2010-01-01

Compressive Sensing with Prior Information Applied to Magnetic Resonance Imaging

Cristiano Jacques Miosso
University of Texas at El Paso, miosso.utep@gmail.com

Compressive Sensing with Prior Information Applied to Magnetic Resonance Imaging

Cristiano Jacques Miosso

Department of Electrical and Computer Engineering

Approved:

Ricardo von Borries, PhD, Chair

Joseph H. Pierluissi, PhD

Patricia A. Nava, PhD

Zainul Abedin, MD, External Member

Patricia D. Witherspoon, PhD

Dean of the Graduate School

To my mother (in memoriam), to whom I owe everything,
and to Estelle, who joined me in this journey.

They inspire me and give me strength.

Compressive Sensing with Prior Information Applied to Magnetic Resonance Imaging

by

Cristiano Jacques Miosso, E.C.E., M.Sc.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Department of Electrical and Computer Engineering

The University of Texas at El Paso

December 2010

Acknowledgment

In developing this PhD research, I was fortunate enough to receive support and encouragement from very special people. Completing my PhD was certainly a great challenge, albeit an extremely rewarding and stimulating one; but the direct and indirect help from friends and relatives made the experience much more enjoyable and enlightening. Therefore, I must show my deepest appreciation for them.

First, I want to mention the great guidance by Dr. Ricardo von Borries of the University of Texas at El Paso, my research advisor. Being his student during these last four years was an amazing experience. His vision of science and research, his insights, and his incredible dedication cause a great impact in the lives of people who have the opportunity to work with him. I am proud to be one of these people.

And I am especially grateful for all the hours of very productive discussion I had with him; in particular, I will never forget our discussions on stochastic processes, linear algebra, and signal processing, as well as how diligently he helped me with the nice theoretical problems we had to analyze. One of the most important qualities of a researcher, in my opinion, is the ability to convert the unforeseen difficulties into opportunities for new insights and for ingenious discoveries. He did precisely that, and always showing genuine optimism and encouragement. I am therefore very grateful for his confidence and support.

Also thanks to Dr. Pierluissi, Dr. Nava and Dr. Abedin, for gently accepting to be part of this research, and for their enriching comments and suggestions. Dr. Pierluissi is also co-director of the High Performance Signal Processing Lab, where I developed the research, whereas Dr. Nava is the principal investigator for the grant of which this research is part.

I want to show my appreciation to Dr. Lush and Dr. Usevitch as well, for many interesting and inspiring conversations about scientific research and about signal processing and communications theory. They were always willing to debate, in-depth, interesting concepts

in their fields, and this helped me to keep a strong interest on my own research. I am also grateful to the UTEP staff. Additionally, a special thanks to Dr. Williams and Mr. Nito Gumataotao, some of the developers of Virgo, the cluster in which I ran most of the simulations in this project.

I am also proud of and grateful for being the recipient of different scholarships, which supported this research. I acknowledge the National Science Foundation (NSF), which supported me for four years, under the grant NSF CNS-0709438, and the Texas Instruments (TI), for the TI Endowed Scholarship, which I received three times. Also thanks to the Krutilek Scholarship, and to the University Research Institute (URI UTEP). Finally, I am very grateful to the National Geospatial-Intelligence Agency (NGIA), which supported me in my last Summer at UTEP, under the grant HM1582-09-1-0015.

In conducting this research, I used many medical images gently provided by the Sierra Medical Center, at El Paso (TX), to which I am very grateful. I thank Dr. Aida Grijalva, Mr. S. Alonzo Armendariz and Mr. Enrique Ordonez, for their assistance with acquiring the images and for their time and help. And I am extremely grateful to Dr. Abedin also in this respect, for he helped me with the long process of getting the needed authorization to collect the images, and oriented me about each step involved. Also, his medical knowledge was very useful when analyzing the images.

While in the US, I was really fortunate to meet people who were always willing to assist me and to help in very practical terms, as well as to share the wonders of a different culture and of different places. Besides Dr. von Borries, who was much more than an advisor and who shared many common interests with me, I want to thank Dr. Lee Rosenthal and her family, as well as Vikram Jayaram, Erich Warren, Leonel Salayandia and Fernanda Leite. My fiancée Estelle and I shared some wonderful moments with them, including some memorable trips. We will miss those moments, but we are positive that we will meet again soon, and share new experiences together.

Additionally, many thanks to my siblings Júnior, Alexandre, and Simone, and to my aunt Marilda. The feeling of being among such altruistic people was and continues to be a source of

great comfort and security. I know I can count on them, as I have many times. Besides, they were always enthusiastic about my endeavors in engineering, music, and research, which helped me to keep alive my own dreams and to remember that I was not alone in those endeavors.

Also thanks to other special friends in Brazil. I am grateful to Luís Gustavo Carvalho and Hugo Gosmann, who always encouraged me during my PhD. And Dr. Adson Rocha, Dr. Adolfo Bauchspiess, Dr. Lucio Martins and Dr. George von Borries have been great supporters of my academic career, for which I thank them. Also, I am extremely honored to now live together with Hilda, Andrey, and Noelle, who have proved to be wonderful friends and an amazing company.

To conclude, I want to give a special acknowledgment to two very important people, who inspire me with their accomplishments, and who have done so much for me.

First, let me thank my beloved fiancée, Estelle. When I came to the US for my doctoral studies, she gladly and courageously accepted to join me. And she saw the opportunity to advance her own studies, as she brilliantly completed her master's degree in Public Health degree. At this stage, I saw a new model in her, that of a very professional and competent researcher, and her constant company was a source of joy for many years of hard work. I am very happy for having found her, and for sharing a life with such a nice and intelligent woman.

And I simply have no words to properly thank all the incentive, guidance and support I received from my mother Alaide, as well as to describe the inspiration she has become. Coming from a very low-income family, not only were she able to graduate and to succeed in life; she always put her children's education at the highest level of priority, and never measured efforts in helping us pursue our goals.

Unfortunately, she never saw me complete my PhD. She passed away just a few months before my defense. Nevertheless, she made it possible. And through all her efforts and unconditional support, she now leaves a trace in these pages and in all my future accomplishments. Together with this acknowledgment, here goes a very special dedication to her.

Abstract

This dissertation introduces the theory of compressive sensing with prior information about a signal's sparse representation. We show, mathematically and in numerical simulations, that prior information improves signal reconstruction, in terms of number of required measurements, computation time and signal-to-noise ratios. Following, we present a set of methods for enhanced magnetic resonance imaging and tomography that can be combined with prior information, for enhanced image quality.

In developing the theory of compressive sensing with prior information, we provide a mathematical proof of the required condition (in terms of number of linear measurements) for reconstruction using the ideal approach of ℓ_0-minimization, as well as the necessary and sufficient conditions for reconstruction using ℓ_1-minimization. We then develop the theory for evaluating the probability of reconstruction when dealing with stochastic signals, both in the case without and with prior information. Furthermore, we compare the proposed models with the empirical probability of reconstruction, by evaluating in Monte-Carlo simulations the percentage of cases in which the conditions for reconstruction with prior information are satisfied, opposed to compressive sensing without prior information. These analyzes show that prior information about the sparse representation's support reduces the number of linear measurements required to reconstruct a signal, both in terms of the theoretical lower bound and of the ℓ_1-minimization procedure, commonly used in the compressive sensing literature.

Regarding practical reconstruction algorithms, we provide an optimization procedure for compressive sensing with prior information based on ℓ_p-minimization. The experimental results show that further improvement can be obtained by combining support prior informa-

tion with the reduction of the p parameter in the ℓ_p-minimization. Furthermore, by using this method we show that partial support information also reduces the total computation time, when using a direct method to solve the inner linear systems. For a fixed number of taken linear measurements, on the other hand, the experiments show that prior information improves the signal-to-noise ratios of the reconstructed signals.

The dissertation also proposes a compressive sensing method for enhanced reconstruction of magnetic resonance (MR) images using a prefiltering strategy in the k-space domain. This method can improve the quality of the reconstruction over a standard compressive sensing approach. In particular, it leads to higher signal-to-noise ratios, and generally to lower reconstruction times. Also, it easily allows for a parallel implementation since different computation stages are independent of each other, which can further reduce the reconstruction times by a factor of up to three, in the tested reconstruction schemes.

Finally, we show how the proposed MR imaging method can be combined with the previous approach of compressive sensing with prior information. The experiments conducted over simulated and real MR images and functional MR images show a further improvement by combining pre-filtering in the k-space domain with the use of support prior information, in terms of visual quality and signal-to-noise ratios.

Table of Contents

List of Figures

List of Tables

1

Introduction

In this research, we introduce the theory of compressive sensing using prior information about a signal's sparse representation. We discuss how this type of information enhances signal reconstruction, first from a theoretical point-of-view and then experimentally, using an efficient novel compressive sensing algorithm. In particular, as we will show, support prior information reduces the theoretical minimum number of measurements required for signal reconstruction, as well as the total computation time. Furthermore, for a fixed number of measurements, it improves the signal-to-noise ratios of the reconstructed signals.

We also propose a compressive sensing method for enhanced reconstruction of magnetic resonance (MR) images using a prefiltering strategy in the k-space domain. This method can improve the quality of the reconstruction over a standard compressive sensing approach. In particular, it leads to higher signal-to-noise ratios, and generally to lower reconstruction times. Also, it easily allows for a parallel implementation since different computation stages are independent of each other, which can further reduce the reconstruction times by a factor of up to three, in the tested reconstruction schemes.

Following, we show how this method can be combined with the previous approach of compressive sensing with prior information; preliminary results show a further improvement

by combining prefiltering in the k-space domain with the use of prior information.

This chapter establishes the main objectives related to this development, as well as the contents of the remaining of the dissertation, and is organized as follows. Section 1.1 presents the context of the research, with emphasis on medical imaging applications, and the statement of purpose. Next, Section 1.2 describes the main contributions of the dissertation and the research questions that we will tackle in evaluating the implemented systems' performance. Finally, Section 1.3 describes the structure of the remaining chapters.

1.1 Context of the Research and Statement of Purpose

In recent years, diverse modalities of imaging techniques have led to important advances in healthcare and basic medical research. Since the discovery that X-rays allow one to see inside a human body, research on medical imaging has focused on improving ways to observe several types of internal tissues and organs without the need for exploratory surgical interventions [53]. Techniques such as magnetic resonance imaging (MRI), X-ray tomography and positron emission tomography (PET) then became important tools to detect diseases and medical conditions, from joint lesions and damaged organs to malignant tumors [96], [50].

Among the recent advances in imaging technology, those in MRI, functional MRI and computed tomography (CT) are particularly important to the development of computer-aided detection and diagnostic systems and to the understanding of physiology [85], [52], [79], [69]. For example, a magnetic resonance piece of equipment based on high static magnetic fields can acquire different types of images with high signal-to-noise ratios and resolution, so that small structures can be observed. Also, functional imaging allows a researcher or a health professional to analyze the behavior of different organs along the time, as a sequence of images registers their functioning; this has an important impact, for instance, on current research about the different regions of the brain and their functions, and about the physiology of the heart and other organs.

However, in MRI and CT systems the data acquisition is inherently slow [58], [57]. Patients or subjects are typically required to stay still in a very small volume and for a few to several minutes as the measurements are taken; indeed, dealing with anxiety and claustrophobia is a special concern of modern imaging equipment. The acquisition time is particularly critical when analyzing moving structures, such as the coronary arteries and the whole heart; in this situation, a common procedure is to synchronize the data acquisition with the movement cycle, so that the measurements are taken when the structures are in approximately the same position. As a result, however, waiting times are introduced, which is specially inconvenient when the patient has to hold the breath during the acquisitions [81].

In medical imaging, thus, the number of acquired measurements is always limited by hardware, software, and physiological constraints. The main challenge is therefore to reconstruct high resolution, high signal-to-noise images from, for example, limited k-space samples in the case of MRI, or limited linear projections in the case of X-ray tomography.

In this context, the recently developed technique of compressive sensing can be very effective in reconstructing the images, when compared to traditional approaches such as filtered backprojection and gridding techniques. In fact, compressive sensing allows different types of signals, known to have sparse representations in transformed domains, to be reconstructed with theoretically perfect accuracy from fewer measurements than previous approaches [65], [5], [26]. The fundamental idea is to acquire few linear measurements from a nonsparse representation, and later apply an appropriate optimization procedure in order to obtain the complete desired signal. Note that compressive sensing can then simplify the signal acquisition, at the cost of a more complex reconstruction stage.

This approach fits appropriately in the context of medical imaging, as new compressive sensing techniques for MRI and CT are being developed. The images to be reconstructed are usually compressible by linear transformation [57], which means that they have a sparse representation in a known transformed domain. Furthermore, the acquired samples correspond to measurements from a nonsparse representation, also a required condition. Compressive sensing is thus a promising tool to cope with the sampling constraints of medical imaging,

especially considering that the cost of the equipment needed for the optimization stages does not add significantly to that of the acquisition equipment itself.

Nevertheless, most reconstruction techniques in compressive sensing are still general with respect to the coefficients in the sparse representation. Indeed, the reconstruction aims at finding the sparsest vector that corresponds to the available measurements, generally with no assumption on the structure or distribution of these coefficients. This approach does not explore any prior information that may exist about the signals to be reconstructed. However, this kind of information can, as we will show, reduce the number of measurements required for reconstruction, or, in practical terms, improve the quality of the output images when the same number of measurements is used.

This research focuses on compressive sensing and MRI methods that can make use of prior information about the sparse representation's support. In developing these techniques, we initially consider the general case of signals with sparse representations, rather than restricting the treatment to medical images. Later, we propose a method for improved reconstruction of MR images, and compare this method to a standard compressive sensing approach [10], [57]. The comparison at first does not take into account any form of prior information; following, we show how the previously studied prior information can be included in the proposed method, and evaluate the additional improvement when using this information.

1.2 Main Research Questions and Contributions

This dissertation presents contributions to the fields of signal reconstruction and magnetic resonance imaging. The main objective is to develop compressive sensing methods that can improve signal reconstruction, by reducing the number of measurements required to attain a certain probability of reconstruction or a certain measure of signal quality (for instance, the signal-to-noise ratio will be relevant in evaluating the efficacy of the MR techniques). While these methods are based on the use of prior information, we also present an independent technique for improved reconstruction of MR images. We then show that this technique can

be combined with a prior-information-based approach, to further improve the final signal-to-noise ratios.

First, we propose a compressive sensing approach that uses prior information about the support of a signal's sparse representation. As a theoretical contribution, we mathematically prove that this type of information reduces the number of linear measurements required to guarantee the reconstruction, by ℓ_0-minimization, of signals that have a sparse representation in a known domain. Following, we present a theorem on the conditions for reconstruction with prior information when using convex optimization (specifically, ℓ_1-minimization). The basic idea is that these conditions should be more feasible than those of the case without prior information. Indeed, we prove that prior information increases the lower bound for the probability of correct reconstruction; as an important corollary, by adding prior information it is possible to reduce the number of used measurements and still keep that probability.

We also propose a practical method for signal reconstruction using prior information and based on the iteratively reweighted least-squares (IRLS). Particularly, we show that this method can use a specific weighting strategy in order to include prior information in an ℓ_p-minimization problem; this strategy increases the probability of reconstruction for the same number of linear measurements. The increase occurs for all tested values of p in the ℓ_p-minimization, so further improvement can be obtained by decreasing p and at the same time using prior information.

Regarding the mentioned MR imaging techniques, we present a prefiltering method for the image reconstruction with compressive sensing. In this method, the measurements corresponding to filtered versions of the image are computed from the available k-space samples; the filtered versions, rather than the desired image, are then reconstructed by an optimization procedure, and an image composition stage builds the final image from the filtered versions. We propose that this approach can improve the quality of the obtained images, by enhancing sparsity through filtering.

The proposed method also allows the inclusion of the prior information previously described, thus leading to a complete reconstruction scheme that combines two strategies:

prefiltering in the k-space domain and the use of prior information about the sparse representation (note that, in the case of medical images, the prior information can be obtained from medical records, from previous iterations in the reconstruction process or, in the case of functional MRI, from previous reconstructed images, as we will explore later). We then evaluate the performance improvement due to these two strategies, first separately, and then combined.

In order to evaluate the proposed compressive sensing and MRI methods, the analyses and experiments to be conducted must tackle the following questions:

1. What optimization problem must be solved in order to reconstruct a sparse signal from limited linear measurements, when using the described prior information about its sparse representation support? What are the sufficient and necessary conditions for the solution of this optimization problem to match the desired signal?

2. What is the reduction of the number of measurements required to reconstruct a sparse signal, as a function of the number of known positions in its sparse representation support?

3. What are the new reconstruction times, as a function of the number of known positions in its sparse representation's support?

4. What are the improvements in the signal-to-noise ratios of the MR images reconstructed using the proposed compressive sensing methods with and without prior information, as a function of the number of k-space samples?

5. What is the effect to the signal-to-noise ratios of combining prior information about the sparse representation support with wrong support locations, in the proposed MRI methods? This measures the system robustness to wrong prior information.

6. What is the effect to the signal-to-noise ratios of the number of known locations that are determined automatically, in the proposed MRI methods? This is related to the

previous question, as the locations determined automatically are prone to contain correct and wrong prior information.

Note that, in assessing the performance of MRI methods, the points above emphasize the signal-to-noise-ratios, the number of k-space samples taken, and the total reconstruction times. Since these figures of merit objectively evaluate physical system parameters, they correspond to level 1 in the hierarchical model of medical imaging efficacy, as proposed by Fryback and Thornbury [32], [87]. Hence, the choice for these particular parameters is based on their adequacy in measuring imaging efficacy, as well as their objective nature, so they can be determined from a signal processing point-of-view. Future research can also benefit from clinical tests to evaluate other levels of efficacy, as we will comment later in the text.

1.3 Structure of the Dissertation

The remaining of the dissertation is organized as follows. First, Chapter 2 presents a literature review of the main topics in the dissertation. We briefly present the state-of-the-art regarding compressive sensing and some modalities of medical imaging techniques, such as tomography and magnetic resonance (MR) imaging.

Next, Chapter 4 introduces the idea of prior information about the signal's support in a sparse representation. We show how this type of prior information can improve signal reconstruction by compressive sensing, an idea that is explored in later chapters in practical reconstruction algorithms and in enhanced MR imaging. Here, the theoretical development initially refers to the case in which the sparse representation and the measurements domain are related by the Fourier transform, as this directly applies to the proposed MR imaging methods later described; nevertheless, we experimentally extend the results to general sparse representations later in the text.

The aforementioned chapter starts with a brief overview of basic compressive sensing ideas, in order to define the required concepts for later development and to establish notation. Then, it presents the theory of how prior information reduces the minimum number

of measurements required to reconstruct a signal that has a sparse representation. Note that this theoretical lower bound corresponds to the reconstruction by ℓ_0-minimization; the more practical case of reconstruction by convex optimization is considered in a later section, where we show the necessary and sufficient conditions for reconstruction with prior information using ℓ_1-minimization. Based on these conditions, we then analyze the probability of reconstruction with prior information as a function of the number of linear measurements, and compare this probability with that of the case without prior information. Finally, the chapter provides a link between compressive sensing with prior information and the fractional Fourier transform.

In order to derive an efficient reconstruction algorithm, in Chapter 5 we show how the proposed prior information can be added to the reconstruction procedure by ℓ_p-minimization. This formulation leads to a weighting strategy for reconstruction by iteratively reweighted least-squares. We then analyze the behavior of the resulting implementation as a function of the number of known locations in the sparse representation and the number of linear measurements. We also evaluate the algorithm's robustness to changes in the additional parameter required by the new weighting strategy, as well as its performance in the presence of incorrect and incomplete prior information. Finally, the chapter presents a comparison of performance using direct and indirect methods with prior information.

Following, in Chapter 6 we propose a method for enhanced reconstruction of MR images using compressive sensing. The method is based on the prefiltering of the input signals in the k-space domain, in order to improve sparsity. Rather than reconstructing the final MR image directly, it then reconstructs filtered versions of it, and a final image composition stage builds the desired image from the filtered versions. After providing a brief review of the necessary concepts of MRI, the chapter fully describes the method, and three different schemes based on specific choices of filters; different configurations can also be used to reconstruct MR images that have sparse representations in different specific domains.

Chapter 7, on the other hand, presents the compressive sensing-based method for MR reconstruction using the proposed prior information, combined with the basic schemes de-

scribed in the previous chapter. First, it discusses how prior information about the signal's support in the sparse representation can be obtained in MR imaging applications, including the case of functional MRI. Next, it details how this type of prior information can be added to the method proposed in Chapter 6. We then illustrate the resulting system's performance in some theoretical examples, in order to compare the reconstruction for different numbers of known locations in the support and also its behavior when wrong prior information is used. The system's performance when applied to real functional MR imaging is considered later. Additionally, Chapter 7 presents a method for the fast computation of the matrix associated to the linear systems that need to be solved at each stage of the reconstruction. This is useful for a direct implementation of the optimization algorithms.

When presenting each of the chapters above, we already describe the first experiments testing the corresponding methods and their results, especially when their discussion motivates or justifies the developments of the other chapters. In Chapter 8, on the other hand, we specifically present additional experiments, for validation and in order to test more detailed aspects and more intricate relations between the different parts of the work. Individual sections present the results of compressive sensing with prior information, of the proposed method for MR Imaging and of MR Imaging with prior information.

Finally, Chapter 9 describes our conclusions and suggestions for future related work.

2

Literature Review

This chapter provides a brief literature review of the fields most closely related to this research. It starts, in Section 2.1, by presenting some of the publications that either originated or provided the most significant contributions to compressive sensing. It also discusses some recent publications that provide an overview or an introduction to this technique.

Following, Section 2.2 describes the works in medical imaging that influenced our investigations, with emphasis on magnetic resonance imaging (MRI), tomography, and methods to assess efficiency in medical imaging techniques. It also cites the works that describe traditional approaches to image reconstruction in MRI and tomography.

Finally, we mention the recent research on magnetic resonance imaging (MRI) and tomography using compressive sensing. The cited works emphasize how the problem of signal reconstruction in MRI and tomography can be posed in terms of compressive sensing, as well as the advantages of using this technique over other traditional methods.

We emphasize that the purpose at this stage is not yet to establish the mathematical foundation or the most technical, implementation-oriented aspects of the mentioned fields. Rather, this chapter sets the context of our research in terms of the recent publications, and indicates the recent scientific findings that support our own research. The technical

background of compressive sensing and MRI appears in the next chapters, when they are required by the proposed theories and methods.

2.1 Compressive Sensing

2.1.1 General Principles and Applications

Most digital signals of interest are highly compressible due to their inherent redundancies. Even high-resolution images can be stored and transmitted in ways that are orders of magnitude more efficient than simply coding every single pixel inside a frame. Linear transformations, for instance, can represent the same information by using fewer coefficients, and storing these coefficients and their locations is sufficient to later reconstruct the image in its original form, with high accuracy.

This type of redundancy has been explored in compression techniques for many years. This strategy aims at reducing the storage or transmission cost, by reducing the number of coefficients with respect to the number of samples initially acquired. It is thus appropriate in applications where signal acquisition is easier or less expensive than storing and transmitting.

In other applications, however, the most costly procedure is the acquisition itself (as an example, in magnetic resonance imaging for medical applications, the acquisition is inherently slow and expensive). In this case, the traditional procedure of acquiring a large number of samples even when significantly fewer could store the same information suggests a waste of resources (especially if after acquisition and imaging, significantly fewer coefficients are actually used). Compressive sensing is a recently-developed technique that allows to simplify the acquisition of signals that are known to have a sparse representation. The idea is not to acquire a large number of samples and later code them in a more efficient way, but rather to acquire fewer measurements in the first stages of the signals' acquisition itself, and in such a way that all the samples can be later computed by solving an optimization problem.

The origins of compressive sensing can be traced to the case of signals that are sparse in the time domain, and are reconstructed from samples in the Fourier domain (we discuss the

concept of sparsity in Chapter 3). In fact, Candès, Romberg et al. showed that if a length-N_p signal is η-sparse in time, meaning that only η time samples are nonzero, with $\eta << N_p$, then 2η Fourier coefficients are sufficient to reconstruct all the other $N_p - \eta$ coefficients, with theoretically perfect accuracy [10], [9] (to be precise, the statement in its original form requires N_p to be a prime number). This idea follows from an important result in harmonic analysis proven by Terence Tao in 2005; he showed in [86] that any square submatrix of the Fourier matrix of prime dimensions is invertible. Based on this statement, Candès, Romberg et al. proved that from a very reduced set of samples taken in the frequency domain, compared to traditional approaches, it is possible to represent all the frequency coefficients without ambiguity, provided that there is sparsity in the time domain.

This result already represents an important simplification, in terms of acquisition (not reconstruction), when compared to more traditional sampling approaches. Indeed, it allows acquiring signals and fully reconstructing them at sub-Nyquist rates, but only if the domain where the sampling occurs and the one where the signal is sparse are related by a Fourier transform.

Similar results were also obtained for general sparse representations. The most immediate difference when applying compressive sensing to these type of signals, where the Fourier transform or its inverse does not provide a sparse representation, is that, in this case, the measurements are not isolated signal samples. In fact, several researchers evaluated the types of measurements that would allow reconstruction in a general case. Candès and Tao introduced the concept of restricted isometry (RI) property [12] (which we review in Chapter 3), and showed that the solution to the reconstruction problem is stable if the so-called measurement matrix satisfies the RI property with respect to the sparsifying transformation. They also showed that if the measurement matrix is the realization of a Gaussian process with independent and identically distributed entries, than the RI property is satisfied with high probability for an arbitrary sparsifying transform.

It is interesting to observe that by using a random Gaussian matrix, one obtains measurements that are a random mixture of the signal's samples, rather than individual time

or frequency samples, as in the previous case. It may seem counter-intuitive that by randomly mixing the signal's samples it may be possible to acquire it using a reduced number of measurements, as compared to a traditional sampling method. However, this type of measurement matrix appears very frequently in the compressive sensing literature [16], [17], [18], [23]. Also, Baraniuk, Davenport et al. proved, in a different manner, that the random Gaussian measurement matrix satisfies the RI property with high probability for an arbitrary sparse representation, based on the Johnson-Lindenstrauss lemma [6].

Today, the focus of compressive sensing research is more on improving the optimization methods for signal reconstruction, in terms of reconstruction times, numbers of required measurements, and robustness to noise, rather than on defining efficient types of measurements (although some important results on this appeared recently, as we exemplify later). In fact, the ideal approach to the reconstruction problem in compressive sensing, in terms of the number of measurements required for signal reconstruction, is ℓ_0-minimization, as already stated in [10] (see also [5] and [23]). However, it can lead to combinatorial complexity, which is inviable even for signals considered relatively small in length. An important discovery in compressive sensing was that, under certain conditions (which we discuss in Chapter 3 and Chapter 4), the ℓ_1-minimization leads to the same result as the ℓ_0-minimization, but it has the advantage of being a convex optimization problem that can be solved by interior-point methods, and has polynomial complexity [72], [8]. Therefore, the ℓ_1-minimization approaches replaces combinatorial complexity by polynomial complexity, at the cost of more measurements being generally required. This result is discussed in detail in several publications [24], [23], [9].

Recent research also focus on other types of algorithms for signal reconstruction in compressive sensing. For instance, the more general ℓ_p-minimization approach allows a reduction of the number of required measurements, in comparison to ℓ_1-minimization, by reducing the parameter p [18], [67], [95]. Greeding pursuit and matching pursuit are common procedures as well; see, for instance, [40], [88], [89], and [90]. Also, see [36], [33], and [35] for discussions on fast implementations of reconstruction algorithms in compressive sensing, and on

convergence issues.

Additionally, many of the reconstruction methods and procedures in recent research focus on improving robustness to noise, especially investigating how additive noise in the limited number of measurements may have a proportionally large effect, after reconstruction, and how to minimize this problem [11]. The effect of noise to the reconstruction of sparse signals from limited measurements in compressive sensing is also investigated in [34].

Other recent developments in compressive sensing refer to the idea of jointly sparse signals, and the related concept of distributed compressive sensing (DSC). This concept exploits inter-signal and intra-signal correlations, in order to improve coding algorithms [25], [28], [31], [38].

We also emphasize that several recent publications in compressive sensing describe the developed theory more broadly, covering not only the most recent and advanced research topics, but also the basic theory starting at a more introductory level. See, for instance, [13], [5], [80] and chapter 13 of [59].

2.1.2 Compressive Sensing and Prior Information

As we mentioned previously, compressive sensing allows discrete-time signals having a sparse representation in some domain to be unambiguously represented by a limited number of linear measurements. While sparsity is the only signal characteristic that is usually assumed during reconstruction from a set of linear measurements [27], [29], other forms of prior information about the signal's structure have been investigated as a way to improve reconstruction. In [26] and [54], for instance, a connected tree structure is assumed in the wavelet domain, which restricts the class of signals that can be reconstructed to be piecewise smooth.

G. Chen, J. Tang and S. Leng, on the other hand, propose a method that uses prior images when reconstructing different frames in computed tomography (CT), using compressive sensing [19]. They suppose a sequence of images that don't change significantly from one frame to the next, so that the previous image is entirely used as prior information when reconstructing the next one. In order to do so, they do the reconstruction by minimizing

an objective function that combines a standard total-variation component with a weighted version of the previous image.

Note that the application in [19] resembles one we propose in magnetic resonance imaging (reconstructing images using prior information from previous frames); however, the type of prior information we introduce is different. While in [19] the prior information is a complete image (meaning the pixel values themselves), in our approach the prior information corresponds only to locations of high coefficients in the sparse representation, as we detail later. In fact, we make no assumption on the values on those locations, so that they can change from one frame to the next (due to noise or to changes in physiological activity).

In a different, nondeterministic formulation, Garcia-Frias and Esnaola show that if a signal is a realization of a stochastic process, prior information about this process (such as second order statistics) can improve the signal's reconstruction from a limited number of measurements [37], [30]. The prior information used in this approach refers to a statistical description of the source that generates the analyzed signals, and their proposed method is similar to joint source-channel coding of digital sources.

This dissertation, on the other hand, introduces a different type of prior information, represented by positions of the support of the signal's sparse representation, in order to reduce the required number of measurements and the computational complexity during the reconstruction stage. This particular aspect of the research already lead to two published articles (other publications are related to the medical imaging methods with and without prior information, as we mention later). First, von Borries, Miosso et al. introduced the basic concept and implemented a series of simulations, exemplifying the reductions in the number of required measurements in the context of DFT-sparse signals [92].

Also, Miosso, von Borries et al. showed how this type of prior information can be efficiently added to a reconstruction procedure by ℓ_p-minimization, based on an iteratively reweighted least squares method (IRLS) [65]. The proposed method improves reconstruction in terms of the number of measurements required to attain a percentage of correct reconstructions (up to a specified tolerance), compared to the IRLS without prior informa-

tion. Furthermore, it reduces the total reconstruction times and the number of iterations, in the tested direct method implementation.

Finally, another journal publication is under preparation [67], mathematically proving the conditions for signal reconstruction using prior information and the final reconstruction probabilities, as a function of the number of known support locations. This aspect of the dissertation is discussed in detail in Chapter 4.

2.2 Computed Tomography and Magnetic Resonance Imaging

In medical applications, magnetic resonance (MR) and computed tomography (CT) imaging aim at obtaining internal images of the body in a noninvasive manner. In this section, we briefly mention recent research on the mathematical problem of reconstructing the image from the set of measurements collected by the CT and MRI scanners. We also mention some references that focus on objective methods for evaluating efficiency of medical imaging techniques.

In magnetic resonance imaging (MRI), the reconstruction problem corresponds to computing a single image or a sequence of images from the set of measurements provided by an MR scanner. A strong static magnetic field and a radio frequency magnetic field are applied to the body to be examined, the first polarizing its internal protons while the second allows for an output magnetic moment.

The literature on medical imaging provides straightforward descriptions of the physics of MRI, and specially of the relation between the acquired measurements and the image to be reconstructed (which is relevant to the implementation of the imaging algorithms). Wright, for instance, describes how the MR signals are generated, and how the output magnetic moment can be made proportional to some local property of interest (corresponding to some clinically relevant anatomy or physiology) [96]. A similar description also appears in [57] and [58].

In the case of static MR images (as opposed to a sequence of images in functional MRI), the resulting signals collected at receiver coils in the scanner are samples of the bidimensional Fourier transform of the image to be reconstructed, and these samples belong to specific trajectories that depend on the type of magnetic pulses applied during the acquisition [57], [58].

Under this condition, the reconstruction problem corresponds to computing an image from a rather incomplete set of Fourier samples, and recent research describes several techniques that exist to handle this task. The interpolation techniques, for instance, base the reconstruction on estimating the complete Fourier transform of the desired MR image over a Cartesian grid, followed by an inverse Fourier transformation. Lehmann, Gonner et al. investigate several interpolation methods in the context of medical imaging [55], [56].

A second class of reconstruction methods is based on gridding algorithms, which start by convolving the available Fourier coefficients with a window function, for smoothing purposes, prior to an interpolation over a Cartesian grid and an inverse Fourier transform; a final stage compensates for the applied window function, yielding the desired image [82], [47], [74].

We also mention the filtered backprojection, which is a standard reconstruction method in nondiffraction tomography and which can be applied in MRI if the available Fourier samples are distributed over radial lines. Kak provides a now standard text on filtered backprojection [50]; he also conducts a comparative study of interpolation and filtered backprojection approaches applied to diffraction tomography [75].

In the case of CT, another numerical approach to the reconstruction problem is based on representing the Radon transform as a discrete operation over the image to be computed, and to modeling this operation in a matrix form. The reconstruction problem can then be viewed as a large-scale matrix equation, whose variable is the desired image, and the known terms are the measured projections and the matrix that models the Radon transform for the considered image dimensions. This approach to tomographic reconstruction, known as algebraic reconstruction technique (ART), is well described in [60]. See also [1] and [22] for a description of some advances in the basic technique, and [2] for the introduction of a refined version of it, called the simultaneous algebraic reconstruction technique (SART). Finally,

since the ART depends on representing the Radon transform as a discrete operation, see [46] for a possible discretization of this transform, based on spline convolutions.

Another promising reconstruction strategy, which has evolved significantly in the last few years, is based on iterative statistical procedures. In fact, although filtered backprojection is the most common tomographic technique in clinical applications, software packages by leading manufacturers now also include tools based on iterative statistical methods [21]. These methods, however, require significantly more computation time than filtered backprojection.

For a summary of the aforementioned traditional numerical methods in tomography and MRI, including backprojection, iterative methods and Fourier methods with interpolation, see [71] and [70]. Also, for a discussion of the convergence of the SART, see [48] and [49].

Compressive sensing provides yet another class of algorithms for the reconstruction of an MR image from the k-space samples. In fact, medical images are typically sparse under appropriate transformations [58], and, as previously mentioned, under this condition, compressive sensing can reduce the number of measurements required for reconstruction (in Chapter 1, we emphasized that signal acquisition in MRI is inherently slow, so minimizing the number of required measurements is of utmost importance). This possibility is well explored in [10] to reconstruct images that are piecewise constant – and therefore sparse under finite differencing – by using total-variation minimization.

Lustig, Donoho et al., on the other hand, proposed methods based on compressive sensing for reconstructing different types of MR images [58], [57]. Lustig also investigated what types of k-space trajectories are most appropriate to improve reconstruction of images having different types of sparsifying transforms. Specifically, he introduced a measure of how well the restricted isometry (RI) property is attained for different choices of sampling sets (k-space trajectories), and associated different sparsifying transformations to the tested trajectories that maximize the RI measures.

It is also interesting to observe that compressive sensing has been investigated as a potential tool in other medical applications, other than imaging. In [3], for instance, S. Aviyente proposed a technique that uses compressive sensing to acquire and compress electroen-

cephalogram (EEG) signals. The technique is based on the fact that EEG signals are sparse under the Gabor transformation. A matching pursuit algorithm reconstructs the signals from a limited number of projections.

In the context of compressive applied to MR imaging, we investigate, among other aspects, the use of prior information to enhance reconstruction, in terms of signal-to-noise ratios and computation times. In the past, prior information has been used in Bayesian approaches to medical imaging applications, when using traditional reconstruction techniques. In [45], for instance, K. M. Hanson and G. W. Wecksung suggest that the Bayesian approach can improve the reconstruction in computed tomography, especially when there is a limitation in terms of the angular range of the acquired measurements. This technique also appears in [44], in more detail and with additional results. Our purpose, on the other hand, is to combine prior information with compressive sensing, as we develop both the underlying theory and practical methods for MR imaging. The objective is to combine the improvement, in the case of limited linear measurements, by compressive sensing over traditional approaches with the improvement by a Bayesian approach.

Evaluating these different types of medical imaging techniques requires a tested method for assessing imaging efficacy. Fryback [32] and Thornbury [87] introduced the hierarchical model of diagnostic imaging systems' efficacy, as a guiding principle for methods' evaluation and comparison; this model is now traditionally used in technical imaging in clinical evaluations. Also, Krupinski reviewed the approach in a more recent publication [53]. The model includes several levels of diagnostic efficacy, and establishes the type of parameters to be assessed at each level. These levels are: technical efficacy, diagnostic accuracy efficacy, diagnostic thinking efficacy, therapeutic efficacy, and patient outcome efficacy. In evaluating our proposed MR method with and without prior information, we focus on the technical efficacy, which assess physical, objectively measurable parameters such as image quality metrics (e.g., signal-to-noise ratios) and reconstruction times. Evaluating the other levels, such as diagnostic accuracy and diagnostic thinking efficacies, is an interesting possibility for future related work, involving clinical trials.

3

Basic Concepts of Compressive Sensing

In digital signal processing applications, many N-dimensional signals can be unambiguously represented by fewer than N coefficients, by applying an appropriate linear transformation. This means that if a such a signal, represented by a length-N vector $\mathbf{x}$, is left-multiplied by an specific, invertible matrix $\mathbf{T}$, the resulting vector $\tilde{\mathbf{x}} = \mathbf{T}\mathbf{x}$ will have most of its N coefficients negligible in magnitude, when compared to the remaining η coefficients.

This fact is exploited, for instance, in signal compression methods. In traditional sampling/compression approaches, once all the N samples of $\mathbf{x}$ are acquired, the linear transformation $\tilde{\mathbf{x}} = \mathbf{T}\mathbf{x}$ is computed. The η most significant coefficients of $\tilde{\mathbf{x}}$ are then coded together with their locations, yielding 2η values to be stored or transmitted, instead of the $N > \eta$ original samples. As $\mathbf{T}$ is an invertible transformation, all the N original samples of $\mathbf{x}$ can later be approximately reconstructed, by composing an N-dimensional vector $\tilde{\mathbf{x}}$ with the η stored values in the corresponding locations, and with zeros in the remaining $N - \eta$ locations; an approximate version of $\mathbf{x}$ is then obtained by computing the inverse transformation $\mathbf{T}^{-1}\tilde{\mathbf{x}}$.

This compression approach, however, requires all the N samples of $\mathbf{x}$ to be initially acquired, even though only η will be later used for reconstruction. Compressive sensing, on the other hand, aims at reducing the number of measurements that are initially acquired. The basic idea is to take only ℓ linear measurements from $\mathbf{x}$, where $\ell < N$, but in such a way that all N samples can be computed from these measurements, with theoretically perfect accuracy. In fact, the reconstruction process is based on indirectly finding the sparsest signal that matches the available measurements, by using an optimization procedure. In this case, the acquisition procedure is simplified, at the cost of a more expensive reconstruction, when compared to the traditional compression approach.

As we will detail later, the key point behind signal reconstruction, in compressive sensing, is that although there are infinite signals that can generate the same $\ell < N$ linear measurements, only one will have the sparsest representation, provided that the number of measurements is high enough. A possible reconstruction method corresponds, therefore, to computing the signal with the lowest possible number of nonzero components that matches the available measurements. Directly computing this sparse representation, which leads to the ℓ_0-minimization that we describe later, can however lead to combinatorial complexity, and other optimization procedures, such as the ℓ_1-minimization and the ℓ_p-minimization, are used instead. We emphasize, at this point, that these methods are well-known tools (although still under development) in the numerical optimization literature, and many recent advances in compressive sensing correspond to showing how to apply them to reconstruct signals from limited measurements that can be obtained in practical contexts.

In this chapter, we provide an overview of the basic concepts of compressive sensing. The objective is to provide the background material before we introduce, in Chapter 4, the idea of support prior information, and to establish the notation we use throughout the text.

Regarding the applications of compressive sensing, we emphasize, both in this chapter and in the remaining of the dissertation, magnetic resonance imaging [58], [57], [91]. However, it is important to emphasize that other applications already appear in the compressive sensing literature as well, such as ground penetrating radar (GPR) imaging [43], photo-acoustic

tomography [76] and image and video coding [97].

We start by briefly describing, in Section 3.1, the concept of sparse representations, which is central to compressive sensing. Next, we discuss, in Section 3.2, how compressive sensing explores the sparsity condition in optimization procedures to reconstruct signals from reduced numbers of measurements. Section 3.3 then describes the restricted isometry property and the incoherence between the measurement vectors and the sparsifying bases; the first is a sufficient condition for the underdetermined reconstruction system to be stable, whereas the second is a related condition that guides the choice of measurement bases for a particular type of sparse representation. Finally, we mention in Section 3.4 the special case in which the sparse representation and the measurement domain are related by a discrete Fourier transform, given its importance to our proposed methods for magnetic resonance imaging.

3.1 Sparse Representations and the Acquisition Process

A length-N discrete-time random signal, represented by an $N \times 1$ vector $\mathbf{x}$, is said to have a sparse representation if and only if there exists a deterministic and invertible matrix $\mathbf{T}_{N \times N}$ such that the transformed vector

$$\tilde{\mathbf{x}} = \mathbf{T}\mathbf{x} \tag{3.1}$$

has most of its N components equal to zero. Under this condition, the total number of nonzero components in $\tilde{\mathbf{x}}$, here represented by η, is called the sparsity of $\tilde{\mathbf{x}}$. We can then write

$$\eta = |\{n \mid \tilde{x}[n] \neq 0\}|,$$

with $|\mathbb{A}|$ the cardinality of set $\mathbb{A}$, and state that $\tilde{\mathbf{x}}$ is an η-sparse vector. Finally, $\mathbf{T}$ is called a sparsifying transformation matrix of $\mathbf{x}$, and $\tilde{\mathbf{x}}$ is called a sparse representation of $\mathbf{x}$ in the domain defined by $\mathbf{T}$.

We emphasize that, throughout this dissertation, $\mathbf{x}$ and $\tilde{\mathbf{x}}$ can represent all the pixels of an image and its transformed version, in vectorized or stacked forms. They will represent, for instance, magnetic resonance images and their bidimensional-domain discrete Fourier transforms, or filtered versions of them.

Note that since the matrix $\mathbf{T}_{N\times N}$ is required to be invertible, if $\tilde{\mathbf{x}}$ is known, $\mathbf{x}$ can be unequivocally determined by computing an inverse transformation,

$$\mathbf{x} = \mathbf{T}^{-1}\tilde{\mathbf{x}}. \tag{3.2}$$

Then, knowing the N distinct entries of $\mathbf{x}$ is equivalent to knowing the $\eta < N$ nonzero entries of $\tilde{\mathbf{x}}$ and their locations, so that the inverse transformation in (3.2) can be computed.

In practice, the condition that $N - \eta$ coefficients of $\tilde{\mathbf{x}}$ are equal to zero is often relaxed; instead, they are required to be close to zero, when compared, in magnitude, with the other η entries in $\tilde{\mathbf{x}}$. In this case, $\mathbf{x}$ is called a compressible signal.

In fact, having a representation $\tilde{\mathbf{x}}$ with most components close to zero, and in a domain defined by a known transformation matrix $\mathbf{T}$, is the basis for the so-called transformation-based compression methods. These methods exploit this type of representation in order to store all the information needed to completely reconstruct a signal $\mathbf{x}$ by coding only η values and their locations in the transformed domain. The basic procedure behind these methods is then: acquire a signal $\mathbf{x}$, at a Nyquist rate, so that all N samples are immediately available; compute the linear transformation $\tilde{\mathbf{x}} = \mathbf{T}\mathbf{x}$, which is a representation of $\mathbf{x}$ in a known domain; locate the η highest coefficients; store (or transmit) these η coefficients and their locations.

With this compression approach, the problem of storing or transmitting N values is replaced by operating on 2η values only, where we assume $N > 2\eta$. Also, the reconstruction of the original N components of $\mathbf{x}$ is straightforward: since the η highest (in magnitude) values of of $\tilde{\mathbf{x}}$ are available, together with their locations, assume zeros in the remaining locations, thus building the full vector $\tilde{\mathbf{x}}$; then, compute the inverse transformation of $\tilde{\mathbf{x}}$, to obtain an approximate reconstructed $\mathbf{x}$ according to (3.2). The error introduced by the whole compression procedure, in this context, corresponds to the total energy of the $N - \eta$ lowest (in magnitude) coefficients of $\tilde{\mathbf{x}}$, supposing an energy-preserving transformation $\mathbf{T}$.

Note that this compression strategy aims at reducing the storage or transmission cost, by reducing the number of coefficients with respect to the number of samples initially acquired. It is thus appropriate in applications where signal acquisition is easier or less expensive than storing and transmitting, and where reconstructing the signals has to be straightforward (in fact, a single inverse transformation).

In other applications, however, the most costly procedure is the acquisition itself. This is the case, for instance, of magnetic resonance imaging (MRI) and tomography. As we discuss in Chapter 1 and Chapter 6, in these applications, the number of measurements that can be acquired is limited by hardware, time, and physiological constraints. The storage of the acquired measurements, on the other hand, is hardly a problem with the technology available for this type of application, even more so when compared to the cost of the expensive medical MR scanners. In this case, the traditional compression procedure of acquiring N samples even when it is known that 2η coefficients could store the same information, after the proper transformation and coding, suggests a waste of resources. Compressive sensing, in a similar context, was developed to deal with the following question: isn't it possible to acquire all the information necessary to reconstruct an N-length signal, by taking fewer than N linear measurements, if we know that this signal is compressible (which means that it can be represented, under transformation, by fewer coefficients)?

Compressive sensing is then aimed at simplifying the acquisition process itself, and still allowing the reconstruction of all the N components of the compressible signal $\mathbf{x}$, at the cost of a computationally more complex reconstruction procedure. In Section 3.2, we briefly review how traditional compressive sensing allows a compressible N-dimensional signal to be reconstructed from fewer linear measurements, with theoretically perfect accuracy. We emphasize that the whole procedure is based on solving an underdetermined linear system (after all, fewer than N measurements are used to compute N values), using as a constraint the fact that the signal is sparse in a known domain.

3.2 Reconstruction from Fewer Linear Measurements by Constrained Optimization

Compressive sensing is a technique that allows the reconstruction of signals with sparse representations in known domains by using limited linear measurements and with theoretically perfect accuracy. The idea is to simplify the acquisition process over standard sampling methods, in such a way that a comparatively small number of measurements is taken, and the full signal is reconstructed later through an optimization procedure.

In fact, consider an N-dimensional complex discrete-time signal $\mathbf{x} : \mathbb{D} \to \mathbb{C}$, where the domain set is given by $\mathbb{D} = \{0, 1, \ldots, N-1\}$ and where $\mathbb{C}$ is the set of complex numbers. If $\mathbf{x}$ is compressible by linear transformation, with a sparse representation given by (3.1), then the theory of compressive sensing states that $\mathbf{x}$ can be fully reconstructed from an ℓ-dimensional vector of measurements given by

$$\mathbf{b} = \mathbf{M}\mathbf{x},$$

with $\mathbf{M}_{\ell \times N}$ the so-called measurement matrix, provided that certain conditions are satisfied and by solving an appropriate optimization problem. For instance, the ℓ_1-minimization approach is a commonly used reconstruction procedure, and if the measurement matrix satisfies the restricted isometry (RI) property with respect to the sparsifying transformation, then the stability of the solution is guaranteed. In Section 3.3 we briefly review the RI property and the related incoherence condition.

Once the measurements $\mathbf{b}$ are available, obtaining $\mathbf{x}$ corresponds to finding the sparsest vector $\tilde{\mathbf{x}}$ that satisfies those measurements, and then computing the inverse transform of $\tilde{\mathbf{x}}$. By defining $\|\tilde{\mathbf{x}}\|_0$ (the ℓ_0 of $\tilde{\mathbf{x}}$) as the number of nonzero components in $\tilde{\mathbf{x}}$, the ideal reconstruction problem (the one that can reconstruct $\mathbf{x}$ unambiguously and with the minimum possible number of measurements) can then be written as [10], [13]

$$\begin{aligned} &\min \|\tilde{\mathbf{x}}\|_0 \\ &\text{subject to } \mathbf{A}\tilde{\mathbf{x}} = \mathbf{b}, \end{aligned} \tag{3.3}$$

where $\mathbf{A} = \mathbf{M}\mathbf{T}^{-1}$.

Note, however, that directly solving for the minimum ℓ_0 can lead to combinatorial complexity, which is not viable even for moderately sized signals [10]. An important discovery in compressive sensing is that, under specific circumstances, the ℓ_1-minimization approach can lead to the same solution as the ℓ_0 minimization [23], [24], [5]. In this case, the combinatorial complexity of the ℓ_0 approach is replaced by a polynomial complexity, by solving

$$\begin{aligned} &\min \|\tilde{\mathbf{x}}\|_1 \\ &\text{subject to } \mathbf{A}\tilde{\mathbf{x}} = \mathbf{b}, \end{aligned} \tag{3.4}$$

which is a convex optimization problem that can be tackled using interior point methods. In practice, reconstructing $\mathbf{x}$ using (3.4) generally requires more measurements than using (3.3) directly, so the computational complexity is reduced at the cost of taking more measurements.

Another possible approach for signal reconstruction in compressive sensing is solving a nonconvex optimization problem, like the ℓ_p-minimization. In this particular case, the problem in (3.4) is generalized by minimizing the ℓ_p of $\tilde{\mathbf{x}}$, instead of ℓ_1, so that (3.4) becomes

$$\begin{aligned} &\min \|\tilde{\mathbf{x}}\|_p^p \\ &\text{subject to } \mathbf{A}\tilde{\mathbf{x}} = \mathbf{b}, \end{aligned} \tag{3.5}$$

with $0 < p \leqslant 1$.

Note that (3.4) is a special case of (3.5), with $p = 1$. Therefore, the ℓ_p-minimization approach includes the ℓ_1-minimization as a special case, but reducing p has some important advantages; for instance, it allows reducing the number of measurements required for reconstruction, with respect to the case $p = 1$ [18], [14]. In Chapter 5, for instance, we use an iteratively reweighted least squares method combined with prior information about the sparse representation's support in order to reconstruct signals in compressive sensing using the ℓ_p-minimization approach; we show that this type of prior information further enhances reconstruction, both in terms of required number of measurements, and in terms of computation times.

3.3 Restricted Isometry Property and Incoherence

Reconstructing a length-N signal from a set of ℓ measurements, with $\ell << N$, is usually an ill-conditioned problem. However, the restricted isometry (RI) property, first defined by Candès and Tao [12] and which we summarize below, is sufficient for a stable solution.

Let $\mathbf{x}$ be a length-N signal with an η-sparse representation $\tilde{\mathbf{x}} = \mathbf{T}\mathbf{x}$. A measurement matrix $\mathbf{M}_{\ell \times N}$ is said to satisfy the RI property with respect to $\mathbf{T}$ and with a tolerance $\epsilon > 0$ if and only if [5], [6]

$$1 - \epsilon \leqslant \frac{\|\mathbf{M}\mathbf{T}^{-1}\mathbf{v}\|_2}{\|\mathbf{v}\|_2} \leqslant 1 + \epsilon, \tag{3.6}$$

for any N-dimensional vector $\mathbf{v}$ that is 3η-sparse.

Note that (3.6) can be interpreted in the following manner: the lower the tolerance ϵ, the closer to 1 the fraction $\|\mathbf{M}\mathbf{T}^{-1}\mathbf{v}\|_2/\|\mathbf{v}\|_2$ must be. Now, $\|\mathbf{M}\mathbf{T}^{-1}\mathbf{v}\|_2/\|\mathbf{v}\|_2$ being close to 1 means that the matrix $\mathbf{M}\mathbf{T}^{-1}$ practically does not modify the norm of any 3η-sparse vector when multiplied by it. Therefore, what the RI property states is that if the measurement matrix $\mathbf{M}$, when multiplied by $\mathbf{T}^{-1}$ times an arbitrary 3η-sparse vector [5], maintains this vector's 2-norm (up to a tolerance ϵ), then the measurements acquired from an η-sparse vector $\mathbf{x}$ using this measurement matrix are sufficient for a stable reconstruction of $\mathbf{x}$. Also, the lower the tolerance we can adopt, the more stable the solution.

The second condition we mentioned, which is related to the RI property, is that of incoherence between the rows of the measurement matrix and the bases that sparsify the signal $\mathbf{x}$. The incoherence condition states that the Hermitian of each row of the measurement matrix does not have a sparse representation in the domain defined by $\mathbf{T}$. In other words, represent the measurement matrix by

$$\mathbf{M} = [\mathbf{m}_1 \ \mathbf{m}_2 \ \dots \ \mathbf{m}_\ell]^H,$$

where $\mathbf{m}_i^H$ is the i-th row of $\mathbf{M}$. The incoherence condition states that, if $\mathbf{T}\mathbf{x}$ is sparse, then $\mathbf{T}\mathbf{m}_i$ should not be sparse, for any i.

We emphasize that the incoherence condition is not sufficient for reconstruction. Rather, it is a necessary for the RI property to hold (the RI property itself being sufficient, and not necessary, for the solution stability).

Note that the RI and the incoherence conditions depend on the sparsifying transform. Hence, at first it seems that a particular measurement matrix has to be chosen according to the type of domain where the signal to be reconstructed is sparse. However, there is a class of measurement matrices that satisfy these conditions, with high probability, for arbitrary sparsifying transformations. Indeed, Candès and Tao showed that if $\mathbf{M}$ is built randomly, with the entries independent and identically distributed with Gaussian distribution, than the RI property is satisfied with high probability, independently on the sparsifying transformation, provided that [12], [10], [13]

$$\ell > \lambda \eta \log\left(\frac{N}{\eta}\right), \tag{3.7}$$

with λ a specified constant. This result was also proven by Baraniuk, Davenport et al. in [6].

Based on this described property, and on (3.7), note that choosing a random $\ell \times N$ matrix with Gaussian independent entries leads to a high probability of reconstruction if the number of measurements ℓ is high enough (we will discuss the relation between the number of measurements and the probabilities of reconstruction, under specific conditions, in Chapter 7). As this works for a general sparsifying transform, this type of random measurement matrix is considered universal, and that is why today it is largely used in compressive sensing applications and research [12], [13], [23], [28], [5].

3.4 The Special Case of Sparse Fourier Transforms

A particular case of interest, which has application in our proposed magnetic resonance imaging method, is that in which the measurement domain and some sparse representation $\tilde{\mathbf{x}}$ are related by a Fourier transform, that is,

$$\tilde{\mathbf{x}} = \mathrm{DFT}(\mathbf{x}), \text{ or} \tag{3.8}$$

$$\tilde{\mathbf{x}} = \text{IDFT}(\mathbf{x}), \tag{3.9}$$

with DFT and IDFT the direct and the inverse discrete Fourier transforms, respectively. In this situation, if all rows of $\mathbf{M}$ are independent spike vectors, then the incoherence property automatically holds. In fact, the Fourier transform of any spike vector is a full vector (which means that the rows of $\mathbf{M}$, which are simple spikes, are not sparse under the same transformation that sparsifies the signal to be reconstructed). Also, for this choice of $\mathbf{M}$ the corresponding measurements are simply individual samples of the vector in the measurement domain, meaning

$$\mathbf{b} = \mathbf{M}\mathbf{x} = \mathbf{x}_{n\in\mathbb{S}} \tag{3.10}$$

where $\mathbf{x}_{n\in\mathbb{S}}$ denotes the ℓ-dimensional vector whose components are individual samples from $\mathbf{x}$, taken in the locations belonging to set $\mathbb{S}$ (note that n belongs to $\mathbb{S}$ if and only if one of the rows of $\mathbf{M}$ has a single nonzero component, located in column n and equal to one).

The condition described by (3.8), (3.9) and (3.10) is of great importance to this research due to the MRI and tomography applications. It is also interesting in the general context of signal reconstruction from undersampled and irregularly sampled data, as the measurements are individual samples, at arbitrary positions, from the desired signal (in fact, $\mathbf{b}$ corresponds to an arbitrary set of samples from $\mathbf{x}$, acquired at a sub-Nyquist rate). It is also worth mentioning that the origins of compressive sensing can be traced to the reconstruction of sparse signals from individual DFT samples [10], [23]; later, the theory was extended to more general sparse representations.

Another interesting aspect of compressive sensing, in the context of DFT-sparse signals and of (3.10), is that this particular context can be viewed as one of the possible generalizations of discrete-time signals' undersampling with reconstructibility (it can also be viewed as a generalization of the sampling of a continuous-time signal, where the samples can be taken nonuniformly and at fewer locations). In fact, when a discrete-time signal is undersampled at regular intervals, the Fourier transform (FT) of the resulting signal is the sum of shifted versions of the original FT [68]. Hence, if the original spectrum has an interval

where the FT is zero, and in such a way that this interval allows the nonzero regions not to overlap with the shifting, the original signal can be recovered by zero-filling the undersampled signal followed by ideal low-pass filtering. Compressive sensing generalizes this idea in two ways: (1) by allowing the undersampling to be done irregularly (or even by 'mixing' the original samples through incoherent projections), (2) by allowing the FT not to be zero in a contiguous interval – instead, the DFT must contain several zeros, which can however be distributed arbitrarily over the spectrum. Note that a further generalization of the idea of undersampling and of continuous-time signals sampling then occurs when we consider again the sparsity in a different domain, rather the in the DFT representation, as the basic requirement of a region of zeros in the FT is itself removed.

In the case of (3.8), (3.9) and (3.10), the optimization problems (3.3) and (3.4) can be written respectively as

$$\begin{aligned} &\min \|\tilde{\mathbf{x}}\|_0 \\ &\text{subject to } \mathcal{F}\left(\tilde{\mathbf{x}}\right)_{n\in\mathbb{S}} = \mathbf{b}, \end{aligned} \tag{3.11}$$

and

$$\begin{aligned} &\min \|\tilde{\mathbf{x}}\|_1 \\ &\text{subject to } \mathcal{F}\left(\tilde{\mathbf{x}}\right)_{n\in\mathbb{S}} = \mathbf{b}, \end{aligned} \tag{3.12}$$

where $\mathcal{F}$ is the inverse of the transformation from $\mathbf{x}$ to $\tilde{\mathbf{x}}$, that is, $\mathcal{F}$ is the IDFT in the case of (3.8) and the DFT in the case of (3.9).

As previously explained, choosing $\mathbf{M}$ as an $\ell \times N$ matrix whose rows are independent spikes automatically guarantees the incoherence condition, for the case of (3.8) and that of (3.9). However, incoherence is not sufficient for reconstruction, as we mentioned in Section 3.3.

In Chapter 4, we analyze in detail the sufficient and necessary conditions for reconstruction by ℓ_1-minimization, specifically in the case of (3.8) and that of (3.9), given its importance to the MRI application. Furthermore, we introduce the concept of prior information about the sparse representation's support and show how it can be incorporated into the reconstruction procedure. This leads to a modification in the conditions for reconstruction, and

we prove that the new conditions are more feasible than those without prior information. In fact, we show that the probabilities of reconstruction increase, for the same number of measurements ℓ, with the number of known support locations in the sparse domain.

4

Compressive Sensing with Prior Information

This chapter summarizes the main proposed theory for compressive sensing with prior information. As previously mentioned in Chapter 3, compressive sensing allows the reconstruction of a signal having a sparse representation in a known domain by using fewer linear measurements than traditional approaches, and with theoretically perfect accuracy. We now show that by using prior information on the support region of the sparse representation, it is possible to reduce the minimum number of measurements that leads to perfect reconstruction.

We start by proposing a generalization of compressive sensing, by showing how information on the sparse domain can be used during the reconstruction procedure. This result is first discussed from a theoretical point of view, when we show the reduction of the required number of samples that must be taken from a signal with a sparse discrete-time Fourier transform (DFT). At this point, the emphasis on the Fourier transform is motivated by our proposed method for magnetic resonance imaging; in fact, as we discuss in Chapter 6, in this method we use a prefiltering strategy to obtain sparse components that are related to the measurements domain by this transform. Following the theoretical formulation, we present

an extensive series of simulations in which signals that are sparse in the DFT domain or in a general random domain are reconstructed with and without using prior information.

Two important aspects are particularly relevant when comparing compressive sensing with and without prior information. The first refers to the number of initial linear measurements ℓ that must be taken from a signal to theoretically guarantee unambiguous reconstruction using compressive sensing, as a function of the number η of non-null components in the sparse representation. The second aspect is how the reconstruction itself can be done, i.e., given the ℓ measurements, how to determine all N samples in the original domain.

In this chapter, we show that the theoretical minimum bound for the number of samples needed to reconstruct a DFT-sparse signal decreases if at least some of the frequencies of the nonzero DFT values are known. We also show how the conditions for a signal to be reconstructible using ℓ_1-minimization change when we use this same type of partial prior information. Next, we prove that the Chebychev lower bound for the probability of attaining these conditions increases with the number of known support locations. This means, for instance, that the probability of reconstruction by ℓ_1-minimization is higher when prior information is used. Finally, we describe simulation tests that verify the practical reduction of the number of needed samples when the reconstruction is conducted using ℓ_1-minimization.

The remaining of the chapter is organized as follows. Section 4.1 extends the basic compressive sensing technique by showing that the theoretical lower limit for the needed number of required measurements ℓ is reduced when information on the support of the sparse domain is used. Also shown is the corresponding generalization of the reconstruction problem, and an interpretation of the fractional Fourier transform (FrFT) in the context of this generalized compressive sensing approach. The FrFT is shown to correspond to one extreme of the proposed technique, whereas the initial compressive sensing appears on the other extreme. Section 4.2 presents the conditions under which an ℓ_1-minimization procedure can be used to reconstruct a signal, supposing there is partial prior information about its support in the sparse representation. It also discusses the Chebychev lower bound for the probability of attaining these conditions. Section 4.3 describes an important numerical issue

regarding the reconstruction problem, both in the original compressive sensing framework and when additional information on the position of the non-null sparse coefficients is used. Section 4.4 then presents the results of two series of extensive numerical simulations, one related to signals that are sparse in the DFT domain, the other to signals for which a different, general orthogonal transform provides a sparse representation. Finally, Section 4.5 presents some remarks on the presented theory and on the first simulations.

In Chapter 8, we present additional experiments that validate the theory and methods presented here. We also discuss in more detail the reduction of the number of measurements when prior information is used in the reconstruction.

4.1 Prior Information on the Sparse Domain's Support

Compressive sensing can be first understood in the context of signals which have sparse DFTs, and later extended to other sparse representations. An initial theorem [10] asserts that if a signal $\mathbf{x}$, with prime length N_p, has only η nonzero samples, then only 2η coefficients from its DFT are theoretically needed to unambiguously reconstruct all N_p samples.

Also, a slightly different procedure can be used for signals which do not have a sparse DFT, but which do have a sparse representation in other domain, say $\hat{\mathbf{x}} = \mathbf{T}\mathbf{x}$ for some known orthonormal transformation matrix $\mathbf{T}$. In this case, instead of just taking isolated samples of $\mathbf{x}$, one must consider *linear measurements* of $\mathbf{x}$ as the input to the compressive sensing scheme. These measurements are defined as projections of $\mathbf{x}$ onto a set of linearly independent N_p-dimensional vectors $\mathbf{v}_i \; \forall \; i \in \{1, 2, \ldots, \ell\}$. Note that these independent vectors are required not to have a sparse representation on the same domain defined by the transformation matrix $\mathbf{T}$, a condition referred to as incoherence [12]. In other words, let the vectors $\mathbf{v}_i$ be such that $\mathbf{T}\mathbf{v}_i$ is not sparse, $\forall \; i \in \{1, 2, \ldots, \ell\}$, meaning that the transformation matrix $\mathbf{T}$ does not provide a sparse representation for any of the vectors $\mathbf{v}_i$. In this

case, the ℓ linear measurements of $\mathbf{x}$ must be given by

$$\mathbf{y}_{\ell\times 1} = \mathbf{M}_{\ell\times N_p}\ \mathbf{x}_{N_p\times 1,}\ \text{with}\ \ell < N_p, \tag{4.1}$$

where $\mathbf{M} = [\mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3\ \ldots \mathbf{v}_\ell]^T$ is the *measurement matrix*, with T the transposition operation. Note that (4.1) represents a generalization of the procedure of taking isolated samples of $\mathbf{x}$ when the DFT of $\mathbf{x}$ is sparse; an appropriate choice of $\mathbf{M}$, with each row a distinct canonic vector, results in $\mathbf{y}$ being a vector of isolated samples of $\mathbf{x}$.

Next, we address the reduction of the number of required measurements by using prior information on the support region of the sparse domain, and the reconstruction by ℓ_1-minimization. In Section 4.2, we show the conditions under which a signal with sparse representation can be reconstructed by ℓ_1-minimization with prior information.

4.1.1 Reduction in the Number of Required Measurements

A critical point is defining how many linear measurements are necessary for the unambiguous determination of the signal $\mathbf{x}$ that generated those measurements. As previously mentioned, if this signal has prime length and is sparse in the DFT domain, so that the linear measurements may be taken as isolated time samples, a theorem provided in [10] shows that the number of signal samples has to be at least twice the number of non-null DFT coefficients. More precisely, Theorem 1.1 in [10] assumes that the signal is sparse in the time domain, and that the samples are taken in the DFT domain; however, given the duality aspect of this transform, the same relation holds when the sparsity occurs in the frequency domain and the samples are taken in time domain. In practice, more coefficients are usually needed as the ℓ_0-minimization to which the lower bound corresponds is substituted by a computationally more tractable optimization approach, such as the ℓ_1-minimization.

In Section 4.2, we show the conditions for a sparse signal to be reconstructible by ℓ_1-minimization when we use prior information about the sparse representation. First, we show that the theoretical lower bound for the number of needed samples is reduced if at least partial information is available, and still supposing the length of the signal to be prime. In

other words, if the positions of at least some nonzero coefficients are known, the number of samples required for unambiguous reconstruction is less than twice the number of nonzero coefficients. This idea was introduced by the authors in [92] and [67] for the DFT-sparse case, but here we provide the formal proof, analyze the new conditions for reconstruction using ℓ_1-minimization, and conduct an extensive set of simulations. Again, in practice more coefficients than the theoretical limit may be required due to the optimization algorithms involved, but even in this situation the required number of samples is fewer than that of the initial compressive sensing approach, as we will discuss in Section 4.4.

The theorem we show, establishing the reduction in the number of measurements required for perfect reconstruction, is a consequence of the invertibility of square submatrices of the $N_p \times N_p$ DFT matrix, where N_p is a prime number. Hence, before our result, we present, in Theorem 1, the statement of this invertibility. An elegant proof of this condition is found in [86]; here we use a different notation, for consistency with the rest of the dissertation.

Theorem 1. *(from Lemma 1.3 in [86], with a different notation) Let W_{N_p} be the N_p-th complex root of unity given by $\exp(-j2\pi/N_p)$, so that the DFT of a length-N_p signal $\mathbf{x}$ is given by $\hat{\mathbf{x}} = \mathbf{F}\,\mathbf{x}$, where $\mathbf{F} = \{W_{N_p}^{nk}\}_{0 \leqslant n,k \leqslant N_p-1}$ is the $N_p \times N_p$ DFT matrix, with $W_{N_p}^{nk}$ the entry of the n-th row and k-th column, and where $\mathbf{x}$ and $\hat{\mathbf{x}}$ are represented as column vectors, meaning that*

$\mathbf{x} = [x[0] \;\; x[1] \;\; x[2] \;\; \ldots \;\; x[N_p-1]]^T$ *and* $\hat{\mathbf{x}} = [\hat{x}[0] \;\; \hat{x}[1] \;\; \hat{x}[2] \;\; \ldots \;\; \hat{x}[N_p-1]]^T$.

If N_p is a prime number and $\mathbb{A}$ and $\mathbb{B}$ are subsets of $\{0, 1, 2, \ldots, N_p-1\}$ that have the same cardinality ℓ, i.e., $|\mathbb{A}| = |\mathbb{B}| = \ell$, then the $\ell \times \ell$ submatrix $\mathbf{F}_{\mathbb{A},\mathbb{B}}$ defined by

$$\mathbf{F}_{\mathbb{A},\mathbb{B}} = \{W_{N_p}^{nk}\}_{\text{all } n \in \mathbb{A},\ \text{all } k \in \mathbb{B}}$$

is invertible.

By using this important result in harmonic analysis, we can prove the following Theorem 2, which establishes the reduction of the number of frequency coefficients needed to reconstruct a time-sparse signal when some of the positions of the non-null time-samples are known. In other words, Theorem 2 assumes that the signal is sparse in the time domain,

whereas the measurements (in this case, isolated coefficients) are taken in the DFT domain. An extension is then provided for the reverse case, namely when the DFT is sparse and samples are taken in time domain.

Theorem 2. *Let* $\hat{\mathbf{x}}: \mathbb{G} \to \mathbb{C}$*, with* $\mathbb{C}$ *the set of complex numbers, represent the DFT of a discrete-time signal with length* N_p*, meaning that* $\mathbb{G} = \{0, 1, 2, \ldots, N_p - 1\}$ *with* N_p *a prime number. Also, let* $\mathbb{L}$ *be a non-empty subset of* $\mathbb{G}$*, with cardinality* $|\mathbb{L}| = \ell \leqslant N_p$*. If we define a second subset* Φ *of* $\mathbb{G}$*, such that its cardinality* $|\Phi| = \varphi$ *satisfies* $0 \leqslant \varphi \leqslant \ell$*, then there exists at most one signal* $\mathbf{x}_s: \mathbb{G} \to \mathbb{C}$ *satisfying the conditions below.*

1. *The support of* $\mathbf{x}_s$*, denoted by* $\text{supp}(\mathbf{x}_s)$*, has at most* $(\ell + \varphi)/2$ *elements and contains the subset* Φ *of* φ *elements:* $|\text{supp}(\mathbf{x}_s)| \leqslant (\ell + \varphi)/2$*, with* $\Phi \subseteq \text{supp}(\mathbf{x}_s)$*.*

2. *The discrete-time Fourier transform of* $\mathbf{x}_s$*, here represented by* $\hat{\mathbf{x}}_s$*, is such that*

$$\hat{x}_s[k] = \hat{x}[k] \;\; \forall\, k \in \mathbb{L},$$

where $\hat{x}_s[k]$ *and* $\hat{x}[k]$ *denote the coefficients of* $\hat{\mathbf{x}}_s$ *and of* $\hat{\mathbf{x}}$ *in position* k*, respectively.*

Proof. Suppose that there exist two signals $\mathbf{x}_s$ and $\mathbf{y}_s$ which satisfy the conditions 1 and 2 in the statement of Theorem 2. We will show that in this case $\mathbf{x}_s = \mathbf{y}_s$, so there is at most one signal under those conditions. By condition 1, the supports of $\mathbf{x}_s$ and $\mathbf{y}_s$ have at most $\dfrac{\ell + \varphi}{2}$ elements each, that is,

$$|\text{supp}(\mathbf{x}_s)| \leqslant \frac{\ell + \varphi}{2} \tag{4.2}$$

and

$$|\text{supp}(\mathbf{y}_s)| \leqslant \frac{\ell + \varphi}{2}. \tag{4.3}$$

Then, consider the set defined by

$$\mathbb{D} = \text{supp}(\mathbf{x}_s) \cup \text{supp}(\mathbf{y}_s),$$

which has cardinality

$$|\mathbb{D}| = |\text{supp}(\mathbf{x}_s)|| + |\text{supp}(\mathbf{y}_s)| - |\text{supp}(\mathbf{x}_s) \cap \text{supp}(\mathbf{y}_s)|. \tag{4.4}$$

By using (4.2) and (4.3) in (4.4), we obtain

$$|\mathbb{D}| \leqslant \frac{\ell+\varphi}{2} + \frac{\ell+\varphi}{2} - |\mathrm{supp}(\mathbf{x}_s) \cap \mathrm{supp}(\mathbf{y}_s)|,$$

which reduces to

$$|\mathbb{D}| \leqslant \ell + \varphi - |\mathrm{supp}(\mathbf{x}_s) \cap \mathrm{supp}(\mathbf{y}_s)|.$$

Now, note that since $\mathbf{x}_s$ and $\mathbf{y}_s$ are supposed to satisfy condition 1 in the theorem, $\Phi \subseteq \mathrm{supp}(\mathbf{x}_s)$ and $\Phi \subseteq \mathrm{supp}(\mathbf{y}_s)$, so $\Phi \subseteq (\mathrm{supp}(\mathbf{x}_s) \cap \mathrm{supp}(\mathbf{y}_s))$. Then,

$$|\mathrm{supp}(\mathbf{x}_s) \cap \mathrm{supp}(\mathbf{y}_s)| \geqslant |\Phi| = \varphi$$

and we obtain

$$|\mathbb{D}| \leqslant \ell + \varphi - \varphi,$$

or simply

$$|\mathbb{D}| \leqslant \ell. \tag{4.5}$$

The inequality in (4.5) indicates that set $\mathbb{D}$ has at most ℓ elements. Then, we can define a set $\mathbb{A}$ of ℓ distinct elements $\{n_1,\ n_2,\ \ldots\ n_l\}$ which includes all elements of $\mathbb{D}$ and $\ell - |\mathbb{D}|$ additional distinct elements from the original set $\mathbb{G}$. In other words, $\mathbb{A}$ is the set of all positions where $\mathbf{x}_s$ or $\mathbf{y}_s$ can be non-zero, plus possibly some other positions where they both are zero. Then, consider the vectors $\mathbf{x}_a$ and $\mathbf{y}_a$ given by $\mathbf{x}_a = [x_s[n_1]\ \ x_s[n_2]\ \ \ldots\ \ x_s[n_l]]^T$ and $\mathbf{y}_a = [y_s[n_1]\ \ y_s[n_2]\ \ \ldots\ \ y_s[n_l]]^T$.

The important aspect of $\mathbf{x}_a$ and $\mathbf{y}_a$ is that they include all non-zero values of $\mathbf{x}_s$ and $\mathbf{y}_s$, respectively. Hence, with $W_{N_p} = \exp(-j2\pi/N_p)$ and defining the square matrix

$$\mathbf{F}_{\mathbb{L},\mathbb{A}} = \{W_{N_p}^{nk}\}_{\text{all } k\in\mathbb{L},\, \text{all } n\in\mathbb{A}},$$

which is an $\ell \times \ell$ submatrix of the DFT matrix $\mathbf{F}$, we conclude that $\hat{\mathbf{x}}_a$ and $\hat{\mathbf{y}}_a$ defined by

$$\hat{\mathbf{x}}_a = \mathbf{F}_{\mathbb{L},\mathbb{A}}\, \mathbf{x}_a \tag{4.6}$$

and

$$\hat{\mathbf{y}}_a = \mathbf{F}_{\mathbb{L},\mathbb{A}}\, \mathbf{y}_a \tag{4.7}$$

are the vectors whose components are the DFT coefficients of $\mathbf{x}_s$ and $\mathbf{y}_s$, evaluated at the position indexes in $\mathbb{L}$. However, note that since $\mathbf{x}_s$ and $\mathbf{y}_s$ are supposed to satisfy the second condition in the theorem, these Fourier coefficients $\hat{\mathbf{x}}_a$ and $\hat{\mathbf{y}}_a$ must be equal to the values of $\hat{x}[k]$ for $k \in \mathbb{L}$, so that

$$\hat{\mathbf{x}}_a = \hat{\mathbf{y}}_a,$$

which from (4.6) and (4.7) leads to

$$\mathbf{F}_{\mathbb{L},\mathbb{A}}\, \mathbf{x}_a = \mathbf{F}_{\mathbb{L},\mathbb{A}}\, \mathbf{y}_a,$$

resulting in

$$\mathbf{F}_{\mathbb{L},\mathbb{A}}(\mathbf{x}_a - \mathbf{y}_a) = \mathbf{0}. \tag{4.8}$$

Finally, observe that $|\mathbb{A}| = |\mathbb{L}| = \ell$; so, $\mathbf{F}_{\mathbb{L},\mathbb{A}}$ satisfies the conditions of Theorem 1 (it is a square submatrix of the DFT matrix) and it is then invertible. Therefore there exists $(\mathbf{F}_{\mathbb{L},\mathbb{A}})^{-1}$ such that $(\mathbf{F}_{\mathbb{L},\mathbb{A}})^{-1}\, \mathbf{F}_{\mathbb{L},\mathbb{A}} = \mathbf{I}$ (the $\ell \times \ell$ identity matrix). Hence, (4.8) leads to

$$\mathbf{x}_a - \mathbf{y}_a = (\mathbf{F}_{\mathbb{L},\mathbb{A}})^{-1} \cdot \mathbf{0},$$

which reduces to

$$\mathbf{x}_a - \mathbf{y}_a = \mathbf{0}$$

and

$$\mathbf{x}_a = \mathbf{y}_a.$$

Now, from the definition of $\mathbf{x}_a$ and $\mathbf{y}_a$, this indicates that all the nonzero elements of signals $\mathbf{x}_s$ and $\mathbf{y}_s$ are pairwise equal. Hence, the signals $\mathbf{x}_s$ and $\mathbf{y}_s$ are equal. This completes the proof. □

4.1.2 Alternative Formulation

It is important to emphasize that, in stating Theorem 2, we represented by $\mathbb{L}$ the set of ℓ positions in the nonsparse domain where the isolated measurements are taken (in this case, the frequencies where the DFT is sampled), while Φ represents the set of φ known positions

of nonzero samples in the sparse domain (time). Theorem 2 then states that, given the DFT coefficients corresponding to the frequency indexes in $\mathbb{L}$, there is at most one prime-length signal, with no more than $(\ell + \varphi)/2$ nonzero time samples such that its DFT matches the given ℓ coefficients, and whose positions for the sparse nonzero samples include the given set Φ of known positions. This statement also means that:

1. ℓ isolated coefficients in the frequency domain are enough to unambiguously determine prime-length signals which are known to have no more than $(\ell+\varphi)/2$ nonzero samples in the time domain, provided that the positions of φ of these nonzero samples are known, while the values of these samples are not required;

2. to reconstruct signals having prime-valued lengths and which are known to be η-sparse in time domain, $2\eta - \varphi$ frequency coefficients are enough, provided that the position of φ nonzero time samples are known.

Observe that, according to Theorem 2, if we take ℓ coefficients from a signal whose representation in time is $(\ell+\varphi)/2$-sparse, and where φ positions of the non-null coefficients are known, then all its DFT coefficients can theoretically be unambiguously reconstructed. In fact, the theorem guarantees the unicity of a vector with at most $(\ell + \varphi)/2$ non-null coefficients (φ of them in the specified known positions), such that its direct DFT matches the ℓ taken coefficients in frequency. Then, this vector can be uniquely determined from only ℓ coefficients, and the whole signal can be reconstructed by taking the inverse DFT.

According to this analysis, the following Theorem 3 is a direct consequence of Theorem 2. In fact, Theorem 3 establishes that the number of coefficients which must be taken from the DFT of a signal is $\ell = 2\eta - \varphi$ (if this signal has an η-sparse time representation), in order to theoretically guarantee reconstruction and if φ positions associated to nonzero time samples are known. This is opposed to $\ell = 2\eta$ coefficients in the initial compressive sensing approach, which does not assume any prior information on the support of the sparse domain.

Theorem 3. *Let* $\mathbf{x}_s : \{0, 1, 2, \ldots, N_p - 1\} \to \mathbb{C}$ *be an* η*-sparse discrete-time signal, meaning that it has at most* η *non-null samples:* $|\text{supp}(\mathbf{x}_s)| \leqslant \eta$*, where* η *is an integer satisfying* $0 < \eta \leqslant N_p/2$*, with* N_p*, the length of* $\mathbf{x}_s$*, a prime number.*

If Φ *is a subset of the support of* $\mathbf{x}_s$ *with* φ *elements, meaning* $\Phi \subseteq \text{supp}(\mathbf{x}_s)$*, with* $|\Phi| = \varphi$*, and if this subset* Φ is known*, then the signal* $\mathbf{x}_s$ *can be unambiguously reconstructed from* $2\eta - \varphi$ *coefficients of its length-*N_p *discrete-time Fourier transform.*

It is also possible to find signals which are sparse not in time, but in the frequency domain. In this case, the same ideas can still be applied in a reverse way, given the duality property of the discrete-time Fourier transform. In this new configuration, $\mathbb{L}$ represents the set of ℓ time samples, with Φ representing the set of φ known positions of nonzero DFT coefficients. Note then that $2\eta - \varphi$ is also the minimum number of time samples required to reconstruct a signal that has a η-sparse DFT.

4.1.3 Reconstruction Using Prior Information

The presented theorems describe the reduction in the required number of samples in the nonsparse domain for unambiguous reconstruction, when φ positions of non-null coefficients in the sparse domain are known. The next step is how to incorporate the information on these known positions when conducting the reconstruction procedure. Note that the important additional point during the reconstruction is to guarantee that the obtained nonzero elements of the signal in the sparse domain will include the φ mentioned positions. In this way, the reduction in the number of required original measurements will follow the unicity of the sparse signal that explains those measurements and at the same time include, among its non-zero coefficients, those in the φ specified positions.

The corresponding adaptation of the reconstruction problem is straightforward. In this problem, the objective is to determine the sparsest vector $\mathbf{x}$ satisfying the condition $\mathbf{M}\mathbf{T}^{-1}\hat{\mathbf{x}} = \mathbf{y}$, with $\mathbf{M}$ the measurement matrix, $\mathbf{T}$ the transformation matrix which leads to the sparse representation of $\mathbf{x}$, and $\mathbf{y}$ the available linear measurements. The new objective is basically the same; the only difference is that we must guarantee that any solution

includes, among the nonzero components, φ in the specified Φ positions. The key point for describing the new optimization problem is the fact that finding the sparsest vector $\hat{\mathbf{x}}$ that includes, among the nonzero components, φ in pre-specified positions is equivalent to minimizing the number of non-null components in the other $N_p - \varphi$ positions. In fact, since those φ positions are guaranteed to include nonzero elements, the concern when conducting the ℓ_0-minimization of $\hat{\mathbf{x}}$ should be to minimize the number of non-null components in the other positions only and, at the same time, satisfying the equality constraint $\mathbf{MT}^{-1}\hat{\mathbf{x}} = \mathbf{y}$.

Hence, by defining $\hat{\mathbf{x}}_{k\notin\Phi}$ as the subvector of all components $\hat{x}_k$ in $\hat{\mathbf{x}}$ such that $k \notin \Phi$, and defining $\|\hat{\mathbf{x}}_{k\notin\Phi}\|_0$ as the number of nonzero components in that subvector, we can write the optimization problem for reconstruction with prior information as

$$\begin{aligned} &\min_{\hat{\mathbf{x}}} \ \|\hat{\mathbf{x}}_{k\notin\Phi}\|_0 \\ &\text{subject to } \ (\text{IDFT}\{\hat{\mathbf{x}}\})_{n\in\mathbb{L}} = \mathbf{x}_{n\in\mathbb{L}}, \end{aligned} \tag{4.9}$$

It is worth mentioning that although the components of $\hat{\mathbf{x}}$ in the locations belonging to the support Φ are not included in the objective function, they are constrained by the equality $(\text{IDFT}\{\hat{\mathbf{x}}\})_{n\in\mathbb{L}} = \mathbf{x}_{n\in\mathbb{L}}$, and they are also computed during the solution to (4.9).

Note, however, that the ideal direct solution to (4.9) based on combinatorial optimization is not viable even for moderately-sized signals [92]. As we will discuss in Section 4.2, we can use the ℓ_1-minimization as an efficient way to compute a solution with minimum number of nonzero components in $\hat{\mathbf{x}}_{k\notin\Phi}$, so that (4.9) can be written as

$$\begin{aligned} &\min_{\hat{\mathbf{x}}} \ \|\hat{\mathbf{x}}_{k\notin\Phi}\|_1 \\ &\text{subject to } \ (\text{IDFT}\{\hat{\mathbf{x}}\})_{n\in\mathbb{L}} = \mathbf{x}_{n\in\mathbb{L}}. \end{aligned} \tag{4.10}$$

In Section 4.2, we present the conditions that are sufficient and necessary for the solution to (4.10) to match the desired signal $\mathbf{x}$, considering the available prior information.

If the signal is sparse in the time domain (as described in Theorem 3), instead of the frequency domain, then the optimization problem leading to the reconstruction is analogous to (4.9) and (4.10). The only difference is that one must minimize the number of non-null components in the time domain, with the condition that φ of the nonzeros will be in the

positions specified by the set Φ and that the DFT of the reconstructed signals must match the ℓ coefficients taken as measurements in the frequency domain. The optimization problem for the sparsity in the time domain then becomes

$$\begin{aligned} \min_{\mathbf{x}} \ & \|\mathbf{x}_{n\notin\Phi}\|_1 \\ \text{subject to } & (\text{DFT}\{\mathbf{x}\})_{k\in\mathbb{L}} = \hat{\mathbf{x}}_{k\in\mathbb{L}}. \end{aligned} \tag{4.11}$$

The basic idea of this section, namely that in compressive sensing the reconstruction can be conducted from fewer measurements, if information on the positions of non-null components in the sparse domain is available, is illustrated in Fig. 4.1. Note that two domains are shown, one in which the considered signal is not sparse (full vector), the other where it has a sparse representation. The two domains are related by the DFT, but the same basic procedure can be used no matter if the sparse representation occurs in time or in frequency. Figure 4.1(a) represents the linear measurements taken in the nonsparse domain in the positions specified by the set $\mathbb{L}$. In the standard compressive sensing approach, if the signal is η-sparse in the corresponding DFT or IDFT domain, 2η measurements would be necessary to avoid ambiguity during reconstruction; here, we showed that if φ positions are known for non-null components in the sparse domain, positions represented by Φ in Fig. 4.1(b), then the number of measurements can be reduced to $\ell = 2\eta - \varphi$.

Now, consider still that the signal $\mathbf{x}$ does not have a sparse representation in the DFT or in the IDFT domain. In this case, if an arbitrary orthogonal transform matrix $\mathbf{T}$ provides a sparse representation, meaning that $\hat{\mathbf{x}} = \mathbf{T}\mathbf{x}$ is sparse, $\mathbf{x}$ can be represented by linear projections on ℓ independent vectors: $\mathbf{y}_{\ell\times 1} = \mathbf{M}_{\ell\times N_p}\mathbf{x}_{N_p\times 1}$. The condition for this representation to hold is that all rows of $\mathbf{M}$ must not have a sparse representation in the domain defined by $\mathbf{T}$ [65]. In other words, the vector obtained by multiplying $\mathbf{T}$ by any row of $\mathbf{M}$ must not be sparse.

In the initial compressive sensing approach, the reconstruction is then conducted by solving

$$\begin{aligned} \min_{\hat{\mathbf{x}}} \ & \|\hat{\mathbf{x}}\|_1 \\ \text{subject to } & \mathbf{M}\mathbf{T}^{-1}\hat{\mathbf{x}} = \mathbf{y}; \end{aligned}$$

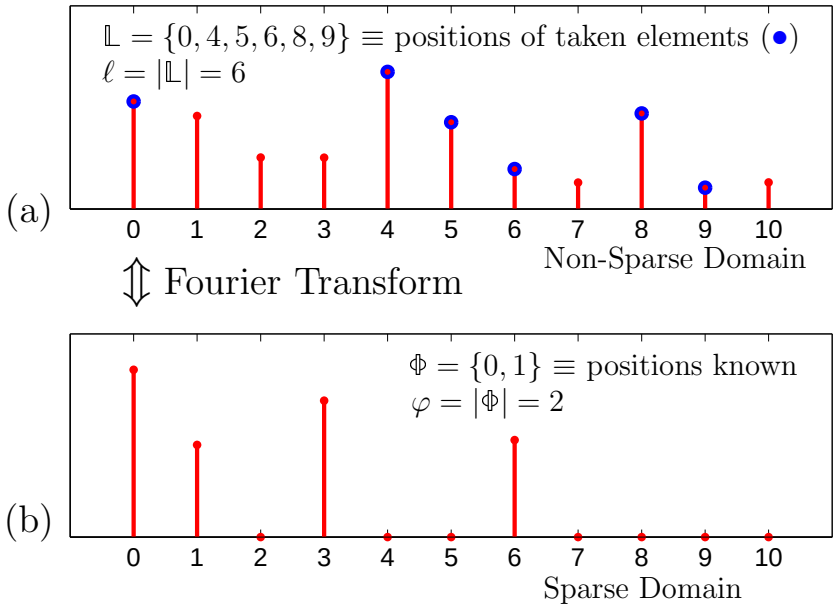


Figure 4.1. Reconstruction using partial information on the sparse representation's support. The top plot shows the magnitude of the elements of a nonsparse vector with prime length $N_p = 11$, and the bottom plot shows the magnitude of the elements of a sparse vector with $\eta = 4$ non-null elements. The two vectors are mapped by the Fourier transform. If $\varphi = 2$ positions of the non-null elements of the sparse vector are known, then we only need to take the values of $\ell = 2\eta - \varphi$ elements of the nonsparse vector to be able to reconstruct the two vectors unambiguously.

note that if it is known that φ of the nonzero components of $\hat{\mathbf{x}}$ are in the positions specified by the set Φ, this information can be included in the optimization problem by solving

$$\begin{aligned} &\min_{\hat{\mathbf{x}}} \ \|\hat{\mathbf{x}}_{k \notin \Phi}\|_1 \\ &\text{subject to } \ \mathbf{M}\mathbf{T}^{-1}\hat{\mathbf{x}} = \mathbf{y}. \end{aligned} \tag{4.12}$$

4.1.4 Relation to the Fractional Discrete-Time Fourier Transform

In Theorem 3, we have shown that if a signal $\mathbf{x}$ is η-sparse, and if φ elements of its support are known, then $\ell = 2\eta - \varphi$ components taken from its DFT are theoretically enough to ensure unambiguous reconstruction. Two special cases can be mentioned regarding this number of required components, and they refer to the two extreme possibilities for the number of known

positions φ. First, if no position in the support of the sparse domain is known ($\varphi = 0$), then the number of DFT coefficients required for unambiguous reconstruction is exactly 2η; this is in accordance with the initial compressive sensing framework, which makes no reference to positions known in the sparse domain.

The second special case refers to *all* positions in the support of the sparse domain being known; in this situation, $\varphi = \eta$ and the number of coefficients required for reconstruction reduces to $\ell = \eta$. This result was expected already from Theorem 1. In fact, if all positions of the support of the sparse domain are known, reconstructing $\mathbf{x}$ is equivalent to determining the time samples $\mathbf{x}_{n\in\Phi}$ in positions Φ (the other samples are all zero); but note that the DFT coefficients in positions specified by the set $\mathbb{L}$ are given by $\hat{\mathbf{x}}_{k\in\mathbb{L}} = \mathbf{F}_{\mathbb{L},\Phi}\mathbf{x}_{n\in\Phi}$. Then, if the number of elements in $\mathbb{L}$ is equal to that in Φ ($\ell = \eta$), the matrix $\mathbf{F}_{\mathbb{L},\Phi}$ is invertible according to Theorem 1, and all the nonzero samples of $\mathbf{x}$ can be determined from $\mathbf{x}_{n\in\Phi} = (\mathbf{F}_{\mathbb{L},\Phi})^{-1}\hat{\mathbf{x}}_{k\in\mathbb{L}}$.

It is also important to observe that this second mentioned special case, namely all the positions in the sparse domain being known (which leads to the number of required DFT samples being equal to the number of nonzero time samples), is an example of what happens in the context of the fractional Fourier transform (FrFT) [4]. In this context, a signal of length N_1 is first zero-padded to $N_2 > N_1$ samples, and the DFT is computed for the extended version $\mathbf{x}$; the FrFT is then defined by the first N_1 obtained coefficients. Note that the zero-padded signal $\mathbf{x}$ has only $\eta = N_1$ non-null components, and can be reconstructed from N_1 DFT coefficients even though its dimension N_2 is higher than N_1.

The idea that the first coefficients of the DFT convey enough information for the reconstruction of the original signal is used in the FrFT framework [4], but it is also a special case of Theorem 3: if η positions of the support of the sparse domain are known (in this case the N_1 first positions), then only η samples of the nonsparse domain are necessary to compute all samples. It is also significant that, based on Theorem 2, any N_1 coefficients can be used. In fact, it is interesting to observe that our proof to this theorem was based on the invertibility of all the square matrices of the DFT matrix, provided that the number of columns and rows in this full matrix is a prime number. In the case of the FrFT, if we take the first

samples of the DFT for the reconstruction, the computation of the signal in the time domain involves the square submatrix defined by the first rows and first columns of the DFT. The compressive sensing framework, on the other hand, is based on any square submatrix of the DFT, and although it requires the signal's dimension to be prime, it otherwise generalizes the FrFT analysis by allowing the zero-padded signal to be reconstructed from any N_1 DFT coefficients.

An example of this association between compressive sensing and the fractional Fourier transform is provided in Fig. 4.2. Figure 4.2(a) shows a length-20 signal $\mathbf{x}$, which is zero-padded in order to yield a length-29 signal. The magnitude of the DFT $\hat{\mathbf{x}}$ of the resulting signal is shown in Fig. 4.2(b), and the 20 first coefficients correspond to the fractional Fourier transform as defined in [4]. According to Theorem 2, if one knows the position of all the $\eta = 20$ non-null coefficients of $\mathbf{x}$, the whole signal can be reconstructed from any 20 samples of $\hat{\mathbf{x}}$. We reconstructed the signal, using compressive sensing, from the first 20 samples in the hatched region in Fig. 4.2(b), as is done in the inverse FrFT. The errors were below 10^{-10}, and then were negligible with respect to the input suggesting a correct reconstruction, using the compressive sensing approach. In this particular case, we could use the procedure used for the standard FrFT.

A second alternative procedure based on compressive sensing and the test signal in Fig. 4.2 can be provided. In this case, we assumed that only 17 positions in the support of $\mathbf{x}$ were known, so that $\ell = 2 \cdot 20 - 17 = 23$ DFT coefficients would be necessary in order to guarantee correct reconstruction according to Theorem 2. In this case, we chose to take the last 23 coefficients — see Fig. 4.2(b); the errors between the reconstructed signal and the original one were below 1.5×10^{-12}, in magnitude.

4.2 Prior Information and ℓ_1-Minimization

In Section 4.1, we have shown that prior information about the locations of φ nonzero components of a η-sparse signal reduces by φ the theoretical minimum number of DFT samples required for reconstruction by ℓ_0-minimization. The minimum limit refers, however,

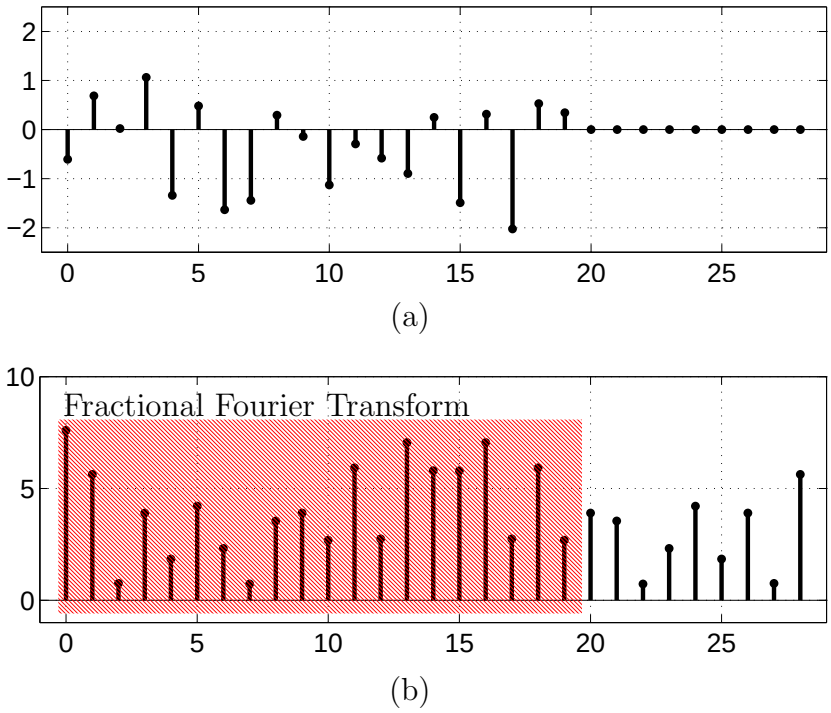


Figure 4.2. Example of the link between the fractional Fourier transform and compressive sensing with prior information. (a) A length-20 signal zero-padded to 29 samples in order to determine the fractional Fourier transform. (b) The 29-point DFT (in magnitude) of the zero-padded signal; the fractional Fourier transform here used corresponds to the coefficients inside the hatched region. The signal $x[n]$ shown in (a) can be reconstructed from the 20 coefficients in the shaded area, by exploiting the fact that x is sparse (9 coefficients are known to be zero) and then using compressive sensing.

to the number of samples that guarantees that the acquired measurements will never be the same for two distinct signals with at most η components and which match the conditions of the prior information. No assumption is made, therefore, to the actual algorithm used to reconstruct the signal from its measurements, and even though the ℓ_0-minimization approach can theoretically attain the minimum limit, this generally is not viable in practical situations, as previously discussed.

An important discovery in compressive sensing is that, given a sufficient number of ℓ linear measurements of a η-sparse signal $\mathbf{x}_s$, the ℓ_1-minimization of a vector with the condition that it gives the same measurements as $\mathbf{x}_s$ will provide a solution that is unique and matches $\mathbf{x}_s$

with high probability [10]; the exact value of this probability depends basically on ℓ, η and on the signal's length. Based on this discovery, the ℓ_1-minimization approach is one of the methods of reconstruction most commonly used in compressive sensing [5]. It is therefore important to analyze the reduction on the required number of measurements due to the use of partial prior information when the ℓ_1-minimization approach is used.

4.2.1 Conditions for Reconstruction

We begin by stating a set of conditions that are necessary and sufficient for the solution of the ℓ_1-minimization problem, with prior information, to be unique and to match the signal from which the linear measurements are taken. In the next section, we discuss the probability of attaining these conditions, as opposed to the corresponding probability when no prior information is available, and argue that it is higher in the first case.

Regarding the reconstruction without prior information, Candès et al. have shown the conditions under which a sparse signal can be computed from limited DFT samples by solving an ℓ_1-minimization problem [10]. In this section we show that for the same signal and set of DFT samples as above, but with the additional prior information that the φ elements in set Φ belong to the support of $\mathbf{x}_s$, the necessary and sufficient conditions for the reconstruction of $\mathbf{x}_s$ based on the ℓ_1-minimization approach are modified. In fact, in this case the solution to (4.11), with $\hat{\mathbf{x}} = \text{DFT}(\mathbf{x}_s)$, is unique and matches $\mathbf{x}_s$ if and only if there exists a function $\mathbf{p}$ with the same domain as $\mathbf{x}_s$ that satisfies the conditions

C1. $p[n] = \text{sgn}(x_s[n])\ \forall\, n \in \text{supp}(\mathbf{x}_s) - \Phi$;

C2. $p[n] = 0\ \forall\, n \in \Phi$;

C3. $|p[n]| < 1\ \forall\, n \notin \text{supp}(\mathbf{x}_s)$.

C4. $\hat{p}[k] = 0\ \forall\, k \notin \mathbb{L}$.

The existence of partial prior information about the support of $\mathbf{x}_s$ modifies the described conditions in the sense that, in the known locations of the support, $\mathbf{p}$ should vanish, whereas

in the case without prior information $\mathbf{p}$ has to match the sign of $\mathbf{x}_s$ inside all the support. In Section 4.2.2, we argue that, for the same number of measurements, the probability of existence of a function satisfying the reconstruction conditions is higher in the case with prior information than when such information is not available. Before that, however, we present Theorem 4, which contains the formal statement of the reconstruction conditions in the case of prior information.

Theorem 4. *Let* $\mathbf{x}_s : \mathbb{G} \to \mathbb{C}$ *be an* η*-sparse discrete-time signal defined on*

$$\mathbb{G} = \{0, 1, \ldots, N-1\},$$

with N *its length and* $\text{supp}(\mathbf{x}_s) = \mathbb{T}$ *its support, so that* $x_s[n] = 0 \; \forall \, n \notin \mathbb{T}$*. Suppose that* Φ *is a subset of* $\mathbb{T}$*, representing* $|\Phi| = \varphi$ *known locations in that support. Also, let* $\mathbb{L} \subset \mathbb{G}$ *be a set of* $\ell = |\mathbb{L}|$ *frequency locations where the discrete Fourier transform of* $\mathbf{x}_s$*, here represented by* $\hat{\mathbf{x}} = \text{DFT}(\mathbf{x}_s)$*, is sampled, so that* $\mathbf{b} = \text{DFT}(\mathbf{x}_s)|_{k \in \mathbb{L}}$ *is the known measurement vector. Finally, suppose that the function* $\mathcal{F}_{\mathbb{T} \to \mathbb{L}}$ *that associates the signals supported in* $\mathbb{T}$ *with their* ℓ *Fourier coefficients located in* $\mathbb{L}$ *is an injection, meaning that for a given* ℓ*-tuple of coefficients there is at most one signal supported in* $\mathbb{T}$ *whose Fourier coefficients at the locations in* $\mathbb{L}$ *match that* ℓ*-tuple. Under these conditions, the optimization problem*

$$\begin{aligned} &\min \; \|\mathbf{x}_{n \notin \Phi}\|_1 \\ &\text{subject to } \; \text{DFT}(\mathbf{x})|_{k \in \mathbb{L}} = \mathbf{b} \end{aligned} \tag{4.13}$$

has a unique solution $\mathbf{x}^*$ *with* $\mathbf{x}^* = \mathbf{x}_s$ *if and only if there exists a function* $\mathbf{p} : \mathbb{G} \to \mathbb{C}$ *that satisfies the following conditions.*

C1. $p[n] = \text{sgn}(x_s[n]) \; \forall \, n \in \mathbb{T} - \Phi$*.*

C2. $p[n] = 0 \; \forall \, n \in \Phi$*.*

C3. $|p[n]| < 1 \; \forall \, n \notin \mathbb{T}$*.*

C4. $\hat{p}[k] = 0 \; \forall \, k \notin \mathbb{L}$*, where* $\hat{\mathbf{p}}$ *denotes the DFT of* $\mathbf{p}$*.*

Proof. There are two main assertions to prove. First, we must show that if there is a function P that satisfies C1, C2, C3, and C4, then the solution $\mathbf{x}^*$ to (4.13) is unique and matches $\mathbf{x}_s$. The second assertion is that if $\mathbf{x}_s$ is the unique solution to (4.13), then there exists a function P that satisfies those conditions.

Consider that there exists a function P for which C1, C2, C3, and C4 hold, and let $\mathbf{x} : \mathbb{G} \to \mathbb{C}$ be an arbitrary signal such that $\text{DFT}(\mathbf{x})|_{k\in\mathbb{L}} = \mathbf{b}$. By definition,

$$\|\mathbf{x}_{n\notin\Phi}\|_1 = \sum_{n\notin\Phi} |x[n]|.$$

Defining $\mathbf{d}$ as the difference between $\mathbf{x}_s$ and $\mathbf{x}$, so that $d[n] = x_s[n] - x[n]\ \forall\, n \in \mathbb{G}$, we then have

$$\|\mathbf{x}_{n\notin\Phi}\|_1 = \sum_{n\notin\Phi} |x_s[n] - d[n]|. \tag{4.14}$$

Now, observe that, from conditions C1, C2, C3, C4, and using the definition of the sign function (sgn), it follows that $|\overline{P[n]}| \leqslant 1$, with the equality holding if and only if $n \in \mathbb{T} - \Phi$. Combined with (4.14), this yields

$$\|\mathbf{x}_{n\notin\Phi}\|_1 \geqslant \sum_{n\notin\Phi} |x_s[n] - d[n]||\overline{P[n]}|, \tag{4.15}$$

where the equality holds if and only if $x_s[n] - d[n] = 0\ \forall\, n \notin \Phi$ such that $|\overline{P[n]}| < 1$, which is equivalent to $d[n] = 0\ \forall\, n \notin \mathbb{T}$.

By further developing (4.15), we then obtain

$$\|\mathbf{x}_{n\notin\Phi}\|_1 \geqslant \sum_{n\notin\Phi} |x_s[n]\overline{P[n]} - d[n]\overline{P[n]}|, \tag{4.16}$$

with the equality holding if and only if $d[n] = 0\ \forall\, n \notin \mathbb{T}$.

Now, since, from C1, $P[n] = \text{sgn}(x_s[n])\ \forall\, n \in \mathbb{T} - \Phi$, we have

$$x_s[n]\overline{P[n]} = x_s[n]\overline{\text{sgn}(x_s[n])} = x_s[n]\frac{\overline{x_s[n]}}{|x_s[n]|} = \frac{|x_s[n]|^2}{|x_s[n]|} = |x_s[n]|,\ \ \forall n \in \mathbb{T} - \Phi.$$

The equality also holds for $n \notin \mathbb{T}$, since $x_s[n] = 0\ \forall\, n \notin \mathbb{T}$, so (4.16) becomes

$$\|\mathbf{x}_{n\notin\Phi}\|_1 = \sum_{n\notin\Phi} ||x_s[n]| - d[n]\overline{P[n]}|textrm, if d[n] = 0\ \forall\, n \notin \mathbb{T}.$$

Using the relation $|c| \geqslant \mathrm{Re}(c)\ \forall\, c \in \mathbb{C}$, we then obtain

$$\|\mathbf{x}_{n\notin\Phi}\|_1 \geqslant \sum_{n\notin\Phi} \mathrm{Re}\Big[|x_s[n]| - d[n]\overline{P[n]}\Big]. \tag{4.17}$$

Note that if the equality in (4.17) holds, then $d[n] = 0\ \forall\, n \notin \mathbb{T}$; indeed, we showed that in (4.16) the equality is true if and only if $d[n] = 0\ \forall\, n \notin \mathbb{T}$, so if this condition is not satisfied, we get the inequality in (4.16) and, together with the condition $|c| \geqslant \mathrm{Re}(c)$, also an inequality in (4.17).

By developing (4.17), we then obtain

$$\|\mathbf{x}_{n\notin\Phi}\|_1 \geqslant \sum_{n\notin\Phi} \left[\mathrm{Re}(|x_s[n]|) - \mathrm{Re}\left(d[n]\overline{P[n]}\right)\right];$$

since $\mathrm{Re}(|x_s[n]|) = |x_s[n]|$, this gives

$$\|\mathbf{x}_{n\notin\Phi}\|_1 \geqslant \|\mathbf{x}_{s,n\notin\Phi}\|_1 - \mathrm{Re}\left(\sum_{n\notin\Phi} d[n]\overline{P[n]}\right). \tag{4.18}$$

Now, observe that

$$\sum_{n\notin\Phi} d[n]\overline{P[n]} = \sum_{n=0}^{N-1} d[n]\overline{P[n]} - \sum_{n\in\Phi} d[n]\overline{P[n]}, \tag{4.19}$$

and, from Parseval's theorem,

$$\sum_{n=0}^{N-1} d[n]\overline{P[n]} = \sum_{k=0}^{N-1} \hat{d}[k]\overline{\hat{P}[k]}; \tag{4.20}$$

where $\hat{\mathbf{d}}$ is the DFT of $\mathbf{d}$. Also, since $\mathbf{d} = \mathbf{x_s} - \mathbf{x}$, then $\hat{d}[k] = \hat{x_s}[k] - \hat{x}[k]\ \forall\, k \in \mathbb{G}$, and since $\hat{x_s}[k] = \hat{x}[k] = b[k]\ \forall\, k \in \mathbb{L}$,

$$\hat{d}[k] = 0\ \forall\, k \in \mathbb{L}; \tag{4.21}$$

furthermore, from condition C4,

$$\overline{\hat{P}[k]} = 0\ \forall\, k \notin \mathbb{L}. \tag{4.22}$$

Using (4.21) and (4.22) in (4.20), we obtain

$$\sum_{n=0}^{N-1} d[n]\overline{P[n]} = 0,$$

so (4.19) becomes

$$\sum_{n\notin\Phi} d[n]\overline{P[n]} = -\sum_{n\in\Phi} d[n]\overline{P[n]} = 0, \tag{4.23}$$

since $\overline{P[n]} = 0 \;\forall\, n \in \Phi$, from condition C2.

Substituting (4.23) into (4.18), we then obtain

$$\|\mathbf{x}_{n\notin\Phi}\|_1 \geqslant \|\mathbf{x}_{s,n\notin\Phi}\|_1, \tag{4.24}$$

where, if the equality holds, $d[n] = 0 \;\forall\, n \notin \mathbb{T}$.

Now, observe that since $\|\mathbf{x}_{n\notin\Phi}\|$ is a convex function of $\mathbf{x}$, a local solution to (4.13) is also the global solution. Furthermore, since $\mathbf{x}_s$ satisfies the equality constraint $\mathrm{DFT}(\mathbf{x}_s)|_{k\in\mathbb{L}} = \mathbf{b}$, a local solution $\mathbf{x}^*$ to (4.13) satisfies $\|\mathbf{x}^*_{n\notin\Phi}\|_1 \leqslant \|\mathbf{x}_{s,n\notin\Phi}\|_1$, since the case $\|\mathbf{x}^*_{n\notin\Phi}\|_1 > \|\mathbf{x}_{s,n\notin\Phi}\|_1$ would imply that $\mathbf{x}^*$ is not a global solution. Hence, from (4.24) we obtain

$$\|\mathbf{x}^*_{n\notin\Phi}\|_1 = \|\mathbf{x}_{s,n\notin\Phi}\|_1,$$

and since the equality in (4.24) implies that $d[n] = 0 \;\forall\, n \notin \mathbb{T}$, we conclude that

$$x^*[n] = x_s[n] = 0 \;\; \forall\, n \notin \mathbb{T}.$$

Hence, $\mathbf{x}^*$ and $\mathbf{x}_s$ have the same support $\mathbb{T} = \mathrm{supp}(\mathbf{x}^*) = \mathrm{supp}(\mathbf{x}_s)$, and they have the same Fourier coefficients inside $\mathbb{L}$ since $\mathrm{DFT}(\mathbf{x}_s)|_{k\in\mathbb{L}} = \mathbf{b}$ and $\mathrm{DFT}(\mathbf{x}^*)|_{k\in\mathbb{L}} = \mathbf{b}$; from the condition that $\mathcal{F}_{\mathbb{T}\to\mathbb{L}}$ is an injection, it then follows that

$$\mathbf{x}^* = \mathbf{x}_s.$$

We have thus shown that if there exists a function P satisfying conditions C1 to C4, then the solution $\mathbf{x}^*$ to (4.13) is unique and equals $\mathbf{x}_s$. Now, we must show that if the solution to (4.13) is unique and matches $\mathbf{x}_s$, then there exists a function satisfying the same conditions. In order to do so, we first consider the affine set $\mathbb{A}$ of all signals whose Fourier coefficients lying in $\mathbb{L}$ match those of $\mathbf{x}_s$, meaning

$$\mathbb{A} = \{\mathbf{y} \,|\, \mathrm{DFT}(\mathbf{y})|_{k\in\mathbb{L}} = \mathbf{b}\},$$

and the set defined by

$$\mathbb{B} = \{\mathbf{y} \,|\, \|\mathbf{y}_{n\notin\Phi}\|_1 = \|\mathbf{x}_{s,n\notin\Phi}\|_1\}.$$

Observe that $\mathbb{A}$ and $\mathbb{B}$ intersect at a single point, namely $\mathbf{x}_s$, since $\mathbf{x}_s$ is the unique solution to (4.13). According to the Hahn-Banach theorem, there exists then a function $P : \mathbb{G} \to \mathbb{C}$ such that

$$\mathbb{A} \subseteq \left\{ \mathbf{y} : \mathbb{G} \to \mathbb{C} | \sum_{n\in\mathbb{G}} \text{Re}\left(y[n]\overline{P[n]}\right) = \|\mathbf{x}_{s,n\notin\Phi}\|_1 \right\} \tag{4.25}$$

and

$$\mathbb{B} \subseteq \left\{ \mathbf{y} : \mathbb{G} \to \mathbb{C} | \sum_{n\in\mathbb{G}} \text{Re}\left(y[n]\overline{P[n]}\right) \leqslant \|\mathbf{x}_{s,n\notin\Phi}\|_1 \right\}. \tag{4.26}$$

We will then show that the function P satisfies the conditions C1 to C4. We start by showing that $P[n] = 0 \; \forall \, n \in \Phi$ (condition C2). In fact, suppose that there is a point $n_0 \in \Phi$ such that $P[n_0] \neq 0$. Consider then the signal $\mathbf{y} : \mathbb{G} \to \mathbb{C}$ defined by

$$y[m] = \begin{cases} \dfrac{1 + \|\mathbf{x}_{s,n\notin\Phi}\|_1 - \sum_{n\neq n_0} x_s[n]\overline{P[n]}}{\overline{P[m]}}, & \text{if } m = n_0, \\ \\ x_s[m], & \text{otherwise.} \end{cases}$$

Note that, since $n_0 \in \Phi$,

$$\sum_{n\notin\Phi} |y[n]| = \sum_{n\notin\Phi} |x_s[n]| = \|\mathbf{x}_{s,n\notin\Phi}\|_1,$$

so $\mathbf{y} \in \mathbb{B}$. Also note that

$$\sum_{n\in\mathbb{G}} \text{Re}\left(y[n]\overline{P[n]}\right) = 1 + \|\mathbf{x}_{s,n\notin\Phi}\|_1 > \|\mathbf{x}_{s,n\notin\Phi}\|_1,$$

which contradicts (4.26) since $\mathbf{y} \in \mathbb{B}$. Hence, the hypothesis that there is some $n_0 \in \Phi$ such that $P[n_0] \neq 0$ leads to a contradiction, and we conclude that

$$P[n] = 0 \; \forall \, n \in \Phi. \tag{4.27}$$

Regarding the remaining values of $P[n]$, we next show that $|P[n]| \leqslant 1 \; \forall \, n \in \mathbb{G} - \Phi$. Suppose, by contradiction, that $|P[n_0]| > 1$ for some $n_0 \in \mathbb{G} - \Phi$. In this case, consider the

signal $\mathbf{y} : \mathbb{G} \to \mathbb{C}$ defined by

$$y[m] = \begin{cases} \text{sgn}(P[n])\|\mathbf{x}_{s,n\notin\Phi}\|_1, & \text{if } m = n_0, \\ 0, & \text{otherwise;} \end{cases}$$

note that $\mathbf{y} \in \mathbb{B}$, since $\sum_{m\notin\Phi}|y[m]| = |y[n_0]| = \|\mathbf{x}_{s,n\notin\Phi}\|_1$. Also, note that, for the same $n_0 \in (\mathbb{G} - \mathbb{C})$,

$$\sum_{n\notin\Phi} \text{Re}\left(y[n]\overline{P[n]}\right) = \|\mathbf{x}_s\|_1|P[n_0]|,$$

so, if $|P[n_0]| > 1$,

$$\sum_{n\notin\Phi} \text{Re}\left(y[n]\overline{P[n]}\right) > \|\mathbf{x}_s\|_1,$$

which contradicts (4.26) since $\mathbf{y} \in \mathbb{B}$. Therefore,

$$|P[n]| \leqslant 1 \ \forall\, n \in \mathbb{G} - \Phi, \tag{4.28}$$

as the hypothesis $|P[n]| > 1$ for some $n \notin \Phi$ leads to a contradiction.

Furthermore, since $\mathbf{x}_s \in \mathbb{A}$ and using (4.25),

$$\sum_{n\in\mathbb{G}} \text{Re}\left(x_s[n]\overline{P[n]}\right) = \|\mathbf{x}_{s,n\notin\Phi}\|_1 = \sum_{n\notin\Phi}|x_s[n]|;$$

by using (4.27) with the identity $|x_s[n]| = x_s[n]\overline{\text{sgn}(x_s[n])}$ and the fact that $x_s[n] = 0 \ \forall\, n \notin \mathbb{T}$, we then obtain

$$\sum_{n\in\mathbb{T}-\Phi} \text{Re}\left(x_s[n]\overline{P[n]}\right) = \sum_{n\in\mathbb{T}-\Phi} x_s[n]\overline{\text{sgn}(x_s[n])},$$

which, since $|P[n]| \leqslant 1 \ \forall\, n \in \mathbb{T} - \Phi$, leads to

$$P[n] = \text{sgn}(x_s[n]) \ \forall\, n \in \mathbb{T} - \Phi, \tag{4.29}$$

thus proving condition C1.

Based on (4.29), we can also prove C3. In fact, note that since $\mathbb{A}$ and $\mathbb{B}$ intersect only at $\mathbf{x}_s$, and since according to (4.25) $\mathbb{A}$ is contained in the set

$$\hat{\mathbb{A}} = \left\{ \mathbf{y} : \mathbb{G} \to \mathbb{C} / \sum_{n\in\mathbb{G}} \text{Re}\left(y[n]\overline{P[n]}\right) = \|\mathbf{x}_{s,n\notin\Phi}\|_1 \right\},$$

then $\hat{\mathbb{A}} \cup \mathbb{B}$ is contained in the minimal facet of $\mathbb{B}$, which is the set of all functions whose supports are contained in the support of $\mathbf{x}_s$, or

$$\hat{\mathbb{A}} \cup \mathbb{B} \subseteq \{\mathbf{y} : \mathbb{G} \to \mathbb{C}/\text{supp}(\mathbf{y}) \subseteq \mathbb{T}\}.$$

Hence, every signal $\mathbf{y}$ that satisfies both the conditions

$$\|\mathbf{y}_{n\notin\Phi}\|_1 = \|\mathbf{x}_{s,n\notin\Phi}\|_1 \tag{4.30}$$

and

$$\sum_{n\in\mathbb{G}} \text{Re}\left(y[n]\overline{P[n]}\right) = \|\mathbf{x}_{s,n\notin\Phi}\|_1 \tag{4.31}$$

must also satisfy

$$y[n] = 0 \; \forall \, n \notin \mathbb{T}. \tag{4.32}$$

Now, we want to show that $|P[n]| < 1 \; \forall \, n \notin \mathbb{T} - \Phi$ (for $n \in \Phi$, we have already shown that $P[n] = 0$). We will assume that there is a point $n_0 \notin \mathbb{T} - \Phi$ such that $|P[n_0]| = 1$ and show that under this condition there exists a function $\mathbf{y}$ satisfying (4.30) and (4.31) that does not satisfy (4.32), which is a contradiction.

Supposing that $|P[n_0]| = 1$ for a $n_0 \notin \mathbb{T} - \Phi$, consider the function $\mathbf{y}$ defined by

$$y[n] = \begin{cases} \text{sgn}(P[n_0])\|\mathbf{x}_{s,n\notin\Phi}\|_1, & \text{if } n = n_0, \\ 0, & \text{otherwise.} \end{cases}$$

Then,

$$\|\mathbf{y}_{n\notin\Phi}\|_1 = \sum_{n\notin\Phi} |y[n]| = |\text{sgn}(P[n_0])\|\mathbf{x}_{s,n\notin\Phi}\|_1| = \|\mathbf{x}_{s,n\notin\Phi}\|_1,$$

meaning that $\mathbf{y}$ satisfies (4.30). Also,

$$\sum_{n\in\mathbb{G}} \text{Re}\left(y[n]\overline{P[n]}\right) = \|\mathbf{x}_{s,n\notin\Phi}\|_1,$$

which shows that $\mathbf{y}$ satisfies (4.31).

We have thus shown that if $|P[n_0]| = 1$ for some $n_0 \notin (\mathbb{T} - \Phi)$, then the function $\mathbf{y}$ satisfies (4.30) and (4.31). Since $\mathbf{y}$ does not satisfy (4.32), as $y[n_0] = \text{sgn}(P[n_0])\|\mathbf{x}_{s,n\notin\Phi}\|_1$, this is a

contradiction. Therefore, $|P[n]| \neq 1 \; \forall \, n \in \mathbb{T} - \Phi$. Together with (4.27) and (4.28), this leads to

$$|P[n]| < 1 \, \forall n \in \mathbb{T},$$

thus proving condition C3.

Finally, we must prove that $P[n]$ satisfies condition C4. Consider an arbitrary frequency location $k_0 \in \mathbb{G} - \mathbb{L}$. Also, let $\mathbf{y_1}$, $\mathbf{y_2}$, $\mathbf{y_3}$ be the signals whose DFTs are defined, respectively, by

$$\hat{y}_1[k] = \begin{cases} \hat{x}_s[k], & \text{if } k \in \mathbb{L}, \\ 0, & \text{otherwise,} \end{cases} \quad \hat{y}_2[k] = \begin{cases} \hat{x}_s[k], & \text{if } k \in \mathbb{L}, \\ 1, & \text{if } k = k_0, \\ 0, & \text{otherwise,} \end{cases} \quad \hat{y}_3[k] = \begin{cases} \hat{x}_s[k], & \text{if } k \in \mathbb{L}, \\ j, & \text{if } k = k_0, \\ 0, & \text{otherwise.} \end{cases}$$

Since $\text{DFT}(\mathbf{y}_1)|_{k \in \mathbb{L}} = \text{DFT}(\mathbf{y}_2)|_{k \in \mathbb{L}} = \text{DFT}(\mathbf{y}_3)|_{k \in \mathbb{L}} = \mathbf{b}$, we have $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3 \in \mathbb{A}$. From (4.25) combined with Parseval's theorem, and with $\hat{P}$ the DFT of P, we therefore obtain

$$\sum_{k \in \mathbb{G}} \text{Re}\left(\hat{y}_1[k]\overline{\hat{P}[k]}\right) = \sum_{k \in \mathbb{L}} \text{Re}\left(\hat{x}_s[k]\overline{\hat{P}[k]}\right) = \|\mathbf{x}_{s,n \notin \Phi}\|_1, \tag{4.33}$$

$$\sum_{k \in \mathbb{G}} \text{Re}\left(\hat{y}_2[k]\overline{\hat{P}[k]}\right) = \sum_{k \in \mathbb{L}} \text{Re}\left(\hat{x}_s[k]\overline{\hat{P}[k]}\right) + \text{Re}\left(\overline{\hat{P}[k_0]}\right) = \|\mathbf{x}_{s,n \notin \Phi}\|_1, \tag{4.34}$$

and

$$\sum_{k \in \mathbb{G}} \text{Re}\left(\hat{y}_3[k]\overline{\hat{P}[k]}\right) = \sum_{k \in \mathbb{L}} \text{Re}\left(\hat{x}_s[k]\overline{\hat{P}[k]}\right) + \text{Re}\left(j\overline{\hat{P}[k_0]}\right) = \|\mathbf{x}_{s,n \notin \Phi}\|_1. \tag{4.35}$$

Finally, comparing (4.33) with (4.34) and (4.33) with (4.35) we obtain, respectively, $\text{Re}\left(\hat{P}[k_0]\right) = 0$ and $\text{Re}\left(j\overline{\hat{P}[k_0]}\right) = \text{Re}\left(\hat{P}[k_0]\right) = 0$, so $\hat{P}[k_0] = 0$. Since this development is valid for any $k_0 \in \mathbb{G} - \mathbb{L}$, we conclude that

$$\hat{P}[k] = 0 \; \forall \, k \notin \mathbb{L},$$

which corresponds to condition C4.

We have thus shown that if the solution $\mathbf{x}^*$ to (4.13) is unique and matches $\mathbf{x}_s$, then there exists a function $P : \mathbb{G} \to \mathbb{C}$ that satisfies C1, C2, C3, and C4. This completes the proof. □

4.2.2 Comparison of the Conditions for Reconstruction

Once we stated the conditions that must happen in order to the solution of (4.11) to match $\mathbf{x}_s$, we need to compare the probability that these conditions are satisfied, in the cases with and without prior information about the signal's support. In this and the next section we discuss this issue.

In the case in which no prior information about the support is implied, Candés et al. presented a function $\mathbf{p}_1$, defined for a given signal $\mathbf{x}_s$ and a given set of sampling locations $\mathbb{L}$, that by construction meets some of the conditions for reconstruction based on ℓ_1-minimization [10]. Specifically, $\mathbf{p}_1$ is guaranteed to match the sign of $\mathbf{x}_s$ in all the positions of its support, and the DFT of $\mathbf{p}_1$ is supported in $\mathbb{L}$. Regarding the remaining necessary condition for reconstruction without prior information, which states that $\mathbf{p}_1$ outside the support of $\mathbf{x}_s$ must be less than 1 in absolute value, they analyze the probability that this is satisfied, as a function of the number of measurements taken and the signal's length and sparsity.

In the case of partial prior information about $\mathbb{T} = \text{supp}(\mathbf{x}_s)$, we can adopt a similar approach, but taking into account that the conditions for the solution to (4.11) to match $\mathbf{x}_s$ are then different, according to Theorem 4. Since the function $\mathbf{p}$ must vanish in the φ locations that are known to belong to $\mathbb{T}$, while matching the sign of $\mathbf{x}_s$ in the remaining $\eta - \varphi$ locations of $\mathbb{T}$, we start by defining an N-dimensional vector $\mathbf{y}$ such that

$$y_n = \begin{cases} \text{sgn}(x_s[n]), & \text{if } n \in \mathbb{T} - \Phi, \\ 0, & \text{otherwise.} \end{cases} \tag{4.36}$$

We also define $\mathcal{K} : \mathbb{C}^N \to \mathbb{C}^\eta$ as the operator that associates any N-dimensional vector to its η components in the locations specified by set $\mathbb{T}$; hence, $\mathcal{K}(\mathbf{y})$ is the η-dimensional vector composed by φ zeros (corresponding to the known positions of $\mathbb{T}$) and by the signs of $\mathbf{x}_s$ in the $\eta - \varphi$ points belonging to $\mathbb{T} - \Phi$. Finally, let $\mathbf{z}$ be an ℓ-dimensional vector such that

$$\mathbf{F}^H_{\mathbb{T}\to\mathbb{L}}\mathbf{z} = \mathcal{K}(\mathbf{y}), \tag{4.37}$$

where $\mathbf{F}_{\mathbb{T}\to\mathbb{L}}$ is the submatrix of the $N \times N$ DFT matrix composed by its columns specified by $\mathbb{T}$ and rows specified by $\mathbb{L}$, and where $\mathbf{F}^H_{\mathbb{T}\to\mathbb{L}}$ is the Hermitian of that submatrix.

Now, let $\mathbf{z}_c$ be the vector whose components at positions in $\mathbb{L}$ match $\mathbf{z}$ and whose remaining $N - \ell$ components are zero. According to (4.37), the IDFT of $\mathbf{z}_c$ sampled at the locations specified by $\mathbb{L}$ equals $\mathcal{K}(\mathbf{y})$. Therefore, if we obtain a vector $\mathbf{z}$ that satisfies (4.37), determine the corresponding $\mathbf{z}_c$ vector by filling with zeros in the $N - \ell$ positions outside $\mathbb{L}$, and compute the IDFT of $\mathbf{z}_c$, we obtain a vector that satisfies conditions C1, C2, and C4 of Theorem 4. In fact, since the IDFT of $\mathbf{z}_c$ matches $\mathcal{K}(\mathbf{y})$ in the locations specified by $\mathbb{T}$, according to (4.36) it follows that this IDFT is zero in the φ known positions of $\mathbb{T}$ (condition C2) and matches the sign of $\mathbf{x}_s$ in the remaining positions of $\mathbb{T}$ (condition C1). Also, since $\mathbf{z}_c$ is obtained by zero filling $\mathbf{z}$ in the locations outside $\mathbb{L}$, it also satisfies condition C4 ($\mathbf{z}_c$ is supported in $\mathbb{L}$).

Observe, however, that (4.37) is as underdetermined system, as $\eta < \ell$. Hence, it can potentially have infinite solutions, and we wish to obtain one for which all the components that outside the locations in $\mathbb{T}$ are less than 1 in absolute value (condition C3). In an attempt to satisfy this condition, we take the solution that has minimum energy, meaning the least squares solution given by $\mathbf{z} = \mathbf{F}_{\mathbb{T}\to\mathbb{L}}(\mathbf{F}^H_{\mathbb{T}\to\mathbb{L}}\mathbf{F}_{\mathbb{T}\to\mathbb{L}})^{-1}\,\mathcal{K}(\mathbf{y})$, supposing that $\mathbf{F}^H_{\mathbb{T}\to\mathbb{L}}\mathbf{F}_{\mathbb{T}\to\mathbb{L}}$ is invertible (which corresponds to the hypothesis, in the statement of Theorem 4, that $\mathcal{F}_{\mathbb{T}\to\mathbb{L}}$ is an injection). Now, zero-filling $\mathbf{z}$ in the locations outside $\mathbb{L}$ and obtaining its IDFT is equivalent to computing

$$\mathbf{p} = \mathbf{F}^H_{\mathbb{L}}\mathbf{F}_{\mathbb{T}\to\mathbb{L}}(\mathbf{F}^H_{\mathbb{T}\to\mathbb{L}}\mathbf{F}_{\mathbb{T}\to\mathbb{L}})^{-1}\,\mathcal{K}(\mathbf{y}), \tag{4.38}$$

where $\mathbf{F}_{\mathbb{L}}$ is the submatrix of the $N \times N$ DFT matrix composed by its rows specified by set $\mathbb{L}$, and $\mathbf{F}^H_{\mathbb{L}}$ is the Hermitian of that submatrix. We emphasize here that for $\varphi = 0$, $\mathbf{p}$ becomes the same testing function defined in [10], where no prior information is assumed.

4.2.3 Probability of Reconstruction

In order to study the probabilities of reconstruction with and without prior information, we address three main points: (i) the change in the variances of the test coefficients, (ii)

the Chebychev lower bounds for the probability of reconstruction, and (iii) the closed form expression for the probability of reconstruction. These points are detailed next.

i) Variances of the test coefficients

As seen above, the function in (4.38) already satisfies, by construction, the conditions C1, C2, C4 of Theorem 4. The additional necessary condition, namely that $p[n]$ must be less than 1, in absolute value, for all $n \notin \text{supp}(\mathbf{x})$, is attained with a certain probability that depends on the signals' statistical properties.

In this section, we prove that the Chebychev lower bound for this probability increases with the number of known positions φ in the sparse representation's support. The exact value of the probability depends on the distributions of the signals' signs ($\text{sgn}(x[n])$); therefore, we analyze the Chebychev bound as it is the sharpest possible bound when only first and second moments of the distributions are available [94], [73], [61], [7].

In order to compare the Chebychev bounds with and without information, we start by defining the vectors $\mathbf{v}$ and $\mathbf{w}$ which are composed by the components of $\mathbf{p}$ which are outside the support, respectively for the cases without and with prior information:

$$\mathbf{v} = \mathbf{p}_v|_{n \notin \text{supp}(\mathbf{x})} \tag{4.39}$$

and

$$\mathbf{w} = \mathbf{p}_w|_{n \notin \text{supp}(\mathbf{x})}\,, \tag{4.40}$$

where $\mathbf{p}_v$ and $\mathbf{p}_w$ are the vectors given by (4.38), respectively without prior information and with φ known locations of prior information. With these definitions, and once conditions C1, C2, and C4 are satisfied, the reconstruction by ℓ_1-minimization is possible, according to Theorem 4, if and only if the components of $\mathbf{v}$ or of $\mathbf{w}$ are all less than 1, in absolute value.

As we show below, if the components of $\mathbf{x}$ have zero mean, also do the components of $\mathbf{v}$ and $\mathbf{w}$. Hence, the probability that these components are between 0 and 1, in absolute value, is a decreasing function of their variances. The following Theorem 5 compares the

variance $\sigma^2_{v_n}$ of a single component v_n of $\mathbf{v}$ with the variance $\sigma^2_{w_n}$ of a single component w_n of $\mathbf{w}$. In particular, it shows that

$$\sigma^2_{v_n} - \sigma^2_{w_n} = \sum_{i=1}^{\varphi} |a_{n,i}|^2, \tag{4.41}$$

where $a_{n,i}$ is a known scalar that depends on n and i. Note from (4.41) that $\sigma^2_{w_n}$ is always lower than $\sigma^2_{v_n}$, provided that $\varphi > 0$. Furthermore, as the difference between $\sigma^2_{v_n}$ and $\sigma^2_{w_n}$ depends on the location n, for the considered set of assumptions, the theorem also provides the average difference for all n.

In stating the theorem, we adopt the following notation. Given a signal length N, $\mathbf{F}$ is the $N \times N$ DFT matrix, whereas $\mathbf{F}_{\mathbb{A},\mathbb{B}}$ is the submatrix of $\mathbf{F}$ which includes its rows specified by set $\mathbb{A}$ and its columns specified by set $\mathbb{B}$. We also define the submatrices

$$\mathbf{G}_1 = \mathbf{F}_{\mathbb{L},\overline{\text{supp}(\mathbf{x})}}, \ \mathbf{G}_2 = \mathbf{F}_{\mathbb{L},\text{supp}(\mathbf{x})}, \text{ and } \mathbf{G}_3 = \mathbf{F}_{\mathbb{L},\Phi}, \tag{4.42}$$

where $\text{supp}(\mathbf{x})$ is the support of $\mathbf{x}$ and set $\overline{\mathbb{A}}$ represents the complement of set $\mathbb{A}$. Additionally, $E(v)$ is the expected value of a random variable v and $E(v_1 v_2^*)$ is the covariance between v_1 and v_2 (assuming that v_1 and v_2 have zero mean). Finally, given a random vector $\mathbf{v}$, $\mathbf{C}_\mathbf{v}$ represents its covariance matrix.

Theorem 5. *Let $\mathbf{x} : \mathbb{D} \to \mathbb{C}$, with $\mathbb{D} = \{0,\ 1,\ \ldots,\ , N-1\}$, be an N-dimensional, η-sparse random vector with DFT $\hat{\mathbf{x}}$ and with a deterministic support $supp(\mathbf{x})$. Suppose that the η nonzero entries of $\mathbf{x}$ are uncorrelated and have zero mean, and that their signs, denoted by $sgn(x[n])$, have variance σ_s^2. Also, let $\mathbb{L}$ be a deterministic set of ℓ locations where $\hat{x}$ is sampled, and Φ be a set of known locations (prior information) in the support of $\mathbf{x}$. Finally, define $\mathbf{p}_v$ as the minimum-energy vector that satisfies conditions C1 to C4 in Theorem 4 without prior information ($\Phi = \emptyset$), and $\mathbf{p}_w$ as the minimum-energy vector that satisfies the same conditions with φ known positions in $supp(\mathbf{x})$.*

If

$$\mathbf{v} = \mathbf{p}_v|_{n \notin supp(\mathbf{x})}$$

and

$$\mathbf{w} = \mathbf{p}_w|_{n \notin supp(\mathbf{x})},$$

then the covariances between v_n *and* v_m *and between* w_n *and* w_m *(*$n, m \in \{1, 2, \ldots, N - \eta\}$*) are given by*

$$E\left(v_n v_m^*\right) = \sigma_s^2 \, \hat{\mathbf{e}}_n^H \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \left(\mathbf{G}_2 \mathbf{G}_2^H\right) \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \hat{\mathbf{e}}_m \tag{4.43}$$

and

$$E\left(w_n w_m^*\right) = \sigma_s^2 \, \hat{\mathbf{e}}_n^H \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \left(\mathbf{G}_2 \mathbf{G}_2^H - \mathbf{G}_3 \mathbf{G}_3^H\right) \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \hat{\mathbf{e}}_m, \tag{4.44}$$

where $\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3$ *are defined according* (4.42)*, and* $\hat{\mathbf{e}}_n$ *is the* n*-th column of* $\mathbf{G}_1$*.*

Under the same conditions, the differences in the variances of each component of $\mathbf{v}$ *and* $\mathbf{w}$ *can be expressed as*

$$\sigma_{v_n}^2 - \sigma_{w_n}^2 = \sum_{i=1}^{\varphi} |a_{n,i}|^2, \tag{4.45}$$

where

$$a_{n,i} = \sigma_s^2 \, \hat{\mathbf{e}}_n^H \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \mathbf{e}_i,$$

with $\mathbf{e}_i$ *the* i*-th column of* $\mathbf{G}_3$*.*

Finally, the trace of the difference between the covariance matrices of $\mathbf{v}$ *and* $\mathbf{w}$*,* T_d*, satisfies the relations*

$$T_d = \sigma_s^2 \sum_{i=1}^{\varphi} \mathbf{e}_i^H \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \mathbf{e}_i \tag{4.46}$$

and

$$T_d > \frac{\varphi(N - \ell)\sigma_s^2}{N}. \tag{4.47}$$

Proof. In the case of $\mathbf{p}_v$, defined in (4.39) and corresponding to no prior information being available, we have, from the definition of the DFT,

$$\hat{p}_v[k] = \sum_{n=0}^{N-1} p_v[n] W_N^{kn},$$

which can be rewritten as

$$\hat{p}_v[k] = \sum_{n \in \text{ supp}(\mathbf{x})} p_v[n] W_N^{kn} + \sum_{n \notin \text{ supp}(\mathbf{x})} p_v[n] W_N^{kn}. \tag{4.48}$$

By using (4.48) with conditions C1 and C4 from Theorem 4, we then obtain

$$\sum_{n \notin \text{ supp}(\mathbf{x})} p_v[n] W_N^{kn} = - \sum_{n \in \text{ supp}(\mathbf{x})} \text{sgn}(x[n]) W_N^{kn} \;\; \forall \; k \notin \mathbb{L}, \tag{4.49}$$

which in matrix form becomes

$$\mathbf{G}_1 \mathbf{v} = -\mathbf{G}_2 \, \mathcal{K}(\mathbf{s}), \tag{4.50}$$

using the notation in (4.42).

Note that (4.50) is an underdetermined system. Taking the minimum-energy solution, we then obtain

$$\mathbf{v} = -\mathbf{G}_1^H \left(\mathbf{G}_1 \mathbf{G}_1^H \right)^{-1} \mathbf{G}_2 \, \mathcal{K}(\mathbf{s}). \tag{4.51}$$

Now, in the case of $\mathbf{p}_w$, defined in (4.40) and corresponding to the set Φ of known support locations,

$$\hat{p}_w[k] = \sum_{n \in \text{ supp}(\mathbf{x}) - \Phi} p_w[n] W_N^{kn} + \sum_{n \notin \text{ supp}(\mathbf{x})} p_w[n] W_N^{kn}, \tag{4.52}$$

since $p_w[n] = 0 \; \forall \, n \in \Phi$, from condition C2. By using (4.52) with conditions C1 and C4, we obtain

$$\sum_{n \notin \text{ supp}(\mathbf{x})} p_w[n] W_N^{kn} = - \sum_{n \in \text{ supp}(\mathbf{x}) - \Phi} \text{sgn}(x[n]) W_N^{kn} \;\; \forall \; k \notin \mathbb{L},$$

which is equivalent to

$$\sum_{n \notin \text{ supp}(\mathbf{x})} p_w[n] W_N^{kn} = - \sum_{n \in \text{ supp}(\mathbf{x})} \text{sgn}(x[n]) W_N^{kn} + \sum_{n \in \Phi} \text{sgn}(x[n]) W_N^{kn} \;\; \forall \; k \notin \mathbb{L}. \tag{4.53}$$

Now, define

$$\mathbf{e}_i = \begin{bmatrix} W_N^{k_1 m_i} & W_N^{k_2 m_i} & \dots & W_N^{k_{N-\ell} m_i} \end{bmatrix}^T,$$

with m_i the i-th element of the set $\Phi = \{m_1,\ m_2,\ \ldots,\ m_\varphi\}$ of known support locations, and $\{k_1,\ k_2,\ \ldots,\ k_{N-\ell}\}$ the set of $N-\ell$ frequency locations that are not sampled. Note then that $\mathbf{e}_i$ is the i-th column of $\mathbf{G}_2 = \mathbf{F}_{\mathbb{L},\mathrm{supp}(\mathbf{x})}$, so that (4.53) can be written in matrix form as

$$\mathbf{G}_1\mathbf{w} = -\mathbf{G}_2\,\mathcal{K}(\mathbf{s}) + \sum_{i=1}^{\varphi} \mathrm{sgn}(x[m_i])\mathbf{e}_i, \tag{4.54}$$

or, equivalently,

$$\mathbf{G}_1\mathbf{w} = -\mathbf{G}_2\,\mathcal{K}(\mathbf{y}), \tag{4.55}$$

where $\mathbf{y}$ is defined according to (4.36).

Taking the minimum-energy solution to (4.55), we then obtain

$$\mathbf{w} = -\mathbf{G}_1^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \mathbf{G}_2\,\mathcal{K}(\mathbf{y}). \tag{4.56}$$

In the particular case in which all the positions are known, $\Phi = \mathrm{supp}(\mathbf{x})$ and hence $\mathbf{y} = \mathbf{0}$, so that (4.56) leads to $\mathbf{w} = [0\ 0\ \ldots\ 0]^T$. This shows that for $\Phi = \mathrm{supp}(\mathbf{x})$, the polynomial defined by (4.38) always satisfies the conditions C1 to C4 in Theorem 4 provided that $\ell \geqslant \eta$. Therefore, $\ell = \eta$ measurements are sufficient for reconstruction when all the support locations are known.

Since the entries of $\mathbf{x}$ are assumed to have zero mean, from (4.51) and (4.56) we conclude that $\mathbf{v}$ and $\mathbf{w}$ also have zero mean, and it then follows that their covariance matrices are given by

$$\mathbf{C}_\mathbf{v} = \mathbf{G}_1^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \mathbf{G}_2\mathbf{C}_{\mathcal{K}(\mathbf{s})}\mathbf{G}_2^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \mathbf{G}_1$$

and

$$\mathbf{C}_\mathbf{w} = \mathbf{G}_1^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \mathbf{G}_2\mathbf{C}_{\mathcal{K}(\mathbf{y})}\mathbf{G}_2^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \mathbf{G}_1.$$

Also, as the entries of $\mathbf{x}$ are assumed uncorrelated and the variances of the components of $\mathcal{K}(\mathbf{s})$ are equal to σ_s^2, $\mathbf{C}_{\mathcal{K}(\mathbf{s})}$ is the identity matrix times σ_s^2. On the other hand, $\mathbf{C}_{\mathcal{K}(\mathbf{y})}$ is the

diagonal matrix with zeros in the diagonal locations corresponding to Φ and σ_s^2 in each of the other diagonal locations. Therefore, we obtain

$$\mathbf{C_v} = \sigma_s^2 \, \mathbf{G}_1^H \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \mathbf{G}_2 \mathbf{G}_2^H \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \mathbf{G}_1$$

and

$$\mathbf{C_w} = \sigma_s^2 \, \mathbf{G}_1^H \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \left(\mathbf{G}_2 \mathbf{G}_2^H - \mathbf{G}_3 \mathbf{G}_3^H\right) \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \mathbf{G}_1,$$

using the notation in (4.42).

Now, by evaluating the elements in row n and column m of $\mathbf{C_v}$ and of $\mathbf{C_w}$, we obtain, respectively

$$E\left(v_n v_m^*\right) = \sigma_s^2 \, \hat{\mathbf{e}}_n^H \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \left(\mathbf{G}_2 \mathbf{G}_2^H\right) \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \hat{\mathbf{e}}_m \tag{4.57}$$

and

$$E\left(w_n w_m^*\right) = \sigma_s^2 \, \hat{\mathbf{e}}_n^H \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \left(\mathbf{G}_2 \mathbf{G}_2^H - \mathbf{G}_3 \mathbf{G}_3^H\right) \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \hat{\mathbf{e}}_m, \tag{4.58}$$

thus proving (4.43) and (4.44).

Now, in order to prove (4.45), we observe from (4.57) and (4.58) that

$$E\left(|v_n|^2\right) = \sigma_s^2 \, \hat{\mathbf{e}}_n^H \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \left(\mathbf{G}_2 \mathbf{G}_2^H\right) \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \hat{\mathbf{e}}_n$$

and

$$E\left(|w_n|^2\right) = \sigma_s^2 \, \hat{\mathbf{e}}_n^H \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \left(\mathbf{G}_2 \mathbf{G}_2^H - \mathbf{G}_3 \mathbf{G}_3^H\right) \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \hat{\mathbf{e}}_n,$$

so

$$E\left(|v_n|^2\right) - E\left(|w_n|^2\right) = \sigma_s^2 \, \hat{\mathbf{e}}_n^H \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \left(\mathbf{G}_3 \mathbf{G}_3^H\right) \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \hat{\mathbf{e}}_n.$$

Note that if $\mathbf{e}_i$ is the i-th column of $\mathbf{G}_3$, then

$$\mathbf{G}_3 \mathbf{G}_3^H = \sum_{i=1}^{\varphi} \mathbf{e}_i \mathbf{e}_i^H,$$

so

$$E\left(|v_n|^2\right) - E\left(|w_n|^2\right) = \sigma_s^2\, \hat{\mathbf{e}}_n^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \left(\sum_{i=1}^{\varphi} \mathbf{e}_i\mathbf{e}_i^H\right) \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \hat{\mathbf{e}}_n,$$

which leads to

$$E\left(|v_n|^2\right) - E\left(|w_n|^2\right) = \sigma_s^2 \sum_{i=1}^{\varphi} \left(\hat{\mathbf{e}}_n^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \mathbf{e}_i\mathbf{e}_i^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \hat{\mathbf{e}}_n\right). \tag{4.59}$$

By defining

$$a_{n,i} = \sigma_s^2\, \hat{\mathbf{e}}_n^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \mathbf{e}_i,$$

we then obtain

$$E\left(|v_n|^2\right) - E\left(|w_n|^2\right) = \sum_{i=1}^{\varphi} \left(a_{n,i}a_{n,i}^*\right), \tag{4.60}$$

with $a_{n,i}^*$ the complex conjugate of $a_{n,i}$.

From (4.60), it then follows that

$$E\left(|v_n|^2\right) - E\left(|w_n|^2\right) = \sum_{i=1}^{\varphi} |a_{n,i}|^2,$$

which corresponds to (4.45).

Finally, we must evaluate the trace T_d of the difference between the covariances of v_n and of w_n. Taking the sum of the diagonal components in (4.59), for all n, we obtain

$$T_d = \sigma_s^2 \sum_{n=1}^{N-\eta} \sum_{i=1}^{\varphi} \hat{\mathbf{e}}_n^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \mathbf{e}_i\mathbf{e}_i^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \hat{\mathbf{e}}_n,$$

which is equivalent to

$$T_d = \sigma_s^2 \sum_{i=1}^{\varphi} \sum_{n=1}^{N-\eta} \hat{\mathbf{e}}_n^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \mathbf{e}_i\mathbf{e}_i^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \hat{\mathbf{e}}_n;$$

now, note that $\hat{\mathbf{e}}_n^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \mathbf{e}_i\mathbf{e}_i^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \hat{\mathbf{e}}_n = \mathbf{e}_i^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \hat{\mathbf{e}}_n\hat{\mathbf{e}}_n^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \mathbf{e}_i$, so

$$T_d = \sigma_s^2 \sum_{i=1}^{\varphi} \sum_{n=1}^{N-\eta} \mathbf{e}_i^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \hat{\mathbf{e}}_n\hat{\mathbf{e}}_n^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \mathbf{e}_i.$$

Therefore,

$$T_d = \sigma_s^2 \sum_{i=1}^{\varphi} \mathbf{e}_i^H \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \sum_{n=1}^{N-\eta} \left(\hat{\mathbf{e}}_n \hat{\mathbf{e}}_n^H\right) \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \mathbf{e}_i.$$

Since $\hat{\mathbf{e}}_n$ is the n-th column of $\mathbf{G}_1$,

$$\sum_{n=1}^{N-\eta} \left(\hat{\mathbf{e}}_n \hat{\mathbf{e}}_n^H\right) = \mathbf{G}_1 \mathbf{G}_1^H,$$

and we obtain

$$T_d = \sigma_s^2 \sum_{i=1}^{\varphi} \mathbf{e}_i^H \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \mathbf{e}_i.$$

By normalizing $\mathbf{e}_i$ in this expression, we get

$$T_d = \sigma_s^2 \sum_{i=1}^{\varphi} \|\mathbf{e}_i\|^2 \left(\frac{\mathbf{e}_i^H}{\|\mathbf{e}_i^H\|} \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \frac{\mathbf{e}_i}{\|\mathbf{e}_i\|} \right). \tag{4.61}$$

Note in (4.61) that the expression between parentheses is a Rayleigh product. Furthermore, since $\left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1}$ is a Hermitian matrix, and $\frac{\mathbf{e}_i}{\|\mathbf{e}_i\|}$ has norm 1, it follows from Courant-Fisher minimax theorem [39] that

$$\frac{\mathbf{e}_i^H}{\|\mathbf{e}_i^H\|} \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \frac{\mathbf{e}_i}{\|\mathbf{e}_i\|} \geqslant \lambda_{\min}, \tag{4.62}$$

where $\lambda_{\min}$ is the minimum eigenvalue of $\mathbf{G}_1 \mathbf{G}_1^H$. Using

$$\hat{\lambda}_{\max} < N$$

we obtain

$$\lambda_{\min} > \frac{1}{N},$$

and from (4.61) and (4.62), it follows that

$$T_d \geqslant \sigma_s^2 \sum_{i=1}^{\varphi} \|\mathbf{e}_i\|^2 \frac{1}{N}.$$

Finally, using $\|\mathbf{e}_i\|^2 = N - \ell$, we obtain

$$T_d \geqslant \frac{\varphi(N-\ell)\sigma_s^2}{N},$$

which completes the proof. □

ii) Chebychev's lower bounds

Now, from Chebychev's inequality [84], [51], we have

$$P\left(|v_n - E(v_n)| \geqslant 1\right) \leqslant \sigma_{v_n}^2,$$

where $P\left(|v_n - E(v_n)| > 1\right)$ denotes the probability that v_n deviates more than one unity from its mean $E(v_n)$, in absolute value. Since $E(v_n) = 0$ under the conditions of Theorem 5, we also obtain

$$P\left(|v_n| < 1\right) \geqslant 1 - \sigma_{v_n}^2.$$

We can then define

$$P_{v_n} = 1 - \sigma_{v_n}^2 \tag{4.63}$$

as the Chebychev lower bound for the probability that each component v_n of $\mathbf{v}$ is in the desired range $[0, 1)$ needed for reconstructibility by ℓ_1-minimization. Analogously,

$$P_{w_n} = 1 - \sigma_{w_n}^2 \tag{4.64}$$

is the Chebychev bound for the probability that w_n is in the range $[0, 1)$.

Using the result of Theorem 5, we then get

$$P_{w_n} = P_{v_n} + \sum_{i=1}^{\varphi} |a_{n,i}|^2, \tag{4.65}$$

which shows that the lower bound for the probability that each component of the testing vector is in absolute value below 1 increases by $\sum_{i=1}^{\varphi} |a_{n,i}|^2$ when using prior information.

Note, however, that this result assumes a deterministic support $\text{supp}(\mathbf{x})$ and a deterministic set of sampling locations $\mathbb{L}$. In a more realistic context, both these sets are also random, but the deterministic model is useful as a step to prove that the probability bound also increases in the random case. In fact, suppose that $\text{supp}(\mathbf{x})$ and $\mathbb{L}$ have uniform distribution

over all possible combinations of, respectively, η and ℓ locations; in this case, the variances of $\mathbf{v}$ and $\mathbf{w}$ are the arithmetic averages of the variances obtained for each possible support set and sampling set (as each combination has the same probability):

$$\tilde{\sigma}^2_{v_n} = \overline{\hat{\mathbf{e}}_n^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \left(\mathbf{G}_2\mathbf{G}_2^H\right) \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \hat{\mathbf{e}}_n}\ \sigma_s^2 \tag{4.66}$$

and

$$\tilde{\sigma}^2_{w_n} = \overline{\hat{\mathbf{e}}_n^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \left(\mathbf{G}_2\mathbf{G}_2^H - \mathbf{G}_3\mathbf{G}_3^H\right) \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \hat{\mathbf{e}}_n}\ \sigma_s^2, \tag{4.67}$$

where scalar $\overline{a}$ denotes the arithmetic average of scalar a for all possible choices of $\text{supp}(\mathbf{x})$ and $\mathbb{L}$.

Furthermore, since for each possible choice of $\text{supp}(\mathbf{x})$ and $\mathbb{L}$ it is already shown that $\sigma^2_{w_n} < \sigma^2_{v_n}$, then the averages of these variances over all possible sets also satisfy the same inequality, and thus

$$\begin{aligned}\overline{\hat{\mathbf{e}}_n^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \left(\mathbf{G}_2\mathbf{G}_2^H - \mathbf{G}_3\mathbf{G}_3^H\right) \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \hat{\mathbf{e}}_n}\ \sigma_s^2 \ <\\ \overline{\hat{\mathbf{e}}_n^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \left(\mathbf{G}_2\mathbf{G}_2^H\right) \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \hat{\mathbf{e}}_n}\ \sigma_s^2.\end{aligned} \tag{4.68}$$

Using (4.63), (4.64), (4.66) and (4.67), we then get

$$\begin{aligned}P_w = P_v + \overline{\hat{\mathbf{e}}_n^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \left(\mathbf{G}_2\mathbf{G}_2^H\right) \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \hat{\mathbf{e}}_n}\ \sigma_s^2 -\\ \overline{\hat{\mathbf{e}}_n^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \left(\mathbf{G}_2\mathbf{G}_2^H - \mathbf{G}_3\mathbf{G}_3^H\right) \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \hat{\mathbf{e}}_n}\ \sigma_s^2,\end{aligned} \tag{4.69}$$

and from (4.68) it follows that $P_w > P_v$.

Additionally, observe that $\mathbb{L}$ and $\text{supp}(\mathbf{x})$ being random with uniform distribution results in covariance matrices $\mathbf{C_v}$ and $\mathbf{C_w}$ that are diagonal. In fact, consider an element $E(v_n v_m^*)$ in the covariance matrix of $\mathbf{v}$, located outside the main diagonal ($n \neq m$). From (4.43), and considering the average for all possible sets $\mathbb{L}$ and $\text{supp}(\mathbf{x})$, it follows that

$$E(v_n v_m^*) = \overline{\mathbf{g}_n^H \mathbf{g}_m}, \tag{4.70}$$

where $\mathbf{g}_i^H$ is the i-th row of the matrix defined by

$$\mathbf{G} = \mathbf{G}_1(\mathbf{G}_1\mathbf{G}_1^H)^{-1}\mathbf{G}_2\ \sigma_s.$$

Note that $E(v_n v_m^*)$ is then the arithmetic average of all possible inner products between the n-th and the m-th rows of $\mathbf{G}$. Also observe that the rows of $\mathbf{G}_1$, and therefore of $\mathbf{G}$, depend on the particular choice of sampling locations $\mathbb{L}$. Furthermore, for a given choice of $\mathbb{L}$, there always exists a different choice such that row n of $\mathbf{G}$ is exactly the symmetric of the row corresponding to the first choice, whereas the row m of $\mathbf{G}$ remains unchanged. As a result, the new inner product between rows n and m (the one corresponding to the second choice of $\mathbb{L}$) is the symmetric of the previous one. The conclusion is that $E(v_n v_m^*) = 0$ for $n \neq m$, as the different values of $\mathbf{g}_n^H \mathbf{g}_m$ in (4.70) cancel each other pairwisely. Similarly, $E(w_n w_m^*) = 0$ for $n \neq m$.

A second consequence of $\mathbb{L}$ and $\text{supp}(\mathbf{x})$ being random and uniformly distributed is that the terms on the diagonal of $\mathbf{C_v}$ become all equal. The same holds for $\mathbf{C_w}$, and hence the omission of the index n in P_w and P_v in (4.69). In fact, if all possible choices of $\mathbb{L}$ and $\text{supp}(\mathbf{x})$ are taken into account, then for different values of n the expression in (4.66) provides the same result, as it corresponds to the average of the same sets of values (observe that even if n chances, all possible values of $\mathbf{G}_1$ and $\mathbf{G}_2$ are considered in the averaging process, thus giving the same result). The same reasoning applies to $\mathbf{C_w}$, based on (4.67).

In summary, under the aforementioned conditions the components of $\mathbf{v}$ become uncorrelated and have equal variance; the same holds for the components of $\mathbf{w}$. Using this idea, note that the trace of the difference between $\mathbf{C_v}$ and $\mathbf{C_w}$, given by (4.46), then becomes the sum of $N - \eta$ equal values, meaning

$$T_d = (N - \eta)(\tilde{\sigma}_v^2 - \tilde{\sigma}_w^2),$$

which together with (4.46) gives

$$\tilde{\sigma}_v^2 - \tilde{\sigma}_w^2 = \frac{\sigma_s^2}{N - \eta} \sum_{i=1}^{\varphi} \mathbf{e}_i^H \left(\mathbf{G}_1 \mathbf{G}_1^H\right)^{-1} \mathbf{e}_i.$$

Also, from (4.47),

$$\tilde{\sigma}_v^2 - \tilde{\sigma}_w^2 > \frac{\varphi(N - \ell)\sigma_s^2}{N(N - \eta)},$$

which combined with (4.63) and (4.64) leads to

$$P_w > P_v + \frac{\varphi(N-\ell)\sigma_s^2}{N(N-\eta)}. \tag{4.71}$$

Inequality (4.71) shows that if φ support locations are known, then the bound for the probability that each component of the testing vector satisfies the remaining condition C3 increases by at least $\varphi(N-\ell)\sigma_s^2/(N(N-\eta))$.

iii) Closed form expression

To conclude, note that the final probability of reconstruction can be computed if we additionally know the distributions of the nonzero components of $\mathbf{x}$, or at least the distributions of $\text{sgn}(x[n]) = s[n]$ (in contrast, the previous results in (18) and (23) do not assume any signal distribution but they only provide bounds for the final probabilities). In fact, suppose that the nonzero components of $\mathbf{x}$, given by $\mathcal{K}(\mathbf{x})$, are iid, and have a complex Gaussian distribution. The signs of these components then have a uniform distribution over the complex unit circle, as the complex Gaussian is symmetric with respect to the origin. Hence, from (4.51) in the proof of Theorem 5, but now considering $\text{supp}(\mathbf{x})$ and $\mathbb{L}$ to be random with uniform distribution,

$$\mathbf{v} = \overline{\mathbf{G}_1^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \mathbf{G}_2\, \mathcal{K}(\mathbf{s})},$$

where the components of $\mathcal{K}(\mathbf{s})$ are uniformly distributed over the unit circle and vector $\overline{\mathbf{a}}$ represents the average of vector $\mathbf{a}$ for all possible choices of $\text{supp}(\mathbf{x})$ and $\mathbb{L}$.

Now, since $\mathbf{v}$ is the result of the averaging $\mathbf{G}_1^H \left(\mathbf{G}_1\mathbf{G}_1^H\right)^{-1} \mathbf{G}_2\, \mathcal{K}(\mathbf{s})$ over a number of combinations that increases combinatorially with N, then, from the central limit theorem, as N increases the real and imaginary parts of the components of $\mathbf{v}$ approach each one a Gaussian distribution with standard deviation $\tilde{\sigma}_v/\sqrt{2}$ (note that we used the fact that if the real and imaginary parts of a random complex variable are iid, then the variance of these parts are half the variance of the complex variable). Therefore, defining $\alpha = \sqrt{2}/\tilde{\sigma}_v$, we have

that

$$\alpha^2 |v[n]|^2 = [\alpha \mathrm{Re}(v[n])]^2 + [\alpha \mathrm{Im}(v[n])]^2$$

is, on the limit, the sum of the squares of two independent variables with standard Gaussian distribution, which leads to

$$\alpha^2 |v[n]|^2 \sim \chi_2^2,$$

with χ_2^2 the chi-square distribution with 2 degrees of freedom.

Therefore, the probability density function (pdf) of $|v[n]|^2$, under the conditions above, is given by

$$f_{|v_n|^2}(\lambda) = \alpha^2 f_{\chi_2^2}\left(\alpha^2 \lambda\right), \tag{4.72}$$

where $f_{\chi_2^2}$ is the pdf of a variable with a chi-square with 2 degrees of freedom, given by [51], [84]

$$f_{\chi_2^2}(\lambda) = \frac{1}{2} \exp\left(-\frac{\lambda}{2}\right). \tag{4.73}$$

From (4.72) and (4.73), we then obtain

$$f_{|v_n|^2}(\lambda) = \frac{1}{\tilde{\sigma}_v^2} \exp\left(-\frac{\lambda}{\tilde{\sigma}_v^2}\right), \tag{4.74}$$

with $\tilde{\sigma}_v^2$ given by (4.66) using $\sigma_s^2 = 1$ (since the sign of $x[n]$ now has a uniform distribution over the unit circle).

In an analogous manner, under the same conditions, but with the added prior information of φ support locations, $\mathbf{w}$ has on the limit also a scaled chi-square distribution, but with a different scale factor to account for the different variance previously determined, so

$$f_{|w_n|^2}(\lambda) = \frac{1}{\tilde{\sigma}_w^2} \exp\left(-\frac{\lambda}{\tilde{\sigma}_w^2}\right), \tag{4.75}$$

with $\tilde{\sigma}_w^2$ given by (4.67), again using $\sigma_s^2 = 1$.

It then follows that the probabilities of $|v_n|$ and $|w_n|$ being less than 1 are, respectively,

$$P\left(|v_n| < 1\right) = 1 - \exp\left(-\frac{1}{\tilde{\sigma}_v^2}\right)$$

and

$$P\left(|w_n| < 1\right) = 1 - \exp\left(-\frac{1}{\tilde{\sigma}_w^2}\right),$$

which we obtain by integrating (4.74) and (4.75) from 0 to 1.

We then conclude by observing that, since under the conditions above the components of $\mathbf{v}$ are independent of each other, the same holding for $\mathbf{w}$, then the probabilities that all the components of $\mathbf{v}$ and of $\mathbf{w}$ are lower than 1 in absolute value are, respectively,

$$P\left(|v_1| < 1 \ \cap \ |v_2| < 1 \ \cap \ \ldots \ \cap \ |v_{N-\eta}| < 1\right) = \left(1 - \exp\left(-\frac{1}{\tilde{\sigma}_v^2}\right)\right)^{N-\eta} \tag{4.76}$$

and

$$P\left(|w_1| < 1 \ \cap \ |w_2| < 1 \ \cap \ \ldots \ \cap \ |w_{N-\eta}| < 1\right) = \left(1 - \exp\left(-\frac{1}{\tilde{\sigma}_w^2}\right)\right)^{N-\eta}, \tag{4.77}$$

which, therefore, are the probabilities that $\mathbf{x}$ is reconstructible by ℓ_1-minimization, under the stated conditions and respectively for the case without prior information and with φ known locations in $\text{supp}(\mathbf{x})$.

Observe that (4.76) and (4.77) assume that N is sufficiently large and that each $\text{sgn}(x[n])$ has, for $n \in \text{supp}(\mathbf{x})$, a uniform distribution over the unit circle in the complex plane. The condition $P_w > P_v$, regarding the Chebychev lower bounds for the probabilities of reconstruction, however, does not make any assumption on the distribution of the sign of $\mathbf{x}$ outside the support, except that the different components are independent. Also, computing $\tilde{\sigma}_v^2$ and $\tilde{\sigma}_w^2$ from (4.66) and (4.67) can become impractical as N increases, due to the large number of possibilities for $\text{supp}(\mathbf{x})$ and for $\mathbb{L}$, but we have observed that the result can converge up to a given tolerance after a limited number of tested random combinations.

As an example, we show in Fig. 4.3 the case of complex signals with length $N = 1024$ and sparsity $\eta = 60$, for different number of samples $\ell = |\mathbb{L}|$ and of known positions φ in $\mathbb{T}$. For each combination of ℓ and φ, we computed, using (4.76) and (4.77), the theoretical probabilities of reconstruction by ℓ_1-minimization. These probabilities appear in Fig. 4.3 as the solid lines, one for each value of φ. To test the proposed ideas, we also conducted a series of Monte-Carlo simulations, consisting of 1000 trials for each combination of φ and ℓ. In each

trial, a random complex signal $\mathbf{x}_s$ with Gaussian distribution, length $N = 1024$ and sparsity $\eta = 60$ was generated, with uniformly distributed random support $\mathbb{T}$. The set $\mathbb{L}$ was also randomly defined for each case, with uniform distribution over all sets of ℓ locations. The probability of $\mathbf{p}$ satisfying condition C3 was then estimated for each combination of ℓ and φ by computing the relative frequency in which all the components of $\mathbf{p}$ outside $\mathbb{T}$ were below 1, in absolute value. As observed in Fig. 4.3, increasing φ increases the probability of $\mathbf{p}$ satisfying the conditions of Theorem 4. Furthermore, the probabilities computed according to (4.76) and (4.77) match those provided by the Monte-Carlo trials.

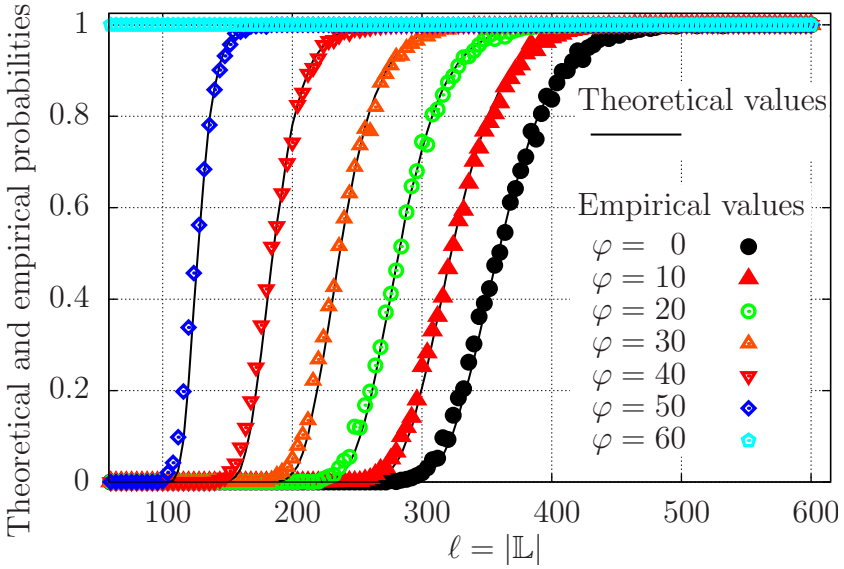


Figure 4.3. Probabilities that the conditions C1 to C4 in Theorem 4 are satisfied by the function $\mathbf{p}$ defined in (4.38) for signals with length $N = 1024$ and sparsity $\eta = 60$, as a function of the number of samples $\ell = |\mathbb{L}|$ and for different numbers of known positions φ in $\mathbb{T}$. The probabilities represented by the solid lines are computed using (4.76) and (4.77) for the different values of φ. For validation, the same probabilities (represented by the individual dots) and are also estimated using a Monte-Carlo simulation, by computing the relative frequency in which the conditions are satisfied during 1000 trials for each combination of ℓ and φ.

4.3 Singular Values of the DFT Submatrices

An important aspect of Theorem 3 is the fact that the established *theoretical* minimum number of linear measurements, $\ell = 2\eta - \varphi$, does not depend upon the length N_p of the signal. In fact, if two distinct signals, $\mathbf{x}_1$ and $\mathbf{x}_2$, have different prime lengths, say N_1 and N_2, but have the same number of non-null components η in the sparse domain and identical numbers of known positions φ for some of these nonzeros, then the minimum number of linear measurements for theoretical unambiguous reconstruction is the same for both signals. In practice, however, the reconstruction by optimization procedures that are computationally more tractable, as compared to the ideal ℓ_0-minimization, affect the number of required samples. Not only this number increases, it also becomes dependent on the signals' length.

Another important practical aspect of the reconstruction of the signals in compressive sensing refers to the finite precision used in the implementation of the optimization algorithms. This finite precision can also affect the reconstruction as the number of measurements decreases or the signal's length increases, as it becomes clear as we analyze the proof to Theorem 2.

In fact, note that the key point to the proof of Theorem 2 is the invertibility of any square submatrix $\mathbf{F}_{\mathbb{L},\mathbb{A}}$ of the DFT matrix $\mathbf{F}_{N_p \times N_p}$ with prime dimensions, where $\mathbb{L}$ and $\mathbb{A}$ are sets of rows and columns extracted from $\mathbf{F}$ such that $|\mathbb{L}| = |\mathbb{A}|$. The number of measurements $\ell = 2\eta - \varphi$ is proved to guarantee unambiguous reconstruction based on the fact that any submatrix with ℓ rows and ℓ columns extracted from the DFT matrix with prime dimensions is invertible, which is equivalent to saying that the columns of the corresponding submatrix are linearly independent.

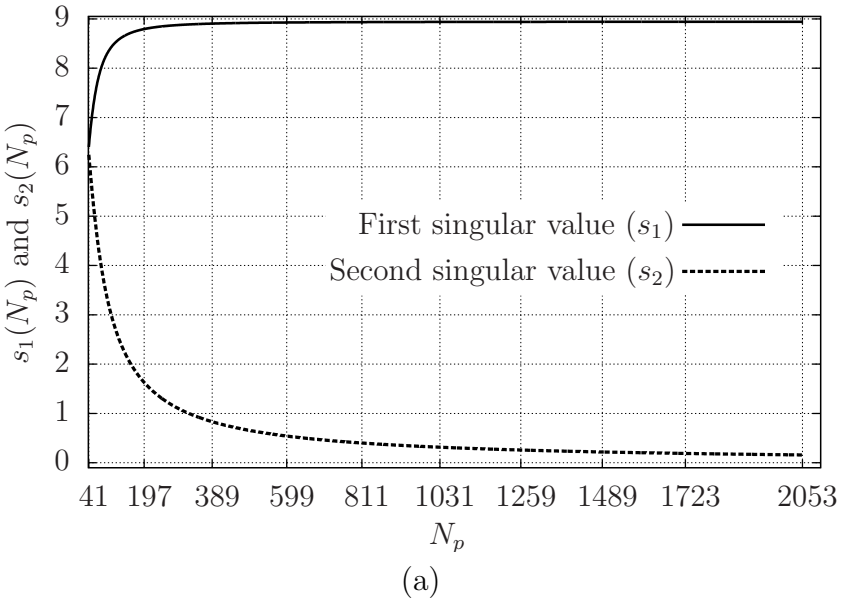
However, a different analysis, in numerical terms, of this invertibility can be carried out using the concept of singular value decomposition. Observe the plots in Fig. 4.4, where we analyze the singular values associated to two segments of columns with fixed length (rows 1 to 40) in a full DFT matrix, as we increase the dimensions of this matrix. In Fig. 4.4(a), we show that as we increase the length N_p of the input signal $\mathbf{x}$, and consequently the dimensions of the full DFT matrix, and keep constant the length $\ell = 40$ (number of measurements taken) of

two isolated segments of columns in this matrix, one of the singular values associated to these sub-columns approaches zero while the other increases. This is also evident in Fig. 4.4(b), where we show that the ratio between the two singular values approaches zero as we increase N_p. Note that a ratio equal to zero would mean that the vectors represented by the columns of the submatrix of the DFT matrix are parallel, and therefore linear dependent. The interpretation of this condition is significant: even though the ratio between the singular values associated to two sub-columns with fixed lengths in the full DFT matrix is never 0 (Theorem 1 guarantees that the submatrices composed by ℓ of these sub-columns are invertible, so two sub-columns cannot be colinear vectors), in numerical terms it can be made arbitrarily close to 0 by increasing the length of the input signal and, correspondingly, the dimensions of the DFT matrix. The direct consequence is that, taking into account the used finite-precision arithmetic, if the length of the signal is sufficiently high compared to the number of linear measurements taken, the submatrix $\mathbf{F}_{\mathbb{L},\mathbb{A}}$ can have two or more columns to be practically colinear vectors, so the matrix will be ill conditioned.

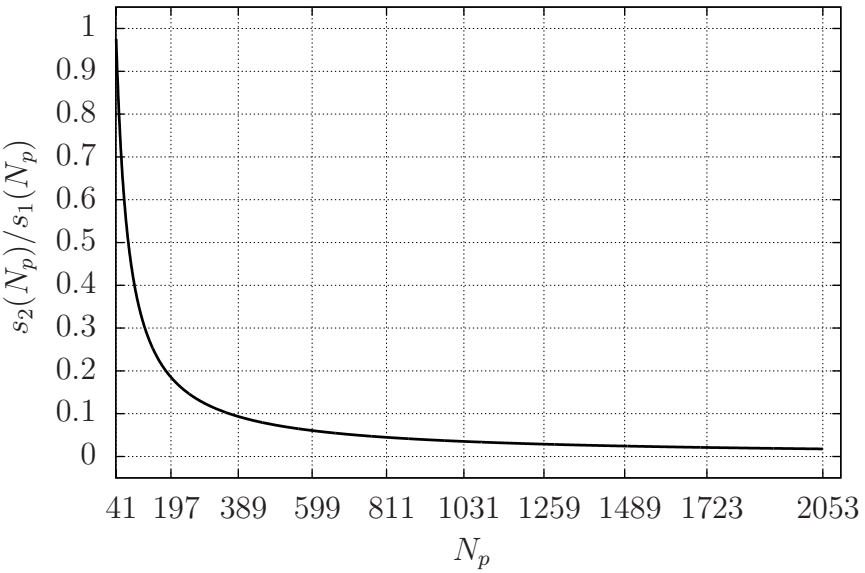
The next important question is whether the ratio between the singular values can be made sufficiently higher than zero in real applications. This problem is analyzed in Fig. 4.5, where we depict the two singular values of two segments of columns (rows 1 to ℓ) as we increase the lengths ℓ of these sub-columns while keeping constant the $N_p \times N_p$ dimension of the full matrix. Note that the increase in the number of linear measurements (corresponding to taking more samples from a DFT- or IDFT-sparse signal) leads to both singular values to increase toward the same limit. In this way, their ratio is made to increase from 0 to 1, meaning that increasing the number of measurements is enough to ensure that the ratio between the singular values departs from 0.

4.4 Results of Numerical Simulations

In order to test the main assertion of this chapter, namely that knowledge about the support of a signal's sparse representation reduces the number of measurements required for perfect reconstruction, we conducted a series of numerical simulations. First, we show the result



(a)



(b)

Figure 4.4. (a) The singular values of the submatrices defined by the first two columns, first 40 rows of the DFT matrix, as a function of its dimension N_p. (b) The ratio of the smallest singular value to the second one, as a function of N_p; note that this ratio can be arbitrarily close to 0, for a sufficiently large value of N_p.

of a DFT-sparse signal reconstructed by ℓ_1-minimization of the frequency coefficients that explain the available time samples, with and without prior information. Next, we provide the

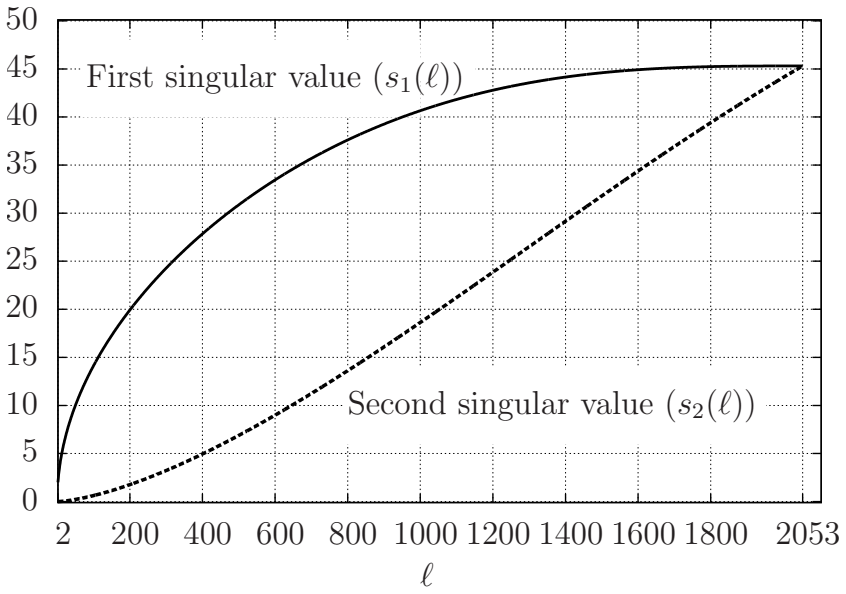


Figure 4.5. The two singular values of the submatrices defined by the first 2 columns, first ℓ rows of the DFT matrix of fixed dimension 2053×2053, as a function of ℓ. Note that both singular values increase with ℓ.

results of a more extensive set of simulations, involving several (1000) DFT-sparse random signals and different numbers of known locations φ in the support of the sparse domain. Finally, we describe the results of a similar kind of simulation, this time involving signals which are not DFT-sparse, but for which a general, randomly defined orthogonal transform provides a sparse representation.

4.4.1 Example of a DFT-Sparse Signal Reconstruction

The first example is illustrated in Fig. 4.6. A random signal $x[n]$ with 15 nonzero DFT coefficients and with length 1301 was created by using a pseudo-random number generator with Gaussian distribution to create 15 complex coefficients. The positions in frequency for these coefficients were determined by taking the integers from 0 to 1300 sampled uniformly at random. The signal $x[n]$ was then computed by taking the inverse DFT of those coefficients.

Two typical examples of reconstruction are illustrated in Fig. 4.6. The absolute values of both $x[n]$ and its reconstructed version $x_r[n]$, obtained by solving (4.10) with $\Phi = \emptyset$ (no prior information) and using only 40 time samples, is shown in Fig. 4.6(a), with its corresponding

DFT shown in Fig. 4.7(a). From the comparison between $x[n]$ and $x_r[n]$, it is clear that the number of samples taken was not enough for reconstruction, since a different signal (which also matches the taken samples) was obtained. Figure 4.6(b), on the other hand, shows the same signal $x[n]$ and its reconstructed version obtained by also minimizing the ℓ_1 of its DFT, but with the additional information on 10 positions of the support domain; the corresponding DFT is shown in Fig. 4.7(b). Note that the values of all the coefficients are unknown; only the positions, and just some of them, are given. In this case, we observe in the figure that the reconstructed signal matches the original one, with all absolute errors less than 10^{-9}.

4.4.2 Simulations using DFT-Sparse Signals

For the more extensive set of simulations, we also generated 1000 distinct signals with 16-sparse DFTs and length 211, using again a pseudo-random generator with normal distribution for the complex coefficients and with uniform distribution for their positions, and taking the inverse DFT of the obtained complex-valued signal. The generated signals were then reconstructed using varying numbers of known positions in the support while solving (4.10). In each numerical test, we also varied the number of samples taken in time (the nonsparse domain), and evaluated for each combination of number of samples and known support positions if the signal could or could not be reconstructed from this information only. As previously mentioned with regard to the example of Figure 4.6, if the signal $x[n]$ can be unambiguously determined, the optimization procedure can reconstruct it with relative errors $e[n] = (x_r[n] - x[n])/x[n]$ on the order of 10^{-9}. We considered as correctly reconstructed the signals for which the maximum absolute relative error was less than a pre-specified tolerance $\tau = 10^{-8}$.

The percentage of correct reconstructions as a function of the number of taken samples is shown in Fig. 4.8, for different values of known positions φ of the support region. Note that while this percentage generally increases with the number of available measurements, as expected, it also increases with φ. In fact, as we increase by 4 the number of known

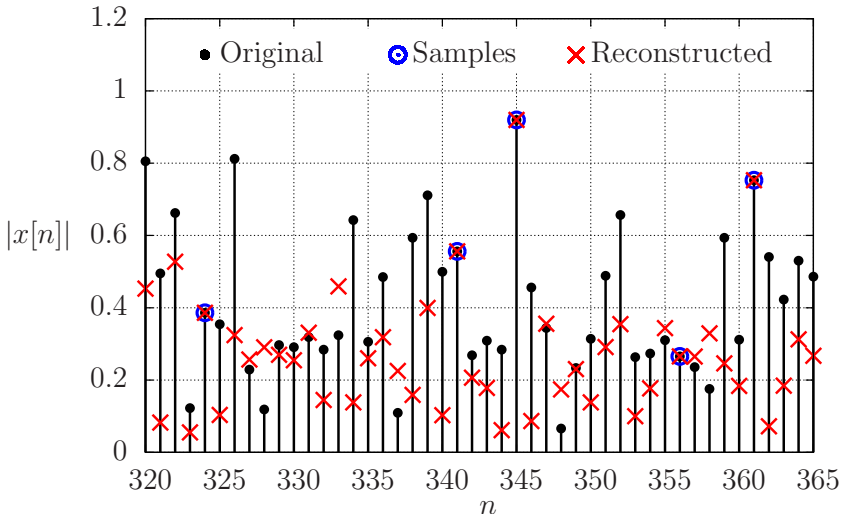
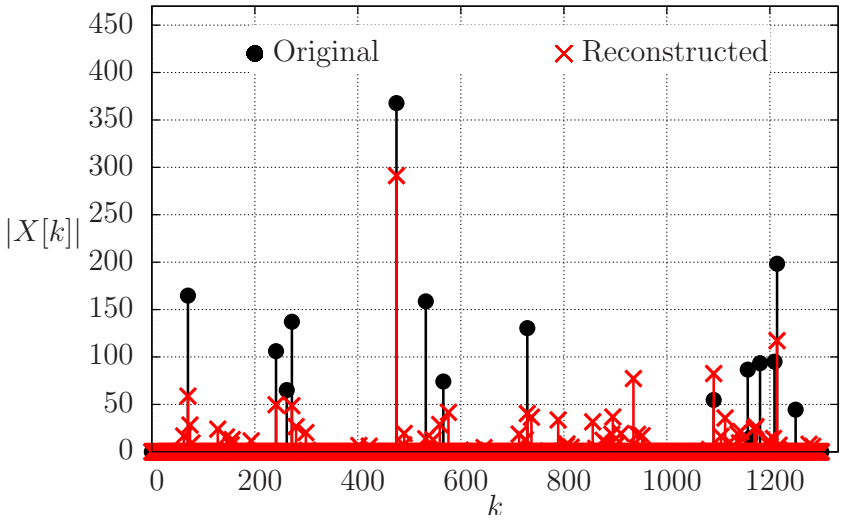
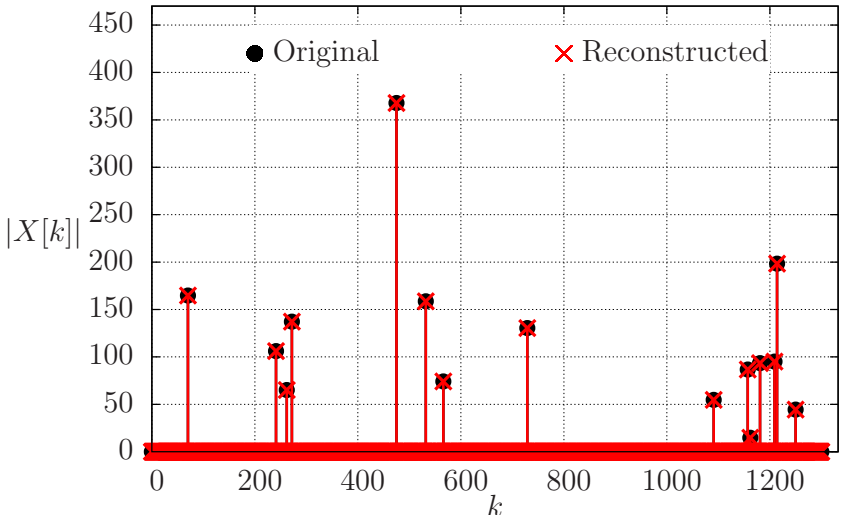


Figure 4.6. Magnitudes of two signals of length 1301 reconstructed from the same 40 samples of a signal with 15-sparse DFT (only 45 samples of each reconstructed signal are shown, for clarity): (a) incorrect reconstruction without using DFT's support information; (b) correct reconstruction using information on the positions of 10 non-null DFT coefficients.

positions of the support, the resulting curve for the percentage of correct reconstructions is shifted to the left, showing that, for the same fixed percentage, the number of required linear measurements is reduced.



(a)



(b)

Figure 4.7. Magnitudes of the DFTs of the reconstructed signals in Figure 4.6 (a) and Figure 4.6 (b), which correspond to the cases without and with prior information, respectively. Only the second DFT (b), based on $\varphi = 10$ known support locations, matches the original signal.

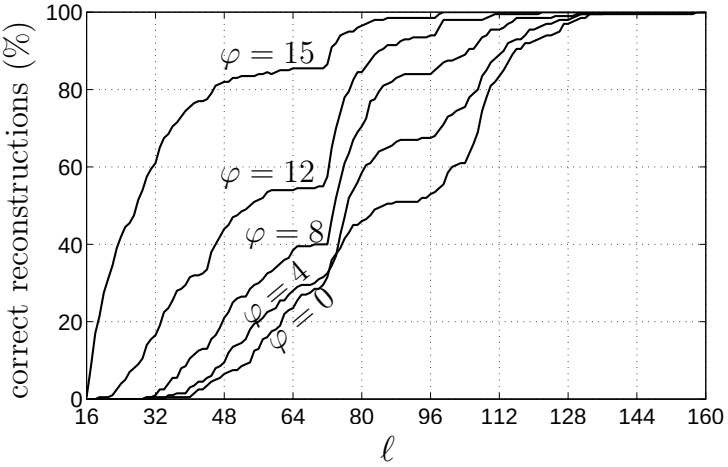


Figure 4.8. Percentages of correct reconstructions (by ℓ_1-minimization in the Fourier domain) of 211-dimensional signals with 16-sparse DFTs, as a function of the number of time samples taken and for different values of known support locations φ.

4.4.3 Simulations Using General-Transform-Sparse Signals

We conducted a second set of simulations, similar to that of Section 4.4.2, with other 1000 randomly generated signals. This time, however, the idea was to evaluate the efficiency of applying (4.12) to reconstruct signals which are not sparse in the DFT domain, but rather in a general transformed one. For that, we defined a random 211×211 matrix, which was orthogonalized and normalized using Gram-Schmidt's technique. This matrix thus represented an arbitrary, orthonormal transformation $\mathbf{T}$ used to provide a more general sparse representation.

The generation of the signals for the simulations was then conducted analogously to the case of Section 4.4.2. For each signal, 16 random positions were determined using a generator with uniform distribution; next the values of the sparse coefficients for those positions were determined using a Gaussian generator, defining the transformed sparse signal $\hat{\mathbf{x}}$; finally, the signal was computed by using the inverse transform, $\mathbf{x} = \mathbf{T}^{-1}\hat{\mathbf{x}}$.

Also note that since the considered signals are not sparse in the DFT domain, but rather on the domain defined by the transformation $\mathbf{T}$, random linear measurements instead of individual samples were taken in time domain, according to (4.1). We then performed the evaluation tests by solving (4.12) for each signal and for different numbers of measurements (ℓ) and of known positions (φ) in the support of the transformed domain. Figure 4.9 shows the percentage of correct signal reconstructions (with maximum relative error of $\tau = 10^{-8}$) as a function of ℓ and for the considered values of φ.

The simulations were repeated for a class of 1000 signals with length 256. The purpose was to illustrate that the reduction on the required number of measurements when using prior information occurs also in this case, even though Theorem 2 assumes the length to be a prime number. In fact, Fig. 4.10 shows a result that is consistent with that of Fig. 4.9. We emphasize here that, while Theorem 2 assumes a prime length when establishing the lower bound for the number of measurements, the ℓ_1-minimization approach for reconstruction, which anyway requires a larger number of measurements, works for any signal length, provided that the number of measurements is high enough for that particular length.

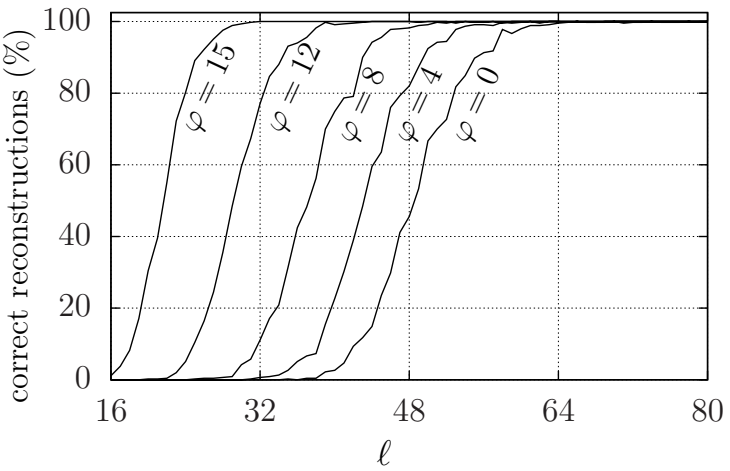


Figure 4.9. Percentages of correct reconstructions (by ℓ_1-minimization in the transformed domain) of 211-dimensional signals with a 16-sparse randomly defined orthogonal transform, as a function of the number of linear measurements taken and for different values of known support locations φ.

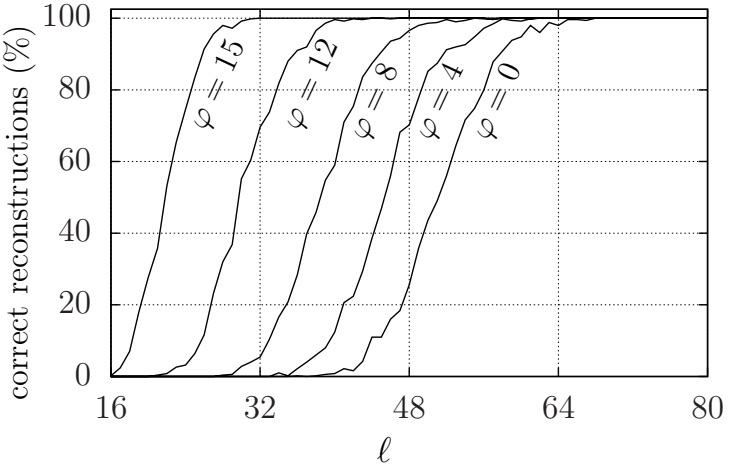


Figure 4.10. Percentages of correct reconstructions (by ℓ_1-minimization in the transformed domain) of 256-dimensional signals with a 16-sparse randomly defined orthogonal transform, as a function of the number of linear measurements taken and for different values of known support locations φ.

Two important observations can be made based on Fig. 4.9. First, the use of information about positions in the support of the sparse domain does allow a reduction on the required number of linear measurements, for a fixed value of correct reconstructions. In fact, all the curves of percentage reconstructions are shifted to the left as the value of φ is increased; also,

the difference between two distinct curves, for a fixed percentage of correct reconstructions, is frequently higher than the theoretical value φ since more than the minimum number of measurements is used, due to the use of a technique based on ℓ_1-minimization. The second observation is that for a given, fixed percentage of correct reconstructions, and for a fixed value of φ, the number of required measurements for Figs. 4.9 and 4.10 was lower than in Fig. 4.8. The reason for this is associated to the use of individual samples in the case of Fig. 4.8 and random measurements computed from *all* the available original samples in the cases of Figs. 4.9 and 4.10. This results in different levels of stability for the used numerical optimization procedures.

4.5 Final Remarks

In this chapter, we provided an extension of the basic compressive sensing framework. Specifically, we showed that the number of linear measurements needed for unambiguous reconstruction of a signal with a sparse representation is reduced if at least some of the positions of non-null components in this representation are known (this extension does not assume any information about the values of the components). This result was approached first from a theoretical point of view, and for signals having a sparse discrete Fourier transform (DFT). We provided a theorem proving that if the positions of φ DFT coefficients are known, the minimum theoretical number of required linear measurements is given by $\ell = 2\eta - \varphi$, thus reduced by φ with respect to the initial compressive sensing approach.

The use of prior information on the support of the sparse domain was shown to be easily incorporated to the reconstruction procedure by ℓ_1-minimization. We established the conditions for a signal having a sparse representation to be reconstructible by ℓ_1-minimization when prior information is available, and compared these conditions with the case with no prior information. Numerical simulations were then conducted based on this procedure, thus allowing to observe the reduction in the number of linear measurements required for unambiguous reconstruction as we increased the available information on the support of the sparse representation.

The use of known spectral characteristics of the input signal showed advantages from practical and theoretical aspects. First, it provided a reduction in the number of required samples from the nonsparse domain in order to still guarantee unambiguous reconstruction. Second, it provided a link between two different frameworks in digital signal processing. In one extreme lies the initial compressive sensing approach, using no prior information at all on the sparse domain and reconstructing it solely from the samples taken in the other domain. In the other extreme we have the fractional Fourier transform, when all the positions in the support of the sparse domain are known and the signal can be reconstructed from just the same number of samples in the nonsparse domain as these positions.

The first simulations presented in this chapter also suggest that a signal with some sparse representation can be reconstructed from fewer random linear measurements than individual time samples, which also require sparsity to happen in the DFT domain. This result was consistent in all simulated cases, namely when prior information on the support was not used in the reconstruction or when different numbers of known positions in this support were used. Naturally, if a discrete-time signal is obtained by traditional sampling and the purpose of the compressive sensing technique is just to reduce the required rate, it would not be viable to take random linear measurements, for they would require all samples to be initially available. Otherwise, if the purpose is just to have the minimum number of linear measurements necessary to reconstruct the signal by ℓ_1-minimization, our results suggest that random measurements should be used instead of isolated time samples.

5

IRLS Method for Compressive Sensing with Prior Information

In Chapter 4, we introduced the concept of prior information about a signal's sparse representation, in compressive sensing. We proved that this type of information reduces the theoretical minimum number of measurements required for reconstruction. In the more practical context of ℓ_1-minimization, we also proved that it improves the lower bound for the probability of reconstruction by convex optimization, and we provided a closed form expression for the probability, under a set of signals' statistical properties.

In this chapter, we provide an efficient, practical algorithm for signal reconstruction using prior information. This algorithm is based on iteratively reweighted least-squares, and defines a special weighting strategy that takes into account the known locations of nonzero components, in the sparse representation. We introduced this approach in [65], and now present it in more detail, with additional experimental results.

Iteratively reweighted least-squares (IRLS) algorithms have been successfully used in compressive sensing to reconstruct sparse signals from incomplete linear measurements taken in nonsparse domains. The underlying optimization problem corresponds to finding the

vector that solves the ℓ_p-minimization while explaining the measurements, and IRLS allows to easily control the used value of p, with effect on the number of required measurements. In this chapter, we propose a weighting strategy in the reconstruction method based on IRLS in order to add prior information on the support of the sparse domain.

Our simulation results show that combining the ℓ_p-minimization approach with the use of prior information in the sparse domain leads to a reduction in the number of linear measurements required for unambiguous reconstruction. This reduction occurs for all values of p, so that a further reduction can be achieved by decreasing p and using prior information (the reductions caused by the two approaches are cumulative). The proposed weighting scheme also reduces the computational complexity with respect to the IRLS with no prior information, both in terms of number of iterations and computation time.

We start by stating the reconstruction problem using prior information, to restrict the class of signals that can be reconstructed and thus allowing fewer linear measurements to represent a signal without generating ambiguity. Next, we show how the reconstruction problem can be tackled using the IRLS and how the prior information can be added to this method through an appropriate weighting scheme. Finally, we show that the resulting algorithm leads to a reduction in the number of required measurements to attain reconstruction, as well as a reduction in the total number of iterations and computation time.

The remaining of the chapter is organized as follows. Section 5.1 presents the idea of adding prior information on the support of the sparse representation to the reconstruction procedure. Following, Section 5.2 presents the problem formulation using IRLS, and specifically proposes a weighting strategy for adding prior information. The corresponding reconstruction procedure is described, leading to the algorithm proposed in Section 5.3. Section 5.4 presents the first experiments and the corresponding numerical results we obtained by applying this algorithm to discrete-time signals having a sparse representation in an arbitrary, random domain. Finally, Section 5.5 presents the chapter's final remarks.

Later on, Chapter 8 further investigates additional important aspects of the proposed algorithm. For instance, it evaluates the obtained signal-to-noise ratios (SNRs) when recon-

structing stochastic signals using different numbers of known support locations. This shows that when prior information increases, the algorithm results in higher SNRs, for the same number of measurements taken. Also, Chapter 8 evaluates the algorithm's robustness to changes in the parameter τ, which we add to the IRLS procedure in order to control how the prior information is weighted in the proposed method. Finally, it evaluates the robustness to wrong prior information, by measuring the effects of adding to the set of prior locations some of those which do not belong to the support. For further details and the results of these analyzes, refer to Sections 8.1.1 and 8.1.2.

5.1 Prior Information and ℓ_p-Minimization

In Chapter 3 and Chapter 4, we described how compressive sensing allows the efficient representation of discrete-time signals that are sparse in some known domain. The main characteristic common to these signals is that their projections on the basis functions of the sparse domain are mostly zero. This distinguishing property ultimately means that a signal of length N can be unambiguously represented by $\eta < N$ values by taking an appropriate transformation and coding both the nonzero coefficients and their positions.

An important discovery in compressive sensing was that random linear measurements are generally valid for the reconstruction of signals that are sparse in some domain [12], [5]. In this context, let $\mathbf{x}$ represent an N-dimensional signal with an η-sparse transform $\hat{\mathbf{x}}$ ($\eta < N$ coefficients of $\hat{\mathbf{x}}$ are nonzero) and $\mathbf{T}$ represent the transformation matrix, meaning

$$\hat{\mathbf{x}} = \mathbf{T}\mathbf{x}.$$

Then, if $\mathbf{M}_{\ell \times N}$, with $\ell < N$, is a random matrix with normal iid entries, then the linear measurements defined by

$$\mathbf{b} = \mathbf{M}\mathbf{x} \tag{5.1}$$

allow the determination of all the N components of $\mathbf{x}$ provided that the amount ℓ of measurements taken is high enough compared to the sparsity of $\hat{\mathbf{x}}$ [12].

Although the ℓ linear measurements defined by (5.1) unambiguously represent the original N samples of $\mathbf{x}$, a problem still remains on how to compute these samples from the available measurements. This problem is commonly referred to as signal reconstruction; it is approached by finding the sparsest vector $\hat{\mathbf{x}}$ such that its inverse transform $\mathbf{T}^{-1}$ generates the same given measurements. The key idea is to indirectly solve the minimization problem

$$\min_{\hat{\mathbf{x}}} \|\hat{\mathbf{x}}\|_0, \quad \text{subject to } \mathbf{M}\mathbf{T}^{-1}\hat{\mathbf{x}} = \mathbf{b}, \tag{5.2}$$

where $\|\hat{\mathbf{x}}\|_0$ corresponds to the number of nonzero components of the vector $\hat{\mathbf{x}}$.

Note that a direct approach to (5.2) leads to combinatorial complexity, which is not viable even for moderately sized signals. An alternative approach that is largely used is the ℓ_1-minimization of $\hat{\mathbf{x}}$, instead of the minimization of the objective function $\|\hat{\mathbf{x}}\|_0$ previously defined, so the reconstruction problem in this case is

$$\min_{\hat{\mathbf{x}}} \|\hat{\mathbf{x}}\|_1, \quad \text{subject to } \mathbf{A}\hat{\mathbf{x}} = \mathbf{b}, \tag{5.3}$$

where $\mathbf{A} = \mathbf{M}\mathbf{T}^{-1}$.

The solution to (5.3) leads to polynomial complexity and it is possibly the most common approach to signal reconstruction in compressive sensing. A second approach is the ℓ_p-minimization of $\hat{\mathbf{x}}$, with $0 < p < 1$, or, equivalently,

$$\min_{\hat{\mathbf{x}}} \frac{1}{2}\|\hat{\mathbf{x}}\|_p^p, \quad \text{subject to } \mathbf{A}\hat{\mathbf{x}} = \mathbf{b}, \tag{5.4}$$

and it has been shown that, by reducing the value of p, it is possible to reduce the number of required linear measurements ℓ with respect to that attained for $p = 1$ [18], [15]. In [77], the minimization of the p-norm-like diversity measures, for $p \leqslant 1$, as well as of the Gaussian and Shannon entropies, is used also to compute sparse solutions to underdetermined systems, in the problem of optimal basis selection; the approach includes the ℓ_p-minimization case when $p > 0$.

A possible approach to (5.4) is based on IRLS, as we will discuss in Section 5.2. An advantage of this method over interior point methods is that it allows the reduction of p in the ℓ_p-minimization in a straightforward manner, so that the same algorithm implemented

for a certain value of p can be used for a different one by changing a single parameter. A different approach is presented in [77], based on a factorization of the gradient of the Lagrangian function and on the successive relaxation of this function.

In this chapter, we propose a particular weighting strategy in the IRLS approach in order to add prior information on the support region of the sparse representation. As we showed in Chapter 4 as well as in [92] and [67], if such information is available it is possible to reduce the amount of taken measurements and still unambiguously reconstruct the underlying signal. In fact, let Φ be the subset of positions in $\{1, 2, \ldots, N\}$ which are known to belong to the support region of $\hat{\mathbf{x}}$, meaning

$$\hat{x}_k \neq 0 \; \forall \, k \in \Phi. \tag{5.5}$$

In Chapter 4 we showed that the information represented by (5.5) can be added to the reconstruction procedure based on the ℓ_1-minimization by solving, instead of (5.3),

$$\min_{\hat{\mathbf{x}}} \sum_{\substack{k=1 \\ k \notin \Phi}}^{N} |\hat{x}_k|, \quad \text{subject to } \mathbf{A}\hat{\mathbf{x}} = \mathbf{b}. \tag{5.6}$$

In the first experimentations in Section 4.4, we solved problem (5.6) using an interior point method approach. The corresponding results show that, for a pre-specified frequency of correct reconstructions, (5.6) leads to a reduction in the number of required measurements, with respect to the case without prior information.

Note that the reason why (5.3) can be replaced by (5.6) when the prior information (5.5) is available comes from the fact that the minimization of the ℓ_1 in (5.3) is actually aimed at finding the sparsest solution that explains the measurements: the ℓ_1 is minimized in order to find the vector $\hat{\mathbf{x}}$ with most null components that satisfies $\mathbf{A}\hat{\mathbf{x}} = \mathbf{b}$. If the positions in Φ are known to contain nonzero components of $\hat{\mathbf{x}}$, then, during the search for a sparse solution to $\mathbf{A}\hat{\mathbf{x}} = \mathbf{b}$, the alternative is to minimize the number of nonzeros in the other positions only (those which do not belong to Φ). The improvement of (5.6) over (5.3) is then related to trying to minimize the number of nonzeros only outside the region where $\hat{\mathbf{x}}$ is already

supposed to be nonzero; hence, (5.6) gives preference to a solution with more zeros outside the specified set Φ.

This chapter shows that the prior information represented by (5.5) can also be added to the IRLS approach to the reconstruction. The corresponding minimization problem, with the prior information, is reformulated as

$$\min_{\hat{\mathbf{x}}} \frac{1}{2} \sum_{\substack{k=1 \\ k \notin \Phi}}^{N} |\hat{x}_k|^p, \quad \text{subject to } \mathbf{A}\hat{\mathbf{x}} = \mathbf{b}. \tag{5.7}$$

Our results show that the prior-information weighting scheme, which allows the reconstruction based on (5.7), leads to a reduction in the number of required measurements with respect to (5.4). This occurs for all used values of p, so a further reduction with respect to (5.3) and (5.4) can be attained by simultaneously using prior information and reducing p.

An important characteristic of (5.7) is that a solution $\hat{\mathbf{x}}$ is not explicitly constrained to be nonzero in the locations specified by Φ; rather, the corresponding values are determined from the equality constraint and the minimization of the objective function associated to the remaining positions. This is specially important in the cases in which the prior information is not perfectly reliable, so that some positions in Φ can actually not belong to the support (in this case, the reconstruction procedure should allow the computation of null elements inside Φ). By solving (5.7), we can still reconstruct the underlying signals, but more measurements may be required compared to the case when no wrong locations are present. In fact the $\hat{\mathbf{x}}$ components in the positions Φ are removed in (5.7) from the minimization function, so if some zero components' locations are mistakenly attributed to Φ the local sparsity of those components is not exploited during the reconstruction (we emphasize that the possibility of reconstructing signals from limited measurements in compressive sensing is based on exploiting the sparsity). Even in this case, however, we observed that if most of the components of Φ belong to the support, our proposed method provides an improvement in the reconstruction, in terms of computational cost and of number of required measurements. More details on this are presented in Chapter 8.

5.2 Problem Formulation

The IRLS method for reconstructing sparse signals from linear measurements is based on iteratively solving (5.4) with a modified objective function, that at each iteration function approaches $\sum_{k=1}^{N}|\hat{x}_k|^p$. More specifically, consider the optimization problem

$$\min_{\hat{\mathbf{x}}} \frac{1}{2}\sum_{k=1}^{N} w_k^{p-2}\hat{x}_k^2, \quad \text{subject to } \mathbf{A}\hat{\mathbf{x}} = \mathbf{b}, \tag{5.8}$$

where w_k is a weighting parameter [18]. Note that (5.8) can be solved in just one iteration, as we will describe, but if the problem is repeatedly solved changing the values of w_k at each time, so that w_k approaches $\hat{x}_k$, the objective function in (5.8) will approach that of (5.7). In fact, let

$$\mathbf{w}^{(m)} = |\hat{\mathbf{x}}^{(m-1)}|,$$

where $\mathbf{w}^{(m)} = [w_1^{(m)} \;\; w_2^{(m)} \;\; \dots \;\; w_N^{(m)}]^T$ is the value of the weighting vector to be used in the m-th iteration and $\hat{\mathbf{x}}^{(m-1)}$ is the $(m-1)$-th iterate. After convergence, $\hat{\mathbf{x}}^{(m-1)}$ will be sufficiently close to $\hat{\mathbf{x}}^{(m)}$, according to a specified tolerance, so

$$\sum_{k=1}^{N}(w_k^{(m)})^{p-2}(\hat{x}_k^{(m)})^2 = \sum_{k=1}^{N}|\hat{x}_k^{(m-1)}|^{p-2}(\hat{x}_k^{(m)})^2$$

will be close to $\sum_{k=1}^{N}|\hat{x}_k^{(m)}|^p$, which is the original objective function.

Now, in order to add the prior information on the positions of nonzero coefficients in the sparse domain, we must do the ℓ_p-minimization over the vector components in the remaining positions only. Also, since the sparse signal must still match all the linear measurements, the equality constraint is the same, and the new minimization problem is

$$\min_{\hat{\mathbf{x}}} \frac{1}{2} \sum_{\substack{k=1 \\ k \notin \Phi}}^{N} |\hat{x}_k|^p, \quad \text{subject to } \mathbf{A}\hat{\mathbf{x}} = \mathbf{b}. \tag{5.9}$$

Now, a local solution to (5.9) can be obtained by iteratively solving

$$\min_{\hat{\mathbf{x}}} \frac{1}{2} \sum_{\substack{k=1 \\ k \notin \Phi}}^{N} w_k^{p-2}\hat{x}_k^2, \quad \text{subject to } \mathbf{A}\hat{\mathbf{x}} = \mathbf{b} \tag{5.10}$$

and changing w_k at each iteration so that $w_k^{p-2}\hat{x}_k^2$ is sufficiently close to $|\hat{x}_k|^p \; \forall \, k \notin \Phi$ after convergence.

By making

$$w_k^{p-2} \approx 0, \;\; \forall \; k \in \Phi, \tag{5.11}$$

(5.10) is posed in the form of (5.8), which has a closed-form solution as indicated below. Since w_k must approach $\hat{x}_k$ for $k \notin \Phi$, we then define

$$w_k^{(m)} = \begin{cases} |\hat{x}_k^{(m-1)}|, & \text{if } k \notin \Phi \\ \tau |\hat{x}_k^{(m-1)}|, & \text{otherwise,} \end{cases} \tag{5.12}$$

where τ is a constant chosen such that τ^{p-2} is small compared to the signal's coefficients, so that $\tau^{p-2}|\hat{x}_k^{(m-1)}|$ is close to zero. In the first simulations evaluating our proposed IRLS method, discussed in Section 5.4, we used $\tau^{p-2} = 10^{-3}$; a discussion of the algorithm's behavior with respect to τ^{p-2} appears in Chapter 8.

A signal $\mathbf{x}$ with sparse representation $\hat{\mathbf{x}}$ can then be reconstructed from a sufficient amount of linear measurements $\mathbf{b}$ by solving for $m = 1, 2, \ldots$ and until convergence

$$\min_{\hat{\mathbf{x}}} \frac{1}{2} \sum_{k=1}^{N} |w_k{}^{(m)}|^{p-2} \hat{x}_k^2, \quad \text{subject to } \mathbf{A}\hat{\mathbf{x}} = \mathbf{b}, \tag{5.13}$$

where $w_k^{(m)}$ is given by (5.12). Our approach to (5.13) using (5.12) leads to the IRLS with prior information. Note that (5.13) corresponds to the minimization of a quadratic form with a linear equality constraint, which leads to the solution

$$\hat{\mathbf{x}}^{(m)} = \mathbf{Q}^{(m)} \mathbf{A}^T \left(\mathbf{A} \mathbf{Q}^{(m)} \mathbf{A}^T \right)^{-1} \mathbf{b}, \tag{5.14}$$

where

$$\mathbf{Q} = \text{diag}(q_1, q_2, \ldots, q_N) \tag{5.15}$$

with

$$q_k = \begin{cases} |\hat{x}_k^{(m-1)}|^{2-p}, & \text{if } k \notin \Phi, \\ \tau^{p-2} |\hat{x}_k^{(m-1)}|^{2-p}, & \text{otherwise.} \end{cases} \tag{5.16}$$

In summary, the sparse representation $\hat{\mathbf{x}}$ can be reconstructed from sufficient linear measurements $\mathbf{b}$ by solving (5.14) for $m = 1, 2, \ldots$, until convergence.

5.3 IRLS with Prior Information

We now summarize the basic steps needed to reconstruct sparse signals with the weighting strategy with prior information in the IRLS. Note that a regularization procedure is needed when defining the components of the main diagonal of $\mathbf{Q}^{(m)}$ according to (5.16). Because of the matrix inversion in (5.14), we must guarantee that these components do not approach zero, so a constant μ is added to $|\hat{x}_k|$ when defining the weights. With this regularization, (5.16) becomes

$$q_k = \begin{cases} (|\hat{x}_k^{(m-1)}| + \mu)^{2-p} & \text{if } k \notin \Phi, \\ \tau^{p-2}(|\hat{x}_k^{(m-1)}| + \mu)^{2-p} & \text{otherwise,} \end{cases} \tag{5.17}$$

where we used $\tau^{p-2} = 10^{-3}$. Our numerical experimentations show the proposed method to be stable regarding the choice of values for τ and the resulting factor τ^{p-2}. In fact, different orders of magnitude of τ were tested, with a large range of values allowing equivalent results. More details on the behavior of the algorithm with respect to this parameter are presented in Chapter 8.

Then, the values of $\hat{\mathbf{x}}$ are updated according to (5.14), starting from an initial value $\hat{\mathbf{x}}^{(0)}$ and until the relative change between the norms of two consecutive iterates is below a specified tolerance. After this, the regularization parameter is then reduced and the iterations in (5.14) are again conducted until convergence. The process is repeated until μ becomes sufficiently small. In our simulations, the initial value $\hat{\mathbf{x}}^{(0)}$ is computed by finding the minimum 2-norm solution to the equality constraint (least-squares solution), so

$$\hat{\mathbf{x}}^{(0)} = \mathbf{Q}^{(0)}\mathbf{A}^H(\mathbf{A}\mathbf{Q}^{(0)}\mathbf{A}^H)^{-1}\mathbf{b}, \tag{5.18}$$

with the initial inverse weight matrix given by $\mathbf{Q}^{(0)} = \text{diag}(q_1^{(0)}, q_2^{(0)}, \ldots, q_N^{(0)})$ where

$$q_k^{(0)} = \begin{cases} 1, & \text{if } k \notin \Phi \\ \tau^{p-2}, & \text{otherwise.} \end{cases} \tag{5.19}$$

In defining the convergence criterion for each iteration stage, our results have shown that the iterative procedure with prior information given by (5.14) can follow the same strategy proposed in [18]. In this scheme, (5.14) is repeated, first with $\mu = 1$, until

$$\frac{||\hat{\mathbf{x}}^{(m)} - \hat{\mathbf{x}}^{(m-1)}||}{1 + ||\hat{\mathbf{x}}^{(m-1)}||} < \frac{\sqrt{\mu}}{100}. \tag{5.20}$$

After (5.20) is attained, μ is reduced by a factor of 10, and the iterative procedure is repeated until $\mu \leqslant 10^{-8}$ [18].

Algorithm 1 below summarizes the procedures for reconstructing a sparse transform $\hat{\mathbf{x}}$ from the linear measurements $\mathbf{b}$ with prior information on the support domain, by solving (5.7). Note that $\mathbf{A} = \mathbf{M}\mathbf{T}^{-1}$; also, the signal $\mathbf{x}$, from which the measurements are taken, can be reconstructed by taking the inverse transform $\mathbf{x} = \mathbf{T}^{-1}\hat{\mathbf{x}}$.

Algorithm 1 IRLS method for signal reconstruction with prior information

Inputs: $p \in (0, 1]$, $\mathbf{A} = \mathbf{MT}^{-1}$, $\mathbf{b}$, Φ, μ (initial regularization parameter), τ (weighting parameter related to prior information).

Output: Sparse representation ($\hat{\mathbf{x}}$) of the desired signal.

Step 1. Define the initial vector of inverse weights, $\mathbf{q}^{(0)}$, and the initial iterate, $\hat{\mathbf{x}}^{(0)}$, using

$$q_k^{(0)} = \begin{cases} 1, & \text{if } k \notin \Phi \\ \tau^{p-2}, & \text{otherwise} \end{cases}$$

and

$$\hat{\mathbf{x}}^{(0)} = \text{diag}(\mathbf{q}^{(0)})\mathbf{A}^H\mathbf{y},$$

where $\mathbf{y}$ is the solution to $\left[\mathbf{A}\text{diag}(\mathbf{q}^{(0)})\mathbf{A}^H\right]\mathbf{y} = \mathbf{b}$.

Step 2. Do the inner loop:

2.1 Initialize $m := 1$.

2.2 Update the vector of inverse weights, $\mathbf{q}^{(m)}$, using

$$q_k^{(m)} = \begin{cases} \left(|\hat{x}_k^{(m-1)}| + \mu\right)^{2-p}, & \text{if } k \notin \Phi \\ \tau^{p-2}\left(|\hat{x}_k^{(m-1)}| + \mu\right)^{2-p}, & \text{otherwise.} \end{cases}$$

2.3 Compute the next iterate, $\hat{\mathbf{x}}^{(m)}$, using

$$\hat{\mathbf{x}}^{(m)} = \text{diag}(\mathbf{q}^{(m)})\mathbf{A}^H\mathbf{y},$$

where $\mathbf{y}$ is the solution to $\left[\mathbf{A}\text{diag}(\mathbf{q}^{(m)})\mathbf{A}^H\right]\mathbf{y} = \mathbf{b}$.

2.4 If

$$\frac{\|\hat{\mathbf{x}}^{(m)} - \hat{\mathbf{x}}^{(m-1)}\|}{1 + \|\hat{\mathbf{x}}^{(m-1)}\|} < \frac{\sqrt{\mu}}{100},$$

go to Step 3; otherwise, let $m := m + 1$ and go to Step 2.2.

Step 3. Update the regularization parameter, $\mu := \mu/10$.

Step 4. If $\mu < 10^{-8}$, finish; else, go to Step 2.

5.4 Simulation Results

Algorithm 1 was first evaluated in different conditions using 500 test signals with length 256 (larger signals are also considered later). We used a procedure to guarantee that each of these signals, although randomly generated, is η-sparse in some specified domain. First, for generality, we defined a random orthogonal transformation matrix $\mathbf{T}$ using Givens decomposition [64]. Each signal was then generated first in the corresponding desired sparse domain; with this purpose, $\eta = 16$ nonzero values were determined using a Gaussian pseudo-random number generator, while their positions were assigned by a generator with uniform distribution. From the sparse vector $\hat{\mathbf{x}}$ thus obtained, the time-domain signal was finally computed by taking the inverse transform $\mathbf{x} = \mathbf{T}^{-1}\hat{\mathbf{x}}$.

Each test consisted on taking an specific number of linear measurements, ℓ, from each of the 500 signals, and applying IRLS method with prior information to reconstruct it using different amounts of known positions φ belonging to the support of the sparse domain. For each possible combination of ℓ and φ, we then evaluated if the signal was correctly reconstructed. In this classification, we considered as correct reconstructions only the cases for which the normalized energy of the error between the original signal and the solution to (5.9) was below a pre-specified tolerance of 10^{-3}.

In Figure 5.1 and Figure 5.2, we show the results using Algorithm 1 with $p = 1$, while in Figure 5.3 and Figure 5.4 we show the results for $p = 0.1$. The percentages of correct reconstructions appear in Figure 5.1 and Figure 5.3 as functions of ℓ and φ. While these percentages generally increase with the amount of available measurements, as expected, they also increase with φ. Indeed, as the number of known support positions increases by 4, the resulting curve for the percentage of correct reconstructions is shifted to the left, showing that for the same fixed percentage, the number of required measurements is reduced. This shows that the information represented by the φ positions is appropriately used by the IRLS method through the weighting scheme given by (5.12).

An important observation regarding the reductions in required measurements when using Algorithm 1 with prior information is that they occur for both values of p (1 and 0.1). This

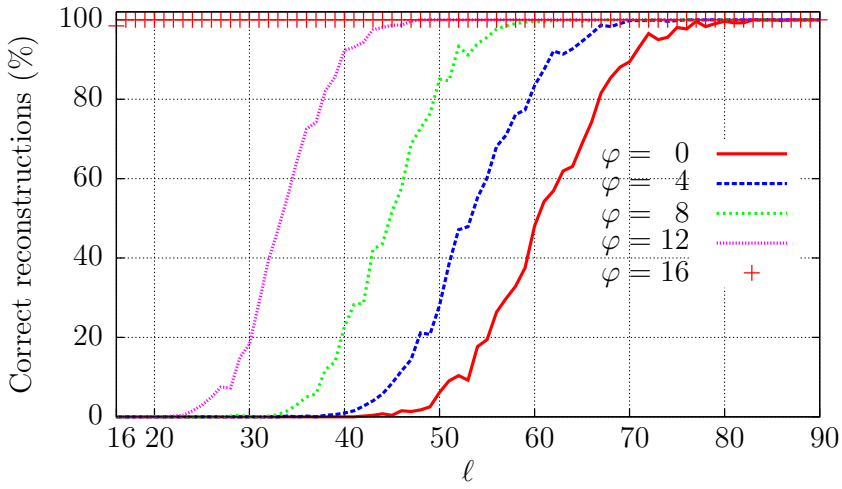


Figure 5.1. Percentages of correct reconstructions using Algorithm 1 with $p = 1$, as a function of the number of samples taken in the nonsparse domain and for different values of known positions (φ) in the support region of the sparse domain. All 500 test signals are of length $N = 256$ and sparsity $\eta = 16$ in an arbitrary, randomly determined transformed domain.

result is also verified for different values of p in the range $0 < p \leqslant 1$, as we will show. In fact, in Figure 5.1 and Figure 5.3, the percentage curves are shifted by the same amount to the left for increasing values of φ. This indicates that, with respect to IRLS with no prior information, a more significant reduction in the amounts of measurements can be attained by reducing p and at the same time using prior information.

We emphasize that the evaluated amounts of correct reconstructions are based on the defined criterion of error (normalized energy of the difference below 10^{-3}), and that the amounts of measurements taken correspond to this criterion. A less strict criterion shifts the curves in Figure 5.1 and Figure 5.3 to the left, meaning less measurements being taken, but the distances between the curves corresponding to different values of φ are preserved. Furthermore, without prior information ($\varphi = 0$) the required number of measurements to attain reconstruction in our simulations matches the results we achieve using the algorithm in [18] for the same reconstruction error.

We have also observed that, by using prior information, Algorithm 1 allows a reduction in both the number of iterations and the total time to convergence. These aspects are

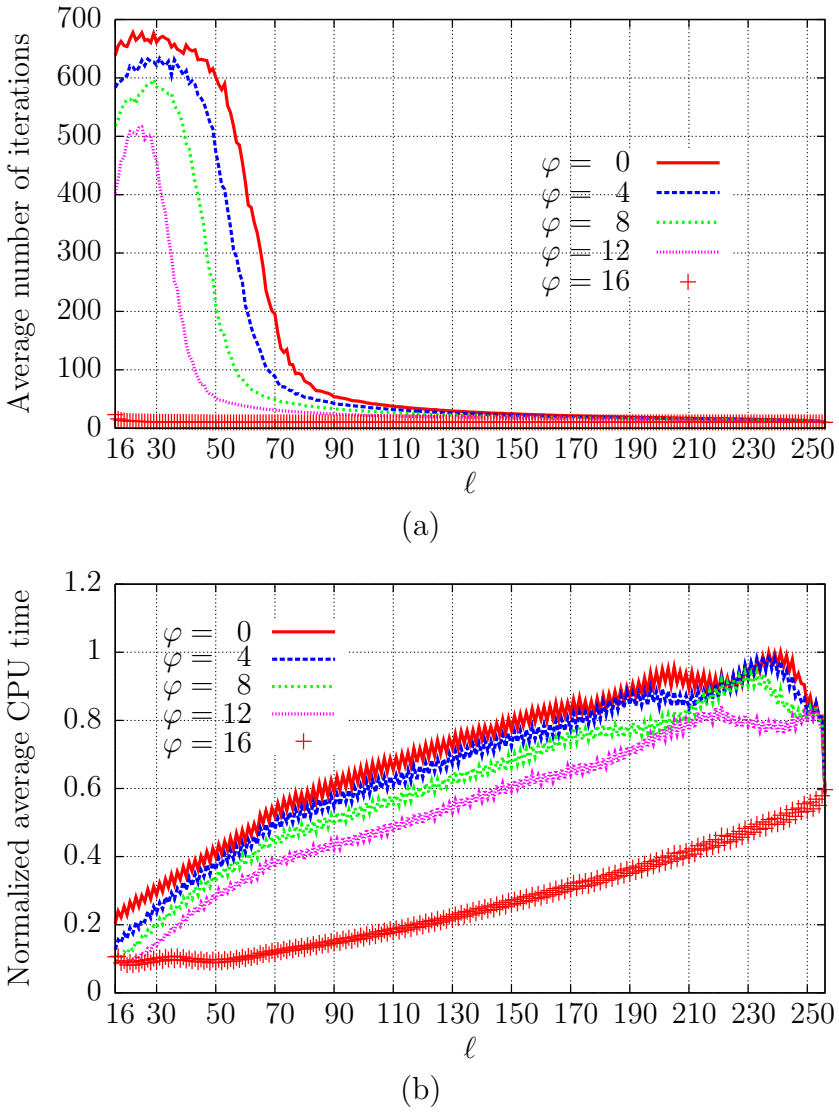


Figure 5.2. Numerical results obtained using Algorithm 1 with $p = 1$, as functions of the number of samples taken in the nonsparse domain and for different values of known positions (φ) in the support region of the sparse domain: (a) Average numbers of iterations; (b) Average times to convergence. All 500 test signals are of length $N = 256$ and sparsity $\eta = 16$ in an arbitrary, randomly determined transformed domain.

illustrated, respectively, in parts (a) and (b) of Figure 5.2 and Figure 5.4. The number of iterations is counted as the number of times (5.14) is executed. Note that, as φ is increased, the curves corresponding to both the number of iterations and the convergence time are shifted to the bottom, meaning a consistent reduction in those quantities for all values of ℓ.

Regarding the reductions in the number of required measurements, the amounts of itera-

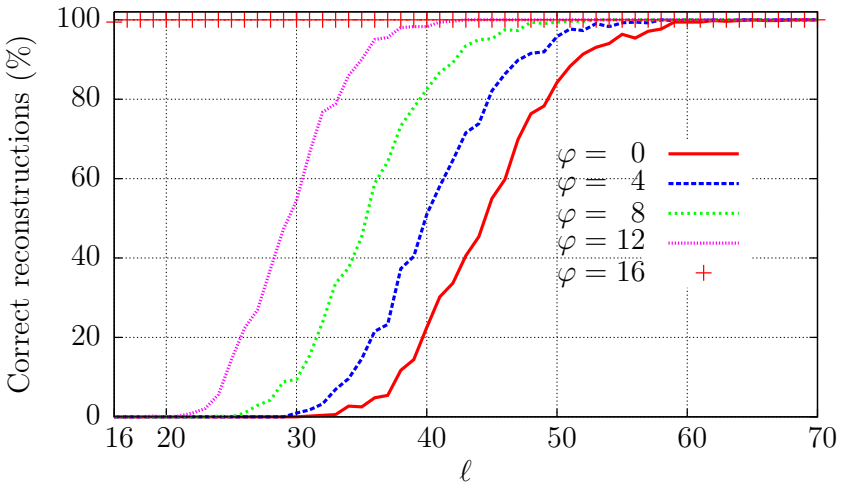


Figure 5.3. Percentages of correct reconstructions when using Algorithm 1 with $p = 0.1$, as a function of the number of samples taken in the nonsparse domain and for different values of known positions (φ) in the support region of the sparse domain. All 500 test signals are of length $N = 256$ and sparsity $\eta = 16$ in an arbitrary, randomly determined transformed domain.

tions, and the computation time when using Algorithm 1 with prior information, we observed that they occur for all tested values of p in the objective function (30 values of p in the range $0 < p \leqslant 1$ were analyzed). In Figure 5.5 and in Figure 5.6, we describe the results for all the tested values of p in the range $0 < p \leqslant 1$ and for a fixed number of linear measurements, $\ell = 2.5\eta = 160$; these results correspond to applying Algorithm 1 to 500 different signals with length $N = 1024$ and sparsity $\eta = 64$. In Figure 5.5, we observe that, independently on the used value of p, a higher value of known positions φ leads to an increase in the percentage of correct reconstructions from the same amount of measurements ℓ, even when a lower value of p already provides an increase with respect to $p = 1$ (the advantage of using the prior information does not vanish when p is decreased to improve reconstruction). Also, Figure 5.6(a) and Figure 5.6(b) show that as φ increases, both the number of iterations and the time to convergence decrease, for all tested values of p. Note in Figure 5.6((b)) that decreasing p has the effect of increasing the time required to reconstruct the signals from the same amount of measurements, but the used prior information reduces both the number of iterations and the convergence time.

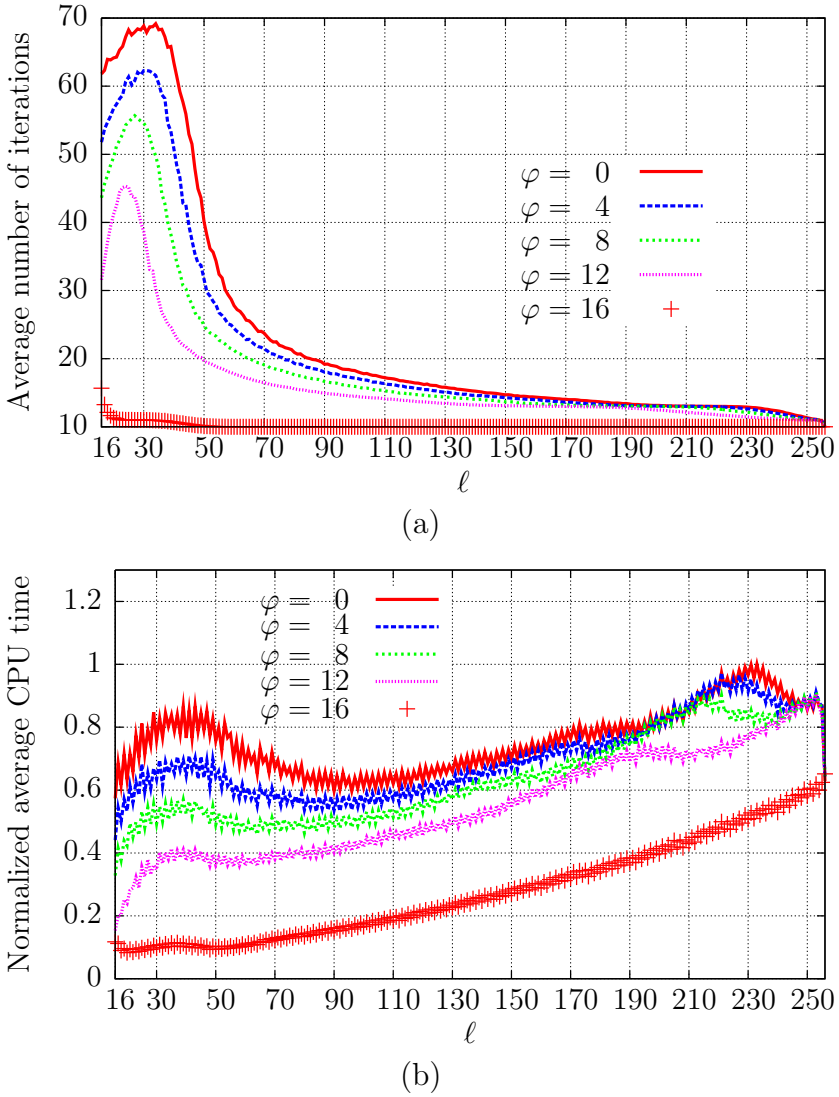


Figure 5.4. Numerical results obtained using Algorithm 1 with $p = 0.1$, as functions of the number of samples taken in the nonsparse domain and for different values of known positions (φ) in the support region of the sparse domain: (a) Average numbers of iterations; (b) Average times to convergence. All 500 test signals are of length $N = 256$ and sparsity $\eta = 16$ in an arbitrary, randomly determined transformed domain.

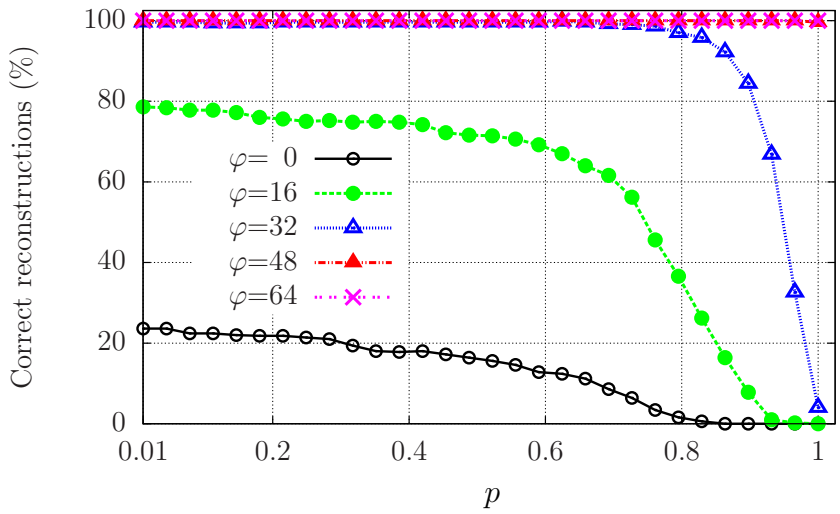
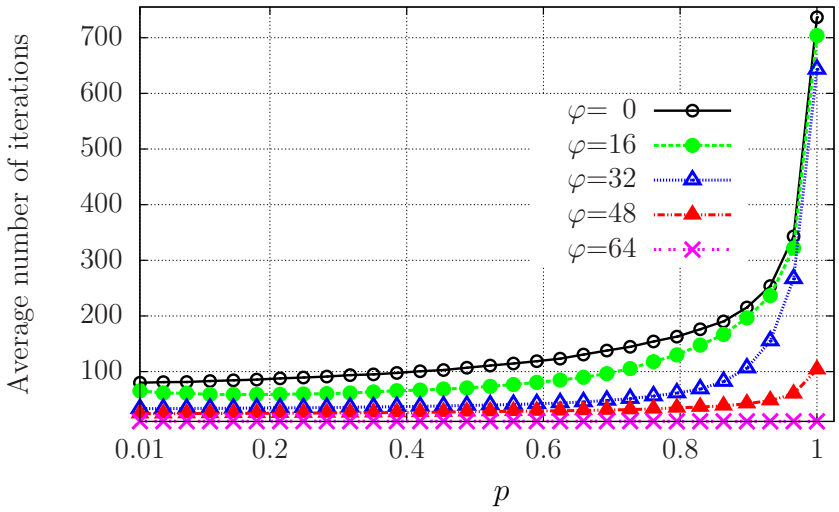
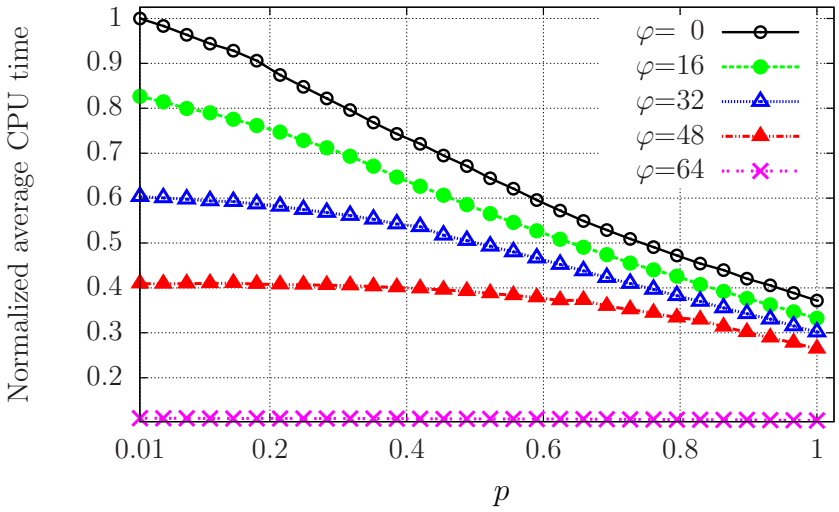


Figure 5.5. Percentages of correct reconstructions when using Algorithm 1, as a function of p and for different values of known positions (φ) in the support region of the sparse representation. All 500 test signals are of length $N = 1024$ and sparsity $\eta = 64$ in an arbitrary, randomly determined transformed domain. In all cases, reconstruction is based on $\ell = 2.5\eta = 160$ measurements.



(a)



(b)

Figure 5.6. Numerical results obtained using Algorithm 1, as functions of p and for different values of known positions (φ) in the support region of the sparse domain: (a) Average numbers of iterations; (b) Average times to convergence. All 500 test signals are of length $N = 1024$ and sparsity $\eta = 64$ in an arbitrary, randomly determined transformed domain. In all cases, reconstruction is based on $\ell = 2.5\eta = 160$ measurements.

5.5 Final Remarks

This chapter proposes a signal reconstruction scheme based on the iteratively reweighted least-squares (IRLS) method for compressive sensing with prior information. The proposed method, related to the definition of the weights matrix used at each iteration, allows the efficient use of information on the support underlying signal's sparse representation.

The first simulation results show that using prior information leads to a reduction in the number of linear measurements required to attain a pre-specified percentage of correct reconstructions. This reduction is directly related to the number of known positions. Furthermore, we also verified a reduction in the number of iterations and computation time when adding prior information to the IRLS. This result was consistent, independently on the number of linear measurements used for reconstruction.

An important observation regarding the IRLS with prior information is that the reduction in the magnitudes of both the number of required measurements and the computational cost when using prior information occurs for all tested values of p in the ℓ_p-minimization. Hence, a further reduction can be attained by reducing p while at the same time using prior information.

In Chapter 8, on the other hand, we show that if the prior information is not perfectly reliable, meaning that some components of the support are misplaced, there is still an improvement in performance with respect to no prior information or to less correct positions being available. The final performance depends on the difference between the numbers or correct and wrong locations.

Also see Chapter 8 for the experiments that evaluate additional aspects of the algorithm, such as its robustness to changes in the new optimization parameters and how the use of prior information in the IRLS approach affects the final SNRs for different numbers of known support information.

6

Method for Improved Reconstruction of Gradient-Sparse MR Images

In this chapter, we propose a compressive sensing method for reconstructing gradient-sparse magnetic resonance (MR) images based on pre-filtering the input signals in the k-space domain. A set of filtered versions of the image is reconstructed using the available k-space samples, and a final reconstruction stage generates the desired image from the filtered versions. Our experiments, conducted over real MR images and angiograms, show that the proposed method improves the reconstruction over the total-variation minimization (which is traditionally used to reconstruct similar types of images), in terms of signal-to-noise ratios and computation times. The proposed method is particularly appropriate for computing MR angiograms, which are typically sparse under the finite-differences operation.

Another advantage of the proposed method is that it allows the incorporation, in the optimization stages, of prior information about the sparse representation's support. In Chapter 4 and Chapter 5, we described in detail this type of prior information; we also proposed an efficient method for signal reconstruction that uses this information to improve reconstruction over traditional compressive sensing. In Chapter 7, on the other hand, we will describe how

the MR methods here introduced can be combined with the prior approach, with further improvement in terms of image quality and signal-to-noise-ratios.

As we described in Chapter 3, Chapter 4 and Chapter 5, compressive sensing allows a signal having a sparse representation in some known domain to be reconstructed using limited linear measurements taken from a nonsparse representation. The particular case of the reconstruction of compressible images from measurements in the frequency domain is of great interest to magnetic resonance (MR) imaging; in this application, the available information provided by the MR scanners is commonly represented by individual samples in the so-called k-space, which corresponds to the Fourier transform of the image to be reconstructed. In this context, we propose a compressive sensing method for reconstructing gradient-sparse MR images based on the preprocessing of the available k-space samples, followed by the reconstruction of filtered versions of the desired image and by a final composition stage that yields a nonfiltered version. Our experimental results show that, for the tested images, the analyzed method improves MR image reconstruction in terms of signal-to-noise ratio (SNR) as compared to the total-variation minimization approach commonly used in compressive sensing for gradient-sparse images.

In this chapter, we propose and evaluate a set of compressive sensing methods for the reconstruction of gradient-sparse MR images, and compare its performance to the reconstruction by TV minimization as described in [10] and [9].

It is worth mentioning that the total variation is relevant for the reconstruction even of images that are not gradient-sparse, in which case an objective function that combines the total variation with the ℓ_1 of the sparse representation can be defined. In most cases, the minimization of this mixed objective function can lead to improved reconstruction as compared to the ℓ_1 minimization alone [58]. Thus, implementing methods that under certain conditions can improve over the total-variation minimization and that can potentially be extended in the future to images that are not gradient-sparse can have an important impact in future MRI systems.

The remaining sections of the chapter are organized as follows. Section 6.1 briefly de-

scribes signal acquisition in MRI systems, and the resulting possible trajectories in the k-space domain. It also describes, in Section 6.1.1, the reconstruction, from k-space samples, of MR images that have a sparse representation in a known domain; the particular case of gradient-sparse images using total-variation is emphasized here. Next, Section 6.2 describes our proposed method, in three different tested schemes. Section 6.3 then describes our first experimental results, for different MR images, and in particular for MR angiograms. Finally, Section 6.4 presents the final chapter's remarks.

In Chapter 8, we discuss additional experiments that evaluate in more detail the methods here proposed. In particular, we consider a larger set of images, and discuss the compromise between reconstruction times and image quality.

6.1 Signal Acquisition in MRI and Compressive Sensing

In magnetic resonance (MR) imaging, the reconstruction problem corresponds to computing a single image or a sequence of images from the set of measurements provided by an appropriate MR scanner. A strong static magnetic field and a radio frequency magnetic field are applied to the body to be examined, the first polarizing its internal protons while the second allows for a magnetic moment that is proportional to some local property of interest [96]. In the desired output image, the luminance at each pixel must then be directly proportional to this property, at a region whose dimensions are inversely proportional to the attained resolution.

To conduct this reconstruction, it is then important to obtain a mathematical model for the measurements, based on the physical properties of the MR scanner, the applied magnetic fields, and the property of interest. In the case of static images the resulting signals collected at receiver coils in the MR scanner are samples of the bidimensional Fourier transform of the image to be reconstructed, and these samples belong to specific trajectories that depend on the type of magnetic pulses applied during the acquisition [57]. The Fourier coefficients

used in the MR reconstruction are usually denominated k-space samples.

Three types of trajectories commonly used in MRI are shown in Figure 6.1. The type of trajectory actually used in a particular MRI system is defined by physiological, time, and hardware constraints. In the case of reconstruction by compressive sensing techniques, the trajectory to be adopted should also take into account the type of transformation that makes the images sparse, as we will describe later.

6.1.1 Image Reconstruction Using Compressive Sensing

Since MR images usually have sparse representations in appropriate transformed domains, compressive sensing provides a viable approach for their reconstruction from k-space samples. Consider a $N_1 \times N_2$ image $\mathbf{X}$ represented in vectorized form by an N-dimensional vector $\mathbf{x}$ ($N = N_1 N_2$), known to have a sparse representation $\hat{\mathbf{x}} = \mathbf{T}\mathbf{x}$. In this case, x can be reconstructed from ℓ linear measurements taken in a nonsparse domain, which we represent by an ℓ-dimensional vector $\mathbf{b} = \mathbf{M}\mathbf{x}$. Note that the measurement matrix $\mathbf{M}$ must model both the bidimensional Fourier transformation in vectorized form (since the samples are taken in the DFT domain) and the selection of coefficients according to the used trajectory. In this

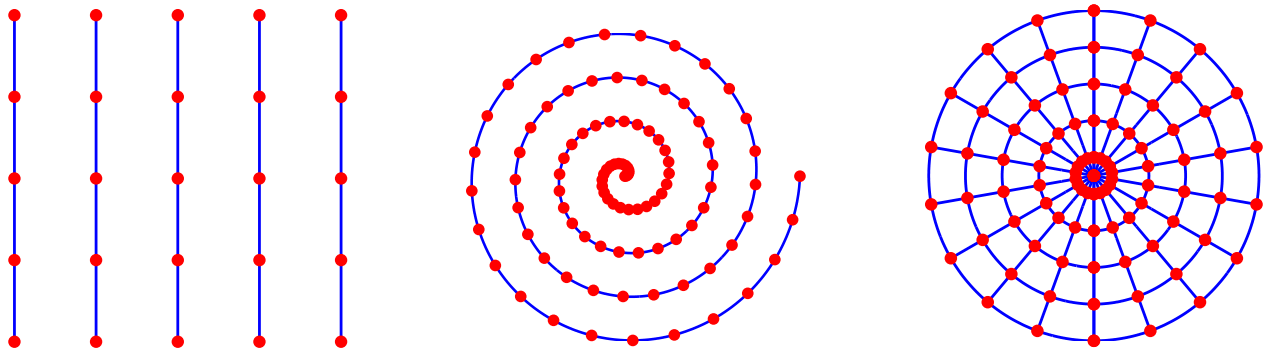


Figure 6.1. Three examples of common k-space trajectories in MRI (the red dots represent the measurements taken): (a) individual columns; (b) spiral trajectory; (c) radial lines.

context, it is possible to reconstruct $\mathbf{x}$ by solving the ℓ_p-minimization problem

$$\begin{aligned} &\min \|\hat{\mathbf{x}}\|_p^p \\ &\text{subject to } \mathbf{M}\mathbf{T}^{-1}\hat{\mathbf{x}} = \mathbf{b}, \end{aligned} \tag{6.1}$$

Substituting $\hat{\mathbf{x}}$ by $\mathbf{T}\mathbf{x}$, (6.1) becomes

$$\begin{aligned} &\min \|\mathbf{T}\mathbf{x}\|_p^p \\ &\text{subject to } \mathbf{M}\mathbf{x} = \mathbf{b}. \end{aligned} \tag{6.2}$$

In the particular case of gradient-sparse images, for which the finite-differences operations over $\mathbf{x}$ generate sparse images, the reconstruction in compressive sensing is commonly carried by minimizing the total variation (TV) of $\mathbf{x}$ [10], so (6.2) becomes

$$\begin{aligned} &\min \|\mathbf{x}\|_{TV} \\ &\text{subject to } \mathbf{M}\mathbf{x} = \mathbf{b}. \end{aligned} \tag{6.3}$$

where $\|\mathbf{x}\|_{TV}$ is the ℓ_1 of the image defined by

$$X_{TV}(n_1, n_2) = \sqrt{|X(n_1, n_2) - X(n_1 - 1, n_2)|^2 + |X(n_1, n_2) - X(n_1, n_2 - 1)|^2}. \tag{6.4}$$

Note that, since $\mathbf{X}$ is supposed sparse under a finite-differencing operator, and since $\mathbf{X}_{TV}$ is obtained by combining the results of a vertical and a horizontal finite-difference operator over $\mathbf{X}$, $\mathbf{X}_{TV}$ is itself sparse. Thus, (6.3) is the minimization of the ℓ_1 of a sparse signal, which follows the convex optimization problem defined in (3.4).

In this research, we propose an alternative method for reconstructing gradient-sparse images in MRI, as opposed to (6.2) and to the particular case $p = 1$ corresponding to the traditional approach in (6.3). The objective of the proposed method is to improve the reconstructed image, in terms of signal-to-noise ratio (SNR), for the same number of projections in the adopted trajectories, as well as to reduce the total reconstruction time. The method can also, in principle, reconstruct medical images in the case of X-ray tomography; indeed, in this type of imaging the available measurements, supposing negligible diffraction, can be mapped to radial lines in the image's bidimensional Fourier domain, by using an unidimensional Fourier transform – an important result in tomography that follows from the slice

Fourier theorem [93], [50]. Hence, the reconstruction problem is, from a mathematical point of view, analogous to the MRI case with the radial k-space trajectories. On the other hand, in the case of the tomographic problem in the presence of diffraction, the mapping from the measurements to the image's bidimensional Fourier transform leads to different types of line [75], so the problem becomes analogous to a MR reconstruction from a different k-space trajectory.

The proposed method starts by computing, before the optimization procedure for reconstruction, the measurements corresponding to one or more filtered versions of the desired gradient-sparse MR image. These filtered versions, or linear combinations of them, are reconstructed, and a final stage composes the MR image by analyzing the filtered images and the original k-space samples. In other words, a set of linear filters is indirectly applied to the desired image by operating on the k-space, but in such a way that the reconstructed filtered images provide the spectral information necessary to obtain the nonfiltered image.

6.2 The Proposed Method

In the proposed method, we compute, before the optimization procedure for reconstruction, the measurements corresponding to one or more filtered versions of the desired gradient-sparse MR image. These filtered versions, or linear combinations of them, are reconstructed, and a final stage composes the MR image by analyzing the filtered images and the original k-space samples. In other words, a set of linear filters is indirectly applied to the desired image by operating on the k-space, but in such a way that the reconstructed filtered images provide the spectral information necessary to obtain the nonfiltered image.

The fundamental idea is that the used filters should be chosen in such a way that they increase the images' sparsity in the pixel domain. Therefore, the reconstruction based on ℓ_p-minimization can be improved over the direct TV-minimization-based reconstruction, for the same number of k-samples. Also, note that, since these samples correspond to Fourier coefficients of the image to be reconstructed, computing the preprocessed measurements associated to linearly filtered versions of this image is straightforward. In fact, let $\mathbf{h}$ be the

bidimensional kernel of one of the linear, shift-invariant filters, and let $\mathbf{H}$ be the Fourier transform of $\mathbf{h}$ computed with the same number of pixels of the image to be reconstructed. If $\mathbf{b}$ is the vector of k-samples of the original image and $\mathbf{b}_f$ is the set of measurements for a filtered version, then

$$\mathbf{b}_f = \mathbf{H}_{k\in\mathcal{T}} \circ \mathbf{b},$$

where $\circ$ represents the element-by-element product and where $\mathbf{H}_{k\in\mathcal{T}}$ is the vector obtained by stacking all the values of $\mathbf{H}$ that belong to the trajectory $\mathcal{T}$ of k-samples.

In order to implement the proposed method, it is then necessary to define an appropriate set of filters to be applied, in the k-space domain, to the available measurements. Each choice of filters, and of possible linear combinations of them, leads to a different reconstruction scheme. We evaluated three schemes, as described below.

6.2.1 Scheme 1

Since we are considering gradient-sparse images, commonly reconstructed in compressive sensing by the TV minimization approach, the first order finite difference filters provide sparse results, when applied to the rows and columns of the images. Let $\mathbf{h}_1$ and $\mathbf{h}_2$ be the kernels of the vertical and horizontal finite difference filters, respectively, so

$$\mathbf{h}_1 = \begin{bmatrix} 1 & -1 \end{bmatrix}^T \text{ and } \mathbf{h}_2 = \begin{bmatrix} 1 & -1 \end{bmatrix}. \tag{6.5}$$

If $\mathbf{X}$ is a gradient-sparse image in matrix form (not vectorized), then

$$\mathbf{Y} = \mathbf{h}_1 * \mathbf{X} \text{ and } \mathbf{Z} = \mathbf{h}_2 * \mathbf{X},$$

with $*$ the convolution operation, denote the results of differencing $\mathbf{X}$ over its rows and columns, respectively. Furthermore, if Y_{n_1,n_2} and Z_{n_1,n_2} are respectively the components of $\mathbf{Y}$ and $\mathbf{Z}$ in row n_1 and column n_2, then the total-variation of $\mathbf{X}$ is given by

$$\|\mathbf{X}\|_{TV} = \sum_{n_1,n_2} \sqrt{Y^2_{n_1,n_2} + Z^2_{n_1,n_2}}, \tag{6.6}$$

so the idea of minimizing the TV of $\mathbf{X}$ in the traditional approach in (6.3) is based on the sparsity of $\mathbf{Y}^2 + \mathbf{Z}^2$.

In the first configuration of our proposed system, shown in Figure 6.2, we compute the k-space samples of $\mathbf{Y}$ and $\mathbf{Z}$ by multiplying the measured samples, $\mathbf{b}$, by the corresponding values of the Fourier transforms $\mathbf{H}_1$ and $\mathbf{H}_2$ of the two kernels $\mathbf{h}_1$ and $\mathbf{h}_2$, in the locations k belonging to $\mathcal{T}$. The obtained measurements are then used to compute the k-samples of a linear combination of $\mathbf{Y}$ and $\mathbf{Z}$, defined by

$$\mathbf{X}_f = \mathbf{Y} + j\mathbf{Z}, \tag{6.7}$$

with j the imaginary unit. The obtained k-space samples corresponding to $\mathbf{X}_f$, denoted by $\mathbf{b_f}$, then allow the reconstruction of $\mathbf{X}_f$ by the described ℓ_p-minimization, which we solved by using iteratively reweighted least squares (IRLS), easily allowing to choose the value of p. Section 6.2.4 describes the choices of parameters we adopted in our experiments when applying the IRLS algorithm.

Note that the motivation for defining $\mathbf{X}_f$ as in (6.7), and of obtaining its k-space measurements $\mathbf{b_f}$ from the original measurements $\mathbf{b}$, is that once $\mathbf{X}_f$ is reconstructed, it is possible to obtain the desired MR image $\mathbf{X}$. The final image composition stage shown in Figure 6.2 takes into account both the reconstructed filtered image $\mathbf{X}_f$ and the original k-space samples in order to obtain the desired image. This stage is described in more detail in Section 6.2.5.

Also, observe that

$$\|\mathbf{x}_f\|_1 = \sum_n |x_f[n]| = \sum_{n_1,n_2} \sqrt{Y^2_{n_1,n_2} + Z^2_{n_1,n_2}},$$

and, from (6.6), it then follows that

$$\|\mathbf{x}_f\|_1 = \|\mathbf{X}\|_{TV}.$$

Therefore, the choice of filters $\mathbf{h}_1$ and $\mathbf{h}_2$ in Figure 6.2 also leads, for $p = 1$, to $\|\mathbf{x}_f\|_p^p$ being equal to the total-variation of the image to be reconstructed. Hence, with $p = 1$ the minimization in Figure 6.2 is closely related to the TV minimization. By changing the value of p in the IRLS method, however, it is possible to improve the reconstruction for the same number of measurements [18], [65]. We also emphasize that the use of the preprocessed k-space samples in order to reconstruct the filtered image, as opposed to the desired image directly,

leads to the sparsity in the pixel domain, during the compressive sensing optimization (hence, not involving direct or inverse transformations at each iteration). Further improvement in the scheme of Figure 6.2 leads to other advantages, as described in Sections 6.2.2 and 6.2.3.

6.2.2 Scheme 2

The method in Figure 6.2 can still be improved by reconstructing separately the filtered versions of $\mathbf{X}$, as shown in Figure 6.3, and with a modified set of filters, as in Section 6.2.3. Since each separate filtered version of $\mathbf{X}$ is supposed sparser than the composite filtered version used in Figure 6.2, the reconstruction can be improved by separately minimizing its ℓ_p.

Although this requires solving more than one optimization problem, as opposed to the standard approach of (6.3) and to Figure 6.2, the different optimizations can be performed in parallel, as they don't depend upon each other's results. Furthermore, our experimental results show that, depending on the system's settings (value of p and used filters), the total computation time can still be lower or at least comparable to the total variation minimization using a log-barrier algorithm.

As we will show in Section 6.3, for the tested images the separate reconstruction of the filtered versions corresponding to $\mathbf{h}_1$ and $\mathbf{h}_2$ leads to an increase in the SNR and to a

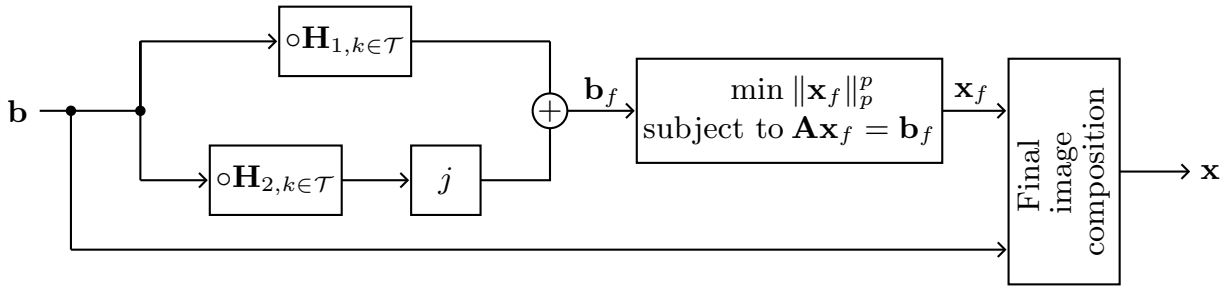


Figure 6.2. Schematic representation of the proposed MRI method, in the first tested form (Scheme 1): prefiltering in the k-space domain of the input measurements $\mathbf{b}$ and reconstruction of the composite filtered version with measurements $\mathbf{b}_f$, before the final image composition.

reduction in the reconstruction time when compared to the system in Figure 6.2, even in the nonparallel implementation.

6.2.3 Scheme 3

Finally, the system in Figure 6.3 can still be used with more than two filters, which can improve reconstruction if the set of filters favors the sparsity of each component and at the same time provides enough spectral information for the final image composition stage. With a total of three filters, we obtain the system in Figure 6.4; in particular, we evaluated the performance for the set of filters in the Haar 2D wavelet decomposition of $\mathbf{X}$, since the $\mathbf{h}_1$ and $\mathbf{h}_2$ kernels of (6.5) are also the Haar 1D analysis filters. The 2D scaling function was not included in the analysis, as it does not generate a sparse version of the input image (its output is a low-resolution version of the input, not a gradient version); rather, the low-pass content needed to fully reconstruct $\mathbf{X}$ is extracted from the original k-space samples. Hence, the set of filters we used in Figure 6.4 is defined by

$$\mathbf{h}_1 = \mathbf{h}_2^T = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \text{ and } \mathbf{h}_3 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

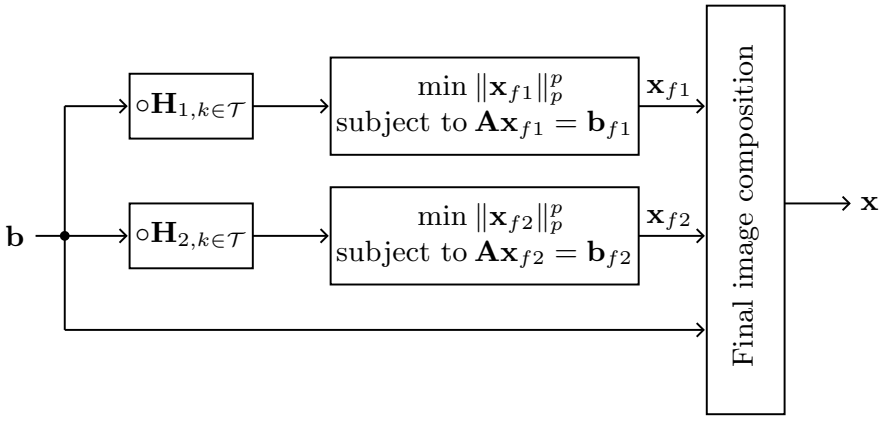


Figure 6.3. Schematic representation of the proposed MRI method, in the second tested form (Scheme 2): prefiltering in the k-space domain of the input measurements $\mathbf{b}$ and separate reconstruction of each filtered version with measurements $\mathbf{b}_{f1}$ and $\mathbf{b}_{f2}$.

It is worth observing that the kernels we tested in the systems of Figure 6.2 and Figure 6.3 correspond to the finite differences in the horizontal and vertical directions, whereas the bidimensional Haar kernels we used in Figure 6.4 also include the finite differences in the diagonal. These filters are also the Haar kernels used in the one-stage wavelet decomposition. In future tests, we want to evaluate the performance of the three schemes in using different sets of wavelets, in order to reconstruct images that cannot be considered sparse under finite differencing.

6.2.4 The Optimization Stages

In evaluating the systems of Figures 6.2, 6.3 and 6.4 we used the iteratively reweighted least squares (IRLS) algorithm to solve the optimization problems. This algorithm is described in detail in [18] and, with prior information, in Chapter 5. Here, we describe the choice of parameters used in our experiments for the case of the MR image reconstructions. Also, we emphasize that these parameters were kept constant during all the tests; the idea was not to tweak the algorithm for each image, but to use the same procedure during all the tests.

The main parameters we defined previously to the implementation of the IRLS were the

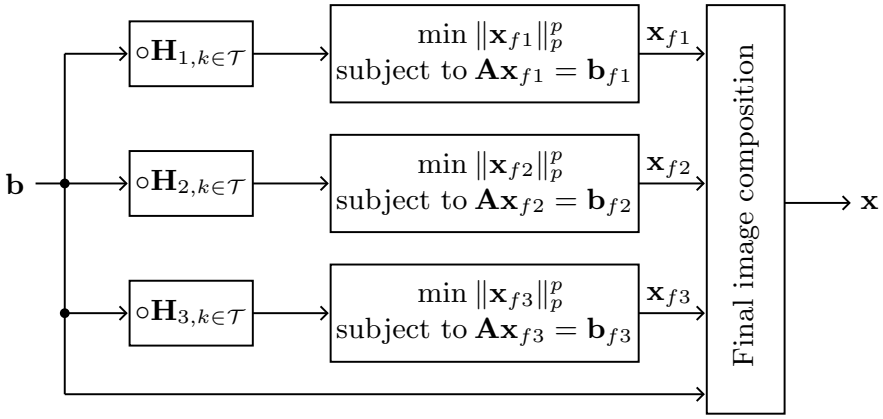


Figure 6.4. Schematic representation of the proposed MRI method, in the third tested form (Scheme 3): prefiltering in the k-space domain of the input measurements $\mathbf{b}$ and separate reconstruction of each filtered version with measurements $\mathbf{b}_{f1}$, $\mathbf{b}_{f2}$ and $\mathbf{b}_{f3}$.

initial regularization parameter μ_0 and the factor f by which the regularization parameter was reduced after each stage of the algorithm (note that a stage is a sequence of iterations for which the regularization parameter is kept constant; the next stage then starts by updating the regularization parameter according to $\mu^{(k)} = f\mu^{(k-1)}$, with $\mu^{(k)}$ the regularization at stage k). We adopted $\mu_0 = 2$ and $f = 0.5$.

As for the stopping condition for each stage, it was defined according to Section 5.3, and we also restricted each stage to a maximum of 10 iterations, and the algorithm to a maximum of 10 stages.

For more information on the IRLS, and on how the parameters above are used in the algorithm, see Chapter 5. Also, an important issue regarding the IRLS algorithm in the proposed MRI method is that the inner linear systems that must be solved at each iteration are large-scale, meaning that the system matrices are too large to be stored or even computed directly by matrix multiplications. Different possible ways to deal with this problem, including the traditional approach of iterative solutions to the linear systems (indirect methods), are discussed in Section 7.2 and Section 7.3.

6.2.5 The Composition Stage

The final image composition stage in the three schemes of Figures 6.2, 6.3 and 6.4 aims at obtaining the desired MR image from the set of reconstructed filtered images and the original measurements in $\mathbf{b}$. Note that the filtered images, corresponding to the finite differences operators applied to the rows, columns, and diagonals of $\mathbf{X}$, provide the high-pass information that complement the original spectral information in $\mathbf{b}$. The DC level of $\mathbf{X}$, for instance, is lost in the filtered images, but it is available in the original measurements $\mathbf{b}$.

In our implementation, $\mathbf{X}$ is reconstructed first in the frequency domain, using the following steps: (1) We divide the frequency components of $\mathbf{X}_{f1}$, $\mathbf{X}_{f2}$ and, in the case of the Scheme 3, $\mathbf{X}_{f3}$ by the corresponding filters $\mathbf{H}_1$, $\mathbf{H}_2$, and $\mathbf{H}_3$ (inverse filtering) — in the particular case of Scheme 1, note that, once $\mathbf{X}_f$ is reconstructed, its real and imaginary parts, $\mathbf{Y}$ and $\mathbf{Z}$, provide the filtered images associated respectively to $\mathbf{H}_1$ and $\mathbf{H}_2$. (2) We

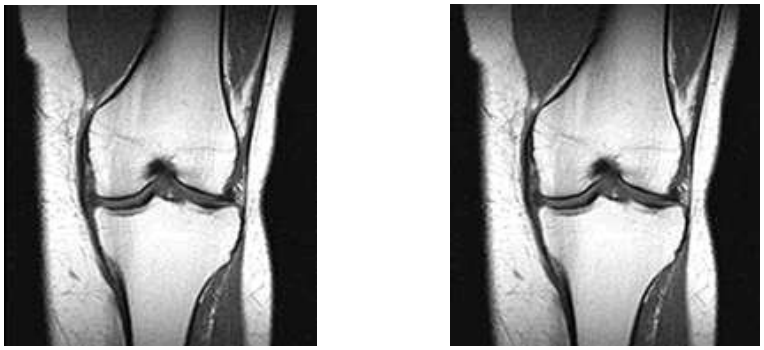
then combine the resulting frequency coefficients of $\mathbf{X}$ given by the above inverse filtering with those in the original k-space trajectory $\mathcal{T}$. (3) Finally, the inverse Fourier transform of the estimated coefficients of $\mathbf{X}$ is computed, providing the representation of $\mathbf{X}$ in the pixel domain.

6.3 First Experimental Results

We evaluated the proposed method for reconstructing real magnetic resonance images (MRIs) and, in particular, magnetic resonance angiograms (MRAs), these being commonly sparse under finite differencing [57]. In each of the experiments, we started by taking a reference image $\mathbf{X}$; the measurements $\mathbf{b}$ were then obtained by computing the Fourier transform of $\mathbf{X}$ over a Cartesian grid and selecting the components closest to a particular k-space trajectory (a procedure similar to that of [10]). For consistency, we always adopted the trajectory $\mathcal{T}$ in Figure 6.1(c), composed of radial lines at uniformly distributed angles (we tested different numbers of radial lines). This particular type of trajectory is appropriate for the reconstruction of gradient-sparse images in compressive sensing [58], but other types could also be used.

We compared the results of the reconstructions using the three schemes in Figures 6.2, 6.3 and 6.4 to those provided by the total-variation (TV) minimization approach, implemented using the log barrier algorithm described in [9]. The comparison took into account the signal-to-noise ratio (SNR) in dB and the total reconstruction time.

As a first example, we show in Figure 6.5 a knee MRI and the corresponding reconstructed version obtained by using Scheme 2 in Figure 6.3, with $p = 0.5$. In this example, 120 radial lines in the k-space were used in the reconstruction, resulting in a 71-dB signal-to-noise ratio. In Figure 6.6, on the other hand, we show a 3.0-Tesla brain MRA, as reconstructed by (a) TV minimization and by (b) the proposed method, using Scheme 3 in Figure 6.4 with $p = 0.5$ (with $r = 72$ radial lines sampled in the k-space, in both cases). The image reconstructed by Scheme 3 appears less blurred than the image reconstructed by TV minimization, although



(a) (b)

Figure 6.5. (a) An example of a knee MR image; (b) the image reconstructed by the proposed method (Scheme 2), from $r = 120$ radial lines when sampling in the k-space. Original reference image (a) retrieved from http://www.diagnosticprofessionals.com/MRI_Info.htm.

the difference is relatively small (corresponding detailed views appear in Figure 6.6(c) and Figure 6.6(d)).

A more detailed numerical evaluation of the reconstruction of the knee MRI of Figure 6.5 appears in Figure 6.7, where we show, for different numbers of radial lines in the k-space, the SNRs of the images reconstructed by the three schemes with $p = 0.5$ and by TV minimization. We also compare, in Figure 6.8, the total reconstruction times (in the case of the proposed method, both for serial and for parallel implementations). Note that Schemes 1, 2, and 3 resulted in higher SNRs, as compared to the TV minimization, for all the tested numbers of radial lines (r). Also, Scheme 2 lead to better results, in terms of SNR, than Schemes 1 and 3; as we will mention, this was not the case for the other tested images, for which the filters used in Scheme 3 provided a better representation and lead to better results than Scheme 2.

Regarding the reconstruction times, we observe in Figure 6.8(a) and Figure 6.8(b) that Schemes 2 and 3 resulted in higher reconstruction times than the TV minimization, for the highest tested values of r and in the case of serial implementation (although for Scheme 1 the reconstruction times were always lower). In the case of the parallel implementations, however, the reconstruction times for the proposed schemes were always lower than in the TV minimization approach. We emphasize that the parallel implementations do not require any modification in the optimization procedures themselves; in fact, as we see in Figures 6.2,

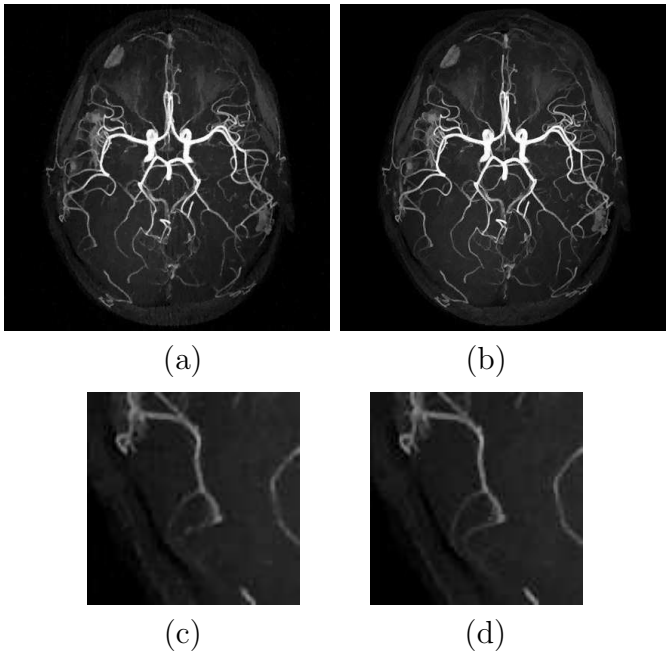


Figure 6.6. Example of a 3.0-Tesla MR angiogram of the brain reconstructed using (a) total-variation minimization and (b) the proposed method (Scheme 3), from the same numbers of radial lines ($r = 72$) when sampling in the k-space; detailed views of (a) and (b) are shown respectively in (c) and (d). Original reference image retrieved from http://wsunews.wsu.edu/Content/Publications/MRI1.jpg.

6.3 and 6.4 the optimizations corresponding to the different filters adopted are independent of each other.

For the brain MRA of Figure 6.6, we also compare the reconstructions, for several values of r, in terms of SNR and total reconstruction times, as shown in Figure 6.9 and Figure 6.10. Observe in this case that Scheme 3 provided higher SNRs than all the other tested methods; this was the case for all the tested images except for the knee MRI, as exemplified in the numerical results plotted in Figure 6.11 for a 2.0-Tesla hand angiogram. We also note that the reconstructions times in the cases of both the brain MRA and the hand MRA (see Figure 6.12) were lower for the proposed method than for the used TV minimization algorithm, even for the serial implementations.

Another important aspect to evaluate in future experiments is the behavior of the implemented method with respect to the parameter p. In Figure 6.13 and Figure 6.14, we present a preliminary comparison of the results of the TV minimization with those by the Scheme

3, for a few values of p and using the brain MRA of Figure 6.6. Observe that, starting with $p = 1$ (for which Scheme 3 already provides higher SNRs than the TV minimization), as we decreased p to 0.5 the resulting SNRs increased. Even lower values of p, however, did not further improve the SNRs, and actually reduced them with respect to $p = 0.5$. This preliminary result lead us to adopt $p = 0.5$ in the main tests above.

Finally, in Table 6.1 we summarize the results, in terms of SNR, for different tested images (including the knee MRI and the brain and hand MRAs). These results refer to Scheme 3 and to the TV minimization, and are consistent with those previously discussed.

6.4 Final Remarks

This chapter described and evaluated a method for the reconstruction of gradient-sparse magnetic resonance images based on the preprocessing of the available k-samples. This method applies one of different possible sets of linear shift-invariant filters during the preprocessing, in order to yield the measurements corresponding to filtered versions of the image. The implemented optimization algorithms then reconstruct these filtered versions using the preprocessed measurements, and a final composition stage builds the desired image from the filtered versions and from the low-pass information in the original measurements.

Our first experimental results show that, for the tested images, the proposed method improves the reconstruction, in terms of signal-to-noise ratio and visual quality, over the total-variation minimization, commonly used for gradient-sparse images. We also observed that the proposed method, when using three preprocessing filters in a serial implementation, leads to reconstruction times that are equivalent or lower than in the total-variation minimization using a log-barrier algorithm (we emphasize, however, that the implementations of both the proposed method and the total-variation minimization were not optimized for time). Additionally, the proposed method also allows for a straightforward parallel implementation, as the reconstructions of the filtered versions of the desired image are independent of each other.

In Chapter 8, we analyze different images as we test the performance of the proposed MRI method. Also, we test the improvement in performance by adding prior information, according to the procedures proposed in Chapter 7.

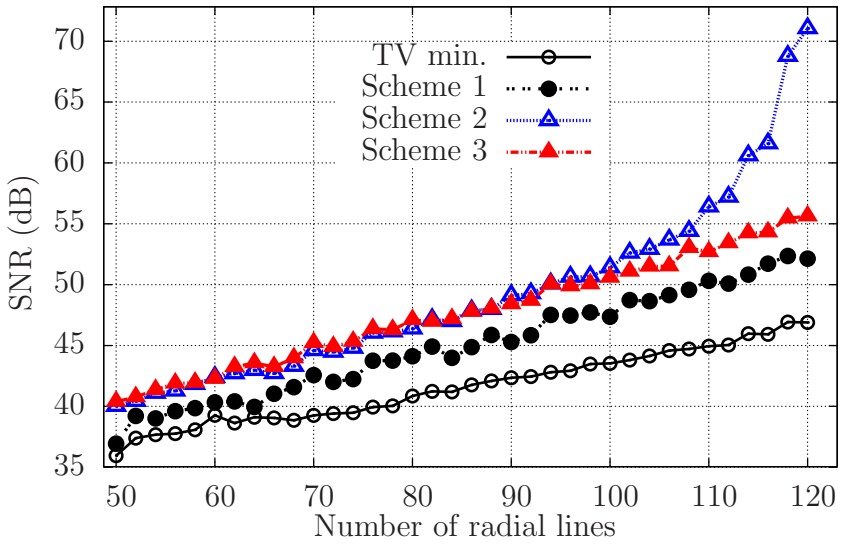


Figure 6.7. Signal-to-noise ratios for a knee MRI when applying the total-variation (TV) minimization and the three different configurations of the proposed method with different numbers of k-space samples. In these tests, we adopted $p = 0.5$ when applying the IRLS algorithm.

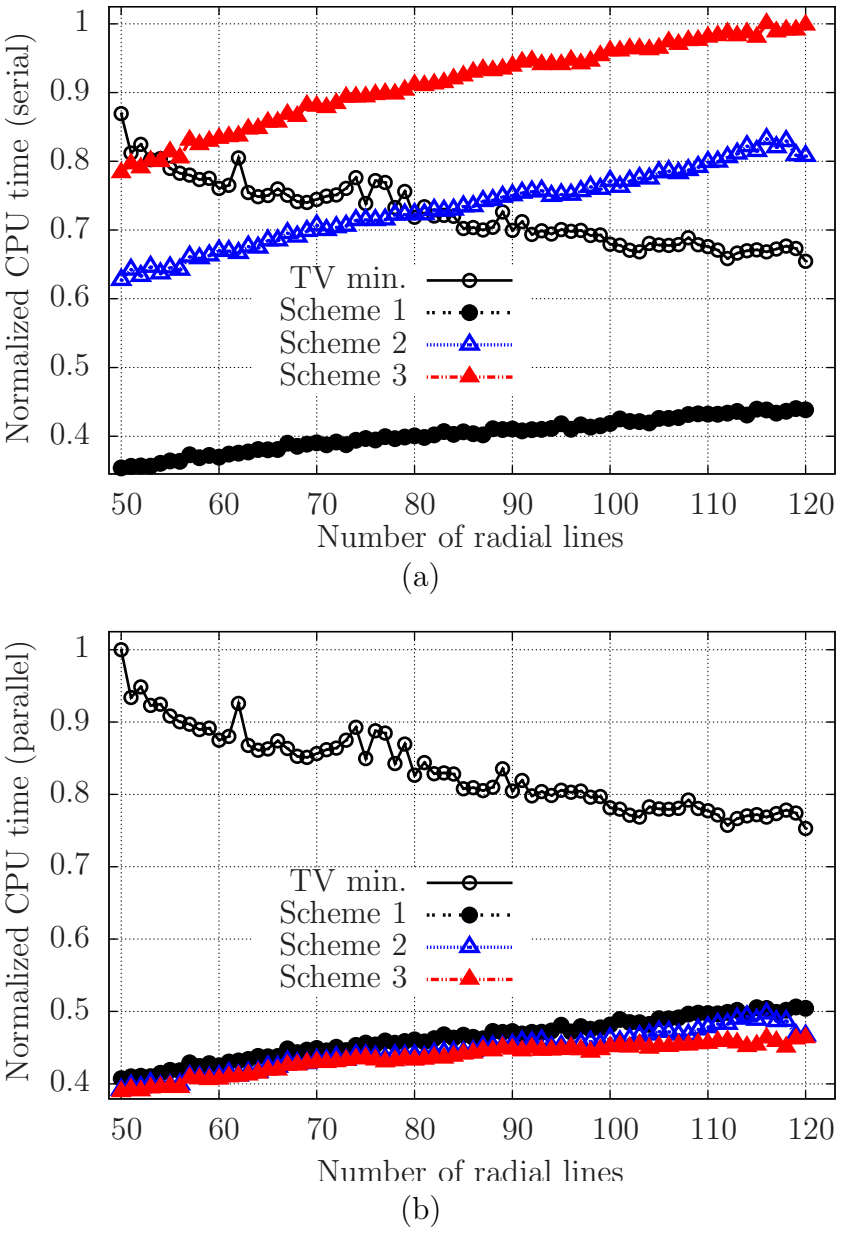


Figure 6.8. Reconstruction times for a knee MRI when applying the total-variation (TV) minimization and the three different configurations of the proposed method with different numbers of k-space samples, using (a) serial and (b) parallel implementations of the proposed method. In these tests, we adopted $p = 0.5$ when applying the IRLS algorithm. In the normalized scale, a CPU time of 1 corresponds to approximately 13 minutes in a PC with a 2.60 GHz processor and 7.80 GB of RAM memory.

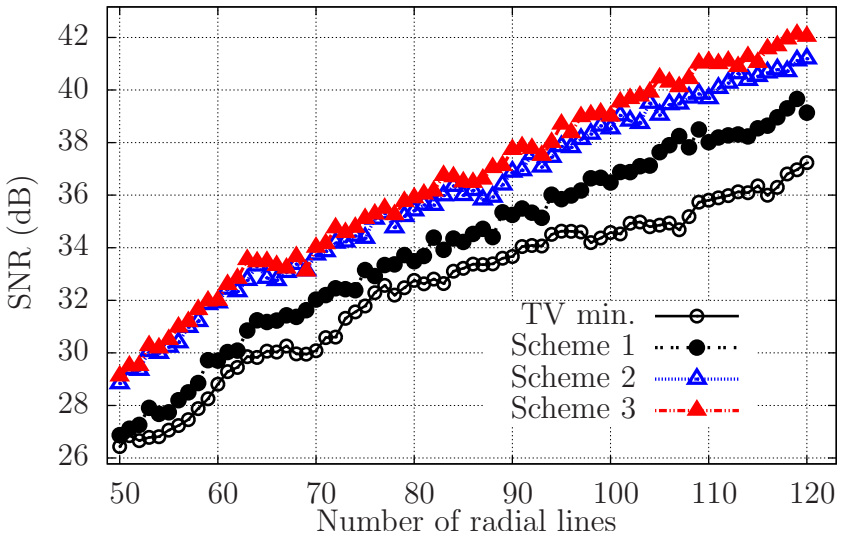


Figure 6.9. Signal-to-noise ratios for Angiogram 2 when applying the total-variation (TV) minimization and the three different configurations of the proposed method with different numbers of k-space samples. In these tests, we adopted $p = 0.5$ when applying the IRLS algorithm.

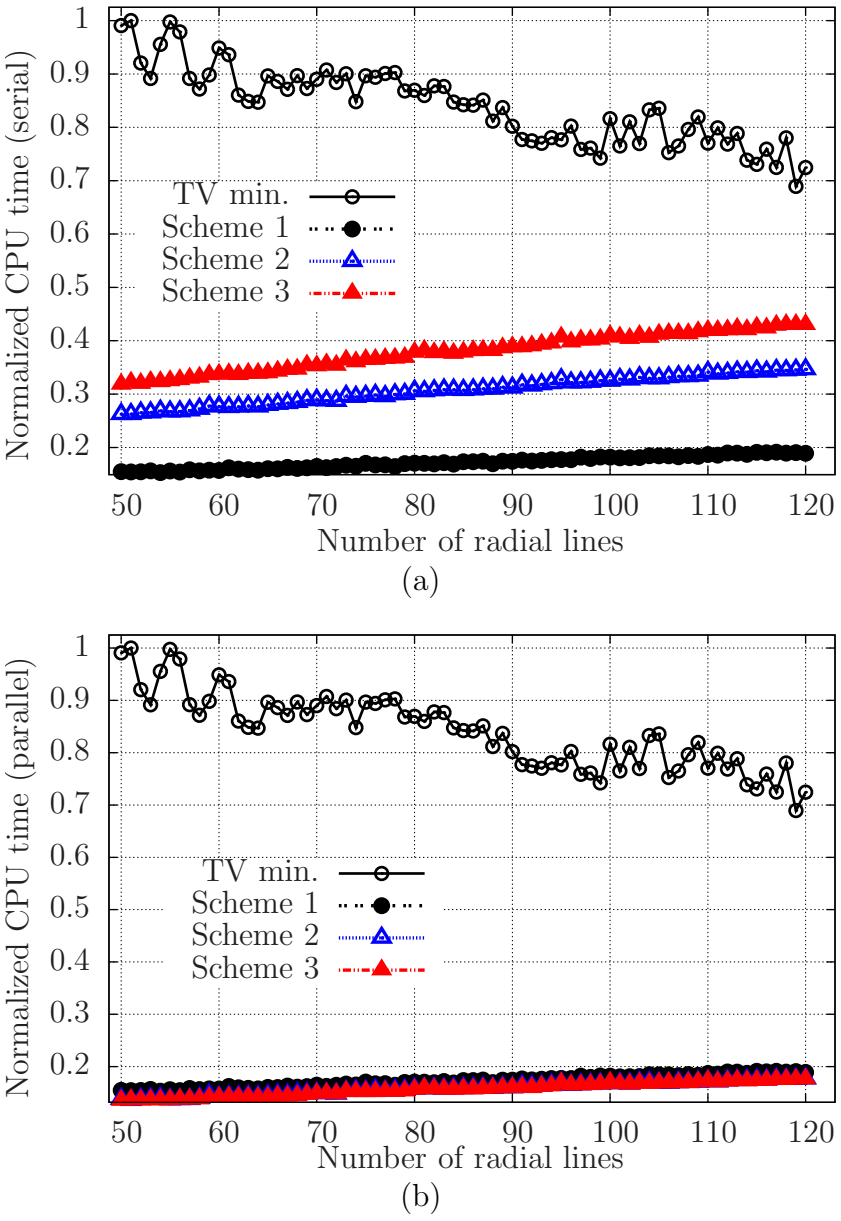


Figure 6.10. Reconstruction times for Angiogram 2 when applying the total-variation (TV) minimization and the three different configurations of the proposed method with different numbers of k-space samples, using (a) serial and (b) parallel implementations of the proposed method. In these tests, we adopted $p = 0.5$ when applying the IRLS algorithm. In the normalized scale, a CPU time of 1 corresponds to approximately 40 minutes in a PC with a 2.60 GHz processor and 7.80 GB of RAM memory.

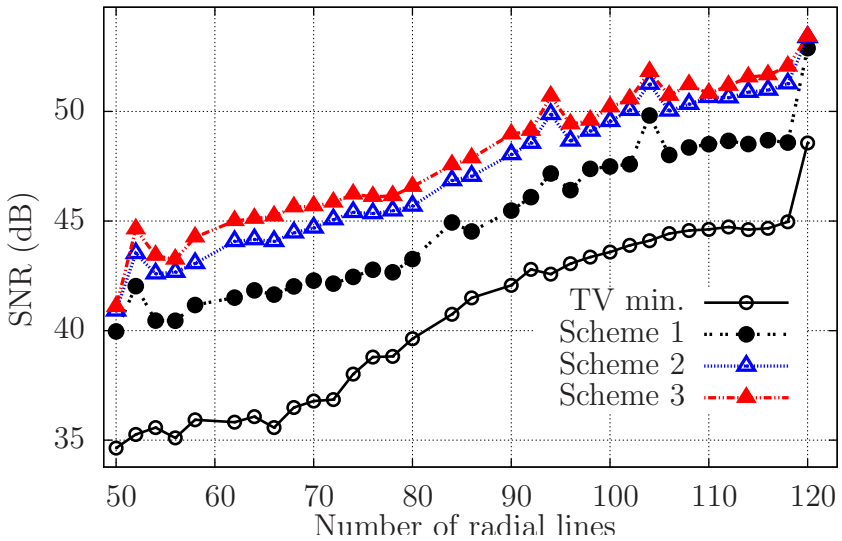


Figure 6.11. Signal-to-noise ratios for a 2.0-Tesla hand angiogram when applying the total-variation (TV) minimization and the three different configurations of the proposed method with different numbers of k-space samples. In these tests, we adopted $p = 0.5$ when applying the IRLS algorithm.

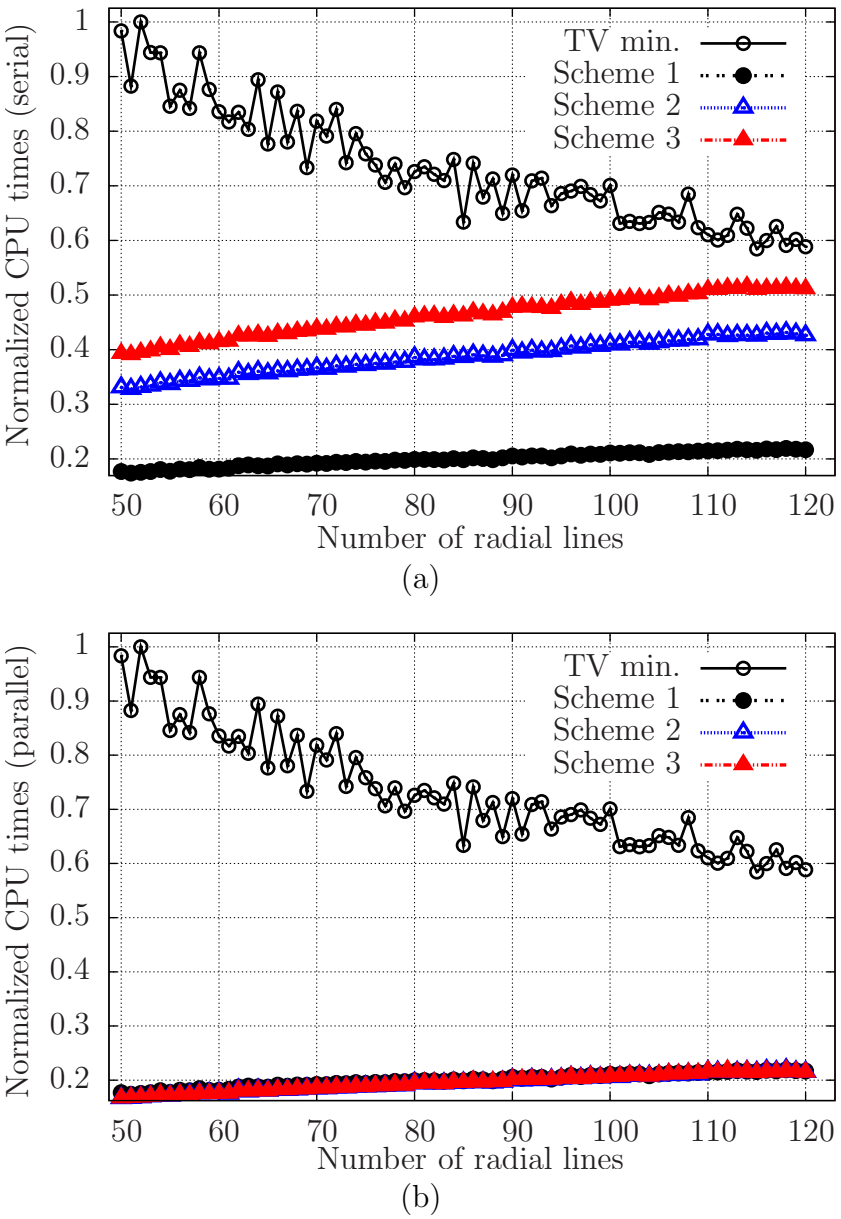


Figure 6.12. Reconstruction times for a 2.0-Tesla hand angiogram when applying the total-variation (TV) minimization and the three different configurations of the proposed method with different numbers of k-space samples, using (a) serial and (b) parallel implementations of the proposed method. In these tests, we adopted $p = 0.5$ when applying the IRLS algorithm. In the normalized scale, a CPU time of 1 corresponds to approximately 50 minutes in a PC with a 2.60 GHz processor and 7.80 GB of RAM memory.

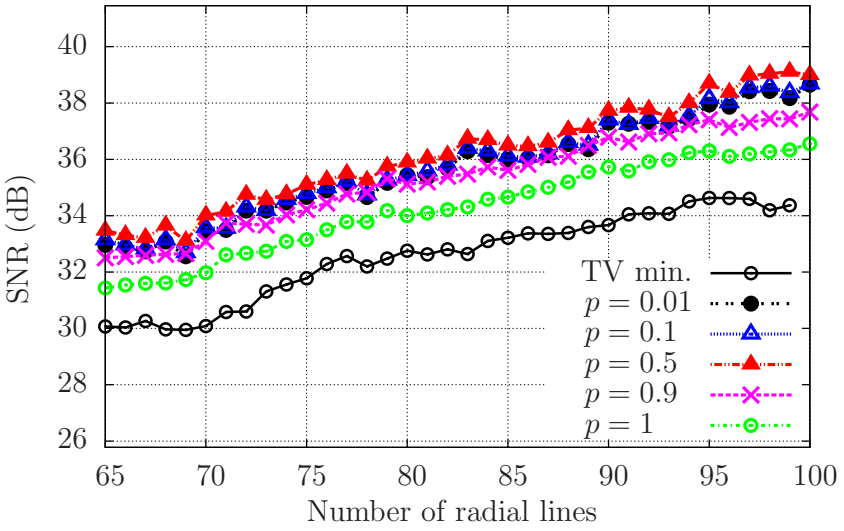
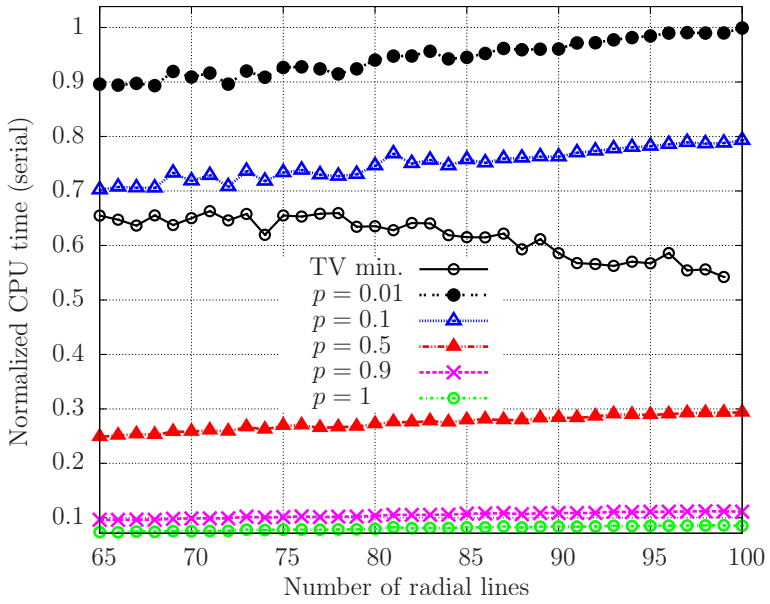
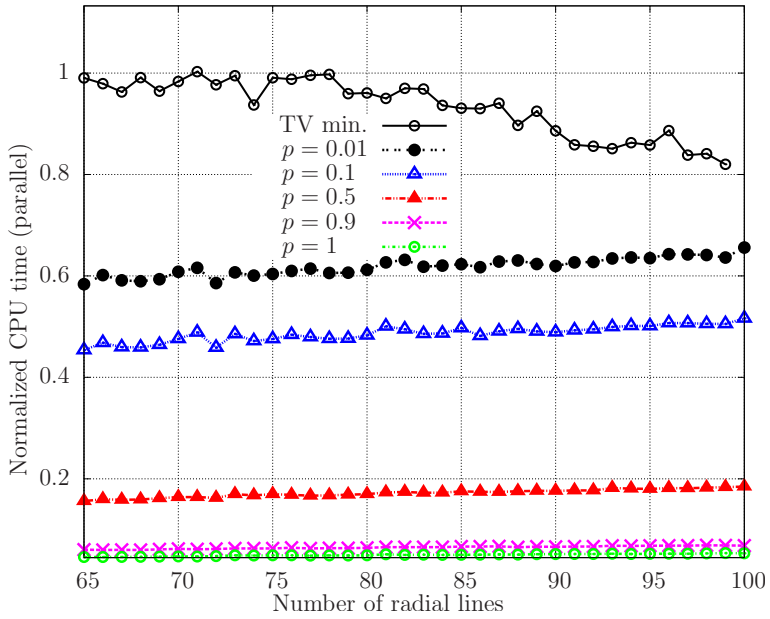


Figure 6.13. Signal-to-noise ratios for Angiogram 2 when applying the total-variation (TV) minimization and Scheme 3 of the proposed method with different values of p and different numbers of k-space samples.



(a)



(b)

Figure 6.14. Reconstruction times for Angiogram 2 when applying the total-variation (TV) minimization and Scheme 3 of the proposed method with different values of p and different numbers of k-space samples: (a) in the serial implementations; (b) in the parallel implementations of the proposed method. In the normalized scale, a CPU time of 1 corresponds to approximately 55 minutes in a PC with a 2.60 GHz processor and 7.80 GB of RAM memory.

Table 6.1. Signal-to-noise ratios (SNRs) of different images reconstructed by total-variation (TV) minimization and by Scheme 3 in Figure 6.4.

Image	Number of radial lines (r) sampled	Output SNR (dB)	
		TV	Scheme 3
Angiogram 1	120	42.2	49.4
Angiogram 2	116	36.0	41.6
Angiogram 3	120	42.0	48.2
Angiogram 4	90	47.7	49.1
Legs MRA	66	35.6	45.2
Hand MRA	120	46.9	55.6
Knee MRI	70	49.1	50.5
Coronary MRA	110	41.0	47.1

7

MR Imaging Using Compressive Sensing with Prior Information

This chapter describes how the MRI method introduced in Chapter 6 can be combined with the use of prior information about the sparse representation's support. As we detail below, since the optimization stages of the proposed method operate on prefiltered k-space samples, the prior information to be extracted before the reconstruction must correspond to the filtered versions of the target images.

We then start by describing, in Section 7.1, how the proposed prior information can be obtained, in the context of MRI. Next, Section 7.2 describes the proposed method combining the prefiltering strategy with the use of prior information. As we mention at this point, the use of prior information with indirect approaches to solve the inner linear systems can increase their instability; therefore, the total reconstruction times can actually increase with respect to the proposed method without prior information (although they were still generally lower than in the traditional TV-minimization using a log-barrier algorithm). We then present, in Section 7.3, a possible strategy for the fast computation of the inner system matrices, which allows a direct method to solve these systems and thus copes with the

instability issue. Finally, the first examples of reconstruction using the proposed MRI method with prior information appear in Section 7.4, where we test the improvements due to prior information using both simulated and real functional MRIs.

In Chapter 8, we describe additional experiments evaluating the method here discussed. In particular, we evaluate its performance as we change the number of automatically extracted prior support locations, and using different types of reference MR images. We also evaluate the system's robustness to wrong prior locations. For more details, see Section 8.3.

7.1 Support Prior Information in MRI

The MRI method proposed in Chapter 6 also allows the inclusion of prior information about the sparse representation's support. This can lead to further improvement in the reconstruction, as we analyze in this chapter.

Note in Figures 6.2, 6.3 and 6.4 that, since the optimization stages operate on the preprocessed k-space samples in order to reconstruct filtered versions of the desired image, the mentioned prior information must correspond to locations of nonzero coefficients in those filtered versions. This information can be extracted from different sources, such as medical records, previous iterations in the optimization procedures, or previous frames in a functional MRI. We emphasize here the case of functional magnetic resonance imaging (fMRI), and discuss how the method can be implemented with prior information in this case.

Typically, in fMRI the functioning of a certain organ or set of organs generates observable changing patterns in a sequence of images, while the patient is required to remain still. Other body parts, mainly external structures, remain mostly unaltered in the images. After some frames are observed, it is possible to infer the basic unchanging structures, and use them as prior information in the next reconstruction stages. This information extraction can be done iteratively, by a trained observer, or automatically, by an algorithm applied to previous frames. We will here adopt the second strategy.

In this situation, consider that at least two frames from an fMRI have already been reconstructed, from k-space samples, by using one of the schemes in Figures 6.2, 6.3 and

6.4, without prior information. From these two frames, suppose that a mask $\mathbf{M}$ is extracted, meaning a binary image that identifies with ones the unchanging structures an with zeros the remaining parts. By applying the filters $\mathbf{h}_1$, $\mathbf{h}_2$, $\mathbf{h}_3$ to $\mathbf{M}$, and finding the resulting highest coefficients, it is then possible to determine the locations of nonzero coefficients that potentially appear in the next frames as well. These locations can then be used in the optimization stages of Figures 6.2, 6.3 and 6.4, according to Algorithm 1, in Section 5.3, to improve the reconstruction of the next frames.

The described method for obtaining prior information for the proposed MRI method appears in Figures 7.1, 7.2 and 7.3, respectively for Scheme 1, Scheme 2 and Scheme 3. Note that there is a different prior information set Φ for each filtered image; thus, $\mathbf{h}_1$ corresponds to Φ_1, $\mathbf{h}_2$ corresponds to Φ_2, and so forth. Even the numbers of known locations can be different for each filtered image, which is appropriate, for example, in case the mask has noticeable patterns in a preferable direction, so more locations φ should be reserved for the filter that favors that direction as well. In the preliminary tests, however, we used the same number of locations for all filters.

Regarding the computation of the mask $\mathbf{M}$, it can also be conducted in different ways, including assisted procedures and automatic ones. The simplest idea is to identify in two consecutive reconstructed frames $\mathbf{X}_{r1}$ and $\mathbf{X}_{r2}$ the regions that change below a specified threshold, as in Figure 7.4. Note that a higher threshold will lead to smaller regions adopted

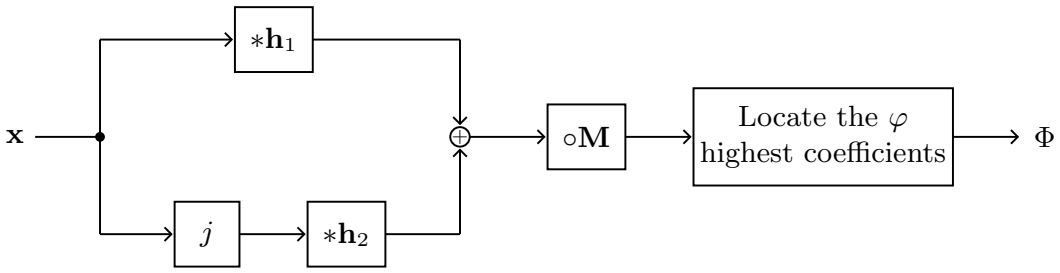


Figure 7.1. Extraction of the prior information from previous frames, in functional MRI using reconstruction Scheme 1. The between-frames mask $\mathbf{M}$ is computed according to Figure 7.4.

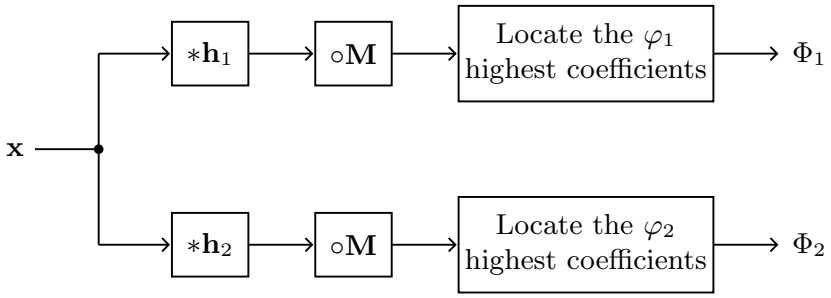


Figure 7.2. Extraction of the prior information from previous frames, in functional MRI using reconstruction Scheme 2. The between-frames mask $\mathbf{M}$ is computed according to Figure 7.4.

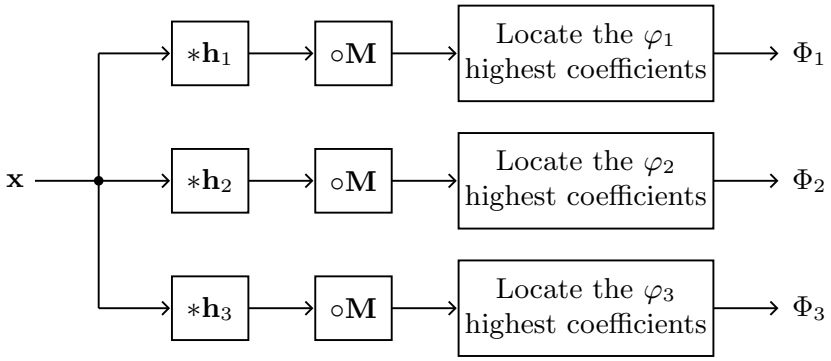


Figure 7.3. Extraction of the prior information from previous frames, in functional MRI using reconstruction Scheme 3. The between-frames mask $\mathbf{M}$ is computed according to Figure 7.4.

as unchanging structures, and vice-versa. Also, the mask can be updated after intervals of frames, to account for slower changes in the fMRI.

We emphasize that an automatic procedure for extracting prior information, as the one here described, is prone to including wrong prior locations as well, meaning locations of filtered components that do not belong to the support. This is specially true when the chosen values of φ_1, φ_2 and φ_3 are high compared to the filtered images' sparsities (which are not known a priori). Other effects can also lead to wrong prior information, as we discuss in Chapter 8, when we also evaluate the method's robustness to this wrong information.

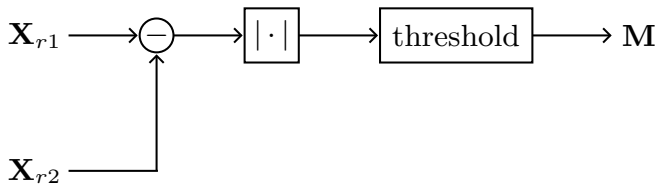


Figure 7.4. Procedure for obtaining the mask $\mathbf{M}$ used to determine the prior information in the proposed functional MRI methods.

7.2 Proposed MRI Methods with Prior Information

Once the needed prior information has been extracted from a sequence of frames, according to the procedures in Section 7.1, it can be efficiently used during the optimization stages of the methods in Figures 6.2, 6.3 and 6.4. The reconstruction itself then follows the methods with prior information that we discussed in details in Chapter 4 and Chapter 5.

Figures 7.5, 7.6 and 7.7 show the three MR imaging schemes that result from the methods in Figures 6.2, 6.3 and 6.4, respectively, combined with the prior information extracted as in Figures 7.1, 7.2 and 7.3. Note that, even for a single fMRI sequence, the prior information sets are different for each used filter. The corresponding optimization stages with prior information can then be conducted using Algorithm 1.

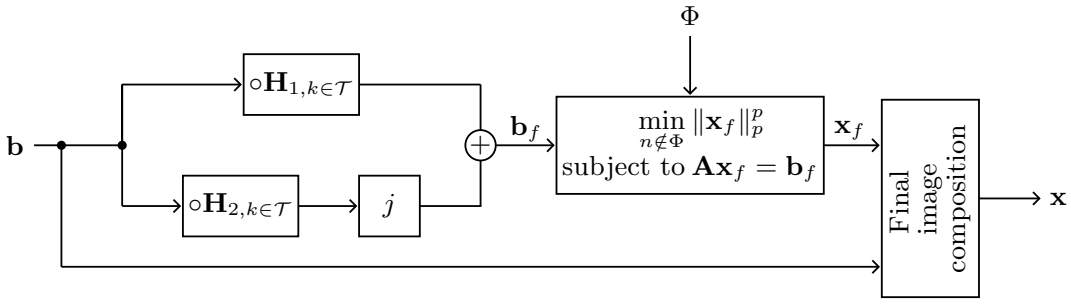


Figure 7.5. Schematic representation of the proposed MRI method with prior information, in the first tested form (Scheme 1): prefiltering in the k-space domain of the input measurements $\mathbf{b}$ and reconstruction of the composite filtered version with measurements $\mathbf{b}_f$, before the final image composition.

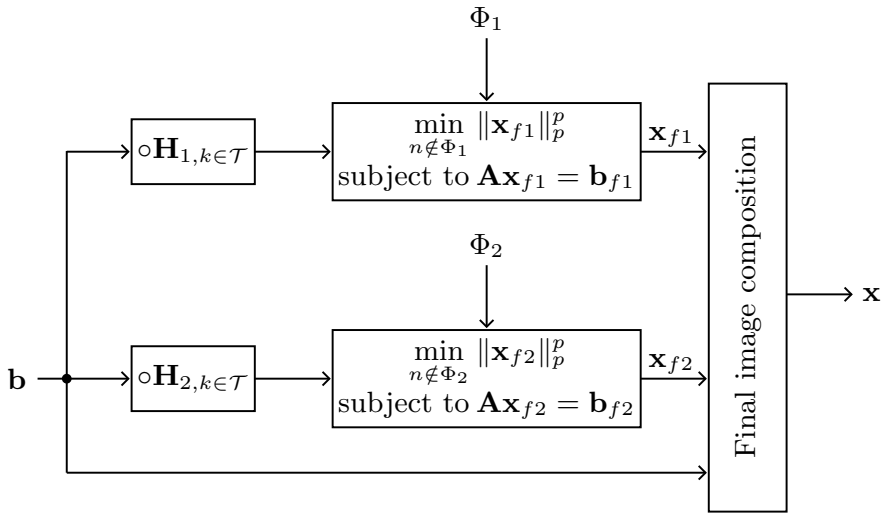


Figure 7.6. Schematic representation of the proposed MRI method with prior information, in the second tested form (Scheme 2): prefiltering in the k-space domain of the input measurements $\mathbf{b}$ and separate reconstruction of each filtered version with measurements $\mathbf{b}_{f1}$ and $\mathbf{b}_{f2}$.

We emphasize that, in implementing the method in Figures 7.5, 7.6 and 7.7, and specifically when doing each ℓ_p-minimization we must solve a sequence of linear systems. Each of these systems leads to an iterate, which is then used to solve the next linear system, until convergence. In fact, in Section 5.3 we describe the IRLS algorithm with prior information, and show that the solution to the minimization problem with prior information can be obtained by solving a sequence of linear systems of the type

$$\mathbf{A}\mathbf{Q}^{(m)}\mathbf{A}^H\mathbf{y} = \mathbf{b}, \tag{7.1}$$

and then computing the m-th iterate

$$x^{(m)} = \mathbf{Q}^{(m)}\mathbf{A}^H\mathbf{b}.$$

Also note that, in the equations above, $\mathbf{A}$ is the matrix such that $\mathbf{A}\mathbf{x} = \mathbf{b}$ is the vector of measurements taken from $\mathbf{x}$, so that $\mathbf{b}$ contains the k-space samples taken from the bidimensional Fourier transform of $\mathbf{x}$, in the adopted trajectory (refer to Chapter 5 and Chapter 6 for details).

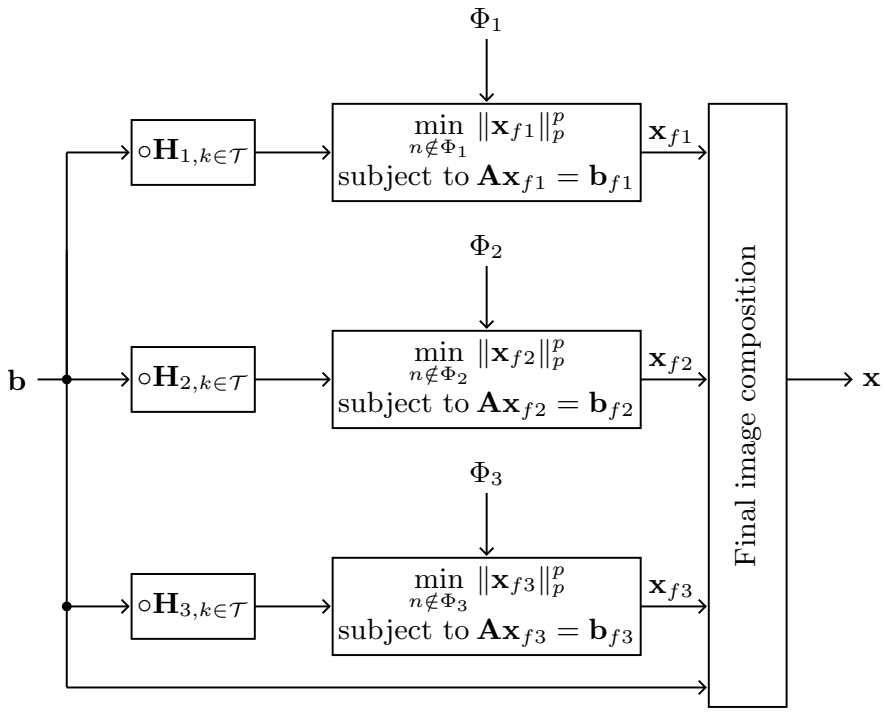


Figure 7.7. Schematic representation of the proposed MRI method with prior information, in the third tested form (Scheme 3): prefiltering in the k-space domain of the input measurements $\mathbf{b}$ and separate reconstruction of each filtered version with measurements $\mathbf{b}_{f1}$, $\mathbf{b}_{f2}$ and $\mathbf{b}_{f3}$.

Now, in solving (7.1), it is not viable to compute and store the whole system matrix $\mathbf{S} = \mathbf{A}\mathbf{Q}^{(m)}\mathbf{A}^H$ at once, and then to solve the system through a direct method such as Gauss elimination or Cholesky factorization. In fact, the matrix $\mathbf{A}$ itself is $\ell \times N$, with ℓ the number of measurements and N the number of pixels in each MR frame; then, for a 1000×1000 image and $\ell = 40,000$ measurements, for instance, $\mathbf{A}$ will have 4×10^{10} components. If each component is stored using 32 bits, this would require more than 1 Gigabytes only to store $\mathbf{A}$.

The most common procedure to deal with such large-scale problems is to use an indirect method, in which the linear system (7.1) itself is solved in an inner iterative procedure. In this case, if it is possible to compute $\mathbf{A}\mathbf{x}$ for any vector $\mathbf{x}$ in an efficient manner and without having to store $\mathbf{A}$, then the system can be solved through a sequence of iterations that use

repeatedly this kind of operations. In fact, we used the conjugate-gradient approach [72], [83] in our solutions; since $\mathbf{Ax}$ is the vector of k-space samples, it can be computed efficiently by taking the bidimensional Fourier transform of $\mathbf{x}$ and then selecting the coefficients located inside the used k-space trajectory (in our case, a set of radial lines).

Using an indirect approach also has its disadvantages, however. As we show in Chapter 8, we have observed that adding prior information combined with indirect methods can result in an increased computation time with respect to our prefiltering method without prior information (although the reconstruction times were still generally lower than in the traditional TV-minimization using a log-barrier algorithm). This issue is analyzed in detail in Section 8.3, but we emphasize that with direct methods, increasing φ always resulted in lower reconstruction times, as discussed in Chapter 5.

In order to obtain a compromise between the reconstruction times and the final image qualities, we opted to increase the value of τ^{p-2} in (5.17) to 10^{-1}, instead of the original $\tau^{p-2} = 10^{-3}$. This reduced the weight of the prior information in the new method, but led to reconstruction times that were still lower than in the TV-minimization, and with an improvement in the final SNRs due to prior information (see Section 8.3).

We also suggest a second approach. If each component of the system matrix $\mathbf{S} = \mathbf{A}\mathbf{Q}^{(m)}\mathbf{A}^H$ can be efficiently computed, then it is possible to solve (7.1) in an efficient manner without having to store $\mathbf{S}$ entirely. In fact, [78], [42] and [41] describe how the Cholesky factorization of a matrix can be obtained by blocks, and even in parallel, if only submatrices of this matrix are available at any given time. Once the Cholesky factorization of $\mathbf{S}$ is done, the solution to the linear system is straightforward, based on sequential substitutions [39] . This approach would have the advantage of allowing lower values of τ^{p-2}, and consequently a higher prior information weight; therefore, it could lead to even higher SNRs and, as Chapter 5 suggests (based on the experiments using direct methods), lower reconstruction times as φ increases.

Implementing this direct solution is out of the scope of this dissertation. However, we show in the next section how each component of $\mathbf{S}$ can be efficiently computed, which allows any of its submatrices to be computed as needed for the methods described in [78], [42]

and [41].

7.3 Efficient Computation of the System Matrix

Consider again the linear system (7.1), which can be written as

$$\mathbf{S}\mathbf{y} = \mathbf{b}, \tag{7.2}$$

with

$$\mathbf{S} = \mathbf{A}\mathbf{Q}^{(m)}\mathbf{A}^H. \tag{7.3}$$

As we mentioned in Section 7.2, for large-scale problems it is not viable to directly compute $\mathbf{S}$ by conducting the matrix multiplications in (7.3). Indeed, the matrix $\mathbf{A}$ is typically too large to be stored directly, so an indirect approach must be devised to obtain $\mathbf{S}$ or to solve (7.2) without the need for explicitly storing this matrix, element-by-element.

In fact, in the case of large-scale problems, it is common to solve the linear systems by using iterative procedures that do not require the main system matrix to be stored in its entirety at any time. For instance, in Chapter 6 we used and iterative procedure based on conjugate gradients to solve each linear system. We emphasize, however, that in some cases direct implementations are still possible, especially in the case of sparse linear systems [20].

In this section, we show that in the MRI problem, all the ℓ^2 components of the $\ell \times \ell$ matrix $\mathbf{S}$ can be computed and stored efficiently (which means, not element-by-element) by performing a single N-point bidimensional DFT (note that N is the number of pixels in the image to be reconstructed, whereas $\ell < N$ is the number of measurements taken in the k-space domain). Given a row n_1 and a column n_2, the corresponding component $S_{n_1 n_2}$ in $\mathbf{S}$ is mapped into a single component of the N-dimensional DFT, so $S_{n_1 n_2}$ can be promptly determined by computing a DFT and applying the mapping function to n_1, n_2. Furthermore, the mapping function is fixed and can be expressed in a closed-form solution, and it can be lately used in a table look-up basis so no further computation than a DFT is necessary to obtain $S_{n_1 n_2}$.

We also emphasize that storing the look-up table requires less storage memory than storing $\mathbf{S}$ itself, so even when $\mathbf{S}$ cannot be stored directly, its components can be made promptly available in a coded way, using an $N \times 1$ vector $\hat{\mathbf{q}}$ and the look-up table, or simply the vector $\hat{\mathbf{q}}$. In the latter case, recomputing the mapping function between the positions (n_1, n_2) in the matrix $\mathbf{S}$ and the position m in $\hat{\mathbf{q}}$ is necessary when building the matrix $\mathbf{S}$.

In order to introduce the main result on the components of $\mathbf{S}$, we first prove the following Lemma 1, which shows how the bidimensional DFT can be computed as a matrix product. In fact, it supposes that the considered image and its DFT are represented in the vectorized form, and then provides the components of the corresponding transformation matrix. In stating this lemma, as well as the main result after it, we use the following notation.

- $\lfloor x \rfloor$ represents the greatest integer less than or equal to x.
- $\langle x \rangle_R = \mathrm{mod}(x, R)$ represents x modulus R.
- If $\mathbf{X}$ is an image, $X_{n_1 n_2}$ represents its component in row n_1 and column n_2.
- If $\mathbf{x}$ is a vector, x_{n_1} represents its component at location n_1.
- If $\mathbf{X}$ is a $R \times R$ image, its vectorized form is the vector $\mathbf{x}$ such that

$$X_{n_1 n_2} = x_{n_1 + (n_2 - 1)R}.$$

  In this case, note that

$$x_m = X_{n_1 n_2}, \text{ where } n_1 = 1 + \langle m - 1 \rangle_R \text{ and } n_2 = 1 + \left\lfloor \frac{m-1}{R} \right\rfloor.$$

- j represents the imaginary unit, so that $j^2 = -1$.
- $W_R = \exp(-j2\pi/R)$.

Lemma 1. *Let $\mathbf{X}$ be an arbitrary $R \times R$ image, with bidimensional DFT given by $\hat{\mathbf{X}}$. If $\mathbf{x}$ and $\hat{\mathbf{x}}$ are the vectorized forms of $\mathbf{X}$ and $\hat{\mathbf{X}}$, so that*

$$X_{n_1 n_2} = x_{n_1 + (n_2 - 1)R}$$

and

$$\hat{X}_{n_1 n_2} = \hat{x}_{n_1+(n_2-1)R},\ \forall\, n_1, n_2 \in \{1, 2, \ldots, R\},$$

then

$$\hat{\mathbf{x}} = \mathbf{F}\mathbf{x},$$

where $\mathbf{F}$ *is the* $R^2 \times R^2$ *matrix with entries*

$$F_{n_1 n_2} = W_R^{\langle n_1-1\rangle_R \langle n_2-1\rangle_R + \left\lfloor \frac{n_1-1}{R} \right\rfloor \left\lfloor \frac{n_2-1}{R} \right\rfloor},\ \forall\, n_1, n_2 \in \{1, 2, \ldots, R^2\}.$$

Proof. By the definition of the bidimensional DFT,

$$\hat{X}_{k_1 k_2} = \sum_{n_1=1}^{R} \sum_{n_2=1}^{R} X_{n_1 n_2} \exp\left(-j\frac{2\pi}{R}(n_1-1)(k_1-1) - j\frac{2\pi}{R}(n_2-1)(k_2-1) \right),$$

which, using the notation $W_R = \exp\left(-j\frac{2\pi}{R}\right)$, can be rewritten as

$$\hat{X}_{k_1 k_2} = \sum_{n_1=1}^{R} \sum_{n_2=1}^{R} X_{n_1 n_2} W_R^{(n_1-1)(k_1-1)+(n_2-1)(k_2-1)}.$$

Since $\hat{X}_{k_1 k_2} = \hat{x}_{k_1+(k_2-1)R}$ and $X_{n_1 n_2} = x_{n_1+(n_2-1)R}$,

$$\hat{x}_{k_1+(k_2-1)R} = \sum_{n_1=1}^{R} \sum_{n_2=1}^{R} x_{n_1+(n_2-1)R} W_R^{(n_1-1)(k_1-1)+(n_2-1)(k_2-1)}.$$

Now, using the change of variables $k_1 + (k_2-1)R = m_1$ and $n_1 + (n_2-1)R = m_2$, so that

$$\begin{cases} k_1 - 1 = \langle m_1 - 1\rangle_R \\ k_2 - 1 = \left\lfloor \dfrac{m_1-1}{R} \right\rfloor \\ n_1 - 1 = \langle m_2 - 1\rangle_R \\ n_2 - 1 = \left\lfloor \dfrac{m_2-1}{R} \right\rfloor, \end{cases}$$

we obtain

$$\hat{x}_{m_1} = \sum_{m_2=1}^{R^2} x_{m_2} W_R^{\langle m_2-1\rangle_R \langle m_1-1\rangle_R + \left\lfloor \frac{m_2-1}{R} \right\rfloor \left\lfloor \frac{m_1-1}{R} \right\rfloor},$$

which in matrix notation can be written as

$$\hat{\mathbf{x}} = \mathbf{F}\mathbf{x},$$

where

$$F_{n_1 n_2} = W_R^{\langle n_1 - 1 \rangle_R \langle n_2 - 1 \rangle_R + \left\lfloor \frac{n_1 - 1}{R} \right\rfloor \left\lfloor \frac{n_2 - 1}{R} \right\rfloor},$$

as we wished to show. □

Using Lemma 1, we can now prove the following Theorem 6, which summarizes how the components of $\mathbf{S}$ can be obtained after the computation of a single N-point bidimensional DFT.

Theorem 6. *Let*

$$\mathbf{S} = \mathbf{A}\mathbf{Q}\mathbf{A}^H, \tag{7.4}$$

with $\mathbf{Q} = \text{diag}(\mathbf{q})$ *a* $R^2 \times R^2$ *diagonal matrix and* $\mathbf{A}$ *an* $\ell \times R^2$ *matrix, with* $\ell < R^2$*. Also, consider the* $R^2 \times R^2$ *matrix* $\mathbf{F}$ *such that*

$$\hat{\mathbf{v}} = \mathbf{F}\mathbf{v}$$

is the vectorized form of the bidimensional DFT of an $R \times R$ *image in vectorized form* $\mathbf{v}$*, with the rows of* $\mathbf{A}$ *obtained by selecting* ℓ *rows of* $\mathbf{F}$*, so that* $\mathbf{A}\mathbf{v}$ *is the vector of* ℓ *samples of* $\mathbf{F}\mathbf{v}$*. If the* n_1*-th row of* $\mathbf{A}$ *is the* $\alpha(n_1)$*-th row of* $\mathbf{F}$*, where* $\alpha(\cdot)$ *is an arbitrary function that defines which DFT coefficients of* $\mathbf{F}\mathbf{v}$ *are sampled, then*

$$S_{n_1 n_2} = \hat{q}_m,$$

where $\hat{\mathbf{q}}$ *is the vectorized form of the 2-D DFT of* $\mathbf{q}$ *and*

$$m = 1 + \Big\langle \langle \alpha(n_1) - 1 \rangle_R - \langle \alpha(n_2) - 1 \rangle_R \Big\rangle_R + \left\langle \left\lfloor \frac{\alpha(n_1) - 1}{R} \right\rfloor - \left\lfloor \frac{\alpha(n_2) - 1}{R} \right\rfloor \right\rangle_R R.$$

Proof. We first consider the expression for $S_{n_1n_2}$. From (7.4), and since $\mathbf{Q} = \text{diag}(\mathbf{q})$, we obtain

$$S_{n_1n_2} = \sum_{r=1}^{R^2} A_{n_1r} q_r A^*_{n_2r}. \tag{7.5}$$

Furthermore, since the n_1-th row of $\mathbf{A}$ is the $\alpha(n_1)$-th row of $\mathbf{F}$, and using the closed-form expression for the components of $\mathbf{F}$ given by Theorem 1,

$$A_{n_1r} = F_{\alpha(n_1)r} = W_R^{\left(\langle \alpha(n_1)-1\rangle_R \langle r-1\rangle_R + \left\lfloor \frac{\alpha(n_1)-1}{R} \right\rfloor \left\lfloor \frac{r-1}{R} \right\rfloor\right)},$$
$$A^*_{n_2r} = F^*_{\alpha(n_2)r} = W_R^{-\left(\langle \alpha(n_2)-1\rangle_R \langle r-1\rangle_R + \left\lfloor \frac{\alpha(n_2)-1}{R} \right\rfloor \left\lfloor \frac{r-1}{R} \right\rfloor\right)},$$

so (7.5) becomes

$$S_{n_1n_2} = \sum_{r=1}^{R^2} W_R^{\gamma(\alpha(n_1),r,R)} W_R^{-\gamma(\alpha(n_2),r,R)} q_r, \tag{7.6}$$

where

$$\gamma(\alpha(n_1), r, R) = \langle \alpha(n_1) - 1\rangle_R \langle r-1\rangle_R + \left\lfloor \frac{\alpha(n_1)-1}{R} \right\rfloor \left\lfloor \frac{r-1}{R} \right\rfloor. \tag{7.7}$$

Now, let $\hat{\mathbf{X}}$ be the 2-D DFT of the image $\mathbf{X}$ whose vectorized form is $\mathbf{q}$, meaning

$$X_{n_1n_2} = q_{n_1+(n_2-1)R}, \ \forall i, j \in \{1, 2, \ldots, R\}. \tag{7.8}$$

By definition,

$$\hat{X}_{k_1k_2} = \sum_{n_1=1}^{R} \sum_{n_2=1}^{R} \exp\left\{-j2\pi \left(\frac{(k_1-1)(n_1-1)}{R} + \frac{(k_2-1)(n_2-1)}{R} \right)\right\} X_{n_1n_2}, \tag{7.9}$$

where j is the imaginary unity. Replacing (7.8) into (7.9) gives

$$\hat{X}_{k_1k_2} = \sum_{n_1=1}^{R} \sum_{n_2=1}^{R} \exp\left\{-j2\pi \left(\frac{(k_1-1)(n_1-1)}{R} + \frac{(k_2-1)(n_2-1)}{R} \right)\right\} q_{n_1+(n_2-1)R}. \tag{7.10}$$

Observe here that $\hat{\mathbf{q}}$ is the vectorized form of $\hat{\mathbf{X}}$, meaning

$$\hat{X}_{k_1k_2} = \hat{q}_{k_1+(k_2-1)R}. \tag{7.11}$$

Using the change of variables

$$n_1 + (n_2 - 1)R = r$$

with n_1 and n_2 integers between 1 and R, it follows from the definition of the mod and $\lfloor \cdot \rfloor$ functions that

$$n_1 = 1 + \langle r - 1 \rangle_R$$

and

$$n_2 = 1 + \left\lfloor \frac{r-1}{R} \right\rfloor, \tag{7.12}$$

so we obtain, from (7.10),

$$\hat{X}_{k_1 k_2} = \sum_{r=1}^{R^2} \exp\left\{ -j2\pi \left(\frac{(k_1 - 1)\langle r - 1 \rangle_R}{R} + \frac{(k_2 - 1)\left\lfloor \frac{r-1}{R} \right\rfloor}{R} \right) \right\} q_r,$$

which can be rewritten as

$$\hat{X}_{k_1 k_2} = \sum_{r=1}^{R^2} W_R^{(k_1-1)\langle r-1 \rangle_R} W_R^{(k_2-1)\left\lfloor \frac{r-1}{R} \right\rfloor} q_r. \tag{7.13}$$

We want to show that there is a component $\hat{X}_{k_1 k_2}$ in $\hat{\mathbf{X}}$ equal to any given component of $\mathbf{S}$. By comparing (7.6) and (7.13), we conclude that

$$S_{n_1 n_2} = \hat{X}_{k_1 k_2}$$

if and only if

$$W_R^{\gamma(\alpha(n_1),k,R)} W_R^{-\gamma(\alpha(n_2),k,R)} = W_R^{(k_1-1)\langle r-1 \rangle_R} W_R^{(k_2-1)\left\lfloor \frac{r-1}{R} \right\rfloor}, \text{ for all } r \in \{1, 2, \ldots, R^2\}. \tag{7.14}$$

Note that the condition in (7.14) is equivalent to

$$\gamma(\alpha(n_1), k, R) - \gamma(\alpha(n_2), k, R) = (k_1 - 1)\langle r - 1 \rangle_R + (k_2 - 1)\left\lfloor \frac{r-1}{R} \right\rfloor + \lambda_r R, \tag{7.15}$$

for all $r \in \{1, 2, \ldots, R^2\}$ and where λ_r is an integer that may depend on r.

Since (7.15) must hold for all $1 \leqslant k \leqslant R^2$, it must hold in particular for $r = 2$ and for $r = R + 1$, so

$$\gamma(\alpha(n_1), 2, R) - \gamma(\alpha(n_2), 2, R) = (k_1 - 1) \langle 1 \rangle_R + (k_2 - 1) \left\lfloor \frac{1}{R} \right\rfloor + \lambda_2 R \tag{7.16}$$

and

$$\gamma(\alpha(n_1), R+1, R) - \gamma(\alpha(n_2), R+1, R) = (k_1 - 1) \langle R \rangle_R + (k_2 - 1) \left\lfloor \frac{R}{R} \right\rfloor + \lambda_{R+1} R. \tag{7.17}$$

Now, observe that $\langle 1 \rangle_R = 1$, $\left\lfloor \frac{1}{R} \right\rfloor = 0$, $\langle R \rangle_R = 0$, and $\left\lfloor \frac{R}{R} \right\rfloor = 1$. Also, from (7.7) it follows that $\gamma(\alpha(n_1), 2, R) = \langle \alpha(n_1) - 1 \rangle_R$ and $\gamma(\alpha(n_2), R+1, R) = \left\lfloor \frac{\alpha(n_2)-1}{R} \right\rfloor$, so (7.16) and (7.17) become respectively

$$k_1 = 1 + \langle \alpha(n_1) - 1 \rangle_R - \langle \alpha(n_2) - 1 \rangle_R - \lambda_2 R \tag{7.18}$$

and

$$k_2 = 1 + \left\lfloor \frac{\alpha(n_1) - 1}{R} \right\rfloor - \left\lfloor \frac{\alpha(n_2) - 1}{R} \right\rfloor - \lambda_{R+1} R. \tag{7.19}$$

Observe still that (7.18) and (7.19) are satisfied respectively for a single integer value of λ_2 and a single integer value of λ_{R+1}, since $1 \leqslant k_1 \leqslant R$ and $1 \leqslant k_2 \leqslant R$. The proof that λ_2 and λ_{R+1} are unique follows from the fact that changing λ_2 or λ_{R+1} modifies the expressions in (7.18) and (7.19) by a multiple of R, thus not allowing more than one solution, to each of these expressions, in the range $\{1, 2, \ldots, R\}$. The solutions to (7.18) and (7.19) in the range $k_1, k_2 \in \{1, 2, \ldots R\}$ are given respectively by

$$k_1 = 1 + \left\langle \langle \alpha(n_1) - 1 \rangle_R - \langle \alpha(n_2) - 1 \rangle_R \right\rangle_R \tag{7.20}$$

and

$$k_2 = 1 + \left\langle \left\lfloor \frac{\alpha(n_1) - 1}{R} \right\rfloor - \left\lfloor \frac{\alpha(n_2) - 1}{R} \right\rfloor \right\rangle_R . \tag{7.21}$$

We have then shown that if (7.15) is satisfied for $r = 2$ and $r = R + 1$, then k_1 and k_2 are be given by (7.18) and (7.19), for uniquely defined values of λ_2 and of λ_{R+1}. We must

still show that if (7.18) and (7.19) are satisfied, then (7.15) holds for all $r \in \{1, 2, \ldots, R^2\}$. In fact, using (7.18) and (7.19) we obtain

$$(k_1 - 1)\langle r-1\rangle_R + (k_2-1)\left\lfloor \frac{r-1}{R}\right\rfloor =$$

$$(\langle \alpha(n_1)-1\rangle_R - \langle \alpha(n_2)-1\rangle_R - \lambda_2 R)\langle r-1\rangle_R$$

$$+\left(\left\lfloor\frac{\alpha(n_1)-1}{R}\right\rfloor - \left\lfloor\frac{\alpha(n_2)-1}{R}\right\rfloor - \lambda_{R+1}R\right)\left\lfloor\frac{r-1}{R}\right\rfloor,$$

or

$$\begin{aligned}(k_1-1)\langle r-1\rangle_R + (k_2-1)\left\lfloor\frac{r-1}{R}\right\rfloor &= \gamma(\alpha(n_1),k,R)\\ -\gamma(\alpha(n_2),k,R)&\left(\lambda_2\langle r-1\rangle_R + \lambda_{R+1}\left\lfloor\frac{r-1}{R}\right\rfloor\right),\end{aligned}$$

which is equivalent to (7.15) if we define

$$\lambda_r = \lambda_2\langle r-1\rangle_R + \lambda_{R+1}\left\lfloor\frac{r-1}{R}\right\rfloor.$$

To complete the proof, we observe that the location in $\hat{\mathbf{q}}$ of the component corresponding to $\hat{X}_{k_1k_2}$ is, according to (7.11), given by

$$m = k_1 + (k_2-1)R. \tag{7.22}$$

Using (7.20), (7.21), and (7.22) we then obtain

$$m = 1 + \left\langle \langle \alpha(n_1)-1\rangle_R - \langle\alpha(n_2)-1\rangle_R\right\rangle_R + \left\langle\left\lfloor\frac{\alpha(n_1)-1}{R}\right\rfloor - \left\lfloor\frac{\alpha(n_2)-1}{R}\right\rfloor\right\rangle_R R,$$

which completes the proof. □

7.4 First Experimental Results

In order to test the proposed fMRI method with prior information, we started by building a functional phantom, meaning a sequence of artificial images with changing internal patterns

(to represent body structures in activity or movement) and static structures. Each artificial image was generated by including 10 different ellipses with random dimensions, random locations, and random pixel values inside a fixed round pattern (note that a single static phantom, like each image here considered, is commonly used in testing MR and tomographic imaging algorithms [50], [10]). Furthermore, we also randomized the pixel levels inside the unmoving parts, to represent nonideal conditions during the reconstruction, as the prior information is based only on the locations, rather than the values, of the fixed structures.

Figure 7.8(a) shows a single frame (number 60) of the simulated functional MR sequence, while Figure 7.8(b) shows the corresponding frame that Scheme 3 reconstructed without prior information and using only 18 radial lines in the k-space. With such a low number of lines, the result includes many visible artifacts, as well as blurred parts. On the other hand, Figure 7.8(c) shows the frame reconstructed from the same 18 radial lines, but this time using Scheme 3 with the prior information extracted from the mask in Figure 7.8(d) according to the method in Figure 7.3. In this particular example, the used prior information noticeably reduces the artifacts. It is worth emphasizing that, in this first example, φ_1, φ_2 and φ_3 in Figure 7.3 were automatically determined in such a way as to include all the support locations with intensities above a comparatively small threshold (10^{-4}).

Table 7.1 shows a frame-by-frame comparison of the reconstructions with and without prior information, for the same simulated fMRI of Figure 7.8. The used prior information is again based on applying the method in Figure 7.3 to the mask in Figure 7.8. Observe that, on an average based on 100 frames, the used prior information improves the reconstruction by 26.4 dB, using the proposed Scheme 3. Also, in this example, the behavior was consistent for all the frames.

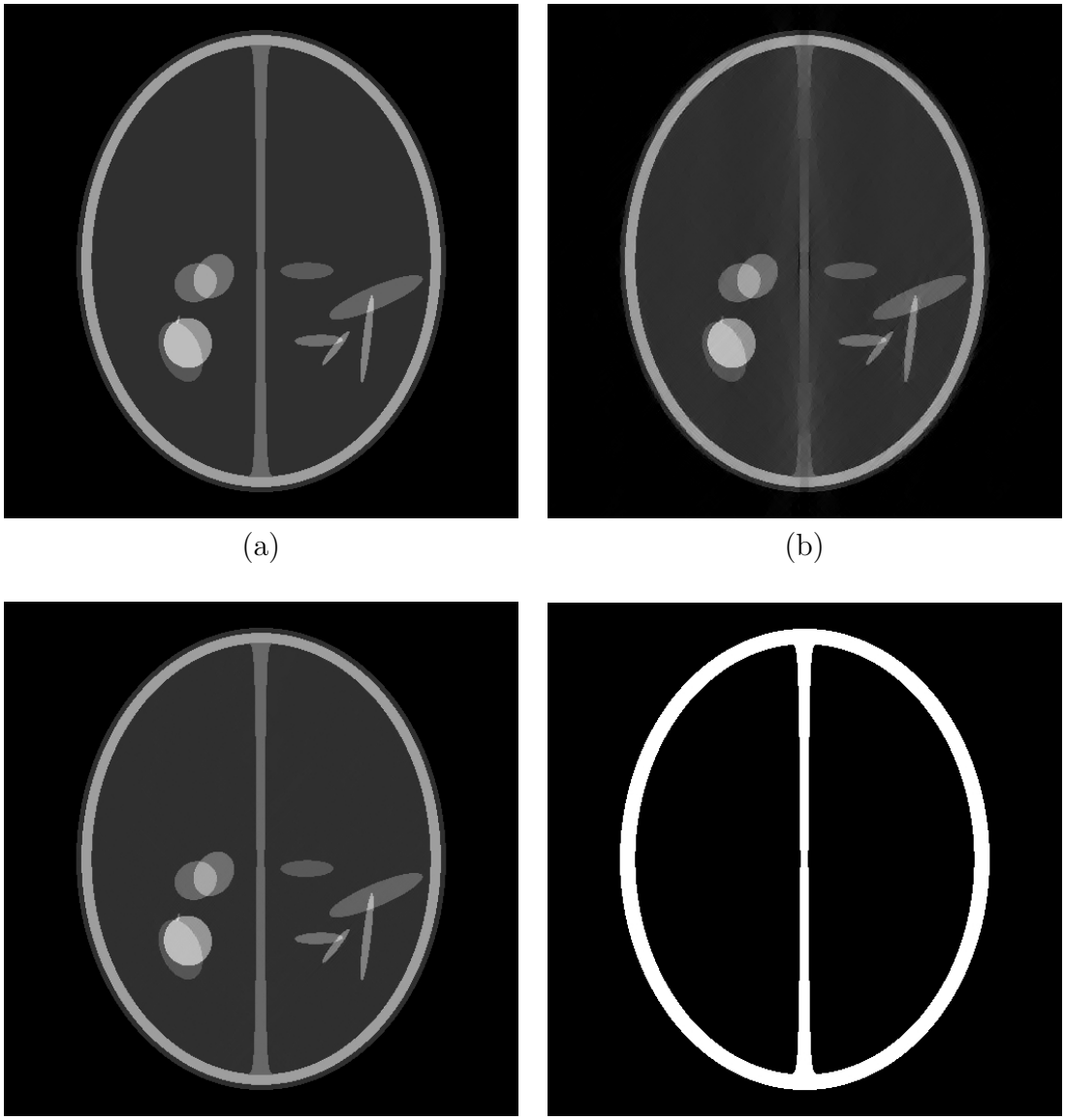


Figure 7.8. A single frame of a simulated functional MRI and reconstructions using 18 radial lines in the k-space domain: (a) reference frame; (b) frame reconstructed using Scheme 3 in Figure 6.4 (no prior information); (c) frame reconstructed using Scheme 3 in Figure 7.7, with the prior information extracted by the method in Figure 7.3; (d) the mask used in the prior information extraction.

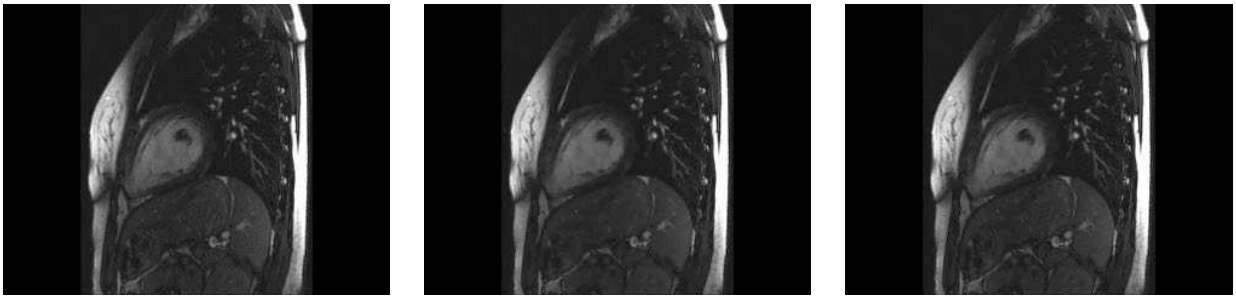
Table 7.1. Signal-to-noise ratios (SNRs) of different frames in the simulated functional MRI of Figure 7.8, reconstructed from 18 radial lines in the k-space. Column SNR_{dB} opi corresponds to SNRs (dB) without prior information using Scheme 3 in Figure 6.4, and column SNR_{dB} wpi corresponds to SNRs with prior information using Scheme 3 in Figure 7.7. The improvement by using prior information is at least 20.4 dB, as indicated in the rightmost column.

Frame number	SNR_{dB} opi	SNR_{dB} wpi	SNR_{dB} improvement
1	21.6	48.9	27.2
5	21.4	49.2	27.8
10	20.9	48.9	28.0
15	20.9	45.5	24.6
20	21.1	42.3	21.2
25	21.5	47.4	25.9
30	21.1	41.5	20.4
35	21.3	51.1	29.8
40	21.7	49.1	27.4
45	21.5	48.6	27.1
50	21.3	49.1	27.8
55	20.7	52.2	31.4
60	21.0	48.3	27.3
65	20.7	48.6	27.9
70	20.9	42.6	21.7
75	20.7	50.3	29.6
80	21.6	48.0	26.4
85	21.2	51.3	30.1
90	21.0	48.2	27.2
95	21.3	45.6	24.3
100	21.4	52.6	31.2
Average for 100 frames	21.1	47.4	26.4

Another example of an fMRI using prior information appears in Figure 7.9, with a reference 76,800-pixel frame compared to the images reconstructed respectively from 60 and 80 radial lines in the k-space. In this case, the prior information is determined for a prespecified value of known locations in the support region ($\varphi = 9000$), and according to Figure 7.1. The reconstruction with prior information then follows Scheme 1 in Figure 7.5.

Regarding the SNRs for this cardiac fMRI example, Table 7.2 compares frame-by-frame the reconstruction using Scheme 1 without prior information and with the prior information extracted according to Figure 7.1. Note that there is an improvement of 4.4 dB by using prior information, and not taking into account the improvement by using Scheme 1 over the TV minimization approach. A similar comparison appears in Table 7.3, also for Scheme 1 with and without prior information, but this time using $r = 80$ radial lines in the k-space, instead of only $r = 60$ radial lines. Note that, as expected, increasing the number of radial lines improves the SNRs, for all frames, but still there is an improvement, frame-by-frame, by using prior information. In fact, in this particular example, the improvement due to prior information is even higher in the case of a larger number of radial lines; a possible reason is that with more radial lines, the prior information extracted from a sequence of frames is also more reliable, so there is less propagation of error when using this prior information to reconstruct a following frame. For more details on this issue, see Chapter 8, Section 8.3.

In Chapter 8, we also compare the resulting SNRs for the method with prior information and as a function of the number of support locations extracted using the schemes in Figures 7.1, 7.2 and 7.3. In a real fMRI, as is the case of Figure 7.9, the expected behavior is that, as φ_1, φ_2 and φ_3 are increased, the resulting SNRs should at first also increase, since more valid prior locations are used in the reconstruction. However, even higher numbers of locations extracted in an automatic manner can result in wrong locations being used in the reconstruction, due to nonideal conditions such as noise and the inclusion of very low



(a) (b) (c)

Figure 7.9. (a) A single 76,800-pixel reference frame of a real cardiac functional MRI and the frame reconstructed from (b) 60 radial lines in the k-space and (c) 80 radial lines in the k-space. The reconstructions use Scheme 1 of Figure 7.5 with prior information given by the method in Figure 7.3. In this example, the reconstruction uses and 9000 elements in the prior information set ($\varphi = 9000$).

coefficients that do not represent the actual image's support. This type of experiment can determine the ideal number of known locations to be used in a particular practical situation, as determined by the type of tissues and organs to be analyzed and by the equipment calibration.

Table 7.2. Signal-to-noise ratios (SNRs) of different frames in the cardiac functional MRI of Figure 7.9, reconstructed from 60 radial lines in the k-space. Column SNR_{dB} opi corresponds to the SNRs (dB) without prior information using Scheme 1 in Figure 6.2, and column SNR_{dB} wpi corresponds to the SNRs with prior information using Scheme 1 in Figure 7.5. The improvement by using prior information is at least 3.5 dB, as indicated on the rightmost column. The reconstruction used $\varphi = 9000$ when extracting the prior information according to Figure 7.1.

Frame number	SNR_{dB} opi	SNR_{dB} wpi	SNR_{dB} improvement
1	34.1	39.5	5.4
5	34.3	39.3	5.0
10	34.4	39.2	4.8
15	34.7	39.2	4.5
20	34.8	39.0	4.2
25	34.5	39.3	4.8
30	34.7	39.1	4.4
35	34.5	39.0	4.5
40	34.8	39.2	4.5
45	34.3	39.2	4.9
50	34.7	39.4	4.7
55	34.6	39.1	4.6
60	34.8	38.9	4.1
65	35.3	39.0	3.8
70	35.4	39.2	3.8
75	35.6	39.0	3.5
80	35.4	39.5	4.1
85	35.3	39.2	3.9
89	35.3	39.3	3.9
Average for 89 frames	34.8	39.2	4.4

Table 7.3. Signal-to-noise ratios (SNRs) of different frames in the cardiac functional MRI of Figure 7.9, reconstructed from 80 radial lines in the k-space. Column SNR_{dB} opi corresponds to the SNRs (dB) without prior information using Scheme 1 in Figure 6.2, and column SNR_{dB} wpi corresponds to the SNRs with prior information using Scheme 1 in Figure 7.5. The improvement by using prior information is at least 3.5 dB, as indicated on the rightmost column. The reconstruction used $\varphi = 9000$ when extracting the prior information according to Figure 7.1.

Frame number	SNR_{dB} opi	SNR_{dB} wpi	SNR_{dB} improvement
1	36.2	42.1	5.9
5	36.5	42.3	5.8
10	36.1	42.9	6.8
15	36.4	42.8	6.5
20	36.0	42.4	6.4
25	35.6	41.9	6.3
30	36.1	41.6	5.5
35	35.8	41.7	6.0
40	35.8	41.8	6.0
45	35.9	41.7	5.8
50	35.9	41.9	6.1
55	36.4	42.3	5.9
60	36.4	41.7	5.3
65	37.1	41.3	4.1
70	36.6	41.6	4.9
75	37.0	41.5	4.4
80	37.0	41.7	4.7
85	36.9	41.8	4.9
89	36.4	41.3	4.9
Average for 89 frames	36.3	41.9	5.6

8

Additional Experimental Results

In the previous chapters, when we presented the proposed theories and methods, we also described and analyzed the first experiments and results, as they were required for the presentation of the following material.

The present chapter describes additional examples and evaluations. It also describes experiments that test detailed aspects of the proposed methods and algorithms. First, Section 8.1 considers the idea of compressive sensing with prior information, introduced in Chapter 4. Starting with additional examples of theoretical and empirical probabilities of reconstruction by ℓ_1-minimization, for signals with different lengths N and sparsity η, it comments on the effect of N to the theoretical model. Additionally, it verifies the impact of prior information in terms of the reduction of the number of measurements required to attain a certain reconstruction probability. This reduction is then compared to that of the theoretical minimum bound for the number of required measurements, which corresponds to reconstruction by ℓ_0-minimization and was introduced in Theorem 3.

Next, Section 8.1.1 evaluates additional aspects of the proposed IRLS method with prior information. It starts by considering the signal-to-noise ratios (SNRs) associated to the reconstruction, as a function of the number of measurements taken and of the number of

known locations in the sparse representation's support. As previously mentioned, the SNR is an important measure of image quality in MRI applications, and thus the importance of measuring the algorithm's performance in terms of this figure of merit. Next, we evaluate the effect of wrong prior information to the number of correct reconstructions, up to a prespecified tolerance. The corresponding tests are important to the assessment of the algorithm's robustness to wrong prior data combined with correct support locations. Still with regard to the IRLS method, Section 8.1.2 evaluates its robustness to changes in the parameters that control the weighting of the available prior information. In fact, this novel algorithm introduces a parameter τ that, associated with the value of p in the ℓ_p-minimization, allows the incorporation of prior information into the iteratively reweighted least squares approach. It is then important to evaluate how the algorithm behaves as the new parameters change.

Following, Section 8.2 tests the proposed methods for MR imaging, introduced in Chapter 6, with an additional set of images. In particular, it compares different compromises between final image quality and reconstruction times, for different types of images.

Finally, Section 8.3 provides further experimentations and discussions related to the proposed MR methods with prior information. Because an automatic method for extracting prior information, in a sequence of MRIs, is always prone to computing some wrong locations, we first evaluate the effect of wrong prior information in function MRI. This effect is compared to that of the basic IRLS algorithm, as evaluated for one-dimensional domain signals in Section 8.1.1. Furthermore, the last section evaluates, for a cardiac fMRI and for several MRIs, the final image qualities and the reconstruction times as a function of the number of prior locations extract automatically.

8.1 Compressive Sensing with Prior Information

In Chapter 4, we have proved that prior information about a sparse representation's support increases the lower bound for the probability of reconstruction by ℓ_1-minimization. Also, we obtained a closed form expression for the reconstruction probability, for the case of random processes that are uncorrelated and symmetrically distributed with respect to the origin.

Using a Monte Carlo simulation, we then obtained empirical values for this reconstruction probability for signals with length $N = 1024$ and sparsity $\eta = 60$, and compared these values with those predicted by the theoretical model.

This section starts by comparing the probabilities of reconstruction for different values φ of known support locations, and considering signals with different lengths and a different sparsity. We then analyze these probability values and those obtained in Chapter 4, in order to evaluate the improvement, due to prior information, from a different perspective. Specifically, based on the computed values, we evaluate the reduction in the number of measurements required to attain certain probabilities, when reconstructing by ℓ_1-minimization. This reduction is compared to the theoretical ideal case of the ℓ_0-minimization, for which, as we proved in Theorem 3, prior information about φ support locations reduces by also φ the number of required measurements.

We first consider signals with length $N = 1024$ and sparsity $\eta = 200$ in the DFT domain. By using (4.76) and (4.77), when evaluated the probabilities that the function $\mathbf{p}$ in (4.38) satisfies all the conditions required, according to Theorem 4, for reconstruction by ℓ_1-minimization. In Figure 8.1(a), the solid lines represent the theoretical values obtained from these equations, for different numbers of measurements ℓ and of known locations φ. Note that as φ increases the corresponding curves shift to the left, indicating a smaller number of measurements required to attain the same probability.

Also, to test again the theoretical model, we conducted a series of Monte-Carlo simulations. For which combination of φ and ℓ, we generated 500 Gaussian signals with uncorrelated entries, and tested for which 500 cases if all the conditions of Theorem 4 were verified. According to the Monte-Carlo method, the relative frequency in which this occurs provides an estimator for the desired probabilities [63], [62]. The empirical probabilities thus obtained appear as the isolated dots in Figure 8.1(a). As it was the case of the preliminary tests in Chapter 4, the empirical values match our theoretical model.

We repeated these tests for signals of length $N = 2048$ and sparsity $\eta = 200$. The corresponding results were consistent with the previous case, as shown in Figure 8.1(b). In fact,

for larger signals it is expected that the theoretical values predict the empirical ones with even greater precision, since the proof of (4.76) and (4.77) uses the central limit theorem and thus assumes a sufficiently high value of N (with closer matches as $N \to \infty$). In any case, in our intended main application, that of magnetic resonance imaging, the analyzed signals are images of usually no fewer that $N = 256 \times 256 = 65535$ pixels, so that testing the theoretical model with $N = 1024$ and $N = 2048$ provides a more than sufficient safety range.

By comparing the different curves in Figure 8.1, also note that the number of measurements ℓ required to attain probability 1 decreases with φ; in fact, for any given probability greater than zero, fewer measurements are needed to attain that probability when φ is increased. It is then possible to assess the improvement by using prior information from a different perspective: that of reduction of the number of measurements required to attain a fixed percentage of reconstructions. Table 8.1 and Table 8.2 show this result for different probabilities, and respectively for $N = 1024$ and $N = 2048$.

From the values in Table 8.1 and Table 8.2, we observe that the larger the increase in φ, the more significant the reduction in the number of measurements ℓ required to attain the same reconstruction probability p. Also, for φ equal or greater than 40, in the observed examples, the reduction in the required number of measurements was more than φ.

This compares favorably with the theoretical reduction corresponding to $p = 1$ and to the ideal case of ℓ_0-minimization (instead of the ℓ_1-minimization procedure here considered). In fact, according to Theorem 3, if ℓ_0-minimization is used, then the number of measurements required to attain $p = 1$ reduces by φ, when φ locations are known. In the case of ℓ_1-minimization, and for larger values of φ, we observed larger reductions (on the other hand, when using ℓ_1-minimization, the reconstruction procedure also has to use more than the ideal minimum number of 2η measurements, to guarantee the conditions of Theorem 4). This result was consistent during all the research.

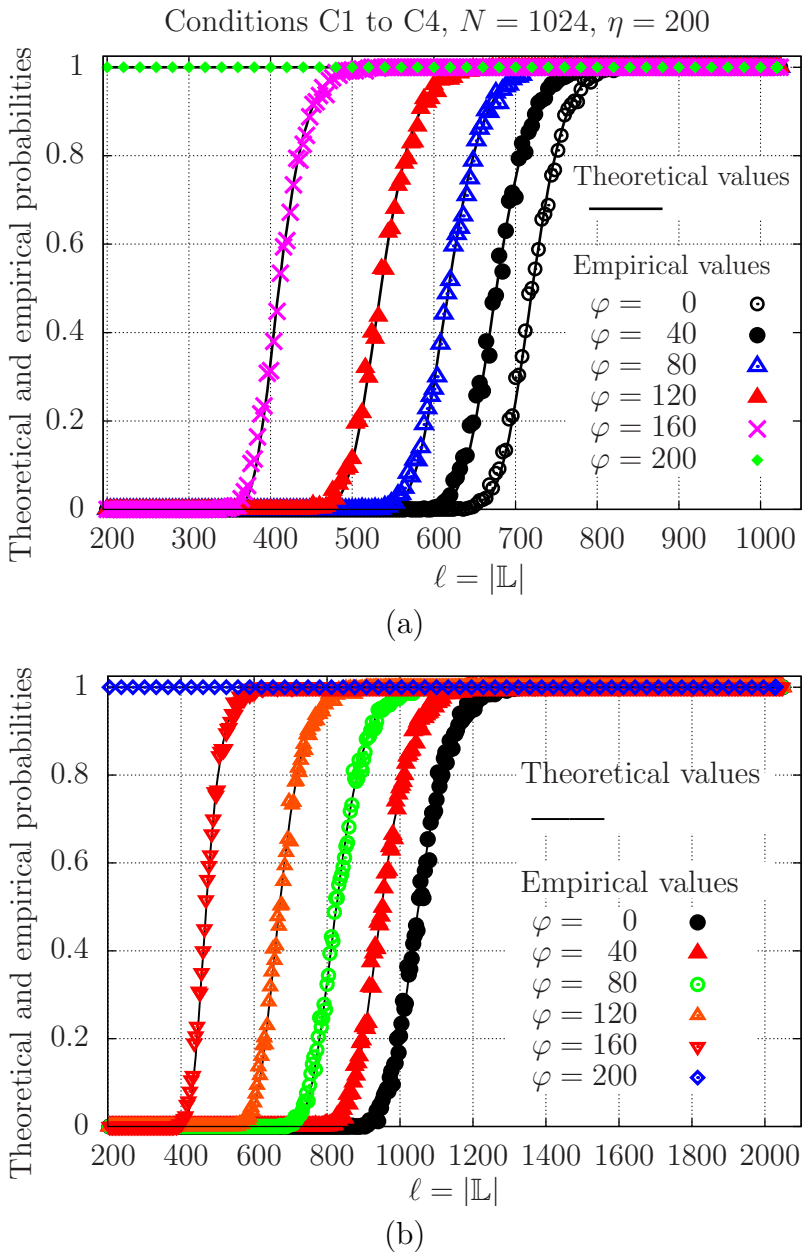


Figure 8.1. Probabilities that the function $\mathbf{p}$ in (4.38) satisfies the conditions C1 to C4 in Theorem 4, for signals with sparsity $\eta = 200$ and with lengths (a) $N = 1024$ and (b) $N = 2048$. The probabilities are evaluated as a function of the number of samples $\ell = |\mathbb{L}|$, and for different numbers of known positions φ in $\mathbb{T}$. The values represented by the solid lines are computed using (4.76) and (4.77) for the different values of φ (theoretical values). For validation, the same probabilities (represented by the individual dots) are also estimated using a Monte-Carlo simulation, by computing the relative frequency in which the conditions happen during 500 trials for each combination of ℓ and φ.

Table 8.1. Number of measurements (ℓ) required to attain different probabilities of reconstruction by ℓ_1-minimization, for different numbers of known locations (φ) in the sparse representation and for signals of length $N = 1024$ and sparsity $\eta = 200$.

Probability of	Required number of measurements (ℓ) for					
reconstruction	$\varphi = 0$	$\varphi = 40$	$\varphi = 80$	$\varphi = 120$	$\varphi = 160$	$\varphi = 200$
0.1	679	636	576	493	379	200
0.2	694	651	586	507	389	200
0.3	703	658	601	519	397	200
0.4	712	667	608	528	405	200
0.5	720	677	618	536	411	200
0.6	729	685	628	543	417	200
0.7	736	695	637	553	425	200
0.8	751	709	649	567	431	200
0.9	766	721	668	581	448	200
1.0	822	778	740	646	500	200

Table 8.2. Number of measurements (ℓ) required to attain different probabilities of reconstruction by ℓ_1-minimization, for different numbers of known locations (φ) in the sparse representation and for signals of length $N = 2048$ and sparsity $\eta = 200$.

Probability of	Required number of measurements (ℓ) for					
reconstruction	$\varphi = 0$	$\varphi = 40$	$\varphi = 80$	$\varphi = 120$	$\varphi = 160$	$\varphi = 200$
0.1	976	880	760	616	436	200
0.2	1000	900	780	632	444	200
0.3	1020	920	796	648	456	200
0.4	1036	936	812	660	464	200
0.5	1052	952	828	672	472	200
0.6	1072	968	844	684	480	200
0.7	1092	988	860	700	492	200
0.8	1116	1012	884	720	504	200
0.9	1152	1048	920	752	524	200
1.0	1684	1616	1512	1344	1028	200

8.1.1 Proposed IRLS method and the effects of correct and wrong prior information

In evaluating the proposed IRLS method with prior information, Chapter 5 focused on the resulting numbers of correct reconstructions, according to a prespecified error tolerance and as a function of the number of known support locations. It also considered the total reconstruction times and numbers of iterations, as they are important measurements of computational complexity. In this chapter, we additionally evaluate three important aspects (the first two are analyzed in this section, while the third one appears in Section 8.1.2):

1. the effect of prior information on the resulting signal-to-noise ratios, when applying the proposed method with prior information, for the same number of measurements;

2. the effect of wrong prior information and

3. the algorithm's robustness to changes in the new parameters, needed to incorporate prior information into the ℓ_p-minimization.

The first two mentioned aspects, considered in this section, are particularly important to the magnetic resonance imaging problem. In fact, and specially in medical applications, the main objective is not only to be able to reduce the number of measurements and still reconstruct the images. The most important aspect is to maximize the images' quality for the number of measurements that can be safely acquired, within a reasonable amount of time, and with the available technology. With this respect, we can assess the impact of prior information, and of the proposed algorithm, by computing the signal-to-noise ratio (an important measure of image quality) as a function of the number of known positions.

Also, when using an automatic procedure to obtain prior information in imaging applications, some of the extracted locations might actually correspond to positions of smaller, rather than higher, values in the sparse representation. In this case, they represent wrong prior information in the proposed approach. It is then important to test the effect of this wrong information combined with correct support locations.

In order to evaluate the effect of the proposed algorithm on the obtained signal-to-noise ratios, we started by adopting an arbitrary, randomly-generated sparsifying transform $\mathbf{T}$, using Givens' decomposition [64]. We then generated 1000 test signals with length $N = 256$ and sparsity $\eta = 16$ in the domain defined by $\mathbf{T}$ (to do so, we selected, for each signal, 16 transformed coefficients chosen randomly with Gaussian distribution, at locations chosen with uniform distribution; the signals in the nonsparse domain, where the measurements are taken, where then computed using the inverse transformation of $\mathbf{T}$). For each signal $\mathbf{x}$, we then took ℓ measurements represented by $\mathbf{b}$ (with ℓ varying from η to N) and applied the proposed method with different numbers of known locations φ in the range from 0 to 16. The reconstructed signals were then compared to the original ones, and the average of the signal-to-noise ratios (SNRs), for each combination of φ and ℓ, was determined over the 1000 test signals. Note that the measurements were taken using $\mathbf{b} = \mathbf{M}\mathbf{x}$, with $\mathbf{M}$ an $\ell \times N$ Gaussian measurement measurement matrix (as this satisfies, with high probability, the restricted isometry property, as discussed in Chapter 3).

Figure 8.2 and Figure 8.3 show the obtained average SNRs, respectively for $p = 1$ and $p = 0.1$ in the ℓ_p-minimization, and for different numbers of known positions φ of measurements ℓ. Note that the behavior of the SNRs with respect to prior information is similar to that previously observed with the percentages of reconstructions. In fact, the obtained SNR values increase with φ, for every considered value of ℓ. Furthermore, as φ increases, the observed increments in the SNRs per added known position become more significant. This behavior was also observed in the proposed magnetic resonance imaging systems, as we have observed in Chapter 6 and will also discuss in other examples in Section 8.2.

In the experiments described in Chapter 5, as well as in Figure 8.2 and Figure 8.3, the given prior information is correct for all given locations. This means that all the components of the given set Φ really belong to the support of the sparse domain, or at least to the set of points that are counted as possibly nonzeros when establishing the signals' sparsity. This could be the case, for instance, of signals that were bandpass filtered, so that the locations corresponding to the passband contain the potentially nonzero coefficients; under

this condition, the set Φ corresponds to the passband.

As stated in Section 5.1, however, it may be the case that the prior information is not perfectly reliable, meaning that some of the components of the considered set Φ belong to the support but others are misplaced, and thus actually associated to null components. In this situation, Algorithm 1 can still reconstruct the underlying signals (the computed components are not constrained to be zero anywhere – even inside the given set Φ), but more measurements may be required compared to the case when no wrong locations are present. In fact, according to (5.9), the elements of Φ are removed from the minimization function, so if some zero components are mistakenly attributed to Φ the local sparsity of those components is not exploited during the reconstruction. Indeed, in our experiments we observed an improvement in performance that depends on the difference between the number of correct positions and wrong positions in Φ. In other words, if c positions in Φ are correct (they belong to the support) and w are wrong (they correspond to zero coefficients), and supposing $c > w$, then the reductions of the number of measurements, computation time,

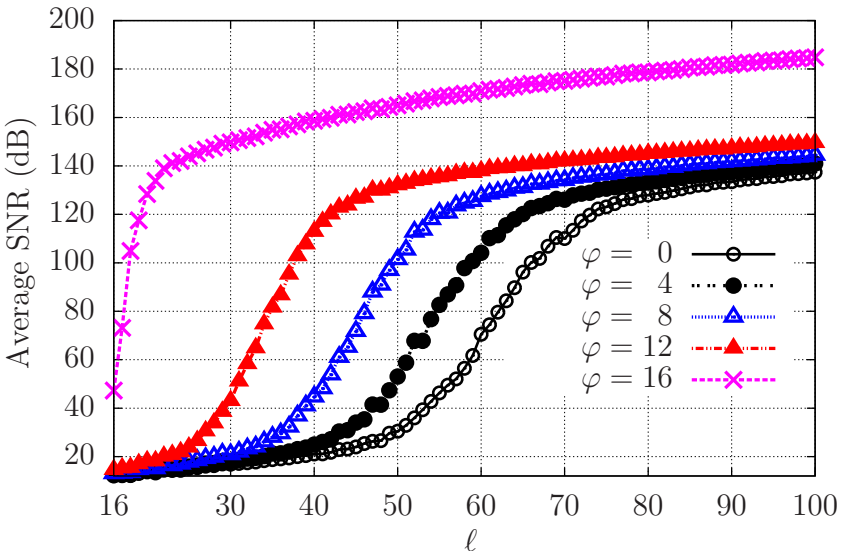


Figure 8.2. Average output SNRs when using the proposed compressive sensing IRLS method with prior information, for $p = 1$ and $\tau^{p-2} = 10^{-3}$. The results are evaluated as a function of the number of samples taken in the nonsparse domain and for different values of known positions (φ) in the sparse representation's support. All 1000 test signals are of length $N = 256$ and sparsity $\eta = 16$ in an arbitrary, randomly determined transformed domain.

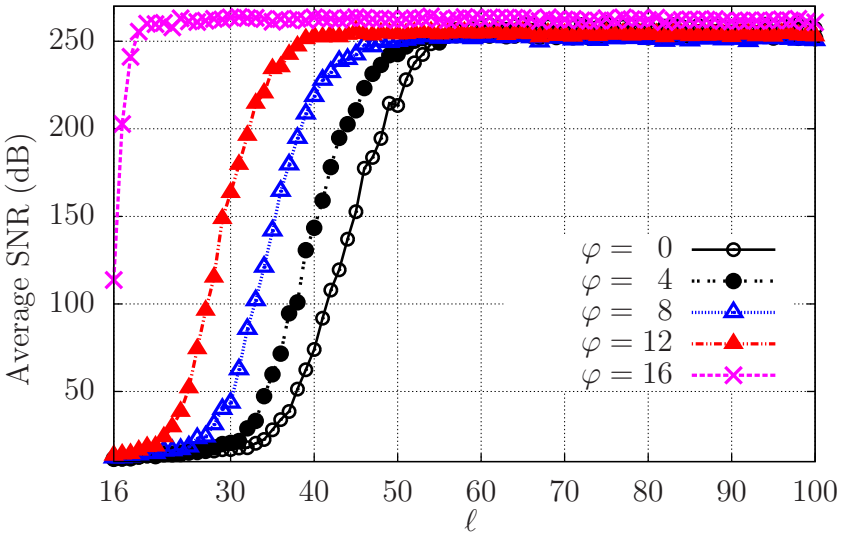


Figure 8.3. Average output SNRs when using the proposed compressive sensing IRLS method with prior information, for $p = 0.1$ and $\tau^{p-2} = 10^{-3}$. The results are evaluated as a function of the number of samples taken in the nonsparse domain and for different values of known positions (φ) in the support region of the sparse domain. All 1000 test signals are of length $N = 256$ and sparsity $\eta = 16$ in an arbitrary, randomly determined transformed domain.

and number of iterations are not worse than in the case where $c - w$ coefficients are correctly assigned to Φ, with no wrong locations (actually, we have observed the first case to lead to better performance).

In Figure 8.4, we exemplify cases in which Φ contains both correct prior information (c locations belonging to the support) and incorrect prior information (w locations that do not belong to the support). Algorithm 1 was applied to 500 signals with length $N = 256$ and sparsity $\eta = 16$, and for different combinations of c and $w < c$. Note that the percentages of correct reconstructions in the cases $(c = 14, w = 2)$, $(c = 11, w = 3)$, $(c = 8, w = 4)$ are, respectively, greater than in the cases $(c = 12, w = 0)$, $(c = 8, w = 0)$, $(c = 4, w = 0)$, suggesting that it is better to have c correct positions and w wrong positions than to have $c - w$ correct positions only. Similarly, in all these conditions the results indicated an improvement with respect to the case with no prior information at all ($c = 0$, $w = 0$), provided that most of the elements in Φ really belong to the support ($c > w$). This result was even more significant in the case of bidimensional-domain signals, as we discuss in Section 8.3.

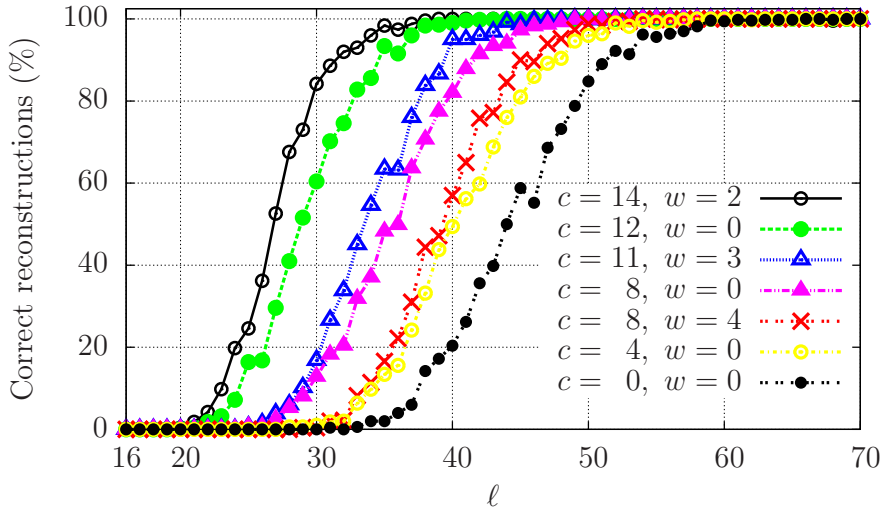


Figure 8.4. Percentages of sparse signals that are correctly reconstructed using Algorithm 1 with partially correct and partially incorrect prior information on the support of the sparse domain. All 500 test signals are of length $N = 256$ and sparsity $\eta = 16$, and ℓ represents the number of linear measurements available for reconstruction. In each line, c is the number of components of the given set Φ that really belong to the support (correct prior information), whereas w is the number of components of Φ that do not belong to the support (wrong prior information).

8.1.2 Robustness to changes in the new parameters in the IRLS method

Finally, it is important to evaluate the sensitivity of the proposed IRLS method with respect to the parameter τ^{p-2}, used in (5.17). Note that τ^{p-2} controls how the prior information is weighted in the iteratively reweighted least squares method. In fact, in this iterative procedure, each iterate component located at a known position is multiplied, after the regularization, by τ^{p-2}, when defining the linear system that will be solved to compute the next iterate. By making τ^{p-2} closer to 1, instead of a smaller value, the effect of prior information is thus reduced.

We conducted a complete set of experiments in which we applied Algorithm 1 with τ^{p-2} ranging from 10^{-12} to 10^2 (although τ^{p-2} should be less than 1, according to Section 5.2, we also included the range $1 \leqslant \tau^{p-2} \leqslant 10^2$ for illustration). These experiments show that a large range of values can be used with equivalent results.

As an example, Figure 8.5 shows the percentages of correct reconstructions (up to the error tolerance defined in Chapter 5), as a function of the tested parameter when applying Algorithm 1 with $p = 0.01$; the remaining parameters were kept as in Figure 5.5. As expected, τ^{p-2} greater than 1 does not allow Algorithm 1 to reconstruct the signals, as shown in the right side of Figure 8.5. On the other hand, too low values of the same parameter also lead to a reduction in the percentage of correct reconstructions, due to the reduced stability of the resulting linear system in (5.14).

On the other hand, Figure 8.6 shows the total number of required iterations and the normalized computation times, as the tested parameter changes. Note that a large range of values of τ^{p-2}, with orders of magnitude between approximately 10^{-6} and 10^{-2}, allows the reconstruction of the same percentage of signals with equivalent computation times and numbers of iterations.

We emphasize that these results refer to the implementation in which a direct method is used to solve the inner linear systems. As we mentioned in Chapter 7, the use of an indirect method leads to a total computation time that is much more sensitive to the choice of τ.

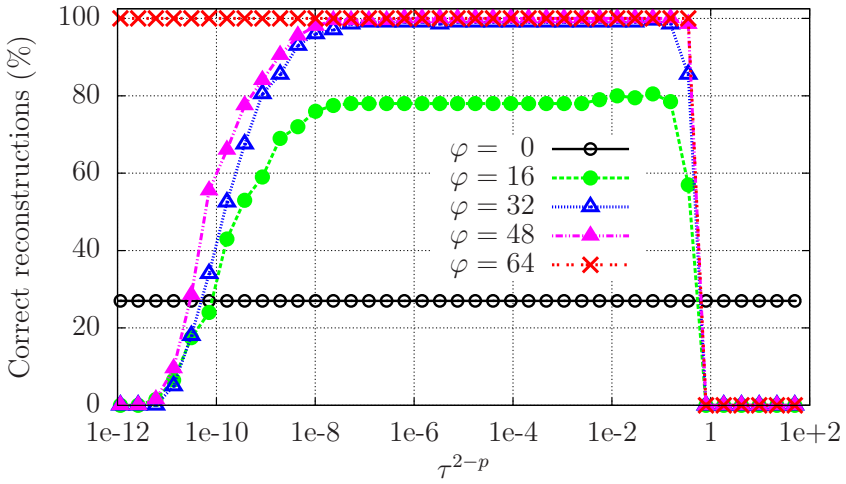
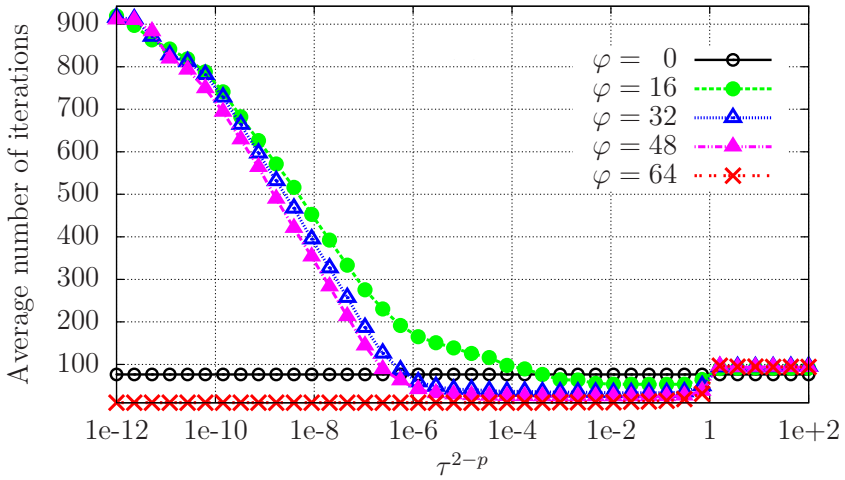
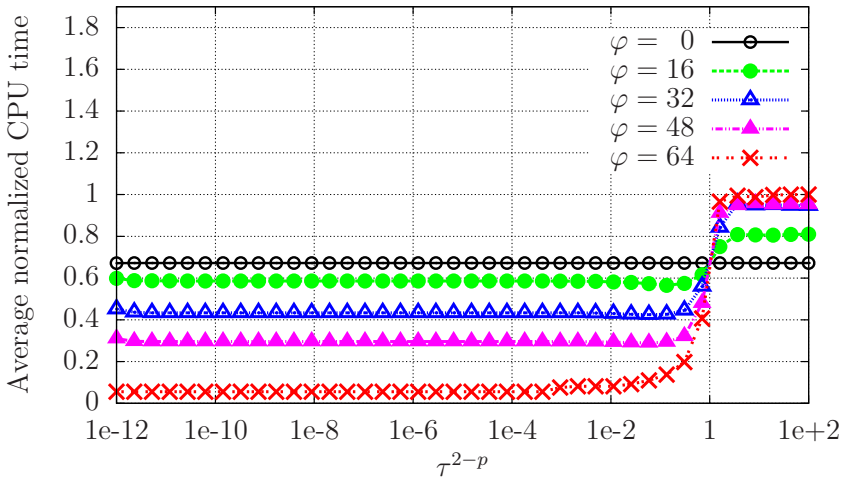


Figure 8.5. Percentages of correct reconstructions when using Algorithm 1, as a function of the used parameter τ^{2-p} and for different values of known positions (φ) in the support region of the sparse domain. All 500 test signals are of length $N = 1024$ and sparsity $\eta = 64$ in an arbitrary, randomly determined transformed domain. In all cases, reconstruction is based on $l = 2.5\eta = 160$ measurements.



(a)



(b)

Figure 8.6. (a) Average number of iterations and (b) average time to convergence when using Algorithm 1, as a function of the used parameter τ^{2-p} and for different values of known locations in the sparse representation's support. All 500 test signals are of length $N = 1024$ and sparsity $\eta = 64$ in an arbitrary, randomly determined transformed domain. In all cases, reconstruction is based on $l = 2.5\eta = 160$ measurements.

Indeed, for smaller values of τ^{p-2}, the linear systems become more unstable and therefore significantly affect the iterative procedures used for solving them in indirect approaches (the instability effect is less noticeable in direct methods). We discuss this issue in more detail in Chapter 9.

8.2 Performance of the Proposed Methods for MR Imaging

In this section, we evaluate the performance of the proposed MR methods when applied to reconstruct a different set of images, other than those of Chapter 6. Here, we show different types of angiograms, as they are usually sparse under finite differencing; thus, the particular choice of filters in Chapter 6 usually provides appropriate sparse representations. In Section 8.3, when evaluating the same methods combined with prior information, we also include other types of MRIs.

We also emphasize that in these tests, we kept constant all the optimization parameters, such as stopping conditions, maximum numbers of stages and error tolerance (see Chapter 5 and Chapter 6 for details). The idea was the evaluate the method's general behavior for different images, rather than to tune it to obtain artificially optimum results for each specific image. However, choosing an appropriate set of parameters for a specific class of MRIs, once a specific application is defined, is an interesting possibility for future work, especially taking into account the different possible compromises between image quality and convergence times, as we will comment on later.

We exemplify here four angiograms. Figure 8.7 shows the Angiogram 1 and the Legs MRA, while Figure 8.8 shows the Brain MRA and the Hand MRI. In each case, we show both a reference image and a single example of reconstruction by the Scheme 3 in Figure 6.4. Note that a zoom of both the reference and the reconstructed images shows that the reconstruction procedure was able to capture fine details even with a reasonably low number of radial lines measured in the k-space, as indicated.

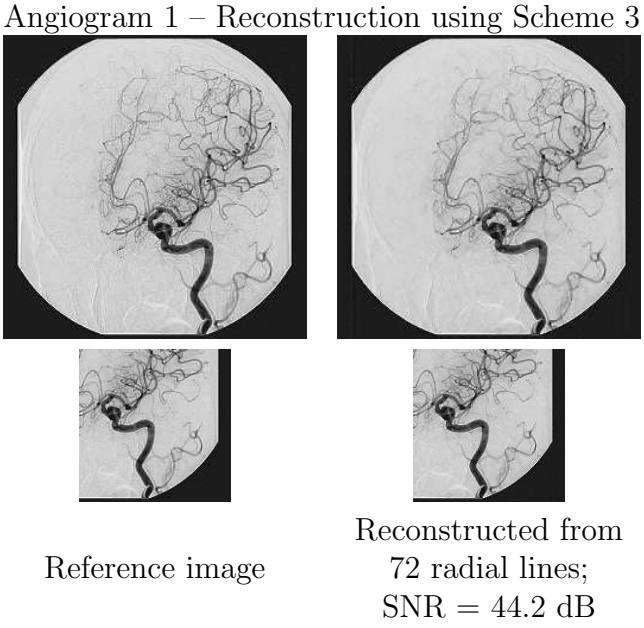


Figure 8.7. Two reference MRIs – (a) Angiogram 1 and (b) Legs MRA – and their reconstructed versions using the proposed Scheme 3, with a detail view.

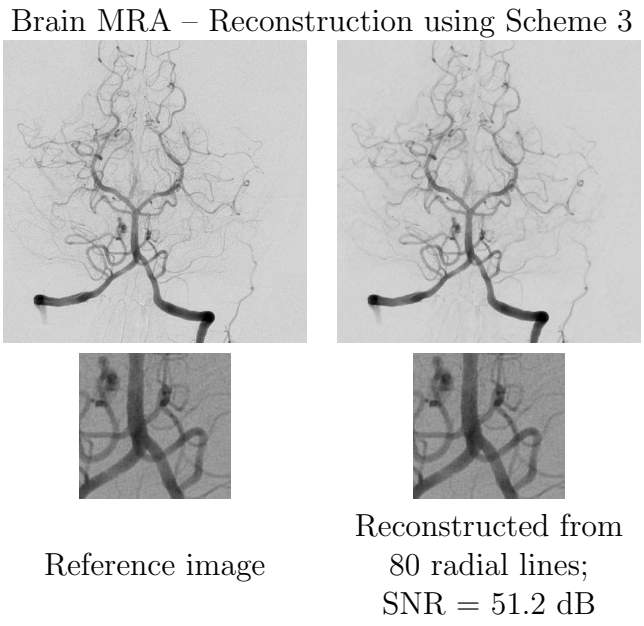


Figure 8.8. Two reference MRIs – (a) Brain MRA and (b) Hand MRA – and their reconstructed versions using the proposed Scheme 3, with a detail view.

In order to compare the proposed method, in the 3 tested schemes, with the standard TV-minimization, we reconstructed the reference images by using a large range of numbers of radial lines, and in each case compared the SNRs and the reconstruction times provided by the four approaches. For the case of the Angiogram 1 in Figure 8.7, for instance, Figure 8.9 and Figure 8.10 show, respectively, the SNRs and the total reconstruction times (in the serial and parallel implementations), as functions of the number of measured radial lines. Note that the results are consistent with those for the set of images in Chapter 6: while the SNRs increase with the number of measurements, they are higher for the proposed schemes (and, in this example, especially for Scheme 3). On the other hand, the proposed method generally provided lower reconstruction times.

Similar results for the Legs MRA appear in Figure 8.11 (SNRs) and in Figure 8.12 (CPU times), while for the Brain MRA they appear in Figure 8.13 (SNRs) and in Figure 8.14 (CPU times). It is interesting to observe, however, that in the case of the Brain MRA the differences between the SNRs provided by the proposed method and those by the TV-minimization approach (Figure 8.13) are lower than in the examples of Figure 8.9 and Figure 8.11 (and also than those of Chapter 6). On the other hand, the reduction in the reconstruction times with the proposed method was higher for the same brain MRA – see Figure 8.14, as compared to Figure 8.10 and Figure 8.12. This suggests that the stopping condition and the error tolerance could be adjusted for the case of the Brain MRA, for instance, so that even higher SNRs could be obtained at the cost of more iterations and not sow low reconstruction times. In other words, it is possible that, by setting the method's parameters for specific types of images, one could obtain different compromises between reconstruction time and output image qualities. The investigation of the choice of parameters according to the desired compromise is an interesting possibility for future related work.

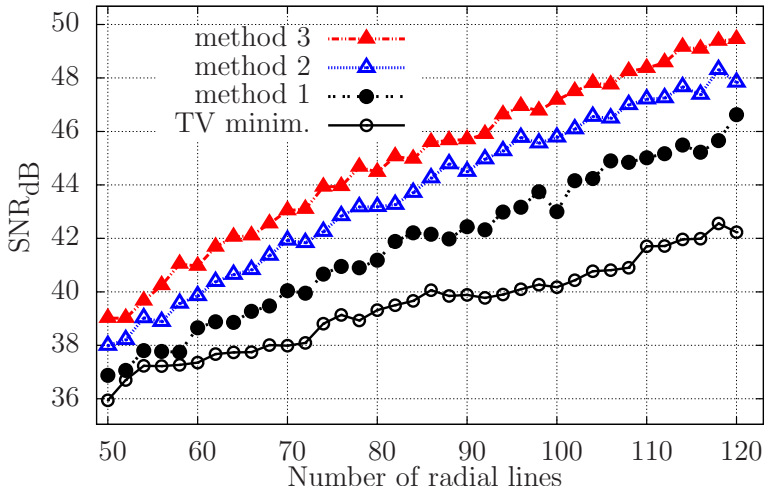


Figure 8.9. Signal-to-noise ratios of the reconstructed MR images corresponding to Angiogram 1, as a function of the number of radial lines measured in the k-space. The results correspond to the proposed method in its three tested forms and to the TV-minimization approach.

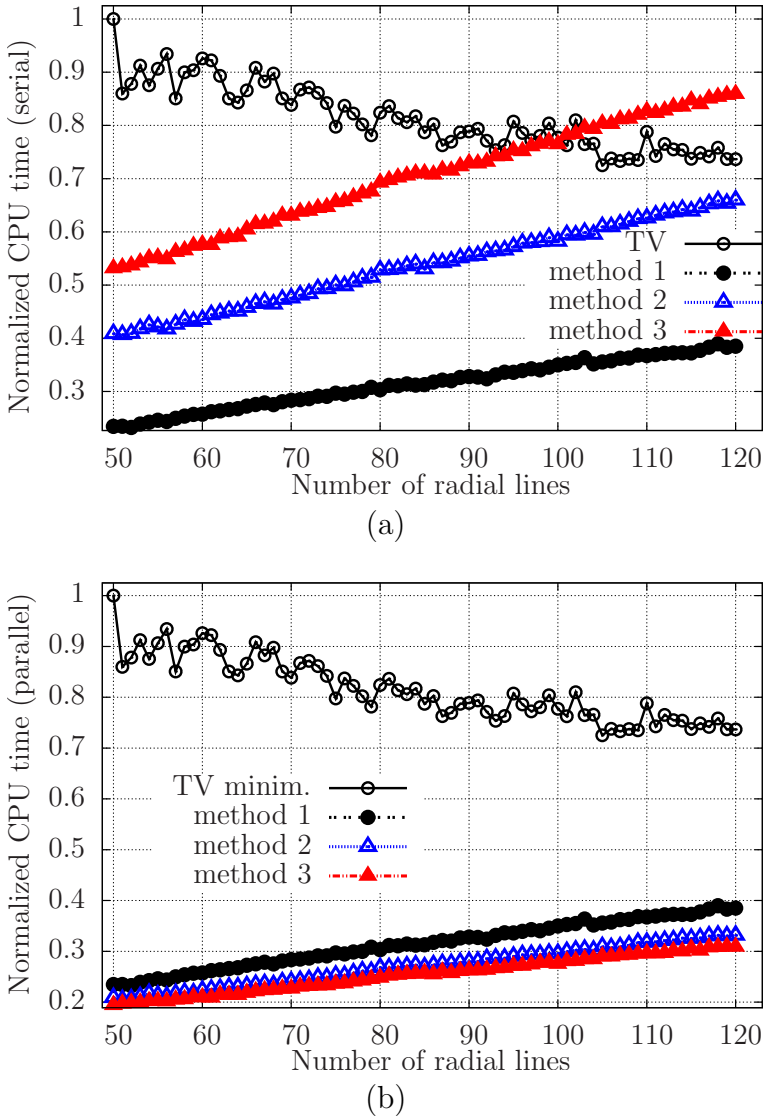


Figure 8.10. Normalized computation times required to reconstruct Angiogram 1, as a function of the number of radial lines measured in the k-space. The results correspond to the TV-minimization approach using a log-barrier algorithm and to the proposed method in its three tested forms, using the (a) serial and (b) the parallel implementation.

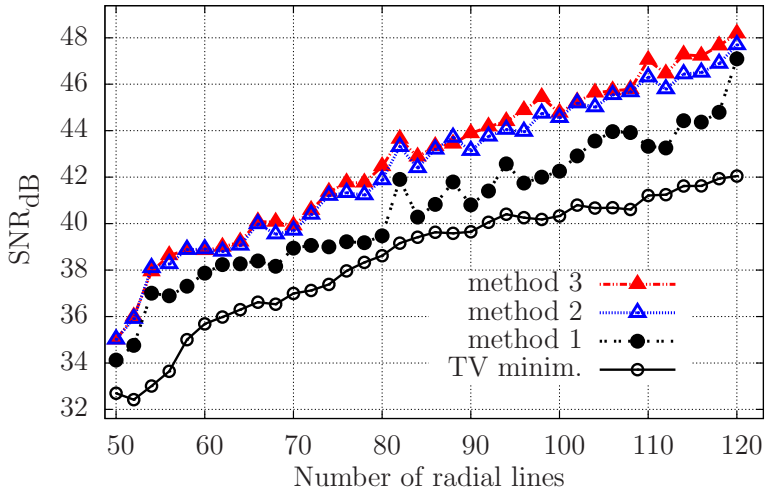


Figure 8.11. Signal-to-noise ratios of the reconstructed MR images corresponding to the Legs MRA, as a function of the number of radial lines measured in the k-space. The results correspond to the proposed method in its three tested forms and to the TV-minimization approach.

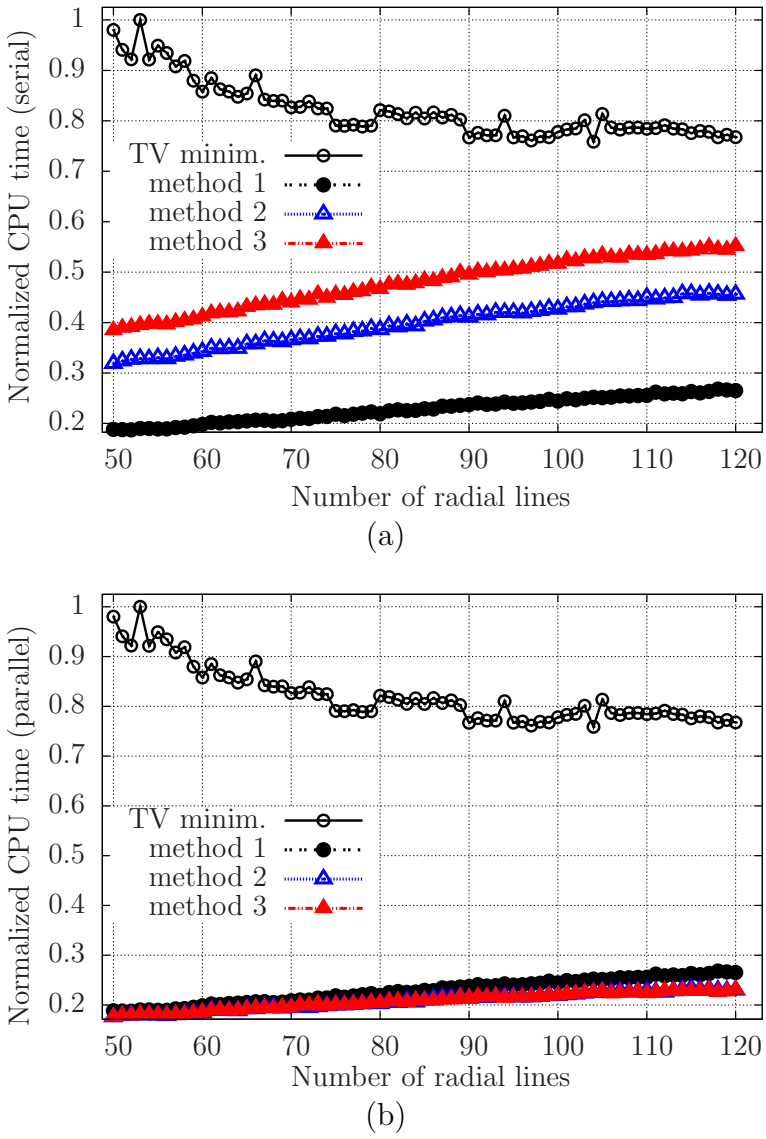


Figure 8.12. Normalized computation times required to reconstruct the Legs MRA, as a function of the number of radial lines measured in the k-space. The results correspond to the TV-minimization approach using a log-barrier algorithm and to the proposed method in its three tested forms, using the (a) serial and (b) the parallel implementation.

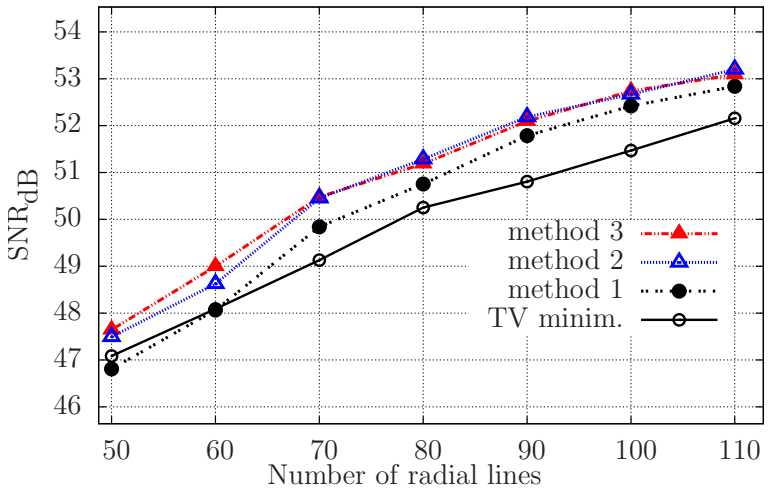


Figure 8.13. Signal-to-noise ratios of the reconstructed MR images corresponding to the Brain MRA, as a function of the number of radial lines measured in the k-space. The results correspond to the proposed method in its three tested forms and to the TV-minimization approach.

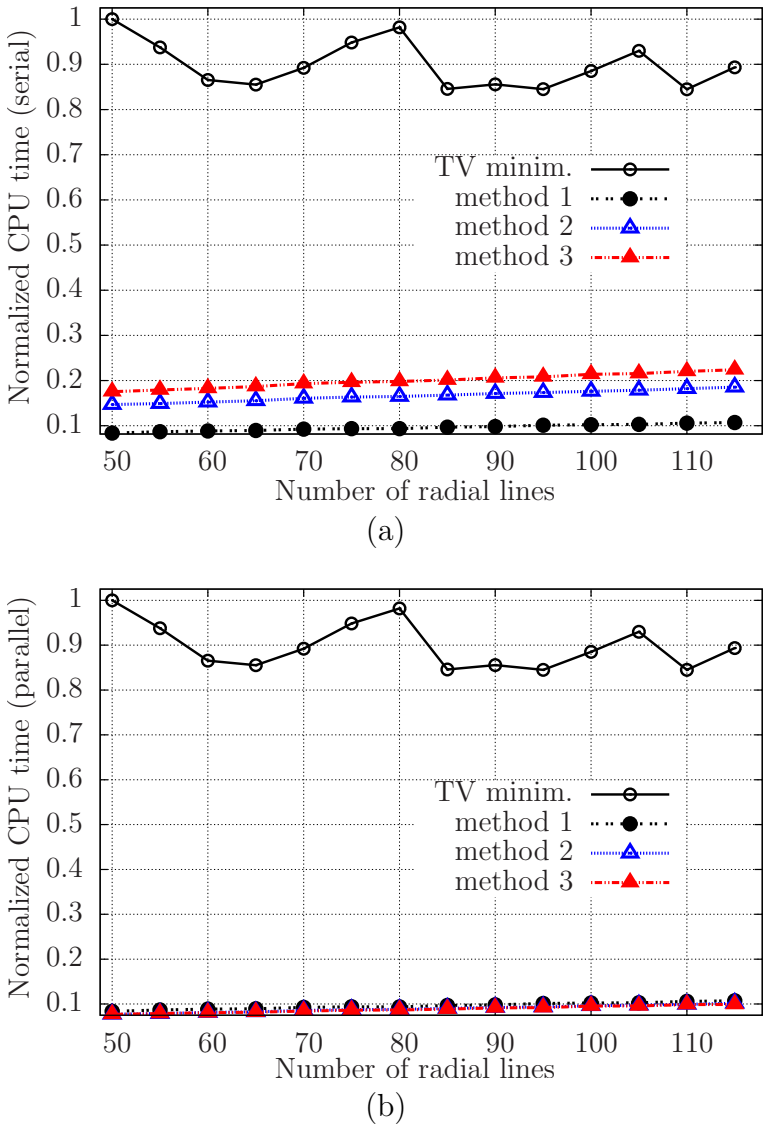


Figure 8.14. Normalized computation times required to reconstruct the Brain MRA, as a function of the number of radial lines measured in the k-space. The results correspond to the TV-minimization approach using a log-barrier algorithm and to the proposed method in its three tested forms, using the (a) serial and (b) the parallel implementation.

8.3 MR Imaging with Prior Information

In this last section, we present some additional results and analyses of experiments that test the propose MRI methods with prior information. Initially, we consider the problem of wrong support locations, specifically in the context of functional MRI. In fact, an automatic procedure for extracting prior information from one frame to the next is always prone to generating wrong support locations, in addition to correct ones. In order to evaluate the system's robustness to this problem, we consider again the simulated functional MRI of Figure 7.8. Note that, as this MRI was generated by defining a fixed image (mask) to which randomly located ellipsoidal patches were added at each frame, we already known all the fixed support locations beforehand. Indeed, we can simply apply, to the fixed mask, one of the sets of filters used in Figure 7.1, Figure 7.2 or Figure 7.3, according to the reconstruction scheme to be used. The positions of the resulting nonzero values are then the correct prior locations to be used in the tests.

For Scheme 1, for instance, we obtained $c = 3794$ prior locations, by applying the filters of Figure 7.1 to the mask in Figure 7.8(d). Besides testing with all the correct $c = 3794$ locations, with also tested with half of them ($c = 1897$) as shown below.

Regarding the wrong prior information, we tested a large range for the number of wrong locations w, which were selected randomly from positions where the filtered frame considered is actually 0.

For each combination of c and w, we then applied the reconstruction Scheme 1 with prior information, shown in Figure 7.5. Figure 8.15 shows the resulting SNRs for frame 60, both for $c = 3794$ and $c = 1897$ correct locations, and as functions of the number of added wrong locations w. It also shows the SNR corresponding to the reconstruction of the same frame without prior information. Additionally, it emphasizes, in the curves for $c = 3794$ and $c = 1897$, the two points where the number of wrong prior locations equals the number of correct ones.

Note, from the two emphasize points in Figure 8.15, that when $w = c$, in the exemplified cases, the SNRs are still higher than without prior information. In fact, the number of prior

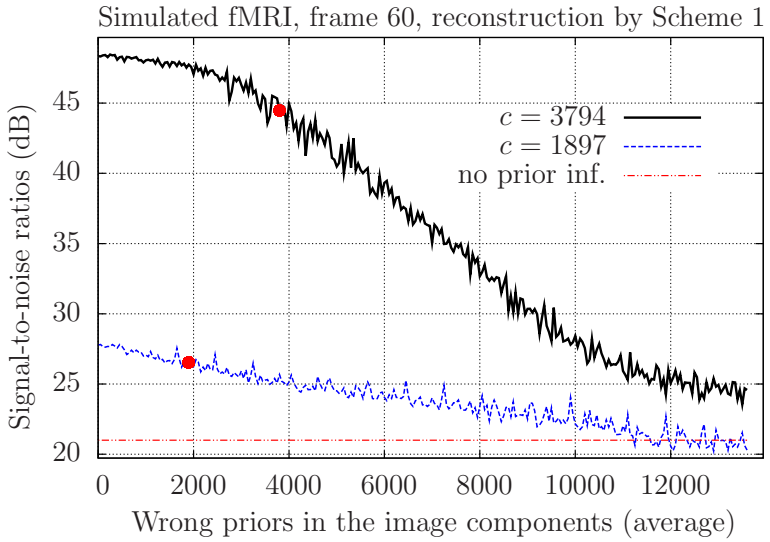


Figure 8.15. Signal-to-noise ratios (SNRs) when reconstructing the simulated fMRI of Figure 7.8, using the prior information about $c = 3794$ and $c = 1897$ correct support locations, and as a function of the number of wrong support locations w. The reconstruction uses the proposed Scheme 1. The bottom curve (red) represents the SNR level when no prior information is used in the reconstruction, while the red dot in each top curve represents the points where the number of wrong locations equals the number of correct locations. Note that when even when $w = c$, the final SNRs are still greater than without prior information.

locations has to be significantly higher than the correct ones before the SNR falls down to the level of no prior information at all. For $c = 1897$, for instance, w has to be higher than approximately 1100 so that using prior information (with wrong locations) performs worse than not using it. Recall that a similar result was also observed in the reconstruction of one-dimensional signals using the proposed IRLS method (see Section 8.1.1), but here we consider the context of the prefiltering strategy, so that both the correct and the wrong locations are obtained by applying the used set of filters to the images' mask.

In an even more realistic context, note also that the mask itself is not known a priori. Therefore, it has to be extracted by analysing a sequence of frames, and determining the regions where the image does not change significantly. To do so, we can apply the method of Figure 7.4 to a real fMRI, as discussed in Chapter 7. The automatically extracted mask is then filtered, as in the case of the simulated fMRI, with the same set of filters used in the reconstruction scheme; the φ highest resulting coefficient (nonzero values) are then used as

prior information.

Note that, in this case, if we take more values than the actual signal's sparsity (which is not known a priori), then wrong positions will also be used during the reconstruction. Wrong positions also appear because of nonexact reconstruction of the frames, which leads to an error propagation as prior information extracted from one frame is used to reconstruct the next one. To evaluate these effects, we applied Scheme 3 with prior information to the cardiac fMRI of Figure 7.9, with φ ranging from 0 to 20% of the total number of pixels in each frame. In Figure 8.16, we show the resulting SNRs, averaged over all the frames, and for $r = 60$ and $r = 80$ radial lines.

Note that, starting with $\varphi = 0$, as φ increases the resulting SNRs also increase. This is the normal effect of prior information, as was tested in the context of one-dimensional signals before and in combination with the proposed MRI systems, both in Section 8.1.1 and in the example of Table 7.3 and Table 7.2. However, note in both curves in Figure 8.16 that, as φ increases above a critical value, the SNRs start to decrease; the reason is the inclusion of

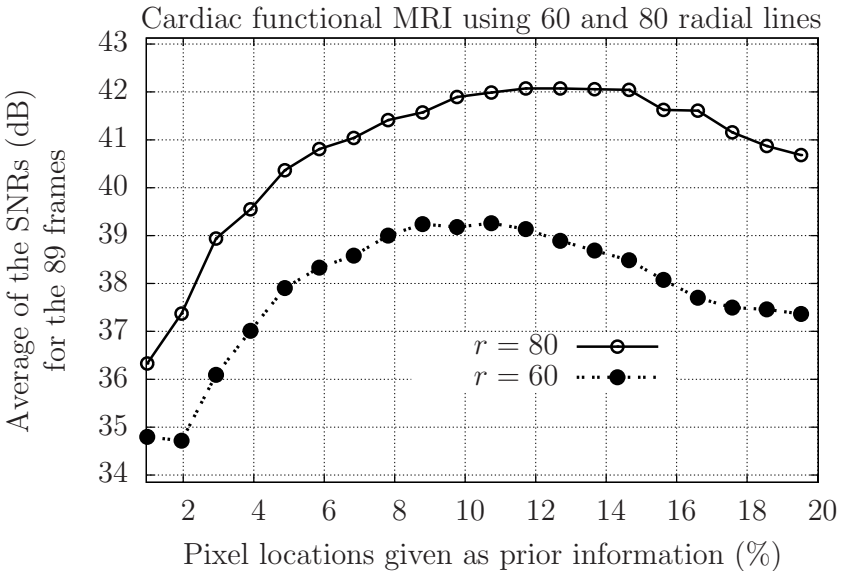


Figure 8.16. Signal-to-noise ratios (SNRs), averaged for 89 frames, when reconstructing the cardiac MRI of Figure 7.9 with prior information, as a function of the number of known support locations φ and for $r = 60$ and $r = 80$ radial lines measured in the k-space. The reconstruction uses the proposed Scheme 3.

wrong prior information when φ is too high, due to the two effects mentioned above. Hence, the same consequence of wrong information observed in Figure 8.4 and in Figure 8.15 also occurs in Figure 8.16 for higher values of φ.

We also emphasize two important additional characteristics in Figure 8.16. First, when φ increases above the critical point of maximum SNR, the observed decrease in the SNRs occurs at a lower rate, compared to the initial increase. The reason is that the wrong prior information has a weaker effect than the correct one, as we observed in Figure 8.15. The second effect is that the critical value of φ is higher when more radial lines are used in the reconstruction. This result was already expected; when more radial lines are available, the reconstruction of each fMRI frame is improved. Therefore, the prior information extracted from each frame is more reliable, resulting in fewer wrong locations, even for a higher value of φ, used to reconstruct the following frames.

Next, we also used the MRI methods with prior information to reconstruct a set of still MR images. The purpose was to evaluate the performance of the reconstruction methods for different kinds of images, and not the performance of the information extraction between frames, as in the previous examples. Hence, for this final set of images we extracted the prior information from the reference images themselves (no error propagation between frames), using the procedures in Figure 7.1, Figure 7.2 and Figure 7.3 with the mask being the set of all pixels in the images. We then measured the performances for different values of φ.

The tested images appear in Figure 8.17, Figure 8.18 and Figure 8.19. Observe that there are two additional angiograms (Head A MRA and Head B MRA) as well as 5 joint MRIs (Ankle A, Ankle B, Ankle C, Ankle D and Ankle E). The three figures show the reference images and the images reconstructed by Scheme 3 without and with prior information. As indicated, the SNRs also increased, in these examples, due to the added prior information.

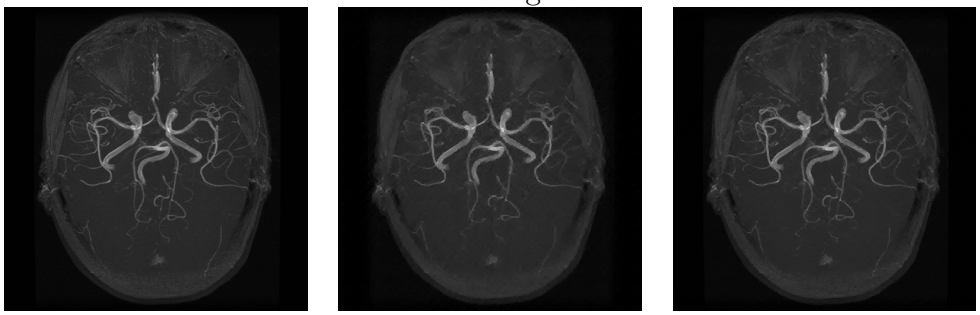
In Figure 8.20, on the other hand, we compare the performance of Scheme 3 for different numbers of radial lines, and as a function of the number of known locations φ extracted using the method in Figure 7.3; the reference image is the Head A MRA of Figure 8.17. First, Figure 8.20(a) shows the resulting SNRs; the found behavior is similar to that obtained for

the functional fMRIs, as shown in Figure 8.16: as φ increases, the SNRs also increase up to a critical point, after which wrong prior information starts to appear and thus the SNRs decrease at a lower rate (compared to the initial increase rate).

The behavior of the reconstruction times in Figure 8.20(b), however, differs from that observed in the case of the one-dimensional signals – see Figure 5.2 and Figure 5.4, for instance. While in the case of Figure 5.2 and Figure 5.4 an increase in φ always resulted in lower reconstruction times, in the case of Figure 8.20(b) simply adding prior information (φ changing from 0 to a low positive value) leads to a significant increase in the reconstruction times. The reason, as previously mentioned, is that to cope with the larger-scale linear systems used in the imaging methods, the direct methods of the basic IRLS tests were replaced by iterative solutions (indirect methods) based on the conjugate gradient approach. Therefore, as these linear systems become more unstable with the inclusion of prior information (the weighting strategy involves multiplying the corresponding systems' components by low weights), the convergence of these solutions become slower. In Chapter 9, we discuss a possible way to cope with this problem, based on Theorem 6.

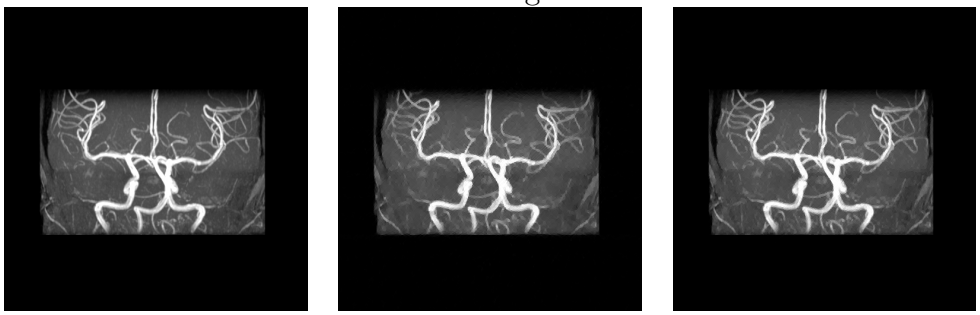
Finally, we also evaluate the performance of the other reconstruction schemes, as compared to Scheme 3, in the presence of prior information. In Figure 8.21 to Figure 8.26, we show the resulting SNRs and reconstruction times for all the tested schemes, as a function of φ and applied to the remaining images in Figure 8.17, Figure 8.18 and Figure 8.19. Note that the behavior of the SNRs and the reconstruction times, described in the example of the Head A MRA, was consistent over all the tested images, and for the three tested schemes.

Head A MRA – Reconstruction using Scheme 3 with 40 radial lines



Reference image | $\varphi = 0$, SNR = 35.5 dB | $\varphi = 60000$, SNR = 38.3 dB

Head B MRA – Reconstruction using Scheme 3 with 40 radial lines



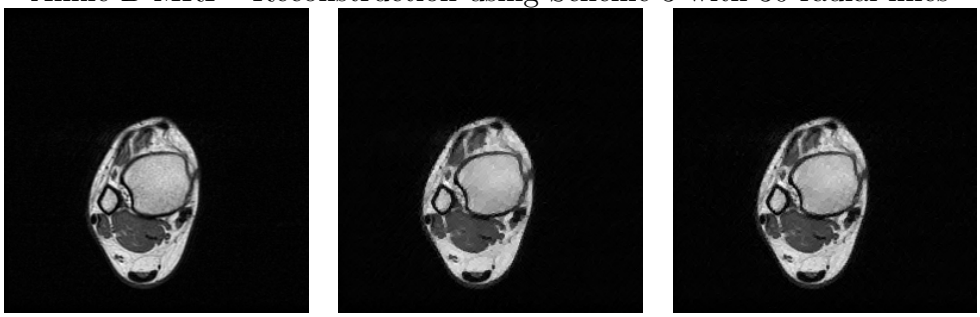
Reference image | $\varphi = 0$, SNR = 35.0 dB | $\varphi = 44000$, SNR = 38.2 dB

Ankle A MRI – Reconstruction using Scheme 3 with 40 radial lines

Reference image | $\varphi = 0$, SNR = 34.0 dB | $\varphi = 20000$, SNR = 37.9 dB

Figure 8.17. Three reference MR images and their versions reconstructed without prior information ($\varphi = 0$) and with partial prior information about the sparse representation's support ($\varphi > 0$). Original reference images provided by the Sierra Medical Center, El Paso, TX.

Ankle B MRI – Reconstruction using Scheme 3 with 30 radial lines



Reference image | $\varphi = 0$, SNR = 33.0 dB | $\varphi = 15000$, SNR = 34.4 dB

Ankle C MRI – Reconstruction using Scheme 3 with 30 radial lines

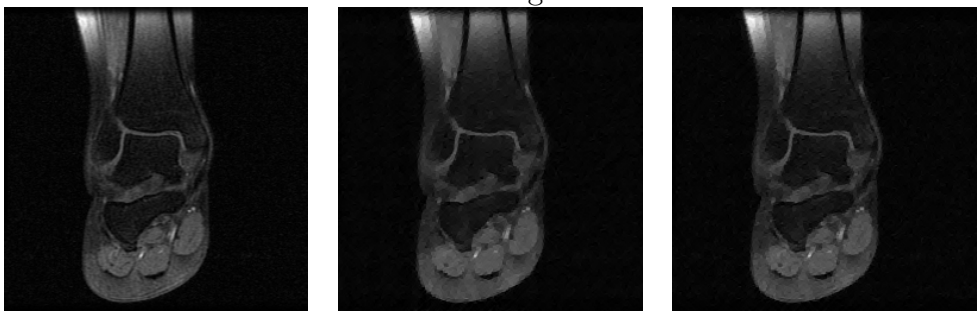
Reference image | $\varphi = 0$, SNR = 29.6 dB | $\varphi = 15000$, SNR = 33.1 dB

Ankle D MRI – Reconstruction using Scheme 3 with 30 radial lines

Reference image | $\varphi = 0$, SNR = 25.0 dB | $\varphi = 24000$, SNR = 31.1 dB

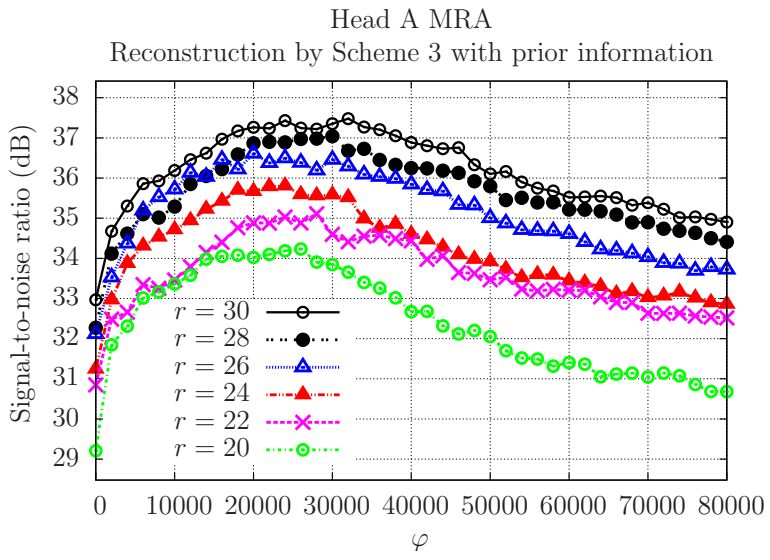
Figure 8.18. Three reference MR images and their versions reconstructed without prior information ($\varphi = 0$) and with partial prior information about the sparse representation's support ($\varphi > 0$). Original reference images provided by the Sierra Medical Center, El Paso, TX.

Ankle E MRI – Reconstruction using Scheme 3 with 30 radial lines

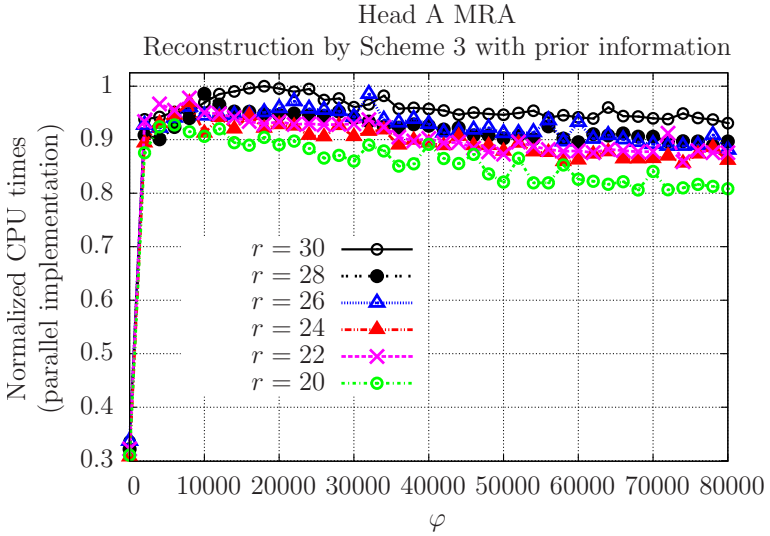


Reference image	$\varphi = 0$ SNR = 29.8 dB	$\varphi = 24000$ SNR = 32.1 dB

Figure 8.19. A reference MR image and its versions reconstructed without prior information ($\varphi = 0$) and with partial prior information about the sparse representation's support ($\varphi > 0$). Original reference images provided by the Sierra Medical Center, El Paso, TX.

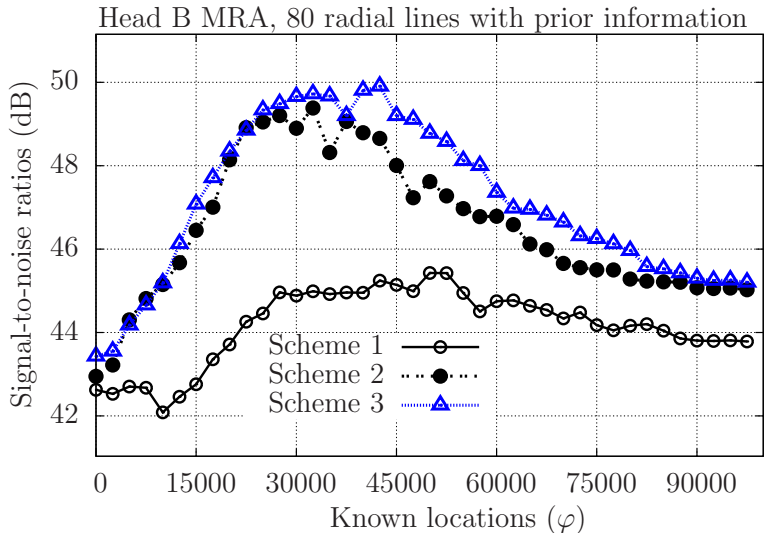


(a)

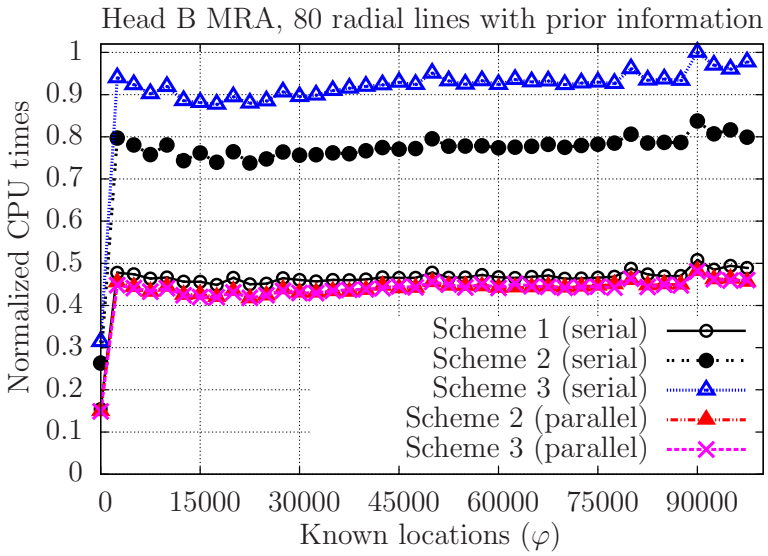


(b)

Figure 8.20. (a) Signal-to-noise ratios and (b) normalized computation times when using the proposed Scheme 3 with prior information to reconstruct the Head A MRA, for different numbers of radial lines r and as a function of the number φ of known support locations. Original reference image provided by the Sierra Medical Center, El Paso, TX.

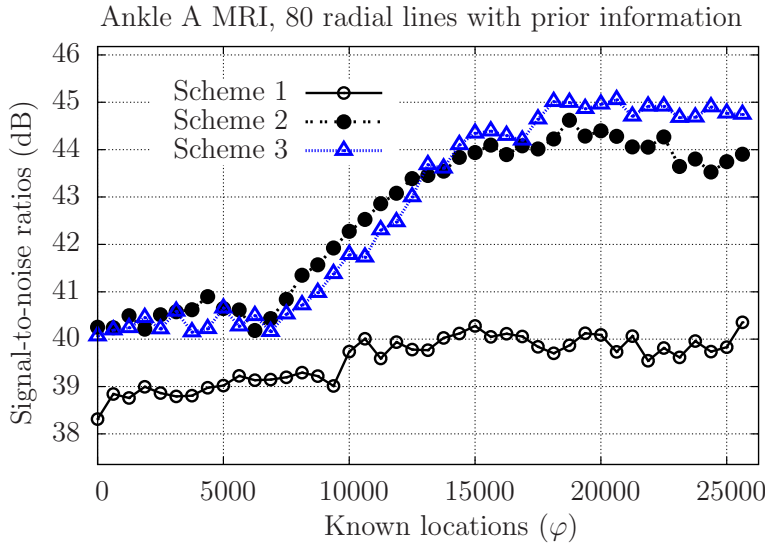


(a)

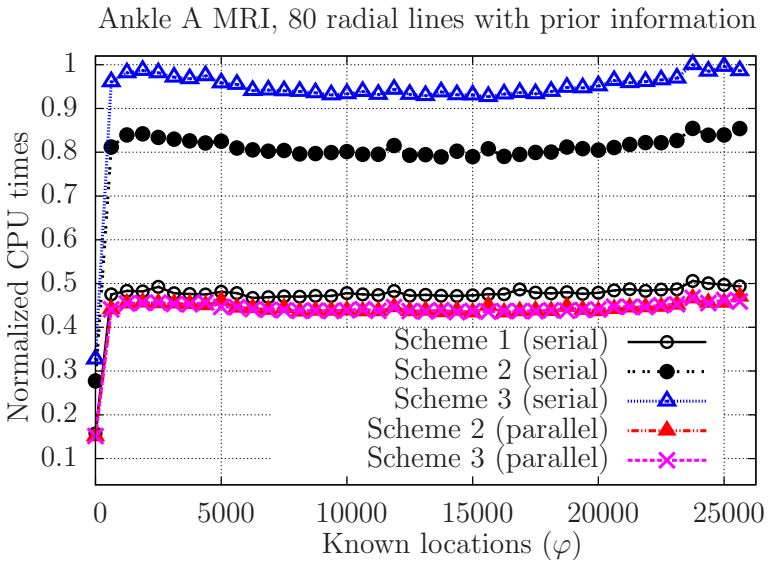


(b)

Figure 8.21. (a) Signal-to-noise ratios and (b) normalized computation times for the proposed MR imaging methods with prior information, when applied to the Head B MRA. Reference image provided by the Sierra Medical Center, El Paso, TX.

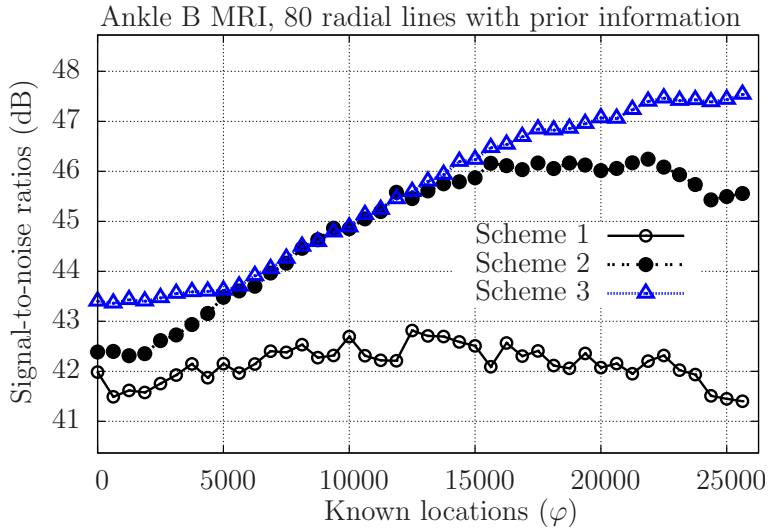


(a)

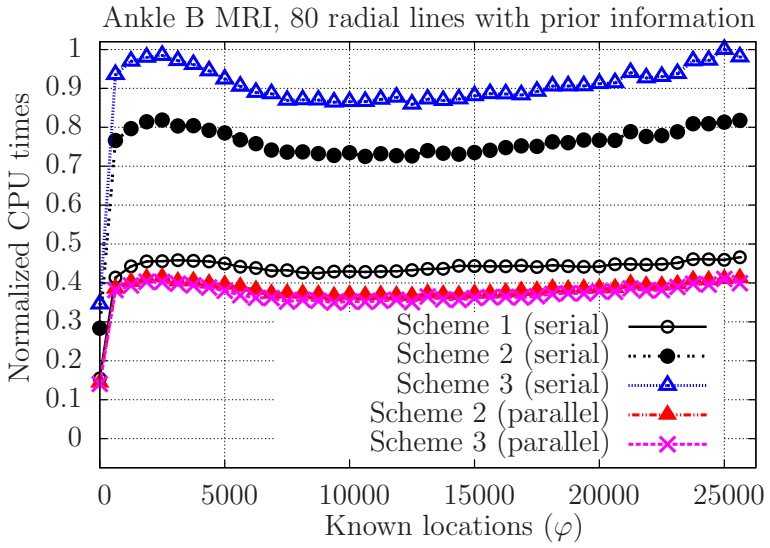


(b)

Figure 8.22. (a) Signal-to-noise ratios and (b) normalized computation times for the proposed MR imaging methods with prior information, when applied to the Ankle A MRI. Reference image provided by the Sierra Medical Center, El Paso, TX.

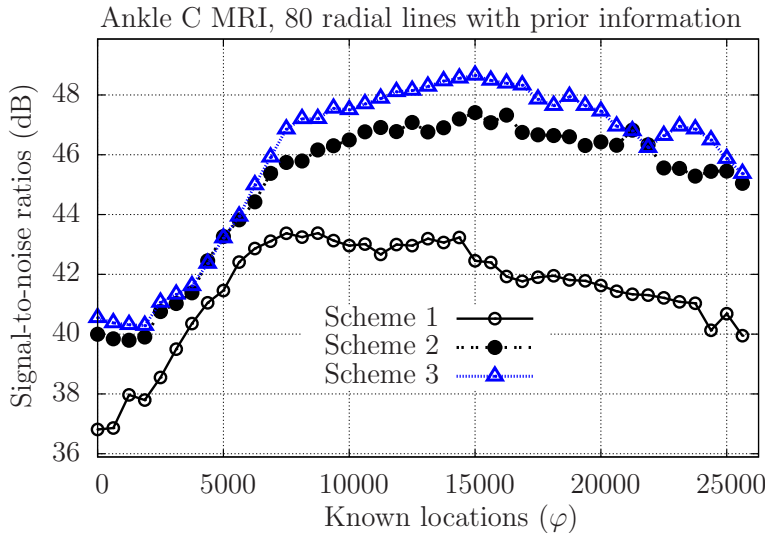


(a)

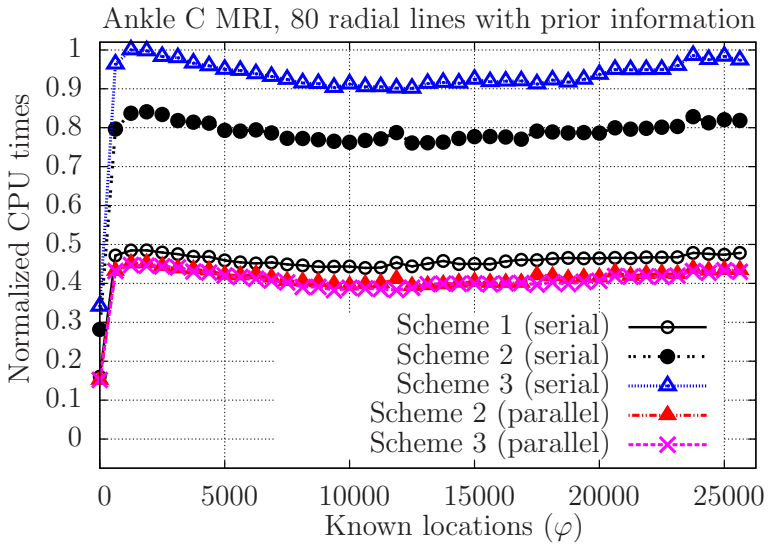


(b)

Figure 8.23. (a) Signal-to-noise ratios and (b) normalized computation times for the proposed MR imaging methods with prior information, when applied to the Ankle B MRI. Reference image provided by the Sierra Medical Center, El Paso, TX.

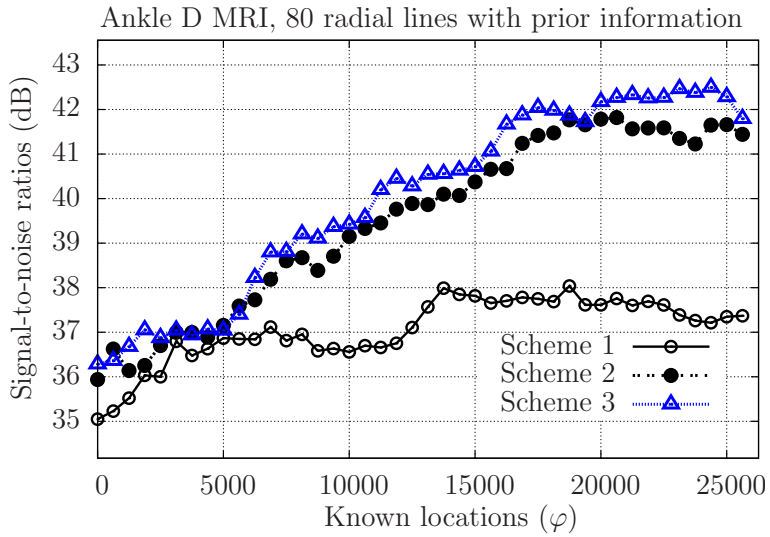


(a)

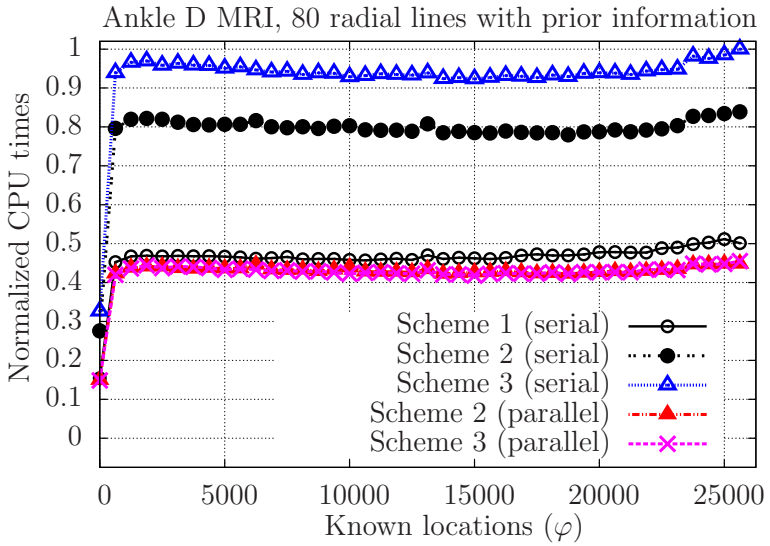


(b)

Figure 8.24. (a) Signal-to-noise ratios and (b) normalized computation times for the proposed MR imaging methods with prior information, when applied to the Ankle C MRI. Reference image provided by the Sierra Medical Center, El Paso, TX.

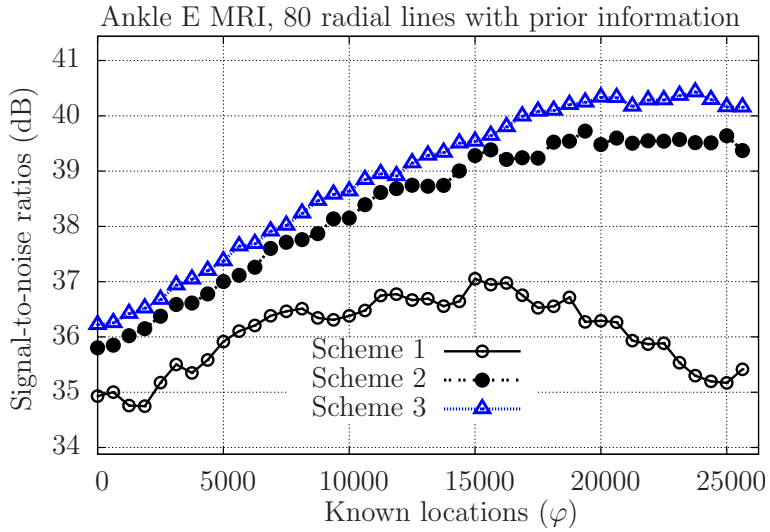


(a)

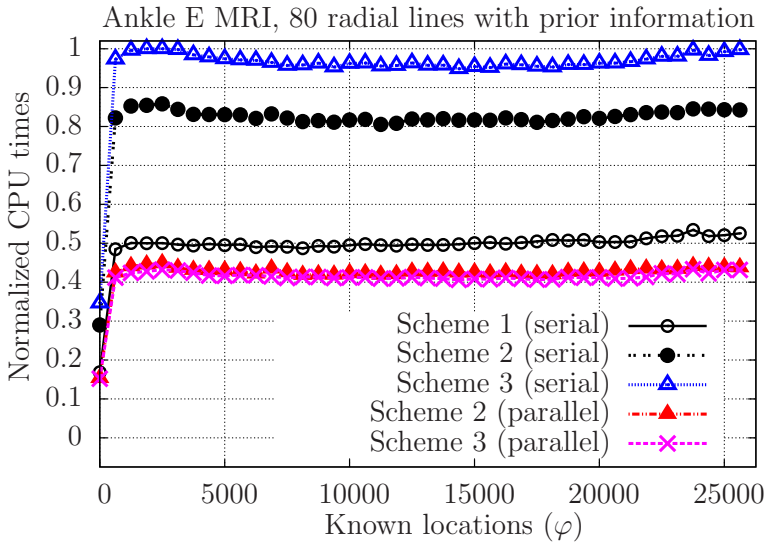


(b)

Figure 8.25. (a) Signal-to-noise ratios and (b) normalized computation times for the proposed MR imaging methods with prior information, when applied to the Ankle D MRI. Reference image provided by the Sierra Medical Center, El Paso, TX.



(a)



(b)

Figure 8.26. (a) Signal-to-noise ratios and (b) normalized computation times for the proposed MR imaging methods with prior information, when applied to the Ankle E MRI. Reference image provided by the Sierra Medical Center, El Paso, TX.

9

Conclusion

This dissertation introduced the theory of compressive sensing with prior information about the sparse representation in order to enhance signal reconstruction. It also presented a set of methods for magnetic resonance (MR) imaging based on compressive sensing, and that can be used with or without prior information. We have then shown that the proposed MR methods and the prior information approach reduce the number of linear measurements required for reconstruction, and improve the output signal-to-noise ratios for the same number of available measurements. In this final chapter, we present our main conclusions regarding the proposed theory and methods, based on the analyzes of the provided theorems and numerical experimentations. We also summarize this research's main contributions, while providing some suggestions for future work related to our findings.

Compressive Sensing with Prior Information

In the first part of the research, we presented the idea of prior information about the support of a signal's sparse representation. The analysis of this type of information represents a contribution both to the theory of signal reconstruction and to practical compressive sensing algorithms. In the first case, we mathematically proved that prior information about the

locations of φ components in the sparse representation reduces by φ the minimum number of linear measurements required for reconstruction. The ℓ_0-minimization approach can reconstruct compressible signals from this minimum set of measurements, but it is generally not viable in real applications due to its computational complexity; the alternative ℓ_1-minimization, for instance, is then more frequently used, and the compressive sensing literature describes the conditions that are necessary and sufficient for the ℓ_1-minimization to provide the same ideal solution. In this context, we extended this analysis by showing how these conditions change when prior information is used in the minimization procedure, and described how this procedure is affected in order to use this information in an optimal manner.

A further contribution at this stage was to show that the new conditions, corresponding to the reconstruction using prior information, are more feasible than in the case without this information. Specifically, we proved that the lower bound for the probability of reconstruction increases with the number of known locations in the sparse representation. Also, we obtained a closed form expression for the reconstruction probability with and without prior information, for random processes that are uncorrelated and symmetrically distributed with respect to the origin. The proposed theoretical model matched the results from Monte-Carlo simulations testing signals of this type. Future research about this subject can benefit from the analysis of random processes with different statistical properties, in order to evaluate the impact of prior information when reconstructing signals found in specific kinds of applications.

Future research on the proposed type of prior information should also investigate its effect when reconstructing signals in the presence of additive noise. Indeed, different studies in compressive sensing focus on reconstruction from noisy measurements, and their effect on the required number of measurements and reconstruction quality. Typically, the equality constraints in the optimization procedures are then replaced by inequality constraints, to provide a specified tolerance to noise. Including prior information in this approach may then provide important results, and an interesting possibility is that prior information might

improve robustness to noise.

Efficient Algorithm for Signal Reconstruction with Prior Information

In terms of a practical compressive sensing method, this research also introduced an efficient algorithm for signal reconstruction with prior information, based on ℓ_p-minimization. In fact, it is already known, in the compressive sensing research, that ℓ_p-minimization can improve reconstruction over ℓ_1-minimization by reducing the p parameter. This dissertation's contribution, in this respect, was to incorporate prior information into an iteratively reweighted least squares (IRLS) method through an appropriate weighting strategy, which takes into account the known locations in the sparse representation's support. We then showed that reducing p does not weight out the effect of prior information, so that further improvement can be achieved by reducing p and at the same time using prior information.

The numerical experimentations using the proposed method with prior information shows a reduction in the number of measurements required to attain a fixed percentage of correct reconstructions (meaning the cases in which the reconstructed signals match the testing ones within a prespecified tolerance). Furthermore, the method leads to a total reconstruction time that is a decreasing function of the number of known support locations, when using a direct method to solve the inner linear systems. Hence, under this condition prior information improves reconstruction without increasing the total cost in terms of computation time, and actually decreasing this cost.

It is interesting to observe, on the other hand, that in the developed MR imaging systems, when using an indirect method based on conjugate gradients, adding prior information in some cases lead to an increase in the total computation times, although the other figures of merit (such as signal-to-noise ratios) still improved. Anyway, the computation times for the proposed magnetic resonance imaging (MRI) methods were still generally lower than with a traditional compressive sensing approach used for the same type of images (the increase was with respect to the three proposed schemes without prior information). We discuss this issue in more detail when providing our concluding remarks on the proposed methods

for MRI with prior information; for now, it is worth emphasizing that the time increases, observed in some instances, are related to the stability of the linear system solved during the reconstructions. As prior information is added, some components of the system matrix approach zero, reducing the stability and thus affecting the convergence time in an indirect approach. In fact, by adjusting the parameter τ that controls the prior information weight, and thus how the system matrix components approach zero, it was possible to reduce the time delays, at the cost of reducing the impact of prior information.

Future research on the proposed IRLS algorithm can thus benefit from a detailed investigation of the effect of prior information on the computation times when using different indirect methods, and for a large range of the controlling parameter τ. Note that prior information itself, not counting the stability effect in the indirect methods, can significantly reduce the computation times, as we systematically observed by using the direct method.

When evaluating the IRLS method with prior information, we considered not only the number of measurements required to attain a certain percentage of correct reconstructions, but also the output signal-to-noise ratios (SNRs) for a fixed number of measurements. Indeed, in medical imaging applications, for instance, the focus is primarily on improving the image quality for the number of measurements that can be safely acquired in a reasonable amount of time and with the available technology. In this evaluation, we then observed that the SNRs significantly increase with prior information. Furthermore, this increase happens at a higher rate as the number of known support locations progressively increases. In other words, as more prior information is added, the global effect on the output SNRs per known location become higher.

We also emphasize that in investigating the proposed IRLS method's performance, we evaluated the figures of merit for a large range of the main configuration parameters, in particular the scalar τ that controls the prior information weighting. While controlling the regularization parameter showed a key hole in allowing convergence, as already discussed in previous research on optimization and compressive sensing, the final performance in terms of SNR and required number of measurements did not change significantly with respect to

τ. This new parameter can be chosen over a large range (typically more than 4 orders of magnitude) without compromising these figures of merit. Its impact was more important in the case of indirect implementations, where, as previously mentioned, it can be used to adjust the compromise between numerical stability and prior information weighting.

Prefiltering in the k-space domain for Enhanced Magnetic Resonance Imaging

Next, this research described and evaluated a method for the reconstruction of gradient-sparse magnetic resonance images based on the preprocessing of the available k-space samples. This method applies one of different possible sets of linear shift-invariant filters during the preprocessing, in order to yield the measurements corresponding to filtered versions of the image. The optimization algorithms, based on the IRLS method, then reconstruct these filtered versions using the preprocessed measurements, and a final composition stage builds the desired image from the filtered versions and from the low-pass information in the original measurements. The core idea is that the preprocessing stage, and the indirect reconstruction of filtered versions instead of the desired image directly, favor the reconstruction by compressive sensing methods, by increasing the objective function's sparsity.

The experimental results show that, for the tested images, the proposed method improves the reconstruction, in terms of SNRs and visual quality, over the total-variation minimization, commonly used for gradient-sparse images. We also observed that the method, when using three preprocessing filters in a serial implementation, leads to reconstruction times that are equivalent or lower than in the total-variation minimization using a log-barrier algorithm. Additionally, it also allows for a straightforward parallel implementation, as the reconstructions of the filtered versions of the desired image are independent of each other. The parallel implementations lead to a further time reduction of up to three times, in the tested schemes.

In a future research on the proposed MRI method, it is interesting to evaluate different sets of filters for the reconstruction of other types of magnetic resonance images, as opposed to those sparse under finite-differencing. Of particular interest are the images that are sparse

under some wavelet transformation, in which case the high-pass filters used in the wavelet decomposition should be tested. This condition is well motivated by the filters we used in the third form of the method, which are the high-pass bidimensional Haar kernels used in the one-stage image decomposition.

Another suggestion for future work on the proposed MR techniques is to evaluate their performance for different types of trajectories in the k-space, as we considered only the case of radial lines. Also, the robustness with respect to noise in the measurements must also be assessed (in our simulations, we did not add noise to the k-space samples, both in the case of the proposed method and in the total-variation minimization).

Magnetic Resonance Imaging using Compressive Sensing with Prior Information

Finally, this dissertation evaluated MR imaging systems that combine the proposed prefiltering strategy with the use of support prior information in the compressive sensing-based reconstruction. After discussing how this prior information in MR systems can be obtained in different contexts, we evaluated its impact on the output SNRs, for different numbers of known support locations and different numbers of k-space radial lines. The simulations show that prior information increases the SNRs, when compared to the same prefiltering schemes without prior information.

In the case of prior information extracted automatically, we observed that as the number of known locations increases, the output SNRs also increase up to a maximum value. After that critical point, which depends on the analyzed image, further increasing the number of extracted locations leads to a reduction in the output image quality. The reason is that after the critical point, when most of the support locations are already determined, further increasing the number of extracted points from the sparse representation leads to locations that actually do not belong to the support. Therefore, wrong prior information is used in the reconstruction process. This indicates that the number of extracted positions should be adjusted for the type of MR images considered and hence for the particular application.

A critical point is then to evaluate the system's robustness to wrong prior information.

In order to do so, we used a set of simulated functional magnetic resonance images (fMRIs), from which different numbers of wrong prior locations and correct prior locations (previously known, by construction) were extracted. We then evaluated the system's performance as a function of the number of wrong prior locations. The simulations showed that although increasing the number of wrong locations reduces the output SNRs, as expected, prior information still improves the reconstruction even when the number of wrong locations is significantly higher then the number of known correct ones. In fact, in the reconstruction with prior information, the positions assumed to belong to the support are removed from the objective function in the minimization process. Therefore, when wrong prior information is supplied, the local sparsity of the corresponding wrong locations is not exploited in the minimization process, which reduces the method's efficiency with respect to the case where no wrong information exists. However, the algorithm can still reconstruct the signals if the number of measurements is high enough, and the improvement provided by the correct prior information was shown to compensate for the wrong ones. Hence, even for a reasonably accurate prior information extraction procedure, using prior information can significantly improve reconstruction in fMRI systems.

In some cases, we have observed that using prior information in the proposed MR method can increase the reconstruction times with respect to the same method without prior information (although these times are still lower compared to the tested TV-minimization approach). This result differs from those from the simulations testing the proposed IRLS method. The reason, as previously mentioned, relates to the use of the indirect method to solve the inner linear systems in the MR problem, which is required by its large-scale characteristic, as opposed to the direct methods in the IRLS tests. Adding prior information reduces these systems' stability, thus requiring more time for the convergence of the iterative procedures used to solve them in an indirect approach. By adjusting the parameter τ that controls the prior information weighting, it was possible to set a compromise between convergence time and the improvement provided by using prior information.

Another interesting possibility for a future related work, therefore, is to evaluate the

viability of a direct implementation to solve the inner linear systems. Large-scale problems represent an important challenge to implementations based on direct methods, as memory requirements to store the system matrices can be prohibitive. A possible approach, however, is to compute the system matrices' components separately as they are needed in a matrix factorization procedure, and store only reduced blocks of those matrices. In this case, a Cholesky factorization, for instance, can be conducted and the systems can be solved by parts, until all the needed variables are computed. This procedure might allow a wider range of the controlling parameter τ, and thus a further reduction of the computation times as observed in the basic IRLS simulations. Besides, recent research on linear algebra and optimization methods describes novel parallel implementations of these matrices factorizations, with a clear benefit in terms of computation time.

In order to allow the future implementation of this type of direct method, we described how each component of the inner system matrices can be computed in an efficient manner, by using bidimensional-domain Fourier transforms of the previous iterates, when solving the related optimization problems. A closed form expression provides each component, so that it can be assessed at any time during the factorizations without the need to store the whole matrices. Comparing the performance and the computation times of a direct solution based on this approach with the indirect methods we used is a unique possibility for a future work.

Bibliography

[1] A.H. Andersen. Algebraic reconstruction in CT from limited views. *IEEE Transactions on Medical Imaging*, 8(1):50–55, 1989.

[2] A.H. Andersen and A.C. Kak. Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm. *Ultrasonic Imaging*, 6(1):81–94, 1984.

[3] S. Aviyente. Compressed sensing framework for EEG compression. In *IEEE/SP 14th Workshop on Statistical Signal Processing, 2007. SSP '07*, pages 181–184, 2007.

[4] D.H. Bailey and P.N. Swarztrauber. The fractional Fourier transform and applications. *SIAM Review*, 33(3):389–404, 1991.

[5] R.G. Baraniuk. Compressive sensing [lecture notes]. *IEEE Signal Processing Magazine*, 24(4):118–121, 2007.

[6] R.G. Baraniuk, M. Davenport, et al. A simple proof of the restricted isometry property for random matrices (aka the Johnson-Lindenstrauss lemma meets compressed sensing). *Constructive Approximation* [Online]. Available at http://www.dsp.ece.rice.edu/cs/JLCSfinalrevision1.pdf, 2007.

[7] Z.W. Birnbaum and A.W. Marshall. Some multivariate Chebyshev inequalities with extensions to continuous parameter processes. *The Annals of Mathematical Statistics*, 32(3):687–703, 1961.

[8] S.P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, UK; New York, 2004.

[9] E.J. Candès and J. Romberg. L1-magic: recovery of sparse signals via convex programming. Available at http://www.acm.caltech.edu/l1magic/l1magic.pdf, October 2005.

[10] E.J. Candès, J. Romberg, et al. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

[11] E.J. Candès, J.K. Romberg, et al. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

[12] E.J. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, 2006.

[13] E.J. Candès and M.B. Wakin. An introduction to compressive sampling [a sensing/sampling paradigm that goes against the common knowledge in data acquisition]. *IEEE Signal Processing Magazine*, 25(2):21–30, 2008.

[14] R. Chartrand. Exact reconstruction of sparse signals via nonconvex minimization. *IEEE Signal Processing Letters*, 14(10):707–710, 2007.

[15] R. Chartrand. Nonconvex compressed sensing and error correction. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007*, volume 3, pages III–889–III–892, 2007.

[16] R. Chartrand. Nonconvex compressive sensing and reconstruction of gradient-sparse images: Random vs. tomographic Fourier sampling. In *15^{th} IEEE International Conference on Image Processing, 2008. ICIP 2008*, pages 2624–2627, 2008.

[17] R. Chartrand. Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, Boston, Massachusetts, 2009.

[18] R. Chartrand and W. Yin. Iteratively reweighted algorithms for compressive sensing. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008*, pages 3869–3872, Las Vegas, Nevada, March 30 – April 4, 2008.

[19] G.H. Chen, J. Tang, et al. Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets. *Medical Physics*, 35(2):660–3, February 2008.

[20] T.A. Davis. *Direct methods for sparse linear systems*. Fundamentals of algorithms. SIAM, Society for Industrial and Applied Mathematics, Philadelphia, 2006.

[21] F.A. de Jonge and K.A. Blokland. Statistical tomographic reconstruction: how many more iterations to go? *European Journal of Nuclear Medicine*, 26(10):1247–50, October 1999.

[22] L. Desbat, S. Rit, et al. Algebraic and analytic reconstruction methods for dynamic tomography. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 726–730, 2007.

[23] D.L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[24] D.L. Donoho. For most large underdetermined systems of linear equations, the minimal ℓ_1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics*, 59(6):797–829, June 2006.

[25] M.F. Duarte, S. Sarvotham, et al. Distributed compressed sensing of jointly sparse signals. In *Conference Record of the Thirty-Ninth Asilomar Conference on Signals, Systems and Computers*, pages 1537–1541, 2005.

[26] M.F. Duarte, M.B. Wakin, et al. Fast reconstruction of piecewise smooth signals from incoherent projections. In *Structure et parcimonie pour la représentation adaptative de signaux, 2005. SPARS'05*, Rennes, France, November 16–18, 2005.

[27] M.F. Duarte, M.B. Wakin, et al. Wavelet-domain compressive signal reconstruction using a hidden Markov tree model. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008*, pages 5137–5140, 2008.

[28] M.R. Duarte, M.B. Wakin, et al. Universal distributed sensing via random projections. In *The Fifth International Conference on Information Processing in Sensor Networks — IPSN*, pages 177–185, 2006.

[29] M. Elad. Sparse representations are most likely to be the sparsest possible. *EURASIP Journal on Applied Signal Processing*, 2006(1):115–115, 2006.

[30] I. Esnaola and J. Garcia-Frias. Exploiting prior knowledge in the recovery of non-sparse signals from noisy random projections. In *41st Annual Conference on Information Sciences and Systems, 2007. CISS '07*, pages 731–731, 2007.

[31] I. Esnaola and J. Garcia-Frias. Distributed compression of correlated signals using random projections. In *Data Compression Conference, 2008. DCC 2008*, pages 322–331, 2008.

[32] D.G. Fryback and J.R. Thornbury. The efficacy of diagnostic imaging. *Medical Decision Making*, 11(2):88–94, 1991.

[33] J.-J. Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Transactions on Information Theory*, 50(6):1341–1344, 2004.

[34] J.-J. Fuchs. Recovery of exact sparse representations in the presence of bounded noise. *IEEE Transactions on Information Theory*, 51(10):3601–3608, 2005.

[35] J.-J. Fuchs. Convergence of a sparse representations algorithm applicable to real or complex data. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):598–605, 2007.

[36] J.-J. Fuchs. Fast implementation of a ℓ_1-ℓ_1 regularized sparse representations algorithm. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009*, pages 3329–3332, 2009.

[37] J. Garcia-Frias and I. Esnaola. Exploiting prior knowledge in the recovery of signals from noisy random projections. In *Data Compression Conference, 2007. DCC '07*, pages 333–342, 2007.

[38] J. Garcia-Frias and I. Esnaola. Distributed compression of correlated real sequences using random projections. In *Information Theory Workshop, 2008. ITW '08. IEEE*, pages 189–193, 2008.

[39] G.H. Golub and C.F. Van Loan. *Matrix computations*. Johns Hopkins studies in the mathematical sciences. Johns Hopkins University Press, Baltimore, 3^{rd} edition, 1996.

[40] R. Gribonval and M. Nielsen. Beyond sparsity: recovering structured representations by ℓ_1-minimization and greedy algorithms. *Journal of Advances in Computational Mathematics*, 28(1):23–41, January 2008.

[41] B.C. Gunter, W.C. Reiley, et al. Implementation of out-of-core Cholesky and QR factorizations with POOCLAPACK. Technical report, University of Texas at Austin, 2000.

[42] B.C. Gunter, W.C. Reiley, et al. Parallel out-of-core Cholesky and QR factorizations with POOCLAPACK. In *Parallel and Distributed Processing Symposium., Proceedings 15^{th} International*, pages 1885–1894, 2001.

[43] A.C. Gurbuz, J.H. McClellan, et al. Compressive sensing for GPR imaging. In *Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers, 2007. ACSSC 2007*, pages 2223–2227, 2007.

[44] K.M. Hanson. Bayesian and related methods in image reconstruction from incomplete data. In Henry Stark, editor, *Image recovery: theory and application*. Academic Press, Orlando, 1987.

[45] K.M. Hanson and G.W. Wecksung. Bayesian approach to limited-angle reconstruction in computed tomography. *Journal of the Optical Society of America*, 73(11):1501–1509, 1983.

[46] S. Horbelt, M. Liebling, et al. Discretization of the Radon transform and of its inverse by spline convolutions. *IEEE Transactions on Medical Imaging*, 21(4):363–376, 2002.

[47] J.I. Jackson, C.H. Meyer, et al. Selection of a convolution function for Fourier inversion using gridding [computerised tomography application]. *IEEE Transactions on Medical Imaging*, 10(3):473–478, 1991.

[48] M. Jiang and G. Wang. Convergence of the simultaneous algebraic reconstruction technique (SART). In *Conference Record of the Thirty-Fifth Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 360–364 vol.1, 2001.

[49] M. Jiang and G. Wang. Convergence of the simultaneous algebraic reconstruction technique (SART). *IEEE Transactions on Image Processing*, 12(8):957–961, 2003.

[50] A.C. Kak and M. Slaney. *Principles of computerized tomographic imaging*. Classics in applied mathematics. Society for Industrial and Applied Mathematics, Philadelphia, 2001.

[51] S.M. Kay. *Intuitive probability and random processes using MATLAB*. Springer, New York, 2006.

[52] C.S. Kidwell, J.L. Saver, et al. Magnetic resonance imaging detection of microbleeds before thrombolysis: an emerging application. *Stroke*, 33(1):95–8, 2002.

[53] E.A. Krupinski and Y. Jiang. Anniversary paper: evaluation of medical imaging systems. *Medical Physics*, 35(2):645–59, Feb 2008.

[54] C. La and M.N. Do. Signal reconstruction using sparse tree representations. In *Proceedings of the SPIE Conference on Wavelet Applications in Signal and Image Processing XI*, San Diego, USA, July 2005.

[55] T.M. Lehmann, C. Gonner, et al. Survey: interpolation methods in medical image processing. *IEEE Transactions on Medical Imaging*, 18(11):1049–75, Nov 1999.

[56] T.M. Lehmann, C. Gonner, et al. Addendum: B-spline interpolation in medical image processing. *IEEE Transactions on Medical Imaging*, 20(7):660–665, 2001.

[57] M. Lustig. *Sparse MRI*. Ph.D thesis, Stanford University, 2008.

[58] M. Lustig, D.L. Donoho, et al. Compressed sensing MRI [a look at how CS can improve on current imaging techniques]. *IEEE Signal Processing Magazine*, 25(2):72–82, 2008.

[59] S.G. Mallat. *A wavelet tour of signal processing: the sparse way*. Elsevier/Academic Press, Amsterdam; Boston, 3^{rd} edition, 2009.

[60] A. Markoe. *Analytic tomography*. Cambridge University Press, New York, 2006.

[61] A.W. Marshall and I. Olkin. Multivariate Chebyshev inequalities. *The Annals of Mathematical Statistics*, 31(4):1001–1014, 1960.

[62] N. Metropolis. The beginning of the Monte Carlo method. *Los Alamos Science*, 15:125–130, 1987.

[63] N. Metropolis and S. Ulam. The Monte Carlo method. *Journal of the American Statistical Association*, 44(247):335–41, September 1949.

[64] C.D. Meyer. *Matrix analysis and applied linear algebra*. Society for Industrial and Applied Mathematics, Philadelphia, 2000.

[65] C.J. Miosso, R. von Borries, et al. Compressive sensing reconstruction with prior information by iteratively reweighted least-squares. *IEEE Transactions on Signal Processing*, 57(6):2424–2431, 2009.

[66] C.J. Miosso, R. von Borries, et al. Compressive sensing method for improved reconstruction of gradient-sparse magnetic resonance images. In *Forty-Third Asilomar Conference on Signals, Systems & Computers*, Pacific Grove, California, 2010.

[67] C.J. Miosso, R. von Borries, et al. Compressive sensing using prior information on the support region of the sparse domain. *Journal paper in preparation*, 2010.

[68] S.K. Mitra. *Digital signal processing.* McGraw-Hill series in electrical and computer engineering. McGraw-Hill/Irwin, Boston, 3^{rd} edition, 2005.

[69] P. Munoz, J.L. Semmlow, et al. Cortical activity in disparity vergence eye-movements: a fMRI study. In *Bioengineering Conference, 1998. Proceedings of the IEEE 24^{th} Annual Northeast*, pages 79–81, 1998.

[70] F. Natterer. Algorithms in tomography. In *The State of the Art in Numerical Analysis.* Clarendon Press, 1997.

[71] F. Natterer. Numerical methods in tomography. *Acta Numerica*, 8:107–143, 1999.

[72] J. Nocedal and S.J. Wright. *Numerical optimization.* Springer, New York, 2^{nd} edition, 2006.

[73] I. Olkin and J.W. Pratt. A multivariate Tchebycheff inequality. *The Annals of Mathematical Statistics*, 29(1):226–234, 1958.

[74] J.D. O'Sullivan. A fast sinc function gridding algorithm for Fourier inversion in computer tomography. *IEEE Transactions on Medical Imaging*, 4(4):200–207, 1985.

[75] S. Pan and A. Kak. A computational study of reconstruction algorithms for diffraction tomography: Interpolation versus filtered-backpropagation. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 31(5):1262–1275, 1983.

[76] J. Provost and F. Lesage. The application of compressed sensing for photo-acoustic tomography. *IEEE Transactions on Medical Imaging*, 28(4):585–594, 2009.

[77] B.D. Rao and K. Kreutz-Delgado. An affine scaling methodology for best basis selection. *IEEE Transactions on Signal Processing*, 47(1):187–200, 1999.

[78] J.K. Reid and J.A. Scott. An out-of-core sparse Cholesky solver. Technical report, Council for the Central Laboratory of the Research Councils, October 2006.

[79] S. Resch, A. Stenstrom, et al. The diagnostic efficacy of magnetic resonance imaging and ultrasonography in Morton's neuroma: a radiological-surgical correlation. *Foot & Ankle International*, 15(2):88–92, 1994.

[80] J. Romberg. Imaging via compressive sampling [introduction to compressive sampling and recovery via convex programming]. *IEEE Signal Processing Magazine*, 25(2):14–20, 2008.

[81] J.M. Santos, C.H. Cunningham, et al. Single breath-hold whole-heart MRA using variable-density spirals at 3T. *Magnetic Resonance in Medicine*, 55(2):371–379, 2006.

[82] H. Schomberg and J. Timmer. The gridding method for image reconstruction by Fourier transformation. *IEEE Transactions on Medical Imaging*, 14(3):596–607, 1995.

[83] J. Shewchuk. An introduction to the conjugate gradient method without the agonizing pain. Technical report, Carnegie Mellon University, 1994.

[84] H. Stark and J.W. Woods. *Probability and random processes with applications to signal processing*. Prentice Hall, Upper Saddle River, N.J., 3^{rd} edition, 2002.

[85] J. Steffener, T.L. Alvarez, et al. Motor learning in the saccadic oculomotor system revealed through fMRI. In *Proceedings of the 29^{th} Annual Bioengineering Conference*, pages 100–101, 2003.

[86] T. Tao. An uncertainty principle for cyclic groups of prime order. *Mathematical Research Letters*, 12(1):121–128, 2005.

[87] J.R. Thornbury. Eugene W. Caldwell lecture. Clinical efficacy of diagnostic imaging: love it or leave it. *Americal Journal of Roentgenology*, 162(1):1–8, 1994.

[88] J.A. Tropp. Greed is good: algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(10):2231–2242, 2004.

[89] J.A. Tropp and A.C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 53(12):4655–4666, 2007.

[90] J.A. Tropp, A.C. Gilbert, et al. Simultaneous sparse approximation via greedy pursuit. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05)*, volume 5, pages v/721–v/724 Vol. 5, 2005.

[91] J.D. Trzasko and A. Manduca. A fixed point method for homotopic ℓ_0-minimization with application to MR image recovery. In *Medical Imaging 2008: Physics of Medical Imaging*, volume 6913, pages 69130F–11, San Diego, CA, USA, 2008. SPIE.

[92] R. von Borries, C.J. Miosso, et al. Compressed sensing using prior information. In *2nd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, 2007. CAMSAP 2007*, pages 121–124, St. Thomas, U.S. Virgin Islands, December 12–14, 2007.

[93] R. von Borries, C.J. Miosso, et al. Directional filter banks for wavelet decomposition of images based on the Radon transform. In *Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers, 2007. ACSSC 2007*, pages 2095–2099, 2007.

[94] P. Whittle. A multivariate generalization of Tchebichev's inequality. *The Quarterly Journal of Mathematics*, 9(1):232–240, January 1, 1958 1958.

[95] B. Wohlberg and P. Rodriguez. An iteratively reweighted norm algorithm for minimization of total variation functionals. *IEEE Signal Processing Letters*, 14(12):948–951, 2007.

[96] G.A. Wright. Magnetic resonance imaging. *IEEE Signal Processing Magazine*, 14(1):56–66, 1997.

[97] Y. Zhang, S. Mei, et al. A multiple description image/video coding method by compressed sensing theory. In *IEEE International Symposium on Circuits and Systems, 2008. ISCAS 2008*, pages 1830–1833, 2008.

Curriculum Vitae

Cristiano Jacques Miosso completed his undergraduate studies in Electrical Engineering at University of Brasilia in 1999, when he received the best-student award from the Electrical Engineering Department. He then joined the master's program at the same institution, and received the Master of Science degree in 2002.

After working for three years as a substitute professor at University of Brasilia, and for five years as a consultant for projects in automation and electronics, Cristiano Miosso joined the doctoral program at UTEP in 2006, where he has served as a research assistant.

Dr. Miosso has been the recipient of several awards, including the best-student award from the Electrical and Computer Engineering Department, at University of Brasilia, and a full scholarship from the Brazilian Ministry of Education, during his master's studies. During his doctoral studies, he also received the Krutilek Scholarship and, for three times, the Texas Instruments Endowed Scholarship. From 2006 to 2010, he received funding from the National Science Foundation (NSF), the National Geospatial-Intelligence Agency (NGA) and the University Research Institute (URI UTEP).

A 4.0-GPA student during both his master's and his doctoral degree studies, Dr. Miosso joined the Phi Kappa Phi honor society in 2008. He is also a member of the Institute of Electrical and Electronics Engineers (IEEE). In parallel to his engineering studies, he pursued his passion for music and piano playing; in 2001 he was the only South-American pianist to take part in the International Festival of Piano in Merano (Italy) and he was a finalist and the third price recipient at the South-American Piano Competition in Goiania (selective for the Gina Bachauer International Competition).

Dr. Miosso's research has been published in several papers, presented at international conferences and in scientific journals, such as The 2007 IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, the 2008 and 2010 Asilomar Conferences on Signals, Systems & Computers, and the IEEE Transactions on Signal Processing.

Dr. Miosso has been invited to teach graduate courses in Electrical Engineering, at University of Brasilia, starting in the first semester of 2011.

His dissertation, entitled "Compressive Sensing with Prior Information Applied to Magnetic Resonance Imaging," was supervised by Dr. Ricardo von Borries, with whom he continues to collaborate in other research projects. His main research interests include statistical digital signal processing, stochastic processes, compressive sensing, medical imaging, and communication theory.

He can be reached through the email miosso@ieee.org.