Sparse Signal Recovery: An Introduction to Compressed Sensing

A Project Report

submitted by

ABHISHEK SAINI

in partial fulfilment of the requirements
for the award of the degree of

MASTER OF TECHNOLOGY

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS.

June 2016

THESIS CERTIFICATE

This is to certify that the thesis titled **Sparse Signal Recovery: An Introduction to Compressed Sensing**, submitted by **Abhishek Saini**, to the Indian Institute of Technology, Madras, for the award of the degree of **Master of Technology**, is a bona fide record of the research work done by him under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Bharath Bhikkaji
Project Guide
Assistant Professor
Dept. of Electrical Engineering
IIT-Madras, 600 036

Place: Chennai

Date: 9 June 2016

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to my thesis guide Dr. Bharath Bhikkaji for his continuous support throughout my thesis. I offer him my earnest appreciation for his guidance, knowledge, patience and motivation. His unique and flexible way to approach any problem made working under him fun and exciting. I could always approach him whenever I ran into trouble. I could not have imagined having a friendlier advisor.

I am honored to have been taught by some of the best teachers during my stay here. They will continue to inspire me in my journeys to come as they have done during my stay here.

The 5 years spent here at IITM would not have been as entertaining if not for my friends. I would like to thank my wing mates, hostel mates and batch mates for making it an enjoyable experience.

Last but not the least, I would like to express my heartfelt gratitude and love to my parents and my sister for their unfailing support and constant encouragement.

ABSTRACT

KEYWORDS: Compressed Sensing; Restricted Isometry Property; Sparse approximation; Sparse Signal Recovery

Consider a linear system of equations $x = D\alpha$ where D is an undetermined $m \times p$ matrix $(m \ll p)$ and $x \in \mathbb{R}^m, \alpha \in \mathbb{R}^p$ where D is called the dictionary or the design matrix. The problem is to estimate the vector α, subject to the constraint that it is sparse. The underlying motivation behind finding such a sparse approximation is that although we observe the signal in a high dimensional space($\mathbb{R}^m$), we assume that the actual signal can be described in a lower dimensional space ($\mathbb{R}^k, k \ll m$).

Since α is sparse, only few of its components are non zero. x can be decomposed as a linear combination of a small number of $m \times 1$ vectors. The subspace spanned by these vectors contains x. Thus they can be seen as a basis of that subspace. These basis vectors are not required to be orthogonal, unlike that in other dimensionality reduction techniques like PCA.

Mathematically the problem can be formulated as,

$$\min_{\alpha \in \mathbb{R}^p} \|\alpha\|_0 \text{ such that } x = D\alpha$$

where $\|\alpha\|_0$ is the l_0 norm which is nothing but the number of non zero components in the vector α

One such paradigm for sparse signal recovery, compressed sensing, will be explored in this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

IITM	Indian Institute of Technology, Madras
RIP	Restricted Isometry Property
CS	Compressed Sensing

CHAPTER 1

Introduction

The rapid advances in technology in the last century has led to an enormous increase in our capacity to store, process and transmit data. Nyquist and Shannon **?** showed in their pioneering work that it is possible for a discrete time signal to capture all the information from a continuous time signal of a finite bandwidth. This allowed the transition of signal processing from the analog to the digital domain. Sampling theorem piggybacked on the success of Moore's Law resulting in a widespread use of high quality sensing systems. These digital sensing systems are more durable, flexible and inexpensive compared to their analog equivalents.

Despite the tremendous success of sampling signals at the Nyquist rate, there are emerging applications where the data acquired is extremely high dimensional. Sampling such high dimensional signals at Nyquist rate, we could end up with a large number of samples or it may simply not be cost effective or even physically possible to build such a sensing system. The acquisition and processing of signals in application areas such as imaging, video, medical imaging, remote surveillance, spectroscopy, and genomic data analysis continues to pose a tremendous challenge **?**.

Compression techniques, used in dealing with high dimensional data, often rely on the motivation that the data has an intrinsically simple structure. For example, transform coding maps natural data like audio or images to a basis that produces a sparse representation of the signal. This approximates the information of the data by retaining only the largest coefficients of the signal. This process is called sparse approximation and is the basis of transform coding schemes, including the JPEG, JPEG2000, MPEG and MP3 standards **?**.

Since, the actual information content is much lower than the dimensionality of the signal, the conventional approach of sampling a signal at Nyquist rate and then discarding a large part, appears to be inherently excessive. The fundamental idea behind CS is that if the signal has an inherent sparse representation, it is possible to sense the data using fewer observations and have the compression inherent to the sensing process. Candès **?** showed that finite dimensional sparse signals and images can be reconstructed accurately and sometimes exactly by solving a simple convex optimization problem.

The foundation for compressed sensing is based on fundamental mathematical ideas from the areas of linear algebra, linear optimization, probability theory, approximation theory and signal processing.

1.1 Chapter Outline

In Chapter 2, the underlying concepts of compressed sensing are explained. Properties like sparsity, incoherence, their relation to compressed sensing as well as the success of l^1 minimization for sparse recovery are described in this chapter.

Chapter 3 gives a mathematical insight into compressed sensing. Properties like the Restricted Isometry Property and conditions for exact recovery of sparse signal are derived in this chapter. It is assumed that the reader has some knowledge of elementary probability theory.

In Chapter 4, Compressed sensing is applied to a practical example of image detection using fewer observations than the dimension of the original signal. Difference in the representation basis of the sparse signal leads to a difference in the quality of recovery.

Chapter 5 concludes the discourse of compressed sensing and ends the thesis with a discussion on possible future developments.

1.2 Contributions of This Thesis

From the year 2004, when the field of compressed sensing was explored for the first time by David Donoho, Emmanuel Candès, Justin Romberg and Terence Tao, CS has now become a popular subject in applied mathematics and signal processing. It has been applied to fields as diverse as biology, magnetic resonance imaging, astronomical imaging, infrared imaging, machine learning, facial recognition and even in geology. This thesis aims to serve as an introductory guide to the enormous field of CS. The wide range of applications of CS means that it is not just electrical engineers with a background in signal processing who may need to learn such a tool. It is aimed at undergraduates who have an elementary understanding of mathematics, specifically probability theory. The area of CS is a relatively recent field of study and this thesis explores the underlying concepts that one who wishes to go deeper into the field must have a clear grasp of. This thesis is essentially divided into three parts. The first part intends to lay the heuristic foundation of the key principles behind CS. The second part builds on the first part and is directed towards the mathematics that governs the principles discussed in the previous part. In the final part, we take a jump from theory into practice and deal with applying CS to the problem of image detection. All the graphs and images have been generated by the author unless stated otherwise and the Matlab codes for generating them have been uploaded on the Internet for anyone to wishes to see, use or learn.

CHAPTER 2

Introduction to Compressed Sensing

Compressed Sensing deals with extracting information about a signal using as few measurements as possible. This can be better viewed by first looking at a simpler problem - the well known puzzle of finding the counterfeit coin.

2.1 Counterfeit Coin Problem

To understand the basic ideas of compressed sensing, we look at the counterfeit coin problem. Suppose we have 8 coins of which one is counterfeit, what is the minimum number of weighings needed to find the counterfeit coin? Clearly we can find the counterfeit by weighing each coin separately but this requires 8 weighings at most. Using binary search, we can always narrow down on the counterfeit in 3 weighings.

An alternate way to look at this is to create a weighing matrix Φ. Element at row k and column j indicates whether the jth coin is being weighed in the kth weighing. 8×1 vector x represents each coin's weight deviation from the nominal. Since we are assuming that only one coin is counterfeit, x has only one non zero element. 3×1 vector b, is the outcome of the weighing. b can take 8 distinct values depending on which of the 8 coin is counterfeit. For N coins, one can deduce which coin is counterfeit by looking at the results of $\log_2 N$ measurements.

Can we extend this result to the case when we have more than one counterfeit coin? This turns out to be an instance of combinatorial group testing where we perform tests on groups of elements to locate elements that have a required property. As an example, group testing and compressed sensing finds application in measuring the gene expression levels. In contrast to conventional DNA microarrays, in which each genetic sensor is designed to respond to a single target, in compressive sensing microarrays, each sensor responds to a set of targets **?**.

To answer the question above, let us first assume that the number of counterfeits m and total number of coins N are such that $m \ll N$. We want to recover x using fewer measurements than N. Our sensing matrix is $\Phi_{m\times N}$ and the measured vector is b. The system of equations $\Phi x = b$ is underdetermined and can thus have infinitely many solutions. To recover x from b we find solutions of $\Phi x = b$ where x has fewest non zero components.

$$\min_{x\in\mathbb{R}^N} \|x\|_0 \text{ such that } \Phi x = b \tag{2.1}$$

The quantity $\|x\|_0$ denotes the number of non zero components of x and is called as the "l^0 norm". Minimizing l^0 norm is a procedure combinatorial in nature and has exponential complexity, and has led researchers to develop alternatives **?**. One such alternative comes from the fact that when x is sufficiently sparse, the l^1 norm minimization and l^0 norm minimization solutions coincide **?** where the l^1 norm is defined as

$$\|x\|_1 = \sum_{i=1}^{N} |x_i|$$

2.2 l^1 Minimization

The success of applying l^1 minimization to recover the signal perfectly has been illustrated by Bryan and Leise **?** who performed l^1 regularization on the counterfeit coin problem for $N = 100$ coins for different number of weighings, with each coin being chosen for every weighing with a probability of 0.5. The probability of successful recovery using l^1 minimization for varying number of counterfeits and for varying number of weighings is shown in Figure **??**.

$$\min_{x\in\mathbb{R}^N} \|x\|_1 \text{ such that } \Phi x = b \tag{2.2}$$

There exist different algorithms to solve the problem of l^1 regularization **?**. Why do we use l^1 minimization and not something else like l^2 minimization? We perform the

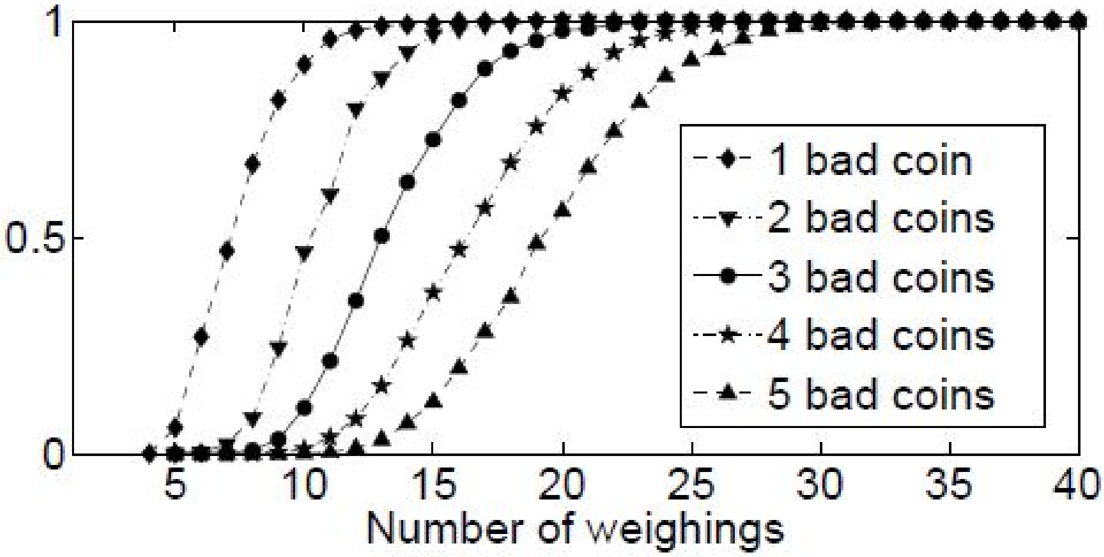


Figure 2.1: Probability of successful recovery for different number of counterfeits and different number of weighings.(Bryan and Leise, 2013)

following 2 experiments to get an insight into why l^1 is the more appropriate choice in our case.

In our first experiment, we generate a matrix $A_{200\times100}$ where each element A_ij is an i.i.d. standard normal variable and a measurement vector $b_{200\times1}$ whose every element is also an i.i.d. standard normal variable. This is an overdetermined system and we fit $x_{100\times1}$ to this data using both l^1 and l^2 regularization. Comparing the results we see that minimizing $\|Ax-b\|_1$ sets a large number of components to zero while minimizing $\|Ax-b\|_2$ doesn't have the same effect. Intuitively this could be understood because l^1 places less weight on larger residuals and more weight on smaller residuals when compared to l^2 and thus causes a large number of residuals to be exactly zero.

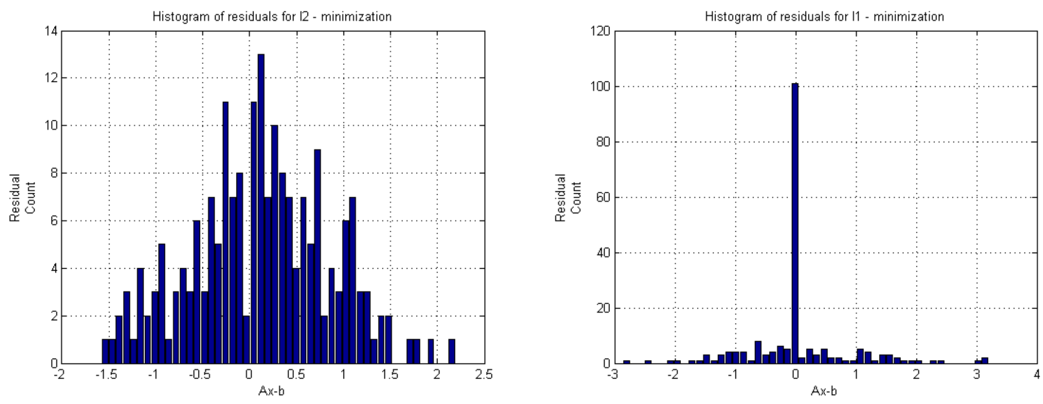


Figure 2.2: Residual distribution $Ax-b$ for l^2 regularization on the left when compared with that of l^1 regularization on the right.

In our second experiment, we generate a matrix $A_{100\times200}$ where each element A_ij is an i.i.d. standard normal variable and a measurement vector $b_{100\times1}$ whose every element is also an i.i.d. standard normal variable. This is an underdetermined system and we find a solution $x_{200\times1}$ to this system using both l^1 and l^2 regularization. Again, this leads to very different kinds of solution. A lot of the components of x are set to zero when l^1 regularization is used.

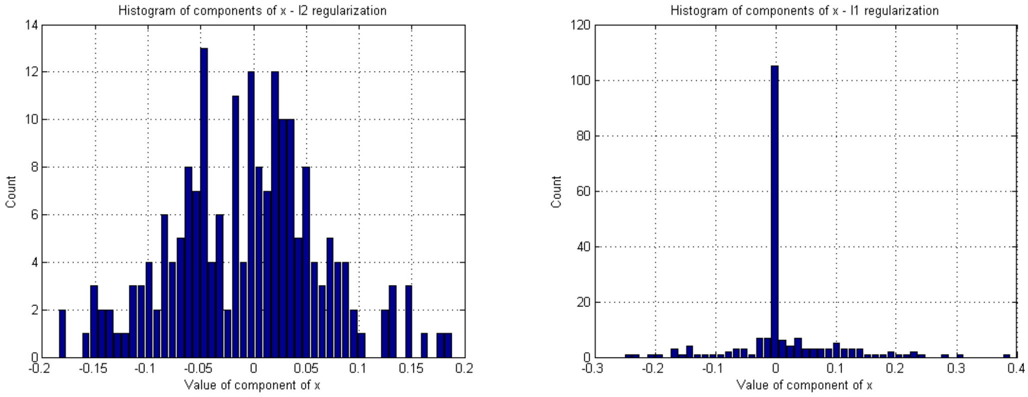


Figure 2.3: Distribution of components of x for l^2 regularization on the left when compared with that of l^1 regularization on the right.

Figure **??** shows, geometrically, why l^1 regularization works while l^2 regularization does not for a three dimensional equivalent of our problem. To find the solution to the plane, increase the size of the l^1 or the l^2 ball until it meets the plane. Clearly, such l^1 solutions have more components set to zero when compared to the l^2 solutions.

To understand CS it is important to first understand the two key concepts on which it relies - sparsity and incoherence. These two conditions are required for the recovery of the signal to be possible.

2.3 Sparsity

Sparsity is a property of the signal. It comes from the fact that signals may have a much lower information content than that suggested by the sampling theorem. The actual number of independent variables may be much smaller than the length of the signal. CS makes use of the knowledge that real world signals usually have a sparse representation when expressed using the right basis Ψ.

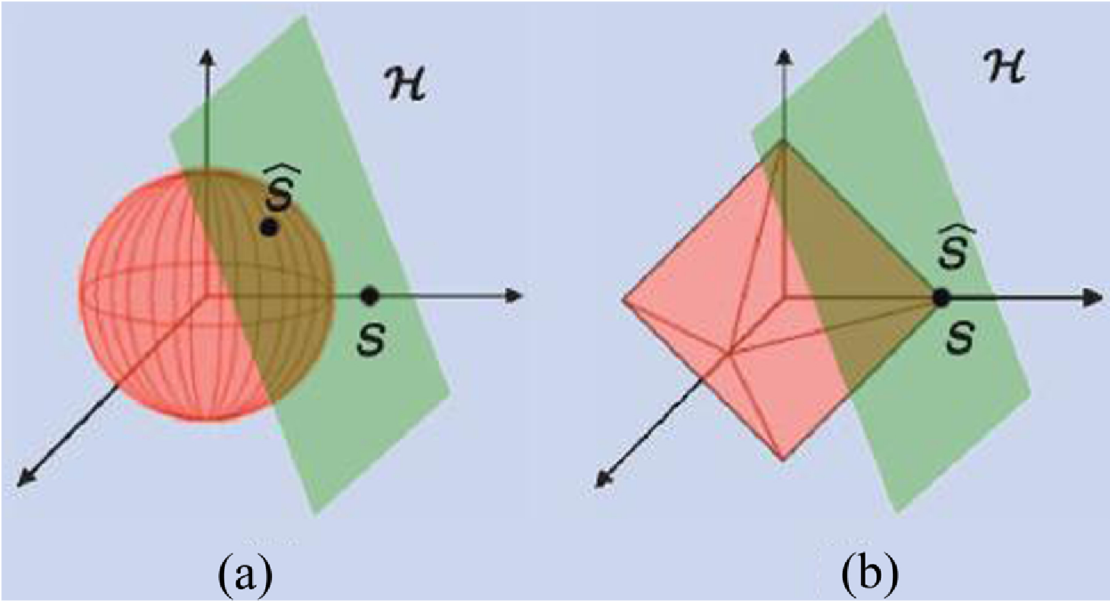


Figure 2.4: (a) Intersection of the plane of possible solutions $\mathcal{H}$ with l^2 ball of minimum radius doesn't lead to a sparse solution. (b) Intersection of the plane of possible solutions $\mathcal{H}$ with l^1 ball of minimum radius leads to a sparse solution. (Baraniuk 2007 **?**)

To illustrate sparsity, we will take an image, transform it to an appropriate orthogonal basis and then throw out the components of this transform that have values lower than a certain threshold. We then perform an inverse transform to get an approximation of the original image. Figure **??** shows the difference between the two images. The transformation applied is the discrete wavelet transform using Deslauriers wavelets. 95.35% of the components were discarded during the compression. The difference in the first thousand wavelet coefficients of the two images is shown in Figure **??** illustrating the negligible effect on quality of setting a major chunk of coefficients to zero. JPEG2000 uses 2D discrete wavelet transform for image compression **?**.

2.4 Incoherence

The basis Ψ is such that the signal has a sparse representation in it. Apart from that, there exists the sensing basis Φ. Sparsity deals with the signal and the basis Ψ whereas incoherence relates the signal to the sensing modality Φ.

CS requires that the signal have a dense representation in the sensing basis Φ. For example, suppose we want to measure a signal that is sparse in time such that it has a few spikes at some unknown time instances. If our sensing modality was also such that

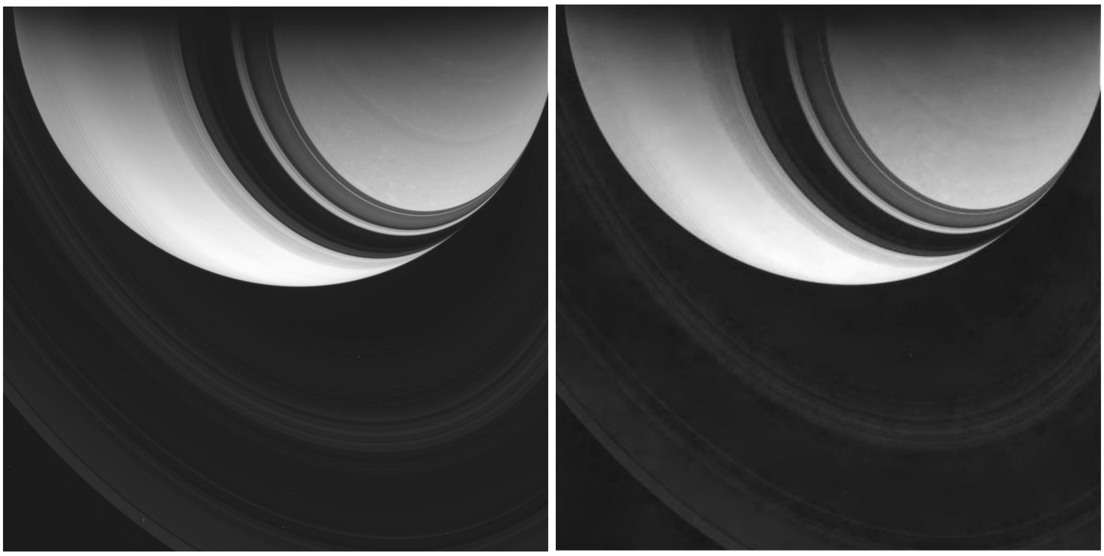


Figure 2.5: Comparison of original image vs compressed image(PSNR 44.02dB) containing only the largest 4.65% of the discrete wavelet transform components done using Deslauriers wavelets.

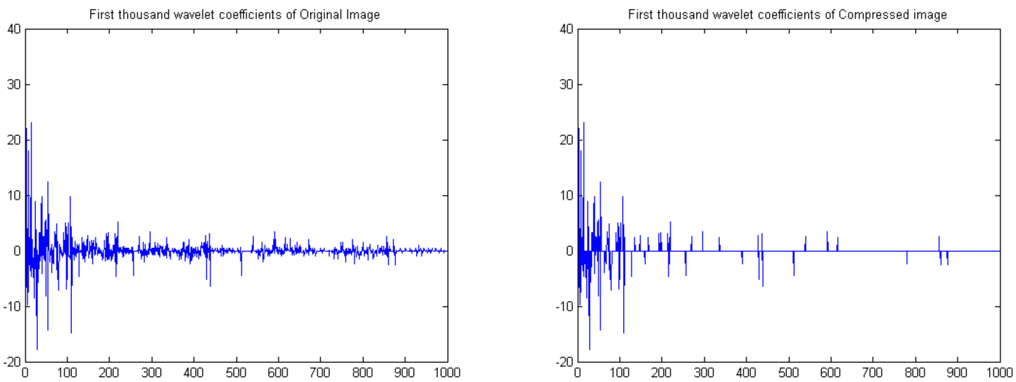


Figure 2.6: Comparison of the first 1000 wavelet coefficients of the original and the compressed image shows how majority of the coefficients are set to zero for the compressed image.

it gave us the values of the signal at a few instances of time, we would need a lot of measurements to be sure that no information is lost. Now consider the case when our sensing basis was the Fourier basis and we had a way to sense the Fourier transform of the previous signal at some frequencies. Then using these measurements, the signal can be recovered exactly. On the other hand if we wanted to sense sinusoids instead of spikes, we would need to sample the signal at a few instances of time since sinusoids have a sparse representation in the Fourier domain.

Intuitively, every measurement should gather a small knowledge about each component. So although the signal is sparse, a few measurements give us sort of a linear combination of every component of the sparse signal.

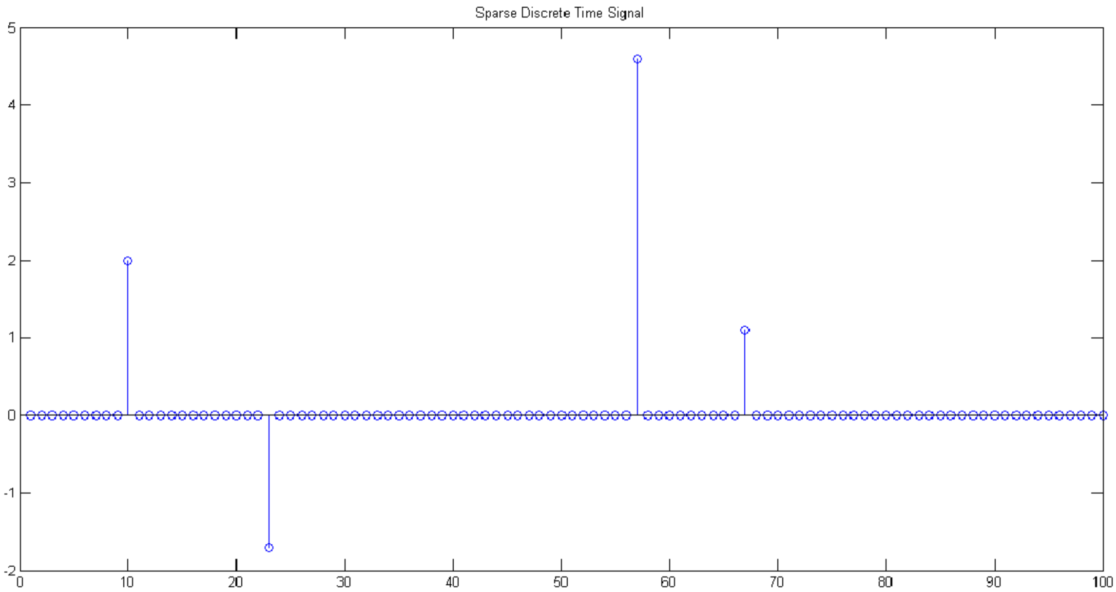


Figure 2.7: Discrete time sparse signal to be recovered using coherent(time) and incoherent(frequency) systems

To illustrate the difference in performing sparse signal recovery between a coherent system and an incoherent system, we perform the following experiment. We create a sparse discrete time signal, X as shown in Figure **??** whose values are $X(10) = 2$, $X(23) = -1.7$, $X(57) = 4.6$, $X(67) = 1.1$ and 0 elsewhere. For the first part, we choose a sensing basis that is incoherent. To do this we sample the discrete Fourier transform of the above signal at 15 different randomly selected frequencies. So our sensing matrix is $\Phi_{15\times 100}$ and our measurement vector is $b_{15\times 1}$. On recovering x by performing l^1 minimization we get the exact solution 998 times out of 1000. One such successful recovery is shown in Figure **??**. For the second part, we choose a sensing basis that is coherent. We sample the signal at 75 different, randomly chosen instances of time. On recovering x by performing l^1 minimization we get the exact solution just 321 times out of 1000.

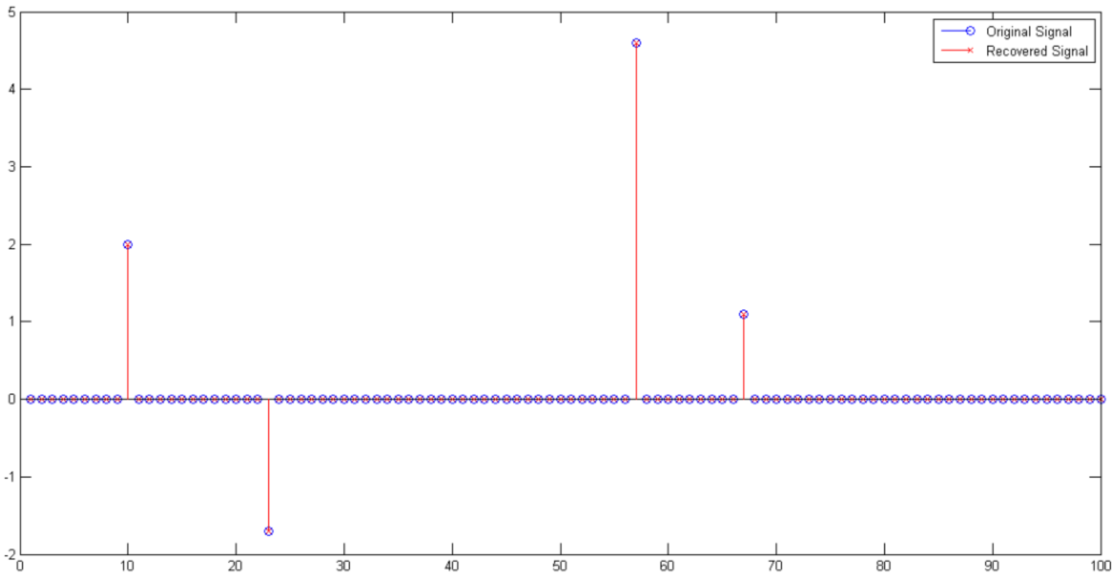


Figure 2.8: Comparison of original signal and signal recovered using just 15 measurements.

Even after sampling at 75 different time instances, exact recovery of the signal happened roughly 30% of the times. To explain this, let's look at the probability that the 4 time instances, $t = 10, 23, 57, 67$ are present in the 75 samples.

$$\mathbb{P}\left(t = 10, 23, 57, 67 \in 75 \text{ random samples}\right) = \frac{\binom{96}{71}}{\binom{100}{75}} \approx 31\%$$

As seen above time and frequency are maximally incoherent. Signals compact in frequency domain (sine waves) will be spread out in the time domain and vice versa (spikes). Random noise is incoherent with almost any signal. We have already seen an instance of using randomness to create an incoherent basis which was used to detect the counterfeits by Bryan and Leise **?**. We will look at other ways to use randomness to generate sensing systems that are incoherent with almost any fixed representation basis Ψ. The key takeaway we should note is that depending on what signal we want to sense, we can design an efficient sensing scheme that minimizes the number of measurements needed to recover the sparse signal.

CHAPTER 3

Mathematical Introduction to Compressed Sensing

In the last chapter, we got an instinctive grasp of CS by looking at the counterfeit coin problem. We gained a better understanding of how l^1 regularization is fundamentally different from other types of regularization techniques and the concepts of sparsity and incoherence were introduced. In this chapter, we will dive deeper into the fundamental mathematical ideas that will help us realize how and why CS works. We will first look at why some matrices created using some random process help us in recovering the sparse solution. Then we will look at the conditions under which l^1 minimization successfully recovers the sparse solution.

3.1 Mathematical Definitions and Assumptions

We wish to recover $x_{N\times 1}$ which is a solution to the underdetermined system $b = \Phi x$, where $\Phi_{m\times N}$, $m \ll N$, is the sensing matrix and $b_{m\times 1}$ is the measurement vector. CS works when the signal we want to recover is sparse in some representation basis Ψ. For the sake of simplicity, we will assume that the signal x, is sparse in the canonical basis. This implies that most of the components of x will be zero. Let the number of non zero components of x be k. Clearly $k \ll N$. We say that a vector is k-sparse if it has at most k non zero components.

As discussed earlier, randomness can be used to generate sensing systems that are incoherent with almost any representation basis. One such way to generate an incoherent sensing system is to let each element Φ_{ij} of the matrix Φ be an i.i.d. Gaussian random variable of mean 0 and variance $1/m$. Thus, $\Phi_{ij} \sim \mathcal{N}(0, 1/m)$. These definitions will be used in the theorems that follow in this chapter.

How can we be sure that the solution recovered is indeed the signal we wish to detect? One important property that we want our system to have is the uniqueness of solution. Otherwise, if we have two distinct signals giving the same measured vector, we will have no way of looking at the measurement to deduce what the original signal was. So let's look at a property that if satisfied by our sensing matrix Φ, we are guaranteed that the solution we are looking for, if it exists, is unique.

3.2 Restricted Isometry Property

The matrix Φ is said to satisfy the Restricted Isometry Property(RIP) of order k with constant δ if there exists some $\delta \in (0, 1)$ such that

$$(1-\delta)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1+\delta)\|x\|_2^2 \tag{3.1}$$

for all k-sparse vectors $x \in \mathbb{R}^N$

Theorem 3.1. *If a matrix Φ satisfies the RIP of order $2k$ for some integer $k \geq 1$, then any k-sparse solution to the system of equations $\Phi x = b$ is unique.*

Proof. To prove it by contradiction let's assume that the statement is false, i.e. there exists two distinct k-sparse solutions x_a and x_b.

$$\Phi x_a = b$$

$$\Phi x_b = b$$

$$\Phi(x_a - x_b) = 0$$

Let $\eta = x_a - x_b$. Thus, $\Phi\eta = 0$. Since x_a and x_b are k-sparse, η is $2k$-sparse. However, matrix Φ satisfies the RIP. Therefore, there exists some $\delta \in (0, 1)$ such that $\Phi\eta \geq (1-\delta)\|\eta\|_2^2$. But $\Phi\eta = 0$.

Assuming the statement to be false leads us to a contradiction. Hence, if a solution exists, it must be unique. □

We want our sensing system Φ to satisfy the RIP so according to Theorem **??** if any solution exists, we know that it is unique. It should be noted that RIP is only a sufficient condition for a unique solution. Verifying that a matrix Φ satisfies the RIP is an NP hard problem **?**. However, certain matrices created using randomness can be shown to satisfy the RIP with high probability. Now, we will prove that the Gaussian matrix as defined in the previous section satisfies the RIP with high probability. To do this, in Theorem **??** a result similar to RIP is shown to hold for any fixed $x \in \mathbb{R}^N$. This result is then extended to all k-sparse unit vectors with the help of Theorems **??**, **??** and **??**.

Theorem 3.2. *For any $\delta \in (0, 1/2)$ and any fixed vector $x \in \mathbb{R}^N$, the inequality*

$$(1 - \delta) \|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta) \|x\|_2^2$$

holds with probability greater than $1 - 2e^{-\delta^2 m/8}$

Proof. We know that each element of the matrix $\Phi_{ij} \sim \mathcal{N}(0, 1/m)$. Let us express the random variable $\|\Phi x\|_2^2$ that we are trying to bound. The ith element of Φx will be $(\Phi x)_i = \sum_{j=1}^N \Phi_{ij} x_j$. Thus $(\Phi x)_i$ is a sum of weighted i.i.d. Gaussian random variables. Therefore, $(\Phi x)_i$ is also a Gaussian with mean $= \sum_{j=1}^N E[\Phi_{ij}] x_j = 0$ and variance $= \sum_{j=1}^N Var[\Phi_{ij}] x_j^2 = \|x\|_2^2 / m$.

Thus,

$$(\Phi x)_i \sim \mathcal{N}\left(0, \|x\|_2^2 / m\right)$$
$$\frac{\sqrt{m}}{\|x\|_2} (\Phi x)_i \sim \mathcal{N}(0, 1)$$

The sum of the squares of m independent standard normal random variables is the chi-squared distribution with m degrees of freedom. Therefore,

$$\frac{m}{\|x\|_2^2} \|\Phi x\|_2^2 \sim \chi^2(m)$$

From Chernoff Bound, we know that for a random variable X and for every $t > 0$,

$$P(X \geq a) \leq \frac{E[e^{t.X}]}{e^{t.a}}$$

To get the upper bound on $\|\Phi x\|_2^2$, we set X as $\frac{m}{\|x\|_2^2} \|\Phi x\|_2^2$, which is nothing but the chi-squared distribution with m degrees of freedom, and a as $(1 + \delta) m$. Also, $E\left[e^{t.X}\right]$ is the moment generating function of X. Moment generating function of chi squared distribution with m degrees of freedom is $(1-2t)^{-m/2}$ Therefore,

$$\begin{aligned} P\left(\|\Phi x\|_2^2 \geq (1+\delta)\|x\|_2^2\right) &\leq \frac{E[e^{t.X}]}{e^{t(1+\delta)m}} \\ &\leq \frac{(1-2t)^{-m/2}}{e^{t(1+\delta)m}} \\ &\leq \left[\frac{e^{-2t(1+\delta)}}{1-2t}\right]^{m/2} \end{aligned} \tag{3.2}$$

We choose t such that the term inside the bracket is minimum. Differentiating the term with respect to t and setting it to zero, we get the minimum.

$$\frac{d}{dt}\left[\frac{e^{-2t(1+\delta)}}{1-2t}\right] = 0$$

Solving, we get,

$$t = \frac{\delta}{2(1+\delta)}$$

Substituting t back in Equation **??** and since $\delta \in (0,1)$, we get,

$$\begin{aligned} P\left(\|\Phi x\|_2^2 \geq (1+\delta)\|x\|_2^2\right) &\leq \left[(1+\delta)e^{-\delta}\right]^{-m/2} \\ &\leq e^{-\left(\delta^2-\delta^3\right)m/4} \\ &\leq e^{-\delta^2 m/8} \end{aligned}$$

Similarly, to get the lower bound of the inequality **??**, use the following variation of Chernoff Bound.

$$P(X \leq a) = P\left(e^{-tX} \geq e^{-ta}\right) \leq \frac{E\left[e^{-tX}\right]}{e^{-ta}}$$

On solving as we did for the upper bound, we get,

$$\begin{aligned} P\left(\|\Phi x\|_2^2 \leq (1-\delta)\|x\|_2^2\right) &\leq \left[(1-\delta)e^{-\delta}\right]^{-m/2} \\ &\leq e^{-\left(\delta^2-\delta^3\right)m/4} \\ &\leq e^{-\delta^2 m/8} \end{aligned}$$

Thus for any fixed vector $x \in \mathbb{R}^N$, the inequality **??** holds with probability greater than $1-2e^{-\delta^2 m/8}$ □

We have proved that the inequality holds with high probability for a single vector x. Let T be a subset of size k of the set $\{1, 2, ..., N\}$. Let us denote by X_T the set of vectors whose elements are zero when their indices are outside T. We will first extend the result of the Theorem **??** to a fixed finite set of k sparse vectors. Then, we shall extend the result to all vectors of the set X_T for any fixed T with $|T| = k$ and finally to all such possible sets T i.e. the set of all k sparse vectors.

Theorem 3.3. *Fix any subset $T \subset \{1, 2, ..., N\}$ such that $|T| = k$. Let A_S be a finite subset of X_T such that that $|A_S| = S$. Then, for any $\delta \in (0, 1/2)$ the inequality*

$$(1 - \delta) \|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta) \|x\|_2^2$$

holds concurrently for all $x \in A_S$ with probability greater than $1 - 2Se^{-\delta^2 m/8}$.

Proof. For a fixed x, the inequality fails to hold with a probability less than $2e^{-\delta^2 m/8}$. We want the inequality to hold for all $x \in A_S$. This is nothing but the complement of the event that the inequality fails to hold for at least one of the vector $x \in A_S$. The event that the inequality fails to hold for at least one of the vector is the union of the events that the inequality fails to hold for each vector. From the union bound we know that, $P(A_1 \cup A_2 \cup ... \cup A_j) \leq P(A_1) + P(A_2) + ... + P(A_j)$ Thus, the probability that the inequality fails to hold for at least one of the vector is bounded above by $2Se^{-\delta^2 m/8}$. Hence, the inequality holds for all $x \in A_S$ with probability greater than $1 - 2Se^{-\delta^2 m/8}$. □

Let $\mathbb{S}^{k-1} = \left\{u \in \mathbb{R}^k : \|u\|_2 = 1\right\}$ denote the unit sphere.

Now let's look at a way to come up with an appropriate finite subset $A_S(\epsilon)$. Choose $\epsilon \in (0, 1)$ and $k \geq 2$. Then at each step i, choose point p_i on the unit sphere such that $\|p_i - p_j\| \geq \epsilon$ for all $j < i$. This process will terminate once you can find no such point on the unit sphere and hence there is a bound on the size of our set A_S. To get this bound, imagine spheres of radius $\epsilon/2$ centered at each of the points in our finite set A_S. Clearly none of these spheres overlap otherwise they would not have been part of our set because of the way we constructed it. Also all these spheres are enclosed in the sphere centered at the origin having radius $1+\epsilon/2$. Thus $S.Vol(r = \epsilon/2) \leq Vol(r = 1+\epsilon/2)$. Also, $Vol(r = R) = Const.R^k$. This gives us $S \leq (1 + 2/\epsilon)^k \leq \left(\frac{3}{\epsilon}\right)^k$

Theorem 3.4. *Fix any subset $T \subset \{1, 2, ..., N\}$ such that $|T| = k$. Then, for any $\delta \in (0, 1/2)$ the inequality*

$$(1-\delta)\left\|x\right\|_2 \leq \left\|\Phi x\right\|_2 \leq (1+\delta)\left\|x\right\|_2$$

holds for all $x \in X_T$ with probability greater than $1 - 2\left(12/\delta\right)^k e^{-\delta^2 m/32}$.

Proof. Note that we are working with the norms and not the square of the norms unlike in our previous theorems. Additionally, it will be enough to prove that the inequality holds with high probability for $\left\|x\right\|_2 = 1$, because Φ is linear. Construct a finite subset of points $A_S\left(\delta/4\right)$ as described in the previous paragraph. $\left|A_S\left(\delta/4\right)\right| \leq \left(12/\delta\right)^k$. Applying Theorem **??** to the finite subset $A_S\left(\delta/4\right)$,

$$1 - \delta/2 \leq \left\|\Phi p_i\right\|_2^2 \leq 1 + \delta/2 \text{ holds } \forall\, p_i \in A_S\left(\delta/4\right)$$

with probability greater than $1 - 2\left(12/\delta\right)^k e^{-\delta^2 m/32}$.

Let us define A to be the smallest number such that

$$\left\|\Phi x\right\|_2 \leq (1+A)\left\|x\right\|_2 \;\forall\, x \in X_T \text{ such that } \left\|x\right\|_2 = 1 \tag{3.3}$$

We intend to show that $A \leq \delta$ with a very high probability. Now for any $x \in X_T$ such that $\left\|x\right\|_2 = 1$, there exists a point $p_i \in A_S\left(\delta/4\right)$ such that $\left\|p_i - x\right\| \leq \delta/4$

$$\begin{aligned}
\left\|\Phi x\right\|_2 &= \left\|\Phi\left(x - p_i\right) + \Phi p_i\right\|_2 \\
&\leq \left\|\Phi\left(x - p_i\right)\right\| + \left\|\Phi p_i\right\| \\
&\leq (1+A)\left\|x - p_i\right\| + \sqrt{1 + \delta/2} \\
&\leq (1+A)\,\delta/4 + 1 + \delta/2
\end{aligned}$$

The above inequality holds true with probability larger than $1 - 2\left(12/\delta\right)^k e^{-\delta^2 m/32}$. Further since the inequality holds for any x, it also holds true for that x for which A was the smallest in **??**. Thus, $1 + A \leq (1+A)\,\delta/4 + 1 + \delta/2$. This simplifies to $A \leq \frac{3\delta}{(4-\delta)}$. Since $\delta \in (0, 1/2)$, $A \leq \delta$ with probability greater than $1 - 2\left(12/\delta\right)^k e^{-\delta^2 m/32}$.

It is straightforward to get the other inequality by using the triangle inequality differently $\|\Phi p_i\|_2 = \|\Phi (p_i - x) + \Phi x\|_2 \leq \|\Phi (p_i - x)\|_2 + \|\Phi x\|_2$. This leads to,

$$\begin{aligned} \|\Phi x\|_2 &\geq \|\Phi p_i\|_2 - \|\Phi (p_i - x)\|_2 \\ &\geq \sqrt{1 - \delta/2} - (1 + A) \|p_i - x\|_2 \\ &\geq 1 - \delta/2 - (1 + \delta)\delta/4 \\ &\geq 1 - \delta \end{aligned}$$

with probability greater than $1 - 2 (12/\delta)^k e^{-\delta^2 m/32}$. □

Theorem 3.5. *For any $\delta \in (0, 1/2)$, the sensing matrix Φ whose elements are i.i.d. standard normal random variables, satisfies the RIP of order k with constant δ with probability greater than $1 - 2 \left(\frac{36Ne}{k\delta}\right)^k e^{-\delta^2 m/288}$.*

Proof. Applying Theorem **??** with $\delta/3$ in place of δ, demonstrates that the inequality

$$(1 - \delta/3) \|x\|_2 \leq \|\Phi x\|_2 \leq (1 + \delta/3) \|x\|_2$$

holds for all $x \in X_T$ with probability greater than $1 - 2 (36/\delta)^k e^{-\delta^2 m/288}$. On squaring,

$$(1 + \delta) \|x\|_2^2 \leq (1 - \delta/3)^2 \|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta/3)^2 \|x\|_2^2 \leq (1 + \delta) \|x\|_2^2$$
$$\therefore (1 + \delta) \|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta) \|x\|_2^2$$

holds with a high probability for a fixed subset $T \subset \{1, 2, ..., N\}$ such that $|T| = k$. There are $\binom{N}{k}$ such subsets and we want the inequality to hold for all of them simultaneously. It fails to hold with probability less than $2 (36/\delta)^k e^{-\delta^2 m/288}$ for one such subset. The event that it fails to hold for at least one such subset is nothing but the union of the events that it fails to hold for each subset. But according to Boole's inequality, the probability of at least one event happening is not greater than sum of probabilities of the individual events. Thus the probability P_F that the inequality fails to hold for at least one such subset is less than $\binom{N}{k}$ times that it fails to hold for one such subset.

$$P_F \leq \binom{N}{k} 2 (36/\delta)^k e^{-\delta^2 m/288}$$

It can be shown that $\binom{N}{k}$ is bounded above by $\left(\frac{Ne}{k}\right)^k$. Thus,

$$P_F \leq 2\left(\frac{36Ne}{k\delta}\right)^k e^{-\delta^2 m/288}$$

$\square$

Theorem **??** shows us that for any combination of N, k, δ we can choose the number of observations m such that the probability that RIP fails to hold for Φ is very small. On rearranging terms, we see that $m = O\left(k \log\left(N/k\delta\right)/\,\delta^2\right)$. Davenport **?** showed that it is not possible for a matrix Φ to satisfy the RIP if the number of observations m is not greater than $C_\delta k \log\left(N/k\right)$. Bounds obtained in Theorem **??** are not exact and sensing matrix Φ can satisfy the RIP for smaller number of observations than that suggested by the theorem.

3.3 Exact Recovery Using l^1 Minimization

We have established conditions under which the sensing matrix Φ satisfies RIP with high probability which is useful in determining the uniqueness of the recovered solution. Now, under what conditions does the l^1 minimization result in exact recovery of the original signal? The following theorem answers this question.

Theorem 3.6. *Let Φ be an $m \times N$ matrix that satisfies RIP of order $3k$ with constant $\delta_{3k} < 1/3$. Let x_0 be a k-sparse vector such that we observe the measured vector $b = \Phi x_0$. Given measured vector b, performing the l^1 minimization*

$$\min_{x \in \mathbb{R}^N} \|x\|_1 \text{ subject to } \Phi x = b$$

exactly recovers the k-sparse vector x_0.

Proof. Let x'' be the solution to

$$\min_{x \in \mathbb{R}^N} \|x\|_1 \text{ subject to } \Phi x = b$$

Let $\eta = x'' - x_0$. Since x'' and x_0 are both solutions to the underdetermined system, $\Phi x_0 = b$ and $\Phi x'' = b$. Thus,

$$\Phi\eta = \Phi\left(x'' - x_0\right) = \Phi x_0 - \Phi x'' = 0 \tag{3.4}$$

Also since x'' is the solution with the least possible l^1 norm, we have,

$$\|x''\|_1 = \|x_0 + \eta\|_1 \leq \|x_0\|_1 \tag{3.5}$$

We will show that when Φ satisfies RIP of order $3k$ with constant $\delta < 1/3$, the above two conditions are contradictory and imply that $\eta = 0$.

For any subset $T \subset 1, 2, ..., N$ and any vector η, η_T is another vector defined to be the same as vector η at indices that are present in the set T and zero at indices that are not present in the set T.

Let T_0 be the set of indices at which x_0 is not zero. Thus its complement T_0^C denotes those indices at which x_0 is zero. Since x_0 is k-sparse, $|T_0| = k$. Let T_1 be the set of indices of the $2k$ elements largest in magnitude in $\eta_{T_0^C}$. Let T_2 be the set of indices of the next $2k$ elements largest in magnitude in $\eta_{T_0^C}$ and define such sets of size $2k$ until T_s. Note that the size of the last such set T_s could be any value between 1 and $2k$ depending on N.

Before proceeding, let us convert the Equation **??** to a form that will be more useful.

$$\begin{aligned}
\|x_0\|_1 \geq \|x''\|_1 &= \|x_0 + \eta\|_1 \\
&= \|x_0 + \eta_{T_0} + \eta_{T_0^C}\|_1 \\
&\geq \|x_0 + \eta_{T_0^C}\|_1 - \|\eta_{T_0}\|_1 && \text{(Triangle inequality)} \\
&\geq \|x_0\|_1 + \|\eta_{T_0^C}\|_1 - \|\eta_{T_0}\|_1 \\
\Rightarrow \|\eta_{T_0}\|_1 &\geq \|\eta_{T_0^C}\|_1
\end{aligned} \tag{3.6}$$

Define $T = T_0 \cup T_1$. Therefore, $T_C = \cup_{i=2}^{s}$. Observe that $\Phi\eta = 0 = \Phi\eta_T + \Phi\eta_{T^C}$. Therefore, $\Phi\eta_T = -\Phi\eta_{T^C}$. Clearly η_T is $3k$-sparse, whereas η_{T_i} is $2k$ sparse for all

$i \geq 1$.

$$\begin{aligned}
\|\eta_{T_0}\|_1 &\leq \sqrt{k}\|\eta_{T_0}\|_2 && \text{(Using Cauchy Schwarz inequality)} \\
&\leq \sqrt{k}\|\eta_T\|_2 && \text{since } \|\eta_T\|_2 = \|\eta_{T_0}\|_2 + \|\eta_{T_1}\|_2 \\
&\leq \frac{\sqrt{k}}{\sqrt{1-\delta}}\|\Phi\eta_T\|_2 && \text{(RIP of order 3k)} \\
&\leq \frac{\sqrt{k}}{\sqrt{1-\delta}}\|\Phi\eta_{T^C}\|_2 && \text{since } \Phi\eta_T = -\Phi\eta_{T^C} \\
&\leq \frac{\sqrt{k}}{\sqrt{1-\delta}}\sum_{i=2}^{s}\|\Phi\eta_{T_i}\|_2 && \text{(Applying the triangle inequality)} \\
&\leq \frac{\sqrt{(1+\delta)\,k}}{\sqrt{1-\delta}}\sum_{i=2}^{s}\|\eta_{T_i}\|_2 && \text{(RIP of order 2k)} \qquad (3.7)
\end{aligned}$$

We have a bound on the l^2 norm on the right hand side. We would like to have terms with l^1 norm instead. By construction, every element in η_{T_i} is larger in magnitude than every element in $\eta_{T_{i+1}}$. Thus, average of the magnitude of elements in η_{T_i}, which is equal to $\frac{\|\eta_{T_i}\|_1}{2k}$, is larger than every element in $\eta_{T_{i+1}}$. Thus,

$$\begin{aligned}
\|\eta_{T_{i+1}}\|_2 &\leq \sqrt{\left(\frac{\|\eta_{T_i}\|_1}{2k}\right)^2 + \left(\frac{\|\eta_{T_i}\|_1}{2k}\right)^2 + \cdots + \left(\frac{\|\eta_{T_i}\|_1}{2k}\right)^2} \\
&\leq \sqrt{\left(\frac{\|\eta_{T_i}\|_1}{2k}\right)^2 \times 2k} \\
&\leq \frac{\|\eta_{T_i}\|_1}{\sqrt{2k}} \qquad (3.8)
\end{aligned}$$

Substituting inequality **??** in **??**, we get,

$$\begin{aligned}
\|\eta_{T_0}\|_1 &\leq \frac{\sqrt{(1+\delta)\,k}}{\sqrt{1-\delta}}\sum_{i=1}^{s-1}\frac{\|\eta_{T_i}\|_1}{\sqrt{2k}} \\
&\leq \sqrt{\frac{1+\delta}{2\,(1-\delta)}}\sum_{i=1}^{s-1}\|\eta_{T_i}\|_1 \\
&\leq \sqrt{\frac{1+\delta}{2\,(1-\delta)}}\|\eta_{T_0^C}\|_1
\end{aligned}$$

However, $\delta < 1/3$ implies, $\sqrt{\frac{1+\delta}{2(1-\delta)}} < 1$. Therefore,

$$\|\eta_{T_0}\|_1 \leq \|\eta_{T_0^C}\|_1$$

This contradicts with Inequality **??** which should hold if $x'' \neq x_0$. Thus $x'' = x_0$ and the signal is recovered exactly. $\square$

This concludes our discussion on the mathematics behind CS. The conditions under which there exists a unique solution to the underdetermined system as well as the conditions under which this solution can be exactly recovered have been detailed in this chapter. For those interested in some more fundamental results about CS, Davenport **?** has listed some of them.

CHAPTER 4

Application of Compressed Sensing

The previous two chapters deal with the theory behind CS. In this chapter, we apply CS principles to detect sparse signals. Many real world signals have a sparse representation in some representation basis. As seen in Section **??** one such group of signals are images which have a sparse representation in basis formed by its wavelet transform. We shall see two more such sparse representation basis of images - the Fourier basis and the discrete cosine transform basis.

4.1 Recovering Images Using CS

4.1.1 Formulating the Problem

In Section **??**, we had assumed the signal to be sparse under the canonical basis. However, images are not sparse in the canonical way. The sparsity is apparent when expressed in a different basis. Let us assume that it is sparse in some representation basis Ψ. Ψ could be the wavelet basis, Fourier basis or the DCT basis. Let our image(signal) be denoted as $x \in \mathbb{R}^N$. The basis $\Psi = [\psi_1 \; \psi_2 \cdots \psi_N]$. The signal in terms of the basis will be $x = \Psi s$ where s is a $N \times 1$ vector. Since s is sparse a lot of its components will be zero. In practice, a lot of the components will be much smaller in magnitude. This concise representation of image signal makes it possible to sense it using CS techniques. Our sensing basis is $\Phi_{m \times N}$ $(m < N)$. As discussed in Section **??** our sensing basis Φ must be incoherent with the representation basis Ψ. The measured $m \times 1$ vector will be $b = \Phi x$.

4.1.2 Mathematical Model of the Problem

Since x is sparse in Ψ, we have to choose the solution that gives the minimum l^1 norm of s. Mathematically it is formulated as follows,

$$\min_{s\in\mathbb{R}^N} \|s\|_1 \text{ such that } \Phi\Psi s = b \tag{4.1}$$

However s may not be strictly sparse. Nevertheless, CS holds in such common settings too. We simply relax the constraint $\Phi\Psi s = b$. Instead,

$$\min_{s\in\mathbb{R}^N} \|s\|_1 \text{ such that } \|\Phi\Psi s - b\|_2 \leq \epsilon \tag{4.2}$$

The above formulation leads to an approximate recovery of the signal **?**.

The problem above is solved using Candès and Romberg's Matlab library l1-magic **?**. The above formulation works by itself when the representation matrix is real(DCT), however, for complex representation matrix(DFT), l1-magic fails. To get l1-magic to work for the complex case, i.e. when our representation basis is the Fourier basis, we formulate the problem using just real numbers instead of complex numbers as follows. We wish to regularize **??** when Ψ is complex. Ψ can be written as $R + Ci$, where R is the real part and C is the imaginary part of Ψ. Also $s \in \mathbb{C}^N$. Thus, we have,

$$\min_{s\in\mathbb{C}^N} \|s\|_1 \text{ such that } \Phi\left(R + iC\right)s = b$$

Since s is complex, the constraint can be written as,

$$\Phi\left(R + iC\right)\left(s_R + is_C\right) = b$$
$$\Phi\left[Rs_R - Cs_c + i\left(Rs_C + Cs_R\right)\right] = b$$

Since b is real, this implies,

$$\Phi\left(Rs_R - Cs_c\right) = b$$
$$\Phi\left(Rs_C + Cs_R\right) = 0$$

Now we have two real constraints instead of a complex constraint. We can merge them into a single real constraint as follows

$$\begin{bmatrix} \Phi & 0 \\ 0 & \Phi \end{bmatrix} \begin{bmatrix} R & -C \\ C & R \end{bmatrix} \begin{bmatrix} s_R \\ s_C \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}$$

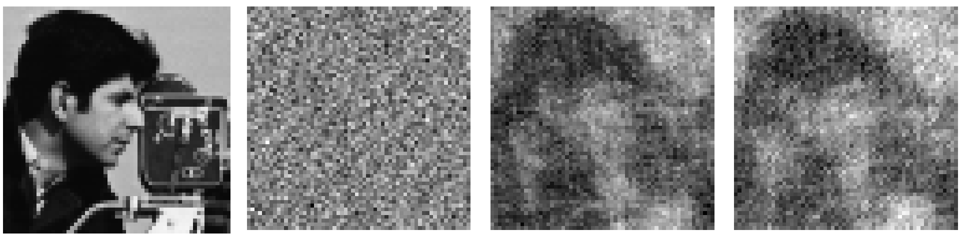
Let $\Phi' = \begin{bmatrix} \Phi & 0 \\ 0 & \Phi \end{bmatrix}$, $\Psi' = \begin{bmatrix} R & -C \\ C & R \end{bmatrix}$, $s' = \begin{bmatrix} s_R \\ s_C \end{bmatrix}$ and $b' = \begin{bmatrix} b \\ 0 \end{bmatrix}$. This gives,

$$\Phi'\Psi's' = b'$$

and this constraint can be implemented using l1-magic since it has no complex numbers in it.

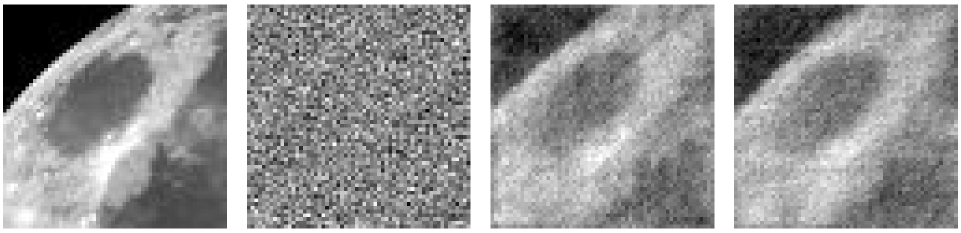
4.2 Results

The simulation was run on three different images. The image was reconstructed using samples $1/4$th of the dimension of the signal. The reconstruction methods used were an l^2 regularization for representation basis (Ψ) as the DCT and l^1 regularization for representation basis as the DCT as well as the DFT.



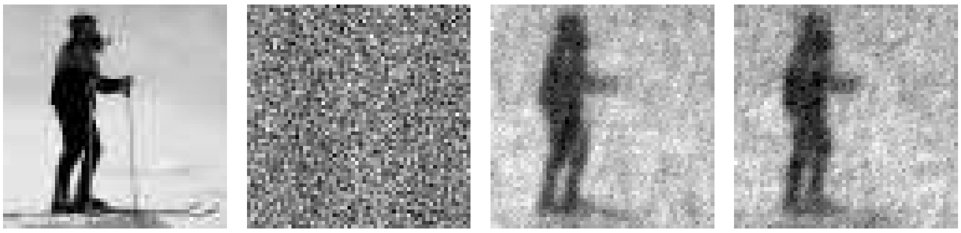
(a) (b) (c) (d)

Figure 4.1: (a) Original image. (b) l^2 recovery. (c) l^1 recovery using the DCT basis. (d) l^1 recovery using the DFT basis, image - cameraman.png



(a) (b) (c) (d)

Figure 4.2: (a) Original image. (b) l^2 recovery. (c) l^1 recovery using the DCT basis. (d) l^1 recovery using the DFT basis, image - moon.jpg



(a) (b) (c) (d)

Figure 4.3: (a) Original image. (b) l^2 recovery. (c) l^1 recovery using the DCT basis. (d) l^1 recovery using the DFT basis, image - ski.jpg

Table 4.1: Comparison in image reconstruction quality for the different recovery methods.

PSNR values(dB) for different images/reconstructions

Reconstruction method	l2	dct l1	dft l1
Image 1 - cameraman.png	8.3893	13.1746	12.1882
Image 2 - moon.jpg	10.3961	17.2207	18.2311
Image 3 - ski.jpg	9.0060	17.6027	17.0713

CHAPTER 5

Conclusion and Discussions

In this thesis, we've looked at compressed sensing, a signal processing technique for reconstruction of sparse signals. The two properties essential for efficient recovery of signals, sparsity and incoherence were discussed in detail in Chapter **??**. This was accompanied by an analysis of l^1 minimization where its characteristics were examined in contrast to the widely used l^2 minimization. Solutions returned by l^1 regularization turned out to be much more sparser than those returned by l^2 regularization. The finer mathematical aspects behind CS were explored in Chapter **??**. First we looked at the Restricted Isometry Property and its implications in finding a unique solution to our underdetermined system of equations. RIP is a sufficient condition for uniqueness of k-sparse solutions, however, it is NP-hard to determine whether a given matrix satisfies RIP. Certain random matrices do satisfy the RIP with a high probability as observed in Theorem **??**. Conditions for exact recovery using l^1 minimization were scrutinized by Theorem **??**. CS was applied using Matlab and the l1-magic library in Chapter **??**. The problem had to be modeled to take into account the complex DFT matrix while using the Fourier basis. The difference in reconstruction for different type of representation basis was examined.

For further topics in CS, **?** applies CS in efficient recovery of higher dimensional signals. **??** deal with finding the solutions of l^1 regularization. For a practical application of CS, take a look at the single pixel camera **?** developed at Rice University. A list of hardware implementation of CS can be found at **?**.

CS finds application in fields varying from biology, geophysics, medical imaging, data compression and channel coding. The field of CS is growing rapidly. It is an interesting branch in applied mathematics that spreads to a wide range of disciplines and will most likely interface with other disciplines soon.

APPENDIX A

Moment Generating Function of a Chi Squared Distribution

The pdf of χ^2 distribution is

$$f_{\chi^2}(x;k) = \begin{cases} \frac{x^{(k/2-1)}e^{-x/2}}{2^{k/2}\Gamma(k/2)} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}\,dx$.

Therefore, assuming $t < 1/2$,

$$\begin{aligned} E[e^{tX}] &= \frac{1}{2^{k/2}\Gamma(k/2)} \int_0^\infty x^{(k-2)/2} e^{-x/2} e^{tx}\,dx \\ &= \frac{1}{2^{k/2}\Gamma(k/2)} \int_0^\infty x^{(k-2)/2} e^{x(t-(1/2))}\,dx \end{aligned}$$

Changing variable, $x(\frac{1}{2} - t) = u$,

$$\begin{aligned} E[e^{tX}] &= \frac{1}{2^{k/2}\Gamma(k/2)} \int_0^\infty \left(\frac{u}{\frac{1}{2}-t}\right)^{(k-2)/2} (e^{-u}) \left(\frac{du}{\frac{1}{2}-t}\right) \\ &= \frac{1}{2^{k/2}\Gamma(k/2)} \frac{1}{(\frac{1}{2}-t)^{k/2}} \int_0^\infty u^{(k-2)/2} e^{-u}\,du. \end{aligned}$$

The integral is nothing but the value of gamma function at $k/2, = \Gamma(k/2)$. Thus,

$$E[e^{tX}] = (1-2t)^{(-k/2)}$$

APPENDIX B

Chernoff Bound

Markov's Inequality

If X is a non negative random variable and $a > 0$, then

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

Proof. For any event E, let I_E be the indicator random variable of E, i.e. $I_E = 1$ when E occurs and $I_E = 0$ when it doesn't.

$$\mathbb{E}(I_E) = 1 \cdot \mathbb{P}(E) + 0 \cdot \mathbb{P}\left(E^C\right) = \mathbb{P}(E) \tag{B.1}$$

For any given $a > 0$, let E be the event such that $X \geq a$, therefore, $I_{(X \geq a)} = 1$ if $X \geq a$ and $I_{(X \geq a)} = 0$ if $X < a$. Clearly,

$$\begin{aligned}
aI_{(X\geq a)} &\leq X \\
\mathbb{E}\left(aI_{(X\geq a)}\right) &\leq \mathbb{E}(X) && \text{since } \mathbb{E} \text{ is a monotonically increasing function} \\
a\mathbb{P}(X \geq a) &\leq \mathbb{E}(X) && \text{linearity of expectations and Equation } \textbf{??}
\end{aligned}$$

□

Chernoff Bound

Applying Markov's inequality to e^{tX}, For every $t > 0$,

$$\mathbb{P}(X \geq a) = \mathbb{P}\left(e^{t \cdot X} \geq e^{t \cdot a}\right) \leq \frac{\mathbb{E}\left(e^{t \cdot X}\right)}{e^{t \cdot a}}$$

REFERENCES

1. A. J. Jerri, "The Shannon sampling theorem - Its various extensions and applications: A tutorial review," Proceedings of the IEEE, 65(11): 1565-1596, Nov. 1977.

2. Y. Eldar and G. Kutyniok, "Compressed Sensing: Theory and Applications", Cambridge University Press, 2012.

3. E. J. Candès, "Compressive sampling", Proc. Int. Congr. Math., vol. 3, pp. 1433-1452, 2006.

4. Dai W, Sheikh M, Milenkovic O, Baraniuk R. Compressive sensing DNA microarrays. EURASIP J Bioinform Syst Biol. 2009.

5. E. J. Candès, and T. Tao, "Decoding by linear programming", IEEE Transactions on Information Theory, 51(12): 4203-4215, 2005.

6. D. L. Donoho, "Compressed sensing", IEEE Transactions on Information Theory, 52(4): 1289-1306, 2006.

7. K. Bryan and T. Leise, "Making Do with Less: An Introduction to Compressed Sensing", SIAM J. Appl. Math., 55(3): 547-566, 2013.

8. J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems", Proceedings of IEEE, Special Issue on Applications of Compressive Sensing Sparse Representation, 98(6): 948-958, 2010.

9. R. Baraniuk, "Compressive Sensing", IEEE Signal Processing Magazine, 24(4):118 - 121, July 2007.

10. D.S. Taubman and M.W. Marcellin, JPEG 2000: Image Compression Fundamentals, Standards and Practice. Norwell, MA: Kluwer, 2001.

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

12. M. Davenport. Random observations on random observations: Sparse signal acquisition and processing. PhD thesis, Rice University, Aug. 2010.

13. E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements" Communications on Pure and Applied Mathematics, 59(8), 2006.

14. E. Candès and J. Romberg, "L1-Magic: Recovery of Sparse Signals via Convex Programming", http://users.ece.gatech.edu/justin/l1magic/downloads/l1magic.pdf, 2005. (Date of access - 29/05/2016).

15. M. F. Duarte and R. G. Baraniuk, "Kronecker compressive sensing", IEEE Trans. Image Processing, 21(2): 494-504, 2012

16. E. Hale, W. Yin, and Y. Zhang, "Fixed-point continuation for l1-minimization: methodology and convergence", SIAM J. Optim., 19(3): 1107-1130, 2008.

17. M. A. T. Figueiredo, R. D. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems", IEEE J. Sel. Top. Signal Proces., 1(4): 586-598, 2007.

18. M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly and R. G. Baraniuk, "Single pixel imaging via compressive sampling", IEEE Signal Process. Mag., 25(2): 83-91, 2008.

19. https://sites.google.com/site/igorcarron2/compressedsensinghardware