Contents

List of Figures

Abstract

In the present world, Nyquist-Shannon criterion poses a heavy burden on the hardware devices (ADCs et al.) and adds to the computational woes. In a normal photograph, all the samples are processed only to throw away 90% of them before you even see the picture. Compressive sensing (CS) seeks to alleviate this problem by sensing only the required samples. It does this by taking into account the inherent sparsity in natural signals. An often ignored truth is that almost upto 90% of signals in real world could be approximated as sparse in some basis. Compressive Signal Processing (CSP) framework deals with problems where there is no need to reconstruct the sparsely sampled signals but only an inference needs to be drawn . We extend this concept into the field of cognitive radio where we detect the channel occupancy. It is observed that, accurate results could be obtained by merely using detection without having to reconstruct the signal thus relaxing the computation constraints. The advantage of this method is mainly due to the fact that most of the complexity in CS lies in the reconstruction or the decoding stage. The model we have developed stands low in complexity while at the same time providing high accuracy. Similarly, it is observed that many communication channels encountered in practice tend to exhibit a sparse structure. Consequently, CS finds its application in sparse channel estimation. CSP has also been used to estimate the frequency response of AWGN channel. We implement orthogonal matching pursuit and MMSE estimator in estimating the channel impulse response. The model is observed to perform almost as well as a traditional channel estimation technique while using lesser samples.

Chapter 1

Introduction

1.1 Shannon Nyquist theorem

The Shannon Nyquist theorem states that every band-limited signal can be recovered from its discretization if its sampling rate is atleast two times its maximum frequency. Therefore according to this theorem there is no way to recover the original baseband signal back if its sampled at a rate less then the Nyquist rate.

$fs = 2 * W$

where fs is the sampling frequency and W is the maximum frequency . We know that the spectrum of a discrete signal is periodic in frequency domain having a period of fs. As seen from the above figure if the sampling frequency is less then W then there is aliasing in the spectrum causing loss of information.

1.2 General Transmission Scheme

Figure above shows the general transmission scheme used in communication.Sampling at nyquist rate we get N samples. Using traditional compression techniques we retain only K samples and discard N-K , these K samples are then transmitted and at the receiver from K samples the original signal is recovered. In the following transmission scheme the main problem that can be observed is that if only K samples have to transmitted whats the need to acquire initially N samples (N>>K) and then discard N-K ; can in some way the k samples be directly obtained and transmitted . The second issue with the above transmission scheme is that decompression of a signal i.e reconstructing the original N length signal from the K samples acquired is a very complex process and in most cases complete reconstruction is not needed and can be replaced with a simple inference block which can be used to detect the required information. The third issue with the above scheme is that in many situations such as channel estimation or cognitive radio, to name some the bandwidth occupied is of the range of mega hertz therefore the ADCs have to work at two times these rates which is practically not feasible. These factors provide us the motivation to use compressive signal processing with

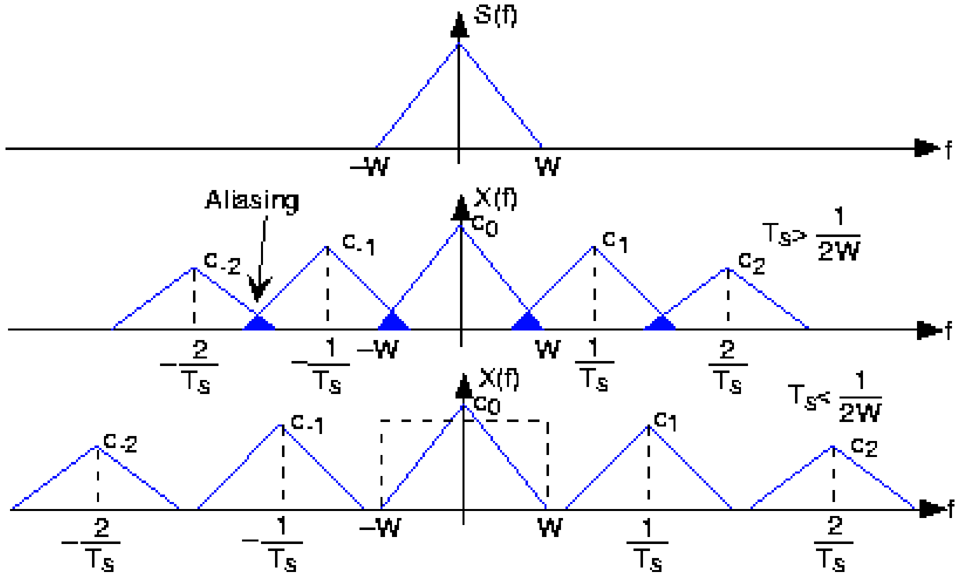


Figure 1.1.1: Nyquist theorem

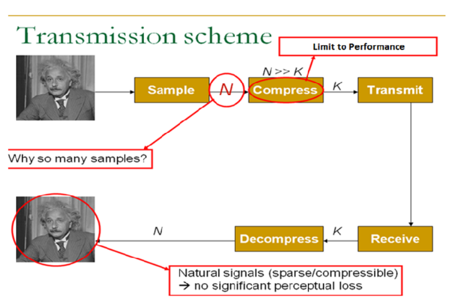


Figure 1.2.1: General Transmission Scheme

its application in detection and estimation.

1.3 Introduction to Compressed Sensing

We are in the midst of a digital revolution that is driving the development and deployment of new kinds of sensing systems with ever-increasing fidelity and resolution. The theoretical foundation of this revolution is the pioneering work of Kotelnikov, Nyquist, Shannon, and Whittaker on sampling continuous-time band-limited signals.Their results demonstrate that signals introduction to Compressed Sensing images, videos, and other data can be exactly recovered from a set of uniformly spaced samples taken at the so-called Nyquist rate of twice the highest frequency present in the signal of interest. Capitalizing on this discovery, much of signal processing has moved from the analog to the digital domain and ridden the wave of Moore's law. Digitization has enabled the creation of sensing and processing systems that are more robust, flexible, cheaper and, consequently, more widely used than their analog counterparts.

As a result of this success, the amount of data generated by sensing systems has grown from a trickle to a torrent. Unfortunately, in many important and emerging applications, the resulting Nyquist rate is so high that we end up with far too many samples. Alternatively, it may simply be too costly, or even physically impossible, to build devices capable of acquiring samples at the necessary rate. Thus, despite extraordinary advances in computational power, 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.

To address the logistical and computational challenges involved in dealing with such high-dimensional data, we often depend on compression, which aims at finding the most concise representation of a signal that is able to achieve a target level of acceptable distortion. One of the most popular techniques for signal compression is known as transform coding, and typically relies on finding a basis or frame that provides sparse or compressible representations for signals in a class of interest.By a sparse representation, we mean that for a signal of length n, we can represent it with k n nonzero coefficients; by a compressible representation, we mean that the signal is well-approximated by a signal with only k nonzero coefficients. Both sparse and compressible signals can be represented with high fidelity by preserving only the values and locations of the largest coefficients of the signal. This process is called sparse approximation, and forms the foundation of transform coding schemes that exploit signal sparsity and compressibility, including the JPEG, JPEG2000, MPEG, and MP3 standards.

Leveraging the concept of transform coding, compressive sensing (CS) has emerged as a new framework for signal acquisition and sensor design. CS enables a potentially large reduction in the sampling and computation costs for sensing signals that have a sparse or compressible representation. While the Nyquist-

Shannon sampling theorem states that a certain minimum number of samples is required in order to perfectly capture an arbitrary bandlimited signal, when the signal is sparse in a known basis we can vastly reduce the number of measurements that need to be stored. Consequently, when sensing sparse signals we might be able to do better than suggested by classical results. This is the fundamental idea behind CS: rather than first sampling at a high rate and then compressing the sampled data, we would like to find ways to directly sense the data in a compressed form i.e. at a lower sampling rate. The field of CS grew out of the work of Candes, Romberg, and Tao and of Donoho, who showed that a finite-dimensional signal having a sparse or compressible representation can be recovered from a small set of linear, nonadaptive measurements.The design of these measurement schemes and their extensions to practical data models and acquisition systems are central challenges in the field of CS.

While this idea has only recently gained significant attraction in the signal processing community, there have been hints in this direction dating back as far as the eighteenth century. In 1795, Prony proposed an algorithm for the estimation of the parameters associated with a small number of complex exponentials sampled in the presence of noise. The next theoretical leap came in the early 1900's, when Caratheodory showed that a positive linear combination of any k sinusoids is uniquely determined by its value at t = 0 and at any other 2k points in time.This represents far fewer samples than the number of Nyquist rate samples when k is small and the range of possible frequencies is large. In the 1990's, this work was generalized by George, Gorodnitsky, and Rao, who studied sparsity in bio magnetic imaging and other contexts. Simultaneously, Bresler, Feng, and Venkataramani proposed a sampling scheme for acquiring certain classes of signals consisting of k components with nonzero bandwidth (as opposed to pure sinusoids) under restrictions on the possible spectral supports, although exact recovery was not guaranteed in general. In the early 2000's Blu, Marziliano, and Vetterli developed sampling methods for certain classes of parametric signals that are governed by only k parameters, showing that these signals can be sampled and recovered from just 2k samples.

A related problem focuses on recovery of a signal from partial observation of its Fourier transform. Beurling proposed a method for extrapolating these observations to determine the entire Fourier transform .One can show that if the signal consists of a finite number of impulses, then Beurling's approach will correctly recover the entire Fourier transform (of this non-bandlimited signal) from any sufficiently large piece of its Fourier transform. His approach to find the signal with smallest l^1 norm among all signals agreeing with the acquired Fourier measurements bears a remarkable resemblance to some of the algorithms used in CS.

More recently, Candes, Romberg, Tao and Donoho showed that a signal having a sparse representation can be recovered exactly from a small set of linear, nonadaptive measurements. This result suggests that it may be possible to sense sparse signals by taking far fewer measurements, hence the name compressed sensing. Note, however, that CS dears from classical sampling in three important respects. First, sampling theory typically considers infinite length, continuous-time signals. In contrast, CS is a mathematical theory focused on measuring nite-dimensional vectors in R^n. Second, rather than sampling the signal at specific points in time, CS systems typically acquire measurements in the form of inner products between the signal and more general test functions. This is in fact in the spirit of modern sampling methods which similarly acquire.signals by more general linear measurements . We will see throughout this report that randomness often plays a key role in the design of these test functions. Thirdly, the two frameworks differ in the manner in which they deal with signal recovery, i.e., the problem of recovering the original signal from the compressive measurements. In the Nyquist-Shannon framework, signal recovery is achieved through sinc interpolation a linear process that requires little computation and has a simple interpretation. In CS, however, signal recovery is typically achieved using highly nonlinear methods.

CS has already had notable impact on several applications. One example is medical imaging where it has enabled speedups by a factor of seven in pediatric MRI while preserving diagnostic quality. Moreover, the broad applicability of this framework has inspired research that extends the CS framework by proposing practical implementations for numerous applications, including sub-Nyquist sampling systems, compressive imaging architectures, and compressive sensor networks .

1.4 Introduction to Compressive Signal Processing (CSP)

Compressive Signal Processing is applying the CS concepts but does not include reconstruction of the signal at the receiver and focuses on the fact that in many cases we replace the reconstruction stage by a detector or estimator block. CSP thus combines CS concepts with basic detection and estimation theory therefore avoiding reconstruction completely.

We have applied CS for the following scenarios:

1. Detect the presence of primary users in cognitive radio, therefore utilizing the spectrum holes in transmitting secondary users in the absence of primary users.
2. Estimating the frequency response of a sparse channel without signal reconstruction
3. Estimation of Channel response of a system using algorithms such as OMP

1.5 Compressive Sensing Analogy

Given :

1. A bag with 100 marbles, nominal mass 10 grams each but one marble may be bad.
2. An electronic balance.

How can we find the bad marble with the fewest weightings?

1.5.0.1 Divide and Conquer

1. Put half the marbles on the scale.
2. Decide if the bad marble is in that subset. Repeat with the subset that (might) contain the bad marble.
3. We find the bad marble in about $log_2(100) \approx 7$ weighings, a big improvement over the expected 50.5 weighings for a sequential approach.
4. If there are k $\geq$2 bad marbles (of unknown weights) the problem gets a lot harder.
5. The sequential approach takes an expected $\frac{101K}{K+1}$ weighings if there are k bad marbles.
6. To do better we need to weigh subsets but how should be choose them?
7. Answer : Randomly!

1.5.0.2 Random Subsets

1. Number the marbles 1 to 100.
2. Choose a random subset of the marbles, each marble included with some probability (say 1/2).
3. Weigh the subset and record.
4. Repeat the last step a total of n times, say with n = 25.
5. Claim: If there are only a few bad marbles (say k = 3) and we weigh n = 25 subsets, we almost certainly have enough information to identify the bad marbles. K
6. Let X_i denote the true deviation of the ith marble from nominal (10 grams). We expect X_i= 0 for most i .
7. Suppose the marbles in the first random subset have indices $i = i_1, i_2, i_{3.}...., i_m (m \approx 50)$ and this subset has mass a_1.
8. Then $X_{i1} + X_{i2} + X_{i3} + ... + X_{im} = a_1 - 10m$
9. A similar equation holds for each of the other random subsets that we choose K
10. Let x_i be our estimate of the true value X_i. We end up with n = 25 linear equations and N = 100 unknowns (the mass deviations), of the form ϕx = b; ϕ is the sensing matrix, matrix with $\phi_{ij} = 1$ if the jth marble was included in the ith subset, and b_j is the mass of the j^{th} subset minus 10 times the number of marbles in the subset K.

1.5.0.3 Random Subset Example

1. For example, with N = 10 marbles and n = 3 weighings we might have $\phi = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$

2. If the first weighing involves marbles 1,4,5,7 the second involves marbles 3,4,5,6,8 the third involves marbles 2,3,4,5,9,10

3. In the marble problem with N = 100, suppose that X_{13} = -0.3; X_{37} = 0.44; and X_{71} = -0.33, and all other $X_i = 0$ (but we don't know this).

4. To estimate the marble masses, we choose random 25 subsets, each of size about 50, and weigh each.

5. We obtain a system x = ϕb: The matrix is 25 rows by 100 columns|there will be at least 75 free variables.

6. Solving for the x_i in any meaningful way looks hopeless.

1.5.0.4 Minimum l^2 norm solution

A traditional approach to "solving" a consistent under determined system of equations Mx = b is to choose that (unique) vectorx^* with the properties that

1. $Mx^* = b$ (x^*actually satisfies the equations)

2. x^*has minimum l^2 norm, that is, if $x^{**} \neq x^*$; such that x^{**} satisfies $Mx^{**} = b$ then $||x^*||_2 < ||x^{**}||_2$ where $||x_2|| = (\sum_i |x_i|^2)^{\frac{1}{2}}$ is the l^2 norm

3. Computing the l^2 minimum solution is a standard calculus/matrix algebra problem. In the present case we find:

- As seen from the above figure l^2 norm provides poor result
- Ideally since x is sparse l^0 norm would give the perfect solution but this is NP hard
- Therefore we compromise and find the l^1 norm solution that is find an x that minimizes $||x||_1 = \sum_{i=1}^{N} |\mathrm{x}_i|$
- This is an easy linear programming problem

1. As seen from the figure l^1 provides an accurate solution.

2. The more weighings we do (for any fixed number of defective marbles) the better chance of success l^1 minimization has.

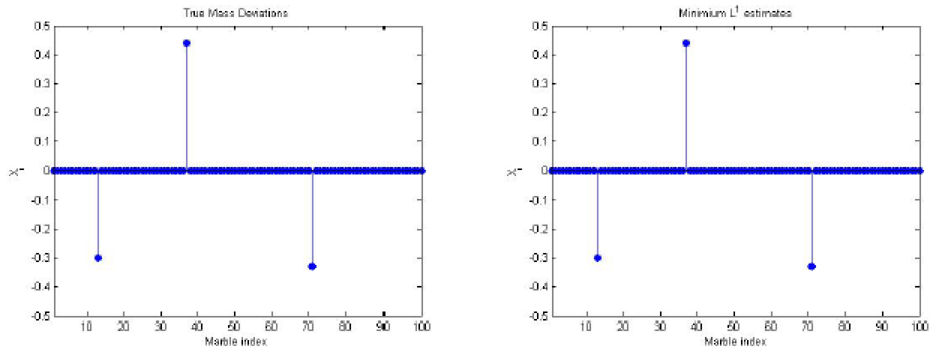


Figure 1.5.1: True Estimate & L1 Estimate

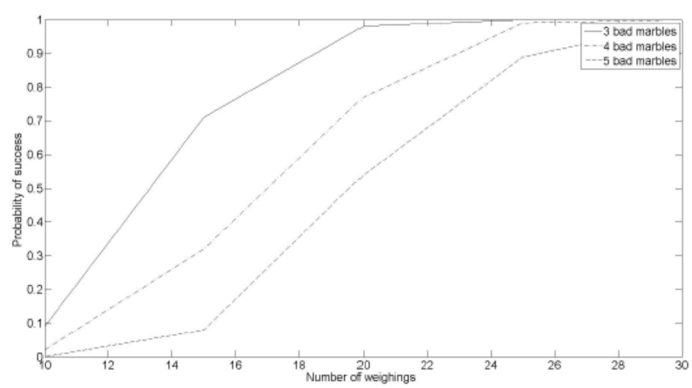


Figure 1.5.2: Marble Weighings

Chapter 2

Compressive Sensing Concepts

2.1 Sparsity

Many natural signals have a concise representation when expressed in convenient basis

$f \in R^N$

Let $\psi = [\varphi_1, \varphi_2, ..., \psi_N]$ be the orthonormal basis

$f(t) = \sum x_i \psi_i(t)$

x coefficients of f

In matrix notation $f = \psi x$

If f is sparse we can discard all but s largest coefficients,therefore

$f_s = \psi x_s$(s largest coefficients of f)

We have$||f - f_s||_{l2} = ||x - x_s||_{l2}$;if x is sparse no information is lost on discarding N-S coefficients.

Then signal is S sparse

Definition of Sparsity : A signal

$$x = (x_i)_{i=1}^n \tag{2.1.1}$$

$\in R^N$is called k sparse if

$$||x_0||_0 = \#\{i\, x_i :\neq 0\} \leq k \tag{2.1.2}$$

$\sum_k$represents all possible k sparse vectors in a vector of length N

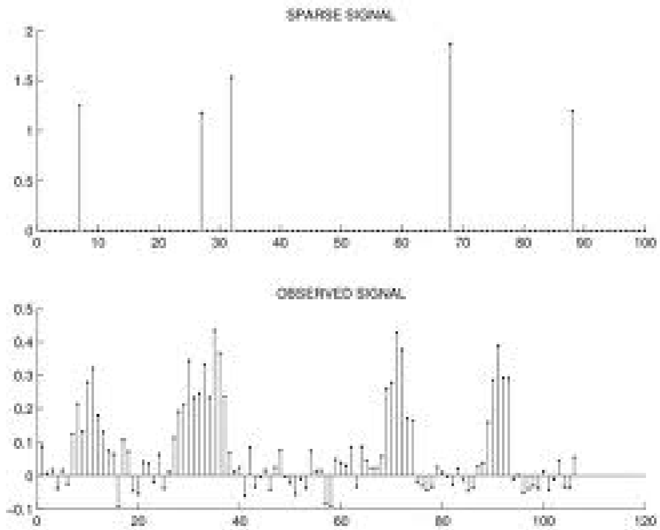


Figure 2.1.1: Sparse Representation of Signal

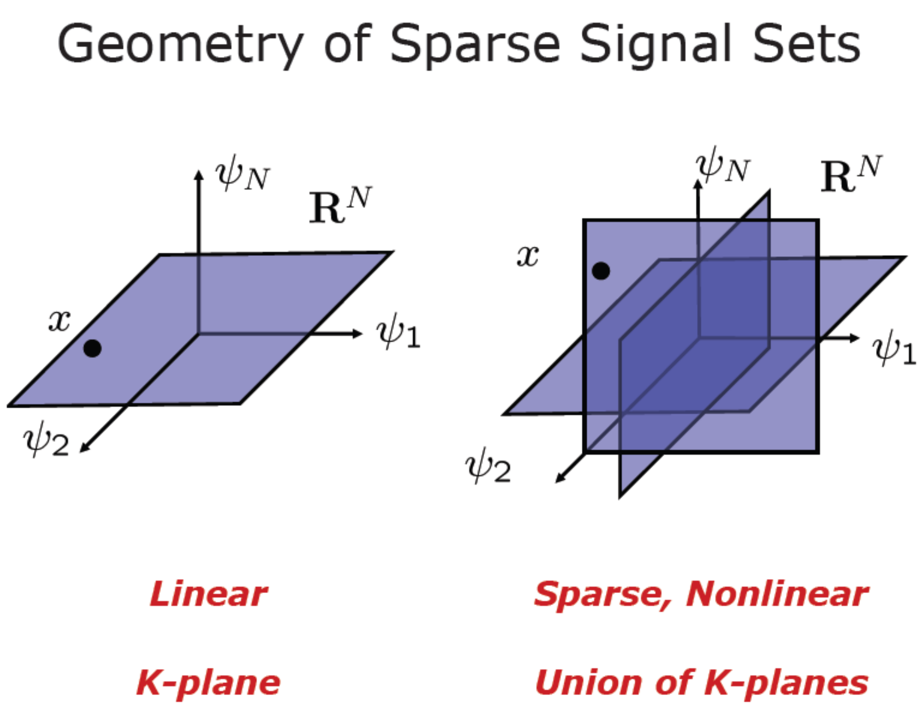


Figure 2.1.2: Geometry of Sparse Signals

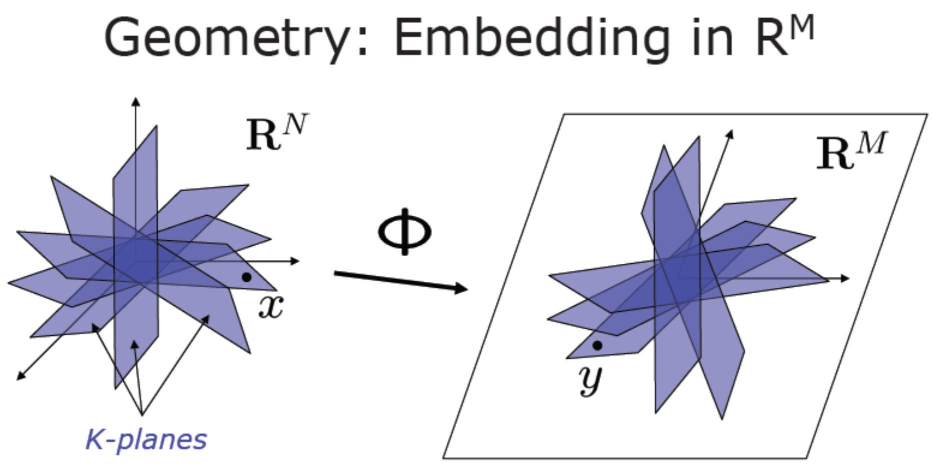


Figure 2.1.3: Geometry of Projection

2.2 Under-determined System

An under determined system consists of few equations and lots of unknowns; characterized by a short and fat matrix.Traditional methods of solving a consistent under determined system $Ax = b$ are to choose an unique vector $x^*$with the properties that :

1.$Ax^* = b$ (x*satisfies the equation)

2.$x*$has minimum l^2norm , that is , x$^{**} \neq x^*$,$Ax^{**} = b$ then $||x^*||_2 < ||x^{**}||_2$

where $||x_2|| = (\sum_i |x_i|^2)^{\frac{1}{2}}$is the l^{2}norm

3.We know in CS that the vector x is sparse due to this factl^2norm gives poor results.Therefore we should

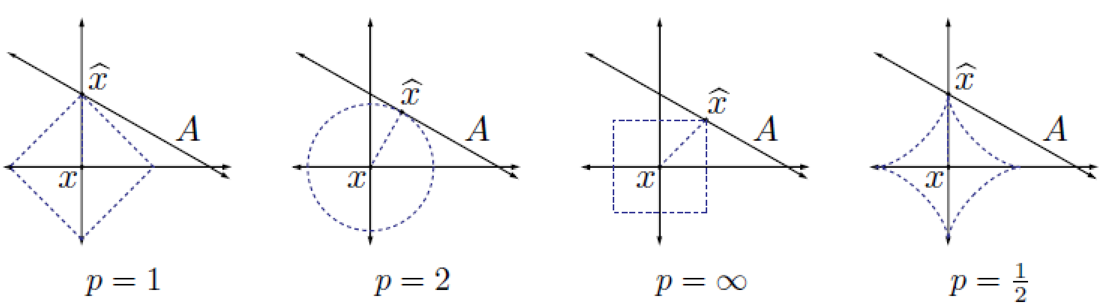


Figure 2.2.1: Solution to an undetermined system using different norms

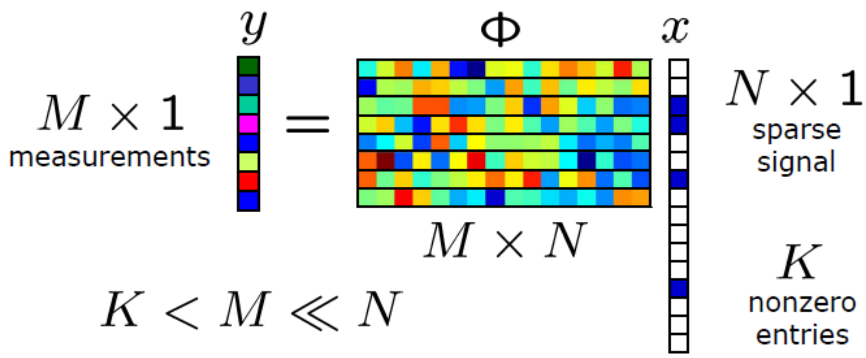


Figure 2.2.2: Under determined System

find a vector x that satisfies $Ax = b$ and x has a few non-zero components as possible, that is, x minimizes the $||x_0||$=$|\{\mathrm{x}_i : x_i \neq 0\}|$ this is known as minimizing the $\mathrm{l}_0$norm.This would provide the ideal solution to our problem but this is NP hard and cannot be implemented

4.Therefore we go for the $\mathrm{l}_1$norm that is find an x that minimizes $||x||_1 = \sum_{i=1}^{N}|\mathrm{x}_i|$ fixed algorithms exist for this (eg: orthogonal matching pursuit-OMP) which can be implemented to determine the solution.

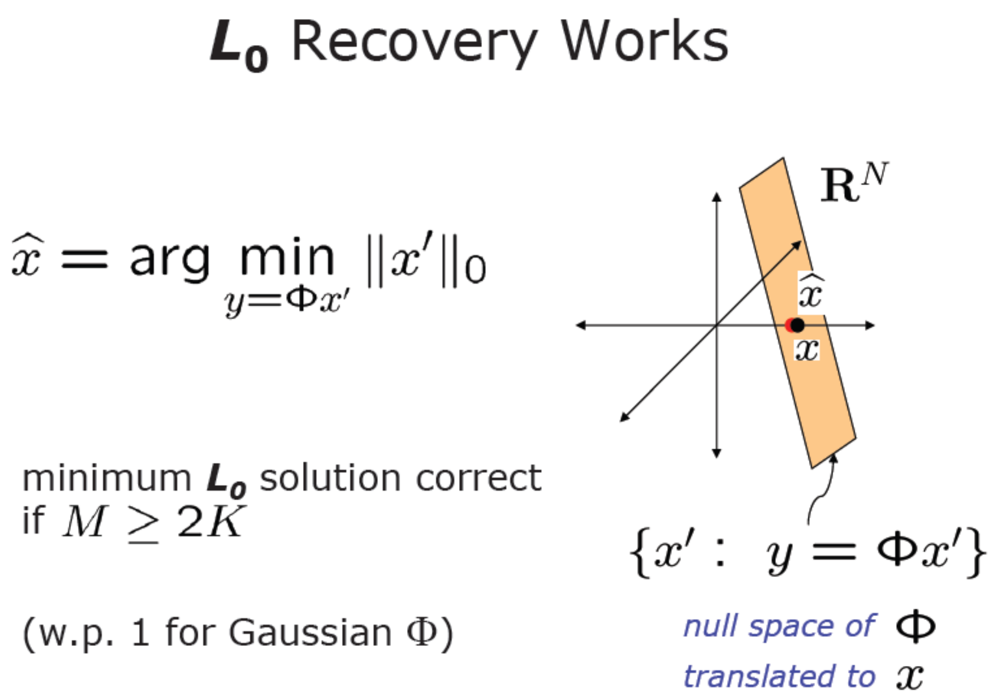


Figure 2.2.3: Geometry of l^0norm

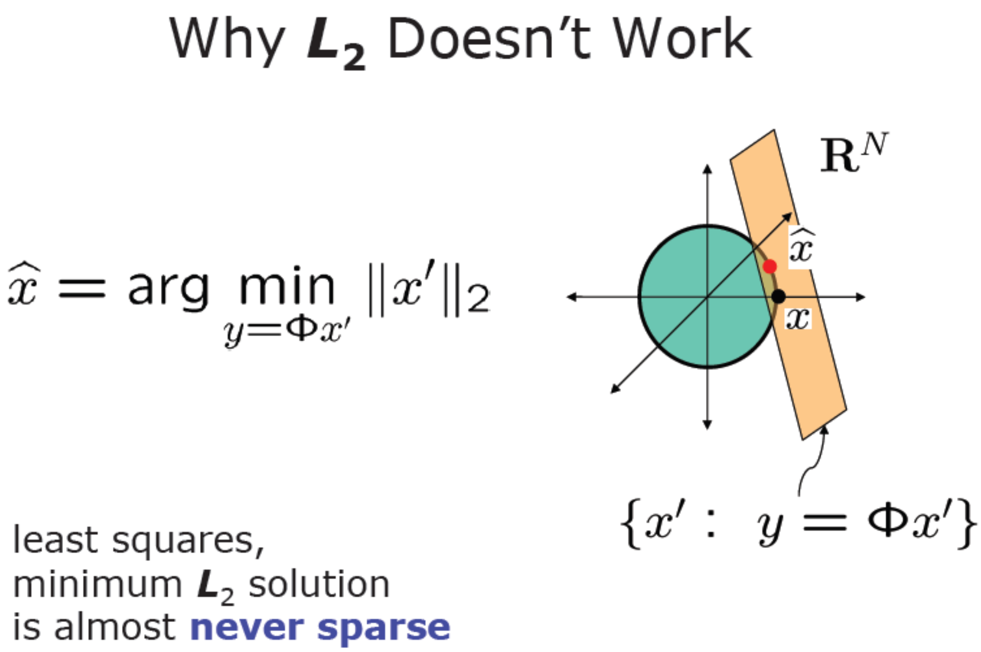


Figure 2.2.4: Geometry of l^2norm

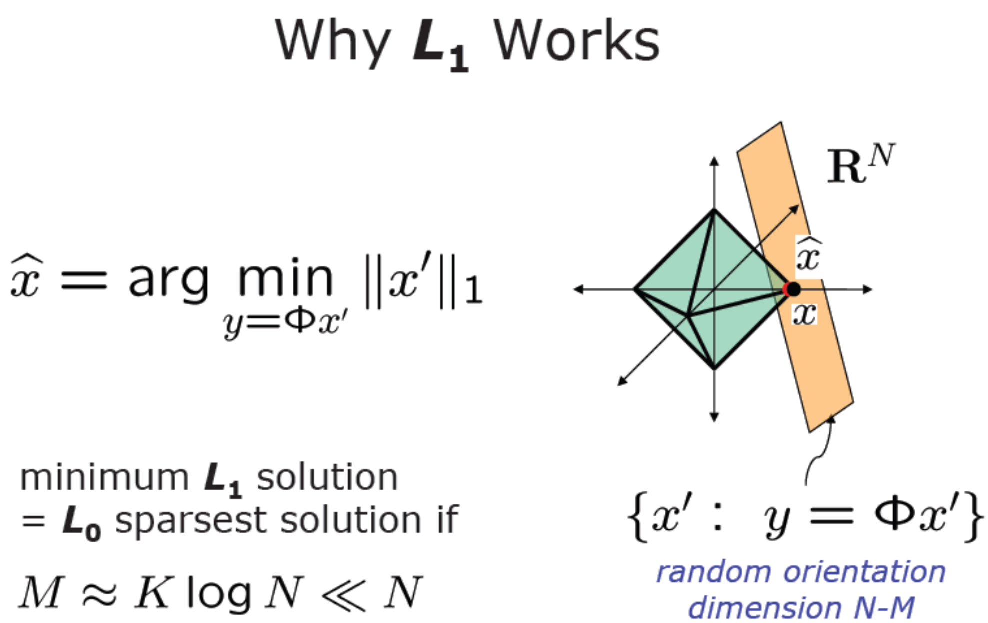


Figure 2.2.5: l^1geometry

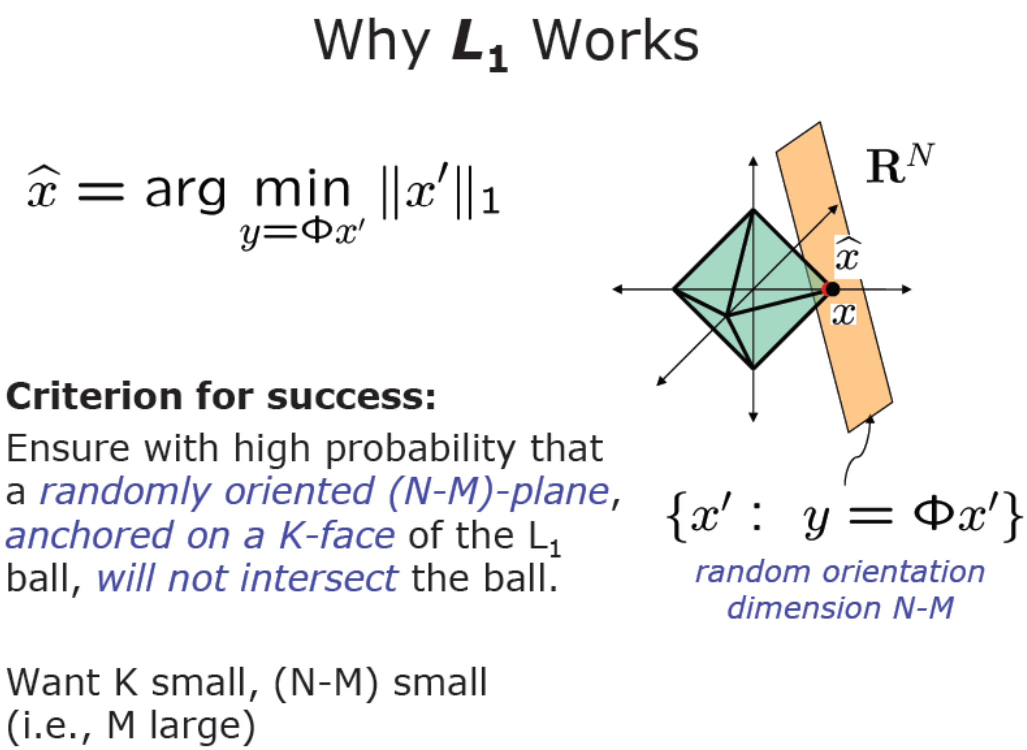


Figure 2.2.6: l^1Geometry

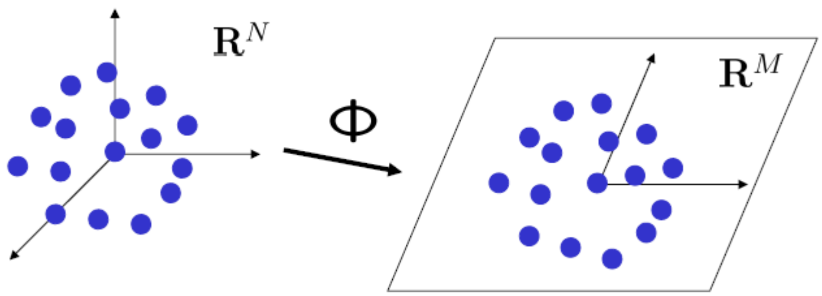


Figure 2.3.1: Transformation from N to M dimension

2.3 Sensing Matrix and its Properties

Consider the system

$$\phi x = b \tag{2.3.1}$$

$\phi$is known as the sensing matrix.An under determined system can be accurately solved using $l_1$minimization only if ϕsatisfies the Restricted Isometry Property (RIP) property.

2.3.0.5 Development of RIP

Let ϕbe a n by N matrix and suppose $\phi x = b$ has a k sparse solution (2.3) (x=X is the only k sparse solution)

Now consider x=$\widetilde{X}$is another solution to the system (2.3) therefore

$$\phi\widetilde{X} = b \tag{2.3.2}$$

(assume x=$\widetilde{X}$is another k sparse solution)

From 2.3 and 2.4 we see that $\phi(X - \widetilde{X}) = 0$; $X - \widetilde{X}$=W;

Therefore $\phi W = 0$; which means W lies in the null space of ϕ

But we know that X,$\widetilde{X}$are k sparse hence W is 2k sparse therefore W$\neq$ 0lies $N(\phi)$

Therefore if k sparse solutions are not unique $N(\phi)$ contains 2k sparse vectors

Hence if construct ϕsuch that $N(\phi)$ does not contain any 2k sparse vectors then we can be assured that any k sparse solution we obtain will be unique.

$\phi x \neq 0$ for all 2k sparse vectors can be alternatively written as

$$c_1||x||_2 \leq ||\phi X||_2 \equiv c_1 \leq ||\phi u||_2 \tag{2.3.3}$$

for all 2k sparse unit vectors

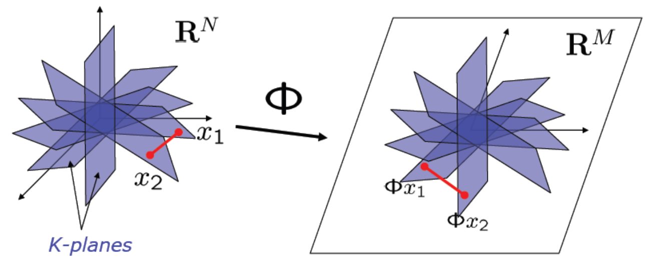


Figure 2.3.2: Projection

$$c_2 \geq ||\phi u| \tag{2.3.4}$$

$_2$ comes automatically

Combining (2.5) and (2.6) we get

$$c_1 \leq ||\phi u||_2 \leq c_2 \tag{2.3.5}$$

for all 2k sparse unit vectors

(2.7) Represents one form of the RIP property and if RIP for 2k sparse vectors holds good then all k sparse vectors can be uniquely decoded.

Multiplying (2.7) by a constant does not change the solvability therefore $2.7 * \frac{2}{c1+c2}$ RIP becomes

$$(1-\delta) \leq ||\phi u||_2 \leq (1+\delta) \tag{2.3.6}$$

; where $\delta = \frac{c2-c1}{c2+c1}$ and $0 \leq \delta \leq 1$.

Therefore RIP holds good for all 2k sparse vectors.

2.3.0.6 Achieving RIP

In the paper Signal Processing with Compressive Measurements by Mark Davenport, Petros Boufounos, Michael Wakin, Richard Baraniuk it is shown that if the entries of the sensing matrix ϕ(M,N) are drawn from an iid (independent and identically distributed) Gaussian random variable $N(0, \frac{1}{M})$ that is having a mean 0 and variance $\frac{1}{M}$ then ϕ satisfies the RIP property for all 2k sparse vectors.

Chapter 3

Detection

3.1 General Problem

In many situations at the receiver complete reconstruction of signal is not required and instead of decompression we can simply implement a detection block.

Let us examine the the detection problem to distinguish between two hypothesis :

$$H0 \,:\, y = \phi n \tag{3.1.1}$$

$$H1 \,: y \; = \phi(s+n) \tag{3.1.2}$$

where $n \sim N(0, \sigma^2 I_N)$ is iid Gaussian noise

$\phi$is known fixed measurement matrix

$s \in \mathbb{R}^N$ is known signal to be detected

Our task is to decide between $H0$ and$H1$ accordingly

3.2 Development of the Compressive Detector

P$_F$= Pr($H_1$is chosen when H_0true)

P$_D$ =Pr($H_1$is choose when H_1true) denote the false alarm and detection rate respectively.

The Neymann-Pearson(NP) detector is the decision rule that maximizes P$_D$subject to the constraint that P$_F \leq \alpha . H_0 and\ H_1$hypothesis have the following density functions

$$f_0(y) = \frac{\exp(-\frac{1}{2} y^T (\sigma^2 \phi \phi^T)^{-1} y)}{|\sigma^2 \phi \phi^T|^{\frac{1}{2}} (2\pi)^{\frac{M}{2}}} \tag{3.2.1}$$

$$f_1(y) = \frac{\exp(-\frac{1}{2}(y-\phi s)^T(\sigma^2\phi\phi^T)^{-1}(y-\phi s))}{|\sigma^2\phi\phi^T|^{\frac{1}{2}}(2\pi)} \tag{3.2.2}$$

The NP optimal decision rule is to compare the ratio $\frac{f_1(y)}{f_0(y)}$ to a threshold η which is the likelihood ratio test

$$\Lambda(y) == \frac{f_1(y)}{f_0(y)} \frac{\geq^{H1}}{\leq_{H0}} \eta \tag{3.2.3}$$

η is chosen such that

$$P_F = \int\limits_{\Lambda(y)>\eta}^{\infty} f_0(y)dy = \alpha \tag{3.2.4}$$

Simplifying (3.5) and taking logarithm we obtain an equivalent test as

$$y^T(\phi\phi^T)^{-1}\phi s \frac{\geq^{H1}}{\leq_{H0}} \sigma^2\log(\eta) + \frac{1}{2}s^T\phi^T(\phi\phi^T)^{-1}\phi s := \gamma \tag{3.2.5}$$

The compressive detector is defined as

$$t := y^T(\phi\phi^T)^{-1}\phi s \tag{3.2.6}$$

It can be Shown that t is the sufficient statistic for our detection problem.t contains the information relevant to distinguish between the 2 hypothesis.

we now define $P_{\phi^T} = \phi^T(\phi\phi^T)^{-1}\phi$as the orthogonal projection onto $R(\phi^T)$ i.e the row space of ϕ, since $P_{\phi^T} = P_{\phi^T}^T$ therefore

$$s^T\phi^T(\phi\phi^T)^{-1}\phi s = ||P_{\phi^T}s||^2 \tag{3.2.7}$$

Using (3.9) we can write

$t \sim N(0, \sigma^2||P_{\phi^T}s||^2)$ under H_0

$t \sim N(||P_{\phi^T}s||^2, \sigma^2||P_{\phi^T}s||^2)$ under H_1

$$P_F = P(t > \gamma|H_0) = Q(\frac{\gamma}{\sigma||P_{\phi^T}s||_2}) \tag{3.2.8}$$

$$P_D = P(t > \gamma|H_1) = Q(\frac{\gamma - ||P_{\phi^T}s||^2)}{\sigma||P_{\phi^T}s||_2}) \tag{3.2.9}$$

where $Q(z) = \int_z^\infty exp(-u^2/2)du$

To determine the threshold, we set $P_F = \alpha$ and thus

$$\gamma = \sigma ||P_{\phi^T} s||_2 Q^{-1}(\alpha) \tag{3.2.10}$$

resulting in

$$P_D(\alpha) = Q(Q^{-1}(\alpha) - ||P_{\phi^T} s||_2 / \sigma) \tag{3.2.11}$$

the $SNR := \frac{||s||_2}{\sigma^2}$

$\phi(M, N)$ is the matrix where M is the number of compressive measurements taken and N is the total length of vector s ; the probability of detection on terms of M,N and SNR can be written as

$$P_D(\alpha) \approx Q(Q^{-1}(\alpha) - \sqrt{M/N}\sqrt{SNR}) \tag{3.2.12}$$

3.3 Compressive Detector Results

3.3.0.7 Effect of M on Pd

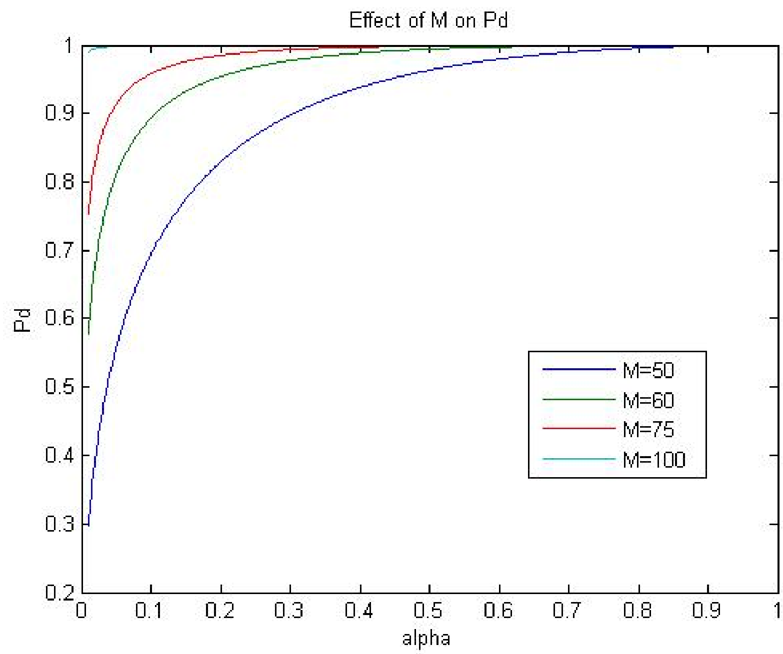


Figure 3.3.1: Pd vs α

The settings for the above simulation were as follows:

Number of Samples 'N' = 1000
Sparsity 'k' = 25

1. M- Number of Measurements
2. Pd- Probability of Detection
3. α- Probability of False Alarm
4. N=1000 (signal length) ; k=25 (sparsity)

3.3.0.8 Effect of M on Pd at various SNR

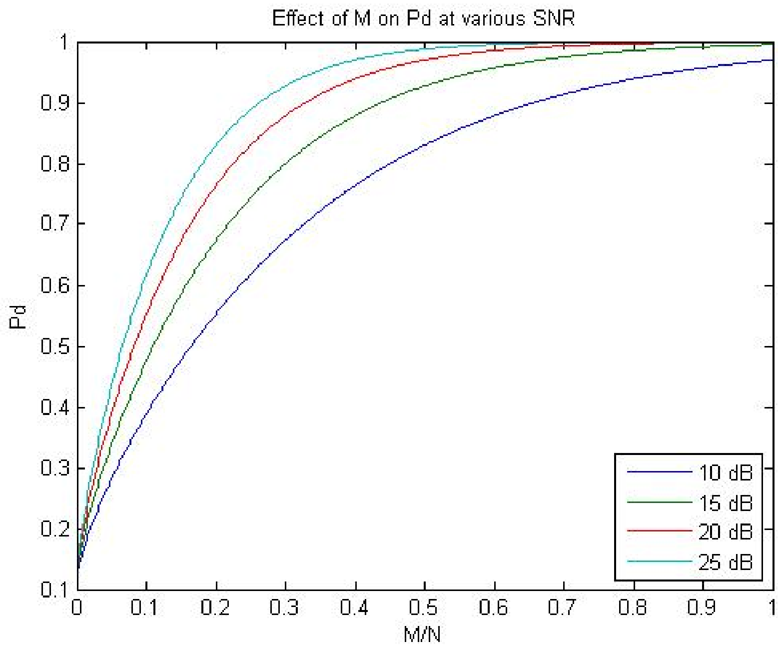


Figure 3.3.2: Pd vs $\frac{M}{N}$
The settings for the above simulation were as follows:
Number of Samples 'N' = 500
Sparsity of signal 'k' = 25
$\alpha = 0.1$

1. M- Number of Measurements
2. Pd-Probability of Detection
3. α=0.1
4. N=500 (signal length) ; k=25 (sparsity)

Chapter 4

Detection in Cognitive Radio

4.1 Introduction

A Cognitive Radio (CR) is a software defined radio(SDR) that additionally senses its environment ,tracks changes and responds accordingly. It is basically a technique for efficient spectral usage.

There are two kinds of users with reference to cognitive radio:

- **Primary users:** Users who have legacy rights/ higher priority on the usage of a specific part of the spectrum.
- **Secondary users:** These are users who make use of the spectrum when the primary users are not using the spectrum or these users use the spectrum in such a way as to not cause any interference to the higher priority primary users.

In order to detect the presence of the primary user, spectrum sensing is performed. It is the process of identifying occupancy in all dimensions of the spectrum space.The idea of CR is based on the observation that at certain times, most of the licensed spectrum is not used by the licensed users Secondary (unlicensed) users detect the spectrum holes (unoccupied spectrum) and utilize the spectrum at the absence of the primary (licensed) users.

As shown above [10], making efficient use of the spectrum holes is the basic idea of cognitive radio.

4.2 Application of CS in CR

Current Cognitive Radios are limited in their operational bandwidth by existing hardware devices, much of the extensive theoretical work on spectrum sensing is impossible to realize in practice over a wide frequency band. To alleviate the sampling bottleneck, we use Compressive Sensing (CS), which allows for the acquisition of sparse signals at sub-Nyquist rates, in conjunction with Cognitive Radio(CR). We have sequentially

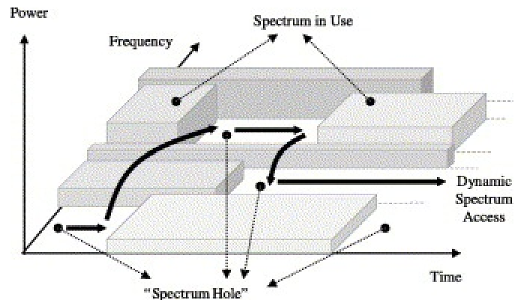


Figure 4.1.1: Spectrum hole

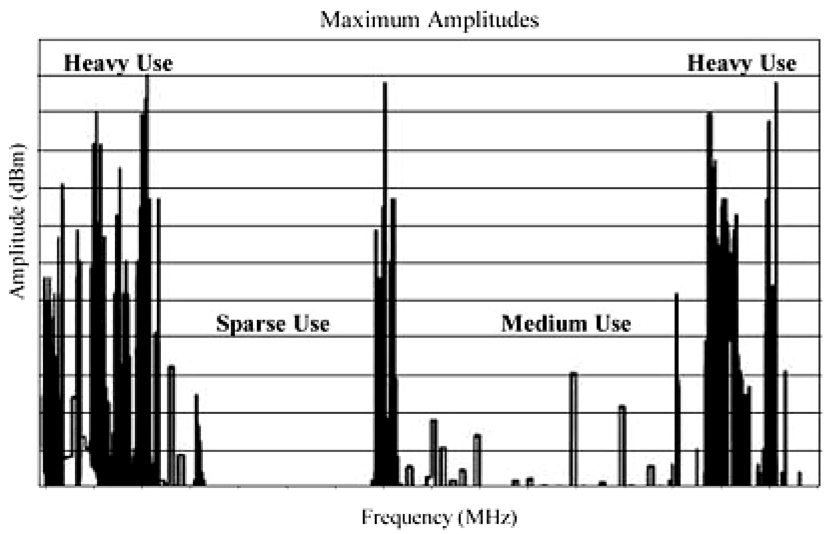


Figure 4.2.1: Sparsity of channels in Cognitive Radio

combined the two techniques: first acquiring compressed samples, then performing CR spectrum sensing on the compressive signal.

As we know, compressive sensing makes use of sparsity for recovering data. In the figure shown above [9], there are frequencies which are heavily used and also frequencies which are not used. Hence this could be considered as a sparse structure . It is this sparsity that is made use of while applying compressive sensing in cognitive radio.

While CS alleviates the bandwidth constraints imposed by front-end ADCs, the resulting increase in computation/complexity is non-trivial, especially in a power-constrained mobile CR. In addition, the computation time introduces significant delay into the spectrum sensing operation. Our key observation is that the CR does not have to reconstruct the entire signal because it is only interested in detecting the presence of Primary Users The fundamental task of a CR is not to analyze the entire signal, rather it is to estimate the presence of Primary Users, thus if we can directly estimate the Primary User signals from the compressive measurements, the reconstruction stage can be completely eliminated. Although, using CS in CR results in an increase in the computation time and complexity, we relax these by just detecting the signal instead of estimating. This results in a significant reduction in complexity.

4.3 Spectrum sensing techniques

The various sensing techniques are as follows:

1. Energy detector based sensing
2. Waveform based sensing
3. Cyclostationary based sensing
4. Matched filtering based sensing
5. Radio Identification based sensing

4.3.0.9 Energy detector based sensing

Energy detector based approach, also known as radiome-try or periodogram, is the most common way of spectrum sensing because of its low computational and implementation complexities . The signal is detected by comparing the output of the energy detector with a threshold which depends on the noise floor. Some of the challenges with energy detector based sensing include selection of the threshold for detecting primary users, inability to differentiate interference from primary users and noise, and poor performance under low signal-to-noise ratio (SNR) values.

Let us assume that the received signal has the following simple form:

$y(n) = s(n) + w(n)$

where $s(n)$ is the signal to be detected, $w(n)$ is the additive white Gaussian noise (AWGN) sample, and n is the sample index. Note that $s(n) = 0$ when there is no transmission by primary user. The decision metric for the energy detector can be written as

$M = \sum_{n=0}^{n=N} |y(n)|^2$

where N is the size of the observation vector. The decision on the occupancy of a band can be obtained by comparing the decision metric M against a fixed threshold λ . This is equivalent to distinguishing between the following two hypotheses:

$H0 : y(n) = w(n)$

$H1 : y(n) = s(n) + w(n)$.

The performance of the detection algorithm can be summarized with two probabilities: probability of detection P_D and probability of false alarm P_F . P_D is the probability of detecting a signal on the considered frequency when it truly is present. Thus, a large detection probability is desired. It can be formulated as

$P_D = Pr(M > \lambda|H1)$. P_F is the probability that the test incorrectly decides that the considered frequency is occupied when it actually is not, and it can be written as

$P_F = Pr(M > \lambda | H0)$. P_F should be kept as small as possible in order to prevent under utilization of transmission opportunities. The decision threshold λ can be selected for finding an optimum balance between P_D and P_F . However, this requires knowledge of noise and detected signal powers. The noise power can be estimated, but the signal power is difficult to estimate as it changes depending on ongoing transmission characteristics and the distance between the cognitive radio and primary user. In practice, the threshold is chosen to obtain a certain false alarm rate. Hence, knowledge of noise variance is sufficient for selection of a threshold.

4.3.0.10 Waveform based sensing :

Known patterns are usually utilized in wireless systems to assist synchronization or for other purposes. Such patterns include preambles, mid-ambles, regularly transmitted pilot patterns, spreading sequences etc. In the presence of a known pattern, sensing can be performed by correlating the received signal with a known copy of itself. This method is only applicable to systems with known signal patterns, and it is termed as waveform-based sensing or coherent sensing. It has been shown that waveform-based sensing outperforms energy detector based sensing in reliability and convergence time. Furthermore, it is shown that the performance of the sensing algorithm increases as the length of the known signal pattern increases.

4.3.0.11 Cyclostationary based sensing:

Cyclostationarity feature detection is a method for detecting primary user transmissions by exploiting the cyclostationarity features of the received signals. Cyclostationary features are caused by the periodicity in the signal or in its statistics like mean and auto-correlation or they can be intentionally induced to assist spectrum sensing. Instead of power spectral density (PSD), cyclic correlation function is used for detecting signals present in a given spectrum. The cyclostationarity based detection algorithms can differentiate noise from primary users signals. This is a result of the fact that noise is wide-sense stationary (WSS) with no correlation while modulated signals are cyclo-stationary with spectral correlation due to the redundancy of signal periodicities. Furthermore, cyclostationarity can be used for distinguishing among different types of transmissions and primary users.

4.3.0.12 Matched filtering based sensing:

Matched filtering is known as the optimum method for detection of primary users when the transmitted signal is known. The main advantage of matched filtering is the short time to achieve a certain probability of false alarm or probability of miss-detection as compared to other methods that are discussed in this section. In fact, the required number of samples grows as O(1/SN R) for a target probability of false alarm at low SNRs for matched-filtering. However, matched filtering requires cognitive radio to demodulate received signals.

27

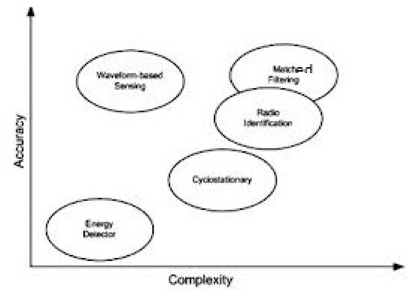


Figure 4.3.1: Comparison of sensing methods

Hence, it requires perfect knowledge of the primary users signaling features such as bandwidth, operating frequency, modulation type and order, pulse shaping, and frame format. Moreover, since cognitive radio needs receivers for all signal types, the implementation complexity of sensing unit is impractically large. Another disadvantage of match filtering is large power consumption as various receiver algorithms need to be executed for detection.

4.3.0.13 Radio identification based sensing:

Matched filtering is known as the optimum method for detection of primary users when the transmitted signal is known. The main advantage of matched filtering is the short time to achieve a certain probability of false alarm or probability of miss-detection as compared to other methods that are discussed in this section. In fact, the required number of samples grows as O(1/SN R) for a target probability of false alarm at low SNRs for matched-filtering. However, matched filtering requires cognitive radio to demodulate received signals. Hence, it requires perfect knowledge of the primary users signaling features such as bandwidth, operating frequency, modulation type and order, pulse shaping, and frame format. Moreover, since cognitive radio needs receivers for all signal types, the implementation complexity of sensing unit is impractically large. Another disadvantage of match filtering is large power consumption as various receiver algorithms need to be executed for detection.

On comparison, we see that Waveform-based sensing is more robust than energy detector and cyclostationarity based methods. Energy detector based sensing is limited but complexity is minimum. Also, Cyclostationary based methods perform worse than energy detector based sensing methods when the noise is stationary.

28

4.4 Problems in Cognitive Radio:

The basic problems in Cognitive Radio are as follows:

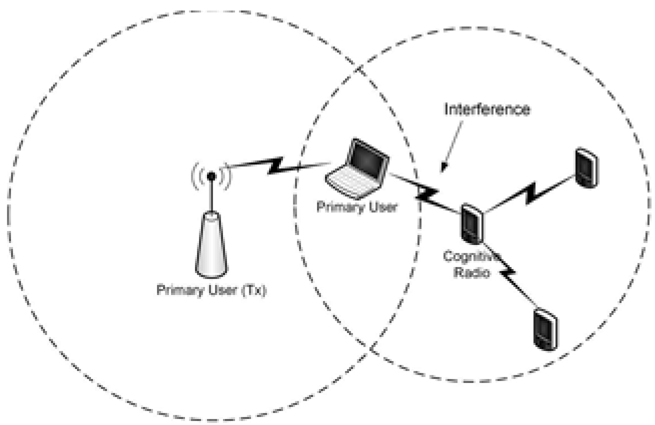
1. Hardware constraints
2. Hidden primary user problem
3. Detecting spread spectrum primary users
4. Sensing duration and frequency
5. Security

4.4.0.14 Hardware constraints:

Spectrum sensing for cognitive radio applications requires high sampling rate, high resolution analog to digital converters (ADCs) with large dynamic range, and high speed signal processors. Noise variance estimation techniques have been popularly used for optimal receiver designs like channel estimation, soft information generation etc., as well as for improved hand-off, power control, and channel allocation techniques The noise/interference estimation problem is easier for these purposes as receivers are tuned to receive signals that are transmitted over a desired bandwidth. Moreover, receivers are capable of processing the narrow band base band signals with reasonably low complexity and low power processors. However, in cognitive radio, terminals are required to process transmission over a much wider band for utilizing any opportunity. Hence, cognitive radio should be able to capture and analyze a relatively larger band for identifying spectrum op-port unities. The large operating bandwidths impose additional requirements on the radio frequencies (RF) components such as antennas and power amplifiers as well. These components should be able to operate over a range of wide operating frequencies. Furthermore, high speed processing units (DSPs or FPGAs) are needed for performing computationally demanding signal processing tasks with relatively low delay.

29

4.4.0.15 Hidden primary user problem:



The hidden primary user problem is similar to the hidden node problem in Carrier Sense Multiple Accessing (CSMA). It can be caused by many factors including severe multipath fading or shadowing observed by secondary users while scanning for primary users transmissions. Fig. 1.4 shows an illustration of a hidden node problem where the dashed circles show the operating ranges of the primary user and the cognitive radio device. Here, cognitive radio device causes unwanted interference to the primary user (receiver) as the primary transmitters signal could not be detected because of the locations of devices. Cooperative sensing is proposed in the literature for handling hidden primary user problem.

4.4.0.16 Detecting spread spectrum primary users:

For commercially available devices, there are two main types of technologies: fixed frequency and spread spectrum. The two major spread spectrum technologies are frequency-hoping spread-spectrum (FHSS) and direct-sequence spread-spectrum (DSSS). Fixed frequency devices operate at a single frequency or channel. An example to such systems is IEEE 802.11a/g based WLAN. FHSS devices change their operational frequencies dynamically to multiple narrow band channels. This is known as hopping and performed according to a sequence that is known by both transmitter and receiver. DSSS devices are similar to FHSS devices, however, they use a single band to spread their energy.

Primary users that use spread spectrum signaling are difficult to detect as the power of the primary user is distributed over a wide frequency range even though the actual information bandwidth is much narrower. This problem can be partially avoided if the hopping pattern is known and perfect synchronization to the signal can be achieved as discussed. However, it is not straightforward to design algorithms that can do the estimation in code dimension.

4.4.0.17 Sensing duration and frequency:

Primary users can claim their frequency bands anytime while cognitive radio is operating on their bands. In order to prevent interference to and from primary license owners, cognitive radio should be able to identify the presence of primary users as quickly as possible and should vacate the band immediately. Hence, sensing methods should be able to identify the presence of primary users within a certain duration. This requirement poses a limit on the performance of sensing algorithm and creates a challenge for cognitive radio design.

4.4.0.18 Security:

In cognitive radio, a selfish or malicious user can modify its air interface to mimic a primary user. Hence, it can mislead the spectrum sensing performed by legitimate primary users. Such a behavior or attack is termed as primary user emulation (PUE) attack.Here, the position of the transmitter is used for identifying an attacker. A more challenging problem is to develop effective countermeasures once an attack is identified. Public key encryption based primary user identification could be used to prevent secondary users masquerading as primary users. Legitimate primary users are required to transmit an encrypted value (signature) along with their transmissions which is generated using a private key. This signature is, then, used for validating the primary user. This method, however, can only be used with digital modulations. Furthermore, secondary users should have the capability to synchronize and demodulate primary users signal.

4.5 Our Algorithm for Cognitive Radio Sensing Via CS:

Consider a scenario with 'n' channels. Let N be the number of samples of the received signal at nyquist rate out of which we take M measurements using compressive sensing, where M<<N.

1. We first generate a signal bank 's'. This contains signal templates of all possible cases when only one channel is transmitting and the rest are not. The Fourier transform of these templates is stored in 's'.
2. The Fourier transform of the received signal 'y' is found using FFT() function
3. The sensing matrix (Phi) is generated using the normrnd() function of mat lab
4. The received signal is sampled using the sensing matrix $Y = \Phi y$ where $\Phi$is an MxN matrix;
5. We then find the test statistic using $t := Y^T(\Phi\Phi^T)^{-1}\Phi s_i$;where $s_i$is i^{th} channel signal template
6. We also define an ideal statistic as $t_{ideal} := s_i^T\Phi^T(\Phi\Phi^T)^{-1}\Phi s_i$;where $s_i$is i^{th} channel signal template
7. We then decide that a primary user is present if t lies within 50% of t_{ideal} by observation.

4.6.0.19 Comparison of CS Detector with Energy Detector

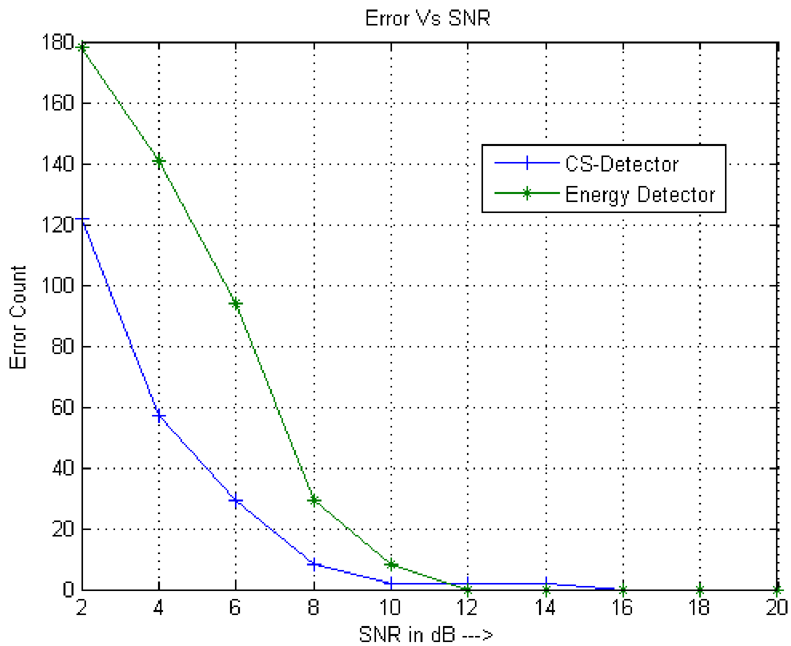


Figure 4.6.1: Error vs SNR

The settings for the above simulation were as follows:

Number of Samples Measured = 400
Total Number of Samples for CS-Detector = 512
Number of Channels = 20

At Lower SNR, our CS-detector outperforms energy detector significantly. At about 12 dB the perfomance converges

4.6 Energy detector

Consider a scenario with 'n' channels. The number of samples taken is N. Let received signal be Y.

1. We find Fourier transform of received signal $Y(\omega)$ using FFT function.

2. The power spectrum is obtained using $|Y(\omega)|^2$.

3. A suitable threshold is arrived at from observation.

4. If $|Y(\omega)|^2$ is greater than the threshold chosen, we say primary user is present at that particular frequency. Secondary user is simulated in the remaining frequencies.

5. Error rates and detection rates are noted.

4.7 Results- Cognitive Radio

In this section we have tabulated the results for Cognitive Radio. We have compared our results with that of cognitive radio implementation using Orthogonal Matching Pursuit and we have shown that our method is far superior both in terms of accuracy and time taken. In Cognitive Radio both these criteria are very important as they both are critical for the Primary User to function efficiently. Minimum accuracy and minimum time taken to detect the presence of Primary user will cause minimum interference to the Primary user.

33

4.7.0.20 Snapshot of Our Simulation

This section contains the snapshots of our simulation:

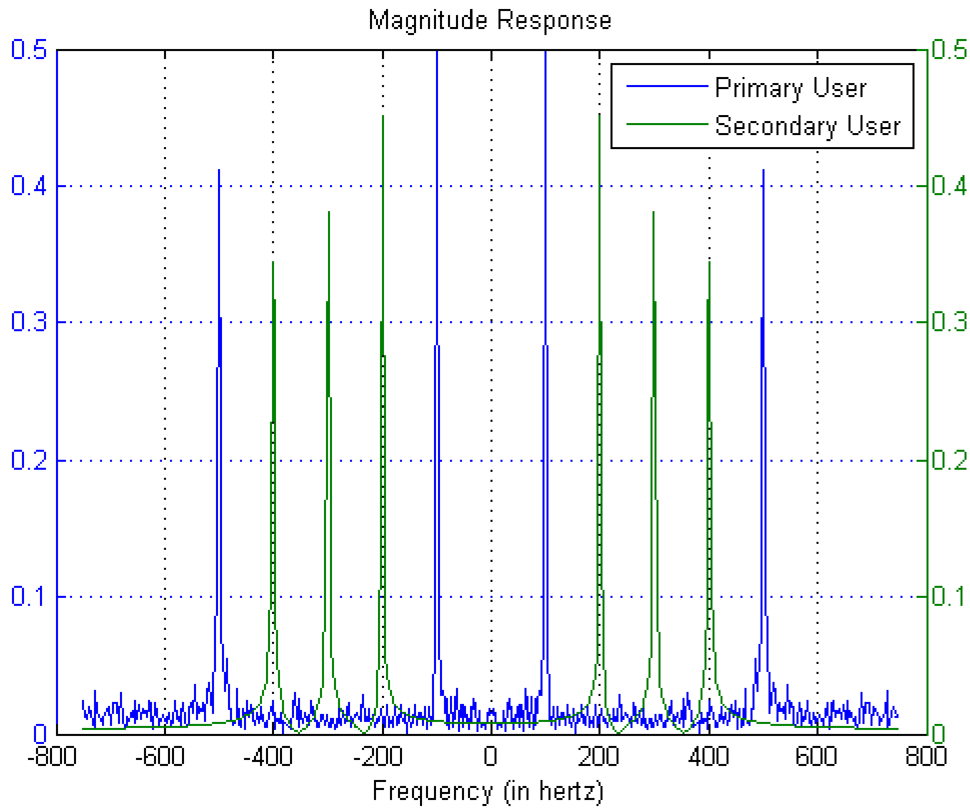


Figure 4.7.1: Simulation Snapshot

The settings for the above simulation were as follows:

SNR = 15 dB
Number of Samples Measured = 200
Total Number of Samples = 512
Number of Channels = 5

1. The blue lines indicate the frequency response of the Primary User that have been detected by our CS-Detector. The green are the transmission made by the Secondary user depending on the detection made.

4.7.0.21 Error Count

In this part we have shown the comparison of Error Count for two methods: CS-Detector and OMP-Estimator:

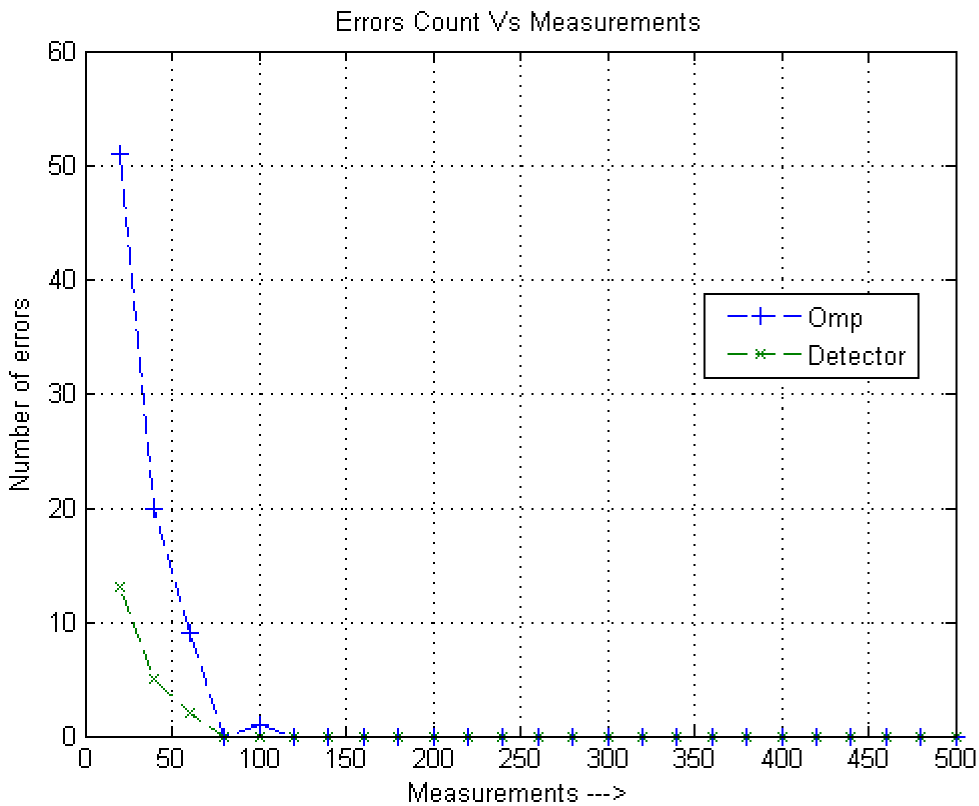


Figure 4.7.2: Errors Vs Measurements: CS-Detector Vs OMP Estimator

The settings for the above simulation were as follows:

SNR = 15 dB

Number of Channels = 5
Total Number of Samples: 512

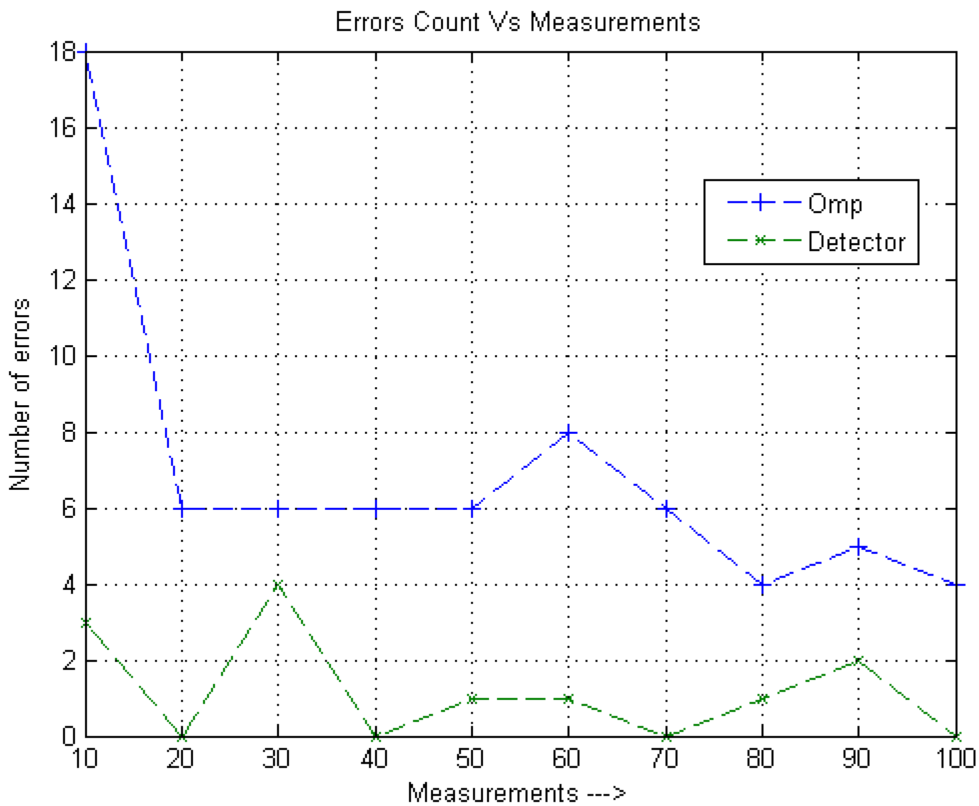


Figure 4.7.3: Error Vs SNR (dB) : CS-Detector vs OMP Estimator

The settings for the above simulation were as follows:

SNR = 15 dB
Number of Samples Measured = 80
Total Number of Samples = 512
Number of Channels = 5

1. **Note**: The error count is taken over 1000 transmissions of the Cognitive radio
2. The abrupt changes between the error counts of OMP and CS-Detector can be attributed to the random nature of the sensing matrix ϕ. Sometimes it will contain the required basis to capture the non-zero elements of the signals while sometimes it may miss out a few. But the significant trend that we have to observe is that the number of errors for OMP is always greater than the number for CS-Detector.

4.7.0.22 Time Taken

In this part we have shown the comparison of Error Count for two methods: CS-Detector and OMP-Estimator:

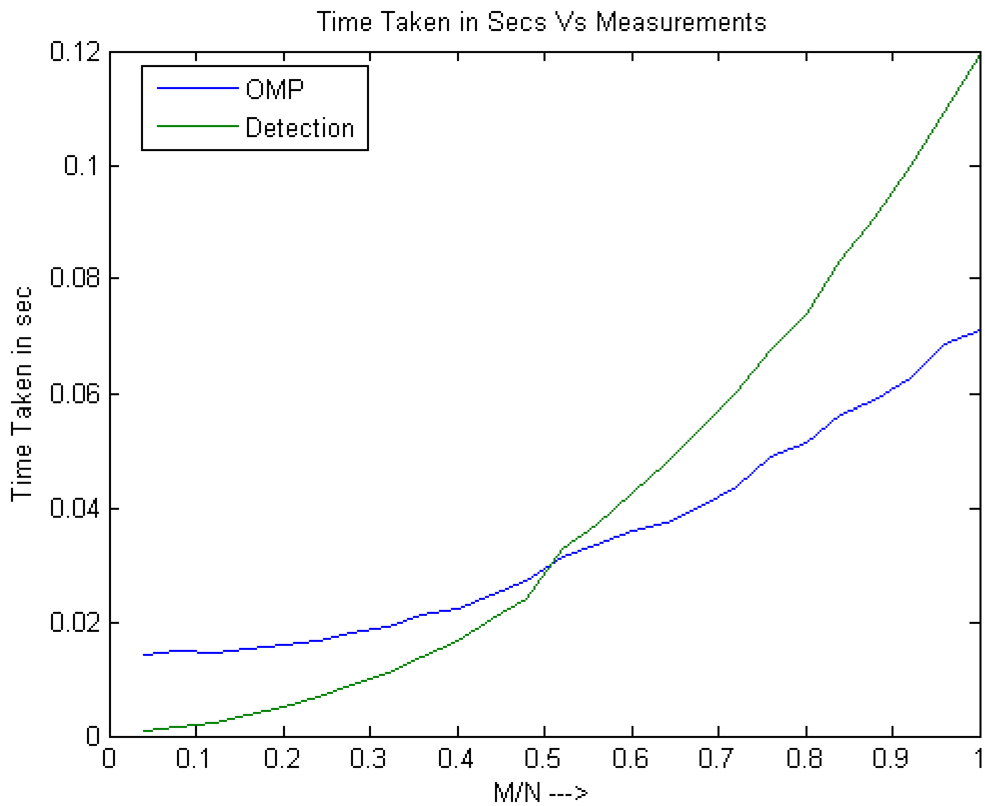


Figure 4.7.4: Time Vs Measurements: CS-Detector Vs OMP Estimator

The settings for the above simulation were as follows:

SNR = 15 dB
Total Number of Samples = 512
Number of Channels = 5

1. We observe that lower number of Measurements the performance of CS-Detector is far superior to that of OMP-Estimator. As the number of Measurements increases and reaches the Nyquist Criterion the Time Taken too increases

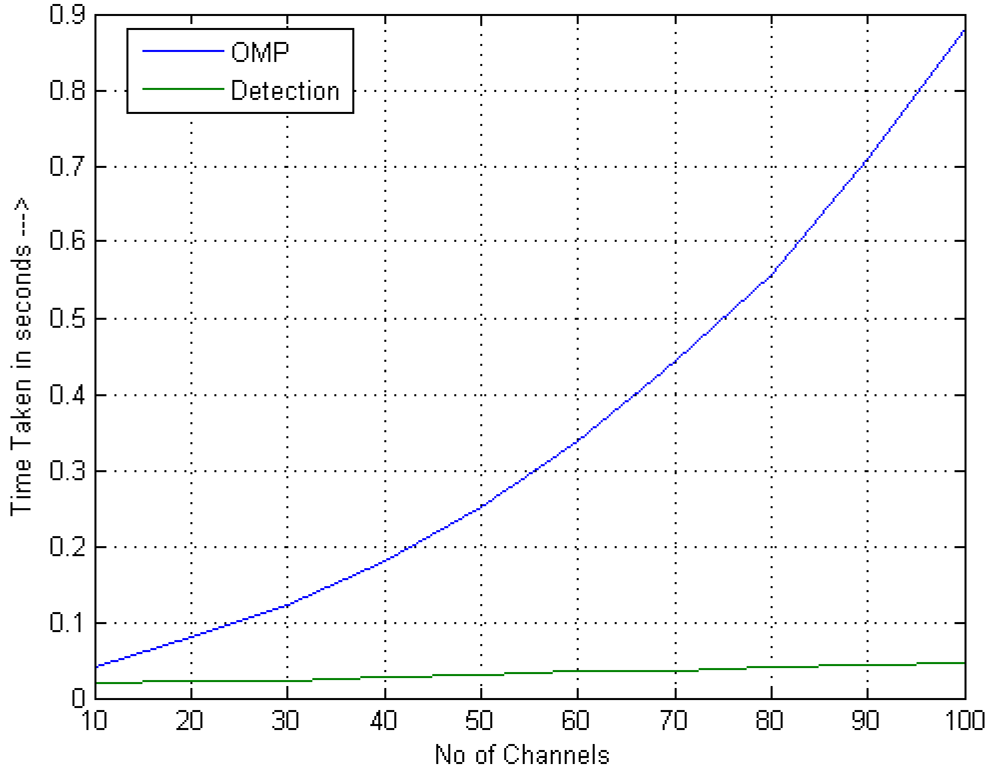


Figure 4.7.5: Time Vs Measurements: CS-Detector Vs OMP Estimator

The settings for the above simulation were as follows:

SNR = 20 dB
Number of Samples Measured = 200
Total Number of Samples = 512
Number of Channels = 5

1. Here we see that as the number of Channels that we seek to detect increases the time taken by OMP increases sharply while that of CS-Detector remains nearly the same. This is great advantage in practical purposes where we may need to detect the presence of a large number of primary users.

4.7.0.23 Performance of our CS-Detector at various settings

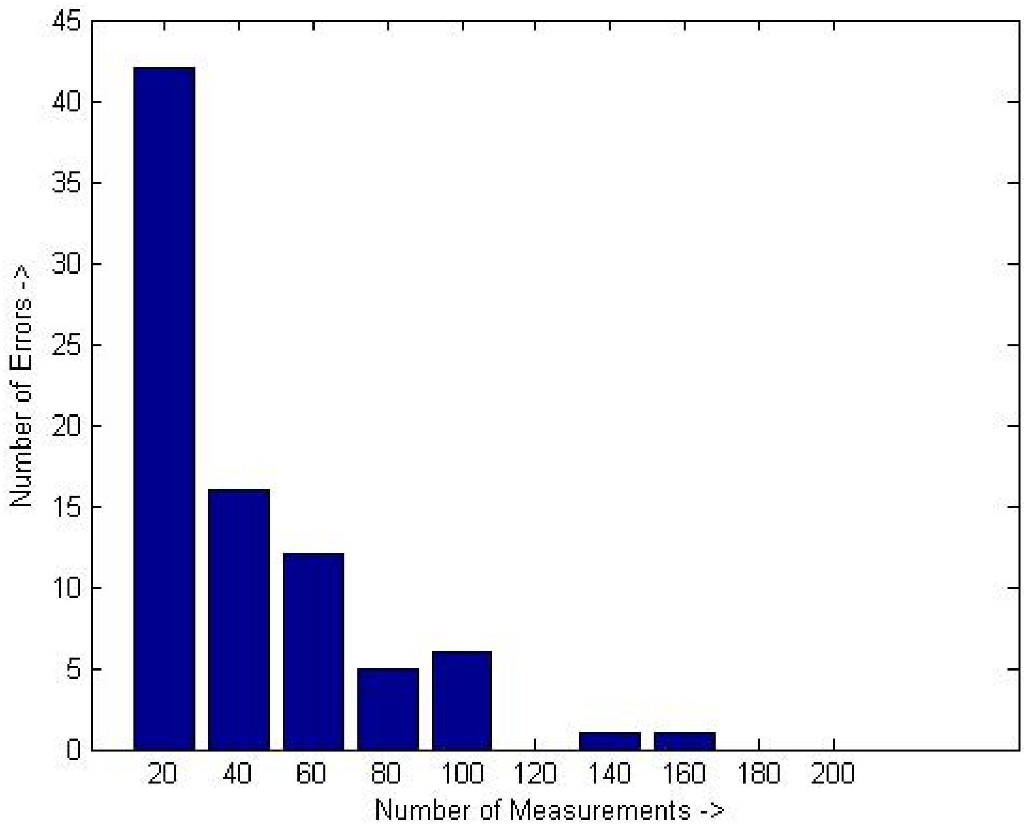


Figure 4.7.6: Number of Errors Vs Number of Measurements

The settings for the above simulation were as follows:

SNR = 4 dB
Total Number of Samples = 512
Number of Channels = 5

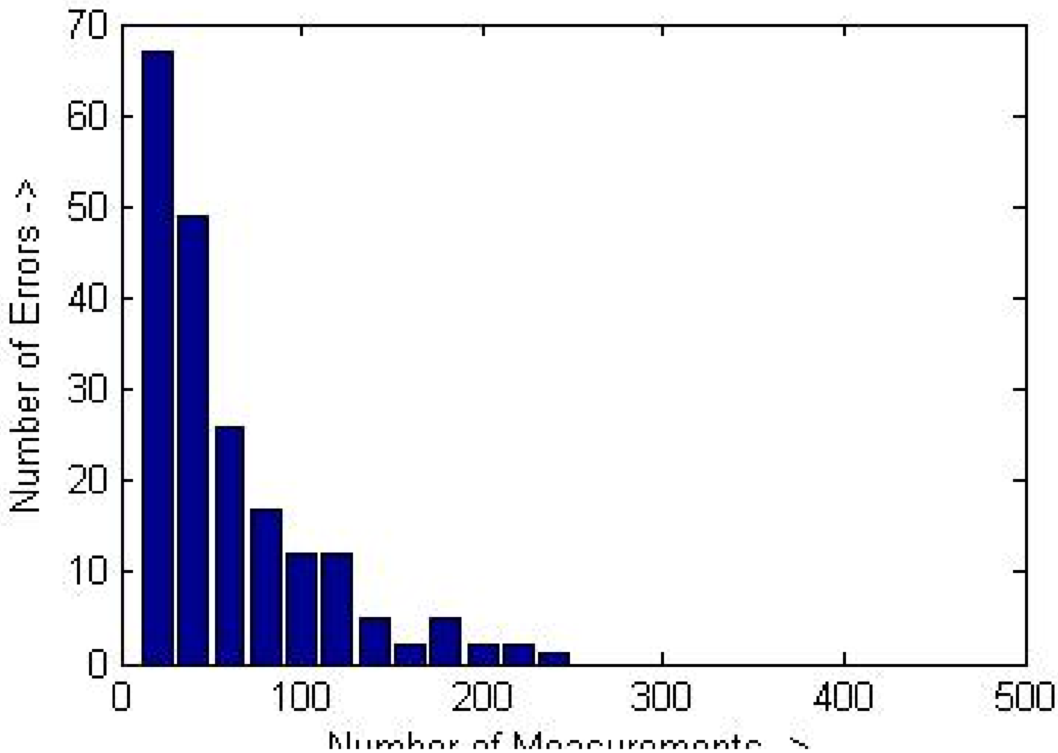


Figure 4.7.7: Number of Errors Vs Number of Measurements

The settings for the above simulation were as follows:

SNR = 4 dB
Total Number of Samples = 512
Number of Channels = 8

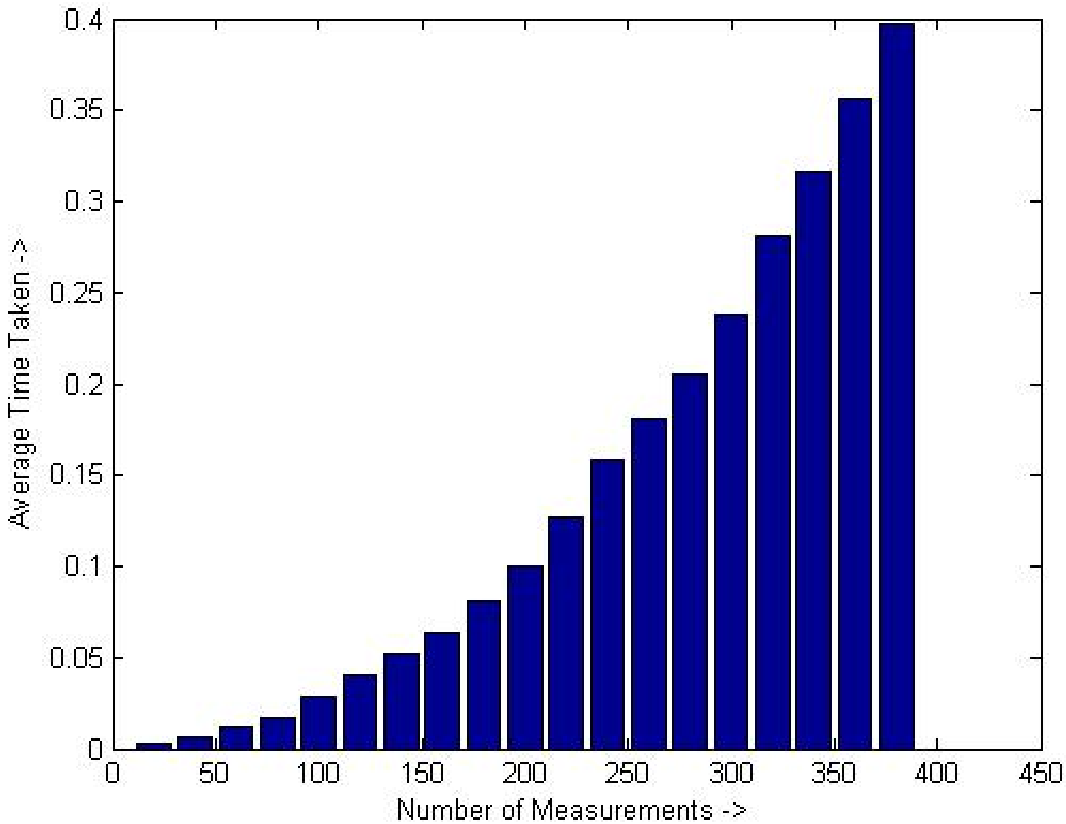


Figure 4.7.8: Time Taken Vs Number of Measurements
SNR = 4 dB
Total Number of Samples = 512
Number of Channels = 4

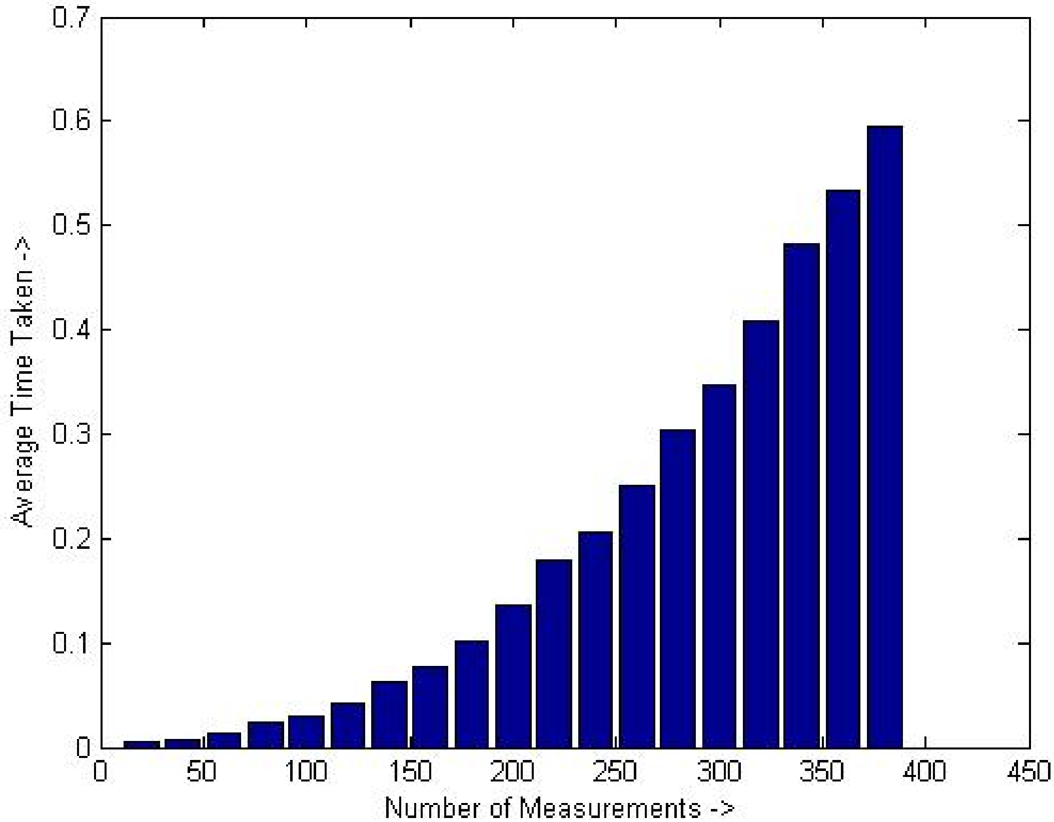


Figure 4.7.9: Time Taken Vs Number of Measurements
SNR = 4 dB
Total Number of Samples = 512
Number of Channels = 8

4.7.0.24 Orthogonal and Non-Orthogonal Carrier Frequencies Over Given Bandwidth

1. CR performance was compared for both orthogonal and non-orthogonal carrier frequencies.
2. No. of Channels=4
3. SNR=3
4. M=100

Type	Average Error / 100 transmissions
Orthogonal	4
Non Orthogonal	26

Figure 4.7.10: No. of errors: Orthogonal vs Non.Orthogonal per 100 Transmissions.

- It is seen that the number of errors increase significantly for non-orthogonal center frequencies

42

Chapter 5

Estimation

5.1 General Estimation Theory

While most of the signal processing applications can be solved using detection and classification in some cases we cannot reduce our task to selecting among finite set of hypothesis rather we are interested in estimating some function of the data.We focus on estimating a linear function of the data from compressive measurements.

Let $y = \phi s$ and we wish to estimate $< l, s >$from the measurements y,where $l \in \mathbb{R}^{\mathbb{N}}$is a fixed test vector.In the case where $\phi$is a random matrix,a natural estimator is essentially the same as the compressive detector.Suppose we would like to estimate a set of L linear functions from y.One potential estimator for this scenario,which is a simple generalization of the compressive detector (3.8) is given by

$$\frac{N}{M} * y^T(\phi\phi^T)^{-1}\phi l_i \tag{5.1.1}$$

for i=1,2,....,$|L|$

Equation (5.1) represents the orthogonalized estimator

Let us consider another estimator ; the most common one given by

$$< y, \phi l_i > \tag{5.1.2}$$

Equation (5.2) is called a direct estimator since it eliminates the orthogonalization step by directly correlating the compressive measurements with ϕl_i.

5.2 Commonly Used Estimators

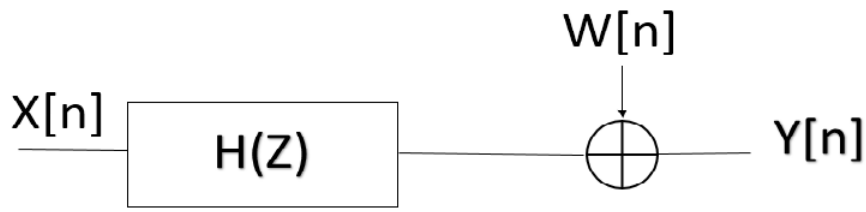


Figure 5.2.1: Block diagram of a noise corrupted signal scenario

5.3 Least Squares Estimator

The method of least squares is about estimating parameters by minimizing the squared discrepancies between observed data, on the one hand, and their expected values on the other. Least squares corresponds to the maximum likelihood criterion if the experimental errors have a normal distribution and can also be derived as a method of moments estimator.

Consider the block diagram 2.1 and the below conditions:

H(Z) = $\sum_{l=0}^{l=L-1}$h(l) z^{-l} L length FIR filter

x[n] : x[0].....x[M-1] —– M length sequence

h[n]: h[0]......h[L-1] —– Filter coefficients

y[n] = h[n] * x[n] + w[n]

$$\begin{bmatrix} Y[0] \\ . \\ . \\ . \\ . \\ . \\ Y[M+L-1] \end{bmatrix} = \begin{bmatrix} x[0] & 0 & 0 & 0 & 0 \\ x[1] & x[0] & 0 & 0 & 0 \\ . & . & x[0] & 0 & 0 \\ . & . & . & . & . \\ x[M-1] & . & . & . & x[0] \\ 0 & x[M-1] & . & . & . \\ 0 & 0 & . & . & x[M-1] \end{bmatrix} \begin{bmatrix} h[0] \\ . \\ . \\ . \\ . \\ . \\ h[L-1] \end{bmatrix} + \begin{bmatrix} w[0] \\ . \\ . \\ . \\ . \\ . \\ w[M+L-1] \end{bmatrix} \tag{5.3.1}$$

Y=XH+W where X is a toeplitz matrix

Estimate of H : $\hat{h}$ = (X^{T}X)$^{-1}$X^{T}Y

5.4 MMSE Estimator (Wiener filter)

Minimum mean square error (MMSE) estimator is an estimation method which minimizes the mean square error (MSE) of the fitted values of a dependent variable, which is a common measure of estimator quality.

Consider the block diagram 2.1 and the below conditions:

H(Z) = $\sum_{l=0}^{l=L-1}$h(l) z^{-l} L length FIR filter

x[n] : x[0].....x[M-1] —– M length sequence

h[n]: h[0]......h[L-1] —– Filter coefficients

y[n] = h[n] * x[n] + w[n]

$$\begin{bmatrix} Y[0] \\ . \\ . \\ . \\ . \\ . \\ Y[M+L-1] \end{bmatrix} = \begin{bmatrix} x[0] & 0 & 0 & 0 & 0 \\ x[1] & x[0] & 0 & 0 & 0 \\ . & . & x[0] & 0 & 0 \\ . & . & . & . & . \\ x[M-1] & . & . & . & x[0] \\ 0 & x[M-1] & . & . & . \\ 0 & 0 & . & . & x[M-1] \end{bmatrix} \begin{bmatrix} h[0] \\ . \\ . \\ . \\ . \\ . \\ h[L-1] \end{bmatrix} + \begin{bmatrix} w[0] \\ . \\ . \\ . \\ . \\ . \\ w[M+L-1] \end{bmatrix}$$

Y=XH+W where X is a toeplitz matrix

Estimate of H : $\hat{h}$ = (X^{T}X+$\sigma^2 I$)$^{-1}$X^{T}Y where $\sigma^2$is the noise variance

5.5 Application of Compressive Estimator to Determine the mean of a Signal

- We use the compressive estimator to find the mean of a signal s
- The signal s is a length N=1000 vector with entries distributed according to a Gaussian Distribution with unit mean and variance.
- l is chosen to be $l = [\frac{1}{N}\frac{1}{N}.......\frac{1}{N}]^T$to compute the mean of s.
- The average estimation error for an orthogonalized estimator is given by

$$|(N/M) * s^T\phi^T\phi l - < l, s > |/||s||_2||l||_2 \tag{5.5.1}$$

- The average estimation error for the direct estimator is given by

$$| < \phi l, \phi s > - < l, s > |/||s||_2||l||_2 \tag{5.5.2}$$

- The error for both the estimators is computed for varying (M/N) values.

46

Chapter 6

Channel Estimation

6.1 Introduction

Channel Estimation is an essential component in radar and data communication.In multipath time varying channel it is necessary to estimate the time-shift and the scale-shift of the multipaths.With recent advances in sparse estimation (Compressive Sensing) new estimation techniques have emerged which give better performances then the traditional techniques.We concentrate on the estimation of multipath channels with sparse impulse responses. Such channels are encountered in many applications such as high definition television (HDTV), communication near a hilly terrain, and underwater acoustic communication near the surf zone.Due to the sparse impulse responses of these channels, traditional estimation techniques such as least-squares result in over-parameterization and thus poor performance of the estimator.Fortunately, the structure of these channels can be exploited using sparse reconstruction algorithms such as Matching Pursuit (MP).We solve the channel estimation problem using the Orthogonal Matching Pursuit algorithm (OMP).The Channel can be sparse in the frequency domain and hence have a sparse frequency response, if this is the case (13) can be used to estimate the frequency response.

6.2 Application of Compressive Estimator in determining the Frequency Response

The Estimator in (5.1) can be extended for analysis of estimation beyond the estimation of scalar-valued linear functions to more general linear operators.Any finite-dimensional linear operator on a signal $x \in \mathbb{R}^{\mathbb{N}}$can be represented as a matrix multiplication Lx where L has size ZxN for some Z.Decomposing L in terms of its rows,this computation can be expressed as

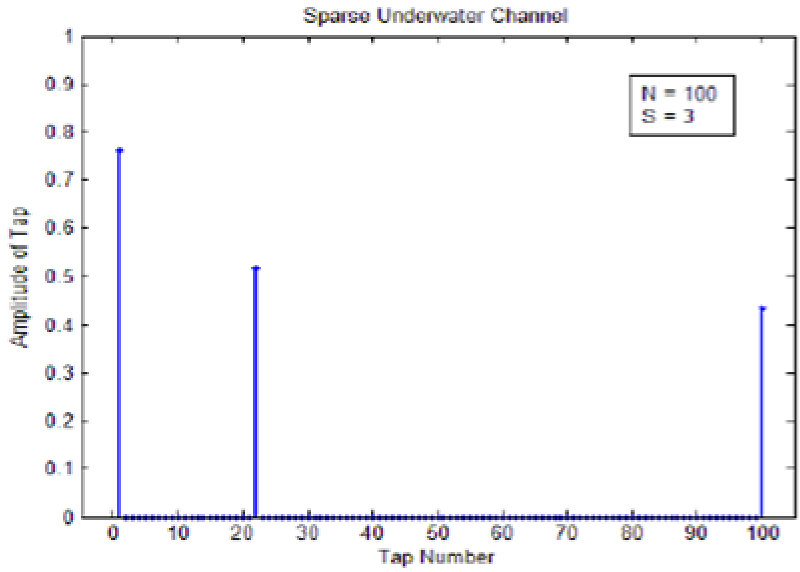


Figure 6.1.1: Sparse Under Water Channel

$$Lx = \begin{bmatrix} l_1^T \\ l_2^T \\ . \\ . \\ . \\ l_Z^T \end{bmatrix} x = \begin{bmatrix} < l_1, x > \\ < l_2, x > \\ . \\ . \\ . \\ < l_Z, x > \end{bmatrix} \tag{6.2.1}$$

We can note the following points

1. If the channel frequency response is sparse we can apply (6.1) in determining it.
2. The estimator finds the mean of a signal therefore the test vectors l_i have to be chosen such that the mean value of the signal at each frequency is found out thus giving us the frequency response
3. This can be achieved if L is taken as the DFT matrix.
4. Each row of the DFT matrix will select a particular row, the estimator will then find the mean of the signal at the frequency value selected by the DFT matrix therefore giving the frequency response.

6.3 Algorithm

1. Generate a random signal X
2. Take the N point FFT of X ; where N is the length of the signal

6.4 Frequency Response Results

6.4.0.25 Simulation Snapshots

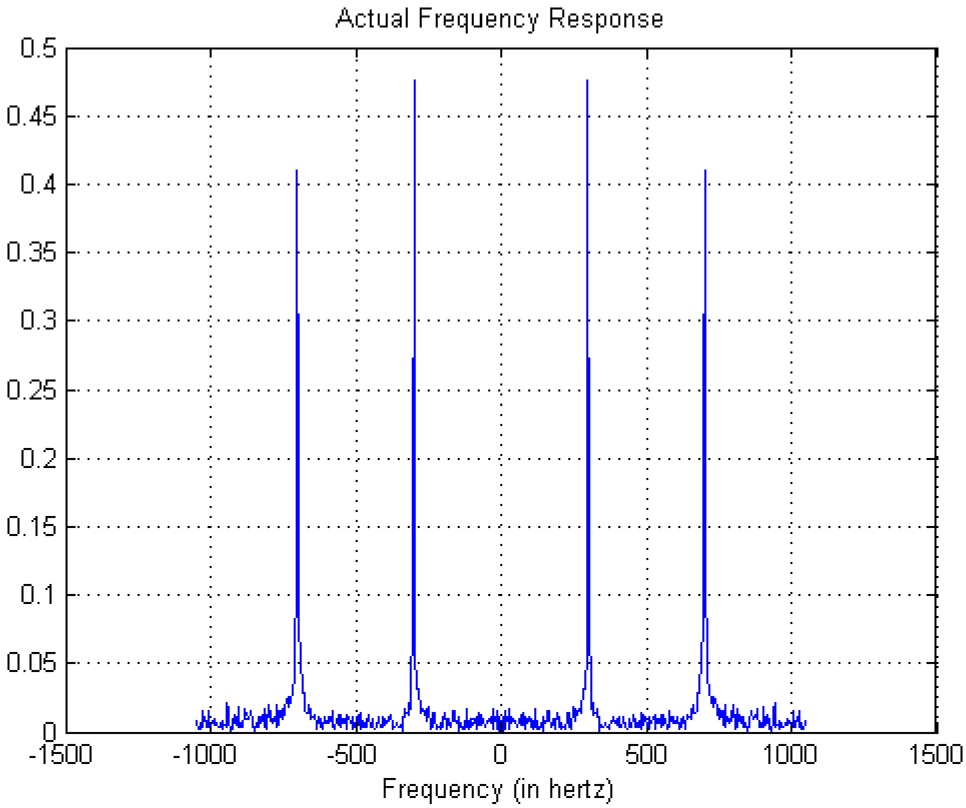


Figure 6.4.1: Actual Frequency Response
SNR=15dB
No. of Measurements=200
No. of Samples=512

3. Perform compressive Sensing on X ; $Y = \phi X$; where ϕ is the sensing matrix and Y is the measurements obtained.

4. The test vector matrix is taken as $l = DFT(N)$; which is the N point DFT matrix

5. From (5.1) and (6.1) $X_{estimate} = \frac{N}{M} * Y^T(\phi\phi^T)^{-1} * \phi l$; where M is the no. of measurements

6. $X_{estimate}$ gives the estimated frequency response.

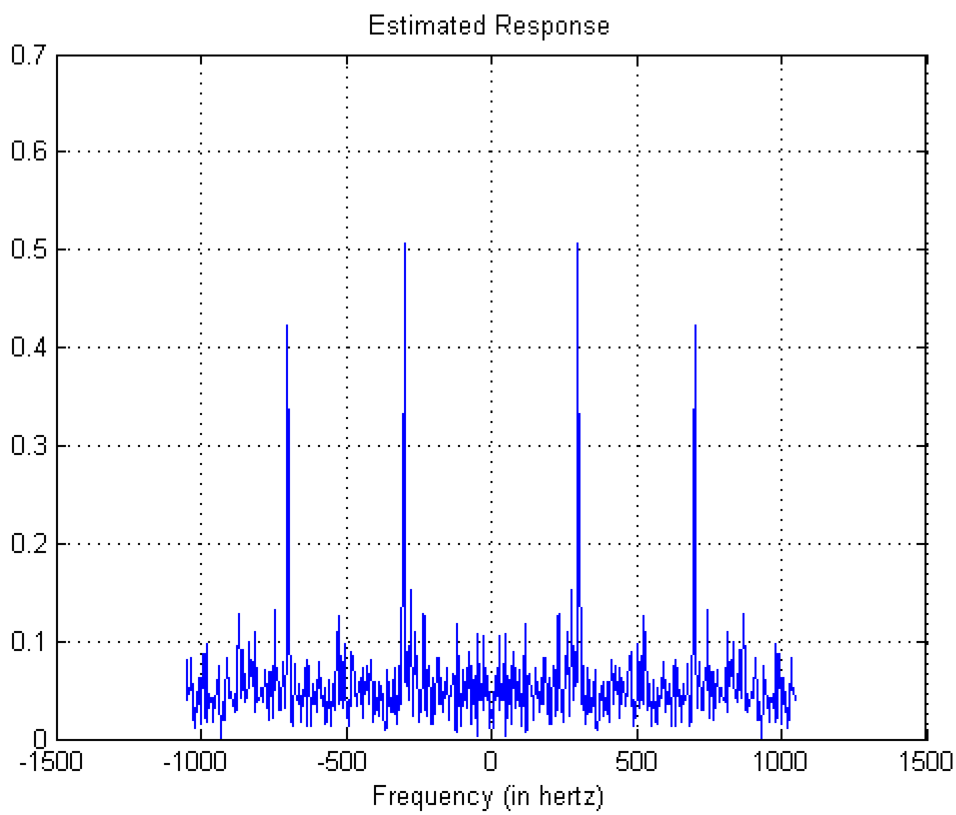


Figure 6.4.2: Estimated Frequency Response
SNR=15dB
No. of Measurements=200
Total no. of Samples=512

6.4.0.26 Errors vs measurements

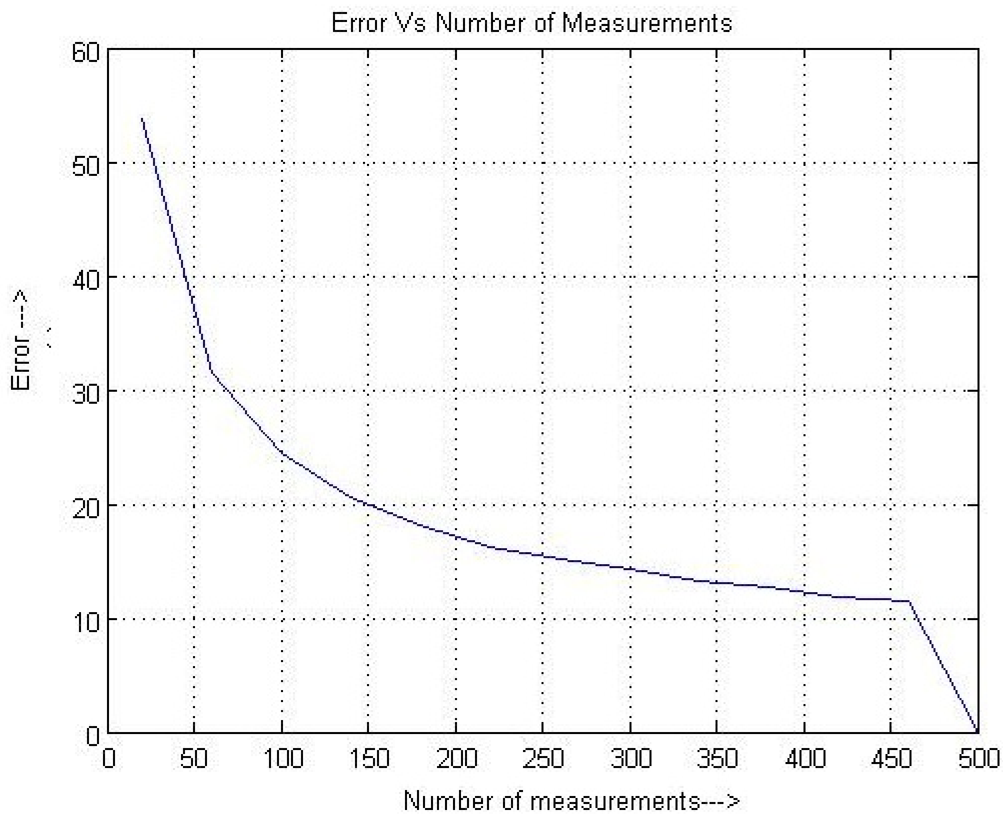


Figure 6.4.3: No.of Errors vs No. of Measurements
SNR=15 dB
Total Number of Measurements=512

6.4.0.27 Errors vs SNR

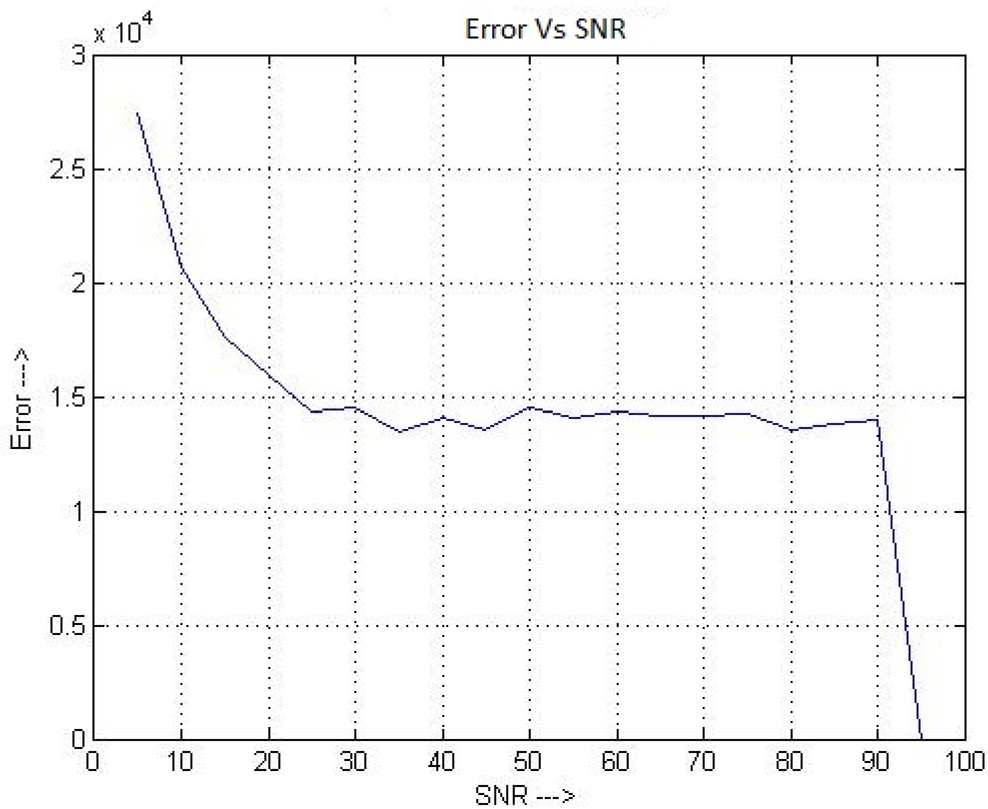


Figure 6.4.4: No.of errors vs SNR
No. of Measurements=15 dB
Total Number of Measurements=512

The variations observed in the middle are due to the random nature of the sensing matrix.

Chapter 7

Orthogonal Matching Pursuit

7.1 Introduction

Orthogonal matching pursuit (OMP) is an algorithm for the recovery of a high-dimensional sparse signal based on a small number of noisy linear measurements. OMP is an iterative greedy algorithm that selects at each step the column which is most correlated with the current residuals.It can be shown that under conditions on the mutual incoherence and the minimum magnitude of the nonzero components of the signal, the support of the signal can be recovered exactly by the OMP algorithm with high probability.

Recovery of a high-dimensional sparse signal based on a small number of linear measurements, possibly corrupted by noise, is a fundamental problem in signal processing. Specifically one considers the following model:

$$Y = \phi X + \eta \tag{7.1.1}$$

where the observation $Y \epsilon \mathbb{R}^{\mathbb{N}}$, the matrix $\phi \epsilon \mathbb{R}^{\mathbb{N} \mathbb{X} \mathbb{P}}$, and the measurement errors $\varepsilon \epsilon \mathbb{R}^{\mathbb{N}}$. Suppose $\phi = (\phi_1, \phi_2, \cdots, \phi_P)$ where ϕ_i denotes the ith column of ϕ . We shall assume that the columns of ϕ are normalized, i.e. $\|\phi_i\|_2 = 1$ for i = 1; 2;$\cdots$; p. The goal is to reconstruct the unknown vector $X \epsilon \mathbb{R}^{\mathbb{P}}$ based on Y and ϕ. A setting that is of significant interest and challenge is that the dimension p of the signal is much larger than the number of measurements n.

For a vector $X \epsilon (X_1, X_2, \cdots, X_P)$, the support of X is defined to be the set supp$(X) = \{i; X_i \neq 0\}$ and X is said to be k-sparse if $| \, supp(X) \, | \leqslant k$. A widely used framework for sparse signal recovery is the Mutual Incoherence Property (MIP) introduced in Donoho and Huo (2001). The mutual incoherence is defined by

$$\mu = \max_{i \neq j} |< \phi_i, \phi_j >| \tag{7.1.2}$$

The MIP requires the mutual incoherence μ to be small. Other conditions used in the compressive sensing literature include the Restricted Isometry Property (RIP) and Exact Recovery Condition (ERC).In contrast to the MIP, these conditions are not computationally feasible to verify for a given matrix ϕ. We consider the orthogonal matching pursuit (OMP) algorithm for the recovery of the support of the k-sparse signal X under the above model

.OMP is an iterative greedy algorithm that selects at each step the column of ϕ which is most correlated with the current residuals. This column is then added into the set of selected columns. The algorithm updates the residuals by projecting the observation Y onto the linear subspace spanned by the columns that have already been selected, and the algorithm then iterates. Compared with other alternative methods, a major advantage of the OMP is its simplicity and fast implementation.

In particular, support recovery has been considered in the noiseless case by Tropp (2004), where it was shown that $\mu < \frac{1}{2k-1}$ is a sufficient condition for recovering a k-sparse X exactly in the noiseless case. We however consider the OMP algorithm in the general setting where noise is present. Note that the residuals after each step in the OMP algorithm are orthogonal to all the selected columns of ϕ, so no column is selected twice and the set of selected columns grows at each step.One of the key components of an iterative procedure like OMP is the stopping rule. Specific stopping rules are given for the OMP algorithm in both bounded noise and Gaussian noise cases. The algorithm is then fully data-driven. It can be shown that under the MIP condition $\mu < \frac{1}{2k-1}$ and a condition on the minimum magnitude of the nonzero coordinates of X, the support of X can be recovered exactly by the OMP algorithm in the bounded noise cases and with high probability in the Gaussian case.In fact all the main results hold under the Exact Recovery Condition (ERC).

In many applications, the focus is often on identifying significant components, i.e., coordinates of X with large magnitude, instead of the often too ambitious goal of recovering the whole support of X exactly. In our simulation, we also consider the problem of identifying large coordinates of X in the case where some of the nonzero coordinates are possibly small. It is shown that in this case the OMP algorithm will still select all the most important components before possibly selecting incorrect ones. In addition, with modified stopping rules, the OMP algorithm can ensure that no zero components are selected.

7.2 The OMP Algorithm

In this section we give a detailed description of the orthogonal matching pursuit (OMP) algorithm. We assume that the columns of ϕ are normalized so that $\|\phi_i\|_2 = 1$ for i = 1; 2; $\cdots$; p. For any subset S$\subseteq \{1, 2, \cdots, p\}$, denote by ϕ(S) a sub matrix of ϕ consisting of the columns ϕ_i with i ϵ S. In our report we shall also call columns of ϕ variables by following the convention in statistics. Thus we use ϕ_i to denote the both ith column of ϕ and the ith variable of the model. Following the same convention, we shall call

ϕ_i a correct variable if the corresponding $X_i \neq 0$ and call ϕ_i an incorrect variable otherwise. With slight abuse of notation, we shall use ϕ(S) to denote both the subset of columns of ϕ with indices in S and the corresponding sub matrix of ϕ.

The OMP algorithm can be stated as follows.

- Step 1: Initialize the residual $r_0 =$ y and initialize the set of selected variable $\phi(c_0) =\phi$;. Let iteration counter i = 1.
- Step 2: Find the variable ϕ_{t_i} that solves the maximization problem $\max_t |\phi_t^{'} r_{i-1}|$ and add the variable ϕ_{t_i} to the set of selected variables. Update $c_i = c_{i-1} U \{t_i\}$
- :Step 3: Let $P_i = \phi(c_i)(\phi(c_i)^t\phi(c_i))^{-1}\phi(c_i)^t$ denote the projection onto the linear space spanned by the elements of $\phi(c_i)$. Update $r_i =$ (I - P_i)y.
- Step 4: If the stopping condition is achieved, stop the algorithm. Otherwise, set i = i + 1 and return to Step 2.

The OMP is a stepwise forward selection algorithm and is easy to implement. A key component of OMP is the stopping rule which depends on the noise structure. In the noiseless case the natural stopping rule is $r_i = 0$. That is, the algorithm stops whenever $r_i = 0$ is achieved.

In our project we have considered the Gaussian noise where $\varepsilon_i \epsilon \eta(0, \sigma^2)$

7.2.1 The OMP Algorithm: Stopping Rules and Properties

The Gaussian noise case is of particular interest in statistics. The results on the bounded noise cases given earlier are directly applicable to the case where noise is Gaussian. This is due to the fact that Gaussian noise is "essentially bounded".

It can be shown that for $\varepsilon\epsilon\eta(0, \sigma^2 I_n)$ case the stopping condition $\|r_i\|_2 \leq \sigma\sqrt{n + 2\sqrt{nlogn}}$ will select the true subset $\phi(T)$with probability $1 - 1/n$. For our project we have assumed the sparsity of the received signal to be k. We performed OMP for k iterations to recover the supports.

7.2.2 Understanding OMP: Why and How it Works

To gain insight on the OMP algorithm and to illustrate the main ideas behind the proofs, it is instructive to provide some technical analysis of the algorithm. The analysis sheds light on how and when the OMP algorithm works properly.

Note that the support T = $\{i : X_i \neq 0\}$ and the set of significant or "correct" variables is ϕ(T) = $\{\phi_i : i\epsilon T\}$. At each step of the OMP algorithm, the residual vector is projected onto the space spanned by the selected variables (columns of ϕ). Suppose the algorithm selects the correct variables at the first

t steps and the set of all selected variables at the current step is $\phi(\mathrm{c}_t)$. Then $\phi(\mathrm{c}_t)$ contains t variables and $\phi(\mathrm{c}_t)\subset\phi(\mathrm{T})$. Recall that $\mathrm{P}_t = \phi(\mathrm{c}_i)(\phi(\mathrm{c}_i)^t\phi(\mathrm{c}_i))^{-1}\phi(\mathrm{c}_i)^t$ is the projection operator onto the linear space spanned by the elements of $\phi(\mathrm{c}_t)$. Then the residual after t steps can be written as

$$r_t = (I - P_t)y = (I - P_t)\phi x + (I - P_t)\epsilon \equiv s_t + n_t \tag{7.2.1}$$

where $s_t = (I - P_t)\phi x$ is the signal part of the residual and $n_t = (I - P_t)_\epsilon$ is the noise part of the residual. Let

$$M_{t,1} = \max_{x \epsilon X(T)} \{|\phi' n_t|\} \tag{7.2.2}$$

$$M_{t,2} = \max_{x \epsilon XX(T)} \{|\phi' s_t|\} \tag{7.2.3}$$

and

$$N_t = \max_{x \epsilon X}\{|x' n_t|\} \tag{7.2.4}$$

It is clear that in order for OMP to select a correct variable at this step, it is necessary to have $\max_{x\epsilon X(T)}\{|x'r_t|\} > \max_{x\epsilon x/X(t)}\{|x'r_t|\}$. A sufficient condition is $M_{t,1} - M_{t,2} > 2N_t$.This is because $M_{t,1} - M_{t,2} > 2N_t$implies

$$\max_{x\epsilon X(T)}\{x'r_t\} \geq M - N_t > M_{t,2} + N_t \geq \max_{x\epsilon X(T)}\{x'r_t\} \tag{7.2.5}$$

7.3.0.3 Average error vs SNR

7.3.1 How We Applied OMP for Channel Estimation

1. Perform compressive sensing on the received signal
2. Apply OMP algorithm to the received signal.
3. Here We assume that we have knowledge about the sparsity of the channel beforehand.
4. After Performing OMP and finding those basis with high correlation to that of received signal we set the remaining basis to be zero.
5. This changes the system from an under-determined to an determined system
6. We are now left with the same number of equations as that of unknowns

7.3 Channel Estimation with OMP results

7.3.0.1 Simulation Screen shots

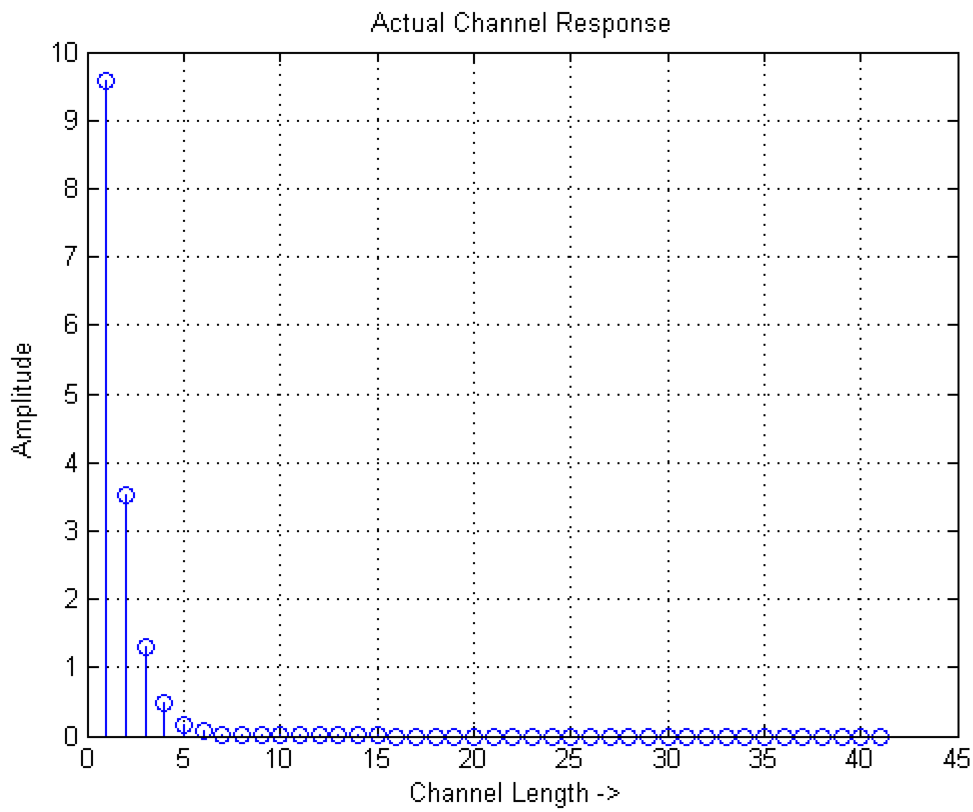


Figure 7.3.1: Actual Impulse Response

SNR of Channel :15
Length of channel response :247
Number of sparse elements in the channel response :15
Length of the pilot :25
The number of samples N: 271
Number of Measurements :150

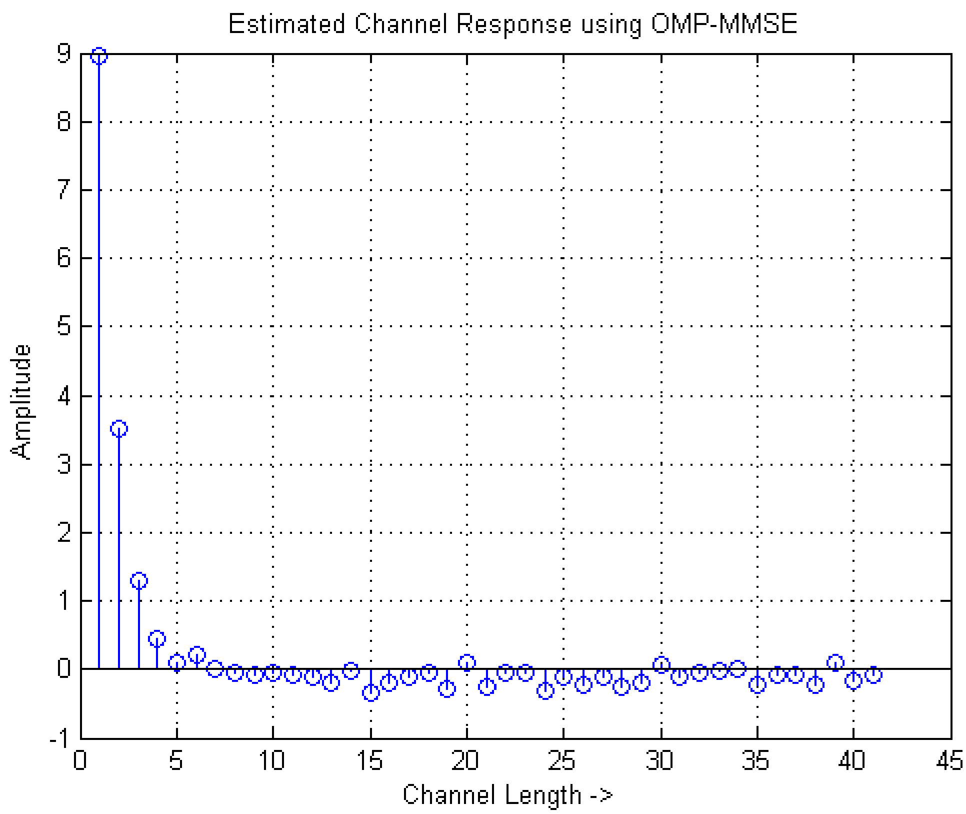


Figure 7.3.2: Estimated Channel Impulse Response
SNR of Channel :15
Length of channel response :247
Number of sparse elements in the channel response :15
Length of the pilot :25
The number of samples N: 271
Number of Measurements :150

7.3.0.2 Average error vs. measurements taken

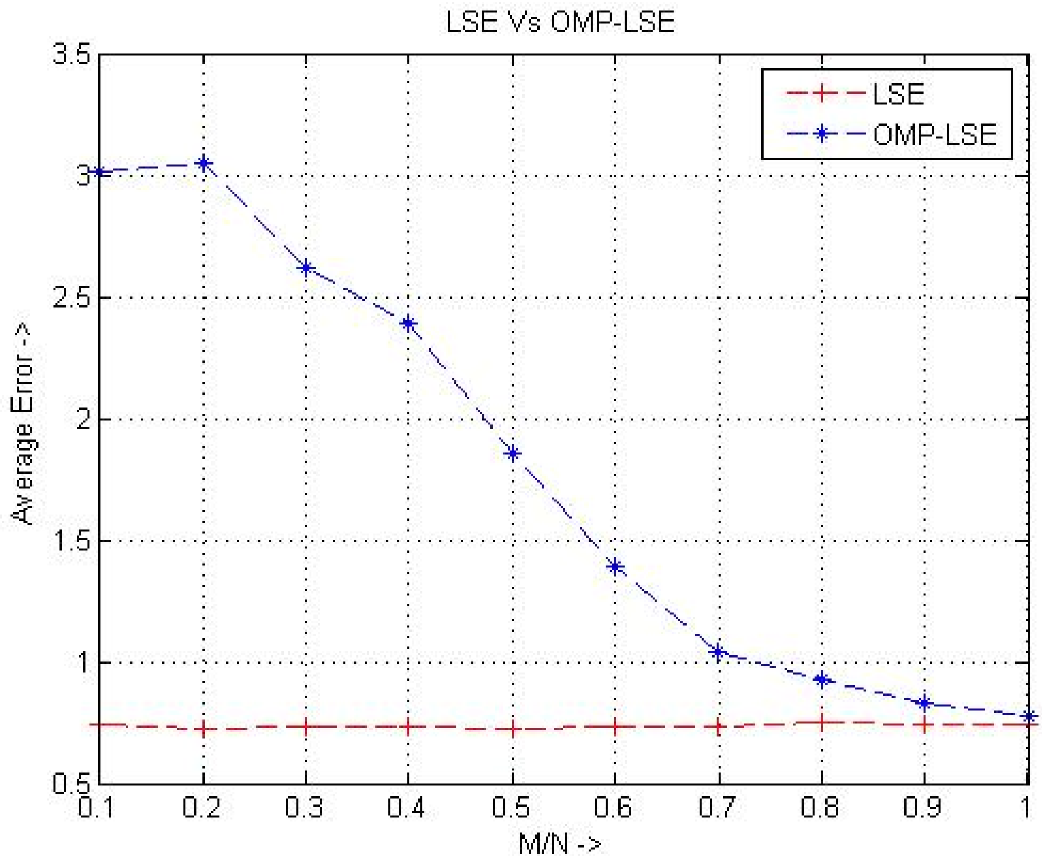


Figure 7.3.3: Average error vs $\frac{M}{N}$ for LSE vs OMP-LSE
Length of channel response :181
Number of Sparse elements in the channel response :10
Length of the pilot :20
SNR of Channel :10 dB

1. N is the no. of measurements
2. M is the signal Length
3. Since OMP involves CS the results depend on M and the average errors reduce as M increases.
4. LSE no.of measurements is fixed therefore with M varying performance does not alter.

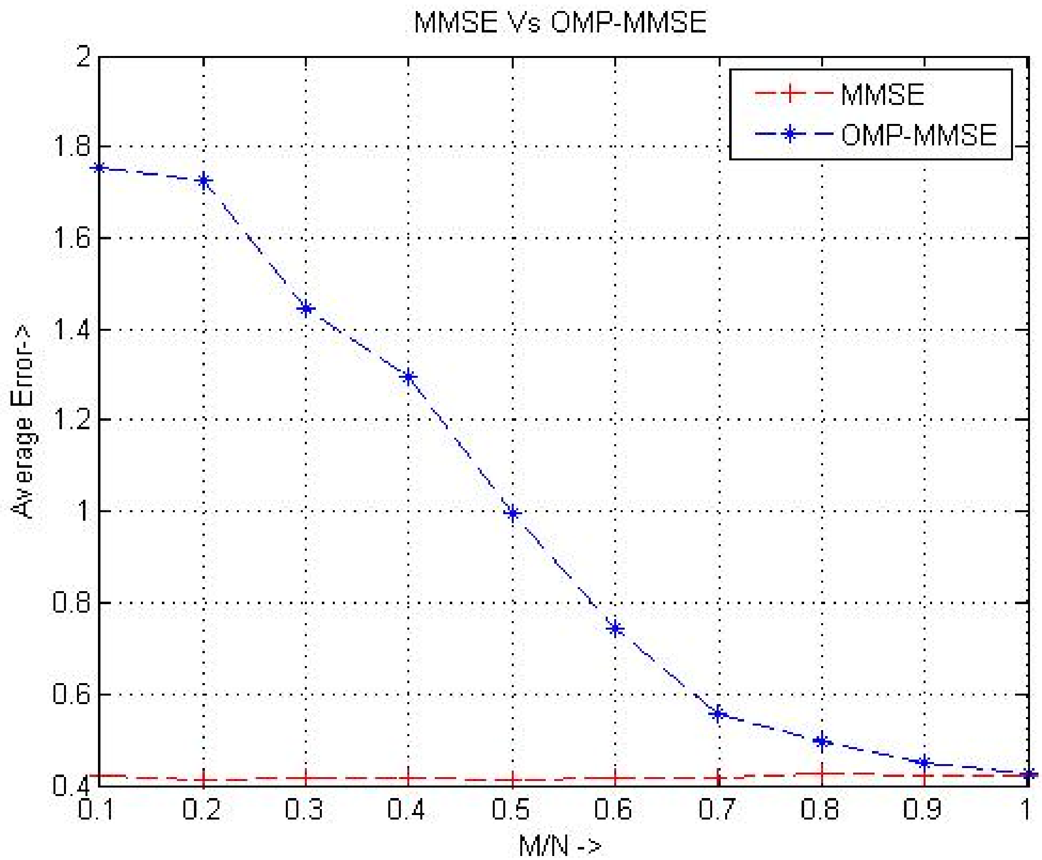


Figure 7.3.4: Average error vs $\frac{M}{N}$ for MMSE vs OMP-MMSE
Length of channel response :181
Number of sparse elements in the channel response :10
Length of the pilot :20
SNR of Channel :10 dB

1. N is the no. of measurements
2. M is the signal Length
3. Since OMP involves CS the results depend on M and the average errors reduce as M increases.
4. MMSE no.of measurements is fixed therefore with M varying performance does not alter.

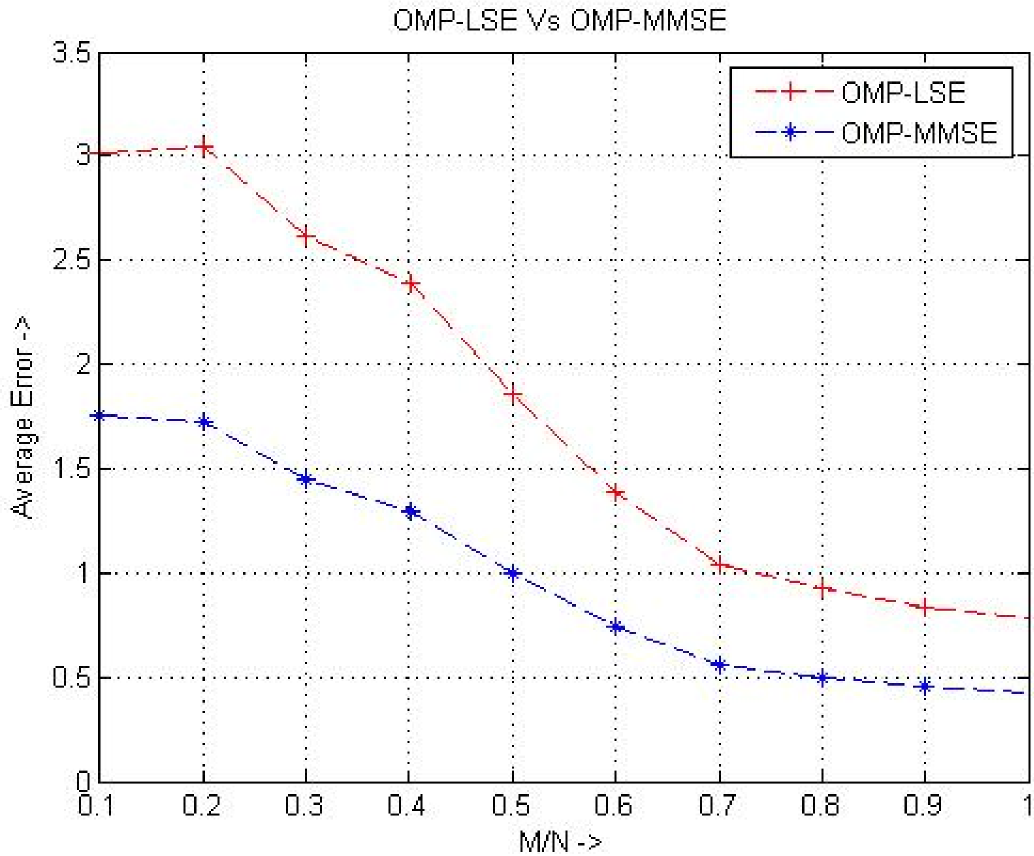


Figure 7.3.5: Average error vs $\frac{M}{N}$ for OMP-LSE vs OMP-MMSE
Length of channel response :181
Number of sparse elements in the channel response :10
Length of the pilot :20
SNR of Channel :10 dB

1. N is the no. of measurements
2. M is the signal Length
3. As seen from the above graph OMP with MMSE for estimating the channel impulse response performs better then OMP with LSE for estimating the channel impulse response

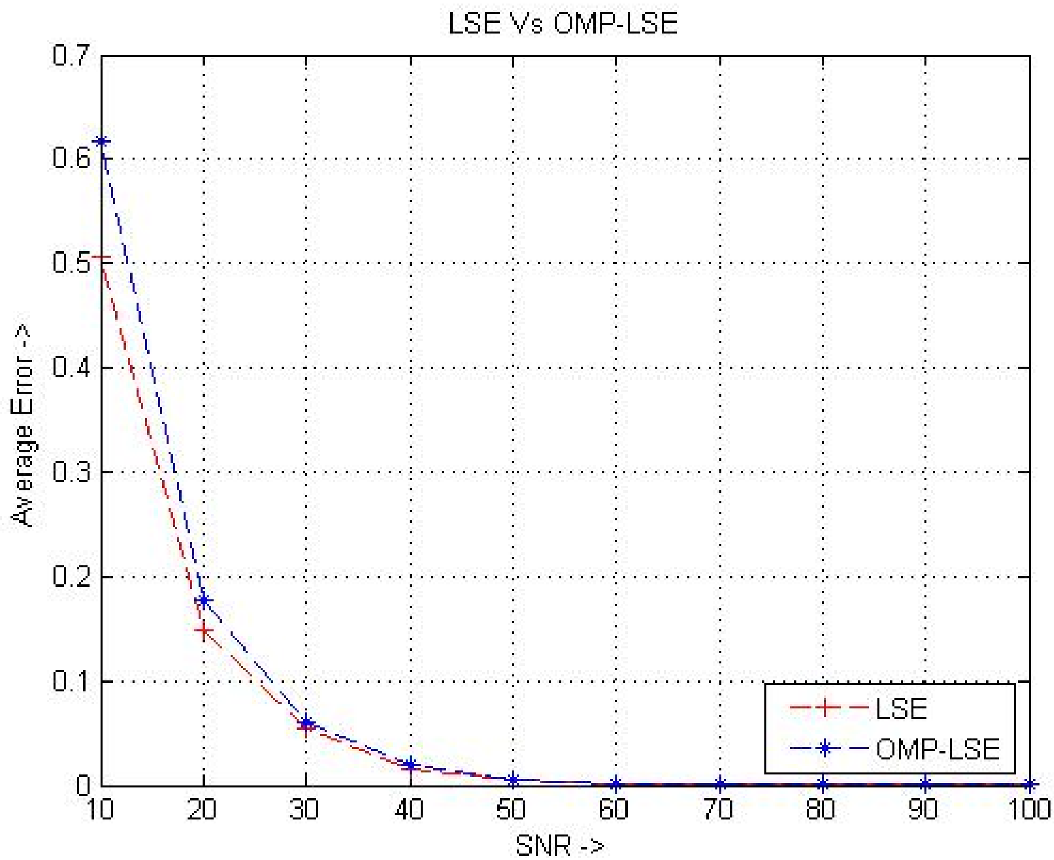


Figure 7.3.6: Average error vs SNR for LSE vs OMP-LSE
Length of channel response :247
Number of sparse elements in the channel response :20
Length of the pilot :20
no.of measurements :100

1. As observed from the graph at very low SNR LSE performs better then OMP-LSE but as SNR reaches around 30 dB the performance becomes almost equal.

2. Therefore at around 30dB we can get the same performance from roughly 20% percent measurements.

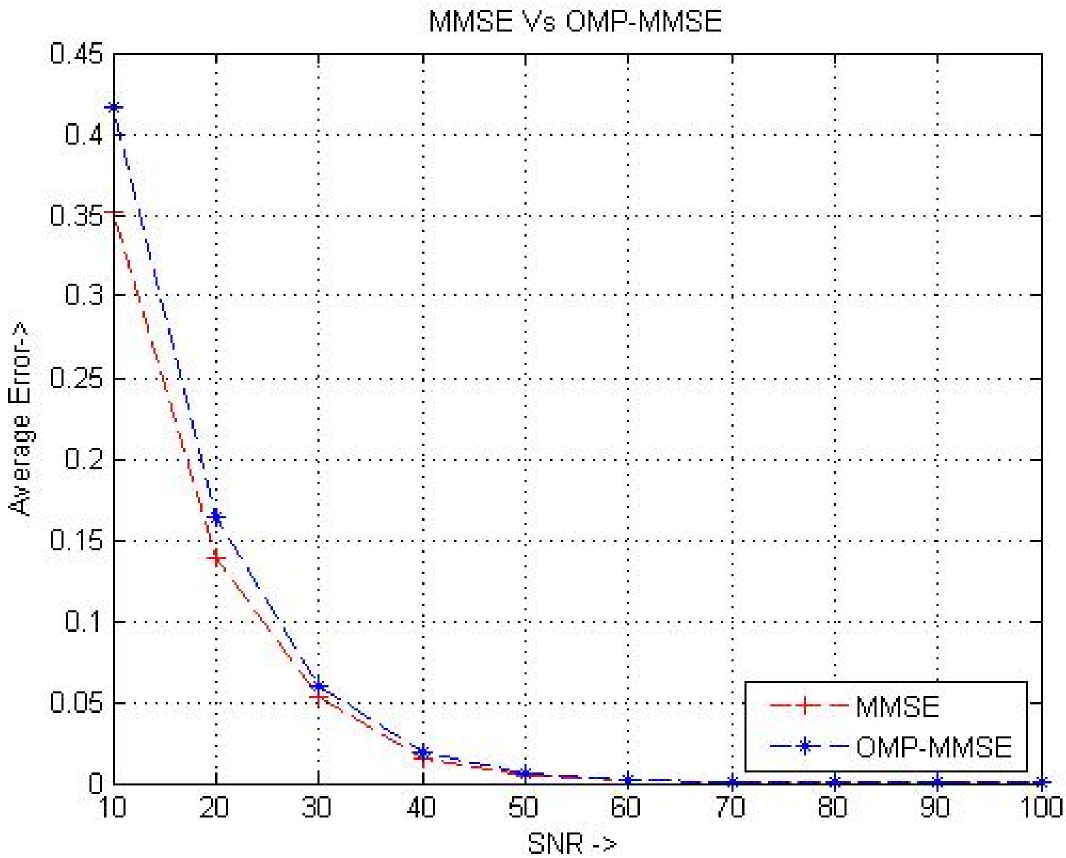


Figure 7.3.7: Average error vs SNR for MMSE vs OMP-MMSE
Length of channel response :247
Number of sparse elements in the channel response :20
Length of the pilot :20
no.of measurements :100

1. As observed from the graph at very low SNR MMSE performs better then OMP-LSE but as SNR reaches around 30 dB the performance becomes almost equal.

2. Therefore at around 30dB we can get the same performance from roughly 20% percent measurements.

3. Comparing above two graphs we observe an expected result that in general MMSE performs better then LSE.

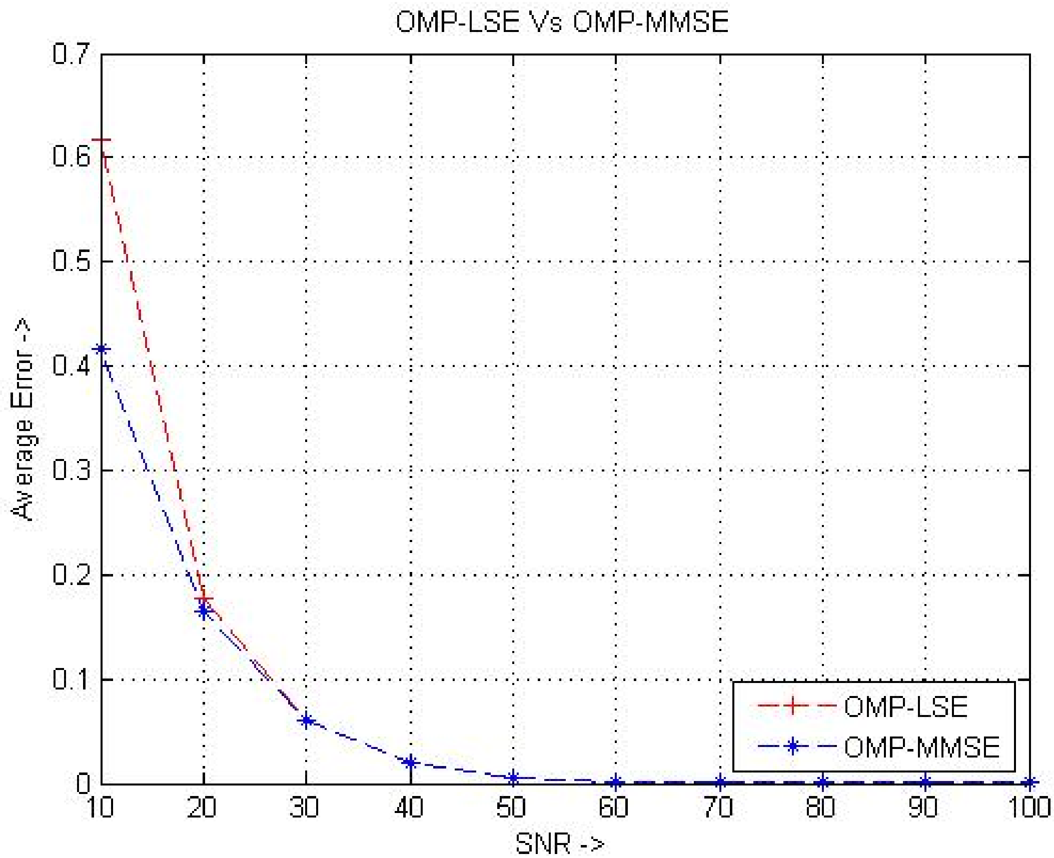


Figure 7.3.8: Average error vs SNR for OMP-MMSE vs OMP-MMSE
Length of channel response :247
Number of sparse elements in the channel response :20
Length of the pilot :20
no.of measurements :100

1. At low SNR values OMP-MMSE outperforms OMP-LSE. This is implicit.
2. As SNR reaches around 30dB the performance of the two becomes comparable

7. Solving this will give us the original signal

8. Once we have recovered these signals we then perform LSE or MMSE estimation to find the channel response

65

Chapter 8

Conclusion and Future work

It is evident from the work done till now, Compressive sensing is still a new concept that has much more to offer to the world. As shown in our results, it is clear that it has the potential to impact a major breakthrough in the signal processing field. The cognitive Radio via compressive sensing implementation developed by us has numerous advantages:

1. For detecting the primary users, our method only uses only 10% (approx.) of samples relative to the traditional methods.

2. As previously mentioned, we do not estimate the signal, we merely detect if there is a primary user or not by appropriate thresholding, thereby eliminating reconstruction of the signal. The benefits of this point is mainly due to the fact that most of the complexity of compressive sensing lies in receiver stage or the decoding stage and not the encoding block.

3. From our simulation results, one can observe that our method performs with low error rates adding to the benefits.

4. We implemented an energy detector (which uses the simplest sensing method) to compare with our model for performance results. From the thorough analysis presented, it is safe to say that our model stands low in complexity while at the same time sits high in accuracy making it tremendously advantageous.

Our model could be further developed by implementing other sensing methods such as waveform based sensing, cyclo-stationary sensing et al to obtain a clear view of the model's performance. Since our model is still in its early stages, it could further be enhanced by incorporating co-operative detection, adaptive threshold, sensing duration, as well as spread spectrum.

The channel estimation via compressive sensing (Orthogonal Matching Pursuit) and Minimum mean square error estimator (MMSE estimator) is an optimal method owing to the added benefits of Orthogonal

Matching Pursuit (OMP) and MMSE estimator. These include:

1. Lesser samples required thereby imposing low hardware constraints.

2. Moreover, it is observed that the sparse channel estimation using compressive sensing is found to perform almost as well as a traditional channel estimation technique.

3. One of the main advantages of OMP is that it is guaranteed to converge in a finite number of steps.

In channel estimation, further research could be directed towards the wavelet approach for sparse channel estimation as this is a promising field. Another potential avenue is the concept of orthogonal deficiency (od) which is a method to assess the orthogonality of the columns of a matrix as shown below:

- od(H) = 1 implies H is singular
- od(H) = 0 implies columns of H are orthogonal

67

Bibliography

[1] E. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” IEEE Trans. Inform. Theory, vol. 52, no. 2, pp. 489–509, 2006.

[2] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, “A simple proof of the restricted isometry property for random matrices,” Const. Approx., vol. 28, no. 3, pp. 253–263, Dec. 2008.

[3] Mark Davenport, Petros Boufounos, Michael Wakin, Richard Baraniuk”Signal Processing with Compressive Measurements”, Inform. Theory, vol. 36, no. 7, 2010.

[4] M. Davenport, M. Wakin, and R. Baraniuk, “Detection and estimation with compressive measurements,” Tech. Rep. TREE0610, Rice University ECE Department, 2006.

[5] Yücek Tevfik,”A survey of spectrum sensing algorithms for cognitive radio applications”, Communications Surveys & Tutorials, IEEE Volume: 11, Issue: 1 Page(s): 116 – 130,First Quarter 2009

[6] Hong, Steven;"Multi-resolution bayesian compressive sensing for cognitive radio primary user detection" ;Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE, Publication Year: 2010, Page(s): 1 - 6

[7] Cai, T. Tony, and Lie Wang. “Orthogonal Matching Pursuit for Sparse Signal Recovery with Noise.” IEEE Transactions on Information Theory 57.7 (2011): 4680–4688.

[8] Yi Zhang, Octavia A. Dobre, Ramachandran Venkatesan, and Cheng Li; ”Compressed Sensing and Its Application to Sparse Channel Estimation for Underwater Acoustic Communications “,

[9] Anirudh M. Rao, B. R. Karthikeyan , Dipayan Mazumdar , Govind R. Kadambi; ENERGY DETECTION TECHNIQUE FOR SPECTRUM. SENSING IN COGNITIVE RADIO. ; SASTECH. 73. Volume 9, Issue 1, April 2010.

[10] Ian F. Akyildiz,Won-Yeol Lee,Mehmet C. Vuran,Shantidev Mohanty; Computer Networks NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey:; Elsevier; September 2006

[11] M.A. Davenport, M.F. Duarte, Y.C. Eldar, and G. Kutyniok, "Introduction to Compressed Sensing," in Compressed Sensing: Theory and Applications, Cambridge University Press, 2012.

[12] J. A. Tropp;"Greed is good: Algorithmic results for sparse approximation";IEEE Trans. Inform. Theory, vol. 50, num. 10, pp. 2231-2242, Oct. 2004

[13] S. Kay, Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory, Prentice Hall, 1998.