2013

On Modified l_1-Minimization Problems in Compressed Sensing

Man Bahadur Basnet
Iowa State University

On modified ℓ-one minimization problems in compressed sensing

by

Man Bahadur Basnet

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Mathematics (Applied Mathematics)

Program of Study Committee:

Fritz Keinert, Co-major Professor

Namrata Vaswani, Co-major Professor

Eric Weber

Jennifer Davidson

Alexander Roitershtein

Iowa State University

Ames, Iowa

2013

DEDICATION

I would like to dedicate this thesis to my parents Jaya and Chandra and to my family (wife Sita, son Manas and daughter Manasi) without whose support, encouragement and love, I would not have been able to complete this work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I thank my advisors Professor Fritz Keinert and Professor Namrata Vaswani for their guidance and support throughout my PhD study at Iowa State University. My advisors' combined passion and capability for theoretical as well as applied research over a broad spectrum of fields have been, and will always be, a great inspiration to me. My special thanks also goes to Professor Keinert who introduced me to this beautiful and new field of research called compressed sensing, and Professor Vaswani for introducing me to the problems studied in this thesis. It is a great pleasure and privilege for me to have worked together with them over the past several years. They were always available to guide and share the wisdom with me whenever I was stuck doing the problem.

It was, of course, difficult at times trying to raise a family of four, supporting parents back home and fulfilling the necessary duties and coursework in the department while still not getting behind doing my research. I admire my advisors for their patience and belief in me during this and several other unintended and uninvited incidents that occurred during my PhD tenure here at Iowa State University.

Thank you to Professor Jennifer Davidson and Professor Eric Weber for their guidance on several occasions and also for serving in my POS committee. I would also like to thank and welcome professor Alex Roitershtein for agreeing to be in my POS committee. My thanks also goes to Professor Clifford Bergman, Professor Wolfgang Kliemann and Dr. Heather Bolles for their valuable career advice and help preparing my job application documents. I also owe a debt of tremendous gratitude to our graduate secretary and record analyst Melanie Erickson for answering every administerial question I encountered during my time at ISU.

Finally, my special thanks go to my family for their patience and unreserved support, which are the cornerstones of my PhD life.

ABSTRACT

Sparse signal modeling has received much attention recently because of its application in medical imaging, group testing and radar technology, among others. Compressed sensing, a recently coined term, has showed us, both in theory and practice, that various signals of interest which are sparse or approximately sparse can be efficiently recovered by using far fewer samples than suggested by Shannon sampling theorem.

Sparsity is the only prior information about an unknown signal assumed in traditional compressed sensing techniques. But in many applications, other kinds of prior information are also available, such as partial knowledge of the support, tree structure of signal and clustering of large coefficients around a small set of coefficients.

In this thesis, we consider compressed sensing problems with prior information on the support of the signal, together with sparsity. We modify regular ℓ_1-minimization problems considered in compressed sensing, using this extra information. We call these *modified ℓ_1-minimization problems*.

We show that partial knowledge of the support helps us to weaken sufficient conditions for the recovery of sparse signals using modified ℓ_1 minimization problems. In case of deterministic compressed sensing, we show that a sharp condition for sparse recovery can be improved using modified ℓ_1 minimization problems. We also derive algebraic necessary and sufficient condition for modified basis pursuit problem and use an open source algorithm known as ℓ_1-homotopy algorithm to perform some numerical experiments and compare the performance of modified Basis Pursuit Denoising with the regular Basis Pursuit Denoising.

CHAPTER 1. INTRODUCTION

1.1 Outline of the Thesis

In chapter 1, we will discuss some background in sampling, sparse recovery, compressed sensing (CS), and state our research problem. In sections 2, 3 and 4 of this chapter, we will define several important concepts such as sparsity, compressibility, coherence, spark, and discuss their importance in CS in general and in this thesis. Compressed sensing and the relevant results will be briefly explained in section 5. In section 6, we explain the motivation behind our research problem and state several research problems considered in this thesis. We introduce problems Basis Pursuit (BP), Basis Pursuit Denoising (BPDN) and Dantzig Selector (DS) as well as their modified versions.

Chapter 2 consists of several sufficient conditions for the uniqueness of solutions of optimization problems mod-BP, mod-BPDN and mod-DS. In section 1, we consider conditions based on restricted isometries. In the second part of chapter 2, we will use coherence of a matrix for the recovery. Random matrices are primarily used for the CS recovery problems, as they have proved to be better than deterministic matrices. In a situation where random measurements are not possible, we have to rely on deterministic matrices, and having coherence as a measure of the quality of a measurement matrix becomes important. We prove that knowing part of the support beforehand helps not only to weaken the sufficient conditions, but it also helps to break the coherence barrier.

In chapter 3, we derive dual and Karush Kuhn Tucker (KKT) optimality conditions for modified ℓ_1- minimization problems. We also prove an algebraic necessary and sufficient conditions for the uniqueness of solution of mod-BP using theory of convex optimization. Chapter 4 will have numerical experiments using an open source solver called ℓ_1 homotopy. We solve mod-BPDN numerically using this solver and compare the performance of BPDN and mod-BPDN. Chapter 5 consists of summary of results, future research projects and the conclusion of the thesis. Finally, appendix A will have MATLAB codes we used to solve mod-BPDN and the bibliography will list all the citations used in this thesis.

1.2 Sampling

The world is analog but information is digital. Because computers process only digital data, we need a medium to go between these two worlds. Thus a signal $x(t)$, which is continuous, becomes a string of numbers through sampling.

$$x(t) \xrightarrow{\text{Sampling}} x[n]$$

This then leads to some important questions: (a) how many samples should we take to be able to reconstruct the original signal, and (b) how should we go about reconstructing the original signal ? This question was answered by Shannon in 1948 and also by many others in different forms and times, some as early as 1915. The following theorem, famously known as Shannon's Sampling Theorem [56], answers both questions.

Theorem 1.2.1. *If a continuous function (signal) $f(t)$ contains no frequencies higher than $f_{\max}$, it is completely determined by giving its ordinates at a series of points spaced by $T \leq \frac{\pi}{f_{\max}}$ and*

$$f(t) = \sum_{k \in Z} f(kT) \operatorname{sinc}\left(\frac{t}{T} - k\right)$$

where $\operatorname{sinc} t = \frac{\sin \pi t}{\pi t}$.

Here $f_{\max} = \frac{\pi}{T}$ is called the Nyquist frequency. Although Shannon's sampling theorem is an elegant piece of mathematical work and has played a crucial role in signal processing [61], it has some flaws:

1. Sometimes it requires too many samples. If the Fourier Transform of a signal has a large support, then it requires too many samples, which will be a problem to process in a computer.

2. It only applies to band-limited signals. No real world signals are band-limited. If we assume that a function is band-limited, then it is entire. Such signals cannot start and stop and hence have to go on forever in time domain. But in fact real signals start and stop and therefore cannot be band-limited. This means that no system that samples data from the real world can do so properly, as you would have to wait an infinite amount of time for your results.

3. For the reconstruction, most of these theorems use the sinc function, because exact interpolation with an ideal low pass filter can be performed with a sinc function. But the sinc function has a considerable amount of energy in an extended interval, and hence it has a much slower rate of convergence. So, in practice, the sinc function is used as a heuristic to get another better interpolating function.

1.3 Sparsity and Compressibility

Sparse signals (having only a few nonzero entries) are of great importance in signal processing for the purpose of compression. The criteria of sparsity has also been used in deconvolution, machine learning, and regularization. Mathematically, a discrete signal $x \in \mathbb{R}^p$ is said to be *s-sparse*, where $s \le p$ and s is an integer, if the support of x, defined by $supp(x) = \{i : x_i \neq 0\}$ contains at most s indices.

Sparsity may be in a canonical basis or transformed basis or a combination of bases. A set of vectors that can be used to form linear combinations to represent other vectors is called a *dictionary*. Thus, a signal $y \in \mathbb{R}^m$ which is sparse in a dictionary represented by an $m \times p$ matrix A can be written as $y = Ax$ with x sparse.

There are many examples of real world sparse signals. For example, a radar signal at an airport has typically zero entries, except for a few spikes locating true positions and velocity of nearby aircraft. A signal coming from a musical instrument may not be sparse, but its Fourier transform may be sparse. So the assumption of sparsity prior is not just a theoretical phenomenon.

Of course not all natural and man-made signals are sparse, but in many situations they can be approximated by sparse signals. Such signals are called *compressible*. A signal $x \in \mathbb{R}^p$ is said to be compressible if the magnitudes of coefficients of x sorted in decreasing order follow the power law $\left|x_{I(i)}\right| \leq C i^{\frac{-1}{r}}, \quad i = 1, 2, 3, .., p$, where r is a constant and $I(i)$ are the indices for the sorted components of x.

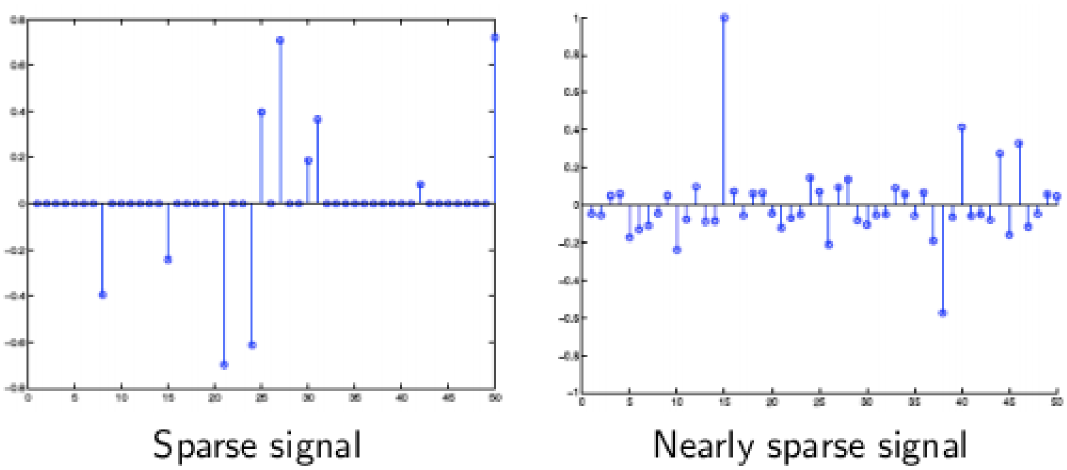


Figure 1.3.1 Sparse and compressible signal

A sparse approximation of such an x with k components is called a *k-term approximation of x* and is denoted by $x_{\max(k)}$. The set of all such $x_{\max(k)}$ is denoted by $\sum_k$. Notice that $\sum_k$ is not a subspace of $\mathbb{R}^p$; it is a union of subspaces in $\mathbb{R}^p$. The image in Figure 1.3.2 below is $\sum_2$ in $\mathbb{R}^3$.

In case of a k-term approximation of x, the error in the given ℓ_q-norm is defined as follows.

$$\sigma_k(x)_q = \min_{\bar{x} \in \sum_K} \|x - \bar{x}\|_q = \|x - x_{\max(k)}\|_q$$

where

$$\|x\|_q = \left(\sum_{i=1}^{p} |x_i|^q \right)^{\frac{1}{q}}$$

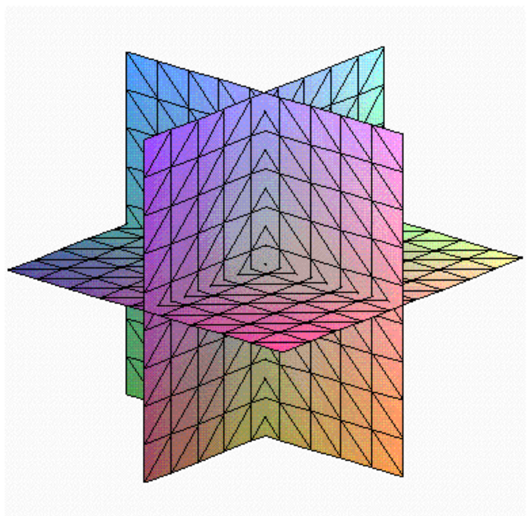


Figure 1.3.2 The set $\sum_2$ of 2-sparse vectors in $\mathbb{R}^3$

The main problems associated with sparse/compressible linear models are the following.

1. Given a signal $x \in \mathbb{R}^p$, how do we find a sparse/compressible representation of x in some basis or in a dictionary D, say $x = D\alpha$?

2. Once we know that x has a such a representation, how do we recover it by measuring it with a matrix A from the linear model $y = AD\alpha$?

We start with a simple model of a signal as a p-tuple $x \in \mathbb{R}^p$ where p is large, and we assume that x is s-sparse, where $s \ll p$. To get information about x we need to sample

x by a matrix say $A = [a_1, \ldots, a_p]$ where $a_1 \ldots, a_p \in \mathbb{R}^m$ are column vectors. Denoting each sample by $y_i = r_i' x$ where r_i' are the rows of A, we can write x as

$$y = \sum_{i=1}^{p} x_i a_i = Ax$$

Ideally, we want to solve this underdetermined system of equations and find x that has the fewest number of non-zeros. The best measure of sparsity is the ℓ_0-quasinorm defined as $\|x\|_0 = \#\{i : x_i \neq 0\}$. The search of sparse x then amounts to solving an optimization problem

$$\underset{x \in \mathbb{R}^p}{\text{minimize}} \ \|x\|_0 \qquad \text{such that} \qquad y = Ax \tag{P_0}$$

That is, among infinitely many x's that satisfy $y = Ax$, we choose the one that has the fewest nonzero components. While ℓ_0-quasinorm gives the best measure for sparsity, it has some flaws. First, ℓ_0 is not a norm. It does not satisfy the homogeneity properly, namely $\|\alpha x\|_0 \neq |\alpha| \|x\|_0$; but also $\|\cdot\|_0$ is not convex. Further, it has been shown [52] that the search for such an x is an NP hard problem.

To avoid these problems, we look for another measure for sparsity which is convex but still induces sparsity of x as close as possible to $\|x\|_0$-quasinorm. The following optimization problem, known as Basis Pursuit [23], does exactly that.

$$\underset{x \in \mathbb{R}^p}{\text{minimize}} \ \|x\|_1 \qquad \text{such that} \qquad y = Ax \tag{BP}$$

Here, $\|x\|_1 = \sum_{i=1}^{p} |x_i|$. Although basis pursuit was studied on or after 1995, the ℓ_1 norm appeared in the literature as early as 1907. The following figure 1.3.3 describes why the ℓ_1 norm promotes sparsity. As we can see from the figure, ℓ_1 norm is the closest norm that makes the problem convex and still preserves the sparsity of the solution among others.

In most practical situations, the true measurements are not possible, so we consider the stable sparse recovery based on the assumption that the measurements are noisy, i.e., $y = Ax + w$ where w is some noise with $\|w\|_2 \leq \epsilon$. If a sparse solution is found, then

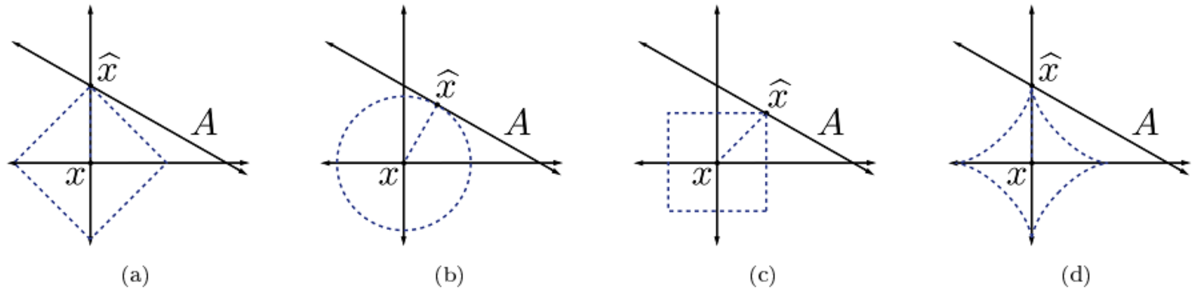


Figure 1.3.3 ℓ_q balls for $q = 1, q = 2, q = \infty$ and $q = \frac{1}{2}$

all other solutions with some sparsity lie very close to it. We state here the optimization problems for noisy cases for the reference.

$$\underset{x \in \mathbb{R}^p}{\text{minimize}} \; \|x\|_0 \qquad \text{such that} \qquad \|y - Ax\|_2 \leq \epsilon \tag{P_0^ϵ}$$

$$\underset{x \in \mathbb{R}^p}{\text{minimize}} \; \|x\|_1 \qquad \text{such that} \qquad \|y - Ax\|_2 \leq \epsilon. \tag{BPDN}$$

Note that BPDN is a quadratic problem.

In 2007, Candès and Tao added one more ℓ_1 minimization problem called Dantzig Selector (DS) [15] which handles noisy measurements and is still a linear programming problem.

$$\underset{x \in \mathbb{R}^p}{\text{minimize}} \; \|x\|_1 \qquad \text{such that} \qquad \|A'(y - Ax)\|_\infty \leq \epsilon. \tag{DS}$$

Some algorithms attempt to find a sparse x by solving the (P_0) problem from scratch without converting to a convex problem. These are mainly based on greedy methods and some nonlinear programming techniques, for example, Matching Pursuit (MP) [50], Orthogonal Matching Pursuit (OMP) [60], Iterative Hard Thresholding (IHT) [7], etc. Although these methods are easy to implement, they tend to fail and do not guarantee the recovery of a sparse vector if the number of nonzero entries of x is large. This thesis revolves around the modified versions of three ℓ_1-minimization problems BP, BPDN, and DS.

1.4 Sparse Recovery

To be able to recover a sparse x by solving one of the optimization problems above, the measurement matrix has to satisfy some special properties. We now define some of these properties of the matrix. More details will be given in later chapters.

Mutual coherence: Let $A_{m\times p}$ be a matrix with normalized columns $a_1, \ldots, a_p$ then the mutual coherence of A, denoted by $\mu(A)$, is defined as

$$\mu(A) = \max_{i\neq j} |\langle a_i, a_j\rangle|.$$

It is clear that $0 \leq \mu(A) \leq 1$. Thus μ measures how spread out the columns of A are, so that y captures unique information about x. We want μ to be as small as possible so that A is close to being an orthonormal matrix. For a matrix A of size $m \times p$, $m \leq p$, the following inequality gives the lower bound for μ and is known as Welch bound [68],

$$\mu \geq \sqrt{\frac{p-m}{m(p-1)}}$$

For a $m \times p$ matrix A with $m < p$ and $\text{rank}(A) = m$, the system $y = Ax$ is underdetermined and hence there are infinitely many solutions. The null space of A, $N(A)$, has dimension $\dim N(A) = p-m$. For any solution x_0 of $Ax = y$, the solution set will take the form $x_0 + N(A)$. Thus, the null space of A plays a crucial role in solving aforementioned optimization problems. We define a new term called the *spark* of a matrix which will help in characterizing the null space of a matrix A.

The *spark* of a given matrix A is the smallest number of columns of A that are linearly dependent. It is clear from the definition that $2 \leq \text{spark}(A) \leq m+1$. While it is not easy to find the spark of a matrix [34], it gives a simple criterion for the uniqueness of the sparse solution of $Ax = b$.

We will see in chapter 2 that various conditions for sparse recovery, such as $\|x\|_0 < \frac{1}{2}\left(1 + \frac{1}{\mu(A)}\right)$ or $\|x\|_0 < \frac{1}{4}\left(1 + \frac{1}{\mu(A)}\right)$, are used for the sufficiency of uniqueness or stability of solution of these ℓ_1 minimization problems.

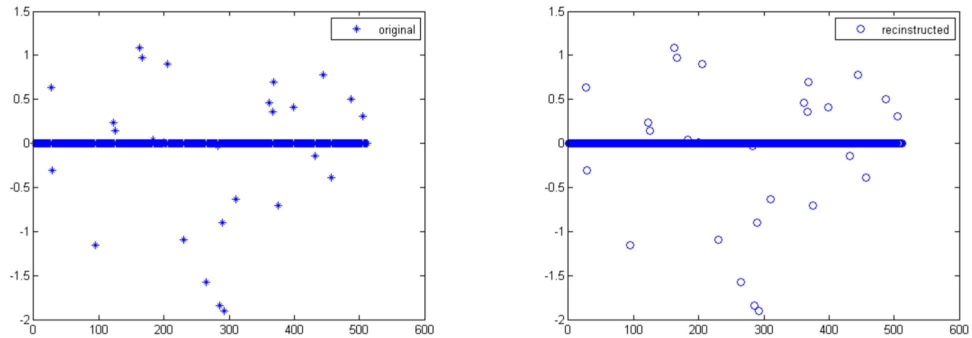
1.5 Compressed Sensing

The concept of sparse representation is one of the central methodologies in signal processing. It has been primarily used as a way to compress data by trying to minimize the number of atoms in the representation. Equipped with sparsity concepts and the results from classical sampling theorem, a stage was set for a new and revolutionary way of sampling based on the sparsity but not on the ambient dimension of a signal. A recent theory called compressed sensing (CS), a ground breaking work by Donoho, Candès, Tao and Romberg [18, 17, 15, 28, 19] states that the lower bound on the sampling rate can be highly reduced as soon as we realize that the signal is sparse, and the measurements are generalized to be any kind of linear measurement.

Compressed sensing uses ideas from various fields such as sampling theory, statistics, measure concentration in probability, inverse problems, approximation theory etc. In the process of measuring a sparse signal $x \in \mathbb{R}^p$, we produce many measurements and only keep nonzero components, thereby wasting many of the measurements. CS overcomes this by combining compression and sensing (measuring) at the same time. It also generalizes many concepts in sparse recovery and sampling theory, such as: 1) sparsity can occur in any basis or redundant dictionary, not just in Fourier basis; 2) sampling can be done inadaptively and linearly; 3) nonlinear reconstruction techniques such as the interior point method and greedy methods can be used for the reconstruction.

Here is an example of a signal reconstructed by using compressed sensing techniques. For this example, $p = 512, s = 28, m = 64$ and F is a 512×512 discrete Fourier matrix. A is the submatrix of F with m randomly chosen rows of F and x is a 512 dimensional vector with s nonzero coefficients chosen randomly and $y = Ax$.

In this short review of compressed sensing, we will use ℓ_1-minimization as a recovery method. In sparse recovery, the quality of a measurement matrix is measured by coherence, which only compares two columns at a time. Another measure of coherence known



(a) Original Signal (b) Reconstructed Signal

Figure 1.5.1 An example of a sparse signal reconstructed using CS

as restricted isometry is used in compressed sensing.

The problems considered in CS have the same form as in sparse recovery such as BP, BPDN; and the reconstruction processes are same as in sparse recovery; however, the measurement matrix is often taken to be a random matrix. The reason behind this is that random matrix is more incoherent than the deterministic matrices. This incoherence of a random matrix (measured by restricted isometries) helps to capture unrelated samples of a sparse signal thereby helping in recovery with fewer samples. Before we define restricted isometry, we will define null space property which is closely related to restricted isometry property.

Null space property (NSP): A matrix A is said to satisfy the null space property of order s if

$$\|h_S\|_1 < \|h_{S^c}\|_1$$

holds for all $h \in N(A)$ and for all S with $|S| \leq s$, where h_S is a vector restricted to indices in S and S^c denotes the complement of S.

This says that the vectors in the null space of A should not be too concentrated on a small subset of indices. NSP readily gives the necessary and sufficient conditions of recovery. The following theorem relates NSP and sparse recovery.

Theorem 1.5.1. *[24] A s-sparse x that satisfies $y = Ax$ is a unique solution of BP iff A satisfies the NSP of order s.*

While NSP gives the necessary and sufficient condition for the recovery, it is hard to verify for a given matrix [35]. Candès and Tao [19] defined a new property of a matrix called restricted isometry property.

Restricted Isometry Property (RIP): A matrix $A \in \mathbb{R}^{m \times p}$ is said to satisfy a Restricted Isometry Property of order s with constant δ_s if

$$(1 - \delta_s) \|c\|_2^2 \leq \|Ac\|_2^2 \leq (1 + \delta_s) \|c\|_2^2$$

for all s-sparse $c \in \mathbb{R}^p$.

RIP states that all column submatrices of A with s columns have to be well conditioned in the sense that eigenvalues of $A'_S A_S$ must be contained in the interval $[1 - \delta_s, 1 + \delta_s]$. The following theorem proved in [35] relates NSP and RIP.

Theorem 1.5.2. *If $A \in \mathbb{R}^{m \times p}$ satisfies RIP with constant $\delta_{2k} < \frac{1}{3}$, then A satisfies NSP of order k.*

Another similar property of the measurement matrix known as the restricted orthogonality property (ROC) is defined as follows.

Restricted Orthogonality Property (ROC): A matrix A satisfies the Restricted Orthogonality Property of order (k, k') with constant $\theta_{k,k'}$ if

$$|\langle Ac, Ac' \rangle| \leq \theta_{k,k'} \|c\|_2 \|c'\|_2$$

where c is k-sparse, c' is k'-sparse, and $\text{supp}(c) \cap \text{supp}(c') = \varnothing$.

Use of RIP and ROP is ubiquitous in compressed sensing. These are primarily used to obtain sufficient conditions for the uniqueness of various ℓ_1 minimization problems.

1.5.1 Matrices for CS

In this subsection we will discuss some measurement matrices which satisfy the Restricted Isometry Property. Most of the known constructions of measurement matrices that satisfy RIP are random matrices. Here are some examples of matrices used in compressed sensing [19, 17, 35].

(i). *Matrix with entries from Gaussian measurements*: Majority of the matrices used in CS are Gaussian random matrices. The following theorem was proved in [19].

Let the entries of the measurement matrix $A \in \mathbb{R}^{m\times p}$ *be independent normal random variables with variance* $\frac{1}{m}$ *and mean* 0 *and* $1 \le s \le p$*. The measurement matrix* A *satisfies RIP of order* s *with RIC* $\delta_s \le \delta \approx 0.01$ *with probability* $1 - O(e^{-\alpha m})$ *for some* $\alpha > 0$ *if* $m \ge Ck\log(p/k)$ *for some constant* C*.*

That is, we only require m measurements with $m \ge Ck\log(p/k)$ if x is s-sparse and entries are from Gaussian distribution.

(ii). *Matrix with entries from Bernoulli measurements*: A matrix A whose entries are taken from Bernoulli's distribution also satisfies RIP with high probability with the same bounds as in the Gaussian measurement matrix case.

(iii). *Structured random matrices (submatrices with rows chosen randomly from known orthogonal matrices)*: The measurement matrix can also come as a randomly chosen submatrix of a large Fourier Matrix with the number of measurements $m \ge Ck\log^{\alpha}(p)$.

(iv). *Deterministic matrices with low coherence*: In a situation where random sampling is not possible, the so-called incoherent dictionaries are also used as measurement matrices. The number of measurements required in this case is $m \ge C\mu^2(A)\log^4(p)$. Recently in [25], DeVore constructed a deterministic measurement matrix, but the RIC of such matrix is worse than the random matrices.

1.6 Problem Definition

In this section we define the problems that are the main content of this thesis.

1.6.1 Background

Compressed sensing problems require an assumption of sparsity or compressibility of a signal in an appropriate basis or in a redundant system. But there can be other forms of prior information that we want to exploit in the model. Some examples of prior information include the following.

- *Partial knowledge of the support*: Let x be an image reconstructed using classical filtered back projection; then, of course, the reconstruction $\hat{x}$ is not a true image but matches with an actual image in many aspects. We can use the support of $\hat{x}$ as a partial knowledge of the support. See [65] for detail.

- *Tree structure*: It is known that the wavelet coefficients of piecewise smooth signals and images tend to live on a rooted and connected tree structure [6].

- *Joint Sparsity:* There may be other instances where we want to exploit the extra information. Signals might have large coefficients clustered around a small set of indices, also called blank sparsity or joint sparsity. Multiple signals having some common values or having a common support are also considered in this model. For details, see [33] etc.

Among these various possibilities of prior information, we assume that partial knowledge of the support is given in advance, i.e., for a s-sparse signal $x \in \mathbb{R}^p$, we assume that k indices where $k \leq p$ are known. Note that we may not know the signal value at those k indices, we are only assuming that the locations of k components of x are known in advance. Here are some examples of where these problems arise.

1. The Wavelet transform of many images has only a few nonzero wavelet coefficients. The known part K of the indices can be the indices of the scaling function coefficients.

2. In a problem of reconstructing an image or signal by an iterative method, the support estimate from the previous step can be used as a set of known indices to make the reconstruction process faster.

3. Support of an approximate solution from some suboptimal methods such as Filtered Back Projection in Tomography can be used as a known part of the support.

There are many benefits of using this extra information and incorporating it into the recovery problems. We modify three main ℓ_1 minimization problems (BP, BPDN and DS) to include the knowledge of the partial support. By doing this, we not only generalize the regular ℓ_1 minimization problems but also expect to get weaker sufficient conditions, better bounds, and flexibility in choosing measurement matrix, thereby, reducing the number of measurements or recovering more sparse signal.

1.6.2 Research Problems

We introduce some notations before we state our research problems. If $K \subset \{1, 2, \ldots, p\}$ is a set of indices, we denote by x_K the vector formed from the corresponding entries of x. The set of indices of the nonzero coefficients of x is called the *support of* x. For a matrix A, A_K is the matrix formed from the corresponding columns of A. We use A' instead of A^T to denote the transpose of a matrix A.

Let A be an $m \times p$ matrix of full rank, with $m < p$, and y a vector in $\mathbb{R}^m$. Let S be the true support of the solution x of mod-BP, and K an estimate of the support. The "unknown part" U consists of indices in S that are not in K. The "error" E consists of the elements in K which are not actually in S. The numbers s, k, e, u will denote the number of indices in S, K, E, U respectively. Thus we have

$$S = (K \cup U) \setminus E,$$

$$K \cap U = \varnothing,$$

$$s = k + u - e.$$

Modified basis pursuit is the problem of solving the optimization problem

$$\underset{x}{\text{minimize}} \ \|x_{K^c}\|_1 \quad \text{subject to } y = Ax. \tag{mod-BP}$$

That is, among all the x that satisfy $y = Ax$, we choose one that has the smallest ℓ_1 norm outside the known part of the support. For $K = \emptyset$, this reduces to regular basis pursuit. For the noisy measurements, $y = Ax + w$ with bounded noise we modify BPDN and DS to get the modified problems

$$\underset{x}{\text{minimize}}\ \|x_{K^c}\|_1 \quad \text{subject to } \|y - Ax\|_2 \le \epsilon \tag{mod-BPDN}$$

and

$$\underset{x}{\text{minimize}}\ \|x_{K^c}\|_1 \quad \text{subject to } \|A'(y - Ax)\|_\infty \le \epsilon \tag{mod-DS}$$

Note that the original Dantzig selector program is considered with Gaussian noise. Here we have modified with the bounded noise, and we will also consider mod-DS with Gaussian noise.

This thesis deals with above three modified ℓ_1 minimization problems. We will discuss sufficient conditions, necessary and sufficient conditions for the existence of unique solution, and numerical performance of these modified problems.

1.6.3 Literature Review of CS with Partially Known Support

The recursive reconstruction problem with partially known support was first studied in [62] and the modification of it was done in [20]. Recent works on recursive reconstruction use various approaches such as Bayesian, model based approaches, etc [46, 55, 71, 70, 38, 44] to estimate sparse signal recursively with time varying support.

The problem of sparse reconstruction with partial knowledge of the support was studied simultaneously in [64, 65] and in [45, 51]. The work of [45] obtains exact recovery thresholds for weighted ℓ_1, similar to those in [27], for the case when a probabilistic prior on the signal support is available. Some related work motivated by [64] include modified OMP [57], modified CoSaMP [21], modified block CS [58], error bounds on modified BPDN [43, 63, 47, 48], better conditions for modified-CS based exact recovery [36], and exact support recovery conditions for multiple measurement vectors (MMV) based recursive recovery [46].

There is other recent work that may also be referred to as recursive sparse reconstruction, but whose goals are quite different from the problem that we discuss in this thesis. This includes (i) homotopy methods, e.g. [54, 67], whose goal is to only speed up the optimization algorithm using homotopy or warm starts and the previous reconstructed signal, but not to reduce the number of measurements required; (ii) [49, 54, 1, 39] which reconstruct a single signal from sequentially arriving measurements; and (iii) [14, 22, 67], which iteratively improve support estimation for a single sparse signal.

In [66] Borries, Miosso and Potes have showed that if a signal is sparse in the discrete Fourier basis, then the number of measurements while reconstructing it by solving BP can be reduced by the number of known indices. Bandeira, Scheinberg and Vicente [4] recently used the modified null space property and modified RIP property of the measurement matrix A to prove that we only require $m \geq O(s + \log(p - k)$ measurements to reconstruct a s-sparse signal when k indices are known. Note that regular BD requires $m \geq O(s \log(p/s))$. Ince, Nacaroglu and Watsuji [42] extended this result for a nonconvex compressed sensing problem by replacing the ℓ_1 norm with the ℓ_q norm $(0 < q < 1)$ and proved that $\delta_{k+(a+1)u} + a^{\frac{1}{2}-\frac{1}{q}} \left(\delta^2_{k+(a+1)u} + \delta^2_{2au}\right) < 1$ with $0 < q < 1$ and $a > 1$ is sufficient for the stable recovery by solving mod-BPDN. This sufficient condition reduces to L. Jacques' sufficient conditions [43] $\delta^2_{2u} + 2\delta_{k+2u} < 1$ when $q = 1$ and $a = 1$.

CHAPTER 2. SUFFICIENT CONDITIONS

It is now well known that compressed sensing offers an efficient method of reconstructing a high dimensional vector, if the vector is sufficiently sparse and the measurement matrix satisfies suitable restricted isometry and restricted orthogonality properties. In this chapter, we derive various sufficient conditions for the uniqueness and stability of solution of modified ℓ_1 minimization problems based on restricted isometries and restricted orthogonality constants. We will also consider coherence as a measure of quality of a measurement matrix for the deterministic compressed sensing problem and derive a sufficient condition which is weaker than the corresponding sufficient condition used for BP and BPDN.

2.1 RIC and ROC Based Sufficient Conditions

We will first discuss some properties of restricted isometry constants (RICs) and restricted orthogonality constants (ROCs) and state some sufficient conditions that are used for regular ℓ_1 minimization problems, before we prove our results.

2.1.1 Some Properties of δ_k and $\theta_{k,k'}$

As defined in chapter 1, the RIC δ_k is the smallest constant satisfying

$$\sqrt{1-\delta_k}\,\|c\|_2 \leq \|Ac\|_2 \leq \sqrt{1+\delta_k}\,\|c\|_2$$

for all k-sparse $c \in \mathbb{R}^p$, and the ROC $\theta_{k,k'}$ is the smallest constant satisfying

$$|\langle Ac, Ac'\rangle| \leq \theta_{k,k'}\|c\|_2\,\|c'\|_2$$

for all k-sparse c and k'-sparse c' with disjoint supports and $k+k' \le p$. We will use the following properties satisfied by these constants.

$$\begin{aligned} &\delta_k \le \delta_\ell \quad \text{if } k \le \ell \\ &\theta_{k,k'} = \theta_{k',k} \le \theta_{\ell,k} = \theta_{k,\ell} \quad \text{if } k' \le \ell \\ &\theta_{k,k'} \le \delta_{k+k'} \le \theta_{k,k'} + \max(\delta_k, \delta'_k) \end{aligned} \tag{2.1.1}$$

$$\theta_{k,ak'} \le \sqrt{a}\theta_{k,k'} \quad \text{if } ak' \text{ is integer, and } a>0 \text{ is a real number.}$$

The first two are obvious from the definitions. The third inequality comes from [19]. The last comes from [11], where it is called the *square root lifting property*.

2.1.2 Some Well Known Theorems in CS

The following theorems are the main foundation of CS theory.

Theorem 2.1.1. *[19] Let x^* be a feasible k-sparse vector for BP, that is, it satisfies $y = Ax^*$. Then x^* is the unique minimizer of BP if $\delta_k + \theta_{k,k} + \theta_{k,2k} < 1$.*

The same authors extended the result for the noisy case as follows.

Theorem 2.1.2. *[18] Let x_0 be k-sparse, feasible for BPDN and $\delta_{3k} + 3\delta_{4k} < 2$. Then the minimizer x^* of BPDN satisfies*

$$\|x^* - x_0\|_2 \le C_k \epsilon$$

where C_k is a constant that depends only on restricted isometry constants.

The following theorem states the similar theorem for the DS optimization problem with Gaussian noise with bound $\epsilon = \sigma\sqrt{2\log p}$ where σ is the standard deviation of components w_i of the noise w in the model $y = Ax + w$.

Theorem 2.1.3. *[15] Let $x_0 \in \mathbb{R}^p$ be k-sparse, feasible for DS and $\delta_{2k} + \theta_{k,2k} < 1$. Choose $\epsilon = \sigma\sqrt{2\log p}$, then with large probability the minimizer x^* of DS satisfies*

$$\|x^* - x_0\|_2^2 \le 2\,k\,C_1^2\,\sigma^2\,\log(p)$$

where $C_1 = \frac{4}{1-\delta_k-\theta_{k,2k}}$.

2.1.3 Existing Sufficient Conditions for Modified Problems

The following is a list of some sufficient conditions used to either exactly recover x by solving mod-BP or approximately recover x by solving mod-BPDN. *Modified compressed sensing* [65] is one of the original papers in which authors used both sparsity and partial knowledge of the support for the recovery of sparse vectors by solving mod-BP. The sufficient conditions used in this paper are

$$\delta_{k+u} < 1$$

and

$$\delta_{2u} + \theta_{u,u} + \theta_{u,2u} + \delta_k + \theta_{k,u}^2 + 2\theta_{k,2u}^2 < 1.$$

It was also demonstrated and proved that these conditions are weaker than the original condition $\delta_s + \theta_{s,s} + \theta_{s,2s} < 1$ for the same purpose when solving BP. A simpler condition

$$\delta_{2u}^2 + 2\delta_{k+2u} < 1$$

was used in [43] to stably recover x by solving mod-BPDN. There have been efforts to use ℓ_q-norm with $0 < q \leq 1$ for the recovery by using the following optimization problem

$$\underset{x \in \mathbb{R}^p}{\text{minimize}} \ \|x\|_q \qquad \text{such that} \qquad y = Ax \tag{Pq}$$

where

$$\|x\|_q = \left(\sum_{i=1}^{p} |x_i|^q \right)^{\frac{1}{q}}.$$

In [42], a sufficient condition

$$\delta_{k+(a+1)u} + a^{\frac{1}{2}-\frac{1}{q}} \left(\delta_{k+(a+1)u}^2 + \delta_{2au}^2 \right) < 1$$

is used to guarantee the uniqueness of solution of (Pq). In this section, we prove that

$$\delta_{k+u+a} + \sqrt{\frac{u}{b}}\, \theta_{k+u+a,b} < 1,$$

where a, b, u are integers with $0 < a < b \leq 4a$, $0 \leq u \leq p-k$ works for both the uniqueness of solution of mod-BP and stability of solution of mod-BPDN and mod-DS. We will also show that our condition is weaker than the conditions used by other authors to solve modified ℓ_1 minimization problems. We will also derive several other conditions by varying a, b and u.

2.1.4 Partitioning the Indices

Our notation is an extension of the notation introduced by Vaswani and Lu [65]. The *support* of a vector x is the set of indices of the nonzero components. S is the true support of the solution vector, and K an estimate of the support (the "known part"). In case x is compressible, S denotes the indices with largest s components (in absolute value) of x. The "unknown part" U consists of indices in S that are not in K. The "error" E consists of the elements in K which are not actually in S. The numbers s, k, e, u denote the size of these sets. We also find it convenient to introduce another set $R = (K \cup U)^C$ (the "rest"). Here is a sketch, where bullets represent the indices of x:

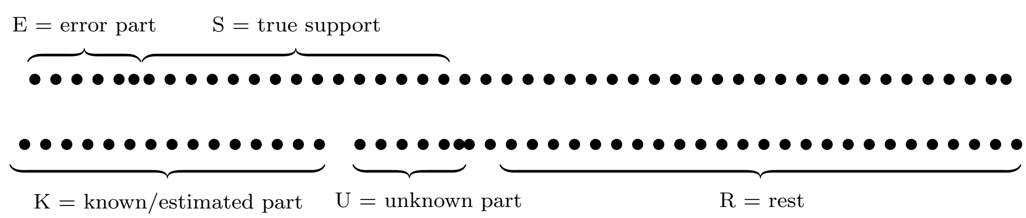


Figure 2.1.1 Support division of x

If $T \subset \{1, 2, \ldots, n\}$ is an index set, the notation x_T means the vector of length $t = |T|$ formed from the entries of x with indices in T. For a matrix A of size $m \times p$, A_T is the matrix of size $m \times t$ formed from the corresponding columns of A. For vectors (but not for matrices) we use x_T sometimes to denote a vector with all coefficients outside T set to 0; this is a vector of length p. The meaning is always clear from the context.

A general vector x is divided into three non-overlapping parts x_K, x_U, x_R. Here is a sketch of our setup:

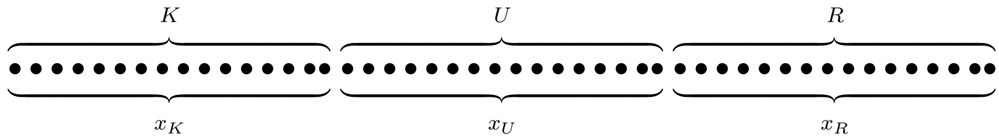


Figure 2.1.2 Support division and corresponding parts of x

We do not use the sets E, S in any of the developments in this paper. E is determined after the problem has been solved, and in turn determines S. We cannot make a statement such as "x is s-sparse" without knowing what e is. We can only make statements about the sparsity of x outside K, such as saying "x_{K^C} is u-sparse". It is true that [65] gives estimates involving e, but only in the combination $s + e$, which really means $k + u$.

If x is a vector in $\mathbb{R}^p$, the notation x_j means the jth component of x. The notation x_{+j} is the vector x with all entries in K set to 0, and all but the largest j entries (in absolute value) outside the set K set to 0. The vector x_{-j} is x with all entries in K set to 0, and the largest j entries outside K set to 0. Thus, for any j, $x_{+j} + x_{-j} = x_{K^C} = x_U + x_R$.

Note: $x_{\pm j}$ may not be uniquely defined, if there are several entries of the same size. This causes no problems. We just use one of the possible choices.

Here is an example. Consider the following vector: Its parts are given in the

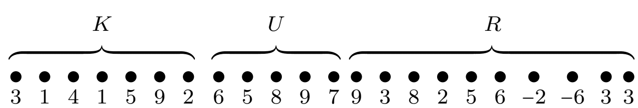


Figure 2.1.3 An example of x with different parts

following table, where zeros are represented by dots, for easier reading.

We will also keep referring to the three modified problems stated in section 1.6.

$$\min_{x \in \mathbb{R}^p} \|x_{K^c}\|_1 \quad \text{subject to } y = Ax \tag{mod-BP}$$

Table 2.1.1 : Separated parts of x from Figure 2.1.3

x_U	.	.	.	.	.	.	.	6	5	8	9	7	.	.	.	.	.	.	.	.	.	.
x_R	.	.	.	.	.	.	.	.	.	.	.	.	9	3	8	2	5	6	-2	-6	3	3
x_{+5}	.	.	.	.	.	.	.	.	.	8	9	7	9	.	8	.	.	.	.	.	.	.
x_{-5}	.	.	.	.	.	.	.	6	5	.	.	.	.	3	.	2	5	6	-2	-6	3	3

That is, we seek an x that has the least ℓ_1 norm outside the known part K and satisfies $y = Ax$. The other two problems are

$$\min_{x \in \mathbb{R}^p} \|x_{K^c}\|_1 \quad \text{subject to } \|y - Ax\|_2 \leq \epsilon \tag{mod-BPDN}$$

$$\min_{x \in \mathbb{R}^p} \|x_{K^c}\|_1 \quad \text{subject to } \|A'(y - Ax)\|_\infty \leq \epsilon \tag{mod-DS}$$

In above two programs, we have considered the bounded noise case. The case when the noise is Gaussian will be treated differently, at the end of subsection 2.1.6.

2.1.5 Some Basic Lemmas

In this subsection, we start with some preliminary lemmas. The first one relates the ℓ_1, ℓ_2 and ℓ_∞ norms of a vector.

Lemma 2.1.4. *If x is a vector with support length L, then*

$$\frac{1}{\sqrt{L}}\|x\|_1 \leq \|x\|_2 \leq \sqrt{L}\|x\|_\infty.$$

Proof. For the first inequality, let $e = \text{sign}(x)$, and apply the Cauchy-Schwartz inequality:

$$\|x\|_1 = \langle x, e\rangle \leq \|x\|_2 \, \|e\|_2 = \sqrt{L}\, \|x\|_2.$$

For the second part, use

$$\|x\|_2^2 = \sum_i |x_i|^2 \leq L \cdot \max(|x_i|)^2 = L\, \|x\|_\infty^2.$$

□

The following is known as the *Shifting Lemma* and is proved in [11].

Lemma 2.1.5. *Let q, r be positive integers satisfying $q \le 3r$. Then any nonincreasing sequence of nonnegative real numbers*

$$a_1 \ge a_2 \ge \cdots \ge a_r \ge b_1 \ge \cdots \ge b_q \ge c_1 \ge \cdots \ge c_r \ge 0$$

satisfies

$$\sqrt{\sum_{i=1}^{q} b_i^2 + \sum_{i=1}^{r} c_i^2} \le \frac{\sum_{i=1}^{r} a_i + \sum_{i=1}^{q} b_i}{\sqrt{q+r}}$$

Stated another way, if the vectors a, b and c satisfy the conditions of the shifting lemma, then

$$\|(b,c)\|_2 \le \frac{\|a\|_1 + \|b\|_1}{\sqrt{q+r}}.$$

For the proofs of the following lemma and the theorems in the next section, we introduce some more notation.

Assume K, U are fixed. Choose some integers a, b at random, with $0 < a < b \le 4a$. An arbitrary vector h is partitioned into h_K, h_U and the rest h_R. We sort h_R in decreasing order, and partition it into h_*, h_1, h_2, $\ldots$, where h_* has length a and each h_i has length b. (Pad h with some zeros at the end, if necessary). Then we subdivide each h_i into h_{i1} of length $b-a$ and h_{i2} of length a. We also define $h_0 = h_K + h_U + h_*$.

This is illustrated in the following table:

Table 2.1.2 : Division of different parts and support of h

<table>
<tr><td rowspan="3">Parts of h:</td><td>h_K</td><td>h_U</td><td colspan="8">h_R</td></tr>
<tr><td colspan="3">h_0</td><td colspan="2">h_1</td><td colspan="2">h_2</td><td colspan="2">h_3</td><td>...</td></tr>
<tr><td>h_K</td><td>h_U</td><td>h_*</td><td>h_{11}</td><td>h_{12}</td><td>h_{21}</td><td>h_{22}</td><td>h_{31}</td><td>h_{32}</td><td>...</td></tr>
<tr><td rowspan="2">Support size:</td><td colspan="3">$k+u+a$</td><td colspan="2">b</td><td colspan="2">b</td><td colspan="2">b</td><td>...</td></tr>
<tr><td>k</td><td>u</td><td>a</td><td>$b-a$</td><td>a</td><td>$b-a$</td><td>a</td><td>$b-a$</td><td>a</td><td>...</td></tr>
</table>

There have been multiple efforts [16, 9, 11] to improve the sufficient conditions based on RIPs and ROPs. One important result is stated in following theorem.

Theorem 2.1.6. *[11] Let x^*, $\hat{x}$, $\breve{x}$ be the minimizers of BP, BPDN and DS (with bounded noise) respectively. If x is a vector feasible for the respective optimization prob-*

lem and if $\delta_{1.25k} + \theta_{1.25k,k} < 1$ *then*

$$\|x^* - x\|_2 \le B\|x_{-k}\|_1$$

$$\|\hat{x} - x\|_2 \le A\epsilon + B\|x_{-k}\|_1$$

$$\|\breve{x} - x\|_2 \le C\epsilon + B\|x_{-k}\|_1$$

where A, B, C *are constants depending on RIC and ROC.*

This is important because this result not only weakens the sufficient condition used for these ℓ_1 minimization problems, but also uses the same condition for all three optimization problems.

Lemma 2.1.7. *Let* $\hat{x}$ *be a solution of one of the modified* ℓ_1 *minimization problems (mod-BP, mod-BPDN and mod-DS), and* x *some vector which is feasible for the same problem.*

Choose an integer u *with* $0 \le u \le p-k$ *at random, and let* $U = supp(x_{+u})$*. Let* $h = \hat{x} - x$*, and partition* h *as above. Then the following estimates hold.*

$$\|h_R\|_1 \le \|h_U\|_1 + 2\|x_{-u}\|_1. \tag{2.1.2}$$

$$\sum_i \|h_i\|_2 \le \sqrt{\frac{u}{b}}\,\|h_0\|_2 + \frac{2}{\sqrt{b}}\|x_{-u}\|_1 \tag{2.1.3}$$

$$\|h\|_2 \le \sqrt{1 + \frac{u}{b}}\|h_0\|_2 + \frac{2}{\sqrt{b}}\|x_{-u}\|_1 \tag{2.1.4}$$

$$\|h\|_2 \le \left(1 + \sqrt{\frac{u}{b}}\right)\|h_0\|_2 + \frac{2}{\sqrt{b}}\|x_{-u}\|_1 \tag{2.1.5}$$

Note: The estimate in (2.1.4) is better than (2.1.5). However, (2.1.5) may lead to estimates that are easier to interpret.

Proof. By the choice of U (the largest u components of x outside K) we have $x_U = x_{+u}$ and $x_R = x_{-u}$.

Since $\hat{x}$ minimizes the ℓ_1 norm on K^C over all feasible vectors, we have

$$\begin{aligned}
\|x_U\|_1 + \|x_R\|_1 = \|x_{K^C}\|_1 \geq \|\hat{x}_{K^C}\|_1 &= \|(x+h)_{K^C}\|_1 \\
&= \|(x+h)_U\|_1 + \|(x+h)_R\|_1 \\
&\geq (\|x_U\|_1 - \|h_U\|_1) + (\|h_R\|_1 - \|x_R\|_1) .
\end{aligned}$$

Solve this for $\|h_R\|_1$:

$$\begin{aligned}
\|h_R\|_1 &\leq \|x_U\|_1 + \|x_R\|_1 - \|x_U\|_1 + \|h_U\|_1 + \|x_R\|_1 \\
&= \|h_U\|_1 + 2\|x_R\|_1 = \|h_U\|_1 + 2\|x_{-u}\|_1.
\end{aligned}$$

This proves (2.1.2).

Since $0 < b - a \leq 3a$, we can use the shifting inequality for the triplets (h_*, h_{11}, h_{12}), (h_{12}, h_{21}, h_{22}), (h_{22}, h_{31}, h_{32}), and so on, to get

$$\|h_1\|_2 \leq \frac{\|h_*\|_1 + \|h_{11}\|_1}{\sqrt{b}}, \quad \|h_2\|_2 \leq \frac{\|h_{12}\|_1 + \|h_{21}\|_1}{\sqrt{b}}, \quad \|h_3\|_2 \leq \frac{\|h_{22}\|_1 + \|h_{31}\|_1}{\sqrt{b}}, \quad \dots$$

Taking the sum of these inequalities, we find

$$\begin{aligned}
\sum_i \|h_i\|_2 &\leq \frac{\|h_*\|_1 + \sum_i \|h_i\|_1}{\sqrt{b}} = \frac{\|h_R\|_1}{\sqrt{b}} \\
&\leq \frac{\|h_U\|_1 + 2\|x_{-u}\|_1}{\sqrt{b}} \qquad \text{by (2.1.2)} \\
&\leq \frac{\sqrt{u}\|h_U\|_2 + 2\|x_{-u}\|_1}{\sqrt{b}} \qquad \text{by lemma 2.1.4} \\
&\leq \frac{\sqrt{u}\|h_0\|_2 + 2\|x_{-u}\|_1}{\sqrt{b}}
\end{aligned}$$

This proves (2.1.3).

The third estimate is

$$\begin{aligned}\|h\|_2^2 &= \|h_0\|_2^2 + \sum_i \|h_i\|_2^2 \\ &\le \|h_0\|_2^2 + \left(\sum_i \|h_i\|_2\right)^2 \\ &\le \|h_0\|_2^2 + \left(\sqrt{\frac{u}{b}}\|h_0\|_2 + \frac{2}{\sqrt{b}}\|x_{-u}\|_1\right)^2 \\ &\le \left(\sqrt{1+\frac{u}{b}}\|h_0\|_2 + \frac{2}{\sqrt{b}}\|x_{-u}\|_1\right)^2.\end{aligned}$$

Here the last inequality can be verified by expanding the squares.

The fourth estimate follows immediately from the third, by

$$\sqrt{1+\frac{u}{b}} \le 1 + \sqrt{\frac{u}{b}},$$

which again can be verified by squaring both sides. □

Lemma 2.1.8. *Assuming that h is partitioned as above, we have*

$$|\langle Ah, Ah_0\rangle| \ge \left(1 - \delta_{k+u+a} - \sqrt{\frac{u}{b}}\,\theta_{k+u+a,b}\right)\|h_0\|_2^2 - \theta_{k+u+a,b}\,\|h_0\|_2\,\frac{2}{\sqrt{b}}\|x_{-u}\|_1. \tag{2.1.6}$$

Proof. This estimate is based on equation (2.1.3).

$$\begin{aligned}|\langle Ah, Ah_0\rangle| &= |\langle A(h_0 + \sum_{i\ge1} h_i), Ah_0\rangle| \\ &\ge \|Ah_0\|_2^2 - \sum_{i\ge1}|\langle Ah_i, Ah_0\rangle| \\ &\ge (1-\delta_{k+u+a})\|h_0\|_2^2 - \theta_{k+u+a,b}\|h_0\|_2\sum_{i\ge1}\|h_i\|_2 \\ &\ge (1-\delta_{k+u+a})\|h_0\|_2^2 - \theta_{k+u+a,b}\|h_0\|_2\left(\sqrt{\frac{u}{b}}\|h_0\|_2 + \frac{2}{\sqrt{b}}\|x_{-u}\|_1\right) \\ &= \left(1-\delta_{k+u+a} - \sqrt{\frac{u}{b}}\theta_{k+u+a,b}\right)\|h_0\|_2^2 - \theta_{k+u+a,b}\,\|h_0\|_2\,\frac{2}{\sqrt{b}}\|x_{-u}\|_1.\end{aligned}$$

□

Equipped with the necessary lemmas, notations and the background, we are now in position to state and prove the main results of this section.

2.1.6 Main Results

In this section, we prove our main results in various theorems. The proofs are similar to [11] and have been modified and changed according to the new problem at hand.

We define

$$A = \frac{2\sqrt{1+\frac{u}{b}}\,\sqrt{1+\delta_{k+u+a}}}{1-\delta_{k+u+a}-\sqrt{\frac{u}{b}}\,\theta_{k+u+a,b}}$$

$$B = \frac{2}{\sqrt{b}}\left(1+\frac{\theta_{k+u+a,b}\sqrt{1+\frac{u}{b}}}{1-\delta_{k+u+a}-\sqrt{\frac{u}{b}}\,\theta_{k+u+a,b}}\right) \tag{2.1.7}$$

$$C = \frac{2\sqrt{1+\frac{u}{b}}\,\sqrt{k+u+a}}{1-\delta_{k+u+a}-\sqrt{\frac{u}{b}}\,\theta_{k+u+a,b}}$$

Theorem 2.1.9. *Suppose that $\hat{x}$ is a solution to mod-BP, and that x is feasible for mod-BP. Choose at random some integers a, b, u with $0 < a < b \le 4a$, $0 \le u \le p-k$.*

If

$$\delta_{k+u+a} + \sqrt{\frac{u}{b}}\,\theta_{k+u+a,b} < 1,$$

then we have

$$\|\hat{x}-x\|_2 \le B\|x_{-u}\|_1.$$

Proof. As before, let $U = \text{supp}(x_{+u})$. Let $h = \hat{x}-x$, then $Ah = A\hat{x} - Ax = 0$, so equation (2.1.6) leads to

$$\|h_0\|_2 \le \frac{2\theta_{k+u+a,b}}{\sqrt{b}\left(1-\delta_{k+u+a}-\sqrt{\frac{u}{b}}\,\theta_{k+u+a,b}\right)}\|x_{-u}\|_1. \tag{2.1.8}$$

Substituting this into equation (2.1.4) leads to

$$\begin{aligned}\|h\|_2 &\le \left(\sqrt{1+\frac{u}{b}}\right)\|h_0\|_2 + \frac{2}{\sqrt{b}}\|x_{-u}\|_1 \\ &\le \left(\sqrt{1+\frac{u}{b}}\right)\frac{2\theta_{k+u+a,b}}{\sqrt{b}\left(1-\delta_{k+u+a}-\sqrt{\frac{u}{b}}\,\theta_{k+u+a,b}\right)}\|x_{-u}\|_1 + \frac{2}{\sqrt{b}}\|x_{-u}\|_1,\end{aligned}$$

which gives the stated result after simplification. □

Note 1: If x_{K^C} is u-sparse, then $x = \hat{x}$.

Note 2: If we use estimate (2.1.5) instead of (2.1.4), we get for the mod-BP case, the estimate [11]

$$\|\hat{x} - x\|_2 \leq \tilde{B}\|x_{-u}\|_{-1} = \frac{1 - \delta_{k+u+a} + \theta_{k+u+a,b}}{1 - \delta_{k+u+a} - \sqrt{\frac{u}{b}}\,\theta_{k+u+a,b}}\,\frac{2}{\sqrt{b}}\|x_{-u}\|_1.$$

The constant $\tilde{B}$ is not as sharp as B.

Theorem 2.1.10. *Suppose that $\breve{x}$ is the solution of mod-BPDN, and x^* is the solution to mod-DS. Let x be a feasible solution of the corresponding problem.*

Choose at random some integers a, b, u with $0 < a < b \leq 4a$, $0 \leq u \leq p - k$.

If

$$\delta_{k+u+a} + \sqrt{\frac{u}{b}}\,\theta_{k+u+a,b} < 1,$$

then we have

$$\|\breve{x} - x\|_2 \leq A\epsilon + B\|x_{-u}\|_1,$$

$$\|x^* - x\|_2 \leq C\epsilon + B\|x_{-u}\|_1.$$

Proof. As before, let $U = \operatorname{supp}(x_{+u})$. For the case of mod-BPDN, we use $h = \breve{x} - x$. Then we have $\|Ah\|_2 \leq 2\epsilon$ and

$$|\langle Ah, Ah_0\rangle| \leq \|Ah\|_2\,\|Ah_0\|_2 \leq 2\epsilon\sqrt{1 + \delta_{k+u+a}}\,\|h_0\|_2.$$

We put this into equation (2.1.6) and the result into (2.1.4) and simplify.

For the case of mod-DS, let $h = x^* - x$. Let A_0 be the matrix of size $p \times (k + u + a)$ formed by the columns of A corresponding to the support of h_0.

$$\begin{aligned}
|\langle Ah, Ah_0\rangle| &= |\langle Ah, A_0 h_0\rangle| \\
&= |\langle A_0'(Ax^* - y - Ax + y), h_0\rangle| \\
&\leq (\|A_0'(Ax^* - y)\|_2 + \|A_0'(Ax - y)\|_2)\,\|h_0\|_2 \\
&\leq \sqrt{k + u + a}\,(\|A_0'(Ax^* - y)\|_\infty + \|A_0'(Ax - y)\|_\infty)\,\|h_0\|_2 \\
&\leq 2\epsilon\sqrt{k + u + a}\|h_0\|_2.
\end{aligned}$$

We put this again into equation (2.1.6), and simplify to get

$$\|h_0\|_2 \le \frac{2\theta_{k+u,b}}{\sqrt{b}\left(1-\delta_{k+u+a}-\sqrt{\frac{u}{b}}\,\theta_{k+u+a,b}\right)}\|x_{-u}\|_1 + \frac{2\epsilon\sqrt{k+u+a}}{1-\delta_{k+u+a}-\sqrt{\frac{u}{b}}\,\theta_{k+u+a,b}}$$

Applying again equation (2.1.4) gives the desired result. □

Note: Theorem 2.1.9 also follows directly from the first part of theorem 2.1.10 by setting $\epsilon = 0$.

Finally, we work on mod-BPDN and mod-DS, considering the case that the noise is Gaussian. The model is $y = Ax + w$, where $w \sim N(0, \sigma^2 I)$.

The following lemma proved in [9] deals with the probability part of the theorem.

Lemma 2.1.11. *If $w \sim N(0, \sigma^2 I)$ then*

$$P\left(\|w\|_2 \le \sigma\sqrt{m + 2\sqrt{m\log m}}\right) \ge 1 - \frac{1}{m}$$

and

$$P\left(\|A'w\|_\infty \le \sigma\sqrt{2\log p}\right) \ge 1 - \frac{1}{2\sqrt{\pi\log p}}$$

The bounds are chosen so as to express the probability in simpler form. The proof of the following theorem follows directly from this lemma and theorem 2.1.10.

Theorem 2.1.12. *Suppose that $\delta_{k+u+a} + \sqrt{\frac{u}{b}}\,\theta_{k+u+a,b} < 1$ for positive integers a, b, k, u defined as above. Then the minimizer $\breve{x}$ of mod-BPDN with $\epsilon = \sigma\sqrt{m + 2\sqrt{m\log m}}$ satisfies*

$$\|\breve{x} - x\|_2 \le A\sigma\sqrt{m + 2\sqrt{m\log m}} + B\|x_{-u}\|_1$$

with probability at least $1 - \frac{1}{m}$, and the minimizer x^* of mod-DS with $\epsilon = \sigma\sqrt{2\log p}$ satisfies

$$\|x^* - x\|_2 \le C\sigma\sqrt{2\log p} + B\|x_{-u}\|_1$$

with probability at least $1 - \frac{1}{2\sqrt{\pi\log p}}$.

2.1.7 Comparison/Discussion

In this section we derive various forms of our sufficient conditions and compare them with other existing sufficient conditions. The original sufficient condition from theorem 2.1.9 is

$$\delta_{k+u+a} + \sqrt{\frac{u}{b}}\,\theta_{k+u+a,b} < 1,\ 0 < a < b \le 4a \tag{2.1.9}$$

Another sufficient condition is

$$\delta_{k+2u} + \theta_{k+2u,u} < 1. \tag{2.1.10}$$

This follows by taking $b = 2u$, $a = u$ in (2.1.9) and using square root lifting inequality:

$$\delta_{k+2u} + \frac{1}{\sqrt{2}}\,\theta_{k+2u,2u} \le \delta_{k+2u} + \frac{\sqrt{2}}{\sqrt{2}}\,\theta_{k+2u,2u} = \delta_{k+2u} + \theta_{k+2u,u} < 1.$$

Compare this condition with Candès and Tao's sufficient condition [15]

$$\delta_{2s} + \theta_{2s,s} = \delta_{2k+2u} + \delta_{2k+2u,k+u} < 1.$$

Using (2.1.10), square root lifting inequality again and the known inequality $\theta_{k,k'} \le \delta_{k+k'}$, we have

$$\begin{aligned} \delta_{k+2u} + \theta_{k+2u,u} &\le \delta_{k+2u} + \sqrt{2}\,\theta_{(k/2)+u,u} \\ &\le \delta_{k+2u} + \sqrt{2}\,\delta_{k+2u} \\ &= (\sqrt{2}+1)\delta_{k+2u} \\ &< 1 \text{ if } \ \delta_{k+2u} < \sqrt{2} - 1. \end{aligned}$$

This proves that the condition

$$\delta_{k+2u} < \sqrt{2} - 1 \approx 0.4142. \tag{2.1.11}$$

is sufficient. Compare this with Vaswani and Lu's [65] condition $\delta_{k+2u} < \frac{1}{5} = 0.2$ and also with Candès' condition [16] $\delta_{2s} = \delta_{2k+2u} < \sqrt{2} - 1 \approx 0.4142$.

Another sufficient condition

$$\delta_{k+2u} + \delta_{k+3u} < 1 \tag{2.1.12}$$

follows directly from (2.1.10) and the inequality $\theta_{k,k'} \le \delta_{k+k'}$.

To get a condition in terms of s and u, substitute $k + u = s$ and rewrite (2.1.10) as

$$\delta_{s+u} + \theta_{s+u,u} < 1. \tag{2.1.13}$$

Note that in general, we expect $u \ll s$ [65]. For $u \le \frac{s}{4}$ or equivalently $u \le \frac{k}{3}$, the condition $\delta_{k+2u} + \theta_{k+2u,u} < 1$ holds if

$$\delta_{1.25s} + \theta_{1.25s,u} < 1 \tag{2.1.14}$$

This is significantly weaker than Candès and Tao's sufficient condition [15] $\delta_{2s} + \theta_{2s,s} < 1$ and Cai's sufficient conditions [11] $\delta_{1.25s} + \theta_{1.25s,s} < 1$.

Sufficient condition involving only δ_{2s} are very popular in the compressed sensing literature. Here, we obtain a sufficient conditions that uses only δ_{2s} . Using $u \le \frac{s}{4}$ in (2.1.10) and using square root lifting property two times, we have

$$\begin{aligned}
\delta_{1.25s} + \theta_{1.25s,u} &\le \delta_{1.25s} + \sqrt{1.25}\,\theta_{s,u} \\
&\le \delta_{1.25s} + \sqrt{1.25}\,\theta_{s,0.25s} \\
&\le \delta_{2s} + \sqrt{1.25}\sqrt{0.25}\,\theta_{s,s} \\
&= \delta_{2s} + \sqrt{0.3125}\,\theta_{s,s} \\
&\le \delta_{2s} + \sqrt{0.3125}\,\delta_{2s} \\
&= (1 + \sqrt{0.3125})\,\delta_{2s}
\end{aligned}$$

Thus, the inequality $\delta_{1.25s}+\theta_{1.25s,u} < 1$ holds if $\delta_{2s} < \frac{1}{1+\sqrt{0.3125}} \approx 0.6414$. This is significantly weaker than Candès' condition [16] $\delta_{2s} < \sqrt{2}-1 \approx 0.4142$ and Cai's condition $\delta_{2s} \le \frac{1}{1+\sqrt{1.25}} \approx 0.4721$ [11].

2.2 Coherence Based Sufficient Conditions

In this section, we derive a sufficient condition based on coherence of a measurement matrix that works for all three modified ℓ_1 minimization problems. Although random matrices are primarily used for CS recovery problems as they have been proved to be better than the deterministic matrices, it is not always possible to measure an unknown vector randomly. To make matter worse, testing whether a matrix satisfies restricted isometry property of a given order is an NP hard problem [5]. On the other hand, coherence just compares the inner products of pairs of columns and hence is easily accessible. The problems we consider here are still mod-BP, mod-BPDN and mod-DS, coming out of the same models we used for RIP based analysis. Namely, an unknown x is measured by a measurement matrix A giving us $y = Ax$ in noiseless and $y = Ax + w$ in noisy case.

2.2.1 Coherence Based Conditions for BP, BPDN, and DS

Recall that coherence of a matrix with normalized columns a_1, a_2, $\ldots$, a_p is given by $\mu = \max_{i \neq j} |\langle a_i, a_j \rangle|$. We note here some important results in sparse recovery based on coherence. Since the problem P_0 is changed to BP, it is natural to ask a question: Under what condition is the solution of BP also a solution of P_0? The following theorem is an answer.

Theorem 2.2.1. *[29, 31] Let $A \in \mathbb{R}^{m \times p}$ be a full rank matrix. If x_0 is a s-sparse solution of $Ax = b$ with $s < \frac{1}{2}\left(1 + \frac{1}{\mu(A)}\right)$, then x_0 is the unique solution of BP and P_0 simultaneously.*

If x_0 is s-sparse, the condition $s < \frac{1}{2}\left(1 + \frac{1}{\mu(A)}\right)$ or equivalently $\mu < \frac{1}{2s-1}$ surprisingly works for both optimization problems P_0 and BP. This condition has also been used to guarantee the recovery of sparse vectors when applying some direct algorithm such as OMP and IHT. And for BPDN, we have the following theorem.

Theorem 2.2.2. *[29, 31] Let x_0 be a s-sparse vector feasible for $y = Ax+w$ with $\|w\|_2 \le \epsilon$. If $s < \frac{1}{4}\left(1 + \frac{1}{\mu(A)}\right)$ then a solution x_0^ϵ of BPDN satisfies*

$$\|x_0^\epsilon - x_0\|_2^2 \le \frac{4\,\epsilon^2}{1 - \mu(A)\,(4s-1)}$$

The sets K, U, S, E have the same meaning as before. To solve one instance of these problems, we assume $E = \varnothing$. As we saw in theorem 2.2.1 the condition $\mu < \frac{1}{2s-1}$ is sufficient to recover a s-sparse vector x by solving BP. For BPDN, a stronger condition $\mu < \frac{1}{4s-1}$ was used by the same author in [30] to approximate a s-sparse x. It has recently been established [12] that the condition $\mu < \frac{1}{2s-1}$ is sufficient to recover x exactly or stably by solving all three minimization problems BP, BPDN and DS (with bounded noise). Authors in [12] also established that $\mu < \frac{1}{2s-1}$ is sharp by giving an example of a matrix with $\mu = \frac{1}{2s-1}$ where the recovery fails.

In this section, we prove that $\mu < \frac{1}{k+2u-1}$ is sufficient to recover a s-sparse x exactly by solving modified form of BP and stably by solving mod-BPDN and mod-DS. Note that $k + 2u < 2s$ hence $\frac{1}{2s-1} < \frac{1}{k+2u-1}$. Thus, the condition $\mu < \frac{1}{2s-1}$ implies $\mu < \frac{1}{k+2u-1}$ and the converse doesn't hold. This proves that our condition is weaker and helps to break the coherence barrier. The condition $\mu < \frac{1}{2s-1}$ gives an upper bound $s < \frac{1}{2}\left(1 + \frac{1}{\mu}\right)$ for the sparsity of a signal that can be recovered by solving BP, BPDN and DS in the deterministic compressed sensing setting. By using $k + u = s$ in $\mu < \frac{1}{k+2u-1}$ we can see that $s < \frac{1}{2}\left(1 + \frac{1}{\mu}\right) + \frac{k}{2}$. This implies that modified ℓ_1-minimization problems help us to recover $\frac{k}{2}$ more sparse signal than the regular ℓ_1-minimization problems.

2.2.2 Some Important Lemmas

We note here two important inequalities that will be used later. For a k-sparse c, the coherence μ satisfies [34]

$$(1 - (k-1)\mu)\|c\|_2^2 \le \|Ac\|_2^2 \le (1 + (k-1)\mu)\|c\|_2^2 \tag{2.2.1}$$

Let x^* be the minimizer of mod-BP or mod-BPDN or mod-DS and x a vector which is feasible for the same problem. We denote the difference by $h = x - x^*$. The sets of indices K, U, R are as above. We also use the fact that if x_{K^c} is u-sparse and $U = \text{supp}(x_{K^c})$, the inequality (2.1.2) simplifies to

$$\|h_R\|_1 \leq \|h_U\|_1 \tag{2.2.2}$$

We use a condition $\mu < \frac{1}{k-1}$ in the following lemma and replace it with stronger condition $\mu < \frac{1}{k+2u-1}$ later.

Lemma 2.2.3. *Let x^* be feasible for mod-BP (or mod-BPDN or mod-DS), and let x_{K^c} be u-sparse with support $U = supp(x_{K^c})$. If $\mu < \frac{1}{k-1}$ and $h = x - x^*$, we have*

$$(1 - (k-1)\mu)\|h_K\|_2^2 \leq |\langle Ah, Ah_K\rangle| + 2\mu\sqrt{uk}\|h_U\|_2 \cdot \|h_K\|_2 \tag{2.2.3}$$

Proof. Proof follows easily by using triangle inequality, definition of μ and lemma 2.1.4

$$\begin{aligned}
|\langle Ah, Ah_K\rangle| &= |\langle Ah_K, Ah_K\rangle + \langle Ah_U, Ah_K\rangle + \langle Ah_R, Ah_K\rangle| \\
&\geq |\langle Ah_K, Ah_K\rangle| - |\langle Ah_U, Ah_K\rangle| - |\langle Ah_R, Ah_K\rangle| \\
&\geq (1-(k-1)\mu)\|h_K\|_2^2 - \sum_{i\in U}\sum_{j\in K} |\langle a_i, a_j\rangle| \cdot |h(i)| \cdot |h(j)| - \sum_{i\in R}\sum_{j\in K} |\langle a_i, a_j\rangle| \cdot |h(i)| \cdot |h(j)| \\
&\geq (1-(k-1)\mu)\|h_K\|_2^2 - \mu\|h_U\|_1\|h_K\|_1 - \mu\|h_R\|_1\|h_K\|_1 \\
&\geq (1-(k-1)\mu)\|h_K\|_2^2 - 2\mu\|h_U\|_1\|h_K\|_1 \\
&\geq (1-(k-1)\mu)\|h_K\|_2^2 - 2\mu\sqrt{uk}\|h_U\|_2\|h_K\|_2
\end{aligned}$$

Hence the lemma follows. □

Lemma 2.2.4. *Let x^* and x be as in lemma 2.2.3. If $\mu < \frac{1}{k-1}$ and $h = x - x^*$, we have the following. If x^* is the minimizer of mod-BP, we have*

$$\|h_K\|_2 \leq \frac{2\mu\sqrt{uk}}{1-(k-1)\mu}\|h_U\|_2, \tag{2.2.4}$$

if x^* is the minimizer of mod-BPDN, we have

$$\|h_K\|_2 \leq \frac{2\mu\sqrt{uk}}{1-(k-1)\mu}\|h_U\|_2 + 2\epsilon\frac{\sqrt{1+(k-1)\mu}}{1-(k-1)\mu} \tag{2.2.5}$$

and if x^* is the minimizer of mod-DS, we have

$$\|h_K\|_2 \leq \frac{2\mu\sqrt{uk}}{1-(k-1)\mu}\|h_U\|_2 + \frac{2\sqrt{k}\epsilon}{1-(k-1)\mu} \tag{2.2.6}$$

Proof. If x^* is the minimizer of mod-BP, we have $Ah = 0$. Using this fact in (2.2.3) gives

$$(1-(k-1)\mu)\|h_K\|_2^2 \leq 2\mu\sqrt{uk}\|h_K\|_2\|h_U\|_2$$

and hence

$$\|h_K\|_2 \leq \frac{2\mu\sqrt{uk}}{1-(k-1)\mu}\|h_U\|_2.$$

If x^* is the minimizer of mod-BPDN then $\|Ah\|_2 \leq 2\epsilon$, so from (2.2.3) again,

$$\begin{aligned}(1-(k-1)\mu)\|h_K\|_2^2 &\leq 2\mu\sqrt{uk}\|h_K\|_2\|h_U\|_2 + 2\epsilon\|Ah_K\|_2 \\ &\leq 2\mu\sqrt{uk}\|h_K\|_2\|h_U\|_2 + 2\epsilon\sqrt{1+(k-1)\mu}\|h_K\|_2\end{aligned}$$

that is,

$$(1-(k-1)\mu)\|h_K\|_2 \leq 2\mu\sqrt{uk}\|h_U\|_2 + 2\epsilon\sqrt{1+(k-1)\mu}$$

Thus,

$$\|h_K\|_2 \leq \frac{2\mu\sqrt{uk}\|h_U\|_2}{1-(k-1)\mu} + \frac{2\epsilon\sqrt{1+(k-1)\mu}}{1-(k-1)\mu}$$

Finally, if x^* minimizes mod-DS, then $Ah = Ax - Ax^* = Ax - y + y - Ax^*$. Note that $\|A'_K(Ax-y)\|_2 \leq \sqrt{k}\epsilon$ and $\|A'_K(Ax^*-y)\|_2 \leq \sqrt{k}\epsilon$. We also have

$$\|A'Ah\|_\infty = \|A'(Ax-y+y-Ax^*)\|_\infty \leq \|A'(Ax-y)\|_\infty + \|A'(y-Ax^*)\|_\infty \leq 2\epsilon$$

So,

$$\begin{aligned}|\langle Ah, Ah_K\rangle| &= |\langle A'_K(Ax-y+y-Ax^*), h_K\rangle| \\ &\leq \left(\|A'_K(Ax-y)\|_2 + \|A'_K(y-Ax^*)\|_2\right)\|h_K\|_2 \\ &\leq 2\sqrt{k}\epsilon\|h_K\|_2\end{aligned}$$

Thus,

$$|\langle Ah, Ah_K \rangle| \leq 2\sqrt{k}\epsilon \|h_K\|_2 \tag{2.2.7}$$

Using inequality (2.2.7) in (2.2.3), we have,

$$\begin{aligned}(1-(k-1)\mu)\|h_K\|_2^2 &\leq 2\mu\sqrt{uk}\|h_K\|_2\|h_U\|_2 + |\langle Ah, Ah_K \rangle| \\ &\leq 2\mu\sqrt{uk}\|h_K\|_2\|h_U\|_2 + 2\sqrt{k}\epsilon\|h_K\|_2\end{aligned}$$

Dividing both side by $(1-(k-1)\mu)\|h_K\|_2$ gives the desired result. □

Lemma 2.2.5. *Let x^* and x be as in lemma 2.2.3. If $\mu < \frac{1}{k+2u-1}$ and $h = x - x^*$ then we have*

$$(1-(2u-1)\mu)\|h_U\|_2^2 \leq \mu\sqrt{uk}\|h_K\|_2\|h_U\|_2 + |\langle Ah, Ah_U \rangle| \tag{2.2.8}$$

Further, if x^* minimizes mod-BP then $h_U = 0$, and if x^* minimizes mod-BPDN then

$$\|h_U\|_2 \leq \alpha(\mu, u, k)\epsilon$$

where

$$\alpha(\mu, u, k) = 2\left[\frac{\mu\sqrt{uk}\,(1+(k-1)\mu) + \sqrt{1+(u-1)\mu}\,(1-(k-1)\mu)}{(\mu+1)\,(1-\mu\,(k+2u-1))}\right] \tag{2.2.9}$$

Proof. Proof of the first part of lemma follows directly from lemma 2.2.3 by replacing K by U. Let x^* be a minimizer of mod-BP. Using (2.2.8), lemma 2.2.3 and $Ah = 0$, we have

$$\begin{aligned}(1-(2u-1)\mu)\|h_U\|_2^2 &\leq \mu\sqrt{uk}\|h_K\|_2\|h_U\|_2 \\ &\leq \mu\sqrt{uk}\,\frac{2\mu\sqrt{uk}\,\|h_U\|_2^2}{1-(k-1)\mu} \\ &\leq \frac{2\mu^2 uk}{(1-(k-1)\mu)}\,\|h_U\|_2^2\end{aligned}$$

Thus,

$$\left[1-(2u-1)\mu - \frac{2\mu^2 uk}{1-(k-1)\mu}\right]\|h_U\|_2^2 \leq 0$$

implying

$$\left[\frac{(\mu+1)\,(1-\mu\,(k+2u-1))}{1-(k-1)\mu}\right]\|h_U\|_2^2 \le 0 \tag{2.2.10}$$

As $\mu < \frac{1}{k+2u-1}$, we must have $1-\mu\,(k+2u-1) > 0$. Thus the number on the left hand side of inequality 2.2.10 is positive. Hence $\|h_U\|_2$ must be zero for the inequality to hold, i.e, $h_U = 0$. This proves the second part of the lemma.

Finally, if x^* is the minimizer of mod-BPDN, using (2.2.5), lemma 2.2.3 and the first part of lemma 2.2.5, we have

$$\begin{aligned}
(1-(2u-1)\mu)\|h_U\|_2^2 &\le \mu\sqrt{uk}\;\|h_K\|_2\|h_U\|_2 + 2\epsilon\|Ah_U\|_2 \\
&\le \mu\sqrt{uk}\left(\frac{2\mu\sqrt{uk}}{1-(k-1)\mu}\|h_U\|_2^2 + \frac{2\epsilon\sqrt{1+(k-1)\mu}}{1-(k-1)\mu}\|h_U\|_2\right) + 2\epsilon\sqrt{1+(u-1)\mu}\|h_U\|_2 \\
&\le \frac{2\mu^2 uk}{(1-(k-1)\mu)}\;\|h_U\|_2^2 + 2\epsilon\left[\frac{\mu\sqrt{uk}\,(1+(k-1)\mu)}{1-(k-1)\mu} + \sqrt{1+(u-1)\mu}\right]\|h_U\|_2
\end{aligned}$$

Therefore,

$$\left[1-(2u-1)\mu - \frac{2\mu^2 uk}{(1-(k-1)\mu)}\right]\|h_U\|_2 \le 2\epsilon\left[\frac{\mu\sqrt{uk\,(1+(k-1)\mu)}}{1-(k-1)\mu} + \sqrt{1+(u-1)\mu}\right]$$

i.e.

$$\left[\frac{(\mu+1)\,(1-\mu\,(k+2u-1))}{1-(k-1)\mu}\right]\|h_U\|_2 \le 2\epsilon\left[\frac{\mu\sqrt{uk\,(1+(k-1)\mu)}}{1-(k-1)\mu} + \sqrt{1+(u-1)\mu}\right]$$

implying

$$\|h_U\|_2 \le \alpha(\mu,u,k)\epsilon$$

where $\alpha(\mu,u,k)$ is as in (2.2.9) Note that for $\mu < \frac{1}{k+2u-1}$, the numbers $1-\mu\,(k+2u-1) > 0$ and $1-\mu\,(k-1) > 0$. This proves the lemma. □

2.2.3 Main Theorems

Theorem 2.2.6. *Let x be feasible for mod-BP and let x_{K^c} be u-sparse. If $\mu < \frac{1}{k+2u-1}$ then mod-BP recovers x exactly.*

Proof. Let x^* be a solution of mod-BP. Define $h = x - x^*$. We will show that $h = 0$.

$$\begin{aligned}
0 &= \|Ah\|_2^2 \\
&= |\langle Ah, Ah\rangle| \\
&= \sum_i \sum_j |\langle a_i, a_j\rangle| \cdot |h(i)| \cdot |h(j)| \\
&= \sum_i |h(i)|^2 + \sum_{i \neq j} |\langle a_i, a_j\rangle| \cdot |h(i)| \cdot |h(j)| \\
&\geq \|h\|_2^2 - \mu \sum_{i \neq j} |h(i)| \cdot |h(j)| \\
&= (1+\mu)\|h\|_2^2 - \mu \sum_i \sum_j |h(i)| \cdot |h(j)| \\
&= (1+\mu)\|h\|_2^2 - \mu\|h\|_1^2
\end{aligned}$$

so,

$$\|h\|_2 \leq \sqrt{\frac{\mu}{\mu+1}}\|h\|_1 \tag{2.2.11}$$

Using (2.2.2), (2.2.11) and (2.2.4), we have,

$$\begin{aligned}
\|h\|_2 &\leq \sqrt{\frac{\mu}{\mu+1}}\|h_R + h_U + h_K\|_1 \\
&\leq \sqrt{\frac{\mu}{\mu+1}}(2\|h_U\|_1 + \|h_K\|_1) \\
&\leq \sqrt{\frac{\mu}{\mu+1}}(2\sqrt{u}\ \|h_U\|_2 + \sqrt{k}\|h_K\|_2) \\
&= \sqrt{\frac{\mu}{\mu+1}}\left(2\sqrt{u}\|h_U\|_2 + \sqrt{k}\frac{2\mu\sqrt{uk}\ \|h_U\|_2}{1-(k-1)\mu}\right) \\
&= 2\sqrt{\frac{\mu u}{\mu+1}}\left(1 + \frac{\mu k}{1-(k-1)\mu}\right)\|h_U\|_2 \\
&= 2\sqrt{\frac{\mu u}{\mu+1}}\left(\frac{1+\mu}{1-(k-1)\mu}\right)\|h_U\|_2 \\
&= \left(\frac{2\sqrt{\mu u(1+\mu)}}{1-(k-1)\mu}\right)\|h_U\|_2
\end{aligned}$$

Since $\mu < \frac{1}{k+2u-1}$, lemma 2.2.5 implies $h_U = 0$ which implies $h = 0$.

This proves the theorem. □

The following theorem deals with noisy case.

Theorem 2.2.7. *Let x be feasible for mod-BPDN and let x_{K^c} be u-sparse. If $\mu < \frac{1}{k+2u-1}$ then the minimizer x^* of mod-BPDN satisfies*

$$\|x - x^*\|_2 \leq C(\mu, u, k)\epsilon$$

where $C(\mu, u, k)$ is a constant depending on μ, u and k.

Proof. Proceeding as in the proof of theorem 2.2.6 and using $\|Ah\|_2 \leq 2\epsilon$, we have,

$$\begin{aligned}
(1+\mu)\|h\|_2^2 &\leq \mu\|h\|_1^2 + 4\epsilon^2 \\
&\leq \mu\|h_K + h_U + h_R\|_1^2 + 4\epsilon^2 \\
&\leq \mu\left(\|h_K\|_1 + 2\|h_U\|_1\right)^2 + 4\epsilon^2 \\
&\leq \mu(\sqrt{k}\|h_K\|_2 + 2\sqrt{u}\|h_U\|_2)^2 + 4\epsilon^2 \\
&\leq \mu\left(\sqrt{k}\frac{2\mu\sqrt{uk}\|h_U\|_2}{1-(k-1)\mu} + 2\epsilon\sqrt{k}\frac{\sqrt{1+(k-1)\mu}}{1-(k-1)\mu} + 2\sqrt{u}\|h_U\|_2\right)^2 + 4\epsilon^2 \\
&= \mu\left[2\sqrt{u}\left(1 + \frac{\mu k}{1-(k-1)\mu}\right)\|h_U\|_2 + 2\epsilon\sqrt{k}\frac{\sqrt{1+(k-1)\mu}}{1-(k-1)\mu}\right]^2 + 4\epsilon^2 \\
&\leq \mu\left[\frac{2\sqrt{u}(1+\mu)}{1-(k-1)\mu}\,\alpha(\mu,u,k)\epsilon + 2\epsilon\frac{\sqrt{k\,(1+(k-1)\mu)}}{1-(k-1)\mu}\right]^2 + 4\epsilon^2 \\
&= \left[1 + \mu\left(\frac{\sqrt{u}(1+\mu)\cdot\alpha(\mu,u,k) + \sqrt{k\,(1+(k-1)\mu)}}{1-(k-1)\mu}\right)^2\right]4\epsilon^2
\end{aligned}$$

This proves that $\|x - x^*\|_2 \leq C(\mu, u, k)\,\epsilon$, where

$$C(\mu,u,k) = 2\sqrt{\frac{1}{1+\mu}\left[1 + \mu\left(\frac{\sqrt{u}(1+\mu)\cdot\alpha(\mu,u,k) + \sqrt{k\,(1+(k-1)\mu)}}{1-(k-1)\mu}\right)^2\right]}$$

and $\alpha(\mu, u, k)$ is given by (2.2.9). This establishes the theorem. □

We now turn to mod-DS.

Theorem 2.2.8. *Let x be feasible for mod-DS and let x_{K^c} be u-sparse. If $\mu < \frac{1}{k+2u-1}$ then a minimizer x^* of mod-DS satisfies*

$$\|x - x^*\|_2 \le D(\mu, u, k)\epsilon$$

where $D(\mu, u, k)$ is a constant depending on μ, u, k.

Proof. Let $h = x - x^*$ as usual. Then $Ah = Ax - Ax^* = Ax - y + y - Ax^*$. Also note that $\|A'_U(Ax - y)\|_2 \le \sqrt{u}\epsilon$, $\|A'_U(Ax^* - y)\|_2 \le \sqrt{u}\epsilon$ and

$$\|A'Ah\|_\infty = \|A'(Ax - y + y - Ax^*)\|_\infty \le \|A'(Ax - y)\|_\infty + \|A'(y - Ax^*)\|_\infty \le 2\epsilon$$

So,

$$\begin{aligned} |\langle Ah, Ah_U\rangle| &= |\langle A'_U(Ax - y + y - Ax^*), h_U\rangle| \\ &\le (\|A'_U(Ax - y)\|_2 + \|A'_U(y - Ax^*)\|_2)\, \|h_U\|_2 \\ &\le 2\sqrt{u}\epsilon\|h_U\|_2 \end{aligned}$$

Again, from lemma 2.2.5

$$\begin{aligned} (1 - (2u - 1)\mu)\|h_U\|_2^2 &\le \mu\sqrt{uk}\|h_K\|_2\|h_U\|_2 + |\langle Ah, Ah_U\rangle| \\ &= \mu\sqrt{uk}\|h_K\|_2\|h_U\|_2 + |\langle A'_U Ah, h_U\rangle| \\ &\le \mu\sqrt{uk}\|h_K\|_2\|h_U\|_2 + 2\sqrt{u}\epsilon\|h_U\|_2 \end{aligned}$$

Canceling $\|h_U\|_2$ and using inequality (2.2.6) implies

$$\begin{aligned}(1-(2u-1)\mu)\|h_U\|_2 &\le \mu\sqrt{uk}\|h_K\|_2 + 2\sqrt{u}\epsilon \\ &\le \mu\sqrt{uk}\left(\frac{2\mu\sqrt{uk}\|h_U\|_2}{1-(k-1)\mu} + \frac{2\sqrt{k}\epsilon}{1-(k-1)\mu)}\right) + 2\sqrt{u}\epsilon \\ &\le \frac{2\mu^2 uk\ \|h_U\|_2}{1-(k-1)\mu} + \left(\frac{2\epsilon\sqrt{u}\mu k}{1-(k-1)\mu} + 2\sqrt{u}\epsilon\right) \\ &= \frac{2\mu^2 uk\ \|h_U\|_2}{1-(k-1)\mu} + \frac{2\epsilon\sqrt{u}(1+\mu)}{1-(k-1)\mu}\end{aligned}$$

i.e.

$$\left[1-(2u-1)\mu - \frac{2\mu^2 uk}{(1-(k-1)\mu)}\right]\|h_U\|_2 \le \frac{2\epsilon\sqrt{u}(1+\mu)}{1-(k-1)\mu}$$

$$\left[\frac{(1+\mu)(1-(k+2u-1)\mu)\ \|h_U\|_2}{(1-(k-1)\mu)}\right] \le \frac{2\epsilon\sqrt{u}(1+\mu)}{1-(k-1)\mu}$$

This proves that

$$\|h_U\|_2 \le \frac{2\epsilon\sqrt{u}}{1-(k+2u-1)\mu} \tag{2.2.12}$$

Also note that

$$\begin{aligned}\|h\|_1 &= \|h_R + h_U + h_K\|_1 \\ &\le 2\|h_U\|_1 + \|h_K\|_1 \\ &\le 2\sqrt{u}\|h_U\|_2 + \sqrt{k}\|h_K\|_2 \\ &\le 2\sqrt{u}\|h_U\|_2 + \frac{2\mu\sqrt{u}\ k\ \|h_U\|_2}{1-(k-1)\mu} + \frac{2\sqrt{k}\epsilon}{1-(k-1)\mu} \\ &\le \frac{2\sqrt{u}(1+\mu)\|h_U\|_2}{1-(k-1)\mu} + \frac{2\sqrt{k}\epsilon}{1-(k-1)\mu} \\ &\le \beta(\mu,u,k)\epsilon\end{aligned}$$

where

$$\beta(\mu,u,k) = \frac{4u(1+\mu)}{(1-(k-1)\mu)(1-(k+2u-1)\mu)} + \frac{2\sqrt{k}}{1-(k-1)\mu} \tag{2.2.13}$$

and the last part of the inequality follows from (2.2.12). Once again, using the proof of theorem 2.2.7,

$$\begin{aligned}
(1+\mu)\|h\|_2^2 &\leq \mu\|h\|_1^2 + \|Ah\|_2^2 \\
&\leq \mu\|h\|_1^2 + |\langle A'Ah, h\rangle| \\
&\leq \mu\|h\|_1^2 + \|A'Ah\|_\infty \|h\|_1 \\
&\leq \mu\left(\beta(\mu,u,k)\epsilon\right)^2 + 2\epsilon^2\beta(\mu,u,k) \\
&\leq \epsilon^2\left(\mu\beta^2 + 2\beta\right)
\end{aligned}$$

Note that we have used $|\langle a,b\rangle| \leq \|a\|_\infty \cdot \|b\|_1$ in the third inequality. This implies that $\|h\|_2 \leq D(\mu,u,k)\epsilon$ where

$$D(\mu,u,k) = \sqrt{\frac{\mu\beta^2 + 2\beta}{1+\mu}}$$

and β as in (2.2.13). Theorem is now proved. □

CHAPTER 3. NECESSARY AND SUFFICIENT CONDITIONS

In this chapter, we derive dual and Karush-Kuhn-Tucker (KKT) optimality conditions for modified ℓ_1-minimization problems. We also prove an algebraic necessary and sufficient conditions for the uniqueness of solution of mod-BP using theory of convex optimization.

3.1 Convex Optimization

In this section, we review the basics of convex optimization, as it is used extensively in this chapter. For details, a standard reference is the book by Boyd and Vandenberghe [8].

A *convex optimization problem* has the form

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le 0, \quad i = 1, 2, \dots, k \\ & Ax = y \end{aligned} \tag{OP}$$

where $x \in \mathbb{R}^p$, f_i are convex functions from $\mathbb{R}^p$ to $\mathbb{R}$, and A is an $m \times p$ matrix. The *Lagrangian* of this problem is defined as

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{k} \lambda_i f_i(x) + \nu'(y - Ax)$$

where $\lambda \in \mathbb{R}^k$, $\nu \in \mathbb{R}^m$.

The *Lagrangian dual function* is defined as

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu).$$

and the *dual optimization problem* corresponding to (OP) is

$$\begin{aligned} \text{maximize} \quad & g(\lambda, \nu) \\ \text{subject to} \quad & \lambda \succcurlyeq 0. \end{aligned} \tag{DP}$$

If x^* and λ^*, ν^* are optimal solutions to primal and dual convex optimization problems, then $f(x^*) = g(\lambda^*, \nu^*)$. This is called *strong duality.*

Lemma 3.1.1. *If the functions f_i are differentiable, then x^*, λ^*, and ν^* are optimal for OP and DP respectively if and only if the* KKT conditions *(Karush-Kuhn-Tucker) are satisfied. These are*

Primal constraints:	$f_i(x^*) \le 0, \quad i = 1, 2, \dots, k$
	$Ax^* = y$
Dual constraints:	$\lambda^* \ge 0$ *where* $\ge$ *is the componentwise inequality.*
Complementary slackness:	$\lambda_i^* f_i(x^*) = 0, \quad i = 1, 2, \dots, k$
Gradient of the Lagrangian with respect to x vanishes:	$\nabla_x L(x^*, \lambda^*, \nu^*) = 0$

Complementary slackness says that for each i, at least one of λ_i^* and $f_i(x^*)$ is zero. The stronger *Goldman-Tucker theorem* [40] states that in the linear programming case (where the f_i are affine linear functions) the existence of an optimal solution implies that there exists an optimal solution where for each i, precisely one of λ_i^* and $f_i(x^*)$ is zero.

One attractive feature of BP and DS is that these problems can be written as linear programming (LP) problems and can be solved readily by existing LP solvers. BPDN is a quadratic programming problem. We will derive LP form of mod-BP, mod-DS and also find the dual of these modified problems. It is possible to derive the dual problems directly from BP and mod-BP, but in order to apply the Goldman-Tucker theorem [69] later we have to reformulate them as linear programs first. We do this for mod-BP and mod-DS; regular BP and DS are then just special cases.

3.2 Convex Optimization View of mod-BP

Recall that mod-BP is the following optimization problem: Given an estimate K of the support of the unknown vector x,

$$\text{minimize } \|x_{K^c}\|_1 \quad \text{subject to } y = Ax. \tag{mod-BP}$$

The equivalent linear program for this is

$$\begin{aligned} \text{minimize} \quad c't \quad \text{subject to} \quad x - t &\preccurlyeq 0 \\ -x - t &\preccurlyeq 0 \\ Ax &= y \end{aligned} \tag{mod-LP}$$

where $c \in \mathbb{R}^p$ is given by

$$c_i = \begin{cases} 0 & \text{if } i \in K, \\ 1 & \text{if } i \notin K. \end{cases}$$

where $\preccurlyeq$ is componentwise inequality.

Lemma 3.2.1. *Mod-BP and mod-LP have same optimal solution in the following sense. A vector x is an optimal solution of mod-BP iff (x, t) is an optimal solution of mod-LP for any t with*

$$\begin{aligned} t_i &= |x_i| \qquad \text{if } i \notin K, \\ t_i &\geq |x_i| \qquad \text{if } i \in K. \end{aligned}$$

Proof. Assume x is optimal for mod-BP, then (x, t) with $t_i \geq |x_i|$ for all i is feasible for mod-LP. To minimize $c't$ we must have $t_i = |x_i|$ for $i \notin K$. If (x, t) is optimum for mod-LP, then $t_i = |x_i|$ for $i \notin K$, and thus $\|x_{K^c}\|_1 = c't$. This proves the lemma. □

3.2.1 Dual and Optimality Conditions of mod-BP

Now, we work on to find the dual of mod-BP. The Lagrangian of mod-LP is

$$\begin{aligned} L(x,t,\lambda_1,\lambda_2,\nu) &= c't + \lambda_1'(x-t) + \lambda_2'(-x-t) + \nu'(y-Ax) \\ &= \nu'y + (\lambda_1' - \lambda_2' - \nu'A)x + (c' - \lambda_1' - \lambda_2')t \end{aligned}$$

The Lagrangian dual function is

$$\begin{aligned} g(\lambda_1,\lambda_2,\nu) &= \min_{x,t} L(x,t,\lambda_1,\lambda_2,\nu) \\ &= \begin{cases} \nu'y & \text{if } c - \lambda_1 - \lambda_2 = 0,\ \lambda_1 - \lambda_2 - A'\nu = 0, \\ -\infty & \text{otherwise.} \end{cases} \end{aligned}$$

The first formulation of dual problem is ,

$$\begin{aligned} \text{maximize} \quad & \nu'y \\ \text{subject to} \quad & \lambda_1 \succcurlyeq 0 \\ & \lambda_2 \succcurlyeq 0 \\ & \lambda_1 - \lambda_2 = A'\nu \\ & \lambda_1 + \lambda_2 = c. \end{aligned} \tag{3.2.1}$$

Simplifying the constraints further:

For $i \in K$, $c_i = 0$. So, $(\lambda_1 + \lambda_2)_i = 0$. Since $(\lambda_1)_i \geq 0$ and $(\lambda_2)_i \geq 0$, we must have, $(\lambda_1)_i = (\lambda_2)_i = 0$. For $i \notin K$, $c_i = 1$ and hence $(\lambda_1)_i + (\lambda_2)_i = 1$. As $(\lambda_1)_i \geq 0$ and $(\lambda_2)_i \geq 0$ we have $-1 \leq (\lambda_1)_i - (\lambda_2)_i \leq 1$ which implies that $-1 \leq (A'\nu)_i \leq 1$.

It is easy to see that for each number in $[-1,1]$, we can find suitable λ_1 and λ_2. The final form of dual of mod-BP can then be written as

$$\begin{aligned} \underset{\nu}{\text{maximize}} \quad \nu'y \quad \text{subject to} \quad & |a_i'\nu| = 0 \ \text{ for } \ i \in K \\ & |a_i'\nu| \leq 1 \ \text{ for } \ i \notin K \end{aligned} \tag{dual mod-LP}$$

Now, we assume pairs (x, t) and $(\lambda_1, \lambda_2, \nu)$ to be optimal for mod-LP and its dual (3.2.1) and find complementary slackness condition and other optimality conditions. The complementary slackness condition becomes

$$(\lambda_1)_i(x_i - t_i) = 0 \quad \text{and} \quad (\lambda_2)_i(-x_i - t_i) = 0 \ \forall i = 1, \dots, p \tag{3.2.2}$$

For $i \in K$, $(\lambda_1)_i = (\lambda_2)_i = 0$ so complementary slackness is automatically satisfied. If $i \notin K$, three scenarios arise, namely, $x_i > 0$, $x_i < 0$ and $x_i = 0$. For $x_i > 0$, $t_i = x_i$ and hence one must have $(\lambda_2)_i = 0$ for (3.2.2) to hold. Now using equality constraints in (3.2.1) gives $(\lambda_1)_i = 1$ implying $a_i'\nu = 1$. Similarly, for $x_i < 0$, we get $(\lambda_2)_i = 1$ implying $a_i'\nu = -1$ and, for $x_i = 0$ equations (3.2.2) is automatically satisfied. To summarize, complementary slackness condition for mod-BP becomes

$$a_i'\nu = \text{sign}(x_i) \quad \text{for} \quad i \in K,\ x_i \neq 0 \tag{3.2.3}$$

$$|a_i'\nu| \leq 1 \qquad \text{for} \quad i \in K, x_i = 0 \tag{3.2.4}$$

Goldman-Tucker theorem [40] says that if there is an optimal solution, there is one with the property that only one of $(\lambda_1)_i$ or $(x_i - t_i)$ is zero and only one of $(\lambda_2)_i$ or $(-x_i - t_i)$ is zero in the complementary slackness condition (3.2.2). For $i \in K$, we have $(\lambda_1)_i = 0 = (\lambda_2)_i$. So, we can increase t on K to make $x_i - t_i \neq 0$ and $-x_i - t_i \neq 0$. For $i \notin K$ and $x_i \neq 0$, we saw that exactly one of these is automatically nonzero. And for $x_i = 0$, we need $(\lambda_1)_i, (\lambda_2)_i > 0$.

Thus, applying Goldman-Tucker theorem in complementary slackness condition, we get

$$a_i'\nu = \text{sign}(x_i) \quad \text{for} \quad i \in K,\ x_i \neq 0 \tag{3.2.5}$$

$$|a_i'\nu| < 1 \qquad \text{for} \quad i \in K, x_i = 0 \tag{3.2.6}$$

The linear programming form LP of BP is the same, with c replaced by $c = (1, 1, \dots, 1)'$. Gradient condition for mod-LP reduces to the last two equations in (3.2.1).

Putting all these preceding discussion in the form of a theorem, we have the following.

Theorem 3.2.2. *x^* is an optimal solution of mod-BP if and only if there exists a ν^* such that x^*, $t^* = |x^*|$ and ν^* are feasible for mod-BP and dual-mod-BP, and*

(a) $a_i'\nu = 0$ for all $i \in K$

(b) $a_i'\nu = sign(x_i^*)$ for all $i \in K$, $x_i^* \neq 0$

(c) $|a_i'\nu| < 1$ for all $i \notin K$, $x_i^* = 0$

3.2.2 Necessary and Sufficient Conditions for Modified Basis Pursuit

The following is the main theorem of this chapter. This says that necessary and sufficient conditions similar to that of basis pursuit can be derived for the modified basis pursuit problem.

Theorem 3.2.3. *Let x be feasible for mod-BP and let x_{K^c} be u-sparse. Assume that $A_{K\cup U}$ has full rank $k+u$, then x^* is the unique solution of $mod-BP$ if and only if there exists a $\nu \in \mathbb{R}^m$ with the following conditions*

(a) $a_i'\nu = 0$ for all $i \in K$

(b) $a_i'\nu = sign(x_i^*)$ for all $i \in U$

(c) $|a_i'\nu| < 1$ for all $i \notin K \cup U$

Proof. We just need to prove that x^* is unique. The other part of the theorem follows from theorem 3.2.2. To that end, let α be another minimum point and the set U, dual vector v^* have been chosen based on x^*.

$$\begin{aligned}
\|\alpha_{K^c}\|_1 &= \sum_{i\in K^c} |\alpha_i| \\
&= \sum_{i\in U} |\alpha_i| + \sum_{i\notin(K\cup U)} |\alpha_i| \\
&\geq \sum_{i\in U} \text{sign}(x_i^*)(x_i^* + \alpha_i - x_i^*) + \sum_{i\notin(K\cup U)} \nu' a_i \alpha_i \\
&= \sum_{i\in U} |x_i^*| + \sum_{i\in U} \nu' a_i(\alpha_i - x_i^*) + \sum_{i\notin(K\cup U)} \nu' a_i \alpha_i \\
&= \sum_{i\in U} |x_i^*| + \sum_{i\in(K\cup U)} \nu' a_i(\alpha_i - x_i^*) + \sum_{i\notin(K\cup U)} \nu' a_i \alpha_i \\
&= \sum_{i\in K^c} |x_i^*| + \nu' A(\alpha - x^*) \\
&= |x_{K^c}^*\|_1
\end{aligned}$$

Now putting the equality sign in above inequalities, we get $\alpha_i = 0$ for all $i \notin (K \cup U)$. Since x_i^* is also zero outside $(K \cup U)$, we have $A_{K\cup U} x_{K\cup U}^* = A_{K\cup U}\alpha_{K\cup U}$ As the rank of $A_{K\cup U}$ is $k + u$, We must have $x_{K\cup U}^* = \alpha_{K\cup U}$. This proves that x^* is unique. □

So, we have derived an algebraic necessary and sufficient condition for the uniqueness of a sparse vector as a solution of modified basis pursuit. We will discuss some preliminary results we have derived on geometric necessary and sufficient conditions in chapter 5.

Author in [32] recently came up with a different necessary and sufficient condition for the recovery of s-sparse x supported on S, using a special set defined as follows.

$$G = \{x \in \mathbb{R}^p : \text{supp}(x) \subset S, \text{rank}(A_S) = s, |a_j' z| < 1 \text{ for all } j \notin S\}$$

By following the same techniques, we define a set G_K as follows

$$G_K = \{x \in \mathbb{R}^p : \text{supp}(x) \subset (T \cup U), \text{rank}(A_{K\cup U}) = k + u, |a_j' z| < 1 \text{ for all } j \notin K \cup U\}$$

where $z = MA_U(A_U' M A_U)^{-1}\text{sign}(x_U^*)$ and $M = I - A_K(A_K' A_K)^{-1} A_K'$. Then we prove the more precise form of the necessity part of the above theorem.

Theorem 3.2.4. *If $x^* \in \bar{G}_K$ then x^* is the unique solution of mod-BP.*

Proof. By theorem 3.2.3, we see that any $x \in G_K$ is the unique solution of mod-BP. Further, as z depends only on the sign and support of x, the set G_K is the union of cones of varying dimension. Hence any vector in $\bar{G}_K$ can be extended to a vector in G_K. Thus, for $x^* \in \bar{G}_K$, there is a v^* with different support from x^* and $x^* + v^* \in G_K$.

If x^* is not the unique optimum, there is $w^* \in \mathbb{R}^p$ such that $Aw^* = Ax^*$ and $\|w^*_{K^c}\|_1 = \|x^*_{K^c}\|_1$.

Define $\hat{v} = w^* + v^*$ then $A\hat{v} = A(w^* + v^*) = A(x^* + v^*)$. Since $x^* + v^* \in G_K$, $x^* + v^*$ is the unique optimum for mod-BP. We must have $\|(x^* + v^*)_{K^c}\|_1 < \|\hat{v}_{K^c}\|_1$

Here

$$\begin{aligned} \|(x^* + v^*)_{K^c}\|_1 &= \|x^*_{K^c}\|_1 + \|v^*_{K^c}\|_1 \\ &= \|w^*_{K^c}\|_1 + \|v^*_{K^c}\|_1 \\ &\geq \|w^*_{K^c} + v^*_{K^c}\|_1 \\ &= \|(w^* + v^*)_{K^c}\|_1 \\ &= \|\hat{v}_{K^c}\|_1 \end{aligned}$$

This implies $\|(x^* + v^*)_{K^c}\|_1 = \|\hat{v}_{K^c}\|_1$ which contradicts the optimality of $x^* + v^*$. So, we must have $x^* + v^* = w^* + v^*$, i.e., $x^* = w^*$. □

3.3 LP Form and Dual of mod-DS

The name Dantzig Selector was chosen to give tribute to the founder of simplex algorithm which solves linear programming. Following the same techniques as in mod-BP case, the problem mod-DS

$$\underset{x}{\text{minimize}} \ \|x_{K^c}\|_1 \quad \text{subject to} \ \ \|A'(y - Ax)\|_\infty \leq \epsilon \qquad \text{(mod-DS)}$$

can be easily written as a linear programming problem

$$\underset{x,t}{\text{minimize}}\; c't \quad \text{subject to} \quad A'(y - Ax) - 1 \cdot \epsilon \preccurlyeq 0 \qquad \text{(LP mod-DS)}$$

$$-A'(y - Ax) - 1 \cdot \epsilon \preccurlyeq 0$$

$$x - t \preccurlyeq 0$$

$$-x - t \preccurlyeq 0$$

where $1 = (1, 1, \ldots, 1)'$ and

$$c_i = \begin{cases} 0 & \text{if } i \in K, \\ 1 & \text{if } i \notin K. \end{cases}$$

The Lagrangian for mod-DS is give by

$$\begin{aligned} L(x, t, \lambda_1, \lambda_2, \lambda_3, \lambda_4) &= c't + \lambda_1'(A'(y - Ax) - 1\epsilon) + \lambda_2'(-A'(y - Ax) - 1\epsilon) + \lambda_3'(x - t) + \lambda_4'(-x - t) \\ &= -\lambda_1' \cdot 1\epsilon - \lambda_2' \cdot 1\epsilon + \lambda_1' A'y - \lambda_2' A'y + (-\lambda_1' A'A + \lambda_2' A'A + \lambda_3' - \lambda_4')x + (-\lambda_3' - \lambda_4' + c')t \end{aligned}$$

Hence the dual function is,

$$\begin{aligned} g(\lambda_1, \lambda_2, \lambda_3, \lambda_4) &= \min_{x,t} L(x, t, \lambda_1, \lambda_2, \lambda_3, \lambda_4) \\ &= \begin{cases} -1'(\lambda_1 + \lambda_2)\epsilon + y'A(\lambda_1 - \lambda_2) & \text{if } A'A(\lambda_1 - \lambda_2) = \lambda_3 - \lambda_4,\;\; \lambda_3 + \lambda_4 = c, \\ -\infty & \text{otherwise.} \end{cases} \end{aligned}$$

The first formulation of dual of mod-DS is

$$\underset{\lambda_1,\lambda_2,\lambda_3,\lambda_4}{\text{maximize}}\; y'A(\lambda_1 - \lambda_2) - \epsilon 1'(\lambda_1 + \lambda_2) \quad \text{subject to} \qquad (3.3.1)$$

$$A'A(\lambda_1 - \lambda_2) = \lambda_3 - \lambda_4$$

$$\lambda_3 + \lambda_4 = c$$

$$\lambda_1 \geq 0,\; \lambda_2 \geq 0,\; \lambda_3 \geq 0,\; \lambda_4 \geq 0$$

The conditions $\lambda_3 + \lambda_4 = c$, $\lambda_3 \geq 0$ and $\lambda_4 \geq 0$ imply that $(\lambda_3)_i = (\lambda_4)_i = 0$ for $i \in K$ and $(\lambda_3)_i - (\lambda_4)_i \in [-1, 1]$ for $i \notin K$. Also note that λ_1 and λ_2 cannot both be nonzero for the optimal solution, otherwise we could decrease $\lambda_1 + \lambda_2$ and increase objective function while keeping $\lambda_1 - \lambda_2$ the same. Define $\nu = \lambda_1 - \lambda_2$ then we have $|\nu_i| = (\lambda_1)_i + (\lambda_2)_i$. Using these in 3.3.1 , we can write the dual of mod-DS as

$$\underset{\nu}{\text{maximize}}\ y' A\nu - \epsilon\|\nu\|_1 \quad \text{subject to} \tag{Dual mod-DS}$$

$$a_i' A\nu = 0 \quad \text{for} \quad i \in K$$

$$|a_i' A\nu| \leq 1 \quad \text{for} \quad i \notin K$$

We now simplify the complementary slackness condition for mod-DS. The complementary slackness for mod-DS can be written as

$$(\lambda_1)_i(x_i - t_i) = 0,\ (\lambda_2)_i(-x_i - t_i) = 0\ \ \forall i = 1 \dots p \tag{3.3.2}$$

and

$$(\lambda_3)_i[a_i'(y - Ax) - \epsilon] = 0,\ (\lambda_4)_i[-a_i'(y - Ax) - \epsilon] = 0\ \ \forall i = 1 \dots p \tag{3.3.3}$$

First we analyze (3.3.2). For $i \in K$, $(\lambda_3)_i = (\lambda_4)_i = 0$ so complementary slackness is automatically satisfied. Goldman-Tucker theorem implies $x_i - t_i \neq 0$ and $-x_i - t_i \neq 0$, which doesn't mean anything here.

If $i \notin K$, three scenarios arise as in the case of mod-LP, namely, $x_i > 0$, $x_i < 0$ and $x_i = 0$. For $x_i > 0$, we have $-x_i - t_i \neq 0$ and hence one must have $(\lambda_4)_i = 0$ implying $(\lambda_3)_i = 0$. This, in turn implies $a_i' A\nu = 1$. Similarly, for $x_i < 0$, we get $x_i - t_i \neq 0$ implying $(\lambda_3)_i = 0$. So, $(\lambda_4)_i = 1$ which then implies $a_i' A\nu = -1$. And, for $x_i = 0$ equations (3.3.3) is automatically satisfied. Goldman-Tucker theorem here indicates that $(\lambda_3)_i > 0, (\lambda_4)_i > 0$ proving that $|a_i' A\nu| < 1$. Similar analysis in 3.3.3 implies the following.

- If $\nu > 0$ then $(\lambda_1)_i > 0$ and $(\lambda_2)_i = 0$. Hence $a_i'(y - Ax) = \epsilon$.

- If $\nu < 0$ then $(\lambda_1)_i = 0$ and $(\lambda_2)_i > 0$. Hence $a_i'(y - Ax) = -\epsilon$, and

- If $\nu = 0$ then $(\lambda_1)_i = 0$ and $(\lambda_2)_i = 0$. Complementary slackness condition (3.3.3) is automatically satisfied.

Gradient condition are simply $\nabla_x L = A'A(\lambda_2 - \lambda_1) + \lambda_3 - \lambda_4 = 0$ and $\nabla_t L - \lambda_3 - \lambda_4 + c = 0$ which are same as the equality constraints in (3.3.1). Finally, Goldman-Tucker theorem in this case implies $a_i'(y - Ax) < \epsilon$ or $-a_i'(y - Ax) < \epsilon$ according as $(\lambda_1)_i = 0$ or $(\lambda_2)_i = 0$ Thus we have $|a_i'(y - Ax)| < \epsilon$. Preceding analysis in a form of a theorem states that

Theorem 3.3.1. *x^* is an optimal solution of mod-DS if and only if there exists a ν^* such that x^*, $t^* = |x^*|$ and ν^* are feasible for mod-DS(LP) and dual-mod-DS, and*

(a) $a_i'A\nu^* = 0$ *for all* $i \in K$

(b) $a_i'A\nu^* = sign(x_i^*)$ *for all* $i \in K,\ x_i^* \neq 0$

(c) $|a_i'A\nu^*| < 1$ *for all* $i \notin K,\ x_i^* = 0$ *and*

(d) $a_i'(y - Ax^*) = \epsilon\ sign(\nu_i^*)$ *for all* $\nu_i^* \neq 0$

(e) $|a_i'(y - Ax^*)| < \epsilon$ *if* $\nu_i^* = 0$

3.4 QP Form and Dual of mod-BPDN

We work with unconstrained version of mod-BPDN, namely,

$$\min_x \frac{1}{2}\|Ax - y\|_2^2 + \lambda\|x_{K^c}\|_1,\ \lambda > 0 \qquad \text{(unconstrained mod-BPDN)}$$

The general form of quadratic programming (QP) is

$$\min_x \frac{1}{2}x'Qx + b'x$$

with one or more inequality constraints. We will first show that *unconstrained mod-BPDN* is a QP. Let

$$x_i = x_i^+ - x_i^- \quad \text{and } |x_i| = x_i^+ + x_i^-$$

where

$$x_i^+ = \max\{x_i, 0\} \quad \text{and } x_i^- = \max\{-x_i, 0\}$$

for all i. Thus, we have

$$x = x^+ - x^- \quad \text{and } \|x_{K^c}\|_1 = c'x^+ + c'x^-$$

where c is defined as above. Objective function of unconstrained mod-BPDN simplifies to

$$\begin{aligned}
\frac{1}{2}\|Ax - y\|_2^2 + \lambda\|x_{K^c}\|_1 &= \frac{1}{2}x'A'Ax - y'Ax + \lambda\|x_{K^c}\|_1 + \frac{1}{2}y'y \\
&= \frac{1}{2}\left((x^+ - x^-)'A'A(x^+ - x^-)\right) - y'A(x^+ - x^-) + \lambda(c'x^+ + c'x^-) + \frac{1}{2}y'y \\
&= \frac{1}{2}\begin{pmatrix} x^+ \\ x^- \end{pmatrix}'\begin{pmatrix} A'A & -A'A \\ -A'A & A'A \end{pmatrix}\begin{pmatrix} x^+ \\ x^- \end{pmatrix} + \begin{pmatrix} \lambda c - A'y \\ \lambda c + A'y \end{pmatrix}'\begin{pmatrix} x^+ \\ x^- \end{pmatrix} + \frac{1}{2}y'y
\end{aligned}$$

Hence unconstrained mod-BPDN is a QP.

Now we show find the dual of unconstrained mod-BPDN. Note that dual of QP is also a QP. Expanding ℓ_2 norm in objective function gives

$$\frac{1}{2}x'A'Ax - y'Ax + \lambda\|x_{K^c}\|_1 + \frac{1}{2}y'y \tag{3.4.1}$$

and has an infimum if

$$A'(Ax - y) + \lambda u = 0$$

where u is in the subgradient of $\partial\|x\|_1$. In other words,

$$u_i = \begin{cases} 0 & \text{if } i \in K, \\ \text{sign}(x_i) & \text{if } x_i \neq 0, i \notin K \\ \in [-1, 1] & \text{if } x_i = 0, i \notin K. \end{cases}$$

Multiplying both sides by x, we get

$$x'A'Ax - x'A'y + \lambda\|x_{K^c}\|_1 = 0.$$

Substituting $x'A'Ax - x'A'y = \lambda\|x_{K^c}\|_1$ in (3.4.1) gives us the dual program,

$$\max_x -\frac{1}{2}\|Ax\|_2^2 + \frac{1}{2}y'y \qquad \text{subject to} \qquad A'(Ax-y) = \lambda u$$

Ignoring the constant term $\frac{1}{2}y'y$, and noting the definition of u, the dual of unconstrained mod-BPDN can be written as

$$\min_x \|Ax\|_2^2 \qquad \text{subject to} \qquad \|A'_{K^c}(Ax-y)\|_\infty \le \lambda,\;\; A'_K(Ax-y) = 0 \;\; \text{(Dual-mod-BPDN)}$$

Let us denote the objective function of mod-BPDN by $Q_\lambda^K(x)$. We have the following lemma.

Lemma 3.4.1. *For any y, A and $\lambda \ge 0$, the mod-BPDN has the following properties.*

(a) mod-BPDN either has a unique solution or there are infinitely many solutions.

(b) If $u(\lambda)$ and $v(\lambda)$ are two distinct solutions of mod-BPDN for a fixed λ then $Au = Av$.

(c) for a fixed $\lambda > 0$, any two solutions u, v of mod-BPDN satisfy $\|u_{K^c}\|_1 = \|v_{K^c}\|_1$

Proof. (a) The objective function $Q_\lambda^K(x)$ is not strictly convex if $A'A$ is not positive definite. So, there may be infinite number of solutions. If the solution is not unique, let $u(\lambda)$ and $v(\lambda)$ be two different solutions for a fixed value of λ. As the solution set of mod-BP is convex, the convex combination $\alpha u(\lambda) + (1 - \alpha v(\lambda))$ for $0 < \alpha < 1$ are also solutions. Hence there are infinitely many solutions.

(b) Note that for an optimal solution $u(\lambda)$ of mod-BPDN, $y \ne Au(\lambda)$. Let $c^* = Q_\lambda^K(u) = Q_\lambda^K(v)$ be the common optimum value. If $Au \ne Av$, define w as the convex combination of u and v as $w = \alpha u + (1-\alpha)v$ for $0 < \alpha < 1$ then

$$\begin{aligned} Q(w) &= Q(\alpha u + (1-\alpha)v) \\ &= \frac{1}{2}\|y - A(\alpha u + (1-\alpha)v)\|_2^2 + \lambda\|(\alpha u + (1-\alpha)v)_{K^c}\|_1 \\ &\le \frac{1}{2}\|\alpha(y - Au) + (1-\alpha)(y - Av)\|_2^2 + \lambda\alpha\|u_{K^c}\|_1 + \lambda(1-\alpha)\|v_{K^c}\|_1 \\ &< \alpha\left(\frac{1}{2}\|(y - Au)\|_2^2 + \lambda\|u_{K^c}\|_1\right) + (1-\alpha)\left(\frac{1}{2}\|(y - Au)\|_2^2 + \lambda\alpha\|u_{K^c}\|_1\right) \\ &= \alpha c^* + (1-\alpha)c^* \\ &= c^* \end{aligned}$$

Inequality above is due to the convexity of the norm $\|.\|_1$ and the strict inequality is due to the strict convexity of the function $\|.\|_2^2$. Thus we have the contradiction that u and v are the optimum of the mod-BPDN. This proves that $Au = Av$.

(c) Using (b), we have $y - Au = y - Av$. Since both solution give the same value of the objective function, we must have $\|u_{K^c}\|_1 = \|v_{K^c}\|_1$ □

CHAPTER 4. NUMERICAL EXPERIMENTS: mod-BPDN.

In this chapter, we discuss numerical solution of one of the modified ℓ_1 minimization problems, namely, mod-BPDN using an open source solver called ℓ_1-homotopy [2]. Homotopy method starts with an initial solution and constructs a desired solution iteratively by tracing a linear path between two consecutive approximate solutions and gradually adjusting the parameter.

4.1 Unconstrained mod-BPDN and Homotopy

$$\underset{x \in \mathbb{R}^p}{\text{minimize}} \; \frac{1}{2}\|y - Ax\|_2^2 + \lambda \|x_{K^c}\|_1, \quad \lambda > 0 \qquad \text{(mod-BPDN-Unconstrained)}$$

By defining $W = \text{diag}\{w_1, w_2, \ldots, w_p\}$ where $w_i = \begin{cases} 0 & \text{if } i \in K \\ \lambda & \text{if } i \in K^c. \end{cases}$ unconstrained mod-BPDN reduces to the following problem discussed in [3]

$$\underset{x \in \mathbb{R}^p}{\text{minimize}} \; \frac{1}{2}\|y - Ax\|_2^2 + \lambda \|Wx\|_1, \quad \lambda > 0 \qquad \text{(weighted-BPDN)}$$

where W is a diagonal matrix of nonnegative weights. Homotopy algorithm starts with a solution of another closely related problem and finds the solution of mod-BPDN iteratively. For details see [3].

4.2 Experiments

We create a matrix A of size 250×500 with entries drawn from Gaussian distribution and normalize the columns. We also generate a s-sparse vector with support S where s is a positive integer in $[60, 120]$ with its nonzero entries uniformly distributed in the range of $[-20, 20]$. The known part K of the support S will have 50% to 100% of the indices in S. After generating x we generate bounded noise w and add it to the measurements Ax giving us $y = Ax + w$. The parameter λ is chosen to be the maximum of pre-defined tolerance $10^{-4} \cdot \|A'y\|_\infty$ and theoretical optimum $\sigma \cdot \sqrt{\log(p)}$ where σ is the standard deviation of each component of error vector w. Then we apply homotopy algorithm to approximate x from the measurements y using both BPDN and mod-BPDN. We run 100 realization of such experiment and plot the average result. We calculate the ratio of ℓ_2 error between x and approximate solution. Here are the details of the plots.

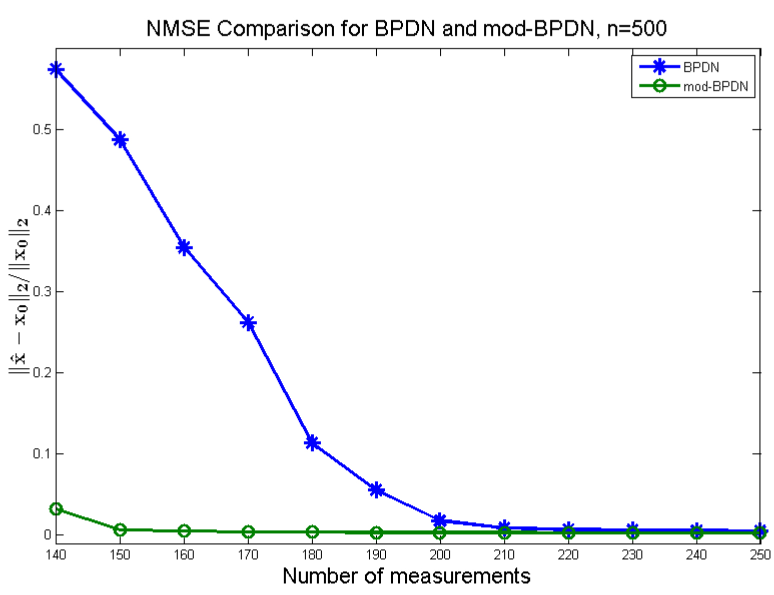


Figure 4.2.1 NMSE comparison, BPDN versus mod-BPDN, $|K|$ is 60% of $|S|$.

As we can see from the figure 4.2.1 mod-BPDN performs significantly better when we know 60% of the indices and it becomes less significant as we increase measurements. The following figure is when $|K| = 80\%$.

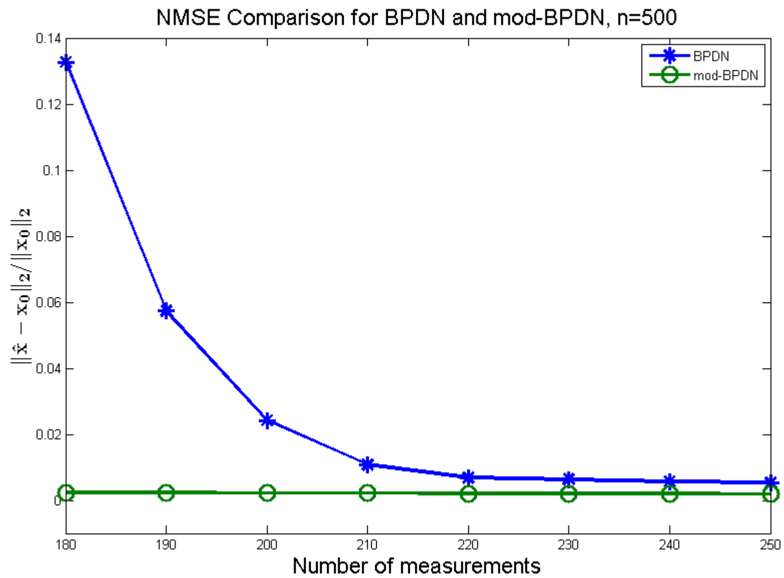


Figure 4.2.2 NMSE comparison, BPDN versus mod-BPDN, $|K|$ is 80% of $|S|$.

In figure 4.2.3, we plot the NMSE of solutions of both BPDN and mod-BPDN versus increasing sparsity. When x is very sparse, difference in performance is minimal but as the sparsity grows the performance of mod-BPDN becomes significantly better than BPDN. In the next figure 4.2.4, we vary $|K|$ from 50% to 100% and show that more knowledge of the support reduces the error significantly.

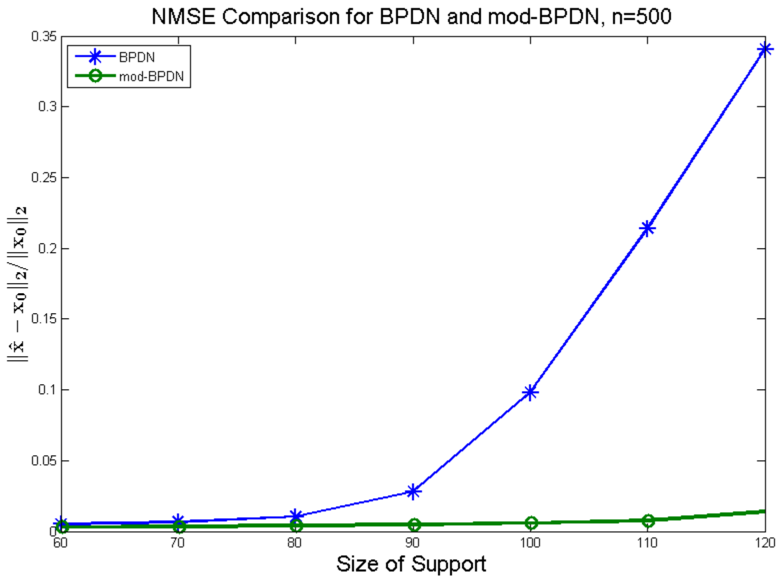


Figure 4.2.3 NMSE comparison, BPDN versus mod-BPDN, $|K|$ is 60% of $|S|$. $|S|$ is increasing by 10 from 60 to 120

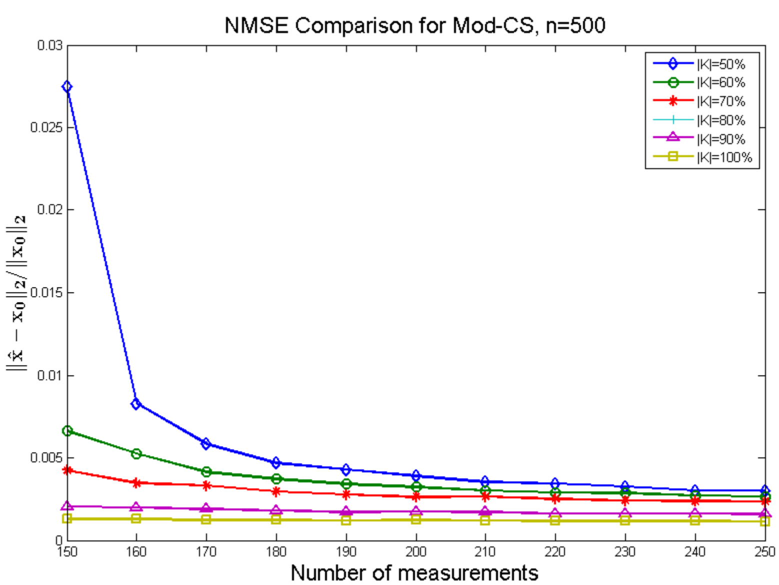


Figure 4.2.4 NMSE comparison, BPDN versus mod-BPDN, $s = 60$ and K is increasing from 50% to 100%

CHAPTER 5. SUMMARY, FUTURE RESEARCH AND CONCLUSION

This chapter consist of summary of results derived during my time at Iowa State University. We will also list some of the interesting problems I will work on as my future research projects.

5.1 Summary of Results

5.1.1 RIC and ROC Based Results

Table 5.1.1 : Notations

Notations				
x	a feasible vector		x_{-u}	$x - x_K - x_U$
S	support of x		$\hat{x}$	$\arg\min_x \{\|x_{K^c}\|_1 : y = Ax\}$
K	known part of S		$\breve{x}$	$\arg\min_x \{\|x_{K^c}\|_1 : \|y - Ax\|_2 \le \epsilon\}$
U	unknown part of S		x^*	$\arg\min_x \{\|x_{K^c}\|_1 : \|A'(y - Ax)\|_\infty \le \epsilon\}$
A, B, C	constants on u, k, δ , θ		s, k, u	$\|S\|$, $\|K\|$, $\|U\|$ resp.

5.1.2 Coherence Based Main Results

We assume x to be feasible and s-sparse here. Other notations are same. C and D are constants depending on μ, u and k.

Table 5.1.2 : Sufficient conditions based on restricted isometry

Results Based on RIC and ROC	
Sufficient Condition	Main Results
$\delta_{k+u+a} + \sqrt{\frac{u}{b}}\,\theta_{k+u+a,b} < 1,$	$\|\hat{x} - x\|_2 \le B\|x_{-u}\|_1$ (mod-BP)
	$\|\breve{x} - x\|_2 \le A\epsilon + B\|x_{-u}\|_1$ (mod-BPDN with bounded noise)
	$\|\breve{x} - x\|_2 \le A\sigma\sqrt{m + 2\sqrt{m\log m}} + B\|x_{-u}\|_1$ w.p. $\ge 1 - \frac{1}{m}$ (mod-BPDN with Gaussian noise)
$0 < a < b \le 4a,\ 0 \le u \le p - k$	$\|x^* - x\|_2 \le C\epsilon + B\|x_{-u}\|_1$ (mod-DS with bounded noise)
	$\|x^* - x\|_2 \le C\sigma\sqrt{2\log p} + B\|x_{-u}\|_1$ w.p. $\ge 1 - \frac{1}{2\sqrt{\pi\log p}}$ (mod-DS with Gaussian noise)

Table 5.1.3 : More sufficient conditions

Other Sufficient Conditions	
For $b = 2u,\ a = u$	1. $\delta_{k+2u} + \theta_{k+2u,u} < 1$
	2. $\delta_{k+2u} < \sqrt{2} - 1 \approx 0.414$
	3. $\delta_{k+2u} + \delta_{k+3u} < 1.$
For $u \le \frac{s}{4}$ or $u \le \frac{k}{3}$	4. $\delta_{2s} < \frac{1}{1+\sqrt{0.3125}} \approx 0.6414$

5.1.3 NSC for mod-BP

Let x be feasible for mod-BP and let x_{K^c} be u-sparse. Assume that $A_{K\cup U}$ has full rank $k + u$, then x^* is the unique solution of $mod - BP$ if and only if there exists a $\nu \in \mathbb{R}^m$ with the following conditions

(a). $a_i'\nu = 0$ for all $i \in K$

(b). $a_i'\nu = \text{sign}(x_i^*)$ for all $i \in U$

(c). $|a_i'\nu| < 1$ for all $i \notin K \cup U$

Table 5.1.4 : Sufficient conditions based on coherence

Results Based on Coherence	
Sufficient Condition	Main Results
$\mu < \frac{1}{k+2u-1}$	$x = \hat{x}$ (mod-BP)
	$\|\breve{x} - x\|_2 \le C(\mu, u, k)\epsilon$ (mod-BPDN)
	$\|x^* - x\|_2 \le D(\mu, u, k)\epsilon$ (mod-DS)

5.2 Future Research

In this subsection, I list some problems that I plan to work on after my Ph.D.

5.2.1 Sharp Bounds for Modified ℓ_1 Minimization Problems

Ever since sufficient conditions based on RIC and ROC were used by Candès and Tao for the recovery of sparse signals, researchers are trying to improve the bounds by trying different approaches. Main problem of my research is also related to one of such paper by Cai et. al [11]. There have been multiple efforts to improve bounds and sufficient conditions [10, 13]. Sharp sufficient conditions for all three ℓ_1-minimization problems which have recently been found [59] is given by

$$\delta_{ts} < \begin{cases} \sqrt{\frac{t-1}{t}} & \text{if } t \geq \frac{4}{3} \\ \frac{4}{4-t} & \text{if } ts \text{ is even and } t < \frac{4}{3} \\ \frac{\sqrt{t^2-\frac{1}{s^2}}}{4-2t+\sqrt{t^2-\frac{1}{s^2}}} & \text{if } ts \text{ is odd and } t < \frac{4}{3} \end{cases}$$

As a first project after my graduation, I will work to establish the corresponding sharp bounds for modified ℓ_1 minimization.

5.2.2 Neighbourliness of a polytope and NSC conditions for mod-BP

The first problem I approached during my research is to find a necessary and sufficient conditions based on convex polytopes and property of such polytopes called neighbourliness. To explain the actual problem, we need some basics of convex geometry.

5.2.2.1 Basics of Convex Polytopes

In this subsection, we will review facts about convex polytopes that are related to this work. Readers are advised to see [41] for the detailed explanation. For a given $a \neq 0 \in \mathbb{R}^p$, $c \in \mathbb{R}$, a set of the form $H = \{x \in \mathbb{R}^p : a'x = c\}$ is called a *hyperplane*. The hyperplane defines two *closed halfspaces* $K_1 = \{x \in \mathbb{R}^p : a'x \leq c\}$ and $K_2 = \{x \in \mathbb{R}^p : a'x \geq c\}$. A set of the form

$$\left\{ \sum_{i=1}^{p} \lambda_i a_i : \lambda_i \text{ are scalars}, a_i \in \mathbb{R}^p \quad \text{for all} \quad i = 1 \ldots p \right\}$$

is called a *linear combination* of $a_1, ..., a_p$. It is called a *convex linear combination* if $0 \leq \lambda_i \leq 1$ for all i, and $\sum_i \lambda_i = 1$. Convex combination becomes *affine combination* if the requirement $0 \leq \lambda_i \leq 1$ is removed. The set of convex linear combinations of $\{a_1, \ldots, a_p\}$ is also called the *convex hull*, written $\text{conv}\{a_1, \ldots, a_p\}$.

A *convex polyhedron* is defined as the intersection of a finite number of closed half spaces. A *polytope* P is a bounded convex polyhedron, or equivalently the convex hull of a finite number of points in $\mathbb{R}^p$.

A hyperplane H is a *supporting hyperplane* for a polytope P if $P \cap H \neq \emptyset$, and P lies completely in one of the closed half spaces defined by H. A set $F \subset P$ is a *face* of P if $F = \emptyset$ or $F = P$ or $F = P \cap H$ for some supporting hyperplane H. A face F is a *proper face* of P if $F \neq \emptyset$, and $F \neq P$. A *vertex* is a face consisting of a single point.

The polytope is the convex hull of its vertices. Two important polytopes are the *hypercube* $\{x : -1 \leq x_i \leq 1 \, \forall i\}$ (the ℓ_∞-ball) and the *crosspolytope* $\{x : \sum_i |x_i| \leq 1\}$ (the ℓ_1-ball). By the *dimension* of polytope P, we mean the dimension of the affine hull of the vertices of P. An r-dimensional polytope is simply called a *r-polytope.* A *r-simplex* in $\mathbb{R}^m$ is defined to be the convex hull of $r + 1$ affinely independent points in $\mathbb{R}^m$. A polytope is said to be *k-neighbourly* if every set of $k+1$ vertices of P forms a proper face of P. Every r-simplex is k-neighbourly for $1 \leq k \leq r$.

Polarity plays a very important role in proving some of the theorems in this paper. For a set $A \subset \mathbb{R}^m$, the *polar set* is defined as

$$A^* = \{\nu \in \mathbb{R}^m : x'\nu \leq 1 \ \forall x \in A\}$$

For example, the hypercube and the crosspolytope are polar to each other. The polar of a polytope P containing 0 in its interior is also a polytope. We use polarity as also a duality. Two polytopes P, P^* are said to be *dual* to each other if there exists a one-to-one map ψ between the faces of P and the faces of P^*, which is inclusion reversing, that is, if F_1 and F_2 are faces of P with $F_1 \subset F_2$, $\psi(F_1)$, and $\psi(F_2)$ are faces of P^* with $\psi(F_1) \supset \psi(F_2)$.

Let P be a polytope and P^* be its polar. For a face F of P define $F^* = \{\nu \in P^* : y'\nu = 1 \quad \forall y \in F\}$. Then F^* is a face of P^*, and the map ψ defined by $\psi(F) = F^*$ is one-one and inclusion reversing, establishing the duality of P and P^*. If F is a face of an m-polytope and F^* is the corresponding face of its polar P^*, then we have $\dim(F) + \dim(F^*) = m - 1$. A polytope is said to be *centrally symmetric* if $-P = P$, equivalently, if it can be written as $P = \text{conv}\{\pm a_1, \pm a_2, ..., \pm a_n\}$ for some a_i in $\mathbb{R}^m$. In the case of a centrally symmetric polytope, P is called k-neighbourly if every set of $k+1$ vertices not including any antipodal pair forms a face of P. The polar of P can be written as $P^* = \{\nu \in \mathbb{R}^m : |A'\nu| \le 1\}$.

5.2.2.2 Previous NSC for mod-BP based on Polytopes

Conditions based on RIPs and coherence give only sufficient conditions. Below we will make a brief review of some necessary and sufficient conditions for the existence of a unique solution of the basis pursuit problem.

Theorem 5.2.1. *[37] Let an s-sparse x^* with support S satisfy $y = Ax^*$. Then x^* is the unique solution of BP if and only if*

(a). rank$(A_S) = s$,

and there exists $\nu \in \mathbb{R}^m$ such that

(b). $a_i'\nu = sign(x_i^*) \ \forall \ i \in S$, and

(c). $|a_i'\nu| < 1 \ \forall \ i \in S^c$.

The sufficiency of Fuchs' condition was proved in his paper [37]. The necessity was recently proved by Dossal [32]. Theorem 3.2.3 gives an alternative, simpler proof.

Theorem 5.2.2. *[26] Let A be a $m \times p$ matrix with $m \ll p$. Whenever $y = Ax^*$ has a solution x^* having at most s nonzero components, x^* is the unique optimal solution of BP if and only if the centrally symmetric polytope $P = conv\{\pm a_1, \pm a_2, .., \pm a_p\}$ is s-neighbourly.*

5.2.2.3 The Problem

Donoho's theorem 5.2.2 says that s-neighbourliness of $P = \text{conv}\{\pm a_1, \pm a_2, .., \pm a_p\}$ is necessary and sufficient for the uniqueness of solution of BP. Because we have an extra knowledge of the support in mod-BP problem, we may not require s-neighbourliness of P. Thus the question arising here is, how neighbourly does P have to be for the uniqueness of solution of mod-BP? As in the spirit of restricted isometry and coherence based analysis, we expect P to have a weaker form of neighbourliness. Before we explain the steps I have taken while trying to answer this question, we note here mod-BP and its dual.

$$\text{minimize } \|x_{K^c}\|_1 \quad \text{subject to } y = Ax. \tag{mod-BP}$$

$$\begin{aligned} \underset{\nu}{\text{maximize}} \quad \nu' y \quad \text{subject to} \quad & |a_i' \nu| = 0 \ \text{ for } \ i \in K \\ & |a_i' \nu| \le 1 \ \text{ for } \ i \notin K \end{aligned} \tag{dual mod-LP}$$

What follows is the details of some of the theorems and lemmas I have proved and proposed to solve the problem at hand. I plan to finish working in this project after my graduation.

It is important to note here that Fuchs' condition is for one particular sparse vector x^*, but in Donoho's condition the emphasis was on proving uniqueness for all s-sparse solutions, not just one specific x^*. We have modified Donoho's condition for a particular x^*, which is given below.

5.2.2.4 Connection Between Fuchs' Condition and Donoho's Condition

We state and give here a direct proof of equivalence of the two.

Proposition 5.2.3. *Let an s-sparse x^* with support S satisfy $y = Ax$. Then the following are equivalent.*

(a) Fuchs' condition

(i) Rank$(A_S) = s$. (ii) There exists $\nu \in \mathbb{R}^m$ such that $A_S'\nu = sign(x_S^*)$ and $|a_i'\nu| < 1 \ \forall\, i \in S^c$

(b) Modified Donoho's condition

$F = conv\{sign(x_i^*)a_i : i \in S\}$ is a $s-1$ simplex, and a face of the centrally symmetric polytope $P = conv\{\pm a_1, \pm a_2, .., \pm a_p\}$.

Proof. Let us first assume that Fuchs' condition holds. We will show that an $m-s$ dimensional face F^* of P^* exists and use duality to prove the existence of F. Consider $F^* = \{\nu \in P^* : A_S'\nu = \text{sign}(x_S^*)\}$. Note that for each $i \in S$, a set of the form $H_i = \{\nu : a_i'\nu = \text{sign}(x_i^*)\}$ is a hyperplane and is nonempty as given by Fuchs' condition. This proves that F^* is also nonempty. Define a vector $d' = 1'A_S'$ and a scalar $\alpha = 1'\text{sign}(x_S^*)$, where 1 is a vector of ones. Now for any $\nu \in F^*$, we have $d'\nu = 1'A_S'\nu = 1'\text{sign}(x_S^*) = \alpha$ and for a $\nu \in P^*$, we have $d'\nu \le \alpha$. This implies that $d'\nu \le \alpha$ is a valid inequality for P^* and the hyperplane $H = \{\nu : d'\nu = \alpha\}$ is a supporting hyperplane for P^*. Note that $d'\nu = \alpha$ if and only if $A_S'\nu = \text{sign}(x_S^*)$. Hence $F^* = P^* \cap H = \{\nu \in P^* : A_S'\nu = \text{sign}(x_S^*)\}$ is a face of P^*.

Since $\text{rank}(A_S) = s$, it holds that $\dim(F^*) + \text{rank}(A_S) = m$ [53]. Hence $\dim(F^*) = m-s$. Now using polar duality there is a face F of P such that $\dim(F) + \dim(F^*) = m-1$. Thus $\dim(F) = s-1$.

Now, we show that F is a simplex. Let $S = \{i_1, i_2, ..., i_s\}$ and define $T_j = S - \{i_1, i_2, ..., i_j\}$ where $j \in \{1, 2, ..., s-1\}$. Define sets of the form $F_j^* = \{\nu \in P^* : A_{T_j}'\nu = \text{sign}(x_{T_j}^*)\}$. Then by the same argument as above, it can be shown that F_j^* is also a face of P^* of dimension $m-s+j$ for $j \in \{1,2,...,s-1\}$ and the relation $F^* \subset F_1^* \subset F_2^* \subset \cdots \subset F_{s-1}^*$ holds. By taking polar in this relation we get faces $F_{s-1} \subset F_{s-2} \subset \cdots \subset F_1 \subset F$ of dimensions $0,1,2,...,s-2$. This proves that F is a simplex. Further more, we can write F^* as $F^* = \{\nu : |A'_S\nu| = 1\} \cap P^*$ where $|.|$ is componentwise and 1 is a vector of ones. Thus by definition of polar, F can be written as $F = \text{conv}\{\sigma_i a_i, \sigma_i \in \{1,-1\}, i \in S\}$.

Conversely, let us assume that Donoho's modified condition holds. Since $F = \text{conv}\{\pm a_i, i \in S\}$ is an $s-1$ dimensional simplex and is a face of P, its polar F^* can be written as $F^* = \{\nu : |A'_S\nu| \leq 1\} \cap P^*$. The fact that G is an $s-1$ dimensional face implies that there exists a ν such that $|a'_i\nu| = 1$ for all $i \in S$ and $|a'_i\nu| < 1$ for all $i \notin S$. This proves (ii) and (iii). For the rank, we note that F^* is $m-s$ dimensional which means that there are exactly s different supporting hyperplanes $\{\nu : |a'_i\nu| = 1, \forall i\}$ having nonempty intersection with P^* and hence $\text{rank}(A_S) = s$. This proves the theorem. □

Lemma 5.2.4. *The feasible set $P_K^* = \{\nu \in \mathbb{R}^m : |A'\nu| \preccurlyeq c\}$ of the dual of mod-BP is a polytope.*

Proof. By definition P_K^* is a polyhedron. Since $P_K^* \subset P^*$, P_K^* is also bounded. □

The proof of the following theorems need verification and checking.

We connect Fuchs' algebraic conditions for mod-BP to a vertex and a face of a polar polytope in the following theorem.

Theorem 5.2.5. *Let x be feasible for mod-BP and let x_{K^c} be u-sparse. Also assume that $A_{K\cup U}$ has full rank $k+u$. Then x^* is the unique optimum of mod-BP if and only if*

(a) A set $F_K^* = \{\nu \in P_K^* : A'_U = sign(x_U^*)\}$ is an $m-u$ dimensional face of P_K^*.

(b) There exists a vector $\nu \in \mathbb{R}^m$ of the form $\nu = MA_U(A'_UMA_U)^{-1}sign(x_U^*)$ where $M = I - A_K(A'_KA_K)^{-1}A'_K$, which lies on some face of F_K^*

Proof. We will show that the conditions (a) and (b) in theorem 3.2.3 are equivalent to the conditions (a) and (b) in theorem 5.2.5.

We prove (a) first. In the proof of proposition 1 we noted that any set of the form $\nu \in P^* : A'_T\nu = \text{sign}(x^*_T)$ where $T \subseteq S$ is an $m - |T|$ dimensional face of the polytope P^* if Fuchs' condition holds. Since $U \subseteq S$ and $A'_U\nu = \text{sign}(x^*_U)$ by (a) of theorem 3, we conclude that $\nu \in P^* : A'_U\nu = \text{sign}(x^*_U)$ is an $m - u$ dimensional face of P^*. Since P^*_K is a polytope and $P^*_K \subset P^*$, the set $P^*_K \cap \{\nu \in P^* : A'_T\nu = \text{sign}(x^*_T)\}$ will be a face of P^*_K only if $T \subseteq U$. Thus for $T = U$ we have $P^*_K \cap \{\nu \in P^* : A'_U\nu = \text{sign}(x^*_U)\} = \{\nu \in P^*_K : A'_U\nu = \text{sign}(x^*_U)\} = F^*_K$ and hence F^*_K is an $m - u$ dimensional face of P^*_K.

Now we prove (b). The condition $A'_K\nu = 0$ in (a) in theorem 3 implies that ν is of the form

$$\nu = \left(I - A_K(A'_KA_K)^{-1}A'_K\right)\beta$$

for some $\beta \in \mathbb{R}^m$. Denoting $I - A_K(A'_KA_K)^{-1}A'_K$ by M and substituting this ν in $A'_U\nu = \text{sign}(x^*_U)$ we have $A'_UM\alpha = \text{sign}(x^*_U)$ and hence $\alpha = (A'_UM)^\dagger\text{sign}(x^*_U)$ where $\dagger$ is the pseudo inverse. Using the formula for the pseudo inverse and noting that facts that $M = M^2 = MM'$, we have

$$\alpha = M'A_U(A'_UMM'A_U)^{-1}\text{sign}(x^*_U) = MA_U(A'_UMA_U)^{-1}\text{sign}(x^*_U).$$

Hence

$$\nu = MA_U(A'_UMA_U)^{-1}\text{sign}(x^*_U)$$

It remains to show that ν lies on some face of F^*_K. It is clear that ν is in F^*_K. If it lies in relative interior we get a contradiction by applying A'_U or A'_K.

For the converse, the existence of $m-u$ dimensional face of F^*_K implies $a'_j\nu = \text{sign}(x^*_j)$ for all $j \in U$ and $|a'_j\nu| < 1$ for all $j \notin U$). By applying A'_U and A'_K gives us the required conditions in (a) in theorem 3.2.3. □

Finally we transfer these conditions to primal polytope. We have used the polar of P^*_K by P_K.

Theorem 5.2.6. *Let x be feasible for mod-BP and let x_{K^c} be u-sparse. Assume that $A_{K\cup U}$ has full rank $k + u$. Then x^* is the unique optimum of mod-BP if and only if*

(a) *The polytope* $P \cap P_K$ *has a* $u-1$ *dimensional face* $F_K = conv\{sign(x_i^* a_i, i \in U$ *which is a simplex*

(b) *The hyperplane* $x'\nu = 1$ *for* $\nu = MA_U(A'_U MA_U)^{-1} sign(x_U^*)$ *is a supporting hyperplane for the face in* (a).

Proof. Again, it suffices to prove that (a) and (b) in theorem 4 are equivalent to (a) and (b) in theorem 4. we proved in theorem 3 that F_K^* is a an $m-u$ dimensional face of both P_K^* and P^*. Hence by duality there exists a face F_K in P and in P_K that has dimension $u-1$. The fact that F_K is a simplex follows from the same argument we used in proposition. This proves (a). Note here that $P_K^* \subset P_K^*$ implies that face of For (a), we note that a supporting hyperplane for P_K^* is also a supporting hyperplane for P^*. So, $P^* \cap F_K^*$ is an $m-u$ dimensional face of P^*. Once again by duality , there is an $u-1$ dimensional face in P which is a simplex. (b) also follows by polar duality. The fact that ν lies on a face F^* implies that F has to lie in such an hyperplane. □

5.3 Conclusion

In this work, we have studied compressed sensing problems known as mod-BP, mod-BPDN and mod-DS by making an extra assumption that we know the part of the support in regular compressed sensing problems. We saw that this generalization not only performs better both in theory and practice, it can also be applied to many applications as discussed in chapter one.

After discussing fundamentals of sparse recovery and compressed sensing, we introduced our research problems in chapter 1. In chapter 2, we proved that modified ℓ_1-minimization programs, namely mod-BP, mod-BPDN and mod-DS, need weaker sufficient conditions (based on RIC and ROC) than regular compressed sensing problems. We also showed that our sufficient conditions are weaker than previously used sufficient conditions for modified compressed sensing problems. At the end of chapter 2, we showed that some knowledge of the support of the signal helps to break the coherence barrier in a deterministic compressed sensing problem.

In chapter 3, we turned to convex optimization theory to analyze modified ℓ_1-minimization problems. We discussed various known necessary and sufficient conditions for mod-BP and proved that previously known sufficient condition known as Fuchs' condition can be generalized to be a necessary and sufficient condition.

In chapter 4, we used an open source algorithm called homotopy algorithm and compare the performance of mod-BPDN and BPDN. We demonstrated by several experiments that numerical results corroborate our findings in chapter 2. Summary of our research project results are presented in chapter 5. We also discussed some of our future projects in this chapter.

APPENDIX A. MATLAB CODE USED FOR THE EXPERIMENTS

These MATLAB code use an open source algorithm [2] called ℓ_1 homotopy package and solve the mod-BPDN problem

$$\underset{x\in\mathbb{R}^p}{\text{minimize}}\ \frac{1}{2}\|y-Ax\|_2^2+\lambda\|x_{K^c}\|_1,\quad \lambda>0 \qquad \text{(mod-BPDN-Unconstrained)}$$

Matlab Code

```
clear;
randn('state',0)
rand('state',0)
addpath utils/
n=500;
s=60;
sigma=0.01;

for k=1:1:100
    x_0=2*(rand(n,1)-0.5)*20;
    idx=randperm(n);
    sup=idx(1:s);
```

```
    sup_c=setdiff((1:n),sup);
    x_0(sup_c)=0;


    % Generating a  n \times m sampling matrix, 250 columns
        will be used for sampling
    A_0 =randn(500,500);

    % Sampling matrix A, weight W and the model with error e
    for i=18:1:25 % we will take 180 samples to 250 samples
       when |K| is 80%
        % and 130 samples to 250 samples when |K| is 60%
        sam_per=0.02*i;
        idx=randperm(n);
        sam_idx=idx(1:int64(sam_per*n));
        A=A_0(sam_idx,:);
        for j=1:1:n
            A(:,j)=A(:,j)/norm(A(:,j));
        end
        e=sigma*(rand(int64(sam_per*n),1)-1)*2;
        y=A*x_0+e;
        err_fun = @(z) (norm(x_0-z)/norm(x_0))^2;


        %BPDN_homotopy, note that lambda(here tau) is chosen to
            be max of
        %predefined 1e-4*max(abs(A'*y) and the theoritical max
```

```
        opts.tau= max(1e-4*max(abs(A'*y)),sigma*sqrt(log(n)));
        opts.debias = 0;
        opts.verbose = 0;
        opts.plots = 0;
        opts.record = 1;
        opts.err_fun = err_fun;
        opts.delx_mode='qr';
        out = l1homotopy(A, y, opts);
        err_cs(k,i)=norm(out.x_out-x_0)/norm(x_0);


        %mod-BPDN_homotopy
        %|K|=50% to 100%
        for j=5:1:10
            k_per=j/10.0;
            idx=randperm(length(sup));
            k_idx=sup(idx(1:length(sup)*k_per));
            weight=ones(n,1);
            weight(k_idx)=1e5;
            weight=diag(weight);
            Aw=A*weight;
            opts.tau= max(1e-4*max(abs(A'*y)),sigma*sqrt(log(n)
                ));
            opts.debias = 0;
            opts.verbose = 0;
            opts.plots = 0;
            opts.record = 1;
```

```
                opts.err_fun = err_fun;
                opts.delx_mode='mil';
                opts.x_orig=zeros(n,1);
                out =l1homotopy(Aw, y, opts);
                x_out=weight*out.x_out;
                err_modcs(k,i,j)=norm(x_out-x_0)/norm(x_0);

            end

        end
end
%Compare BPDN and mod-BPDN with |K|=60%,80%
tx=18:1:25;%tx=13:1:25;
figure
plot((0.02*tx)*500,mean(err_BPDN(:,tx)),'*-',(0.02*tx)*500,mean
    (err_modBPDN(:,tx,8)),'o-');
legend('BPDN','mod-BPDN')
title(['NMSE Comparison for BPDN and mod-BPDN, ', 'n=',num2str
    (n)],'FontSize',16);
xlim([180 250]);
ylim([-0.01 0.14]);
xlabel('Number of measurements','FontSize',16);
ylabel('$\mathbf{\|\hat{x}-x_0\|_2/\|x_0\|_2}$','interpreter','
    latex','FontSize',16);
%performance of mod-BPDN with K=50%-100%
tx=15:25;
figure
```

```
plot((0.02*tx)*500,mean(err_modBPDN(:,tx,5)),'-',(0.02*tx)*500,mean(err_modBPDN(:,tx,6)),'o-', (0.02*tx)*500,mean(err_modBPDN(:,tx,7)),'*-', (0.05*tx)*500,mean(err_modBPDN(:,tx,8)),'+-',(0.02*tx)*500,mean(err_modBPDN(:,tx,9)),'^-',(0.02*tx)*500,mean(err_modBPDN(:,tx,10)));
legend('|K|=50%','|K|=60%','|K|=70%','|K|=80%','|K|=90%','|K|=100%')
title(['NMSE Comparison for Mod-CS, ', 'n=',num2str(n)],'FontSize',16);
xlim([150 250]);
xlabel('Number of measurements','FontSize',16);
ylabel('$\mathbf{\|\hat{x}-x_0\|_2/\|x_0\|_2}$','interpreter','latex','FontSize',16);
%Include the following to vary support
for i=6:1:12
            s=i*10;
             idx=randperm(n);
        sup=idx(1:s);
        sup_c=setdiff((1:n),sup);
        x_0=2*(rand(n,1)-0.5)*20;
        x_0(sup_c)=0;


        e=sigma*(rand(int64(sam_per*n),1)-1)*2;
        y=A*x_0+e;
        err_fun = @(z) (norm(x_0-z)/norm(x_0))^2;
end
```

```
%And the plot is (sparsity varies from 60 to 120
tx=6:12;
figure
plot((10*tx),mean(err_BPDN(:,tx)),'*-',(10*tx+0.1),mean(
    err_modBPDN(:,tx,5)),'o-');
legend('BPDN','mod-BPDN')
title(['NMSE Comparison for BPDN and mod-BPDN, ', 'n=',num2str
    (n)],'FontSize',16);
xlim([60 120]);
xlabel('Size of Support','FontSize',16);
ylabel('$\mathbf{\|\hat{x}-x_0\|_2/\|x_0\|_2}$','interpreter','
    latex','FontSize',16);
```

BIBLIOGRAPHY

[1] D. Angelosante and G.B. Giannakis. RLS-weighted lasso for adaptive estimation of sparse signals. *IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009.*, pages 3245–3248, 2009.

[2] Muhammad Salman Asif. ℓ_1 homotopy package @ONLINE. http://users.ece.gatech.edu/~sasif/homotopy/, June 2013.

[3] Muhammad Salman Asif and Justin K. Romberg. Sparse recovery of streaming signals using ℓ_1–homotopy. *CoRR*, abs/1306.3331, 2013.

[4] Afonso S. Bandeira, Katya Scheinberg, and Luís N. Vicente. On partial sparse recovery. *CoRR*, abs/1304.2809, 2013.

[5] A.S. Bandeira, E. Dobriban, D.G. Mixon, and W.F. Sawin. Certifying the restricted isometry property is hard. *Information Theory, IEEE Transactions on*, 59(6):3448–3450, 2013.

[6] Richard G. Baraniuk, Volkan Cevher, Marco F. Duarte, and Chinmay Hegde. Model-based compressive sensing. *IEEE Trans. Inf. Theor.*, 56(4):1982–2001, April 2010.

[7] Thomas Blumensath and Mike E. Davies. Iterative hard thresholding for compressed sensing. *Appl. Comp. Harm. Anal.*

[8] Stephen Boyd and Lieven Vandenberghe. *Convex optimization.* Cambridge University Press, Cambridge, 2004.

[9] T. Tony Cai, Guangwu Xu, and Jun Zhang. On recovery of sparse signals via ℓ_1 minimization. *IEEE Trans. Inform. Theory*, 55(7):3388–3397, 2009.

[10] T. Tony Cai and Anru Zhang. Compressed sensing and affine rank minimization under restricted isometry. *CoRR*, abs/1304.3531, 2013.

[11] Tony Cai, Lie Wang, and Guangwu Xu. Shifting inequality and recovery of sparse signals. *IEEE Trans. Signal Process.*, 58(3, part 1):1300–1308, 2010.

[12] Tony Cai, Lie Wang, and Guangwu Xu. Stable recovery of sparse signals and an oracle inequality. *IEEE Trans. Inf. Theor.*, 56(7):3516–3522, July 2010.

[13] Tony Cai and Anru Zhang. Sharp RIP bound for sparse signal and low-rank matrix recovery. *CoRR*, abs/1302.1236, 2013.

[14] E.J. Candès, M.B Wakin, and S.P. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier Analysis and Applications*, 14 (5-6):877–905, 2008.

[15] Emmanuel Candès and Terence Tao. The Dantzig selector: statistical estimation when p is much larger than n. *Ann. Statist.*, 35(6):2313–2351, 2007.

[16] Emmanuel J. Candès. The restricted isometry property and its implications for compressed sensing. *C. R. Math. Acad. Sci. Paris*, 346(9-10):589–592, 2008.

[17] Emmanuel J. Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[18] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.

[19] Emmanuel J. Candès and Terence Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.

[20] A. Carmi, P. Gurfil, and D. Kanevsky. Methods for sparse signal recovery using Kalman filtering with embedded pseudo-measurement norms and quasi-norms. *IEEE Trans. Sig. Proc.*, pages 2405–2409, April 2010.

[21] R. Carrillo, L.F. Polania, and K. Barner. Iterative algorithms for compressed sensing with patially known support. In *ICASSP*, 2010.

[22] R. Chartrand and W. Yin. Iteratively reweighted algorithms for compressive sensing. *ICASSP*, 2008.

[23] S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Camp.*, 20(1):33–61, 1998.

[24] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *J. Amer. Math. Soc.*, 22(1):211–231, 2009.

[25] Ronald A. DeVore. Deterministic constructions of compressed sensing matrices. *J. Complex.*, 23(4-6):918–925, August 2007.

[26] D. Donoho. Neighborly polytopes and sparse solutions of underdetermined linear equations. Technical report 2005-04, Stanford Univ., Stat. Dept., January 2005.

[27] D. Donoho. For most large underdetermined systems of linear equations, the minimal ℓ_1 norm solution is also the sparsest solution. *Comm. Pure and App. Math.*, 59(6):797–829, June 2006.

[28] David L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[29] David L. Donoho and Michael Elad. Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ_1 minimization. In *Proc. Natl Acad. Sci. USA 100 2197–2202*.

[30] David L. Donoho, Michael Elad, and Vladimir N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Iform. Theory*, 52(1):6–18, 2006.

[31] David L. Donoho and Xiaoming Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 2001.

[32] Charles Dossal. A necessary and sufficient condition for exact sparse recovery by ℓ_1 minimization. *C. R. Math. Acad. Sci. Paris*, 350(1-2):117–120, 2012.

[33] Marco F. Duarte, Shriram Sarvotham, Dror Baron, Michael B. Wakin, and Richard G. Baraniuk. Distributed compressed sensing of jointly sparse signals. In *In Asilomar Conf. Signals, Sys., Comput*, pages 1537–1541, 2005.

[34] M. Elad. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, New York, NY, 2010.

[35] M. Fornasier. *Theoretical Foundations and Numerical Methods for Sparse Recovery*. Radon Series on Computational and Applied Mathematics. De Gruyter, 2010.

[36] Michael P. Friedlander, Hassan Mansour, Rayan Saab, and Özgür Yilmaz. Recovering compressively sampled signals using partial support information. *IEEE Trans. Inform. Theory*, 58(2):1122–1134, 2012.

[37] Jean-Jacques Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Trans. Inform. Theory*, 50(6):1341–1344, 2004.

[38] J. Garcia-Frias and I. Esnaola. Exploiting prior knowledge in the recovery of signals from noisy random projections. In *Data Compression Conference*, 2007.

[39] Pierre J. Garrigues and Laurent El Ghaoui. A homotopy algorithm for the lasso with online observations. *Adv. Neural Info. Proc. Sys. (NIPS)*, 2008.

[40] A. J. Goldman and A. W. Tucker. Theory of linear programming. In *Linear inequalities and related systems*, Annals of Mathematics Studies, no. 38, pages 53–97. Princeton University Press, Princeton, N.J., 1956.

[41] Branko Grünbaum. *Convex polytopes*, volume 221 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 2003. Prepared and with a preface by Volker Kaibel, Victor Klee and Günter M. Ziegler.

[42] Taner Ince, Arif Nacaroglu, and Nurdal Watsuji. Nonconvex compressed sensing with partially known signal support. *Signal Processing*, 93(1):338–344, 2013.

[43] Laurent Jacques. A short note on compressed sensing with partially known signal support. *Signal Processing*, 90(12):3308–3312, 2010.

[44] Shihao Ji, Ya Xue, and L. Carin. Bayesian compressive sensing. *Signal Processing, IEEE Transactions on*, 56(6):2346–2356, 2008.

[45] M. Amin Khajehnejad, Weiyu Xu, A. Salman Avestimehr, and Babak Hassibi. 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, 2009.

[46] Jongmin Kim, Ok Kyun Lee, and Jong Chul Ye. Dynamic sparse support tracking with multiple measurement vectors using compressive music. In *ICASSP*, pages 2717–2720, 2012.

[47] W. Lu and N. Vaswani. Modified BPDN for noisy compressive sensing with partially known support. *ICASSP*, 2010.

[48] W. Lu and N. Vaswani. Regularized modified BPDN for noisy sparse reconstruction with partial erroneous support and signal value knowledge. *IEEE Trans. Sig. Proc.*, January 2012.

[49] D. M. Malioutov, S. Sanghavi, and A. S. Willsky. Compressed sensing with sequential observations. In *Proc. of ICASSP, Las Vegas, NV*, 2008.

[50] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Processing*, 41(12):3397–3415, 1993.

[51] C. J. Miosso, R. von Borries, M. Argez, L. Valazquez, C. Quintero, and C. Potes. Compressive sensing reconstruction with prior information by iteratively reweighted least-squares. *IEEE Trans. Sig. Proc.*, 57 (6):2424–2431, June 2009.

[52] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Comput.*, 24(2):227–234, 1995.

[53] George L. Nemhauser and Laurence A. Wolsey. *Integer and combinatorial optimization*. Wiley-Interscience, New York, NY, USA, 1988.

[54] M. Salman Asif and J. Romberg. Dynamic updating for sparse time varying signals. In *Information Sciences and Systems, 2009. CISS 2009. 43rd Annual Conference on*, pages 3–8, 2009.

[55] D. Sejdinovic, C. Andrieu, and R. Piechocki. Bayesian sequential compressed sensing in sparse dynamical systems. In *Allerton Conf. Communication, Control, and Computing*, 2010.

[56] Claude E. Shannon. Communication in the presence of noise. *Proc. of the Inst. of Radio Engg.*, 37(13):10–21, 1949.

[57] V. Stankovic, L. Stankovic, and S. Cheng. Compressive image sampling with side information. *ICIP*, 2009.

[58] M. Stojnic. Block-length dependent thresholds for ℓ_2/ℓ_1-optimization in block-sparse compressed sensing. In *ICASSP*, 2010.

[59] Cai Tony and Zhang Anru. Sparse representation of a polytope and recovery of sparse signals and low-rank matrices. Technical report, University of Pennsylvania, Department of Statistics, 06 2013.

[60] J. Tropp and A. Gilbert. Signal recovery from partial information via orthogonal matchng pursuit. *IEEE Trans. Inform. Theory*, 53(12):4655–4666, 2007.

[61] M. Unser. Sampling-50 years after Shannon. *Proceedings of the IEEE*, 88(4):569–587, 2000.

[62] N. Vaswani. Kalman filtered compressed sensing. *IEEE Intl. Conf. Image Proc. (ICIP)*, 2008.

[63] N. Vaswani. Stability (over time) of Modified-CS for Recursive Causal Sparse Reconstruction. In *Allerton Conf. Communication, Control, and Computing*, 2010.

[64] N. Vaswani and W. Lu. Modified-CS: Modifying compressive sensing for problems with partially known support. *ISIT*, 2009.

[65] Namrata Vaswani and Wei Lu. Modified-CS: Modifying compressive sensing for problems with partially known support. *IEEE Trans. Signal Process.*, 58(9):4595–4607, 2010. Supplementary material available online.

[66] R. Von Borries, C.J. Miosso, and C. Potes. Compressed sensing using prior information. *Computational Advances in Multi-Sensor Adaptive Processing, 2007. CAMPSAP 2007. 2nd IEEE International Workshop on*, pages 121–124, 2007.

[67] Yilun Wang and Wotao Yin. Sparse signal reconstruction via iterative support detection. *SIAM J. Imaging Sci.*, 3(3):462–491, 2010.

[68] L. Welch. Lower bounds on the maximum cross correlation of signals. *IEEE Trans. Inform. Theory*, 20(3):397–399, 1974.

[69] Stephen J. Wright. *Primal-dual interior-point methods*. SIAM, 1997.

[70] Z. Zhang and B. D. Rao. Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning. *IEEE J. Sel. Topics Sig. Proc., Special Issue on Adaptive Sparse Representation of Data and Applications in Signal and Image Processing*, 5(5):912–926, Sept 2011.

[71] J. Ziniel, L. C. Potter, and P. Schniter. Tracking and smoothing of time-varying sparse signals via approximate belief propagation. In *Asilomar Conf. on Sig. Sys. Comp.*, 2010.