlier in Section 3.4.1, the class of fractional Laplacian operators $(-\Delta)^{\gamma/2}$, with $\gamma > 0$, can be chosen as the whitening operator. These operators allow us to generate self-similar stochastic processes that are used for modeling of fractal-type images [70, 71, 72]. These isotropic differential operators defined in Fourier domain by

$$\mathcal{F}\left\{(-\Delta)^{\gamma/2}g\right\}(\boldsymbol{\omega}) = \|\boldsymbol{\omega}\|_2^{\gamma}\mathcal{F}\{g\}(\boldsymbol{\omega}),$$

where g has a well-defined Fourier transform. The fractional Laplacian is a linear, self-adjoint, and continuous operator with translation-, rotation-, and scaling-invariance properties. Specifically, for g having a well-defined gradient ∇g, the following relation is established:

$$\begin{aligned}\|\nabla g\|_2^2 &= \|\boldsymbol{\omega}\|_2^2\|\mathcal{F}\{g\}\|_2^2 \\ &= \|(-\Delta)^{1/2}g\|_2^2, \end{aligned} \tag{4.17}$$

where the first equality follows from squared L_2-norms being homogeneous of order 1 and the Parseval relation. The algorithmic implication of (4.17) is valuable:

Under the decoupling simplification, the quadratic regularizer used in the experiments provides the computational means to perform the MAP/MMSE estimation of a self-similar process with the underlying continuous-domain innovation being Gaussian. Despite all these enticing mathematical properties, fractional Laplacian-based regularizations (both ℓ_2- and ℓ_1-types) either do not overperform [18, 101], or provide improvement for a narrow class of images [87] in comparison to their gradient-based counterparts. In any manner, the implicit assumption in our formulation that the gradient of the image is independent from one pixel location to the next is quite likely to be an oversimplification [102].

Theoretical sophistication and exactness are thus abandoned for practical simplicity and insight in anticipation of designing algorithms that are applicable to real data. By limiting our attention to a particular operator, the essential results and constructive ideas are conveyed more intuitively. These points will be more apparent when we propose our optical phase imaging algorithm for unstained cell samples in Chapter 5 and properly extend TV regularization for vector fields in Chapter 6.

Chapter 5

Variational Phase Imaging

5.1 Overview

In this chapter,[1] we introduce a variational phase retrieval algorithm for imaging transparent objects that are optically-thin. Our approach is based on the so-called transport-of-intensity equation (TIE). The model puts forward a differential equation that relates the phase of an optical field to the variation of its intensity along the direction of propagation. In practical terms, TIE allows one to use a conventional light microscope and record a set of defocus images to obtain the phase map of the sample that is being imaged. Accordingly, we first investigate the effect of the defocus distance on the retrieved phase image. The key outcome of our analysis is that the chosen defocus value has direct influence on the range of spatial frequencies of the reconstructed phase image. Based on this observation, we propose a phase reconstruction algorithm that is nonlinear and utilizes weighted norms. The method combines different ranges of spatial frequencies—depending on the defocus value of the measurements—in a regularized fashion. Our simulations outperform commonly used linear and nonlinear TIE solvers. We also illustrate and validate our method on real microscopy data of HeLa cells. An important practical outcome of the proposed method is that the reconstructed phase images are greatly in line with those produced by dedicated phase microscopes.

---

[1] This chapter is based on our paper [103].

5.2 Introduction

The generic problem of imaging transparent objects is highly relevant to biological research as most cells and thin tissue samples do not absorb light and produce images with very low contrast when observed under a standard bright-field microscope (see Figure 5.1(a)) [104]. To reveal specific structures in the sample, one can apply staining or use fluorescent dyes and biomarkers. These exogenous contrast agents then allow for the use of advanced light microscopy techniques that include the super resolution ones such as stimulated-emission depletion microscopy (STED) [105] and single-molecule localization microscopy (SMLM) [106, 107]. The high-end modalities, however, require careful sample preparation and are not well-suited to image live cells (especially over extended periods of time) since the contrast agents can be phototoxic [108]. Also, the observation of the global tissue or cell morphology becomes harder [109]. For these reasons, label-free phase imaging techniques are needed especially when minimal manipulation of the cell is required (such as in stem cell and drug discovery studies) [110].

While transparent objects have insignificant absorption, they do introduce phase shifts—on the incident light field—due to variations in the optical path length in the sample. Knowing that specimens change the phase of the light wave, the information about the distribution of the refractive index (hence, about the global morphology and the structure of the specimen) is encoded in the phase. Unfortunately, the optical phase shifts are lost during the acquisition since the detectors are sensitive only to the intensity of the field. Two conventional imaging modalities that translate the phase shifts into detectable intensity differences are phase contrast microscopy (PC) [111] and differential interference contrast (DIC) microscopy [112]. However, both DIC and PC images are challenging to segment (and to track) due to shade and halo artifacts (see Figure 5.1). This essentially limits their use since segmentation is a key step in cell biology [113]. Another well-established technique for phase imaging is digital holography microscopy (DHM) [114, 20]. Applying holographic imaging principles to microscopy imaging, DHM images yield quantitative phase information, which distinguishes them from DIC and PC images. Basically, DHM allows one to obtain the actual shape map of the transparent samples. The apparent advantage is that the upcoming task of segmenting cells (and other specific structures of interest) is more convenient. The downside of DHM is that it is a hardware-based solution of higher cost compared to bright-field, DIC, and PC microscopies [115].

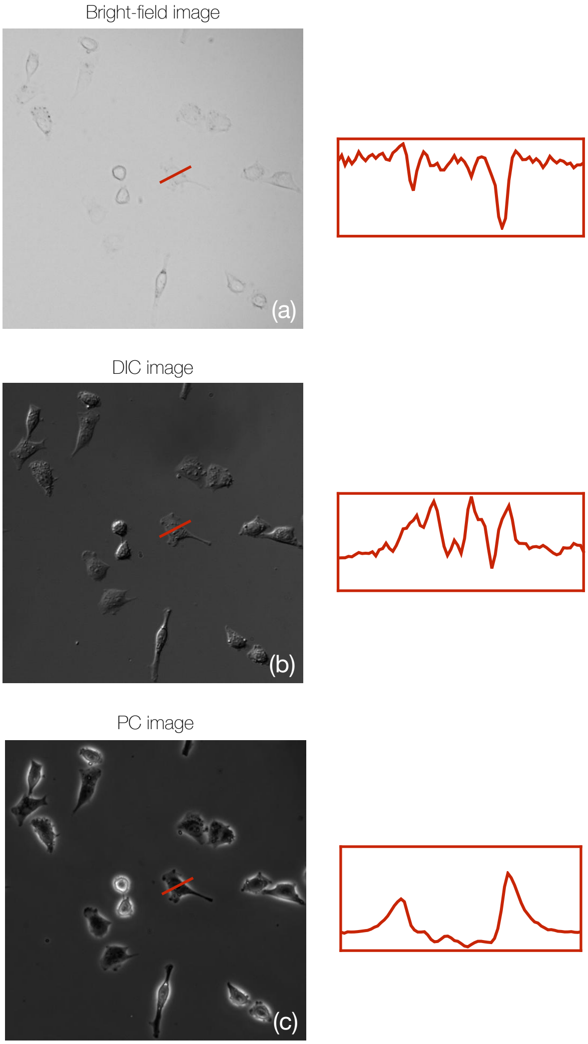


Figure 5.1: Imaging the same HeLa cells with different microscopes. Infocus images for the same field of view are obtained by using (a) bright-field, (b) DIC, and (c) PC microscopes. Normalized intensity profiles corresponding to the same line segment are given at the right hand side of each image.

In this chapter, we consider a low-cost label-free phase imaging approach that uses the transport-of-intensity equation (TIE) [11]. This method has significant advantages: it is a computational method that can be used with a basic bright-field [116] microscope; the resulting phase dispenses with the unwrapping task required by interferometric methods such as DHM [117]. As we shall see later, TIE provides us with a linear mathematical formalism that relates the spatial phase map of the sample to the derivative of its intensity map along the propagation direction. In a practical setup, the phase is recovered by using a set of images acquired at various positions along the optical axis. Hence, TIE simply requires a stack of defocus measurements to obtain the phase information at the infocus plane (i.e., the axial location of the sample). Allowing phase images to be obtained by using numerical methods—rather than implementing specialised hardware modifications—TIE-based imaging is a viable tool for electron microscopy [118] and X-ray imaging [119]. In addition, TIE can be applied to a partially coherent source, even though it has been initially derived for coherent illumination [116]. This makes it applicable to a DIC microscope [120].

TIE has been extensively studied in the optics community, where researchers have explored ways to obtain better estimates of the axial derivative of the intensity. The simplest approach has been to capture two images that are slightly above and below the focus plane, followed by a centered finite difference. Various refined models using additional defocused images have also been proposed. In [121], the authors consider computing higher-order terms in the Taylor development of the intensity along the optical axis. The use of finite differences results in a simple linear combination of defocused images, but the estimate tends to be sensitive to noise. An alternative approach is based on a pixel-wise polynomial fitting of the intensity along the optical axis [121]. Again, it is carried out at the expense of acquiring more images. The authors of [122] have presented a framework in which they estimate the axial derivative of the intensity through fitting in the spatial frequency domain. In [123], the estimation by polynomial fitting is generalized via the Savitzky-Golay differentiation filter.

In addition to such approaches, one may combine multiple TIE solutions. In [124], different TIE reconstructions for distinct defocus distances are fused by using linear filtering operations by investigating the the validity of first order finite difference approximations for a given noise level. The authors have designed improved filters by considering the pointwise variance of the phase in the frequency domain in [125]. The designed filters are found to be a function of the noise variance so that the applicability of the method has been limited. The method of [124] has been extended in [126] by using combining various TIE solutions in the spatial frequency domain. What is common in all of these methods is that they are all linear reconstruction algorithms and that the TIE is solved directly by inverting the model. The latter suggests the use of regularization to stabilize the solutions. In this regard, Tikhonov-type [127] regularizations have been considered.

Rather recently, TIE has also drawn attention of the image processing community. TV and its nonconvex variants, which agree with MAP reconstructions based on Student's t-priors, have been applied to simulated data in [128]. In [129], the authors have showed the applicability of TV regularization for real data reconstructions.

5.2.1 Contributions

On a fundamental level, TIE suffers from two problems. First, when the model is inverted, the noise in the measurements is integrated because TIE is a differential form. This typically produces low-frequency errors (such as "cloudy" artifacts) in the reconstructions. Second, the computation of the axial intensity derivative by using centered finite differences necessitates small defocus distances. Deviations from this requirement introduce further errors. As we shall show in the coming sections, these problems are tightly related. Extending our previous research [129], we propose a nonlinear variational method that provides a unifying framework for the resolution of TIE. Our main contributions are as follows:

- A joint model that is able to combine the spectral information coming from different defocus distances within a regularized reconstruction framework. The model is built upon identifying reliable frequency ranges as the defocus distance changes.

- The proposal of an iterative algorithm that is based on the alternating minimization concepts. The method is highly modular and takes full advantage of the underlying structure of TIE.

- A detailed comparison of our method with common TIE reconstruction algorithms. We show that we achieve better phase reconstructions for both simulated and experimental data. The reconstructed phase images are also validated by comparing them with DHM acquisitions.

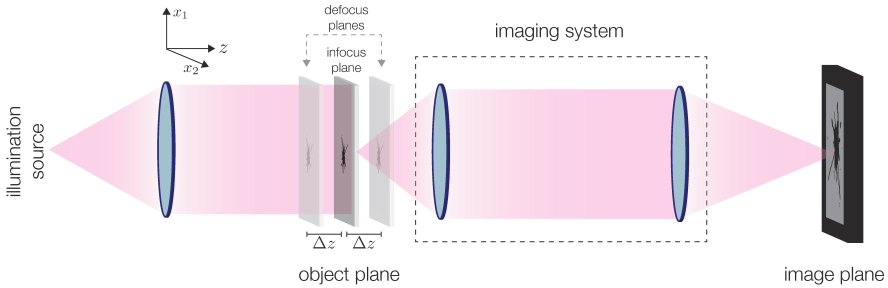


Figure 5.2: Representation of the measurement model (and the corresponding problem geometry) in this chapter. A transparent object is located at the input plane of an ideal imaging system. Note that a magnified version of the object field is generated at the image plane. Defocus images are captured by moving the object symmetrically around the focus position. The phase map at the infocus plane is then obtained by using the defocus images in (5.10).

5.2.2 Outline

The chapter is organized as follows: In Section 5.3, we first explain the underlying physical and mathematical frameworks of TIE that constitute the core of our approach. We shall then formulate a complementary approach known as the contrast transfer function. By considering the two models together, we establish a relationship between the defocus distance and the range of the spatial frequencies (of a transparent object) for which the reconstruction is potentially of high quality. Following this observation and our considerations in Chapter 4, we cast the phase reconstruction task as an inverse problem and propose an iterative algorithm for its resolution in Section 5.4. Finally, we provide numerical simulations, real data experiments, and discuss our results in Section 5.5.

5.3 Physical Models

Consider the problem geometry given in Figure 5.2. We define $(\boldsymbol{x}, z)$ as our coordinate vector where $\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2$ denotes the spatial location on a transverse plane that is perpendicular to the optical axis z. The monochromatic plane wave $\mathrm{e}^{\mathrm{j}kz}$ illuminates a *thin object* that lies on a bounded domain in $\mathbb{R}^2$. The wave function after traversing the object is written as

$$U(\boldsymbol{x}, z) = U_A(\boldsymbol{x}, z)\mathrm{e}^{\mathrm{j}kz}, \tag{5.1}$$

where U_A represents the so-called *complex amplitude* of U, $k = 2\pi/\nu$ is the wave number with ν being the illumination wavelength. The expression of the complex amplitude is given by

$$U_A(\boldsymbol{x}, z) = \sqrt{I(\boldsymbol{x}, z)}\mathrm{e}^{\mathrm{j}\phi(\boldsymbol{x}, z)}, \tag{5.2}$$

where the real-valued functions I and ϕ are the intensity and the phase, respectively.

For convenience, we shall assume that the object is located at the axial position $z = 0$. Therefore, the wave field at the object plane is specified by

$$O(\boldsymbol{x}) = U_A(\boldsymbol{x}, 0) = \sqrt{I_0(\boldsymbol{x})}\mathrm{e}^{\mathrm{j}\phi_0(\boldsymbol{x})}, \tag{5.3}$$

where $\phi_0(\boldsymbol{x}) = \phi(\boldsymbol{x}, 0)$ corresponds to the spatial phase map of the object. In practice, one is able to record the intensity map of a light field (specifically, its average over a certain time) and the phase information is lost (as a result of the fast oscillations at visible frequencies). This necessitates establishing a relationship between the phase and the intensity maps of U_A where only the latter is measured.

5.3.1 Transport-of-Intensity Equation

Suppose the propagation of U_A is dominant along the z axis (i.e., the paraxial approximation is valid). Then, the physics of U_A is governed by the paraxial wave equation

$$\left(\nabla_\perp^2 + 2\mathrm{j}k\frac{\partial}{\partial z}\right) U_A(\boldsymbol{x}, z) = 0, \tag{5.4}$$

where $\nabla_\perp^2$ is the transverse Laplacian operator defined by $\nabla_\perp^2 = \left(\partial^2/\partial x_1^2 + \partial^2/\partial x_2^2\right)$.

By multiplying (5.4) by U_A^* on the left-hand side, and separating the real and the imaginary parts, Teague [11] has derived two equations. In particular, the imaginary part[2] specifies the *transport-of-intensity equation* (TIE)

$$-k\frac{\partial}{\partial z}I(\boldsymbol{x},z) = \nabla_\perp \cdot \left(I(\boldsymbol{x},z)\nabla_\perp \phi(\boldsymbol{x},z)\right), \tag{5.5}$$

where $\nabla_\perp = [\partial/\partial x_1 \;\; \partial/\partial x_2]^{\mathrm{T}}$ is the transverse gradient operator and $\cdot$ denotes the dot product. We see that (5.5) is an elliptical second-order partial differential equation that links the phase information to the axial derivative of the intensity of the field. Practically, TIE bears an important outcome: the spatial phase map ϕ is computed by measuring the intensity I and its axial derivative $\partial I(\boldsymbol{x},z)/\partial z$, where the latter can be approximated by finite differences.

We now investigate TIE in relation to our problem settings. We consider an object that is at the input plane of an ideal magnification system, generating a dilated version of the object field at the image plane. As mentioned in Section 5.2, we are interested in imaging unstained biological samples. Such class of objects are modeled as *phase-only* objects, meaning that they do not significantly absorb or scatter the illuminating field [109]. We further assume that the illumination is uniform so that $I_0(\boldsymbol{x})$ is constant: $I_0(\boldsymbol{x}) = J_0$. Therefore, (5.5) is rewritten as

$$-\frac{k}{J_0}\frac{\partial}{\partial z}I(\boldsymbol{x},z)\bigg|_{z=0} = \nabla_\perp^2 \phi_0(\boldsymbol{x}). \tag{5.6}$$

To simplify our notation, let us define

$$y(\boldsymbol{x}) = \frac{I(\boldsymbol{x},\Delta z) - I(\boldsymbol{x},-\Delta z)}{J_0} \tag{5.7}$$

as our measurement, obtained by a centered finite difference around $z = 0$ for $\partial I/\partial z$. We then specify TIE in the Fourier domain as

$$\widehat{y}(\boldsymbol{\omega}) = 4\pi\nu\Delta z\|\boldsymbol{\omega}\|_2^2\widehat{\phi_0}(\boldsymbol{\omega}), \tag{5.8}$$

---

[2] The real part provides us with the so-called transport-of-phase equation (TPE). However, TPE requires the measurement of the axial derivative of the phase $\partial\phi/\partial z$. This makes TPE impractical as ϕ itself is unknown.

where the Fourier transform of a function $f : \mathbb{R}^2 \to \mathbb{C}$ is given by

$$\begin{aligned}\widehat{f}(\boldsymbol{\omega}) &= \mathcal{F}\left\{f\right\}(\boldsymbol{\omega}) \\ &= \int_{\mathbb{R}^2} f(\boldsymbol{x}) \mathrm{e}^{-\mathrm{j}2\pi\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}} \, \mathrm{d}\boldsymbol{x}. \end{aligned} \tag{5.9}$$

Consequently, the sought phase is obtained by

$$\phi_0 = \mathcal{F}^{-1}\left\{\widehat{y}\,\widehat{h}_{\mathrm{TIE}}^{-1}\right\}, \tag{5.10}$$

where

$$\widehat{h}_{\mathrm{TIE}}(\boldsymbol{\omega}) = \begin{cases} \zeta, & \boldsymbol{\omega} = \boldsymbol{0} \\ 4\pi\nu\Delta z\|\boldsymbol{\omega}\|_2^2, & \text{otherwise,} \end{cases} \tag{5.11}$$

with $\zeta > 0$ being introduced to deal with the singularity at the origin. This makes the inverse filter $\widehat{h}_{\mathrm{TIE}}^{-1}$ well-defined. In practical terms, (5.10) explains that a conventional reconstruction of the phase requires one to capture three images (the infocus image recorded at $z = 0$ and the two defocus images recorded at $z = \pm\Delta z$), and to use a Fourier-domain filtering operation.

We now would like to discuss how the defocus distance affects the reconstruction performance. From a mathematical point of view, Δz should be as small as possible so that $\partial I/\partial z$ is well-approximated. However, in practice, the intuitive appeal of choosing a very small defocus is suboptimal. To see this, let us assume that the measurement y contains some amount of additive white noise that does not depend on Δz. The spectrum of the noise in the reconstructed phase ϕ_0 is *shaped* by $\widehat{h}_{\mathrm{TIE}}^{-1}$ which is essentially a two fold integrator (since $\widehat{h}_{\mathrm{TIE}} \propto \|\boldsymbol{\omega}\|_2^2$). Therefore, the phase image (especially the lower spatial frequencies) is perturbed more as Δz gets smaller (see Figure 5.3). This aspect of TIE is well-known and discussed in several works [11, 130, 124, 131].

In summary, the key observation is that TIE-based methods resolve lower spatial frequencies better by increasing the defocus distance since noise-induced artifacts are reduced. However, for large Δz, the implicit linearity in the finite difference approximation breaks and one obtains coarser $\partial I/\partial z$ estimates. It is thus reasonable to use large defocus measurements for reconstructing low frequencies as we rely on small defocus measurements for the high ones. Unfortunately, (5.8) by itself does not provide us with further hints about how to combine these two regimes. To have a better understanding, we shall explore a complementary formulation.

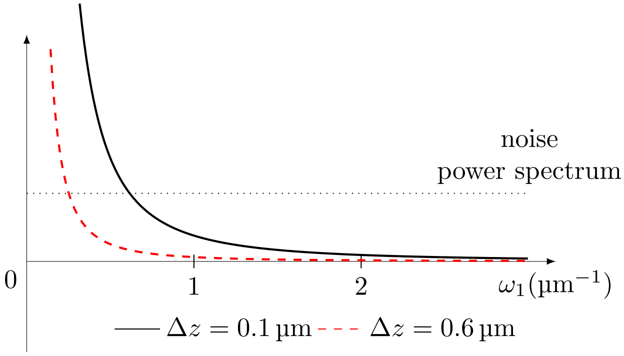


Figure 5.3: Radial profile of the inverse transfer function $\widehat{h}_{\mathrm{TIE}}^{-1}$ for $\nu = 632\,\mathrm{nm}$ using two distinct defocus distances. For simplicity, the profile is shown only for positive spatial frequencies.

5.3.2 Contrast Transfer Function

In the sequel, we base our approach on the principles of wave propagation. Under the paraxial approximation, the intensity of the wave field is expressed as

$$\begin{aligned} I(\boldsymbol{x}, z) &= |U(\boldsymbol{x}, z)|^2 \\ &= |O(\boldsymbol{x}) * p(\boldsymbol{x}, z)|^2, \end{aligned} \tag{5.12}$$

where $*$ denotes the convolution operator and

$$p(\boldsymbol{x}, z) = \frac{\mathrm{e}^{\mathrm{j}kz}}{\mathrm{j}\nu z} \exp\left(\frac{\mathrm{j}k}{2z}\|\boldsymbol{x}\|_2^2\right) \tag{5.13}$$

represents the impulse response of the Fresnel propagation [132]. We remark that

$$\hat{p}(\boldsymbol{\omega}, z) = \mathrm{e}^{\mathrm{j}kz}\mathrm{e}^{-\mathrm{j}\pi\nu z\|\boldsymbol{\omega}\|_2^2}. \tag{5.14}$$

In effect, the term $\exp(-\mathrm{j}\pi\nu z\|\boldsymbol{\omega}\|_2^2)$ represents the optical transfer function for the propagation. The complex amplitude is given by

$$\widehat{U_A}(\boldsymbol{\omega}, z) = \widehat{O}(\boldsymbol{\omega}, z)\mathrm{e}^{-\mathrm{j}\pi\nu z\|\boldsymbol{\omega}\|_2^2}.$$

As we are interested in imaging thin biological samples, we assume that ϕ_0 is small. Further, we continue with the assumption that the object is uniformly illuminated. This enables us to approximate the object field $O(\boldsymbol{x})$ by its first-order Taylor expansion that is given by

$$O(\boldsymbol{x}) \approx \sqrt{J_0}\,(1 + \mathrm{j}\phi_0(\boldsymbol{x}))\,. \tag{5.15}$$

Then,

$$\widehat{U_A}(\boldsymbol{\omega}, z) \approx \sqrt{J_0}\,\Big(\delta(\boldsymbol{\omega}) + \widehat{\phi_0}(\boldsymbol{\omega})\,\big(\mathrm{j}\cos(\pi\nu z\|\boldsymbol{\omega}\|_2^2) + \sin(\pi\nu z\|\boldsymbol{\omega}\|_2^2)\big)\Big).$$

Also, we remark that

$$\widehat{I}(\boldsymbol{\omega}, z) = \Big(\widehat{U_A}^{\,*}(-\cdot, z) * \widehat{U_A}(\cdot, z)\Big)\,(\boldsymbol{\omega}) \tag{5.16}$$

and that $\widehat{\phi_0}$ is Hermitian-symmetric since ϕ_0 is real-valued. Then, by developing (5.16) up to the first-order term in ϕ_0 (i.e., neglecting the higher-order terms that include the convolution $\widehat{\phi_0} * \widehat{\phi_0}$), we obtain

$$\widehat{I}(\boldsymbol{\omega}, z) = J_0\,\Big(\delta(\boldsymbol{\omega}) + 2\sin(\pi\nu z\|\boldsymbol{\omega}\|_2^2)\widehat{\phi_0}(\boldsymbol{\omega})\Big), \tag{5.17}$$

which is known as the *contrast transfer function* (CTF) [133]. As suggested by its name, the CTF explains how the phase and the propagation distance is changing the image contrast. For instance, in the present case, it points out that the visibility of phase-only objects is increased by introducing a suitable defocus (see Figure 5.8(a)). Evaluating (5.17) at points $z = \Delta z$ and $z = -\Delta z$ yields the TIE-like equation

$$\widehat{b}(\boldsymbol{\omega}) = 4\sin(\pi\nu\Delta z\|\boldsymbol{\omega}\|_2^2)\widehat{\phi_0}(\boldsymbol{\omega}). \tag{5.18}$$

Similar to $\widehat{h}_{\mathrm{TIE}}$ in (5.10), we define

$$\widehat{h}_{\mathrm{CTF}}(\boldsymbol{\omega}) = \begin{cases} \zeta, & \boldsymbol{\omega} = \mathbf{0} \\ 4\sin(\pi\nu\Delta z\|\boldsymbol{\omega}\|_2^2), & \text{otherwise.} \end{cases} \tag{5.19}$$

Considering $\widehat{h}_{\mathrm{TIE}}$ and $\widehat{h}_{\mathrm{CTF}}$, let us further develop our approach for the phase retrieval problem. We make two essential observations:

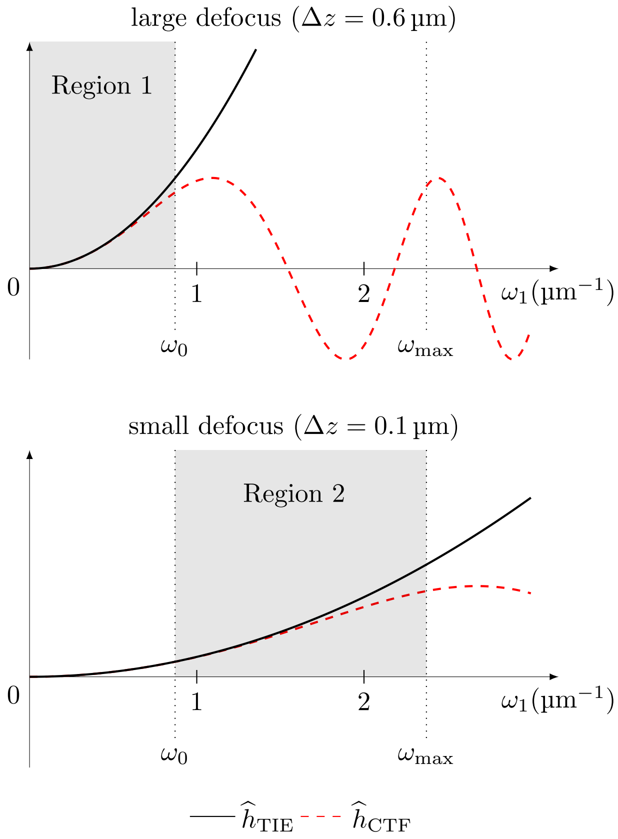


Figure 5.4: Radial profile of the transfer functions for $\nu = 632\,\mathrm{nm}$. For simplicity, the profile is shown only for positive spatial frequencies.

– We notice that the models are indistinguishable since

$$4\sin(\pi\nu\Delta z\|\boldsymbol{\omega}\|_2^2) \approx 4\pi\nu\Delta z\|\boldsymbol{\omega}\|_2^2,$$

when Δz is sufficiently small. Therefore, if small defocus distances are considered for imaging weak phase objects, both TIE and CTF models result in similar frameworks. This implies that finite difference approach—used for linearization of the intensity variation along z—is consistent with the CTF formulation. In plain words, the mathematical approximation incorporated in the TIE has a physical correspondence based on the paraxial wave propagation.

– In contrast, as Δz gets larger, the models agree up to a certain frequency ω_0, and differ afterwards (see Figure 5.4). The validity of utilizing TIE with large defocus measurements is hence enforced in the region where it agrees with the CTF model. Above ω_0, the two models tend to disagree and it is therefore safer to exclude these frequencies.

Based on this, we shall formulate a joint approach for the phase reconstruction problem that takes the best of two worlds. Large defocus intensity images are used to recover low frequencies up to ω_0 (denoted as Region 1 in Figure 4). For the rest of the spectrum (denoted as Region 2 in Figure 4), we rely on the small defocus images.

5.4 Reconstruction Algorithm

In this section, we put our phase recovery problem in a variational framework. Our effort towards this aim is initiated by the formulation of the data fidelity term. Guided by our developments in the previous section, we shall work with two defocus distances and formulate a weighted reconstruction. The derived data term is then combined with a regularization functional which will be chosen in accordance with our investigations in Chapter 4. In the sequel, we start our developments by constructing the spectral weighting filters to effectively combine low- and high-frequency components of the phase.

5.4.1 Spectral Weighting Filters

Considering the physical principles explained in Section 5.3, we introduce two isotropic functions that are appropriate for the spectral weighting of the underlying phase image. We denote these functions by W_{LP} and W_{HP}, for low-pass and high-pass weighting, respectively. They are defined as follows:

$$W_{\mathrm{LP}}(\boldsymbol{\omega}) = \begin{cases} 1, & \|\boldsymbol{\omega}\|_2 < \omega_0 - \frac{L}{2} \\ g(\boldsymbol{\omega}), & \omega_0 - \frac{L}{2} < \|\boldsymbol{\omega}\|_2 < \omega_0 + \frac{L}{2} \\ 0, & \|\boldsymbol{\omega}\|_2 > \omega_0 + \frac{L}{2}, \end{cases} \tag{5.20}$$

where

$$g(\boldsymbol{\omega}) = \frac{1}{2}\left(1 + \cos\left(\frac{\pi}{L}\left(\|\boldsymbol{\omega}\|_2 - \omega_0 + \frac{L}{2}\right)\right)\right) \tag{5.21}$$

and

$$W_{\mathrm{HP}}(\boldsymbol{\omega}) = 1 - W_{\mathrm{LP}}(\boldsymbol{\omega}). \tag{5.22}$$

The parameter ω_0 denotes the cutoff frequency; the values of W_{HP} and W_{LP} are equal for $\|\boldsymbol{\omega}\|_2 = \omega_0$. Note that (5.21) provides a smooth transition zone of width L around ω_0. A graphical representation of our filters is seen in Figure 5.5.

It is noteworthy that ω_0 specifies the limit spatial frequency up to which the phase reconstruction based on a large defocus is reliable. For the frequencies that are higher than ω_0, we shall rely on the measurements obtained with a small defocus. Considering (5.11) and (5.19), we define ω_0 as

$$\omega_0 = \sqrt{\frac{\theta_0}{\pi\nu\Delta z}}, \tag{5.23}$$

where θ_0 is the critical value after which the difference between $\widehat{h}_{\mathrm{TIE}}$ and $\widehat{h}_{\mathrm{CTF}}$ is not negligible. A typical choice is $\theta_0 = \pi/10$.

5.4.2 Discrete Formulation

We consider that two distinct defocus values, Δz_1 and Δz_2, with $\Delta z_1 < \Delta z_2$, are used. Let y_1 and y_2 represent the corresponding measurements computed via (5.7) for Δz_1 and Δz_2, respectively. Since the measurements are only known at discrete pixel locations, we collect their sample values in the vectors[3] $\mathbf{y}_1$, $\mathbf{y}_2 \in \mathbb{R}^N$ with N being the total number of pixels on the detector. Then, in the noise-free

---

[3] The vectorization of the measurements are assumed to be lexicographically ordered.

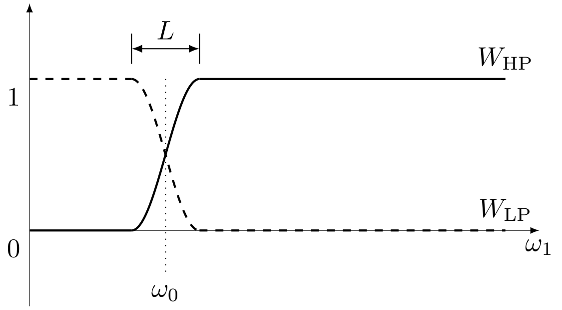


Figure 5.5: Radial profile of the spectral filters used to weight low and high frequencies of the measurements, computed for two defocus distances.

scenario, (5.8) implies that

$$\begin{aligned}\mathbf{y}_1 &= \mathbf{H}_1\mathbf{\Phi},\\ \mathbf{y}_2 &= \mathbf{H}_2\mathbf{\Phi},\end{aligned} \tag{5.24}$$

where $\mathbf{\Phi} \in \mathbb{R}^N$ is the discretized version of the original phase map and $\mathbf{H}_1$, $\mathbf{H}_2 \in \mathbb{R}^{N\times N}$ correspond to h_{TIE} for Δz_1 and Δz_2, respectively (thus, they represent Laplacian operators up to different multiplicative factors). Note that these operators are self-adjoint in the sense that $\mathbf{H}_1^{\mathrm{T}} = \mathbf{H}_1$ and $\mathbf{H}_2^{\mathrm{T}} = \mathbf{H}_2$.

We then define our weighting matrices as

$$\begin{aligned}\mathbf{W}_1 &= \mathbf{F}^{-1}\mathbf{\Sigma}_{\mathrm{HP}}\mathbf{F},\\ \mathbf{W}_2 &= \mathbf{F}^{-1}\mathbf{\Sigma}_{\mathrm{LP}}\mathbf{F},\end{aligned} \tag{5.25}$$

where $\mathbf{F}$ is the DFT matrix and $\mathbf{\Sigma}_{\mathrm{LP}}$, $\mathbf{\Sigma}_{\mathrm{HP}} \in \mathbb{R}^{N\times N}$ are diagonal matrices whose entries are the discrete samples of W_{LP} and W_{HP}, respectively. In practice, a small positive constant can be added to the diagonal elements. We remark that such construction makes $\mathbf{W}_1$ and $\mathbf{W}_2$ positive-definite matrices.

We propose the data-fidelity term

$$\mathcal{D}(\mathbf{\Phi};\mathbf{y}_1;\mathbf{y}_2) =\frac{1}{2}\|\mathbf{H}_1\mathbf{\Phi} - \mathbf{y}_1\|^2_{\mathbf{W}_1} + \frac{1}{2}\|\mathbf{H}_2\mathbf{\Phi} - \mathbf{y}_2\|^2_{\mathbf{W}_2}, \tag{5.26}$$

where $\|\cdot\|^2_{\mathbf{W}_j}$ is the weighted norm defined as $\langle \mathbf{W}_j \cdot, \cdot \rangle$ for $j = 1, 2$. This corresponds to a maximum-likelihood (ML) functional under the hypothesis that the measurements are degraded by Gaussian noise [134].

Based on this formalism, we aim at obtaining a phase reconstruction $\boldsymbol{\Phi}^\star$ such that

$$\boldsymbol{\Phi}^\star = \arg\min_{\boldsymbol{\Phi}} \ (\mathcal{D}(\boldsymbol{\Phi}; \mathbf{y}_1; \mathbf{y}_2) + \tau \mathcal{R}(\boldsymbol{\Phi})). \tag{5.27}$$

In (5.27), $\mathcal{R}$ is the regularization functional of the form

$$\mathcal{R}(\boldsymbol{\Phi}) = \sum_{\boldsymbol{k} \in \Omega} \|[\mathbf{L}\boldsymbol{\Phi}]_{\boldsymbol{k}}\|_2^p, \tag{5.28}$$

where Ω is the index set of all pixel locations, p is a positive scalar value, and $\mathbf{L}$ is the discrete analogue of the transverse gradient operator. Thus, $[\mathbf{L}\boldsymbol{\Phi}]_{\boldsymbol{k}} \in \mathbb{R}^2$ represents the first-order finite differences of $\boldsymbol{\Phi}$ along the vertical and horizontal directions at location $\boldsymbol{k}$. We note that (5.28) is in line with the discrete regularizer given in (4.2). The proposed data term selectively incorporates the high- and low-frequency information coming from the measurements taken at $\pm\Delta z_1$ and $\pm\Delta z_2$, respectively. The regularizer imposes smoothness on the reconstructions, where the strength of the smoothness is controlled by the regularization parameter $\tau > 0$.

Let us explain some special cases of the proposed framework. Discarding the spectral weighting (i.e., $\mathbf{W}_1$ and $\mathbf{W}_2$ are identity operators) and setting $p = 2$, one obtains the classical Tikhonov regularization, which can be seen as the conventional way of solving TIE. More importantly, if weighting is nontrivial and Tikhonov regularization is considered, the solution is given by

$$\boldsymbol{\Phi}^\star_{\text{Tik}} = \underbrace{\left(\mathbf{H}_1\mathbf{W}_1\mathbf{H}_1 + \mathbf{H}_2\mathbf{W}_2\mathbf{H}_2 + 2\tau\mathbf{L}^{\mathrm{T}}\mathbf{L}\right)}_{\mathbf{A}}{}^{-1} \underbrace{\left(\mathbf{H}_1\mathbf{W}_1\mathbf{y}_1 + \mathbf{H}_2\mathbf{W}_2\mathbf{y}_2\right)}_{\mathbf{b}}. \tag{5.29}$$

Assuming periodic boundary conditions, the good news is that the matrix $\mathbf{A}$ becomes BCCB by construction so that the solution $\boldsymbol{\Phi}^\star_{\text{Tik}} = \mathbf{A}^{-1}\mathbf{b}$ is computed directly by using FFTs as we have explained in Section 4.3.1. More specifically, as $\tau \to 0$ (i.e., no regularization), one recovers a linear phase reconstruction method that is in spirit of [123, 124, 125, 126]. We note that the matrix inversion in (5.29) is well-defined, even in the absence of regularization.

In cell imaging applications, it is highly desirable that the specimens be well-isolated from the background and that the background itself is homogeneous [115]. As we have highlighted in Section 4.4, TV regularization—obtained by setting $p = 1$—is known to preserve the discontinuities better than Tikhonov-type regularizations as it provides us with piecewise smooth reconstructions. To attain the mentioned attributes in the final phase image, we combine our data-fidelity term with a TV regularizer and propose the nonlinear optimization problem

$$\mathbf{\Phi}^\star = \arg\min_{\mathbf{\Phi}} \left(\frac{1}{2}\|\mathbf{H}_1\mathbf{\Phi} - \mathbf{y}_1\|^2_{\mathbf{W}_1} + \frac{1}{2}\|\mathbf{H}_2\mathbf{\Phi} - \mathbf{y}_2\|^2_{\mathbf{W}_2} + \tau \sum_{\boldsymbol{k}\in\Omega} \|[\mathbf{L\Phi}]_{\boldsymbol{k}}\|_2\right). \quad (5.30)$$

In the sequel, we are going to solve our specific phase reconstruction problem using generic optimization tools such as ADMM and proximal mapping, which have been discussed in Chapter 4.

5.4.3 Optimization Algorithm

We first cast (5.30) as a constrained optimization problem given by

$$\begin{aligned}\mathbf{\Phi}^\star = &\arg\min_{\mathbf{\Phi}} \left(\frac{1}{2}\|\mathbf{H}_1\mathbf{\Phi} - \mathbf{y}_1\|^2_{\mathbf{W}_1} + \frac{1}{2}\|\mathbf{H}_2\mathbf{\Phi} - \mathbf{y}_2\|^2_{\mathbf{W}_2}\right.\\ &\left. + \tau \sum_{\boldsymbol{k}\in\Omega} \|\mathbf{u}[\boldsymbol{k}]\|_2\right) \text{ s.t. } \mathbf{u} = \mathbf{L\Phi}, \end{aligned} \quad (5.31)$$

where $\mathbf{u}$ is an auxiliary variable. To solve (5.31), we introduce the associated augmented Lagrangian functional

$$\begin{aligned}\mathcal{L}_{\mathcal{A}}(\mathbf{\Phi}, \mathbf{u}, \boldsymbol{\alpha}) = &\frac{1}{2}\|\mathbf{H}_1\mathbf{\Phi} - \mathbf{y}_1\|^2_{\mathbf{W}_1} + \frac{1}{2}\|\mathbf{H}_2\mathbf{\Phi} - \mathbf{y}_2\|^2_{\mathbf{W}_2}\\ &+ \tau \sum_{\boldsymbol{k}\in\Omega} \|\mathbf{u}[\boldsymbol{k}]\|_2\\ &- \boldsymbol{\alpha}^{\mathrm{T}}(\mathbf{u} - \mathbf{L\Phi}) + \frac{\mu}{2}\|\mathbf{u} - \mathbf{L\Phi}\|^2_2, \end{aligned} \quad (5.32)$$

where $\boldsymbol{\alpha}$ is the Lagrange multiplier and $\mu > 0$ is the penalty parameter. We then use ADMM and individually treat $\mathcal{L}_{\mathcal{A}}$ over each of its arguments while the others are assumed to be fixed. This results in an iterative framework whose steps are

$$\mathbf{u}^{t+1} = \arg\min_{\mathbf{u}} \quad \mathcal{L}_{\mathcal{A}}(\boldsymbol{\Phi}^t, \mathbf{u}, \boldsymbol{\alpha}^t), \tag{5.33a}$$

$$\boldsymbol{\Phi}^{t+1} = \arg\min_{\boldsymbol{\Phi}} \quad \mathcal{L}_{\mathcal{A}}(\boldsymbol{\Phi}, \mathbf{u}^{t+1}, \boldsymbol{\alpha}^t), \tag{5.33b}$$

$$\boldsymbol{\alpha}^{t+1} = \boldsymbol{\alpha}^t - \mu(\mathbf{u}^{t+1} - \mathbf{L}\boldsymbol{\Phi}^{t+1}). \tag{5.33c}$$

We note that the minimization over $\mathbf{u}$ in (5.33a) is separable and amounts to the proximal operator associated with $\|\cdot\|_2$. This implies that the solution is obtained by solving the N minimization problems

$$\mathbf{u}^{t+1}[\boldsymbol{k}] = \arg\min_{\mathbf{u}[\boldsymbol{k}]} \Big(\tau\|\mathbf{u}[\boldsymbol{k}]\|_2 + \frac{\mu}{2}\|\mathbf{u}[\boldsymbol{k}] - \mathbf{z}^t[\boldsymbol{k}]\|_2^2\Big), \; \forall \boldsymbol{k} \in \Omega, \tag{5.34}$$

with the closed-form solution being (see Section 4.3.3)

$$\mathbf{u}^{t+1}[\boldsymbol{k}] = \max\left(||\mathbf{z}^t[\boldsymbol{k}]||_2 - \frac{\tau}{\mu}, 0\right) \frac{\mathbf{z}^t[\boldsymbol{k}]}{||\mathbf{z}^t[\boldsymbol{k}]||_2}, \tag{5.35}$$

where $\mathbf{z}^t = \mathbf{L}\boldsymbol{\Phi}^t + \frac{\boldsymbol{\alpha}^t}{\mu}$. As the minimization is implemented for each pixel location in parallel, $\mathbf{u}^{t+1}$ is computed efficiently.

The second sub–problem (5.33b) has the form of a standard quadratic minimization. Investigating first-order optimality conditions yields that

$$\boldsymbol{\Phi}^{t+1} = \left(\mathbf{H}_1\mathbf{W}_1\mathbf{H}_1 + \mathbf{H}_2\mathbf{W}_2\mathbf{H}_2 + \mu\mathbf{L}^{\mathrm{T}}\mathbf{L}\right)^{-1}\mathbf{b}^t, \tag{5.36}$$

where

$$\mathbf{b}^t = \left(\mathbf{H}_1\mathbf{W}_1\mathbf{y}_1 + \mathbf{H}_2\mathbf{W}_2\mathbf{y}_2 + \mu\mathbf{L}^{\mathrm{T}}\left(\mathbf{u}^{t+1} - \frac{\boldsymbol{\alpha}^t}{\mu}\right)\right). \tag{5.37}$$

It is actually seen that a Tikhonov-type minimization is computed. Similar to (5.29), this problem is solved directly using FFTs. The last step (5.33c) is a standard refinement of the Lagrange multiplier.

Algorithm 6: Proposed phase reconstruction algorithm according to (5.30)

```
input  : First defocus value $\Delta z_1$, second defocus value $\Delta z_2$,
         $\partial I/\partial z$ approximations $\mathbf{y}_1$, $\mathbf{y}_2$ (at $\Delta z_1$ and $\Delta z_2$, respectively),
         maximum spatial frequency $\omega_{\max}$, critical value $\theta_0$, wavelength $\nu$,
         transition width $L$, regularization parameter $\tau$, penalty parameter $\mu$
output: Reconstructed phase map $\mathbf{\Phi}^\star$

$\mathbf{H}_1 \leftarrow$ constructTieKernel($\Delta z_1$);
$\mathbf{H}_2 \leftarrow$ constructTieKernel($\Delta z_2$);
$\omega_0 \leftarrow \sqrt{\theta_0/(\pi\nu\Delta z)}$;
$(\mathbf{W}_1, \mathbf{W}_2) \leftarrow$ constructSpectralFilters($\omega_0, L, \omega_{\max}$); (see (5.20))
$\widehat{\mathbf{E}}_{\mathbf{A}} \leftarrow$ computeEigenvalues($\mathbf{H}_1\mathbf{W}_1\mathbf{H}_1 + \mathbf{H}_2\mathbf{W}_2\mathbf{H}_2 + \mu\mathbf{L}^{\mathrm{T}}\mathbf{L}$);
$t \leftarrow 0$; $\mathbf{s}^t \leftarrow \mathbf{0}$; $\mathbf{u}^t \leftarrow \mathbf{0}$; $\boldsymbol{\alpha}^t \leftarrow \mathbf{0}$; $\mathbf{b} \leftarrow \mathbf{H}_1\mathbf{W}_1\mathbf{y}_1 + \mathbf{H}_2\mathbf{W}_2\mathbf{y}_2$;
repeat
    $\mathbf{z}^t \leftarrow \mathbf{L}\mathbf{s}^t + \boldsymbol{\alpha}^t/\mu$;
    for $\boldsymbol{k} \in \Omega$
        $\mathbf{u}^{t+1}[\boldsymbol{k}] \leftarrow \max\left(||\mathbf{z}^t[\boldsymbol{k}]||_2 - \frac{\tau}{\mu}, 0\right) \frac{\mathbf{z}^t[\boldsymbol{k}]}{||\mathbf{z}^t[\boldsymbol{k}]||_2}$; (proximal mapping)
    end for
    $\mathbf{b}^t \leftarrow \mathbf{b} + \mu\mathbf{L}^{\mathrm{T}}\left(\mathbf{u}^{t+1} - \boldsymbol{\alpha}^t/\mu\right)$;
    $\mathbf{\Phi}^{t+1} \leftarrow \mathbf{A}^{-1}\mathbf{b}^t$; (use Algorithm 4)
    $\boldsymbol{\alpha}^{t+1} \leftarrow \boldsymbol{\alpha}^t + \mu(\mathbf{L}\mathbf{s}^{t+1} - \mathbf{u}^{t+1})$;
    $t \leftarrow t + 1$;
until stopping criteria
return $\mathbf{\Phi}^t$;
```

5.5 Experiments

We illustrate the utility of our phase reconstruction method by performing experiments in simulated and practical configurations. In every experiment, the images are extended using periodic boundary conditions. The ζ parameter in (5.11) is set to 1 so that the filter $\widehat{h}_{\mathrm{TIE}}$ is mean-preserving. The θ_0 parameter in (5.23) is set to $\pi/10$, and the associated width L is manually tuned according to the maximal frequency present in the signal. For synthetic data, this maximal frequency is set to the inverse of the pixel size; for real data experiments, it is the minimum of the inverse effective pixel size (i.e., the camera pixel size divided by the magnification) and the resolution imposed by the diffraction limit (i.e., twice the numerical aperture divided by the wavelength). For ADMM-based reconstructions, we fix $\mu = 10\tau$. We also fix the stopping criterion as reaching either a relative ℓ_2-normed difference of 10^{-4} between two successive iterates, or a maximum of 250 iterations.

5.5.1 Synthetic Data

The original object field is assumed to have a unit intensity, and the spatial phase map is given by the set of two-dimensional images shown in Figures 5.6 and 5.7. [4] All the ground-truth phase maps are of size (256×256) pixels and have values in the range of $[0, 1]$ radians. We set the pixel size to 2 µm and the wavelength ν to 632 nm. The original object field is propagated to axial distances of ±50 µm (small defocus) and ±300 µm (large defocus) by using a Fresnel diffraction kernel [135]. We generate the observed intensities by taking the square modulus of the propagated complex fields. Finally, the intensity images are degraded by various levels of AWGN. The standard deviation of the noise is chosen such that a given signal-to-noise ratio (SNR) is achieved. An example defocus stack is illustrated in Figure 5.8(a).

We measure the quality of the reconstructed phase maps by computing the regressed SNR—the regression is there to get rid of additive and multiplicative constants. Between the ground-truth map $\mathbf{\Phi}$ and a reconstructed one $\mathbf{\Phi}^\star$, this measure is defined as

$$\mathrm{SNR}(\mathbf{\Phi}, \mathbf{\Phi}^\star) = \max_{a,b \in \mathbb{R}} 10 \log \left(\frac{\|\mathbf{\Phi}\|_2^2}{\|\mathbf{\Phi} - (a\mathbf{\Phi}^\star + b)\|_2^2} \right). \tag{5.38}$$

---

[4] The biological images originate from the cell image database and are available at http://www.cellimagelibrary.org.

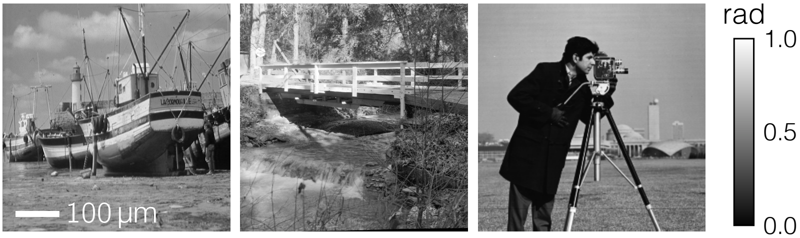


Figure 5.6: The set of standard images representing the ground-truth spatial phase maps used in the simulations. From left to right, they are referred to as 1) Boat, 2) Bridge, 3) Cameraman.

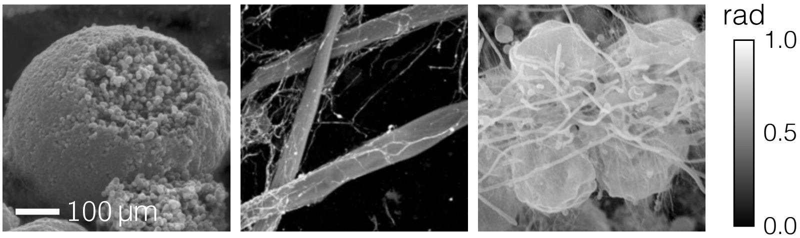


Figure 5.7: The set of cell images representing the ground-truth spatial phase maps used in the simulations. From left to right, they are referred to as, 1) CIL214, 2) CIL38921, and 3) CIL39789.

The performance of our nonlinear method is compared against the following algorithms:

1. Tikhonov regularization based TIE (TIE-Tik): We adapt a linear reconstruction scheme that incorporates a Tikhonov regularization (i.e., $p = 2$ in (5.28)). The method uses three images (the infocus image and two symmetrically defocused images) and does not apply any spectral weighting scheme. We apply the method separately with each set of measurements acquired at $\pm 50\,\mu\text{m}$ and $\pm 300\,\mu\text{m}$ defocus.

Table 5.1: Reconstruction performance of TIE-based phase retrieval algorithms: Simulation results for the set of standard test images. Numbers are given in decibel unit.

	Input SNR	$\Delta z = 50\,\mu\text{m}$		$\Delta z = 300\,\mu\text{m}$		Composite TIE-Tik	Proposed method
		TIE-Tik	TIE-TV	TIE-Tik	TIE-TV		
Boat	20	10.72	12.26	12.85	14.52	13.47	**15.22**
	25	11.64	13.86	14.52	15.63	15.94	**17.19**
	30	12.34	15.24	15.84	16.32	18.10	**18.70**
	35	14.16	16.74	16.78	16.88	19.88	**20.01**
	40	16.09	17.98	17.09	17.11	20.69	**20.73**
	45	18.08	19.20	17.19	17.20	20.95	**20.96**
	50	19.99	20.19	17.23	17.23	**21.04**	**21.04**
Bridge	20	8.59	10.56	10.91	11.99	11.69	**13.11**
	25	9.51	11.68	12.39	12.99	13.95	**14.88**
	30	10.30	12.40	13.27	13.44	15.43	**15.80**
	35	12.40	13.73	13.99	14.01	16.77	**16.84**
	40	13.67	14.74	14.20	14.22	17.24	**17.28**
	45	15.68	15.96	14.28	14.28	17.39	**17.40**
	50	16.86	16.89	14.30	14.31	**17.44**	**17.44**
Cameraman	20	8.00	13.83	13.17	13.91	14.37	**15.89**
	25	9.77	15.54	14.11	14.26	16.00	**16.67**
	30	12.57	16.38	14.43	14.48	16.82	**17.08**
	35	14.77	16.79	14.56	14.55	17.09	**17.18**
	40	16.46	17.23	14.61	14.60	17.18	**17.24**
	45	17.52	**17.60**	14.63	14.62	17.25	17.29
	50	17.70	**17.72**	14.63	14.62	17.24	17.27

2. TV regularization based TIE (TIE-TV): The method is similar to TIE-Tik except that we make the reconstruction nonlinear by using a TV regularizer (i.e., $p = 1$ in (5.28)). Similarly, the algorithm uses three images (without spectral weighting) and is tested with both small and large defocus cases.

3. Composite TIE-Tik: We adapt the reconstruction scheme given in (5.29). Composite TIE-Tik uses five images. It linearly combines high and low frequency components (together with a Tikhonov regularization) of the phase

maps obtained by using small and large defocus measurements, respectively.The algorithm is compatible with the ones proposed in [123, 124, 125, 126].

As described in Section 5.4, all linear reconstructions are computed directly. The nonlinear ones are solved iteratively using the proposed ADMM-based algorithm. For each reconstruction algorithm, we tune the regularization parameter to achieve the best-possible SNR performance using an oracle. The output SNRs given in Tables 5.1 and 5.2 are averaged over 10 realizations for a reliable comparison.

For the simulated measurements, we see that the proposed method outperforms the other algorithms in almost all of the cases. Especially for moderate and high levels of noise, our framework significantly improves the reconstruction quality. A visual inspection of the reconstructed phase maps (see Figure 5.8) demonstrates that our method is able to reconstruct the high-frequency components accurately. However, the reconstructions using intensity images taken at large defocus distances are notably blurred. We also see that our algorithm produces much fewer artifacts in the low-frequency regions. Reconstruction errors are more visible for the other methods. As one gets to very low levels of noise, TIE-TV (for small defocus) and composite TIE-Tik methods become competitive, and TIE-TV reconstruction can yield the best results for certain configurations. This is explained by the decreased presence of noise-induced errors (i.e., low-frequency artifacts). Next, we assess our method in experimental settings to corroborate our simulations.

5.5.2 Real Data

We imaged[5] paraformaldehyde-fixed and unstained HeLa cells at room temperature ($\sim 22°$C). Acquisitions were performed on a Zeiss Axio Observer Z1 microscope (Carl Zeiss AG, Jena, Germany) equipped with a Leica HCX PL Fluotar 40×0.75 NA objective (Leica Microsystems GmbH, Wetzlar, Germany). The camera pixel size is 6.5 µm and the illumination wavelength is 684.5 nm. The defocus images were recorded with distances of ±2 µm and ±10 µm from the best focal position.

A ROI of size (512×512) pixels is chosen. Based on the previous results, we compare the reconstruction performance of TIE-Tik, TIE-TV, composite TIE-Tik, and the proposed method in a qualitative manner. We note that both TIE-Tik and

---

[5] We would like to thank Benjamin Rappaz from the Biomolecular Screening Facility, EPFL, and Etienne Shaffer from the Microsystems Laboratory 1, EPFL, for their help in acquiring the experimental data.

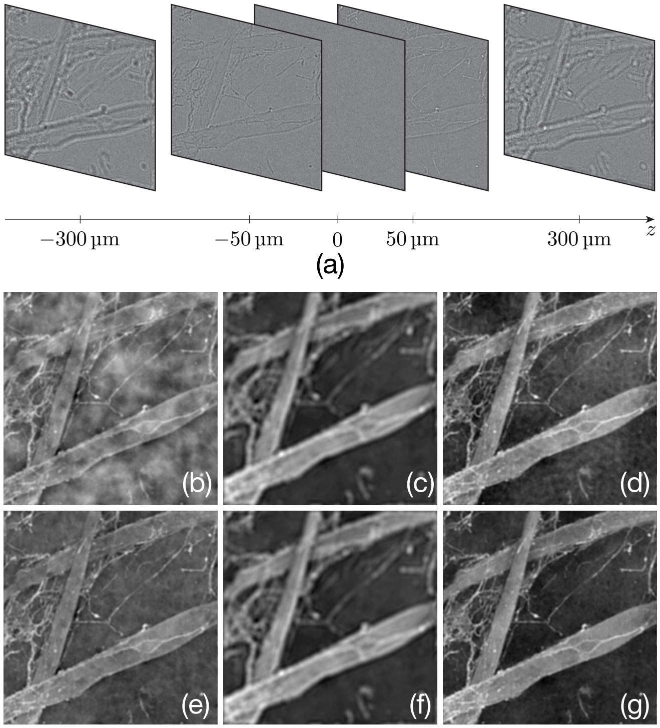


Figure 5.8: Illustrative example of a phase-reconstruction simulation, where the object field (at $z = 0$) has unit intensity and the phase is encoded by the CIL38921 image: (a) Simulated measurements (SNR is 20 dB), (b) TIE-Tik reconstruction using the measurements at 50 µm (SNR is 4.01 dB), (c) TIE-Tik reconstruction using the measurements at 300 µm (SNR is 8.67 dB), (d) composite TIE-Tik reconstruction (SNR is 9.72 dB), (e) TIE-TV reconstruction using the measurements at 50 µm (SNR is 8.31 dB), (f) TIE-TV reconstruction using the measurements at 50 µm (SNR is 11.25 dB), and (g) proposed method (SNR is 13.20 dB).

Table 5.2: Reconstruction performance of TIE-based phase retrieval algorithms: Simulation results for the set of cell images. Numbers are given in decibel unit.

	Input SNR	$\Delta z = 50\,\mu\text{m}$ TIE-Tik	$\Delta z = 50\,\mu\text{m}$ TIE-TV	$\Delta z = 300\,\mu\text{m}$ TIE-Tik	$\Delta z = 300\,\mu\text{m}$ TIE-TV	Composite TIE-Tik	Proposed method
CIL214	20	9.52	11.87	13.18	14.82	13.54	**15.12**
	25	10.60	13.52	14.59	15.86	15.72	**16.94**
	30	12.06	14.86	16.05	16.73	18.46	**19.22**
	35	13.34	15.87	17.22	17.37	21.16	**21.42**
	40	15.81	17.81	17.92	17.96	23.23	**23.35**
	45	18.11	19.68	18.08	18.07	24.05	**24.09**
	50	21.82	21.97	18.20	18.19	24.48	**24.49**
CIL38921	20	3.85	8.30	8.49	11.15	9.43	**12.63**
	25	5.30	10.89	10.64	12.11	13.13	**15.55**
	30	7.10	13.41	11.51	12.48	15.41	**17.88**
	35	9.75	15.51	12.46	12.80	18.52	**19.77**
	40	13.06	17.54	12.87	12.95	20.65	**21.11**
	45	15.34	18.86	12.96	12.99	21.35	**21.53**
	50	18.33	20.51	13.02	13.02	21.81	**21.84**
CIL39789	20	11.92	14.64	16.29	18.89	16.57	**19.24**
	25	12.79	17.22	18.17	20.46	19.18	**22.15**
	30	14.22	19.44	19.81	21.34	21.77	**24.58**
	35	16.12	21.47	21.56	22.14	24.96	**26.22**
	40	18.48	23.67	22.35	22.48	27.09	**27.38**
	45	21.42	25.55	22.61	22.65	28.25	**28.40**
	50	25.93	27.67	22.73	22.75	28.65	**28.68**

TIE-TV methods use the measurements recorded at ±2 µm. Since the ground-truth phase is not available, the regularization parameters are manually tuned (by paying attention to image contrast and physiological relevance) for all considered methods. The algorithmic settings of Section 5.5.1 are kept the same for the reconstruction.

All TIE reconstructions improve the visibility of the cell compared to the in-focus bright-field image. However, a closer examination reveals that the proposed approach enhances the homogeneity of the background better than its competitors. It allows one to better distinguish the cell membrane (see Figure 5.9). These aspects suggest that the proposed TIE-based approach is suitable for imaging phase-only objects.

5.5.3 Validation

Our final step is to compare different phase imaging techniques. To have a reference, the HeLa cells are imaged with a digital holographic microscope (DHM) that allows to directly obtain the complex object field. The hologram image is acquired in off-axis configuration on a T-1001 microscope (LynceeTec SA, Lausanne, Switzerland) using a Leica HI PLAN I 10×0.22 NA objective (Leica Microsystems GmbH, Wetzlar, Germany). The DHM image is then reconstructed numerically using the algorithm in [114]. A stack of bright-field images is acquired using the same objective. We then choose to apply our TIE formalism to differential interference contrast (DIC) images. The applicability of TIE to DIC imaging has been proposed in [120], where the authors consider DIC image formation together with thin phase-only object assumption (see [120] for further details). Therefore, we also acquire DIC images using an A-Plan 10×0.25 NA with a DIC analyser II. Both image stacks include the infocus image as well as images recorded at $\pm 2\,\mu\text{m}$ and $\pm 10\,\mu\text{m}$ defocus distances. Since the DHM measurements are performed on a different setup, we have been unable to align the exact same ROI. Instead, we choose a common ROI of size (256×256) in all images. We perform TIE reconstructions using our method on the chosen ROI.

We remark that the bright-field image has very low contrast, as explained in Section 5.2. Meanwhile, DIC microscopes increase the contrast (also the visibility) of the cells; the final image can be roughly seen as the directional derivative of the phase. Note that DHM records a hologram of the object (i.e., single image acquisition) and produces the phase image by demodulating the hologram, which is a linear operation. However, DHM requires dedicated hardware and is much costlier compared than bright-field and DIC microscopes. By qualitatively looking at the results given in Figure 5.10, we see that the two TIE reconstructions render the morphology of the cells faithfully in accordance with the DHM image. We remark that TIE with DIC images produces sharper results than its counterpart using bright-field images. It is also seen that the TIE reconstructions have homogeneous backgrounds. Hence, TIE can be considered as a cost-effective alternative to DHM.

As last illustration of the potential of our technique, we consider the problem of cell segmentation (i.e., automatic delineation of the cell boundaries). We choose a

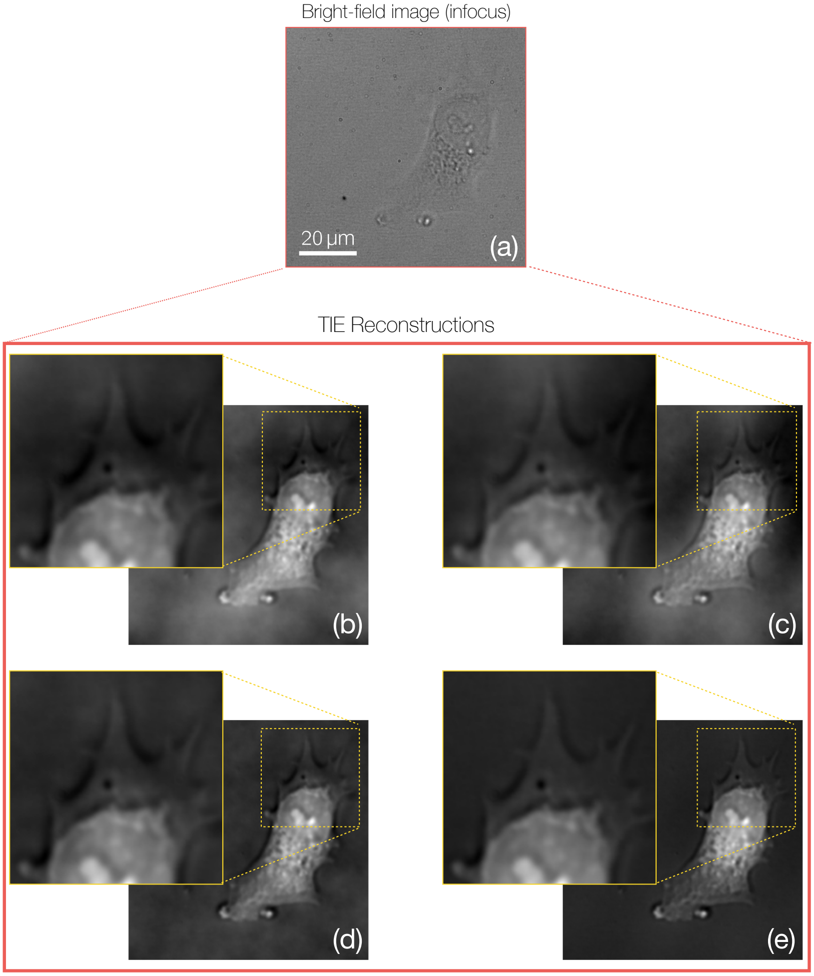


Figure 5.9: TIE reconstruction results of different algorithms for experimentally acquired bright-field images of a HeLa cell: (a) Infocus bright-field image, (b) TIE-Tik reconstruction with defocus images acquired at ±2 μm, (c) composite TIE-Tik reconstruction, (d) TIE-TV reconstruction with defocus images acquired at ±2 μm, and (e) the proposed phase reconstruction method.

larger ROI and use the bright-field images for phase reconstruction. We apply the watershed segmentation algorithm.[6] The segmentation results given in Figure 5.11 are seen to be satisfactory. By applying this basic segmentation algorithm to bright-field images, it is not possible to achieve the same results as the bright-field images lack contrast.

5.6 Summary

The purpose of this chapter has been to utilize our grasp of sparse image reconstruction algorithms for the optical phase retrieval problem. The main effort has concentrated on the development of the data fidelity term. Our problem-specific design has been initiated by the implications of using TIE with measurements at different defocus distances—in terms of the spatial frequencies of the reconstructed phase map–and has led to the formulation of a weighted phase-reconstruction algorithm. Combined with a TV regularizer, the proposed approach nonlinearly combines two phase map that originate from a set of measurements at distinct defocus distances. We have illustrated that our method improves the performance of phase reconstruction, as compared to previous algorithms. We have also shown that the scheme is applicable to experimental data. An important aspect is that obtained real data reconstructions are observed to be in line with the ones produced by DHM which is a well-established dedicated phase microscope. In this regard, our numerical phase reconstruction algorithm is appreciated a practical low-cost substitute since it operates with a bright-field microscope.

---

[6] Available at http://bigwww.epfl.ch/sage/soft/watershed/

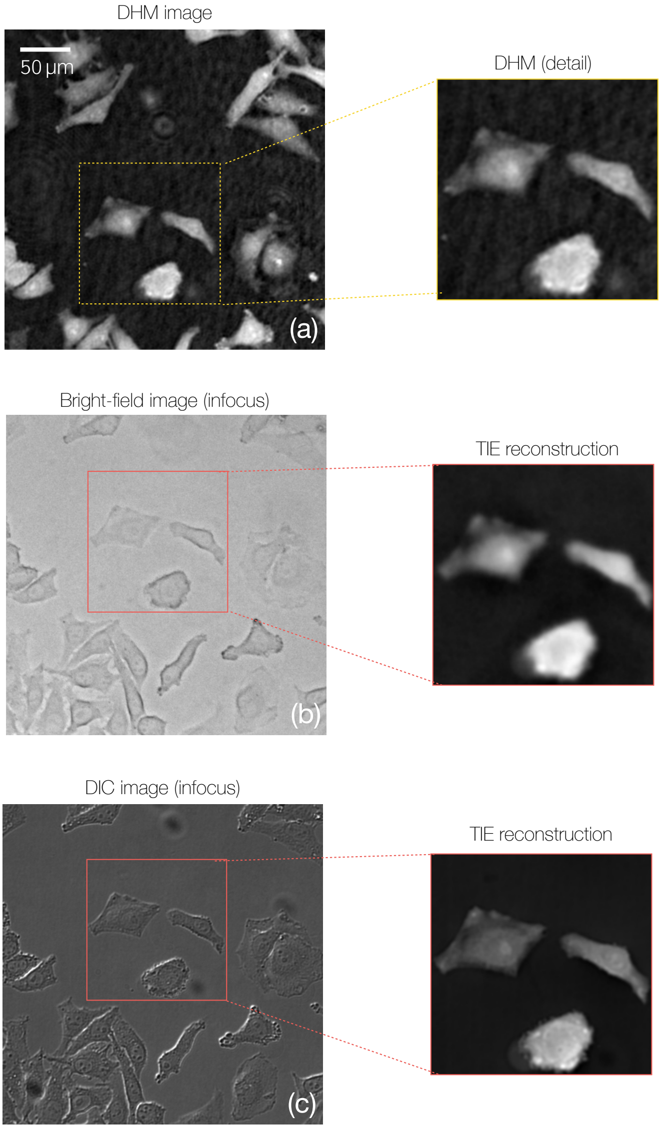


Figure 5.10: Validation of the proposed TIE reconstruction method on the same sample: (a) Reference digital holographic microscope (DHM) image of HeLa cells, (b) result of the TIE reconstruction using bright-field images, and (c) result of the TIE reconstruction using differential interference contrast (DIC) images.

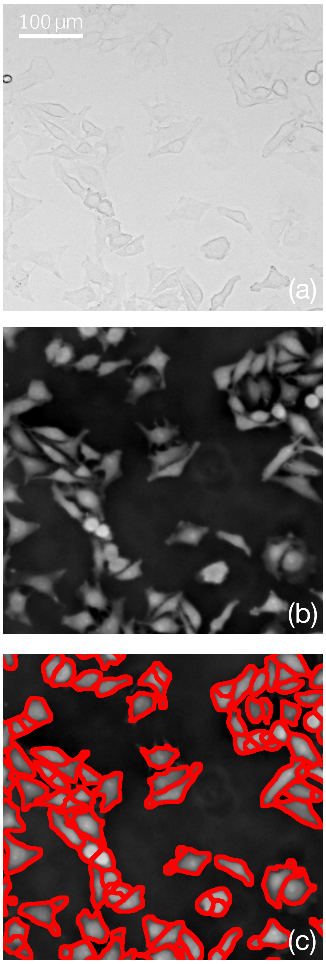


Figure 5.11: Delineation of HeLa cells: (a) Infocus bright-field image, (b) phase image obtained by using the proposed TIE reconstruction, and (c) result of watershed algorithm.

Chapter 6

Sparse Reconstruction of Vector Fields

6.1 Overview

In this chapter,[1] we consider regularized reconstruction of vector fields, extending our previous considerations which have capitalized on scalar-valued images. Motivated by the TV regularization, we propose a class of regularization functionals that penalize the singular values of the Jacobian of a given vector field. Our framework generalizes some well-known variants of TV for vector-valued functions and is invariant with respect to fundamental transformations such as translation, scaling, and rotation. We put special emphasis on the use of a particular member of our class of regularizers—called the nuclear total variation (TVN)—that imposes sparsity on the singular values. Using Legendre-Fenchel duality arguments, we derive an efficient algorithm to solve the consequent optimization problem. Our numerical experiments on phantom data show that TVN improves the denoising performance in comparison to existing vectorial extensions of total variation and curl-divergence regularizations. Finally, we illustrate the practical relevance of the proposed scheme by enhancing the streamline visualization of an experimentally-acquired phase-contrast MRI recording.

---

[1] This chapter is based on our paper [136].

6.2 Introduction

Regularized reconstruction of vector data is becoming a prominent subject of research in image processing. This is partially due to the appearance of vector fields as the appropriate mathematical representation of objects of interest (such as the displacement field in motion estimation, or the deformation field for image registration). More importantly, vector quantities are also measured directly thanks to recent advances in imaging technologies. Particularly, the development of new modalities and techniques has centered around flow fields, putting emphasis on flow-sensitive imaging. For instance, the optical measurement technique known as the particle image velocimetry (PIV) provides instantaneous velocity vector measurements in fluid flows [137]. Tomographic extensions of PIV are capable of measuring three-dimensional velocity fields and are functional tools for analyzing organization of complex turbulent flows [138]. For *in-vivo* medical imaging, phase-contrast MRI (PC MRI) [2] is used for the acquisition of time-dependent blood flow with full volumetric coverage [12]. PC MRI is non-invasive and can be set to measure the blood flow in a specific vascular region of interest. This makes it highly suitable for developing diagnostic strategies for cardiovascular diseases by assessing the generation of complex physiological flow phenomena [139]. The acquired data can also be combined with computational fluid dynamics (CFD) simulations for evaluating surgical outcomes and unfavorable shear stress along the aged or diseased central arteries [140]. By virtue of the increasing presence of flow-sensitive imaging devices (and their important applications in medical imaging), we shall specifically be interested in flow field imaging.

As for all imaging systems, the quality of the measurements are degraded by noise and other imperfections (these could be background phase contributions from eddy-currents or velocity aliasing in case of PC MRI). This brings to the forefront the need for efficient denoising algorithms that can remove these perturbations efficiently. Further, these algorithms are useful for data visualization and quantitative analysis (if further need be). In the development of algorithms for flow maps, the fundamental importance of computational efficiency is far more pronounced than the case of scalar image reconstruction problems. For example, if a three-dimensional flow field regularization is considered, we have three components per voxel, which requires one to store $3N^3$ values for a volume of size $(N \times N \times N)$.

---

[2] In MRI community, different names such as 4D flow MRI and velocity-encoded MRI are used for the same imaging system.

Variational denoising algorithms have been investigated noticeably from two main perspectives:

- Since the flow field measurements are typically related in a fundamental way to some physical phenomenon, one approach has been to impose certain physical constraints on the solution. Following this reasoning, curl and divergence operators are frequently used since they control the rotational and laminar characteristics, respectively. In the context of quadratic regularization, these regularization operators have been considered in combination with ℓ_2-norms for a wide span of applications [141, 142, 143]. Similar to Tikhonov-type image regularization methods, these methods are found to oversmooth the flow discontinuities occurring at interfaces between different fluids and object boundaries. Consistent with the recent trend in biomedical image processing—which favors sparse regularization—researchers also combined these operators with vector ℓ_1-norms [144]. The latter have been effectively used for regularizing flows that exhibit discontinuities [145, 146]; and they are found to typically overperform their quadratic counterparts in terms of denoising performance.

- In the second type of approach, the multi-channel data (for example, color and hyperspectral images) is simply viewed as a vector-valued function. Following the success of TV regularization for scalar images, penalization of the variations of every component of the vector field in a separable way has been considered [147]. While this framework is easy to work with (in terms of algorithm design), it is limited by the fact that it does not take into account the dependencies (whether they are physical or not) that might exist among the different components of the vector data. The pivotal design goal is then set to effectively couple the information coming from different channels as the discontinuities are preserved. Researchers have investigated alternative vectorial extensions of TV regularization that introduce a coupling between the components. Among them, the most popular has been the so-called vectorial total variation (VTV) [148]. More recently, the authors have introduced another regularizer that extends TV by penalizing only the maximum variation of the field at each spatial location [149]. Other methods have used the structure tensor as the regularization operator [150].

6.2.1 Contributions

In the present chapter, we are interested in a reconstruction framework that is well-suited for flow fields with discontinuities. Our guiding principle is to penalize the vector-variations through the consideration of local geometry. To that end, we propose a family of regularizers that penalizes the singular values of the Jacobian operator. The generic numerical problem to solve in our setting is therefore nonlinear. Thus, we identify the need for efficient algorithms.

The main contributions of this chapter are:

- The formulation of a generalized regularization scheme that is appropriate for flow-field denoising. We show that the proposed class of regularizers are invariant to translation, scaling and rotation, and we highlight connections with some well-known vectorial TV extensions.

- The proposal of a particular regularizer (termed nuclear total variation) that penalizes the nuclear norm of the Jacobian evaluated at every spatial location of the flow.

- The derivation of an efficient optimization algorithm based on duality principles. This yields an algorithm that is sufficiently fast to process large volumes of data.

- The experimental demostration that TVN achieves better denoising performance than the existing vectorial TV and curl-divergence models. We further apply the framework to a real PC MRI data of blood flow in the human aorta.

6.2.2 Outline

The chapter is organized as follows: In Section 6.3, we present the relevant mathematical framework for flow-field regularization and revisit the existing schemes. Next in Section 6.4, we introduce a class of regularization functionals that involves the Schatten p-norms of the Jacobian operator. After showing invariance properties, we establish connections between our framework and the previous approaches on vectorial TV. Following this, we specify TVN regularization and explain its effect on the reconstructed flow. We then provide a numerical algorithm that handles the occurring optimization problem in Section 6.5. Finally, we perform experiments on simulated and real data in Section 6.6.

6.3 Flow-Field Regularization

We represent a flow field as a d-dimensional vector field with d components. Such an object is denoted by the vector function $\mathbf{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \dots, f_d(\boldsymbol{x}))$ over $\mathbb{R}^d$.

Similar to our developments in previous chapters, we consider the generic linear observation model, where a vector field $\mathbf{f} : \mathbb{R}^d \to \mathbb{R}^d$ is measured by a system (its physical response is given by the linear operator $\mathbf{H}$) to produce noisy vector-valued measurements

$$\mathbf{y} = \mathbf{Hf} + \mathbf{n}, \tag{6.1}$$

where $\mathbf{n}$ represents the measurement noise. The reconstructed flow is obtained by minimizing an energy so that the problem at hand takes the form of the optimization problem

$$\mathbf{f}^\star = \arg\min_{\mathbf{f}} \quad \mathcal{D}(\mathbf{f}; \mathbf{y}) + \tau \mathcal{R}(\mathbf{f}). \tag{6.2}$$

For the rest of this chapter, we consider the simplest measurement operator (that is the idendity), corresponding to $\mathbf{y} = \mathbf{f} + \mathbf{n}$. We specify a quadratic functional for the fidelity term $\mathcal{D}$ and ultimately consider the generic regularized least-squares problem:

$$\mathbf{f}^\star = \arg\min_{\mathbf{f}} \tfrac{1}{2} \|\mathbf{y} - \mathbf{f}\|_2^2 + \tau \mathcal{R}(\mathbf{f}). \tag{6.3}$$

The general framework established above provides us with a convenient setting for assessing the regularization functional $\mathcal{R}$ (whose contribution is weighted via the regularization parameter $\tau > 0$). Next, we review the existing regularization schemes.

6.3.1 Curl- and Divergence-Based Methods

When the measurements of a flow field is related to some physical phenomenon, it is preferable to formulate a reconstruction framework that controls the physical properties of reconstructed flow. This can be achieved by separately imposing constraints on curl and divergence of the solution. Fundamentally, the irrotational and incompressible characteristics of fluid flows are governed by these operators. For regularization purposes, recent applications concentrate on combining them with L_1-norms[3] following the sparsity-promoting techniques [144]. This results in

---

[3] Vector L_1-norms are defined as the scalar L_1-norm of the magnitude of the vector field. We refer the reader to [144] for further details.

the regularization functional that is of the form

$$\mathcal{R}_{\mathrm{CD}}(\mathbf{f}) = \tau_c \int_{\mathbb{R}^d} |\mathbf{curl}\, \mathbf{f}(\boldsymbol{x})|\, \mathrm{d}\boldsymbol{x} + \tau_d \int_{\mathbb{R}^d} |\mathrm{div}\, \mathbf{f}(\boldsymbol{x})|\, \mathrm{d}\boldsymbol{x}, \tag{6.4}$$

where $|\cdot|$ denotes the absolute value of a scalar, or the magnitude of a vector, as appropriate.

6.3.2 Gradient-Based Methods

We have emphasized many times in the thesis that TV regularization is a popular choice for a wide range of imaging applications. The most favorable property of TV is that it tends to preserve signal discontinuities. Let us remind that the classical version of TV for scalar fields, $f : \mathbb{R}^d \to \mathbb{R}$, is

$$\mathrm{TV}(f) = \int_{\mathbb{R}^d} \|\nabla f(\boldsymbol{x})\|_2 \, \mathrm{d}\boldsymbol{x}. \tag{6.5}$$

The commonality of all extensions of TV is that the definition for vector fields should coincide with the scalar one (6.5) for $d = 1$. The simplest and most straightforward extension of TV involves the penalization of the total intensity variation of every vector component in a separable manner. This leads to the following definition of the so-called *separable* TV [147]:

$$\mathrm{TV}_S(\mathbf{f}) = \sum_{i=1}^{d} \int_{\mathbb{R}^d} \|\nabla f_i(\boldsymbol{x})\|_2 \, \mathrm{d}\boldsymbol{x}. \tag{6.6}$$

While (6.6) is easy to work with, it does not take into account the dependencies (whether they are physical or not) that might exist among the different components of the vector data. Alternative extensions that provide a coupling between the components have also been studied. Among them, the most popular one is the *vectorial total variation* (VTV) [148] which is defined as

$$\mathrm{VTV}(\mathbf{f}) = \int_{\mathbb{R}^d} \left(\sum_{i=1}^{d} \|\nabla f_i(\boldsymbol{x})\|_2^2 \right)^{1/2} \mathrm{d}\boldsymbol{x}. \tag{6.7}$$

6.4 Jacobian-Based Regularization

After reviewing previous attempts, we now propose a class of regularization functionals to solve the flow field denoising problem in (6.3). To establish the desired connection with the standard TV functional, let us first note that the main component of TV is the magnitude of the image gradient which essentially captures the intensity variations that are later penalized. This basically brings the task of choosing the appropriate regularization operator for the flow field case. The Jacobian operator defined as

$$\mathbf{Jf} = \left[\nabla f_1 \dots \nabla f_d\right]^{\mathrm{T}},$$

provides us with the natural extension of the gradient in regards to the following proposition which specifies the Taylor's theorem for vector-valued functions:

Proposition 5. *Let $\boldsymbol{x}_0$ be a point in $\mathbb{R}^N$ and $\mathbf{f} : \mathbb{R}^N \to \mathbb{R}^N$ be differentiable at $\boldsymbol{x}_0$. Then the $\mathbf{Jf}(\boldsymbol{x}_0)$ is the best linear approximation of near $\boldsymbol{x}_0$ in the sense that*

$$\mathbf{f}(\boldsymbol{x}) = \mathbf{f}(\boldsymbol{x}_0) + \mathbf{Jf}(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) + o(\|\boldsymbol{x} - \boldsymbol{x}_0\|_2)$$

for $\boldsymbol{x}$ close to $\boldsymbol{x}_0$.

For flow fields, the Jacobian evaluated at a spatial location $\boldsymbol{x}$ corresponds to a matrix of size $d \times d$ that embodies all possible first-order derivatives of all field components at that specific point. It hence carries the information encoded in the divergence and curl of the field. More importantly, the information about the strength of the flow field variations is encoded in the d singular values of $\mathbf{Jf}$. To see this, we remark that

$$\sqrt{\sigma\left(\mathbf{J}^{\mathrm{T}}\mathbf{Jf}\right)} = \sigma\left(\mathbf{Jf}\right),$$

where σ denotes the singular values. Note that the expression $\mathbf{J}^{\mathrm{T}}\mathbf{J}$ reduces to the squared magnitude of the gradient (essentially the key element of TV) if $\mathbf{f}$ is considered to be a scalar function. Accordingly, the directions of these variations are encoded in the corresponding singular vectors. The crucial outcome is that a vectorial extension of TV should penalize the singular values of the Jacobian. Secondly, since the Jacobian is a matrix-valued operator, it is natural to work with the matrix norms in the design of the regularizer. In line with these remarks, we propose the generic regularizer of the form

$$\mathrm{TV}_p\left(\mathbf{f}\right) = \int_{\mathbb{R}^d} \left\|\mathbf{Jf}\left(\mathbf{x}\right)\right\|_{\mathcal{S}_p} \mathrm{d}\mathbf{x}, \quad \forall p \geq 1 \tag{6.8}$$

where $\|\cdot\|_{\mathcal{S}_p}$ is the Schatten p-norm of a matrix.

Definition 6. *Let* $\mathbf{A} \in \mathbb{R}^{N_1 \times N_2}$ *be a matrix with the singular value decomposition*

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}},$$

where $\mathbf{U} \in \mathbb{R}^{N_1 \times N_1}$ *and* $\mathbf{V} \in \mathbb{R}^{N_2 \times N_2}$ *are unitary matrices containing the singular vectors of* $\mathbf{A}$*, and* $\mathbf{\Sigma} \in \mathbb{R}^{N_1 \times N_2}$ *with non-negative real numbers on the diagonal consisting of the singular values of* $\mathbf{A}$*. Then the Schatten p-norm of* $\mathbf{A}$ *is defined as*

$$\|\mathbf{A}\|_{\mathcal{S}_p} = \left(\sum_{j=1}^{\min(N_1, N_2)} \sigma_j(\mathbf{A})^p \right)^{\frac{1}{p}}, \tag{6.9}$$

where $p \geq 1$ *and* $\sigma_j(\mathbf{A})$ *denotes the* j*th singular value of* $\mathbf{A}$ *corresponding to the* (j, j) *entry of* $\mathbf{\Sigma}$*.*

By looking at Definition 6, we see that Schatten p-norms correspond to computing the standard ℓ_p-norm of a vector that is composed of the singular values of the matrix argument. One major advantage of (6.8) is that it is a convex function of $\mathbf{f}$ since Schatten p-norms are convex [151].

It is noteworthy that if a scalar-valued function is considered in (6.8), the Jacobian reduces to the gradient and the ℓ_p norm (for any $p \geq 1$) of its singular value is equal to the gradient magnitude. Therefore, all the regularizers of the form (6.8) are valid vectorial TV extensions. Next, we show that the TV_p regularizers satisfy the following invariance properties, which are essential for any regularizer applied on flow fields.

Proposition 6. *The regularizer* TV_p*, defined as in* (6.8)*, is invariant under translation, scaling (up to a multiplicative factor), and rotation, where the rotation of a flow field* $\mathbf{f}$ *by some orthogonal matrix* ξ *is given by* $\mathbf{f} \mapsto \xi^{\mathrm{T}}\mathbf{f}(\xi\, \cdot)$*.*

Proof. First, we observe that the Fourier transform of the Jacobian operator is

$$\mathcal{F}\{\mathbf{J}\mathbf{f}\}(\boldsymbol{\omega}) = \hat{\mathbf{f}}(\boldsymbol{\omega})\mathrm{j}\boldsymbol{\omega}^{\mathrm{T}},$$

where $\hat{\mathbf{f}}(\boldsymbol{\omega}) = (\hat{f}_1(\boldsymbol{\omega}), \ldots, \hat{f}_d(\boldsymbol{\omega}))$ is the Fourier transform of $\mathbf{f}$. For the scaling invariance, we note that the scaling operator $\mathrm{S}_a : \mathbf{f} \mapsto \mathbf{f}(\cdot/a)$ commutes with the

Jacobian (up to a multiplicative constant) as it holds that

$$\begin{aligned}\mathcal{F}\{\mathbf{J}\{\mathrm{S}_a\mathbf{f}\}\}(\boldsymbol{\omega}) &= |a|^d\hat{\mathbf{f}}(a\boldsymbol{\omega})\mathrm{j}\boldsymbol{\omega}^{\mathrm{T}}\\ &= (1/a)\left(|a|^d\hat{\mathbf{f}}(a\boldsymbol{\omega})\mathrm{j}(a\boldsymbol{\omega})^{\mathrm{T}}\right)\\ &= (1/a)\,\mathcal{F}\{\mathrm{S}_a\{\mathbf{J}\mathbf{f}\}\}(\boldsymbol{\omega}).\end{aligned}$$

Since the Schatten p-norms are 1-homogeneous functions, we obtain that

$$\begin{aligned}\mathrm{TV}_p(\mathrm{S}_a\mathbf{f}) &= \frac{1}{|a|}\int_{\mathbb{R}^d}\|\{\mathbf{J}\mathbf{f}\}(\boldsymbol{x}/a)\|_{\mathcal{S}_p}\,\mathrm{d}\boldsymbol{x}\\ &= |a|^{d-1}\int_{\mathbb{R}^d}\|\mathbf{J}\mathbf{f}(\boldsymbol{q})\|_{\mathcal{S}_p}\,\mathrm{d}\boldsymbol{q}\\ &= |a|^{d-1}\,\mathrm{TV}_p(\mathbf{f}),\end{aligned}$$

where the second equality follows from a simple change of variables.

As for rotation by a matrix ξ, we have (R_ξ is the $d\times d$ rotation matrix)

$$\begin{aligned}\mathcal{F}\{\mathbf{J}\{\mathrm{R}_\xi\mathbf{f}\}\}(\boldsymbol{\omega}) &= \xi\hat{\mathbf{f}}(\xi\boldsymbol{\omega})\mathrm{j}\boldsymbol{\omega}^{\mathrm{T}}\\ &= \xi\hat{\mathbf{f}}(\xi\boldsymbol{\omega})\mathrm{j}\boldsymbol{\omega}^{\mathrm{T}}\xi^{\mathrm{T}}\xi\\ &= \xi\hat{\mathbf{f}}(\xi\boldsymbol{\omega})\mathrm{j}(\xi\boldsymbol{\omega})^{\mathrm{T}}\xi\\ &= \mathcal{F}\{\mathrm{R}_\xi\{\mathbf{J}\mathbf{f}\}\xi\}(\boldsymbol{\omega}).\end{aligned}$$

Then, we write that

$$\begin{aligned}\mathrm{TV}_p(\mathrm{R}_\xi\mathbf{f}) &= \int_{\mathbb{R}^d}\|\mathbf{J}\{\mathrm{R}_\xi\mathbf{f}\}(\boldsymbol{x})\|_{\mathcal{S}_p}\,\mathrm{d}\boldsymbol{x}\\ &= \int_{\mathbb{R}^d}\left\|\xi^{\mathrm{T}}\{\mathbf{J}\mathbf{f}\}(\xi\boldsymbol{x})\xi\right\|_{\mathcal{S}_p}\,\mathrm{d}\boldsymbol{x}\\ &= \int_{\mathbb{R}^d}\|\mathbf{J}\mathbf{f}(\xi\boldsymbol{x})\|_{\mathcal{S}_p}\,\mathrm{d}\boldsymbol{x},\end{aligned}$$

since the Schatten norms are unitarily invariant [151]. Now, applying a change of variable $\boldsymbol{q}=\xi\boldsymbol{x}$, with $\mathrm{d}\boldsymbol{q}=|\det\xi|\mathrm{d}\boldsymbol{x}=\mathrm{d}\boldsymbol{x}$, we arrive at the desired result:

$$\mathrm{TV}_p(\mathrm{R}_\xi\mathbf{f})=\mathrm{TV}_p(\mathbf{f})$$

Translation invariance is straight forward to show by using a change of variable. □

6.4.1 Connections with the Existing Methods

Having built the foundations of our regularizers, we now point out some interesting connections with the existing versions of vectorial TV regularizers. Essentially, we shall be investigating the most popular cases of the family in the sequel. That being said, let us start our consideration by setting $p = 2$ in (6.8). In this case, one operates with $\mathcal{S}_2$-norm which is also called the Frobenius norm.

Proposition 7. *Let* TV_p *defined as in* (6.8). *Then, for* $p = 2$*, we recover the VTV [148] given in* (6.7).

Proof. The proof is immediate. We write that

$$\begin{aligned}\mathrm{TV}_2(\mathbf{f}) &= \int_{\mathbb{R}^d} \|\mathbf{Jf}(\boldsymbol{x})\|_F \,\mathrm{d}\boldsymbol{x} \\ &= \int_{\mathbb{R}^d} \left(\sum_{j=1}^{d} \sigma_j(\mathbf{Jf}(\boldsymbol{x}))^2 \right)^{1/2} \mathrm{d}\boldsymbol{x} \\ &= \int_{\mathbb{R}^d} \left(\sum_{j=1}^{d} \mathrm{eig}(\mathbf{J}^{\mathrm{T}}\mathbf{Jf}(\boldsymbol{x})) \right)^{1/2} \mathrm{d}\boldsymbol{x} \\ &= \int_{\mathbb{R}^d} \mathrm{trace}\big(\mathbf{J}^{\mathrm{T}}\mathbf{Jf}(\boldsymbol{x})\big)^{1/2} \mathrm{d}\boldsymbol{x} \\ &= \int_{\mathbb{R}^d} \left(\sum_{j=1}^{d} \|\nabla f_j(\boldsymbol{x})\|_2^2 \right)^{1/2} \mathrm{d}\boldsymbol{x} \\ &= \mathrm{VTV}(\mathbf{f}),\end{aligned}$$

where $\|\cdot\|_F$ is the Frobenius norm. □

As a second case, we consider $p = \infty$ and use $\mathcal{S}_\infty$-norm (also called the spectral norm). We then obtain the so-called *natural vectorial* TV (TV_J) of [149].

Proposition 8. *Let* TV_p *defined as in* (6.8). *Then, for* $p = \infty$*, we recover the* TV_J *defined as*

$$\mathrm{TV}_\infty(\mathbf{f}) = \int_{\mathbb{R}^d} \|\mathbf{Jf}(\boldsymbol{x})\|_S \,\mathrm{d}\boldsymbol{x} = \mathrm{TV}_J(\mathbf{f}), \tag{6.10}$$

with $\|\cdot\|_S$ *being the spectral norm.*

Proof. The result is attained by definition. See Proposition 3.1 in [149]. □

6.4.2 Nuclear Total Variation

We see that the introduced TV_p family provides key links with the already-known methods. Based on our preceding progression, it is further possible to identify another vectorial extension of TV by choosing $p = 1$.We call this regularization functional the *nuclear total variation* (TVN); it is given by

$$\mathrm{TV}_1 = \int_{\mathbb{R}^d} \|\mathbf{Jf}\,(\mathbf{x})\|_N \,\mathrm{d}\mathbf{x} = \mathrm{TVN}\,(\mathbf{f}), \tag{6.11}$$

where $\|\cdot\|_N$ is the nuclear norm.

From definition (6.11), we observe that TVN exerts a coupling between the flow field components by imposing an ℓ_1-penalty on the singular values of the Jacobian. This leads to the deduction that TVN promotes flow field reconstructions with sparse singular values for the Jacobian at each spatial location. To see why the nuclear norm is expected to be a better choice than the other norms, we first need to understand more the attributes enforced on the reconstructed flow. Let us first state the following which are relevant to gain the desired understanding.

Definition 7. *Let* g *be a scalar-valued function whose domain is convex. The convex envelope of* g *(on its domain) is defined as the largest convex function* g_{env} *such that* $g_{\mathrm{env}}(\boldsymbol{x}) \leq g(\boldsymbol{x})$ *for all* $\boldsymbol{x} \in \mathrm{dom}(g)$.

Theorem 5 (Convex envelop of rank [152])**.** *Let* $g(\mathbf{A}) = \mathrm{rank}(\mathbf{A})$ *and* $\mathrm{dom}(g) = \{\mathbf{A} \in \mathbb{R}^{N_1 \times N_2} | \|\mathbf{A}\|_S \leq M\}$. *Then the convex envelope of* g *on its domain is given by the scaled-form of the nuclear norm that is* $g_{\mathrm{env}}(\mathbf{A}) = \frac{1}{M}\|\mathbf{A}\|_N$.

First of all, as a result of imposing sparsity on the singular values of the Jacobian, we anticipate that the variation at the dominant orientation (this is expected to be the flow itself) will be kept, whereas the small variations (they are expected to be the noise) will be reduced. In other words, directional features at the dominant orientation are preserved in the reconstructions. Secondly, we see that TVN regularization favors flow field reconstructions where the Jacobian matrix at each location is of low rank as a result of Theorem 5. We note that the rank of the Jacobian (i.e., the number of linearly independent columns) is directly related to the gradient vector of each component since they form the columns of the Jacobian matrix. TVN promotes solutions for which the gradient vectors of the flow components are aligned, which also implies that directional behavior of the flow is preserved. What is more important is that TVN preserves the discontinuities at the flow boundaries since Jacobian matrix of low rank is also appropriately satisfied at fluid interfaces. This behavior is illustrated in Figure 6.1.

The discussed properties of TVN make it convenient for applications in blood flow imaging as we shall demonstrate in our numerical experiments. We note that nuclear norm minimization is frequently used for rank minimization and matrix completion problems on account of Theorem 5 [153]. Before, we proceed with the derivation of the optimization algorithm, we would like to comment on the use of TVN for multi-channel images.

TVN for Multi-Channel Images

Let us consider a two-dimensional image with two channels that is represented by the vector function $\mathbf{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), f_2(\boldsymbol{x}))$. Knowing that TVN encourages the gradient vectors of each channel to be aligned (i.e., low rank Jacobian), we gain a straightforward intuition on the implications of using such regularizer.

For illustrative purposes, we consider a spatial location that is located on an edge in f_1. In the present case, TVN regularization encourages ∇f_2 to be parallel to ∇f_1. This means that it encourages an edge in f_2 that is correlated with the one in f_1 in terms of the location. Also, this behavior is not strict as $\nabla f_2 = 0$ (i.e., f_2 is smooth at the considered location) is also linearly dependent on ∇f_1 (see Figure 6.2). This qualifies TVN as suitable regularizer for multi-channel images if common edge locations are desired. We remark that TVN can be derived as a special case of the regularization family introduced in [150], which penalizes the rooted eigenvalues of the structure tensor of an image. In the specific context of multi-channel image restoration problems, we also note that TVN has been independently proposed and investigated in [154].

6.5 Reconstruction Algorithm

In the sequel, we consider the discrete version of TVN and describe a fast algorithm for solving (6.3). Specifically, we obtain the denoised flow field as the minimizer of the following strictly convex energy functional

$$\mathbf{f}^\star = \arg\min_{\mathbf{f}} \tfrac{1}{2} \left\|\mathbf{y} - \mathbf{f}\right\|_2^2 + \tau \left\|\mathbf{J}\mathbf{f}\right\|_{1,N}, \tag{6.12}$$

where $\mathbf{f}$ has been vectorized and is of size $\mathbb{R}^{k\cdot d}$, with k denoting the cardinality of the discrete index set (*e.g.* the number of voxels in 3-D). Further, we introduce

$$\left\|\mathbf{J}\mathbf{f}\right\|_{1,N} = \sum_{j=1}^{k} \left\|\left(\mathbf{J}\mathbf{f}\right)_j\right\|_N$$

as a compact notation for the discrete TVN that employs the mixed ℓ_1-nuclear norm and the discrete Jacobian $\mathbf{J} : \mathbb{R}^{k\cdot d} \mapsto \mathbb{R}^{k\times d\times d} \triangleq \mathcal{X}$. Invoking [150, Lemma 1] and considering Legendre-Fenchel duality [43, 39], we derive the following dual definition for the discrete TVN:

$$\begin{aligned}\mathrm{TVN}\left(\mathbf{f}\right) &= \max_{\boldsymbol{\xi}\in\mathcal{X},\boldsymbol{\xi}_j\in\mathcal{B}_S^{d\times d}} \langle \mathbf{f}\,,\, \mathbf{J}^*\boldsymbol{\xi}\rangle \\ &= \max_{\boldsymbol{\xi}\in\mathcal{X},\boldsymbol{\xi}_j\in\mathcal{B}_S^{d\times d}} \sum_{i=1}^{d}\sum_{j=1}^{k} \left(f_i\right)_j \left(\nabla^*\boldsymbol{\xi}_i\right)_j \end{aligned} \tag{6.13}$$

where ∇^* and $\mathbf{J}^*$ are the adjoints of the discrete gradient and Jacobian, respectively. In (6.13), $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k) \in \mathcal{X}$ is a dual variable with $\left(\boldsymbol{\xi}_i\right)_j$ referring to the ith row of the matrix $\boldsymbol{\xi}_j \in \mathbb{R}^{d\times d}$, while $\mathcal{B}_S^{d\times d} = \left\{\mathbf{A} \in \mathbb{R}^{d\times d} : \left\|\mathbf{A}\right\|_S \leq 1\right\}$ is the spectral unit-norm ball.

Using the min-max theorem, we rewrite (6.12) in the equivalent form

$$\max_{\boldsymbol{\xi}\in\mathcal{X},\boldsymbol{\xi}_j\in\mathcal{B}_S^{d\times d}} \min_{\mathbf{f}} \tfrac{1}{2} \left(\left\|\mathbf{f} - \mathbf{u}\right\|_2^2 + \left\|\mathbf{y}\right\|_2^2\right) - \left\|\mathbf{u}\right\|_2^2 \tag{6.14}$$

where $\mathbf{u} = (\mathbf{y} - \tau\mathbf{J}^*\boldsymbol{\xi})$. Based on this development, the solution is derived in closed form as $\mathbf{f}^\star = (\mathbf{y} - \tau\mathbf{J}^*\boldsymbol{\xi}^\star)$, where $\boldsymbol{\xi}^\star$ corresponds to the maximizer

$$\boldsymbol{\xi}^\star = \arg\max_{\boldsymbol{\xi}\in\mathcal{X},\boldsymbol{\xi}_j\in\mathcal{B}_S^{d\times d}} \left\|\mathbf{y}\right\|_2^2 - \left\|\mathbf{y} - \tau\mathbf{J}^*\boldsymbol{\xi}\right\|_2^2. \tag{6.15}$$

Algorithm 7: Proposed TVN denoising algorithm according to (6.12)

```
input  : Noisy field $\mathbf{y}$, regularization parameter $\tau > 0$, and the projection
         onto $\mathcal{B}_{S^{d\times d}}$ projectionDualNormBall()
output: Denoised flow field $\mathbf{f}^\star$

$\boldsymbol{\psi}_1 \leftarrow \mathbf{0}$, $\boldsymbol{\xi}_0 \leftarrow \mathbf{0}$, $t_1 \leftarrow 1$, $n \leftarrow 1$;
repeat
    $\boldsymbol{\xi}_n \leftarrow$ projectionDualNormBall($\boldsymbol{\psi}_n + (1/(12\tau))\mathbf{J}\left(\mathbf{y} - \tau\mathbf{J}^*\boldsymbol{\psi}_n\right)$);
    $t_{n+1} \leftarrow \frac{1+\sqrt{1+4t_n^2}}{2}$;
    $\boldsymbol{\psi}_{n+1} \leftarrow \boldsymbol{\xi}_n + \left(\frac{t_n - 1}{t_{n+1}}\right)(\boldsymbol{\xi}_n - \boldsymbol{\xi}_{n-1})$;
    $n \leftarrow n + 1$;
until stopping criteria
return $(\mathbf{y} - \tau\mathbf{J}^*\boldsymbol{\xi}_{n-1})$;
```

Since the function in (6.15) is smooth with well-defined gradient we compute $\boldsymbol{\xi}^\star$ using an accelerated projected gradient ascent based on Nesterov's method [47]. The details of the approach are given in Algorithm 1. We note that in Algorithm 1 the operation Π_S refers to the independent projection of each of the k matrices onto the $\mathcal{B}_S^{d\times d}$ unit ball. To perform this operation, we rely on the result of [150, Proposition 1] which provides a connection between matrix and vector projections. For a matrix $\mathbf{X}$ with singular value decomposition $\mathrm{SVD}(\mathbf{X}) = \mathbf{USV}^{\mathrm{T}}$, the projection is performed as $\Pi_{\mathrm{S}}(\mathbf{X}) = \mathbf{XVS}^{+}\tilde{\mathbf{S}}\mathbf{V}^{\mathrm{T}}$. Here, $\mathbf{S}^{+}$ is the pseudo-inverse of $\mathbf{S}$ and $\tilde{\mathbf{S}}$ is the diagonal matrix which contains the projected singular values of $\mathbf{S}$ onto the ℓ_∞ unit-norm ball. This projection sets to one the singular values that exceed this value while leaving the rest untouched. It is noteworthy that SVD can be performed very efficiently for 3D flow fields.

6.6 Experiments

Based on the above developments, we now conduct experiments for simulated and real data, where all visualizations are generated with ParaView (Kitware Inc.) [155]. In all the experiments, the flow fields are extended using periodic boundary condi-

Table 6.1: Denoising performance of several regularization approaches: Simulation results for the set of simulated flow fields considered in this chapter. Numbers are given in decibel unit.

	Input SNR	Reconstruction SNR				
		CDR	TV_S	VTV	TV_J	TVN
Tube	0	14.47	**19.66**	17.64	16.42	19.19
	10	22.67	**27.09**	24.62	23.37	26.47
	20	30.72	**34.89**	31.96	30.75	34.20
	30	39.29	**43.38**	40.09	39.05	42.89
Torus	0	13.49	17.59	16.25	14.63	**18.07**
	10	20.37	25.30	23.63	22.24	**26.08**
	20	29.17	33.66	31.78	30.48	**34.63**
	30	38.28	42.34	40.40	39.42	**43.60**
Gradient	0	15.19	19.51	19.64	18.43	**20.77**
	10	23.24	26.82	27.72	26.81	**28.39**
	20	31.86	35.30	36.54	35.77	**36.80**
	30	41.07	44.40	45.73	45.03	**45.89**
Blood	0	12.70	15.80	15.33	13.86	**16.55**
	10	19.34	22.44	22.43	21.35	**23.45**
	20	27.44	29.92	30.35	29.62	**30.92**
	30	36.37	38.42	39.01	38.44	**39.24**

tions and consider regularization of volumetric ($d = 3$) flows.

6.6.1 Synthetic Data

We generate a dataset composed of four different three-dimensional phantom models (see Figure 6.3). The measurements are obtained by degrading the data with different levels of additive white Gaussian noise. The denoising performance of our TVN regularization is compared against the following methods:

1. Curl and divergence regularization (CDR): We use the regularization given

in (6.4) that penalizes the ℓ_1-norms of the curl and the divergence (thus rotational and incompressible features) of the field.

2. Separable TV regularization (TV_S): This method refers to the regularizer given in (6.6). It applies scalar TV regularization separately to each flow component. Hence, it treats these components independently and does not correlate them.

3. Vectorial TV regularization (VTV): We refer to the regularization given in (6.7). We note that the method is recovered within our framework by setting $p = 2$ in (6.8). It introduces a coupling between the flow components unlike TV_S.

4. Natural vectorial TV regularization (TV_J): We adopt the regularization given in (6.10). Similar to VTV, the method is derived by choosing $p = \infty$ in (6.8). The method penalizes the sum of the maximum singular values of the Jacobian over the whole set of voxels.

For all methods, we use the same optimization algorithm that combines the duality arguments with Nesterov's method [47]. Note that such algorithms have been developed in [156] for CDR, in [49] for TV_S, in [157] for VTV, and in [149] for TV_J. In all cases, the stopping criterion is set to either reaching a relative ℓ_2-normed difference of 10^{-4} between two successive estimates, or a maximum of 500 iterations. At each experiment, for all of the algorithms, the regularization parameter is optimized for the best-possible SNR performance using an oracle.

By inspecting the results given in Table 6.1, we see that the TVN outperforms CDR and the vectorial TV methods for most of the simulated fields considered in the experiments. This demonstrates the ability of our regularization scheme to preserve the discontinuities at the boundaries (see Figure 7.1). One surprising result provided by the experiments is that the performance of TV_S is highly competitive. Even though, this model simply does not capture the vectorial nature of the flow data, it achieves the best SNR performance for the Tube phantom. This is explained by the fact that Tube is a separable phantom (superposition of 2D flow fields) by construction. We also see that TV_S is outperformed by TVN for the Torus flow (which essentially encapsulates Tube with a circular flow). This observation further supports that TV_S provides better denoising results (compared to the other algorithms) for Tube due to the particular structure of the underlying flow.

6.6.2 Real Data

As a supplement to our *in silico* experiments, we considered a multidirectional phase-contrast MRI dataset in the region between the proximal and thoracic aorta of a 25-year-old healthy male volunteer. PC MRI data was acquired with a sagittal oblique 3D slab covering the entire aorta, using a navigator-gated, ECG-gated RF-spoiled gradient echo (GRE) sequence [158]. The dataset was acquired on a 3T clinical MR scanner (MAGNETOM Trio, Siemens AG, Healthcare Sector, Erlangen, Germany). The sequence was motion compensated and the following imaging parameters were used: TR/TE 5.2/2.59; flip angle 15°; velocity encoding 150 cm/sec; matrix 224×138×24; field of view 450 mm, and acquired voxel size 2.0 mm^3.

To regularize the data, we assess for the time point of peak ascending aortic flow. We use TVN with 250 iterations where we manually calibrate the regularization parameter. Streamlines are generated for both pre- and post-regularization states by using identical number of seed points, seed area and integration length.

A qualitative examination of Figure 6.5 shows that the amount of streamline artifacts is decreased especially in the ventral side of the arch. Velocity field discontinuities are reduced with direct positive impact on the estimation of the first order velocity derivatives that are required for the calculation of important flow parameters such as vorticity and flow helicity [158, 159]. Descending-aorta velocity field becomes less noisy and more coherent, similar to what is expected physiologically. Furthermore all the above improvements are achieved without affecting the magnitude of the velocity field. Instantaneous ascending aortic flow for the initial dataset is 415ml/s and is decreased only by 5 ml/s for the regularized dataset corresponding to a minute decrease of 1.2%. These aspects suggest that our method is beneficial for the visualization of aortic hemodynamic phenomena without affecting the magnitude of the field.

6.7 Discussion

In this chapter, our main goal has been to extend TV regularization to vector fields. For this particular goal, we have considered the problem of flow field denoising and have proposed a regularizer that penalizes the nuclear norm of the Jacobian of the field. To illustrate the efficiency of our regularizer, we first conducted denoising experiments on different phantom data involving rapid transitions at the flow boundaries. We observed that the proposed method is superior in terms of SNR performance (up to 1.3 dB) than the curl-divergence regularizer and the existing vectorial extensions of TV in general. We also used our reconstruction algorithm for enhancing streamline visualizations of a real phase-contrast MRI recording.

As a result of our numerical experiments, we have seen that efficient algorithms are designed by extending the principles of scalar image reconstruction presented in the earlier chapters of the thesis. We have also seen that the simplest TV-based vector field regularizer (namely, separable TV) can provide the best denoising performance in certain cases. Consistent with our previous remarks in Chapter 4, this observation emphasizes the point that the regularization functional needs to be chosen in accordance with the features of the underlying vector field. In simpler words, prior to any attempt for variational reconstruction of such data (especially in practical configurations), it is necessary to evaluate whether the characteristics encouraged by the regularizer match the structure of the data, which is to be processed, or not.

Our design efforts in this chapter have been driven by deterministic principles. Similar to case of TV regularization, exact continuous-domain stochastic justification of TVN is not straightforward. The reassuring news is that such an interpretation is possible in the case of a quadratic regularization (as in the Tikhonov case). Based on our explanations in Section 4.5 and Proposition 7, we establish the following relation:

$$\begin{aligned}
\mathrm{TV}_2^2\left(\mathbf{f}\right) &= \int_{\mathbb{R}^d} \|\mathbf{Jf}\left(\boldsymbol{x}\right)\|_F^2 \,\mathrm{d}\boldsymbol{x} = \sum_{j=1}^{d} \int_{\mathbb{R}^d} \|\nabla f_j(\boldsymbol{x})\|_2^2 \mathrm{d}\boldsymbol{x} \\
&= \sum_{j=1}^{d} \|\nabla f_j\|_2^2 = \sum_{j=1}^{d} \|(-\Delta)^{1/2} f_j\|_2^2,
\end{aligned}$$

We see that (once again under the decoupling simplification) the quadratic member of the family TV_2^2 computes the MAP/MMSE estimation of a self-similar vector process whose components are independent. For both of these components, the underlying continuous-domain innovation is Gaussian.

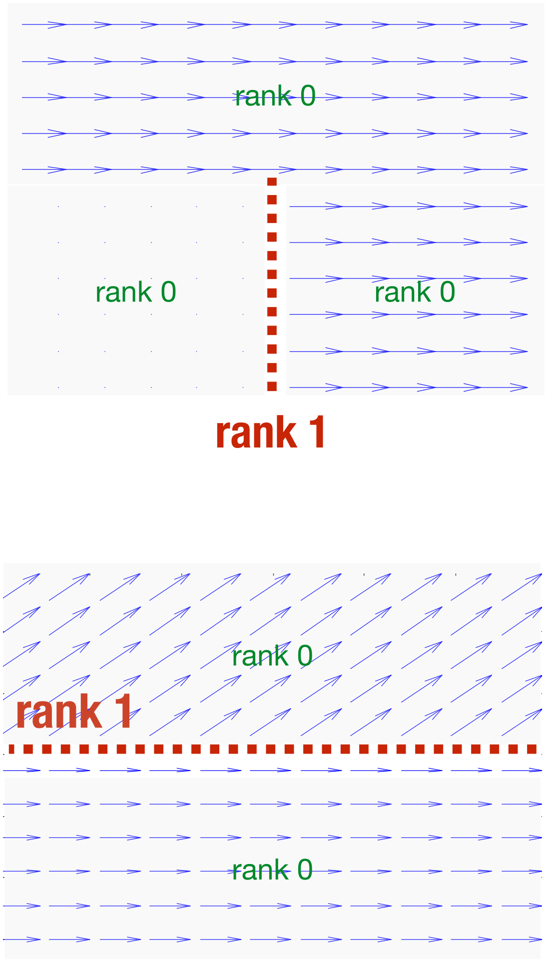


Figure 6.1: Illustrative examples of two-dimensional flow fields where the rank of the Jacobian is less than 2. Note that the rank of the Jacobian of the field is 1 at the flow boundaries indicated by dashed lines.

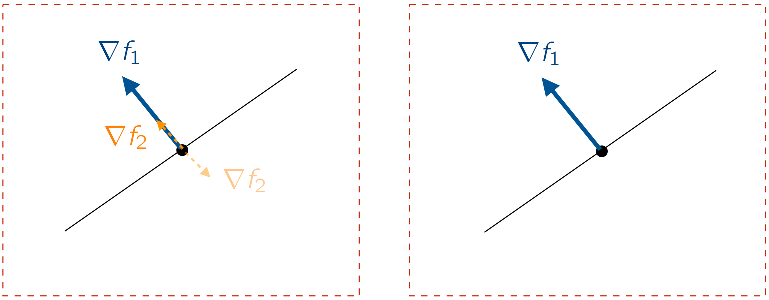


Figure 6.2: Illustrative examples of three configurations of ∇f_2 being linearly dependent on the ∇f_1. We note that in the last case $\nabla f_2 = 0$.

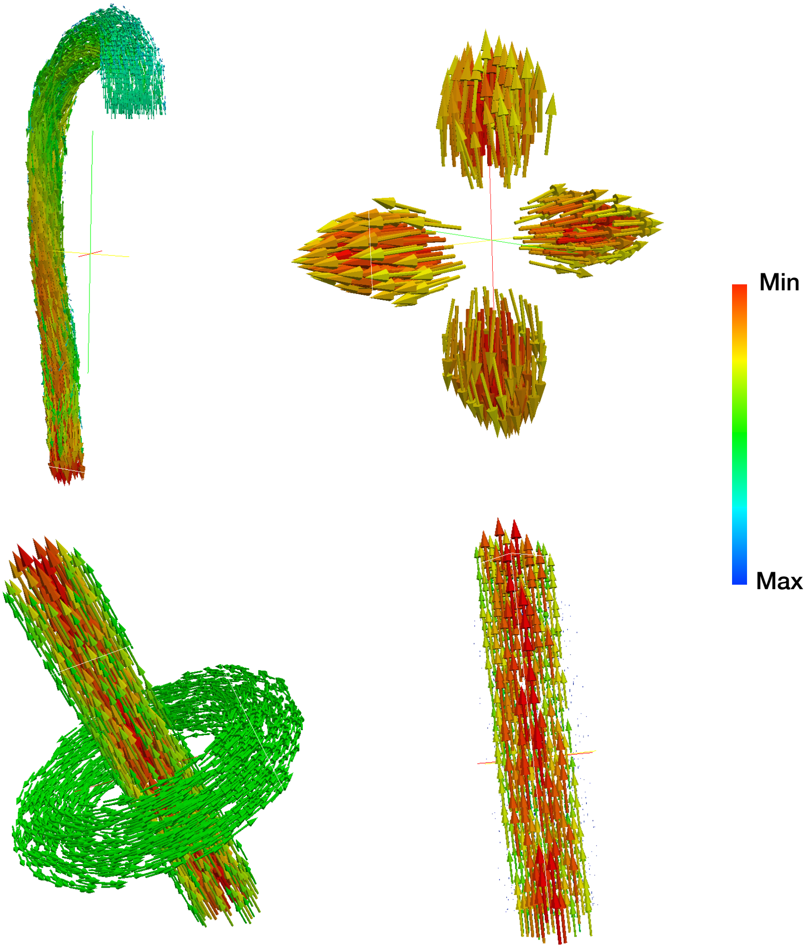


Figure 6.3: The set of cell images representing the ground-truth spatial phase maps used in the simulations. From top-left to bottom-right, they are referred to as, 1) Blood flow, 2) Gradient, 3) Torus, and 4) Tube.

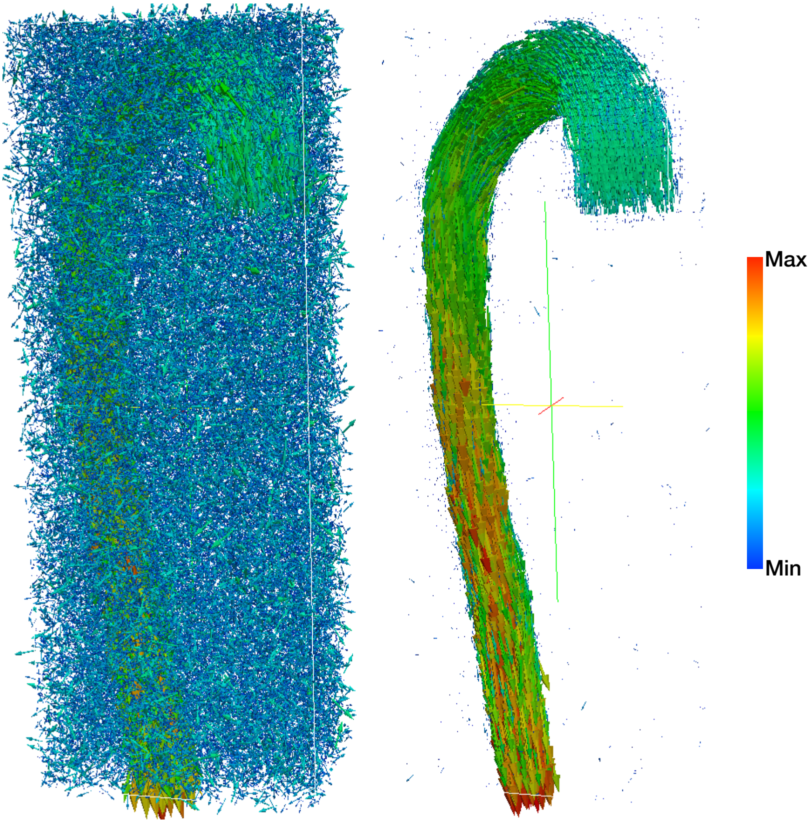


Figure 6.4: Denoising of the simulated blood flow: noisy flow (left, SNR= 0 dB), and the denoised flow (right, SNR= 16.55 dB) by using the proposed method.

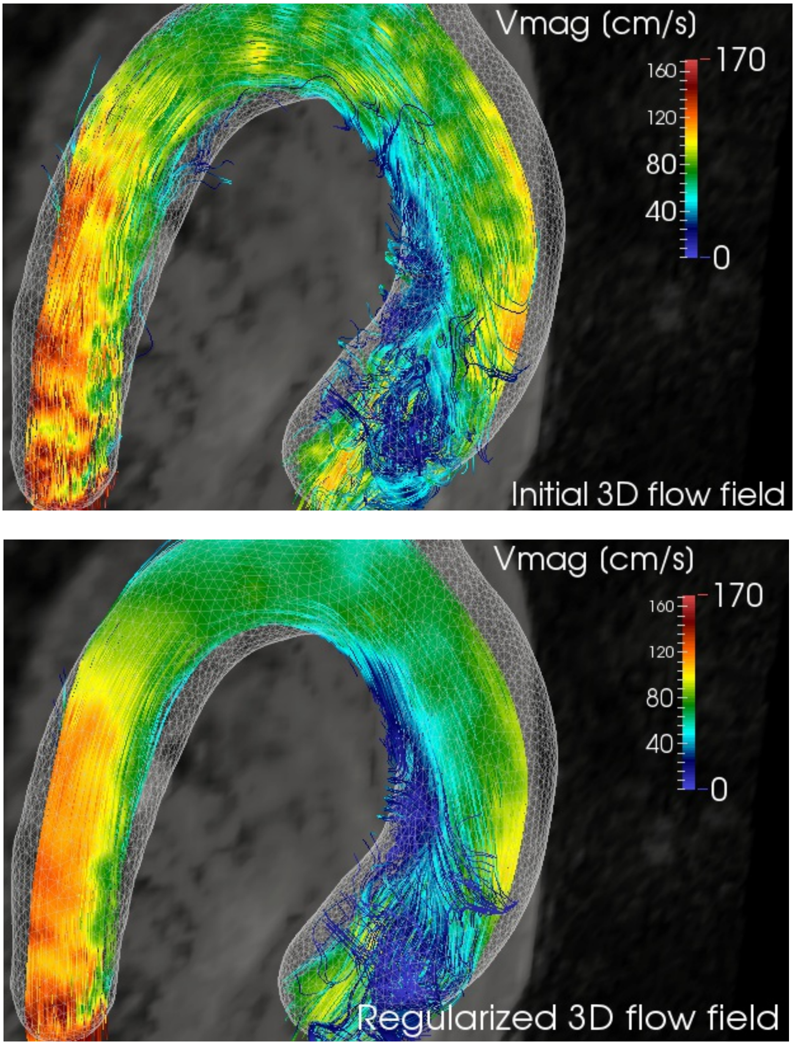


Figure 6.5: Enhancement of streamline visualizations of a real phase-contrast MRI recording: Original aortic blood flow data in the aortic arch (top) and the data after processing with the proposed method (bottom). See text for further details of the experiment.

Chapter 7

Non-Iterative Model-Based Sparse Flow Reconstruction

7.1 Overview

In this chapter,[1] we focus on the regularization of incompressible flows that have zero divergence. In particular, we propose an efficient construction of wavelet frames in any number of dimensions that are divergence-free. The key mathematical concept of the proposed method is to apply the Leray projector, which is scale-invariant, to a standard wavelet frame. We prove that the projected wavelets retain the basic characteristics (decay rate and order of vanishing moments) of the initial wavelets. Since the Leray projector is also shift-invariant, it is defined as a Fourier multiplier, and our construction is implemented efficiently using the fast Fourier transform. Based on these developments, we then formulate a wavelet-based denoising algorithm that is solved directly. In order to illustrate the practicality of the method, we present vector field denoising experiments, where our simulations show superior performance compared to the currently used divergence-free wavelet designs.

---

[1] This chapter is based on our paper [160].

7.2 Introduction

As we have mentioned in Chapter 6, phase-contract MRI (PC MRI) provides three-dimensional and time-resolved blood flow measurements [12], yet often suffers from noise-induced artifacts. The artifacts are more pronounced when the spatial Fourier space is heavily sub-sampled to increase the time resolution. To reduce these inaccuracies, one designs denoising methods based on the physical principles governing the underlying flow, which are not directly imposed in the raw acquisitions. In this type of approaches, an often exploited property is the physical prior knowledge of the blood flow being *incompressible* in large arteries [146].

The mass conservation principle implies that the velocity fields representing incompressible fluid flows have zero divergence. Therefore, these physical systems are often represented using divergence-free wavelets. Not only are they useful for the multi-scale analysis of such phenomena, but also fundamentally prominent for studying the incompressible solutions of the Navier-Stokes equations [161]. Accordingly, they are used for the numerical simulation of incompressible flows [162].

The relevance of divergence-free wavelets for PC MRI imaging has been considered in [163]. Modeling the blood flow incompressible, noise-like errors are reduced by identifying (and eventually suppressing) non-divergence-free components. Similar ideas have recently been applied to volumetric particle image velocimetry (vPIV), which is also capable of measuring three-dimensional velocity fields. In vPIV imaging of incompressible flow, the effects of the measurement noise is removed by eliminating spurious divergence through a redundant atomic signal decomposition [164].

The design of divergence-free wavelets was considered first by Battle and Federbush [165] and Lemarié-Rieusset [166]. The development in [165] generated orthogonal wavelets (in the 2- and 3-dimensional cases) with exponential decay. On the other hand, the construction in [166] resulted in compactly-supported biorthogonal wavelets. Their construction is based on two pairs of wavelets and scaling functions satisfying certain integration and differentiation relations. An efficient (tensor-product based) implementation of these wavelets was developed by Deriaz and Perrier [167]. Stevenson [168] and Kadri-Harouna and Perrier [169] have extended these constructions to domains with free-slip boundary conditions. Other basis constructions have also been considered [170].

7.2.1 Contributions

All of the previous schemes are rather technical as they are concerned with divergence-free *wavelet basis* functions. In this chapter, we loosen this constraint by allowing *redundancy.*

Based on this relaxation, our contributions are listed as follows:

- We propose a simple yet elegant approach for constructing divergence-free wavelet frames. Our design involves projecting a wavelet frame onto the space of zero divergence vector-valued functions.
- We prove that the decay rate and order of vanishing moments of the initial wavelets are inherited by their divergence-free counterparts.
- We employ our framework for the vector field denoising problem both in the two- and three-dimensional settings. We show that our method achieves better denoising performance than its existing counterpart that is bi-orthogonal by design.

In the thesis, we shall concentrate on divergence-free *tight wavelet frames* since they are advantageous for applications. However our method applies for constructing divergence-free wavelet frames in a more general sense.

7.2.2 Outline

In the remainder of the chapter, we first provide the mathematical notions that are necessary (especially the Leray projection operator which is the key element in the construction) in Section 7.3. In Section 7.4, we explain the construction of divergence-free wavelet frames. This is followed by our proof regarding the decay rates of the designed wavelets. After providing the implementation details, we consider wavelet-based sparse denoising of incompressible vector fields and provide simulation results in Section 7.6.

7.3 Mathematical Framework

7.3.1 Basic Definitions

We denote the space of square integrable vector fields (or vector-valued functions) on $\mathbb{R}^d$ by $\left(L_2(\mathbb{R}^d)\right)^d$. A vector field $\boldsymbol{f} \in \left(L_2(\mathbb{R}^d)\right)^d$ is represented as $\boldsymbol{f} = (f_1\, f_2\, \ldots\, f_d)^{\mathrm{T}}$, where each component is a function in $L_2(\mathbb{R}^d)$. The inner product is defined as follows:

$$\begin{aligned}\langle \boldsymbol{f}, \boldsymbol{g} \rangle &= \sum_{n=1}^{d} \langle f_n, g_n \rangle \\ &= \sum_{n=1}^{d} \int_{\mathbb{R}^d} f_n(\boldsymbol{x}) g_n(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}. \end{aligned} \tag{7.1}$$

The Fourier transform of a vector field is the component-wise Fourier transform

$$\begin{aligned}\mathcal{F}\{\boldsymbol{f}\}(\boldsymbol{\omega}) &= \widehat{\boldsymbol{f}}(\boldsymbol{\omega}) \\ &= \left(\widehat{f}_1(\boldsymbol{\omega})\, \widehat{f}_2(\boldsymbol{\omega})\, \ldots\, \widehat{f}_d(\boldsymbol{\omega})\right)^{\mathrm{T}}, \end{aligned} \tag{7.2}$$

where

$$\widehat{f}_n(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} f_n(\boldsymbol{x}) \mathrm{e}^{-2\pi \mathrm{j} \boldsymbol{x}\cdot\boldsymbol{\omega}} \mathrm{d}\boldsymbol{x}. \tag{7.3}$$

The divergence of $\boldsymbol{f} \in \left(L_2(\mathbb{R}^d)\right)^d$ is given by

$$\begin{aligned}\operatorname{div} \boldsymbol{f} &= \nabla \cdot \boldsymbol{f} \\ &= \sum_{n=1}^{d} \frac{\partial f_n}{\partial x_n} \\ &= \left(\frac{\partial}{\partial x_1}\, \frac{\partial}{\partial x_2}\, \cdots\, \frac{\partial}{\partial x_d}\right) (f_1\, f_2\, \ldots\, f_d)^{\mathrm{T}}. \end{aligned} \tag{7.4}$$

We denote the collection of divergence-free, square-integrable vector fields as

$$\mathcal{H}(\mathbb{R}^d) = \left\{\boldsymbol{f} \in \left(L_2(\mathbb{R}^d)\right)^d : \operatorname{div} \boldsymbol{f} \in L_2(\mathbb{R}^d), \operatorname{div} \boldsymbol{f} = 0\right\}. \tag{7.5}$$

7.3.2 The Leray Projector

After setting up the basic mathematical notions, we now present the central element of our design that is the Leray projector. This operator $\mathcal{P} : \left(L_2(\mathbb{R}^d)\right)^d \to \left(L_2(\mathbb{R}^d)\right)^d$ is the unique orthogonal projector that maps a vector field to its divergence-free version [171]. It is defined in the Fourier domain as

$$\mathcal{F}\left\{\mathcal{P}\boldsymbol{f}\right\}(\boldsymbol{\omega}) = (\mathrm{I} - \mathrm{P}(\boldsymbol{\omega}))\widehat{\boldsymbol{f}}(\boldsymbol{\omega}), \tag{7.6}$$

where I is the $d \times d$ identity matrix and P is the $d \times d$ matrix

$$\mathrm{P}(\boldsymbol{\omega}) = |\boldsymbol{\omega}|^{-2}(\omega_1\, \omega_2\, \ldots\, \omega_d)^{\mathrm{T}}(\omega_1\, \omega_2\, \ldots\, \omega_d). \tag{7.7}$$

Proposition 9. *The Leray projector $\mathcal{P}$ is self-adjoint and is invariant to translation and scaling in the sense that*

$$\mathcal{P}\{\boldsymbol{f}(\cdot/a - \mathbf{b})\}(\mathbf{x}) = \mathcal{P}\{\boldsymbol{f}\}(\mathbf{x}/a - \mathbf{b}),$$

for any $\boldsymbol{f} \in \left(L_2(\mathbb{R}^d)\right)^d$, $a \in \mathbb{R}^+$, and $\mathbf{b} \in \mathbb{R}^d$.

The outcome of Proposition 9 is crucial as it implies that the Leray projector maps a wavelet to another wavelet. In the sequel, we shall use this for our construction of divergence-free wavelet frames.

7.4 Divergence-Free Wavelets

The construction of a band-limited mother wavelet ϕ for a tight wavelet frame

$$\left\{2^{jd/2}\phi\left(2^j\boldsymbol{x} - \boldsymbol{k}\right) : j \in \mathbb{Z}, \boldsymbol{k} \in \mathbb{Z}^d\right\} \tag{7.8}$$

of $L_2(\mathbb{R}^d)$ is straightforward, cf. Theorem 12.2.1 of [172]. Recall that a tight wavelet frame satisfies an energy preservation relationship

$$\|f\|^2_{L_2(\mathbb{R}^d)} = \sum_{j\in\mathbb{Z},\boldsymbol{k}\in\mathbb{Z}^d} \left|\left\langle f, 2^{jd/2}\phi\left(2^j\cdot - \boldsymbol{k}\right)\right\rangle\right|^2, \tag{7.9}$$

and it is dual to itself

$$f(\boldsymbol{x}) = \sum_{j\in\mathbb{Z},\boldsymbol{k}\in\mathbb{Z}^d} \left\langle f, 2^{jd/2}\phi\left(2^j\cdot -\boldsymbol{k}\right)\right\rangle 2^{jd/2}\phi\left(2^j\boldsymbol{x}-\boldsymbol{k}\right). \tag{7.10}$$

From a scalar-valued frame, we can easily construct band-limited tight frames of $\left(L_2(\mathbb{R}^d)\right)^d$. The vector-valued frames are obtained as follows. For $n \in \mathbb{N}_{\le d}$, let $\boldsymbol{e}_n$ denote the nth unit vector in $\mathbb{R}^d$.

Proposition 10. *The collection*

$$\left\{\boldsymbol{\psi}_{j,\boldsymbol{k},n} := 2^{jd/2}\phi\left(2^j\cdot -\boldsymbol{k}\right)\boldsymbol{e}_n : j\in\mathbb{Z}, \boldsymbol{k}\in\mathbb{Z}^d, n\in\mathbb{N}_{\le d}\right\} \tag{7.11}$$

is a vector-valued tight frame of $\left(L_2(\mathbb{R}^d)\right)^d$.

The next step is to derive a divergence-free tight frame by applying the Leray projector.

Proposition 11. *The collection*

$$\left\{\mathcal{P}\boldsymbol{\psi}_{j,\boldsymbol{k},n} : j\in\mathbb{Z}, \boldsymbol{k}\in\mathbb{Z}^d, n\in\mathbb{N}_{\le d}\right\} \tag{7.12}$$

is a tight frame of $\mathcal{H}(\mathbb{R}^d)$. *In particular,*

$$\boldsymbol{f}(\boldsymbol{x}) = \sum_{j\in\mathbb{Z},\boldsymbol{k}\in\mathbb{Z}^d,n\in\mathbb{N}_{\le d}} \left\langle \boldsymbol{f}, \mathcal{P}\boldsymbol{\psi}_{j,\boldsymbol{k},n}\right\rangle \mathcal{P}\boldsymbol{\psi}_{j,\boldsymbol{k},n}. \tag{7.13}$$

Proof. Let $\boldsymbol{f}\in\mathcal{H}(\mathbb{R}^d)$. Since the Leray projector is self-adjoint, we have

$$\begin{aligned}
\|\boldsymbol{f}\|^2_{\mathcal{H}(\mathbb{R}^d)} &= \|\boldsymbol{f}\|^2_{(L_2(\mathbb{R}^d))^d} \\
&= \|\mathcal{P}\boldsymbol{f}\|^2_{(L_2(\mathbb{R}^d))^d} \\
&= \sum_{j\in\mathbb{Z},\boldsymbol{k}\in\mathbb{Z}^d,n\in\mathbb{N}_{\le d}} \left|\left\langle \mathcal{P}\boldsymbol{f}, \boldsymbol{\psi}_{j,\boldsymbol{k},n}\right\rangle\right|^2 \\
&= \sum_{j\in\mathbb{Z},\boldsymbol{k}\in\mathbb{Z}^d,n\in\mathbb{N}_{\le d}} \left|\left\langle \boldsymbol{f}, \mathcal{P}\boldsymbol{\psi}_{j,\boldsymbol{k},n}\right\rangle\right|^2.
\end{aligned} \tag{7.14}$$

□

7.4.1 Decay estimates

An important property of our construction is that we preserve the advantageous properties of the initial wavelet ϕ. Both vanishing moments and decay rates remain unchanged.

Theorem 6. *Let ϕ be a differentiable function with vanishing moments of order $N \geq 1$ such that ϕ and its derivatives satisfy the decay estimates*

1. $|\phi(\boldsymbol{x})| \leq C(1+|\boldsymbol{x}|)^{-d-N+\epsilon}$
2. $|\mathrm{D}^{\boldsymbol{\alpha}}\phi(\boldsymbol{x})| \leq C(1+|\boldsymbol{x}|)^{-d-N-1+\epsilon}, \quad |\boldsymbol{\alpha}| = 1,$

for some $C > 0$ and $0 \leq \epsilon < 1$. Then the components of the divergence-free wavelets $\mathcal{P}\boldsymbol{\psi}_{0,\mathbf{0},n}$ have the same number of vanishing moments and similar decay to ϕ, i.e., for $n = 1, \dots, d$

$$\left(\boldsymbol{e}_n \cdot \mathcal{P}\boldsymbol{\psi}_{0,\mathbf{0},n}\right)(\boldsymbol{x}) \leq C(1+|\boldsymbol{x}|)^{-d-N+\epsilon'} \tag{7.15}$$

for some $0 \leq \epsilon' < 1$.

Proof. This result follows from the analysis of singular-integral operators that was presented in [173]. As each component is similar, we only consider $n = 1$. In the Fourier domain,

$$\begin{aligned}\mathcal{F}\left\{\mathcal{P}\boldsymbol{\psi}_{0,\mathbf{0},1}\right\}(\boldsymbol{\omega}) &= (\mathrm{I}-\mathrm{P}(\boldsymbol{\omega}))\boldsymbol{e}_1\widehat{\phi}(\boldsymbol{\omega}) \\ &= -\widehat{\phi}(\boldsymbol{\omega})\,|\boldsymbol{\omega}|^{-2}\sum_{n=1}^{d}\omega_1\omega_n.\end{aligned} \tag{7.16}$$

Since $\left||\boldsymbol{\omega}|^{-2}\sum_{n=1}^{d}\omega_1\omega_n\right| \leq 1$, the vanishing moments are preserved. Furthermore, $\sum_{n=1}^{d}\omega_1\omega_n$ is a homogeneous harmonic polynomial of degree 2, so we can decompose the multiplier in the spherical harmonics of degree 2 $\{Y_{2,m}\}_{m=1}^{d}$ as

$$|\boldsymbol{\omega}|^{-2}\sum_{n=1}^{d}\omega_1\omega_n = \sum_{m=1}^{d}c_m Y_{2,m}\left(\frac{\boldsymbol{\omega}}{|\boldsymbol{\omega}|}\right) \tag{7.17}$$

for some coefficients c_m. The decay estimates now follow from [174, Theorem 3.2]. □

7.4.2 Wavelet Implementation Details

The particular wavelet ϕ that we use for implementation is a Meyer-type mother wavelet. It is defined in the Fourier domain by a radial profile function h, i.e., $\widehat{\phi}(\boldsymbol{\omega}) = h(|\boldsymbol{\omega}|)$, where

$$h(r) = \begin{cases} \cos\left(2\pi q_m\left(r\right)\right), & \frac{1}{8} < r \leq \frac{1}{4} \\ \sin\left(2\pi q_m\left(\frac{r}{2}\right)\right), & \frac{1}{4} < r \leq \frac{1}{2} \\ 0, & \text{otherwise} \end{cases} \tag{7.18}$$

and q_m is a degree $2m+1$ polynomial satisfying the following: $q_m(1/8) = 0$, $q_m(1/4) = 0$, and $q_m^{(j)}(1/8) = q_m^{(j)}(1/4) = 0$ for $j = 1, \ldots, m$, cf. [175]. In our experiments, we set the parameter $m = 2$. Our divergence-free wavelets are constructed as described above. We generate the scalar-valued tight frame associated with ϕ; extend it to a vector-valued frame; and finally, apply the Leray projector to each frame element. As for the implementation of our frame, we follow the pyramid construction of [76]. Our method is efficient as it has a complexity of $\mathcal{O}(\mathrm{N}^3 \log \mathrm{N})$ with N^3 being the number of voxels.

7.5 Sparse Vector Field Regularization

7.5.1 Variational formulation

We now consider the problem of restoring incompressible fields from noisy measurements. Our goal is to denoise a discrete vector field denoted by $\mathbf{f}$ from the observation

$$\mathbf{y} = \mathbf{f} + \mathbf{n},$$

where $\mathbf{n}$ is assumed to be additive white Gaussian noise (AWGN). Using a wavelet tight frame, the unknown field is expressed as $\mathbf{f} = \mathbf{W}^{\mathrm{T}}\mathbf{w}$, where $\mathbf{W}$ expands the field in the divergence-free frame. Based on the developments in the thesis, we formulate a sparsity-driven denoising algorithm using our wavelet construction. To that end, we consider the following optimization problem:

$$\mathbf{f}^{\star} = \mathbf{W}^{\mathrm{T}} \underbrace{\arg\min_{\mathbf{w}} \left\{ \frac{1}{2} \|\tilde{\mathbf{w}} - \mathbf{w}\|_2^2 + \tau \|\mathbf{w}\|_1 \right\}}_{\mathbf{w}^{\star}}, \tag{7.19}$$

where $\tilde{\mathbf{w}}$ denotes the wavelet coefficients of the noisy field $\mathbf{y}$ and the equality is obtained since the construction is tight. Note that (7.19) is separable and the solution is provided by the proximal mapping of the ℓ_1-norm, which is a soft-thresholding function. Thus, we have

$$\mathbf{w}^{\star} = \eta_{\tau}(\tilde{\mathbf{w}}), \tag{7.20}$$

where $\eta_{\tau}(\cdot)$ ($\tau > 0$ being the threshold value) represents the soft-thresholding that is applied componentwise [56]. The denoised vector field is then given by $\mathbf{f}^{\star} = \mathbf{W}^{\mathrm{T}}\tilde{\mathbf{w}}$.

7.6 Numerical Results

For our experiments, we generate two different band-limited vector field phantoms that have zero divergence (see Figures 7.4 and 7.5). Our 2-dimensional model (called Circle) is of size $128^2 \times 2$, whereas the volumetric one (called Torus) is of size $64^3 \times 3$. The noisy measurements are obtained by degrading the data with AWGN to achieve various levels of signal-to-noise ratio (SNR). We measure the denoising performance of a given denoising method in terms of ΔSNR, which simply denotes the improvement provided. Also, for each wavelet-based method, the soft-thresholding operator is applied only to the wavelet (or detail) coefficients. We note that all visualizations are rendered with ParaView (Kitware Inc.) [155].

7.6.1 Scale-Adapted Thresholding

First, we investigate the effect of suitably modifying the threshold value for multi-scale decompositions. We use our the divergence-free frame $\mathcal{P}\boldsymbol{\psi}_{j,\boldsymbol{k},n}$ with 3-scale decomposition for denoising the Circle phantom. To make the thresholding scale-adapted, we compute appropriate multiplicative factors by expanding a white Gaussian noise (with a known variance) and measuring the average signal power at each scale. We compare this strategy to simply applying the same threshold value to all scales. By looking at Figure 7.2, we see that the scale-adapted thresholding significantly improves the denoising performance of our method.

7.6.2 Comparison with the Bi-Orthogonal Design

In the second part of our experiments, we compare our method against the bi-orthogonal divergence-free wavelets of [167].[2] Since our method is redundant (redundancy factor 3), we also include the cycle-spun version of the bi-orthogonal divergence-free wavelets (redundancy factor 8). As a baseline method, Leray projection, which is implemented as a Fourier domain filtering using (7.6), is incorporated in the comparisons.

The denoising is done as follows. As a pre-processing step, we apply a low-pass filter to the initial noisy data. Recall that our phantoms are band-limited, and the pre-filtering is used for all of the algorithms in question to make the comparison fair. For wavelet methods, 3-scale signal decomposition is performed, scale-adapted thresholding is used, and the threshold values are optimized for the best-possible SNR performance using an oracle.

By inspecting the results given in Figure 7.3, we see that the proposed divergence-free frame outperforms the other methods for the simulated field considered in the experiment. Since the construction given in [167] is bi-orthogonal, the wavelet transform does not preserve energy. In other words, the minimization of the ℓ_2-normed error in the wavelet domain does not imply the minimization of the ℓ_2-normed error in the signal domain. The noise statistics also change in the transform domain. However, since our reconstruction is tight, the energy is preserved, and thus the said drawback of bi-orthogonal methods is not observed. This shows the effectiveness of our approach for denoising incompressible vector fields (see Figure 7.5 for visual inspection).

7.7 Summary

We proposed a construction of divergence-free wavelet frames. We based our approach on the observation that applying the Leray projector $\mathcal{P}$ to a standard frame produces a divergence-free version. Furthermore, we used the singular-integral interpretation of $\mathcal{P}$ to show that the derived frame maintains the desirable attributes (decay rate and order of vanishing moments) of the original wavelets.

Based on a particular collection of wavelet frames of $L_2(\mathbb{R}^d)$, we proposed divergence-free frames of $(L_2(\mathbb{R}^d))^d$. The effectiveness of the two- and three-

---

[2] The software is available at http://www.eecs.berkeley.edu/∼mlustig/Software.html.

dimensional variants was illustrated on generated band-limited phantoms. These experiments hint at the potential of our wavelets for denoising impressible flow data, which is modeled as divergence-free.

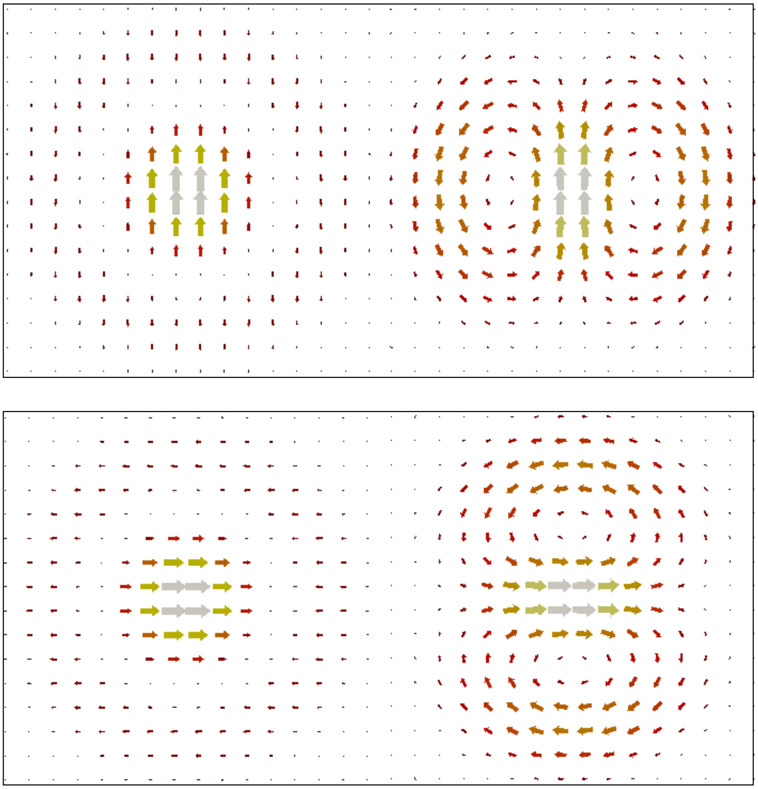


Figure 7.1: Two-dimensional divergence-free wavelet frame. Vector plot representations of $\boldsymbol{\psi}_{1,\mathbf{0},1}$ (top-left) and its Leray projection $\mathcal{P}\boldsymbol{\psi}_{1,\mathbf{0},1}$ (top-right) are given. Similarly, $\boldsymbol{\psi}_{1,\mathbf{0},2}$ (bottom-left) and its Leray projection $\mathcal{P}\boldsymbol{\psi}_{1,\mathbf{0},2}$ (bottom-right) are also illustrated. Note that the thickness of the vectors encode the magnitude.

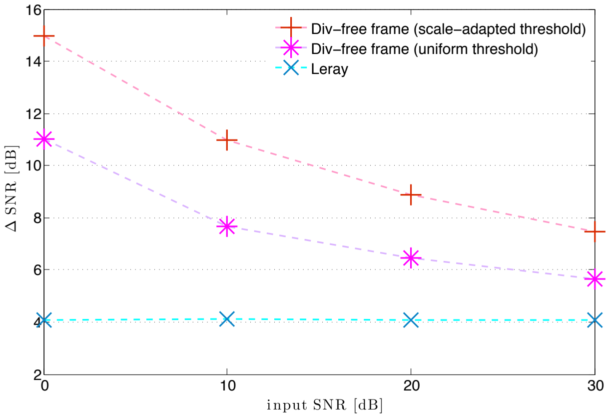


Figure 7.2: The effect of adapting the threshold parameter with respect to the decomposition scales. The denoising simulations are performed for Circle phantom. See text for further details.

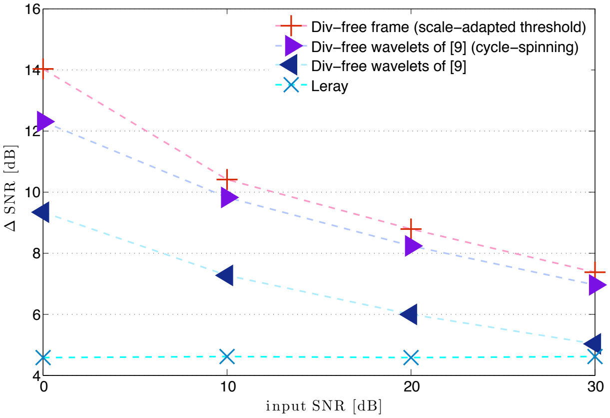


Figure 7.3: Comparison of the denoising performance of different methods. Simulations are performed for Torus phantom.

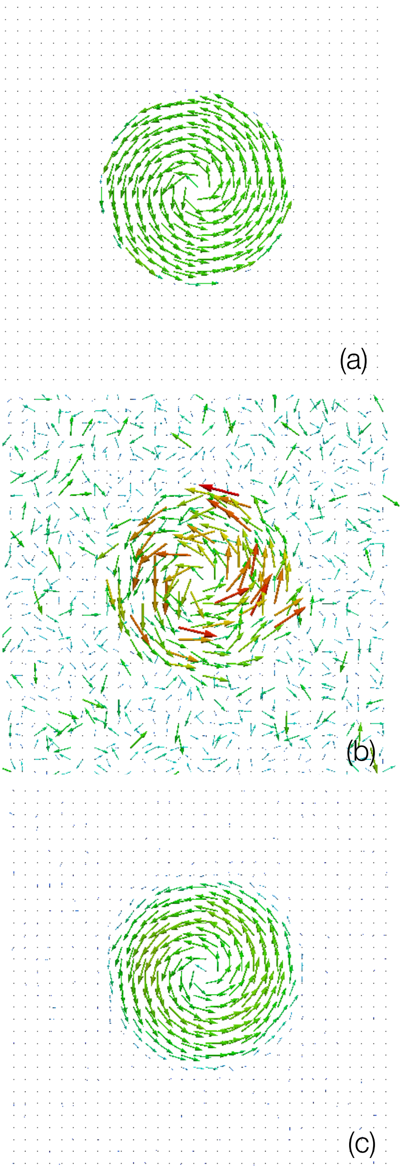


Figure 7.4: Illustration of the denoising results. Original Circle phantom is shown in (a). Noisy vector field (provided in (b) and its SNR value is 0 dB) is reconstructed by using our divergence-free wavelet frame. The SNR of the reconstructed field is 14.94 dB for (c).

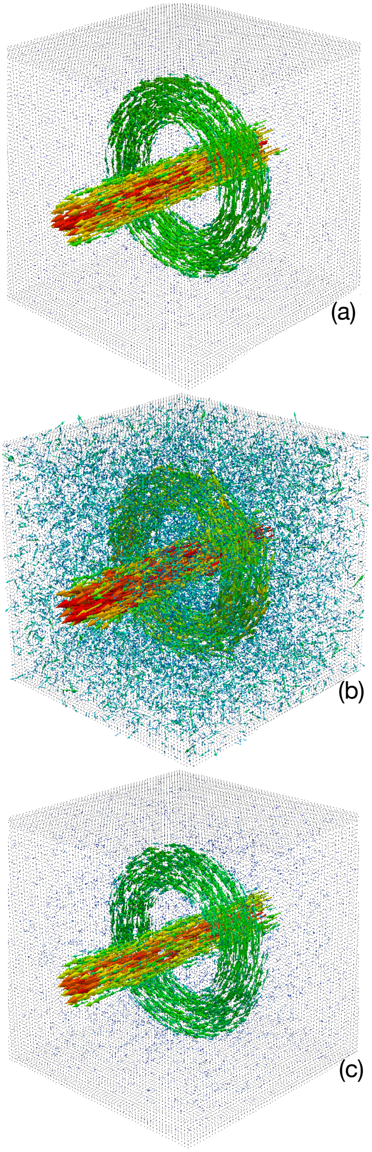


Figure 7.5: Illustration of the denoising results. Original Torus phantom is shown in (a). Noisy vector field (provided in (b) and its SNR value is 0 dB) is reconstructed by using our divergence-free wavelet frame. The SNR of the reconstructed field is 14.08 dB for (c).

Chapter 8

Conclusion

In the thesis, we have presented a class of novel and competitive methods to solve linear inverse problems. We have demonstrated the practical aspect of these ideas in regards to biomedical imaging. In the remainder of the chapter, we first recapitulate our main contributions and results. This is then followed by an outline of potential directions for future research.

8.1 Summary of Results

Stochastic justification of sparsity-promoting regularizations: We provided a consolidating statistical interpretation of the regularized solutions of inverse problems. In doing so, we combined a proper discretization method and a continuous-domain stochastic signal model. The crucial theoretical consequence of our framework is that the class of admissible prior distributions (for the discrete representation of the underlying signal signal) is confined to infinitely divisible distributions. Remarkably, the MAP estimators based on the proposed formalism justifies the use of both the classical Tikhonov-type methods and the sparsity-promoting regularizations that are currently of use.

In practical terms, we studied the problems of deconvolution, MR image reconstruction from partial Fourier coefficients, and image reconstruction from X-ray tomograms. We compared the reconstruction performance of regularization models of increasing sparsity. We emphasized the particular observation that sparsity-promoting algorithms are powerful tools for solving biomedical image reconstruction problems.

A novel phase retrieval algorithm for imaging unstained cells: We presented a practical phase retrieval algorithm for imaging label-free and thin biological samples by using a standard bright-field microscope. The model uses a focal stack of intensity images to have an approximation of the intensity variation of the infocus optical field along the axial direction. The infocus spatial phase is then obtained by relying on the transport-of-intensity equation (TIE). Our fundamental contribution is a sparsity-based reconstruction algorithm that nonlinearly combines different ranges of spatial frequencies depending on the defocus value of the intensity measurements. We showed that the proposed method outperforms commonly used TIE-based phase reconstruction algorithms in both simulated and real data configurations. Further, we compared the obtained phase maps with the ones acquired by using a digital holographic microscope (DHM). The latter confirms the validity of the proposed phase reconstruction technique.

Sparsity-based reconstruction of vector fields: We developed efficient regularization schemes for vector fields. Particularly, we considered the problem of flow field denoising and have proposed a regularizer that penalizes the nuclear norm of the Jacobian of the field. Our simulations illustrate that our regularizer is significantly superior (in terms of SNR performance) than the curl- and divergence-based regularizers as well as the well-known vectorial extensions of TV. We also used our reconstruction algorithm to enhance streamline visualizations of a real 4D flow MRI recording.

We also provided an efficient construction of divergence-free tight frames of $(L_2(\mathbb{R}^d))^d$. Based on our design, we developed a wavelet-based denoising scheme that is suitable for incompressible flow data, which is modeled as divergence-free. The effectiveness of the method over its bi-orthogonal variant is illustrated in two- and three-dimensional simulations.

8.2 Outlook

New regularization models for image reconstruction: The statistical analysis of natural images is a subject of research in image processing. In this perspective, it is empirically observed that natural images are compressible in wavelet bases [6] and tend to exhibit fractal properties [176]. Fageot et al. [72] have used the innovation modeling approach to investigate a general class of (finite-variance) self-similar and sparse random processes. Further, they have shown that the coarse-scale wavelet coefficients of these processes are asymptotically Gaussian, provided the wavelet has enough vanishing moments. The theoretical outcome of the model can thus be investigated for potentially improving the performance of wavelet-based image reconstruction schemes. To that end, we can modify the shrinkage functions (appearing in ADMM-based algorithms) that would take the predicted Gaussianity property into account.

Three-Dimensional Phase Microscopy: Phase images are of great interest in biology. In particular, quantitative phase measurements provide the computational means to calculate the spatial distributions of the refractive index of a cell, which are essential in morphology studies [109]. Therefore, the need for a technique that allows one to obtain the refractive index map in three dimensions is emphasized. To provide a simple and robust solution, one can try to extend the TIE formalism to a volumetric setting. The first stage of such an extension is to properly formulate the phase shifts introduced by the thick sample on the incident light field. An encouraging step towards this direction has already been taken in [177], where the authors discuss a suitable model under Born approximation. The latter can be used as the forward model in the context of a regularization approach.

Spatio-Temporal Flow Reconstruction: As mentioned in the thesis, PC MRI provides us with the time-resolved blood flow measurements with volumetric coverage. This necessitates the formulation of an appropriate spatio-temporal regularization method for denoising PC MRI data. Therefore, the practical importance of the proposed TVN-based framework can be further improved by adapting it to time-lapse flow imaging. In the development of the temporal regularization, we can be guided by [178], where it is explained that the flow velocity waveforms in the ascending and descending aorta are varying smoothly through the cardiac cycle. More importantly, all ideas developed for the denoising problem can and should be carried over to the general problem of PC MRI data reconstruction from undersampled Fourier coefficients. In such attempts, we can utilize TVN (and/or divergence-free wavelets) with our generic ADMM-based image reconstruction algorithm.

Bibliography

[1] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*, Taylor & Francis, 1998.

[2] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, IEEE Press, 1988.

[3] O. Scherzer, *Handbook of Mathematical Methods in Imaging*, Springer, 2011.

[4] M. Bertero and M. Piana, “Inverse problems in biomedical imaging: Modeling and methods of solution,” in *Complex Systems in Biomedicine*, pp. 1–33. Springer, 2006.

[5] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, 1993.

[6] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 2008.

[7] M. Lustig, D. Donoho, and J. M. Pauly, “Sparse MRI: The application of compressed sensing for rapid MR imaging,” *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–95, December 2007.

[8] C. Bouman and K. Sauer, “A generalized Gaussian image model for edge-preserving MAP estimation,” *IEEE Transactions on Image Processing*, vol. 2, no. 3, pp. 296–310, July 1993.

[9] S. D. Babacan, R. Molina, and A.K. Katsaggelos, “Bayesian compressive sensing using Laplace priors,” *IEEE Transactions on Image Processing*, vol. 19, no. 1, pp. 53–64, January 2010.

[10] M. Unser and P. D. Tafti, *An Introduction to Sparse Stochastic Processes*, Cambridge University Press, 2014.

[11] M. R. Teague, "Deterministic phase retrieval: A Green's function solution," *Journal of the Optical Society of America*, vol. 73, no. 11, pp. 1434–1441, November 1983.

[12] M. Markl, A. Frydrychowicz, S. Kozerke, M. Hope, and O. Wieben, "4D flow MRI," *Journal of Magnetic Resonance Imaging*, vol. 36, no. 5, pp. 1015–1036, November 2012.

[13] M. A. T. Figueiredo and J. M. Bioucas-Dias, "Restoration of Poissonian images using alternating direction optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3133–3145, December 2010.

[14] C. R. Vogel, *Computational Methods for Inverse Problems*, SIAM, 2002.

[15] G. H. Golub, P. C. Hansen, and D. P. O'Leary, "Tikhonov regularization and total least squares," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 1, pp. 185–194, 1999.

[16] D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley & Sons, 1997.

[17] M. Guerquin-Kern, *Wavelet-Based Reconstruction for Magnetic Resonance Imaging*, Ph.D. thesis, Swiss Federal Institute of Technology Lausanne (EPFL), July 2012.

[18] S. Lefkimmiatis, A. Bourquard, and M. Unser, "Hessian-based norm regularization for image restoration with biomedical applications," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 983–995, March 2012.

[19] X. Zhang and E. Y. Lam, "Edge-preserving sectional image reconstruction in optical scanning holography," *Journal of Optical Society of America A*, vol. 27, no. 7, pp. 1630–1637, July 2010.

[20] M. M. Marim, M. Atlan, E. Angelini, and J.-C. Olivo-Marin, "Compressed sensing with off-axis frequency-shifting holography," *Optics Letters*, vol. 35, no. 6, pp. 871–873, March 2010.

[21] J.-C. Baritaux, K. Hassler, M. Bucher, S. Sanyal, and M. Unser, “Sparsity-driven reconstruction for FDOT with anatomical priors,” *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1143–1153, May 2011.

[22] A. Szameit, Y. Shechtman, E. Osherovich, E. Bullkich, P. Sidorenko, H. Dana, S. Steiner, E. B. Kley, S. Gazit, T. Cohen-Hyams, S. Shoham, M. Zibulevsky, I. Yavneh, Y. C. Eldar, O. Cohen, and M. Segev, “Sparsity-based single-shot subwavelength coherent diffractive imaging,” *Nature Materials*, vol. 11, no. 5, pp. 455–459, May 2012.

[23] B. F. Logan, *Properties of High-Pass Signals*, Ph.D. thesis, Columbia University, November 1965.

[24] D. L. Donoho and B. F. Logan, “Signal recovery and the large sieve,” *SIAM Journal on Applied Mathematics*, vol. 52, no. 2, pp. 577–591, 1992.

[25] B. K. Natarajan, “Sparse approximate solutions to linear systems,” *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, 1995.

[26] E. J. Candès and T. Tao, “Near-optimal signal recovery from random projections: Universal encoding strategies?,” *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, December 2006.

[27] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, 2006.

[28] E. J. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Transcations on Information Theory*, vol. 52, no. 2, pp. 489–509, February 2006.

[29] M. Elad, M. A. T. Figueiredo, and M. Yi, “On the role of sparse and redundant representations in image processing,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 972–982, June 2010.

[30] J.-C. Baritaux, *Sparsity-Inducing Reconstruction Methods for Fluorescence Diffuse Optical Tomography*, Ph.D. thesis, Swiss Federal Institute of Technology Lausanne (EPFL), March 2012.

[31] M. Nikolova, “Weakly constrained minimization: Application to the estimation of images and signals involving constant regions,” *Journal of Mathematical Imaging and Vision*, vol. 21, no. 2, pp. 155–175, September 2004.

[32] U. S. Kamilov, E. Bostan, and M. Unser, “Wavelet shrinkage with consistent cycle spinning generalizes total variation denoising,” *IEEE Signal Processing Letters*, vol. 19, no. 4, pp. 187–190, April 2012.

[33] D. L. Donoho, “De-noising by soft-thresholding,” *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[34] M. Vetterli, J. Kovačević, and V. K. Goyal, *Fourier and Wavelet Signal Processing*, Cambridge University Press, 2014.

[35] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D*, vol. 60, no. 1-4, pp. 259–268, November 1992.

[36] Y. Wang, J. Yang, W. Yin, and Y. Zhang, “A new alternating minimization algorithm for total variation image reconstruction,” *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.

[37] E. Candès, M. B. Wakin, and S. Boyd, “Enhancing sparsity by reweighted ℓ_1 minimization,” *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, December 2008.

[38] M. Nikolova, M. K. Ng, and C.-P. Tam, “Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction,” *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3073–3088, December 2010.

[39] A. Chambolle and T. Pock, “A first-order primal-dual algorithm for convex problems with applications to imaging,” *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, May 2011.

[40] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, “Optimization with sparsity-inducing penalties,” *Foundations and Trends in Machine Learning*, vol. 4, no. 1, pp. 1–106, 2012.

[41] L. Condat, "A generic proximal algorithm for convex optimization–Application to total variation minimization," *IEEE Signal Processing Letters*, vol. 21, no. 8, pp. 985–989, August 2014.

[42] P. L. Combettes, L. Condat, J.-C. Pesquet, and B. C. Vũ, "A forward-backward view of some primal-dual optimization methods in image recovery," in *Proceedings of the 21st IEEE International Conference on Image Processing (ICIP'14)*, Paris, France, October 27-30, 2014, pp. 4141–4145.

[43] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pp. 185–212. Springer, 2011.

[44] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 906–916, August 2003.

[45] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, November 2004.

[46] A. Beck and M. Tebouelle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 183–202, 2009.

[47] Y. Nesterov, "A method of solving a convex programming problem with convergence rate $O(1/k^2)$," *Soviet Mathematics Doklady*, vol. 27, no. 2, pp. 372–376, 1983.

[48] M. Guerquin-Kern, M. Häberlin, K. P. Prüssmann, and M. Unser, "A fast wavelet-based reconstruction method for magnetic resonance imaging," *IEEE Transactions on Medical Imaging*, vol. 30, no. 9, pp. 1649–1660, September 2011.

[49] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2419–2434, November 2009.

[50] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, 1999.

[51] 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*, vol. 19, no. 9, pp. 2345–2356, September 2010.

[52] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 681–695, March 2011.

[53] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[54] E. Bostan, U. S. Kamilov, M. Nilchian, and M. Unser, "Sparse stochastic processes and discretization of linear inverse problems," *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2699–2710, July 2013.

[55] I. Rish and G. Y. Grabarnik, *Sparse Modeling: Theory, Algorithms, and Applications*, CRC Press, 2014.

[56] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, September 1994.

[57] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," in *Wavelets and Statistics*, pp. 125–150. Springer, 1995.

[58] U. S. Kamilov, P. Pad, A. Amini, and M. Unser, "MMSE estimation of sparse Lévy processes," *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 137–147, January 2013.

[59] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in Medicine and Biology*, vol. 53, no. 17, pp. 47–77, August 2008.

[60] J. F. Claerbout and F. Muir, "Robust modeling with erratic data," *Geophysics*, vol. 38, no. 5, pp. 826–844, October 1973.

[61] H. L. Taylor, S. C. Banks, and J. F. McCoy, "Deconvolution with the ℓ_1 norm," *Geophysics*, vol. 44, no. 1, pp. 39–52, January 1979.

[62] M. Zibulevsky and M. Elad, "L1-L2 optimization in signal and image processing," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 76–88, May 2010.

[63] H. Choi and R. Baraniuk, "Wavelet statistical models and Besov spaces," in *Proceedings of the SPIE Conference on Wavelet Applications in Signal Processing*, Denver CO, USA, July 18, 1999, vol. 3813, pp. 489–501.

[64] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, December 2007.

[65] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-Laplacian priors," in *Proceedings of the 23rd Annual Conference on Neural Information Processing Systems (NIPS'09)*, Vancouver BC, Canada, December 7-12, 2009, pp. 1033–1041.

[66] M. Unser, P. D. Tafti, and Q. Sun, "A unified formulation of Gaussian versus sparse stochastic processes—Part I: Continuous-domain theory," *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1945–1962, March 2014.

[67] M Unser, "Sampling—50 Years after Shannon," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 569–587, April 2000.

[68] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 1991.

[69] Q. Sun and M. Unser, "Left-inverses of fractional Laplacian and sparse stochastic processes," *Advances in Computational Mathematics*, vol. 36, no. 3, pp. 399–441, April 2012.

[70] B. B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman, 1983.

[71] J. Huang, *Statistics of Natural Images and Models*, Ph.D. thesis, Brown Univesity, May 2000.

[72] J. Fageot, E. Bostan, and M. Unser, "Wavelet statistics of sparse and self-similar images," *SIAM Journal on Imaging Sciences*, vol. 8, no. 4, pp. 2951–2975, 2015.

[73] I. Gelfand and N. Y. Vilenkin, *Generalized Functions Vol. 4. Applications of Harmonic Analysis*, Academic Press, 1964.

[74] K. Sato, *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press, 1994.

[75] F. W. Steutel and K. V. Harn, *Infinite Divisibility of Probability Distributions on the Real Line*, Marcel Dekker, 2004.

[76] M. Unser, P. D. Tafti, A. Amini, and H. Kirshner, "A unified formulation of Gaussian vs. sparse stochastic processes–Part II: Discrete-domain theory," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 3036–3051, May 2014.

[77] A. Amini, M. Unser, and F. Marvasti, "Compressibility of deterministic and random infinite sequences," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5193–5201, November 2011.

[78] R. Gribonval, V. Cevher, and M. E. Davies, "Compressible distributions for high-dimensional statistics," *IEEE Transactions on Information Theory*, vol. 58, no. 8, pp. 5016–5034, August 2012.

[79] A. Amini, U. S. Kamilov, E. Bostan, and M. Unser, "Bayesian estimation for continuous-time sparse stochastic processes," *IEEE Transactions on Signal Processing*, vol. 61, no. 4, pp. 907–920, February 2013.

[80] D. Wipf and S. Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 317–329, April 2010.

[81] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proceedings of the 33rd IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'08)*, Las Vegas NV, USA, 31 March-4 April 2008, pp. 3869–3872.

[82] Y. Zhang and N. Kingsbury, “Fast L0-based sparse signal recovery,” in *Proceesings of the 20th IEEE International Workshop on Machine Learning for Signal Processing (MLSP’10)*, Kittila, Finland, 29 August-1 Septembre 2010, pp. 403–408.

[83] M. A. T. Figueiredo, “Adaptive sparseness using Jeffrey’s prior,” in *Proceedings of the 14th Annual Conference on Neural Information Processing Systems (NIPS’01)*, Vancouver BC, Canada, December 3-8, 2001, pp. 697–704.

[84] M. A. T. Figueiredo and R. D. Nowak, “Wavelet-based image estimation: An empirical Bayes approach using Jeffrey’s noninformative prior,” *IEEE Transactions on Image Processing*, vol. 10, no. 9, pp. 1322–1331, September 2001.

[85] S. Ramani and J. A. Fessler, “A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 31, no. 3, pp. 677–688, March 2012.

[86] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, 2006.

[87] E. Bostan, J. Fageot, U. S. Kamilov, and M. Unser, “MAP estimators for self-similar sparse stochastic models,” in *Proceedings of the Tenth International Workshop on Sampling Theory and Applications (SampTA’13)*, Bremen, Germany, July 1-5, 2013, pp. 197–199.

[88] C. Hansen, J. G. Nagy, and D. P. O’Leary, *Deblurring Images: Matrices, Spectra, and Filtering*, SIAM, 2006.

[89] Y. Nesterov, *Introductory Lectures on Convex Optimization*, Springer, 2004.

[90] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[91] J. Yang, W. Yin, Y. Zhang, and Y. Wang, “A fast algorithm for edge-preserving variational multichannel image restoration,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 569–592, 2009.

[92] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, “Sparse reconstruction by separable approximation,” *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2479–2493, July 2009.

[93] C. Preza, M. I. Miller, and J.-A. Conchello, “Image reconstruction for 3D light microscopy with a regularized linear method incorporating a smoothness prior,” in *Proceedings of the SPIE Symposium on Electronic Imaging*, San Jose CA, USA, January 31, 1993, vol. 1905, pp. 129–139.

[94] N. Dey, L. Blanc-Féraud, C. Zimmer, P. Roux, Z. Kam, J.-C. Olivo-Marin, and J. Zerubia, “Richardson-Lucy algorithm with total variation regularization for 3D confocal microscope deconvolution,” *Microscopy Research and Technique*, vol. 69, no. 4, pp. 260–266, April 2006.

[95] K. T. Block, M. Uecker, and J. Frahm, “Undersampled radial MRI with multiple coils: Iterative image reconstruction using a total variation constraints,” *Magnetic Resonance in Medicine*, vol. 57, no. 6, pp. 1086–1098, June 2007.

[96] J. Trzasko and A. Manduca, “Highly undersampled magnetic resonance image reconstruction via homotopic ℓ_0-minimization,” *IEEE Transactions on Medical Imaging*, vol. 28, no. 1, pp. 106–121, January 2009.

[97] A. Entezari, M. Nilchian, and M. Unser, “A box spline calculus for the discretization of computed tomography reconstruction problems,” *IEEE Transactions on Medical Imaging*, vol. 31, no. 7, pp. 1289–1306, August 2012.

[98] M. Nilchian, *High Performance Reconstruction Framework for Straight Ray Tomography: From Micro to Nano Resolution Imaging*, Ph.D. thesis, Swiss Federal Institute of Technology Lausanne (EPFL), May 2015.

[99] J. Wang, T. Li, H. Lu, and Z. Liang, “Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography,” *IEEE Transactions on Medical Imaging*, vol. 25, no. 10, pp. 1272–1283, October 2006.

[100] X.-Q. Zhang and F. Jacques, “Constrained total variation minimization and application in computerized tomography,” in *Proceedings of the 5th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR’05)*, St. Augustine FL, USA, 2005, pp. 456–472.

[101] H. Yue and M. Jacob, “Higher degree total variation (HDTV) regularization for image recovery,” *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2559–2571, May 2012.

[102] A. J. Nasrallah and L. D. Griffin, "Gradient direction dependencies in natural images," *Spatial Vision*, vol. 20, no. 3, pp. 277–299, 2007.

[103] E. Bostan, E. Froustey, M. Nilchian, D. Sage, and M. Unser, "Variational phase imaging using the transport-of-intensity equation," *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 807–817, February 2016.

[104] D. A. Boas, C. Pitris, and N. Ramanujam, *Handbook of Biomedical Optics*, CRC press, 2011.

[105] S. W. Hell and J. Wichmann, "Breaking the diffraction resolution limit by stimulated emission: Stimulated-emission-depletion fluorescence microscopy," *Optics Letters*, vol. 19, no. 11, pp. 780–782, June 1994.

[106] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. L.-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science*, vol. 313, no. 5793, pp. 1642–1645, September 2006.

[107] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nature Methods*, vol. 3, no. 10, pp. 793–796, August 2006.

[108] D. J. Stephens and V. J. Allan, "Light microscopy techniques for live cell imaging," *Science*, vol. 300, no. 5616, pp. 82–86, 2003.

[109] G. Popescu, *Quantitative Phase Imaging of Cells and Tissues*, McGraw-Hill, 2011.

[110] J. Marrison, L. Räty, P. Marriott, and P. O'Toole, "Ptychography–a label free, high-contrast imaging technique for live cells using quantitative phase information," *Scientific reports*, vol. 3, no. 2369, pp. 1–7, August 2013.

[111] F. Zernike, "Phase contrast, a new method for the microscopic observation of transparent objects," *Physica*, vol. 9, no. 7, pp. 686–698, December 1942.

[112] G. Nomarski, "Nouveau dispositif pour l'observation en contraste de phase différentiel," *Journal de Physique et le Radium*, vol. 16, no. 5, pp. S88–S88, 1955.

[113] E. Meijering, "Cell segmentation: 50 years down the road," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 140–145, September 2012.

[114] P. Marquet, B. Rappaz, P. J. Magistretti, É. Cuche, Y. Emery, T. Colomb, and C. Depeursinge, "Digital holographic microscopy: A noninvasive contrast imaging technique allowing quantitative visualization of living cells with subwavelength axial accuracy," *Optics Letters*, vol. 30, no. 5, pp. 468–470, March 2005.

[115] B. Rappaz, B. Breton, E. Shaffer, and G. Turcatti, "Digital holographic microscopy: A quantitative label-free microscopy technique for phenotypic screening," *Combinatorial Chemistry & High Throughput Screening*, vol. 17, no. 1, pp. 80–88, January 2014.

[116] D. Paganin and K. A. Nugent, "Noninterferometric phase imaging with partially coherent light," *Physical Review Letters*, vol. 80, no. 12, pp. 2586–2589, March 1998.

[117] A. Barty, K. A. Nugent, D. Paganin, and A. Roberts, "Quantitative optical phase microscopy," *Optics Letters*, vol. 23, no. 11, pp. 817–819, June 1998.

[118] M. Beleggia, M. A. Schofield, V. V. Volkov, and Y. Zhu, "On the transport of intensity technique for phase retrieval," *Ultramicroscopy*, vol. 102, no. 1, pp. 37–49, 2004.

[119] L. J. Allen and M. P. Oxley, "Phase retrieval from series of images obtained by defocus variation," *Optics Communications*, vol. 199, no. 1, pp. 65–75, 2001.

[120] S. S. Kou, L. Waller, G. Barbastathis, and C. J. R. Sheppard, "Transport-of-intensity approach to differential interference contrast (TI-DIC) microscopy for quantitative phase imaging," *Optics Letters*, vol. 35, no. 3, pp. 447–449, 2010.

[121] L. Waller, L. Tian, and G. Barbastathis, "Transport of intensity phase-amplitude imaging with higher order intensity derivatives," *Optics Express*, vol. 18, no. 12, pp. 12552–12561, June 2010.

[122] Z. Jingshan, R. A. Claus, J. Dauwels, L. Tian, and L. Waller, “Transport of intensity phase imaging by intensity spectrum fitting of exponentially spaced defocus planes,” *Optics Express*, vol. 22, no. 9, pp. 10661–10674, May 2014.

[123] C. Zuo, Q. Chen, Y. Yu, and A. Asundi, “Transport-of-intensity phase imaging using Savitzky-Golay differentiation filter-theory and applications,” *Optics Express*, vol. 21, no. 5, pp. 5346–5362, March 2013.

[124] D. Paganin, A. Barty, P. J. McMahon, and K. A. Nugent, “Quantitative phase-amplitude microscopy. III. The effects of noise,” *Journal of Microscopy*, vol. 214, no. 1, pp. 51–61, April 2004.

[125] C.-Y. Chou, Y. Huang, D. Shi, and M. A. Anastasio, “Image reconstruction in quantitative X-ray phase-contrast imaging employing multiple measurements,” *Optics Express*, vol. 15, no. 16, pp. 10002–10025, August 2007.

[126] J. Martinez-Carranza, K. Falaggis, and T. Kozacki, “Multi-filter transport of intensity equation solver with equalized noise sensitivity,” *Optics Express*, vol. 23, no. 18, pp. 23092–23107, September 2015.

[127] M. Mitome, K. Ishizuka, and Y. Bando, “Quantitativeness of phase measurement by transport of intensity equation,” *Journal of Electron Microscopy*, vol. 59, no. 1, pp. 33–41, February 2010.

[128] L. Tian, J. C. Petruccelli, and G. Barbastathis, “Nonlinear diffusion regularization for transport of intensity phase imaging,” *Optics Letters*, vol. 37, no. 19, pp. 4131–4133, October 2012.

[129] E. Bostan, E. Froustey, B. Rappaz, E. Shaffer, D. Sage, and M. Unser, “Phase retrieval by using transport-of-intensity equation and differential interference contrast microscopy,” in *Proceedings of the 21st IEEE International Conference on Image Processing (ICIP’14)*, Paris, France, October 27-30, 2014, pp. 3939–3943.

[130] M. Soto, E. Acosta, and S. Ríos, “Performance analysis of curvature sensors: Optimum positioning of the measurement planes,” *Optics Express*, vol. 11, no. 20, pp. 2577–2588, October 2003.

[131] J. Martinez-Carranza, K. Falaggis, and T. Kozacki, “Optimum plane selection for transport-of-intensity-equation-based solvers,” *Applied Optics*, vol. 53, no. 30, pp. 7050–7058, October 2014.

[132] J. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, 2008.

[133] F. Joachim, *Three-dimensional Electron Microscopy of Macromolecular Assemblies*, Academic Press, 1996.

[134] F. Aguet, D. Van De Ville, and M. Unser, “Model-based 2.5-D deconvolution for extended depth of field in brightfield microscopy,” *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1144–1153, July 2008.

[135] M. Liebling, T. Blu, and M. Unser, “Fresnelets: New multiresolution wavelet bases for digital holography,” *IEEE Transactions on Image Processing*, vol. 12, no. 1, pp. 29–43, January 2003.

[136] E. Bostan, S. Lefkimmiatis, O. Vardoulis, N. Stergiopulos, and M. Unser, “Improved variational denoising of flow fields with application to phase-contrast MRI data,” *IEEE Signal Processing Letters*, vol. 22, no. 6, pp. 762–766, June 2015.

[137] J. G. Santiago, S. T. Wereley, C. D. Meinhart, D. J. Beebe, and R. J. Adrian, “A particle image velocimetry system for microfluidics,” *Experiments in Fluids*, vol. 25, no. 4, pp. 316–319, September 1998.

[138] F. Scarano, “Tomographic PIV: Principles and practice,” *Measurement Science and Technology*, vol. 24, no. 1, pp. 1–28, January 2013.

[139] A. Frydrychowicz, R. Arnold, D. Hirtler, C. Schlensak, A. F. Stalder, J. Hennig, M. Langer, and M. Markl, “Multidirectional flow analysis by cardiovascular magnetic resonance in aneurysm development following repair of aortic coarctation,” *Journal of Cardiovascular Magnetic Resonance*, vol. 10, no. 1, pp. 30, June 2008.

[140] O. Vardoulis, E. Coppens, B. Martin, P. Reymond, P. Tozzi, and N. Stergiopulos, “Impact of aortic grafts on arterial pressure: A computational fluid dynamics study,” *European Journal of Vascular and Endovascular Surgery*, vol. 42, no. 5, pp. 704–710, November 2011.

[141] S. N. Gupta and J. L. Prince, “On div-curl regularization for motion estimation in 3-d volumetric imaging,” in *Proceedings of the 3rd IEEE International Conference on Image Processing (ICIP’96)*, Lausanne, Switzerland, September 16-19 1996, vol. 1, pp. 929–932.

[142] D. Suter and F. Chen, “Left ventricular motion reconstruction based on elastic vector splines,” *IEEE Transactions on Medical Imaging*, vol. 19, no. 4, pp. 295–305, April 2000.

[143] M. Arigovindan, M. Sühling, C. Jansen, P. Hunziker, and M. Unser, “Full motion and flow field recovery from echo doppler data,” *IEEE Transactions on Medical Imaging*, vol. 26, no. 1, pp. 31–45, January 2007.

[144] P. D. Tafti and M. Unser, “On regularized reconstruction of vector fields,” *IEEE Transactions on Image Processing*, vol. 20, no. 11, pp. 3163–3178, November 2011.

[145] P. D. Tafti, R. Delgado-Gonzalo, A. F. Stalder, and M. Unser, “Variational enhancement and denoising of flow field images,” in *Proceedings of the 8th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI’11)*, Chicago IL, USA, March 30-April 2, 2011, pp. 1061–1064.

[146] E. Bostan, O. Vardoulis, D. Piccini, P. D.Tafti, N. Stergiopulos, and M. Unser, “Spatio-temporal regularization of flow-fields,” in *Proceedings of the 10th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI’13)*, San Francisco CA, USA, April 7-11, 2013, pp. 836–839.

[147] H. Attouch, G. Buttazzo, and G. Michaille, *Variational Analysis in Sobolev and BV Spaces: Applications to PDEs and Optimization*, SIAM, 2006.

[148] P. Blomgren and T. F. Chan, “Color TV: total variation methods for restoration of vector-valued images,” *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 304–309, March 1998.

[149] B. Goldluecke, E. Strekalovskiy, and D. Cremers, “The natural vectorial total variation which arises from geometric measure theory,” *SIAM Journal on Imaging Sciences*, vol. 5, no. 2, pp. 537–563, 2012.

[150] S. Lefkimmiatis, J. Ward, and M. Unser, "Hessian Schatten-norm regularization for linear inverse problems," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1873–1888, May 2013.

[151] A. H. Roger and J. R. Charles, *Matrix Analysis*, Cambridge University Press, 2012.

[152] M. Fazel, *Matrix Rank Minimization with Applications*, Ph.D. thesis, Stanford University, March 2002.

[153] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, December 2009.

[154] K. M. Holt, "Total nuclear variation and Jacobian extensions of total variation for vector fields," *IEEE Transactions on Image Processing*, vol. 23, no. 9, pp. 3975–3989, September 2014.

[155] J. Ahrens, B. Geveci, and C. Law, "Paraview: An end-user tool for large-data visualization," in *The Visualization Handbook*, pp. 717–732. Elsevier, 2005.

[156] E. Bostan, P. D. Tafti, and M. Unser, "A dual algorithm for L_1-regularized reconstruction of vector fields," in *Proceedings of the 9th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'12)*, Barcelona, Spain, May 2-5, 2012, pp. 1579–1582.

[157] X. Bresson and T. F. Chan, "Fast dual minimization of the vectorial total variation norm and applications to color image processing," *Inverse Problems and Imaging*, vol. 2, no. 4, pp. 455–484, November 2008.

[158] M. Markl, P. Kilner, and T. Ebbers, "Comprehensive 4d velocity mapping of the heart and great vessels by cardiovascular magnetic resonance," *Journal of Cardiovascular Magnetic Resonance*, vol. 13, no. 1, pp. 7, January 2011.

[159] U. Morbiducci, R. Ponzini, G. Rizzo, M. Cadioli, A. Esposito, F. D. Cobelli, A. D. Maschio, F. M. Montevecchi, and A. Redaelli, "In vivo quantification of helical blood flow in human aorta by time-resolved three-dimensional cine phase contrast magnetic resonance imaging," *Annals of Biomedical Engineering*, vol. 37, no. 3, pp. 516–531, March 2009.

[160] E. Bostan, M. Unser, and J.P. Ward, “Divergence-free wavelet frames,” *IEEE Signal Processing Letters*, vol. 22, no. 8, pp. 1142–1145, August 2015.

[161] X. Zhou and Y. He, “Using divergence free wavelets for the numerical solution of the 2-D stationary Navier–Stokes equations,” *Applied Mathematics and Computation*, vol. 163, no. 2, pp. 593–607, April 2005.

[162] K. Urban, *Wavelets in Numerical Simulation: Problem Adapted Construction and Applications*, vol. 22, Springer, 2002.

[163] F. Ong, M. Uecker, U. Tariq, A. Hsiao, M. T. Alley, S. S. Vasanawala, and M. Lustig, “Robust 4D flow denoising using divergence-free wavelet transform,” *Magnetic Resonance in Medicine*, vol. 73, no. 2, pp. 828–842, February 2015.

[164] D. Schiavazzi, F. Coletti, G. Iaccarino, and J. K. Eaton, “A matching pursuit approach to solenoidal filtering of three-dimensional velocity measurements,” *Journal of Computational Physics*, vol. 263, no. 1, pp. 206–221, April 2014.

[165] G. Battle and P. Federbush, “Divergence-free vector wavelets,” *The Michigan Mathematical Journal*, vol. 40, no. 1, pp. 181–195, 1993.

[166] P. G. Lemarié-Rieusset, “Analyses multi-résolutions non orthogonales, commutation entre projecteurs et dérivation et ondelettes vecteurs á divergence nulle,” *Revista Matemática Iberoamericana*, vol. 8, no. 2, pp. 221–237, 1992.

[167] E. Deriaz and V. Perrier, “Divergence-free and curl-free wavelets in two dimensions and three dimensions: Application to turbulent flows,” *Journal of Turbulence*, vol. 7, no. 3, pp. 1–37, 2006.

[168] R. Stevenson, “Divergence-free wavelet bases on the hypercube: Free-slip boundary conditions, and applications for solving the instationary Stokes equations,” *Mathematics of Computation*, vol. 80, no. 275, pp. 1499–1523, July 2011.

[169] S. Kadri-Harouna and V. Perrier, “Effective construction of divergence-free wavelets on the square,” *Journal of Computational and Applied Mathematics*, vol. 240, pp. 74–86, March 2013.

[170] K. Urban, "On divergence-free wavelets," *Advances in Computational Mathematics*, vol. 4, no. 1, pp. 51–81, December 1995.

[171] A. J. Majda and A. L. Bertozzi, *Vorticity and Incompressible Flow*, vol. 27, Cambridge University Press, 2002.

[172] O. Christensen, *An Introduction to Frames and Riesz Bases*, Birkhäuser, 2003.

[173] J. P. Ward, K. N. Chaudhury, and M. Unser, "Decay properties of Riesz transforms and steerable wavelets," *SIAM Journal of Imaging Sciences*, vol. 6, no. 2, pp. 984–998, 2013.

[174] J. P. Ward and M. Unser, "Harmonic singular integrals and steerable wavelets in $L_2(\mathbb{R}^d)$," *Applied and Computational Harmonic Analysis*, vol. 36, no. 2, pp. 183–197, March 2014.

[175] S. Held, M. Storath, P. Massopust, and B. Forster, "Steerable wavelet frames based on the Riesz transform," *IEEE Transactions on Image Processing*, vol. 19, no. 3, pp. 653–667, February 2010.

[176] D. Mumford and B. Gidas, "Stochastic models for generic images," *Quarterly of Applied Mathematics*, vol. 59, no. 1, pp. 85–112, March 2001.

[177] K. G. Phillips, S. L. Jacques, and O. J. T. McCarty, "Measurement of single cell refractive index, dry mass, volume, and density using a transillumination microscope," *Physical Review Letters*, vol. 109, no. 11, pp. 118105, September 2012.

[178] W. W. Nichols, M. F. O'Rouke, and C. Vlachopoulos, *McDonald's Blood Flow in Arteries: Theoretical,Experimental and Clinical Principles*, Hodder Arnold, 2011.

Curriculum Vitæ

EMRAH BOSTAN

Avenue de Sévelin 2B 1007, Lausanne, Switzerland
(+41) 78 813 98 62 | emrah.bostan@gmail.com | bigwww.epfl.ch/bostan

EDUCATION

Swiss Federal Institute of Technology in Lausanne (EPFL), Lausanne, Switzerland

- Ph.D. in Electrical Engineering 2011 - 2016
  Thesis : « Sparsity-Based Data Reconstruction Models for Biomedical Imaging »
  Advisor: Prof. Michael Unser

- M.Sc. in Electrical Engineering (specialization in Information Technologies) 2009 - 2011
  Thesis : « Efficient Algorithms for Vector Field Reconstruction »
  Advisor: Prof. Michael Unser

Istanbul Technical University (ITU), Istanbul, Turkey

- B.Sc. in Telecommunication Engineering 2005 - 2009

RESEARCH INTERESTS

Computational Biomicroscopy	- Deconvolution for fluorescence imaging - Compressed sensing for diffraction tomography - Phase retrieval techniques - Super-resolution microscopy
Inverse Problems in Medical Imaging	- Image reconstruction for X-ray CT - Denoising of 4D Flow-sensitive MRI
Algorithm Design	- Variational models for image enhancement - Convex optimization techniques - Bayesian inference
Statistical Modeling of Images	- Continuous-domain stochastic models - Discretization of linear inverse problems - Multiscale image analysis

PROFESSIONAL EXPERIENCE

EPFL

Research and Teaching Assistant in Biomedical Imaging Group 2011 - ongoing

- Developed theory and software solutions for inverse problems in *biomedical imaging, iterative data reconstruction, and statistical data modeling.* Defined and supervised research projects, all of which involved the design and implementation of specific tools for image processing/analysis applications. Some projects were conducted in collaboration with other EPFL labs.

Summer Intern in Laboratory for Information and Inference Systems 2010

- Designed low-dimensional data models for better compression of high-dimensional data that feature structured geometric and combinatorial foundations. Developed combinatorial optimization techniques that significantly reduced the computational complexity compared to the state-of-the-art.

ITU

Researcher in Electromagnetic Research Group 2009

- Worked on a tomographic model for non-destructively measuring the thickness of dielectric coatings on metal surfaces and performed proof-of-concept experiments in the anechoic chamber. I also delivered a software package implementing the developed theory.

PUBLICATIONS

Journal Papers

1. M. Nilchian, **E. Bostan**, Z. Wang, M. Stampanoni, and M. Unser, « Joint Absorption and Phase Retrieval in Grating-Based X-ray Radiography, » Optics Express, *in press*.

2. **E. Bostan,** E. Froustey, M. Nilchian, D. Sage, M. Unser, « Variational Phase Imaging Using the Transport-of-Intensity Equation, » IEEE Transactions on Image Processing, vol. 25, no. 2, pp. 807-817, February 2016.

3. J. Fageot, **E. Bostan,** M. Unser, « Wavelet Statistics of Sparse And Self-Similar Images, » SIAM Journal on Imaging Sciences, vol. 8, no. 4, pp. 2951-2975, December 2015.

4. **E. Bostan,** M. Unser, J. P. Ward, « Divergence-free Wavelet Frames, » IEEE Signal Processing Letters, vol. 22, no. 8, pp. 1142-1146, August 2015.
   ★ Recognized as part of the « Top 10% » papers at IEEE International Conference on Image Processing 2015.

5. **E. Bostan,** S. Lefkimmiatis, O. Vardoulis, N. Stergiopulos, M. Unser, « Improved Variational Denoising of Flow Fields with Application to Phase-Contrast MRI Data, » IEEE Signal Processing Letters, vol. 22, no. 6, pp. 762-766, June 2015.
   ★ Recognized as part of the « Top 10% » papers at IEEE International Conference on Image Processing 2015.

6. U. S. Kamilov, **E. Bostan,** and M. Unser, « Variational Justification of Cycle Spinning for Wavelet-Based Solutions of Inverse Problems, » IEEE Signal Processing Letters, vol. 21, no. 11, pp. 1326-1330, November 2014.

7. **E. Bostan,** U. S. Kamilov, M. Nilchian, and M. Unser, « Sparse Stochastic Processes and Discretization of Linear Inverse Problems, » IEEE Transactions on Image Processing, vol. 22, no. 7, pp. 2699-2710, July 2013.

8. A. Kazerouni, U. S. Kamilov, **E. Bostan,** and M. Unser, « Bayesian Denoising: From MAP to MMSE Using Consistent Cycle Spinning, » IEEE Signal Processing Letters, vol. 20, no. 3, March 2013.

9. A. Bourquard, N. Pavillon, **E. Bostan,** C. Depeursinge, M. Unser, « A Practical Inverse-Problem Approach to Digital Holographic Reconstruction, » Optics Express, vol. 21, no. 3, pp. 3417-3433, February, 2013.

10. A. Amini, U. S. Kamilov, **E. Bostan,** and M. Unser, « Bayesian Estimation for Continuous-Time Sparse Stochastic Processes, » IEEE Transactions on Signal Processing, vol. 61, no. 4, pp. 907-920, February 2013.

11. U. S. Kamilov, **E. Bostan,** and M. Unser, « Wavelet Shrinkage with Consistent Cycle Spinning Generalizes Total Variation Denoising, » IEEE Signal Processing Letters, vol. 19, no. 4, pp. 187-190, April 2012.

Conference Papers

1. P. Tohidi, **E. Bostan**, P. Pad, and M. Unser, « MMSE Denoising of Sparse and Non-Gaussian AR(1) Processes, » Proceedings of the 41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'16), Shanghai,China, March 20-25, 2016, *in press*.

2. **E. Bostan,** E. Froustey, B. Rappaz, E. Shaffer, D. Sage, M. Unser, « Phase Retrieval by using Transport-of-Intensity Equation and Differential Interference Contrast Microscopy," Proceedings of the 21st IEEE International Conference on Image Processing (ICIP'14), Paris, France, October 27-30, 2014, pp 3939-3943.
    ★ Recognized as part of the « Top 10% » papers.

3. J. Fageot, **E. Bostan,** M. Unser, « Statistics of Wavelet Coefficients For Sparse Self-Similar Images, » Proceedings of the 21st IEEE International Conference on Image Processing (ICIP'14), Paris, France, October 27-30, 2014, pp. 6096-6100.

4. E. Froustey, **E. Bostan,** S. Lefkimmiatis, M. Unser, « Digital Phase Reconstruction via Iterative Solutions of Transport-of-Intensity Equation, » Proceedings of the 13th IEEE Workshop on Information Optics (WIO'14), Neuchâtel NE, Switzerland, July 7-11, 2014, pp. 1-3.

5. **E. Bostan,** J. Fageot, U. S. Kamilov, and M. Unser, « MAP Estimators for Self-Similar Sparse Stochastic Models, » Proceedings of the 10th International Conference on Sampling Theory and Applications (SampTA'13), Bremen, Germany, July 1-5, 2013, pp. 197-199.

6. **E. Bostan,** U. S. Kamilov, M. Nilchian, and M. Unser, « Consistent Discretization of Linear Inverse Problems using Sparse Stochastic Processes, » 5th Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS'13), July 8-11, 2013, Lausanne.

7. **E. Bostan,** O. Vardoulis, D. Piccini, P. D. Tafti, N. Stergiopulos, and M. Unser, « Spatio-Temporal Regularization of Flow-Fields, » Proceedings of the 10th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'13),San Francisco CA, USA, April 7-11, 2013, pp. 824-827.

8. P.D. Tafti, **E. Bostan,** and M. Unser, « Variational Decomposition of Vector Fields in the Presence of Noise, » Proceedings of the 10th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'13), San Francisco CA, USA, April 7-11, 2013, pp. 1162-1165.

9. U. S. Kamilov, A. Bourquard, **E. Bostan,** and M. Unser, « Autocalibrated Signal Reconstruction from Linear Measurements using Adaptive GAMP, » Proceedings of the 38th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'13),Vancouver BC, Canada, May 26-31, 2013, pp. 5925-5928.

10. B. Tekin, U. S. Kamilov, **E. Bostan,** and M. Unser, « Benefits of Consistency in Image Denoising with Steerable Wavelets, » Proceedings of the 38th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'13), Vancouver BC, Canada, May 26-31, 2013, pp. 1355-1358.

11. **E. Bostan,** U. Kamilov, M. Unser, « Reconstruction of Biomedical Images and Sparse Stochastic Modeling, » Proceedings of the 9th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'12), Barcelona, Spain, May 2-5, 2012, pp. 880-883.

12. **E. Bostan,** P.D. Tafti, M. Unser, « A Dual Algorithm for L1-Regularized Reconstruction of Vector Fields, » Proceedings of the 9th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'12), Barcelona, Spain, May 2-5, 2012, pp. 1579-1582.

13. U. Kamilov, **E. Bostan,** M. Unser, « Generalized Total Variation Denoising via Augmented Lagrangian Cycle Spinning with Haar Wavelets, » Proceedings of the 37th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'12), Japan, March 25-30, 2012, pp. 909-912.

14. O. Ozdemir, H. Haddar, C. Fidan, **E. Bostan,** « Thickness Reconstruction of Dielectric Coatings by the Use of Higher Order Impedance Boundary Conditions, » 6th Workshop on Advanced Computational Electromagnetics (ACE'10), July 5-7, 2010, Zurich

TEACHING & SUPERVISION

- Teaching Assistant (TA) for the following courses at EPFL:
  Signals and Systems I:
  Autumn 2011, Autumn 2012, Autumn 2013, Autumn 2014
  Signals and Systems II:
  Spring 2012, Spring 2013, Spring 2014

- Supervised the following M. Sc. students at EPFL:
  1. Bugra Tekin (co-advised with Ulugbek Kamilov and Prof. Michael Unser), 2012
     Semester Project: « Solving Inverse Problems with Sparsifying Transforms »

  2. Emmanuel Froustey (co-advised with Prof. Michael Unser), 2013
     Semester Project: « Optical Flow Estimation under Sparsity Constraints »

     Master Project: « A practical inverse problem approach for phase imaging »
  3. Pascal Bienz (co-advised with Stamatis Lefkimmiatis and Prof. Michael Unser), 2013
     Master Project: « Coil Sensitivity Estimation for Parallel Magnetic Resonance Imaging »

  4. David Nguyen (co-advised with Prof. Michael Unser), 2013
     Semester Project: « Transport-of-intensity approach for quantitative phase imaging »

  6. Florence Gavin (co-advised with John Ward and Prof. Michael Unser), 2014
     Semester Project: « Flow field enhancement with divergence-free wavelets »

7. Christopher Finelli (co-advised with Denis Fortun and Prof. Michael Unser), 2015
   Master Project: « Variational approaches for optical flow estimation »

8. Pouria Tohidi (co-advised with Pedram Pad and Prof. Michael Unser), 2015
   Summer Research Project: « MMSE Denosing of Sparse and Non-Gaussian AR(1) Processes »

TECHNICAL SKILLS

- Coding Languages: C, C++
- Computation tools: MATLAB, Mathematica, Maple
- Technologies, Libraries: ImageJ, Fiji, Icy, Paraview, VTK, OpenCV, SVN, Git
- Environments: Mac OS X, GNU/Linux, Windows

HONORS & AWARDS

- Nomination for a Ph.D. Thesis Award of EPFL
- Member of the TA team that has won the Education Award of EPFL's Life Sciences Section (2013)
- EPFL Excellence Scholarship for Master Studies (2009-2011)
- Alcatel-Lucent Graduation Award at Undergraduate Level (2009)
- Graduated as Salutatorian at Undergraduate Level (2009)
- ITU High Honor Student for all eight semesters (2005-2009)
- ITU Merit Scholarship at Undergraduate Level (2005)

PROFESSIONAL ACTIVITIES

- Reviewer for the following journals:
  IEEE Transactions on Image Processing, IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Signal Processing, IEEE Transactions on Medical Imaging, IEEE Signal Processing Letters, Elsevier Journal of Visual Communication and Image Representation, and BioMedical Engineering OnLine

- Reviewer for the following conferences:
  IEEE ICASSP 2013, IEEE ISBI 2013, SampTA 2013, IEEE ISBI 2014, and IEEE ISBI 2016

- IEEE Student Member (2011 - ongoing)

PERSONAL

- Languages: Turkish (native), English (fluent), French (intermediate)
- Photography: I am interested in traditional film photography as well as darkroom printing. You can find the link to my portfolio at my website.