

**Nonconvex l_p Minimization for Compressed
Sensing under General Perturbations**

Ph.D. Thesis
in
Electrical and Electronics Engineering
University of Gaziantep

Supervisor
Prof. Dr. Arif Nacaroglu

By
Taner İNCE
September 2012

© 2012 [Taner İnce]

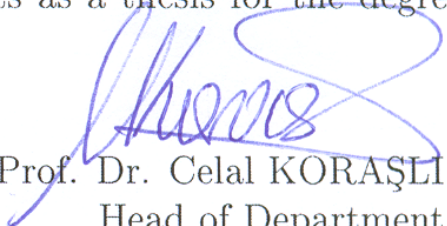
T.C.
GAZIANTEP UNIVERSITY
GRADUATE SCHOOL OF
NATURAL AND APPLIED SCIENCES
ELECTRICAL AND ELECTRONICS ENGINEERING

Name of the Thesis : Nonconvex l_p Minimization for Compressed Sensing
under General Perturbations
Name of the Student : Taner İNCE
Exam Date : 14.09.2012

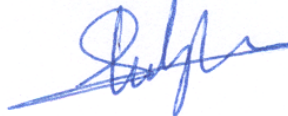
Approval of the Graduate School of Natural and Applied Sciences.

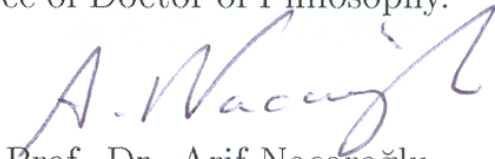

Prof. Dr. Ramazan KOÇ
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.


Prof. Dr. Celal KORAŞLI
Head of Department

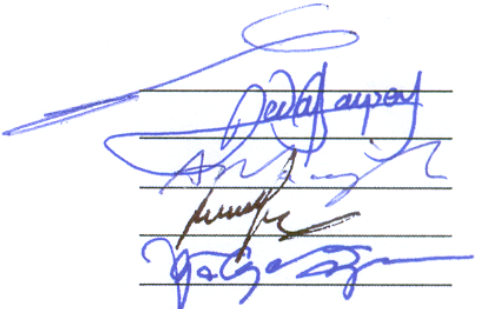
This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Doctor of Philosophy.


Asst. Prof. Dr. Nurdal Watsuji
Co-Supervisor


Prof. Dr. Arif Nacaroglu
Major Supervisor

Examining Committee Members:

Prof. Dr. Cüneyt GÜZELİŞ
Prof. Dr. Sedat BAYSEÇ
Prof. Dr. Arif NACAROĞLU
Prof. Dr. Rauf MİRZABABAYEV
Assoc. Prof. Dr. Gölge ÖĞÜCÜ



I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Taner İnce

ABSTRACT

NONCONVEX l_p MINIMIZATION FOR COMPRESSED SENSING UNDER GENERAL PERTURBATIONS

İNCE, Taner

Ph.D. in Electrical and Electronics Engineering

Supervisor: Prof. Dr. Arif Nacaroglu

Co-Supervisor: Asst. Prof. Dr. Nurdal Watsuji

September 2012, 84 pages

Compressed sensing (CS) provides a framework for acquisition of signals far below the Nyquist rate if it is represented as sparse or compressible on an orthonormal basis. Recovering sparse and compressible signals using l_p minimization with $p < 1$ when some part of the support of the signal is known a priori is studied. A sparse reconstruction method based on l_p minimization with partially known set is proposed and recovery conditions are given. Error bound noise constant and error bound compressibility constants are obtained for sparse and compressible signal cases. Theoretical results show that l_p minimization with partially known support is stable and robust. Experimental results are presented to expose the modification of l_p minimization improves performance and need fewer samples to reconstruct the signal. Also l_p minimization with $p < 1$ under both additive and multiplicative noise in compressed sensing is studied. The results are based on the restricted isometry constant and relative perturbations. The exact reconstruction of a signal is not possible under additive and multiplicative noise. However simulation results show that under multiplicative noise, l_p minimization performs better than l_1 minimization for average reconstruction error with varying parameters such as noise level, sparsity and measurement level.

Key words: Compressed sensing, restricted isometry property, l_p minimization, sparse signal recovery, partially known support, perturbation, l_1 minimization

ÖZET

GENEL BOZULMALAR ALTINDA SIKIŞTIRMALI ALGILAMA İÇİN DIŞBÜKEY l_p KÜÇÜMSEME

İNCE, Taner

Doktora Tezi, Elektrik ve Elektronik. Müh.

Tez Yöneticisi: Prof. Dr. Arif Nacaroğlu

Tez Yönetici Yardımcısı: Yrd. Doç. Dr. Nurdal Watsuji

Eylül 2012, 84 sayfa

Sıkıştırılabilir algılama, eğer bir işaret dik bir dönüşümle seyrek veya sıkıştırılabilir olarak ifade edilebilirse o işaretin Nyquist hızının altında örneklenmesine olanak sağlamaktadır. Kısmi destek kümesi bilinen seyrek veya sıkıştırılabilir işaretlerin l_p küçümseme ile geri çatılması çalışılmaktadır. Kısmi destek kümesi l_p küçümseme tabanlı seyrek geri çatılma metodu sunulmuştur ve geri kazanım şartları verilmiştir. Seyrek ve sıkıştırılabilir işaretler için hata sınır sabiti ve hata sıkıştırma sabiti elde edilmiştir. Teorik sonuçlara göre l_p küçümsemenin kararlı ve gürbüz olduğu gözlemlenmiştir. Deneysel sonuçlara göre l_p küçümseme daha az sayıda örnek kullanarak işaretin geri çatılmasına olanak sağlamaktadır. Buna ek olarak $p < 1$ değerli l_p küçümseme metodunun ekleyici ve çarpıcı gürültü altındaki performansı incelenmiştir. Sonuçlar kısıtlı izometri özelliği ve göreceli bozulmalara göre verilmiştir. İşaretin çözümü kesin olmamasına rağmen simulasyon sonuçlarına göre çarpıcı gürültü altında l_p küçümseme l_1 küçümsemeye göre geri kazanım hata oranına göre daha iyi sonuç vermektedir. Bu parametreler gürültü seviyesi, seyreklik ve ölçüm seviyesidir.

Anahtar kelimeler: Sıkıştırılabilir algılama, kısıtlı izometri özelliği, l_p küçümseme, seyrek işaret kazanımı, kısmi bilinen destek, bozukluk, l_1 küçümseme

I dedicate this work to my parents.....

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Prof. Dr. Arif Nacaroglu, for his excellent guidance, caring, patience, and providing me with an excellent atmosphere for doing research.

I would also like to thank my parents. They were always supporting me and encouraging me with their best wishes.

TABLE OF CONTENTS

CHAPTER

1	INTRODUCTION	1
1.1	Thesis Motivation	3
2	COMPRESSED SENSING	5
2.1	Undersampled Situations	5
2.1.1	Signal Recovery From Undersampled Measurements	6
2.1.2	Basis Pursuit	6
2.1.3	Restricted Isometry Property	7
2.1.4	Noiseless Recovery	8
2.1.5	Noisy Recovery	9
2.2	Greedy Algorithms	10
2.2.1	Orthogonal Matching Pursuit	10
2.2.2	OMP Algorithm	11
2.2.3	Compressive Sampling Matching Pursuit (COSAMP)	12
2.2.4	COSAMP Algorithm	13
2.3	Simulation Results	14
2.4	Applications	18
2.5	Summary	19
3	NONCONVEX COMPRESSED SENSING WITH PARTIALLY KNOWN SIGNAL SUPPORT	20
3.1	Introduction	20
3.2	Nonconvex Compressed Sensing	21
3.3	CS With Partially Known Signal Support	24
3.4	$l_{p<1}$ Minimization With Partially Known Support	25
3.5	Experimental Results	29
3.5.1	Exact Recovery Results	31
3.5.2	The Sparse Case	33
3.5.3	The Compressible Case	35
4	NONCONVEX COMPRESSED SENSING FOR GENERAL PERTURBATIONS	39
4.1	Introduction	39
4.1.1	Notations and Symbols	40
4.2	Completely Perturbed l_1 Minimization	40
4.2.1	RIP for A	40
4.2.2	Stability From Completely Perturbed Observations	41

4.2.3	Derivation of Total Perturbation Bound $\varepsilon'_{\Phi,K,y}$	41
4.3	$l_{p<1}$ Minimization Under General Perturbations	42
4.4	Stability and Instance Optimality in Nonconvex Compressed Sensing	47
4.5	Stability and Instance Optimality in Completely Perturbed $l_{p<1}$ Minimization	50
4.6	Simulation Results	52
5	CONCLUSION	60
 APPENDIX		
A	DERIVATION OF IRLS	66
B	CODES FOR SIMULATIONS IN CHAPTER 3	67
B.1	MATLAB CODES	67
B.1.1	IRLS	67
B.1.2	IRLS with partially known support	69
B.1.3	Script for Simulation Results (Sparse Case)	70
B.1.4	Script for Simulation Results (Compressible case)	72
C	CODES FOR SIMULATIONS IN CHAPTER 4	75
C.1	MATLAB CODES	75
C.1.1	Noisy IRLS	75
C.1.2	Script for varying sparsity level K	77
C.1.3	Script for varying sparsity level $\varepsilon_{\Phi,K,y}$	78
C.1.4	Script for varying measurement level under additive Gaus- sian noise	79
PUBLICATIONS		82

LIST OF TABLES

Table

3.1	Commonly used nonconvex penalty functions	21
-----	---	----

LIST OF FIGURES

Figure

1.1	Typical Compression Scheme	1
1.2	(a) 256×256 cameraman image (b) Restored image using only 8000 largest coefficients	3
1.3	Wavelet coefficients of cameraman image	3
2.1	Illustration of CS process with random gaussian measurement matrix Φ and discrete cosine transform (DCT) matrix Ψ	6
2.2	Exact reconstruction frequency versus measurement level using Basis Pursuit	14
2.3	Exact reconstruction frequency versus measurement level using OMP	15
2.4	Exact reconstruction frequency versus measurement level using COSAMP	15
2.5	Exact reconstruction frequency versus measurement level for sparsity level $K=20$ using three algorithms	16
2.6	Exact reconstruction frequency versus sparsity using BP	16
2.7	Exact reconstruction frequency versus sparsity using OMP	17
2.8	Exact reconstruction frequency versus sparsity using COSAMP	17
2.9	Exact reconstruction frequency versus sparsity for measurement level $M = 100$ using three algorithms	18
3.1	Penalty functions in Table I (a) SCAD (b) Zhang (c) $l_{p<1}$ (d) Log	22
3.2	(a) $C_{a,k,s,p}$ versus p , (b) $D_{a,k,s,p}$ versus p , for different values of a	29
3.3	(a) Reconstruction error to the p^{th} power versus p for various s for $M = 256$, (b) Theoretical error bound for $s = 60, k = 20$, (c) Theoretical error bound for $s = 90, k = 30$, (d) Theoretical error bound for $s = 120, k = 40$	31
3.4	Exact reconstruction frequency of modified $l_{p<1}$ minimization and MOD-CS for (a) $s = 0$, (b) $s = 12$, (c) $s = 20$	32
3.5	Exact reconstruction frequency of modified $l_{p<1}$ minimization and MOD-CS for (a) $M = 80$, (b) $M = 100$ and (c) $M = 120$	33
3.6	Performance of $l_{p<1}$ minimization with partially known support and MOD-CS in terms of SNR for sparse signal, varying the number of measurements for (a) $s=4$, (b) $s=16$ and (c) $s=24$	34
3.7	Performance of $l_{p<1}$ minimization with partially known support and MOD-CS in terms of SNR for sparse signals, varying the support size s for (a) $M=80$, (b) $M=120$ and (c) $M=160$	35

3.8	Performance of $l_{p<1}$ minimization with partially known support and MOD-CS in terms of SNR for compressible signals, varying the number of measurements for (a) $s=4$, (b) $s=16$ and (c) $s=24$. The coefficients decay with a power $\tau = 1.5$	36
3.9	Performance of $l_{p<1}$ minimization with partially known support and MOD-CS in terms of SNR for compressible signals, varying the support size s for (a) $M=100$, (b) $M=120$ and (c) $M=140$. The coefficients decay with a power $\tau = 1.5$	37
3.10	Performance of $l_{p<1}$ minimization with partially known support and MOD-CS in terms of SNR for cardiac image, varying the number of measurements ($s=64$).	38
4.1	Error bound noise constant ($C_1^{(1)}$) versus p	45
4.2	Error bound noise compressibility constant ($C_1^{(2)}$) versus p	45
4.3	Average relative error versus ε_Φ for (a) $K = 10$, (b) $K = 20$, (c) $K = 30$	54
4.4	Average relative error versus ε_Φ with measurement noise $\sigma = 0.05$ for (a) $K = 10$, (b) $K = 20$, (c) $K = 30$	55
4.5	Average relative error versus measurement level M for (a) $\varepsilon_\Phi=0$, (b) $\varepsilon_\Phi=0.05$ and (c) $\varepsilon_\Phi=0.1$	56
4.6	Average relative error versus measurement level M for $\varepsilon_\Phi=0.05$ with measurement noise, (a) $\sigma = 0.01$ (b) $\sigma = 0.05$	57
4.7	Average relative error versus sparsity level K for (a) $\varepsilon_\Phi=0$, (b) $\varepsilon_\Phi=0.05$ and (c) $\varepsilon_\Phi=0.1$	58
4.8	Average relative error versus sparsity level K for (a) $\varepsilon_\Phi=0.05$ and measurement noise $\sigma = 0.05$	58
4.9	Average relative error versus p with measurement noise $\sigma = 0.05$	59

LIST OF SYMBOLS

CS: Compressed Sensing

Φ : Measurement Matrix

Ψ : Orthonormal Basis

K : Signal sparsity

y : Measurement vector

x : General signal

\hat{x} : Reconstructed signal

e : Measurement noise

E : Perturbation matrix

A : Perturbed matrix

BP: Basis pursuit

BPDN: Basis pursuit denoising

COSAMP: Compressive Sampling matching pursuit

OMP: Orthogonal matching pursuit

ROMP: Regularized Orthogonal matching pursuit

DFT: Discrete Fourier transform

DCT: Discrete cosine transform

DWT: Discrete wavelet transform

T : Partially known set

s : Size of $|T|$

T_0 : K largest components of x

RIP: Restricted isometry property

RIC: Restricted isometry constant

δ_K : Restricted isometry constant

ϵ : Noise level

$\|x\|_0$: l_0 norm of x

$\|x\|_1$: l_1 norm of x

$\|x\|_2$: l_2 norm of x

$\|x\|_p$: l_p norm of x

r : Residual vector

$\epsilon'_{\Phi, K, y}$: Total noise parameter

$\mathcal{O}(n)$: Order of n

CHAPTER 1

INTRODUCTION

The classical Nyquist sampling theorem [1] suggests that the sampling rate must be at least twice the maximum frequency content in the signal in order to reconstruct the original signal. In numerous applications such as digital cameras, the Nyquist rate can be so high that many samples result in that compression becomes a necessity for storage or transmission. Standard Analog to Digital Converters (ADC) follows this rule when sampling a signal. However, if a signal has a bandwidth around GHz levels or not bandlimited, then standard ADC are not capable of sampling this kind of signals, such as Ultra Wide Band Signals (UWB). Therefore sampling this kind of signals pushes the performance limits of ADCs around GHz levels, which is not acceptable in nowadays if sampling frequencies of the ADCs are compared.

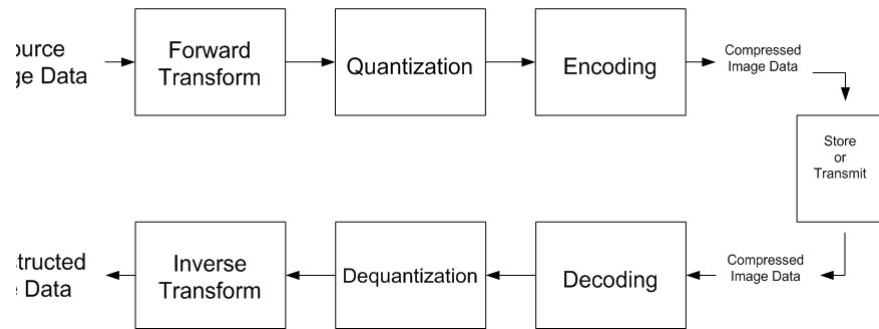


Figure 1.1: Typical Compression Scheme

In data acquisition systems transform coding [2] plays a very important role. In a typical compression scheme given in Fig. 1.1, the image is acquired from a digital camera or an acquisition device, then all the transform coefficients (Wavelet, Discrete Cosine Transform, etc) are calculated and then small coefficients are discarded, finally the remaining coefficients are encoded in an adaptive manner based the compression scheme such as JPEG, JPEG2000 etc.

But this "sample and then encode" framework has some disadvantages. First, initial samples may be quite large for even if the number of large coefficients is small, second, even if the number of large coefficients is small, all the transform coefficients must be calculated. As a third disadvantage, locations of large coefficients must be encoded.

Many natural signals have sparse representations when transformed into a convenient basis. Suppose that signal $x \in \mathbb{R}^N$ such as the image in Fig. 1.2 (a) can be expanded using an orthonormal basis such as wavelet DCT or any other convenient basis $\Psi = [\psi_1 \ \psi_2 \dots \psi_N]$ as follows:

$$u(t) = \sum_{i=1}^N x_i \psi_i(t) \quad (1.1)$$

where x is the coefficient vector of u . u can be expressed as Ψx where Ψ is the $N \times N$ matrix with $\psi_1 \ \psi_2 \dots \psi_N$ as its columns. When a signal has a sparse representation, small coefficients can be discarded without much perceptual loss. Suppose that $u_K := \Psi x_K$ is represented as where x_K is the vector of the largest K coefficients. This type of signals called K -sparse meaning that it has at most K nonzero entries. Since the columns of Ψ constitutes an orthonormal basis we have

$$\|u - u_K\|_2 = \|x - x_K\|_2 \quad (1.2)$$

and if x is sparse or compressible meaning that the magnitude of coefficients in x decay quickly, then u is well-approximated by u_K and therefore $u - u_K$ is small. In Fig. 1.2 (a) 256×256 camera is shown and its wavelet coefficients are shown in Fig. 1.3. Many of the wavelet coefficients of this image are small and most of the information of the image is concentrated on a few of large coefficients. In Fig. 1.2 (b) image is reconstructed using only 8000 coefficients out of 65536. As it is noticed the difference between Fig. 1.2 (a) and Fig. 1.2 (b) is hardly noticeable. This is the basic principle of most modern transform coding techniques including JPEG, JPEG2000 [2]. Therefore sparsity is a fundamental modelling tool which provides efficient signal processing such as statistical estimation, classification, data compression etc.



(a)

(b)

Figure 1.2: (a) 256×256 cameraman image (b) Restored image using only 8000 largest coefficients

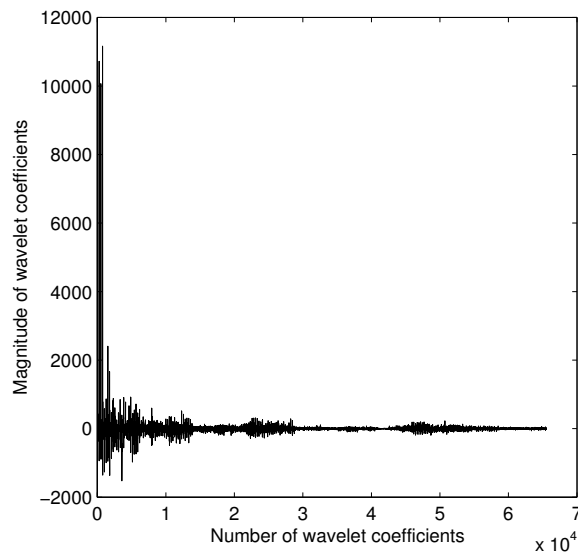


Figure 1.3: Wavelet coefficients of cameraman image

A new signal processing area, Compressive Sampling / Compressed Sensing (CS) which emerged in the works [3–6], addresses these disadvantages by acquiring the signal in compressed form without acquiring all samples, and sampling it far below the Nyquist rate [1].

1.1 Thesis Motivation

In Chapter 2 the detailed literature search is presented and basic concepts of compressed sensing and the algorithms used in the recovery process are given.

Basically, necessary conditions for l_1 minimization and commonly used greedy algorithms are explained. Also simulation results are given for l_1 minimization and greedy algorithms are given at the end of Chapter 2.

Chapter 3 is related with the sparse approximation problem using nonconvex compressed sensing when some part of the signal is known a priori. Background on nonconvex approximation problem and a literature survey is also given in this chapter. Theorems are proved to show the validity of nonconvex compressed sensing with partially known support. Simulation results are given at the end of this chapter and compared to l_1 minimization with partially known support. Simulations are carried on with different parameter sets. These are partially known signal support level and measurement level. Also validation of the theoretical results are presented for the cardiac signal by varying the nonconvexity parameter p .

Chapter 4 deals with the perturbation analysis in nonconvex compressed sensing. Additive and multiplicative noise is considered for nonconvex compressed sensing. Theoretical and simulation results are given in this chapter and compared to l_1 minimization. The simulations are realized for different sets of parameters such as measurement level, sparsity level, perturbation level, and nonconvexity parameter.

CHAPTER 2

COMPRESSED SENSING

Compressed Sensing/Compressive Sampling (CS) uses the fact that any natural signal is sparse or compressible, if it is expressed on some transform basis Ψ (Wavelet, Fourier, etc). For example, a periodical sum of a finite number of sinusoidal functions has a sparse representation if they are expressed in Fourier basis. In general, signals are of interest may not be sparse, but they may be thought approximately as a sparse set. For example transform coefficients of natural signals decay geometrically in magnitude. In this chapter background on CS and algorithms used in the recovery process are given and these algorithms are run for different sparsity and measurement levels.

2.1 Undersampled Situations

Consider the problem of recovering a vector $u \in \mathbb{R}^N$ from M linear measurements y about u of the form (2.1)

$$y = \Phi u \tag{2.1}$$

Then, replacing $u = \Psi x$ into (2.1), y can be written as

$$y = \Phi u = \Phi \Psi x = \Theta x \tag{2.2}$$

where $\Theta = \Phi \Psi$ is an $M \times N$ measurement with $M \ll N$ meaning that it has far fewer rows than columns. This situation can be demonstrated in Fig. 2.1

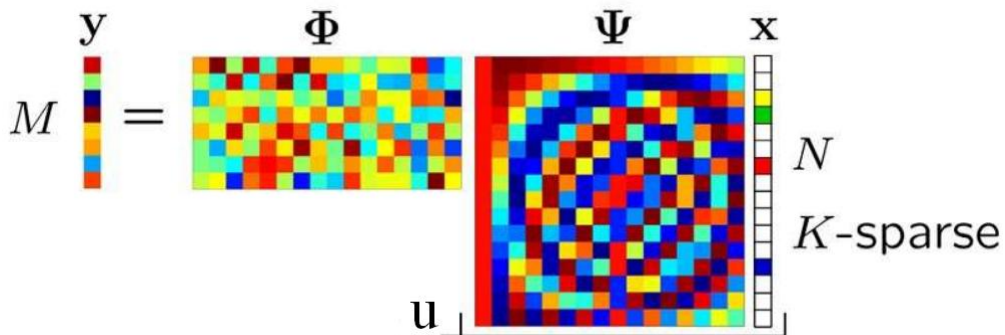


Figure 2.1: Illustration of CS process with random gaussian measurement matrix Φ and discrete cosine transform (DCT) matrix Ψ .

2.1.1 Signal Recovery From Undersampled Measurements

Reconstruction algorithms can be divided into 3 broad categories. These are convex relaxation by replacing intractable combinatorial problem (2.3) with (2.5) and solve using convex optimization techniques, greedy pursuits and non-convex optimization. Greedy pursuits try to find a one or more components that produces the greatest improvement in quality. Nonconvex optimization problems relaxes the l_0 minimization problem in (2.3) to a nonconvex problem (l_p minimization) to obtain a global minimum [7]. There also exist reconstruction algorithms such as Bayesian framework [8] and brute force.

2.1.2 Basis Pursuit

The system of equations in (2.2) is an undetermined system $M \ll N$, where the number of equations are far fewer than the number of unknowns. This occurs in many areas of signal processing problems. This system appears to be ill-conditioned. However, x is K -sparse and locations of the nonzero terms in x are known, then the problem can be solved provided $M \geq K$.

In order to obtain an exact solution of this problem, one of the simplest way to recover signal from its measurements $y = \Phi\Psi x$ is to solve the l_0 minimization. l_0 norm counts the nonzero terms in reconstructed vector \hat{x} .

$$\min \|\hat{x}\|_0 \quad \text{subject to} \quad \Phi\Psi\hat{x} = y \quad (2.3)$$

(2.3) can reconstruct a K -sparse signal with high probability using only $M = K + 1$ iid (independent and identically distributed) Gaussian measurements. Solution must be sparse basis which means that it must have only K nonzero terms. In theory l_0 minimization works perfect. However, it is NP-Hard in

general, it requires an exhaustive search of all $\binom{N}{K}$ possible locations of nonzero entries in x .

It is possible to propose a minimum energy solution

$$\min \|\hat{x}\|_2 \quad \text{subject to} \quad \Phi\Psi\hat{x} = y \quad (2.4)$$

to solve the system. This optimization problem has the closed form solution $\hat{x} = \Theta^T(\Theta\Theta^T)^{-1}y$. But the solution can be a nonsparse \hat{x} with many nonzero components and one can almost never find a K -sparse solution.

The problem in (2.3) is NP-Hard and combinatorially intractable, however Donoho et. al [9] showed that for certain measurement matrices Φ , this problem is equivalent to

$$\min \|\hat{x}\|_1 \quad \text{subject to} \quad \Phi\Psi\hat{x} = y \quad (2.5)$$

which is called Basis Pursuit (BP).

2.1.3 Restricted Isometry Property

In order to recover a sparse signal stably from far fewer observations by solving (2.5), the matrix $\Phi\Psi$ must satisfy the Restricted Isometry Property (RIP). Candes and Tao proved that if certain matrices satisfy the RIP then programs (2.3) and (2.5) are equivalent [4].

Definition 2.1.1. *For each integer $K = 1, 2, \dots$ define the isometry constant δ_K of a matrix $\Phi\Psi$ as the smallest number such that*

$$(1 - \delta_K)\|x\|_2^2 \leq \|\Phi\Psi x\|_2^2 \leq (1 + \delta_K)\|x\|_2^2 \quad (2.6)$$

which holds for all sparse vectors x . δ_K is called restricted isometry constant (RIC).

If δ_K is small, then the matrix $\Phi\Psi$ is approximately an orthonormal matrix. It is difficult to check whether a measurement matrix satisfies RIP or not. However in the literature it has been shown that certain matrices satisfy the RIP property with high probability such as, Gaussian, Bernoulli and partial Fourier matrices. Candes and Tao [4] showed that if a measurement matrix satisfies the restricted isometry condition, BP recovers all sparse signals exactly. In other publication of Candes the results in [4] are sharpened and recovery conditions for noiseless measurement and noisy measurements are given [10].

Verifying the RIP for a matrix may be a difficult task. It requires the combinatorial search for all submatrices by selecting K columns. it is difficult to check $\binom{N}{K}$ columns, also calculation of spectral norm of a random matrix is

not easy generally. However some certain matrices having special distribution can hold RIP with high probability. In [11] it is shown that $\|\Phi x\|_2^2$ is strongly concentrated about its expected value as

$$\Pr(|\|\Phi x\|_2^2 - \|x\|_2^2| \geq \epsilon \|x\|_2^2) \leq 2e^{-Mc_0(\epsilon)}, \quad 0 < \epsilon < 1 \quad (2.7)$$

where the probability is taken over all $M \times N$ matrices Φ and $c_0(\epsilon)$ is a positive constant that depends on ϵ . Some examples of matrices that can satisfy (2.7) are Gaussian and Bernoulli matrices.

An $M \times N$ random matrix whose entries $\phi_{i,j}$ are distributed as

$$\phi_{i,j} \mathcal{N}(0, \frac{1}{M}) \quad (2.8)$$

satisfy (2.7) with $c_0(\epsilon) = \epsilon^2/4 - \epsilon^3/6$. Also matrices where the entries are realizations of Bernoulli distributions like

$$\phi_{i,j} := \begin{cases} +1/\sqrt{M} & \text{with probability } \frac{1}{2}, \\ -1/\sqrt{M} & \text{with probability } \frac{1}{2}, \end{cases} \quad (2.9)$$

or distributions such as

$$\phi_{i,j} := \begin{cases} +\sqrt{3/M} & \text{with probability } \frac{1}{6}, \\ 0 & \text{with probability } \frac{2}{3}, \\ -\sqrt{3/M} & \text{with probability } \frac{1}{6}, \end{cases} \quad (2.10)$$

satisfy (2.7) with $c_0(\epsilon) = \epsilon^2/4 - \epsilon^3/6$.

Theorem 2.1.1. [11] *Suppose that M , N , and $0 < \delta < 1$ are given. If the probability distribution generating the $M \times N$ matrices Φ satisfies (2.7), then there exists constants $c_1, c_2 > 0$ depending only on δ such that the RIP (2.6) holds for Φ with the prescribed δ and any $K \leq c_1 M / \log(N/K)$ with probability not less than $1 - 2e^{-c_2 M}$.*

It is proved in [11] that RIP holds for Φ with a high probability when the matrix is drawn according to the one of the distributions (2.8), (2.9), and (2.10).

2.1.4 Noiseless Recovery

Theorem 2.1.2. *Suppose that x_K is the best K -term approximation to x and $\delta_{2K} < \sqrt{2} - 1$ then the solution \hat{x} to (2.5) obeys*

$$\|\hat{x} - x\|_2 \leq C_0 K^{-1/2} \|x - x_K\|_1 \quad (2.11)$$

for some constant C_0 . In particular, if x is K -sparse then the recovery is exact.

2.1.5 Noisy Recovery

Since in any acquisition device, measured data may be corrupted by many sources, so one needs to reconstruct the signal in any noisy environment. In CS theory [12], supposing that the measurements are corrupted at the output of the acquisition device, it is modeled as

$$y = \Phi\Psi x + e \quad (2.12)$$

where e is the noise term. Then program in (2.5) is relaxed to

$$\min\|\hat{x}\|_1 \quad \text{subject to} \quad \|\Phi\Psi\hat{x} - y\|_2 \leq \epsilon \quad (2.13)$$

and is called Basis Pursuit Denoising (BPDN).

Theorem 2.1.3. *Suppose that x_K is the best K -term approximation to x and $\delta_{2K} < \sqrt{2} - 1$ then the solution \hat{x} to (2.13) obeys*

$$\|\hat{x} - x\|_2 \leq C_0\epsilon + C_1K^{-1/2}\|x - x_K\|_1 \quad (2.14)$$

for constants C_0 and C_1 [10], within the noise level $\|e\|_2 \leq \epsilon$. where

$$C_0 = \frac{4\sqrt{1 + \delta_{2K}}}{1 - (\sqrt{2} + 1)\delta_{2K}}, \quad C_1 = \frac{2(1 + (\sqrt{2} - 1)\delta_{2K})}{1 - (\sqrt{2} + 1)\delta_{2K}} \quad (2.15)$$

If x is K -sparse which means that it has only K components and there is no error in the measurements then the recovery will be exact.

When $\delta_K = 0.2$ the error in (2.14) is less than $4.2K^{-1/2}\|x - x_K\|_1 + 8.5\epsilon$. In the following chapters orthonormal transformation matrix Ψ is assumed to be identity matrix I in order to simplify the notations ($\Psi = I$). The RIP property also holds for any orthonormal basis.

In [3], Donoho asked a question "Can we not just directly measure the part that will not end up being thrown away?" and showed the possibility of recovering a sparse signal from nonadaptive measurements which is very small than the signal dimension. If a signal x has a sparse representation on some orthonormal basis (e.g., Wavelet, Fourier) then it is possible to recover the expansion coefficients using only $M = \mathcal{O}(K \log(N/K))$ non-adaptive measurements, M is the measurement number, N is the signal dimension and K is the sparsity level.

Another important study about CS is the signal reconstruction from corrupted measurements [4]. Candes considers the problem of recovering a sparse signal with corrupted measurements. He studied to recover an input signal $x \in R^N$ from $y = \Phi x + e$ measurements, where Φ is M by N sensing matrix with $M \ll N$, and e is the error term. Then under suitable conditions

on sensing matrix Φ , the input x is the unique solution to the l_1 minimization problem given as $\min_{\hat{x} \in R^N} \|y - \Phi \hat{x}\|_1$ provided that the norm of error vector is not too large.

In an other publication [5], Candes considers the problem of reconstructing a signal from highly incomplete frequency samples. For a discrete signal f and a randomly chosen set of frequencies Ω , if f is composed of superposition of $|T_0|$ spikes expressed as $f(t) = \sum_{\tau \in T_0} f(\tau) \delta(t - \tau)$ satisfying $|T_0| \leq C_M (\log N)^{-1} (|\Omega|)$ for some constant $C_M > 0$, again f can be reconstructed as the solution of l_1 minimization problem $\min_g \sum_0^{N-1} |g(t)|$, subject to $\hat{g}(\omega) = \hat{f}(\omega)$.

In order to consider how compressive sampling works, it is suitable to give an example in signal processing area. Suppose that we are given an incomplete set of frequency samples of a discrete signal x of length N . Our aim is to reconstruct the signal x given only K samples in the Fourier domain such that

$$y_k = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} x_t e^{-j2\pi\omega_k t/N} \quad (2.16)$$

where the visible frequencies ω_k are a subset Ω that has a size K . In this problem, the sensing matrix Φ is chosen as N by N DFT matrix. Then Candes [5] showed that it is always (almost) possible to recover the signal x exactly by solving the BP problem. Assume that signal x is K -sparse and that we are given M Fourier coefficients with frequencies selected uniformly at random. Suppose that the number of observations obeys

$$M > K \log(N) \quad (2.17)$$

then BP reconstructs x exactly with a high probability. The first important result is that the information is not lost even if just K frequency coefficients are measured. As a second result, nonzero coordinates and amplitudes of the signal x are all completely unknown.

2.2 Greedy Algorithms

Greedy algorithms find the support (location of nonzero entries) of the sparse vector x iteratively. When the support of x is computed correctly then pseudo-inverse of the measurement matrix related to corresponding columns can be used to reconstruct the original signal x .

2.2.1 Orthogonal Matching Pursuit

Mallat and Zhang [13] introduced Orthogonal Matching Pursuit (OMP) and analyzed by Gilbert and Tropp [14]. OMP is a greedy algorithm and for

an N dimensional signal with K nonzero entry it reconstructs signal x from $\mathcal{O}(K \log(N))$ random measurements.

Suppose x is an arbitrary K -sparse signal in R^N . Construct an $M \times N$ measurement matrix Φ and observe the M measurements as $y = \Phi x$ and columns of Φ are denoted as $\varphi_1, \varphi_2 \dots \varphi_N$. Since x has only K nonzero components, the data vector $y = \Phi x$ is a linear combination of K columns from Φ . In the language of sparse approximation, we say that x has K -term representation over the dictionary Φ .

In order to recover K -sparse signal x , question is to determine which columns of Φ participate in the measurement vector. The idea is to collect columns in a greedy manner. At each iteration, a column of Φ that is most strongly correlated with the remaining part of y is selected. Then contribution due to this column is subtracted from y and iterated on the residual. After K iterations the algorithm will have identified the correct set of columns.

Once the correct set of columns, denoted I , are found, the estimate of sparse signal x can be found using

$$\begin{aligned} x &= \Phi_I^\dagger y \\ \Phi_I^\dagger &\triangleq (\Phi_I^* \Phi_I)^{-1} \Phi_I^* \end{aligned} \tag{2.18}$$

OMP is fast in theory and practice but its recovery guarantees are not as strong as that of Basis Pursuit.

Theorem 2.2.1. *Fix $\delta \in (0, 0.36)$ and let Φ be an $M \times N$ Gaussian measurement matrix with $M > CK \log(N/\delta)$. Let x be a K -sparse signal in R^N . Then with high probability exceeding $1 - 2\delta$, OMP correctly reconstructs the signal x from its measurements Φx .*

2.2.2 OMP Algorithm

INPUT:

- An $M \times N$ measurement matrix Φ
- An M dimensional data vector y
- The sparsity level K of the ideal signal

OUTPUT:

- An estimate \hat{x} in R^N for the ideal signal
- A set Λ_K containing the K elements from $\{1, 2, \dots, N\}$

- An M -dimensional approximation a_K of the data y
- A K -dimensional residual $r_K = y - a_K$ of the data y

PROCEDURE:

- (1) Initialize residual $r_0 = y$, the index set $\Lambda_0 = \emptyset$
- (2) Find the index λ_t that solves easy optimization problem
 $\lambda_t = \arg \max_{j=1,2,\dots,N} |\langle r_{t-1}, \varphi_j \rangle|$ if the maximum occurs for multiple indices, break the tie deterministically.
- (3) Augment the index set and the matrix of chosen atoms: $\Lambda_t = \Lambda_{t-1} \cup \{\lambda_t\}$
and $\Phi_t = [\Phi_{t-1} \ \varphi_{\lambda_t}]$
- (4) Solve a least squares problem to obtain a new signal estimate: $v_t = \arg \min_x \|y - \Phi_t v\|_2$
- (5) Calculate the new approximation of the data and the new residual

$$\begin{aligned} a_t &= \Phi_t v_t \\ r_t &= y - v_t \end{aligned}$$

- (6) Increment t , and return step 2 if $t < K$
- (7) The estimate \hat{x} for the ideal signal has nonzero indices at the components listed in Λ_K . The value of the estimate \hat{x} in component λ_j equals the j^{th} component of v_t

2.2.3 Compressive Sampling Matching Pursuit (COSAMP)

The difficulty of signal reconstruction comes from the identification of the locations of the largest components in the signal. Tropp and Needell developed a new OMP based algorithm namely COSAMP [15]. It uses an approach inspired by the Restricted Isometry Property. Suppose the measurement matrix Φ has Restricted Isometry Property $\delta_K \ll 1$. For a K -sparse signal x , measurements $u = \Phi^* \Phi x$ can be used as a proxy for the signal because the energy in each set of K components of u approximates the energy in the corresponding K components of x . The largest K entries of the proxy u points toward the largest K entries of the signal x . Because the samples have the form $y = \Phi x$, proxy can be obtained by multiplying the measurements by Φ^* . The algorithm uses this idea iteratively to approximate the original signal. The steps in the algorithm are

- (1) **Identification.** The algorithm forms a proxy of the residual from the current samples and locates the largest components of the proxy.
- (2) **Support Merger.** The set of newly identified components is united with the set of components that appear in the current approximation.
- (3) **Estimation.** The algorithm solves a least squares problem to approximate the target signal on the merged set of components.
- (4) **Pruning.** The algorithm produces a new approximation by retaining only the largest entries in the least-squares signal approximation.
- (5) **Sample Update.** Finally, the samples are updated so that they reflect the residual, the part of the signal that has not been approximated.

Theorem 2.2.2. *Suppose that Φ is an $M \times N$ measurement matrix satisfying the restricted isometry condition $\delta_{2K} < c$. Let $y = \Phi x + e$ be measurements contaminated with arbitrary noise e . For a given precision parameter η , the algorithm COSAMP produces a K -sparse approximation \hat{x} that satisfies*

$$\|x - \hat{x}\|_2 \leq C \max \left\{ \eta, \frac{1}{\sqrt{K}} \|x - x_{K/2}\|_2 + \|e\|_2 \right\} \quad (2.19)$$

where $x_{K/2}$ is a best- $(K/2)$ approximation to x .

2.2.4 COSAMP Algorithm

INPUT:

- Measurement Matrix Φ , noisy measurement vector y , sparsity level K

OUTPUT:

- A K -sparse approximation a of the target signal

PROCEDURE:

- (1) $a^0 \leftarrow 0$ //Trivial Initial approximation
- (2) $v \leftarrow y$ //Current samples =input samples
- (3) $k \leftarrow 0$
- (4) repeat
 - $k \leftarrow k + 1$
 - $u = \Phi^* \Phi v$ //Form signal proxy
 - $\Omega \leftarrow \text{supp}(u_{2s})$ //Identify large components

```

 $T \leftarrow \Omega \cup \text{supp}(a^{k-1})$  //Merge Supports
 $b|_T \leftarrow \Phi_T^\dagger y$  //Signal estimation by least squares
 $b|_{T^c} \leftarrow 0$ 
 $a^k \leftarrow b_s$  //Prune to obtain next approximation
 $v \leftarrow y - \Phi a^k$  //Update current samples
Until halting criterion true

```

2.3 Simulation Results

In this section above mentioned algorithms are simulated to compare the OMP, COSAMP and Basis Pursuit. 100 trials are performed for each set of sparsity level and measurement levels (K, M) and selected signal dimension $N = 256$. Probabilities of exact reconstructions are plotted as a function of measurement levels for different sparsity levels. Figure 3.1.1 through Figure 3.1.3 shows the performances of three algorithms as a number of measurement levels for different sparsity levels. Figure 3.1.4 shows the Performance of different algorithms for a fixed sparsity levels. It is clear that Basis Pursuit outperforms OMP and COSAMP in measurement levels. Basis Pursuit requires fewer measurements than OMP and COSAMP. Also OMP requires more measurement than COSAMP. If run-times of the algorithms are compared, COSAMP has the best run time compared with others, especially much shorter run time than Basis Pursuit.

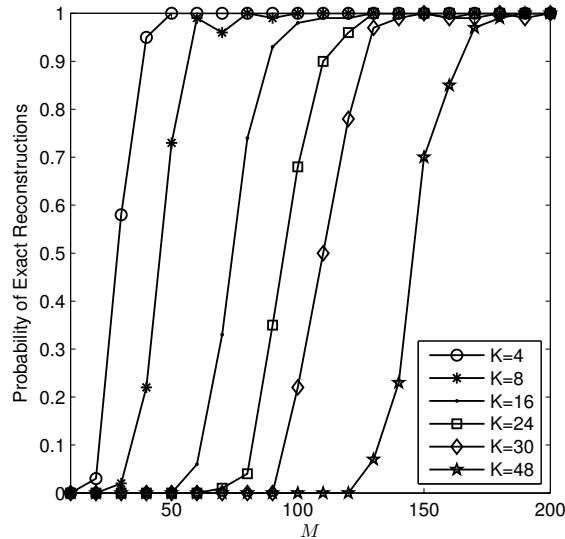


Figure 2.2: Exact reconstruction frequency versus measurement level using Basis Pursuit

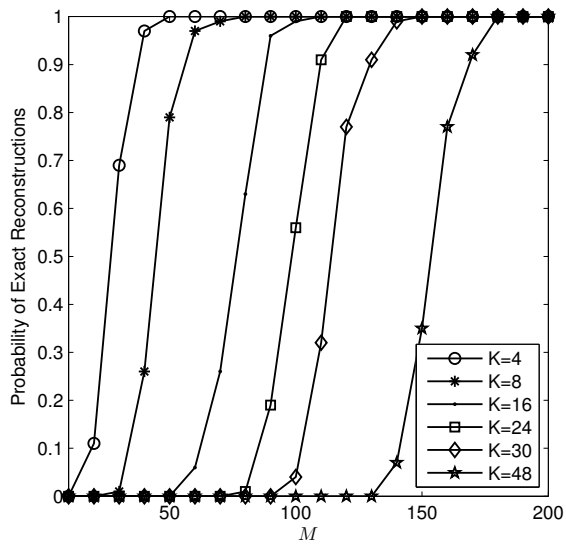


Figure 2.3: Exact reconstruction frequency versus measurement level using OMP

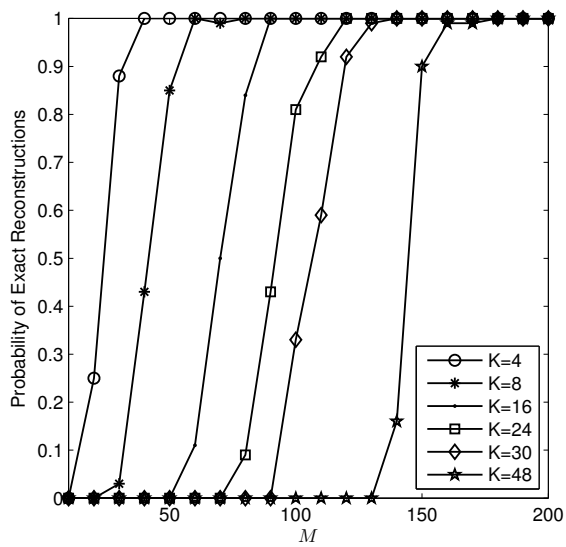


Figure 2.4: Exact reconstruction frequency versus measurement level using COSAMP

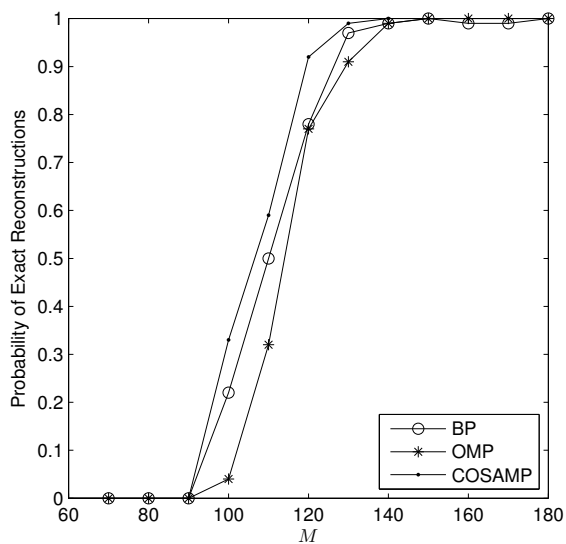


Figure 2.5: Exact reconstruction frequency versus measurement level for sparsity level $K=20$ using three algorithms

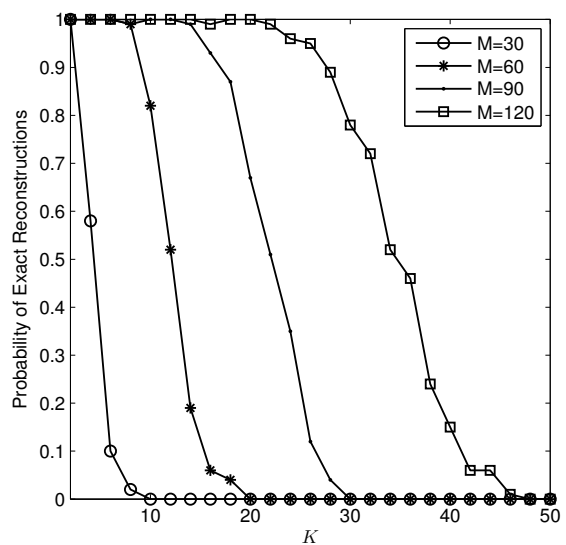


Figure 2.6: Exact reconstruction frequency versus sparsity using BP

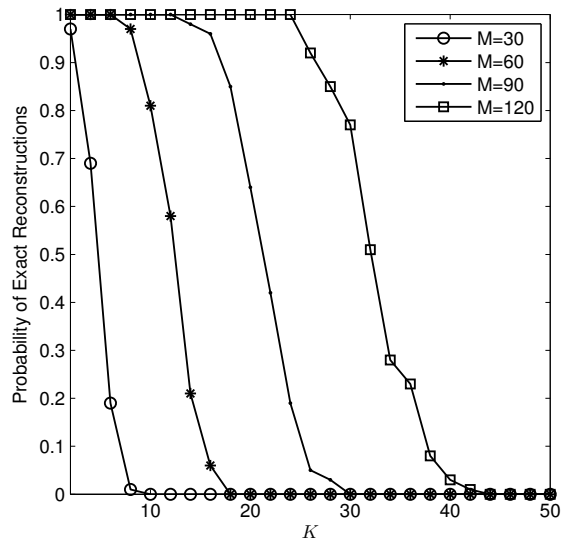


Figure 2.7: Exact reconstruction frequency versus sparsity using OMP

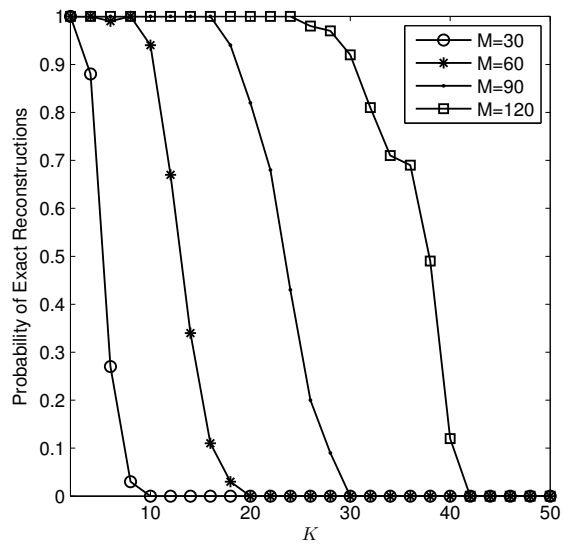


Figure 2.8: Exact reconstruction frequency versus sparsity using COSAMP

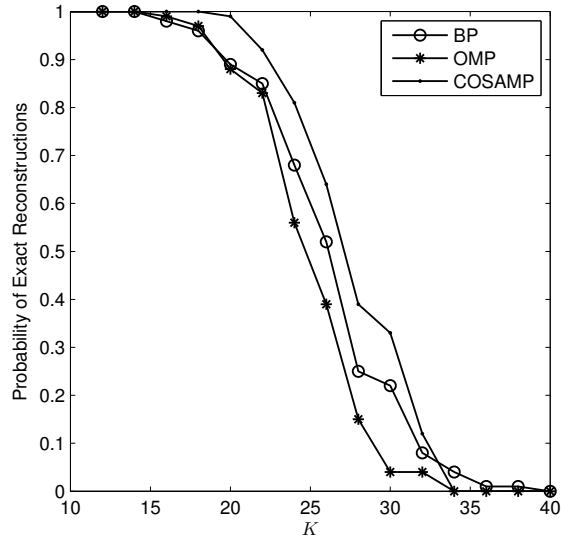


Figure 2.9: Exact reconstruction frequency versus sparsity for measurement level $M = 100$ using three algorithms

2.4 Applications

CS can be used to capture a signal using a small number of measurements compared to signal dimension. This leads CS to use in many applications of signal processing area [16].

- **Data compression.** In some situations, the sparsity basis Ψ is not be able to known at the encoder or impractical to implement for data compression. However a randomly designed Φ can be considered a universal encoding strategy. This property is useful in distributed sensor networks.
- **Channel Coding.** CS can be used to design error correcting codes to prevent errors during transmission.
- **Inverse problems.**
- **Data acquisition.** In ADC systems sometimes it may be difficult to obtain the full collection of samples of an analog signal. Therefore CS is powerful a sensing mechanism that allows to sample signal below Nyquist rate. A digital camera has millions of digital sensors, as soon as after the photo is taken; huge amount of information is coded (transform coding) in to a small size such as around Kilobytes levels. Duarte et. al [17] used the results of Compressed Sensing (CS) to build a simple, cheap, single pixel camera based on Digital Micromirror Device (DMD). His motivation on this study is just to show, how CS is powerful sensing mechanism

in signal processing area. Instead of calculating all the transformation coefficients as in JPEG or other transform coding approaches, the proposed architecture captures the image in already compressed form.

2.5 Summary

OMP works only for the case where Φ is Gaussian matrix, whereas BP works for a more general class of matrices. Also BP works correctly for all signals once the measurement matrix satisfies the restricted isometry property. The advantage of OMP however is that, it has faster runtime than Basis Pursuit. There are other versions of OMP based algorithms namely, Regularized Orthogonal Matching Pursuit (ROMP) [18], Stagewise Orthogonal Matching Pursuit (StOMP) [19]. COSAMP [15] combines both the advantages of optimization based approaches and greedy approaches.

CHAPTER 3

NONCONVEX COMPRESSED SENSING WITH PARTIALLY KNOWN SIGNAL SUPPORT

3.1 Introduction

Recovering compressively sampled signals with partially known support has been previously studied. Borries et al. [20] showed that it is possible to decrease the number of compressive measurements by the size of partially known support, if the signal has a sparse discrete fourier transform (DFT). Khajehnejad et al. [21] have also studied the partially known compressed sensing (CS) problem using a probabilistic approach of the known part of the signal. In a recent study of Vaswani [22], recovery conditions of noise free compressive sensing over the complement of the known set is studied and is called modified-CS (MOD-CS). Their results are weaker than the standard l_1 minimization conditions of [12] and they also extended the study to noisy CS problem with partially known support using a regularized MOD-CS approach [23]. Furthermore Jacques [24] improved the results in [22] to the case of compressible signals and noisy measurements by demonstrating that the solution is $l_2 - l_1$ instance optimal. More recently Friedlander et al. [25] studied weighted l_1 minimization approach when support of the signal is partially known. They showed that if at least 50% of the partial support information is available, then weighted l_1 minimization is stable and robust under weaker conditions than standard l_1 minimization.

Our work has shown that l_p minimization with partially known signal support for $p < 1$, denoted by $l_{p<1}$, exhibits similar stability and robustness compared to l_1 minimization with partially known signal support [24]. Recovery error is bounded by two terms; one is related to measurement noise and the other with *best k -term* approximation of residual defined as the difference between the original signal and the signal with known support.

3.2 Nonconvex Compressed Sensing

In the literature there exists convex and nonconvex penalty functions for recovering sparse vectors using

$$\min \frac{1}{2} \|y - \Phi x\|_2 + \sum_{i=1}^N g_\lambda(|x_i|) \quad (3.1)$$

which includes a nonsmooth and nonconvex penalty function $g_\lambda(\cdot)$. These penalty functions are summarized in Table 1 and in Fig. 3.1.

Table 3.1: Commonly used nonconvex penalty functions

Penalty	Formula
SCAD	$g_\lambda(x_i) = \begin{cases} \lambda x_i & x_i \leq \lambda \\ \frac{- x_i ^2 + 2a\lambda x_i - \lambda^2}{2(a-1)} & \lambda < x_i \leq a\lambda \\ \frac{(a+1)\lambda^2}{2} & x_i \geq a\lambda \end{cases}$
Zhang	$g_\lambda(x_i) = \begin{cases} \lambda x_i & \text{if } x_i < \eta \\ \lambda\eta & \text{otherwise} \end{cases}$
l_p	$g_\lambda(x_i) = \lambda x_i ^p, \quad 0 < p < 1$
Log	$g_\lambda(x_i) = \lambda \log(x_i + \varepsilon) - \lambda \log(\varepsilon)$

Nonconvex minimization methods have been studied in literature previously. Nikolova et al. studied image restoration and reconstruction problem using nonconvex regularized least squares for different penalty functions including l_p penalty with $p < 1$ [26–28], actually not related with CS approach, they try to restore images with neat edges. In [29] sparse recovery problem is solved using nonconvex penalty functions like $l_{p<1}$ or Smoothly Clipped Absolute Deviation (SCAD) penalty which is proposed by Fan and Li [30] instead of l_1 norm. In [31] Zhang penalty is proposed in solution of (3.1). This penalty is composed of two stages, first stage corresponds to original Lasso and second stage is a modified Lasso problem where large parameters are not penalized anymore.

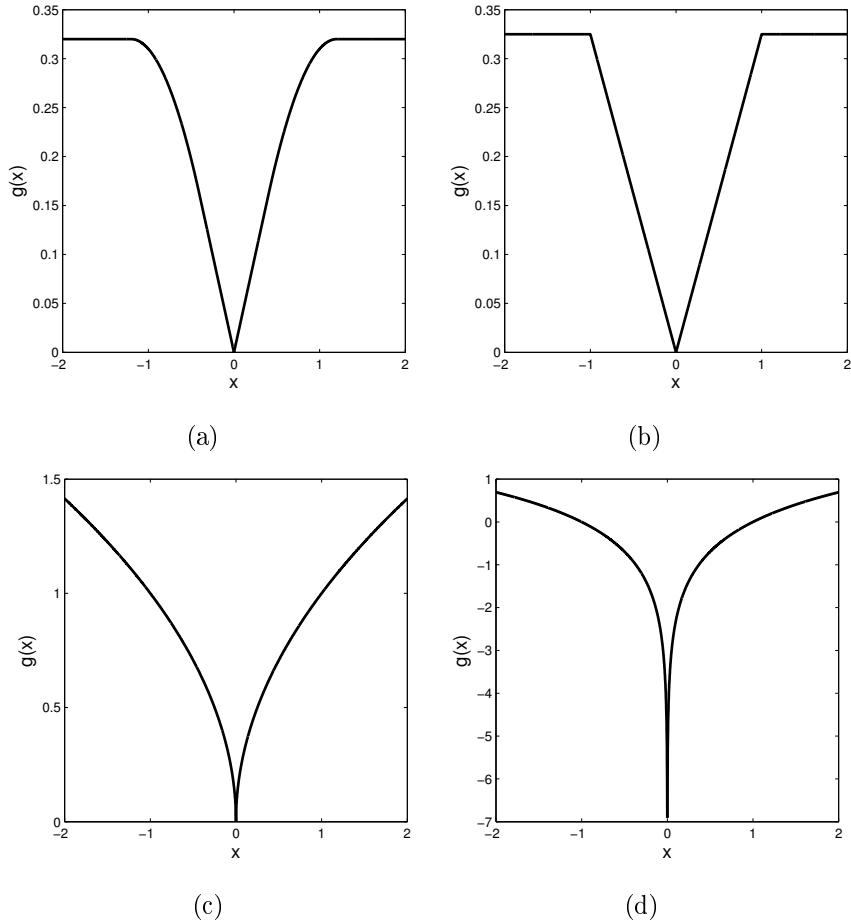


Figure 3.1: Penalty functions in Table I (a) SCAD (b) Zhang (c) $l_{p<1}$ (d) Log

Rao and Kreutz-Delgado replaced the objective l_1 norm with $l_{p<1}$ norm and solved $l_{p<1}$ minimization using Iteratively Reweighted Least Squares (IRLS) [32]. Chartrand applied a regularization strategy in solution of IRLS and has shown that $l_{p<1}$ minimization reconstructs sparse signal exactly with fewer measurements compared to BP [33]. Saab et al. [34, 35] studied the stability of $l_{p<1}$ minimization for the sparse and compressible signals when measurements contain some additive noise and gave the error bounds on the reconstruction error.

Theorem 3.2.1. [7] Let Φ be an $M \times N$ matrix. Let $x \in \mathbb{C}^N$ and let $K = \|x\|_0$ be the size of the support of x . Let $p \in [0, 1]$, $b > 1$, and $a = b^{p/(2-p)}$. Suppose that Φ satisfies

$$\delta_{aK} + b\delta_{(a+1)K} < b - 1 \quad (3.2)$$

then unique minimizer to

$$\min \|\hat{x}\|_p \quad \text{subject to} \quad y = \Phi \hat{x} \quad (3.3)$$

is exactly x .

Interpretation of Theorem 3.2.1 is important when $p = 0.5$ and $a = 3$, $\delta_{3K} + 27\delta_{4K} < 26$ is required for perfect reconstruction. In [7] solution of (3.3) is computed using a regularized IRLS approach. Although it is not guaranteed to obtain a global minimum, it has been shown that computed local minimizers are actually global [7, 36].

In an other publication Saab et.al [34] provided theoretical and numerical results to recover sparse and compressible signals from incomplete and noisy measurements. Results of Candes et. al [12] are extended to the $p < 1$ case. Results indicate that $l_{p < 1}$ minimization provides better theoretical guarantees in terms of stability and robustness compared to l_1 minimization.

Theorem 3.2.2. [34] *Assume that x is an arbitrary signal and if*

$$\delta_{aK} + a^{2/p-1}\delta_{(a+1)K} < a^{2/p-1} - 1 \quad (3.4)$$

holds for some $a > 1$ and $ak \in \mathbb{Z}^+$. Then solution \hat{x} to

$$\min \|\hat{x}\|_p^p \quad \text{subject to} \quad \|y - \Phi\hat{x}\|_2 \leq \epsilon \quad (3.5)$$

for $p < 1$ obeys

$$\|\hat{x} - x\|_2^p < C_{K,a,p}^{(1)}\epsilon^p + C_{K,a,p}^{(2)} \frac{\|x - x_K\|_p^p}{K^{1-p/2}} \quad (3.6)$$

for some constants $C_{K,a,p}^{(1)}$ and $C_{K,a,p}^{(2)}$

Theorem 3.2.2 states that, one can recover sparse and compressible signal stably and robustly.

Chartrand and Staneva studied to obtain the number of Gaussian measurements needed to reconstruct the sparse signal.

Theorem 3.2.3. [36] *Let Φ be an $M \times N$ matrix whose elements are iid random variables distributed normally with mean zero and variance σ_2 , where $M < N$. Then there are constants $C_1(p)$ and $C_2(p)$ such that whenever $0 < p \leq 1$ and*

$$M \geq C_1(p)K + pC_2(p)K \log(N/K), \quad (3.7)$$

the following is true with probability exceeding $1 - 1/\binom{N}{K}$ for any $x \in \mathbb{R}^N$ with sparsity $\|x\|_0 = K$, x is unique solution to

$$\min \|\hat{x}\|_p^p \quad \text{subject to} \quad y = \Phi\hat{x} \quad (3.8)$$

Numerical results in [36] are shown to obtain much lower measurement numbers although Theorem 3.2.3 requires more.

3.3 CS With Partially Known Signal Support

Vaswani et. al. [22] studied the problem of reconstructing a sparse signal from a limited number of measurements when some part of the support of the signal is known a priori for noiseless measurements and the problem is called Modified CS (MOD-CS). Numerical and theoretical studies have shown that CS with partially known support can reduce the required number of measurements for exact recovery. They obtained sufficient conditions for exact reconstruction for the MOD-CS. This technique is then used to reconstruct a series of dynamic MR images frame by frame. Their results are much weaker than those needed for CS. BP problem is solved over the complement of the partially known set T given by

$$\min \|\hat{x}_{T^c}\|_1 \quad \text{subject to} \quad y = \Phi x \quad (3.9)$$

The results in [22] are then modified by the Jacques [24] for noisy measurements and reconstruction error bound is given.

Theorem 3.3.1. [24] *Suppose that the partially known signal support of signal x is given as T and size of $|T| = s$. Let us assume that the matrix Φ satisfies the RIP with $\delta_{2k}^2 + 2\delta_{s+2k} < 1$ then solution to program*

$$\min \|\hat{x}_{T^c}\|_1 \quad \text{subject to} \quad \|y - \Phi \hat{x}\|_2 \leq \epsilon \quad (3.10)$$

obeys

$$\|x - \hat{x}\|_2 \leq C_{s,k}\epsilon + D_{s,k}e_0(r; k) \quad (3.11)$$

where r is residual $r = x - x_T$, and $e_0(r; k) = k^{-1/2}\|r - r_k\|_1$ is the compressibility error at k -term of r . When $k \ll s$, if $\delta_{2k} = 0.02$ and $\delta_{s+2k} = 0.2$, then $C_{s,k} < 7.32$ and $D_{s,k} < 3.35$.

Some observations can be given for Theorem 3.3.1, if there is no knowledge about the signal support, i.e., $T = \emptyset$ and $s = 0$, we find the previous conditions of [10], $\delta_{2k} < \sqrt{2} - 1$. Also the condition $\delta_{2k}^2 + 2\delta_{s+2k} < 1$ is satisfied if $\delta_{s+2k} < \sqrt{2} - 1$ since $\delta_{2k} < \delta_{s+2k}$. If the signal is exactly sparse and no noise in the measurements, perfect reconstruction is guaranteed as in [22].

In another study about partially known CS problem is the weighted l_1 minimization problem [25]. Friedlander et.al studied weighted l_1 minimization problem for signal reconstruction when partial support information is available. If at least 50% of the partial support information is available, then weighted l_1 minimization is stable and robust under weaker conditions than the standard l_1 minimization.

Theorem 3.3.2. [25] Let x be in \mathbb{R}^N and let x_K be its best k -term approximation, supported on T_0 . Let partially known part of the signal is denoted by $|T| = \rho K$ and $|T \cap T_0| = \alpha \rho K$. Suppose that ak is a positive integer, with $a \geq (1 - \alpha)\rho$, $a > 1$, and if

$$\delta_{aK} + \frac{a}{(\omega + (1 - \omega)\sqrt{(1 + \rho - 2\alpha\rho)})} \delta_{(a+1)K} < \frac{a}{(\omega + (1 - \omega)\sqrt{(1 + \rho - 2\alpha\rho)})} - 1 \quad (3.12)$$

for some given $0 \leq \omega \leq 1$. Then solution \hat{x} to

$$\min \|\hat{x}\|_{1,\omega} \quad \text{subject to} \quad \|y - \Phi \hat{x}\|_2 \leq \epsilon \quad (3.13)$$

obeys

$$\|x - \hat{x}\|_2 \leq C'_0 \epsilon + C'_1 K^{-1/2} (\omega \|x - x_K\|_1 + (1 - \omega) \|x_{T^c \cap T_0^c}\|_1) \quad (3.14)$$

$\|\hat{x}\|_{1,\omega} := \sum_i \omega_i |\hat{x}_i|$ is the weighted l_1 norm. The main idea is to choose ω such that the entries of x expected to be large are penalized less in this optimization problem.

3.4 $l_{p<1}$ Minimization With Partially Known Support

In this section we study sparse recovery problem from compressive measurements using nonconvex optimization when some part of the support of the signal is known a priori. We present theoretical results and simulations for $l_{p<1}$ minimization with partial support information. The support of the signal may exist from prior knowledge in various applications [17, 22, 37]. For instance, support estimation of the previous time instant may be used to reconstruct time sequences of sparse signals iteratively. This problem arises in applications such as real-time dynamic MRI reconstruction, real time single pixel camera imaging or video compression. It is possible to have many non-zero wavelet or DCT coefficients which carry most of the energy of the signal [38]. In these cases it is useful to incorporate prior information for reconstructing signals from compressive measurements.

Theoretical results are presented by showing that it is possible to recover sparse and compressible signals using $l_{p<1}$ minimization when partial support information is available. The stability of recovering x from y is also related with the RIP of the measurement matrix. The following two theorems summarize our results.

Theorem 3.4.1. [39] Suppose that complement of the partially known set of the signal x is denoted as T^c and the size of partially known set T is $|T| = s$. If

$$\delta_{s+(a+1)k} + a^{1/2-1/p} (\delta_{s+(a+1)k}^2 + \delta_{2ak}^2)^{1/2} < 1 \quad (3.15)$$

is satisfied for some constant $a > 1$ and ak is a positive integer then the solution \hat{x} to

$$\min \|\hat{x}_{T^c}\|_p^p \quad \text{subject to} \quad \|y - \Phi \hat{x}\|_2 \leq \epsilon \quad (3.16)$$

for $p < 1$ obeys

$$\|\hat{x} - x\|_2^p < C_{a,k,s,p} \epsilon^p + D_{a,k,s,p} \frac{\|r - r_k\|_p^p}{k^{1-p/2}} \quad (3.17)$$

where r is defined as the residual $r = x - x_T$ and r_k is the best k -term approximation to r . The constants $C_{a,k,s,p}$ and $D_{a,k,s,p}$ are found as

$$C_{a,k,s,p} = 2^p \frac{(1 + a^{p/2-1})(1 + \delta_{s+(a+1)k})^{p/2}}{(1 - \delta_{s+(a+1)k})^p - a^{p/2-1}(\delta_{s+(a+1)k}^2 + \delta_{2ak}^2)^{p/2}} \quad (3.18)$$

$$D_{a,k,s,p} = 2a^{p/2-1} \left[1 + \frac{(1 + a^{p/2-1})(\delta_{s+(a+1)k}^2 + \delta_{2ak}^2)^{p/2}}{(1 - \delta_{s+(a+1)k})^p - a^{p/2-1}(\delta_{s+(a+1)k}^2 + \delta_{2ak}^2)^{p/2}} \right] \quad (3.19)$$

Recovery error (to the p^{th} power) is bounded by the sum of two terms, first one is proportional to the measurement error and the second is related to best k -term approximation error of the residual.

Proof. Let x be an arbitrary signal and $0 < p < 1$. T is a known set and T_0 be the locations of the k -largest coefficients of the residual with $T_0 \cap T = \emptyset$. First, let us write $\hat{x} = x + h$ with $h \in R^N$. Our aim is to bound $\|h\|_2$ given that $\|\Phi h\|_2 \leq 2\epsilon$. Let $\bar{T}_0 = T \cup T_0$ and $\bar{T}_{01} = T \cup T_0 \cup T_1$. We begin by dividing h into a sum of sparse vectors with disjoint sets. Then divide \bar{T}_0^c into sets T_1, T_2, \dots where $|T_j| = L$ for $j \geq 1$. T_1 is the location of L largest coefficients of $h_{\bar{T}_0^c}$; T_2 is the locations of the second L largest coefficients of $h_{\bar{T}_0^c}$ and so on. Since $T^c = T_0 \cup \bar{T}_0^c$ and $\|\cdot\|_p^p$ satisfies the triangle inequality

$$\begin{aligned} \|x_{T^c}\|_p^p &\geq \|x_{T^c} + h_{T^c}\|_p^p = \|x_{T_0} + h_{T_0}\|_p^p + \|x_{\bar{T}_0^c} + h_{\bar{T}_0^c}\|_p^p \\ &\geq \|x_{T_0}\|_p^p - \|h_{T_0}\|_p^p - \|x_{\bar{T}_0^c}\|_p^p + \|h_{\bar{T}_0^c}\|_p^p \end{aligned}$$

then

$$\begin{aligned} \|h_{\bar{T}_0^c}\|_p^p &\leq \|x_{T^c}\|_p^p + \|x_{\bar{T}_0^c}\|_p^p + \|h_{T_0}\|_p^p - \|x_{T_0}\|_p^p \\ &= 2\|x_{\bar{T}_0^c}\|_p^p + \|h_{T_0}\|_p^p \\ &= 2\|r - r_{T_0}\|_p^p + \|h_{T_0}\|_p^p \end{aligned}$$

For each $u \in T_j$ and $v \in T_{j-1}$ $|h(u)| \leq |h(v)|$ so

$$|h(u)|^p \leq \|h_{T_{j-1}}\|_p^p / L$$

Then

$$\begin{aligned} |h(u)|^2 &= \|h_{T_{j-1}}\|_p^2 / L^{2/p} \\ \|h_{T_j}\|_2^2 &\leq L^{1-2/p} \|h_{T_{j-1}}\|_p^2 \\ \|h_{T_j}\|_2^p &\leq L^{p/2-1} \|h_{T_{j-1}}\|_p^p \end{aligned}$$

and thus

$$\begin{aligned} \|h_{\bar{T}_0^c}\|_2^p &\leq \sum_{j \geq 2} \|h_{T_j}\|_2^p \leq L^{p/2-1} \sum_{j \geq 1} \|h_{T_j}\|_p^p = L^{p/2-1} \|h_{\bar{T}_0^c}\|_p^p \\ &= L^{p/2-1} (2\|r - r_{T_0}\|_p^p + \|h_{T_0}\|_p^p) \end{aligned} \quad (3.20)$$

Using Hölder's inequality for $\|h_{T_0}\|_p^p$ to convert back from l_p to l_2

$$\|h_{T_0}\|_p^p \leq (|T_0|)^{1-p/2} \|h_{\bar{T}_0}\|_2^p \quad (3.21)$$

is obtained and (3.20) becomes

$$\|h_{\bar{T}_0^c}\|_2^p \leq \sum_{j \geq 2} \|h_{T_j}\|_2^p \leq L^{p/2-1} \left((|T_0|)^{1-p/2} \|h_{\bar{T}_0}\|_2^p + 2\|r - r_{T_0}\|_p^p \right) \quad (3.22)$$

Now we need to control the size of $\|h_{\bar{T}_0}\|_2^p$. Observe that $\Phi h_{\bar{T}_0} = \Phi h - \sum_{j \geq 2} \Phi h_{T_j}$ therefore

$$\begin{aligned} \|\Phi h_{\bar{T}_0}\|_2^2 &= \langle \Phi h_{\bar{T}_0}, \Phi h_{\bar{T}_0} \rangle = \langle \Phi h_{\bar{T}_0}, \Phi h \rangle - \langle \Phi h_{\bar{T}_0}, \sum_{j \geq 2} \Phi h_{T_j} \rangle \\ &\leq \|\Phi h_{\bar{T}_0}\|_2 \|\Phi h\|_2 + \sum_{j \geq 2} |\langle \Phi h_{\bar{T}_0}, \Phi h_{T_j} \rangle| \end{aligned}$$

Taking the p^{th} power and using the fact that for any $b, c \geq 0$, and $0 < p < 1$, $b^p + c^p > (b + c)^p$ we have

$$\|\Phi h_{\bar{T}_0}\|_2^{2p} \leq \|\Phi h_{\bar{T}_0}\|_2^p \|\Phi h\|_2^p + \sum_{j \geq 2} (|\langle \Phi h_{\bar{T}_0}, \Phi h_{T_j} \rangle|)^p \quad (3.23)$$

Observing that $(1 - \delta_{L+|T_0|+|T|})^{p/2} \|h_{\bar{T}_0}\|_2^p \leq \|\Phi h_{\bar{T}_0}\|_2^p$ and $\|\Phi h_{\bar{T}_0}\|_2^p \leq (2\epsilon)^p$, (3.23) becomes

$$\begin{aligned} (1 - \delta_{L+|T_0|+|T|})^p (\|h_{\bar{T}_0}\|_2^p)^2 &\leq (2\epsilon)^p (1 + \delta_{L+|T_0|+|T|})^{p/2} \|h_{\bar{T}_0}\|_2^p \\ &\quad + (\delta_{L+|T_0|+|T|} \|h_{\bar{T}_0}\|_2 + \delta_{2L} \|h_{\bar{T}_1}\|_2)^p \times \sum_{j \geq 2} \|h_{T_j}\|_2^p \\ &\leq (2\epsilon)^p (1 + \delta_{L+|T_0|+|T|})^{p/2} \|h_{\bar{T}_0}\|_2^p \\ &\quad + (\delta_{L+|T_0|+|T|}^2 + \delta_{2L}^2)^{p/2} \|h_{\bar{T}_0}\|_2^p \times \sum_{j \geq 2} \|h_{T_j}\|_2^p \end{aligned}$$

Using (3.21) and (3.22) we have

$$(1 - \delta_{L+|T_0|+|T|})^p \|h_{\bar{T}_{01}}\|_2^p \leq (2\epsilon)^p (1 + \delta_{L+|T_0|+|T|})^{p/2} + (\delta_{L+|T_0|+|T|}^2 + \delta_{2L}^2)^{p/2} L^{p/2-1} \times \left(2\|r - r_{T_0}\|_p^p + (|T_0|)^{1-p/2} \|h_{\bar{T}_{01}}\|_2^p \right) \quad (3.24)$$

Then setting $L = ak$, $|T_0| = k$ and $|T| = s$ in (3.24) we obtain

$$\|h_{\bar{T}_{01}}\|_2^p \leq \alpha \epsilon^p + \beta \frac{\|r - r_k\|_p^p}{k^{1-p/2}} \quad (3.25)$$

with constants

$$\alpha = \frac{2^p (1 + \delta_{s+(a+1)k})^{p/2}}{(1 - \delta_{s+(a+1)k})^p - a^{p/2-1} (\delta_{s+(a+1)k}^2 + \delta_{2ak}^2)^{p/2}}$$

$$\beta = \frac{2(\delta_{s+(a+1)k}^2 + \delta_{2ak}^2)^{p/2} a^{p/2-1}}{(1 - \delta_{s+(a+1)k})^p - a^{p/2-1} (\delta_{s+(a+1)k}^2 + \delta_{2ak}^2)^{p/2}}$$

Substituting (3.22) and (3.25) into $\|h\|_2^p \leq \|h_{\bar{T}_{01}^c}\|_2^p + \|h_{\bar{T}_{01}}\|_2^p$ results in

$$\|h\|_2^p \leq C_{a,k,s,p} \epsilon^p + D_{a,k,s,p} \frac{\|r - r_k\|_p^p}{k^{1-p/2}}$$

where the constants $C_{a,k,s,p}$ and $D_{a,k,s,p}$ are as in (3.18) and (3.19) in Theorem 3.4.1. Denominator of these constants must be positive implying the condition given in (3.15). \square

Theorem 3.4.2. [39] *Let x be a strictly sparse signal and suppose that support of the signal is partially known. If (3.15) is satisfied for some constant $a > 1$ and ak is a positive integer, then the solution \hat{x} to (3.16) for $p < 1$ obeys*

$$\|\hat{x} - x\|_2 < (C_{a,k,s,p})^{1/p} \epsilon$$

where $C_{a,k,s,p}$ is given in (3.18)

Proof. The proof can easily be deduced from the proof of Theorem 3.4.1 by setting $r = r_k$. Since we assume x is sparse, the term related with $D_{a,k,s,p}$ in (3.17) goes to zero and $C_{a,k,s,p}$ in Theorem 3.4.2 is obtained. \square

Remark 3.4.1. *In Theorems 3.4.1 and 3.4.2 necessary conditions are provided for recovering sparse and compressible signals with partially known support. The constants $C_{a,k,s,p}$ and $D_{a,k,s,p}$ determines the upper bounds on the recovery error.*

Remark 3.4.2. *If $p = 1$ and $a = 1$ we obtain precisely the same constants $C_{s,k} = 4\sqrt{1 + \delta_{s+2k}} / (1 - \delta_{s+2k} - \mu_{s,k})$ and $D_{s,k} = 2(1 + \mu_{s,k} - \delta_{s+2k}) / (1 - \delta_{s+2k} - \mu_{s,k})$ and reconstruction conditions $1 - \delta_{s+2k} > \mu_{s,k}$ with $\mu_{s,k} = \sqrt{\delta_{s+2k}^2 + \delta_{2k}^2}$ given in [24].*

Remark 3.4.3. If there is no partial support information with $a = 1$ and $p = 1$, i.e., $T = \emptyset$ and $s = 0$ then $r = x$, we find the reconstruction conditions and constants given in [10]

Remark 3.4.4. If $p = 1$ and $a = 1$ when $k \ll s$, the constants $C_{a,k,s,p} < 7.32$ and $D_{a,k,s,p} < 3.35$ for $\delta_{s+2k} = 0.2$ and $\delta_{2k} = 0.02$, which is the same constants given in [24].

Remark 3.4.5. In Theorem 3.4.1, the compressibility of the signal is determined by the compressibility error $r - r_k$.

The constants $C_{a,k,s,p}$ and $D_{a,k,s,p}$ depend on s which is related to partially known support level, on k which reflects the degree of compressibility of the residual r , on p determined by the recovery algorithm and on a which is a free parameter that (3.15) holds. In Fig. 3.2 (a) and (b), $C_{a,k,s,p}$ and $D_{a,k,s,p}$ are plotted versus p for different values of a . $\delta_{s+(a+1)k} = 0.2$ and $\delta_{2ak} = 0.1$ are fixed for $k \ll s$. As shown in Fig. 3.2 (a), error bound noise constant ($C_{a,k,s,p}$) is lower for smaller p ($p < 1$) and for bigger a and in Fig. 3.2 (b), error bound compressibility constant ($D_{a,k,s,p}$) is lower for bigger a .

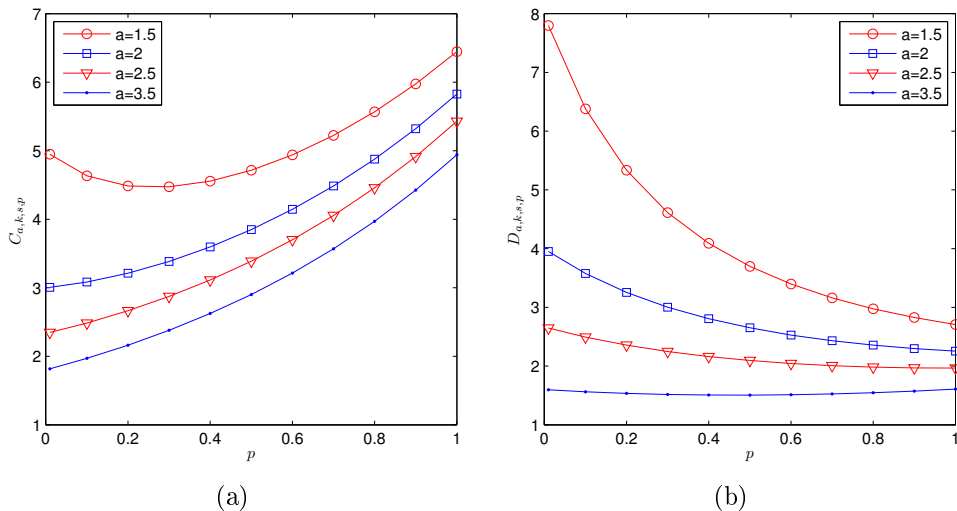


Figure 3.2: (a) $C_{a,k,s,p}$ versus p , (b) $D_{a,k,s,p}$ versus p , for different values of a

3.5 Experimental Results

In this section we present the results of several numerical experiments to explore the performance of $l_{p<1}$ minimization with partially known support for sparse and compressible signals. The problem in (3.16) is nonconvex and it

is possible to have several local minima on the feasible set. Therefore IRLS approach (3.26) is used in the solution of (3.16).

$$\min \sum_{i=1}^N w_i \hat{x}_i^2 \quad \text{subject to} \quad \Phi \hat{x} = y \quad (3.26)$$

The closed form of the solution is given as, giving the next iterate $\hat{x}^{(n+1)}$:

$$\hat{x}^{(n+1)} = Q_n \Phi^T (\Phi Q_n \Phi^T)^{-1} y$$

where Q_n is diagonal matrix with elements $1/w_i$. Choose the weights $w_i = ((\hat{x}_i^{(n)})^2 + \gamma)^{p/2-1}$ where γ is initially a large constant added to avoid division by zero whenever $\hat{x}_i^{(n)} = 0$ since $p - 2$ is negative. The value of γ is decreased according to rule $\gamma_{n+1} = 0.99\gamma_n$ and iteration is continued until γ becomes very small. In order to include partially known support in the solution, the elements on the diagonal of Q_n whose positions are in T need to have larger weight than the others. This is satisfied by multiplying these elements with 1000 times the largest element on the diagonal.

In order to show that the recovery error in the simulations does below the given theoretical error bound given in (3.17), $N = 1024$ dimensional compressible 32×32 cardiac image [22] and zero mean Gaussian matrix are used. The simulations are performed by selecting $s = 60$, $s = 90$ and $s = 120$ best term approximations. Measurement level M is set to 256 and this process is repeated 50 times. In Fig. 3.3 (a) $\|\hat{x} - x\|_2^p$ is plotted versus p for $s = 60$, $s = 90$ and $s = 120$. The theoretical error bound is plotted in Fig. 3.3 (b)-(d) for different values of a and k providing that the condition in (3.15) holds. Also k is assumed to be less than s and hence RIC of the measurement is fixed as $\delta_{s+(a+1)k} = 0.2$ and $\delta_{2ak} = 0.1$. It can be observed that $\|\hat{x} - x\|_2^p$ is lower than the theoretical error bounds for $s = 60$, $s = 90$ and $s = 120$.

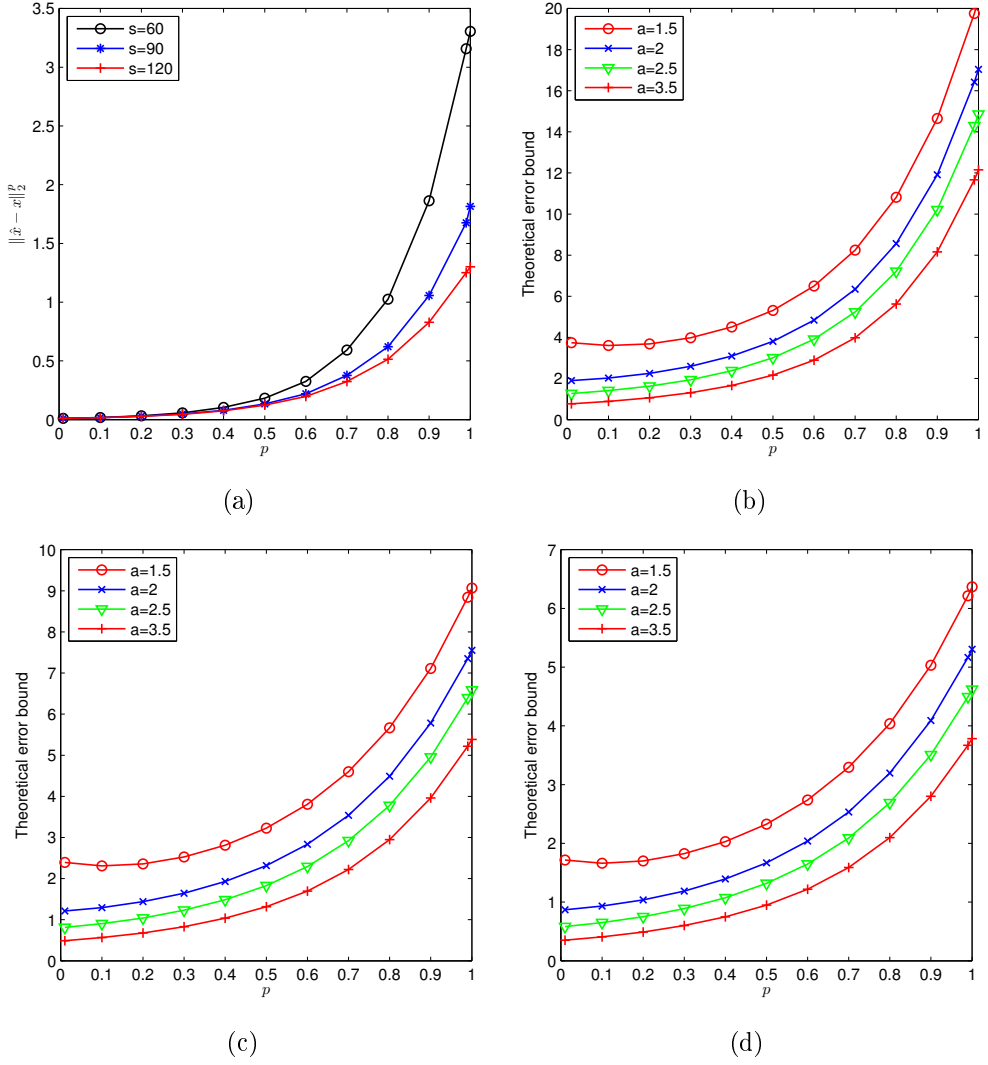


Figure 3.3: (a) Reconstruction error to the p^{th} power versus p for various s for $M = 256$, (b) Theoretical error bound for $s = 60, k = 20$, (c) Theoretical error bound for $s = 90, k = 30$, (d) Theoretical error bound for $s = 120, k = 40$

3.5.1 Exact Recovery Results

100 trials are performed and correct reconstructions out of 100 trials are recorded and it is defined as the exact reconstruction frequency. If $\|x - \hat{x}\|_2 \leq 10^{-3}$, we accept the recovery of x is exact. In Fig. 3.4 we compare the performance of $l_{p < 1}$ minimization with partially known support for different values of p with modified compressed sensing approach (MOD-CS) [22] for different number of measurement levels. Fig. 3.4 (a)-(c) are plotted for $s = 0, s = 12$ and (c) $s = 20$ known support levels, respectively. It is observed that exact reconstruction frequency is more accurate for $p < 1$ for the same measurement levels.

In Fig. 3.5 (a)-(c) we analyze the effect of partially known support with

measurement levels $M = 80$, $M = 100$ and $M = 120$, respectively. We plotted exact reconstruction frequency increasing the known support in steps of $s = 4$. $p < 1$ gives better estimates compared to MOD-CS for all known support levels.

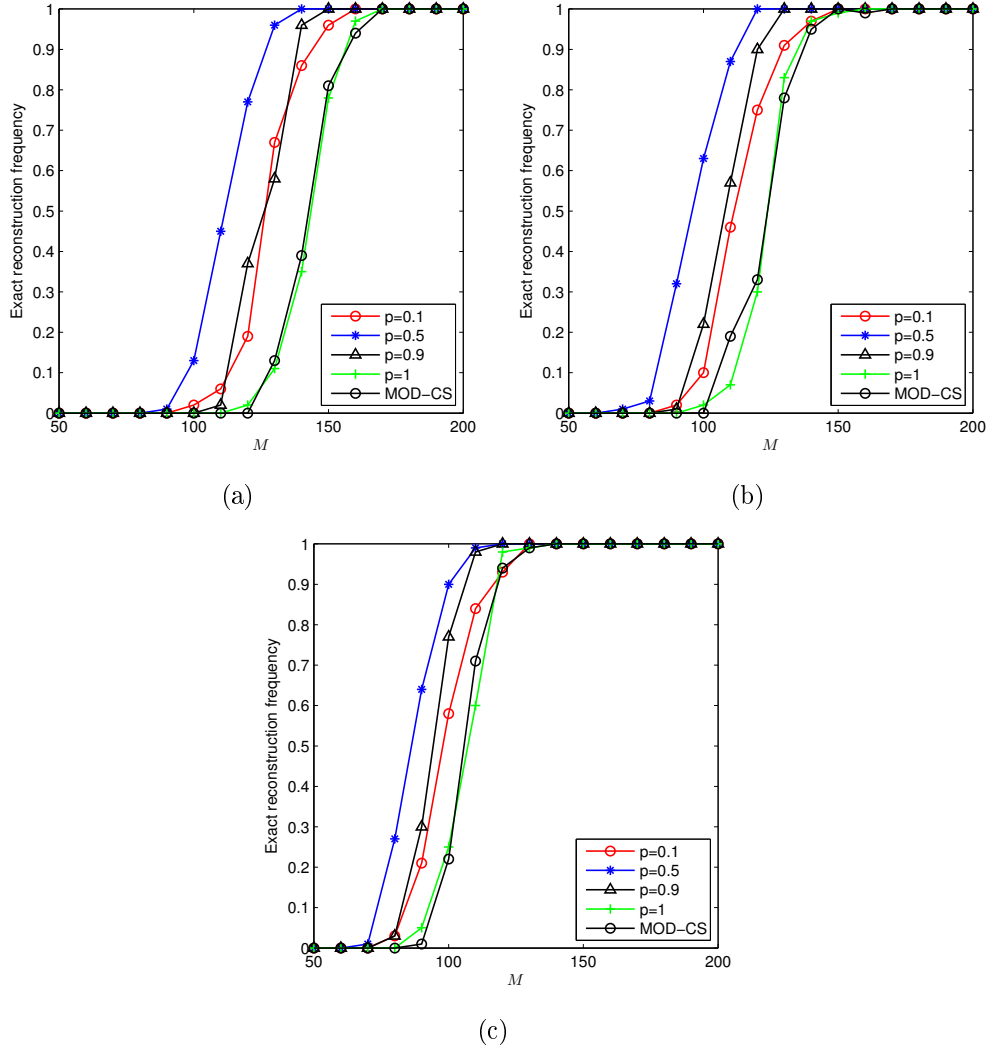


Figure 3.4: Exact reconstruction frequency of modified $l_{p < 1}$ minimization and MOD-CS for (a) $s = 0$, (b) $s = 12$, (c) $s = 20$.

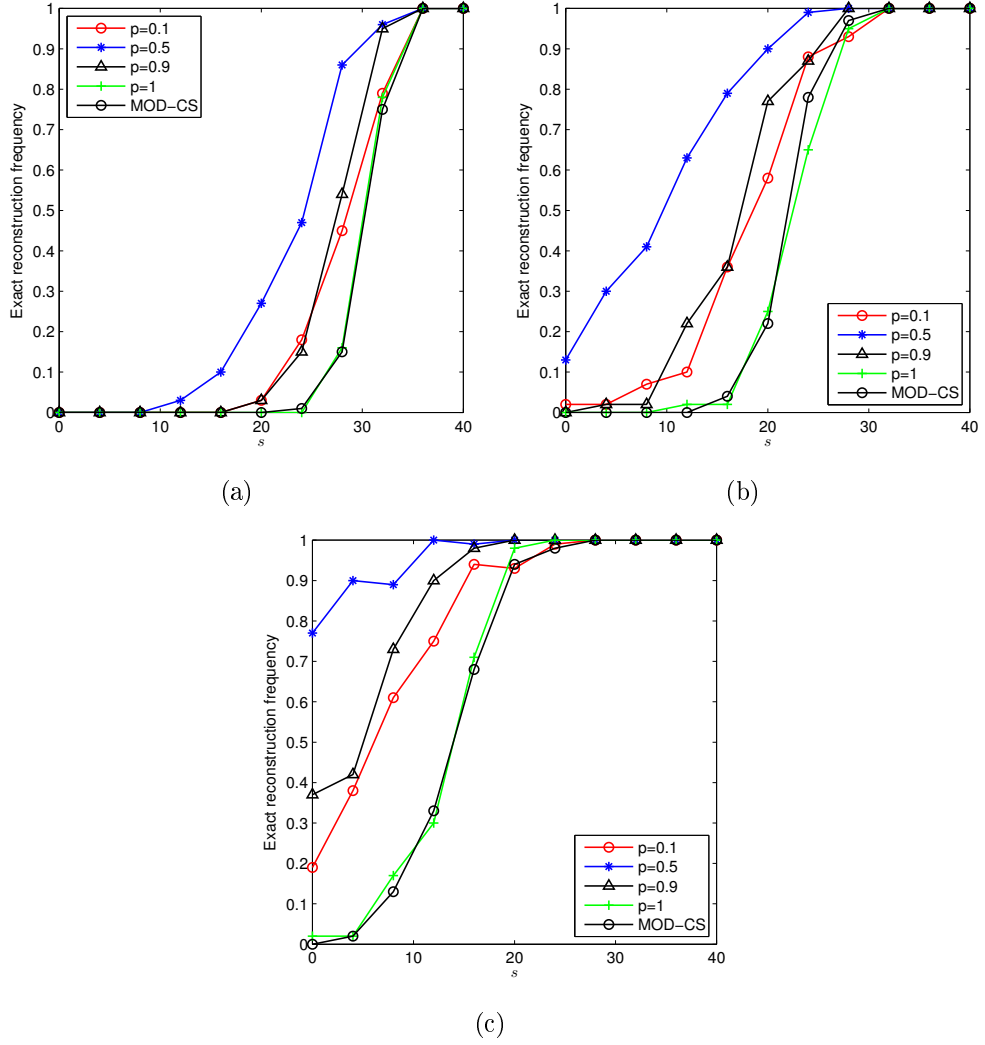


Figure 3.5: Exact reconstruction frequency of modified $l_{p<1}$ minimization and MOD-CS for (a) $M = 80$, (b) $M = 100$ and (c) $M = 120$.

3.5.2 The Sparse Case

We choose signal dimension as $N = 512$ and generate N -dimensional vector x with 40 nonzero entries. The location of the non-zero entries are selected randomly using standard Gaussian distribution. We generate $M \times N$ measurement matrix Φ with i.i.d. entries. We let $y = \Phi x$ and run IRLS algorithm to find the local minima of (3.16). Each experiment is repeated 50 times.

We compare the performance of $l_{p<1}$ minimization with partially known support for different values of p with MOD-CS [22] and average reconstruction signal to noise ratio (SNR) is calculated as

$$\text{SNR} = 20 \log_{10} \left(\frac{\|x\|_2}{\|x - \hat{x}\|_2} \right)$$

In Fig. 3.6 (a)-(c) SNR is plotted versus measurement level M for $s = 4$, $s = 16$ and $s = 24$, respectively. It is evident that increased known support causes

better reconstruction. Also it is observed that reconstruction SNR is higher for $p < 1$ compared to MOD-CS for same measurement levels.

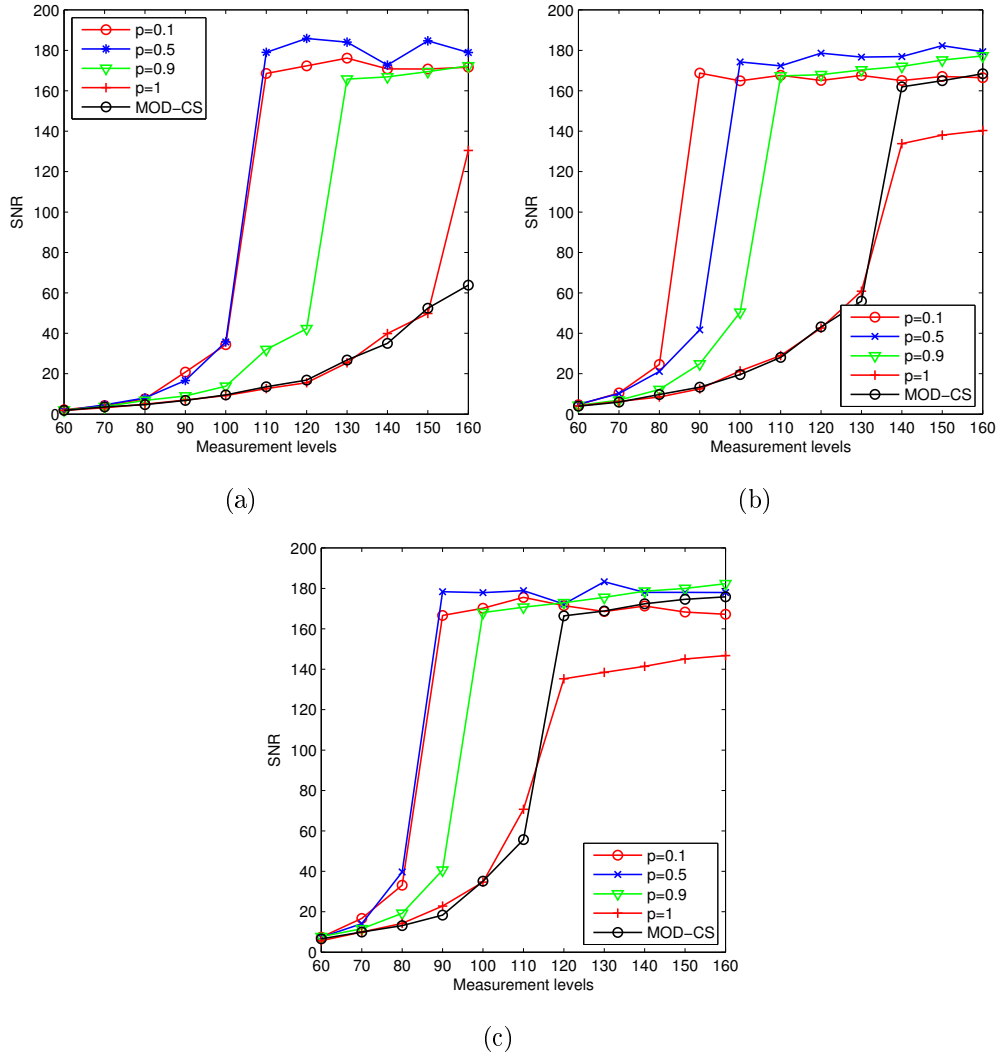


Figure 3.6: Performance of $l_{p < 1}$ minimization with partially known support and MOD-CS in terms of SNR for sparse signal, varying the number of measurements for (a) $s=4$, (b) $s=16$ and (c) $s=24$.

In Fig. 3.7 (a)-(c) measurement level is changed as 80, 120 and 160, respectively for different values of p on domain of known support. As it is expected SNR is increased with increased level of known support.

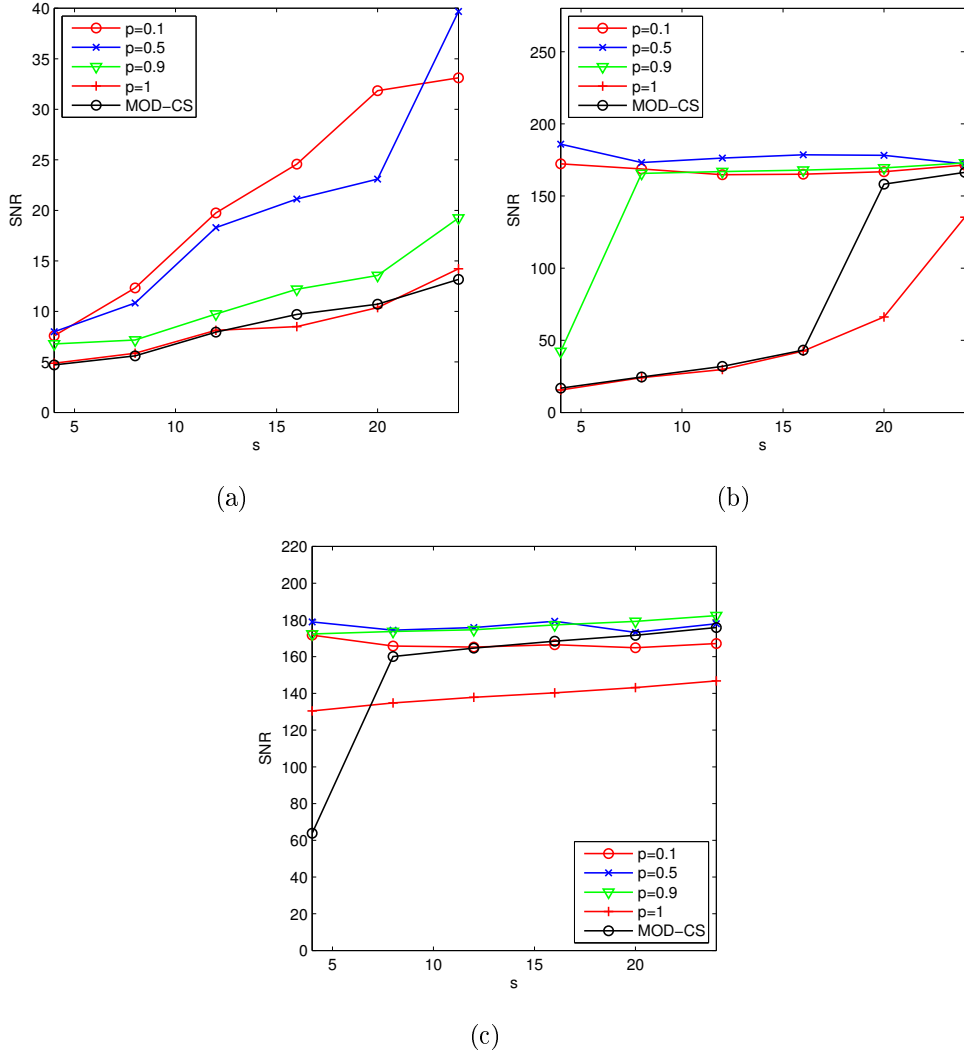


Figure 3.7: Performance of $l_{p<1}$ minimization with partially known support and MOD-CS in terms of SNR for sparse signals, varying the support size s for (a) $M=80$, (b) $M=120$ and (c) $M=160$.

3.5.3 The Compressible Case

For compressible case, we generate x whose coefficients decay like $j^{-\tau}$ where $j \in \{1, \dots, N\}$ and $\tau > 1$ with $N = 512$. In Fig. 3.8 (a)-(c) we analyze the effect of partially known support for different number of measurement levels for a compressible signal with $\tau = 1.5$. We choose best 40-term approximation of the compressible signal x . We plotted reconstruction SNR varying the number of measurements for $s = 4$, $s = 16$ and $s = 24$.

In Fig. 3.9 (a)-(c) performance of $l_{p<1}$ minimization with partially known support and MOD-CS in terms of SNR for compressible signals, varying the support size s for $M = 100$, $M = 120$ and $M = 140$ is obtained. As it is obvious in all cases $l_{p<1}$ minimization with partially known support gives better SNR for

lower s .

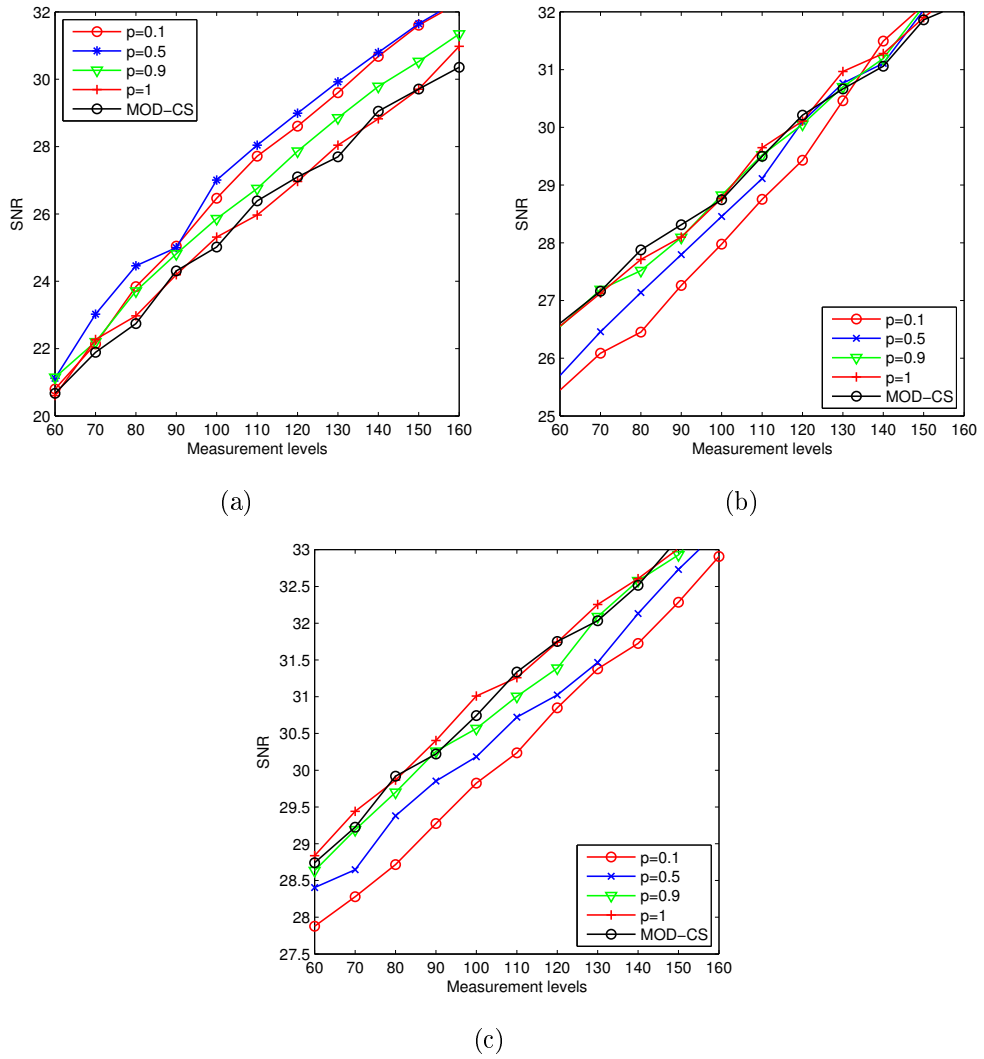


Figure 3.8: Performance of $l_{p < 1}$ minimization with partially known support and MOD-CS in terms of SNR for compressible signals, varying the number of measurements for (a) $s=4$, (b) $s=16$ and (c) $s=24$. The coefficients decay with a power $\tau = 1.5$.

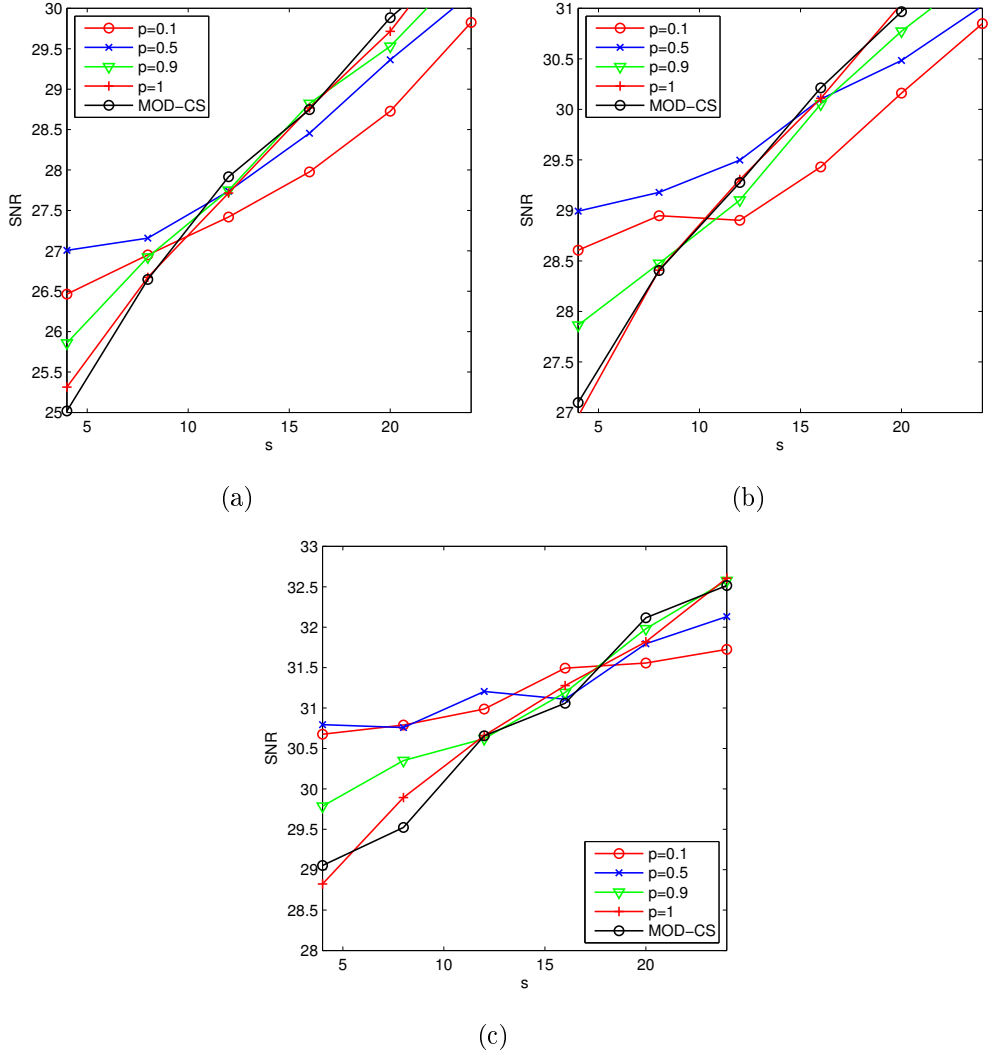


Figure 3.9: Performance of $l_{p<1}$ minimization with partially known support and MOD-CS in terms of SNR for compressible signals, varying the support size s for (a) $M=100$, (b) $M=120$ and (c) $M=140$. The coefficients decay with a power $\tau = 1.5$.

As the last experiment, we want to demonstrate the performance of the $l_{p<1}$ minimization with partially known support for real compressible signal. We choose cardiac image as our simulation data used in [22] owing to its structure for sparse decomposition. Two-level Daubechies-4 2D-DWT as sparsifying basis is used for 32×32 cardiac image. Support size of the image is 107 and the set of indexes of the approximation coefficients are selected as known part $s = |T| = 64$. Reconstruction SNR is shown in Fig. 3.10 for MOD-CS and $l_{p<1}$ minimization with partially known support for different p values. For this experiment the performance of $l_{p<1}$ minimization with partially known support is almost the same as MOD-CS when measurement level is relatively low. However reconstruction SNR is higher for increased levels of measurement level M .

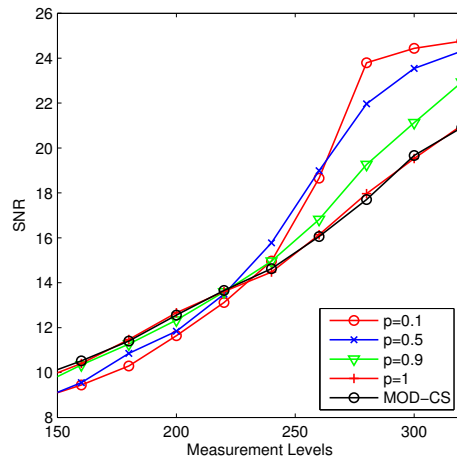


Figure 3.10: Performance of $l_{p<1}$ minimization with partially known support and MOD-CS in terms of SNR for cardiac image, varying the number of measurements ($s=64$).

CHAPTER 4

NONCONVEX COMPRESSED SENSING FOR GENERAL PERTURBATIONS

4.1 Introduction

In the CS community most of the research is directed toward to removing the additive noise from the measurements. However, when the matrix used to decode is different than the matrix used to encode, there exists a multiplicative noise in addition to the additive noise term in the measurements. Let A is decoding matrix such that

$$A = \Phi + E \tag{4.1}$$

where E models the distortions in the measurement matrix Φ . Substituting (4.1) in $\hat{y} = \Phi x + e$ causes an extra noise term Ex in the measurements in addition to additive noise. This type of noise can be encountered in several applications, for example when implementing decoding matrix in a sensor, or E can be modeled as the errors resulting from the transmission channel, which can be encountered in telecommunications, radar [40], source separation [41] and other scenarios. Therefore investigation of the multiplicative noise in compressed sensing is an important issue. Herman and Strohmer [42] analyzed the performance of BP under multiplicative and additive noise. Their framework is based on the relative error bounds of the measurement matrix and additive noise and then Herman and Needell [15] extended to greedy algorithms namely Compressive Sampling Matching Pursuit(COSAMP). Their results show that under reasonable assumptions stable reconstruction of the signal is limited by the noise level in the observation. Ding, Chen and Gu [43] analyzed the performance of Orthogonal Matching Pursuit [14] under General Perturbations. Exact recovery of support set of x can be guaranteed under suitable conditions, these conditions are based on RIP, relative perturbations and the smallest nonzero entry. In this chapter a completely perturbed nonconvex CS scenario is considered for $l_{p<1}$ minimization. The main result is that $l_{p<1}$ minimization for completely perturbed scenario is better than BP in terms of l_2 reconstruction error.

4.1.1 Notations and Symbols

Throughout this section, we adopt the notations and symbols used in [42]. Let the relative perturbations E and e are bounded by

$$\frac{\|E\|_2}{\|\Phi\|_2} \leq \varepsilon_\Phi, \quad \frac{\|E\|_2^{(K)}}{\|\Phi\|_2^{(K)}} \leq \varepsilon_\Phi^{(K)}, \quad \frac{\|e\|_2}{\|y\|_2} \leq \varepsilon_y. \quad (4.2)$$

$\|\cdot\|_2$ denotes the spectral norm and $\|\cdot\|_2^{(K)}$ represents the largest spectral norm taken over all K -column submatrices. α_K and β_K in (4.3) define the signals tail relative to its head, x_K is the best K -term approximation to x and x_{K^c} is the remaining terms.

$$\alpha_K := \frac{\|x_{K^c}\|_2}{\|x_K\|_2}, \quad \beta_K := \frac{\|x_{K^c}\|_1}{\sqrt{K}\|x_K\|_2} \quad (4.3)$$

and $\kappa_\Phi^{(K)}$ and γ_Φ are given as

$$\kappa_\Phi^{(K)} := \frac{\sqrt{1 + \delta_K}}{\sqrt{1 - \delta_K}}, \quad \gamma_\Phi := \frac{\|\Phi\|_2}{\sqrt{1 - \delta_K}}. \quad (4.4)$$

4.2 Completely Perturbed l_1 Minimization

4.2.1 RIP for A

Assume the completely perturbed situation with $e, E \neq 0$. In this case the problem in (2.13) is generalized to

$$\min \|\hat{x}\|_1 \quad \text{subject to} \quad \|\hat{y} - A\hat{x}\|_2 \leq \varepsilon'_{\Phi, K, y} \quad (4.5)$$

with a corrupted measurement matrix $A = \Phi + E$ for some $\varepsilon'_{\Phi, K, y}$ [42]. RIP condition and reconstruction error bounds are given in in [42].

Theorem 4.2.1. *Given the RIC δ_K associated with matrix Φ and the relative perturbation $\varepsilon_\Phi^{(K)}$ associated with matrix E in (4.2), fix the constant*

$$\hat{\delta}_{K, \max} := (1 + \delta_K)(1 + \varepsilon_\Phi^{(K)})^2 - 1. \quad (4.6)$$

Then the RIC $\delta_K < \hat{\delta}_{K, \max}$ is the smallest nonnegative number such that

$$(1 - \hat{\delta}_K)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \hat{\delta}_K)\|x\|_2^2 \quad (4.7)$$

holds for any K -sparse vector x .

4.2.2 Stability From Completely Perturbed Observations

Theorem 4.2.2. [42] Fix the relative perturbations $\varepsilon_\Phi, \varepsilon_\Phi^{(K)}, \varepsilon_\Phi^{(2K)}$, and ε_y be in (4.2). Assume the RIC for matrix Φ satisfies

$$\delta_{2K} < \frac{\sqrt{2}}{(1 + \varepsilon_\Phi^{(2K)})^2} - 1 \quad (4.8)$$

and that general signal x satisfies

$$\alpha_K + \beta_K < \frac{1}{\kappa_\Phi^{(K)}} \quad (4.9)$$

Set the total noise parameter

$$\varepsilon'_{\Phi, K, y} = \left(\frac{\varepsilon_\Phi^K \kappa_\Phi^{(K)} + \varepsilon_\Phi \gamma_\Phi \alpha_K}{1 - \kappa_\Phi^{(K)} (\alpha_K + \beta_K)} + \varepsilon_y \right) \|y\|_2 \quad (4.10)$$

Then, given $\hat{y} = \Phi x + e$ and $A = \Phi + E$, the solution of BP problem (4.5) obeys

$$\|\hat{x} - x\|_2 \leq K^{-1/2} C_0 \|x - x_K\|_1 + C_1 \varepsilon'_{\Phi, K, y} \quad (4.11)$$

where

$$C_0 = \frac{2(1 + (\sqrt{2} - 1))[(1 + \delta_{2K})(1 + \varepsilon_\Phi^{(2K)})^2 - 1]}{1 - (\sqrt{2} + 1)[(1 + \delta_{2K})(1 + \varepsilon_\Phi^{(2K)})^2 - 1]} \quad (4.12)$$

$$C_1 = \frac{4\sqrt{1 + \delta_{2K}}(1 + \varepsilon_\Phi^{(2K)})}{1 - (\sqrt{2} + 1)[(1 + \delta_{2K})(1 + \varepsilon_\Phi^{(2K)})^2 - 1]} \quad (4.13)$$

Theorem 4.2.2 generalizes the results in [10] in (2.14). If matrix Φ is unperturbed, then $E = 0$ and $\varepsilon_\Phi = \varepsilon_\Phi^{(K)} = 0$. Then $\hat{\delta}_K = \delta_K$ in (4.6) and RIP conditions for A and Φ are same.

4.2.3 Derivation of Total Perturbation Bound $\varepsilon'_{\Phi, K, y}$

Proposition 4.2.1. Assume that the matrix Φ satisfies the upper bound of the RIP in (4.7). Then, for every signal x we have

$$\|\Phi x\|_2 \leq \sqrt{1 + \delta_K} (\|x\|_2 + \frac{1}{\sqrt{K}} \|x\|_1) \quad (4.14)$$

Lemma 4.2.1. Assume the condition (4.9). Then for general signal x , its image under Φ can be bounded below by the positive quantity

$$\|\Phi x\|_2 \geq \sqrt{1 - \delta_K} \left(\|x_K\|_2 - \kappa_\Phi^{(K)} (\|x_{K^c}\|_2 + \frac{\|x_{K^c}\|_1}{\sqrt{K}}) \right) \quad (4.15)$$

Lemma 4.2.2. *Assume the condition (4.9) and set*

$$\varepsilon'_{\Phi, K, y} = \left(\frac{\varepsilon_{\Phi}^K \kappa_{\Phi}^{(K)} + \varepsilon_{\Phi} \gamma_{\Phi} \alpha_K}{1 - \kappa_{\Phi}^{(K)} (\alpha_K + \beta_K)} + \varepsilon_y \right) \|y\|_2 \quad (4.16)$$

where ε_{Φ} , $\varepsilon_{\Phi}^{(K)}$, ε_y defined in (4.2), and α_K and β_K in (4.3), then total perturbation obeys

$$\|Ex\|_2 + \|e\|_2 \leq \varepsilon'_{\Phi, K, y} \quad (4.17)$$

Proof. First divide the multiplicative noise term by $\|y\|_2$ and then apply Lemma 4.2.1

$$\begin{aligned} \frac{\|Ex\|_2}{\|\Phi x\|_2} &\leq \frac{(\|E\|_2^{(K)} \|x_K\|_2 + \|E\|_2 \|x_{K^c}\|_2) \cdot \frac{1}{\sqrt{1-\delta_K}}}{\|x_K\|_2 - \kappa_{\Phi}^{(K)} (\|x_{K^c}\|_2 + \|x_{K^c}\|_1 / \sqrt{K})} \\ &= \frac{(\|E\|_2^{(K)} + \|E\|_2 \alpha_K) \cdot \frac{1}{\sqrt{1-\delta_K}}}{1 - \kappa_{\Phi}^{(K)} (\alpha_K + \beta_K)} \\ &\leq \left(\frac{\varepsilon_{\Phi}^K \kappa_{\Phi}^{(K)} + \varepsilon_{\Phi} \gamma_{\Phi} \alpha_K}{1 - \kappa_{\Phi}^{(K)} (\alpha_K + \beta_K)} \right). \end{aligned} \quad (4.18)$$

Including the contribution from the additive noise term completes the proof [42]. \square

4.3 $l_{p<1}$ Minimization Under General Perturbations

Problem in (3.3) is nonconvex when $p < 1$. However Chartrand [33] showed with extensive numerical results that exact signal reconstruction is possible and required number of measurements for exact reconstruction is fewer and implementation is simpler compared to BP.

Saab et al. [34] studied the stability of $l_{p<1}$ minimization for the sparse and compressible signals in the presence of additive noise and gave the error bounds on the reconstruction error. Their results show that $l_{p<1}$ minimization is better than the BP in terms of stability and robustness with decreasing values of p . The following two theorems summarize our results.

Theorem 4.3.1. *Let $\sum_K^N := \{x \in R^N : \|x\|_0 \leq K\}$ represent the set of all K -sparse signals in R^N and the measurements as given in (2.12), for some constant $a > 1$ and aK is a positive integer and if (4.19) and (4.20) satisfy*

$$\alpha_K + \beta_K < \frac{1}{\kappa_{\Phi}^{(K)}} \quad (4.19)$$

$$(1 + \delta_{aK})(1 + \varepsilon_{\Phi}^{(aK)})^2 + a^{2/p-1}(1 + \delta_{(a+1)K})(1 + \varepsilon_{\Phi}^{(a+1)K})^2 < a^{2/p} \quad (4.20)$$

and the total perturbation is defined as [42]

$$\varepsilon'_{\Phi,K,y} = \left(\frac{\varepsilon_{\Phi}^K \kappa_{\Phi}^{(K)} + \varepsilon_{\Phi} \gamma_{\Phi} \alpha_K}{1 - \kappa_{\Phi}^{(K)} (\alpha_K + \beta_K)} + \varepsilon_y \right) \|y\|_2 \quad (4.21)$$

then program

$$\min \|\hat{x}\|_p^p \quad \text{subject to} \quad \|\hat{y} - A\hat{x}\|_2 \leq \varepsilon'_{\Phi,K,y} \quad (4.22)$$

obeys

$$\|\hat{x} - x\|_2^p < C_0^{(1)} \varepsilon'_{\Phi,K,y}$$

where $C_0^{(1)}$ is given as

$$C_0^{(1)} = \frac{2\sqrt{1 + \frac{1}{a^{(2/p)-1}((2/p)-1)}}}{\left[\left(2 - \left(1 + \delta_{(a+1)K} \right) \left(1 + \varepsilon_{\Phi}^{(a+1)K} \right)^2 \right)^{p/2} - \left(1 + \delta_{aK} \right)^{p/2} \left(1 + \varepsilon_{\Phi}^{(aK)} \right)^p a^{(p/2)-1} \right]} \quad (4.23)$$

Since x is sparse, total perturbation bound given in (4.21) will reduce to

$$\varepsilon'_{\Phi,K,y} = \left(\varepsilon_{\Phi}^K \kappa_{\Phi}^{(K)} + \varepsilon_y \right) \|y\|_2$$

for α_K and β_K are zero.

Theorem 4.3.2. *If x is not sparse (general case) and (4.19) and (4.20) are satisfied then program (4.22) obeys*

$$\|x - \hat{x}\|_2^p \leq C_1^{(1)} (\varepsilon'_{\Phi,K,y})^p + C_1^{(2)} a^{(p/2)-1} \|x - x_K\|_p^p$$

where

$$C_1^{(1)} = \frac{2^p \left[1 + a^{(p/2)-1} (2/p - 1)^{-p/2} \right]}{\left[\left(2 - \left(1 + \delta_{(a+1)K} \right) \left(1 + \varepsilon_{\Phi}^{(a+1)K} \right)^2 \right)^{p/2} - \left(1 + \delta_{aK} \right)^{p/2} \left(1 + \varepsilon_{\Phi}^{(aK)} \right)^p a^{(p/2)-1} \right]} \quad (4.24)$$

and

$$\begin{aligned} C_1^{(2)} &= 2 \frac{\left(\frac{p}{2-p} \right)^{p/2}}{a^{1-p/2}} \times \\ &= \left[1 + \frac{\left(1 + a^{(p/2)-1} \right) \left(1 + \delta_{aK} \right)^{p/2} \left(1 + \varepsilon_{\Phi}^{(aK)} \right)^p}{\left[\left(2 - \left(1 + \delta_{(a+1)K} \right) \left(1 + \varepsilon_{\Phi}^{(a+1)K} \right)^2 \right)^{p/2} - \left(1 + \delta_{aK} \right)^{p/2} \left(1 + \varepsilon_{\Phi}^{(aK)} \right)^p a^{(p/2)-1} \right]} \right] \end{aligned} \quad (4.25)$$

Remark 4.3.1. *Theorem 4.3.1 and 4.3.2 generalizes the results of [34]. For unperturbed decoding matrix Φ , $E = 0$ and $\varepsilon_{\Phi}^{(aK)} = \varepsilon_{\Phi}^{((a+1)K)} = 0$ and hence (4.23), (4.24) and (4.25) reduces to*

$$C_0^{(1)} = \frac{2\sqrt{1 + \frac{1}{a^{(2/p)-1}((2/p)-1)}}}{[(1 - \delta_{(a+1)K})^{p/2} - (1 + \delta_{aK})^{p/2}a^{(p/2)-1}]}$$

$$C_1^{(1)} = 2^p \frac{[1 + a^{(p/2)-1}((2/p) - 1)^{-p/2}]}{[(1 - \delta_{(a+1)K})^{p/2} - (1 + \delta_{aK})^{p/2}a^{(p/2)-1}]}$$

and

$$C_1^{(2)} = 2 \frac{((\frac{p}{2-p})^{p/2})}{a^{1-p/2}} \left[1 + \frac{(1 + a^{(p/2)-1})(1 + \delta_{aK})^{p/2}}{[(1 - \delta_{(a+1)K})^{p/2} - (1 + \delta_{aK})^{p/2}a^{(p/2)-1}]} \right]$$

respectively as in [34].

Remark 4.3.2. *If $p = 1$ and there is no perturbation on the measurement matrix i. e., $\varepsilon_{\Phi}^{(aK)} = \varepsilon_{\Phi}^{((a+1)K)} = 0$ we obtain precisely the same constants in [12].*

Theorems 4.3.1 and 4.3.2 considers a total noise term $\varepsilon'_{\Phi,K,y}$ which represents the additive and multiplicative noise simultaneously.

The constants $C_1^{(1)}$ and $C_2^{(1)}$ depend on K which reflects the degree of compressibility of signal x , on p determined by the recovery algorithm and on a which is a free parameter that (4.20) holds. $\delta_{(a+1)K}$ and δ_{aK} are fixed to 0.2 and 0.1, respectively in the figures 4.1 and 4.2. In Fig. 4.1 (a) and (b), $C_1^{(1)}$ is plotted versus p for $a = 3$ and $a = 4$ for $\varepsilon_{\Phi}^{(aK)} = \varepsilon_{\Phi}^{((a+1)K)} = 0, 0.05, 0.1$. Similarly in Fig. 4.2 (a) and (b) $C_1^{(1)}$ is plotted versus p for $a = 3$ and $a = 4$ for the same perturbation levels. As it is observed from the figures error bound noise constant ($C_1^{(1)}$) and error bound compressibility constant ($C_1^{(2)}$) is lower for smaller p ($p < 1$) and for bigger a . Also these constants are increasing with increasing values of $\varepsilon_{\Phi}^{(aK)}$ and $\varepsilon_{\Phi}^{((a+1)K)}$.

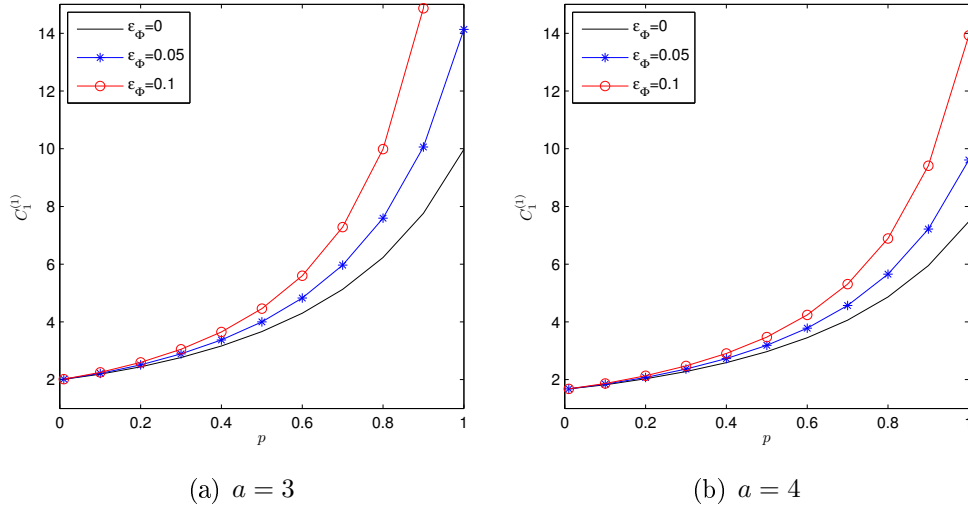


Figure 4.1: Error bound noise constant ($C_1^{(1)}$) versus p

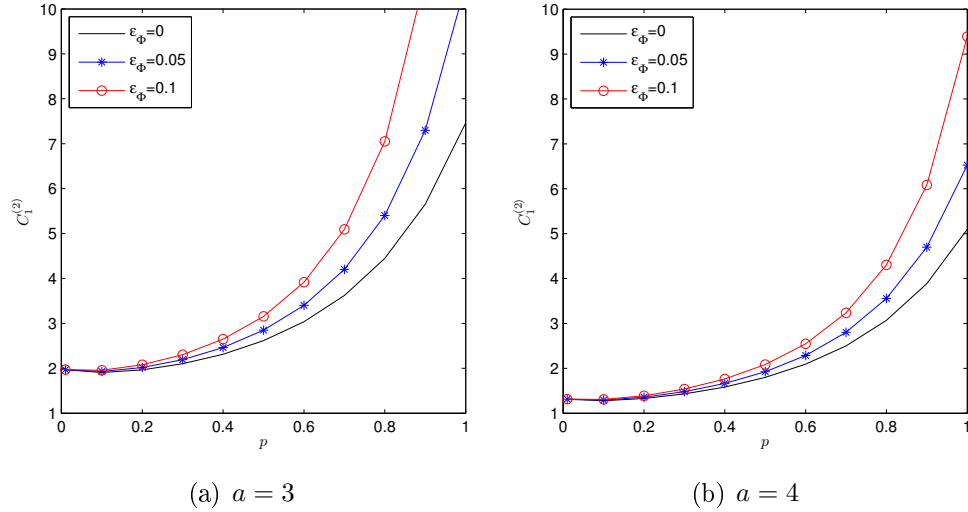


Figure 4.2: Error bound noise compressibility constant ($C_1^{(2)}$) versus p

Proof. Let \hat{x} is the solution of (4.22) and total perturbation is given as in (4.21). Define the difference $h = \hat{x} - x$ between the original and reconstructed signals. Divide h into h_{T_0} and $h_{T_0^c}$ where T_0 is the set of K -largest components of x and T_0^c is the rest. Decompose T_0^c in to sets T_1, T_2, \dots, T_j for $j > 1$. Let T_{01} denotes the set union $T_0 \cup T_1$. Our proof is analogous to proof in [34] and [12], but we take into account the perturbed measurement matrix and modified RIC. Since x and

\hat{x} are feasible then from the triangle inequality

$$\begin{aligned}
\|Ah\|_2^p &= \|Ah_{T_{01}} + \sum_{j \geq 2} Ah_{T_j}\|_2^p \\
&= \|Ah_{T_{01}}\|_2^p - \left\| \sum_{j \geq 2} Ah_{T_j} \right\|_2^p \\
&\geq (1 - \hat{\delta}_{L+|T_0|})^{p/2} \|h_{T_{01}}\|_2^p - (1 + \hat{\delta}_{L+|T_0|})^{p/2} \sum_{j \geq 2} \|h_{T_j}\|_2^p \quad (4.26)
\end{aligned}$$

Using the result of Theorem 1 in [42], (4.26) becomes

$$\begin{aligned}
\|Ah\|_2^p &\geq (2 - (1 + \delta_{L+|T_0|})(1 + \varepsilon_{\Phi}^{L+|T_0|})^2)^{p/2} \|h_{T_{01}}\|_2^p \\
&\quad - (1 + \delta_L)^{p/2} (1 + \delta_L)^{p/2} (1 + \varepsilon_{\Phi}^L)^p \sum_{j \geq 2} \|h_{T_j}\|_2^p \quad (4.27)
\end{aligned}$$

writing $\|h_{T_{01}}\|_2^p$ and $\sum_{j \geq 2} \|h_{T_j}\|_2^p$ in terms of $\|h\|_2$. Since [34]

$$\sum_{j \geq 2} \|h_{T_j}\|_2^p \leq L^{p/2-1} \sum_{j \geq 1} \|h_{T_j}\|_p^p \leq L^{p/2-1} \|h_{T_0^c}\|_p^p$$

and

$$\|h_{T_0^c}\|_p^p \leq \|h_{T_0}\|_p^p + 2\|x_{T_0^c}\|_p^p$$

then

$$\sum_{j \geq 2} \|h_{T_j}\|_2^p \leq L^{p/2-1} \left(\|h_{T_0}\|_p^p + 2\|x_{T_0^c}\|_p^p \right) \leq (\zeta)^{1-p/2} \left(\|h_{T_{01}}\|_2^p + 2|T_0|^{p/2-1} \|x_{T_0^c}\|_p^p \right) \quad (4.28)$$

where $\zeta = \frac{|T_0|}{L}$ and $\|h_{T_0}\|_p^p \leq |T_0|^{1-p/2} (\|h_{T_0}\|_2^p)$. Combining (4.27) and (4.28), the triangle inequality becomes

$$\begin{aligned}
\|Ah\|_2^p &\geq \left[\left(2 - (1 + \delta_{L+|T_0|})(1 + \varepsilon_{\Phi}^{L+|T_0|})^2 \right)^{p/2} - \left((1 + \delta_L)^{p/2} (1 + \varepsilon_{\Phi}^L)^p \right) \zeta^{1-p/2} \right] \times \\
&\quad \|h_{T_{01}}\|_2^p - 2\zeta^{1-p/2} |T_0|^{p/2-1} (1 + \delta_L)^{p/2} (1 + \varepsilon_{\Phi}^L)^p \|x_{T_0^c}\|_p^p
\end{aligned}$$

Also

$$\|Ah\|_2 \leq 2\varepsilon'_{\Phi, K, y}$$

then $\|h_{T_{01}}\|_2^p$ is bounded as

$$\|h_{T_{01}}\|_2^p \leq \frac{\left[(2\varepsilon'_{\Phi, K, y})^p + 2\zeta^{1-p/2} \left((1 + \delta_L)^{p/2} (1 + \varepsilon_{\Phi}^L)^p \right) \frac{\|x_{T_0^c}\|_p^p}{|T_0|^{1-p/2}} \right]}{\left[\left(2 - (1 + \delta_{L+|T_0|})(1 + \varepsilon_{\Phi}^{L+|T_0|})^2 \right)^{p/2} - \left((1 + \delta_L)^{p/2} (1 + \varepsilon_{\Phi}^L)^p \right) \zeta^{1-p/2} \right]} \quad (4.29)$$

At this point, denominator of (4.29) must be greater than zero which impose the condition given in (4.20).

Using the same arguments given in [12]

$$\|h\|_2^2 \leq \left[(1 + \zeta^{1-p/2}(2/p-1)^{-p/2}) \|h_{T_{01}}\|_2^p + 2\zeta^{1-p/2}(2/p-1)^{-p/2} \frac{\|x_{T_0^c}\|_p^p}{|T_0|^{1-p/2}} \right]^{2/p}$$

Setting $|T_0| = K$ and $L = aK$ we obtain

$$\|h\|_2^p \leq C_1^{(1)} (\varepsilon'_{\Phi, K, y})^p + C_1^{(2)} a^{(p/2)-1} \|x - x_K\|_p^p$$

Since $\|x_{T_0^c}\|$ is zero for sparse case, the proof of Theorem 4.3.2 is easily modified for the proof of Theorem 4.3.1. \square

4.4 Stability and Instance Optimality in Nonconvex Compressed Sensing

In this section theoretical results are presented to show that it is possible to recover sparse and compressible signals robustly and stably using $l_{p<1}$ minimization when additive noise exists. The stability of recovering x from y is also related with the RIP of the measurement matrix. We begin with a stability and robustness theorem for $l_{p<1}$ minimization that generalizes the theorem in [10].

Theorem 4.4.1. *Suppose that x_K is the best K -term approximation to x and the measurements are given as in (2.2), if*

$$\delta_{(a+1)K} + a^{1/2-1/p} (\delta_{(a+1)K}^2 + \delta_{2aK}^2)^{1/2} < 1 \quad (4.30)$$

is satisfied for some constant $a > 1$ and aK is a positive integer then the solution \hat{x} to

$$\min \|\hat{x}\|_p^p \quad \text{subject to} \quad \|\hat{y} - \Phi \hat{x}\|_2 \leq \epsilon \quad (4.31)$$

obeys

$$\|\hat{x} - x\|_2^p < C_{a,K,p} \epsilon^p + D_{a,K,p} \frac{\|x - x_K\|_p^p}{K^{1-p/2}}$$

The constants $C_{a,K,p}$ and $D_{a,K,p}$ are

$$C_{a,K,p} = 2^p \frac{(1 + a^{p/2-1})(1 + \delta_{(a+1)K})^{p/2}}{(1 - \delta_{(a+1)K})^p - a^{p/2-1}(\delta_{(a+1)K}^2 + \delta_{2aK}^2)^{p/2}} \quad (4.32)$$

$$D_{a,K,p} = 2a^{p/2-1} \left[1 + \frac{(1 + a^{p/2-1})(\delta_{(a+1)K}^2 + \delta_{2aK}^2)^{p/2}}{(1 - \delta_{(a+1)K})^p - a^{p/2-1}(\delta_{(a+1)K}^2 + \delta_{2aK}^2)^{p/2}} \right] \quad (4.33)$$

Recovery error (to the p^{th} power) is bounded by the sum of two terms, first one is proportional to total perturbation term and the second is related to best K -term approximation error of signal.

Proof. Let x be an arbitrary signal and $0 < p < 1$. T_0 be the locations of the K -largest coefficients of x . Let us write $\hat{x} = x + h$ with $h \in R^n$. Our aim is to bound $\|h\|_2$ given that $\|\Phi h\|_2 \leq 2\epsilon$. T_{01} denotes $T_0 \cup T_1$. We begin by dividing h into a sum of sparse vectors with disjoint set and T_0^c into sets T_1, T_2, \dots where $|T_j| = L$ for $j \geq 1$. T_1 is the locations of L largest coefficients of $h_{T_0^c}$; T_2 is the locations of the second L largest coefficients of $h_{T_0^c}$ and so on. Since $\|\cdot\|_p^p$ satisfies the triangle inequality

$$\begin{aligned} \|\hat{x}\|_p^p &\geq \|x + h\|_p^p = \|x_{T_0} + h_{T_0}\|_p^p + \|x_{T_0^c} + h_{T_0^c}\|_p^p \\ &\geq \|x_{T_0}\|_p^p - \|h_{T_0}\|_p^p - \|x_{T_0^c}\|_p^p + \|h_{T_0^c}\|_p^p \end{aligned}$$

then

$$\begin{aligned} \|h_{T_0^c}\|_p^p &\leq \|x\|_p^p + \|x_{T_0^c}\|_p^p + \|h_{T_0}\|_p^p - \|x_{T_0}\|_p^p \\ &= 2\|x_{T_0^c}\|_p^p + \|h_{T_0}\|_p^p \\ &= 2\|x - x_{T_0}\|_p^p + \|h_{T_0}\|_p^p \end{aligned}$$

For each $u \in T_j$ and $v \in T_{j-1}$ $|h(u)| \leq |h(v)|$ so

$$|h(u)|^p \leq \|h_{T_{j-1}}\|_p^p / L$$

Then

$$\begin{aligned} |h(u)|^2 &= \|h_{T_{j-1}}\|_p^2 / L^{2/p} \\ \|h_{T_j}\|_2^2 &\leq L^{1-2/p} \|h_{T_{j-1}}\|_p^2 \\ \|h_{T_j}\|_2^p &\leq L^{p/2-1} \|h_{T_{j-1}}\|_p^p \end{aligned}$$

and thus

$$\begin{aligned} \|h_{T_{01}^c}\|_2^p &\leq \sum_{j \geq 2} \|h_{T_j}\|_2^p \leq L^{p/2-1} \sum_{j \geq 1} \|h_{T_j}\|_p^p = L^{p/2-1} \|h_{T_0^c}\|_p^p \\ &= L^{p/2-1} (2\|x - x_{T_0}\|_p^p + \|h_{T_0}\|_p^p) \end{aligned} \quad (4.34)$$

Using Hölder's inequality for $\|h_{T_0}\|_p^p$ to convert back from $l_{p < 1}$ to l_2

$$\|h_{T_0}\|_p^p \leq (|T_0|)^{1-p/2} \|h_{T_{01}}\|_2^p$$

is obtained and (4.34) becomes

$$\|h_{T_{01}^c}\|_2^p \leq \sum_{j \geq 2} \|h_{T_j}\|_2^p \leq L^{p/2-1} \left((|T_0|)^{1-p/2} \|h_{T_{01}}\|_2^p + 2\|x - x_{T_0}\|_p^p \right) \quad (4.35)$$

Now we need to control the size of $\|h_{T_{01}}\|_2^p$. Observing that $\Phi h_{T_{01}} = \Phi h - \sum_{j \geq 2} \Phi h_{T_j}$ therefore

$$\begin{aligned} \|\Phi h_{T_{01}}\|_2^2 &= \langle \Phi h_{T_{01}}, \Phi h_{T_{01}} \rangle = \langle \Phi h_{T_{01}}, \Phi h \rangle - \langle \Phi h_{T_{01}}, \sum_{j \geq 2} \Phi h_{T_j} \rangle \\ &\leq \|\Phi h_{T_{01}}\|_2 \|\Phi h\|_2 + \sum_{j \geq 2} |\langle \Phi h_{T_{01}}, \Phi h_{T_j} \rangle| \end{aligned}$$

Taking the p^{th} power and using the fact that for any $b, c \geq 0$, and $0 < p < 1$, $b^p + c^p > (b + c)^p$ we have

$$\|\Phi h_{T_{01}}\|_2^{2p} \leq \|\Phi h_{T_{01}}\|_2^p \|\Phi h\|_2^p + \sum_{j \geq 2} (|\langle \Phi h_{T_{01}}, \Phi h_{T_j} \rangle|)^p \quad (4.36)$$

Observing that $(1 - \delta_{L+|T_0|})^{p/2} \|h_{T_{01}}\|_2^p \leq \|\Phi h_{T_{01}}\|_2^p$ and $\|\Phi h_{T_{01}}\|_2^p \leq (2\epsilon)^p$, (4.36) becomes

$$\begin{aligned} (1 - \delta_{L+|T_0|})^p (\|h_{T_{01}}\|_2^p)^2 &\leq (2\epsilon)^p (1 + \delta_{L+|T_0|})^{p/2} \|h_{T_{01}}\|_2^p \\ &\quad + (\delta_{L+|T_0|} \|h_{T_0}\|_2 + \delta_{2L} \|h_{T_1}\|_2)^p \times \sum_{j \geq 2} \|h_{T_j}\|_2^p \\ &\leq (2\epsilon)^p (1 + \delta_{L+|T_0|})^{p/2} \|h_{T_{01}}\|_2^p \\ &\quad + (\delta_{L+|T_0|}^2 + \delta_{2L}^2)^{p/2} \|h_{T_{01}}\|_2^p \times \sum_{j \geq 2} \|h_{T_j}\|_2^p \end{aligned}$$

Using (4.35) we have

$$\begin{aligned} (1 - \delta_{L+|T_0|})^p \|h_{T_{01}}\|_2^p &\leq (2\epsilon)^p (1 + \delta_{L+|T_0|})^{p/2} \\ &\quad + (\delta_{L+|T_0|}^2 + \delta_{2L}^2)^{p/2} L^{p/2-1} \left(2\|x - x_{T_0}\|_p^p + (|T_0|)^{1-p/2} \|h_{T_{01}}\|_2^p \right) \end{aligned} \quad (4.37)$$

Then setting $L = aK$, $|T_0| = K$ in (4.37) we obtain

$$\|h_{T_{01}}\|_2^p \leq \theta \epsilon^p + \zeta \frac{\|x - x_K\|_p^p}{K^{1-p/2}} \quad (4.38)$$

with constants

$$\begin{aligned} \theta &= \frac{2^p (1 + \delta_{(a+1)K})^{p/2}}{(1 - \delta_{(a+1)K})^p - a^{p/2-1} (\delta_{(a+1)K}^2 + \delta_{2aK}^2)^{p/2}} \\ \zeta &= \frac{2(\delta_{(a+1)K}^2 + \delta_{2aK}^2)^{p/2} a^{p/2-1}}{(1 - \delta_{(a+1)K})^p - a^{p/2-1} (\delta_{(a+1)K}^2 + \delta_{2aK}^2)^{p/2}} \end{aligned}$$

Substituting (4.35) and (4.38) into $\|h\|_2^p \leq \|h_{T_{01}^c}\|_2^p + \|h_{T_{01}}\|_2^p$ results in

$$\|h\|_2^p \leq C_{a,K,p} \epsilon^p + D_{a,K,p} \frac{\|x - x_K\|_p^p}{K^{1-p/2}}$$

where the constants $C_{a,K,p}$ and $D_{a,K,p}$ are as in (4.32) and (4.33) in Theorem 4.4.1. Denominator of these constants must be positive implying the condition given in (4.30). \square

Theorem 4.4.2. *Let x be a strictly sparse signal. If (4.30) is satisfied for some constant $a > 1$ and aK is a positive integer, then the solution \hat{x} to (4.31) obeys*

$$\|\hat{x} - x\|_2 < (C_{a,K,p})^{1/p} \epsilon$$

where $C_{a,K,p}$ is given in (4.32)

Proof. The proof can easily be deduced from the proof of Theorem 4.4.1 by setting $x = x_K$. Since we assume x is sparse, the term related with $D_{a,K,p}$ in (4.33) goes to zero and $C_{a,K,p}$ in Theorem 4.4.1 is obtained. \square

In Theorem 4.4.1 and Corollary 4.4.2 necessary conditions are provided for recovering sparse and compressible signals using $l_{p<1}$ minimization. The constants $C_{a,K,p}$ and $D_{a,K,p}$ determines the upper bounds on the recovery error.

Remark 4.4.1. *If $p = 1$ and $a = 1$, we obtain precisely the same constants $C_{a,K,p}$ and $D_{a,K,p}$ and reconstruction condition $\delta_{2K} < \sqrt{2} - 1$ given in [10] and $C_{a,K,p} < 8.48$ and $D_{a,K,p} < 4.19$ for $\delta_{2K} = 0.2$.*

Remark 4.4.2. *It is sufficient to ensure the stable and robust recovery of signal x from measurements y if the measurement matrix Φ satisfies a stronger condition*

$$\delta_{2aK} < \frac{1}{1 + \sqrt{2}a^{1/2-1/p}} \quad (4.39)$$

with the constants $C_{a,K,p}$ and $D_{a,K,p}$ given in (4.32) and (4.33) since $\delta_{2aK} < \delta_{(a+1)K}$.

4.5 Stability and Instance Optimality in Completely Perturbed $l_{p<1}$ Minimization

In previous section reconstruction conditions and error bounds for $l_{p<1}$ minimization under additive noise is given. In this section a completely perturbed CS scenario is considered for $l_{p<1}$ minimization. Suppose that the measurement matrix is corrupted as in (4.1) in addition to additive noise in the measurements.

Theorem 4.5.1. *Let the relative perturbations ε_Φ , $\varepsilon_\Phi^{(2aK)}$, and ε_y be in (4.2) and signal x satisfies $\alpha_K + \beta_K < \frac{1}{\kappa_\Phi^{(K)}}$. If*

$$\delta_{2aK} < \frac{2 + \sqrt{2}a^{1/2-1/p}}{(1 + \sqrt{2}a^{1/2-1/p})(1 + \varepsilon_\Phi^{(2aK)})^2} - 1 \quad (4.40)$$

is satisfied for some constant $a > 1$ and aK is a positive integer then the solution \hat{z} to

$$\min \|\hat{z}\|_p^p \quad \text{subject to} \quad \|\hat{y} - A\hat{z}\|_2 \leq \varepsilon'_{\Phi,K,y} \quad (4.41)$$

obeys

$$\|\hat{z} - x\|_2^p < \hat{C}_{a,K,p}(\varepsilon'_{\Phi,K,y})^p + \hat{D}_{a,K,p} \frac{\|x - x_K\|_p^p}{K^{1-p/2}}$$

with total noise parameter [42]

$$\varepsilon'_{\Phi,K,y} = \left(\frac{\varepsilon_\Phi^{(K)} \kappa_\Phi^{(K)} + \varepsilon_\Phi \gamma_\Phi \alpha_K}{1 - \kappa_\Phi^{(K)} (\alpha_K + \beta_K)} + \varepsilon_y \right) \|y\|_2$$

The constants $\hat{C}_{a,K,p}$ and $\hat{D}_{a,K,p}$ are

$$\hat{C}_{a,K,p} = 2^p \frac{(1 + a^{p/2-1})(1 + \delta_{(a+1)K})^{p/2} (1 + \varepsilon_{\Phi}^{(a+1)K})^p}{\Delta} \quad (4.42)$$

$$\hat{D}_{a,K,p} = 2a^{p/2-1} \left[1 + \frac{(1 + a^{p/2-1}) \left[((1 + \delta_{(a+1)K})(1 + \varepsilon_{\Phi}^{(a+1)K})^2 - 1)^2 + ((1 + \delta_{2aK})(1 + \varepsilon_{\Phi}^{2aK})^2 - 1)^2 \right]^{p/2}}{\Delta} \right] \quad (4.43)$$

where

$$\begin{aligned} \Delta &= \left[2 - (1 + \delta_{(a+1)K})(1 + \varepsilon_{\Phi}^{(a+1)K})^2 \right]^p \\ &\quad - a^{p/2-1} \left[((1 + \delta_{(a+1)K})(1 + \varepsilon_{\Phi}^{(a+1)K})^2 - 1)^2 + ((1 + \delta_{2aK})(1 + \varepsilon_{\Phi}^{2aK})^2 - 1)^2 \right]^{p/2} \end{aligned} \quad (4.44)$$

Proof. The proof of Theorem 4.5.1 is similar to proof of theorem 4.4.1 with decoding matrix Φ is replaced with A and δ_K is replaced with $\hat{\delta}_K$. Let the reconstructed vector be $\hat{z} = x + h$ where h is the perturbation due to e and E . Also it is observed

$$\|A(\hat{z} - x)\|_2 \leq \|A\hat{z} - \hat{y}\| + \|\hat{y} - Ax\| \leq 2\varepsilon'_{\Phi,K,y} \quad (4.45)$$

where $\varepsilon'_{\Phi,K,y}$ is total noise parameter term. We obtain reconstruction condition

$$\hat{\delta}_{2aK} < \frac{1}{1 + \sqrt{2}a^{1/2-1/p}} \quad (4.46)$$

by imposing $\hat{\delta}_{(a+1)K} < \hat{\delta}_{2aK}$. Substituting the condition [42]

$$\hat{\delta}_{2aK,max} := (1 + \delta_{2aK})(1 + \varepsilon_{\Phi}^{(2aK)})^2 - 1$$

into (4.46) we obtain the reconstruction condition in (4.40). Following the same procedures in proof of theorem 4.4.1 we obtain the constants $\hat{C}_{a,K,p}$ and $\hat{D}_{a,K,p}$ in (4.42) and (4.43). \square

Theorem 4.5.2. *Let x be a strictly sparse signal. If (4.40) is satisfied for some constant $a > 1$ and aK is a positive integer, then the solution \hat{z} to (4.41) obeys*

$$\|\hat{z} - x\|_2 < \hat{C}_{a,K,p} \left((\kappa_{\Phi}^{(K)} \varepsilon_{\Phi}^{(K)} + \varepsilon_y) \|y\|_2 \right)^p$$

where $\hat{C}_{a,K,p}$ is given in (4.42)

Proof. Since x is sparse, total perturbation bound given in (4.31) will reduce to

$$\varepsilon'_{\Phi,K,y} = (\kappa_{\Phi}^{(K)} \varepsilon_{\Phi}^{(K)} + \varepsilon_y)$$

due to the terms α_K and β_K goes to zero. \square

Remark 4.5.1. If $p = 1$ and $a = 1$, we obtain precisely the same constants given in [42] and reconstruction condition $\delta_{2K} < \frac{\sqrt{2}}{(1+\varepsilon_{\Phi}^{(2K)})^2} - 1$.

Remark 4.5.2. If $p = 1$ and $a = 1$ and there is no perturbation in the measurement matrix, i.e. $E = 0$, we obtain precisely the same constants $\hat{C}_{a,K,p} = 4\sqrt{1+\delta_{2K}}/(1-(\sqrt{2}+1)\delta_{2K})$ and $\hat{D}_{a,K,p} = 2(1+(\sqrt{2}-1)\delta_{2K})/(1-(\sqrt{2}+1)\delta_{2K})$ and reconstruction condition $\delta_{2K} < \sqrt{2} - 1$ [10].

4.6 Simulation Results

In this section the simulation results are presented for $l_{p<1}$ minimization under multiplicative noise. No measurement noise added so that aim is to show the effect of perturbation in the measurement matrix. The performance of $l_{p<1}$ minimization and BP are compared under perturbed measurement matrix. MATLAB is used in all simulations. Regularized Iteratively Reweighted Least Squares method is used in the solution of (4.22) where $l_{p<1}$ objective function in (4.22) is replaced by a weighted l_2 norm [33]

$$\min \sum_{i=1}^N w_i \hat{x}_i^2 \quad \text{subject to} \quad A\hat{x} = y \quad (4.47)$$

The closed form of the solution is given as, giving the next iterate $\hat{x}^{(n+1)}$:

$$\hat{x}^{(n+1)} = Q_n A^T (A Q_n A^T)^{-1} y$$

where Q_n is diagonal matrix with elements $1/w_i$. Choose the weights $w_i = ((\hat{x}_i^{(n)})^2 + \gamma)^{p/2-1}$ where γ is initially a large constant added to avoid division by zero whenever $\hat{x}_i^{(n)} = 0$ since $p - 2$ is negative and to prevent local minima in early iterations. The value of γ is decreased according to rule $\gamma_{n+1} = 0.99\gamma_n$ and iteration is continued until $\gamma < 10^{-8}$ or $\|\hat{x}_n - \hat{x}^{n-1}\|_2 < 10^{-8}$ is satisfied [33].

Adding Gaussian noise with $e \sim \mathcal{N}(0, \sigma^2)$ in (2.12), measurements are corrupted and unconstrained $l_{p<1}$ minimization

$$\min \|\hat{x}\|_p^p + \frac{1}{\lambda} \|\hat{y} - A\hat{x}\|_2 \quad (4.48)$$

is solved. The parameter λ is adjusted manually. Starting with an initial $\hat{x}^{(1)}$ and using the necessary conditions of Euler-Lagrange equation, iterative solution of (4.48) giving the next iterate $\hat{x}^{(n+1)}$ is given as

$$\hat{x}^{(n+1)} = \left(A^T A + \lambda Q_n^{-1} \right)^{-1} A^T \hat{y} \quad (4.49)$$

where Q_n is defined previously. Minimization is repeated until the constraint in (4.22) is active.

In each parameter set 100 trials are performed and in each trial an $M \times N$ Gaussian matrix Φ with independent normally entries $\mathcal{N}(0, \sigma^2)$ is created. Standard deviation is set to $1/\sqrt{M}$. Different random Gaussian matrix E is created and it is scaled so that $\|E\|_2 = \varepsilon_\Phi \|\Phi\|_2$. Perturbed matrix A is formed as $\Phi + E$ and it is used in the solution of $l_{p<1}$ minimization, BP and BPDN. SPGL1 [44] toolbox is used for BP and BPDN. Normalized reconstruction error $\|x - \hat{x}\|/\|x\|_2$ of 100 trials are averaged for different parameter sets.

In the first experiment we show the effect of increase in the relative perturbation ε_Φ in reconstructing signal x varying the perturbation level ε_Φ for measurement level $M = 128$ and the results are given in Fig. 4.3 (a)-(c) for noiseless measurements. At low sparsity levels the performance of BP and $l_{p<1}$ minimization almost same but increased sparsity level causes $l_{p<1}$ minimization to obtain better estimate compared to BP. For the noisy measurements unconstrained $l_{p<1}$ minimization and BPDN are solved for comparison. In Fig. 4.4 (a)-(c) average relative error versus ε_Φ is shown when measurements are corrupted by adding Gaussian noise with $\sigma = 0.05$.

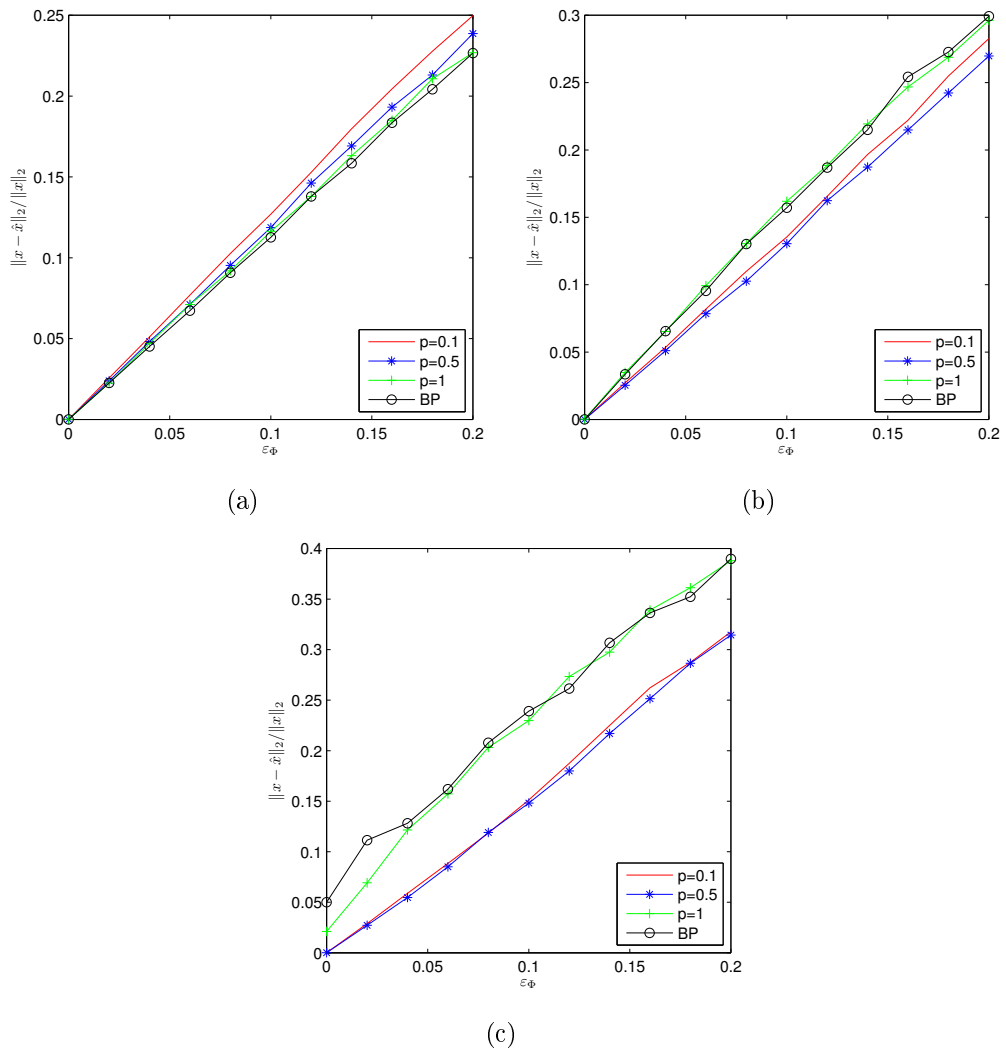


Figure 4.3: Average relative error versus ε_Φ for (a) $K = 10$, (b) $K = 20$, (c) $K = 30$.

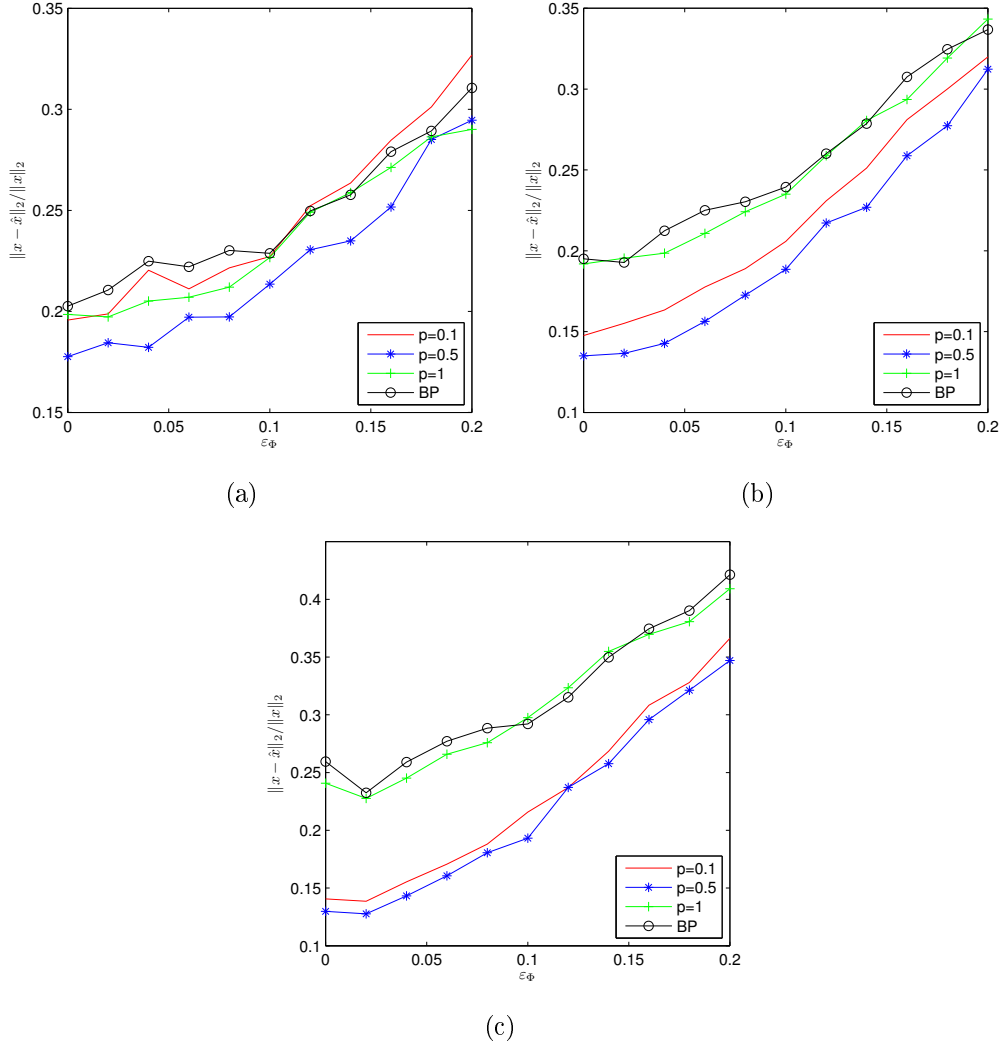


Figure 4.4: Average relative error versus ε_Φ with measurement noise $\sigma = 0.05$ for (a) $K = 10$, (b) $K = 20$, (c) $K = 30$.

In Fig. 4.5 (a)-(c) average relative error is plotted versus measurement level M for $\varepsilon_\Phi = 0$, $\varepsilon_\Phi = 0.05$ and $\varepsilon_\Phi = 0.1$, respectively. It is observed that average relative error is lower for $p < 1$ compared to BP for same measurement levels. Also values $p < 1$ gives lower error compared to $p = 1$. In Fig. 4.6 (a) and (b) average relative error versus measurement level is presented for perturbed measurement matrix A with measurement noise $\sigma = 0.01$ and $\sigma = 0.05$, respectively. As it is obvious, $l_{p < 1}$ minimization gives lower error compared to BPDN for lower measurement levels.

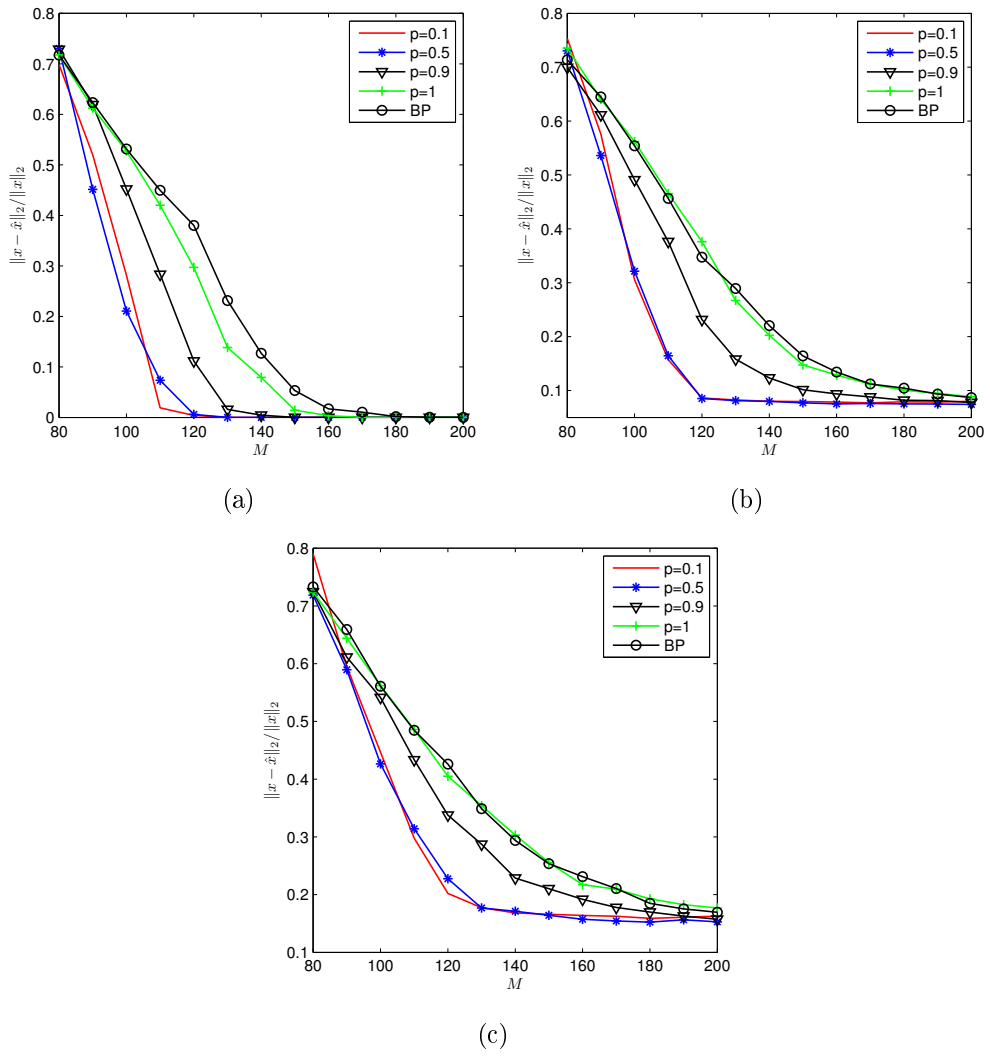


Figure 4.5: Average relative error versus measurement level M for (a) $\varepsilon_\Phi=0$, (b) $\varepsilon_\Phi=0.05$ and (c) $\varepsilon_\Phi=0.1$.

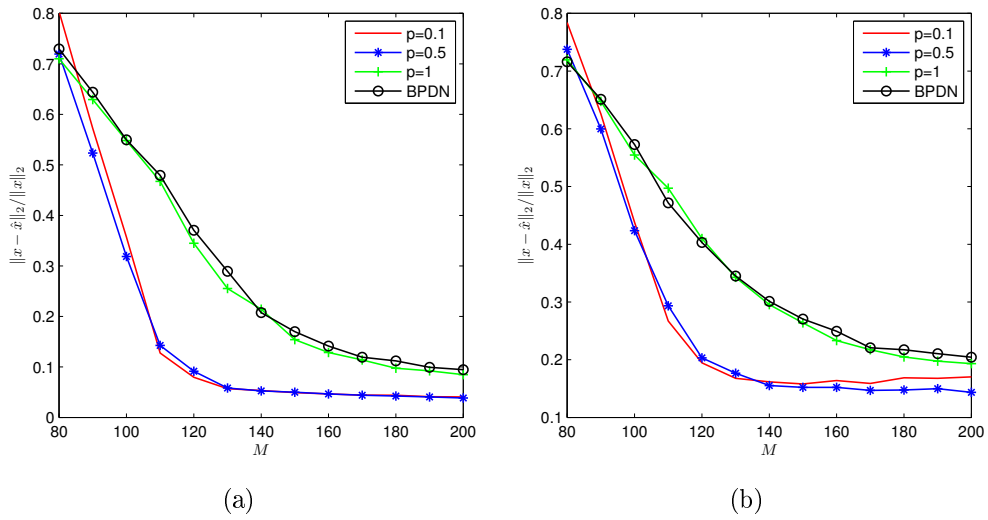
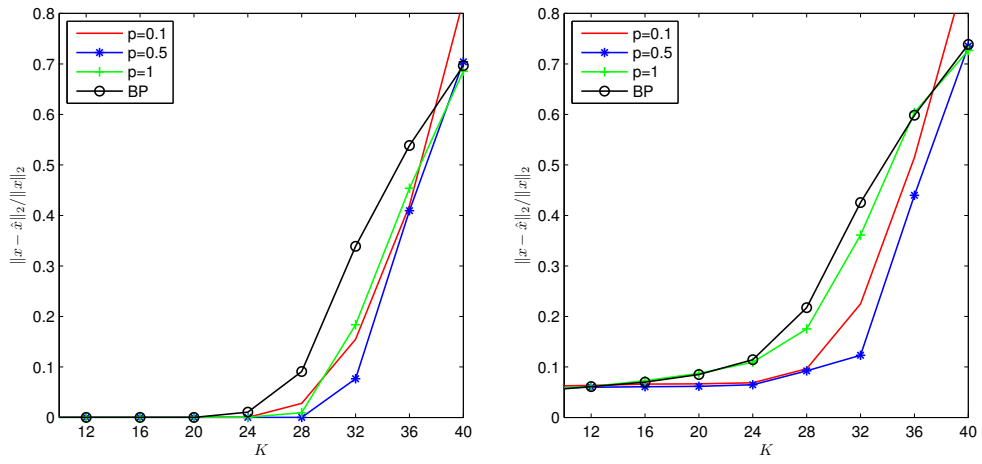


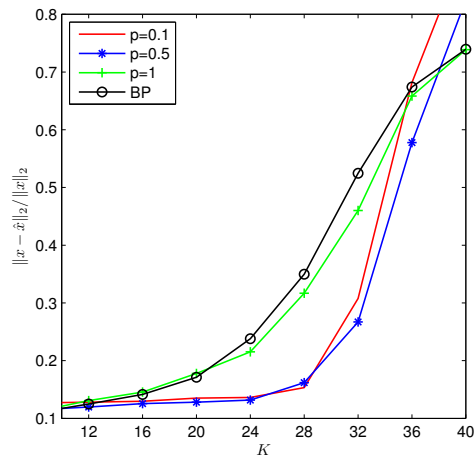
Figure 4.6: Average relative error versus measurement level M for $\varepsilon_\Phi=0.05$ with measurement noise, (a) $\sigma = 0.01$ (b) $\sigma = 0.05$

Effect of increasing sparsity level K is given in Fig. 4.7 (a)-(c) for $\varepsilon_\Phi = 0$, $\varepsilon_\Phi = 0.05$ and $\varepsilon_\Phi = 0.1$, respectively. As it is seen from the curves, BP and $l_{p<1}$ minimization gives almost the same reconstruction errors for low sparsity levels. Nevertheless, reconstruction error remains lower in $l_{p<1}$ minimization compared to BP for higher values of sparsity levels. Similar results can also be observed for the noisy measurements given in Fig. 4.8.



(a)

(b)



(c)

Figure 4.7: Average relative error versus sparsity level K for (a) $\varepsilon_\Phi=0$, (b) $\varepsilon_\Phi=0.05$ and (c) $\varepsilon_\Phi=0.1$.

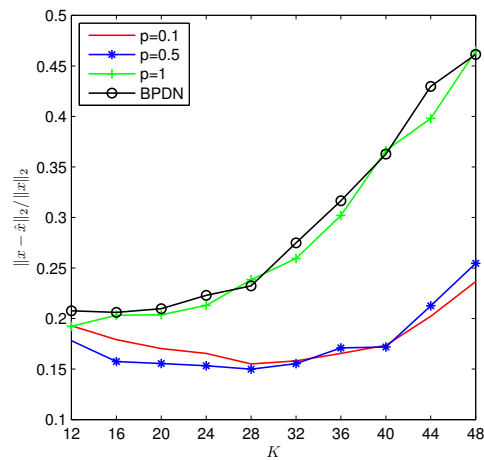


Figure 4.8: Average relative error versus sparsity level K for (a) $\varepsilon_\Phi=0.05$ and measurement noise $\sigma = 0.05$

As a last experiment, effect of p in reconstructing sparse signal x under multiplicative noise and measurement noise is investigated. Measurements are corrupted with Gaussian noise with $\sigma = 0.05$ and measurement matrix is perturbed with $\varepsilon_{\Phi} = 0.05$. Then p is increased in steps of 0.1 including $p = 0.01$ and average error is recorded and results are presented in Fig. 4.9. Clearly p values smaller than 1 has better signal estimate compared to $p = 1$. Also error increases with increased levels of ε_{Φ} .

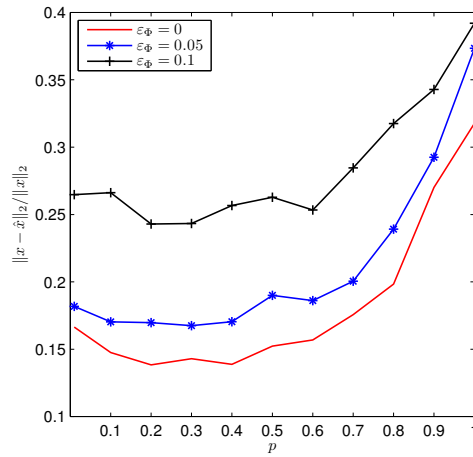


Figure 4.9: Average relative error versus p with measurement noise $\sigma = 0.05$

CHAPTER 5

CONCLUSION

In this thesis theoretical analysis and experimental results in nonconvex compressed sensing have been reported. In this chapter, a brief summary of the investigation and the results of thesis work are presented. In addition some future works are mentioned in the light of the results found.

In the third chapter a literature survey in nonconvex compressed sensing is given and a modified l_p minimization that includes partially known support in the recovery process is proposed. The theorems for modified l_p minimization are presented for both sparse and compressible signals when partial known signal support is available. Necessary conditions related with restricted isometry constant of the measurement matrix for stable and robust recovery of modified l_p minimization are derived. The derivation of coefficients determining the upper bounds of the reconstruction error are given in theorems. Numerical results show that the modification of l_p minimization improves the performance, thereby requiring fewer measurements to reconstruct the signal. The potential applications of nonconvex compressed sensing with partially known support are real-time dynamic MRI reconstruction, video imaging or compression/decompression and also reconstruction for time sequences of sparse signals.

Chapter 4 presented the general perturbation in nonconvex compressed sensing. l_p minimization for completely perturbed CS scenario is studied for both sparse and general case. The work in this chapter extends the previous results of l_p minimization by including both the additive and multiplicative noise in the measurements. Theorems reveal that under suitable conditions reconstruction error is bounded with total noise parameter, relative perturbations and RIC of matrix Φ . The error bound noise constant and error bound compressibility constant in theorems determine the worst case recovery error in the reconstruction. These constants are based on the sparsity of the signal, p value, relative perturbations and the parameter a . The multiplicative and additive noise terms are not exactly known therefore their worst case relative perturbations are used. In theorems the reconstruction errors caused from the multiplicative and additive noise terms are bounded with total noise term and best- K term approximation error of the signal. Simulation results show that reconstruction with perturbed

matrix using l_p minimization performs better than l_1 minimization in terms of average reconstruction error for different parameters such as, relative perturbation, sparsity level, measurement level and measurement noise.

The presented theorems in this thesis fulfills the requirements of nonconvex compressed sensing for l_p minimization with partially known signal support and under general perturbations. This study can be extended to compressive source separation, hyperspectral imaging, etc.

LIST OF REFERENCES

- [1] Alan V. Oppenheim and Ronald W. Schaffer, 2009. *Discrete-Time Signal Processing*. Prentice Hall Press, Upper Saddle River, NJ, USA, 3rd edition.
- [2] M. W. Marcellin and D. S. Taubman, 2001. *JPEG 2000: Image compression fundamentals, standards and practice*. Kluwer.
- [3] D.L. Donoho. (2006). Compressed sensing. *Information Theory, IEEE Transactions on*, 52 (4) 1289 –1306.
- [4] E.J. Candes and T. Tao. (2005). Decoding by linear programming. *Information Theory, IEEE Transactions on*, 51 (12) 4203 – 4215.
- [5] E.J. Candes, J. Romberg, and T. Tao. (2006). Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52 (2) 489 – 509.
- [6] E.J. Candes and T. Tao. (2006). Near-optimal signal recovery from random projections: Universal encoding strategies? *Information Theory, IEEE Transactions on*, 52 (12) 5406 –5425.
- [7] R. Chartrand. (2007). Exact reconstruction of sparse signals via nonconvex minimization. *Signal Processing Letters, IEEE*, 14 (10) 707 –710.
- [8] Shihao Ji, Ya Xue, and L. Carin. (2008). Bayesian compressive sensing. *Signal Processing, IEEE Transactions on*, 56 (6) 2346 –2356.
- [9] David L. Donoho and Philip B. Stark. (1989). Uncertainty principles and signal recovery. *SIAM J. Appl. Math.*, 49 (3) 906–931.
- [10] E Candes. (2008). The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346 (9-10) 589–592.
- [11] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. (2008). A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28 253–263.

- [12] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59 (8) 1207–1223.
- [13] S.G. Mallat and Zhifeng Zhang. (1993). Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on*, 41 (12) 3397–3415.
- [14] Joel A. Tropp, Anna, and C. Gilbert. (2007). Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inform. Theory*, 53 4655–4666.
- [15] Deanna Needell and Joel A. Tropp. (2010). Cosamp: iterative signal recovery from incomplete and inaccurate samples. *Commun. ACM*, 53 93–100.
- [16] E.J. Candes and M.B. Wakin. (2008). An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25 (2) 21–30.
- [17] M.F. Duarte, M.A. Davenport, D. Takhar, J.N. Laska, Ting Sun, K.F. Kelly, and R.G. Baraniuk. (2008). Single-pixel imaging via compressive sampling. *Signal Processing Magazine, IEEE*, 25 (2) 83–91.
- [18] D. Needell and R. Vershynin. (2010). Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. *Selected Topics in Signal Processing, IEEE Journal of*, 4 (2) 310–316.
- [19] D.L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck. (2012). Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *Information Theory, IEEE Transactions on*, 58 (2) 1094–1121.
- [20] R. von Borries, C.J. Miosso, and C. Potes. (2007). Compressed sensing using prior information. In *Computational Advances in Multi-Sensor Adaptive Processing, 2007. CAMPSAP 2007. 2nd IEEE International Workshop on*, pages 121–124.
- [21] M. Amin Khajehnejad, Weiyu Xu, A. Salman Avestimehr, and Babak Hassibi. (2009). Weighted l_1 minimization for sparse recovery with prior information. In *Proceedings of the 2009 IEEE international conference on Symposium on Information Theory - Volume 1, ISIT'09*, pages 483–487, Piscataway, NJ, USA, (2009). IEEE Press.
- [22] N. Vaswani and Wei Lu. (2010). Modified-cs: Modifying compressive sensing for problems with partially known support. *Signal Processing, IEEE Transactions on*, 58 (9) 4595–4607.

- [23] Wei Lu and N. Vaswani. (2010). Exact reconstruction conditions and error bounds for regularized modified basis pursuit (reg-modified-bp). In *Signals, Systems and Computers (ASILOMAR), 2010 Conference Record of the Forty Fourth Asilomar Conference on*, pages 763–767.
- [24] Laurent Jacques. (2010). A short note on compressed sensing with partially known signal support. *Signal Processing*, 90 (12) 3308–3312.
- [25] M.P. Friedlander, H. Mansour, R. Saab, and O. Yilmaz. (2012). Recovering compressively sampled signals using partial support information. *Information Theory, IEEE Transactions on*, 58 (2) 1122–1134.
- [26] Mila Nikolova. (2005). Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares. *Multiscale Modeling and Simulation*, 4 (3) 960–991.
- [27] Mila Nikolova, Michael K. Ng, Shuqin Zhang, and Wai-Ki Ching. (2008). Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization. *SIAM Journal on Imaging Sciences*, 1 (1) 2–25.
- [28] M. Nikolova, M.K. Ng, and Chi-Pan Tam. (2010). Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction. *Image Processing, IEEE Transactions on*, 19 (12) 3073–3088.
- [29] G. Gasso, A. Rakotomamonjy, and S. Canu. (2009). Recovering sparse signals with a certain family of nonconvex penalties and dc programming. *Signal Processing, IEEE Transactions on*, 57 (12) 4686–4698.
- [30] Jianqing Fan and Li R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96 1348–1360.
- [31] Tong Zhang. 2009. Some sharp performance bounds for least squares regression with L1 regularization. *Ann. Statist.*, 37 (5A) 2109–2144.
- [32] B.D. Rao and K. Kreutz-Delgado. (1999). An affine scaling methodology for best basis selection. *Signal Processing, IEEE Transactions on*, 47 (1) 187–200.
- [33] R. Chartrand and Wotao Yin. (2008). Iteratively reweighted algorithms for compressive sensing. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 3869–3872.

- [34] R. Saab, R. Chartrand, and O. Yilmaz. (2008). Stable sparse approximations via nonconvex optimization. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 3885 –3888.
- [35] Rayan Saab and Ozgur Yilmaz. (2010). Sparse recovery by non-convex optimization instance optimality. *Applied and Computational Harmonic Analysis*, 29 (1) 30 – 48.
- [36] Rick Chartrand and Valentina Staneva. (2008). Restricted isometry properties and nonconvex compressive sensing. *Inverse Problems*, 24.
- [37] R. Robucci, Leung Kin Chiu, J. Gray, J. Romberg, P. Hasler, and D. Anderson. (2008). Compressive sensing on a cmos separable transform image sensor. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, pages 5125 –5128.
- [38] Stephane Mallat, 2008. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*. Academic Press, 3rd edition.
- [39] Taner Ince, Arif Nacaroglu, and Nurdal Watsuji. (2012). Nonconvex compressed sensing with partially known signal support. *Signal Processing*, (0) –.
- [40] M.A. Herman and T. Strohmer. (2009). High-resolution radar via compressed sensing. *Signal Processing, IEEE Transactions on*, 57 (6) 2275 – 2284.
- [41] Thomas Blumensath and Mike Davies. (2007). Compressed sensing and source separation. In *in International Conference on Independent Component Analysis and Blind Source Separation*.
- [42] M.A. Herman and T. Strohmer. (2010). General deviants: An analysis of perturbations in compressed sensing. *Selected Topics in Signal Processing, IEEE Journal of*, 4 (2) 342 –349.
- [43] Jie Ding, Laming Chen, and Yuantao Gu. (2011). Performance analysis of orthogonal matching pursuit under general perturbations. *CoRR*, abs/1106.3373.
- [44] van and M. P. Friedlander. SPGL1: A solver for large-scale sparse reconstruction. (2007).

APPENDIX A

DERIVATION OF IRLS

Define the Lagrangian

$$L(x, \lambda) = E^p(x) + \lambda^T(\Phi x - y) \quad (\text{A.1})$$

where

$$E^p = \|x\|_p^p = \sum_{i=1}^N (x(i))^p \quad 0 \leq p \leq 1 \quad (\text{A.2})$$

Stationary points of Lagrangian are

$$\nabla_x L(x_*, \lambda_*) = \nabla_x E^p(x_*) + \Phi^T \lambda_* = 0 \quad (\text{A.3})$$

$$\nabla_\lambda L(x_*, \lambda_*) = \Phi x_* - y \quad (\text{A.4})$$

then

$$\nabla_x E^p(x_*) = |p|Q(x)x \quad (\text{A.5})$$

where $Q(x) = \text{diag}(|x_i|^{p-2})$. Substituting (A.5) into (A.3)

$$|p|Q(x_*)x_* + \Phi^T \lambda_* - \Phi x_* - y = 0 \quad (\text{A.6})$$

From this

$$x_* = -\frac{1}{p}Q^{-1}(x_*)\Phi^T \lambda_* \quad (\text{A.7})$$

Substituting (A.7) into the first equation of (A.6) and solving for λ_*

$$\lambda_* = -|p|(\Phi Q^{-1}(x_*)\Phi^T)^{-1}y \quad (\text{A.8})$$

Then x_* in (A.7) becomes

$$x_* = Q^{-1}(x)\Phi^T(\Phi Q^{-1}(x_*)\Phi^T)^{-1}y \quad (\text{A.9})$$

in iterative manner

$$x^{n+1} = Q_n^{-1}\Phi^T(\Phi Q_n^{-1}\Phi^T)^{-1}y \quad (\text{A.10})$$

where Q_n^{-1} is diagonal matrix with elements $w_i = |x^{(n-1)}|^{2-p}$

APPENDIX B

CODES FOR SIMULATIONS IN CHAPTER 3

B.1 MATLAB CODES

B.1.1 IRLS

Iteratively reweighted least squares (IRLS)

```
function [u,flag] = irls1d(Phi,b,p)
% Inputs:
%      Phi: forward matrix from sources to data
%      b: data vector
%      p: power, no more than 1
% Outputs:
%      u: the reconstructed solution of Phi * u = b
%      flag: 0 - success; 1 - max # itr reached
%

epsilon = 1; % initial epsilon
eps_min = 1e-8; % minimally allowed epsilon
max_itr = 1000; % max # of iterations allowed

%% preparation
PhiT = Phi.';
n = size(Phi,2);

%% initialization
method = 3;
switch method
    case 1
        % method 1: MATLAB mldivide solution, basis solution
        u = Phi\b;
    case 2
        % method 2: least squares solution
        [u,flag_tmp] = lsqr(Phi,b);
    case 3
        % method 3: u = ones(n,1);
```

```

    u = ones(n,1);
end

%% main iterations
for itr = 1 : max_itr'
    u_prev = u;

    %% main update
    Q = spdiags( (u.*u + epsilon).^( 1 - p / 2 ), 0, n, n);
    method = 1;
    switch method
        case 1
            % method 1: MATLAB mldivide
            u = Q*PhiT*((Phi*Q*PhiT)\b);
        case 2
            % method 2: wpcg, a correct version of MATLAB's...
            % pcg by Zaiwen Wen
            [y,flag_tmp] = wpcg(Phi*Q*PhiT,b,1e-8,100);
            u=Q*PhiT*y;
    end

    %% update epsilon
    if norm(u-u_prev)<sqrt(epsilon)/100
        if epsilon < eps_min - 1e-15; break; end
        epsilon = epsilon / 10;
    end

end

%% post-process
if itr == max_itr
    fprintf('IRLS : max # itr %i reached\n',max_itr);
    flag=1; % flag of max # itr reached
else
    fprintf('IRLS : normal stop, itr=%i last_eps=%4.2e\n'...
        ,itr,epsilon);
    flag=0; % flag of success
end

end
end

```

B.1.2 IRLS with partially known support

```
function [u,flag] = irls_pks(Phi,b,p,T0)
% Inputs:
%     Phi: forward matrix from sources to data
%     b: data vector
%     p: power, no more than 1
%     T0: partially known support
% Outputs:
%     u: the reconstructed solution of Phi * u = b
%     flag: 0 - success; 1 - max # itr reached
%

epsilon = 1; % initial epsilon
eps_min = 1e-8; % minimally allowed epsilon
max_itr = 1000; % max # of iterations allowed

%% preparation
PhiT = Phi';
n = size(Phi,2);

%% initialization
method = 3;
switch method
    case 1
        % method 1: MATLAB mldivide solution, basis solution
        u = Phi\b;
    case 2
        % method 2: least squares solution
        [u,flag_tmp] = lsqr(Phi,b); %#ok<NASGU>
    case 3
        u = ones(n,1);

%% main iterations
for itr = 1 : max_itr'
    u_prev = u;

    %% main update
    if T0~=0;
        Q = spdiags( (u.*u + epsilon^2).^( 1 - p / 2 ), 0, n, n);
        for i=1:size(T0,2)
            Q(T0(i),T0(i))=10^4;
        end
    else if T0==0
```

```

        Q = spdiags( (u.*u + epsilon^2).^( 1 - p / 2 ), 0, n, n);
    end
end

method = 1;
switch method
    case 1
        % method 1: MATLAB mldivide
        u = Q*PhiT*((Phi*Q*PhiT)\b);

    case 2
        % method 2: wpcg, a correct version of MATLAB's pcg ...
        %by Zaiwen Wen
        [y,flag_tmp] = wpcg(Phi*Q*PhiT,b,1e-8,100); %#ok<NASGU>
        u=Q*PhiT*y;
    end

%% update epsilon
if norm(u-u_prev)<sqrt(epsilon)/100
    if epsilon < eps_min - 1e-15; break; end
    epsilon = epsilon / 10;
end

end

%% post-process
if itr == max_itr
%    fprintf('IRLS : max # itr %i reached\n',max_itr);
    flag=1; % flag of max # itr reached
else
%    fprintf('IRLS : normal stop, itr=%i last_eps=%4.2e\n'...
%,itr,epsilon);
    flag=0; % flag of success
end

end
end

```

B.1.3 Script for Simulation Results (Sparse Case)

```

p=[0.01 0.1:0.1:1];
sVals=40; % Sparsity levels
mVals=60:20:200; %Measurement levels
dVals=512; %dimension
numTrials=20; %Number of trials per parameter set

```

```

%partially known support percentage
% percentage=0:0.25:1;
Ti=sVals;
%Set Variable lengths and Data Collection
nump=length(p);
% per=length(percentage);
nums=length(sVals);
numm=length(mVals);
numd=length(dVals);
error = zeros(nump, numm, numd);
numCorrect_lp = zeros(numm, Ti, nump);
numCorrect_modcs= zeros(numm, Ti);

error_bp=zeros(numm, Ti, numTrials);
error_lp = zeros(numm, Ti, nump, numTrials);
error_lp_ppower = zeros(numm, Ti, nump, numTrials);

s = sVals;

d = dVals;
z = randperm(d);
x = zeros(d, 1);
x(z(1:sVals)) = (randn(sVals,1));

T=z(1:sVals);
Ti=length(T);

for im=1:numm
for Tindex=4:4:Ti
for pp=1:nump
for trial=1:numTrials

phi = randn(mVals(im), d)/sqrt(mVals(im));
y = phi*x;

if pp==1
xhat_modcs=Modifiedcs_static(phi, y, T(1:Tindex));
xhat_lp_pks=irls_pks(phi, y, p(pp), T(1:Tindex));
error_lp(im, Tindex, pp, trial) = norm(xhat_lp_pks-x)/norm(x);
error_lp_ppower(im, Tindex, pp, trial) = (norm(xhat_lp_pks-x))^p(pp);
error_bp(im, Tindex, trial) = norm(xhat_modcs-x)/norm(x);

if norm(xhat_lp_pks(1:dVals)-x) <= 10^(-3)
numCorrect_lp(im, Tindex, pp) = numCorrect_lp(im, Tindex, pp) + 1;
end

```

```

if norm(xhat_modcs(1:dVals)-x) <= 10^(-3)
numCorrect_modcs(im,Tindex) = numCorrect_modcs(im,Tindex) + 1;
end

elseif pp~=1

xhat_lp_pks=irls_pks(phi,y,p(pp),T(1:Tindex));
error_lp_ppower(im,Tindex,pp,trial) = (norm(xhat_lp_pks-x))^p(pp);
error_lp(im,Tindex,pp,trial) = norm(xhat_lp_pks-x)/norm(x);

if norm(xhat_lp_pks(1:dVals)-x) <= 10^(-3)
numCorrect_lp(im,Tindex,pp) = numCorrect_lp(im,Tindex,pp) + 1;
end
end
fid=fopen('result.txt','wt');
fprintf(fid,'im %d\nTindex %d\npp %d\ntrial %d\n',im,Tindex,pp,trial);
fclose(fid);

end

end
end
end

```

B.1.4 Script for Simulation Results (Compressible case)

```

p=[0.01 0.1:0.1:0.9 0.95 1];
sVals=40; % Sparsity levels
mVals=60:10:200; %Measurement levels
dVals=512; %dimension
numTrials=50; %Number of trials per parameter set

%partially known support percentage
Ti=sVals;
%Set Variable lengths and Data Collection
nump=length(p);
% per=length(percentage);
numm=length(sVals);
nummm=length(mVals);
numd=length(dVals);
error = zeros(nump,nummm, numd);
numCorrect_lp = zeros(nummm,Ti,nump);
numCorrect_modcs= zeros(nummm,Ti);

```

```

error_bp=zeros(numm,Ti,numTrials);
error_lp = zeros(numm,Ti,nump,numTrials);
error_lp_ppower = zeros(numm,Ti,nump,numTrials);

s = sVals;

d = dVals;
z = randperm(d);
x = zeros(d, 1);
x=(z.^-1.5)';

[xx, index]=sort(x,'descend');

T=[(index(1:40))];
Ti=length(T);

for im=1:numm
for Tindex=4:4:Ti
for pp=1:nump
for trial=1:numTrials

phi = randn(mVals(im),d)/sqrt(mVals(im));
y = phi*x;

if pp==1
xhat_modcs=Modifiedcs_static(phi,y,T(1:Tindex));
xhat_lp_pks=irls_pks(phi,y,p(pp),T(1:Tindex));
error_lp(im,Tindex,pp,trial) = norm(xhat_lp_pks-x)/norm(x);
error_lp_ppower(im,Tindex,pp,trial) = (norm(xhat_lp_pks-x))^p(pp);
error_bp(im,Tindex,trial) = norm(xhat_modcs-x)/norm(x);

if norm(xhat_lp_pks(1:dVals)-x) <= 10^(-3)
numCorrect_lp(im,Tindex,pp) = numCorrect_lp(im,Tindex,pp) + 1;
end

if norm(xhat_modcs(1:dVals)-x) <= 10^(-3)
numCorrect_modcs(im,Tindex) = numCorrect_modcs(im,Tindex) + 1;
end

elseif pp~=1

xhat_lp_pks=irls_pks(phi,y,p(pp),T(1:Tindex));
error_lp_ppower(im,Tindex,pp,trial) = (norm(xhat_lp_pks-x))^p(pp);
error_lp(im,Tindex,pp,trial) = norm(xhat_lp_pks-x)/norm(x);

```



```
if norm(xhat_lp_pks(1:dVals)-x) <= 10^(-3)
numCorrect_lp(im,Tindex,pp) = numCorrect_lp(im,Tindex,pp) + 1;
end

end

fid=fopen('result.txt','wt');
fprintf(fid,'im %d\nTindex %d\npp %d\ntrial %d\n',im,Tindex,pp,trial);
fclose(fid);

end

end
end
end
```

APPENDIX C

CODES FOR SIMULATIONS IN CHAPTER 4

C.1 MATLAB CODES

C.1.1 Noisy IRLS

```
function [u,flag] = irls1d_mu(Phi,b,p,lambda)
    % Inputs:
    %     Phi: forward matrix from sources to data
    %     b: data vector
    %     p: power, no more than 1
    % Outputs:
    %     u: the reconstructed solution of Phi * u = b
    %     flag: 0 - success; 1 - max # itr reached

    epsilon = 1; % initial epsilon
    eps_min = 1e-8; % minimally allowed epsilon
    max_itr = 1000; % max # of iterations allowed

    %% preparation
    PhiT = Phi.';
    n = size(Phi,2);

    %% initialization
    method = 1;
    switch method
        case 1
            % method 1: MATLAB mldivide solution, basis solution
            u = Phi\b;
        case 2
            % method 2: least squares solution
            [u,flag_tmp] = lsqr(Phi,b);
        case 3
            % method 3: u = ones(n,1);
            u = ones(n,1);
    end
```

```

%% main iterations
for itr = 1 : max_itr'
    u_prev = u;

    %% main update
    Q = spdiags( p*lambda*(u.*u + epsilon).^( p/2 -1 ), 0, n, n);
    method = 1;
    switch method
        case 1
            % method 1: MATLAB mldivide
            %           u = Q*PhiT*((Phi*Q*PhiT)\b);
            u=(PhiT*Phi+Q)\(PhiT*b);

        case 2
            % method 2: wpcg, a correct version of MATLAB's pcg ...
            %by Zaiwen Wen
            [y,flag_tmp] = wpcg(Phi*Q*PhiT,b,1e-8,100);
            u=Q*PhiT*y;
    end

    %% update epsilon
    if norm(u-u_prev)<sqrt(epsilon)/100
        if epsilon < eps_min - 1e-15;
            break;
        end
        epsilon = epsilon / 10;
    end

end

%% post-process
if itr == max_itr
    %   fprintf('IRLS : max # itr %i reached\n',max_itr);
    flag=1; % flag of max # itr reached
else
    %   fprintf('IRLS : normal stop, itr=%i last_eps=%4.2e\n'...
    %,itr,epsilon);
    flag=0; % flag of success
end

end

```

C.1.2 Script for varying sparsity level K

```
sVals=2:2:50; % Sparsity levels
mVals=[80 120 160 200]; %Measurement levels
n=512; %ambient dimension
numTrials=100; %Number of trials per parameter set

perturbation_epsA=[0 0.05 0.1];
p=[0.01 0.1 0.5 0.9 1];

pert=length(perturbation_epsA);
nums=length(sVals);
numm=length(mVals);
numd=length(n);
nump=length(p);

error_irlsld = zeros (pert, nums, numm, nump);
error_bp = zeros (pert, nums, numm);

out_irlsld_signal=zeros (n, pert, nums, numm, nump);
out_bp_signal=zeros (n, pert, nums, numm, nump);
signal=zeros (n, pert, nums, numm, nump);

for ipert=1:pert
    for is=1:nums
        for im=1:numm
            s = sVals(is);
            m = mVals(im);
            for pi=1:nump
                for trial=1:numTrials
                    A = randn (m, n) /sqrt (m);
                    F=randn (m, n);
                    spettr_F=sqrt (max (eig (F*F')));
                    spettr_A=sqrt (max (eig (A*A')));
                    F=F/spettr_F;
                    E=F*perturbation_epsA(ipert)*spettr_A;
                    phi=A+E;
                    z = randperm(n);
                    x = zeros (n, 1);
                    x(z(1:s)) =sign(randn(s, 1));
                    y=A*x;
                    %x01 = phi'*y;
                    xhat_irls = irlsld(phi, y, p(pi));
                    error_irlsld(ipert, is, im, pi, trial) =...
```

```

        norm(xhat_irls-x)/norm(x);

    if pi==1
        opts = spgSetParms('verbosity',0);
        xhat_bp = spg_bp(phi,y,opts);
        error_bp(ipert,is,im,trial) = norm(xhat_bp-x)/norm(x);

    end

    fid=fopen('result.txt','wt');
    fprintf(fid,'ipert %d\nis %d\nim %d\npi %d\ntrial %d\n'...
        ,ipert,is,im,pi,trial);
    fclose(fid);
    clc

end
end
end
end
end
end

```

C.1.3 Script for varying sparsity level $\varepsilon_{\Phi,K,y}$

```

sVals=10:10:30; % Sparsity levels
mVals=128; %Measurement levels
n=512; %ambient dimension
numTrials=100; %Number of trials per parameter set

perturbation_epsA=0:0.02:0.2;
% perturbation_epsA=[0 0.05 0.1 0.2];
% p=[0.1 0.5 0.9 1];
p=[0.01 0.1:0.1:0.9 1];

pert=length(perturbation_epsA);
nums=length(sVals);
numm=length(mVals);
numd=length(n);
nump=length(p);

for ipert=1:pert
    for is=1:nums
        for im=1:numm
            s = sVals(is);
            m = mVals(im);

```

```

for pi=1:nump
    for trial=1:numTrials
        A = randn(m,n)/sqrt(m);
        F=randn(m,n);
        spettr_F=sqrt(max(eig(F*F')));
        spettr_A=sqrt(max(eig(A*A')));
        F=F/spettr_F;
        E=F*perturbation_epsA(ipert)*spettr_A;
        phi=A+E;
        z = randperm(n);
        x = zeros(n, 1);
        x(z(1:s)) =randn(s,1);
        y=A*x;
        %x01 = phi'*y;
        xhat_irls = irls1d(phi,y,p(pi));
        error_irls1d(ipert,is,im,pi,trial) =...
            norm(xhat_irls-x)/norm(x);

        if pi==1
            opts = spgSetParms('verbosity',0);
            xhat_bp = spg_bp(phi,y,opts);
            error_bp(ipert,is,im,trial) = norm(xhat_bp-x)/norm(x);

        end

        end

        fid=fopen('result.txt','wt');
        fprintf(fid,'ipert %d\nis %d\nim %d\npi %d\ntrial %d\n'...
            ,ipert,is,im,pi,trial);
        fclose(fid);

    end

end

end

end

end

```

C.1.4 Script for varying measurement level under additive Gaussian noise

```

sVals=40; % Sparsity levels
mVals=80:10:200; %Measurement levels
n=512; %ambient dimension
numTrials=100; %Number of trials per parameter set

```

```

perturbation_epsA=0.05;
p=[0.1 0.5 1];

noise_sigma=[0.01 0.05];
lambda=1e-2;

pert=length(perturbation_epsA);
nums=length(sVals);
numm=length(mVals);
numd=length(n);
nump=length(p);

for ipert=1:pert
    for isigma=1:length(noise_sigma)
        for im=1:numm
            s = sVals;
            m = mVals(im);
            for pi=1:nump
                for trial=1:numTrials
                    A = randn(m,n)/sqrt(m);
                    F=randn(m,n);
                    spettr_F=sqrt(max(eig(F*F')));
                    spettr_A=sqrt(max(eig(A*A')));
                    F=F/spettr_F;
                    E=F*perturbation_epsA(ipert)*spettr_A;
                    phi=A+E;

                    e = noise_sigma(isigma)*randn(m,1);
                    z = randperm(n);
                    x = zeros(n, 1);
                    x(z(1:sVals)) =(randn(sVals,1));
                    y=A*x+e;

                    %x01 = phi'*y;
                    xhat_irls = irls1d_mu(phi,y,p(pi),lambda);
                    error_irls1d(ipert,isigma,im,pi,trial) = ...
                    norm(xhat_irls-x)/norm(x);

                    if pi==1
                        sigma=0.01;
                        opts = spgSetParams('verbosity',0);
                        xhat_bpdn = spg_bpdn(phi,y,sigma,opts);
                        error_bpdn(ipert,isigma,im,trial) =...
                        norm(xhat_bpdn-x)/norm(x);
                    end
                end
            end
        end
    end
end

```

```
fid=fopen('result.txt','wt');
fprintf(fid,'ipert %d\nisigma %d\nim %d\npi %d\ntrial %d\n'...
,ipert,isigma,im,pi,trial);
fclose(fid);
    clc
    end
end
end
end
end
```


PUBLICATIONS

- [1] T. Ince, A. Nacaroglu, N.Watsuji, "Nonconvex compressed sensing with partially known signal support", *Signal Processing*.
- [2] T. Ince, A. Nacaroglu, N.Watsuji, "Nonconvex Compressed Sensing for General Perturbations", *Digital Signal Processing*, (under review, revised).
- [3] T. Ince, A. Nacaroglu, "On the Perturbation of Measurement Matrix in Nonconvex Compressed Sensing", *Applied and Computational Harmonic Analysis*, (Under review).
- [4] T. Ince, A. Nacaroglu, N. Watsuji, "Nonconvex compressed sensing with partially known support" *IEEE 20th Signal Processing and Communications Applications Conference*, April 2012.
- [5] T. Ince, A. Nacaroglu, N. Watsuji, "Iteratively Reweighted Least Squares Minimization For Sparsely Corrupted Measurements" *IEEE 19th Signal Processing and Communications Applications Conference*, April 2011.
- [6] T. Ince, A. Nacaroglu, N. Watsuji, "Performance of Analog to Information Converter Under Noisy Conditions" *IEEE 18th Signal Processing and Communications Applications Conference*, April 2010.

CIRRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: INCE, Taner

Nationality: Turkish (TC)

Date and Place of Birth: 25 April 1980, Germany

Phone: +90 342 3172143

email:tince@gantep.edu.tr

EDUCATION

Degree	Institution	Year of Graduation
MS	University of Gaziantep	2006
BS	University of Gaziantep	2003

WORK EXPERIENCE

Year	Place	Enrollment
2004-present	University of Gaziantep	Research Assistant