ROBUST WATERMARKING OF COMPRESSIVE SENSED MEASUREMENTS UNDER NOISE ATTACKS

by

MEHMET YAMAÇ

B.S., Electrical and Electronics Engineering, Anadolu University, 2009

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Electrical and Electronics Engineering
Boğaziçi University
2014

ACKNOWLEDGEMENTS

First of all, I would like to express my deep gratitude to Prof. Bülent Sankur, my thesis supervisor for his unlimited support and impressive ideas that changed my perspective.

Çağatay Dikici, my thesis supervisor, introduced me to this new area. He worked late into the night with me.

In addition I would like to thank every member of BUSIM. We spent great time in a very friendly atmosphere. Respectively, I would like to thank Oya Çeliktutan Dikici for sharing her coffee in weekend studies, Erinç Dikici for teaching me the documentation for the legislation, and last Cem Subakan for keeping company until late at night studies.

I would like to thank my thesis committee, Assoc. Prof. Murat Saraçlar, Prof. Emin Anarim, Assoc. Prof. Mehmet Akar and Assoc. Prof. Güneş Karabulut Kurt for valuable suggestions to improve this thesis.

A special thanks to Çişel Aras, İpek Şen and Betül Duranoğlu who enlist their assistance to me in any case.

A grateful acknowledgement also goes to my friends, Artun Oyman and Can Altay for the support they provided during the preparation of this thesis.

Finally, I would like to thank my prospective partner Nergiz Solak. If Nergiz's motivation and support had not existed I would still have been in the beginning of this thesis.

ABSTRACT

ROBUST WATERMARKING OF COMPRESSIVE SENSED MEASUREMENTS UNDER NOISE ATTACKS

Traditional methods in data acquisition follow Shannon /Nyquist sampling theorem; one must sample a band-limited signal by at least two times faster than the signal bandwidth to exactly reconstruct the signal. However, signals that we encounter in many applications are sparse in some proper bases, and they can be represented by very few coefficients. Therefore, the number of samples obtained in Shannon/Nyquist framework is much more than it is required to reconstruct the signal. Compressive Sensing (CS) shows that certain signals can be captured from far fewer samples as compared to conventional methods, and they can be reconstructed by developing effective non-linear reconstruction algorithms. The methods of data embedding and data hiding (watermarking) for signal that is sampled based on Shannon /Nyquist sampling theorem is well advanced. However, watermarking methods for compressive sensing measurements is investigating in this thesis. The measurements that is sampled using CS methods, which is superiors to traditional methods, are required to be transmitted or stored. In addition to efficiently transmitting or storing CS-based measurements, one may wish to embed a watermark onto these measurements. Hence, the copyright information or meta-data can be embedded onto CS measurements. Such a watermarking scheme must satisfy the following properties: (i) the watermark information must be decoded exactly, and (ii) the reconstruction of the signal must not suffer at the decoder side. In this work, we propose a watermark algorithm that embeds information directly onto Compressive Sensed Measurements of a sparse signal. Proposed watermarking algorithm outperforms the classical ℓ_2 and ℓ_1 minimization algorithms in terms of watermark embedding capacity. Experimental results show that the proposed algorithm has a robust reconstruction performance under Gaussian and impulsive noise.

ÖZET

ŞIKIŞTIRILMIŞ ALGILANAN SİNYALLERİN GÜRBÜZ DAMGALANMASI

Shannon/Nyquist teoremine dayanan geleneksel sinyal örnekleme yöntemlerine göre bir sinyalin geriçatılabilmesi için, sinyal bant genişliğinin en az iki katı kadar örneklenmelidir. Fakat pratikte karşılaştığımız çoğu sinyal aslında bir başka alt uzayda seyrektir, diğer bir deyişle çok daha az bileşenle ifade edilir. Oysa, Shannon/Nyquist teoremine dayalı örnekleme sonucunda elde edilen örnekler, sinyali geriçatmak için gereken örneklerden çok daha fazladır. Sıkıştırılmış Algılama (Compressive Sensing: CS) seyrek sinyallerin Nyquist oranından çok daha düşük bir oranda bile örneklenebileceğini ve doğrusal olmayan geriçatma algoritmaları tasarlanarak geriçatılabileceğini göstermiştir. Shannon/Nyquist örneklemesiyle elde edilen işaretlerin içine bilgi gömme ve saklama (işaret damgalama) teknikleri çok gelişmiştir. Oysa sıkıştırılmış örneklemeli işaretler için damgalama yöntemleri henüz ele alınmaktadır. Klasik örnekleme yöntemlerinden daha etkin olan bu yeni yöntemle elde edilen sinyal kaydedilebilir ve de bir başka ortama iletilebilir. İşte bu durumda, sıkıştırılmış algılama yöntemi ile elde edilen örnekleri doğrudan damgalamak (watermarking) gerekebilir. Böylece telif hakkı bilgileri veya kullanıcıya ait üst veriler, damga bilgisi olarak CS örneklerine gömülebilir. Böyle bir damgalama yönteminde aranacak özellikler (i) damga çözücü, gürültü eklenmiş CS örneklerinden gürbüz bir şekilde damga bilgisini doğru çözebilmeli; ve de (ii) gürültülü örneklerden geriçatılan seyrek sinyal orjinal sinyale yakın olmalıdır. Bu çalışmada, doğrudan Sıkıştırılmış Algı ile örneklenen seyrek bir sinyale bilgi gömmek için bir damgalama yöntemi önerilmektedir. Önerilen damgalama yöntemi halihazırdaki l_2 ve l_1 enküçültme geriçatım algoritmalarına oranla daha fazla damga bilgisi gömme kapasitesine sahiptir. Deneysel sonuçlar, önerilen yöntemin Gauss ve dürtün gürültü altında gürbüz geriçatım başarımı yaptığını göstermiştir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS

$\mathcal{B}^n(r)$	Unit Euclidean ball of radius r
$\mathbb{E}\{\cdot\}$	Expected value
$\mathcal{N}(A)$	Null space of matrix A
$\mathcal{N}(\mu, \sigma)$	Gaussian distribution with mean μ and standard deviation σ
P_j	jth convex problem
$\Pr(a > b)$	Probability of a being higher than b
$\mathcal{R}$	Maximum achievable embedding rate
S	Discrete version of $s(t)$
$Spark(A)$	Minimum number of linearly dependent columns of A
$s(t)$	Continuous-time bandlimited signal
$\\|x\\|_{l_0^N}$	l_0 norm
y	Measurement vector
Λ	Subset of indices indexing non-zero coefficients of the sparse vector x_Λ
$\mu(U)$	Mutual Coherence
Φ	Measurement matrix
ϕ_k	kth vector of Φ
ψ	Sparcifying matrix

LIST OF ACRONYMS/ABBREVIATIONS

AWGN	Additive White Gaussian Noise
BP	Basis Pursuit
BPDN	Basis Pursuit Denoising
CS	Compressive Sensing
DCT	Discrete Cosine Transform
DHT	Discrete Hartley Transform
DS	Danzig Selector
ECG	Electrocardiography
EEG	Electroencephalography
GPRS	Gradient Projection for Sparce Reconstruction
JPEG	Joint Photographic Experts Group
JL	Johnson-Lindenstrauss
LASSO	Least Absolute Shrinkage and Selection Operator
NSP	Null Space Property
OMP	Orthogonal Matching Pursuit
PSNR	Peak Signal-to-noise Ratio
RIC	Restricted Isometry Constant
RIP	Restricted Isometry Property
SCOP	Second Order Cone Program
WBSN	Wireless Body Sensor Networks

1. INTRODUCTION

The traditional data acquisition systems make use of the well-known Shannon/ Nyquist theorem, which states that a signal must be sampled at least two times faster than the signal bandwidth in order to be exactly reconstructed. Compressive Sensing (CS) goes against the traditional data acquisition methods and claims that a sparse signal can be reconstructed with far fewer random measurements with polynomial time algorithms using ℓ_1 minimization [1–4]. Indeed, conventional sampling methods based on Shannon/Nyquist are wasteful, because we first over-sample and reduce the samples via compression methods. Many signals we encounter in real-world applications can be well-approximated by sparse ones in a proper basis. In fact, lossy compression techniques use this property and compress data before storage or transmission by keeping only largest coefficients of a signal in a proper basis. Compressive Sensing paradigm tries to combine acquisition and compression processes into one step. Mathematically speaking, a signal $x \in \mathbb{R}^N$ is k-sparse if and only if at most k entries of x are non-zero. Let $s \in \mathbb{R}^N$ be a k-sparse signal in the ψ space that is expressed as a linear combination of N orthonormal vectors that form a basis ψ, such that

$$s = \psi x, \tag{1.1}$$

where at most k of the coefficients of x are non-zero. Instead of using traditional data acquisition and compression methods as in Shannon/Nyquist sampling theorem; the data is transformed to a basis where it can be represented sparsely using $x = \psi^{\mathrm{T}} s$. Notice that the specific basis, ψ, should not to be known, so that one can sample data directly from the signal using CS as

$$y = \Phi s, \tag{1.2}$$

where $y \in \mathbb{R}^m$ is the measurement vector and Φ is an $m \times N$ measurement matrix. It is desired for CS to have $m \ll N$ so that CS methods have an advantage over the traditional Nyquist-Shannon based data acquisition.

In addition to data acquisition, the method that is used to reconstruct the original signal in compressive sensing, also differs from the procedure that is used in the Shannon/Nyquist framework [5]. In the Shannon/Nyquist framework the signal is recovered using sinc interpolation, which requires a small amount of computation. On the other hand, signal recovery in compressive sensing framework, is generally achieved by non-linear reconstruction methods that are relatively costly. In this sense, one may wish to store or transmit the measurements without reconstructing the original signal. Indeed, there have been increasing research efforts to handle the sensing problem and the reconstruction problem separately in compressive sensing framework. Recent studies [6, 7] propose a system in which measurements are obtained using compressive sensing methods and transmitted away from measurement devices. We will briefly introduce these studies in Chapter 5 as examples of possible application areas of our work.

On the other hand, one may wish to embed a hidden message about the signal within the samples. Watermarking can be defined as the process of embedding information, called watermark within the signal of interest [8]. The problem of watermarking sparse signals has recently studied by Sheikh and Baraniuk [9] in compressive sensing framework. In the watermarking part, they first achieve the sparse representation of the signal of interest by using a convenient transform (e.g. DCT for images). Then, they spread out the watermark information onto these sparse coefficients. In the decoding part they reconstruct the sparse signal by using reconstruction methods in compressive sensing framework. If we have already either the whole signal or a sparse representation of the signal of interest, this approach is a convenient strategy before storage or transmission. However, as we discussed above, some applications [6, 7] require to transmit the compressive sensing measurements without reconstruction. In such cases, data hiding onto CS measurements may be required.

In this thesis, we propose a new watermarking scheme that embeds a watermark directly onto the CS measurements. This enables embedding of watermark information while sensing. Using the proposed watermarking scheme [10], we first insert a watermark sequence, w, onto compressive sensed measurements of a sparse signal x

under Gaussian noise. This is the case when the watermarked signal is contaminated by additive white Gaussian noise (AWGN), but otherwise the channel is free of errors. Second, we extend our algorithm in [11] to handle impulsive noise, where we combat the presence of both the AWGN and channel errors in the form of impulsive noise.

This thesis is structured as follows: Chapter 2 gives a brief review of compressive sensing and discuss the stability properties of convex optimization problems in compressive sensing framework, such as ℓ_1 minimization. In Chapter 3, a detailed analysis will be done on random measurements matrices. It will be discussed that random matrices satisfy restricted isometry property, which implies stability of solution of ℓ_1 minimization problem, with high probability. Chapter 4 provides a brief overview on the sparse signal reconstruction algorithms. In Chapter 5, we present our watermarking scheme and reconstruction method in the presence of (i) additive white Gaussian noise, and (ii) sum of an unbounded impulsive noise and an additive white Gaussian noise. In same chapter, we will also give the guarantee condition of the proposed watermarking and reconstruction scheme. In Chapter 6, we will propose a fast embedding scheme in addition to a fast joint signal reconstruction and watermark recovery algorithm. In this section, a subset of rows of the noiselets basis will be used as measurement matrix, and embedding processes will also use an orthonormal basis to make the watermarking scheme that is proposed in Chapter 5, more convenient fot real-time, large scale applications.

2. COMPRESSIVE SENSING

Nyquist-Shannon sampling theorem states that a continuous-time bandlimited signal $s(t)$ can be represented by N uniformly spaced samples taken at least two times faster than the signal bandwidth. Let $S \in \mathbb{R}^N$ be the discrete version of $s(t)$, then compressive sensing (CS) seeks for a way to take samples which are obtained using linear mapping

$$y_k = \langle \phi_k, s \rangle \tag{2.1}$$

over the continuous time signal $s(t)$, and to approximate $S \in \mathbb{R}^N$ by recovering from these $m << N$ measurements $\{y_k\}_{k=1}^{m}$.

We would like to mention that while it is possible to develop CS theory in continuum setting [12], we will hereafter consider our signal in discrete time, since most of the works in CS have been developed in discrete setting (assuming that $s(t) := S \in \mathbb{R}^N$). The linear mapping in (2.1) can be represented by an $m \times N$ matrix $\Phi = [\phi_1, \phi_2, .., \phi_m]^T$, where $\{\phi_k\}_{k=1}^{m}$ are N-long vectors, resulting in

$$y = \Phi S. \tag{2.2}$$

This is obviously an underdetermined system of linear equations. Since we have fewer equations than the number of unknown parameter to be estimated, this system has either no solution or infinitely many solution. Under the assumption that Φ is full row rank, a solution exists, but not unique. In order to achieve a unique solution, some prior information is needed. One formulation of this is the least-norm solution, where the objective functional $\|S\|$ is desired to be small [13] when the optimization problem can be defined as follows

$$(P_j) : \min_S \ \ \|S\| \ \text{ subject to } \Phi S = y, \tag{2.3}$$

where $\|.\|$ is the norm. If the objective is chosen as Euclidean or ℓ_2-norm, Equation (2.3) turns to be

$$(P_2) : \min_{S} \ \ \|S\|_{\ell_2^N}^2 \text{ subject to } \Phi S = y. \tag{2.4}$$

Introducing Lagrange multipliers λ, we define auxiliary function

$$\Upsilon(S) = \|S\|_{\ell_2^N}^2 + \lambda^T(\Phi S - y). \tag{2.5}$$

By solving $\frac{d\Upsilon(S)}{dS} = 0$, we get

$$S^* = -\frac{1}{2}\Phi^T\lambda. \tag{2.6}$$

Plugging S^* into the constraint $\Phi S = y$ gives us $\lambda^* = -2(\Phi\Phi^T)^{-1}y$. Then, the optimal solution S^* of (P_2) can be found via

$$S^* = \Phi^{\mathrm{T}}(\Phi\Phi^{\mathrm{T}})^{-1}y \tag{2.7}$$

under the assumption that $\mathrm{rank}(\Phi) = m < N$ so that the matrix $\Phi\Phi^T$ is invertible.

2.1. Compressible Signals, Sparsity and Compressed Sensing

A great majority of the signals that we encounter in practical applications exhibit a rapid decay when expressed in an appropriate basis. In other words, let $S \in \mathbb{R}^N$ be our signal of interest and Ψ be a basis so that S has a unique representation in that basis such that $S = \Psi x$. Let us rearrange the elements of the vector x in descending order of magnitude i.e, $|x|^{(1)} \geq |x|^{(2)} \geq ... \geq |x|^{(N)}$. If this basis Ψ is properly chosen according to the signal of interest, $S \in \mathbb{R}^N$, then these sorted magnitudes approach zero very fast like a power decay [14]. For instance, consider the wavelet transform, most coefficients of a typical natural image in wavelet basis are negligibly small, as shown in Figure 2.1. In fact this is the idea that gives birth to most of the lossy compression techniques such as JPEG [15] , JPEG 2000 [16], etc. Many transform coding based compression techniques like JPEG keep only large coefficients which constitute of most of the energy of the signal, and get rid of small ones. Although these techniques are widespread and standard in applications, one may need to look beyond the Nyquist-Shannon scheme in niche applications. This brings us to the following question: Is it possible to compress the signals at the beginning ?

(a) (b)

Figure 2.1. Image as a typical compressible signal in the wavelet domain. (a) Original image. (b) Wavelet coefficients. Small coefficients are represented by dark pixels.

One strategy could be to choose the rows of the measurement matrix in the domain in which we can compress our signal of interest [17]. For example, consider our signal $S \in \mathbb{R}^N$ to be a vectorized image. The approach can be to pick the row vectors ϕ_k of measurement matrix Φ as cosine transform vectors starting from the most energetic, usually low spatial frequencies [7, 18]. Then, one can reconstruct vectorized image S by using linear reconstruction (2.7). By capturing m of important DCT coefficients, one can decrease the number of measurements, m that is required to reconstruct S reliably. However, there is a drawback of this method. It always projects the image onto a fixed subspace and it is not adaptive to changes in the signal structure [7, 18].

In the light of this observation, we seek an effective measurement system which is adaptive to changes in signal statistics. However, changing the measurement matrix Φ for every signal S would be very costly. As consequence, the measurement matrix Φ should be

(i) non-adaptive in the sense that it should not depend on the signal S,

(ii) flexible in the sense that it must be able to capture important information without knowledge of signal structure. The answer to the question of whether it is possible to design such a measurement system, was given by Donoho [19] and Candes *et al.* [20] leading to the the Compressive Sensing idea. To satisfy (i), CS measurement procedure exploits a fixed measurement matrix Φ that is independent of the signal $S \in \mathbb{R}^N$. Then, measurements are directly and linearly sensed using this one and only one measurement matrix Φ, $y = \Phi S$. We also have to assume that S has a unique representation in a suitable basis such that $S = \Psi x$. Now, we can re-formulate measurements as follows

$$y = \Phi S = \Phi \Psi x = Ax, \tag{2.8}$$

where $A = \Phi\Psi$. If Φ has certain properties, which will be discussed later, A may have the ability to capture the important information without knowledge of which coefficients of the vector x is active. Then, it also satisfies (ii). Before discussing reconstruction procedure, let us give the formal definition of sparse signals;

Definition 2.1 (k-Sparse Signal). *Let $x \in \mathbb{R}^N$ be our signal of interest. This signal x is entitled as k-sparse if it has at most k non-zero coefficients, i.e., $\|x\|_{\ell_0^N} \leq k$, where*

$$\|x\|_{\ell_0^N} = \lim_{p \to 0} \sum_{i=1}^{N} |x_i|^p = \#\{j : x_j \neq 0\}. \tag{2.9}$$

Assuming that $\|x\|_{\ell_0^N} \leq k$, the signal $S \in \mathbb{R}^N$ also called k-sparse if

$$S = \Psi x, \tag{2.10}$$

where Ψ is sparsifying basis that is typically orthonormal. For convenience we will denote the set of all k-sparse signals in $\mathbb{R}^N$ by

$$\Sigma_k = \{x : \|x\|_{\ell_0^N} \leq k\}. \tag{2.11}$$

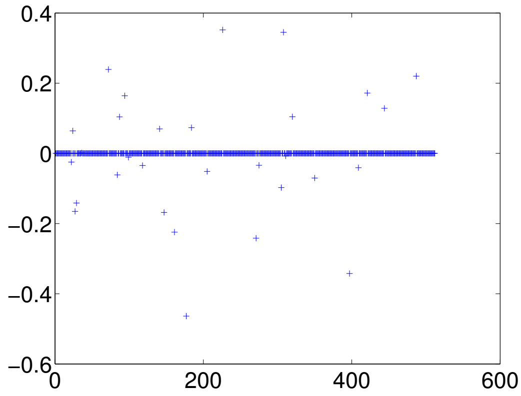


Figure 2.2. An example of k-sparse signal, $k = 30$, the rest of $500 - 30 = 470$ samples are all zero, which, when plotted gives the appearance of a blue line.

In this study, we will mostly assume that the signal of interest has a unique and strictly sparse representation in a proper basis, $S = \Psi x$ such that $\|x\|_{\ell_0^N} \leq k$. From this prior information, it is easy to say that ℓ_2-norm minimization (2.7) can never give the desired sparse solution of x, since the penalty term is on the energy, but not on sparsity. To achieve the sparsest solution $\hat{x}$ of x, we can determine the objective as $\|x\|_{\ell_0^N}$, and define the optimization problem as follows

$$(P_0) : \min_x \ \|x\|^2_{\ell_0^N} \text{ subject to } Ax = y. \tag{2.12}$$

Since ℓ_0-norm minimization requires combinatorial search, P_0 can be relaxed to P_1 [21] as follows

$$(P_1) : \min_x \ \|x\|^2_{\ell_1^N} \text{ subject to } Ax = y. \tag{2.13}$$

2.2. Uniqueness of solution for (P_0) and (P_1)

Let x_i, $x_j \in \Sigma_k$ be two distinct vectors. By observing that if $Ax_i = Ax_j$ then, $A(x_i - x_j) = 0$ with $x_i - x_j \in \Sigma_{2k}$, then we cannot distinguish x_i from x_j by any method. Consequently, we can say that if we wish any k sparse solution $\hat{x}$ to be unique, $\mathcal{N}(A)$

must not have any vector in Σ_{2k}, where $\mathcal{N}(A)$ denotes the Null space of matrix A,

$$\mathcal{N}(A) = \{z : Az = 0\}. \tag{2.14}$$

This property was formally defined by Donoho-Elad [22] in the concept of spark of a matrix.

Theorem 2.2 ([22]). *Let A be an $m \times N$ matrix. The solution of P_0 is guaranteed to be unique if*

$$2k < \mathit{Spark(A)}, \tag{2.15}$$

where Spark(A) is minimum number of linearly dependent columns of A.

This implies the requirements $m \geq 2k$, since $\mathit{spark}(A) \in [2, m+1]$.

Now we wish also the solution of P_1 to be unique. Before analysing the uniqueness conditions of P_1, let us give following definition that we will use from now on. Let $x \in \mathbb{R}^N$ be an arbitrary vector. We define $\Lambda \subset \{1, 2, 3, ..., N\}$ is a subset of indices indexing non-zero coefficients of the sparse vector x_Λ and $|\Lambda|$ denotes number of these non-zero coefficients, where $|.|$ denote the cardinality. The complement of set Λ is also defined as $\Lambda^c = \{1, 2, 3, ..., N\} \setminus \Lambda$. In other words, when we define $x_\Lambda \in \mathbb{R}^N$, we mean x_Λ is a $|\Lambda|$-sparse vector and non-zero coefficients of this vector are indexed by the set Λ.

Let us assume that we have a $|\Lambda|$-sparse vector x whose non-zero coefficients are indexed by set Λ. Given an arbitrary vector $z \in \mathbb{R}^N$, $z \neq x$, satisfying $Ax = Az$. If we wish the solution x of P_1 to be unique, then the condition $\|x\|_{\ell_1^N} \leq \|z\|_{\ell_1^N}$ should be satisfied. Since $Ax = Az$, one can define a vector $h : x - z \in \mathcal{N}(A)$. If $\|h_\Lambda\|_{\ell_1^N} \leq \|h_{\Lambda^c}\|_{\ell_1^N}$

then we can get

$$\|x\|_{\ell_1^N} \le \|x - z_\Lambda\|_{\ell_1^N} + \|h_\Lambda\|_{\ell_1^N} = \|h_\Lambda\|_{\ell_1^N} + \|z_\Lambda\|_{\ell_1^N} \tag{2.16}$$

$$\le \|h_{\Lambda^c}\|_{\ell_1^N} + \|z_\Lambda\|_{\ell_1^N} = \|-z_{\Lambda^c}\|_{\ell_1^N} + \|z_\Lambda\|_{\ell_1^N} \tag{2.17}$$

$$\le \|z\|_{\ell_1^N} \tag{2.18}$$

which satisfies $\|x\|_{\ell_1^N} \le \|z\|_{\ell_1^N}$. Indeed, this simplified version of the proof of the Null Space Property [23] is reproduced to give an intuition on the Null Space Property. A complete proof is in [23]. Now let us give more formal definition for the Null Space Property and uniqueness condition of P_1;

Definition 2.3 (Null Space Property)**.** *Let A be a $m \times N$ matrix. A satisfies the Null Space Property (NSP) of order k, if*

$$\|h_\Lambda\|_{\ell_1^N} \le \|h_{\Lambda^c}\|_{\ell_1^N} \tag{2.19}$$

holds for all $h \in \mathcal{N}(A)$, where $|\Lambda| \le k$.

In plain English, we mentioned that $\mathcal{N}(A)$ should not have vectors which are sparse. Now, Null Space Property implies that $\mathcal{N}(A)$ should also not have too compressible vectors [5].

Theorem 2.4 ([24])**.** *If an $m \times N$ matrix A satisfies the NSP of order k, then the solution $\hat{x}$ of P_1 is unique for all $x \in \mathbb{R}^N$ with $x \in \Sigma_k$.*

2.3. Stability of Solution for (P_1)

We have presented the conditions under which the ℓ_1 minimization is exact in noise free and exactly sparse case if the measurement matrix has NSP. In this section we will discuss the condition in which the signal is corrupted by a noise. We start by defining the convex optimization problem in the presence of noise:

Let us define the bounded noise as z with $\|z\|_{\ell_2^m} \le \epsilon$. Suppose that measurements are erroneous such that $y = Ax + z$. Then a k-sparse signal $\hat{x}$ can be estimated solving the following optimization problem

$$(P_1^{\epsilon}) : \min_{x} \ \|x\|_{\ell_1^N} \quad \text{s.t.} \quad \|y - Ax\|_{\ell_2^m} \leq \epsilon. \tag{2.20}$$

However there are some works that use stronger NSP to prove stability of the solution in the presence of noise [25], we will discuss a more common used property, Restricted Isometry Property. Restricted Isometry Property implies stability of (P_1^{ϵ}). It also implies stability when we deal with approximately sparse signals, i.e, compressible signals. We mean by stability that a small change in the measurement vector should not lead to a substantial change in the recovered signal. Mathematically speaking, let perturbed measurements be form of $y = Ax + z_{\epsilon}$, where z_{ϵ} is a perturbation. Then, it is expected from a stable solution $\hat{x}$ of the true signal x to obey

$$\|x - \hat{x}\|_{\ell_2^N} \leq \kappa \|z_{\epsilon}\|_{\ell_2^N} \tag{2.21}$$

with a small positive constant κ .

Definition 2.5 (Restricted Isometry Property)**.** *Let A be an $m \times N$ matrix and let $\delta_k \in (0, 1)$ be the smallest quantity such that*

$$(1 - \delta_k) \|x\|_{\ell_2^N}^2 \leq \|Ax\|_{\ell_2^m}^2 \leq (1 + \delta_k) \|x\|_{\ell_2^N}^2 \quad \textit{for all } x_k \in \Sigma_k \tag{2.22}$$

then the matrix A satisfies the Restricted Isometry Property (RIP) of order of k with the restricted isometry constant δ_k.

The idea is that RIP requires all $m \times k$ submatrices of A being close to orthonormal matrices. In the literature much effort has been given to find sufficient conditions of stability including restricted isometry constant. Elad [26] has discussed the stability properties of solution to (P_0). However, we will only consider stability condition of the recovery in (P_1^{ϵ}) from now on. In early works of Candes *et al.* [27], it is shown that the solution $\hat{x}$ of (P_1^ϵ) is unique as soon as $\delta_{3k}+3\delta_{4k} < 2$. Later Candes [28] has showed that stable recovery of k-sparse signals is guaranteed with $\delta_{2k} < \sqrt{2}-1$. Notwithstanding these results, efforts to improve this sufficient condition [29] continue.

The following stability condition for (P_1^ϵ), a consequence of the RIP condition, will be useful in the sequel.

Theorem 2.6 ([28]). *Assume that an $m \times N$ matrix A holds RIP of order k with $\delta_{2k} < \sqrt{2}-1$. Suppose also that signal of interest x is strictly sparse. Then, the solution $\hat{x}$ of P_1^ϵ obeys*

$$\|x-\hat{x}\|_{\ell_1^N} \leq 4\frac{\sqrt{1+\delta_{2k}}}{1-(1+\sqrt{2})\delta_{2k}}\epsilon. \tag{2.23}$$

In the case the measurements $y = Ax$ are corrupted by Additive White Gaussian Noise (AWGN) rather than bounded noise, then, Theorem 2.6 enables the following lemma.

Lemma 2.7. *Suppose that an $m \times N$ measurement matrix A satisfies the RIP of order $2k$ with $\delta_{2k} < \sqrt{2}-1$. Suppose also measurements are the form $y = Ax + z$ where z is the noise pattern with i.i.d elements z_i drawn from $\sim \mathcal{N}(0, \sigma_G^2)$. Then when $\epsilon = (1+\gamma)\sqrt{m}\sigma_G$, the solution $\hat{x}$ of (P_1^ϵ) obeys*

$$\|x-\hat{x}\|_{\ell_2^N} \leq 4\frac{\sqrt{1+\delta_{2k}}}{1-(1+\sqrt{2})\delta_{2k}}(1+\gamma)\sqrt{m}\sigma_G \tag{2.24}$$

with probability at least $1-e^{-\frac{3m}{4}\gamma^2}$ where $0 < \gamma < 1$.

Proof. Let $z_1, ...z_m$ be i.i.d Gaussian random variables with zero mean and variance σ_G^2. If one shows that

$$\Pr(\|z\|_{\ell_2^m} \geq (1+\gamma)\sqrt{m}\sigma_G) \leq e^{-3/4m\gamma^2} \tag{2.25}$$

holds for every $z = (z_1, ..., z_M) \in \mathcal{R}^m$ with a given $0 < \gamma < 1$, then Lemma 2.7 can be obtained applying Theorem 2.6. Therefore, it is sufficient to prove (2.25) holds. To begin with, it is obvious that

$$\Pr(\|z\|_{\ell_2^m} \geq (1+\gamma)\sqrt{m}\sigma_G) = \Pr(\|z\|^2_{\ell_2^m} \geq (1+\gamma)^2 m\sigma_G^2). \tag{2.26}$$

Let us define $W = \|z\|^2_{\ell_2^m} = z_1^2 + z_2^2 + ... + z_m^2$. For simplicity let us define $\alpha = (1+\epsilon)^2 m\sigma_G^2$ and $\beta = \sigma_G^2$. Now, the aim is to find a bound on probability that the positive random variable, W exceeds the given value, α. It is easy to see that $z_1^2, z_2^2, ..., z_m^2$ are independent random variables. Therefore, we can conclude that the standard Chernoff-style method is applicable. The Markov inequality yields

$$\Pr(W \geq \alpha) = \Pr(e^{tW} \geq e^{t\alpha}) \leq \frac{\mathbb{E}(e^{tW})}{e^{t\alpha}}, \tag{2.27}$$

for any $t > 0$. By using independence, we obtain

$$\mathbb{E}(e^{tW}) = \mathbb{E}(e^{t\sum_{i=1}^{m} y_i^2}) = \prod_{i=1}^{m} \mathbb{E}(e^{ty_i^2}), \tag{2.28}$$

where

$$\mathbb{E}(e^{ty_i^2}) = \frac{1}{\sqrt{2\pi\beta}} \int_{-\infty}^{\infty} e^{t\xi^2} e^{-\frac{1}{2}\frac{\xi^2}{\beta}}\, \mathrm{d}\xi = \frac{1}{\sqrt{2\pi\beta}} \int_{-\infty}^{\infty} e^{-\xi^2(\frac{1}{2\beta}-t)}\, \mathrm{d}\xi. \tag{2.29}$$

Let $z = \xi\sqrt{\frac{1}{\beta} - 2t}$, then we have

$$\mathbb{E}(e^{ty_i^2}) = \frac{1}{\sqrt{2\pi}\sqrt{1-2t\beta}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2}\, \mathrm{d}z = \frac{1}{\sqrt{1-2t\beta}}. \tag{2.30}$$

By using (2.28), (2.29), (2.30), we obtain

$$\Pr(e^{tW} \geq e^{t\alpha}) \leq e^{-t\alpha}(1-2t\beta)^{-m/2}. \tag{2.31}$$

By minimizing the right-hand side of the equation with respect to t, we get $t = (\frac{1}{2\beta} - \frac{m}{2\alpha})$.

Then Equation (2.31) turns to

$$\Pr(e^{tW} \geq e^{t\alpha}) \leq e^{-t\alpha}(1-2t\beta)^{-m/2} = (\frac{m\beta}{\alpha})^{-m/2} e^{-(\frac{1}{2\beta}-\frac{m}{2\alpha})\alpha}. \tag{2.32}$$

When we plugging $\alpha = (1+\gamma)^2 m\sigma_G^2$ and $\beta = \sigma_G^2$, we get

$$e^{-\frac{m}{2}((1+\gamma)^2-1)} e^{-\frac{m}{2}\log\frac{1}{(1+\gamma)^2}} = e^{-\frac{m}{2}(2\gamma+\gamma^2-2\log(1+\gamma))} \leq e^{-\frac{m}{2}(3\gamma^2/2)}, \tag{2.33}$$

where we have used the inequality $\log(1+a) \leq a - a^2/4$, which is valid for all $0 \leq a \leq 1$. □

2.4. Matrices that satisfies the RIP

Random matrices satisfy the RIP of order of k with high probability. A simple proof of the RIP for the random matrices was given by Baraniuk *et al.* [30]. Technique they used to prove RIP for random matrices partly comes from famous Johnson-Lindenstrauss Lemma [31]. Given a set D of points in $\mathbb{R}^N$. Johnson-Lindenstrauss Lemma states that one can find a map that embeds these points into a lower-dimensional Euclidean space $\mathbb{R}^m$ while preserving relative distance of any of two of these points. The key ideas and Johnson-Lindenstauss Lemma will be investigated in details in following chapter.

3. RESTRICTED ISOMETRY PROPERTY (RIP)

In this chapter, we will discuss that the random matrices satisfy the Restricted Isometry Property. We will start by presenting the Johnson-Lindenstrauss Lemma which is directly related to the Restricted Isometry Property. We will use the same methodology introduced by Baraniuk *et al.* [30] to prove that random matrices satisfies the RIP. Although we use the key ideas presented in [30], we will reproduce the results by giving detailed explanations. These slightly new results will be used to prove the stability condition of our proposed algorithm, which will be introduced in Chapter 5.

3.1. The Johnson-Lindenstrauss (JL) Lemma

Johnson-Lindenstrauss is a valuable formulation that is used to provide a low-dimensional representation of the data. Low-dimensional embedding is crucial in many applications such as machine learning, manifold learning and compressive sensing. In particularly, dimensional reduction can increase the speed of most algorithms that are used in machine learning. Besides, in compressive sensing framework, working in low-dimensional space without much distortion between pairwise distance is a must.

Given a set D of points in $\mathbb{R}^N$, we wish to construct an embedding scheme that embeds these points into a lower-dimensional Euclidean space $\mathbb{R}^m$ while preserving relative distances of pair of these points. The original formulation of Johnson-Lindestrauss Lemma is as follows:

Lemma 3.1 (Johnson-Lindenstrauss [31])**.** *For any $0 < \epsilon < 1$ and every set of D of n points in $\mathbb{R}^N$, if m is a positive integer such that $m \geq m_0 = O(\epsilon^{-2} \log n)$, there exists a Lipschitz map $f : \mathbb{R}^N \to \mathbb{R}^m$ such that*

$$(1 - \epsilon) \|u - v\|^2_{\ell_2^N} \leq \|f(u) - f(v)\|^2_{\ell_2^m} \leq (1 + \epsilon) \|u - v\|^2_{\ell_2^N} \tag{3.1}$$

for all $u, v \in D$.

Up to now, much effort has been spent to improve the bound given in the Johnson-Lindestrauss Lemma in (3.1) for a given size of m. Matousek's work [32] contains a brief survey about the extensions and proofs of Johnson-Lindenstrauss Lemma. A simplified version of the original proof of Johnson-Lindestrauss Lemma was given by Frankl and Meahara [33], where the size of m is also tightened and given as $m \geq m_0 = [9(\epsilon^2 - 2\epsilon/3)^{-1} \log n] + 1$. Their proof is based on showing that the squared length of the projection of a fixed unit vector on a random m-dimensional subspace is concentrated around $\frac{m}{N}$. Afterwards, Indyk and Motwani [34] used Gaussian random projections and some elementary probabilistic techniques to improve the m_0 constant. Achlioptas [35] showed that the same bound as $m \geq 4(\epsilon^2/2 - \epsilon^3/3)^{-1} \log n$ as it can be achieved not only with a Gaussian measurement matrix, but also with a matrix consisting of random $\{-1, +1\}$ elements. Dasgupta and Gupta [36] gave an alternate proof simpler than that of Indyk and Motwani [34] and of Achlioptas [35]. They presented the same bounds with those works:

Theorem 3.2 (Dasgupta-Gupta [36]). *For any* $0 < \epsilon < 1$ *and every set of* D *of* n *points in* $\mathbb{R}^N$*, if* m *is a positive integer such that*

$$m \geq 4(\epsilon^2/2 - \epsilon^3/3)^{-1} \log n, \tag{3.2}$$

there exists a Lipschitz map $f : \mathbb{R}^N \to \mathbb{R}^m$ *such that*

$$(1-\epsilon)\left\|u-v\right\|_{\ell_2^N}^2 \leq \left\|f(u)-f(v)\right\|_{\ell_2^m}^2 \leq (1+\epsilon)\left\|u-v\right\|_{\ell_2^N}^2$$

for all $u, v \in D$.

The approach common to all these proof of a JL-embedding is as follows:

(i) Choose a pair $u, v \in D$ and prove that squared norm of the projection of the vector $u - v$ is concentrated around its expected value with high probability. Define the probability that the squared norm of the projection of the vector does not satisfy (3.2) as failure probability, p_e.

(ii) Apply union bound for $\binom{n}{2}$ pairs $u, v \in D$ and prove that the probability of getting a JL-embedding which is bounded by $1 - \binom{n}{2}p_e$ is not zero.

3.2. Concentration of Measure Inequality for Random Matrices

Let the mapping f be a linear mapping, then it is convenient to use $m \times N$ matrix ϕ to represent this mapping. For any arbitrary variable $x \in \mathbb{R}^N$, expected value of random variable $\|\phi x\|^2_{\ell_2^m}$ can be calculated via

$$\mathbb{E}(\|\phi x\|^2_{\ell_2^m}) = \|x\|^2_{\ell_2^N} , \tag{3.3}$$

where $var(\phi_{i,j}) = \frac{1}{m}$.

Proof. We can express $\|\phi x\|^2_{\ell_2^m}$ as $\|\phi x\|^2_{\ell_2^m} = \sum_{i=1}^{m} < \phi_i, x >^2$ where ϕ_i denote the rows of the matrix ϕ. Then

$$\begin{aligned}\mathbb{E}(\|\phi x\|^2_{\ell_2^m}) &= \mathbb{E}(\sum_{i=1}^{m} < \phi_i, x >^2) = \sum_{i=1}^{m} \mathbb{E}(< \phi_i, x >^2) = \sum_{i=1}^{m} \mathbb{E}(\sum_{j=1}^{N} x_j^2 \phi_{i,j}^2) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{N} x_j^2 \mathbb{E}(\phi_{i,j}^2) = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{N} x_j^2 = \frac{1}{m}\sum_{i=1}^{m} \|x\|^2_{\ell_2^N} = \|x\|^2_{\ell_2^N} .\end{aligned} \tag{3.4}$$

□

Lemma 3.3. *Given any arbitrary vector $x \in \mathbb{R}^N$. Let ϕ be an $m \times N$ matrix whose entries are independently drawn from Gaussian distribution $\phi_{i,j} \sim \mathcal{N}(0, \frac{1}{m})$. Then,*

$$Pr(\|\phi x\|^2_{\ell_2^m} \geq (1+\epsilon)\,\|x\|^2_{\ell_2^N}) \leq \exp(-\frac{m}{2}(\epsilon^2/2 - \epsilon^3/3)), \tag{3.5}$$

and

$$Pr(\|\phi x\|^2_{\ell_2^m} \leq (1-\epsilon)\,\|x\|^2_{\ell_2^N}) \leq \exp(-\frac{m}{4}\epsilon^2). \tag{3.6}$$

Proof. To begin with, it is easy to show that $\mathbb{E}(\|\phi x\|_{\ell_2^m}^2) = \|x\|_{\ell_2^N}^2$ was shown in 3.3. Define $y_i =< \phi_i, x >$ and $W = \|\phi x\|_{\ell_2^m}^2 = y_1^2 + y_2^2 + y_3^2 .. + y_m^2$ where ϕ_i denote the rows of the matrix ϕ. Under the assumption that ϕ has orthogonal rows, it is obvious that y_i has distribution $\mathcal{N}(0, \frac{\|x\|_{\ell_2^N}^2}{m})$. For simplicity, define $\alpha = (1+\epsilon)\,\|x\|_{\ell_2^N}^2$ and $\beta = \frac{\|x\|_{\ell_2^N}^2}{m}$. Then, we will make use of proof of the upper bound and the Chernoff bound to prove upper tail: Since

$$\Pr(W \geq \alpha) = \Pr(e^{tW} \geq e^{t\alpha}), \tag{3.7}$$

for any $t > 0$, using Markov inequality, we obtain

$$\Pr(e^{tW} \geq e^{t\alpha}) \leq \frac{\mathbb{E}(e^{tW})}{e^{t\alpha}}. \tag{3.8}$$

Considering the independence of y_i's, we have

$$\mathbb{E}(e^{tW}) = \mathbb{E}(e^{t\sum\limits_{i=1}^{m} y_i^2}) = \prod_{i=1}^{m} \mathbb{E}(e^{ty_i^2}), \tag{3.9}$$

where

$$\mathbb{E}(e^{ty_i^2}) = \frac{1}{\sqrt{2\pi\beta}} \int_{-\infty}^{\infty} e^{t\xi^2} e^{-\frac{1}{2}\frac{\xi^2}{\beta}}\, \mathrm{d}\xi = \frac{1}{\sqrt{2\pi\beta}} \int_{-\infty}^{\infty} e^{-\xi^2(\frac{1}{2\beta}-t)}\, \mathrm{d}\xi. \tag{3.10}$$

Define $z = \xi\sqrt{\frac{1}{\beta} - 2t}$, then (3.10) becomes

$$\mathbb{E}(e^{ty_i^2}) = \frac{1}{\sqrt{2\pi}\sqrt{1-2t\beta}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2}\, \mathrm{d}z = \frac{1}{\sqrt{1-2t\beta}}. \tag{3.11}$$

Rearranging (3.8), (3.9), (3.11) leads to

$$\Pr(e^{tW} \geq e^{t\alpha}) \leq e^{-t\alpha}(1-2t\beta)^{-m/2}. \tag{3.12}$$

By minimizing the right-hand side of the equation with respect to t, we get $t = (\frac{1}{2\beta} - \frac{m}{2\alpha})$.

Then, Equation (3.12) turns to be

$$\Pr(e^{tW} \geq e^{t\alpha}) \leq (1-2t\beta)^{-m/2} e^{-t\alpha} = (\frac{m\beta}{\alpha})^{-m/2} e^{-(\frac{1}{2\beta}-\frac{m}{2\alpha})\alpha}. \tag{3.13}$$

When we plug in $\alpha = (1+\epsilon)\left\|x\right\|_{\ell_2^N}^2$ and $\beta = \frac{\|x\|_{\ell_2^N}^2}{m}$, we get

$$= e^{-\frac{m}{2}\log\frac{1}{(1+\epsilon)}} e^{-\frac{m}{2}\epsilon} = e^{-\frac{m}{2}(\epsilon - \log(1+\epsilon))} \leq e^{-\frac{m}{2}(\epsilon^2/2 - \epsilon^3/3)}, \tag{3.14}$$

where we have used the inequality $\log(1+a) \leq a - a^2/2 + a^3/3$, which is valid for all $a \geq 0$.

Now we will give the proof of the lower bound: Define $\lambda = (1-\epsilon)\left\|x\right\|_{\ell_2^N}^2$, then

$$\begin{aligned} \Pr(W \leq \lambda) &= \Pr(\lambda - W \geq 0) \\ &= \Pr(e^{-tW} e^{t\lambda} \geq 1), \end{aligned} \tag{3.15}$$

for $t > 0$. Then by using Markov's inequality we obtain

$$= \Pr(e^{-tW} e^{t\lambda} \geq 1) \leq \mathbb{E}(e^{-tW}) e^{t\lambda}. \tag{3.16}$$

If we use $-t$ instead of t in (3.9) and using (3.11), we get

$$\mathbb{E}(e^{-tW}) = \mathbb{E}(e^{-t\sum_{i=1}^{m} y_i^2}) = \prod_{i=1}^{m} \mathbb{E}(e^{-ty_i^2}) = (1+2t\beta)^{-m/2}. \tag{3.17}$$

Then we have

$$\Pr(W \leq \lambda) \leq e^{t\lambda}(1+2t\beta)^{-m/2}. \tag{3.18}$$

Optimizing in t we get $t = (\frac{m}{2\lambda} - \frac{1}{2\beta})$. Then (3.18) turns to be

$$\Pr(W \leq \lambda) \leq (1 + 2t\beta)^{-m/2} e^{t\lambda} = (\frac{m\beta}{\lambda})^{-m/2} e^{(\frac{m}{2\lambda} - \frac{1}{2\beta})\lambda}. \tag{3.19}$$

When we plugging $\lambda = (1-\epsilon) \left\| x \right\|_{\ell_2^N}^2$ and $\beta = \frac{\left\| x \right\|_{\ell_2^N}^2}{m}$, we get

$$= e^{\frac{m}{2}\epsilon} e^{-\frac{m}{2}\log\frac{1}{(1-\epsilon)}} = e^{-\frac{m}{2}(-\epsilon - \log(1-\epsilon))} \leq e^{-\frac{m\epsilon^2}{4}}, \tag{3.20}$$

where we have used the inequality $\log(1-a) \leq -a - a^2/2$, which is valid for all $a \in [0, 1)$. □

Lemma 3.4. *Given an arbitrary vector $x \in \mathbb{R}^N$, let ϕ be an $m \times N$ matrix whose entries are drawn from any distribution satisfying equality given in* (3.3)*. Then,*

$$Pr((1-\epsilon) \left\| x \right\|_{\ell_2^N}^2 \leq \left\| \phi x \right\|_{\ell_2^m}^2 \leq (1+\epsilon) \left\| x \right\|_{\ell_2^N}^2) \leq 1 - 2\exp(-mC(\epsilon)) \tag{3.21}$$

for $0 < \epsilon$, where $C(\epsilon)$ is positive constant depending on ϵ and distribution of entries of matrix ϕ.

For instance it is easy to see, from Lemma 3.3, for a random matrix ϕ whose entries are independently drawn from Gaussian distribution with $\phi_{i,j} \sim \mathcal{N}(0, \frac{1}{m})$, that we have $C(\epsilon) = \epsilon^2/4 - \epsilon^3/6$. In the literature much effort has been spent to find $C(\epsilon)$. As we mentioned before, Achlioptas [35] has proved that we also get $C(\epsilon) = \epsilon^2/4 - \epsilon^3/6$ for the following distributions:

$$\phi_{i,j} = \begin{cases} \frac{+1}{\sqrt{m}} & \text{with probability } 1/2, \\ \frac{-1}{\sqrt{m}} & \text{with probability } 1/2, \end{cases} \tag{3.22}$$

$$\phi_{i,j} = \sqrt{3} \times \begin{cases} \frac{+1}{\sqrt{m}} & \text{with probability } 1/6, \\ 0 & \text{with probability } 2/3, \\ \frac{-1}{\sqrt{m}} & \text{with probability } 1/6. \end{cases} \tag{3.23}$$

3.3. RIP via JL Lemma

In this section, we will show that Johnson-Lindenstrauss Lemma implies that Restricted Isometry Property which is sufficient property for reconstruction of compressed sensed signals. Let us recall the Restricted Isometry Property:

Definition: Define a matrix ϕ that satisfies equality given in (3.3). Then, the matrix ϕ satisfies RIP of order k if there exists a $\delta_k \in (0,1)$ such that

$$(1-\delta_k)\|x\|_2^2 \leq \|\phi x\|_2^2 \leq (1+\delta_k)\|x\|_2^2 \tag{3.24}$$

holds for all $x \in \Sigma_k$ where $\Sigma_k = \{x : \|x\|_0 \leq k\}$. Following lemma and theorem will be proven based on work of Baraniuk *et al.* [30].

Lemma 3.5. *Given $0 < \delta < 1$. Define the set of χ_Λ of unit norm vectors for fixed $|\Lambda|$, where $\Lambda \subset \{1, 2, 3, ..., N\}$ is a subset of indices indexing non-zero coefficients of the sparse vector x and $|\Lambda|$ denotes number of these non-zero coefficients. Let ϕ be the $m \times N$ measurement matrix that satisfies the inequality given in* (3.21)*, then we have*

$$(1-\delta)\,\|x\|_{\ell_2^N}^2 \leq \|\phi x\|_{\ell_2^m}^2 \leq (1+\delta)\,\|x\|_{\ell_2^N}^2 \textit{ for every } x \in \chi_\Lambda \tag{3.25}$$

for any set Λ such that $|\Lambda| \leq k$ with probability at least

$$1 - 2\left(\frac{12}{\sqrt{\delta}}\right)^k e^{-C(\delta/4)m}. \tag{3.26}$$

This lemma implies that for a fixed set of k-sparse signals, the defined measurement matrix A satisfies the RIP with probability in (3.26). We will give the proof of Lemma 3.5 which is the last step before achieving the final result that implies that defined measurement matrix, A, satisfies the RIP of order k for all $x \in \Sigma_k$. Before we prove this lemma, let us give the definition of ϵ-net:

Definition 3.6 (Covering numbers, Nets [37])**.** *Let $(X, \|.\|)$ be a metric space. Define $\epsilon > 0$. Then a ϵ-net of X is a subset N_ϵ of X such that every $x \in X$ can be approximated by some point $y \in N_\epsilon$, so that $\|x - y\| \leq \epsilon$. The minimum cardinality of ϵ-net of X is called the covering number of X which is denoted as $N(X, \epsilon)$.*

Lemma 3.7. *Let S^{n-1} be unit Euclidean sphere and let $\|.\|_2$ be Euclidean norm. Define $0 < \epsilon < 1$. Then,*

$$N(S^{n-1}, \epsilon) \leq \left(\frac{3}{\epsilon}\right)^n. \tag{3.27}$$

Proof. Let $\mathcal{B}^n(r)$ be a unit Euclidean ball of radius r and let N_ϵ be a maximal ϵ-separated of S^{n-1}, i.e so that $\|x - y\|_2 \geq \epsilon$ for all $x_i, x_j \in N_\epsilon$, $x_i \neq x_j$. Then it implies that the balls of radii of $\epsilon/2$ centered at the points in N_ϵ are disjoint and all contained in $\mathcal{B}^n(1 + \epsilon/2)$. Then we have

$$|N_\epsilon| \, \text{Vol}(\mathcal{B}^n(\epsilon/2)) \leq \text{Vol}(\mathcal{B}^n(1 + \epsilon/2)). \tag{3.28}$$

Since $\text{Vol}(a\mathcal{B}^n) = a^n \text{Vol}(\mathcal{B}^n)$, equation becomes

$$|N_\epsilon| \leq \frac{(1 + \epsilon/2)^n \text{Vol}(\mathcal{B}^n)}{(\epsilon/2)^n \text{Vol}(\mathcal{B}^n)} = \left(1 + \frac{2}{\epsilon}\right)^n. \tag{3.29}$$

Considering $0 < \epsilon < 1$ we have

$$|N_\epsilon| \leq \left(1 + \frac{2}{\epsilon}\right)^n \leq \left(\frac{3}{\epsilon}\right)^n, \tag{3.30}$$

which completes the proof. □

Now we are ready to prove Lemma 3.5:

Proof of Lemma 3.5. If the matrix ϕ approximately preserves the norm of $\frac{x}{\|x\|}$, it also approximately preserves the norm of x, since the measurement operator ϕ is linear. Define a set of D_Λ of points such that $D_\Lambda \subseteq \chi_\Lambda$,

$$\|x - d\|_2 \leq \frac{\sqrt{\delta}}{4}, \tag{3.31}$$

for all $d \in D_\Lambda$ with $\|d\|_2 = 1$ and for all $x \in \chi_\Lambda$ with $\|x\|_2 = 1$. Johnson-Lindenstrauss property claims that a linear mapping represented by an $m \times N$ matrix ϕ preserves the norm of any point $d \in D_\Lambda$ by the distortion factor at most ϵ with probability ρ where the probability ρ depends on distribution of matrix ϕ. For instance, ρ can be given as

$$\rho \geq 1 - 2\exp(-\frac{m}{2}(\epsilon^2/2 - \epsilon^3/4)), \tag{3.32}$$

for a random matrix of size $m \times N$ whose entries are drawn from any distribution that is given in (3.22), or in (3.23) as well as Gaussian distribution with $\phi_{i,j} \sim \mathcal{N}(0, \frac{1}{m})$. Considering this we set $\epsilon = \frac{\delta}{4}$ in inequality given in (3.21), then we have

$$(1 - \delta/4)\|d\|^2_{\ell_2^N} \leq \|\phi d\|^2_{\ell_2^m} \leq (1 + \delta/4)\|d\|^2_{\ell_2^N} \text{ for every } d \in D_\Lambda, \tag{3.33}$$

and it implies that

$$(1 - \sqrt{\delta}/2)\|d\|_{\ell_2^N} \leq \|\phi d\|_{\ell_2^m} \leq (1 + \sqrt{\delta}/2)\|d\|_{\ell_2^N} \text{ for every } d \in D_\Lambda. \tag{3.34}$$

Now, let us define a smallest positive number Δ such that

$$(1 - \Delta)\|x\|^2_{\ell_2^N} \leq \|\phi x\|^2_{\ell_2^m} \leq (1 + \Delta)\|x\|^2_{\ell_2^N} \text{ for every } x \in \chi_\Lambda, \tag{3.35}$$

and it implies that

$$(1 - \sqrt{\Delta})\|x\|_{\ell_2^N} \leq \|\phi x\|_{\ell_2^m} \leq (1 + \sqrt{\Delta})\|x\|_{\ell_2^N} \text{ for every } x \in \chi_\Lambda. \tag{3.36}$$

Our aim is to show that

$$\Delta \leq \delta \text{ for every } x \in \chi_\Lambda \tag{3.37}$$

in the case we choose a set of D_Λ of points that satisfy the inequality given in (3.36) holds with high probability when we set $\|x - d\|_2 \leq \sqrt{\delta}/4$ for all $d \in D_\Lambda$. To do this, let us write following triangle inequality:

$$\|\phi x\|_{\ell_2^m} \leq \|\phi(x-d)\|_{\ell_2^m} + \|\phi d\|_{\ell_2^m} \leq (1+\sqrt{\Delta})\sqrt{\delta}/4 + 1 + \sqrt{\delta}/2, \tag{3.38}$$

for any x $\in \chi_\Lambda$ with $\|x\|_{\ell_2^N} = 1$ and a fixed $d \in D_\Lambda$ with $\|d\|_{\ell_2^N} = 1$. Since Δ is the smallest positive number that satisfy inequality given in (3.35), we have

$$1 + \sqrt{\Delta} \leq (1+\sqrt{\Delta})\sqrt{\delta}/4 + 1 + \sqrt{\delta}/2. \tag{3.39}$$

which yields

$$\sqrt{\Delta} \leq \frac{3\sqrt{\delta}}{4-\sqrt{\delta}}. \tag{3.40}$$

Because of the fact that $0 < \sqrt{\delta} < 1$, one can re-write this equation as

$$\sqrt{\Delta} \leq \frac{3\sqrt{\delta}}{4-\sqrt{\delta}} \leq \sqrt{\delta}. \tag{3.41}$$

This proves upper inequality. To prove lower inequality, let us write following triangular inequality:

$$\|\phi x\|_{\ell_2^m} \geq \|\phi d\|_{\ell_2^m} - \|\phi(x-d)\|_{\ell_2^m} \geq 1 - \sqrt{\delta}/2 - (1+\sqrt{\Delta})\sqrt{\delta}/4 \geq 1 - \sqrt{\delta}. \tag{3.42}$$

We have proved that $\Delta \leq \delta$ for every $x \in \chi_\Lambda$ in (3.35), then we have

$$(1-\sqrt{\delta})\, \|x\|_{\ell_2^N} \leq \|\phi x\|_{\ell_2^m} \leq (1+\sqrt{\delta})\, \|x\|_{\ell_2^N} \text{ for every } x \in \chi_\Lambda. \tag{3.43}$$

It implies that

$$(1-\delta)\,\|x\|_{\ell_2^N}^2 \leq \|\phi x\|_{\ell_2^m}^2 \leq (1+\delta)\,\|x\|_{\ell_2^N}^2 \text{ for every } x \in \chi_\Lambda. \tag{3.44}$$

Then we have proved that for any $x \in \chi_\Lambda$ with $\|x\|_{\ell_2^N} = 1$ and a fixed $d \in D_\Lambda$ with $\|d\|_{\ell_2^N} = 1$, we have (3.44) when we set $\|x-d\|_2 \leq \sqrt{\delta}/4$ if inequality given in Equation (3.33) is satisfied. Considering Lemma (3.5), one can choose a set D_Λ of $(\frac{12}{\sqrt{\delta}})^k$ of points in χ_Λ. If we apply union bound, we can say that the measurement matrix ϕ acts in a norm preserving way on each fixed k-dimensional subspace χ_Λ with probability at least $1-2(\frac{12}{\sqrt{\delta}})^k e^{-C(\delta/4)m}$. □

Now it is time to extend the given result from fixed k-dimensional subspace to all possible k-dimensional signals:

Theorem 3.8 ([30]). *Let ϕ be the $m \times N$ measurement matrix that satisfies the inequality given in* (3.21). *Define* $0 < \delta < 1$. *If*

$$k \leq C_1 \frac{m}{\log(N/k)}, \tag{3.45}$$

then ϕ satisfies the Restricted Isometry Property of order k with probability $\geq 1-2e^{-C_2 m}$ where C_1 and C_2 are constants depending on δ.

We have $(\frac{12}{\sqrt{\delta}})^k$ of points in a fixed χ_Λ. If we extend this result for $\binom{N}{k}$ possible k-dimensional subspaces we have set D_T of $(\frac{eN}{k})^k(\frac{12}{\sqrt{\delta}})^k$ of points such that

$$D_T = \bigcup_{\Lambda:|\Lambda|\leq k} D_\Lambda. \tag{3.46}$$

Since we can use Sterling's approximation $k! \geq (k/e)^k$ to bound number $\binom{N}{k}$ by

$$\binom{N}{k} = \frac{N(N-1)(N-2)(N-3)..(N-k+1)}{k!} \leq \frac{N^k}{k!} \leq (\frac{eN}{k})^k. \tag{3.47}$$

Applying union bound over D_T, we can say ϕ satisfies RIP of order k with probability

$$\begin{aligned} &\geq 1 - 2(\frac{eN}{k})^k(\frac{12}{\sqrt{\delta}})^k e^{-C(\delta/4)m} \\ &= 1 - 2e^{-C(\delta/4)m + k[\log(N/k)+\log(12/\delta)+1]} \geq 1 - 2e^{-C_2 m} \end{aligned} \tag{3.48}$$

whenever we choose C_1 sufficiently small provided that $C_2 \leq C(\delta/4) - C_1 \frac{[\log(N/k)+\log(12/\delta)+1]}{\log(N/k)}$ for $k \leq C_1 m/\log(N/k)$.

As a conclusion to this chapter, we have proved that random matrices satisfies the RIP with high probability. These results will be used to design a watermarking scheme in Chapter 5.

4. SPARSE SIGNAL RECOVERY

In this chapter, we will start by giving the formulation of the sparse signal recovery problems for (i) error-free case, (ii) the case that measurements are corrupted by a bounded noise, and (iii) the case in which measurements are corrupted by impulsive noise. We review the popular algorithms to be used in recovery of sparse signals for these three cases. Particularly, we will restrict our attention two different type of methods. First, convex optimization algorithms are proven to provide stable recovery, and furthermore, quite a few numerical solvers for them are available. Second, greedy methods yield very fast algorithms, however existing greedy algorithms provide relatively weaker performance guarantees.

We will also discuss that convex relaxation methods require fewer measurements than greedy algorithms do. Then it will be clarified that is why we focus on convex relaxation methods for the rest of this thesis.

4.1. Formulations of Sparse Signal Recovery Problems

For an underdetermined system of equations

$$y = Ax, \tag{4.1}$$

the problem of finding a solution $\hat{x}$ of x can be formulated as follows

$$\text{Find x subject to } y = Ax. \tag{4.2}$$

Basic linear algebra says that this problem has infinitely many solutions under the assumption that A is full row rank. As detailed in Chapter 2, at least one more requirement, sparsity of x, is needed to achieve a unique solution. Since we seek for the sparsest solution $\hat{x}$, the problem is referred to as sparse approximation. Then (4.2) can be recast as an ℓ_0-minimization problem

$$\min_{x} \ \|x\|_{\ell_0^N} \quad \text{s.t.} \ \ y = Ax. \tag{4.3}$$

Unfortunately, solving (4.3) requires combinatorial search and it is NP hard problem. There are five common approaches to overcome this problem [38] : (i) Convex Relaxation, (ii) Greedy Algorithms, (iii) Bayesian Framework, (iv) Non-convex optimization, and (v) Brute force.

4.1.1. Convex Relaxation

Although the convex relaxation technique has recently become more popular with increased interest in Compressive Sensing, the first usages of the method date back to 1970s in the fields of geophysics and statistics [39, 40]. As we discussed in Chapter 2, the relaxation of (4.2), is commonly known as Basis Pursuit (BP),

$$\min_{x} \ \|x\|_{\ell_1^N} \quad \text{s.t.} \ \ y = Ax. \tag{4.4}$$

In a more realistic scenario, the measurements may be corrupted by a noise:

$$y = Ax + z. \tag{4.5}$$

We have shown in Theorem 3.8 that when the matrix A satisfies the RIP of order 2k with $\delta_{2k} \leq \sqrt{2} - 1$, then $m \geq O((k)\log(\frac{N}{k}))$ measurements suffice to recover the signal exactly with overwhelming probability. If the measurements are corrupted by noise provided that $\|z\|_{\ell_2} \leq \epsilon$, then the solution $\hat{x}$ can be found via

$$\hat{x} = \arg\min_{x} \ \|x\|_{\ell_1} \quad \text{s.t.} \quad \|y - Ax\|_{\ell_2} \leq \epsilon. \tag{4.6}$$

The optimization problem defined in Equation (4.6) is known as Basis Pursuit Denoising (BPDN) [41]. Recall that if the matrix A satisfies RIP of order of $2k$ with

$\delta_{2k} < \sqrt{2} - 1$, it is possible to approximate x with a bounded error :

$$\|x - \hat{x}\|_{\ell_2} \leq C_0 \epsilon, \tag{4.7}$$

where C_0 depends on δ_{2k} [28] (see Chapter 2). Convex relaxation models in the case where the measurement vector, $y = Ax$ is corrupted by some noise with bounded energy has also been considered in [42, 43].

4.1.2. Recovery under Impulsive Noise using ℓ_1 minimization

In a realistic scenario, measurement vector can also be corrupted by impulsive noise due, e.g., to shot noise, malfunctioning hardware, transmission errors [44], and malicious attacks. In this scenario, portions of the measurement vector can be completely corrupted. Impulsive noise has typically very large variance, and this may lead to a large reconstruction error in Equation (4.7). Impulsive noise model has been investigated in [45], where the authors used probabilistic approach and proposed a non-convex optimization program to recover the signal. In [27], impulsive noise model has been investigated, but in the context of error correction coding. In [44], noise model that is sparse in a proper basis has been considered in wide range of amplitudes and error rates. In this study, the measurement vector that is corrupted by a noise is modelled as

$$y_n = Ax + \Omega z = \begin{bmatrix} A \mid \Omega \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix}, \tag{4.8}$$

where the signal x is k-sparse, the noise z is L-sparse (hence impulsive) , and the matrix Ω is $m \times m$ unit basis. In this model, one can solve the following optimization problem

$$\begin{bmatrix} \hat{x} \\ \hat{z} \end{bmatrix} = \arg \min_{[x \;\, z]^{\mathrm{T}}} \left\| \begin{bmatrix} x \\ z \end{bmatrix} \right\|_{\ell_1} \quad \text{s.t.} \;\; y_n - \begin{bmatrix} A \mid \Omega \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = 0. \tag{4.9}$$

With a close look, one can easily see that (4.9) also can be referred to as BP. In noisy case, (4.9) can relaxed to

$$\begin{bmatrix} \hat{x} \\ \hat{z} \end{bmatrix} = \arg \min_{[x \ z]^{\mathrm{T}}} \left\| \begin{bmatrix} x \\ z \end{bmatrix} \right\|_{\ell_1} \quad \text{s.t.} \quad \left\| y_n - \begin{bmatrix} A \mid \Omega \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} \right\|_{\ell_2} \leq \epsilon \tag{4.10}$$

which is also a BPDN.

4.2. Recovery Algorithms

4.2.1. ℓ_1 minimization algorithms

After modelling the convex relaxation problems for different noise models, now we will briefly discuss on reconstruction algorithms. In the situation where x, y and A are real, the optimization problem (4.4), basis pursuit, can be adjusted as linear program (LP) with equality constraint using interior-point methods [13]. In complex-valued case of (4.4), the problem can be solved as second order cone program (SCOP). Interested readers may also find a Matlab software, CVX, in aforementioned work [13]. In noisy case (4.6), BPDN, can also be recast as (SCOP) with quadratic constraint. By the way, although we primarily focus on the case (4.6), BPDN, an equivalent formulation can be given by

$$\min_{x} \ \frac{1}{2} \left\| Ax - y \right\|_{\ell_2}^2 + \lambda \left\| x \right\|_{\ell_1} . \tag{4.11}$$

This formulation is frequently referred to as the Least Absolute Shrinkage and Selection Operator (LASSO) [46]. The regularization parameter λ plays a role in controlling the sparsity of the solution $\hat{x}$ in (4.11). There has been much effort in developing an effective algorithm to solve (4.11). For instance, a faster gradient projection algorithm has been recently proposed [47]. This algorithm is known as Gradient Projection for Sparse Reconstruction (GPRS). Another interesting formulation of ℓ_1 minimization remains, the solution of the following ℓ_1 regularization convex problem with conic constraint,

$$\min_{x} \; \|x\|_{\ell_1} \quad \text{s.t.} \quad \left\|A^T(y - Ax)\right\|_{\ell_\infty} \leq \gamma \tag{4.12}$$

is known as Danzig Selector (DS) [48]. A discussion on equivalance between LASSO and DS can be found in [49]. For the sake of completion we refer to the ℓ_1-magic solver that deals with BP to DS, the code is available on the webpage: http://users.ece.gatech.edu / justin/l1magic/.

4.2.2. Greedy Methods

Greedy algorithms try to estimate the sparse signal coefficients iteratively. In each iteration, they re-identify and improve both the support set and estimation of the signal, $\hat{x}$. Despite some improvement in recent years, the theoretical analysis on the guarantee conditions of these methods are still shaky [5]. Because of this reason, we will not use greedy methods in this thesis. Nevertheless, we will give a brief overview in this section. The simplest and best known greedy algorithm is Orthogonal Matching Pursuit [50]. The OMP algorithm is given in Figure 4.1 for interested readers.

A recent work provides a recovery limit in term of restricted isometry constant (RIC):

Theorem 4.1 ([51])**.** *Let $y = Ax$, where x is a k-sparse signal. Let also measurement matrix A satisfy the RIP of order k with δ_k. Then the OMP algorithm recovers x exactly from the observation vector, y, provided that*

$$\delta_{k+1} \leq \frac{1}{\sqrt{k}+1}. \tag{4.13}$$

Theorem 4.1 gives us a theoretical upper bound on δ_{k+1} to guarantee of recovery of a k-sparse signal, x. Unfortunately, it leads the RIC to be relatively small. Therefore it is applicable when the number of measurements, m, is relatively high.

Input: y, A;

Initialize: residual: $r^0 = y$, $\hat{x}^0 = 0$, support: $\Lambda^0 = \emptyset$, $i = 0$;

Determine: ϵ;

repeat

$i \Leftarrow i + 1$;

$q \Leftarrow A^T r^{i-1}$;

$k^i \Leftarrow \arg\max_j (|q_1|, |q_2|, ..., |q_j|, ..., |q_N|)$;

$\Lambda^i \Leftarrow \Lambda^{i-1} \bigcup \{k^i\}$;

$\hat{x}_{\Lambda^i} \Leftarrow \arg\min_x \|y - A_{\Lambda^i} x\|_{\ell_1^m}$;

$r^i = y - A_{\Lambda^i} \hat{x}$;

until $\|r^i\|_{\ell_2^m} \leq \epsilon$.

Output: $\hat{x}$

Figure 4.1. OMP Algorithm.

5. WATERMARKING OF COMPRESSIVE SENSING MEASUREMENTS

In many applications, such as avionics and imagery, it is desirable to reduce the number of sensors. Compressive sensing paradigm can be a remedy to this problem. One extreme example of compressive sensing is the single pixel camera built by Takhar *et al.* [52]. Moreover, we may wish to reduce the power consumed by the sensors. Another interesting application of CS is wireless body sensor networks (WBSN) [6] for health monitoring applications. Doctors want to monitor patients away from hospitals using some wearable devices which send meaningful biosensory data such as ECG, blood pressure, EEG, etc. While these devices need to be power efficient, portable, fail safe to patient's privacy and security, we also want to minimize the volume of transmitted data and extend the battery lifetime. Recently, Mamaghain *et al.* [6] have proposed a system, where sensors sample ECG signals using compressive sensing, then they transmit wirelessly these measurements to a remote monitoring center.

One may also wish to embed meta data on these compressive sensing measurements. For example, in an WBSN application, embedding of patient's information may be required. We could attach patient meta data to his/her biosensory data using some digital watermarking scheme. Digital watermarking can be defined as the process of embedding information, called watermark, in a media signal discreetly and robustly [8]. Most algorithms in the literature embed watermark data in (i) Spatial Domain, (ii) Spectral Domain, (iii) Hybrid domain [53]. Using additive watermarking techniques where a noise carrier is used, or substitutive techniques such as quantization index modulation, may both lead to additional uncertainty in the CS framework and compromise reconstruction. On the other hand working in spectral domain and hybrid domain is out of question, since measurements y have high information content and are not compressible. In the sequel, we will investigate how one can embed some additional information directly onto CS measurements and to reconstruct the signal

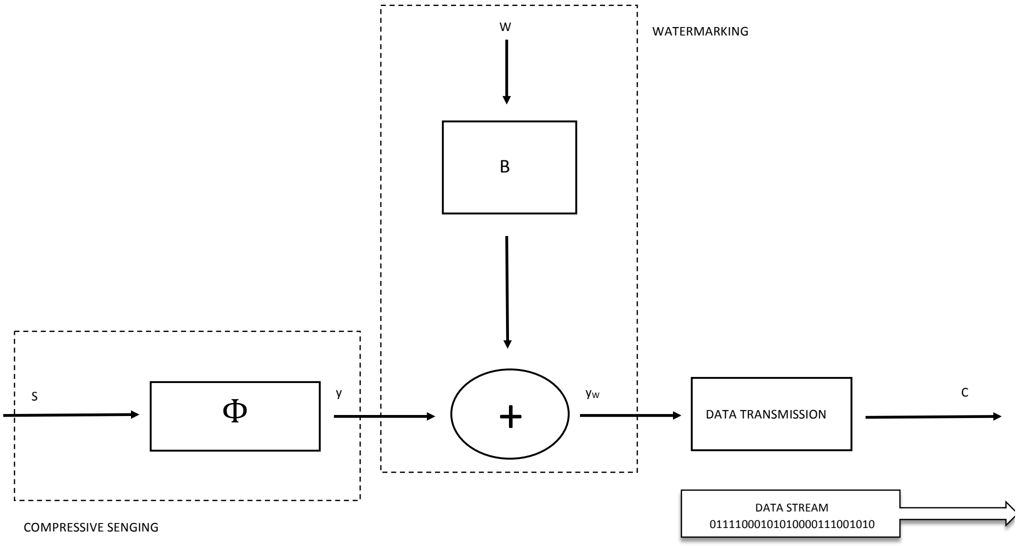


Figure 5.1. Block diagram represents watermarking while sensing.

without loosing any information while recovering the data.

5.1. Data Hiding

Let $w \in \{+a, -a\}^M$ be the watermark sequence. One of the simplest ways of embedding this data onto compressive sensing measurements, $y = Ax$, is linear encoding of watermark and directly adding these encoded messages onto measurements,

$$y_w = Ax + Bw \tag{5.1}$$

where, y_w is watermarked measurements and B is an $m \times M$ encoding matrix generated by a seed, known to both encoder and decoder with $\|Bw\|_{\ell_2^N} \leq P_E$. We have added a power constraint $\|Bw\|_{\ell_2^N} \leq P_E$ to restrict extra bit budget, where P_E is embedding power. A pictorial representation of the embedding scheme can be found in Figure 5.2. In order to meet the embedding power constraint $\|Bw\|_{\ell_2^N} \leq P_E$, the amplitude a is

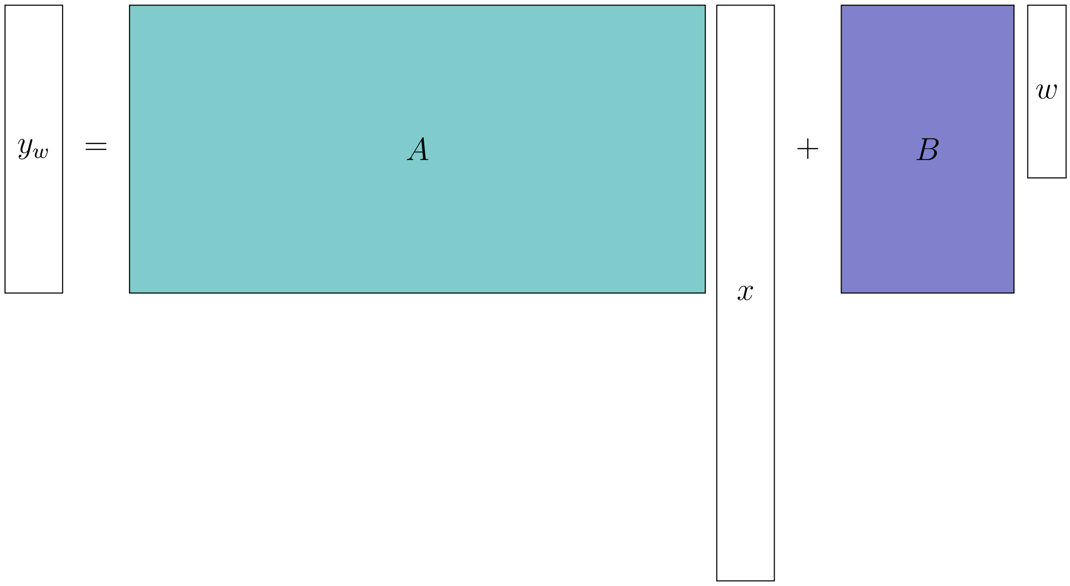


Figure 5.2. Pictorial representation of data hiding in the compressive sensing framework.

chosen accordingly. In a realistic scenario, the noise occurs in the process of the quantization of measurements [54]. Along with quantization error, the measurement process itself may be noisy. Additionally, the watermarked measurements can be altered by a malicious user or by channel imperfections, therefore the decoder receives

$$y_n = y_w + z = Ax + Bw + z, \tag{5.2}$$

where z is an unknown noise pattern.

5.2. Joint Signal Reconstruction and Watermark Recovery

The watermarked compressive sensing measurement problem is formulated as follows: Recover the k-sparse x signal and M bit watermark w from the knowledge of noisy watermarked measurements, y_n. In the case where the noise term vanishes, the recovery of both x and w should be exact; otherwise it should be a close approximation to the true signal, x.

As for the watermark information we assume that it cannot tolerate any loss, therefore recovering of w should be exact. In the light of these requirements, we look for maximum achievable embedding rate $\mathcal{R} = M/m$ in bits/measurement for $\text{Prob}(w \neq \hat{w}) \to 0$. In the meantime, we are interested in the reconstruction of x with a small mean square error $\mathbb{E}\{\|x - \hat{x}\|_2\}$.

5.3. Reconstruction under Additive White Gaussian Noise (AWGN)

The watermarked compressive sensed measurements can be corrupted by additive Gaussian noise and this can be modelled as $z \sim \mathcal{N}(0, \sigma^2 I)$. We can rearrange Equation (5.2) as follows

$$y_n = C \begin{bmatrix} x \\ w \end{bmatrix} + z, \tag{5.3}$$

where $C = [A|B]$.

5.3.1. Classical ℓ_2 Minimization

One classical solution of underdetermined system of equations as in (5.3) can be the minimum norm solution. The solution $\hat{x}$ and $\hat{w}$ that minimizes $\|A\hat{x} + B\hat{w} - y_n\|_{\ell_2}$ is

$$\begin{bmatrix} \hat{x} \\ \hat{w} \end{bmatrix} = C^{\mathrm{T}}(CC^{\mathrm{T}})^{-1} y_n. \tag{5.4}$$

5.3.2. ℓ_1 Minimization

Alternatively, as in Section 2.1, the problem can be formed as BPDN, which is solved by minimizing a $(k+M)$-sparse vector $[x \quad w]^{\mathrm{T}}$ of length $N + M$. This can be done in two steps. First, the vector $[\tilde{x} \;\; \tilde{w}]^{\mathrm{T}}$ can be found via

$$\begin{bmatrix} \tilde{x} \\ \tilde{w} \end{bmatrix} = \arg \min_{[x \;\; w]^{\mathrm{T}}} \left\| \begin{bmatrix} x \\ w \end{bmatrix} \right\|_{\ell_1} \quad \text{s.t.} \quad \left\| y_n - C \begin{bmatrix} x \\ w \end{bmatrix} \right\|_{\ell_2} \leq \epsilon_1. \tag{5.5}$$

If C satisfies the RIP of order $2(k+M)$ with $\delta_{2(k+M)} < \sqrt{2} - 1$, the N-long, k-sparse x and the M-long dense w can be reconstructed approximately by using

$$m \geq O((k+M) \log((N+M)/(k+M))) \tag{5.6}$$

measurements with bounded error such that

$$\left\| \begin{bmatrix} x \\ w \end{bmatrix} - \begin{bmatrix} \tilde{x} \\ \tilde{w} \end{bmatrix} \right\|_{\ell_2} \leq C_1 \epsilon_1, \tag{5.7}$$

where C_1 is a constant that depends on $\delta_{2(k+M)}$. Secondly, an additional step can be performed to increase the maximum achievable embedding rate $\mathcal{R} = M/m$ in bits/measurement when $\mathrm{Prob}(w \neq \mathrm{sgn}(\tilde{w})) \to 0$. Since it is known that watermark information $w_i \in \{-a, +a\}$, $i \in \{1, 2, .., M\}$, $\hat{w}$ can be estimated by thresholding $\tilde{w}$ as follows

$$\hat{w}_i = a * \mathrm{sgn}(\tilde{w}_i). \tag{5.8}$$

After the estimation of $\hat{w}$ using Equation (5.8), we can estimate $\hat{x}$ by solving

$$\hat{x} = \arg \min_{x} \; \|x\|_{\ell_1} \quad \text{s.t.} \quad \|(y_n - B\hat{w}) - Ax\|_{\ell_2} \leq \epsilon_1. \tag{5.9}$$

By using Equation (5.5), (5.8) and (5.9), it is possible to reconstruct k-sparse signal $\hat{x}$ and watermark information $\hat{w}$. It is relatively intuitive that the optimization problem given in Equation (5.5) is not the optimal set-up for our problem, since w is dense. Because of this reason, we propose the following robust recovery approach to decode the watermarking scheme in Equation (5.2).

5.3.3. Proposed Reconstruction Algorithm

In the recovery of the signal and the watermark from noisy measurements, we wish (i) the estimation $\hat{w}$ of watermark to be exact, (ii) the uncertainty of estimation $\hat{x}$ of the signal not to exceed the uncertainty level given in Lemma 2.7.

Thus, it is easy to see that if (i) is satisfied, (ii) is also satisfied using ℓ_1 minimization. Then, we should firstly estimate $\hat{w}$. Since as in (5.9) and (5.8), one can remove the watermark signal, this situation correspond to reversible watermarking in the literature [55–57]. But estimating $\hat{w}$ requires pre-estimation $\tilde{x}$ of x. Mathematically speaking, let us construct a $p \times m$ matrix F which annihilates the watermark information part such that $FB = 0$. Then, if we apply the matrix F to the noisy measurements, y_n, we obtain

$$\tilde{y} = Fy_n = F(Ax + Bw + z_g) = FAx + Fz_g, \tag{5.10}$$

where $p = m - M$. Let also our annihilation matrix F have orthogonal rows and also $\|F_i\|^2_{\ell_2^m} = \frac{m}{p}$ where F_i denote the rows of the matrix F for $i = 1, ..., p$. We would like to mention here that possessing orthogonal rows is not a must but it provides extra convenience, which will be clarified later. The constraint $\|F_i\|^2_{\ell_2^m} = \frac{m}{p}$ is also not a necessity, but it enables us to handle this system mathematically without much effort. After these preliminaries, note that this system is also an underdetermined system of equations with noise pattern Fz. Then, pre-estimation (or first-tier estimation) $\tilde{x}$ of x can be found via

$$\tilde{x} = \arg\min_{x} \|x\|_{\ell_1^N} \quad \text{s.t.} \quad \|\tilde{y} - FAx\|_{\ell_2^p} \leq \epsilon. \tag{5.11}$$

After finding the pre-estimation $\tilde{x}$ of x, computing pre-estimation $\tilde{w}$ of w is possible using least squares,

$$\tilde{w} = (B^{\mathrm{T}}B)^{-1}B^{\mathrm{T}}(y_n - A\tilde{x}). \tag{5.12}$$

At this point, the final estimation $\hat{w}$ is obtained as

$$\hat{w}_i = a * \mathrm{sgn}(\tilde{w}_i), \tag{5.13}$$

and the final estimation $\hat{x}$ of x can be done via

$$\hat{x} = \arg\min_x \ \|x\|_{\ell_1^N} \quad \text{s.t.} \quad \|(y_n - B\hat{w}) - Ax\|_{\ell_2^m} \leq \epsilon. \tag{5.14}$$

Input: y_n, A, B;

Determine: ϵ

1. Apply F to y_n : $\tilde{y} = Fy_n$

2. Estimate $\tilde{x}$: $\tilde{x} = \arg\min_x \|x\|_{\ell_1^N}$ s.t. $\|\tilde{y} - FAx\|_{\ell_2^p} \leq \epsilon$

3. Estimate $\tilde{w}$: $\tilde{w} = (B^{\mathrm{T}}B)^{-1}B^{\mathrm{T}}(y_n - A\tilde{x})$

4. Threshold $\tilde{w}$: $\hat{w}_i = a * \mathrm{sgn}(\tilde{w}_i)$

5. Estimate $\hat{x}$: $\hat{x} = \arg\min_x \ \|x\|_{\ell_1^N}$ s.t. $\|(y_n - B\hat{w}) - Ax\|_{\ell_2^m} \leq \epsilon$

Output: $\hat{x}$, $\hat{w}$

Figure 5.3. The proposed reconstruction algorithm of the signal and the watermark.

5.3.4. Stability of The Proposed Reconstruction Algorithm

In this section, we will define the proper bounds to make the proposed algorithm be stable.

Theorem 5.1. *Let A be an $m \times N$ matrix satisfying RIP order of $2k$ with $\delta_{2k}(A) < \sqrt{2} - 1$, where $\delta_{2k}(A)$ is restricted isometry constant of the matrix A. Let also F be an $p \times m$ matrix in the null space of the watermark sequence modulation matrix B*

with orthogonal rows such that $\|F_i\|^2_{\ell_2^m} = \frac{m}{p}$ *where* F_i *denotes one of the rows of the matrix* F*. Suppose the matrix* FA *satisfies the RIP of order* $2k$ *with* $\delta_{2k}(FA) < \sqrt{2}-1$*. The watermarked and noisy measurements are the form* $y_n = Ax + Bw + z$ *such that* $\|w\|_{\ell_2^M} = \upsilon$*, where* z *is the noise pattern with i.i.d elements* z_i *drawn from* $\sim \mathcal{N}(0, \sigma_G^2)$*. Then the proposed algorithm recovers* M *bits watermark signal exactly with probability*

$$Prob(w \neq \hat{w}) = 1 - e^{-\frac{3p}{4}\gamma^2}, \tag{5.15}$$

provided that

$$M \leq \frac{\upsilon^2}{C_0^2 \epsilon^2}. \tag{5.16}$$

Furthermore, it also approximates x *with a bounded error*

$$\|x - \hat{x}\|_{\ell_2^N} \leq 4 \frac{\sqrt{1+\delta_{2k}(A)}}{1-(1+\sqrt{2})\delta_{2k}(A)}(1+\gamma)\sqrt{m}\sigma_G, \tag{5.17}$$

where C_0 *is a positive constant that depends on* $\delta_{2k}(A)$ *and* $\delta_{2k}(FA)$*.*

Proof. Define a vector $f = Fz$, then elements of this vector will be $f_i =< F_i, z >$, where F_i denotes the ith row of the $p \times m$ matrix F. Assume that F has orthogonal rows and z is a $m \times 1$ vector that consists of i.i.d Gaussian random variables with $z_i \sim \mathcal{N}(0, \sigma_G^2)$. It is obvious that f_i has distribution $\mathcal{N}(0, \frac{m}{p}\sigma_G^2)$, since $\mathbb{E}(f_i) = 0$, and using i.i.d property of z we get

$$\begin{aligned}\mathbb{E}(f_i^2) &= \mathbb{E}(< F_i, z >^2) = \mathbb{E}((\sum_{j=1}^{m} F_{i,j}z_j)^2) = \mathbb{E}(\sum_{j=1}^{m}(F_{i,j}z_j)^2 + \sum_{l=1}^{m}\sum_{\substack{t=1\\t\neq l}}^{m} z_l z_t F_{i,l}F_{i,t}) \\ &= \sum_{j=1}^{m} F_{i,j}^2 \mathbb{E}(z_j^2) + \sum_{l=1}^{m}\sum_{\substack{t=1\\t\neq l}}^{m} \mathbb{E}(z_l z_t)F_{i,l}F_{i,t} = \sigma_G^2 \sum_{j=1}^{m} F_{i,l}^2 = \sigma_G^2 \|F_i\|^2_{\ell_2^m} = \frac{m}{p}\sigma_G^2,\end{aligned} \tag{5.18}$$

where $\mathbb{E}(z_l z_t) = 0$ for $t \neq l$. Then by using inequality given in (2.25), we obtains

$$\Pr(\|Fz\|_{\ell_2^p} \geq (1+\gamma)\frac{\sqrt{m}}{\sqrt{p}}\sqrt{p}\sigma_G) \leq e^{-\frac{3p}{4}\gamma^2}. \quad (5.19)$$

Using Lemma 2.7, we can show that the solution $\tilde{x}$ to (5.11) obeys

$$\|x - \tilde{x}\|_{\ell_2^N} \leq C_3\epsilon \quad (5.20)$$

with probability at least $1-e^{-\frac{3p}{4}\gamma^2}$. In other words we have first shown that the effect of the noise Fz on the first tier solution in (5.11) can be discarded with high probability, $1 - e^{-\frac{3p}{4}\gamma^2}$, so that the deviation of the final (second tier) solution $\hat{x}$ from the true solution can be upperbounded as in (5.17).

At this point we want to find a bound for uncertainty before least squares solution to 5.12. That is to say, we look for $\|y_n - A\tilde{x} - Bw\|_{\ell_2^m}$. Using $y_n = Ax + Bw + z$, we get

$$\begin{aligned}\|y_n - A\tilde{x} - Bw\|_{\ell_2^m} &= \|A(x-\tilde{x}) + z\|_{\ell_2^m} \\ &\leq \|A(x-\tilde{x})\|_{\ell_2^m} + \|z\|_{\ell_2^m} \leq (C_3(\sqrt{1+\delta_{2k}(A)} + C_4) + 1)\epsilon, \quad (5.21)\end{aligned}$$

where the last inequality comes from the triangular inequality. Let us define $e_r = A(x - \tilde{x}) + z$. Then the solution $\tilde{w}$ to (5.12) obeys

$$w - \tilde{w} = (B^{\mathrm{T}}B)^{-1}B^{\mathrm{T}}e_r,$$

and since the eigenvalues of $B^{\mathrm{T}}B$ are well-behaved, we have

$$\|w - \tilde{w}\|_{\ell_2^M} \approx \left\|B^{\mathrm{T}}e_r\right\|_{\ell_2^M} \approx \|e_r\|_{\ell_2^M} \leq C_0\epsilon, \quad (5.22)$$

where $C_0 = C_3(\sqrt{1+\delta_{2k}(A)} + C_4) + 1$. Notice that the reconstruction error of x in the p-dimensional space as in (5.11) has its reflection on the reconstruction of w. The term $\sqrt{1+\delta_{2k}(A)}$ comes from upper bound of RIP of A in (2.22). One can claim that the vector $h = x - \tilde{x}$ may not necessarily be $2k$-sparse, since the solution $\tilde{x}$ to (5.13) is erroneous and not k-sparse. But it is known that the vector h should be approximately sparse and not much violate the upper bound $\sqrt{1+\delta_{2k}(A)}$. Here we use a constant term C_4 to represent such possible violations without computing exact value of C_4.

Now, we would like to point out that the uncertainty on estimation $\tilde{w}$ of w which is given in Equation (5.22), is the worst-case scenario. We mean by the worst-case scenario that it is the situation when we do not know anything about error e_r except its norm-bound, $\|e_r\|_{\ell_2^M} \le C_0\epsilon$. Under this assumption, Equation (5.22) implies that

$$\|w - \tilde{w}\|_{\ell_\infty^M} \le \|w - \tilde{w}\|_{\ell_2^M} \le C_0\epsilon. \tag{5.23}$$

Therefore we can conclude that, if

$$|w_i| \ge C_0\epsilon, \tag{5.24}$$

the solution $\hat{w}$ to (5.13) is exact with full probability. Considering $\|w\|_{\ell_2^M} = \upsilon$, inequality (5.24) turns to

$$|w_i| = \frac{\upsilon}{\sqrt{M}} \ge C_0\epsilon \tag{5.25}$$

and it gives Equation (5.16). Under the condition of the exact watermark recovery, using Lemma 2.7, the solution $\hat{x}$ to (5.14) obeys

$$\|x - \hat{x}\|_{\ell_2^N} \le 4\frac{\sqrt{1+\delta_{2k}}}{1-(1+\sqrt{2})\delta_{2k}}(1+\gamma)\sqrt{m}\sigma_G \tag{5.26}$$

which completes the proof. □

5.3.5. Does FA satisfy the RIP ?

In the discussion of the stability properties of the algorithm to recover watermark and signal jointly, we have tacitly assumed that the $m - M \times m$ matrix FA satisfies the RIP of order $2k$. We know that the matrix A satisfies RIP of order of $2k$, and we left multiply with the null space of B, i.e., such that FA. We must show that this new matrix FA still satisfies the RIP of order $2k$.

Lemma 5.2. *Let A be a $m \times N$ matrix with i.i.d elements $A_{i,j}$ drawn from $\mathcal{N}(0, \frac{1}{m})$ and let F be an $p \times m$ matrix with orthogonal rows such that $\|F_i\|^2_{\ell_2^m} = \frac{m}{p}$ where F_i denote the rows of the matrix F for $i = 1, ..., p$. If*

$$p \geq O(k \log(N/k)), \tag{5.27}$$

then the matrix FA satisfies the Restricted Isometry Property of order k with probability $\geq 1 - 2e^{-O(p)}$.

Proof. Let $A_{C(i)}$ be the columns of the matrix A and let F_i be the rows of the matrix F. Then the elements of the matrix FA will be $Z_{i,j} =< F_i, A_{C(j)} >$ independent Gaussian random variables. We will prove only $\mathbb{E}(Z^2_{i,j}) = \frac{1}{p}$. It can be calculated via

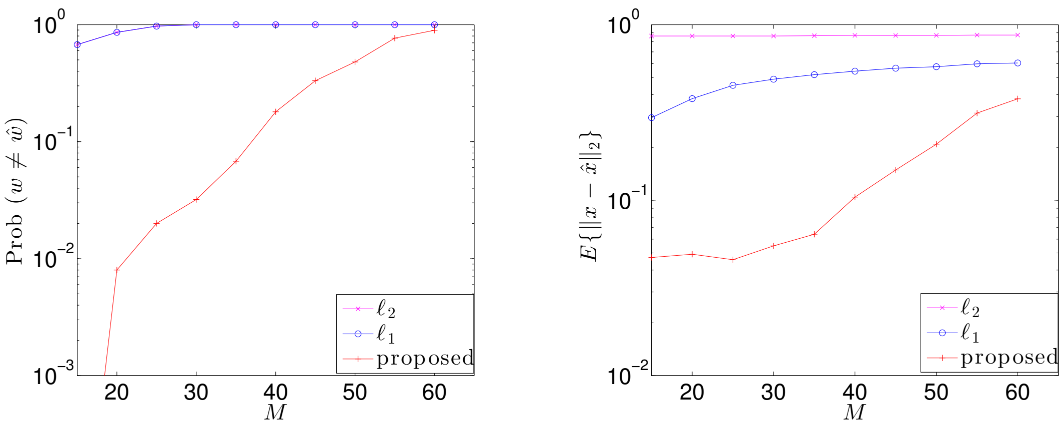
$$\begin{aligned} \mathbb{E}(Z^2_{i,j}) = \mathbb{E}(< F_i, A_{C(j)} >^2) = \mathbb{E}(\sum_{l=1}^{m} F^2_{i,l} A^2_{l,j}) = \sum_{l=1}^{m} F^2_{i,l} \mathbb{E}(A^2_{l,j}) \\ = \frac{1}{m} \sum_{l=1}^{m} F^2_{i,l} = \frac{1}{m} \|F_i\|^2_{\ell_2^m} = \frac{1}{p}. \end{aligned} \tag{5.28}$$

We can therefore use Theorem 3.8 to end up the poof. □

5.3.6. Experiments and Results

In this section, we present performance results of the proposed decoding algorithms for the case where the watermarked measurements are corrupted by AWGN. In our experiments, we generate x, w vectors such that $\|x\|_{\ell_2} = 1$, $\|w\|_{\ell_2} = 0.25$ ($a = \frac{0.25}{\sqrt{M}}$), and produce measurement matrix A as a Gaussian random matrix. We also generate a F with orthogonal rows. Then, columns of the encoding matrix B are obtained from span of null space of F. Each experiment is conducted 250 times, and corresponding $\text{Prob}(w \neq \hat{w})$ and $\mathbb{E}\{\|x-\hat{x}\|_2\}$ values are reported.

Experiment 1: In each try, we produce a $k = 30$-sparse, $N = 512$ sample long unit norm synthetic signal. The measurement matrix, A, is produced as Gaussian random matrix with $A_{i,j} \sim \mathcal{N}(0, 0.3/145)$. Then we obtain measurements, y, by using this measurement matrix. M-bit length watermark is embedded on to 145-measurement for each try. We look for maximum achievable embedding rate $\mathcal{R}$ in bits/measurement when $\text{Prob}(w \neq \hat{w}) \to 0$. We are interested in keeping reconstruction error $\mathbb{E}\{\|x-\hat{x}\|_2\}$ at reasonable level such that $\mathbb{E}\{\|x-\hat{x}\|_2\} \leq \varrho$.

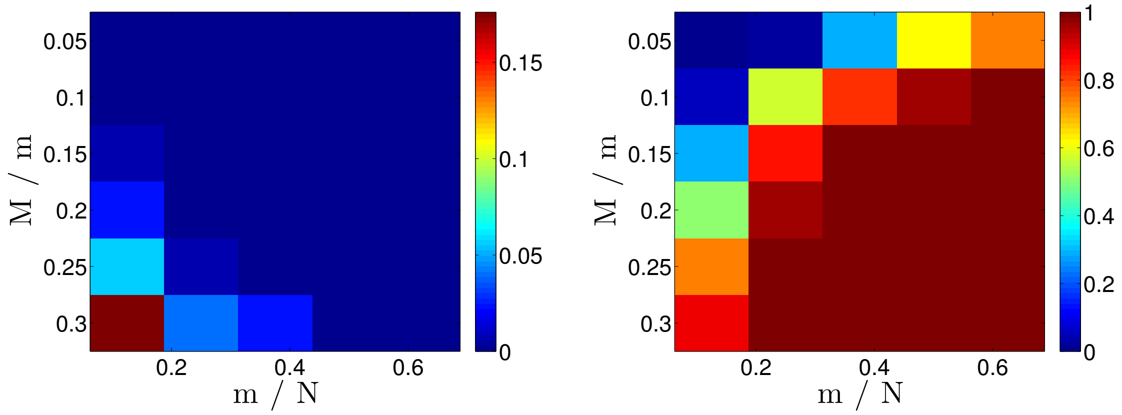


(a) (b)

Figure 5.4. Performance comparison of algorithms for the case in which measurements are corrupted by AWGN (a) M vs. $\text{Prob}(w \neq \hat{w})$. (b) M vs. $\mathbb{E}\{\|x-\hat{x}\|_2\}$.

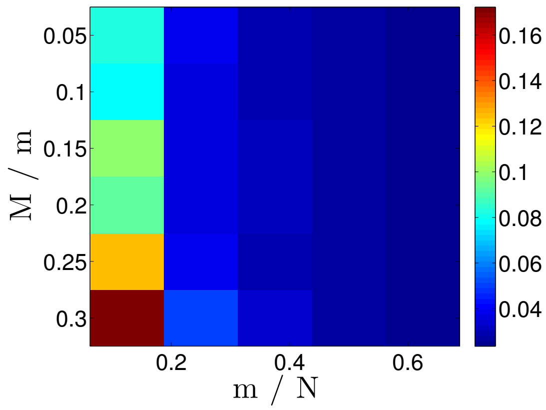
We consider the Gaussian noise with 32dB signal-to-noise ration (SNR). SNR in dB is defined as $20\log_{10}(\frac{\|Ax+Bw\|_2}{\|z\|_2})$. Fig. 5.4 shows that the proposed method achieves $\mathcal{R} \leq 20/145$ bit/measurement performance when $\text{Prob}(w \neq \hat{w}) \to 0$. Furthermore, for the maximum achievable rate, the expected mean squared error of the reconstruction error is bounded by $\mathbb{E}\{\|x-\hat{x}\|_2\} \leq 4 * 10^{-2}$ at $M \leq \frac{20}{145}$ as seen in Fig. 5.4. Therefore, the proposed method outperforms both ℓ_1 and ℓ_2 decoding algorithms.

Experiment 2: Varying number of measurements: This time we also change the number of measurements and compare the performance results of the two algorithms: ℓ_1 minimization and proposed method. In each try, we produce a $m/6$-sparse, $N = 512$ sample long unit norm signal. The measurement matrix, A, is produced as a Gaussian random matrix. Performance results are compared in term of both the embedding rate, $\frac{M}{m}$ and measurement rates $\frac{m}{N}$. Figure 5.5 shows that the proposed method achieves $\mathcal{R} \leq 0.15$ bit/measurement, $\text{Prob}(w \neq \hat{w}) \to 0$, when $\frac{m}{N} = 0.2$. On the other hand, ℓ_1 achieves $\mathcal{R} \leq 0.1$ bit/measurement. When we increase the number of measurement, it is shown that ℓ_1 minimization algorithm fails. Since we fixed $k = m/6$, and increasing in m leads to increasing in k, and if we look (5.6), it is understood that minimum number of measurement to succeed is higher in classical ℓ_1 algorithm.

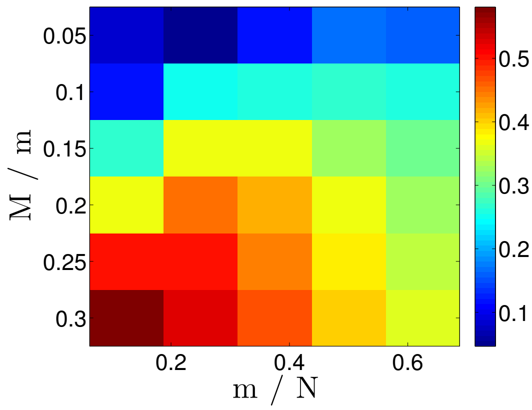


(a) (b)

Figure 5.5. Heat-maps of watermark error probability over measurement rate m/N and embedding rate M/m under AWGN, blue pixels represents $\text{Prob}(w \neq \hat{w}) \to 0$. (a) Performance of proposed algorithm in $\text{Prob}(w \neq \hat{w})$. (b) Performance of ℓ_1 minimization algorithm in $\text{Prob}(w \neq \hat{w})$.



(a)



(b)

Figure 5.6. Heat-maps of mean square reconstruction error over measurement rate m/N and embedding rate M/m under AWGN, blue pixels represents smaller values in $\mathbb{E}\{\|x-\hat{x}\|_2\}$. (a) Performance of proposed algorithm in $\mathbb{E}\{\|x-\hat{x}\|_2\}$. (b) Performance of ℓ_1 minimization algorithm in $\mathbb{E}\{\|x-\hat{x}\|_2\}$.

5.4. Decoding under contamination of both Impulsive and Gaussian Noise (AWGN)

If watermarked measurements are corrupted by the additive combination of both impulsive noise and additive white Gaussian noise, we can model this system as

$$y_n = Ax + Bw + z_I + z_G, \tag{5.29}$$

where $z_G \sim \mathcal{N}(0, \sigma_G^2 I)$ and z_I is L-sparse noise.

5.4.1. ℓ_1 Minimization

One possible solution for this system is to use ℓ_1 minimization of a $(k+M+L)$-nearly sparse vector $[x \quad w \quad z_I]^{\mathrm{T}}$ of length $N+M+m$. This can be done in two steps.

First, the vector $[\tilde{x} \;\; \tilde{w} \;\; \tilde{z}_I]^{\mathrm{T}}$ can be found via

$$\begin{bmatrix} \tilde{x} \\ \tilde{w} \\ \tilde{z}_I \end{bmatrix} = \arg \min_{[x \;\, w \;\, z]^{\mathrm{T}}} \left\| \begin{bmatrix} x \\ w \\ z \end{bmatrix} \right\|_{\ell_1} \quad \text{s.t.} \quad \left\| y_n - \begin{bmatrix} A \mid B \mid I \end{bmatrix} \begin{bmatrix} x \\ w \\ z \end{bmatrix} \right\|_{\ell_2} \leq \epsilon_1. \tag{5.30}$$

Secondly, an additional step can be performed as we did in Section 5.3.3. Since it is known that watermark information w_i is either $-a$ or $+a$, $\hat{w}$ can be estimated as $\hat{w}_i = a * \mathrm{sgn}(\tilde{w}_i)$. Then, $\hat{x}$ can be estimated via

$$\hat{x} = \arg \min_{x} \;\; \|x\|_{\ell_1} \quad \text{s.t.} \quad \|(y_n - B\hat{w} - \tilde{z}_I) - Ax\|_{\ell_2} \leq \epsilon_1. \tag{5.31}$$

However, even if it is possible to reconstruct k-sparse signal x and M bit watermark w using Equation (5.30) and (5.31), it is not optimal set-up as we have discussed before, because w is not a sparse vector.

5.4.2. Proposed Method

The proposed decoding algorithm can be decomposed into three steps: a) We construct a matrix F which annihilates the matrix B on the left, i.e., $FB = 0$. Then applying F to $y_n = Ax+Bw+z_I+z_G$, gives $\tilde{y} = F(Ax+Bw+z_I+z_G) = FAx+Fz_I+\tilde{z}$, where $\tilde{z} = Fz_G$. Then, $[\tilde{x} \;\; \tilde{z}_I]^{\mathrm{T}}$ can be estimated via

$$\begin{bmatrix} \tilde{x} \\ \tilde{z}_I \end{bmatrix} = \arg \min_{[x \;\, z]^{\mathrm{T}}} \left\| \begin{bmatrix} x \\ z \end{bmatrix} \right\|_{\ell_1} \quad \text{s.t.} \quad \left\| \tilde{y} - \begin{bmatrix} FA \mid F \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} \right\|_{\ell_2} \leq \epsilon_1. \tag{5.32}$$

b) In the second step, since B is a tall matrix, if we use $\tilde{x}$ that is found in step one, watermark $\tilde{w}$ can be estimated using least-squares method as follows:

$$\tilde{w} = (B^{\mathrm{T}}B)^{-1}B^{\mathrm{T}}(y_n - A\tilde{x} - \tilde{z}_I). \tag{5.33}$$

c) Finally, since it is known that watermark information w_i is either $-a$ or $+a$, $\hat{w}$ can be estimated using $\hat{w}_i = a * \text{sgn}(\tilde{w}_i)$ and $\hat{x}$ can be found via

$$\hat{x} = \arg\min_{x} \ \|x\|_{\ell_1} \quad \text{s.t.} \quad \|(y_n - B\hat{w} - \tilde{z}_I) - Ax\|_{\ell_2} \leq \epsilon_1. \tag{5.34}$$

Input: y_n, A, B;

Determine: ϵ_1

1. Apply F to y_n : $\tilde{y} = F y_n$

2. Estimate $\tilde{x}$: $\begin{bmatrix} \tilde{x} \\ \tilde{z}_I \end{bmatrix} = \arg\min_{[x \ z]^{\mathrm{T}}} \left\| \begin{bmatrix} x \\ z \end{bmatrix} \right\|_{\ell_1}$ s.t. $\left\| \tilde{y} - \begin{bmatrix} FA \mid F \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} \right\|_{\ell_2} \leq \epsilon_1$;

3. Estimate $\tilde{w}$: $\tilde{w} = (B^{\mathrm{T}}B)^{-1}B^{\mathrm{T}}(y_n - A\tilde{x} - \tilde{z}_I)$

4. Threshold $\tilde{w}$: $\hat{w}_i = a * \text{sgn}(\tilde{w}_i)$

5. Estimate $\hat{x} = \arg\min_x \ \|x\|_{\ell_1}$ s.t. $\|(y_n - B\hat{w} - \tilde{z}_I) - Ax\|_{\ell_2} \leq \epsilon_1$

Output: $\hat{x}$, $\hat{w}$

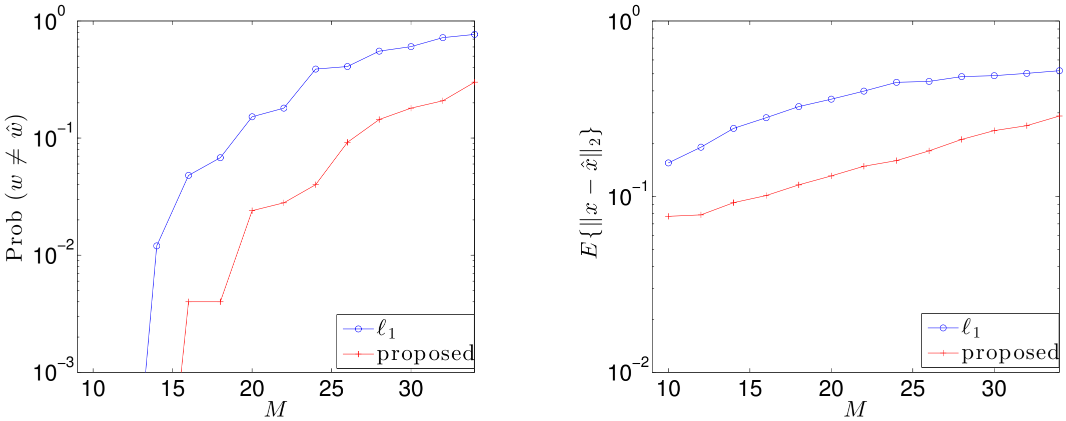
Figure 5.7. Modified Reconstruction Algorithm against Impulsive Noise.

5.4.3. Experiments and Results for the Measurements are Corrupted by Impulsive Noise

In this section, we present performance results of the proposed decoding algorithms for the case where the watermarked measurements are corrupted by sum of AWGN and Impulsive Noises. In our experiments, we generate x, w vectors such that $\|x\|_{\ell_2} = 1$, $\|w\|_{\ell_2} = 0.25$ ($a = \frac{0.25}{\sqrt{M}}$), and produce measurement matrix A as a Gaussian random matrix. We also generate a F with orthogonal rows. Then, columns of the encoding matrix B are obtained from span of null space of F. Each experiment is conducted 250 times, and corresponding $\text{Prob}(w \neq \hat{w})$ and $\mathbb{E}\{\|x - \hat{x}\|_2\}$ values are reported.

Experiment 1: The first experiment is done similar to Experiment 1 in Section

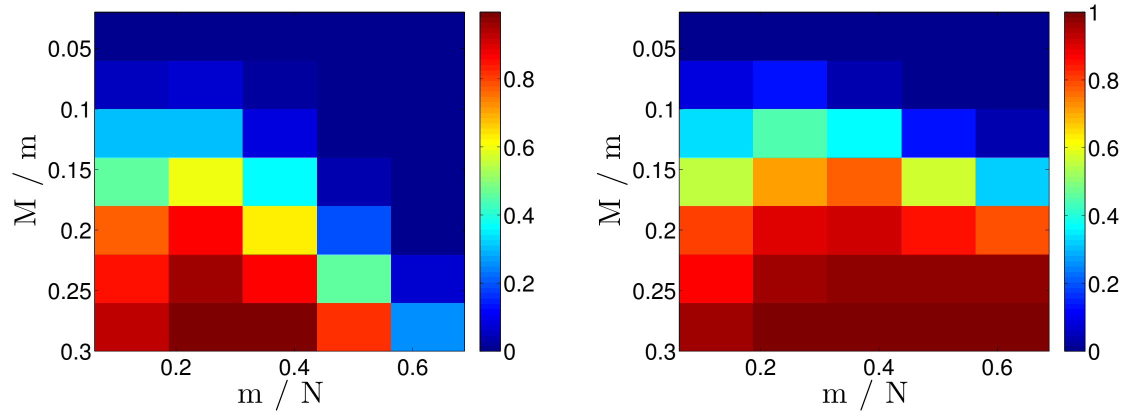
5.3.6. But, this time a impulsive noise is applied in addition to AWGN. Number of measurements, m, is fixed to 145. M bit long watermark is embedded onto these measurements. The sparsity, k, is also fixed as 30 for a $N = 512$ length signal. We set the noise levels in Equation (5.29) as $\sigma_G = 0$ and $\|z_I\|_{\ell_2} = 10$ where z_I is $L = 5$-sparse impulsive noise. Fig. 5.8 shows that the proposed method in Section 5.4.2 achieves $\mathcal{R} \leq 16/145$ bit/measurement and Fig. 5.8 shows $\mathbb{E}\{\|x - \hat{x}\|_2\} \leq 6 * 10^{-2}$.



(a) (b)

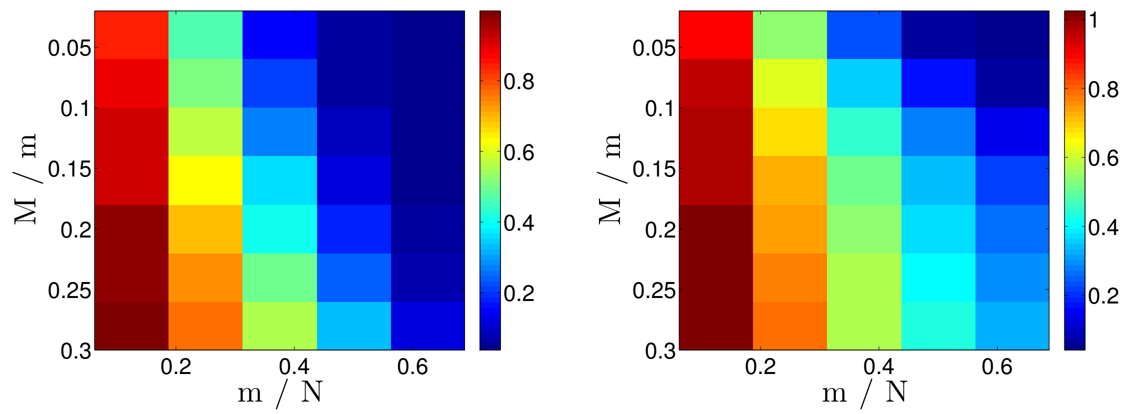
Figure 5.8. Performance comparison of algorithms for the case in which measurements are corrupted by impulsive noise (a) M vs. Prob$(w \neq \hat{w})$. (b) M vs. $\mathbb{E}\{\|x - \hat{x}\|_2\}$.

Experiment 2: This experiment is also conducted similarly with the Experiment 2 in Section 5.3.6. The number of measurements, m, is also varying compared to previous experiment. We compare ℓ_1 minimization algorithm and proposed method in terms of both the embedding rate, $\frac{M}{m}$, and measurement rates $\frac{m}{N}$. But this time, we set $\sigma_G = 0$ to apply only sparse noise. AWGN noise is eliminated for this experiment, since ℓ_1 minimization has already failed under AWGN, see Figure 5.5. As it is seen in Figure 5.9 and Figure 5.10, ℓ_1 minimization performs a bit better compared to the situation that AWGN exists. However, our proposed method still outperforms the ℓ_1 method.



(a) (b)

Figure 5.9. Heat-maps of watermark error probability over measurement rate m/N and embedding rate M/m under 5-sparse impulsive noise, blue pixels represents $\text{Prob}(w \neq \hat{w}) \rightarrow 0$. (a) Performance of proposed algorithm in $\text{Prob}(w \neq \hat{w})$. (b) Performance of ℓ_1 minimization algorithm in $\text{Prob}(w \neq \hat{w})$.



(a) (b)

Figure 5.10. Heat-maps of mean square reconstruction error over measurement rate m/N and embedding rate M/m under 5-sparse impulsive noise, blue pixels represents smaller values in $\mathbb{E}\{\|x - \hat{x}\|_2\}$. (a) Performance of proposed algorithm in $\mathbb{E}\{\|x - \hat{x}\|_2\}$. (b) Performance of ℓ_1 minimization algorithm in $\mathbb{E}\{\|x - \hat{x}\|_2\}$.

6. REGULARIZING THE ENCODING AND MEASUREMENT MATRICES FOR A FAST COMPRESSIVE SENSING AND WATERMARKING SCHEME

In this chapter, we describe a fast compressive sensing and watermarking scheme. We will show that although embedding and reconstruction algorithm remains the same as described in Chapter 5, regularizing the encoding matrix and measurement matrix yields a fast algorithm for large scale, real-time applications. At the end of this chapter we introduce an effective method to adjust the measurement and reconstruction processes for real-time applications. Although we will intensely focus on a specific type of signal, images, the idea can be applicable to other type of signals.

From the outset, we have discussed random matrices and their usage in CS. As a matter of fact, choosing the measurement matrix as a random matrix is optimum preference to a certain degree. The number of required measurement, m, is nearly optimal when random matrices are used [58]. However, when it comes to computational complexity and memory usage on reconstruction part, working with random matrices is cumbersome. As an example, consider we have a 512×512 pixels image to sense. Assume that we will take 90000 measurements from $N = 512^2$ length vectorized image. In this case, the measurement matrix will be 90000×262144, that will require more than 3 gigabytes to store. On the other hand, the reconstruction algorithms presented in Chapter 4 require to compute transpose of the measurement matrix, Φ, several times. This brings a computational burden to the system when measurement matrices are of large size. One approach to surmount this problem is to pick a sensing matrix which consists of randomly chosen rows from an orthonormal basis such as Fourier basis [20]. Although, constructing such a sensing scheme leads to fast implementations on sensing and reconstruction part, guaranteeing theoretically its RIP is rather challenging [59]. In such cases, it is more preferable to deal with a easily computable guarantee condition such as ”coherence”.

To clarify the concept of coherence, let us recall the reconstruction problem in the typical compressive sensing setup: Given m-length observation of an N dimensional signal S, reconstruct k-sparse representation x of the signal S, provided that $m \leq N$. For sampling, we now discuss a special form of the measurement matrix given in (2.2). Mathematically speaking, let $\Omega_1 \in \{1, 2, 3, ..., N\}$ be a subset of indices indexing the location of the chosen rows from an orthonormal basis, Θ, with $|\Omega_1| = m$. In this case, the measurement vector, y will be

$$y = \Theta_{\Omega_1} S = \Theta_{\Omega_1} \psi x = Ax, \tag{6.1}$$

where Θ_{Ω_1} is the measurement matrix consisting of the rows picked from Θ indexed by Ω_1. Obviously, if the location of the non-zero coefficients is unknown, sensing individual coefficients of the vector x may not be a proper strategy (It is equivalent to adjust the matrix A as identity matrix), since this sampling method requires $m = N$ measurements to collect enough information about the signal. Therefore, it is expected from a good sensing scheme that the rows of the matrix should not be too much concentrated so that each element of a row of the measurement matrix can touch every coefficients of the signal when multiplying, therefore every measurement can contain enough information about whole signal. In other words, the rows of the matrix A should be as flat as possible, that is the rows of A should not be sparse or compressible. More formally, let an $N \times N$ matrix $U = \Theta\psi$. Define a functional $\mu(U)$ as

$$\mu(U) = \max_{i,j} |U_{i,j}| \tag{6.2}$$

which is used to loosely quantify how overly distributed the rows of U are [60]. Since $U_{i,j} = \left\langle \Theta_i^T, \psi_j \right\rangle$, the parameter $\mu(U)$ is formally defined follows:

Definition 6.1 (Mutual Coherence [61])**.** *The mutual coherence between the sampling basis, Θ, and the sparsifying basis, ψ, is defined as (both unit norm)*

$$\mu(\Theta, \psi) = \max_{1 \leq i,j \leq N} \left| \left\langle \Theta_i^T, \psi_j \right\rangle \right|. \tag{6.3}$$

By elementary linear algebra, it can be seen that $\frac{1}{\sqrt{N}} \leq \mu(\Theta, \psi) \leq 1$. As we discussed above, compressive sensing framework requires small $\mu(\Theta, \psi)$, or in other words *incoherent basis pairs.* The pair spikes and sinusoids can be given as an example of incoherent pairs [61]. Indeed, this pair of basis sets, are maximally incoherent, since $\mu(Spikes, Sinusoids) = \frac{1}{\sqrt{N}}$. Candes *et al.* [60] give a lower bound on the number of measurement in terms of coherence between sparsifying basis and sampling basis:

Theorem 6.2 ([60])**.** *Given a fixed $S \in \mathbb{R}^N$, which has k-sparse representation, x in a basis ψ. Pick a subset Ω_1 of measurements domain from the sampling basis, Θ, with $|\Omega_1| = m$. If*

$$m \geq C.N.\mu^2(\Theta, \psi).k.\log N, \tag{6.4}$$

where C is a positive constant. Then, the reconstruction, $\hat{x}$ of the signal x is exact via

$$\hat{x} = \min_x \quad \|x\|_{\ell_1^N} \quad s.t. \quad y = \Theta_{\Omega_1}\psi x \tag{6.5}$$

with overwhelming probability.

6.1. Noiselets as Representation Basis

As another example for incoherent pairs, it is known that the sparsifying basis, wavelet basis, is incoherent with noiselets [62]. Noiselets are noise-like functions that are completely uncompressible with wavelet decompositions. Therefore, noiselet basis is maximally incoherent with wavelet basis. The coherence between noiselets and Haar wavelet is $\sqrt{\frac{2}{N}}$ [63]. The coherence of noiselets between Doubechies $D4$ and $D8$ also given as, respectively, $\sqrt{\frac{2.2}{N}}$ and $\sqrt{\frac{2.9}{N}}$ [1]. Besides being incoherent with wavelets, noiselets also have fast implementations. These properties make the noiselets a good choice for a sensing basis especially for large-scale applications. As an example for large-scale signals, we consider the 512×512 image shown in Figure 6.1. This test image is synthetic: This image was transformed into wavelet domain, and the largest coefficients are kept and remaining ones are zeroed out, where $N = 512^2$. Then, an image was reconstructed by taking the inverse wavelet transform of the resulting coefficients. This constitutes our k-sparse signal, $k \leq 0.05 * N = 13107$. Then, 70000 rows are randomly chosen from the rows of Noiselet basis to constitute a measurement matrix, Θ_{Ω_1}. It is recorded that the reconstruction $\hat{x}$ of the wavelet coefficients are exact by using (6.5). The reconstructed image is also shown in Figure 6.1.

(a)

(b)

Figure 6.1. Signal recovery from Noiselet measurements (a) $N = 512^2$ length $k = N * 0.05 = 13107$-sparse synthetic image. (b) Reconstructed image from 70000 Noiselet measurements. Image is recovered with $PSNR > 52dB$

6.2. The Fast Watermarking Scheme using Noiselets

We have discussed the necessity of using a orthonormal matrix as sampling basis in order to attain a low complexity sensing framework while dealing with large scale, real-time signals. In this section, we describe a method to adjust the embedding scheme that is proposed in Chapter 5.

The idea is as follows: Choose the encoding matrix as a subset of columns from another orthonormal matrix; Thus, we choose the columns of the encoding matrix from the columns of a orthonormal basis, Ξ. These columns are uniformly randomly picked with a subset $\Omega_2 \subset \{1, 2, 3, .., m\}$ of indices indexing the location of the chosen columns, with $|\Omega_2| = M$. From now on, we define this encoding matrix as Ξ_{Ω_2}. Therefore, the

m length watermarked measurements in (5.2) turn to be

$$y_w = \Theta_{\Omega_1} S + \Xi_{\Omega_2} w, \tag{6.6}$$

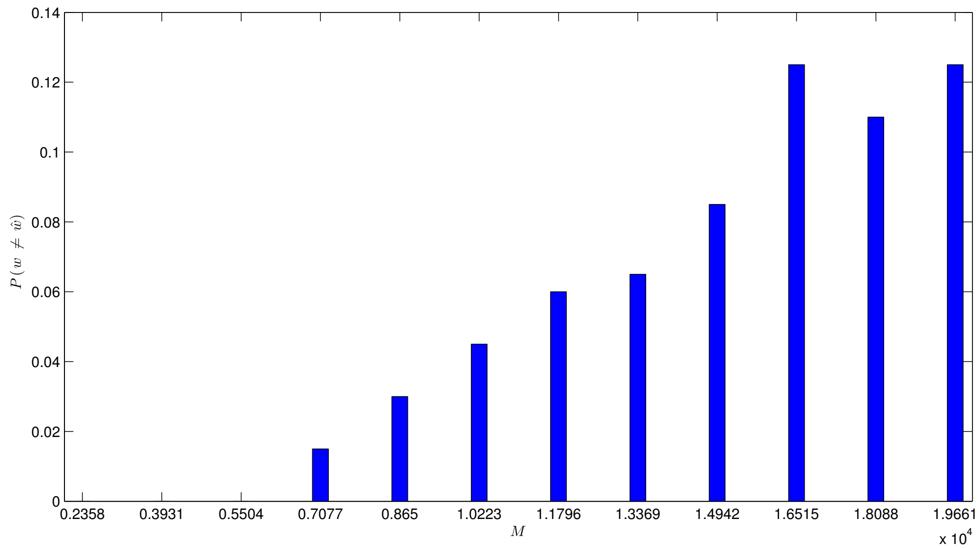
where S is the N length signal to sense and w is M length watermark that we wish to embed.

In the reconstruction stage, the noisy measurements, $y_n = y_w + z$, are received. Let us recall the proposed reconstruction algorithm of the signal and of the watermark (5.3). In that algorithm we began with constructing a annihilator matrix which annihilate the watermarking information before finding first-tier estimation $\tilde{x}$ of x. This matrix was left annihilator of the encoding matrix B. This time constructing this annihilator matrix is also straightforward:

Suppose that $\Omega_3 \subset \Omega_2^c$, where $\Omega_2^c = \{1, 2, 3, .., m\} \backslash \Omega_2$. Let F_{Ω_3} be the $p \times m$ matrix corresponding to rows of F_{Ω_3} are chosen from the columns of the orthonormal basis Ξ indexed by Ω_3. Then it is apparent that the matrix F_{Ω_3} will be the left annihilator of the matrix Ξ_{Ω_2}, i.e., such that $F_{\Omega_3} \Xi_{\Omega_2} = 0$. In plain English, one may use a subset of columns of a basis Ξ as encoding matrix. Then a subset from unused columns can be used as the rows of the annihilator matrix.

The advantage of adjusting the measurement matrix, the encoding matrix and annihilator matrix can be explained in two steps: (i) this scheme requires low-complexity sensing and reconstruction processes, (ii) encoder and decoder sections do not need to store whole matrices; knowledge of Ω_1, Ω_2 and Ω_3 is sufficient for reconstruction.

As an example we choose encoding basis Ξ as Discrete Hartley Transform (DHT) and sensing basis Θ as noiselelet basis. We tested the decoding algorithm for the case where the watermarked measurements are corrupted by AWGN. In our experiments we use 512×512 Lena Image, $S \in \mathbb{R}^N$, where $N = 512^2$. Observation vector are obtained by using noiselet basis, i.e, such that $y = \Theta_{\Omega_1} S$. We fixed the number of the noiselet measurement to $m = N * 0.3$. And also norms of the signal and watermark are normalized to $\|S\|_{\ell_2} = 1$, $\|w\|_{\ell_2} = 0.25$, respectively. M-bit length watermark is embedded onto measurements, by using DHT for each try. We look for maximum achievable embedding rate $\mathcal{R}$ in bits/measurement when $\text{Prob}(w \neq \hat{w}) \to 0$. We consider the Gaussian noise with 32dB signal-to-noise ration (SNR). SNR in dB is defined as $20\log_{10}(\frac{\|\Theta_{\Omega_1}S+\Xi_{\Omega_2}w\|_2}{\|z\|_2})$. Figure 6.2 shows the $\text{Prob}(w \neq \hat{w})$, while decreasing the number of bit M. For each value of M, 200 trials are conducted.



(a)

Figure 6.2. Performance of the proposed decoding algorithm under AWGN. M vs. $\text{Prob}(w \neq \hat{w})$. $\text{Prob}(w \neq \hat{w})$ is calculated after 200 trials.

7. CONCLUSION

Most of the real-world signals we encounter in applications have low information content. In other words, signals can be well approximated by sparse signals in a proper basis. CS uses this fact and attempt to sense signals by using far fewer measurements than the conventional acquisition systems uses. Although data acquisition process has low-complexity by using linear measurements, reconstruction process in which generally non-linear reconstruction algorithms are used can be relatively costly. Because of this reason, there has been an interest in sending the compressive sensing measurements directly to a receiver instead of recovering the signal in sensor part. In such cases, one may also wish to embed useful information called watermark on compressively sensed measurements. In this thesis, we investigated possibility of such an embedding. To our knowledge, this work is the first study investigating such a scheme.

In Chapter 5, we proposed a straightforward embedding scheme in which watermarks information is spread over the compressive sensed measurements by using a encoded matrix. In the same chapter we also proposed a reconstruction algorithm that recovers jointly the signal and the watermark. This algorithm simply uses the existing reconstruction algorithms which are discussed in Chapter 4, successively.

Theoretical justifications of guarantee of the proposed algorithm was also investigated in Section 5.3.4. The guarantee condition of the proposed algorithm was roughly given based on restricted isometry property which is discussed in details in Chapter 3. This theoretical justification shows us that proposed type of embedding scheme is possible and experimental results verify this results.

In Chapter 6 we adjusted measurement, watermarking and reconstruction processes to make the proposed scheme suitable for real-time, large scale applications. As an example, a specific measurement process via noiselet functions was overviewed. In addition to measurement process, watermark embedding process also revised by setting the encoding matrices as a subset of the columns of a determined orthonormal basis.

APPENDIX A: PUBLICATIONS

1. Yamac, M., C. Dikici and B. Sankur, "Watermarking of Compressive Sensed Measurements", SPARS 2013, Lausanne, 2013.

2. Yamac, M., C. Dikici and B. Sankur, "Robust watermarking of Compressive Sensed signals", Signal Processing and Communications Applications Conference (SIU), 2013 21st, pp. 1-4, 2013.

3. Yamac, M., C. Dikici and B. Sankur, "Robust Watermarking of Compressive Sensed Measurements Under Impulsive and Gaussian Attacks", EUSIPCO 2013, Marrakech, 2013.

REFERENCES

1. Candes, E. J. and M. B. Wakin, "An Introduction To Compressive Sampling", *IEEE Signal Processing Magazine*, Vol. 25, No. 2, pp. 21–30, 2008.

2. Baraniuk, R., "Compressive Sensing [Lecture Notes]", *Signal Processing Magazine, IEEE*, Vol. 24, No. 4, pp. 118–121, 2007.

3. Kutyniok, G., "Compressed Sensing: Theory and Applications", *Clinical Orthopaedics and Related Research (CORR)*, Vol. abs/1203.3815, 2012.

4. Bruckstein, A. M., D. L. Donoho and M. Elad, "From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images.", *SIAM Review*, Vol. 51, No. 1, pp. 34–81, 2009.

5. Davenport, M. A., M. F. Duarte, Y. C. Eldar and G. Kutyniok, *Introduction to Compressed Sensing*, Cambridge University Press, 2012.

6. Mamaghanian, H., N. Khaled, D. Atienza and P. Vandergheynst, "Compressed Sensing for Real-Time Energy-Efficient ECG Compression on Wireless Body Sensor Nodes", *Biomedical Engineering, IEEE Transactions on*, Vol. 58, No. 9, pp. 2456–2466, 2011.

7. Robucci, R. W., J. D. Gray, L. K. Chiu, J. K. Romberg and P. E. Hasler, "Compressive Sensing on a CMOS Separable-Transform Image Sensor.", *Proceedings of the IEEE*, Vol. 98, No. 6, pp. 1089–1101, 2010.

8. Cox, I. J., M. L. Miller and J. A. Bloom., *Digital Watermarking and Steganography*, The Morgan Kaufmann Series in Multimedia Information and Systems, Morgan Kaufmann Publishers, Burlington (Mass.), 2008.

9. Sheikh, M. and R. Baraniuk, "Blind Error-Free Detection of Transform-Domainwatermarks", *Image Processing, 2007. ICIP 2007. IEEE International*

Conference on, Vol. 5, pp. V – 453–V – 456, 2007.

10. Yamac, M., C. Dikici and B. Sankur, "Watermarking of Compressive Sensed Measurements", *SPARS 2013, Lausanne*, 2013.

11. Yamac, M., C. Dikici and B. Sankur, "Robust Watermarking of Compressive Sensed Measurements Under Impulsive and Gaussian Attacks", *EUSIPCO 2013, Marrakech*, 2013.

12. Adcock, B. and A. C. Hansen, "Generalized Sampling and Infinite-Dimensional Compressed Sensing", *Technical report NA2011/02, DAMTP, University of Cambridge*, 2011.

13. Boyd, S. and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, NY, USA, 2004.

14. Mallat, S., *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*, Academic Press, 3rd edn., 2008.

15. Pennebaker, W. B. and J. L. Mitchell, *JPEG Still Image Data Compression Standard*, Kluwer Academic Publishers, Norwell, MA, USA, 1st edn., 1992.

16. Wallace, G. K., "The JPEG still picture compression standard", *Communications of the ACM*, pp. 30–44, 1991.

17. Lee, J., A. Bandyopadhyay, I. Faik Baskaya, R. Robucci and P. Hasler, "Image Processing System using a Programmable Transform Imager", *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, Vol. 5, pp. v–101, IEEE, 2005.

18. Romberg, J., "Imaging via Compressive Sampling", *Signal Processing Magazine, IEEE*, Vol. 25, No. 2, pp. 14–20, 2008.

19. Donoho, D. L., "Compressed Sensing", *IEEE Trans. Inform. Theory*, Vol. 52, pp.

1289–1306, 2006.

20. Candès, E. J., J. K. Romberg and T. Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information.", *IEEE Transactions on Information Theory*, Vol. 52, No. 2, pp. 489–509, 2006.

21. Tropp, J. A., "Just relax: Convex Programming Methods for Identifying Sparse Signals in Noise", *Information Theory, IEEE Transactions on*, Vol. 52, No. 3, pp. 1030–1051, 2006.

22. Donoho, D. L. and M. Elad, "Optimally Sparse Representation in General (Nonorthogonal) Dictionaries via ℓ_1 minimization", *Proc. Natl Acad. Sci. USA 100 2197–202*, 2005.

23. Rauhut, H., "Compressive Sensing and Structured Random Matrices", *Theoretical Foundations and Numerical Methods for Sparse Recovery*, Vol. 9, pp. 1–92, 2010.

24. Cohen, A., W. Dahmen and R. DeVore, "Compressed Sensing and Best k-term Approximation", *Journal of the American Mathematical Society*, Vol. 22, No. 1, pp. 211–231, 2009.

25. Zhang, Y., "Theory of Compressive Sensing via ℓ_1-Minimization: a Non-RIP Analysis and Extensions", *Journal of the Operations Research Society of China*, pp. 1–27, 2013.

26. Elad, M., *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer Publishing Company, Incorporated, 1st edn., 2010.

27. Candès, E. J. and T. Tao, "Decoding by Linear Programming", *IEEE Transactions on Information Theory*, Vol. 51, No. 12, pp. 4203–4215, 2005.

28. Candès, E. J., "The Restricted Isometry Property and Its Implications for Compressed Sensing", *Comptes Rendus Mathematique*, Vol. 346, No. 9, pp. 589–592,

2008.

29. Foucart, S., "A Note on Guaranteed Sparse Recovery via ℓ_1-Minimization", *Applied and Computational Harmonic Analysis*, Vol. 29, No. 1, pp. 97 – 103, 2010.

30. Baraniuk, R., M. Davenport, R. DeVore and M. Wakin, "A Simple Proof of the Restricted Isometry Property for Random Matrices", *Constructive Approximation*, Vol. 28, No. 3, pp. 253–263, 2008.

31. Johnson, W. and J. Lindenstrauss, "Extensions of Lipschitz Mappings into a Hilbert Space", *Conference in Modern Analysis and Probability (New Haven, Conn., 1982)*, Vol. 26, pp. 189–206, 1984.

32. Matoušek, J., "On Variants of the Johnson–Lindenstrauss Lemma", *Random Structures & Algorithms*, Vol. 33, No. 2, pp. 142–156, 2008.

33. Frankl, P. and H. Maehara, "The Johnson-Lindenstrauss Lemma and the Sphericity of Some Graphs", *Journal of Combinatorial Theory, Series B*, Vol. 44, No. 3, pp. 355–362, 1988.

34. Indyk, P. and R. Motwani, "Approximate Nearest Neighbors: towards Removing the Curse of Dimensionality", *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, pp. 604–613, ACM, 1998.

35. Achlioptas, D., "Database-Friendly Random Projections: Johnson-Lindenstrauss with Binary Coins", *Journal of Computer and System Sciences*, Vol. 66, No. 4, pp. 671–687, Jun. 2003.

36. Dasgupta, S. and A. Gupta, "An Elementary Proof of a Theorem of Johnson and Lindenstrauss", *Random Structures & Algorithms*, Vol. 22, No. 1, pp. 60–65, 2003.

37. Vershynin, R., "Introduction to the Non-Asymptotic Analysis of Random Matrices", *Clinical Orthopaedics and Related Research (CORR)*, Vol. abs/1011.3027, 2010.

38. Tropp, J. and S. Wright, "Computational Methods for Sparse Solution of Linear Inverse Problems", *Proceedings of the IEEE*, Vol. 98, No. 6, pp. 948–958, 2010.

39. Claerbout, J. F. and F. Muir, "Robust Modeling with Erratic Data", *Geophysics*, Vol. 38, No. 5, pp. 826–844, 1973.

40. Banks, T. H. L., Stephen C., "A Modification to the Discrete l_1 Linear Approximation Algorithm of Barrodale and Roberts", *SIAM Journal on Scientific and Statistical Computing*, Vol. 1, No. 5, pp. 826–844, 1980.

41. Chen, S. S., D. L. Donoho, Michael and A. Saunders, "Atomic Decomposition by Basis Pursuit", *SIAM Journal on Scientific Computing*, Vol. 20, pp. 33–61, 1998.

42. Candès, E. J., J. K. Romberg and T. Tao, "Stable Signal Recovery from Incomplete and Inaccurate Measurements", *Communications on Pure and Applied Mathematics*, Vol. 59, No. 8, pp. 1207–1223, 2006.

43. Haupt, J. and R. Nowak, "Signal Reconstruction from Noisy Random Projections", *Information Theory, IEEE Transactions on*, Vol. 52, No. 9, pp. 4036–4048, 2006.

44. Laska, J. N., M. A. Davenport and R. G. Baraniuk, "Exact Signal Recovery from Sparsely Corrupted Measurements through the Pursuit of Justice", *Proceedings of the 43rd Asilomar Conference on Signals, Systems and Computers*, Asilomar'09, pp. 1556–1560, 2009.

45. Carrillo, R. E., K. E. Barner and T. C. Aysal, "Robust Sampling and Reconstruction Methods for Sparse Signals in the Presence of Impulsive Noise", *IEEE Journal on Selected Topics in Signal Processing*, Vol. 4, No. 2, pp. 392–408, 2010.

46. Tibshirani, R., "Regression Shrinkage and Selection Via the Lasso", *Journal of the Royal Statistical Society, Series B*, Vol. 58, pp. 267–288, 1994.

47. Figueiredo, M. A. T., R. Nowak and S. Wright, "Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems",

IEEE Journal on Selected Topics in Signal Processing, Vol. 1, No. 4, pp. 586–597, 2007.

48. Candes, E. and T. Tao, "The Dantzig Selector: Statistical Estimation when p is Much Larger Than n", *The Annals of Statistics*, Vol. 35, No. 6, pp. 2313–2351, 2007.

49. Asif, M. S. and J. Romberg, "On the Lasso and Santzig Selector Equivalence", *Information Sciences and Systems (CISS), 2010 44th Annual Conference on*, pp. 1–6, IEEE, 2010.

50. Cai, T. T. and L. Wang, "Orthogonal Matching Pursuit for Sparse Signal Recovery With Noise", *Information Theory, IEEE Transactions on*, Vol. 57, No. 7, pp. 4680–4688, 2011.

51. Wang, J. and B. Shim, "On the Recovery Limit of Sparse Signals Using Orthogonal Matching Pursuit", *Signal Processing, IEEE Transactions on*, Vol. 60, No. 9, pp. 4973–4976.

52. Takhar, D., J. N. Laska, M. B. Wakin, M. F. Duarte, D. Baron, S. Sarvotham, K. F. Kelly and R. G. Baraniuk, "A new Compressive Imaging Camera Architecture using Optical-Domain Compression", *Electronic Imaging 2006*, pp. 606509–606509, International Society for Optics and Photonics, 2006.

53. Barni, M. and F. Bartolini, *Watermarking Systems Engineering: Enabling Digital Assets Security and Other Applications*, CRC Press, 2004.

54. Kamilov, U., V. K. Goyal and S. Rangan, "Optimal Quantization for Compressive Sensing under Message Passing Reconstruction", *Clinical Orthopaedics and Related Research (CORR)*, Vol. abs/1102.4652, 2011.

55. Lee, S., C. D. Yoo and T. Kalker, "Reversible Image Watermarking Based on Integer-to-Integer Wavelet Transform.", *IEEE Transactions on Information Foren-*

sics and Security, Vol. 2, pp. 321–330, 2007.

56. Hong, W., T.-S. Chen and H.-Y. Wu, "An Improved Reversible Data Hiding in Encrypted Images using Side Match", *Signal Processing Letters, IEEE*, Vol. 19, No. 4, pp. 199–202, 2012.

57. Coatrieux, G., W. Pan, N. Cuppens-Boulahia, F. Cuppens and C. Roux, "Reversible Watermarking Based on Invariant Image Classification and Dynamic Histogram Shifting", *IEEE Transactions on Information Forensics and Security*, Vol. 8, No. 1, pp. 111–120, 2013.

58. Donoho, D. L., "For Most Large Underdetermined Systems of Linear Equations the Minimal ℓ_1-norm Solution is also the Sparsest Solution", *Communications on Pure and Applied Mathematics*, Vol. 59, No. 6, pp. 797–829, 2006.

59. Candès, E. J. and Y. Plan, "A Probabilistic and RIPless Theory of Compressed Sensing", *Clinical Orthopaedics and Related Research (CORR)*, Vol. abs/1011.3854, 2010.

60. Candes, E. and J. Romberg, "Sparsity and Incoherence in Compressive Sampling", *Inverse Problems*, Vol. 23, No. 3, p. 969, 2007.

61. Donoho, D. L. and X. Huo, "Uncertainty Principles and Ideal Atomic Decomposition", *IEEE Transactions on Information Theory*, Vol. 47, No. 7, pp. 2845–2862, 2001.

62. Coifman, R., F. Geshwind and Y. Meyer, "Noiselets", *Applied and Computational Harmonic Analysis*, Vol. 10, No. 1, pp. 27–44, 2001.

63. Tuma, T., P. Hurley *et al.*, "On the Incoherence of Noiselet and Haar Bases", *SAMPTA'09, International Conference on Sampling Theory and Applications*, 2009.