SIGNAL RECONSTRUCTION FROM COMPRESSIVE SAMPLES

Dornoosh Zonoobi

NATIONAL UNIVERSITY OF SINGAPORE
2012

SIGNAL RECONSTRUCTION FROM COMPRESSIVE SAMPLES

Dornoosh Zonoobi

(M.Eng. (Computer Engineering), NUS)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2012

This thesis is dedicated to

Masood

Acknowledgments

There are many people whom I wish to thank for the help and support they have given me throughout my PhD studies. Foremost, I would like to express my sincere gratitude to my supervisor A/P Ashraf Kassim. I thank him for his patience, care and encouragement that carried me on through all the difficult times, and for his invaluable guidance, insights and suggestions which helped me throughout this work. His nice personality has made me always feel at home in Singapore, even though my motherland is thousands of miles away. Besides my advisor, I would like to thank Prof. Y.V.Venkatesh for his insightful comments, and valuable discussions.

I also take this opportunity to thank all my friend in Singapore, whose presences and fun-loving spirits made this journey all the more memorable.

Last but not the least, I would like to thank my family, Danesh, maman Mahnoosh, maman Soraya, baba Farshid, baba Ghasem, Saeed ,Negin and Naji, for always being there when I needed them most, and for supporting me and Masood through all these years. Maman, baba, Danesh, I am so lucky to have you. Most importantly to my amazing husband and best friend Masood, your love and patience made this possible. This thesis is dedicated to you.

Summary

The emerging field of Compressive Sensing (CS), is a novel sampling paradigm that exploits the sparsity/compressibility of signals to reconstruct them from far fewer samples, than what is required by the traditional Shannon-Nyquist sampling theorem. Unlike conventional methods, which use linear sinc interpolation to recover signals/images from the acquired samples, CS relies on non-linear optimization-based methods to find the sparsest signal among the set of all feasible solutions.

In the current literature of compressive sampling, sparsest signal corresponds to the one with the minimum value of ℓ_0 norm (e.g. number of nonzero elements). However, it is acknowledged that solving the equivalent optimization problem is computationally unwieldy in view of its NP-hard nature. Therefore, in the majority of CS literature, the reconstruction is done using convex ℓ_1-based optimization.

This thesis presents some efficient and practical methodologies for reconstruction of high dimensional signals from compressive measurements that overcome the current limitations of state-of-the-art CS recovery methods. The key contributions include: Developing a stochastic-based method for achieving as close an approximation to ℓ_0-norm as is computationally feasible in signal reconstruction from compressive samples; Exploring properties of the Gini index (GI) as an

sparsity measure in the problem of signal/image reconstruction; Demonstrating the robustness and reliability of GI as an alternative to the currently popular ℓ_p norm-based (for $0 < p \leq 1$) sparsity measures, through extensive experiments.

In the case of time-varying signals, a novel approach for recursively reconstructing sequences of sparse signals is proposed, where sparsity changes smoothly with time. In this approach reconstructed signal of the previous time instant is used to extract a probability model. This priori-knowledge is then incorporated into the reconstruction of the next time instant signal, to significantly reduce the number of needed samples, compared to other state-of-art CS methods. Lastly, the application of the developed method in low power ECG wireless-enabled monitoring devices and medical imaging modalities, is tested.

Contents

List of Figures

List of Tables

List of symbols

$\mathcal{R}^n$	n dimensional Euclidean space
z	Number of zero elements
$\mathbf{s}$	Original signal/image of interest
Σ_k^n	set of all $k-$sparse signals belonging to $\mathcal{R}^n$
$\mathbf{\Phi}$	Measurement/Sampling matrix
$\mathbf{y}$	Compressive measurements
$\mathbf{\Psi}$	Sparsifying Transform
η	Noise vector
$\mathbf{S}$	Sparse approximation of $\mathbf{s}$
$\hat{\mathbf{s}}$	Reconstructed signal
$\mathbf{f}^{(t)}$	Time-varying signal of interest
$\mathcal{S}^{(t)}$	Support set of $\mathbf{f}^{(t)}$
$\mathbf{y}^{(t)}$	Time-varying compressive measurements
$\mathbf{v}$	Null space vector

Acronyms

CS	Compressive Sensing
DCT	Discrete Cosine Transforms
DWT	Discrete Wavelet Transform
GI	Gini Index
i.i.d.	Independent and Identically Distributed
IHT	Iterative Hard Thresholding
MP	Matching Pursuit
MSE	Mean Squared Error
OMP	Orthogonal Matching Pursuit
PSNR	Peak Signal-to-Noise Ratio
RIP	Restricted Isometery Property
RWℓ_1	Re-weighted ℓ_1 minimization
SNR	Signal-to-Noise Ratio
SPSA	Simultaneous Perturbation Stochastic Approximation

Chapter 1

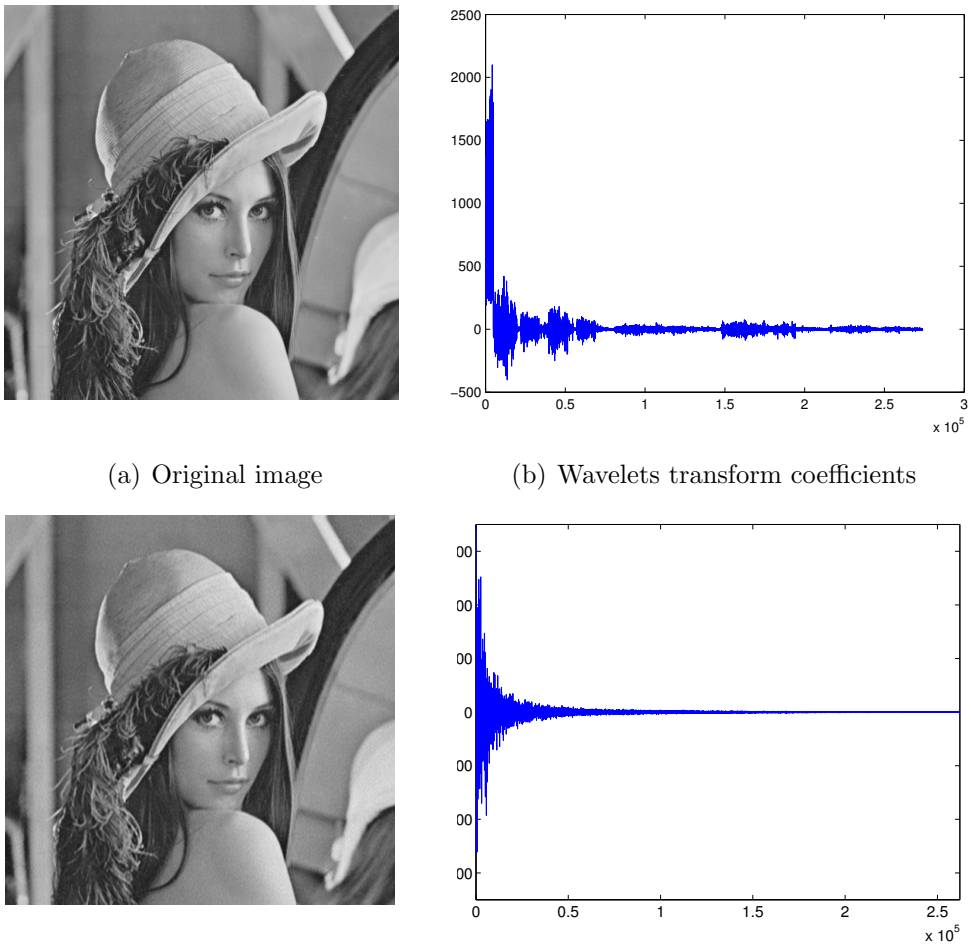
Introduction

1.1 Sparse Representations

The first fundamental step in any digital signal processing system is to discretize the signal of interest. This is called sampling. Conventional approaches to sampling are based on Shannon's sampling theorem (also known as Nyquist rate) which asserts that the sampling rate must be at least twice the maximum frequency present in the signal. Currently, nearly all signal acquisition protocols embrace this rate of sampling. Even in cases where the signal of interest is naturally not band-limited, an anti-aliasing low-pass filter is usually applied first to band limit the signal before sampling.

In many important applications, the signals of interest are sampled at high data rates, resulting in large quantities of data, that need to be processed, transmitted and/or stored. Compression Algorithms are hence heavily relied upon to reduce the dimension/quantities of the acquired data.

Compression algorithms are based on the fact that many natural signals/images are *sparse* or *compressible*. JPEG, for example is based on sparsity of signals in the Discrete Cosine Transforms (DCT), while the sparsity in Discrete Wavelet Transform (DWT) is the heart of JPEG2000.



(a) Original image

(b) Wavelets transform coefficients

(c) Compressed image (setting 90% of DCT coefficients to zero)

(d) DCT transform coefficients

Figure 1.1: Lena image and its sparse transforms.

Intuitively, a signal ($\mathbf{s} \in \mathcal{R}^n$) is considered sparse if it has small amount of nonzero entries. More specifically, $\mathbf{s}$ is considered *approximately* sparse or

compressible, if it can be expanded in terms of a proper basis ($\mathbf{\Psi} \in \mathcal{R}^{n \times n}$), such that $\mathbf{S} = \mathbf{\Psi s}$ satisfies a power law decay as:

$$\mathbf{S}_{[i]} \leq C i^{-1/p}$$

where $\mathbf{S}_{[i]}$ is the i^{th} largest coefficient of $\mathbf{S}$, C is a constant and $0 < p < 1$. Note that smaller values of p correspond to faster decay in the magnitude of the coefficients.

Consider, for example, the Lena image in figure 1.1(a) and its DWT and DCT transforms in 1.1(b) and 1.1(c), respectively. Although nearly all the image pixels have nonzero values, most of the discrete cosine and wavelet transform coefficients are small, and relatively few large coefficients capture most of the image energy in both transform domains.

An immediate perception is that, a sparse signal is built upon much smaller degrees of freedom compared to what is suggested by its original length (n). A standard compression strategy for such signal is to first map it into its *sparsifying* transform domain (for instance a wavelet basis) and then encode those few significant coefficients and store them, for later decoding and reconstruction of the signal of interest.

It is evident that in such conventional approaches, complete information of the signal of interest has to be acquired first through possibly a difficult or time-consuming measurement process after which much of the acquired data is thrown away to obtain the compressed version. This seems to be a huge waste of resources.

One might ask whether there is better way of obtaining the compressed version

of the signal somewhat more efficiently. Measuring directly the large coefficients is impossible since their locations are not know a priori.

1.2 Compressive Sensing

Compressive Sensing (CS) [2, 3] is a novel sensing/sampling paradigm that provides a way of obtaining the compressed version of a signal using only a small number of linear and non-adaptive measurements. CS exploits the sparsity/compressibility, which exists in many natural signals, to reconstruct them from far fewer samples, *measurements*, than traditional methods require.

A successful application of CS is based on three fundamental grounds [4]:

- **Sparse basis:** signal/image to be sampled must have a (approximately) sparse representation in a known transform domain.
- **Incoherent sampling domain:** signal/image of interest should be linearly sampled in a domain which is incoherent with the sparse domain. This means that aliasing artifacts in a linear reconstruction caused by undersampling must be noise-like in the sparsifying transform domain.
- **Nonlinear reconstruction:** unlike conventional methods, CS use nonlinear reconstruction methods which enforces both sparsity of the signal representation and consistency of the reconstruction with the acquired samples.

CS contrasts with conventional sampling theory in two important aspects: First, is in the sampling phase where rather than sampling at specific points

in time, CS typically acquires m weighted linear combination of samples where $m \ll n$. The measurements are acquired as $\mathbf{y} = \mathbf{\Phi s}$, where $\mathbf{\Phi}$ is an $m \times n$ *measurement matrix*. Noise may also be present in the measurement process and therefore we will also consider a more general model $\mathbf{y} = \mathbf{\Phi s} + \eta$, where η is the measurement noise [5].

Secondly, the two sampling paradigms differ mainly in the way they recover the original signal back from the taken samples. In the Nyquist framework, signal is recovered through a linear process that uses sinc interpolation. This is of course only feasible since the number of samples are large enough. In CS, however, number of samples are much smaller than the signal dimension and therefore recovering the original signal can be viewed as solving an under determined system of linear equations which is not possible in general. However, CS makes use of the constraint, that the initial signal is sparse in a known domain and employs non-linear optimization-based methods to search for the sparsest signal that satisfies the measurements (1.1).

$$\begin{aligned} &\arg\max_{\hat{\mathbf{s}} \in \mathcal{R}^n} \text{sparsity } \{(\mathbf{\Psi}\hat{\mathbf{s}})\} \\ &\text{s.t.} \quad \mathbf{\Phi}\hat{\mathbf{s}} = \mathbf{y} \end{aligned} \tag{1.1}$$

In the case where measurements are corrupted with noise, one may change the constraint in (1.1) to $\|\mathbf{\Phi}\hat{\mathbf{s}} - \mathbf{y}\| \leq \epsilon$.

It is evident that to be able to find the sparsest signal, one should first be able to measure and quantify sparsity and this measure has a profound effect on the performance of any CS-based algorithm. In the literature of CS, sparsity is

fundamentally quantified by an intuitive choice, namely ℓ_0 norm[1], which is simply number of the non-zero elements present in a signal (1.2).

$$\|\mathbf{s}\|_0 = \{|i| : \mathbf{s}_i \neq 0\} \tag{1.2}$$

It is believed that the smaller value of ℓ_0 norm in a signal means the sparser that signal is. Therefore, in order to find the sparsest signal, CS methods search for the signal with the least value of ℓ_0 norm, solving (1.3).

$$\begin{aligned} &\arg\min_{\hat{\mathbf{s}}\in\mathcal{R}^n} \ \|\hat{\mathbf{s}}\|_0 \\ &\text{s.t.} \quad \mathbf{\Phi}\hat{\mathbf{s}} = \mathbf{y} \end{aligned} \tag{1.3}$$

However, solving the above minimization problem is known to be computationally unwieldy in view of its combinatorial nature. As a consequence, CS methods are compelled to resort to an alternative convex norm as an approximation to ℓ_0 norm, namely ℓ_1 norm defined as (1.4).

$$\|\mathbf{s}\|_1 = (\sum_{i=1}^{n} |\mathbf{s}_i|) \tag{1.4}$$

It is hoped that by substituting this convex norm and solving (1.5), one can

[1]In CS literature, the norm terminology has invariably been used for ℓ_0 even though it is known that it does not satisfy the triangle inequality. It is a deliberate (but understood) misuse of terminology.

recover a result which is (approximately) equivalent to (1.3).

$$\arg \min_{\hat{\mathbf{s}} \in \mathcal{R}^n} \ \|\hat{\mathbf{s}}\|_1 \quad \text{s.t.} \quad \mathbf{\Phi}\hat{\mathbf{s}} = \mathbf{y} \tag{1.5}$$

It have been established in CS literature [3] [4] that the combinatorial problem (1.3) and its relaxation (1.5), are equivalent provided that $\mathbf{s}$ is sparse enough and $\mathbf{\Phi}$ satisfies certain properties.

1.3 Motivation and Objectives

The emerging field of CS, has attracted considerable attention in recent years. CS has been shown to have a great impact on wide areas of application such as compressive imaging [6, 7], machine learning [8], data streaming [9] and etc. However, despite the extensive recent work that has been done in this field, there are still many challenges to be overcome and yet to be thoroughly studied. Some of these issues, form the basis of the research work presented, in this thesis.

1.3.1 Recovery Algorithms

As mentioned earlier, the non-linear recovery algorithm (decoder) is a crucial part in CS performance. Currently the convex ℓ_1 minimization decoder is the mostly used approach in all CS-based applications, as it offers theoretical recovery guarantees and is stable under measurement noises [10]. However as it will be discussed in more details in Chapter 2, there are two main limitations to this

approach:

- Its underlying convex optimization requires relatively high memory usage and computational cost [11] [12]. For example, the linear programming, needed to solve (1.5), is of the complexity order of n^4 [13], which is not optimally efficient for applications involving large dimension processing. To deal with this problem of the convex-based methods, a class of greedy algorithms has been proposed in CS literature [14]. These methods, at their best provide similar reconstruction performance to those of ℓ_1 minimization based approaches.

- It was empirically observed that treating the NP-hard optimization problem in (1.3) by approximating it to a convex optimization problem (1.5), do not always lead to satisfactory reconstruction of signals. The key difference between the ℓ_1 and ℓ_0 norms is the dependence of ℓ_1 on magnitude. Unlike ℓ_0 norm that penalizes all nonzero elements equally, in ℓ_1 norm larger coefficients are penalized more heavily than smaller coefficients. Therefore, in some scenarios, the ℓ_1 does not provide a close enough approximation of the ℓ_0. This accounts for the choice, in recent literature (e.g. [15, 16]), of the non-convex objective functions, which are a closer approximation of the ℓ_0 than the ℓ_1 norm. These algorithms, though known to be superior to ℓ_1 minimization in terms of the reconstruction performance, are often of a higher computational complexity.

The above observations is the motivation to propose a *stochastic-based recovery algorithm*, as a fast and robust approach to achieve a closer approximation

to ℓ_0-norm than the one of ℓ_1 method. The essential feature of the proposed method, which makes it efficient in difficult multivariate optimization problems, is its underlying gradient approximation that requires only two objective function measurements per iteration regardless of the dimension of the optimization problem. Moreover, due to its stochastic nature, it is stable in view of the noisy inputs.

1.3.2 Measures of Sparsity

CS relies chiefly on finding the sparsest signal and currently the sparsity is measured by an intuitive choice which is the number of non-zero elements of a signal. However, when dealing with real data, the ℓ_0 measure is inapplicable due to the presence of the noise. Therefore, as an alternative, ℓ_0 is modified to ℓ_0^ϵ, in which coefficients having a value smaller than (a threshold) ϵ are considered to be zero. Clearly, the value of the threshold is crucial in the performance of this measure, and its theoretical results. Such a situation is not desirable as the threshold depends on the nature of the signal we are sampling, but we have assumed that, apart from the fact that the signal is sparse, no prior knowledge is available. Moreover, finding the signal with the minimal value of ℓ_0 is known to be computationally unwieldy in view of its NP-hard nature. As a consequence, majority of CS literature, utilize the ℓ_p norms with $0 < p \leq 1$ as an approximation to the ℓ_0 norm.

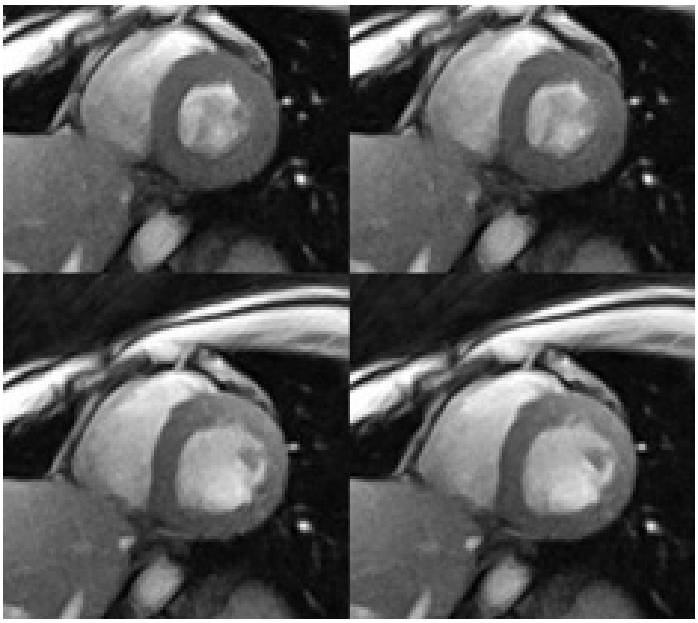
Clearly, choice of sparsity measure has a profound impact on the CS performance. However, not much has been done on studying, analysis and comparison the performance of different sparsity measures. This motivates us to provide a

thorough analysis of different sparsity measures as applied to the problem of signal reconstruction from compressive samples, in Chapter 3 of this thesis. In this chapter, through various examples, it is shown that despite the intense focus on ℓ_0 norm as sparsity measure in the CS field, it may not be the best choice. It is shown that this measure, together with its ℓ_1 and ℓ_p approximations, exhibit some properties which are counter to the intuitive understanding of sparsity. One example of such properties, is the dependence of ℓ_1 and ℓ_p on the size and total energy of the signal. This motivates us to explore the use of an alternative measure of sparsity, namely *Gini index* (GI), in the context of CS.

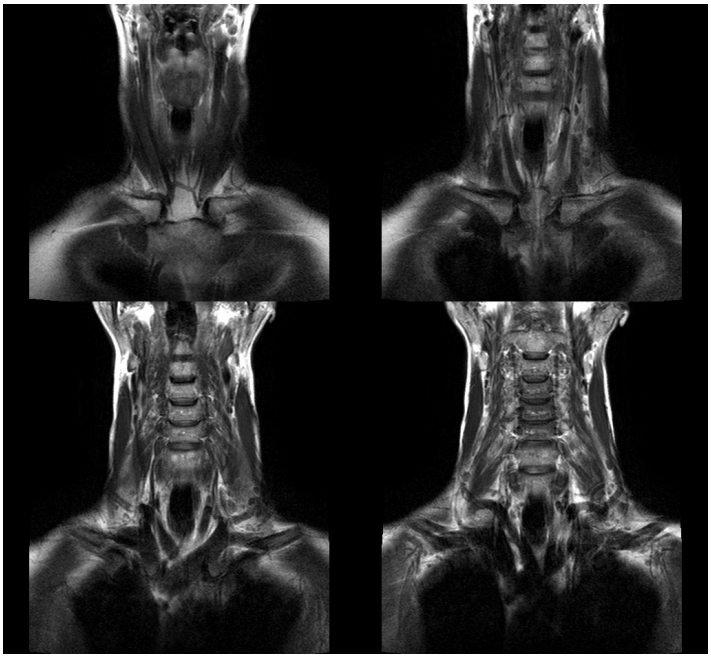
1.3.3 Reconstruction of Dynamic Signals

There are many important real-world applications in which the signal of interest, and therefore compressive measurements, are varying with time. In many of these applications the rate of change is so slow and smooth that the signal at each time instant is closely related to the one at the previous time. Examples of such applications include ECG signals, realtime MRI images, frames of CCTV video sequences and etc. (see figure 1.2).

In these applications we are not only interested in reconstruction of a single signal, but we need to recover the whole sequence. Intuitively the redundancy which exists in the signals of the adjacent time windows should be used to further reduce the number of needed samples for reconstruction.



(a) Sample frames of a real-time MRI



(b) Sequence of Neck MRI images

Figure 1.2: Sequences of sparse images.

While extensive work has been done in the CS literature on the reconstruction of signals, much of them limit their scope to merely reconstruction of static ones [17]. The current body of work in CS literature is mainly focused on general sampling and reconstructing techniques that is not specifically designed for any class of signals. Conventional CS methods are, therefore, not well suited for estimating time varying sparse signals from a series of changing compressive measurements.

Chapter 4 of this thesis presents a new method specifically designed for reconstruction of sequences of dynamic sparse signals. The proposed approach incorporates a priori knowledge, of the current signal, which is extracted from the reconstructed signal of the previous time window, into the recovery process.

1.4 Main Contributions and Organization

This thesis presents some key new developments in the field of compressive sensing, which are efficient and practical methodologies for reconstruction of high dimensional signals from compressive measurements. The key contributions are as follows:

- **Chapter 2:** First some of the existing sparse recovery algorithms and their reconstruction guarantees are reviewed. Then a fast and robust recovery method, well suited to the problem of high-dimensional signal reconstruction, is presented in which ℓ_p-norm $(0 < p \leq 1)$ is used to achieve a close approximation to ℓ_0-norm. By way of thorough comparison of its performance against the one of the other convex, greedy and non-convex based

CS methods, it is shown that the proposed scheme can achieve perfect reconstruction using smaller number of samples. Moreover, the robustness of this approach in dealing noisy measurement, is tested.

- **Chapter 3:** A thorough study and comparison of the currently used ℓ_p-norm ($0 \leq p \leq 1$) sparsity measures vs. the Gini Index, is presented and the superiority of GI is demonstrated through numerous example and experiments. Moreover, the GI is successfully incorporated into an optimization algorithm for signal reconstruction from compressive samples for a significantly improved performance. In addition, the robustness of the proposed approach in dealing with noisy measurements is tested.

- **Chapter 4:** This chapter deals with the problem of time-varying signal reconstruction. A novel weighted-ℓ_1-based method is reported which uses the signal of the previous time instance to extract an estimated probability model for the signal of interest, and then incorporates this model into the reconstruction process. The proposed method is shown to significantly reduce the number of samples needed for perfect recovery of the original signals. Moreover, it is demonstrated that the proposed method can be quite beneficial to use in low power ECG monitoring devices and real-time MRI modalities.

- **Chapter 5:** The thesis is concluded with a summary of the results, discussion of ongoing work, and directions for future research.

1.5 Notations

Throughout this thesis, vectors are denoted by small boldface letters (e.g. $\mathbf{s}, \mathbf{y}$). $\mathbf{y}_i$ is used to refer to the i^{th} element of $\mathbf{y}$ and $\mathbf{\Phi}_i$ to refer to the i^{th} column of $\mathbf{\Phi}$. Scalars are shown by small regular letters (e.g. n, k) and matrices are denoted by bold capital Greek letters (e.g. $\mathbf{\Phi}, \mathbf{\Psi}$). Superscript (t) added to a vector/matrix refers to that of time/iteration t. Notation $\mathbf{y}|_{\mathcal{S}}$ is used to denote the sub-vector containing the elements of $\mathbf{y}$ with indices belonging to $\mathcal{S}$.Moreover, cardinality of a vector $\mathbf{y}$ is denoted by $|(\mathbf{y})|$.

Chapter 2

Sparse Signal Recovery

2.1 Introduction

Let the original signal/image be $\mathbf{s} \in \mathcal{R}^n$ and $\mathbf{S}$ be the a linear transform of it, i.e. $\mathbf{S} = \mathbf{\Psi s}$, that has z zeros. It is said that $\mathbf{S}$ is $k-$sparse, where $k = n - z$. When dealing with real data, $\mathbf{S}$ is called *approximately* $k-$sparse, if it can be closely approximated by a $k-$sparse vector $\mathbf{S}^*$, such that $\|\mathbf{S} - \mathbf{S}^*\|_2 < \epsilon$.

Throughout this section for simplicity, and without loss of generality[1], it is assumed $\mathbf{\Psi} = I$. Σ_k^n is also defined to be the set of all $k-$sparse signals, i.e. $\Sigma_k := \{x \in \mathcal{R}^n : |supp(x)| \leq k\}$.

Consider a measurement system that acquires m linear measurements:

$$\mathbf{y} = \mathbf{\Phi s}$$

with cardinality of $|\mathbf{y}| = m$.

[1]For an orthonormal basis $\mathbf{\Psi}$, the null space of $\mathbf{\Phi\Psi}^{-1}$ is a rotation of $\mathbf{\Phi}$, and such a rotation does not alter the success rate of CS recovery [10].

The null space of $\mathbf{\Phi}$ is defined as :

$$\mathcal{N}(\mathbf{\Phi}) := \{x \in \mathcal{R}^n : \mathbf{\Phi}x = 0\} \tag{2.1}$$

The matrix $\mathbf{\Phi}$, which is of size $m \times n$, represents a dimensionality reduction operator, i.e., it maps signals from $\mathcal{R}^n$, where n is generally large, into $\mathcal{R}^m$, where m is typically much smaller than n. In this case, measurements $\mathbf{y}$ are referred to as *compressive measurements*. It is also assumed that the measurement matrix is incoherent with the sparsity basis.

In the absence of some additional information concerning $\mathbf{s}$, it is obligatory to ensure that $\mathbf{\Phi}$ is invertible, in which case the original signal can be simply recovered via $\mathbf{s} = \mathbf{\Phi}^{-1}\mathbf{y}$. Unfortunately this requires full measurements (setting $m = n$).

The case where $m \leq n$ number of compressive samples is employed to reconstruct the original signal $\mathbf{s}$, leads to the problem of solving an under-determined systems of linear equations. Using the fact that the signal of interest is sparse, CS methodology is to search for the sparsest signal that satisfies the compressive measurements.

Equivalently, it is needed to find a $\hat{\mathbf{s}}$ which is an estimate of $\mathbf{s}$, having minimum number of nonzero elements, as a solution to the following optimization problem:

$$\begin{aligned} &\arg\min_{\hat{\mathbf{s}} \in \mathcal{R}^n} \ \|\hat{\mathbf{s}}\|_0 \\ &\text{s.t.} \quad \mathbf{\Phi}\hat{\mathbf{s}} = \mathbf{y} \end{aligned} \tag{2.2}$$

where $\|\hat{\mathbf{s}}\|_0 = |supp(\hat{\mathbf{s}})|$ and $supp(\hat{\mathbf{s}}) = \{i : \hat{\mathbf{s}}_i \neq 0\}$.

In is evident, however, that in many real world scenarios $\mathbf{s}$ can only be reasonably approximated with a k-sparse signal, in lieu of $\mathbf{s}$ of being truly k-sparse. Moreover, adding even a small amount of noise to observations, $\mathbf{y}$, will make (2.2) incapable of finding $\mathbf{s}$. To introduce some tolerance for noise and other errors, as well as robustness to approximately sparse signals, we would typically rather solve a variation to (2.2) presented in (2.3).

$$\begin{aligned} &\arg\min_{\hat{\mathbf{s}}\in\mathcal{R}^n} \ \|\hat{\mathbf{s}}\|_0 \\ &\text{s.t.} \quad \|\mathbf{\Phi}\hat{\mathbf{s}} - \mathbf{y}\|_2 \leq \epsilon \end{aligned} \tag{2.3}$$

Necessary and Sufficient reconstruction condition:

Theorem 1.1: [18] Every k-sparse vector $\in \Sigma_k^n$ is the unique solution of the ℓ_0-minimization problem (2.2) with $\mathbf{y} = \mathbf{\Phi}\mathbf{s}$ if and only if $\mathbf{\Phi} \in \mathcal{R}^{m\times n}$ satisfies the following null space property:

$$\Sigma_{2k}^n \cap \mathcal{N}(\mathbf{\Phi}) = \{0\} \tag{2.4}$$

where Σ_{2k}^n is the set of all $2k$-sparse vectors in $\mathcal{R}^n$.

Theorem 1.1 basically means that If $\mathbf{s}$ is $k-$sparse and the rank of $\mathbf{\Phi}$ is larger than $2k$, then the solution to (2.2) must be the signal $\mathbf{s}$.

It is also proven that for many random matrices $\mathbf{\Phi} \in \mathcal{R}^{m\times n}$ solving (2.2) one perfectly recovers all $k-$sparse signals $\mathbf{s}$ obeying $m \geq 2k$ with probability near one [19].

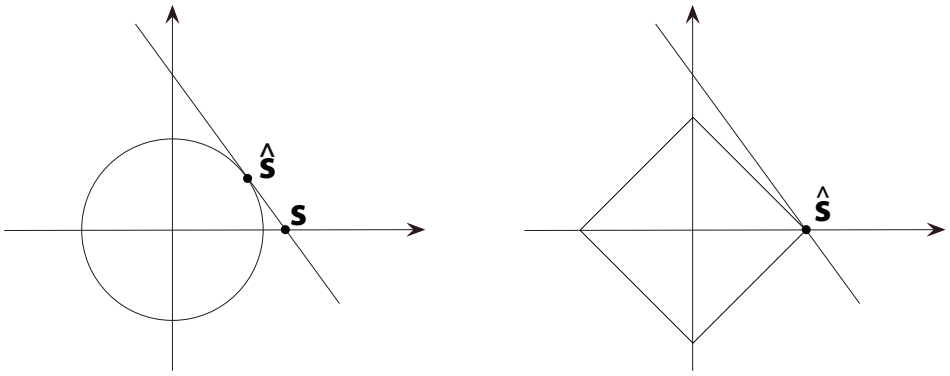
However, this result is of little practical use as solving the minimization problems (2.3) and (2.2) is acknowledged to be computationally unwieldy in view of

their non-convex and combinatorial nature.

2.1.1 Convex Relaxation

One classical alternative to solving the above inverse problem is by using least squares; that is, the vector with smallest ℓ_2 norm (energy) is selected [5]:

$$\begin{aligned} &\arg\min_{\hat{\mathbf{s}}\in\mathcal{R}^n} \ \|\hat{\mathbf{s}}\|_2 \\ &\text{s.t.} \quad \|\mathbf{\Phi}\hat{\mathbf{s}} - \mathbf{y}\|_2 \leq \epsilon \end{aligned} \tag{2.5}$$



(a) (b)

Figure 2.1: A sparse $k-$sparse vector, $\mathbf{s}$ and its recovery via (a) ℓ_2 minimization (b) ℓ_1 minimization.

There is even a convenient closed-form solution $\hat{\mathbf{s}} = (\mathbf{\Phi}^T(\mathbf{\Phi}\mathbf{\Phi}^T)^{-1}\mathbf{y})$. But unfortunately when the original vector $\mathbf{s}$ is k-sparse, ℓ_2 minimization will almost never find it. What is obtained instead is a non-sparse vector with plenty of ringing. This is due to the roundness nature of ℓ_2-norm ball, see figure 2.1.1(b). As

a consequence, the literature on compressive sampling primarily proposes another alternative convex minimization (2.6), where ℓ_1 norm of the $\hat{\mathbf{s}}$ has been used as an approximation to ℓ_0 (see, for instance, [2, 3, 20]).

$$\begin{aligned} &\arg\min_{\hat{\mathbf{s}}\in\mathcal{R}^n} \ \|\hat{\mathbf{s}}\|_1 \\ &\text{s.t.} \quad \|\mathbf{\Phi}\hat{\mathbf{s}} - \mathbf{y}\|_2 \leq \epsilon \end{aligned} \tag{2.6}$$

This problem is not only convex, but, in its noise-free settings, can actually be rewritten as a linear program and solved efficiently [21]. In the presence of noise, the minimization becomes a convex problem with *conic constraint*, for which accurate and efficient solvers exist.

It have been established in CS literature that the combinatorial problem (2.3) and its relaxation (2.6), are equivalent provided that $\mathbf{s}$ is sparse enough and $\mathbf{\Phi}$ satisfies certain properties (e.g. [3, 4]).

Necessary and Sufficient reconstruction condition:

Definition 1.1: A matrix $\mathbf{\Phi} \in \mathcal{R}^{m\times n}$ is said to satisfy the null space property of order k if for all subsets $S \subset n$ with $|S| = k$ it holds:

$$\|\mathbf{v}|_S\|_1 < \|\mathbf{v}|_{\hat{S}}\|_1 \ \ for \ \ \forall \mathbf{v} \in \mathcal{N}(\mathbf{\Phi}) \tag{2.7}$$

Theorem 1.2: [22] Every k-sparse vector $\in \Sigma_k^n$ is the unique solution of the ℓ_1-minimization problem (2.6) with $\mathbf{y} = \mathbf{\Phi}\mathbf{s}$ if and only if $\mathbf{\Phi}$ satisfies the null space property of order k.

The null space property is generally difficult to verify directly. Instead, the *restricted isometry property* [21], which was introduced by E. Candes and T. Tao in [2], has become very popular in compressive sensing.

Definition 1.2: Restricted Isometery Property (RIP): A matrix $\mathbf{\Phi} \in \mathcal{R}^{m \times n}$ is said to satisfy RIP of order k with constant δ_k if for any $\mathbf{s} \in \Sigma_k^n$:

$$(1 - \delta_k)\|\mathbf{s}\|_2 \leq \|\mathbf{\Phi s}\|_2 \leq (1 + \delta_k)\|\mathbf{s}\|_2 \tag{2.8}$$

Sufficient reconstruction condition:

Theorem 1.3: [2, 4] If $\mathbf{\Phi}$ satisfies the RIP with $\delta_{2k} < \sqrt{2} - 1$ then it also satisfies the null space property of order k.

This means that any $\mathbf{s} \in \Sigma_k^n$ can be recovered from its measurement $\mathbf{y} = \mathbf{\Phi s}$, via solving ℓ_1-minimization problem. However, it should be noted that RIP is a sufficient but not necessary condition.

It is acknowledged that checking whether a measurement matrix satisfies a certain RIP is computationally intensive, and becomes rapidly intractable as the size of the matrix increases [18, 23, 24]. However, there are certain important classes of matrices for which this property is verifiable. For example in the case where $\mathbf{\Phi}$ is chosen randomly from a *independent and identically distributed* (i.i.d.) Gaussian distribution, if the number of observations obeys $m \leq ck(\log n)$, for some constant $c > 0$, then by using the ℓ_1 minimization approach, the signal of interest can be recovered with a probability that exceeds $1 - O(n^{-\delta})$ [20].

Empirically, for such measurement matrix and n in the range of a few hundred

to a few thousand, the ℓ_1-norm minimization method can be expected to recover signals more than 50% of the time if the size of the number of samples obeys $m \geq 4k$, and if $m \geq 8k$, then the recovery rate is above 90%.

2.1.2 Greedy Algorithms

The ℓ_1-minimization approach is based on an optimization, which has relatively high complexity. It is acknowledged that even solving the linear program for ℓ_1 optimization has polynomial running time (often $O(n^3)$ [13]). This could be infeasible in applications where n is quite large. To deal with the complexity of running time and storage requirement of the convex-based methods, another class of greedy-based methods has been proposed in CS literature [14].

There are two broad main categories of greedy methods in CS, namely *greedy pursuits* [25, 26] and *Iterative thresholdings* [5, 11, 12].

Input: $\mathbf{y}$
$r^{(0)} = y, \hat{\mathbf{s}}^{(0)} = \mathbf{0}$
while $i \leq \max_{iteration}$ **do**
  $g^{(i)} = \mathbf{\Phi}^T r^{(i-1)}$;
  $j^{(i)} = \arg\max_j |g_j^{(i)}| / \|\mathbf{\Phi}_j\|_2$;
  $\hat{\mathbf{s}}^{(i)} = \hat{\mathbf{s}}^{(i-1)} + g_{j^i}^{(i)} / \|\mathbf{\Phi}_{j^i}\|_2^2$;
  $r^{(i)} = r^{(i-1)} - \mathbf{\Phi}_{j^i} g_{j^i}^{(i)} / \|\mathbf{\Phi}_{j^i}\|_2^2$;
end
Output: $\hat{\mathbf{s}}$

Algorithm 1: Matching Pursuit algorithm [5].

Greedy pursuits algorithms, such as Matching Pursuit (MP) [25] and Orthogonal Matching Persuit (OMP) [26], all iteratively build an estimate to $\mathbf{s}$ and they

share two fundamental steps, as summarized in Algorithm 1:

- (1) elements selection from column of $\mathbf{\Phi}$ to add to the support set which is initially empty
- (2) residual error update

These methods often result in extremely fast and efficient algorithms that are applicable to large dimensional data sets. It should be noted however, that number of samples needed for a perfect recovery is often higher than what is needed in ℓ_1 minimization [5].

Another group, which include Iterative Hard Thresholding (IHT) method [12], Subspace Pursiut [27] and CS Matching Pursiut [28], are called thresholding algorithms as they remove the nonzero elements in each iteration. This is summarized in Algorithm 2. These methods are easy to implement and relatively fast. Moreover, their reconstruction performance guarantee is similar to those of ℓ_1 minimization-based methods.

Input: $\mathbf{y}$
$\hat{\mathbf{s}}^{(0)} = \mathbf{0}$
while $i \leq \max_{iteration}$ **do**
  $\hat{\mathbf{s}}^{(i+1)} = \mathbf{H}_{\epsilon}(\hat{\mathbf{s}}^{(i)} + \mu\mathbf{\Phi}^T(\mathbf{y} - \mathbf{\Phi}\hat{\mathbf{s}}^{(i)}))$;
end
Output: $\hat{\mathbf{s}}$

Algorithm 2: Iterative Hard Thresholding algorithm [12].

2.1.3 Non-convex Alternative to ℓ_1 norm

The key difference between the ℓ_1 and ℓ_0 norm is the dependence of ℓ_1 on magnitude. Unlike ℓ_0 norm that penalizes all nonzero elements equally, in ℓ_1 norm larger coefficients are penalized more heavily than smaller coefficients. It is observed that the ℓ_1 norm is very sensitive to high amplitudes of non-zero elements since it may, in order to keep the ℓ_1 norm small, *choose* a signal with a higher number of nonzero elements with small magnitudes than the one with the correct smaller number of non-zero elements with high values, despite the fact that the latter is sparser.

For example, consider a signal $\mathbf{s}$ of size $n = 503$ which has only 10 non-zero elements (figure 2.2). Since $n = 503$ is a prime number, we should theoretically be able to reconstruct [2] the sparse solution uniquely with $|m| \geq 2k$. 30 random samples is then taken from its Fourier spectrum, and use them in the ℓ_1 norm minimization method to reconstruct the signal. From the result shown in figure 2.2, it can be seen that the ℓ_1 method does not reconstruct the signal with any satisfactory accuracy, and, moreover, is not as sparse as the original one. It has, in fact, 29 non-zero elements and the peak values are also reduced.

There have been some attempts in the CS literature to use non-convex regularizers in lieu of the convex ℓ_1 minimization, in order to further reduce the minimum required number of samples, for a perfect recovery [29]. Chartrand [15] suggested the potential of solving ℓ_p-minimization (2.9) to improve sparse signal

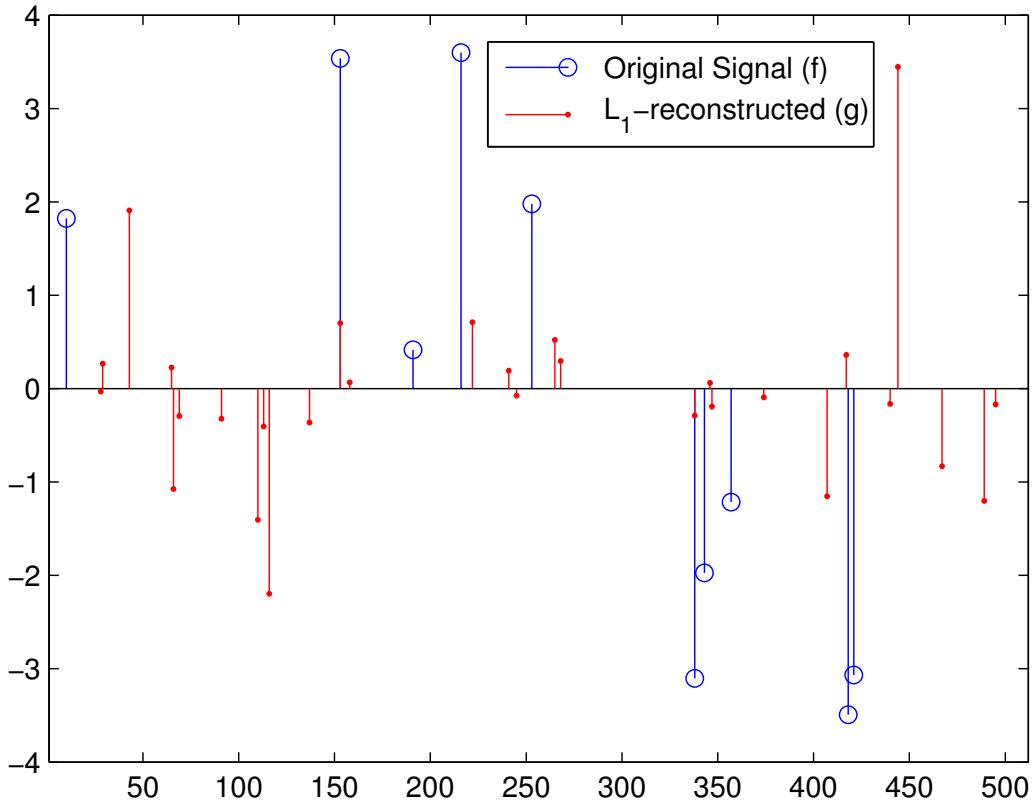


Figure 2.2: Illustrative Example: ℓ_1 norm reconstruction, $n = 503$, $k = 10$, $m = 30$, $\|\mathbf{s}\|_1 = 24.20$, $\|\hat{\mathbf{s}}\|_1 = 19.02$

reconstruction.

$$\begin{aligned} &\arg\min_{\hat{\mathbf{s}}\in\mathcal{R}^n} \ \|\hat{\mathbf{s}}\|_p \\ &\text{s.t.} \quad \mathbf{\Phi}\hat{\mathbf{s}} = \mathbf{y} \end{aligned} \tag{2.9}$$

where $p \in (0, 1)$ and $\|\mathbf{s}\|_p = (\sum_{i=1}^{n} |\mathbf{s}_i|^p)^{1/p}$.

Figure 2.2 depicts the ℓ_p norm ball for different values of p, from which it is evi-

dent that ℓ_p norm is a non-convex (but locally convex) function. It can be seen that ℓ_p approaches ℓ_0 as p gets closer to 0.

Necessary and Sufficient reconstruction condition:

Definition 1.1: A matrix $\mathbf{\Phi} \in \mathcal{R}^{m \times n}$ is said to satisfy the null space property of order k if for all subsets $S \subset n$ with $|S| = k$ it holds:

$$\|\mathbf{v}|_S\|_p < \|\mathbf{v}|_{\hat{S}}\|_p \quad for \quad \forall \mathbf{v} \in \mathcal{N}(\mathbf{\Phi}) \tag{2.10}$$

Theorem 1.2: [15] Every k-sparse vector $\in \Sigma_k^n$ is the unique solution of the ℓ_p-minimization problem (2.6) with $\mathbf{y} = \mathbf{\Phi s}$ if and only if $\mathbf{\Phi}$ satisfies the null space property of order k [15].

It should be noted that the above results are associated with the case where a global minima of (2.9) is achieved. In practice, solving (2.9) for $p < 1$ is non-trivial and computationally hard, as the cost function is non-convex and has many local minimas. On the other hand, it has been empirically demonstrated in CS literature that even finding a local minima of (2.9) can result in exact reconstruction of signals with many fewer measurements, compared with solving the conventional ℓ_1 problem [15]. Moreover, it has been shown that the least-squares solution often serves as a initialization point that is sufficiently close to the global optimum [29].

To find the minima of (2.9), Chartrand [15] proposed a projected gradient descent based algorithm using $(\sum_i (\mathbf{s}_i^2 + \epsilon^2)^{p/2})^{1/p}$, which is a smoothed approximation to ℓ_p, as the cost function. The value of ϵ is set to a large value initially

and then solution of each iteration serves as a starting point for the next iteration with a smaller value of ϵ.

Input: $\mathbf{y}$
Output: $\hat{\mathbf{s}}$
while $i \leq \max_{iteration}$ **do**
  find $\hat{\mathbf{s}}^{(i)} = \arg\min \|\mathbf{W^{(i)}s}\|_1 \;\; s.t. \;\; \mathbf{y} = \mathbf{\Phi s}$;
  $\mathbf{W^{(i+1)}} = 1/(|\hat{\mathbf{s}}^{(i)}| + \epsilon^{(i)})^{1-p}$;
end

Algorithm 3: ℓ_q algorithm [29]

Another approach that has been proposed in [19,29,30] is to solve (2.9) using a sequence of re-weighted ℓ_1 optimizations, where weight is selected as a diagonal positive definite matrix which is a function of the previous solution and/or iteration (i.e $W^{(i)} = f(\hat{\mathbf{s}}^{(i-1)})$). Two important such methods, namely Re-weighted ℓ_1 minimization ($RW\ell_1$) [19] and ℓ_q minimization [29], which were proposed in parallel and independent of this work, are summarized in Algorithms 3 and 4.

It have been empirically reported that these re-weighted ℓ_1-based algorithms significantly outperform the ℓ_1 minimization-based methods, in terms of minimum number of required samples for perfect recovery. However, their complexity and storage requirement are even higher than those of the conventional ℓ_1 minimization. This makes them less applicable in dealing with large-scale and/or real-time problems such as image reconstruction [5].

The reminder of this chapter is organized as follows: Section 2.2 contains an outline of the proposed SPSA-based approach as applied to the reconstruction of images from compressive samples. Section 2.3 presents experimental results on the reconstruction of different images, along with a comparison of the performance

Input: $\mathbf{y}$
Output: $\hat{\mathbf{s}}$
while $i \leq \max_{iteration}$ **do**
  find $\hat{\mathbf{s}}^{(i)} = \arg\min \|\mathbf{W^{(i)}s}\|_1 \;\; s.t. \;\; \mathbf{y} = \mathbf{\Phi s}$;
  $\mathbf{W^{(i+1)}} = 1/(|\hat{\mathbf{s}}^{(i)}|)$;
end

Algorithm 4: Re-weighted ℓ_1 ($RW\ell_1$) minimization algorithm [19]

of the proposed approach with that of others in the literature.

2.2 Proposed SPSA-ℓ_p Algorithm

The high computational and memory requirements of the ℓ_1-based methods, and their conservative approximation to ℓ_0-norm, motivate us to propose the application of *Simultaneous Perturbation Stochastic Approximation* (SPSA) [31–33] by employing ℓ_p norm ($p < 1$), to the problem of high dimensional signal reconstruction from compressive samples.

SPSA method, first introduced in [31] and more fully analyzed in [32], is a powerful tool in dealing with high-dimensional optimization problems. The main difference of SPSA with the conventional optimization methods is that instead of calculating the actual gradient at each point, it relies on estimation of gradient , without any *direct* reference to it. The SPSA approximates the gradient using only two performance function observations per iteration, regardless of the dimension of the signal, n. The two observations are made by simultaneously randomly varying all the variables of the minimization problem. This feature makes SPSA fast and efficient, particularly in dealing with high-dimensional problems and noisy

measurements. Moreover, it has been shown that the SPSA algorithm can achieve convergence in probability to a global optimum under fairly general conditions [34] [33]. Its reliance on estimation of gradient instead of the actual value, works like injected noise into the procedure, which allows the algorithm to escape the local minimas and premature convergence, in early iterations [33].

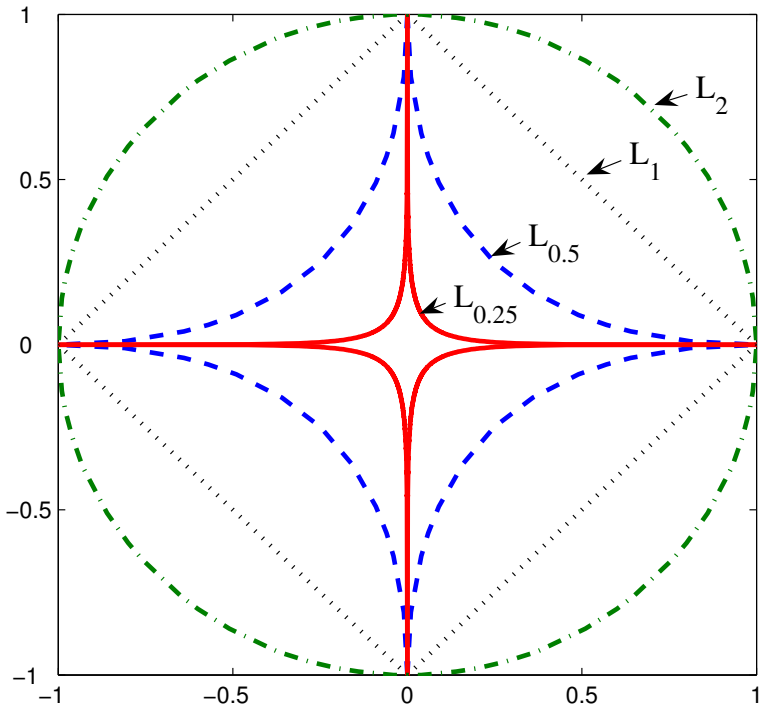


Figure 2.3: Level sets of different norms.

Let $\mathcal{F}(\theta)$ be the function to be minimized over θ. It is assumed that measurements of $\mathcal{F}(\theta)$ are available at various values of θ. Starting from a $\theta^{(\mathbf{0})}$, at each step, all elements of θ are simultaneously perturbed according to a distribution vector ($\delta^{(\mathbf{k})}$), whose elements are generated by Bernoulli distribution. Then $\mathcal{F}(\theta^{(\mathbf{k})} + \beta^{(k)}\delta^{(\mathbf{k})})$ and $\mathcal{F}(\theta^{(k)} - \beta^{(k)}\delta^{(\mathbf{k})})$ are evaluated to obtain an estimation of

the gradient at θ. θ is then updated as follows:

$$\theta^{(k+1)} = \theta^{(k)} - \alpha^{(k)} \frac{\mathcal{F}(\theta^{(k)} + \beta^{(k)}\delta^{(\mathbf{k})}) - \mathcal{F}(\theta^{(k)} - \beta^{(k)}\delta^{(\mathbf{k})})}{2\beta^{(k)}} \delta^{(-\mathbf{k})} \tag{2.11}$$

where, $\alpha^{(k)}$ is a gain sequence defined as $\alpha^{(k)} := \frac{\alpha}{(B+k+1)^{\gamma}}$, $\beta^{(k)}$ is a small positive time-varying constant $\beta^{(k)} := \frac{\beta}{(k+1)^{\lambda}}$. The parameters $B, \alpha, \beta, \lambda, \gamma$ are to be set by the user. The algorithm is terminated when it converges to a solution.

The standard SPSA which invokes the Lagrangian multiplier method to include constraints. In contrast, a different strategy is employed to deal with constraints for both noise-less and noisy measurements: in each iteration, θ is projected onto the set of feasible solutions.

In the case of measurements free from noise, let the set of θ that satisfies the observations be denoted by $\mathcal{O} = \{\theta : \mathbf{\Phi}\theta = \mathbf{y}\}$. Then, with $\mathbf{\Phi}^{\dagger}$ denoting the pseudo-inverse of $\mathbf{\Phi}$, $P(\theta) = (\theta - \mathbf{\Phi}^{\dagger}(\mathbf{\Phi}\theta - \mathbf{y}))$ is the nearest point to θ on $\mathcal{O}$, in other words, $P(\theta)$ is the projection of θ onto set $\mathcal{O}$.

The problem of finding the correct sparse solution to (2.9) is equivalent to the problem of minimizing the ℓ_0 norm. Therefore, the function $\mathcal{F}$ to be minimized in SPSA should ideally be the ℓ_0 norm. However, as mentioned in section 2.1 the ℓ_0 norm is difficult to handle. Therefore, $\mathcal{F}$ is set to be ℓ_p-norm, where p is gradually decreasing. It is empirically observed that starting from $p = 1$ and decreasing p gradually helps the algorithm scape local minimas. The algorithms starts with setting $p = 1$, the convergence point of the algorithm with this value of p is then used as the initial point for the algorithm using a smaller value of p.

Input: $\mathbf{y}$
Output: $\hat{\mathbf{s}}$
$\hat{\mathbf{s}} \leftarrow 0$;
$p \leftarrow 1$;
while $\hat{\mathbf{s}}|_{test} \neq \mathbf{y}|_{test}$ ***and*** $p \geq p_{min}$ **do**
  $k \leftarrow 0$;
  $\theta^{(k)} \leftarrow \hat{\mathbf{s}}$ **while** $\|\theta^{(k+1)} - \theta^{(k)}\|_2 > \epsilon$ ***and*** $k < \max_{it}$ **do**
    $\alpha^{(k)} \leftarrow \frac{\alpha}{(B+k+1)^{\gamma}}$, $\beta^{(k)} \leftarrow \frac{\beta}{(k+1)^{\lambda}}$;
    $\theta^{(k+)} \leftarrow \theta^{(k)} + \beta^{(k)}\delta^{(\mathbf{k})}$;
    $\theta^{(k-)} \leftarrow \theta^{(k)} - \beta^{(k)}\delta^{(\mathbf{k})}$;
    $\grave{\theta}^{(k+1)} \leftarrow \theta^{(k)} - \alpha^{(k)}\frac{\|\theta^{(k+)}\|_p - \|\theta^{(k-)}\|_p}{2\beta^{(k)}}\delta^{(-k)}$;
    $\theta^{(k+1)} \leftarrow \grave{\theta}^{(k+1)} - \mathbf{\Phi}^{\dagger}(\mathbf{\Phi}\grave{\theta}^{(k+1)} - \mathbf{y}))$;
    $k \leftarrow k + 1$;
  end
  $\hat{\mathbf{s}} \leftarrow \theta^{(k)}$;
  $p \leftarrow p - 0.1$;
end

Algorithm 5: SPSA ℓ_p-minimization algorithm

This procedure continues till a desired accuracy is reached or p gets smaller than the allowed p_{min}. As it will be demonstrated in section 2.3.1, with an experiment, even a slight decrease in the value of p from 1 to 0.9, significantly improves the reconstruction performance of the algorithm. However, it is noticed that decreasing p beyond 0.5, does not result in any substantial gain in the reconstruction performance (See figure 2.4). Therefore, p_{min} is set to 0.5.

Further, by way of checking the reconstruction accuracy, an online mode of verification of the reconstruction performance is employed by using a set of test samples as *test samples* ($\mathbf{y}|_{test} \subset \mathbf{y}$) which is not used in the minimization process.

After reconstruction, the computed values are checked with the test samples. If the (norm) difference between them is high or, in other words, if the desired accuracy in the final result is not reached, the value of p is decreased, otherwise the algorithm is terminated. The proposed SPSA-ℓ_p algorithm as applied to the problem is given in Algorithm 5.

2.2.1 Choice of SPSA parameters

Variables γ and λ in (2.11) are set to 0.602 and 0.101, respectively [32]. The constant B is recommended to be set as 10% (or less) of the maximum number of expected SPSA iterations. The parameter β is typically set to the standard deviation of the measurement noise. When the noise is small, β is chosen as 0.01. It is recommended that α be chosen such that the product $\alpha^{(0)} \frac{\|\theta^{(0)}+\beta^{(0)}\delta^{(0)}\|_p - \|\theta^{(0)}-\beta^{(0)}\delta^{(0)}\|_p}{2\beta^{(0)}}$ is approximately equal to the smallest desired step size during the early iterations. Of all the SPSA parameters, selecting α requires the most effort, because it is often difficult to know ahead of time what a good initial step size should be. If the value of α is chosen to be too small, it can significantly increase the number of SPSA iterations required to reach the minima, while large values of α lead to unstable, diverging solutions. However, it has been shown through experimental evaluation that a wide range of α values will yield satisfactory results [31]. Finally, the vector $\delta^{(k)}$ is so chosen as to have a Bernoulli distribution [35].

2.3 Experimental Results

The SPSA-ℓ_p algorithm is applied to both 1D and 2D synthetic signals/images. In addition, noisy samples are used for image reconstruction to test the robustness of the algorithm.

2.3.1 One-Dimensional Signals

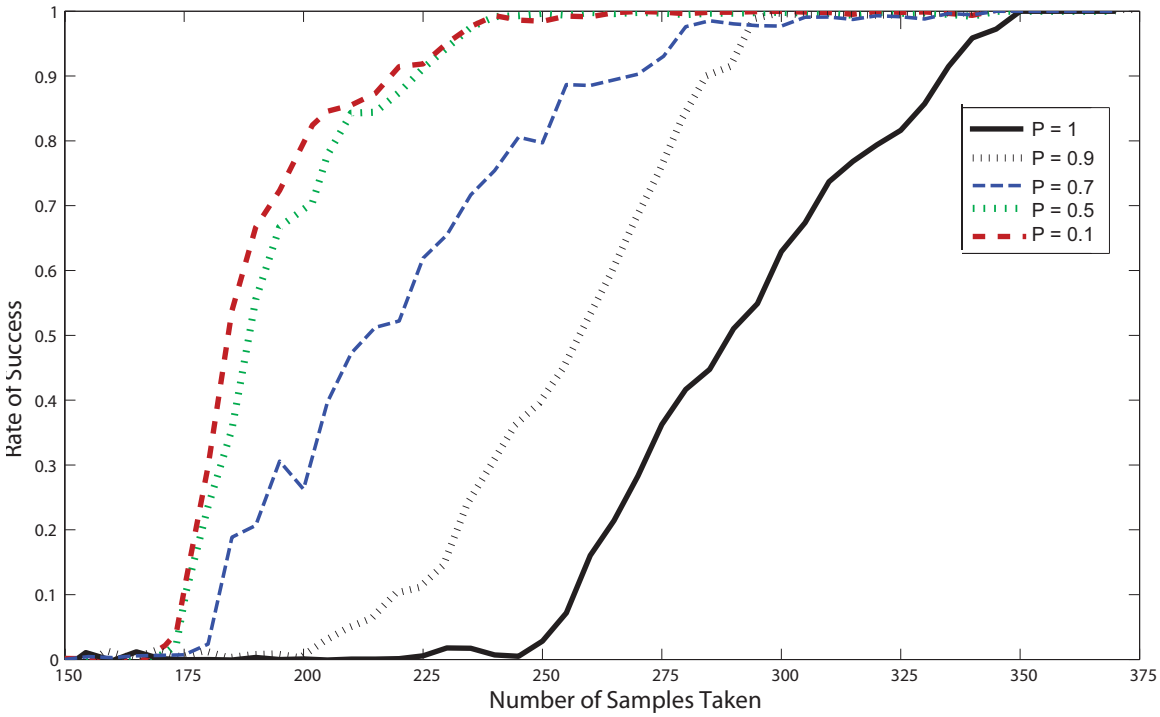


Figure 2.4: Comparison of reconstruction performance for different values of $p \leq 1$.

In the first experiment the effect of the different values of p on the minimum number of needed samples for perfect recovery of random signals with fixed sparsity, is explored. To this end, first 100 random signals of size $n = 512$ with $k = 100$

(which means that the percentage of the number of zeros in each signal is 78%) are generated. Then m number of samples (where m is arbitrary and between 150 to 350) is used to measure the signals. For each choice of m, the entries of the sampling matrices $\mathbf{\Phi}$ of size $m \times n$, are randomly selected from a mean-zero Gaussian distribution. The value of p is then set to be 1, 0.9, 0.7, 0.5 and 0.1. For each choice of p, all the 100 random signals are reconstructed using Algorithm 5. A reconstruction is considered successful, if MSE (2.12) is less than 10^{-4}. It was noted that the algorithm convergence rate decreases as p approaches zero.

The rate of success vs. number of samples taken is depicted in figure 2.4. It can be observed that with $p = 1$ around 350 samples are needed for perfect reconstruction of all signals. However, with a slight decrease in the value of p from 1 to 0.9, significantly fewer measurements are required ($m = 275$). Decreasing p even further, results in smaller number of required samples (250 and 225 for $p = 0.7$ and $p = 0.5$, respectively). However, it is noticed that decreasing p beyond 0.5, does not result in any substantial gain in the reconstruction performance. Therefore, in all other experiments of this section, p_{min} is set to 0.5.

Next, the performance of the proposed SPSA-$\ell_{0.5}$ is tested on two syntactic scenarios, where the number of samples obeys the constraints, $\frac{m}{k} = 2.87$ and$\frac{m}{k} = 2.5$, against the ℓ_1 method. Figures 2.6, 2.7, 2.8 and 2.9 show the results of ℓ_1 reconstruction along with those of SPSA-$\ell_{0.5}$ for $1D$ signals of size 512. It is observed that the proposed method outperforms the ℓ_1 approach in recovering the original signal more accurately. Since ℓ_1 norm is not very sensitive to small elements, its reconstruction results contain many elements of small magnitudes. Moreover, it fails to reconstruct the signal peaks.

$$MSE(\hat{\mathbf{s}}) = \frac{\sum_{k=1}^{n}(\hat{\mathbf{s}}_i - \mathbf{s}_i)^2}{n} \tag{2.12}$$

Next an extensive comparison of the recovery performance of the proposed method (SPSA-ℓ_p) with that of the currently used minimization-based CS methods, namely the ℓ_1-minimization ([36]), the Re-weighted ℓ_1 minimization ($RW\ell_1$ [19]) and the ℓ_q method [29], is presented. It should be noted that the proposed method is an iterative one with a substantially less complexity compared to the above mentioned methods. Therefore the proposed algorithm is also compared against another iterative and less complex method, namely Iterative Hard Thresholding (IHT) [12].

The setup of the experiment is similar to those of the first experiment. In 100 signal of size $n = 512$, 100 elements are randomly selected to have a random nonzero value , chosen from a mean-zero, unit-variance Gaussian distribution. 150 to 350 linear samples are then used to measure each signal. For each choice of m, the entries of the sampling matrices A of size $m \times n$, are randomly selected from a mean-zero Gaussian distribution. A signal is considered to be successfully recovered when error is below 10^{-4}. In ℓ_q method $q \in [0, 0.1, 0.2, ..., 0.9]$ and $\epsilon^{(k)} = 1/2^k$. Success rate for each reconstruction method vs. number of samples taken is reported in figure 2.3.1. It can be seen that the proposed method method significantly outperformed the ℓ_1, IHT and $RW\ell_1$ methods, in terms of smaller number of samples needed for perfect reconstruction of all 100 signals. Its performance is also superior, but comparable, to the one of ℓ_q method.

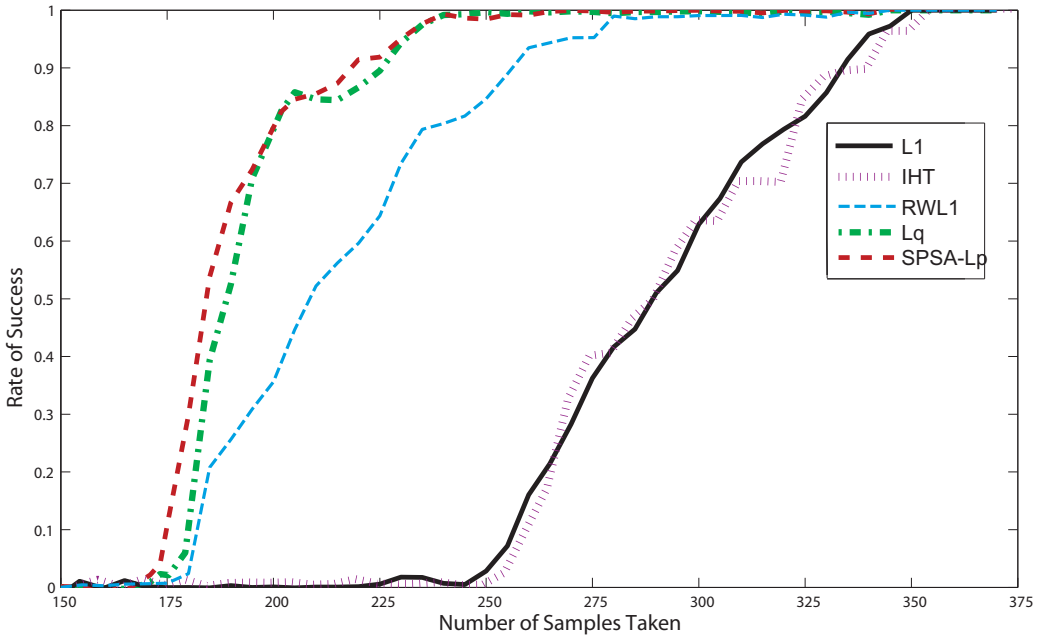
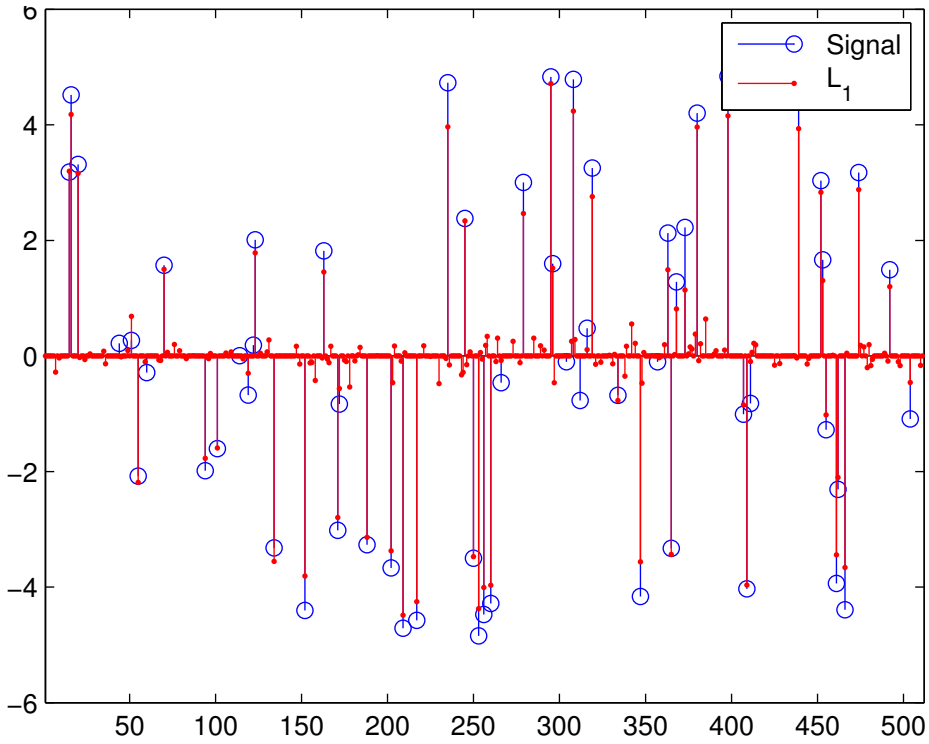
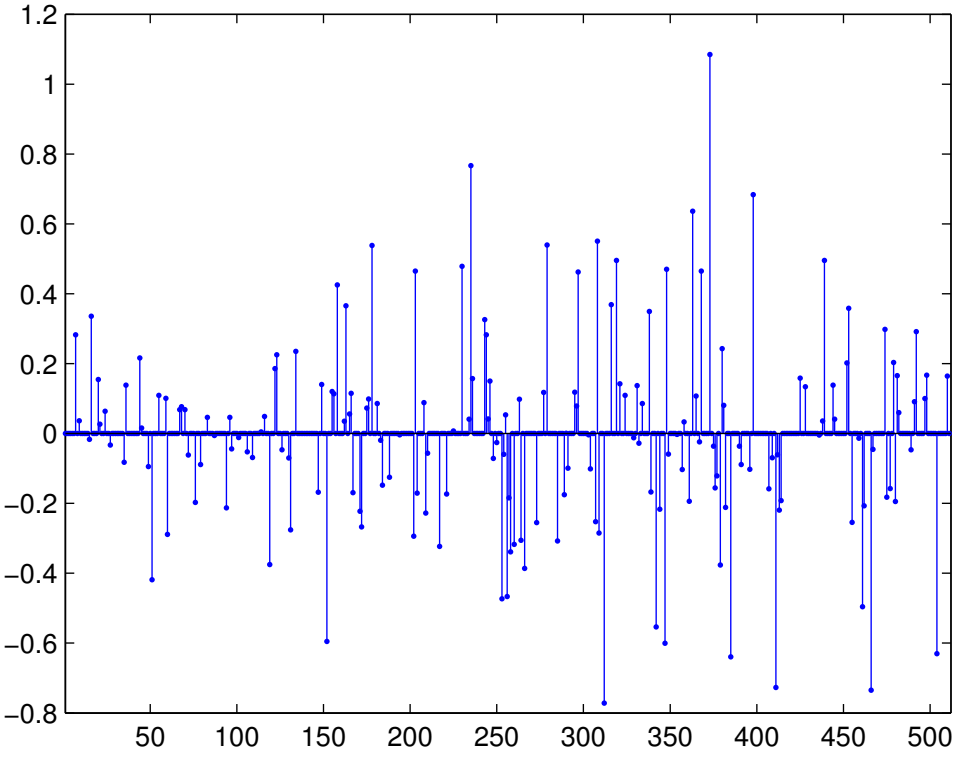


Figure 2.5: Comparison of perfect reconstruction rate for random signals with $n = 512$ and $k = 100$, vs. number of samples taken.

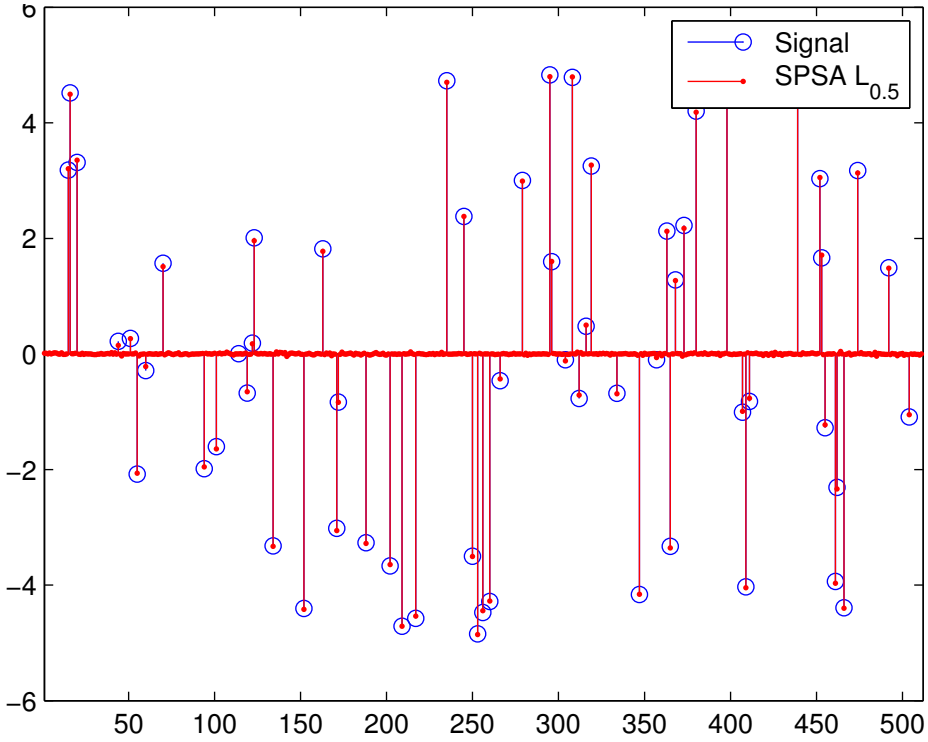


(a) ℓ_1, MSE=0.0276

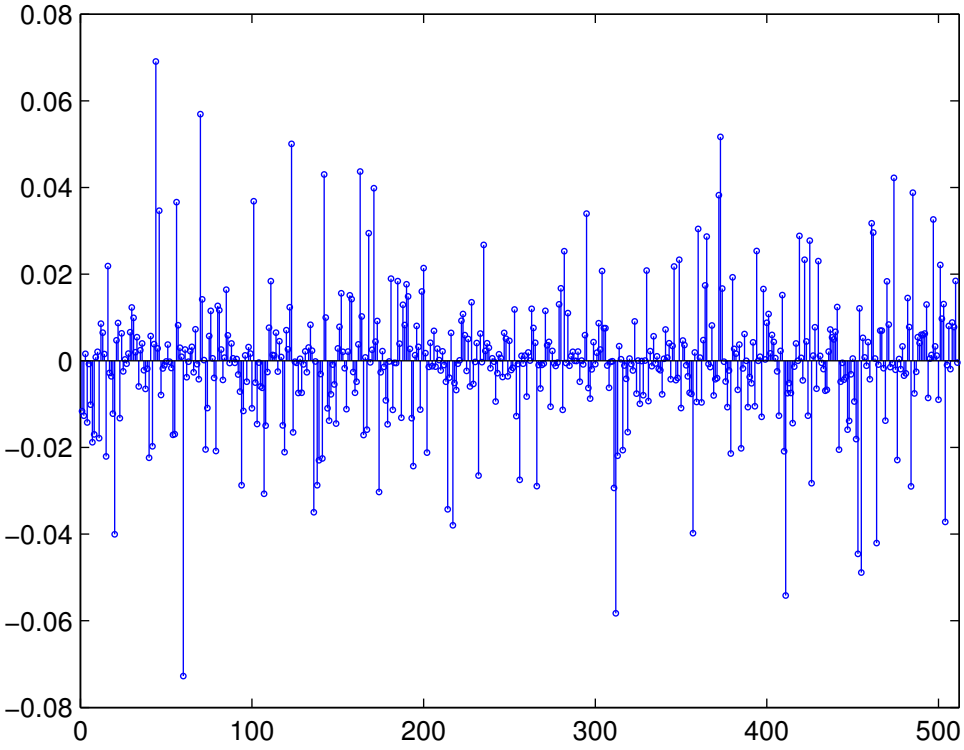


(b) point-wise error ℓ_1, MSE=0.0276

Figure 2.6: ℓ_1 reconstruction for $n = 512$, $k = 60$, $m = 172$

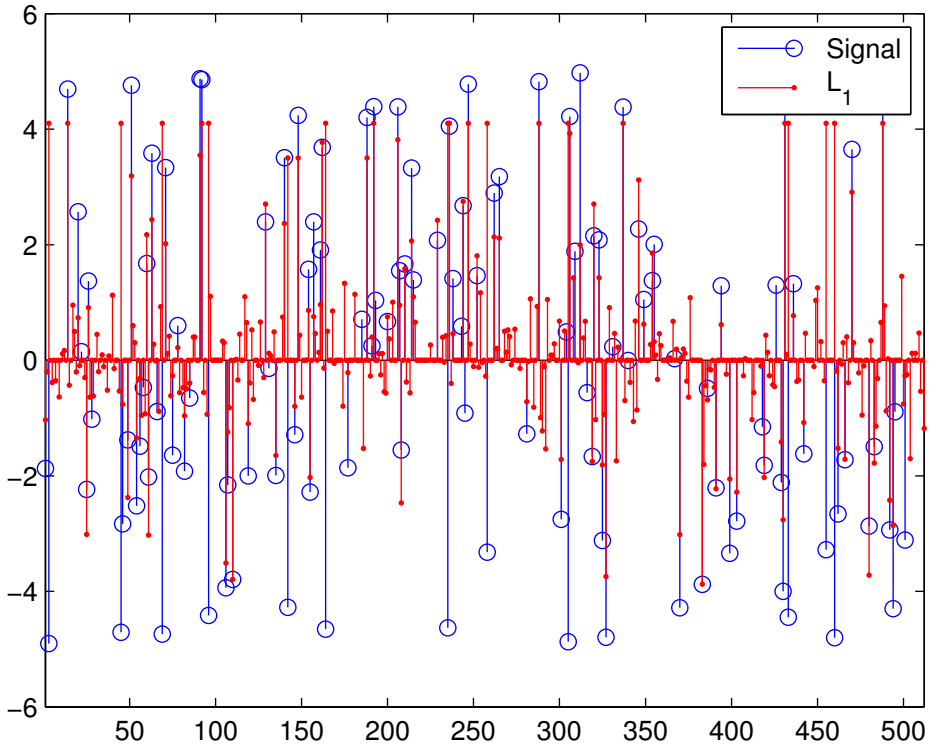


(a) $\ell_{0.5}$-SPSA, MSE=2.1192e-4

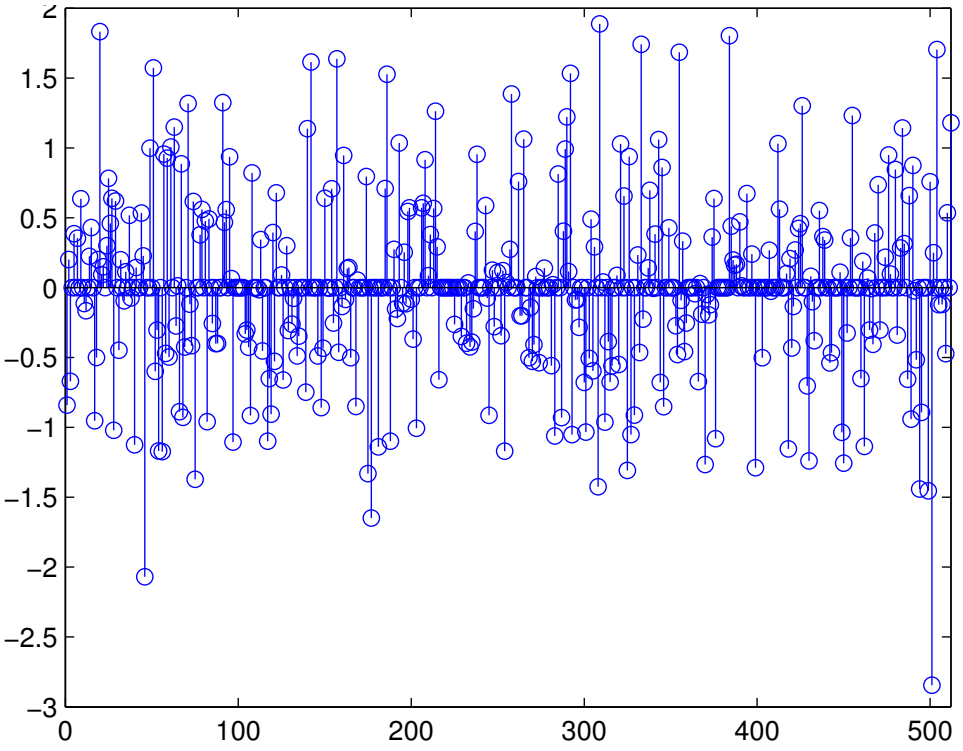


(b) point-wise error $\ell_{0.5}$-SPSA, MSE=2.1192e-4

Figure 2.7: SPSA-$\ell_{0.5}$ reconstruction for $n = 512$, $k = 60$, $m = 172$

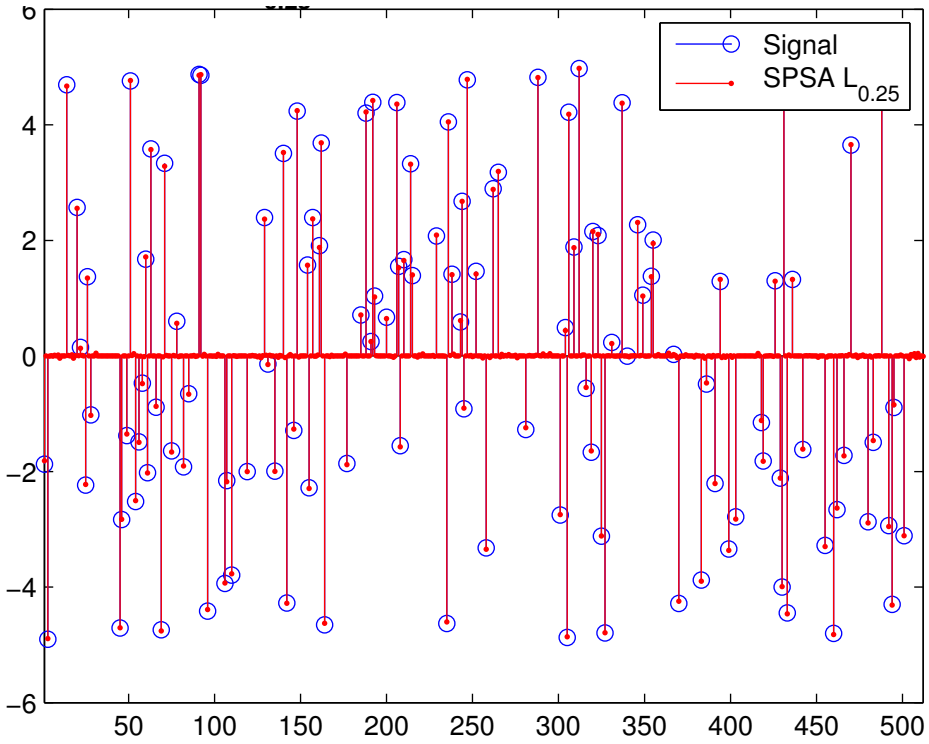


(a) ℓ_1, MSE=0.3329

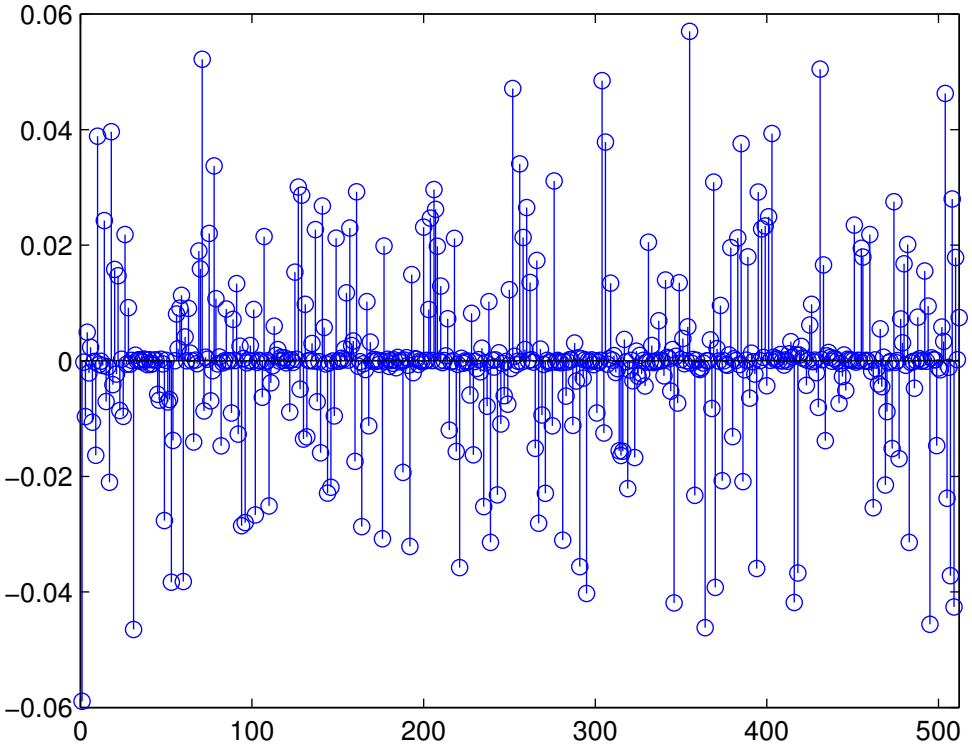


(b) point-wise error ℓ_1, MSE=0.3329

Figure 2.8: ℓ_1 reconstruction for $n = 512$, $k = 120$, $m = 300$



(a) $\ell_{0.5}$-SPSA, MSE=2.0743e-4



(b) point-wise error $\ell_{0.5}$-SPSA, MSE=2.0743e-4

Figure 2.9: SPSA-$\ell_{0.5}$ reconstruction for $n = 512$, $k = 120$, $m = 300$

2.3.2 Two-Dimentional Images

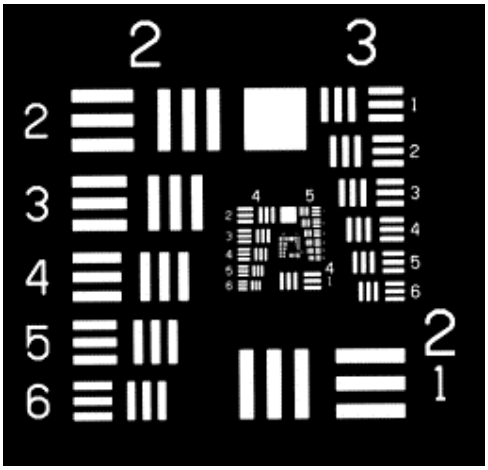
The algorithm is applied to a set of gray scale medical and natural images of size 512×512 (see figures 2.10 and 2.11). It is known that these images, as any other natural image, are quite sparse in wavelet domain, therefore reconstruction has been carried out in this domain. Samples are taken along random radial lines in Fourier domain (figure 2.11(d) shows a sampling mask).

Table 2.1 shows the PSNR (2.13) of the reconstructed image using the proposed method (SPSA-ℓ_p), compared with that of the currently used CS methods, namely the ℓ_1-minimization (using Nesta [36]), $RW\ell_1$ [19] and the ℓ_q method [29] (with $q \in [0, 0.1, 0.2, ..., 0.9]$). It is clear that the SPSA-$\ell_{0.5}$-based reconstruction is superior to reconstruction using the ℓ_1 and $RW\ell_1$ methods. Moreover, its performance is generally better, but comparable, with that of the ℓ_q method.

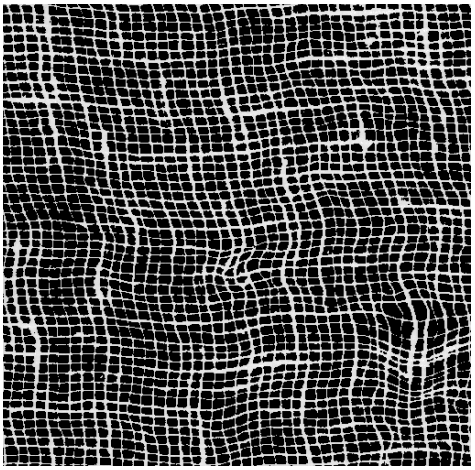
$$PSNR(\hat{\mathbf{s}}) = 20\log_{10}(\frac{255}{\sqrt{(MSE(\hat{\mathbf{s}}))}}) \tag{2.13}$$

Input Data		Reconstructed image PSNR			
Image name	% of Samples	$RW\ell_1$	ℓ_1	ℓ_q	SPSA-$\ell_{0.5}$
Boat	50%	32.41	32.23	32.84	33.12
Hill	39%	25.52	25.52	30.43	30.90
MRI	27%	35.90	34.11	37.90	39.27
Lena	46%	35.16	34.34	35.16	35.16
Peppers	39%	29.92	29.66	27.95	28.52
CT Image	27%	32.61	33.61	32.03	32.81
USAF Target Image	8%	14.99	14.99	48.23	51.31
Texture Image	7%	19.31	10.20	25.01	24.81

Table 2.1: PSNR of the reconstructed images.



(a) USAF Target



(b) Texture

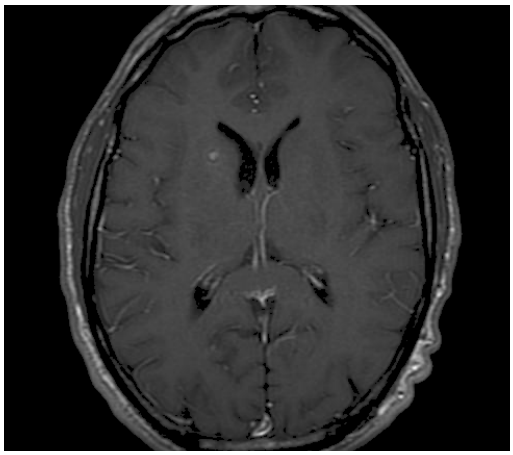
(c) Peppers

(d) Hill

(e) Boat

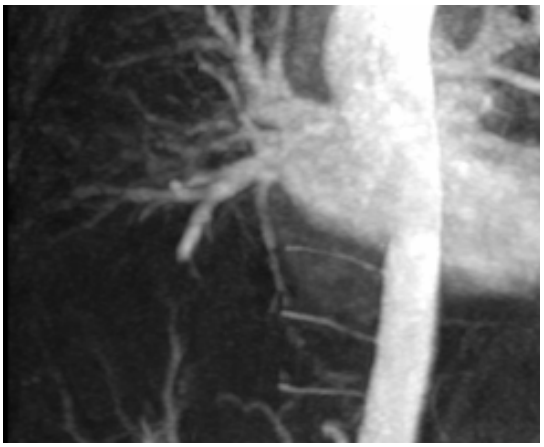
(f) Man

Figure 2.10: Original test images [1].

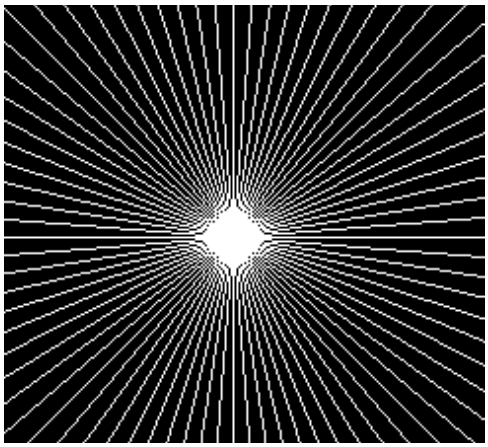


(a) MRI

(b) Lena



(c) CT



(d) Sampling Mask

Figure 2.11: (a-c) Original test images [1] (d) Sampling Mask.

2.3.3 Robustness

The robustness of the proposed method is also tested as applied to the earlier test images which are now corrupted with white Gaussian noise. As before, sampling is done along radial lines in the Fourier domain, and reconstruction, in the wavelet domain. Table 2.2 shows the PSNR of the noisy image and of the reconstructed images using the different methods. It can be seen that the superiority of the proposed method is even more evident when measurements are corrupted with noise.

Input Data		PSNR			
Image name	% of samples	noisy image	ℓ_1	ℓ_q	SPSA-$\ell_{0.5}$
Boat	50%	25.36	24.91	26.14	27.06
Hill	39%	26.12	26.54	26.78	26.78
MRI	27%	27.11	28.01	27.01	29.04
Lena	46%	25.12	25.34	23.69	23.89
Peppers	39%	25.55	23.64	23.12	26.09
CT Image	27%	26.73	26.12	26.73	27.43
USAF Target Image	8%	25.19	26.57	26.37	27.22
Texture Image	7%	23.62	26.12	26.87	26.51

Table 2.2: PSNR of the reconstructed noisy images.

2.4 Summary

In this chapter the non-convex ℓ_p quasi-norm ($0 < p < 1$) is introduced, as an alternative to the convex ℓ_1 minimization. A novel adaptation of SPSA to solve sparse recovery problems using ℓ_p norm, is also proposed. The reconstruction performance of the proposed method is extensively studied in both noise-free and noisy settings and the results are compared against the ones of the currently used methods in CS literature. It is shown that the proposed method outperforms other methods, in terms of smaller number of needed samples for perfect reconstruction.

Chapter 3

Measures of Sparsity

3.1 Introduction

In signal representation, *practical* sparsity can be defined in many ways. For instance, a signal is sparse if its ℓ_0-norm, i.e. number of the non-zero coefficients, is small compared to its dimension. If the original signal $\mathbf{s}$ needs n samples for its complete specification but has z zeros in it, it is said to be $k-$sparse, where $k = n - z$.

But in the case of real signals, this definition may not be practical. Alternatively, a signal is sparse, if its *energy* is concentrated in a small number of coefficients of its representation. In other words, when $\mathbf{s}$ has z elements with small magnitudes, we can extend the above definition to "approximate $k-$sparsity": if there exists a $\mathbf{s}^* \in \mathcal{R}^n$ which is $k-$sparse, and $\inf ||(\mathbf{s} - \mathbf{s}^*)||_p$ is small, where the subscript $p \leq 1$ denotes the ℓ_p (pseudo-) norm, then $\mathbf{s}$ is approximately $k-$sparse.

As it was mentioned in previous chapters, the core of CS is the search for

the sparsest signal. It is evident that to be able to to so, one should first be able to measure and quantify sparsity and this measure has a profound effect on the performance of the CS-based algorithms. Intuitively, a sparsity measure should depend on the relative distribution of energy among the coefficients, as a fraction of the total energy, and not be calculated based solely on the absolute value of each coefficient. In fact, a good measure should be a weighted sum of coefficients of signal representation, based on the importance of a particular coefficient in the overall sparsity. As a consequence, any slight change in the value of a coefficient will affect sparsity only relative to the weight of that coefficient, which is a desirable property. More explicitly, large coefficients should have a smaller weight compared to the small ones so that they do not influence the sparsity measure in a way that does not respond to the changes of the smaller coefficients.

In most of the current literature on compressive sampling, sparsity is measured using the ℓ_p $(0 \leq p \leq 1)$ norm of a vector. However, as it will be demonstrated in section 3.1.1, with the help of various examples, the ℓ_0, ℓ_1 and ℓ_p norms quantify sparsity in a way that runs counter to an intuitive understanding of sparsity. This observation that the norm-based sparsity measures, do not exhibit some desirable properties [37], serves as the motivation to explore the use of the Gini index (GI) as an alternative sparsity measure in the problem of signal/image reconstruction from compressive measures.

3.1.1 Gini index

Gini Index [37], which is a commonly used measure of income/wealth inequality in the literature of economics, is defined as follows:

Given a vector $\mathbf{s} = [s_1, ..., s_n]$, with its elements re-ordered and represented by $f_{[k]}$ for $k = 1, 2, \cdots, n$, where $|s_{[1]}| \leq |s_{[2]}|, \cdots, \leq |s_{[n]}|$, then

$$\text{Gini}\ (\mathbf{s}) \ = \ 1 \ - \ 2\sum_{k=1}^{n} \frac{|s_{[k]}|}{||\mathbf{s}||_1} \left(\frac{n-k+1/2}{n} \right) \tag{3.1}$$

where $||\mathbf{s}||_1$ is the ℓ_1 norm of $\mathbf{s}$.

Remark 1: Gini index is a quasi-convex function on $|\mathbf{s}|$.

Proof: GI, which has been defined in (3.1), can also be expressed as follows:

$$\text{GI}\ (\mathbf{s}) \ = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} ||s_i| - |s_j||}{2n||\mathbf{s}||_1} \tag{3.2}$$

From (3.2), we have

$$\text{GI}\ (\mathbf{s}) \ = \frac{2\sum_{i=1}^{n} \sum_{j=i+1}^{n} |||s_i| - ||s_j||}{2n||\mathbf{s}||_1} = \frac{2\left(||s_1| - |s_2|| + \cdots + ||s_{n-1}| - |s_n||\right)}{2n||\mathbf{s}||_1}$$

Suppose that $|s_a|$ is the $k-$th element in the sorted vector then $|s_a| = |s_{[k]}|$. In the numerator $|s_a|$ is compared with the other $(n-1)$ elements. It is clear that if $|s_a| = |s_{[k]}|$, then $(k-1)$ of $\mathbf{s}$ elements are smaller than $|s_a|$, while $(n-k)$ elements

of $\mathbf{s}$ are greater. Therefore, we have

$$\begin{aligned}&||s_1| - |s_2|| + \cdots + ||s_1| - |s_n|| + \cdots + ||s_{n-1}| - |s_n||\\ &= \sum_{k=1}^{n} (k-1)|s_{[k]}| - (n-k)|s_{[k]}|.\end{aligned}$$

Then it follows that

$$\begin{aligned}\text{GI}\,(\mathbf{s}) &= \frac{2\sum_{i=1}^{n}\sum_{j=i+1}^{n}||s_i| - |s_j||}{2n||\mathbf{s}||_1} = \frac{2\sum_{k=1}^{n}(k-1)|s_{[k]}| - (n-k)|s_{[k]}|}{2n||\mathbf{s}||_1}\\ &= \frac{2\sum_{k=1}^{n}k|s_{[k]}| - ||\mathbf{s}||_1 - n||\mathbf{s}||_1}{n||\mathbf{s}||_1} = \frac{n||\mathbf{s}||_1 + 2\sum_{k=1}^{n}k|s_{_{[k]}}| - ||\mathbf{s}||_1 - 2n||\mathbf{s}||_1}{n||\mathbf{s}||_1}\\ &= 1 - \frac{2\sum_{k=1}^{n}k|s_{[k]}| - ||\mathbf{s}||_1 - 2n||\mathbf{s}||_1}{n||\mathbf{s}||_1}\end{aligned}$$

which is the same as (3.1).

To show that $GI(\mathbf{s})$ is quasi-convex in $|\mathbf{s}|$, it is sufficient to show that the sublevel sets of Eqn. (3.2) are convex sets. The $c-$sublevel set can be written as:

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\Big||s_i| - |s_j|\Big|}{2n||\mathbf{s}||_1} \leq c$$

We can rewrite this as :

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\Big||s_i| - |s_j|\Big| - 2cn||\mathbf{s}||_1 \leq 0 \tag{3.3}$$

Since the first term on the left hand side of (3.3) can be rewritten as a point-wise maximum of linear expressions, it is convex. The second term is linear. Therefore, the above expression is convex.

3.1.2 GI vs. ℓ_p

An important advantage of the Gini index over the conventional norm measures is that it *is normalized*, and assumes values between 0 and 1 for any vector. Further, it is 0 for the least sparse signal with all the coefficients having an equal amount of energy; and 1 for the most sparse one which has all the energy concentrated in just one coefficient. This gives us a meaningful measure, exhibiting the sparsity of the distribution.

Moreover, unlike other norm measures, the value of this index is independent of the size of the vector, thereby enabling us to compare the sparsity of vectors of different sizes. It is also scale-invariant (i.e. $GI(\alpha\mathbf{s}) = GI(\mathbf{s})$), which means that multiplying all the coefficients of the image representation by a constant does not affect its sparsity (see table 3.1). Such a property, which is indeed desirable in the context of image reconstruction from sparse samples, is clearly not satisfied if we use the ℓ_p (pseudo-) norm as a sparsity measure ($\|\alpha x\|_p = \alpha\|x\|_p$). GI is independent of the total energy of the signal and as a consequence, it is ideally suited for comparing the sparsity of a signal in different transform domains. When applied to the problem of signal reconstruction from compressive samples, the Gini index facilitates the discovery of the sparsest domain of transform, if there is any.

Original Signal	ℓ_1	$\ell_{0.5}$	GI
[1 1 1 1 1 1 1 1 1 1]	10	100	0
[1 1 1]	3	9	0
$5 \times$ [1 1 1 1 1 1 1 1 1 1]	50	500	0

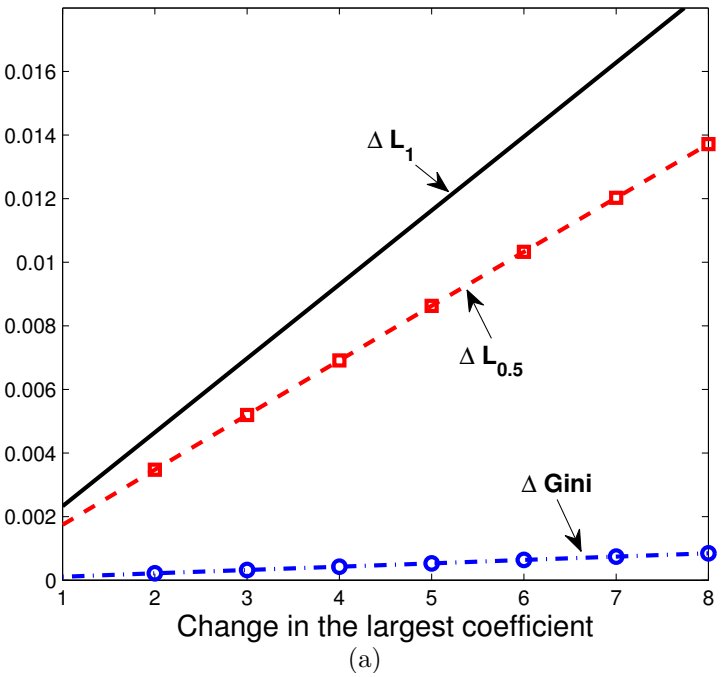
Table 3.1: GI vs. ℓ_1 and $\ell_{0.5}$.

For example, consider the phantom image (figures 3.8) of size 512×512. The Gini index of this image in different transform domains is presented in table 3.5, from which it can be easily concluded that the phantom image is most sparse in the gradient domain and least sparse in the DFT domain. Moreover, with such a definition, it turns out that there is no need to define explicitly approximate sparsity measures: when the Gini index is large (i.e., close to 1), then the signal has only a few values which are dominant; and when the Gini index is small, the signal has very few dominant values.

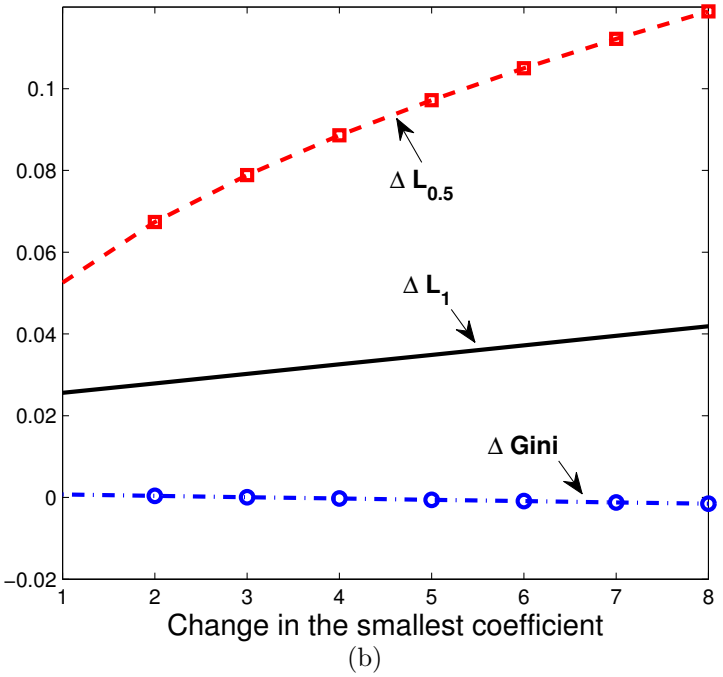
Moreover, it is observed that ℓ_0 does not take the energy distribution of the non-zero elements into account. Consider the two signals $\mathbf{X_1} = [10, 1, 1, 0]$ and $\mathbf{X_2} = [4, 4, 4, 0]$. Their sparsity according to ℓ_0 is the same for these two signals. However, intuitively, the sparsity of $\mathbf{X_1}$ should be more than $\mathbf{X_2}$, since most of the signal energy is concentrated in just one element

Next,through some examples, the effect of changing the magnitude of a single coefficient on the Gini index and also on the norm measures, is examined. This enables us to have a better understanding of the contribution of each coefficient in the sparsity measures and also of how GI compares to the other norm-based measures.

1) **Effect of increasing the peak value of a signal:** The following is intuitively expected: if there is a rise in the magnitude of the peak value, sparsity should not decrease; in other words, it should either remain the same or increase. This is more intuitive when expressed in terms of wealth distribution in a society, any increase in the wealth of rich people will widen the wealth gap and increase the sparsity.



(a)



(b)

Figure 3.1: Percentage of change in sparsity measures vs. change in (a) the largest and (b) the smallest coefficient magnitude.

As an example for two signals given by $\mathbf{X_1} = [100, 10, 0, 0]$ and $\mathbf{X_2} = [120, 10, 0, 0]$, it is apparent that the sparsity of $\mathbf{X_2}$ is not less than the sparsity of $\mathbf{X_1}$. However, as it can be seen in table 3.2, ℓ_1 and $\ell_{0.5}$, unlike the Gini index, identify the latter one to be less sparse.

	GI	ℓ_1	$\ell_{0.5}$
X_1	0.7045	110	173.24
X_2	0.7115	130	199.28

Table 3.2: Sparsity measures for X_1 & X_2.

In figure 3.1(a), the magnitudes of change in the highest coefficient have been plotted against the percentage of respective changes in each measure. It can be seen that the norm measures tend to rise with the increase in the peak values of the signal, which means that the measured sparsity is reduced. This conclusion is *not* correct. In contrast, the rise in the Gini index implies *correctly,* that the signal is getting sparser with the increase in the peak value.

2) **Effect of increasing the value of the smallest coefficient:** It is intuitive that sparsity should decrease with the rise of a small coefficient; In the context of wealth distribution rising magnitude of a small coefficient corresponds to the increase in the wealth of poor people, which clearly reduces the difference (sparsity) between poor and rich people.

Figure 3.1 (b), shows the contribution of the rise in the value of the smallest coefficient in the sparsity measures. It can be seen that all measures, correctly judge that the sparsity is reduced. However, it can be observed

that $\ell_{0.5}$ is quite sensitive to the change in the value of the small coefficient, which makes this measure vulnerable to noise.

Domain	GI	ℓ_1	$\ell_{0.5}$
Original signal	0.9474	440	3.49E3
noisy signal	0.9237	451.06	9.71E3

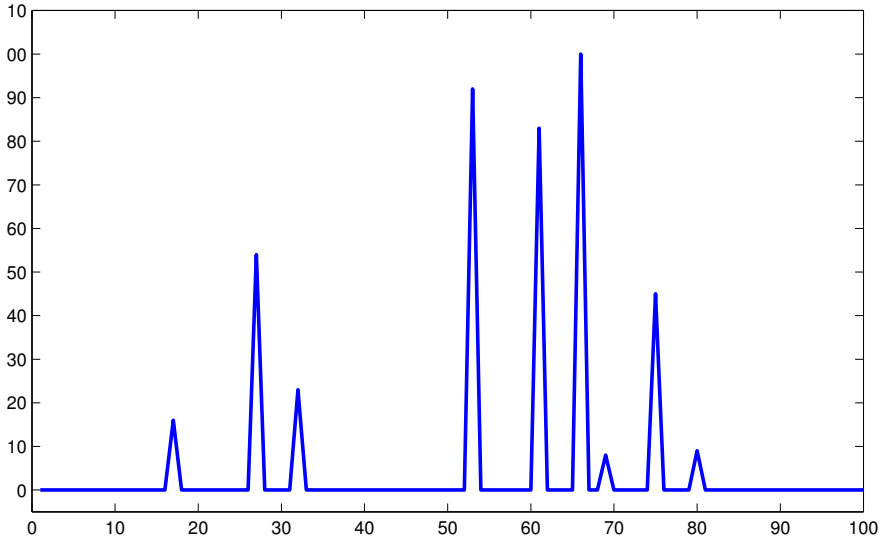
Table 3.3: Sparsity measures for signal in figure 3.2.

Figure 3.2 (b) shows a random signal corrupted with white Gaussian noise (SNR=10 dBW). Sparsity measures of this signal vs. its original are presented in table 3.3. It can be seen that the value of the $\ell_{0.5}$ has increased dramatically for the noisy signal which implies that the noisy signal is much less sparse compared to its original one.

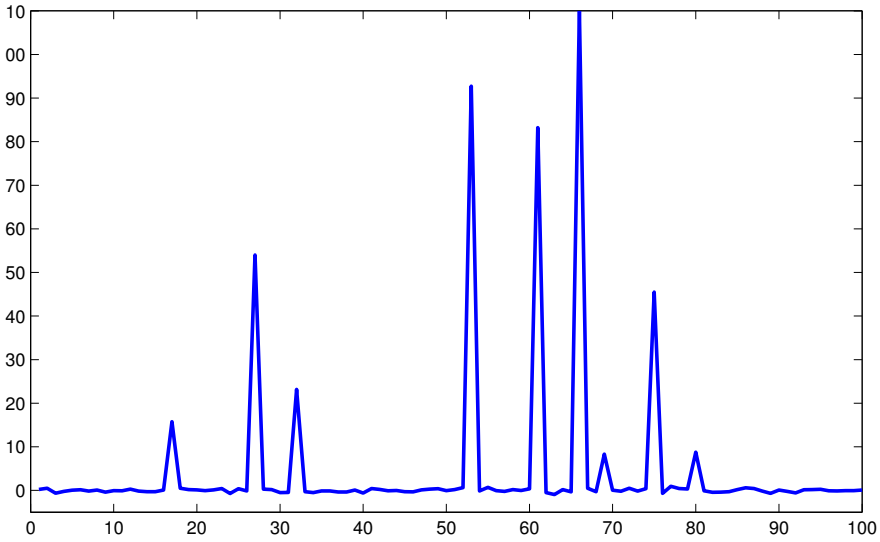
The above observations serve as motivation for the use of the Gini index as a sparsity measure for signal reconstruction from sparse samples. For a more detailed discussion on the concept of sparsity and comparison of different measures of sparsity as applied to different classes of problems, see [37].

It should be noted that, there seem to be no mathematical results that deal with conditions on the measurement matrix $\mathbf{\Phi}$ for recovering uniquely the original signals from compressive samples when the Gini-index is invoked as a sparsity measure. The existing constraints of null space and restricted isometry properties (RIP) of $\mathbf{\Phi}$ for unique reconstruction of $k-$sparse signals from compressive samples seem to be no longer applicable here.

Moreover, it should be noted that the Gini index, which is neither a norm nor, even, a *pseudo- norm* (like the ℓ_p for $0 < p < 1$), has some undesirable



(a) Original Signal



(b) noisy Signal, SnR=10

Figure 3.2: A random sparse signal.

characteristics, too. For instance, the Gini indices of two signals cannot be algebraically manipulated to get the Gini index of a composite signal. Further, we cannot analyze the sparsity of a signal by decomposing it into smaller segments, and computing the sparsity of each for summation later.

3.2 Proposed approach

Based on the Gini index as a sparsity measure, two classes of problems of reconstruction from compressive samples, can be formulated : in the first class, the Gini index of the signal is given, i.e., $\text{Gini}\ (\mathbf{\Psi s}) = \gamma_0$; and, in the other, it is not.

Problem 1: Given the measurements $\mathbf{s}$ find a vector $\hat{\mathbf{s}}$ (which is an estimate of $\mathbf{s}$), as a solution to the following equations:

$$\begin{aligned} &\text{Gini}\ (\mathbf{\Psi}\hat{\mathbf{s}}) = \gamma_0 \\ &\text{such that}\ \ \mathbf{\Phi}\hat{\mathbf{s}} = \mathbf{y} \end{aligned} \tag{3.4}$$

Problem 2: Given the measurements $\mathbf{y}$, find a vector $\hat{\mathbf{s}}$ (which is an estimate of $\mathbf{s}$), as a solution to the following optimization problem:

$$\begin{aligned} &\arg\max_{\hat{\mathbf{s}} \in R^n} \text{Gini}\ (\mathbf{\Psi}\hat{\mathbf{s}}) \\ &\text{such that}\ \ \mathbf{\Phi}\hat{\mathbf{s}} = \mathbf{y} \end{aligned} \tag{3.5}$$

It appears that Problem 1 is meaningful only when some prior knowledge of the sparsity of signals to be reconstructed is available. Therefore, Problem 2 is the one considered in the next section, where the proposed approach for the reconstruction of signals from sparse samples by maximizing the Gini index is presented.

When the measurement is corrupted by noise, the signal reconstruction problem can be formulated as a corollary to the above.

Corollary: Given the measurements $\mathbf{y} = \mathbf{\Phi s} + \eta$, where η is the noise vector, satisfying the inequality $||\eta||_2 \leq \epsilon$, find a vector $\hat{\mathbf{s}}$ (which is an estimate of $\mathbf{s}$), as a solution to the following optimization problem:

$$\begin{aligned} &\arg\max_{\hat{\mathbf{s}} \in \mathcal{R}^n} \text{Gini}\ (\mathbf{\Psi}\hat{\mathbf{s}}) \\ &\text{such that}\ \ ||\mathbf{\Phi}\hat{\mathbf{s}} - \mathbf{y}||_2 \leq \epsilon \end{aligned} \tag{3.6}$$

The goal is to find the sparsest solution, in the sense of the Gini index, among the set of admissible signals (i.e., the samples). To this end, we make use of the "Simultaneous Perturbation Stochastic Approximation" (SPSA) to find a solution to (3.5). It should be noted that the subject of maximization of quasi-convex functions is well known, and some general conditions for global solutions have been discussed in [38], [39].

In the reconstruction of real images, any optimization algorithm must be able to contend with the high dimension of the problem and noisiness in the measurements of the chosen objective function. SPSA has been found to be efficient for

this purpose by providing a satisfactory solution using a relatively small number of measurements of the objective function. SPSA uses observations of the desired performance function (which, in our application, is the Gini index of the signal) without a *direct* reference to its gradient. It approximates the gradient using only two performance function observations per iteration, regardless of the dimension of the signal. The two observations are made by simultaneously varying randomly all the variables in the performance function. This is the essential feature of SPSA, which provides its power and relative ease of use in difficult multivariate optimization problems.

Due to its stochastic nature, SPSA allows for noisy inputs as measurements of an image corrupted by noise. In contrast, many image reconstruction algorithms of the literature are affected somewhat unpredictably by noise (unless special care is taken), leading to problems of convergence and a dramatic decrease in efficiency.

Further, in the case of the SPSA, convergence conditions for both local [31] and global optimization in the face of multiple, local optima under fairly general conditions, including piecewise differentiability of the objective function and Lipschitz continuity of its gradient, have been established in the literature [33,34].

SPSA Algorithm:

Assume that $\mathbf{s}$ is the original sparse vector which is being measured using a matrix $\mathbf{\Phi}$, and $\mathbf{y}$ is the set of observations, $\mathbf{y} = \mathbf{\Phi s}$ and $\mathbf{GI}(\mathbf{\Psi}\theta)$ is the function to be maximized over θ. It should be noted that GI is piecewise differentiable and its gradient is Lipschitz continuous. It is assumed that measurements of $\mathbf{GI}(\mathbf{\Psi}\theta)$ are available at various values of θ.

Let $\mathcal{O}$ be the set of all θ that satisfy the observations, $\mathcal{O} = \{\theta | A\theta = y\}$, and $P(\theta)$ be the nearest point to θ on $\mathcal{O}$ or, in other words, projection of θ into set $\mathcal{O}$. In the case of measurements free from noise, let the set of θ that satisfies the observations be denoted by $\mathcal{O} = \{\theta : \mathbf{\Phi}\theta = \mathbf{y}\}$. Then, with $\mathbf{\Phi}^\dagger$ denoting the pseudo-inverse of $\mathbf{\Phi}$, $P(\theta) = (\theta - \mathbf{\Phi}^\dagger(\mathbf{\Phi}\theta - \mathbf{y}))$ is the nearest point to θ on $\mathcal{O}$, in other words, $P(\theta)$ is the projection of θ onto set $\mathcal{O}$. For the initialization, the algorithm stars from multiple random points and the best answer is stored.

Input: $\mathbf{y}$
Output: $\hat{\mathbf{s}}$
$k \leftarrow 0$;
$\theta^{(k)} \leftarrow \theta_0$
while $\|\hat{\mathbf{s}}|_{test} - \mathbf{y}|_{test}\|_2 < \epsilon$ ***and*** $\|\theta^{(k+1)} - \theta^{(k)}\|_2 > \epsilon$ ***and*** $k < \max_{it}$ **do**
  $\alpha^{(k)} \leftarrow \frac{\alpha}{(B+k+1)^\gamma}$, $\beta^{(k)} \leftarrow \frac{\beta}{(k+1)^\lambda}$;
  $\theta^{(k+)} \leftarrow \theta^{(k)} + \beta^{(k)}\delta^{(\mathbf{k})}$;
  $\theta^{(k-)} \leftarrow \theta^{(k)} - \beta^{(k)}\delta^{(\mathbf{k})}$;
  $\grave{\theta}^{(k+1)} \leftarrow \theta^{(k)} + \alpha^{(k)}\frac{\mathbf{GI}(\mathbf{\Psi}\theta^{(k+)}) - \mathbf{GI}(\mathbf{\Psi}\theta^{(k-)})}{2\beta^{(k)}}\delta^{(\mathbf{k})}$;
  $\theta^{(k+1)} \leftarrow \grave{\theta}^{(k+1)} - \mathbf{\Phi}^\dagger(\mathbf{\Phi}\grave{\theta}^{(k+1)} - \mathbf{y}))$;
  $k \leftarrow k + 1$;
end
$\hat{\mathbf{s}} \leftarrow \theta^{(k)}$;

Algorithm 6: SPSA GI-maximization algorithm

In the case of measurements with noise, the set $\mathcal{O}$ becomes $\mathcal{O}_\epsilon = \{\theta : ||\mathbf{\Phi}\theta - \mathbf{y}||_r < \epsilon\}$, and $P(\theta) = \arg\min_{\mathbf{x}}\{||\mathbf{x} - \theta||_2 \ s.t. \ \mathbf{x} \in \mathcal{O}_\epsilon\}$.

Starting from the point θ_0, at each step, all elements of $\theta^{(k)}$ are perturbed simultaneously according to a distribution vector ($\delta^{(k)}$) whose elements are gen-

erated by a Bernoulli distribution. We then evaluate $\mathbf{GI}(\mathbf{\Psi}(\theta^{(k)} + \beta^{(k)}\delta^{(k)}))$ and $\mathbf{GI}(\mathbf{\Psi}(\theta^{(k)} - \beta^{(k)}\delta^{(k)}))$, and update θ as follows:

$$\theta^{(k+1)} = P\left(\theta^{(k)} + \alpha^{(k)}\frac{\mathbf{GI}(\mathbf{\Psi}(\theta^{(k)} + \beta^{(k)}\delta^{(k)})) - \mathbf{GI}(\mathbf{\Psi}(\theta^{(k)} - \beta^{(k)}\delta^{(\mathbf{k})}))}{2\beta^{(k)}}(\delta^{(k)})^{-1}\right) \tag{3.7}$$

where the gain sequence $\alpha^{(k)} := \frac{\alpha}{(B+k+1)^{\gamma}}$, and the small positive time-varying constant $\beta^{(k)} := \frac{\beta}{(k+1)^{\lambda}}$ and $P(\theta) = (\theta - A^{\dagger}(A\theta - \mathbf{s}))$. Here $A^{\dagger}$ is the pseudo-inverse. The algorithm terminates when it converges to a solution or when the maximum number of iterations is reached.

3.3 Experimental Results

The proposed algorithm is applied to both 1D and 2D synthetic signals/images. In addition, to test the robustness of the algorithm, images are reconstructed from noisy samples.

3.3.1 One-Dimensional Signals

Consider a simple low dimensional example taken from [30]. It has been shown in [30] that ℓ_p fails to reconstruct any 1-sparse signal using the following sampling

(i.e., measurement) matrix:

$$\mathbf{\Phi} = \begin{bmatrix} \frac{1}{\sqrt{1+2^{1-2/p}}} & \frac{2^{1/p-1}}{\sqrt{1+2^{1-2/p}}} & \frac{2^{1/p-1}}{\sqrt{1+2^{1-2/p}}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}$$

Let $\mathbf{s} = [1 \;\; 0 \;\; 0]^T$ and $p = 0.5$.

Then $\mathbf{\Phi} = \begin{bmatrix} 0.3333 & 6 & 6 \\ 0 & 0.7071 & -0.7071 \end{bmatrix}$ and $\mathbf{y} = \mathbf{\Phi s} = [0.3333 \;\; 0]^T$.

The reconstruction results based on the Gini index and $\ell_{0.5}$ norm are compared in table 3.4. It can be seen that the Gini index can recover signals *even* in those cases where the ℓ_p norm-based recovery fails.

Table 3.4: 1D example

Original Signal	$[1 \;\; 0 \;\; 0]^T$	ℓ_1	$\ell_{0.5}$	Gini
$\ell_{0.5}$ Reconstruction	$[0 \;\; 0.0278 \;\; 0.0278]^T$	0.0556	0.1111	0.3333
GI-based Reconstruction	$[1 \;\; 0 \;\; 0]^T$	1	1	0.6667

In the second experiment, the goal is to find empirical bounds on the number of measurements needed for perfect recovery of signals using GI maximization. 100 random signals are generated of size $n = 1000$ with $k = 100$, which means that the percentage of the number of zeros in each signal is 90. Figure 3.3 shows one such random signal. m number of samples (where m is arbitrary) is then used to measure the signals. For each choice of m, the entries of the sampling matrices A, of size $m \times N$, have been randomly selected from a mean-

zero Gaussian distribution. With the GI as the sparsity measure, the signals are reconstructed, and compared with those obtained by employing norms as the sparsity measures (Note that in all of the experiments, the minimization of $\ell_{0.5}$-norm and ℓ_1 has been done using, the algorithm in [40] and NESTA [36], respectively and a reconstruction is considered perfect if MSE (2.12) is less than 10^{-3}).

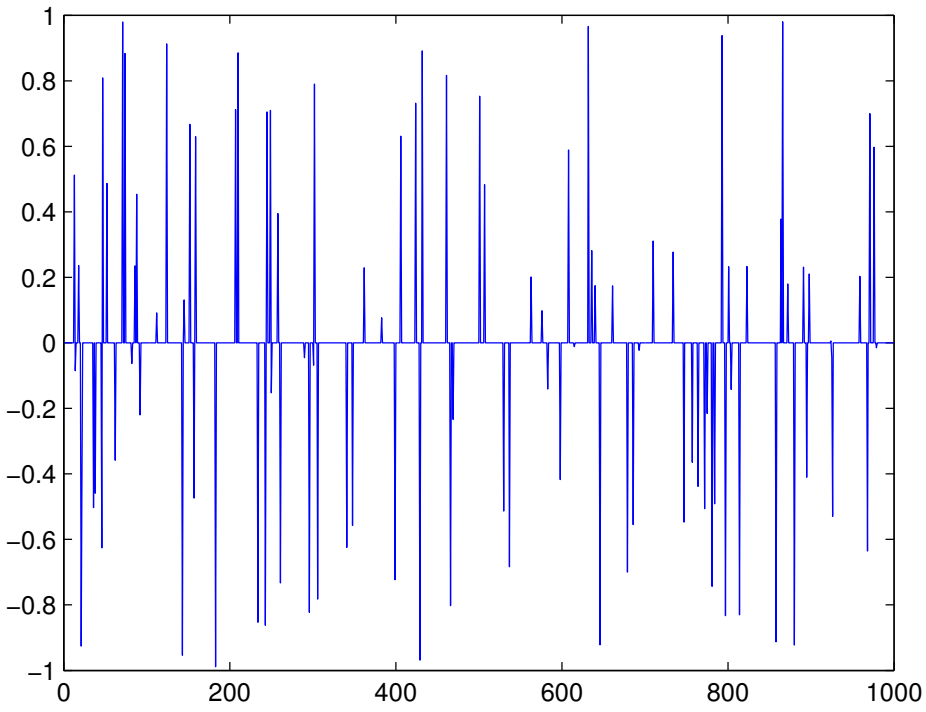


Figure 3.3: Example of a random signal ($n = 1000$, $k = 100$).

Figure 3.4 shows the percentage of success in perfect signal recovery vs. the number of measurements taken, for all measures of sparsity. It is observed that the proposed method outperforms the ℓ_1 and ℓ_p approaches in recovering the original

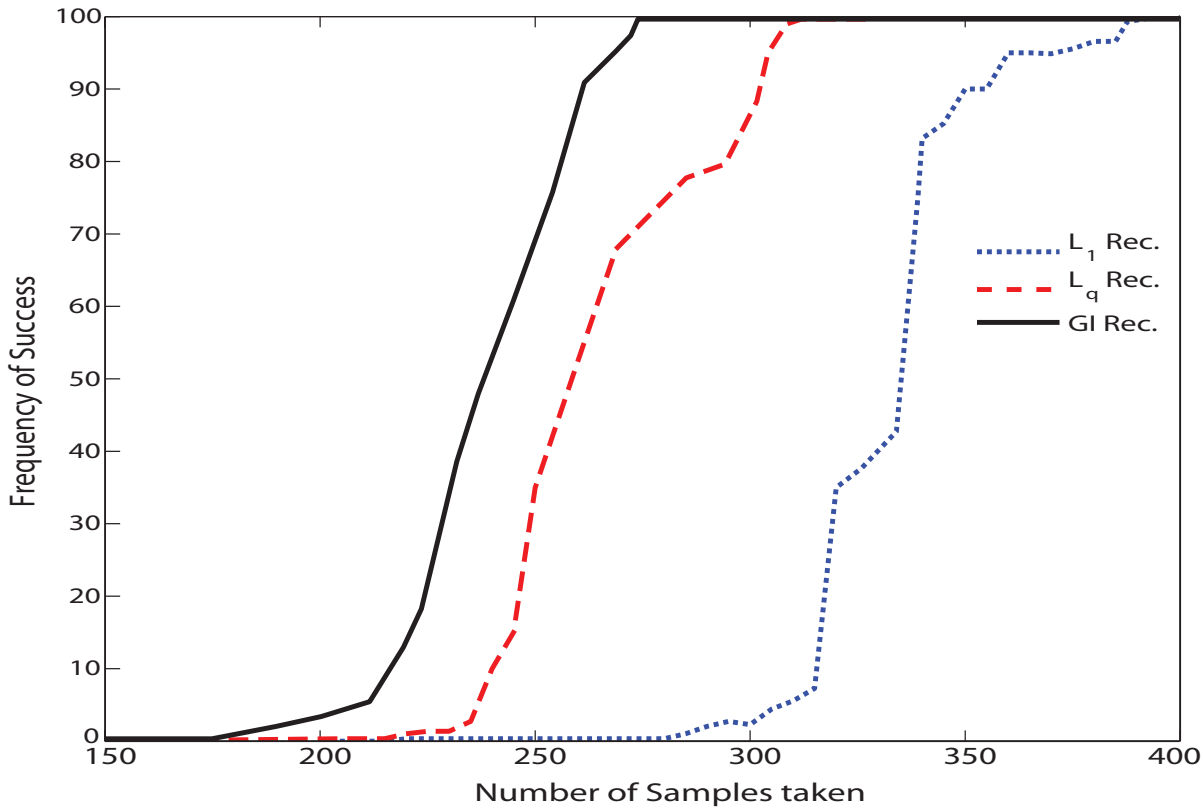
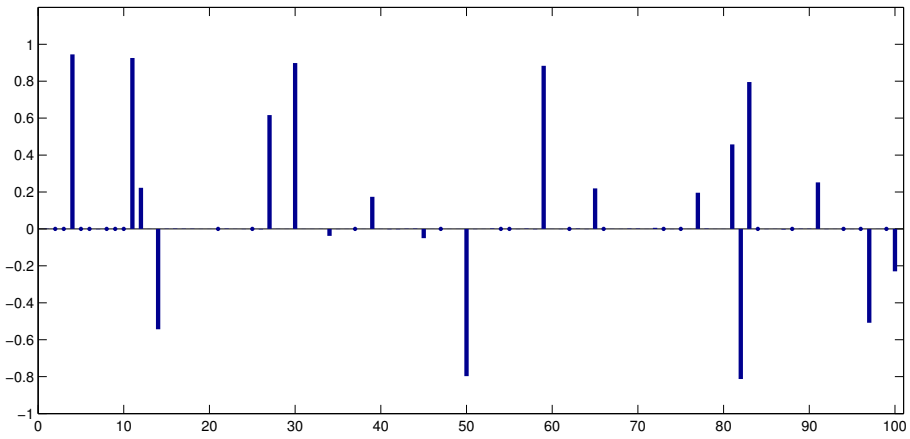


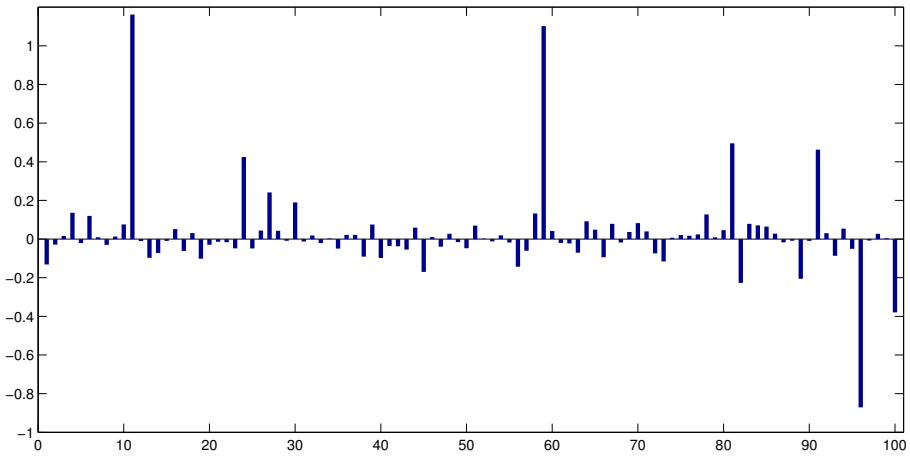
Figure 3.4: Percentage of prefect recovery vs. number of samples taken.

signal.

To further illustrate the performance of the proposed method, the reconstruction result of ℓ_1, ℓ_p vs. GI is depicted in figures 3.5 and 3.6 for one of these random signals over a segment of size 100. It can be seen that GI has reconstructed the signal while other two methods have not been successful. ℓ_1 fails to recover many of the signal peaks due to its tendency to keep the nonzero elements as small as possible. ℓ_p reconstructed signal is sparser than to ℓ_1, however it is not as sparse as the original signal.

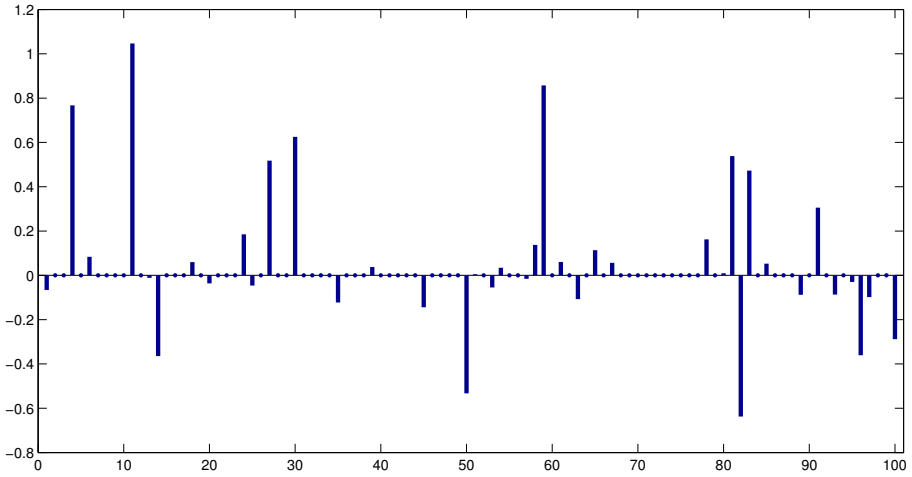


(a) One segment of the original signal

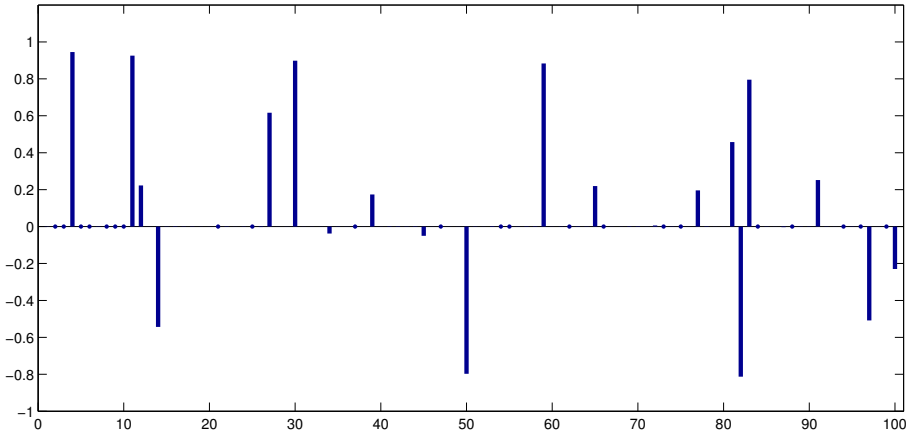


(b) ℓ_1 reconstruction

Figure 3.5: ℓ_1 reconstruction for a segment of a random signal ($n = 1000, k = 100$) using 300 random samples.



(a) ℓ_p reconstruction



(b) Gini reconstruction

Figure 3.6: ℓ_p and GI reconstruction for a segment of a random signal ($n = 1000, k = 100$) using 300 random samples.

3.3.2 Two-Dimensional Images

See figure 3.7 for the original images. The reconstructed images, as obtained from the proposed SPSA algorithm, are compared with those obtained from other currently used methods in the literature (on compressive sampling), employing norms as the sparsity measures. As before, the minimization of $\ell_{0.5}$-norm has been done using [40].

	Domain				
Image	Spatial	DCT	DFT	Wavelets	Gradient
Phantom	0.6977	0.7600	0.5664	0.8887	**0.9846**
Boat	0.1922	0.6948	0.6508	**0.8832**	0.6282
Hill	0.2448	0.6570	0.6150	**0.8781**	0.6036
Man	0.2450	0.6626	0.6099	**0.8838**	0.6099
MRI	0.5560	0.7132	0.6736	**0.9389**	0.8350
Peppers	0.2467	0.6996	0.6552	**0.8970**	0.6771
CT image	0.4534	0.8449	0.6098	**0.9155**	0.8165

Table 3.5: Comparison of the GI of different transform domains for the test images.

	Domain				
Image	Spatial	DCT	DFT	Wavelets	Gradient
Phantom	3.2e09	1.6e09	1.7e12	1.1e08	**9.6e06**
Boat	8.5e12	4.5e11	2.8e14	5.5e11	**2.7e11**
Hill	7.9e12	4.8e11	3.1e14	5.6e11	**2.9e11**
Man	7.3e12	4.7e11	2.9e14	4.9e11	**2.6e11**
MRI	1.2e12	2.4e11	1.5e14	7.9e10	**5.1e10**
Peppers	4.9e11	3.510	1.1e13	3.9e10	**1.8e10**
CT image	3.1e11	1.6e10	4.19e12	1.7e10	**1.2e10**

Table 3.6: Comparison of the $\ell_{0.5}$ of different transform domains for the test images.

Table 3.5 shows the values of Gini index, while table 3.6 shows the $\ell_{0.5}$ norm of each test image, in different representation domains. It is interesting to observe that the Gini index, while apparently giving a more transparent understanding of the measure of sparsity, shows the wavelet domain to be the sparsest for all the test images, excepting the phantom image for which its gradient domain is the sparsest.

In contrast, when the sparsity measure is computed using the $\ell_{0.5}$ norm, the gradient domain is the sparsest for all the test images. As far as the examples of images below are concerned, it is to be noted that measurement samples are randomly taken from radial lines in the Fourier domain.

Table 3.5 shows that the MRI-phantom image of size 256×256 is sparse in its gradient domain (see figure 3.8(b)). This imaged is reconstructed using only the randomly chosen samples along 9 radial lines in its Fourier domain, by maximizing the Gini index of its gradient magnitude. In figures 3.8 and 3.9, the result of the proposed method has been compared with the results of ℓ_1 and TV norm minimization and [16]. It can be seen the GI-based method outperforms the others in reconstructing the image from only 9 radial lines.

The proposed algorithm is also applied to a set of medical and natural images. As mentioned above, according to the Gini index, the wavelet domain is the sparsest domain for these images. Therefore, the Gini index of the wavelet coefficients is maximized in Wavelets domain subject to the observations in the Fourier domain. Results are compared with those obtained from minimizing ℓ_1 and $\ell_{0.5}$ norms of the wavelets coefficients. Table 3.7 presents the PSNR (2.13) of the reconstructed images using different approaches. It is observed that, for

all the test images, the Gini index-based reconstruction is superior, in terms of a significantly improved PSNR, to reconstruction from norms minimization.

Input Data		Reconstructed image PSNR			
Image name	% of Samples	TV min.	ℓ_1 min.	ℓ_0 min.	GI max.
Boat	50%	31.77	32.23	32.84	34.71
Hill	39%	31.51	25.52	30.57	33.12
Man	38%	30.35	30.55	30.52	34.64
MRI	27%	32.66	34.11	36.84	42.98
Peppers	39%	32.90	29.66	30.96	31.82
CT image	27%	36.07	33.61	32.61	35.81

Table 3.7: PSNR of the reconstructed images.

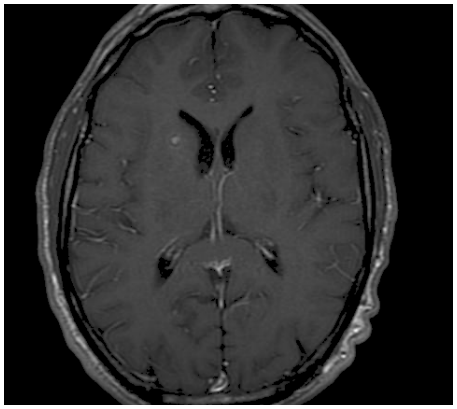
In contrast, the $\ell_{0.5}$ norm suggests that the gradient domain is the sparsest for medical and natural images. This is the reason for including the TV minimization reconstruction results for comparison. According to table 3.7, the GI-based method is inferior to TV minimization in terms of improved PSNR for the CT and Pepper test images. However, for some of the images, such as the Hill, the performance of the TV minimization method is comparable to ours. In fact, figures 3.10, 3.11 and 3.12 show that the perceptual contrast is better in the GI-based method than that of TV minimization. Moreover, TV minimization results in blurred images, affecting the visual quality, which, in medical images, may lead to loss of important minutiae, such as small vessels.

(a) Boat, 512×512

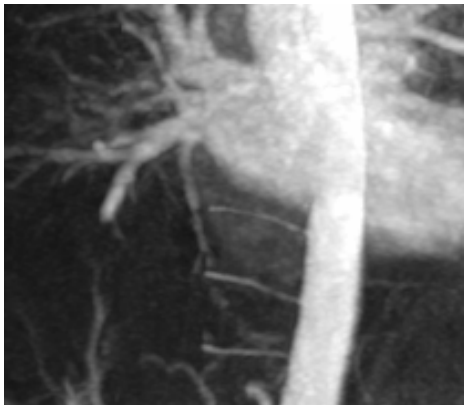
(b) Man, 512×512

(c) Peppers, 256×256

(d) Hill, 512×512

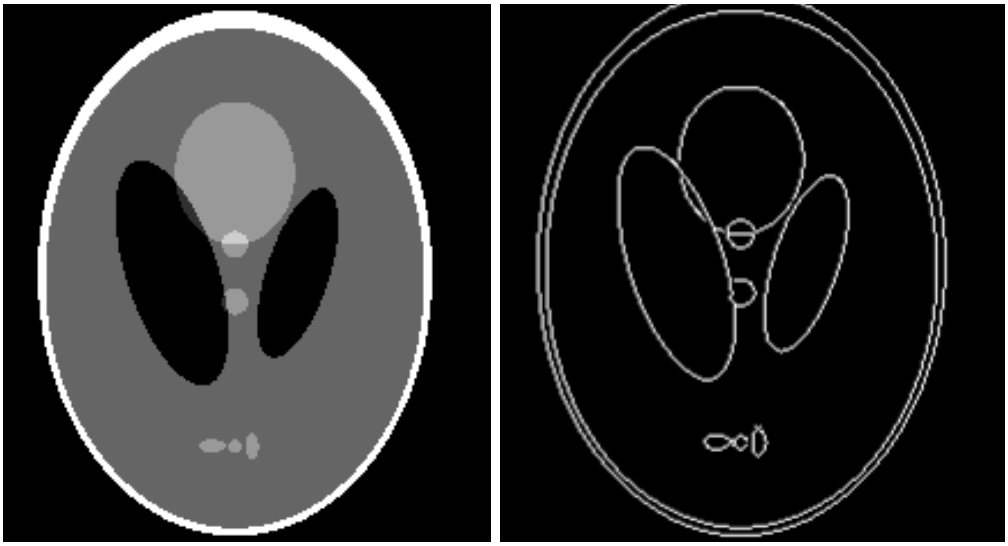


(e) MRI , 512×512

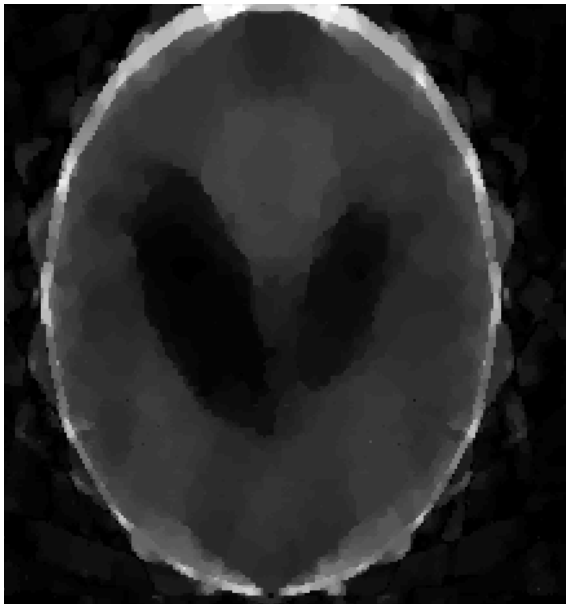


(f) CT , 256×256

Figure 3.7: Original test images [1].

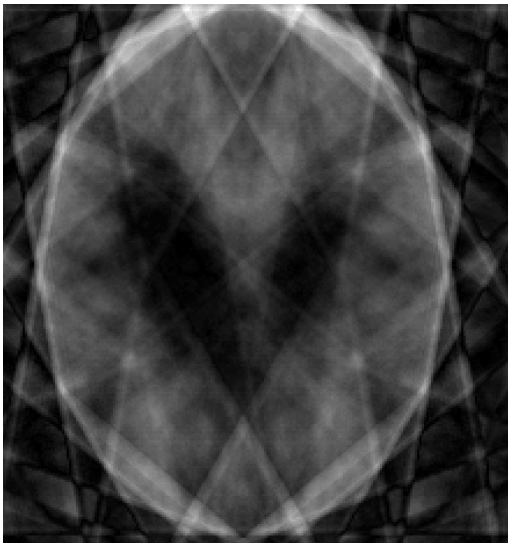


(a) Original Phantom (b) Gradient of Phantom image



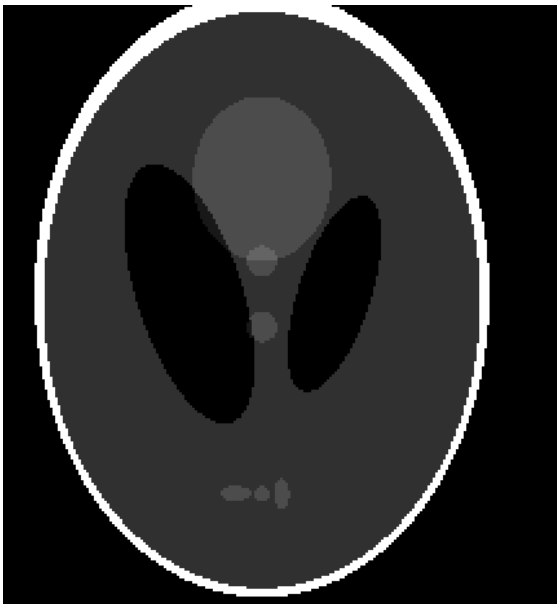
(c) Total Variation reconstruction

Figure 3.8: Phantom image of size $n = 256 \times 256$, sampled over 9 radial lines.



(a) ℓ_1 reconstruction

(b) ℓ_p reconstruction



(c) GI Reconstruction

Figure 3.9: Phantom image of size $n = 256 \times 256$, sampled over 9 radial lines.

(a) TV reconstruction

(b) GI reconstruction

Figure 3.10: TV vs. GI reconstruction of the Hill image (zoomed in).

(a) TV reconstruction

(b) GI reconstruction

Figure 3.11: TV vs. GI reconstruction of the Peppers image (zoomed in).

(a) TV reconstruction

(b) GI reconstruction

Figure 3.12: TV vs. GI reconstruction of the CT image.

3.3.3 Robustness

To test the robustness of the Gini index as a measure of sparsity, 100 random signals are generated of size $n = 1000$ with $k = 100$, which means that 90% of each signal component is randomly chosen to be zero. The signals have been then corrupted with different levels of white Gaussian noise. $m = 280$ number of samples (where m is arbitrary) is used to reconstruct the noisy signals. Entries of the sampling matrices $\mathbf{\Phi}$ of size $m \times n$ has been randomly selected from a mean-zero Gaussian distribution. As explained above, the Gini index is maximized using the SPSA algorithm for reconstruction. The reconstructed signals are compared with those obtained from the currently used methods in the literature. Figure 3.13 shows the MSE (2.12) of the reconstructed signals vs. SNR of the noise. It is observed that GI-based method is superior to the ℓ_1 and ℓ_p norm-based approaches, in general, and more significantly so when a higher amount of noise is present. Moreover, it is observed in figure 3.13 that in presence of higher amount of noise, ℓ_1 norm-based reconstruction sometimes outperforms that of the $\ell_{0.5}$ norm. This could be due to the sensitivity of $\ell_{0.5}$ to the small values of the signal which could make this norm unstable in the presence of noise.

Finally, the performance of the GI-based method is tested against those of TV ℓ_1 and $\ell_{0.5}$ norms, as applied to the earlier test images which are now corrupted with white Gaussian noise. Sampling is in the Fourier domain, and reconstruction, in the wavelet domain. Table 3.8 shows the PSNR of the noisy image and of the reconstructed images using the different methods. It can be seen that the Gini index-based reconstruction is superior to those of norms minimization in terms of a significantly improved PSNR for all the test images.

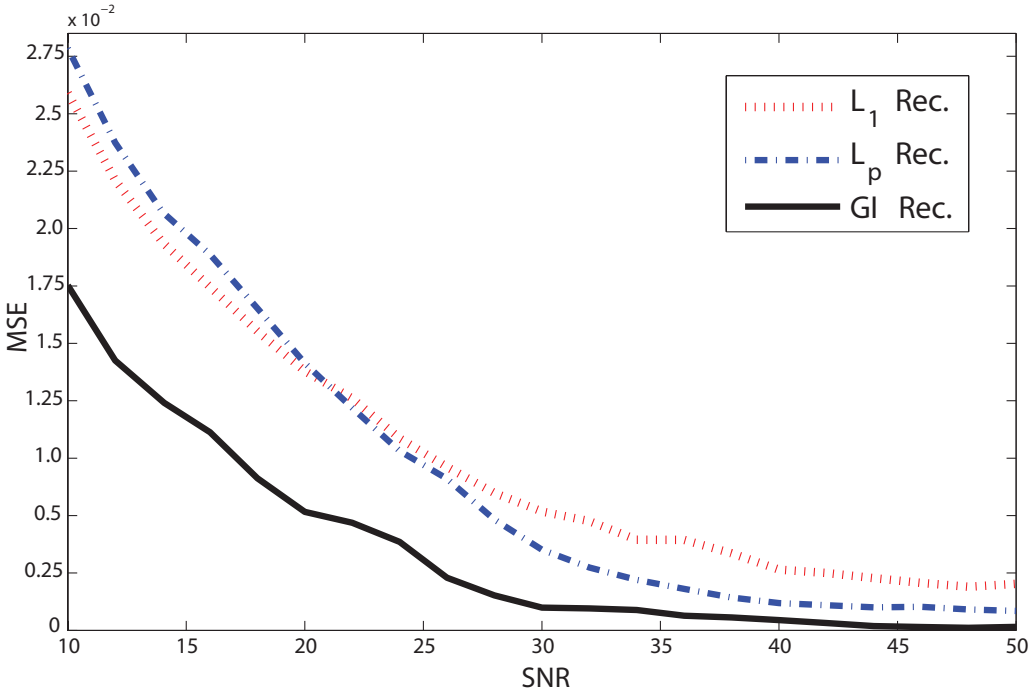


Figure 3.13: MSE of the reconstructed signals vs. SNR of the noise.

Input Data		Reconstructed image PSNR			
Image name	% of Samples	noisy image	TV min.	ℓ_0 min.	GI max.
Boat	50%	25.36	26.92	27.23	29.81
Hill	39%	26.12	27.82	26.78	27.65
Man	38%	26.44	25.72	27.86	29.01
MRI	27%	27.11	30.42	27.33	32.28
Peppers	39%	25.55	25.81	24.28	25.83
CT image	27%	26.73	28.91	28.61	29.10

Table 3.8: PSNR of the reconstructed noisy images.

3.4 Summary

In this chapter, the use of the Gini index (GI) as a sparsity measure in reconstruction of signals/images from compressive samples, is explored. Through several examples, it is illustrated that the GI, which is a quasi-convex function of its arguments, is a more reliable and robust alternative to the currently popular ℓ_p (pseudo-) norm-based (for $0 < p \leq 1$) sparsity measures. Furthermore, the proposed GI-based stochastic optimization method to reconstruction signals and images is shown to be superior over other commonly used norm-based minimization methods, in both noise-free and noisy settings.

Chapter 4

Reconstruction of Sequences of Sparse Signals

4.1 Introduction

In this chapter the problem of reconstruction of time varying sparse signals from a series of linear measurements, particularly those closely related such as video and MRI signals, is investigated.

In its noiseless formulation, the problem can be posed as follows: let $\mathbf{f}^{(t)} := [f_1^{(t)}, f_2^{(t)}, \cdots, f_n^{(t)}] \in \mathcal{R}^n$ be the slowly time-varying signal of interest at epoch t and $\mathcal{S}^{(t)} := \{k \in \{1, \cdots, n\} : f_k^{(t)} \neq 0\}$ denote its support. At each time instant, the signal of interest is measured using a sampling matrix ($\mathbf{\Phi}^{(t)}$ of size $m \times n$) and $\mathbf{y}^{(t)} = \mathbf{\Phi}^{(t)}\mathbf{f}^{(t)}$ is the observation vector, which is assumed to be incoherent with respect to the sparsity basis of the signal. The measurement matrix could be fixed (i.e. $\mathbf{\Phi}^{(t)} = \mathbf{\Phi}$) or it may be changing with time.

The problem, at each time t, is then to recover the original signal, $\mathbf{f}^{(t)}$, from the corresponding compressive samples ($\mathbf{y^{(t)}}$), assuming that the signal of interest is sparse. The static version of the above problem has been thoroughly studied, however, few works have focused on sequences of sparse signals.

A naive approach for reconstruction of the dynamic signal, would be to use the simple CS at each time frame separately. Equivalently, at each time instant, t, we find a $\mathbf{g}$ which is an estimate of $\mathbf{f}^{(t)}$, as a solution to the following optimization problem:

$$\begin{aligned} &\mathbf{f}^{(t)} = \arg\min\{\left\|\hat{\mathbf{s}}\right\|_{\ell_1}\} \\ &\text{subject to } \mathbf{\Phi^{(t)}}\hat{\mathbf{s}} = \mathbf{y}^{(t)} \end{aligned} \tag{4.1}$$

From the CS literature, it is known that for perfect recovery of $\mathbf{f}^{(t)}$, we need $m \gtrsim |\mathcal{S}^{(t)}| \log n$ compressive samples, where $|\mathcal{S}^{(t)}|$ is the cardinality of the support at time t [2]. However this approach does not make any use of the fact that the signal of interest changes slowly over time.

In [41], on the other hand, the authors use this property to propose a way to dynamically update the solution of the above minimization (4.1), without directly solving it. The proposed dynamic update scheme systematically breaks down the solution update into a small number of linear steps. However, this approach merely uses past reconstructions to speed up the current optimization and does not improve the reconstruction error and therefore the number of samples needed is equal to the conventional CS [17].

Authors of [10], discussed the problem of reconstructing a signal when some

priori information exists about the signal (their method is referred to as *priori-CS*). Their method can be applied to the problem of sparse sequence reconstruction as follows:

$$\begin{aligned} \mathbf{f}^{(t)} &= \arg\min\{\left\|\hat{\mathbf{s}}\right\|_{\ell_1}\} \\ \text{subject to } & \mathbf{\Phi^{(t)}}\hat{\mathbf{s}} = \mathbf{y}^{(t)}, \\ & \|\hat{\mathbf{s}} - \mathbf{f}^{(t-1)}\|_{\ell_1} \leq \epsilon \end{aligned} \tag{4.2}$$

The assumption is that ℓ_1 of difference signal, ($\|\mathbf{f}^{(t)} - \mathbf{f}^{(t-1)}\|_{\ell_1}$) would be small, if the support is changing slowly over time. However, this assumption is not always valid as values and locations of the non-zero elements (spikes) of a sparse signal will typically change over time.

Recently Vaswani and Lu [17], proposed the *modified-CS* which uses the support of the previous time instant ($\mathcal{S}^{(t-1)}$) as an estimated support of the signal of interest ($\mathbf{f}^{(t)}$) at current time and then use this estimate for reconstruction of $\mathbf{f}^{(t)}$, by finding a signal which satisfies the observations and is sparsest outside $\mathcal{S}^{(t-1)}$. This is equivalent to solving the following optimization problem:

$$\begin{aligned} \mathbf{f}^{(t)} &= \arg\min\{\left\|\hat{\mathbf{s}}|_{\hat{\mathcal{S}}^{(t-1)}}\right\|_{\ell_1}\} \\ \text{subject to } & \mathbf{\Phi^{(t)}}\hat{\mathbf{s}} = \mathbf{y}^{(t)} \end{aligned} \tag{4.3}$$

where $\hat{\mathcal{S}}^{(t-1)} := \{k \in \{1, \cdots, n\} : f_k^{(t-1)} = 0\}$, is the complement of $\mathcal{S}^{(t-1)}$ and $\hat{\mathbf{s}}|_{\hat{\mathcal{S}}^{(t-1)}} := \{g_i : i \in \hat{\mathcal{S}}^{(t-1)}\}$.

It is shown in [17], that under fairly general conditions, the number of samples needed would be less than the conventional CS. The main assumption in this method is that non-zero elements, stay on the same location in the signal/image of the subsequent time instant. However, it is observed that in many real world scenarios (e.g. figures 4.12 and 4.13), non-zero elements do not stay in the extract location and move to its vicinity.

In next section a method is introduced which is able to achieve exact reconstruction from even fewer number of samples than the modified-CS by extracting more priori information from the previous reconstructed signal than just the estimated support. In section 4.2, we provide details of the proposed approach for extracting priori knowledge from reconstructed signals of previous time instants and explain how the extracted priori knowledge is incorporated into the reconstruction process. The experimental results are presented and analyzed in section 4.3, where the application of the developed method in wire-less enabled ECG sensors and real MRI imaging modalities, is explored.

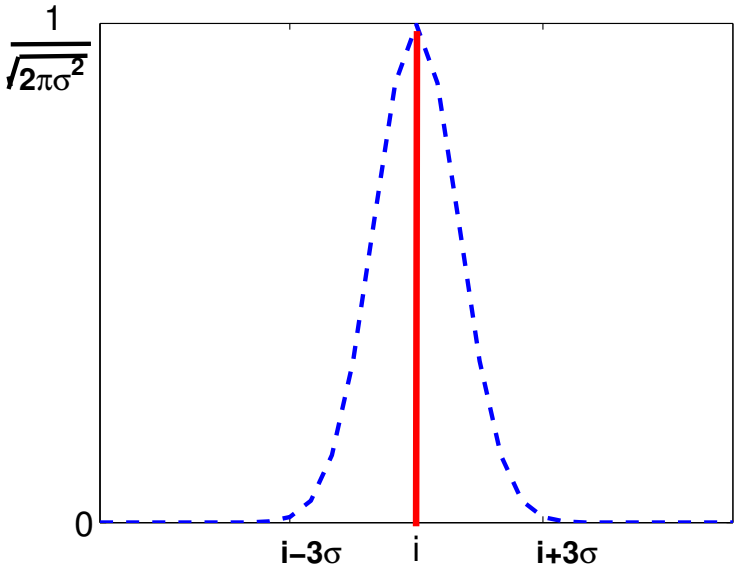
4.2 Proposed Approach

To guide the reconstruction process, the proposed approach makes use of critical a priori knowledge including the estimated support of the signal of interest. The idea is based on the observation that in many real world signals & images, the pattern of sparsity changes slowly and smoothly with time. For example from the sequence of MRI images in figures 4.13 and 4.12, it can be seen that sparsity, which is in the spatial domain, changes smoothly from each slice to the other. Therefore,

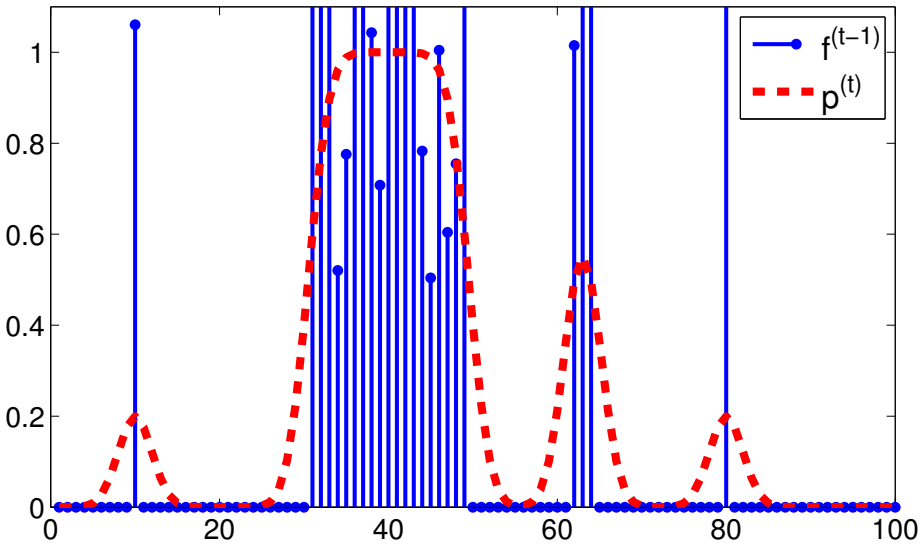
the conjecture is that we should be able to extract some priori information about sparsity of $\mathbf{f}^{(t)}$ from $\mathbf{f}^{(t-1)}$. More specifically, we try to estimate the probability of each element of $\mathbf{f}^{(t)}$ having a non-zero value, from $\mathbf{f}^{(t-1)}$. For the signal of interest at time t ($\mathbf{f}^{(t)}$), let p_i be the probability of its i^{th} element having a non-zero value, $p_i := P(f_i > 0)$ and let $\mathbf{p} := [p_1, p_2, \cdots, p_n] \in \mathcal{R}^n$ be the sparsity probability vector. In the proposed method, probability vector $\mathbf{p}$ for the current signal ($\mathbf{f}^{(t)}$) is estimated from the reconstructed signal of the previous time instant ($\mathbf{f}^{(t-1)}$) and then is used to aid the reconstruction of $\mathbf{f}^{(t)}$. This is discussed in detail in the following sub-sections.

4.2.1 Estimation of sparsity probability vector

In this sub-section, estimation of the sparsity probability model for a signal at time t, from the reconstructed signal at time $t-1$, is discussed. Based on the assumption that sparsity changes smoothly with time, given a spike in the signal at time $t-1$, there is a good chance that either it remains in the same location, or shifts to some point in the same vicinity in the next time frame (t). Similarly, at time t it is expected that zeros appear in the vicinity of zeros at time $t-1$. More specifically, suppose there is a spike in the i^{th} location of $\mathbf{f}^{(t-1)}$. In the next time instant, there is a very high probability this spike remains in the same location but also some possibility that it moves to some other point in the vicinity. Thus, the probability of the spike appearing at the each location decreases as we get farther from (i). This motivated us to use a Gaussian distribution (figure 4.1(a)) to provide an estimate of the probability of the progression of a spike in the next time frame.



(a) Probability of progression of a spike in the next time instant.



(b) Estimation of $\mathbf{p}^{(t)}$ from $\mathbf{f}^{(\mathbf{t-1})}$, with $\sigma = 2$.

Figure 4.1: Illustration of estimation of sparsity probability.

Thus the probability of j^{th} element of $\mathbf{f}^{(t)}$ being non-zero ($p_j^{(t)}$) is the accumulated probability of the spikes of $\mathbf{f}^{(t-1)}$ moving to location j at time t as follows:

$$p_j^{(t)} = \sum_{i \in \mathcal{S}^{(t-1)}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(j-i)^2/2\sigma^2} \tag{4.4}$$

where $\mathcal{S}^{(t-1)}$ is the support of $\mathbf{f}^{(t-1)}$ and σ^2 is the variance of the Gaussian distribution which is set proportional to the signal's rate of change with time. Figure 4.1(b) shows a syntactic signal $\mathbf{f}^{(t-1)}$ together with the probability of the elements of $\mathbf{f}^{(t)}$ being non-zero in the next time frame ($\mathbf{p}^{(t)}$) which is the dashed red line. From this figure and equation (4.4), it can be seen that the maximum value of $p_j^{(t)}$ is 1. This coincides with the locations where in the previous time instant, $f_j^{(t-1)}$ and all elements in its vicinity ($[j-3\sigma, j+3\sigma]$) are non-zero. Similarly, $p_j^{(t)}$ is 0, if there is not any spike in this span.

The above formulation can be easily extended for 2D images using a 2D Gaussian distribution. As an example, figure 4.2 shows the estimated sparsity probability for an MRI image of $t = 3$ (figure 4.2(c)) from the previous frame $t = 2$ (figure 4.2(a)). It should be mentioned that if no priori knowledge on the sparsity of the signal of interest is available, including at $t = 1$ where there is no previous time instant to estimate the probability from, p_i is set to 0.5.

(a) $f^{(2)}$ (b) $f^{(3)}$

(c) Estimated Sparsity of $f^{(3)}$ from $f^{(2)}$

Figure 4.2: Illustration of estimated sparsity model for an MRI image.

4.2.2 Reconstruction using the sparsity probability model

In this sub-section, the problem of recovering a signal $\mathbf{f}$ using its sparsity probability model, is discussed. More specifically, let $\mathbf{p} := [p_1, p_2, \cdots, p_n] \in \mathcal{R}^n$ be the priori knowledge of the signal of interest's sparsity, where $p_i := P(f_i > 0)$ is the probability of f_i having a non-zero value. It is clear that if $p_i = 0$, then f_i is always 0, while if $p_i = 1$, then it is known beforehand that $\mathbf{f}$ has a spike at location i (though its value is unknown). $p_i = 0.5$ basically means that no priori information of f_i sparsity is available and it is as likely as not to have a non-zero value.

In order to incorporate the probability model of the signal into the process of reconstruction, a weighted ℓ_1 norm minimization(4.5) is proposed, where the weights are adjusted according to the probability of each entry being non-zero:

$$\min ||\mathbf{W}\hat{\mathbf{s}}||_{\ell_1} \text{ subject to } \mathbf{\Phi^{(t)}}\hat{\mathbf{s}} = \mathbf{y} \tag{4.5}$$

where $\mathbf{W} = diag([w_1, w_2, \cdots, w_n])$. Intuitively, a smaller weight should be given to those entries with higher probability of being non-zero while those elements with small probability should be penalized with larger weights. Intuitively, to reward and penalize the elements uniformly, a linear function is used. Thus, the choice of the weight for each element is:

$$w_i = 2(1 - p_i) \tag{4.6}$$

Figure 4.3 shows the chosen weight with respect to the value of the probability.

It can be seen that as the probability of an element being non-zero, increases, its weight decreases accordingly. notice that if $p_i = 0.5$ then $w_i = 1$ and $w_i < 1$ when $p_i > 0.5$ (similarly if $p_i < 0.5$ then $w_i > 1$).

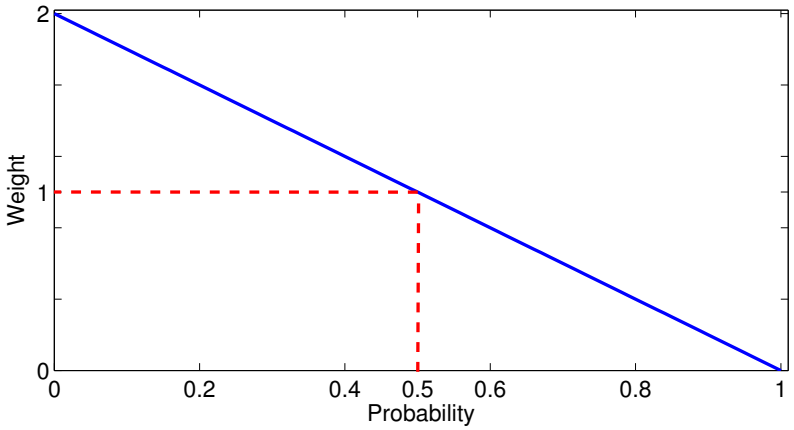


Figure 4.3: Chosen weight (w_i) vs. probability (p_i).

It should be noted that the weighted ℓ_1 approach (4.5), could be seen as a generalized ℓ_1 minimization, since when no priori information of the sparsity is available (which means for all elements $p_i = 0.5$), it reduces to the conventional ℓ_1 minimization as all weights would be equal.

Remark 2: It is possible to draw a parallel between the proposed method and the modified-CS [17], where it is assumed that a part of the support is known a priori. If we set the probability of the known part of the support $p_i = 1$, then it follows that $w_i = 0$ and for the other elements of the signal, since there is no information available on their sparsity, we set $p_i = 0.5$ and therefore $w_i = 1$. Thus, solving the weighted ℓ_1 in (4.5) with the above described weights is in fact equivalent to solving (4.3).

Condition for perfect reconstruction using weighted ℓ_1

In this sub-section a necessary and sufficient condition of the sampling matrix null space for the unique recovery of a sparse signal, is described. The proof is inspired by [10].

Lemma: Let the signal of interest be $\mathbf{x}^*$ with the support $\mathcal{S}$, solving the weighted ℓ_1 minimization (4.5) we can recover $\mathbf{x}^*$ uniquely if and only if

$$\|(\mathbf{Wr})|_{\hat{\mathcal{S}}}\|_{\mathbf{1}} > \|(\mathbf{Wr})|_{\mathcal{S}}\|_{\mathbf{1}} \tag{4.7}$$

where $\mathbf{r} \in null(A) \setminus \{0\}$ and $\hat{\mathcal{S}}$ is the complement of the support of x^*.

Proof: $\mathbf{x}^*$ is recovered uniquely from (4.5) if for all $\mathbf{x}$ that $\mathbf{Ax} = \mathbf{y}$ we have $\|\mathbf{Wx}^*\|_{\mathbf{1}} < \|\mathbf{Wx}\|_{\mathbf{1}}$. Since $\mathbf{x} = \mathbf{x}^* + \mathbf{r}$ then

$$\begin{aligned}
\|\mathbf{Wx}\|_{\mathbf{1}} = \|\mathbf{W}(\mathbf{x}^* + \mathbf{r})\|_{\mathbf{1}} &= \|(\mathbf{Wx}^*)|_{\mathcal{S}} + (\mathbf{Wr})|_{\mathcal{S}}\|_{\mathbf{1}} + \|(\mathbf{Wx}^*)|_{\hat{\mathcal{S}}} + (\mathbf{Wr})|_{\hat{\mathcal{S}}}\|_{\mathbf{1}} \\
&= \|(\mathbf{Wx}^*)|_{\mathcal{S}} + (\mathbf{Wr})|_{\mathcal{S}}\|_{\mathbf{1}} + \|(\mathbf{Wr})|_{\hat{\mathcal{S}}}\|_{\mathbf{1}} \\
&> \|(\mathbf{Wx}^*)\|_{\mathbf{1}} - \|(\mathbf{Wr})|_{\mathcal{S}}\|_{\mathbf{1}} + \|(\mathbf{Wr})|_{\hat{\mathcal{S}}}\|_{\mathbf{1}}
\end{aligned}$$

Therefore $\|(\mathbf{Wx})\|_{\mathbf{1}} - \|(\mathbf{Wx}^*)\|_{\mathbf{1}} > \|(\mathbf{Wr})|_{\hat{\mathcal{S}}}\|_{\mathbf{1}} - \|(\mathbf{Wr})|_{\mathcal{S}}\|_{\mathbf{1}}$. If $\|(\mathbf{Wr})|_{\hat{\mathcal{S}}}\|_{\mathbf{1}} - \|(\mathbf{Wr})|_{\mathcal{S}}\|_{\mathbf{1}} > \mathbf{0}$ then $\|\mathbf{Wx}\|_{\mathbf{1}} - \|\mathbf{Wx}^*\|_{\mathbf{1}} > \mathbf{0}$ which means $\|\mathbf{Wx}\|_{\mathbf{1}} > \|\mathbf{Wx}^*\|_{\mathbf{1}}$ and x^* is the unique minimizer of (4.5).

Now we show necessity by contradiction. Assume for $\mathbf{r} \in null(A) \setminus \{0\}$, we have $\|(\mathbf{Wr})|_{\hat{\mathcal{S}}}\|_{\mathbf{1}} \leq \|(\mathbf{Wr})|_{\mathcal{S}}\|_{\mathbf{1}}$. Now let $\mathbf{x}^*|_{\mathcal{S}} = -\mathbf{r}|_{\mathcal{S}}$ and $\mathbf{x}^*|_{\hat{\mathcal{S}}} = 0$. Then, $\|\mathbf{W}(\mathbf{x}^* + \mathbf{r})\|_{\mathbf{1}} = \|(\mathbf{Wr})|_{\hat{\mathcal{S}}}\|_{\mathbf{1}} \leq \|(\mathbf{Wr})|_{\mathcal{S}}\|_{\mathbf{1}} = \|\mathbf{Wx}^*\|$. Therefore x^* can not be recovered uniquely from (4.5).

The null space condition in equation (4.7) is equal to

$$\sum_{i\in\hat{\mathcal{S}}} w_i|r_i| > \sum_{j\in\mathcal{S}} w_j|r_j|$$

and with the weights set according to (4.6), the above condition becomes:

$$\sum_{i\in\hat{\mathcal{S}}} \big(1-p_i\big)|r_i| > \sum_{j\in\mathcal{S}} \big(1-p_j\big)|r_j| \tag{4.8}$$

If we compare this to the null-space property of the conventional ℓ_1 minimization which is

$$\sum_{i\in\hat{\mathcal{S}}} |r_i| > \sum_{j\in\mathcal{S}} |r_j|$$

and recalling that $\big(1-p_i\big) > 1$ if $p_i < 0.5$ and $\big(1-p_i\big) < 1$ if $p_i > 0.5$, it can be seen that conditions for weighted ℓ_1 recovery is more relaxed than the ℓ_1 provided that spikes appear in the locations with the highest probabilities.

4.2.3 Reconstruction of the sequences of sparse signals

In the earlier sub-sections, we introduced the proposed method for estimating the sparsity probability model and explained its use for reconstruction. We call the method for reconstruction of sequences of sparse signals, the *Weighted CS*, and it is summarized in Algorithm 7 below.

Remark 3: It should be noted that at time $t = 1$, if no priori knowledge of the signal is available, we solve using ℓ_1 minimization thus requiring more samples to achieve a perfect reconstruction.

Input: $\mathbf{\Phi^{(t)}}$ and $\mathbf{y}^{(t)}$.
Output: $\mathbf{g}^{(t)}$.
while $t \leq t_{max}$ **do**

1) If $t = 1$ then $\mathbf{g}^{(1)} = \arg\min\{\left\|\hat{\mathbf{s}}\right\|_{\ell_1}$ subject to $\mathbf{\Phi}^{(1)}\hat{\mathbf{s}} = \mathbf{y}^{(1)}\}$;

2) $t = t + 1$;

3) $\mathcal{S}^{(t-1)} := \{k \in \{1, \cdots, n\} : s_k^{(t-1)} \neq 0\}$

4) Compute $\mathbf{p}^{(t)} := [p_1^{(t)}, \cdots, p_n^{(t)}]$ from (4.4).

5) Compute $\mathbf{W}^{(t)}$ from (4.6).

6) $\mathbf{g}^{(t)} = \arg\min\{\left\|\mathbf{W}^{(t)}\hat{\mathbf{s}}\right\|_{\ell_1}$ subject to $\mathbf{\Phi^{(t)}}\hat{\mathbf{s}} = \mathbf{y}^{(t)}\}$;

7) Go to step 2

end

Algorithm 7: Reconstruction of a sequences of sparse signals varying with time using Weighted-CS.

4.3 Experimental Results

We carried out a number of experiments on synthetic 1D data, real ECG signals and MRI images, in order to compare the reconstruction ability of the proposed Weighted CS method with that of ℓ_1 minimization [36], ℓ_p minimization , priori-CS [10] and the modified-CS [17].

4.3.1 One-Dimensional Synthetic Signals

Using a probability model of size n = 100, we randomly selected 70 elements and set them to 0.01 while the remaining 30 elements were set to a random number between zero and one (figure 4.4). Next, 100 random signals were generated

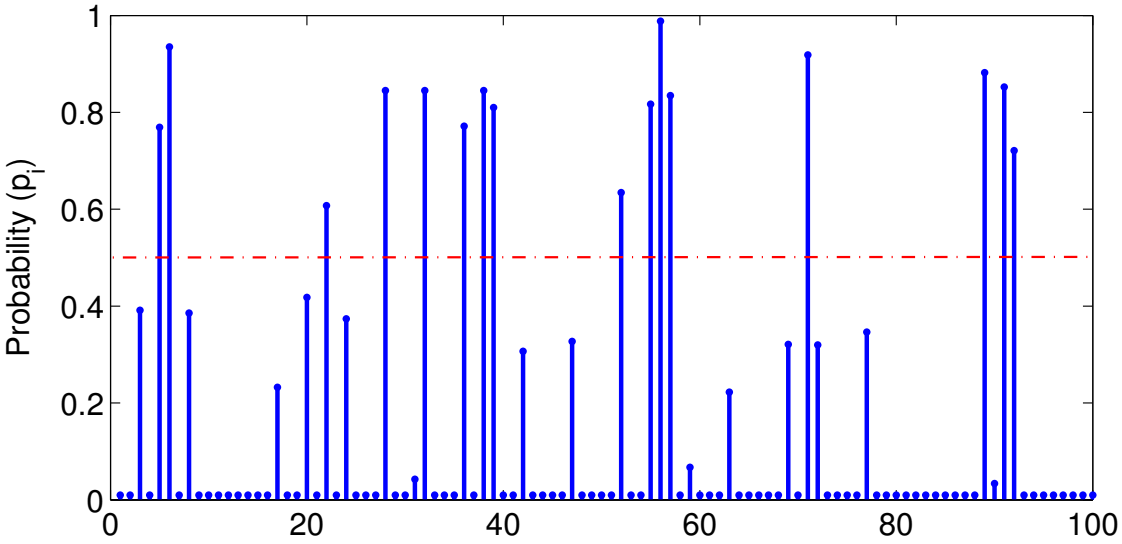


Figure 4.4: Probability of having a non-zero value.

according to this probability model, with the probability of each element having a non-zero value. The value of the non-zero spikes are also chosen randomly. We then use different number of samples (m) and for each m, we randomly select the elements of the sampling matrix A (of size $m \times n$) from a zero-mean Gaussian distribution. The probability of exact reconstruction for each method is estimated by counting the number of times that the error is less than 10^{-5}.

It should be noted that in the modified-CS method, elements with $p_i > 0.5$ are considered as the known support. Frequency of success for each sampling rate is shown figure 4.5. It can be seen that Weighted-CS outperforms the other methods in achieving perfect reconstruction with a smaller number of samples.

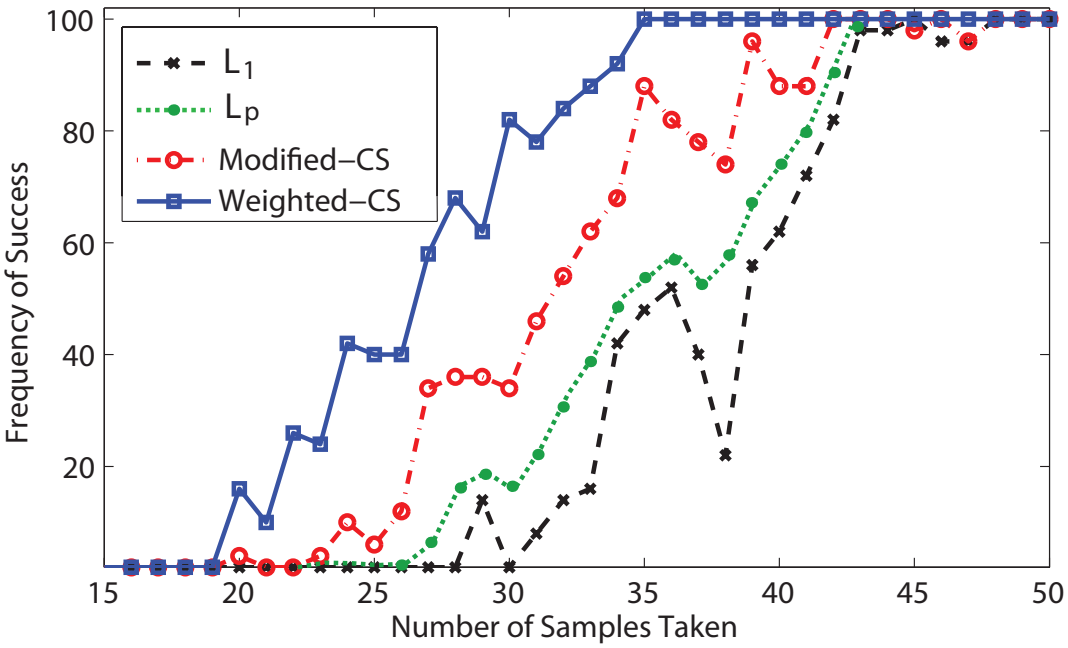


Figure 4.5: Percentage of perfect recovery vs. number of samples taken.

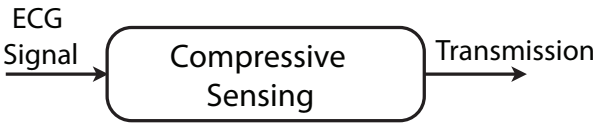
4.3.2 ECG remote sensing

The electrocardiograph (ECG) is considered as one of the most important diagnostic tools for assessing the electrical and muscular functions of the heart and thus have an important role in the battle against cardiovascular diseases which are among the top causes of death in the world.

While resting ECG monitoring is a standard practice in hospitals, major efforts have been underway to realize wireless-enabled low-power ECG monitoring. However, these ambulatory monitoring devices face many technical challenges including limited wireless connectivity, short battery life and etc [42].

State-of-the art ECG monitors fall short because either they transmit the uncompressed ECG data over wireless network, which puts much pressure on

(a) Conventional devices



(b) Compressive Sensing based.

Figure 4.6: Block diagram of ECG monitoring devices.

wireless links; or they compress the data in a compression unit after collecting and storing the full data on the chip, which leads to bulkiness and power inefficiency. In these approaches the signal of interest is first fully sampled according to the Nyquist rate. It is then compressed by using the thresholding-based Wavelet transform (DWT) algorithm [43] before being encrypted for transmission (figure 4.6(a)).

An alternative approach however, would be to use CS, based on the fact that the ECG signal is largely compressible and thus sparse. Compressive sensing employs linear sampling operators that map the signal into a relatively small dimensional space which result in a small number of measurements that can be wirelessly transmitted to the remote tele-cardiology center. The full signal can be recovered then from a much smaller set of measurements than the number of Nyquist-rate samples using complex, yet computationally feasible and numerically stable, non-linear methods. In addition to unifying the sampling and compres-

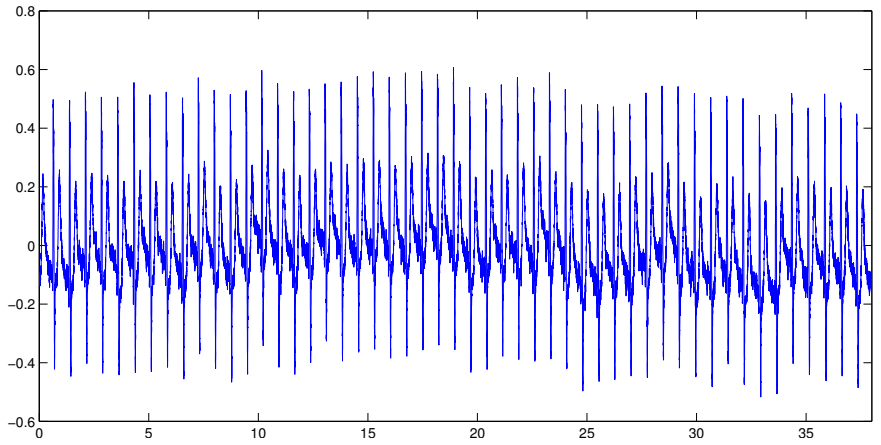
sion phases, under fairly general conditions, the measurements obtained using compressive sensing are considered to be encrypted [44] thus providing a computational guarantee of secrecy. Hence, compressive sensing basically swaps the burden of computation from the sampling device to the receiver end where more resources are available.

In recent works [45], [42], it has been found that compressed sensing based methods used in ECG sampling devices are inferior in terms of compression performance compared to wavelet transform based compression methods. Unsurprisingly, these compressive sensing based methods were found to exhibit the best overall energy efficiency. However, these recent works were carried out using ℓ_1 minimization with pursuit algorithm [2] as the default basis without attempting to exploit the highly structured nature and quasi-periodic property of ECG signals [42].

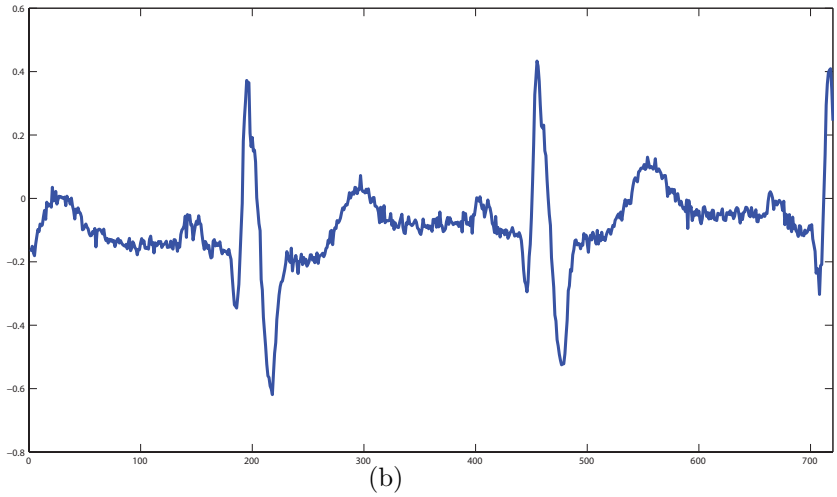
Digitized ECG signals are generally processed in non-overlapping windows of n samples (figure 4.7) due to the limited on-chip memory and real-time computing constraints. We use the Gini index, which has been shown to be an effective measure of sparsity [46], to find the sparse transform domain for ECG signals. Table 4.1 shows the Gini index values in different transform domains for an ECG signal. It can be seen that ECG is sparsest in Wavelet domain (since the Gini index is closest to 1). However it is known that Wavelets and the spatial domain which is sampling domain, are not incoherent. Therefore, we select DCT transform which is the second sparsest domain, as the sparse basis.

	Spatial	DCT	DFT	Wavelets
GI	0.2922	0.6948	0.7909	**0.8332**

Table 4.1: Comparison of the Gini Index (GI) of an ECG signal in different transform domains .



(a)



(b)

Figure 4.7: (a) An ECG signal (38 second); (b) A window (2 second) of an ECG signal.

Figure 4.8 shows the DCT transform of an ECG signal in three consecutive windows. It can be seen that frequencies which exist in these windows are very similar to each other and if there is a change from one window to the other, the change is smooth.

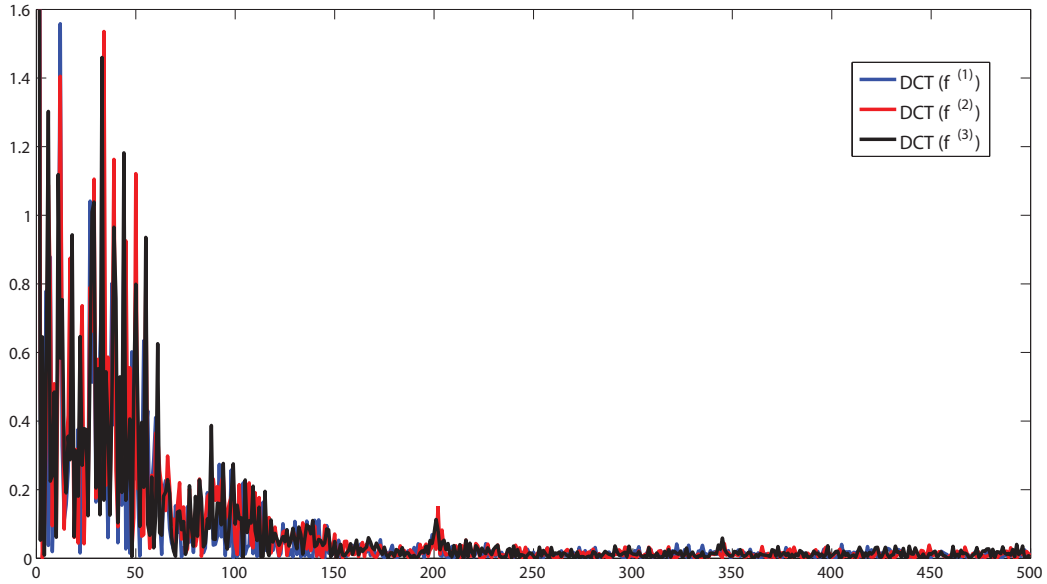


Figure 4.8: DCT transform of 3 ECG windows of size $N = 720$.

Twenty-four ECG recordings from the MIT-BIH Arrhythmia database [47] have been used to compare the performance of the Weighted-CS with the other state-of-the-art compressive sensing and non-compressive sensing based algorithms. The ECG signals were digitized at above-Nyquist sampling rate of 360 Hz for the duration of 38 seconds. The processing window of 2 seconds with length of n = 720 ECG samples, was used for all algorithms (figure 4.7).

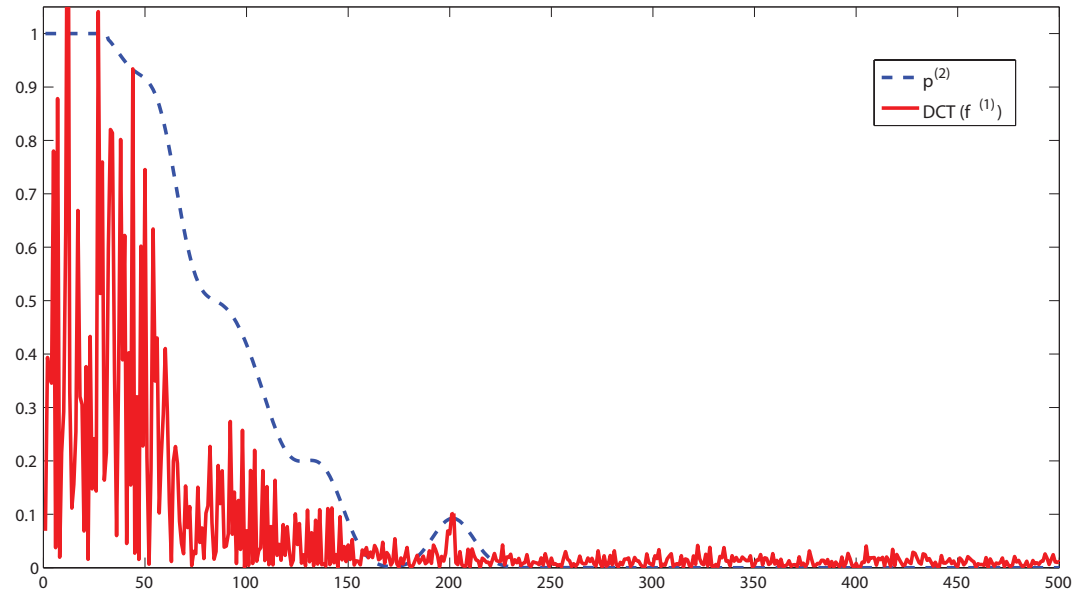


Figure 4.9: Estimation of $\mathbf{p}^{(t)}$ from $DCT(\mathbf{f^{(t-1)}})$, with $\sigma = 10$.

$$PRD = \frac{\left\|\hat{\mathbf{s}} - \mathbf{f}^{(t)}\right\|_{\ell_2}}{\left\|\mathbf{f}^{(t)}\right\|_{\ell_2}} * 100 \tag{4.9}$$

The ECG signals are then reconstructed from different number of samples using Algorithm 7 and the results are compared with two compressive sensing-based methods, namely modified-CS and ℓ_1 minimization. We have also compared the results against the state-of-the-art ECG Discrete Wavelet Transform (DWT) compression technique [43] which has been shown to outperform the embedded zerotree wavelet (EZW) and the set partitioning in hierarchical trees (SPIHT) based methods at a lower cost [48].

It should be noted that in modified-CS and Weighted-CS, reconstruction is

based on the priori knowledge which is extracted from the reconstructed signal of the previous time instant. Therefore, in the very first time window of each ECG signal we use 70% of the samples so that we can fully recover the signal without using any priori knowledge. The percentage of samples taken as indicated in figure 4.10 is for the second window onwards.

The performance measure used is the percentage root-mean-square difference (PRD) (Eqn.(4.9)) which is widely used to quantify the error the original and reconstructed signals. The average reconstruction error for 24 signals over 15 windows are presented in figure 4.10 for all methods. It can be seen that Weighted-CS performance is comparable with DWT and superior to the other methods in terms of PRD. It should be noted, however that DWT method tends to eliminate the high frequency elements of the signal, which could potentially lead to loss of some important diagnostic features. As an example, figure 4.11 shows a segment of the reconstructed signals from 20% of samples using DWT and Weighted-CS methods. It can be seen that DWT fails to fully recover the amplitude of the peaks which is considered as an important feature for cardiac condition detection [49].

It should be also noted that while the percentage of samples indicated in figure 4.10 for DWT refers only to the number of samples that need to be transmitted, the whole signal should be first stored on the memory of the sampling device for the Wavelet transform operation. On the other hand, the proposed method only requires the compressive measurements to be stored and transmitted.

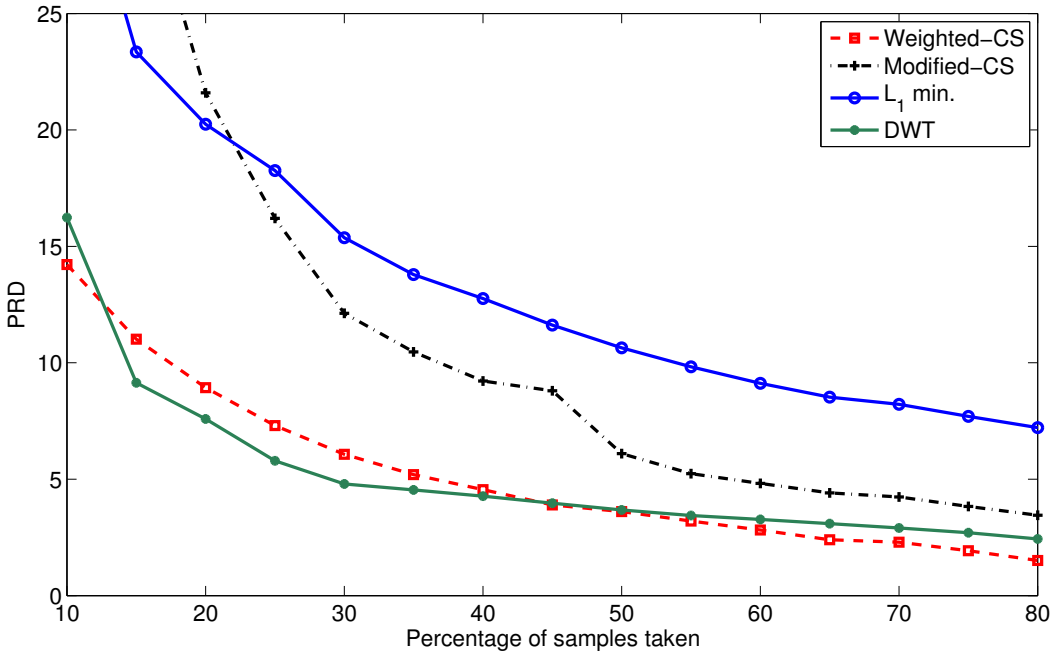


Figure 4.10: PRD vs. number of samples taken for ECG signals.

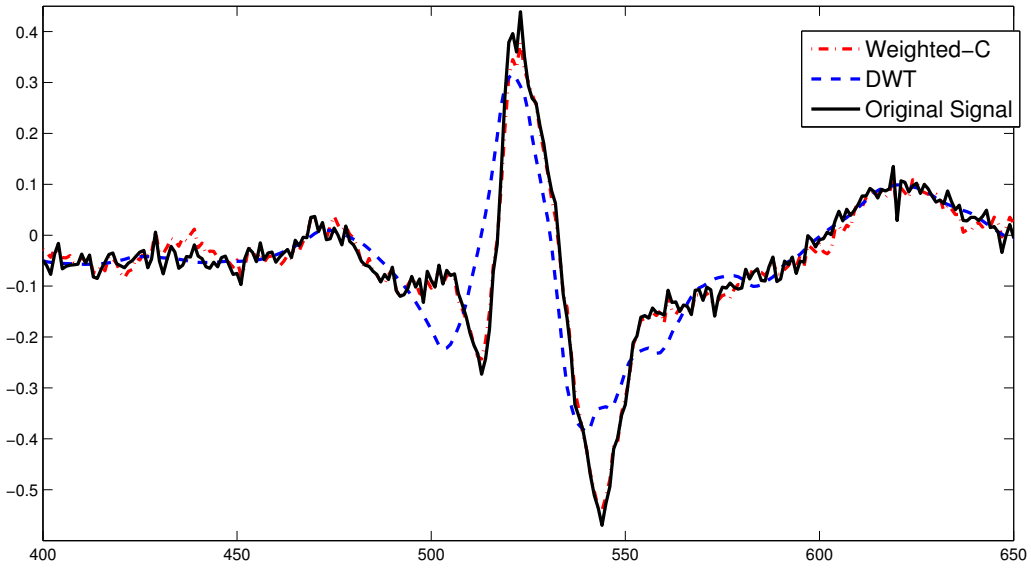


Figure 4.11: A segment of the reconstructed signals.

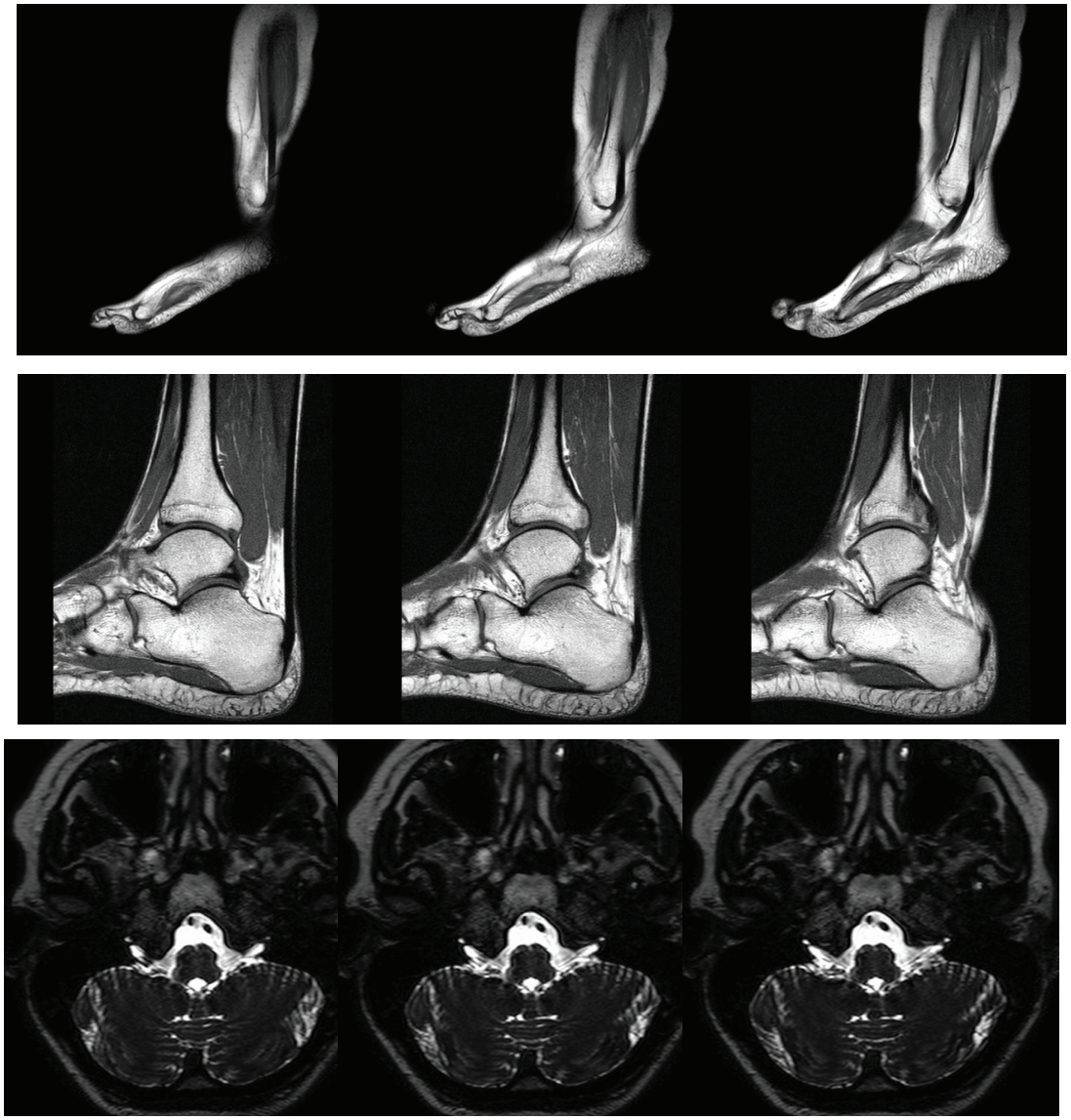


Figure 4.12: Sample images of MRI sequences.

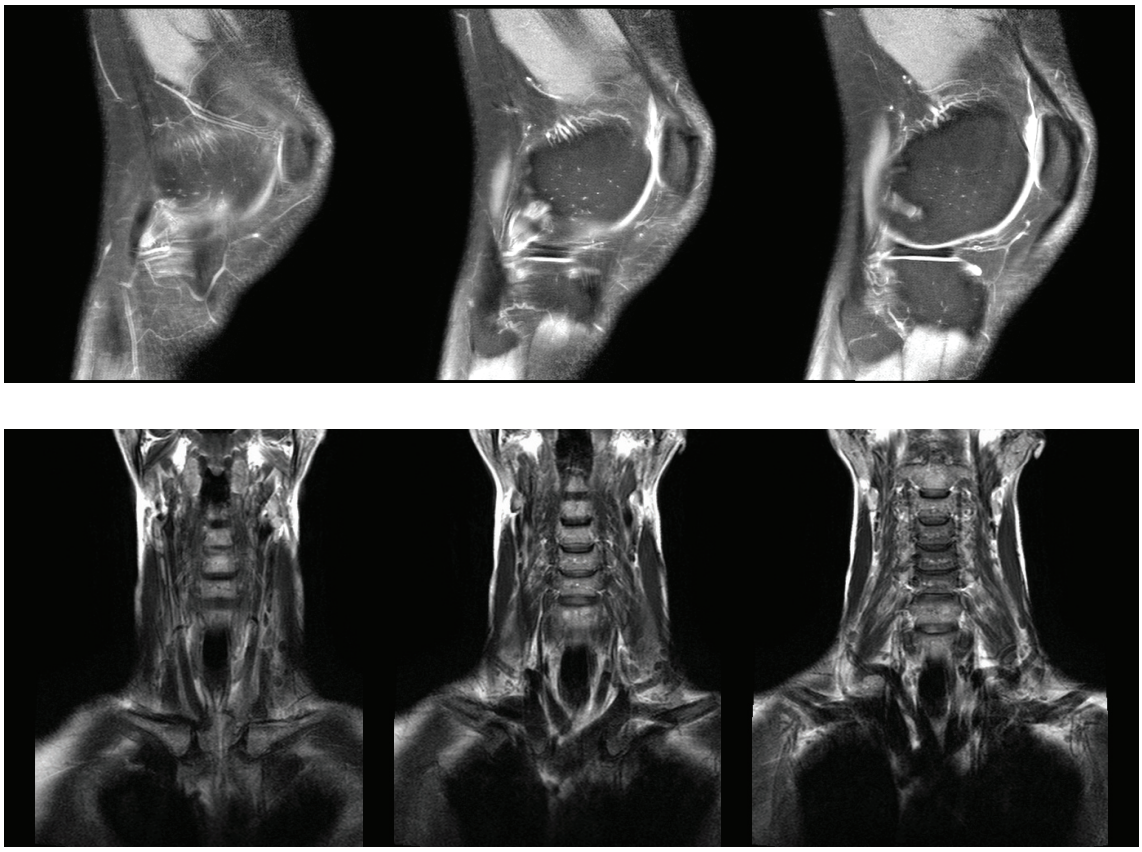


Figure 4.13: Sample images of MRI sequences.

4.3.3 Sequences of MRI Images

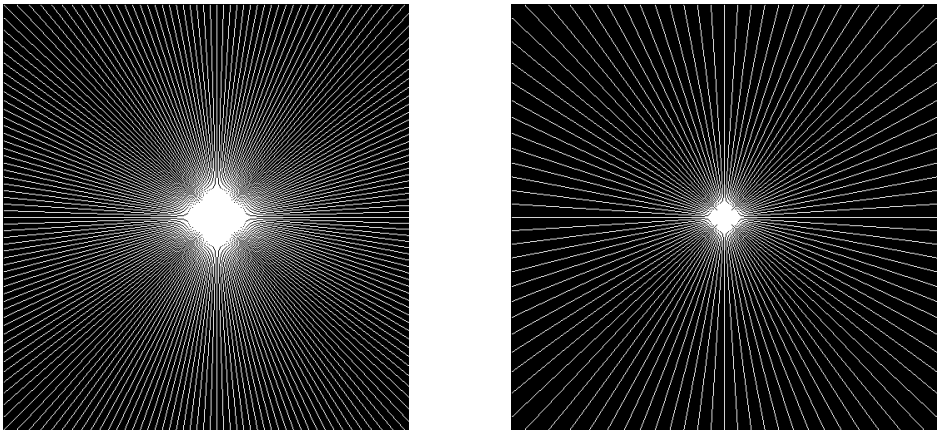
Magnetic Resonance Imaging (MRI) is an essential medical imaging tool burdened by its intrinsic slow data acquisition process. Since data acquisition is sequential in MR imaging modalities, the scan time (time to get enough data to accurately reconstruct one frame) is reduced if fewer measurements are needed for reconstruction. Therefore, goal of many researchers is to employ a smaller set of samples than normally required to reconstruct the original images. However, when k-space is under sampled, the Nyquist criterion is violated, and conventional Fourier reconstructions exhibit aliasing artifacts.

Much of the current CS-based works in the literature are, mostly concerned with reconstruction of static MRI images. However, in many important and challenging MRI applications, such as volumetric (3D) MRI or real-time MRI (refers to the continuous monitoring of moving objects in real time), instead of just one image slice, we are dealing with sequences of MRI images which are closely related to each other. Real-time MRI, for example, currently is only possible with low image quality or limited rate of slices per second due to the time-consuming scanning process [50]. To the best of our knowledge, only a few works have developed reconstruction methods for sequences of MRI images (e.g. [51], [50]). These observations motivate us to apply the Weighted-CS, to the problem of sequences of MRI images.

We tested the proposed algorithm on 5 sequences of MRI images of the foot, knee, ankle, neck (all of size $512 \times 512 \times 20$) and skull base ($512 \times 512 \times 40$), obtained from [52]. Figures 4.13 and 4.12, show some of these MRI images.

For the first image in each sequence, $t = 1$, since no priori knowledge is

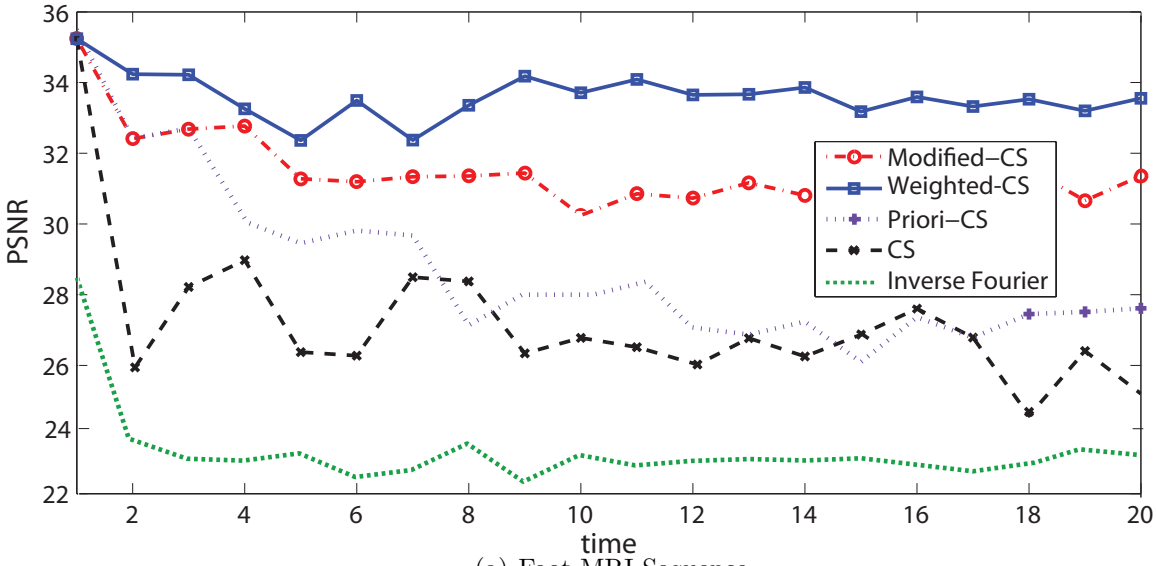
available, 30% of samples are taken along 150 radial lines in the Fourier domain (see sampling mask in figure 4.14(a)) while in the successive frames, only 10% of samples are taken along 50 radial lines (figure 4.14(b)) . The reconstruction is carried out in the Wavelet domain which is assumed to be the sparse domain.



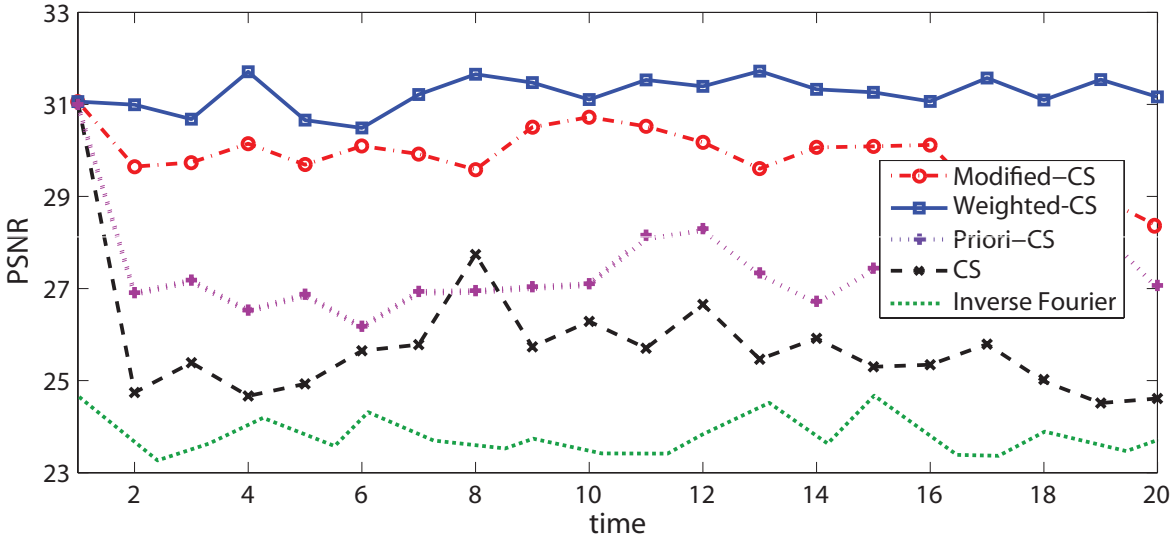
(a) (b)

Figure 4.14: Sampling mask for (a) t=1 (b) subsequent frames.

Figures 4.15, 4.16 and 4.17, compares the reconstruction performance of the Weighted-CS with σ set to 12 for MRI sequences with those obtained using the modified-CS, priori-CS and ℓ_1 minimization (CS). It can be seen that the proposed method out-performs the other methods significantly in terms of the improved PSNR.

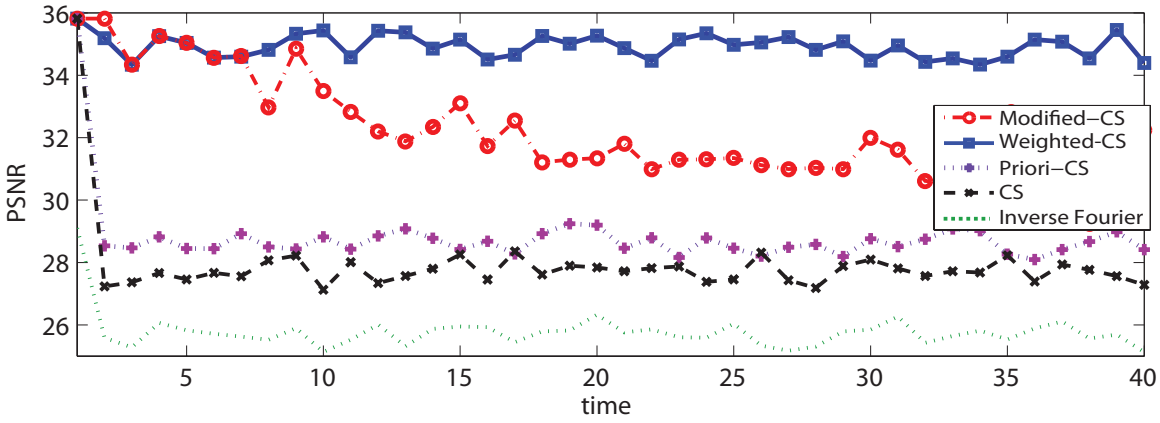


(a) Foot MRI Sequence

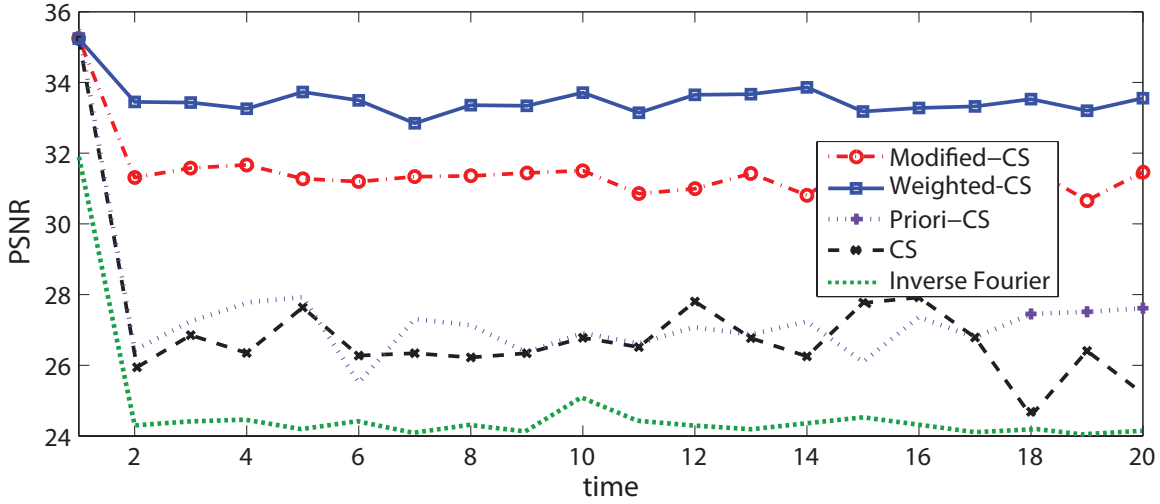


(b) Knee MRI Sequence

Figure 4.15: PSNR of the reconstructed images vs. time.

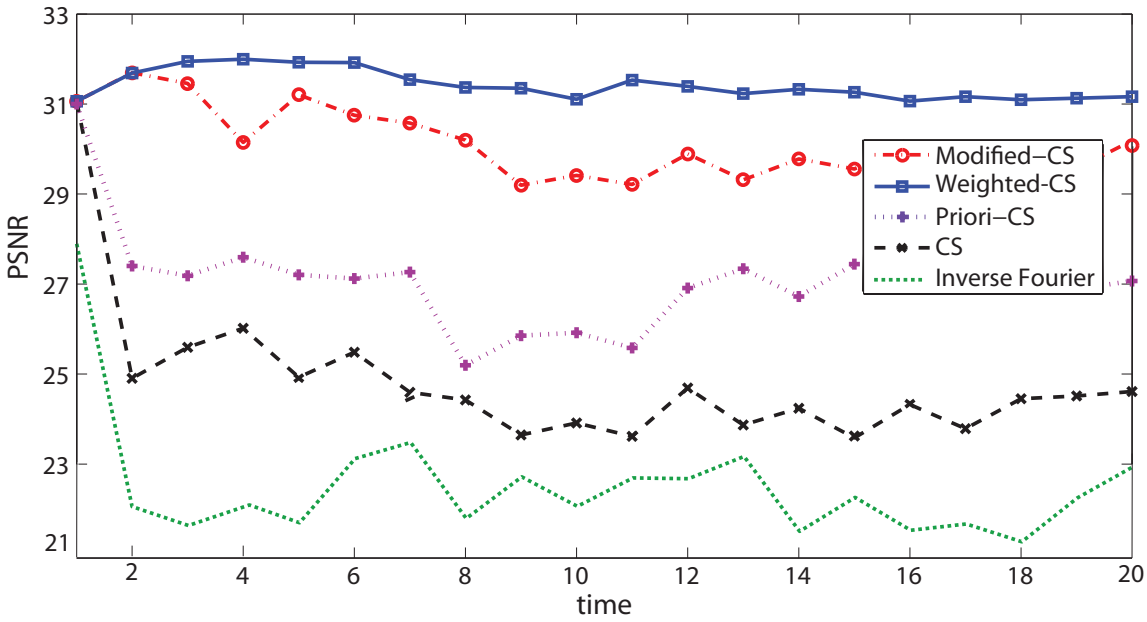


(a) Skull MRI Sequence



(b) Neck MRI Sequence

Figure 4.16: PSNR of the reconstructed images vs. time.



(a) Ankle MRI Sequence

Figure 4.17: PSNR of the reconstructed images vs. time.

4.3.4 Choice of σ

It should be noted that when σ approaches zero, the Gaussian distribution narrows to a spike and thus the proposed Weighted-CS method retrieves the same results as the modified-CS. Figures 4.20 and 4.19, show the reconstruction error as a function of σ for the ECG signals and MRI images used in experiments discussed in sections 3.2 and 3.3. It can be seen that as σ increases, the performance improves and is the lowest possible for a range of σ up to a certain point beyond which the error begins to increase and becomes significantly high as σ becomes too large. Choosing a large σ is however counter to the initial assumption that sparsity changes smoothly with time. It can be seen from the figures that if σ is

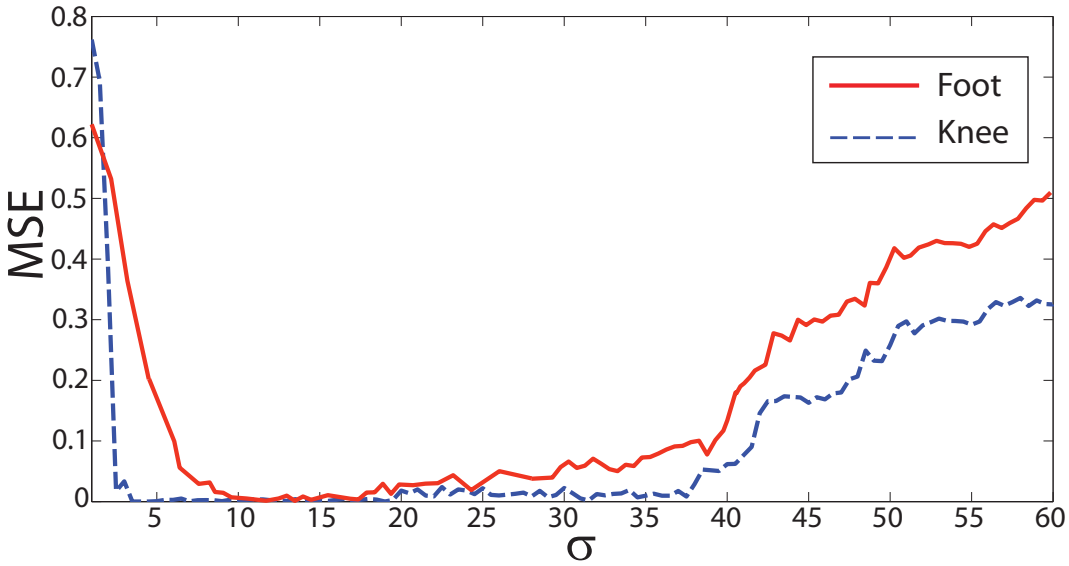
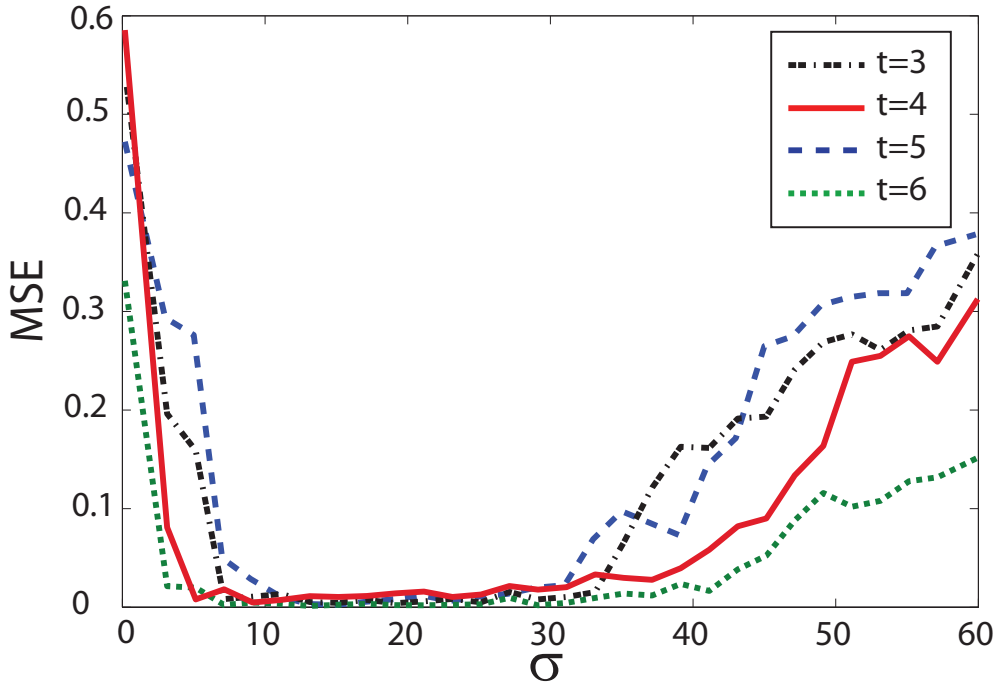


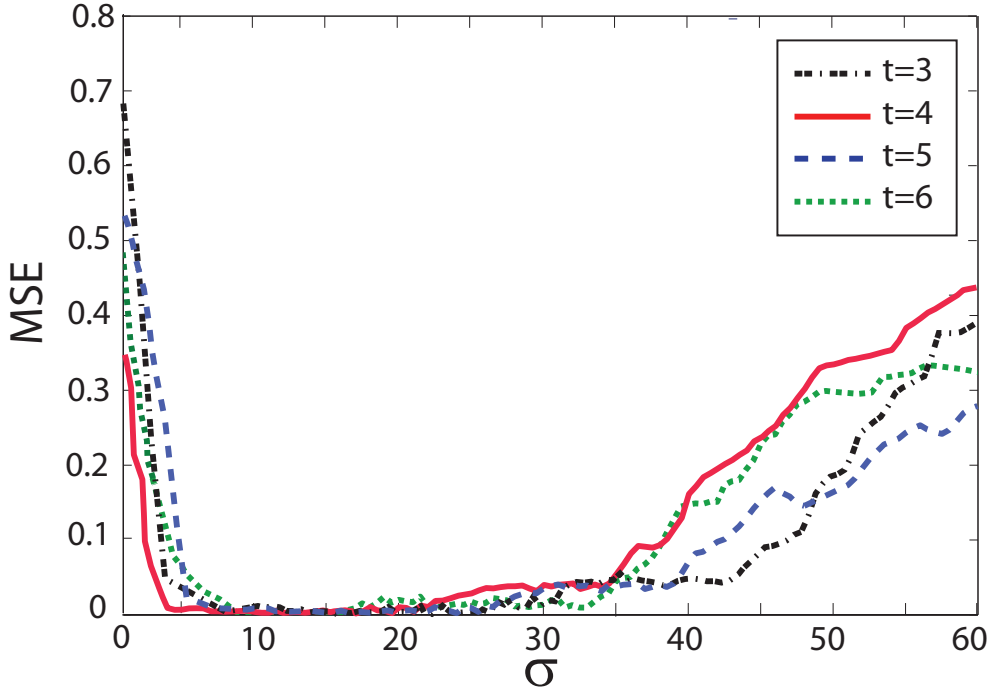
Figure 4.18: MSE of the reconstructed foot/knee images at $t = 2$ vs. σ.

chosen in a reasonable range, the algorithm is not very sensitive to value of σ. For a particular class of signals or images, a single choice of σ would work well for all signals/images in that class.

Figure 4.20(a) shows the reconstruction MSE for one ECG dataset. It can be seen that any $\sigma \in [12, 17]$ would lead to a reasonable reconstruction MSE of less than 0.001. Based on this plot, we choose σ to be 15 for all ECG signals in the experiments. Judging from the reconstruction MSE of the other ECG datasets presented in figure 4.20(b), it is evident that this value of σ leads to minimal error in the other ECG datasets as well. For the MRI images, figure 4.18 shows the MSE error of foot/knee images at $t = 2$. Here any $\sigma \in [10, 18]$ would lead to a reasonable reconstruction MSE of less than 0.001 and so we select a σ of 12 to be used in all succeeding foot/knee MRI images. Figure 4.19(a) and (b), confirm that this value is indeed a proper choice for the other image frames.

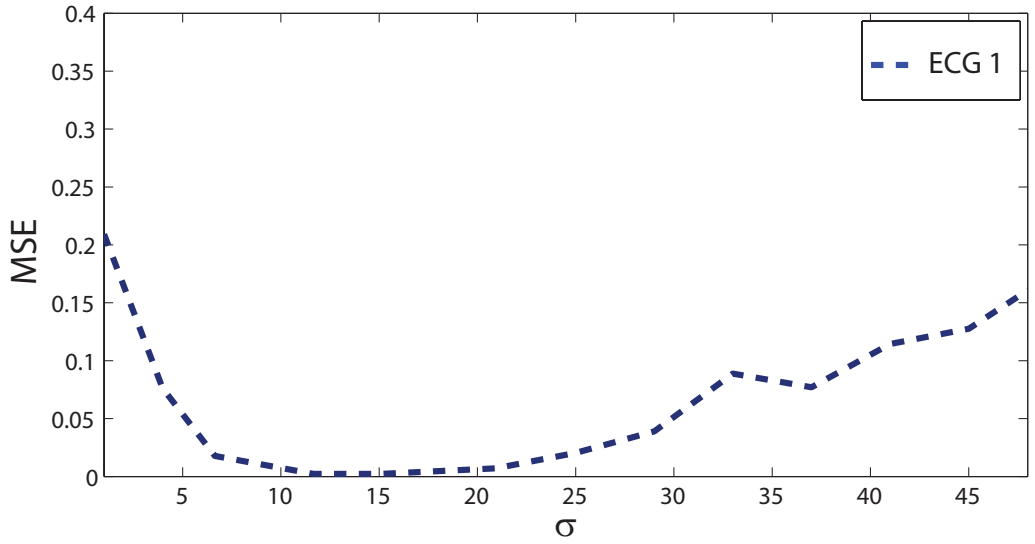


(a) Foot images

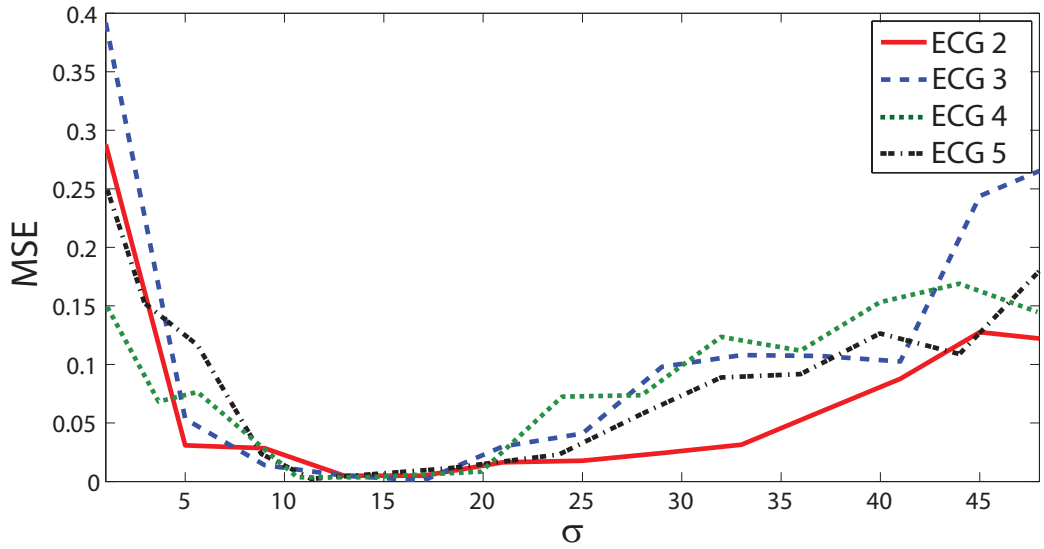


(b) Knee images

Figure 4.19: MSE of the reconstructed signal vs. σ.



(a) ECG1



(b) ECG 2-5

Figure 4.20: MSE of the reconstructed signal vs. σ.

4.4 Summary

In this chapter, problem of reconstructing time sequences of sparse signals from a limited number of linear observations was investigated. Based on the assumption that sparsity changes smoothly with time, a Weighted-CS method, is proposed. This new method extracts some a priori information from the reconstructed signal of the previous time instant and then uses a weighted-ℓ_1 minimization to incorporate this information into the reconstruction process. It is shown, through extensive experiments that the proposed Weighted-CS method can achieve a significant reduction in the number of samples needed compared to the other state-of-art CS algorithms. Lastly, the application of the Weighted-CS in ECG monitoring and real-MRI sampling is explored. Through extensive experimental results on ECG signals and MRI images, it is shown that the proposed algorithm achieves superior performance compared to other the state-of-the-art methods, while imposing significantly less load on the sampling device in terms of memory and energy consumption.

Chapter 5

Conclusions and Future Work

In this thesis, efficient and practical methods for reconstruction of high dimensional signals from compressive measurements have been presented. This work opens up some interesting issues for further investigation.

This final chapter concludes the thesis with a review of the main research contributions and description of the proposed new directions for future research.

5.1 Recovery algorithms

From the detailed review of the recent signal reconstruction methods from compressive samples presented in chapter 2, it is evident that the current widely used ℓ_1-based recovery method, suffers from two main limitations:

- The underlying convex optimization, requires relatively high memory usage and computational cost [11] [12].
- The ℓ_1 norm does not provide a sufficiently close approximation to ℓ_0 [15].

To tackle the above mentioned constraints, an iterative stochastic approximation-based method as a fast and robust algorithm to achieve as close as possible to ℓ_0-norm, using ℓ_p-norm ($0 < p \leq 1$), is proposed and developed. The performance of this algorithm is validated by experimental results on 1D signals and images and is compared against the state-of-the-art convex, greedy and non-convex based CS methods. Moreover, this approach is shown to be robust to noise.

Future work: Much of the work in the field of compressive sensing is focused on reconstruction algorithms and not much has been done in the field of choosing/designing the sampling domain and matrices. Throughout this thesis random matrices whose entries are drawn i.i.d. from Gaussian distribution have been used to sample the signals of interest. It would be of practical interest to investigate the effect of sampling matrices, drawn from other random distributions, on the reconstruction results. Moreover, an interesting and challenging would be to explore the synthesis of measuring metrics for compressive sampling of signals belonging to a special class, using some test signals belonging to that class. It is expected that using a specially designed sampling domain, increases the performance of the proposed algorithms significantly.

5.2 Measures of Sparsity

The Gini index (GI) is explored as an alternative measure of sparsity to the currently popular ℓ_p(pseudo-) norm-based (for $0 < p \leq 1$) measures and its superior robustness and reliability is demonstrated, through extensive experiments. A GI-

based stochastic optimization method is proposed and developed to reconstruct signals from compressive measurements, from significantly smaller number of samples required in the current state-of-the-art CS methods. The robustness of the proposed algorithm in dealing with measurements corrupted with noise is also demonstrated.

Future work: As it is mentioned in chapter 3, the existing constraints of null space (NS) and restricted isometry properties (RIP) of $\mathbf{\Phi}$ for unique reconstruction [3], [18] of k-sparse signals from compressive samples are not applicable to GI. A natural next step of this work would be to develop mathematical results that deal with conditions on the measurement matrix $\mathbf{\Phi}$ for recovering uniquely the original signals from compressive samples when GI is invoked as a sparsity measure. In view of the promising experimental results and and quasi-convexity of GI, it is conjectured that more flexible and less restrictive NS- and RIP-like properties of $\mathbf{\Phi}$ may lead to a unique reconstruction of k-sparse signals. This is left as a challenging problem.

Moreover, through extensive experiments on random signals, the successful recovery rate vs. number of samples taken using the GI (figure 3.4), is illustrated. An interesting and challenging problem, would be to develop the theoretical bounds on the number of needed samples, under which exact reconstruction is possible with GI as a sparsity measure.

5.3 Reconstruction of Dynamic Signals

For reconstructing sequences of time-varying sparse signals, a method based on the assumption that sparsity changes slowly and smoothly with time is proposed and developed. This method, first extracts a probability model of the signal of interest from the reconstructed signal of the previous time instant. It then incorporates this extracted priori-knowledge into the reconstruction process, using a weighted-ℓ_1-based method. Compared against the other state-of-art CS methods, it is shown that the proposed approach requires significantly smaller number samples for a perfect reconstruction. Lastly, the potential application of the developed method in low power ECG wireless-enabled monitoring devices and medical imaging modalities is examined and promising results are obtained.

Future work: One immediate extension of this work would be to use different measures of sparsity such as ℓ_p, $0 < p < 1$ [15] or GI [46], to further reduce the number of samples needed. It should be mentioned that incorporation of priori-knowledge into the GI-maximization process is specifically challenging, as the extension of the weighted-ℓ_1 method to GI index is not trivial. This is left as a challenging problem.

Another interesting challenge would be to investigate the incorporation of the extracted probability model into the measuring process instead of the reconstruction process. This can be done by designing adaptive measurement matrices, that are denser in the locations where higher amount of change is expected.

In the work presented in chapter 4, the extracted probability model is used to reconstruct the signal of interest from small number of samples, however it should be possible to use the extracted model for other purposes such as back-

ground/foreground detection or object tracking in image sequences. Other potential areas of applications, such as the ones mentioned, are to be explored.

Motivated by the encouraging simulation results obtained for the ECG signals, the next step would be actual real-time implementation of the proposed method on embedded platforms.

Bibliography

[1] "The waterloo fractal coding and analysis group." [Online]. Available: http://links.uwaterloo.ca/Repository.html

[2] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406 – 5425, 2006.

[3] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289 – 1306, 2006.

[4] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489 – 509, 2006.

[5] Y. C. Eldar and G. Kutyniok, Eds., *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[6] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling: Building simpler, smaller, and less-expensive digital cameras," *IEEE Signal Processing Magzine*, no. 3, pp. 83–91, 2008.

[7] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing mri: A look at how cs can improve on current imaging techniques," *IEEE Signal Processing Magzine*, no. 3, pp. 72–82, 2008.

[8] M. H. Mahoor, M. Zhou, K. Veon, S. M. Mavadati, and J. Cohn, "Facial action unit recognition with sparse representation," in *IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 336–342, 2011.

[9] T. Bu, A. C. J. Cao, and P. Lee, "A fast and compact method for unveiling significant patterns in high speed networks," in *IEEE International Conference on Computer Communications*, pp. 336–342, 2006.

[10] Y. Zhang, "On theory of compressive sensing via l1 minimization: Simple derivations and extensions," Rice University, CAAM Technical Report TR08-11, 2008.

[11] M. D. T. Blumensath, "Iterative thresholding for sparse approximations," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 629–654, 2008.

[12] ——, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[14] S. Jafarpour, W. Xu, B. Hassibi, and R. Calderbank, "Efficient and robust compressed sensing using optimized expander graphs," *IEEE Transactions on Information Theory*, vol. 55, no. 9, pp. 4299 –4308, 2009.

[15] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 707–710, 2007.

[16] J.Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic $l0$-minimization," *IEEE Transactions on Medical Imaging*, vol. 28, no. 1, pp. 106–121, 2009.

[17] N. Vaswani and W. Lu, "Modified-cs: Modifying compressive sensing for problems with partially known support," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4595 –4607, 2010.

[18] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 3, p. 035020, 2008.

[19] E.J.Candes, M.B.Wakin, and S.P.Boyd, "Enhancing sparsity by re-weighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

[20] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21 – 30, 2008.

[21] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203 – 4215, 2004.

[22] D. Donoho and Y. Tsaig, "Extensions of compressed sensing," *Signal Processing*, vol. 86, no. 3, pp. 549–71, 2006.

[23] J. D. Blanchard, C. Cartis, and J. Tanner, "Compressed sensing: How sharp is the restricted isometry property?" *SIAM Rev.*, vol. 53, no. 1, pp. 105–125, 2011.

[24] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, pp. 253–263, 2008.

[25] B. Shim and J. Wang, "On the recovery limit of sparse signals using orthogonal matching pursuit," *IEEE Transactions on Signal Processing*, vol. 60, no. 99, pp. 4973–4976, 2012.

[26] M. Davenport and M. Wakin, "Analysis of orthogonal matching pursuit using the restricted isometry property," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4395 –4401, 2010.

[27] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Foundations of Computational Mathematics*, vol. 9, no. 3, pp. 317–334, 2009.

[28] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2008.

[29] S. Foucart and M.-J. Lai, "Sparsest solutions of underdetermined linear systems via l_q minimization for $0 \leq q \leq 1$," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 395–407, 2009.

[30] R. G. M.E. Davies, "Restricted isometry constants where l_p sparse recovery can fail for $0 \leq p \leq 1$," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2203–2214, 2009.

[31] J. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Transactions on Automatic Control*, vol. 37, no. 3, pp. 332 – 41, 1992.

[32] P. Sadegh and J. Spall, "Optimal random perturbations for stochastic approximation using a simultaneous perturbation gradient approximation," in *Proceedings of the 1997 American Control Conference*, vol. 6, Evanston, IL, USA, pp. 3582 – 3586, 1997.

[33] J. Maryak and D. Chin, "Global random optimization by simultaneous perturbation stochastic approximation," *IEEE Transactions on Automatic Control*, vol. 53, no. 3, pp. 780 – 783, 2008.

[34] ——, "Global random optimization by simultaneous perturbation stochastic approximation," *Johns Hopkins APL Technical Digest*, vol. 25, no. 2, pp. 91 – 100, 2005.

[35] C. Renjifo, D. Barsic, C. Carmen, K. Norman, and G. S. Peacock, "Improving radial basis function kernel classification through incremental learning and automatic parameter selection," *Neurocomputing*, vol. 72, no. 1, pp. 3–14, 2008.

[36] S.Becker, J. Bobin, and E. J. Candes, "NESTA: A fast and accurate first-order method for sparse recovery," http://www.acm.caltech.edu/~nesta/nesta.html, 2009.

[37] N. Hurley and S. Rickard, "Comparing measures of sparsity," *IEEE Transactions on Information Theory*, vol. 55, no. 10, pp. 4723 – 4741, 2009.

[38] R.Enkhbat and T. Ibaraki, "On the max. and min. of a quasiconvex function," in *Proc. of the Second Int. Conf. on Nonlinear Analysis and Convex Analysis*, Japan, 2003, pp. 37–48.

[39] E. Rentsen, *Quasiconvex Programming and its Applications.* Lap Lambert Academic Publishing, 2009.

[40] Y. V. Venkatesh, A. A. Kassim, and D. Zonoobi, "Medical image reconstruction from sparse samples using simultaneous perturbation stochastic optimization," in 17th IEEE International Conference on Image Processing (ICIP), 3369-3372, 2010.

[41] M. Asif and J. Romberg, "Dynamic updating for sparse time varying signals," in *Information Sciences and Systems, 2009. CISS 2009. 43rd Annual Conference on*, pp. 3 –8, 2009.

[42] H. Mamaghanian, N. Khaled, D. Atienza, and P. Vandergheynst, "Compressed sensing for real-time energy efficient ecg compression on wireless body sensor nodes," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 9, pp. 2456–2465, 2011.

[43] R. Benzid, F. Marir, A. Boussaad, M. Benyousef, and D. Arar, "Fixed percentage of wavelet coefficients to be zeroed for ecg compression," *Electronics Letters*, vol. 39, no. 11, pp. 830–831, 2003.

[44] Y. Rachlin and D. Baron, "The secrecy of compressed sensing measurements," in *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, pp. 813–817, 2008.

[45] S. Aviyente, "Compressed sensing framework for eeg compression," in *Proceedings of the IEEE Workshop on Stat. Signal Proc. (SSP07)*, pp. 181–184, 2007.

[46] D. Zonoobi, A. A. Kassim, and Y. V. Venkatesh, "Gini index as sparsity measure for signal reconstruction from compressive samples," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 927 –932, sept. 2011.

[47] H.-M. D. of Health Sciences and Technology, "Mit-bih database distribution." [Online]. Available: http://ecg.mit.edu.

[48] A. A. Kassim, Y. Niu, and D. Zonoobi, "Wavelet packet transform basis selection method for set partitioning in hierarchical trees," *Journal of Electronic Imaging*, vol. 17, no. 3, pp. 33 007–33 007, 2008.

[49] E. J. Berbari, *Principles of Electrocardiography The Biomedical Engineering Handbook: Second Edition*, J. D. Bronzino, Ed. CRC Press LLC, 2000.

[50] U. Gamper, P. Boesiger, and S. Kozerkey, "Compressed sensing in dynamic MRI," *Magnetic Resonance in Medicine*, vol. 59, no. 2, pp. 365–373, 2008.

[51] W. Lu and N. Vaswani, "Modified compressive sensing for real-time dynamic MR imaging," in *ICIP*, 2009, pp. 3045–3048.

[52] "Magnetic resonance technology information portal." [Online]. Available: http://www.mr-tip.com

List of Publications

- **Dornoosh Zonoobi**, Ashraf A. Kassim, "On the reconstruction of sequences of sparse signals - The Weighted-CS", **Elsevier Journal of Visual Communication** & **Image Representation**: vol. 24, no. 2, pp. 196-202, (2013).

- **Dornoosh Zonoobi**, Ashraf A. Kassim, Y. V. Venkatesh, "Gini Index as Sparsity Measure for Signal Reconstruction from Compressive Samples", **IEEE Journal of Selected Topics in Signal processing** : vol. 5, pp. 927-932, (2011).

- **Dornoosh Zonoobi**, Ashraf A. Kassim, "ECG monitoring using Weighted-Compressive Sensing", Submitted to **IEEE Trans. on Information Technology in Biomedicine** (2012).

- **Dornoosh Zonoobi**, Ashraf A. Kassim, "Weighted-CS for reconstruction of highly under-sampled dynamic MRI sequences", **Asia-Pacific Signal** & **Information Processing Association Annual Summit and Conference (APSIPA ASC)**, (2012).

- Y. V. Venkatesh, Ashraf A. Kassim, **Dornoosh Zonoobi**, "Medical im-

age reconstruction from sparse samples using Simultaneous Perturbation Stochastic Optimization", **IEEE International Conference on Image Processing (ICIP)**: pp. 3369-3372, (2010).