GALATASARAY UNIVERSITY

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

A NOVEL DESIGN METHOD FOR COMPRESSIVE SENSING MATRICES

Utku Kabuli AYTAÇ

August 2020

A NOVEL DESIGN METHOD
FOR COMPRESSIVE SENSING MATRICES

(SIKIŞTIRMALI ALGILAMA MATRİSLERİ
İÇİN YENİ TASARIM YÖNTEMİ)

by

U T K U K A B U L İ A Y T A Ç, B.S.

Thesis

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

MASTER OF SCIENCE

in

MATHEMATICS

in the

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

of

GALATASARAY UNIVERSITY

August 2020

This is to certify that the thesis entitled

A Novel Design Method For Compressive Sensing Matrices

prepared by **Utku Kabuli AYTAÇ** in partial fulfillment of the requirements for the degree of **Master of Science in Mathematics** at the **Galatasaray University** is approved by the

Examining Committee :

Prof. Dr. A. Muhammed ULUDAĞ (Supervisor)
Department of Mathematics
Galatasaray University ____________________

Assoc. Prof. Ayşegül YILDIZ ULUS
Department of Mathematics
Galatasaray University ____________________

Prof. Dr. Ali Aydın SELÇUK
Department of Computer Engineering
TOBB University ____________________

Date: ____________________

Acknowledgements

I would like to begin with expressing my gratitude to my advisor Prof. Dr. A. Muhammed ULUDAĞ for his guidance, patience and encouragement at every stage during the last two years. Without his generous support, this thesis would not have been possible.

I would like to thank professor Hakan AYRAL for his support and his great contribution to my academic life.

I am also grateful to my family, and my friends for believing in me throughout my master's studies. Their positivity and support gave me the motivation I needed to succeed.

12 August 2020
Utku Kabuli AYTAÇ

Table of Contents

LIST OF SYMBOLS . . . . **vi**

LIST OF FIGURES . . . . **vii**

ABSTRACT . . . . **ix**

ÖZET . . . . **x**

1 Introduction **1**

1.1 Sampling 2

1.2 Sparsity 3

2 Literature Review **5**

3 Preliminaries **7**

4 Sparse Recovery **10**

5 Some Sensing Matrix Generation Methods **14**

6 Recovery Algorithms **16**

6.1 Optimization Methods 16

6.1.1 Null Space Property 19

6.1.2 Restricted Isometry Property 22

6.1.3 Experiments 24

6.1.4 Relation between Null Space Property and Restricted Isometry Property 24
6.2 Greedy Methods 32
6.2.1 Experiments 33
6.3 Thresholding-Based Methods 34
6.3.1 Experiments 35

7 Conclusion 36
7.1 A Novel Design Method for Sensing Matrices 36

REFERENCES 37

APPENDICES 41
7.2 Sensing Matrix Generation 41
7.2.1 Gaussian Matrix 41
7.2.2 Bernoulli Matrix 41
7.2.3 Hadamard Matrix 41
7.2.4 Toeplitz Matrix 42
7.3 Restricted Isometry Test Using Different Sensing Matrix Design Methods 42
7.4 Novel Design Method For Sensing Matrices 44
7.5 Experiment to Compare Sensing Matrices Using OMP and IHT Algorithms 45

BIOGRAPHICAL SKETCH 48

LIST OF SYMBOLS

$\mathbb{N}$	: $\{0,1,2,3,\ldots\}$ the set of natural numbers
Σ_k	: The set of all k-sparse vectors
$\|v\|_0$	: Number of nonzero coefficients of v
$\mathcal{N}(A)$	: Null Space of A
$\|v\|_1$	: ℓ_1-norm of v
$[N]$	: the set $\{1,2,3,\ldots N\}$
NSP	: Null Space Property
RIP	: Restricted Isometry Property

List of Figures

Figure 1.1: Schematic representation of a single-pixel camera (Image courtesy of Rice University) 1

Figure 6.1: Unit spheres in $\mathbb{R}^2$ for the ℓ_p norms with $p = 1, 2, \infty$ and for the ℓ_p quasinorm with $p = \frac{1}{2}$. 17

Figure 6.2: Best approximation of a point in $\mathbb{R}^2$ by a one-dimensional subspace using the ℓ_p norms for $p = 1, 2, \infty$ and the ℓ_p quasinorm with $p = 2$. 17

Figure 6.3: Example for ℓ_1 norm approximation 19

Figure 6.4: RIP test using the Gaussian Matrix 24

Figure 6.5: RIP test using the Hadamard Matrix 25

Figure 6.6: Recovery using the Random Gaussian Matrix and the Orthogonal Matching Pursuit algorithm 33

Abstract

In this thesis, we studied the mathematical foundations of Compressive Sensing. Compressive Sensing is an area which gives us more ability than Nyquist-Shannon Theorem with an extra condition: Sparsity. If a signal is sparse, we can recover the signal using fewer measurements than the required in Nyquist-Shannon Theorem.

Firstly, we examined the necessary conditions to use compressive sensing for recovery. Sparsity is the key to use compressive sensing for signal recovery. Besides, we look into the relationship between sparsity of a signal and sensing matrices.

Then, we look into recovery algorithms, sensing matrix design methods and properties of sensing matrices. Two most imporant properties of sensing matrices are Null Space Property and Restricted Isometry Property. We also examined the relationship between Null Space Property and Restricted Isometry Property.

Later, we made experiments using different sensing matrix generation methods. Lastly, we propose a novel design for sensing matrix generation and compared the results of these experiments with the other sensing matrix design methods using different recovery algorithms.

Keywords: Compressive Sensing, Signal Recovery, Null Space Property, Restricted Isometry Property.

Özet

Bu tez çalışmasında sıkıştırmalı algılamanın matematiksel temelleri üzerine çalıştık. Sıkıştırmalı algılama bize seyreklik koşuluyla beraber Nyquist-Shannon Teoreminden daha iyi bir sonuç verir. Bir sinyal seyrekse Nyquist-Shannon Teoreminde gerektiğinden daha az bir ölçümle bu orijinal sinyali kurtarabiliriz.

İlk olarak, sıkıştırmalı algılama için gerek koşulları inceledik. Seyreklik, sinyal kurtarmada sıkıştırmalı algılamayı kullanmanın anahtarıdır. Bunun yanında, seyreklik ile sıkıştırmalı algılama matrisleri arasındaki ilişkiyi inceledik.

Ardından, kurtarma algoritmaları, sıkıştırmalı algılama matris tasarımları ve sıkıştırmalı algılama matris özelliklerini inceledik. Sıkıştırmalı algılama matrislerinin en önemli iki özelliği Kısıtlı İzometri Özelliği ve Hiçlik Uzayı Özelliğidir. Bu iki özellik arasındaki ilişkiyi de inceledik.

Daha sonra, farklı sıkıştırmalı algılama üretme yöntemleri kullanarak deneyler yaptık. Son olarak, yeni bir sıkıştırmalı algılama matrisi üretme yöntemi önerdik ve sonuçlarını diğer sıkıştırmalı algılama matris üretme yöntemleriyle farklı kurtarma algoritmaları aracılığıyla deneyerek kıyasladık.

Anahtar Sözcükler : Sıkıştırmalı Algılama, Sinyal Kurtarma, Hiçlik Uzayı Özelliği, Kısıtlı İzometri Özelliği.

1 INTRODUCTION

Compressive sensing was introduced in two papers [7] by Donoho and [2] by Candès, Romberg and Tao in 2006. Compressive sensing is being used in many areas such as applied mathematics, computer science. There are a lot of applications in biology, medicine and radar technology.

An example for technological application of compressive sensing is single-pixel camera. The idea behind using compressive sensing techniques in single-pixel camera is correlate in hardware a real-world image. Another example from applications in biology is Magnetic Resonance Imaging. MRI is a common technology used for various tasks such as scan anomalies of the brain, follow-up tumors, breast cancer screening. In traditional approaches, we need time (several minutes or hours) to get a high-resolution image. For example, childen are too impatient to sit still more than two minutes. Thanks to compressive sensing, we can get high-resolution images using fewer samples than the traditional methods.

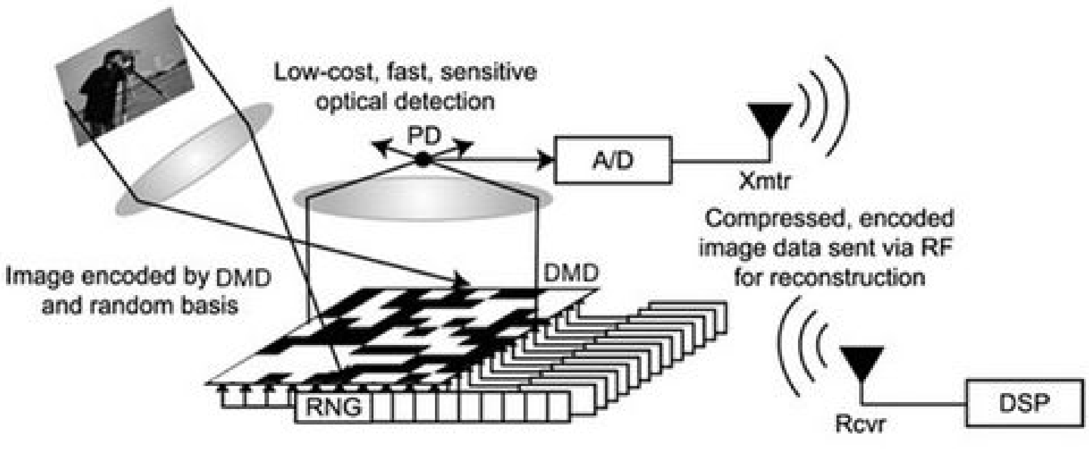

Figure 1.1: Schematic representation of a single-pixel camera (Image courtesy of Rice University)

Today, we are in the middle of a digital revolution and we are more concerned with data than in previous years. The data generated by us is rapidly rising every year. One of the problems is how to store that amount of data. One solution to this problem is sampling. In this approach, we don't store the whole data, bu we take some samples and then store them. One needs good algorithms to recover data from these samples.

Compressive sensing gives us a better way to reconstruct original signals with less samples than the conventional data acquisition methods. It helps us about the storage problem.

Compressive sensing is useful only with sparse signals. Compressive sensing gives us a better rate than Nyquist rate through the sparsity of the signal. This is a efficient method of signal processing using in acquisition and reconstruction of a sparse signal. Compressive sensing is important in the areas in which we can not measure data as much as we want. A sparse signal can be reconstructed from small set of linear measurements using Compressive Sensing.

Imagine we have a signal x of length N. It means that we have N measurements of x. For some reason, we can only obtain $m < N$ measurements. We represent m measurements with the vector y. We use the matrix A to recover the vector x using y. We call A the sensing matrix which helps us to recover x from y with $x = A \times y$.

We can find infinitely many solution matrices A satisfying $y = A \times x$. To choose the best solution, we use the algorithms. We can categorize the algorithms for choosing the best solution in 3 categories:

- Optimization Methods
- Greedy Methods
- Thresholding-Based Methods

1.1 Sampling

Sampling is the reduction of a continuous-time signal to a discrete-time signal. A sample can be a value or a set of values. The sampling theorem let us connect continuous signals and discrete signals.

1.2 Sparsity

A sparse signal is a signal which is represented as a linear combination of relatively few base elements in a basis. In mathematics, we take signals as vectors.
Firstly, we use $[N]$ for the set $\{1,2,3,4...N\}$. Let $x \in \mathbb{R}^N$ be a (high-dimensional) vector.

Definition 1.2.1. *(Support) The support of a vector $x \in \mathbb{R}^N$ is defined as:*

$$\text{supp}(x) = \{j \in [N] \ : \ x_j \neq 0\}$$

Definition 1.2.2. *(s-sparse) A vector x is called s-sparse if at most s of its element is nonzero, i.e.*

$$\|x\|_0 := \text{card}(\text{supp}(x)) \leq s$$

The set of all k-sparse vectors is denoted by

$$\Sigma_k = \{x \ : \ \|x\|_0 \leq k\}.$$

Definition 1.2.3. *(p-compressible with constant) Let $1 \leq p < \infty$ and $r > 0$. A vector $x = (x_i)_{i=1}^n \in \mathbb{R}^N$ is called p-compressible with constant C and rate r, if*

$$\sigma_k(x)_p := \min_{\tilde{x} \in \sum_k} \|x - \tilde{x}\|_p \leq C \times k^{-r} \quad \textit{for any } k \in \{1,...,n\}$$

Definition 1.2.4. *(Sensing matrix) Suppose we want to recover a signal $x \in \mathbb{R}^n$. Let $a_1, a_2, ..., a_m \in \mathbb{R}^N$, then measurements are linear combinations of the entries of x:*

$$y_i = \langle a_i, x \rangle, \quad i = 1,2,...,m$$

where $\langle \cdot, \cdot \rangle$ is the inner product and m is the number of measurements. We will store a_i's in a matrix, denoted $A \in \mathbb{R}^{m \times N}$, as follows

$$A = \begin{bmatrix} a_1^t \\ a_2^t \\ a_3^t \\ \vdots \\ a_m^t \end{bmatrix}$$

which is called the sensing matrix and a_i^t is the conjugate of a_i.

We define $y = Ax$ as

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{bmatrix}$$

Problem: Let $A \in \mathbb{R}^{m \times N}$ a matrix and $y \in \mathbb{R}^m$, solve the minimization problem

$$\min_{\bar{x} \in \mathbb{R}^N} \|\bar{x}\|_0 \quad s.t. \quad A\bar{x} = y \tag{P_0}$$

We assume that $\bar{x} \in \mathbb{R}^N$ is sparse and in this case, there are infinite many solutions. The problem is how to choose $x = \bar{x}$.
We have two main questions:

1. Is there any conditions about the sensing matrix?

2. Which algorithm should we use?

2 LITERATURE REVIEW

Kotelnikov, Nyquist, Shannon and Whittaker worked on sampling continuous-time signals. Their stuides [12], [13], [15] and [17] are the theoretical foundation of today's digital revolution about sensing systems. The results of these studies shows us that signals, images or videos can be exactly recovered from a set of uniformly spaced samples taken at Nyquist rate which is the twice the highest frequency present in the original signal. This is called the Nyquist-Shannon Theorem. The Nyquist-Shannon Theorem provides a bridge between discrete-time signals and continuous-time signals. We can take the discrete-time signals as discrete functions and the continuous-time signals as continuous functions. This theorem gives us a sufficient condition for a sample rate to transform a discrete function to continuous function. The Nyquist-Shannon Theorem honours Harry Nyquist and Claude Shannon. The Nyquist-Shannon Theorem stated as follows: "If a continuous time signal contains no frequency components higher than W hz, then it can be completely determined by uniform samples taken at a rate f_s samples per second where $f_s \geq 2W$ or, in term of the sampling period $T \leq \frac{1}{2W}$".

Compressive sensing was introduced in two papers [7] by Donoho and [2] by Candès, Romberg and Tao in 2006. The main idea behind the Compressive Sensing is to get a better rate than the Nyquist-Shannon rate. The Nyquist-Shannon Theorem states that a certain minimum number of samples is required to recover an arbitrary signal but when the signal is sparse, we can reduce the number of required measurements. Sparsity is the working condition for Compressive Sensing.

The sparse recovery problem can be traced back to earlier papers from the 1990s such as [6] by Donoho and Starck. Two prominent papers [8] by Donoho and Elad and [9] by Donoho and Huo was published in early 2000s. The studies of Emmanuel Candès, Justing Romberg, Terrence Tao and David Donoho had a large impact on the progress of this field. They showed that a finite-dimensional signal having a sparse or compressible representation can be recovered from a small amount of measurements.

The Compressive Sensing approach has many recovery algorithms such as optimization methods, greedy methods and thresholding-based methods. Basis Pursuit is an example of optimization methods and mathching pursuit algorithms are categorized in greedy methods. In [16], J. A. Tropp proved that a greedy algorithm called Orthogonal Matching Pursuit (OMP) can recover a signal with k non-zero entries in dimension n from random lienar measurements of that signal. J. A. Tropp stated that OMP is an effective alternative to Basis Pursuit for signal recovery.

3 PRELIMINARIES

Definition 3.0.1. *A vector space V is a set over a field $\mathbb{K}$ equiped with addition and scalar multiplication which satisfies the following properties:*

1. *Commutativity of addition:* $u+v=v+u$.
2. *Associativity of addition:* $u+(v+w)=(u+v)+w$.
3. *Additive identity: There exists a zero element* $0\in V$ *such that* $u+0=0+u=u$.
4. *Additive inverse: For each* $u\in V$ *there is* u' *such that* $u+u'=0=u'+u$.
5. *Distribution of scalar multiplication with respect to field addition:*
   $(\alpha+\beta)u=(\alpha u)+(\beta u)$
6. *Distribution of scalar multiplication with respect to vector addition:*
   $\alpha(u+v)=\alpha u+\alpha v$.
7. *Associativity of scalar multiplication:* $(\alpha\beta)u=\alpha(\beta u)$
8. *Identity element of scalar multiplication: There is a scalar* $1\in\mathbb{K}$ *such that* $1u=u$.

where $\alpha,\beta\in\mathbb{K}$ *and* $u,v,w\in V$.

Definition 3.0.2. *Let U,V two vector spaces. A function $T:U\to V$ is called a linear transformation if it satisfies:*

1. $T(u+w)=T(u)+T(w)$.
2. $T(\alpha u)=\alpha T(u)$.

where $u,w\in U$ *and* $\alpha\in\mathbb{K}$.

Definition 3.0.3. *The Null Space of the linear transformation T is a set of vectors such that* $T(u)=0$.

Definition 3.0.4. *An inner product on a vector space V is a function which takes two vectors u and v of V to a number $\langle u,v\rangle\in\mathbb{K}$ satisfying the following properties:*

1. *positivity:* $\langle u,u\rangle \geq 0$ *for all* $u \in V$.

2. *definitiveness:* $\langle u,u\rangle = 0$ *if and only if* $u = 0$.

3. *linearity in the first argument:*

$$\langle u+v,w\rangle = \langle u,w\rangle + \langle v,w\rangle \quad \text{for all } u,v,w \in V$$

$$\langle \alpha u,v\rangle = \alpha\langle u,v\rangle \quad \text{for all } \alpha \in \mathbb{K} \text{ and all } u,v \in V$$

4. *conjugate symmetry:* $\langle u,v\rangle = \overline{\langle v,u\rangle}$ *for all* $u,v \in V$.

Definition 3.0.5. *For* $v \in V$*, we define the* <u>*norm*</u> *of* v *by*

$$\|v\| = \sqrt{\langle v,v\rangle}$$

Cauchy-Schwarz Inequality: If $u,v \in V$ then

$$|\langle u,v\rangle| \leq \|u\|\|v\|.$$

Definition 3.0.6. *For a vector* $x \in \mathbb{K}^N$*, the usual* p*-norm is denoted as*

$$\|x\|_p = \left(\sum_{l=1}^{N} |x_l|^p\right)^{1/p}, \quad 1 \leq p < \infty$$

$$\|x\|_\infty = \max_{l\in[N]} |x_l|.$$

Let $A = (a_{jk}) \in \mathbb{K}^{m\times n}$ be an $m \times n$ matrix, we denote the conjugate transpose of A as $A^* = (\overline{a_{kj}})$.
The operator norm of a matrix from ℓ_p into ℓ_p is

$$\|A\|_{p\to p} := \max_{\|x\|_p=1} \|Ax\|_p.$$

We can write an explicit expression of the operator norm of A for $p = 1,2,\infty$:

$$\|A\|_{1\to 1} = \max_{k\in[N]} \sum_{j=1}^{m} |a_{jk}|,$$

$$\|A\|_{\infty\to\infty} = \max_{k\in[m]} \sum_{k=1}^{N} |a_{jk}|,$$

$$\|A\|_{2\to 2} = \sigma_{\max}(A) = \sqrt{\lambda_{\max}(A^*A)},$$

where $\lambda_{\max}(A^*A) > 0$ is the largest eigenvalue of A^*A. Note that $\|A\|_{1\to1} = \|A^*\|_{\infty\to\infty}$.

Definition 3.0.7. *Let A be a matrix. A is* <u>*hermitian*</u> *(or self-adjoint) if its conjugate transpose is equal to itself:*

$$a_{ij} = \overline{a_{ji}},$$

where a_{ij} is the element in the i-th row and j-th column of A.

For a hermitian matrix $A = A^*$,

$$\|A\|_{2\to2} = \sup_{\|x\|_2=1} |\langle Ax, x\rangle|.$$

4 SPARSE RECOVERY

We will study the conditions of sensing matrix and the sparsity of the original vector x for exact recovery. In this chapter, we will focus on ℓ_1-minimization which is one of the most popular methods. We follow mainly the book [10] in this chapter.

Definition 4.0.1. *Let A be an $m \times N$ matrix. The* spark *of A is the smallest number k such that there exists a set of k linearly dependent columns of A. We note it as* spark*(A). Formally, we can write it as,*

$$\text{spark}(A) = \min_{x \neq 0} \|x\|_0 \quad s.t. \quad Ax = 0,$$

where x is a nonzero vector and $\|x\|_0$ denotes its number of nonzero coefficients.

The notion of spark was first introduced by Donoho and Elad in [8]. We say that the spark of a matrix A is 1 if there is a zero column. Also if the rank of A is full, the spark of A is $+\infty$.

Example 4.0.1. *Let*

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

be an 5×5 matrix. This matrix has no zero column but it has a set of 2 columns which are linearly dependent. So the spark of A is 2.

The word spark comes from the verbal fusion of the word “sparse” and “rank”.
We denote the null space of A as $\mathcal{N}(A)$.

Lemma 4.0.1. *Let A be an $m \times N$ matrix. Then*

$$\text{spark}(A) = \min\{k : \mathcal{N}(A) \cap \Sigma_k \neq \{0\}\},$$

and spark*(A) $\in [2, m+1]$.*

Lemma 4.0.2. *For any non-zero $2k$-sparse vector h, there are two different k-sparse vectors x and $\tilde{x}$ such that h can be written as the difference of these two vectors. In other words, for any non-zero vector h, $h \in \Sigma_{2k}$ if and only if $h = x - \tilde{x}$ for some $x, \tilde{x} \in \Sigma_k$.*

Proof. $(\Rightarrow)$
Let $h = (h_1, h_2, h_3, \ldots)$ be a vector in Σ_{2k} where every $h_i \in \mathbb{R}$ is an entry of h. Let S be the support of the vector h. As h is a non-zero vector there is at least one entry which is not zero and $S \neq \emptyset$. As h is $2k$-sparse, it has at most $2k$ entries which are not zero. So $1 \leq |S| \leq 2k$.
Let S_1, S_2 be subsets of S such that:

$$S_1 \cup S_2 = S,\ S_1 \cap S_2 = \emptyset \quad \text{and} \quad |S_1| \leq k,\ |S_2| \leq k.$$

We also assume $S_1 \neq \emptyset$, $S_2 \neq \emptyset$.
Let $x = (x_1, x_2, x_3, \ldots, x_n)$ and $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \ldots, \tilde{x}_n)$ be vectors in Σ_k where:

$$x_i = \begin{cases} h_i & i \in S_1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and } \tilde{x}_i = \begin{cases} -h_i & i \in S_2 \\ 0 & \text{otherwise} \end{cases}$$

So, by substraction $h = x - \tilde{x}$ where $x, \tilde{x} \in \Sigma_k$ and $h \in \Sigma_{2k}$.
$(\Leftarrow)$
Assume $x \neq \tilde{x}$ and $x, \tilde{x} \in \Sigma_k$. Let $h = x - \tilde{x}$ be a vector, so its entries are $h_i = x_i - \tilde{x}_i$. We want to show that $h \in \Sigma_{2k}$. Let the sets K, K_1 and K_2 be the supports of h, x and $\tilde{x}$ respectively. As $x, \tilde{x} \in \Sigma_k$, the cardinality of K_1 and K_2 is less than or equal to k. We need to show that $K \subseteq K_1 \cup K_2$.
Suppose $i \in K$. As $h_i = x_i - \tilde{x}_i$, $x_i \neq 0$ and $\tilde{x}_i \neq 0$, then $h_i \neq 0$. So we can say that $i \in K_1$ or $i \in K_2$ therefore $i \in K_1 \cup K_2$. Hence $K \subseteq K_1 \cup K_2$. Since $K_1 \cup K_2$ has at most $2k$ non-zero entries, $h \in \Sigma_{2k}$.
So $|K| \leq |K_1| + |K_2| \leq k + k = 2k$ which means $h \in \Sigma_{2k}$.

Theorem 4.0.3. *Let A be an $m \times N$ matrix and let $k \in \mathbb{N}$. Then the following conditions are equivalent:*

(i) If a solution x of (P_0) satisfies $\|x\|_0 \leq k$, then this is the unique solution.

(ii) $k < \text{spark}(A)/2$

Proof. $(i) \Rightarrow (ii)$: We will prove it by contradiction.
If (ii) does not hold, we can say that there exists $h \in \mathcal{N}(A)$, $h \neq 0$ such that $\|h\|_0 \leq 2k$. So, there exists x and $\tilde{x}$ satisfying $h = x - \tilde{x}$ and $\|x\|_0, \|\tilde{x}\|_0 \leq k$ but $Ax = A\tilde{x}$. It is a contradiction to (i).

$(ii) \Rightarrow (i)$: Let us assume x and $\tilde{x}$ two vectors which are satisfying $y = Ax = A\tilde{x}$ and $\|x\|_0, \|\tilde{x}\|_0 \leq k$. Thus $x - \tilde{x} \in \mathcal{N}(A)$ and $\|x - \tilde{x}\|_0 \leq 2k <$ spark(A). By the previous lemma, we can say that $x - \tilde{x} = 0$ and it implies (i).

Theorem 4.0.4. *Let A be an $m \times N$ matrix. The following properties are equivalent:*

1. *Every s-sparse vector $x \in \mathbb{R}^N$ is the unique s-sparse solution of $Az = Ax$, that is, if $Ax = Az$ and both x and z are s-sparse, then $x = z$.*
2. *The null space $\mathcal{N}(A)$ does not contain any 2s-sparse vector other than the zero vector, that is, $\mathcal{N}(A) \cap \{z \in \mathbb{R}^N : \|z\|_0 \leq 2s\} = \{0\}$*
3. *For every $S \subset [N]$ with $card(S) \leq 2s$, the submatrix A_S is injective as a map from $\mathbb{R}^S$ to $\mathbb{R}^m$.*
4. *Every set of $2s$ columns of A is linearly independent.*

Recall that the Vandermonde matrix is a matrix with the terms of a geometric progression in each row, an $m \times N$ matrix:

$$V = \begin{bmatrix} 1 & \alpha_1 & \alpha_1^2 & \alpha_1^3 & \dots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \alpha_2^2 & \alpha_2^3 & \dots & \alpha_2^{n-1} \\ 1 & \alpha_3 & \alpha_3^2 & \alpha_3^3 & \dots & \alpha_3^{n-1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_m & \alpha_m^2 & \alpha_m^3 & \dots & \alpha_m^{n-1} \end{bmatrix}$$

The following proposition is well-known.

Proposition 4.0.1. *The determinant of a square Vandermonde matrix (i.e. $m = n$) can be written as*

$$\det(V) = \prod_{1 \leq i < j \leq n} (\alpha_j - \alpha_i).$$

Theorem 4.0.5. *For any integer $N \geq 2s$, there exists an $m \times N$ sensing matrix A with $m = 2s$ rows such that every s-sparse vector $x \in \mathbb{R}^N$ can be recovered from its measurement vector $y = Ax \in \mathbb{R}^m$ as a solution of (P_0).*

Proof. Let $t_N > \cdots > t_2 > t_1 > 0$ be real numbers and $A \in \mathbb{C}^{m \times N}$ be a matrix with $m = 2s$

$$A = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ t_1 & t_2 & t_3 & \dots & t_N \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_1^{2s-1} & t_2^{2s-1} & t_3^{2s-1} & \dots & t_N^{2s-1} \end{bmatrix}$$

Let $S = \{j_1 < \cdots < j_{2s}\}$ be an index set. The square matrix A_S is the transpose of a Vandormonde matrix. We can calculate determinant of the matrix A_S.

$$\det(A_S) = \begin{vmatrix} 1 & 1 & 1 & \dots & 1 \\ t_1 & t_2 & t_3 & \dots & t_N \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_1^{2s-1} & t_2^{2s-1} & t_3^{2s-1} & \dots & t_N^{2s-1} \end{vmatrix} = \prod_{k<l} (t_{j_l} - t_{j_k}) > 0.$$

We can say that A_S is invertible. So every s-sparse vector $x \in \mathbb{C}^N$ is the unique s-sparse solution vector that satisfies $Az = Ax$. Hence, we can recover it as the unique solution (P_0).

5 SOME SENSING MATRIX GENERATION METHODS

In this chapter, we will examine the following sensing matrix generation methods: Gaussian, Bernoulli, Hadamard, Toeplitz and Circulant matrices. Gaussian, Bernoulli and Hadamard matrices are random matrices, Toeplitz and Circulant matrices are deterministic matrices [1]. We will make some experiments with these types of matrices using the Orthogonal Matching Pursuit and Iterative Hard Thresholding recovery algorithm.

Bernoulli, Gaussian and Hadamard matrices generated as $N \times N$ matrices with the given distribution. But in Compressive Sensing, we need that the sensing matrix has a lot more columns than the rows. We choose m random rows from the generated matrices.

Definition 5.0.1. *A random Gaussian matrix is a matrix, each of whose entries is a random variable with normal distribution.*

Definition 5.0.2. *A random Bernoulli Matrix is a matrix, whose entries are* ± 1 *with equal probability.*

Definition 5.0.3. *A Hadamard matrix is a square matrix whose entries are* ± 1 *and whose rows are mutually orthogonal.*

Hadamard matrices were invented by James Joseph Sylvester in 1867. Let H be a Hadamard matrix of order n, then the following matrix is a Hadamard matrix of order $2n$

$$\begin{bmatrix} H & H \\ H & -H \end{bmatrix}.$$

Definition 5.0.4. *A Toeplitz matrix is in the following form*

$$\begin{bmatrix} a_0 & a_{-1} & a_{-2} & \dots & \dots & a_{-(n-1)} \\ a_1 & a_0 & a_{-1} & \ddots & \ddots & \vdots \\ a_2 & a_1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & a_{-1} & a_{-2} \\ \vdots & \ddots & \ddots & \ddots & a_0 & a_{-1} \\ a_{n-1} & \dots & \dots & a_2 & a_1 & a_0 \end{bmatrix}$$

where $A_{i,j} = a_{j-i}$.

Definition 5.0.5. *A Circulant matrix is in the following form*

$$\begin{bmatrix} c_0 & c_{n-1} & c_{n-2} & \dots & c_2 & c_1 \\ c_1 & c_0 & c_{n-1} & \ddots & \ddots & c_2 \\ c_2 & c_1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & c_{n-2} \\ c_{n-2} & \ddots & \ddots & \ddots & c_0 & c_{n-1} \\ c_{n-1} & c_{n-2} & \dots & \dots & c_1 & c_0 \end{bmatrix}$$

which is a special kind of Toeplitx matrix. In each column, instead of a random number, we use the last element of previous column.

6 RECOVERY ALGORITHMS

In this chapter, we will study some recovery algorithms to find the original vector x. These algorithms are divided in three categories:

- Optimization Methods
- Greedy Methods
- Thresholding-Based Methods

6.1 Optimization Methods

An *optimization problem* is a problem of type

$$\underset{x \in \mathbb{R}^N}{\text{minimize}}\ F_0(x) \quad \text{subject to } F_i(x) \le b_i,\ i \in [n],$$

where the function $F_0 : \mathbb{R}^N \to \mathbb{R}$ are called constraint functions. If $F_0, F_1, ..., F_n$ are all convex functions, then the problem is called a *convex optimization problem*.
Our sparse recovery problem is an optimization problem. We can write it as:

$$\text{minimize } \|z\|_0 \quad \text{subject to } Az = y \tag{P_0}$$

The main idea of Chen, Donoho and Saunders in the fundamental paper [3] was to substitute the ℓ_0 norm by the closest convex norm, which is the ℓ_1 norm. This method is called convex-relaxation:

$$\text{minimize } \|z\|_1 \quad \text{subject to } Az = y \tag{P_1}$$

This idea says that the solution of (P_1) coincides with the solution of (P_1). The figure gives us a geometric intuition why the solution of (P_1) coincides with the solution of (P_1).
In Figure 6.1, we can see unit spheres for the ℓ_p norms with $p = 1, 2, \infty$ and for the ℓ_p quasinorm with $p = \frac{1}{2}$ in $\mathbb{R}^2$ [5].

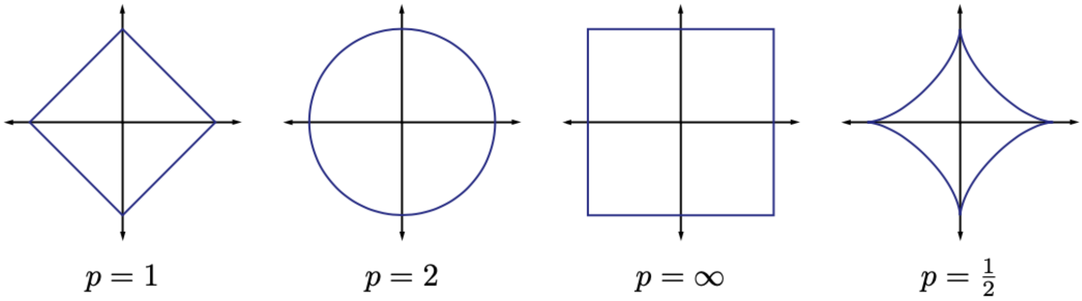

Figure 6.1: Unit spheres in $\mathbb{R}^2$ for the ℓ_p norms with $p = 1, 2, \infty$ and for the ℓ_p quasinorm with $p = \frac{1}{2}$.

Solving an underdetermined system is impossible if we don't have another additional information. For Compressive Sensing, sparsity is the additional information to solve the underdetermined system.

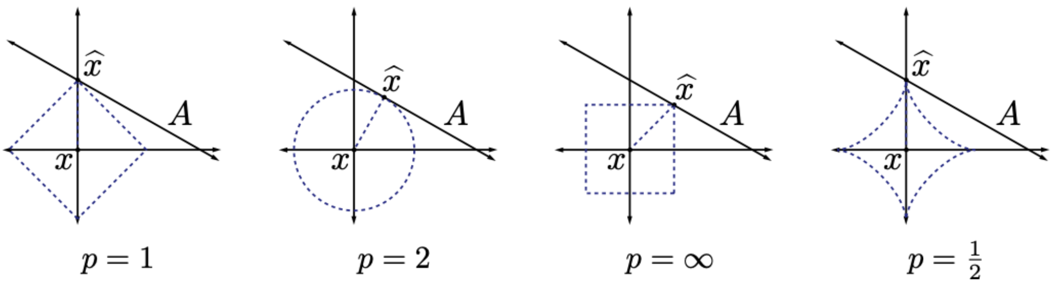

Figure 6.2: Best approximation of a point in $\mathbb{R}^2$ by a one-dimensional subspace using the ℓ_p norms for $p = 1, 2, \infty$ and the ℓ_p quasinorm with $p = 2$.

In Figure 6.2, we can examine the optimization problem from a geometric view [5]. From this view, an optimization problem is blowing up the ℓ_p ball until they touch the hyperplane $Ax = y$. The intersection point is the optimal solution of sparse recovery problem. We can see that for $p = 2$ and $p = \infty$ the intersection points are not on the coordinate axis so the intersection point is not sparse. So, for $p \leq 1$, ℓ_p norm ensure the sparsity of the solution. Also, we know that for $p < 1$, the ℓ_p balls are not convex, we use the ℓ_1 norm for sparse recovery problem.

Basis Pursuit

Input: sensing matrix A, measurement vector y
Instruction:

$$x^\sharp = \text{argmin}\ \|z\|_1 \quad \text{subject to } Az = y$$

Output: the vector $x^\sharp$

argmin function gives us the vector z whose ℓ_1-norm is the minimum.

Theorem 6.1.1. *Let $A \in \mathbb{R}^{m\times N}$ be a sensing matrix with columns $a_1,...,a_N$. Assuming the uniqueness of a minimizer $x^\sharp$ of*

$$\underset{z\in\mathbb{R}^N}{\text{minimize}} \|z\|_1 \quad \text{subject to } Az = y$$

the system $\{a_j, j \in supp(x^\sharp)\}$ is linearly independent, and in particular

$$\|x\|_0 = \text{card}(\text{supp}(x^\sharp)) \leq m.$$

Proof. We will prove this theorem by contradiction. Assume that the system $\{a_j, j \in S\}$ is linearly dependent where $S = \text{supp}(x^\sharp)$. So, there is a nonzero vector $v \in \mathbb{R}^N$ supported on S which satisfies $Av = 0$. We can say that for any $t \neq 0$,

$$\|x^\sharp\|_1 < \|x^\sharp + tv\|_1 = \sum_{j\in S} |x^\sharp + tv_j|.$$

We can write it as below:

$$\sum_{j\in S} |x^\sharp + tv_j| = \sum_{j\in S} \text{sgn}(x_j^\sharp + tv_j)(x_j^\sharp + tv_j).$$

If we take $|t|$ small enough, we have

$$\text{sgn}(x^\sharp + tv_j) = \text{sgn}(x_j^\sharp) \quad \text{for all } j \in S.$$

Hence

$$\|x^\sharp\|_1 < \sum_{j\in S} \mathrm{sgn}(x_j^\sharp)(x_j^\sharp + tv_j) = \sum_{j\in S} \mathrm{sgn}(x_j^\sharp)(x_j^\sharp) + t\sum_{j\in S} \mathrm{sgn}(x_j^\sharp)(v_j)$$
$$= \|x^\sharp\|_1 + t\sum_{j\in S} \mathrm{sgn}(x_j^\sharp)v_j.$$

We can always choose a small $t \neq 0$ such that $t\sum_{j\in S} \mathrm{sgn}(x_j^\sharp)v_j \leq 0$, so it is a contradiction.

Let $v \in \mathbb{C}^N$, $S \subset [N]$ and v_S be the restriction of v on S.

6.1.1 Null Space Property

Null Space Property is an important specification of sensing matrices. Let A be an $m\times N$ matrix. Recall that the null space $\mathcal{N}(A)$ of the matrix A is defined as:

$$\mathcal{N}(A) = \{x \in \mathbb{R}^N \ : \ Ax = 0\}.$$

To recover all k-sparse signals x from $y = Ax$, the matrix A must uniquely represent all $x \in \Sigma_k$. It means:

$$\forall x, \tilde{x} \in \Sigma_k, \quad x \neq \tilde{x} \implies Ax \neq A\tilde{x}.$$

Otherwise we can not recover the original vector from y.

Example 6.1.1. *Let A be our sensing matrix defined as:*

$$A = \begin{bmatrix} 1 & 3 \end{bmatrix}$$

The Null Space of this matrix is $\mathcal{N}(A) = \{(x,y) \ : \ x+3y = 0\}$. Let $x = [0;\ 1]$ be the original vector.

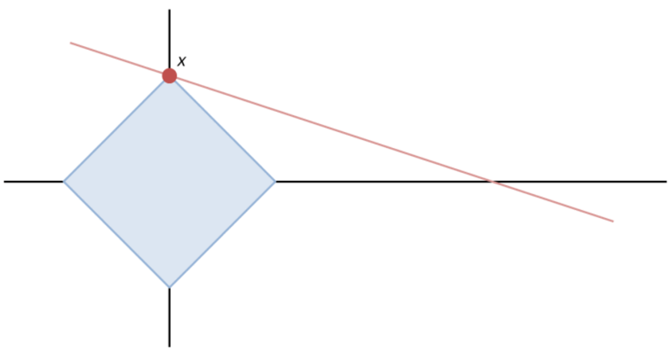

Figure 6.3: Example for ℓ_1 norm approximation

The red dot x in the Figure 6.3 is the original vector x. The blue square is the unit sphere for ℓ_1 norm.

Definition 6.1.1. *(Null Space Property) Let A be an $m \times N$ matrix. We say that A satisfies the Null Space Property relative to a set $S \subset [N]$ if $\|v_S\|_1 < \|v_{\bar{S}}\|_1$ for all $v \in \mathcal{N}(A) \backslash \{0\}$.*

Theorem 6.1.2. *Let A be an $m \times N$ matrix. Every vector $x \in \mathbb{R}^N$ supported on a set S is the unique solution of (P_1) with $y = Ax$ if and only if A satisfies the Null Space Property relative to S.*

Proof. Given an indexed set S, Assume that every vector $x \in \mathbb{R}^N$ supported on S is the unique minimizer of $\|z\|_1$ subject to $Az = Ax$. Thus for any vector $v \in \mathcal{N}A \backslash \{0\}$, the vector v_S is the unique minimizer of $\|z\|_1$ subject to $Az = Av_S$. We have $A(-v_{S^c}) = Av_S$ and $-v_{S^c} \neq v_S$ because $A(v_{S^c} + v_S) = Av = 0$ and $v \neq 0$. We can conclude that $\|v_S|_1 < \|v_{S^c}\|_1$. This establishes the Null Space Property relative to S.
Conversely, let us assume that the Null Space Property relative to S holds.
Given a vector $x \in \mathbb{R}^N$ supported on S and a vector $z \in \mathbb{R}^N$, $z \neq x$ satisfying $Az = Ax$, we consider the vector $v = x - z \in \mathcal{N}(A) \backslash \{0\}$.

$$\begin{aligned} \|x\|_1 &\leq \|x - z_S\|_1 + \|z_S\|_1 \\ &= \|v_S\|_1 + \|z_S\|_1 \\ &< \|v_{S^c}\|_1 + \|z_S\|_1 \\ &= \| - z_{S^c}\|_1 + \|z_S\|_1 \\ &= \|z\|_1. \end{aligned}$$

So, we can say that the vector x is the unique solution of (P_1) with $y = Ax$.

This theorem first appeared explicitly in [11]. We show the term NSP in [4] which was suggested by A. Cohen, W. Dahmen and R. DeVore. The Null Space Property is usually difficult to show directly. Instead of showing the Null Space Property we will use the Restricted Isometry Property which will be introduced in the next chapter. Restricted isometry property is much popular than the Null Space Property.
There is another definition of the Null Space Property using ℓ_2-norm. We will use it later to establish a connection between Null Space Property and Restricted Isometry Property.

Definition 6.1.2. *Let A be an $m \times N$ matrix. The matrix A satisfies the Null Space Property of k if there exists a contant $C \in \mathbb{R}$ such that for all $v \in \mathcal{N}(A)$,*

$$\|v_S\|_2 \leq C \frac{\|v_{S^c}\|_1}{\sqrt{k}} \quad \textit{for all } S \subset [N] \textit{ with } |S| \leq k.$$

Let $\Delta : \mathbb{R}^m \to \mathbb{R}^N$ be a linear map which will be our recovery method. The following property of this method guarantees exact recovery of all sparse signals in Σ_k:

$$\begin{gathered} \|\Delta(Ax) - x\|_2 \leq C\frac{\sigma_k(x)_1}{\sqrt{k}}, \quad \text{for all } x \in \mathbb{R}^n, \\ \text{where } \sigma_k(x)_p = \inf_{\tilde{x}\in\Sigma_k} \|x - \tilde{x}\|_p \quad \text{for } p = 1. \end{gathered} \tag{6.1}$$

If $x \in \Sigma_k$ then we can say that $\sigma_k(x)_1 = 0$. It implies that $\|\Delta(Ax) - x\|_2$ is zero and $\Delta(Ax) = x$.

Theorem 6.1.3. *Let A be an $m \times N$ sensing matrix and let $\Delta : \mathbb{R}^m \to \mathbb{R}^n$ be a recovery algorithm. If the sensing matrix A and the recovery algorithm Δ satisfies the property 6.1 then A satisfies the Null Space Property of order $2k$.*

Proof. Let $v \in \mathcal{N}(A)$ be a vector and we define S for the indices of $2k$ largest entries of the vector v. We can write $v = v_S + v_{S^c}$.
We separate the set S to S_0 and S_1 such that both of them has k elements. As $|S| = 2k$, we can write $v_S = v_{S_0} + v_{S_1}$. We define the vector $x = v_{S_1} + v_{S^c}$ and we will write the vector v as the difference of two vectors: $v = x - \tilde{x}$ where $\tilde{x} = -v_{S_0}$.
So, by construction of $\tilde{x}$, $\tilde{x}$ is a k-sparse vector as $|S_0| = k$.
We apply Equation (6.1) which guarantees the exact recovery of all vectors in Σ_k. We get $\tilde{x} = \Delta(A\tilde{x})$ and as $v \in \mathcal{N}(A)$:

$$Av = A(x - \tilde{x}) = 0.$$

So $A(x) = A(\tilde{x})$ and we have $\tilde{x} = \Delta(A(x))$.
We can say that $\|v_s\|_2 \leq \|v\|_2 = \|x - \tilde{x}\|_2 = \|x - \Delta(Ax)\|_2$.
We apply Equation (6.1) and we get:

$$\|x - \Delta(Ax)\|_2 \leq \frac{C\sigma_k(x)_1}{\sqrt{k}}.$$

And we extend by $\sqrt{2}$ the right side of the equation:

$$\|x - \Delta(Ax)\|_2 \leq \frac{C\sqrt{2}\sigma_k(x)_1}{\sqrt{2k}}.$$

We have $\sigma_k(x)_1 = \inf_{\tilde{x}\in\Sigma_k} \|x - \tilde{x}\|_1$. As $x = v_{S_1} + v_{S^c}$ and $v_{S_1} \in \Sigma_k$, we also have $\sigma_k(x)_1 \leq \|x - v_{S-1}\|_1 = \|v_{S^c}\|_1$. We finally get:

$$\|v_S\|_2 \leq \frac{C\sqrt{2}\|v_{S^c}\|_1}{\sqrt{2k}}.$$

So, the sensing matrix A satisfies the Null Space Property of order $2k$.

6.1.2 Restricted Isometry Property

Definition 6.1.3. *Let $A \in \mathbb{C}^{m\times N}$, the restricted isometry constant δ_s is defined as the smallest δ_S such that*

$$(1-\delta_s)\|x\|_2^2 \le \|Ax\|_2^2 \le (1+\delta_s)\|x\|_2^2 \tag{6.2}$$

for all s-sparse $x \in \mathbb{C}^N$.

A matrix A satisfies the RIP if δ_s is between 0 and 1.
In the RIP, we have bounds around 1 but in practice, we can always use the scalars α and β with $0 < \alpha < \beta < \infty$:

$$\alpha\|x\|_2^2 \le \|Ax\|_2^2 \le \beta\|x\|_2^2.$$

To examine the scalars α and β, we write $A = \frac{1}{\theta}(\theta A)$ where $\theta A = \tilde{A}$ and θ is a scalar. We can write the equation as follows:

$$\alpha\theta^2\|x\|_2^2 \le \|\tilde{A}\|_2^2 \le \beta\theta^2\|x\|_2^2.$$

We have $\alpha\theta^2 = 1-\sigma_k$ and $\beta\theta^2 = 1+\sigma_k$. It gives us $(\alpha+\beta)\theta^2 = 2$ and we obtain $\theta = \sqrt{\frac{2}{\alpha+\beta}}$.
Now, we will get δ_k using θ.

$$\alpha\theta^2 = 1-\delta_k \implies \alpha\frac{2}{\alpha+\beta} = 1-\delta_k \implies \frac{2\alpha}{\alpha+\beta} - 1 = -\delta_k$$
$$\implies \delta_k = \frac{\beta-\alpha}{\alpha+\beta}.$$

This gives us our desired bounds of δ_k: $0 < \delta_k < 1$.
It is important to notice that if a sensing matrix A satisfies the Restricted Isometry Property of order k with the Restricted Isometry Constant $0 < \delta_k < 1$ then the matrix A satisfies the Restricted Isometry Property of order k' with $k' < k$ and $\delta_{k'} < \delta_k$.
In the next section, we will see the relation between the Null Space Property and the Restricted Isometry Constant. Before this, we will examine some properties of Restriced Isometry Contants.

Proposition 6.1.1. *Let $A \in \mathbb{C}^N$ with isometry contants δ_S.*

1. *The restricted isometry contants are ordered, $\delta_1 \le \delta_2 \le \delta_3 \le \dots$*

2. *It holds*

$$\delta_s = \max_{S\subset[N],|S|\le s} \|A_S^* A_S - \mathrm{Id}\|_{2\to 2}$$
$$= \sup_{x\in T_s} |\langle (A^*A - \mathrm{Id})x, x\rangle|,$$

where $T_s = \{x \in \mathbb{C}^N, \|x\|_2 = 1, \|x\|_0 \le s\}$.

3. *Let* $u, v \in \mathbb{C}^N$ *with adjoint supports,* $\operatorname{supp} u \cap \operatorname{supp} v = \emptyset$. *Let* $s = |\operatorname{supp} u| + |\operatorname{supp} v|$. *Then*

$$|\langle Au, Av\rangle| \le \delta_s \|u\|_2 \|v\|_2.$$

Proof. First property is clear since an s-sparse vector is also $s+1$-sparse.
For the second property, we can denote the following equation by subtracting $\|x\|_2$ from Equation 6.2:

$$|\|Ax\|_2^2 - \|x\|_2^2 \le \delta_s \|x\|_2^2 \quad \text{for all } S \subset [N], |S| \le s, \text{ for all } x \in \mathbb{C}^N, \operatorname{supp} x \subset S$$

We can write the term on the left hand side as $|\langle (A^*A - \mathrm{Id})x, x\rangle|$. We take the supremum over all $x \in \mathbb{C}^N$ with $\operatorname{supp} x \subset S$ and unit norm $\|x\|_2 = 1$. It gives us the operator norm $\|A_S^* A_S - \mathrm{Id}\|_{2\to 2}$. We take the maxiumum of the expression over all subsets S with $|S| \le s$ and it completes the proof.
For the third property, we denote $S = \operatorname{supp} u, T = \operatorname{supp} v$ and let $\tilde{u}, \tilde{v}$ denote the vector u, v restricted to their supports. We can write

$$\langle Au, Av\rangle = \tilde{u}^* A_S^* A_T \tilde{v}.$$

We can write this equality as

$$\tilde{u}^* A_S^* A_T \tilde{v} = (\tilde{u}^*, 0_T^*) A_{S\cup T}^* A_{S\cup T} (0_S^*, \tilde{v}^*)^*.$$

where 0_S is the zero vector restricted on S. Since the supports of u and v are disjoint, $(\tilde{u}^*, 0_T^*)\,\mathrm{Id}(0_S^*, \tilde{v}^*)^* = 0$ and we can write

$$\tilde{u}^* A_S^* A_T \tilde{v} = (\tilde{u}^*, 0_T^*)(A_{S\cup T}^* A_{S\cup T} - \mathrm{Id})(0_S^*, \tilde{v}^*)^*.$$

So, we can write

$$|\langle Au, Av\rangle| \le \|A_{S\cup T}^* A_{S\cup T} - \mathrm{Id}\|_{2\to 2} \|u\|_2 \|v\|_2.$$

We apply the second property of this proposition and complete the proof.

6.1.3 Experiments

We made some experiments to test the Restricted Isometry Property using Matlab. In the first experiment, we tried to measure the Restricted Isometry Property. We choose $N = 256$, $m = 16$, $k = 6$ where N is the length of original vector, m is the length of measurement vector and k is the sparsity level. First, we generate two original vectors and a $m \times N$ matrix whose elements are normally distributed random numbers. Then, we calculate the measurement vectors using this matrix. Lastly, we calculate the ratio of the difference between original vectors to the difference between measurement vectors using ℓ_1 norm. We repeated these calculations 100.000 times and the following graphic is the the number of ratios. 91.98% of the calculations are between 0.1 and 0.3. The lowest one is 0.0976 and the biggest one is 0.6811.

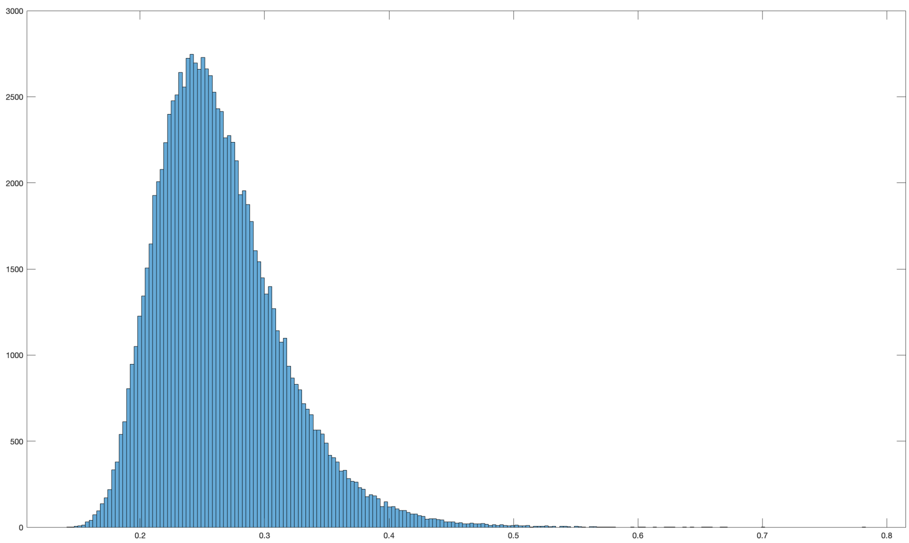

Figure 6.4: RIP test using the Gaussian Matrix

I repeated the experiment using Hadamard matrix. It gives 84.25% of the calculations are between 0.1 and 0.3. The lowest one is 0.1546 and the biggest one is 0.8137.

6.1.4 Relation between Null Space Property and Restricted Isometry Property

We will see that the Restricted Isometry Property is a stronger condition than the Null Space Property. We will prove this condition for both definitions of the Null Space Property. Firstly, we will use the first definition of the Null Space Property. Secondly, we will give the main theorem that used the second definition of the Null Space Property

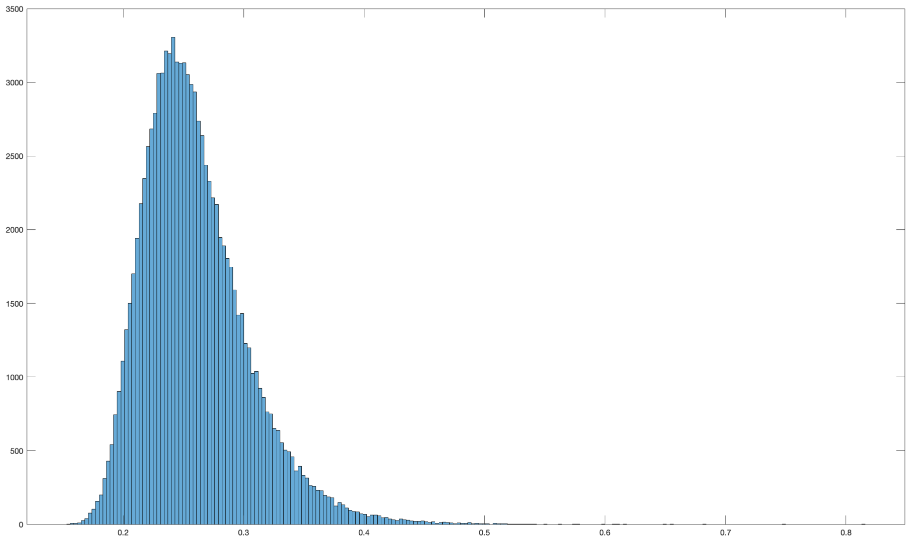

Figure 6.5: RIP test using the Hadamard Matrix

and then we will examine and prove the required lemmas. Lastly, we prove the main theorem.

Theorem 6.1.4. *Suppose the restricted isometry constant* δ_{2s} *of a matrix* $A \in \mathbb{C}^{m \times N}$ *satisfies*

$$\delta_{2s} < \frac{1}{3}$$

then the Null Space Property of order s is satisfied. In particular, every s-sparse vector $x \in \mathbb{C}^N$ *is recovered by* ℓ_1*-minimization.*

Proof. Let $v \in \mathcal{N}(A)$ be given. We define the set S_0 as the index set of s largest absolute entries of the vector v. We partition the complement of S_0 as $S_0^c = S_1 \cup S_2 \cup \ldots$ where S_1 is the index set of s largest absolute entries of $[N] \backslash S_0$ and S_2 is the index set of s largest absolute entries of $[N] \backslash (S_0 \cup S_1)$ etc.
As $v \in \mathcal{N}(A)$, we can write $A(v_{S_0}) = -A(v_{S_1} + v_{S_2} + \ldots)$.

$$\begin{aligned}
\|v_{S_0}\|_2^2 &\leq \frac{1}{1-\delta_{2s}} \|A(v_{S_0})\|_2^2 \\
&= \frac{1}{1-\delta_{2s}} \langle A(v_{S_0}), A(-v_{S_1}) + A(-v_{S_2}) + \ldots \rangle \\
&= \frac{1}{1-\delta_{2s}} \sum_{k \geq 1} \langle A(v_{S_0}, A(-v_{S_k}) \rangle.
\end{aligned}$$

We can apply the third property of the previous proposition:

$$\langle A(v_{S_0}), A(-v_{S_k}) \leq \delta_{2s}\|v_{S_0}\|_2\|v_{S_k}\|_2.$$

So;

$$\|v_{S_0}\|_2^2 \leq \frac{\delta_{2s}}{1-\delta_{2s}} \sum_{k\geq 1} \|v_{S_k}\|_2\|v_{S_0}\|_2.$$

We divide the expression by $\|v_{S_0}\|_2$:

$$\|v_{S_0}\|_2 \leq \frac{\delta_{2s}}{1-\delta_{2s}} \sum_{k\geq 1} \|v_{S_k}\|_2.$$

As s entries of v_{S_k} don't exceed s entries of $v_{S_{k-1}}$ for $k \geq 1$, we have

$$|| < \frac{1}{2} \sum_{l\in S_{k-1}} |_l| \text{ for all } j \in S_k,$$

and therefore

$$\|v_{S_k}\|_2 = \Big(\sum_{j\in S_k} |_j|^2\Big)^{\frac{1}{2}} \leq \frac{1}{\sqrt{s}}\|v_{S_{k-1}}\|_1.$$

We obtain by the Cauchy-Schwarz inequality

$$\begin{aligned}\|v_{S_0}\|_1 &\leq \sqrt{s}\|v_{S_0}\|_2 \leq \frac{\delta_{2s}}{1-\delta_{2s}}\|v_{S_{k-1}}\|_1 \\ &\leq \frac{\delta_{2s}}{1-\delta_{2s}}(\|v_{S_0}\|_1 + \|v_{S_0^c}\|_1)\end{aligned}$$

Since $\delta_{2s} < \frac{1}{3}$, one has $1-\delta_{2s} < \frac{2}{3}$ and $\frac{\delta_{2s}}{1-\delta_{2s}} < \frac{1}{2}$.
We can write $\|v_{S_0}\|_1 < \frac{1}{2}\|v_{S_0}\|_1 + \frac{1}{2}\|v^c_{S_0}\|_1$, so $\|v_{S_0}\|_1 < \|v_{S_0^c}\|_1$.

Theorem 6.1.5. *Let A be an $m\times N$ sensing matrix. Suppose that the matrix A satisfies the Restricted Isometry Property of order $2k$ with $\delta_{2k} < \sqrt{2}-1 \leq \frac{1}{2}$. Then the matrix A satisfies the Null Space Property of order $2k$ with constant:*

$$C = \frac{\sqrt{2}\delta_{2k}}{1-(1+\sqrt{2})\delta_{2k}}.$$

Lemma 6.1.6. *Let v be a k-sparse vector. Then:*

$$\frac{\|v\|_1}{\sqrt{k}} \leq \|v\|_2 \leq \sqrt{k}\|v\|_\infty.$$

Proof. For any vector v, we can denote $\|v\|_1 = |\langle v, \mathrm{sgn}(v)\rangle|$. We apply Cauchy-Schwarz inequality and we obtain:

$$|\langle v, \operatorname{sgn}(v)\rangle| \leq \|v\|_2 \|\operatorname{sgn}(v)\|_2.$$

So, we have also:

$$\|v\|_1 \leq \|v\|_2 \|\operatorname{sgn}(v)\|_2.$$

As $v \in \Sigma_k$, it has no more than k non-zero entries and $\operatorname{sgn}(v)$ has no more than k non-zero entries equal to ± 1. So, we have $\|\operatorname{sgn}(v)\|_2 \leq \sqrt{k}$. We obtain

$$\frac{\|v\|_1}{\sqrt{k}} \leq \|v\|_2,$$

which is our lower bound. By definition we know that $\|v\|_2 = (\sum_{i=1}^{n}(v_i)^2)^{\frac{1}{2}}$ and $\|u\|_\infty = \max_{i=1,2,\ldots,n}|u_i|$. So $\|u\|_\infty$ is an upper bound for each non-zero entry of u and we get the upper bound:

$$\|v\|_2 \leq \sqrt{k}\|v\|_\infty.$$

We conclude:

$$\frac{\|v\|_1}{\sqrt{k}} \leq \|v\|_2 \leq \sqrt{k}\|v\|_\infty.$$

Lemma 6.1.7. *Let A be an $m \times n$ matrix satisfying Restricted Isometry Property of order $2k$. Let $v \in \mathbb{R}^n$ be a non-zero vector. Let $S_0 \in [N]$ be any set such that $|S_0| \leq k$. We define S_1 as the index set of k largest entries of S_0^c and we set $S = S_0 \cup S_1$. Then, we have the following equation:*

$$\|v_S\|_2 \leq \alpha \frac{\|v_{S_0}\|_1}{\sqrt{k}} + \beta \frac{|\langle Av_S, Av\rangle|}{\|v_S\|_2},$$

where $\alpha = \frac{\sqrt{2}\delta_{2k}}{1-\delta_{2k}}$ and $\beta = \frac{1}{1-\delta_{2k}}$.

To prove this lemma, we need the following lemmas.

Lemma 6.1.8. *Suppose u, v are orthogonal vectors, then*

$$\|u\|_2 + \|v\|_2 \leq \sqrt{2}\|u+v\|_2.$$

Proof. We define a 2×1 vector $w = [\|u\|_2, \|v\|_2]^t \in \mathbb{R}^2$. We apply Lemma (6.1.6) with $k = 2$ and we have

$$\|w\|_1 \leq \sqrt{2}\|w\|_2.$$

Using $(a+b)^2 \leq 2(a^2+b^2)$, we have

$$\|u\|_2 + \|v\|_2 \leq \sqrt{2}\sqrt{\|u\|_2^2 + \|v\|_2^2}.$$

As u and v are orthogonal vectors, $\|u\|_2^2 + \|v\|_2^2 = \|u+v\|_2^2$, so we have

$$\|u\|_2 + \|v\|_2 \leq \sqrt{2}\|u+v\|_2.$$

Lemma 6.1.9. *Let A be an $m \times n$ matrix satisfying the Restricted Isometry Property of order $2k$, then for vector $u, v \in \Sigma_k$ with disjoint support,*

$$|\langle Au, Av \rangle| \leq \delta_{2k}\|u\|_2\|v\|_2.$$

Proof. Suppose $u, v \in \Sigma_k$ vectors with disjoint support, also that $\|u\|_2 = \|v\|_2 = 1$ and $u \perp v$. As supports of u and v are disjoint, $u+v \in \Sigma_{2k}$ and $\|u \pm v\|_2^2 = 2$. We apply the Restricted Isometry Property and we have

$$\|u \pm v\|_2^2(1-\delta_{2k}) \leq \|Au \pm Av\|_2^2 \leq \|u \pm v\|_2^2(1+\delta_{2k}).$$

So

$$2(1-\delta_{2k}) \leq \|Au \pm Av\|_2^2 \leq 2(1+\delta_2 k).$$

We know that

$$\|Au + Av\|_2^2 = \langle Au+Av, Au+Av \rangle = \langle Au, Au \rangle + \langle Au, Av \rangle + \langle Av, Au \rangle + \langle Av, Av \rangle,$$

and

$$\|Au - Av\|_2^2 = \langle Au-Av, Au-Av \rangle = \langle Au, Au \rangle - \langle Au, Av \rangle - \langle Av, Au \rangle + \langle Av, Av \rangle.$$

By taking the sum of these equations,

$$|\langle Au, Av \rangle| = \frac{1}{4}\left|\|Au+Av\|_2^2 - \|Au-Av\|_2^2\right| \leq \delta_{2k}.$$

We have shown that if $\|u\|_2 = \|v\|_2 = 1$ and they have disjoint supports and they are k-sparse then $|\langle Au, Av \rangle| \leq \delta_{2k}$.
Suppose $u, v \in \Sigma_k$ are non-zero with disjoint support and let $u_0 = \frac{u}{\|u\|}, v_0 = \frac{v}{\|v\|}$ so

$\|u_0\| = \|v_0\| = 1$ with disjoint support, then

$$|\langle Au_0, Av_0\rangle| = \frac{|\langle Au, Av\rangle|}{\|u\|\|v\|}.$$

So, we have

$$|\langle Au, Av\rangle| \leq \delta_{2k}\|u\|_2\|v\|_2.$$

Lemma 6.1.10. *Let $S_0 \subset [N]$ such that $|S_0| \leq k$. Let $v \in \mathbb{R}^n$ be a vector, we define S_1 as the set of the k largest absolute entries' index of $v_{S_0^c}$, a set S_2 as the set of the next k largest absolute entries' index and so on. Then, we have*

$$\sum_{j\geq 2}\|v_{S_j}\|_2 \leq \frac{\|v_{S_0^c}\|_1}{\sqrt{k}}.$$

Proof. By Lemma 6.1.6, we know that:

$$\sum_{j\geq 2}\|v_{S_j}\|_2 \leq \sqrt{k}\sum_{j\geq 2}\|v_{S_j}\|_\infty.$$

As S_j's are decreasing, we can say that for $j \geq 2$ $\|v_{S_j}\|_\infty \leq \frac{\|v_{S_{j-1}}\|_1}{k}$. So,

$$\sum_{j\geq 2}\|v_{S_j}\|_2 \leq \sqrt{k}\sum_{j\geq 2}\|v_{S_j}\|_\infty \leq \frac{1}{\sqrt{k}}\sum_{j\geq 1}\|v_{S_j}\|_1 = \frac{\|v_{S_0^c}\|_1}{\sqrt{k}}.$$

as $S_0^c = S_1 \cup S_2 \cup S_3 \cup \ldots \cup S_n$.

Lemma 6.1.11. *Let A be an $m \times n$ matrix satisfying Restricted Isometry Property of order $2k$ an let $v \in \mathbb{R}^n$ be a non-zero vector. Let $S_0 \subset [N]$ be a set such that $|S_0| \leq k$. We define the set S_1 as the index set of k absolute largest entries of $v_{S_0^c}$ and $S = S_0 \cup S_1$. Then*

$$\|v_S\|_2 \leq \alpha\frac{\|v_{S_0^c}\|_1}{\sqrt{k}} + \beta\frac{|\langle Av_S, Av\rangle|}{\|v_S\|_2}$$

where

$$\alpha = \frac{\sqrt{2}\delta_{2k}}{1-\delta_{2k}}, \quad \beta = \frac{1}{1-\delta_{2k}}.$$

Proof. We know that $v_S \in \Sigma_{2k}$ and as A satisfies Restricted Isometry Property of order $2k$ we have

$$(1-\delta_{2k})\|v_S\|_2^2 \leq \|Av_S\|_2^2.$$

We define S_j as in the previous lemma. By the definition $v_S = v - \sum_{j\geq 2} v_{S_j} = v_{S_0} + v_{S_1}$ and by linearity of the matrix A, $Av_S = Av - \sum_{j\geq 2} Av_{S_j}$. We have

$$(1-\delta_{2k})\|v_S\|_2^2 \leq \langle Av_S, Av_S\rangle = \langle Av_S, Av - \sum_{j\geq 2} Av_{S_j}\rangle.$$

$$(1-\delta_{2k})\|v_S\|_2^2 \leq \langle Av_S, Av\rangle - \langle Av_S, \sum_{j\geq 2} Av_{S_j}\rangle.$$

We know that $v_{S_i}, v_{S_j} \in \Sigma_k$ for $i \neq j$ have disjoint supports and they are orthogonal:

$$|\langle Av_{S_i}, Av_{S_j}\rangle| \leq \delta_{2k}\|v_{S_i}\|_2\|v_{S_j}\|_2$$

for any $i \neq j$.
Lemma (6.1.8) gives us $\|v_{S_0}\|_2 + \|v_{S_1}\|_2 \leq \sqrt{2}\|v_S\|_2$ and we have

$$\begin{aligned}
\left|\langle Av_S, \sum_{j\geq 2} Av_{S_j}\rangle\right| &= \left|\sum_{j\geq 2}\langle Av_{S_0}, Av_{S_j}\rangle + \sum_{j\geq 2}\langle Av_{S_1}, Av_{S_j}\rangle\right| \\
&\leq \sum_{j\geq 2}|\langle Av_{S_0}, Av_{S_j}\rangle| + \sum_{j\geq 2}|\langle Av_{S_1}, Av_{S_j}\rangle| \\
&\leq \delta_{2k}\|v_{S_0}\|_2 \sum_{j\geq 2}\|v_{S_j}\|_2 + \delta_{2k}\|v_{S_1}\|_2 \sum_{j\geq 2}\|v_{S_j}\|_2 \\
&= \delta_{2k}\sum_{j\geq 2}\|v_{S_j}\|_2(\|v_{S_0}\|_2 + \|v_{S_1}\|_2) \\
&\leq \sqrt{2}\delta_{2k}\|v_S\|_2 \sum_{j\geq 2}\|v_{S_j}\|_2.
\end{aligned}$$

We use Lemma (6.1.10):

$$\left|\langle Av_S, \sum_{j\geq 2} Av_{S_j}\right| \leq \sqrt{2}\delta_{2k}\|v_S\|_2 \frac{\|v_{S_0^c}\|_1}{\sqrt{k}}.$$

So, we have:

$$\begin{aligned}
(1-\delta_{2k})\|v_S\|_2^2 &\leq \left|\langle Av_S, Av\rangle - \langle Av_S, \sum_{j\geq 2} Av_{S_j}\rangle\right| \\
&\leq |\langle Av_S, Av\rangle| + \left|\langle Av_S, \sum_{j\geq 2} Av_{S_j}\rangle\right| \\
&\leq |\langle Av_S, Av\rangle| + \sqrt{2}\delta_{2k}\|v_S\|_2 \frac{\|v_{S_0^c}\|_1}{\sqrt{k}}.
\end{aligned}$$

We divide by $(1-\delta_{2k})\|v_S\|_2$:

$$\|v_S\|_2 \leq \sqrt{2}\delta_{2k}\frac{\|v_{S_0^c}\|_1}{\sqrt{k}(1-\delta_{2k}} + \frac{|\langle Av_S, Av\rangle|}{\|v_S\|_2(1-\delta_{2k}}.$$

We set:

$$\frac{\sqrt{2}\delta_{2k}}{1-\delta_{2k}} = \alpha, \quad \frac{1}{1-\delta_{2k}} = \beta.$$

So, we have:

$$\|v_S\|_2 \leq \alpha\frac{\|v_{S_0^c}\|_1}{\sqrt{k}} + \beta\frac{|\langle Av_S, Av\rangle|}{\|v_S\|_2}.$$

Notice that if $\delta_{2k} < \sqrt{2}-1$ then $\alpha < 1$, we will use it in the proof of next lemma.

Theorem 6.1.12. *Let A be an $m \times n$ matrix satisfying the Restricted Isometry Property of order $2k$ with $\sqrt{2}-1$. A satisfies the Null Space Property of order $2k$ with constant*

$$C = \frac{\sqrt{2}\delta_{2k}}{1-(1+\sqrt{2})\delta_{2k}}.$$

Proof. We know that the matrix A satisfies the Null Space Property of order $2k$ if there exists a non-negative constant C such that for all $v \in \mathcal{N}(A)$,

$$\|v_S\|_2 < C'\frac{\|v_{S^c}\|_1}{\sqrt{2k}} \leq C\frac{\|v_{S^c}\|_1}{\sqrt{k}} \quad \text{where} \quad |S| \leq 2k.$$

We need to show that for the set S as the index of $2k$ largest entries of v, $\|v_S\|_2 < C\frac{\|v_{S^c}\|_1}{\sqrt{k}}$.
We define S_0 the index set of k largest entries of v and we apply Lemma 6.1.11.
As $Av = 0$, the term $\beta\frac{|\langle Av_S, Av\rangle|}{\|v_S\|_2}$ is zero, so we have:

$$\|v_S\|_2 \leq \alpha\frac{\|v_{S_0^c}\|_1}{\sqrt{k}}.$$

We use Lemma 6.1.6 and we have:

$$\|v_{S_0^c}\|_1 \leq \sqrt{k}\|v_{S_0^c}\|_2.$$

We know that $S = S_0 \cup S_1$ where S_0 and S_1 are disjoint with $|S_0| = |S_1| = k$. We use Lemma 6.1.6

$$\|v_{S_0^c}\|_1 = \|v_{S_1}\|_1 + \|v_{S^c}\|_1 \leq \sqrt{k}\|v_{S_1}\|_2 + \|v_{S^c}\|_1.$$

So, we have:

$$\|v_S\|_2 \leq \alpha(\|v_{S_1}\|_2 + \frac{\|vS^c\|_1}{\sqrt{k}}).$$

We know that $\|v_{S_1}\|_2 \leq \|v_S\|_2$, so we have:

$$\|v_S\|_2 - \alpha\|v_S\|_2 \leq \alpha\frac{\|v_{S^c}\|_1}{\sqrt{k}}$$
$$\Rightarrow (1-\alpha)\|v_S\|_2 \leq \alpha\frac{\|v_{S^c}\|_1}{\sqrt{k}}.$$

We know that $\delta_{2k} < \sqrt{2} - 1$ by assumption, so $\alpha < 1$. We can divide by $(1-\alpha)$ and we get:

$$\|v_S\|_2 \leq \frac{\alpha}{1-\alpha}\frac{\|v_{S_0^c}\|_1}{\sqrt{k}}.$$

Hence, we proved the theorem with:

$$C = \frac{\alpha}{1-\alpha} = \frac{\sqrt{2}\delta_{2k}}{1-(1+\sqrt{2})\delta_{2k}}.$$

6.2 Greedy Methods

In this section, we study the orthogonal matching pursuit algorithm which is the most well-known greedy approach. We call S^n the target support and x^n the target vector which is supported on S^n. In each iteration of this algorithm, the index is added to the target support and the target is updated.

Orthogonal Matching Pursuit

Input: sensing matrix A, measurement vector y
Initialization: $S^0 = \emptyset$, $x^0 = 0$.
Iteration: repeat until a stopping criterion is met $n = \bar{n}$:

$$S^{n+1} = S^n \cup \{j_{n+1}\}, \quad j_{n+1} := \text{argmax}\{|A^*(y - Ax^n))_j|\},$$

$$x^{n+1} = \underset{z\in\mathbb{C}^{\mathbb{N}}}{\text{argmin}}\{\|y - Az\|_2, \text{supp}(z) \subset S^{n+1}\}$$

Output: the $\bar{n}$-sparse vector $x^\sharp = x^{\bar{n}}$

6.2.1 Experiments

We made some experiments using the Orthogonal Matching Pursuit algorithm. In the first experiment, we generate a random 256-dimensional 5-sparse vector x and a sensing matrix $A \in \mathbb{R}^{25\times256}$. Then, we calculate the measurement vector $y \in \mathbb{R}^{25\times1}$. We use the Orthogonal Matching Pursuit algorithm to recover x. So my variables of this experiment are $m = 25$, $N = 256$, $s = 5$.

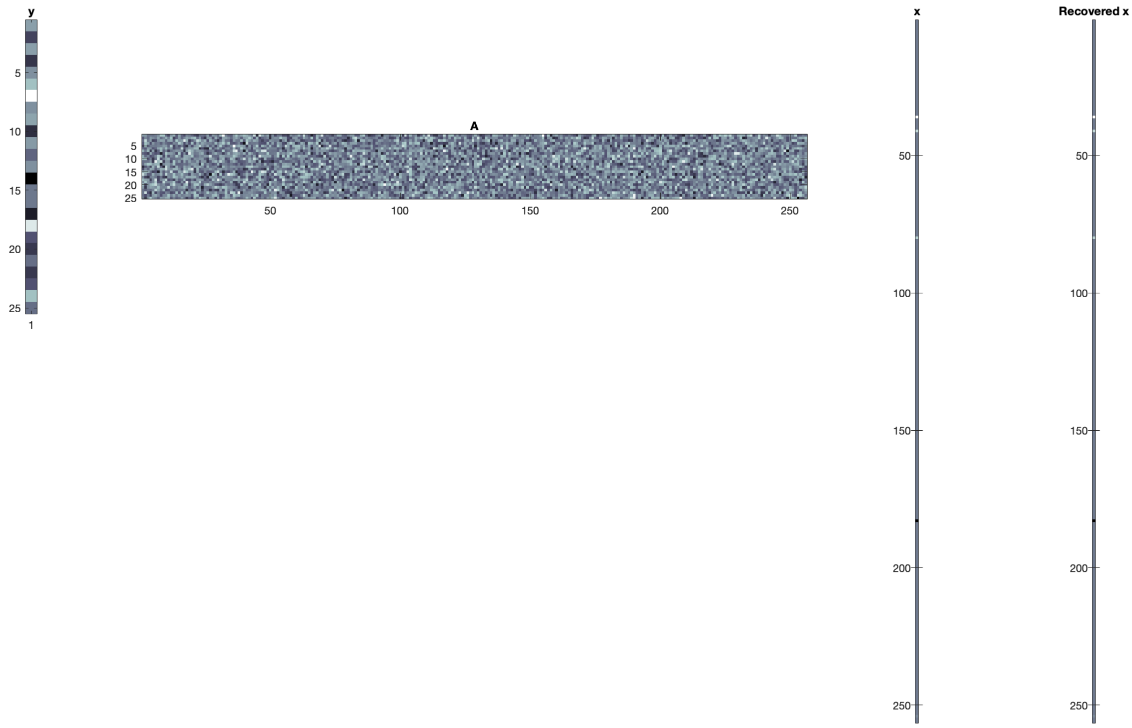

Figure 6.6: Recovery using the Random Gaussian Matrix and the Orthogonal Matching Pursuit algorithm

As shown in the figure above, the original vector x and the recovered x are exactly the same. We compare also the difference between these vectors using ℓ_1-norm and ℓ_2-norm. The difference calculated using ℓ_1-norm between the original vector x and the recovered vector is 1.249×10^{-15}. If we use ℓ_2-norm, the difference would be 7.494×10^{-16}.
We repeat this experiment using Gaussian matrices, Bernoulli matrices, Hadamard matrices and Toeplitz matrices 10000 times.

Sensing Matrix	**Error using ℓ_1**	**Error using ℓ_2**
Gaussian	0.2058	0.1555
Bernoulli	0.2069	0.1558
Hadamard	0.1678	0.1373
Toeplitz	0.5195	0.3911

Compared to the first experiment which have been only repeated once, these averages are very high. We also control if there is some error. We stored all the records in a list and it shows me that the errors are generally (8804 of 10000 for ℓ_1-norm) less than 1×10^{-4} but also there are some greater errors. For example, in one of the 10000 repetitions using Gaussian Matrix, the error was 2.6513.

6.3 Thresholding-Based Methods

We need to define two operators:

$$L_s(z) := \text{index set of } s \text{ largest absolute entries of } z \in \mathbb{C}^N$$

$$H_s(z) := z_{L_s(z)}$$

The operator H_s is called *hard thresholding operator* of order s. The operator H_s keeps the s largest absolute entries of the vector $z \in \mathbb{C}^N$ and set the others to zero.
The *basic thresholding algorithm* firstly determine the support of the s-sparse vector $x \in \mathbb{C}^N$ which we don't know yet. Then, it recovers the vector $x \in \mathbb{C}^N$ from the measurement vector $y = Ax \in \mathbb{C}^N$ as the indices of s largest absolute entries of A^*y and finds the vector with this support that best fits the measurement.

Basic Thresholding

Input: sensing matrix A, measurement vector y, sparsity level s.
Instruction:

$$S^\sharp = L_s(A^*y),$$

$$x^\sharp = \underset{z\in\mathbb{C}^{\mathbb{N}}}{\operatorname{argmin}} \left\{ \|y - Az\|_2,\ \operatorname{supp}(z) \subset S^\sharp \right\}$$

Output: the s-sparse vector $x^\sharp$.

Proposition 6.3.1. *Let $x \in \mathbb{C}^N$ be a vector supported on the set S. x can be recovered from $y = Ax$ using the basic thresholding algorithm if and only if*

$$\min_{j\in S} |(A^*y)_j| > \max_{l\in\overline{S}} |(A^*y)_l|.$$

Proof. We can say that x can be recovered if and only if the set $S^\sharp$ defined in basic thresholding algorithm and the set S are the same. It means that if any entry of A^*y on S is greater than any entry of A^*y on $\overline{S}$.

Iterative Hard Thresholding

Input: sensing matrix A, measurement vector y, sparsity level s.
Initialization: s-sparse vector x^0, typically $x^0 = 0$.
Iteration: repeat until a stopping criterion is met at $n = \overline{n}$:

$$x^{n+1} = H_s(x^n + A^*(y - Ax^n)).$$

Output: the s-sparse vector $x^\sharp = x^{\overline{n}}$.

6.3.1 Experiments

I repeated the experiment that I did with Orhogonal Matching Pursuit algorithm with Iterative Hard Thresholding algorithm. Compared to Orthogonal Matching Pursuit algorithm, using Iterative Hard Thresholding gives us bigger errors using ℓ_1-norm and ℓ_2-norm.

Sensing Matrix	**Error using ℓ_1**	**Error using ℓ_2**
Gaussian	1.0044	0.9885
Bernoulli	0.7522	0.5764
Hadamard	0.7607	0.5815
Toeplitz	1.2042	0.9390

In 8137 of 10000 experiments, the errors calculated using ℓ_1-norm and Gaussian matrices as sensing matrices are less than 1×10^{-4}. The greatest error using ℓ_1-norm is 2.5356. As the results of OMP experiments and IHT experiments, there is no big difference between them. But Orthogonal Matching Pursuit algorithm provided us better recoveries than Iterative Hard Thresholding algorithm.

7 CONCLUSION

In this chapter, we propose a novel design method to generate a sensing matrix. First, we will explain the new method and then we will make experiments with the matrices generated using the novel design method. Also, we will compare the recovery success' of this method for sensing matrices and the others.

7.1 A Novel Design Method for Sensing Matrices

The novel method is a deterministic method for generating sensing matrices. Firstly, we generate a random $n \times n$ matrix using the normal distribution whose entries are between 0 and 1. Let us call it *base matrix*, I use the word "base" to indicate that we use it to generate the sensing matrix. We need a real number c to use it in generation process of the matrix. We use the *base matrix* to generate a $n^2 \times n^2$ matrix and a random number $c \in \mathbb{R}$ between 0 and 1. Firstly, We divide the $n^2 \times n^2$ matrix to n^2 pieces of $n \times n$ matrix. We generate the $n^2 \times n^2$ matrix by looking the entries of $n \times n$ matrix. Then, as the generated matrix is a square matrix, we choose m random rows from this matrix.

The advantage of this method is instead of storing a $n^2 \times n^2$ matrix, it is enough to store an $n \times n$ matrix and the algorithm that generates $n^2 \times n^2$ matrix from $n \times n$ matrix.

I will explain this method with an example. Let B be a 2×2 random matrix whose entries are between 0 and 1 as follows:

$$B = \begin{bmatrix} 0.4 & 0.6 \\ 0.75 & 0.35 \end{bmatrix}$$

We define the matrix $B' = 1 - B$ and we take the constant $c = 0.5$.

$$B' = \begin{bmatrix} 0.6 & 0.4 \\ 0.25 & 0.65 \end{bmatrix}$$

We generate the 4×4 matrix A using the matrix B and B'. First, we denote matrix A as follows

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix}$$

We look at the matrix A like

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}$$

where $A_{i,j}$'s are 2×2 matrices.

We look at the $b_{1,1}$ and if it is bigger than the contant c then $A_{1,1} = B$ if not $A_{1,1} = B'$. In this example, $b_{1,1},\ b_{2,2} < c$ so $A_{1,1} = A_{2,2} = B'$ and $b_{1,2},\ b_{2,1} > c$ so $A_{1,2} = A_{2,1} = B$. Our generated matrix would be

$$A = \begin{bmatrix} 0.6 & 0.4 & 0.4 & 0.6 \\ 0.25 & 0.65 & 0.75 & 0.35 \\ 0.4 & 0.6 & 0.6 & 0.4 \\ 0.75 & 0.35 & 0.25 & 0.65 \end{bmatrix}$$

To terminate this process, we need to choose m random rows of from this matrix. Let me choose the first and the fourth rows, so our generated sensing matrix would be:

$$A = \begin{bmatrix} 0.6 & 0.4 & 0.4 & 0.6 \\ 0.75 & 0.35 & 0.25 & 0.65 \end{bmatrix}.$$

We repeated the experiments that I did before with the Orthogonal Matching Pursuit and Iterative Hard Thresholding algorithms using the novel design method for sensing matrices. The results are below:

Sensing Matrix	**Error using ℓ_1**	**Error using ℓ_2**
Gaussian	0.2058	0.1555
Bernoulli	0.2069	0.1558
Hadamard	0.1678	0.1373
Toeplitz	0.5195	0.3911
Novel Design Method	0.9331	0.7197

Table 7.1: Results using the OMP algorithm

Sensing Matrix	**Error using ℓ_1**	**Error using ℓ_2**
Gaussian	1.0044	0.9885
Bernoulli	0.7522	0.5764
Hadamard	0.7607	0.5815
Toeplitz	1.2042	0.9390
Novel Design Method	1.2643	0.9838

Table 7.2: Results using the IHT algorithm

REFERENCES

Y. Arjoune, N. Kaabouch, H. el ghazi, and A. Tamtaoui. A performance comparison of measurement matrices in compressive sensing. *International Journal of Communication Systems*, 10 2017.

E. J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1998.

A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *American Mathematical Society*, 22:211–231, 01 2009.

M. Davenport, M. Duarte, Y. Eldar, and G. Kutyniok. Introduction to compressed sensing. *Preprint*, 93, 01 2012.

D. Donoho and P. Stark. Uncertainty principles and signal recovery. *Siam Journal on Mathematical Analysis - SIAM J MATH ANAL*, 49:906–931, 06 1989.

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

D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, 2003.

D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47(7):2845–2862, 2001.

S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Birkhäuser Basel, 2013.

R. Gribonval and M. Nielsen. Highly sparse representations from dictionaries are unique and independent of the sparseness measure. *Applied and Computational Harmonic Analysis*, 22(3):335 – 355, 2007.

V. Kotelnikov. On the capacity of ether and wire in telecommunications. In *Proceedings of the First All-Union Congress on Technical Reconstruction of Communication, Union Energy Committee*, 1933.

H. Nyquist. Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47(2):617–644, 1928.

M. Safarpour, I. Hautala, and O. Silvén. Iht mine_m, 01 2020.

C. E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.

J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 53(12):4655–4666, 2007.

E. T. Whittaker. Xviii.—on the functions which are represented by the expansions of the interpolation-theory. *Proceedings of the Royal Society of Edinburgh*, 35:181–194, 1915.

S. Yoo. Matching pursuit (mp) / orthogonal matching pursuit (omp). https://github.com/seunghwanyoo/omp, 2016.

Appendix

7.2 Sensing Matrix Generation

7.2.1 Gaussian Matrix

The code below generates a 25×256 Gaussian Matrix.

```
N=256;
m=25;
A = randn(m,N);
```

Listing 7.1: Gaussian Matrix Generation

7.2.2 Bernoulli Matrix

The code below generates a 25×256 Bernoulli Matrix. We use $p = 0.5$ to get equal probability of 0 and 1.

```
N=256;
m=25;
p = 0.5;          %probability of success

A = rand(m,N);
A = double(A<p);
```

Listing 7.2: Gaussian Matrix Generation

7.2.3 Hadamard Matrix

The code below generates a 25×256 Hadamard Matrix using hadamard function in Matlab.

```
N=256;
m=25;

H = hadamard(N);
rows = randperm(N,m);
```

```
A = H(rows ,:);
```

Listing 7.3: Hadamard Matrix Generation

7.2.4 Toeplitz Matrix

The code below generates a 25×256 Toeplitz Matrix using toeplitz function in Matlab.

```
N=256;
m=25;

T = toeplitz(rand(1,N));
rows = randperm(N,m);
A = T(rows ,:);
```

Listing 7.4: Gaussian Matrix Generation

7.3 Restricted Isometry Test Using Different Sensing Matrix Design Methods

The code below calculates the recovery success using Gaussian Matrix as sensing matrix and Iterative Hard Thresholding algorithm as recovery algorithm. It repeats the recovery process 1000 times and calculates the average error using l_1-norm and l_2-norm.

```
n = 16;
N = n*n;
m = 16;
k = 6;

% Gaussian
%M = randn(m,N);
%M = randn(m,N) > 0;


% Hadamard
M = hadamard(N);
M = M(randperm(N,m) ,:);
```

```
% Toeplitz
%T = toeplitz(rand(1,N));
%rows = randperm(N,m);
%M = T(rows,:);


% My Custom Matrix
%M = my_custom_matrix_v2(m,n);

d = [];

for q = 1:100000

    xa = zeros(N,1);
    support = randperm(N,k);
    xa(support) = (randn(size(support)));

    xb = zeros(N,1);
    support = randperm(N,k);
    xb(support) = (randn(size(support)));

    ya = M*xa;
    yb = M*xb;

    d(q) = norm(xa-xb)/norm(ya-yb);

end

ric_max = max(d)
ric_min = min(d)
d2 = d(d >= 0.1 & d <= 0.3);
dc = length(d2)/100000
```

```
histogram(d)
```

Listing 7.5: Matlab code for the RIP using different sensing matrix design methods

7.4 Novel Design Method For Sensing Matrices

The code below calculates the recovery success using Gaussian Matrix as sensing matrix and Iterative Hard Thresholding algorithm as recovery algorithm. It repeats the recovery process 1000 times and calculates the average error using l_1-norm and l_2-norm.

```
function M = my_custom_matrix_v2(m,N)
M = zeros(N^2);
support = randperm(N^2,m);

B = rand(N);
B_t = 1-B;

k = 1;

s = rand();

for d=1:N:N^2
    for c=1:N:N^2
        if B(k)<s
            M(c:c+N-1,d:d+N-1) = B_t;
        else
            M(c:c+N-1,d:d+N-1) = B;
        end
        k = k+1;
    end
end
M = M(support,:);
```

Listing 7.6: Matlab code for novel design method for sensing matrices

7.5 Experiment to Compare Sensing Matrices Using OMP and IHT Algorithms

The code below calculates the recovery success using Gaussian Matrix as sensing matrix and Iterative Hard Thresholding algorithm as recovery algorithm. It repeats the recovery process 1000 times and calculates the average error using l_1-norm and l_2-norm. I get the Orthogonal Matching Pursuit algorithm Matlab implementation from [18] and Iterative Hard Thresholding algorithm Matlab implementation from [14].

```
clear
n=16;
N=n*n;
m=25;
k=5;

c = 10000;

l1_sum = 0;
l2_sum = 0;

l1es = [];
l2es = [];

l1_suc_rate = 0;
l2_suc_rate = 0;

for i=1:c
    % Gaussian Normal Distribution

    A = randn(m,N);
    A = A/norm(A,2);


    % Random Bernoulli
```

```
%       p = 0.5;          %probability of success
%       A = rand(m,N);
%       A = double(A<p);
%       A = A*2 - 1;


    % Hadamard

%       H = hadamard(N);
%       rows = randperm(N,m);
%       A = H(rows,:);


    % Toeplitz

%       T = toeplitz(rand(1,N));
%       rows = randperm(N,m);
%       A = T(rows,:);


    % My Custom Matrix

%       A = my_custom_matrix_v2(m,n);


    % Construct x for system Ax = b
    x = zeros(N,1);
    support = randperm(N,k);
    %x(support) = sign(randn(size(support)));
    x(support) = randn(size(support));
    y = A*x;

    %x_rec = omp(A,y,k);
    x_rec = iht_mine(y,A,k,N);
```

```
    %l1es = [l1es norm(x-x_rec ,1)/norm(x,1)];
    %l2es = [l2es norm(x-x_rec ,2)/norm(x,2)];

    l1_sum = l1_sum + norm(x-x_rec ,1)/norm(x,1);
    l2_sum = l2_sum + norm(x-x_rec ,2)/norm(x,2);
end

l1_avg = l1_sum / c
l2_avg = l2_sum / c
```

Listing 7.7: Matlab code for the experiment to compare sensing matrices using Orthogonal Matching Pursuit and Iterative Hard Thresholding Algorithms

BIOGRAPHICAL SKETCH

Utku Kabuli AYTAÇ

İstanbul, Turkey, 1994

Education

2013-2017	Bachelor of Science in MATHEMATICS **Galatasaray University**, Istanbul, Turkey
2008-2013	High School Diploma in SCIENCE DEPARTMENT **İstanbul Saint-Joseph French High School**, İstanbul, Turkey