SUCCESSFUL AND FAILING MATRICES FOR l_1-RECOVERY OF SPARSE VECTORS

Aras T. Asaad

Department of Applied Computing

The University of Buckingham

Supervisor:

Dr. Stuart Hall

Thesis submitted for the Degree of MSc in Mathematics to the School of Science in the University of Buckingham

December 2012

Buckingham, United Kingdom

Declaration of Originality

I hereby declare that the thesis here presented is my own work. All work appearing in it is my own except that which is referenced.

Abstract

In this thesis we give an overview of the notion of compressed sensing together with some special types of compressed sensing matrices. We then investigate the Restricted Isometry property and the Null Space property which are two of the most well-known properties of compressed sensing matrices needed for sparse signal recovery. We show that when the Restricted Isometry constant is 'small enough' then we can recover sparse vectors by l_1-minimization. Whereas if the Restricted Isometry constant is 'large', we show that l_1-minimization fails to recover all sparse vectors.

Acknowledgements

In the name of God, the most gracious and the most merciful.

All praises to Allah for the strengths and his blessing in completing this thesis.

Next, I wish to express my deepest gratitude to my supervisor Dr. Stuart James Hall, for his constant support, patience, constructive comments and valuable supervision. I consider myself very lucky to be supervised by a person of such a wonderful personality and enthusiasm.

Not forgotten, my appreciation also goes to Prof. Sabah Jassim for his constant support, advice and discussions that we had about this work.

I am grateful to the Kurdistan Regional Government (KRG) for funding this MSc and the Ministry of higher education and scientific research in my country Kurdistan for nominating me for the Human capacity development programme (HCDP) scholarship award.

My gratitude also goes to my employer and former University, Koya University, for giving me study leave permission.

Sincere thanks to all my friends and college mates especially, Alan, Makki, Abdulbasit, Rasper, Azhin, Halgurd, Hawkar, Muhammed, Sherko, Bawar, Ranj, Briar and others for their moral support during my study. Thanks for the friendship and memories.

Last but not least, my deepest gratitude goes to my parents Mr.Tawfeeq and Mrs.Bahiyah and my two beloved brothers Kewan and Hardy. Also not forgetting my wife, Suriyah M. Ahmed for her love and care. To those who indirectly contributed in this research, your kindness means a lot to me. Thank you very much.

Aras T Asaad

November - 2012

Contents

Table of Figures

Chapter 1

1 INTRODUCTION

Nowadays people live in the centre of a digital revolution that is driving the development and deployment of new kinds of sensing systems with ever-increasing accuracy and resolution. By looking at the typical way that signals are processed, one finds that signals are captured at a rate far from the information rate needed for the traditional technique in which signals are sampled at a rate satisfying the Nyquist-Shannon theorem. This theorem (technique) states that to avoid losing information when capturing a signal, one must sample the signal two times faster than the signal bandwidth (i.e. twice the highest frequency) [2]. On the other hand, when it comes to storing this signal, in images or videos for example, eventually a lot of the captured information will be thrown away in order to compress the signal to fit the available storage/processing capacity. Therefore, only a small proportion of the captured signal will be kept as compared with the amount captured. Because of this people started asking: can we just measure at the information rate? This is exactly the question that the new acquiring paradigm known as *compressed sensing* is trying to answer [6]. In this thesis, the mathematical concepts of compressed sensing, which is a classical problem in linear algebra, will be presented first together with some particular applications of this new sensing scheme. Furthermore, properties of the sensing matrices (also known as dictionaries) that compressed sensing work by will be discussed. We shall also highlight the difficulties that exist, in terms of both computation and construction, in designing dictionaries that satisfy these properties. Finally, the form of the matrices where compressed sensing theory cannot be verified will be discussed in details with some motivations to build compressed sensing matrices. This work contains some basic but interesting results, e.g. theorem (2.1) and theorem (2.3) together with a detailed exposition of results given in [11] for constructing failing compressed sensing matrices.

1.1 Solution of Linear Equations

In this chapter we give a brief introduction on solutions of linear equations in general and then we focus on the particular case of underdetermined systems where we have infinitely many solutions. As a motivational introduction to compressed sensing, we give some real-world applications for which the corresponding underdetermined systems have infinitely many solutions.

We start by considering a system of M linear equations with N unknowns which has the form

$$y_1 = \varphi_{11}x_1 + \varphi_{12}x_2 + \cdots + \varphi_{1N}x_N$$

$$y_2 = \varphi_{21}x_1 + \varphi_{22}x_2 + \cdots + \varphi_{2N}x_N$$

$$\vdots$$

$$y_M = \varphi_{M1}x_1 + \varphi_{M2}x_2 + \cdots + \varphi_{MN}x_N.$$

The above system can also be written in the form of matrices as below:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1N} \\ \varphi_{21} & \varphi_{22} & \cdots & \varphi_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \varphi_{M1} & \varphi_{M2} & \cdots & \varphi_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}.$$

More concisely we can say that

$$y = \Phi x \tag{1.1}$$

where $\Phi \in \mathbb{R}^{M\times N}$ is the coefficient matrix, $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \in \mathbb{R}^N$ and $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} \in \mathbb{R}^M$ are vectors.

In general, there exist a solution to (1.1) if and only if $y \in \mathcal{R}(\Phi)$, where $\mathcal{R}(\Phi)$ is the range of Φ. If (1.1) has a solution, then we have the following two possibilities:

1- $M \geq N$: The solution is unique iff N equals the rank of the coefficient matrix Φ.
2- $M < N$, (1.1) has infinitely many solutions.

The last case is known as an **underdetermined** system of equations. To sum up, basically system (1.1) is either inconsistent (i.e. there is no solution to it) or consistent (i.e. there is a solution to the system (unique or infinite)). We are interested in focusing on underdetermined systems where there are infinitely many solutions, especially when the number of rows is significantly less than the number of columns (i.e. $M \ll N$). This interest relates to the situation when; we have some data (i.e. a high dimensional vector) and we have some measurements (i.e. a linear combination of coordinates) but the number of measurements is not enough to reconstruct the original data (or at least it looks like we do not have enough measurements to figure out what the data is). This is the principle that *Compressed Sensing* works by. Compressed Sensing is about the conditions under which we can solve the underdetermined system and the methods by which we can reconstruct/sense the data. This principle of compressed sensing is very important for the applications in which measurements are expensive or very limited.

In the rest of the introduction, we give a brief description of some compressed sensing applications. Firstly, we describe the Netflix problem as an instance of recommender systems which yields an underdetermined system of equations that can naturally be solved by compressed sensing techniques. In the second example, we recast traits that are produced by some genes as an underdetermined system which has a special structure that enables applying compressed sensing. Lastly, the role of compressed sensing in Magnetic Resonance Imaging scanner is discussed.

The dissertation's outline is as follows. Chapter 1 is an introduction, presenting solution to linear systems of equations, in general, and three examples of compressed sensing applications. It also contains some basic definitions and general discussion about l_0-norm and l_1-norm. The restricted isometry property, which is a sufficient condition that compressed sensing matrices should satisfy and its relation with null space property is discussed in Chapter 2. Finally chapter 3 interprets matrices that fail to satisfy the Null Space Property, and motivations to construct a general form of these matrices.

1.2 Compressed sensing applications

The term 'compressed sensing' first coined by David Donoho in [12]. It is a field which has been growing slowly in the last two decades and exploded in the last six years. The list of real-world applications of compressed sensing growing fast and includes recommender systems (e.g. Netflix problem), genetics, Medical Imaging, Signal processing, recovery of missing data, computational biology, machine learning and many more. We selected MRI, the Netflix problem [1] and genomic to discuss in more detail.

1.2.1 Netflix Prize

The Netflix prize[1] is one of the most famous applications of CS. The Netflix has thousands of movies and millions of customers. The Netflix prize problem is about being able to predict

[1] Netflix is a company that rents out movies to costumers, at the beginning they were sending the movies to the costumers' home but now they have an online website whereby you can watch thousands of movies.

a very large rating matrix whose rows are indexed by movies and columns are indexed by costumers and the (i, j) ratings entry the hypothetical value assigned by customer i to the j-th movie. One would like to complete this matrix so that Netflix might recommend titles that any particular user is likely to be willing to order. Of course, not every costumer has rented all movies in Netflix, therefore only a small fraction of entries are actually known. However, if one makes an assumption that most costumers' rating preference is determined by only a small number of characteristics of the movies, (e.g. genre, year, director, actor/actresses etc), then the matrix should be of (approximately) low-rank. Hence, we end up with a highly underdetermined system of equations that has a special structure where CS paradigm could lead to a solution.

The next example has been paraphrased from [8].

1.2.2 Compressed sensing in Genetics

Another real-world application of CS is in genetics. We begin by considering system (1.1) such that $M \ll N$ (i.e. number of rows is much less than the number of columns). Doctors are interested in finding the location on the genome which is responsible for a trait, for instance cholesterol level. For modelling this problem mathematically, assume y to be the number of measured cholesterol level of patient i, and Φ be a matrix whose rows (i) present the i-th patients and columns (j) is the locations of genome responsible of producing trait, see figure (1). Let us assume locations on the genome that possibly have an influence on cholesterol level to be 500,000 and we have only a few thousand people to study. Therefore, it is a highly underdetermine system of equations to solve. Nonetheless, there are few genes that might be responsible for cholesterol level and in this case we might be able to assume x is sparse. Then the solution will be sparse as well. Roughly speaking sparse means that there are lots of zeros in the solution and we give the formal definition of sparse solution later in this Chapter.

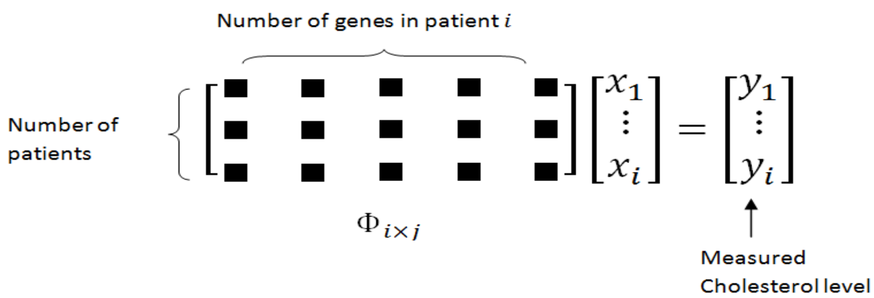

Figure 1: Mathematical model for determining locations on genes that produce trait.

1.2.3 Compressed Sensing and Magnetic Resonance Imaging (MRI)

The theory of compressed sensing was initially inspired by a problem in Medical imaging, specifically in Magnetic Resonance Imaging (MRI), see figure (2). The problem was to speed up the acquisition time in MRI scanning which has serious limitations. However, the speed at which data can be collected in MRI is fundamentally limited by physical (gradient amplitude and slew-rate) and physiological (nerve stimulation) constraints. For instance, when a patient is inside MRI scanner device, he or she should not move otherwise the image collected by the scanner will be blurred. Therefore, many researches are seeking for methods to reduce the amount of data acquired without degrading the image quality. In other words, since we cannot reduce the relaxation time we might think of taking fewer samples to create a good resolution image. Now, since most of MRI images after transforming to an appropriate domain are sparse, CS can be used for this purpose.

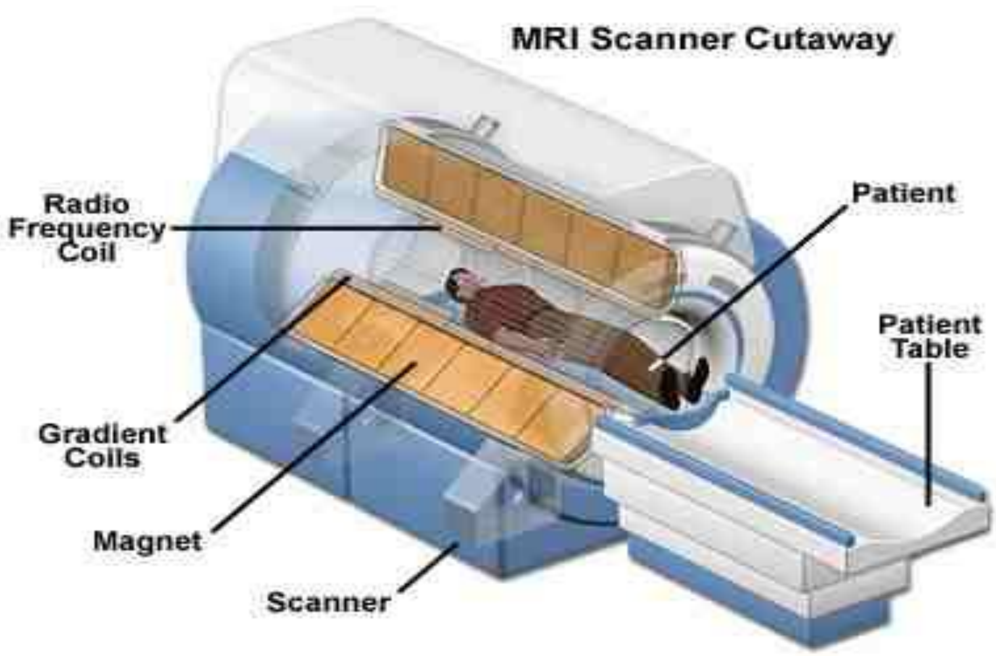

Figure 2: What is inside Magnetic Resonance Scanning [9].

1.3 l_1 -Minimization versus l_0- Minimization

In this sub-section, we start by defining l_0-norm which we need in order to define the sparse solution. Later, the reason behind minimizing l_1-norm instead of l_0-norm will be discussed.

We also give an example to show that minimizing l_1-norm may not be the same as minimizing to l_0-norm, and present a theorem by E.J.Candes, which states that under some conditions (that we discuss later in this chapter) l_1-minimization is equivalent to l_0-minimization.

Definition (1.1) **[$\boldsymbol{l_0}$-norm]:**

Let $x = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \in \mathbb{R}^N$ be a vector, then l_0-norm denoted $\|\cdot\|_{l_0}$, for a vector x is

$$\|x\|_{l_0} = \sum_{i=1}^{N} [1 \; if \; x_i \neq 0, 0 \; if \; x_i = 0].$$

Note that l_0-norm is not quite a norm according to the mathematical definition of norm, because the following axiom of norm does not hold

$$\|\alpha v\| = |\alpha| \|v\|$$

when α is a negative scalar, and v is an element of a vector space V. Next, we define the sparsest solution by using definition (1.1).

Definition (1.2) **[Sparsest Solution]:**

The solution of an underdetermined system of equations $y = \Phi x$ with the smallest (minimum) l_0-norm is called the sparsest solution, and is denoted $x_{l_0}^*$, i.e.

$$\|x^*\|_{l_0} = \min_{\substack{x \in \mathbb{R}^N \\ y = \Phi x}} \|x\|_{l_0} .$$

Solving an underdetermined system whose solution is sparse is like finding the sparsest solution that explains the data. Mathematically, we can cast this as an optimization problem where we try to find among all vector x such that $y = \Phi x$ the one that has the smallest l_0-norm (i.e. we want to find the smallest number of columns of Φ such that the linear combination of these columns yields y), and it is known to NP-hard problem because no deterministic algorithm is known to solve the problem in a polynomial time. Roughly speaking, there are two classes of problems, ‘P’ (i.e. easy to find) versus ‘NP’ (i.e. easy to check) which is about investigation of which kind of problems can be solve by computers, and which types cannot. Basically, P-class problems are "easy" for computers to solve; that is, solutions to these problems can be computed in a reasonable amount of time compared to the complexity of the problem. Meanwhile, for NP-hard problems, a solution might be very hard to find perhaps requiring an unknown number of years' worth of computation, but once found, it is easily checked[2]. Therefore, instead of minimizing l_0-norm we try to minimize what is known as l_1-norm because this is a 'P' problem. In this manner, we need to state the definition of l_1-norm.

Definition (1.3) [$\boldsymbol{l_1}$**-norm**]:

Let $x = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \in \mathbb{R}^N$ be a vector, then l_1-norm for the vector x is

$$\|x\|_{l_1} = \sum_{i=1}^{N} |x_i|$$

and, given a system (1.1)

$$\|x^*\|_{l_1} = \min_{\substack{x \in \mathbb{R}^N \\ y = \Phi x}} \|x\|_{l_1}$$

where $x^*_{l_1}$ is a minimization of l_1-norm. In general,

$$x^*_{l_1} \neq x^*_{l_0} .$$

Let us take an example to illustrate this. Consider the underdetermined system below

$$\begin{bmatrix} \sqrt{3} & \sqrt{3} & \sqrt{3} \\ 0 & -2\sqrt{5} & \sqrt{5} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{1.2}$$

Then the vector

$$x = \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}$$

solves the above system. Let $\Omega = \{1\}$ and $\Omega^c = \{2,3\}$ where $\Omega \subseteq \{1,2,3\}$ for $N = 3$ and

[2] Clay Mathematics Institute in U.S listed 'P versus NP' as one of its millennium problems.

$x_\Omega = x_{\{1\}} = \begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix}$. This implies that $x_\Omega = \begin{bmatrix} -3 \\ 0 \\ 0 \end{bmatrix}$ and $x_{\Omega^c} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$. Next, we find l_1-norm and l_0-norm for both x_Ω and x_{Ω^c} . Now,

$$\|x_\Omega\|_{l_1} = 3 \quad \text{and} \quad \|x_\Omega\|_{l_0} = 1.$$

But

$$\|x_{\Omega^c}\|_{l_1} = 3 \quad \text{and} \quad \|x_{\Omega^c}\|_{l_0} = 2\,.$$

Therefore, $x^*_{l_1} \neq x^*_{l_0}$. In other words, $x^*_{l_0}$ is a unique minimizer of (1.2) system but $x^*_{l_1} = 3$ is not the unique minimizer of (1.2). We can take $x_\Omega = -z_{\Omega^c}$ which also minimizes (1.2). This implies that

$$\Phi(z_\Omega + z_{\Omega^c}) = 0.$$

Then

$$\Phi z_\Omega = -\Phi z_{\Omega^c}$$

and now

$$\mathrm{y} = \Phi z_\Omega.$$

Hence $x^*_{l_1}$ is not unique in general. Moreover, under some conditions when Φ is '*nice*' then $x^*_{l_1} = x^*_{l_0}$ [7]. In order to state the theorem whereby $x^*_{l_1} = x^*_{l_0}$, we need two definitions which are restricted isometry property and sparse vectors.

Definition (1.4) [***k*-sparse vectors]:**

Let $x \in \mathbb{R}^N$ be a vector. Then x is said to be a k-sparse vector if

$$\|x\|_{l_0} \leq k$$

In other words, a vector x is said to be k-sparse if at most k coefficients of x_i are non-zero.

In order to be able to define the property which is known as *Restricted Isometry Property* (RIP) we need to define what is known as l_2-norm.

Definition (1.4) [$\boldsymbol{l_2}$**-norm**]:

Let $x = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \in \mathbb{R}^N$ be a vector, then l_2-norm for the vector x is

$$\|x\|_{l_2} = \sqrt{\sum_{i=1}^{N} {x_i}^2}.$$

Next, we define Restricted Isometry, a property which was first introduced by E. Candes and T.Tao in [4].

Definition (1.5) [**Restricted Isometry Property (RIP)]:**

An $M \times N$ matrix Φ is said to have the restricted isometry property (RIP) of order k if there exist $\delta_k \in (0,1)$ such that

$$(1 - \delta_k)\|x\|_{l_2} \leq \|\Phi x\|_{l_2} \leq (1 + \delta_k)\|x\|_{l_2} \tag{1.3}$$

for all k-sparse vectors x. Also, one can define the Restricted Isometry Constant (RIC) as the smallest number δ_k such that (1.3) holds for every k-sparse vector x.

Note that in (1.3), Φ is a big matrix such that the number of its rows are much less than the number of its columns (i.e. $M \ll N$). This means that the matrix Φ cannot act like an isometry on arbitrary vector because Φ has a huge null space, denoted $\mathcal{N}(\Phi)$. In other words, there are many x's (arbitrary vectors) for which Φx will be zero. In terms of RIP definition, this means that we cannot have a lower bound in (1.3). But if we assume that the RIC δ_k is not too 'big' in (1.3), then it means that by selecting k-columns of Φ then if x is k-sparse we want to preserve a norm. Another way of expressing RIP definition is that by extracting k-columns of Φ, sub-matrices will be well-conditioned. In fact the condition number in (1.3) is simply $\left(\frac{1+\delta_k}{1-\delta_k}\right)$. Geometrically, RIP gives that k-columns of Φ are not orthogonal if one represent them in space but they are not too far from orthogonal. Therefore, by assuming that the solution of (1.1) is k-sparse then it is possible to recover x if the sparse vectors lie away from the null space of Φ ($\mathcal{N}(\Phi)$) because if $\mathcal{N}(\Phi)$ contains sparse vectors then $\Phi x = 0$ so there is nothing we can do. We previously showed that $x_{l_0}^* \neq x_{l_1}^*$, but in 2008 E.J. Candes showed that $x_{l_1}^* = x_{l_0}^*$ by the following theorem:

***Theorem** (1.1) (Candes, 2008 [7]):*

If the matrix Φ has the property of RIP with the restricted isometry constant $\delta_{2k} < \sqrt{2} - 1$, then for all k-sparse vectors x such that $y = \Phi x$, the solution of $x_{l_1}^* = x_{l_0}^*$.

In the next chapter, we will discuss proof of theorem (1.1) after providing appropriate lemmas that we need to construct the proof. Also, in chapter 2, we discuss a known technique for computing RIP and we propose a method in this manner which requires less computational cost. Finally, by the end of chapter 2 we see that when the restricted isometry constant is small 'enough' then we can find a sparse solution (unique sparse solution) of system (1.1) by using l_1-minimization. Whereas in chapter 3 we will show that when the restricted isometry constant reaches 1 (more concisely when $\delta_{2k} \approx 0.7071$) then l_1-minimization cannot recover **all** sparse vectors.

Chapter 2

2 Restricted Isometry Property and Null Space Property

In this chapter we aim to show that the restricted isometry property does imply the null space property (NSP) (the formal definition of null space property will be given later in this chapter). This relation will be showed by the following theorem:

Theorem *(2.1) ([10] Devanport et al., Theorem 1.5)**:** Suppose that $\Phi \in \mathbb{R}^{M\times N}$ is a matrix satisfying the RIP of order $2k$ with the restricted isometry constant (RIC) $\delta_{2k} < \sqrt{2} - 1$. Then Φ satisfies the NSP of order $2k$ with constant*

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

The idea of the proof is based on the same approach that given in [10], but we give a full explanation of the proof. We shall first prove a few appropriate lemmas and proposition that we need to construct the proof of theorem (1.1). The first step is to show that for any vector x, we have

$$\|x_\Lambda\|_{l_2} \le C\frac{\|x_{\Lambda^c}\|_{l_1}}{\sqrt{k}}$$

for the case where Λ is the index set corresponding to the $2k$ largest entry of x, Λ^c is the complement of Λ. Lemma (2.8) is another essential step of the proof states that if a matrix Φ has RIP of order $2k$ then any nonzero vector $x \in \mathbb{R}^N$ has the following form

$$\|x_\Lambda\|_{l_2} \le \alpha\frac{\left\|x_{\Lambda_0^c}\right\|_{l_1}}{\sqrt{k}} + \beta\frac{|\langle\Phi x_\Lambda, \Phi x\rangle|}{\|x_\Lambda\|_{l_2}}$$

where $\alpha = \frac{\sqrt{2}\delta_{2k}}{1-\delta_{2k}}$, $\beta = \frac{1}{1-\delta_{2k}}$, $\Lambda_0 \subseteq \{1,2,\dots,N\}$ and whose cardinality $|\Lambda_0| \le k$. In particular, we apply lemma (2.8) to the case where $x \in \mathcal{N}(\Phi)$. Then lemma (2.7) will be applied to $\left\|x_{\Lambda_0^c}\right\|_{l_1}$. The proof of the theorem requires many other technical results about the various norms computations. Furthermore, the techniques of computing RIP of matrices will be discussed later in this chapter but we start by proposing a method whereby RIP of a matrix can be computed more efficiently in comparison using the definition of RIP.

***Proposition** (2.1): Let $\Phi \in \mathbb{R}^{M\times N}$ be a matrix, then Φ has RIP of order k if there exist a RIC $\delta_k \in (0,1)$ such that*

$$(1-\delta_k)\|x\|_{l_2} \leq \|\Sigma x\|_{l_2} \leq (1+\delta_k)\|x\|_{l_2}$$

for all k-sparse vectors x, and Σ is the diagonal $M \times N$ matrix whose entries are the first singular values of $\Phi^{\mathrm{T}}\Phi$.

We spilt the proof of this proposition in to two cases; firstly we show that when $M = N$ then the problem of computing RIP of matrices will be reduced to computing Eigenvectors, secondly for underdetermined systems (i.e. when $M < N$) we simply construct the inequality in proposition (2.1) by considering what is known as *Singular Value Decomposition (SVD)* method together with the fact that two of the matrices that SVD procedure produces are preserving norms.

Finally, we end this chapter by giving a brief discussion about geometrical behaviour of RIP of order 1, i.e. when the number of nonzero component of the vector x is 1. Then we show that when Φ has RIP of order $2k$ with $\delta_{2k} < 1$ then any k-sparse solution to system (1.1) is a unique solution.

Before going to discuss the relation between RIP and NSP we need to define NSP. For stating NSP we need to define some particular notation. In this manner, let $\Phi \in \mathbb{R}^{M\times N}$ be a given matrix, $\Lambda \subseteq \{1,2,\dots,N\}$ and $|\Lambda| \leq k$, then for any vector $x = (x_1, x_2, \dots, x_N)$ the vector x_Λ is obtained from x by keeping all coordinates whose indices are in Λ and replace all other coordinates by 0. For instance the subset $\Lambda = \{1,3,5,10\}$. Then the vector x according to the index set Λ will be as follow:

$$x_\Lambda = x_{\{1,3,5,10\}} = (x_1, 0, x_3, 0, x_5, 0,0,0,0, x_{10}, 0, \dots).$$

In words, vector x_Λ contains four non-zero entries in the first, third, fifth and tenth row otherwise the entries are all zero up to the twentieth entry. Furthermore, the vector x in terms of the complement of Λ (i.e. Λ^c) will be as follow

$$x_{\Lambda^c} = x_{\{2,4,6,7,8,9,11,\dots\}} = (0, x_2, 0, x_4, 0, x_6, x_7, x_8, x_9, 0, x_{11}, \dots).$$

We now turn to define another property of matrices called the Null space property (NSP). We give a definition of NSP and then we discuss the relation between RIP and NSP.

Definition (2.1) **[Null Space Property]:**

An $M \times N$ matrix Φ is said to have the Null Space Property (NSP) of order k for $C \in (0,1)$ if

$$\|x_\Lambda\|_{l_1} \leq C\|x_{\Lambda^c}\|_{l_1} \tag{2.1}$$

where $\Lambda \subseteq \{1,2,\dots,N\}, |\Lambda| \leq k,$ and $x \in \mathcal{N}(\Phi)$.

2.1 Relation between RIP and NSP

In order to be able to state the relation between RIP and NSP, we need a series of lemmas. We start by considering the Cauchy-Schwarz inequality and Triangle inequality and we show the proof of the second one.

Lemma (2.1) **[Cauchy-Schwarz inequality]:**

For any vectors x and y in $\mathbb{R}^N$

$$|\langle x,y\rangle| \leq \|x\|_{l_2}\|y\|_{l_2} \tag{2.2}$$

where $|\langle x,y\rangle|: \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ is the absolute value of the inner product $\langle x,y\rangle$. Equality holds if and only if x and y are linearly dependent, i.e. when $x = ay$ for some scalar a then above inequality becomes

$$|\langle x,y\rangle| = \|x\|_{l_2}\|y\|_{l_2}.$$

This lemma will be used later in the proof of lemma (2.5) and lemma (3.2).

Lemma (2.2)**:** Let $A \subseteq B \subseteq \{1,2,\cdots,N\}$ be two sets, then for any vector $x \in \mathbb{R}^N$

$$\|x_A\|_{l_2} \leq \|x_B\|_{l_2}\ .$$

Proof: $\|x_A\|_{l_2} = (\sum_{i\in A}(x_i)^2)^{\frac{1}{2}} \leq (\sum_{i\in B}(x_i)^2)^{\frac{1}{2}} + (\sum_{i\in A-B}(x_i)^2)^{\frac{1}{2}} = \|x_B\|_{l_2}$.

Lemma (2.3) **[Triangle Inequality]:**

For any n-vectors $x \in \mathbb{R}^N$ and any norm $\|\cdot\|$,

$$\|x_1 + x_2 + \cdots + x_n\| \leq \sum_{i=1}^{n} \|x_i\| .$$

***Proof*:** The proof follows by mathematical induction.

■

Next, we show that the cardinality of two disjoint sets is less than or equal to sum of the cardinality of the first set and second set minus the cardinality of first set intersect with the second one.

Lemma (2.4)**:** If A and B are sets, then

$$|\mathrm{A} \cup \mathrm{B}| = |A| + |B| - |\mathrm{A} \cap \mathrm{B}|.$$

Hence $|\mathrm{A} \cup \mathrm{B}| \leq |A| + |B|$ if and only if A and B are disjoint sets.

***Proof*:** Since the set of elements in both A and B are $A - B$ and $B - A$ respectively, then

$$\mathrm{A} \cup \mathrm{B} = (A - B) \cup (B - A) \cup (A \cap \mathrm{B}).$$

Equating the cardinalities of the two sides of above equation, we obtain

$$|\mathrm{A} \cup \mathrm{B}| = |A - B| + |B - A| + |A \cap \mathrm{B}|. \tag{2.3}$$

Now

$$|A - B| = |A| - |A \cap \mathrm{B}|. \tag{2.4}$$

Similarly, we have

$$|B - A| = |B| - |A \cap \mathrm{B}|. \tag{2.5}$$

By plugging equation (2.4) and (2.5) into the right hand-side of equation (2.3) we obtain:

$$|\mathrm{A} \cup \mathrm{B}| = |A| + |B| - |A \cap \mathrm{B}|.$$

■

The following lemma describes the relation between l_1-norm and l_2-norm of a k-sparse vector. This lemma will be used later in the proof of theorem (2.1).

Lemma (2.5) ([10], lemma 1.2)**:** Consider $x \in \Sigma_k$, where Σ_k is the set of all k-sparse vectors. Then

$$\frac{\|x\|_{l_1}}{\sqrt{k}} \leq \|x\|_{l_2} . \tag{2.6}$$

Proof: Let $\mathcal{A} = (x_1\ x_2\ ...\ x_N)$ and $\mathcal{B} = (sign(x_1),\ sign(x_2), ...\ ,\ sign(x_N))$ where

$sign(x)$:$\mathbb{R} \to \mathbb{R}$ is a function

$$sign(x) = \begin{cases} 1 \quad if \ \ x > 0 \\ -1 \ \ if \ \ x < 0 \\ 0 \quad if \ x = 0 \end{cases}$$

therefore

$$\sum_{i \in \mathrm{N}} |sign(\ x_{\mathrm{i}})|^2 \leq k\ .$$

Taking the inner product of $\mathcal{A}$ and $\mathcal{B}$ gives

$$|\langle \mathcal{A}, \mathcal{B} \rangle| = |x_1 sign(x_1)| + |x_2 sign(x_2)| + \ldots + |x_N sign(x_N)|.$$

Because $|x\ sign(x)| = |x|$, then we can write the above expression as

$$|\langle \mathcal{A}, \mathcal{B} \rangle| \ = |x_1| + |x_2| + ... + |x_N|.$$

But $|x_1| + |x_2| + ... + |x_N| = \sum_{i=1}^{N} |x_i|$, therefore

$$|\langle \mathcal{A}, \mathcal{B} \rangle| = \sum_{i=1}^{N} |x_i|\ .$$

By the definition of l_1-norm, $|\langle \mathcal{A}, \mathcal{B} \rangle| = \|x\|_{l_1}$ and apply lemma (2.1) on $|\langle x, y \rangle|$ to obtain

$$\|x\|_{l_1} = |\langle \mathcal{A}, \mathcal{B} \rangle| \leq \|\mathcal{A}\|_{l_2}\ \|\mathcal{B}\|_{l_2} \leq \|x\|_{l_2}\ \|sign(x)\|_{l_2}.$$

Since $\|sign(x)\|_{l_2} = \sqrt{\sum_{i \in N} |sign(x_i)|^2} = \sqrt{\sum_{i \in N} (1)} = \sqrt{|N|} \leq \sqrt{k}$, then we obtain

$$\|x\|_{l_1} \leq \sqrt{k}\ \|x\|_{l_2}$$

and this implies

$$\frac{\|x\|_{l_1}}{\sqrt{k}} \leq \|x\|_{l_2}.$$

■

Next we show another key lemma which we need in the proof of theorem (2.1). Roughly speaking, the next lemma will describe the relation between l_1-norm and l_2-norm in terms of the largest index set.

Lemma (2.6)**:** Let $x \in \mathbb{R}^N$ be a vector and let Λ be a set consisting of the indices of the entries with the $2k$ largest modului. Let Υ be a subsets of $\{1,2,\ldots,N\}$ such that $|\Lambda| = 2k$ and $|\Upsilon| \leq 2k$. Then

$$\|x_\Lambda\|_{l_1} \geq \|x_\Upsilon\|_{l_1} \tag{2.7}$$

$$\|x_\Lambda\|_{l_2} \geq \|x_\Upsilon\|_{l_2} \tag{2.8}$$

$$\|x_{\Lambda^c}\|_{l_2} \leq \|x_{\Upsilon^c}\|_{l_2} \tag{2.9}$$

Proof: Consider the definition (1.3) (i.e.l_1-norm)

$$\|x_\Lambda\|_{l_1} = \sum_{i\in\Lambda}|x_i| \tag{2.10}$$

$$\|x_\Upsilon\|_{l_1} = \sum_{i\in\Upsilon}|x_i| \tag{2.11}$$

By Comparing (2.10) and (2.11), we can say that the largest magnitude in expression (2.10) is greater than the largest magnitude in (2.11) by the assumption that x_Λ has the largest modulus. Also, by the definition of l_2-norm

$$\|x_\Lambda\|_{l_2} = \sqrt{\sum_{i\in\Lambda} {x_i}^2} \tag{2.12}$$

$$\|x_\Upsilon\|_{l_2} = \sqrt{\sum_{i\in\Upsilon} {x_i}^2} \tag{2.13}$$

By comparing (2.12) and (2.13), we conclude that the largest element in $\sum_{i\in\Lambda} {x_i}^2$ is greater than or equal to the largest element in $\sum_{i\in\Upsilon} {x_i}^2$, and the second largest element in $\sum_{i\in\Lambda} {x_i}^2$ is greater than the second largest element in $\sum_{i\in\Upsilon} {x_i}^2$ and so on for the rest of elements in both (2.12) and (2.13). The reason behind this is because the assumption x_Λ with index set Λ that has the $2k$ largest modulus assures that elements in $\sum_{i\in\Lambda} {x_i}^2$ are the largest elements. Moreover, taking the complement of Λ will produce the $2k$ smallest number; i.e. $|\Lambda|^c$ will have the smallest size. In contrast, the complement of $|\Upsilon|$ is the set $|\Upsilon|^c > 2k$, i.e $|\Upsilon|^c$ is greater than the $2k$ largest element. Therefore, $\|x_{\Lambda^c}\|_{l_2} \leq \|x_{\Upsilon^c}\|_{l_2}$.

■

Next we state a general result by lemma (2.7), whereby if we take l_p-norm of a vector (for instance $x \in \mathbb{R}^N$) in correspond to two disjoint sets, then l_p-norm of x with respect to union of the two disjoint sets will be the same as summing l_p-norm of x according to the first index set and l_p-norm of x with respect to the second index set. Before stating lemma (2.7), we need to define l_p-norm.

Definition (2.2) **[l_p-norm]**: Let $x \in \mathbb{R}^N$ be a vector, then l_p-norm for the vector x is

$$\|x\|_{l_p} = \left(\sum_{i=1}^{N} |x_i|^p\right)^{\frac{1}{p}}$$

Lemma (2.7)**:** Let A and B be two subsets of $\{1,2,\dots,N\}$ and $A \cap B = \emptyset$. Define x to be a vector in $\mathbb{R}^N$ then

$$\|x_{A\cup B}\|_{l_p}^p = \|x_A\|_{l_p}^p + \|x_B\|_{l_p}^p.$$

Proof: By the definition of l_p norm

$$\|x\|_{l_p}^p = \sum_{i=1}^{N} |x_i|^p \ .$$

Because the union of A and B is the set $\{1,2,\dots,N\}$, then we can write the l_p- norm of a vector $x_{A\cup B}$ as

$$\|x_{A\cup B}\|_{l_p} = (\sum_{i=1}^{N} |x_i|^p)^{\frac{1}{p}} \ .$$

Since

$$(\sum_{i=1}^{N} |x_i|^p)^{\frac{1}{p}} = (\ |x_1|^p)^{\frac{1}{p}} + (\ |x_2|^p)^{\frac{1}{p}} + \cdots + (\ |x_N|^p)^{\frac{1}{p}} \ .$$

Then we can re-write $\|x_{A\cup B}\|_{l_p}^p$ as

$$\|x_{A\cup B}\|_{l_p}^p = |x_1|^p + |x_2|^p + \cdots + |x_N|^p \ . \qquad (2.14)$$

Now, we will count all the elements in the set A, and we also count all elements in the set B without counting the repeated elements in both A and B. Now, we can write equation (2.14) as follow

$$\|x_{A\cup B}\|_{l_p}^p = \sum_{i\in A} |x_i|^p + \sum_{i\in B} |x_i|^p .$$

By the definition of l_p-norm: $\sum_{i\in A}|x_i|^p = \|x_A\|_{l_p}^p$ and $\sum_{i\in B}|x_i|^p = \|h_B\|_{l_p}^p$

Hence

$$\|x_{A\cup B}\|_{l_p}^p = \|x_A\|_{l_p}^p + \|x_B\|_{l_p}^p.$$

■

Next, we state yet another lemma which we need in order to show the relation between RIP and NSP.

Lemma (2.8) ([10], lemma 1.3)**:** Suppose that $\Phi \in \mathbb{R}^{M\times N}$ satisfies the RIP of order $2k$, and let $x \in \mathbb{R}^N$, $x \neq 0$ be arbitrary. Let $\Lambda_0 \subseteq \{1,2,\dots,N\}$ such that $|\Lambda_0| \leq k$. Define Λ_1 to be the set corresponding to the k entries of $x_{\Lambda_0^c}$ with largest magnitude, and set $\Lambda = \Lambda_0 \cup \Lambda_1$. Then

$$\|x_\Lambda\|_{l_2} \leq \alpha \frac{\|x_{\Lambda_0^c}\|_{l_1}}{\sqrt{k}} + \beta \frac{|\langle \Phi x_\Lambda, \Phi x\rangle|}{\|x_\Lambda\|_{l_2}}$$

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

Note that lemma (2.8) holds for arbitrary x. In order to start stating the theorem which illustrates that RIP implies NSP, we need to apply lemma (2.8) to the case where $x \in \mathcal{N}(\Phi)$. The proof of this lemma is a technical proof which can be found in [10]. Keep in mind that previously we showed a series of lemmas that give us the ability to construct the proof of the relation between RIP and NSP, given by theorem (2.1).

Theorem (2.1) ([10] Devanport et al., Theorem 1.5)**:** Suppose that $\Phi \in \mathbb{R}^{M\times N}$ is a matrix satisfying the RIP of order $2k$ with the restricted isometry constant (RIC) $\delta_{2k} < \sqrt{2} - 1$. Then Φ satisfies the NSP of order $2k$ with constant

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

Proof : Suppose $x \in \mathcal{N}(\Phi)$. It is sufficient to show that

$$\|x_\Lambda\|_{l_2} \le C\,\frac{\|x_{\Lambda^c}\|_{l_1}}{\sqrt{k}} \qquad (2.15)$$

holds for the case where Λ is the index set corresponding to the $2k$ largest entry of x , i.e. $|\Lambda| = 2k$. (2.15) implies the NSP definition, given by definition (2.1), simply by multiplying both sides of (2.15) by $\sqrt{k}$ then apply lemma (2.5) to the left hand side of (2.15) to get

$$\|x_\Lambda\|_{l_1} \le \sqrt{k}\|x_\Lambda\|_{l_2} \le C\|x_{\Lambda^c}\|_{l_1}.$$

Hence clearly we get (2.15). Another way of expressing (2.15) is that, if $\|x_\Lambda\|_{l_2} \le C\frac{\|x_{\Lambda^c}\|_{l_1}}{\sqrt{k}}$ and take $L \subseteq \{1,\dots,N\}$ such that $|L| \le 2k$, then by applying lemma (2.5) we obtain

$$\|x_{\mathrm{L}}\|_{l_2} \le \|x_\Lambda\|_{l_2} \le C\,\frac{\|x_{\Lambda^c}\|_{l_1}}{\sqrt{k}} \le C\,\frac{\|x_{\mathrm{L}^c}\|_{l_1}}{\sqrt{k}}\,.$$

Hence, clearly

$$\|x_{\mathrm{L}}\|_{l_2} \le C\,\frac{\|x_{\mathrm{L}^c}\|_{l_1}}{\sqrt{k}}\,.$$

Next, take Λ_0 to be the index set corresponding to the k largest entry of x.

We can apply lemma (2.8) on x as follow

$$\|x_\Lambda\|_{l_2} \le \alpha\,\frac{\left\|x_{\Lambda_0^c}\right\|_{l_1}}{\sqrt{k}} + \beta\,\frac{|\langle \Phi x_\Lambda, \Phi x\rangle|}{\|x_\Lambda\|_{l_2}} \qquad (2.16)$$

Since $x \in \mathcal{N}(\Phi)$, then $\Phi x = 0$. Therefore, $\langle \Phi x_\Lambda, \Phi x\rangle = 0$ and this implies

$$\beta\,\frac{|\langle \Phi x_\Lambda, \Phi x\rangle|}{\|x_\Lambda\|_{l_2}} = 0\,.$$

Therefore equation (2.16) becomes

$$\|x_{\Lambda}\|_{l_2} \leq \alpha \frac{\|x_{\Lambda_0^c}\|_{l_1}}{\sqrt{k}} . \tag{2.17}$$

Now, $\Lambda_1 \subseteq \Lambda_0^c$ and $\Lambda = \Lambda_0 \cup \Lambda_1$ (by lemma (2.8)).

Next, in order to be able to use lemma (2.7) on $x_{\Lambda_0^c}$, we need to show that $\Lambda_0^c = \Lambda_1 \cup \Lambda^c$ and $\Lambda_1 \cap \Lambda^c = \emptyset$. Since $\Lambda_1 \cup {\Lambda_1}^c = \{1,2,\dots,N\}$ and $\Lambda_0^c \cup \Lambda_1 = \Lambda_0^c$, then the intersection between the universal set $\{1,2,\dots,N\}$ and the set $\Lambda_0^c \cup \Lambda_1$ will produce the set Λ_0^c , i.e

$$(\Lambda_1 \cup {\Lambda_1}^c) \cap (\Lambda_0^c \cup \Lambda_1) = \Lambda_0^c .$$

Now, $\Lambda_0^c = (\Lambda_1 \cup {\Lambda_1}^c) \cap (\Lambda_1 \cup \Lambda_0^c)$ and by applying the distributive law, we obtain

$$\Lambda_0^c = (\Lambda_1 \cap \Lambda_1) \cup ({\Lambda_1}^c \cap \Lambda_0^c) .$$

Since $\Lambda_1 \cup \Lambda_1 = \Lambda_1$, then we can rewrite Λ_0^c as follow

$$\Lambda_0^c = \Lambda_1 \cup (\Lambda_0^c \cap {\Lambda_1}^c) .$$

But $\Lambda^c = \Lambda_0^c \cap {\Lambda_1}^c$, then $\Lambda_0^c = \Lambda_1 \cup \Lambda^c$ hence

$$\Lambda_0^c = \Lambda_1 \cup \Lambda^c . \tag{2.18}$$

Also $\Lambda_1 \cap \Lambda^c = \Lambda_1 \cap (\Lambda_0^c \cap {\Lambda_1}^c)$ (because $\Lambda^c = \Lambda_0^c \cap {\Lambda_1}^c$). Next, by applying the distributive law, we get

$$\Lambda_1 \cap \Lambda^c = (\Lambda_1 \cap \Lambda_0^c) \cap (\Lambda_1 \cap {\Lambda_1}^c).$$

Since $\Lambda_1 \subseteq \Lambda_0^c$ then $(\Lambda_1 \cap \Lambda_0^c) = \Lambda_1$ this implies that

$$\Lambda_1 \cap \Lambda^c = \Lambda_1 \cap \emptyset = \emptyset.$$

Therefore,

$$\Lambda_1 \cap \Lambda^c = \emptyset. \tag{2.19}$$

Now, equation (2.18) and equation (2.19) enable us to use lemma (2.7) on x, which implies

$$\|x_{\Lambda_0^c}\|_{l_1} = \|x_{\Lambda_1}\|_{l_1} + \|x_{\Lambda^c}\|_{l_1} \tag{2.20}$$

which is not equal to zero in general, i.e. Λ_1 is the set corresponding to the k largest entry of Λ_0^{c} , this means that Λ^{c} is a non empty set correspond to the other k largest entry of Λ_0^{c} . Next, apply lemma (2.5) on $\left\|x_{\Lambda_1}\right\|_{l_1}$ in equation (2.20) and we get

$$\left\|x_{\Lambda_0^{\mathrm{c}}}\right\|_{l_1} \leq \sqrt{k}\, \left\|x_{\Lambda_1}\right\|_{l_2} + \|x_{\Lambda^{\mathrm{c}}}\|_{l_1} . \tag{2.21}$$

Substituting equation (2.21) in equation (2.17) we obtain

$$\|x_\Lambda\|_{l_2} \leq \alpha\, \frac{\sqrt{k}\, \left\|x_{\Lambda_1}\right\|_{l_2} + \|x_{\Lambda^{\mathrm{c}}}\|_{l_1}}{\sqrt{k}} .$$

This implies

$$\|x_\Lambda\|_{l_2} \leq \alpha \left\|x_{\Lambda_1}\right\|_{l_2} + \alpha \frac{\|x_{\Lambda^{\mathrm{c}}}\|_{l_1}}{\sqrt{k}} .$$

Next, we subtract $\alpha \left\|x_{\Lambda_1}\right\|_{l_2}$ from both sides

$$\|x_\Lambda\|_{l_2} - \alpha \left\|x_{\Lambda_1}\right\|_{l_2} \leq \alpha \frac{\|x_{\Lambda^{\mathrm{c}}}\|_{l_1}}{\sqrt{k}} .$$

Then by using the fact that $\Lambda = \Lambda_0 \cup \Lambda_1$ we obtain

$$(1-\alpha)\|x_\Lambda\|_{l_2} \leq \alpha \frac{\|x_{\Lambda^{\mathrm{c}}}\|_{l_1}}{\sqrt{k}} .$$

The assumption $\delta_{2k} < \sqrt{2} - 1$ assures that $\alpha < 1$, and thus we may divide by $(1-\alpha)$ without changing the direction of the inequality

$$\|x_\Lambda\|_{l_2} \leq \frac{\alpha}{(1-\alpha)} \frac{\|x_{\Lambda^{\mathrm{c}}}\|_{l_1}}{\sqrt{k}} .$$

Hence

$$\|x_\Lambda\|_{\mathrm{l}_2} \leq \mathrm{C} \frac{\|x_{\Lambda^{\mathrm{c}}}\|_{l_1}}{\sqrt{k}}$$

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

■

To sum up, RIP does imply the NSP but the converse is not true. Actually, RIP is significantly more restrictive.

2.2 Computing restricted isometry property (RIP)

The restricted isometry property definition, given by definition (1.5), requires that every subset of the columns of Φ with certain cardinality, should behaves approximately like an orthonormal system. In this chapter we present a known technique for which the restricted isometry property can be computed for any matrix. Likewise, we proposed a method for checking restricted isometry property of underdetermined system of equations.

Computing the restricted isometry property (RIP) of a matrix is about computing the value of Φx in system (1.3) and then taking the Euclidean norm of it. Basically, if a matrix Φ is a square matrix, then computing RIP of Φ is equal with finding the eigenvalues of Φ. Note that RIP of a square matrix $\Phi \in \mathbb{R}^{N\times N}$ is

$$(1-\delta_k)\|x\|_{l_2} \leq \|\Phi x\|_{l_2} \leq (1+\delta_k)\|x\|_{l_2} .$$

For an eigenvector x, $\Phi x = \lambda x$ where λ is the corresponding eigenvalue. Therefore, the above inequality becomes

$$(1-\delta_k)\|x\|_{l_2} \leq \|\lambda x\|_{l_2} \leq (1+\delta_k)\|x\|_{l_2}.$$

Computing eigenvalues does not make any sense other than square matrices. Thus, to determine the RIP of non-square matrices, we need to know what is called as **Singular Value Decomposition (SVD).** For non-square matrices we are interested in the case where the number of rows is less than the number of columns. But before discussing the SVD of matrices, we need another definition known as *singular value*.

Definition (2.3) **[Singular value]:**

Let $\Phi \in \mathbb{R}^{M\times N}$ matrix, then the singular values $\sigma_1, \sigma_2, \dots, \sigma_N$ of Φ are the square roots of the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_N$ of $\Phi^T\Phi$, arranged in a decreasing order, i.e.

$$\sigma_i = \sqrt{\lambda_i} \text{ for } 1 \leq i \leq N.$$

Next, we define the singular value decomposition of any $M \times N$ matrix by using definition (2.3). Then we propose another method for computing RIP that needs less computation in comparison with singular value decomposition method which may have computational implications.

Definition (2.4) **[Singular value decomposition (SVD)]:**

Let Φ be an $M \times N$ matrix with rank r. Then there exist an $M \times N$ matrix $\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$,where

D is the diagonal matrix whose diagonal entries are the first r singular values of Φ, and there exist an $M \times M$ orthogonal matrix U and an $N \times N$ orthogonal matrix V such that

$$\Phi = U \Sigma V^T \tag{2.22}$$

where $U = [u_1 \quad \cdots \quad u_M]$ such that $u_i = \frac{1}{\sigma_i} \Phi v_i$ for $i = 1, \dots, N$ and $V = [v_1 \quad \cdots \quad v_N]$ are orthonormal basis of $\mathbb{R}^N$ for eigenvectors of $\Phi^T \Phi$, and V^T is the transpose of V.

To illustrate SVD in more detail, we select an underdetermined matrix which we are interested in and compute its Singular value decomposition. Roughly speaking, we can find SVD of any matrix through the following 5-step algorithm:

Step (1): Compute $\mathrm{M} = \Phi^{\mathrm{T}} \Phi$, where Φ is an input matrix.

Step (2): Compute the Eigenvalue and Eigenvectors of M.

Step (3): Determine the singular values of M and then sort them in a decreasing order.

Step (4): Compute U , V and Σ where U and V are orthogonal $M \times M$ and $N \times N$ matrices respectively, such that

$$\mathrm{U} = [\mathrm{u}_1 \quad \mathrm{u}_2 \quad \cdots \quad \mathrm{u}_M] \text{ and } \mathrm{V} = [\mathrm{v}_1 \quad \mathrm{v}_2 \quad \cdots \quad \mathrm{v}_N]$$

where $\mathrm{u_i} = \frac{1}{\sigma_i} \Phi \mathrm{v}_i$ for $1 \le i \le N$ and $\sigma_i \neq 0$, $\mathrm{v_i}$ is the unit eigenvector and σ_i is the singular value.

Step (5): Computing the matrix $\Phi = \mathrm{U} \Sigma \mathrm{V}^{\mathrm{T}}$ where Σ is the diagonal $M \times N$ matrix whose entries are the first singular values of M.

Example: If $\mathrm{A} = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$ then the SVD of A is

$$\mathrm{A} = \mathrm{U\Sigma V^T} = \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \begin{bmatrix} \frac{-2}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{\sqrt{18}} & \frac{-1}{\sqrt{18}} & \frac{4}{\sqrt{18}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}.$$

Computing the RIP of underdetermined system is about computing the Euclidean norm of ($\mathrm{U\Sigma V^T}x$) in (1.3) because by substituting (2.22) back in (1.3) we end up with the following inequality

$$(1-\delta_k)\|x\|_{l_2} \leq \|\mathrm{U\Sigma V^T}x\|_{l_2} \leq (1+\delta_k)\|x\|_{l_2}\ . \qquad (2.23)$$

Finally, we conclude a proposition (given by proposition (2.1)), whereby we can compute SVD of underdetermined matrices in a better way in comparison with (2.23). Next, we give the proof of proposition (2.1).

Proof (of Proposition 2.1): Let $\Phi \in \mathbb{R}^{M\times N}$ be a given RIP matrix of order k with $\delta_k \in (0,1)$ such that

$$(1-\delta_k)\|x\|_{l_2} \leq \|\Phi x\|_{l_2} \leq (1+\delta_k)\|x\|_{l_2}. \qquad (2.24)$$

If $\Phi = \mathrm{U\Sigma V^T}$ is the SVD decomposition of Φ then

$$\|\Phi x\|_{l_2} = \|\mathrm{U\Sigma V^T}x\|_{l_2}.$$

Since U, V are orthogonal matrices, then they are preserving norms such that

$$\|Ux\|_{l_2} = \sqrt{(\mathrm{Ux})^2}$$

for some vector $x \in \mathbb{R}^N$, and we are now calculating $(\mathrm{Ux})^2$ whereby we can rewrite above formula as

$$(\mathrm{U}x \cdot \mathrm{U}x)^{1/2}$$

and the last formula can be written as

$$(x^T\ \mathrm{U^T U}x)^{1/2}\ .$$

Since U is orthogonal (i.e. $\mathrm{U^T U} = I$ where I is the identity matrix), we obtain

$$(x\,x)^{1/2} = \|x\|_{l_2} \ .$$

Thus

$$\|\mathrm{U}x\|_{l_2} = \|x\|_{l_2} \tag{2.25}$$

and similarly

$$\|\mathrm{V}^{\mathrm{T}}x\|_{l_2} = \|x\|_{l_2} \ . \tag{2.26}$$

By using (2.25) and (2.26) we get

$$\|\Phi x\|_{l_2} = \|\mathrm{U}\Sigma\mathrm{V}^{\mathrm{T}}x\|_{l_2} = \|\mathrm{U}\Sigma x\|_{l_2} = \|\Sigma x\|_{l_2}.$$

As we previously mentioned that U, V are orthogonal matrices, therefore they are both preserving norm. Hence, (2.24) becomes

$$(1 - \delta_{\mathrm{k}})\|x\|_{l_2} \leq \|\Sigma x\|_{l_2} \leq (1 + \delta_{\mathrm{k}})\|x\|_{l_2} \ .$$

■

Next we are going to discuss the geometrical behaviour of RIP and uniqueness of the sparse solution of system (1.1).

2.3 Geometrical behaviour of RIP

The shape of the set of matrices that has RIP of order 1 is a result of multiplying an annulus by itself *N*-times by the following theorem.

Theorem (2.3)**:** Let $M_{\delta,1}$ be a set of all matrices which have the RIP of order 1 and a restricted isometry constant (RIC) δ_k,then

$$M_{\delta,1} = \underbrace{\mathfrak{D}(\delta_1) \times \mathfrak{D}(\delta_1) \times \ldots \times \mathfrak{D}(\delta_1)}_{N-times}$$

where $\mathfrak{D}(\delta_1)$ is the annulus $\{x \in \mathbb{R}^N \,|\, (1 - \delta_k) \leq \|x\|_{l_2} \leq (1 + \delta_k)\}$. In other words, the matrix's columns are in $\underbrace{\mathfrak{D}(\delta_1) \times \mathfrak{D}(\delta_1) \times \ldots \times \mathfrak{D}(\delta_1)}_{N-times}$.

Proof: Since $M_{\delta,1}$ has RIP of order 1, then by definition of RIP $\exists \delta_k \in (0,1)$ such that

$$(1 - \delta_k)\|x\|_{l_2} \leq \|\Phi x\|_{l_2} \leq (1 + \delta_k)\|x\|_{l_2} \tag{2.27}$$

for all $x \in \Sigma_1 = \{x |\ \|x\|_{l_0} \leq 1\}$.

$$\text{Now, let } x_q = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \longrightarrow \text{q}^{\text{th}} \text{ entry}$$

In other words, x_q is a vector which has only one non-zero element in the q$^{\text{th}}$ column.

Therefore, $\|x_q\|_{l_2} = \sqrt{(1)^2} = 1$ and equation (2.27) becomes

$$(1 - \delta_k) \leq \|\Phi x_q\|_{l_2} \leq (1 + \delta_k).$$

It implies that

$$(1 - \delta_k) \leq \|q^{th} \text{ column of } \Phi\,\|_{l_2} \leq (1 + \delta_k).$$

Then there is a map between the matrix Φ and the annulus such that the first column in the matrix Φ maps the point in the first annulus and the second column in the matrix Φ maps the point in the second annulus and so on.

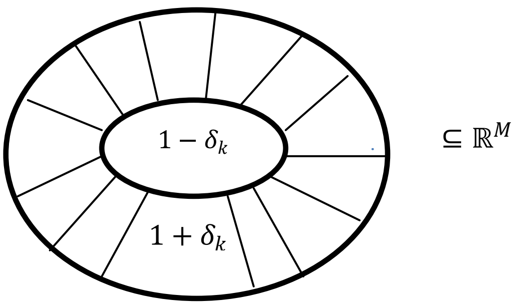

Figure 3: Annulus

Finally, by theorem (2.4) we show that if there is a sparse solution of the system (1.1) then it is a unique sparse solution but before that we need a lemma whereby we show that l_0-norm has the property of sub-linearity.

Lemma (2.9) **[sub-linearity of zero-norm]:** For any vectors x and y in $\mathbb{R}^N$

$$\|x + y\|_{l_0} \leq \|x\|_{l_0} + \|y\|_{l_0}.$$

Proof: We start by re-writing the definition of l_0-norm as follow:

$$\|x\|_{l_0} = \sum_{i=1}^{N} [x_i]$$

where $[\cdot]: \mathbb{R} \to \mathbb{R}$ is a function

$$[x] = \begin{cases} 1 & if\ x \neq 0 \\ 0 & if\ x = 0 \end{cases} .$$

Now, in the case where both x and y are zero it is obvious that

$$[x + y] = [x] + [y].$$

Whereas if one of them not zero, i.e. if $x \neq 0$ or $y \neq 0$ then

$$[x + y] \leq [x] + [y].$$

Taking the summation of the above inequality implies

$$\|x + y\|_{l_0} \leq \|x\|_{l_0} + \|y\|_{l_0}$$

■

Next, we proof the uniqueness of sparse solution of system (1.1) by the following theorem.

Theorem (2.4) [8]**:** Let Φ be an $M \times N$ matrix which has RIP of order $2k$ with the RIC $\delta_{2k} < 1$, then any k-sparse solution of system (1.1) is the unique such solution.

Proof: Suppose we have two k-sparse solutions of $y = \Phi x$, say x_1 and x_2 then

$\Phi x_1 = \Phi x_2 = y$ which implies $\Phi x_1 - \Phi x_2 = 0$ and $\Phi(x_1 - x_2) = 0$

Now, because x_1 and x_2 are both k-sparse solution of $y = \Phi x$, then one can apply lemma (2.9) as follow

$$\|x_1 + (-x_2)\|_{l_0} \leq \|x_1\|_{l_0} + (-\|x_2\|_{l_0}).$$

Then

$$\|x_1 + (-x_2)\|_{l_0} \leq \|x_1\|_{l_0} + (-\|x_2\|_{l_0}) \leq k + k = 2k$$

which means $(x_1 - x_2)$ is $2k$-sparse solution of $y = \Phi x$. Therefore, we can apply the restricted isometry property definition because we concluded that $x_1 - x_2$ is $2k$-sparse and we just take the lower bound of the RIP inequality because the lower bound is crucial, i.e.

$$(1 - \delta_{2k}) \|x_1 - x_2\|_{l_2} \leq \|\Phi(x_1 - x_2)\|_{l_2}. \tag{2.28}$$

Previously we assumed that $\delta_{2k} < 1$ which means that $(1 - \delta_{2k})$ is a positive quantity. Therefore, if we divide both sides of equation (2.28) by $(1 - \delta_{2k})$ then the sign of the inequality is not going to change, as follow

$$\frac{(1 - \delta_{2k}) \|x_1 - x_2\|_{l_2}}{(1 - \delta_{2k})} \leq \frac{\|\Phi(x_1 - x_2)\|_{l_2}}{(1 - \delta_{2k})}$$

and we obtain

$$\|x_1 - x_2\|_{l_2} \leq \frac{\|\Phi(x_1 - x_2)\|_{l_2}}{(1 - \delta_{2s})}.$$

Since $\Phi(x_1 - x_2) = 0$, then we get $\|x_1 - x_2\|_{l_2} \leq 0$

But, by the definition of norm $\|x_1 - x_2\|_{l_2} \geq 0$ iff $x_1 - x_2 = 0$ which implies

$x_1 - x_2 = 0$ iff $x_1 = x_2$.

■

Chapter 3

In this chapter we aim to construct matrices with the restricted isometry constant close to $\frac{1}{\sqrt{2}} \approx 0.7071$ where sparse recovery (l_1-minimization) fails to at least one k-sparse vector. The goal is to understand how much improvement is possible over the best known positive results which relate restricted isometry constant to sparse l_1 -recovery. We give an exposition of M. E. Davies and R. Gribonval's result in [11] but we specialise to l_1-recovery instead of l_p-recovery that was generally dealt with in [11]. Also we extend the proofs of lemmas and propositions that are given in [14]. In their paper, Davies and Gribonval showed that when RIC δ_{2k} close to $\frac{1}{\sqrt{2}}$ then there exist matrices such that l_1-recovery cannot recover all sparse vectors by the following theorem:

Theorem (3.1)[Davies and Gribonval [11]]: *For any $\varepsilon > 0$ there exist an integer k and a matrix Φ with RIC $\delta_{2k} \leq \frac{1}{\sqrt{2}} + \varepsilon$ for which l_1-recovery fails on some k-sparse vector.*

The main idea of the proof of theorem (3.1) is to first reduce the search for the failing matrices that we are interested in to what are called *minimally redundant row orthonormal* matrices. By minimally redundant matrices we mean, a matrix $\Phi \in \mathbb{R}^{\mathrm{M}\times\mathrm{N}}$ such that $\mathrm{M} = \mathrm{N} - 1$. A row orthonormal matrix means that the rows of the matrix are orthonormal. In order to find matrices for which the l_1-minimization fails to recover at least one k-sparse vector with small RIC δ_{2k}, we will be looking for l_1-failing matrices with largest possible asymmetric RIC σ_{2k}^2 (we give the definition of Asymmetric RIC σ_{2k}^2 later in this chapter, by definition 3.2). Then the l_1-failing matrices with largest σ_{2k}^2 can be searched within the restricted set of l_1-failing minimally redundant row orthonormal matrices. Furthermore, for the minimally redundant row orthonormal matrices, σ_{2k}^2 is completely determined by the unit vector z which spans the null space of the matrix Φ, i.e. $\mathcal{N}(\Phi)$. We then recast the problem of selecting l_1-failing minimally redundant row orthonormal matrices with maximal σ_{2k}^2 to an optimization problem where we wish to select a unit norm vector z that allows l_1-failing recovery for k-sparse vectors. This modification can be done by changing the signs and swapping appropriate columns of the matrix Φ. Finally we show how to construct the matrix Φ from the vector z.

Because of theorem (2.1), RIP is one of the commonly used frameworks for sparse recovery via l_1-minimization. E.Candes in [7] showed that if RIC $\delta_{2k} < \sqrt{2} - 1 = 0.4142$ then every k-sparse vector can be uniquely recovered via l_1-minimization. Later, this bound has been improved by Foucart and Lai in [13] to $\delta_{2k} < 0.4531$ and then to $\delta_{2k} < 0.472$ in [3]. Therefore, the main question that arises here is: how large we can set $\delta_{2k} \leq 1$ so that we can recover every k-sparse vector via l_1-recovery? By the end of this chapter we show that δ_{2k} cannot be bigger than 0.7071.

3 Matrices that fail to satisfy the null space property

In this chapter we are interested in sparse solutions to system (1.1), i.e.

$$y = \Phi x \tag{3.1}$$

where $\Phi \in \mathbb{R}^{\mathrm{M \times N}}$ such that $M < N$, and x_Ω denotes a vector that is equal to some x on some index set Ω and zero elsewhere. Furthermore, the vector x_Ω is $|\Omega|$-sparse and we say that the support of the vector x lies within Ω whenever $x_\Omega = x$. In chapter 2, we introduced the NSP, given by definition (2.3), whereby for any non-zero vector $z \in \mathcal{N}(\Phi)$ and a constant $C \in (0,1)$ we had

$$\|z_\Omega\|_{l_1} \leq C \|z_{\Omega^c}\|_{l_1}.$$

This implies that

$$\|\mathrm{z}_\Omega\|_{l_1} < \|\mathrm{z}_{\Omega^c}\|_{l_1} \tag{3.2}$$

is also true that any vector y whose support lies within Ω, can be uniquely recovered by the following optimization problem

$$x_{l_1}^* = \min_x \|x\|_{l_1} \text{ such that } \Phi x = \mathrm{y}. \tag{3.3}$$

Furthermore, this particular definition of NSP (i.e. (3.3)) is tight because the inequality (3.2) does not hold for some $\mathrm{z} \in \mathcal{N}(\Phi)$ then the vector $x := z_\Omega$ is supported on Ω but it is not the unique minimiser for (3.3), for instance we can take another vector to minimize (3.3) such as $\tilde{x} = -z_{\Omega^c}$. Then $\Phi(z_\Omega + z_{\Omega^c}) = \Phi \mathrm{z} = 0$ and $\Phi z_\Omega = -\Phi z_{\Omega^c} = \Phi \tilde{x}$. This is a property that we will call 'l_1 failure' from now on, i.e. when

$$\|z_\Omega\|_{l_1} \geq \|z_{\Omega^c}\|_{l_1}.$$

Actually, in the worst case scenario we have

$$\|\tilde{x}\|_{l_1} = \|z_{\Omega^c}\|_{l_1} < \|z_\Omega\|_{l_1} = \|x\|_{l_1}.$$

In this chapter we will work with a slightly stronger condition than the usual RIP, introduced in chapter 2, by considering the *Unit Spectral Norm Matrix* and later we define the asymmetric RIC.

Definition (3.1) **[Spectral norm]:**

Let $\Phi \in \mathbb{R}^{\mathrm{M}\times\mathrm{N}}$ be a matrix and let a vector $x \in \mathbb{R}^N$. Then the spectral norm of Φ is

$$|||\Phi||| := \sup_{x\neq 0} \frac{\|\Phi x\|_{l_2}}{\|x\|_{l_2}}.$$

Then *Unit spectral norm matrices* are matrices with

$$|||\Phi||| = 1\ .$$

Definition (3.2): [**Asymmetric RIC**]

Let $\Phi \in \mathbb{R}^{\mathrm{M}\times\mathrm{N}}$ be a unit spectral norm matrix. Then the asymmetric RIC $\sigma_k^2(\Phi)$ is defined by

$$\sigma_k^2(\Phi) := \min_{x_\Omega \neq 0, |\Omega| \leq \mathrm{k}} \frac{\|\Phi x_\Omega\|_{l_2}}{\|x_\Omega\|_{l_2}}$$

We begin by a lemma which shows how the asymmetric RIC relates to previously defined RIC definition when the matrix in the question is rescaled.

Lemma (3.1): Let $\Phi \in \mathbb{R}^{\mathrm{M}\times\mathrm{N}}$ be a unit spectral norm matrix and $\sigma_k^2(\Phi)$ be the asymmetric RIC. Define $\Psi_k := \sqrt{\frac{2}{1+\sigma_k^2(\Phi)}}\,\Phi$ as a re-scaled matrix of $\sigma_k^2(\Phi)$. Then the RIC of a re-scaled matrix Ψ_k , $\delta_k(\Psi_k)$ satisfies

$$\delta_k(\Psi_k) \leq \frac{1-\sigma_k^2(\Phi)}{1+\sigma_k^2(\Phi)}$$

Proof: We start by considering the definition of asymmetric RIC together with unit spectral norm matrix definition to obtain

$$\sigma_k^2(\Phi)\|x\|_{l_2}^2 \leq \|\Phi x\|_{l_2}^2 \leq \|x\|_{l_2}^2.$$

Multiplying above inequality by λ^2 we obtain

$$\lambda^2\sigma_k^2(\Phi)\|x\|_{l_2}^2 \leq \lambda^2\|\Phi x\|_{l_2}^2 \leq \lambda^2\|x\|_{l_2}^2 .$$

But the RIC is the smallest number δ_k such that

$$(1-\delta_k)\|x\|_{l_2}^2 \leq \|\Phi x\|_{l_2}^2 \leq (1+\delta_k)\|x\|_{l_2}^2$$

holds for every k-sparse vector y allowing rescaling. In order to find the RIC we need to find $\delta_k(\Psi_k)$ such that

$$\lambda^2\sigma_k^2(\Phi) = (1-\delta_k(\Psi_k)) \tag{3.4}$$

and

$$\lambda^2 = (1+\delta_k(\Psi_k)) . \tag{3.5}$$

After substituting $\lambda^2 = \frac{2}{1+\sigma_k^2}$ in (3.4) and (3.5) we get two simultaneous equations

$$\frac{2\sigma_k^2(\Phi)}{1+\sigma_k^2(\Phi)} = \left(1-\delta_k(\Psi_k)\right) \tag{3.6}$$

$$\frac{2}{1+\sigma_k^2(\Phi)} = \left(1+\delta_k(\Psi_k)\right). \tag{3.7}$$

Then from the above two equations we obtain the rescaled RIC $\delta_k(\Psi_k)$ as follow:

From equation (3.6)

$$\frac{2\sigma_k^2(\Phi)}{1+\sigma_k^2(\Phi)} = \left(1-\delta_k(\Psi_k)\right)$$

which implies

$$\delta_k(\Psi_k) = 1 - \frac{2\sigma_k^2(\Phi)}{1+\sigma_k^2(\Phi)} = \frac{1-\sigma_k^2(\Phi)}{1+\sigma_k^2(\Phi)}.$$

On the other hand, from equation (3.7)

$$\delta_k(\Psi_k) = \frac{2}{1+\sigma_k^2(\Phi)} - 1 = \frac{1-\sigma_k^2(\Phi)}{1+\sigma_k^2(\Phi)}.$$

Then the RIC $\delta_k(\Psi_k)$ is

$$\delta_k(\Psi_k) = \frac{1-\sigma_k^2(\Phi)}{1+\sigma_k^2(\Phi)} .$$

Hence the RIC is smallest constant so that

$$\delta_k(\Psi_k) \leq \frac{1-\sigma_k^2(\Phi)}{1+\sigma_k^2(\Phi)}.$$

■

Our goal is to construct matrices where l_1-recovery fails for at least one k-sparse vector and in order to find l_1-failing matrices with minimum RIC δ_{2k}, we will be searching for l_1-failing matrices with the largest σ_{2k}^2. We now reduce the search for l_1-failing matrices to minimally redundant unit spectral norm matrices by proposition (3.1).

The proposition below shows that l_1-failing unit spectral norm matrices with largest σ_{2k}^2 (k is an integer such that $2k < N$) can be searched within the restricted set of minimally redundant row orthonormal matrices where l_1-minimization fails to recover at least one sparse vector. Furthermore, proposition (3.1) shows that minimally redundant row orthonormal matrices are optimal among unit spectral norm matrices.

Proposition (3.1) [Davies and Gribonval [11], Proposition 1]: Let $\Phi \in \mathbb{R}^{\mathrm{M}\times\mathrm{N}}$ be an arbitrary unit spectral norm matrix which is l_1-failing for some k-sparse vector with $\mathrm{M} < N$, then there exists a minimally redundant row orthonormal (unit spectral norm) matrix $\Phi^* \in \mathbb{R}^{(\mathrm{N}-1)\times\mathrm{N}}$ which is l_1-failing for the same k-sparse vector such that for every m

$$\sigma_m^2(\Phi) \leq \sigma_m^2(\Phi^*).$$

Proof: We start by applying slightly a different form of the singular value decomposition on Φ whereby we obtain $\Phi = \mathrm{V}\Sigma\mathrm{U}^{\mathrm{T}}$ where $\mathrm{V} \in \mathbb{R}^{\mathrm{M}\times\mathrm{M}}$ and $\mathrm{U}^{\mathrm{T}} \in \mathbb{R}^{\mathrm{M}\times\mathrm{N}}$ are row orthonormal, and $\Sigma \in \mathbb{R}^{\mathrm{M}\times\mathrm{M}}$ is a diagonal matrix whose entries are singular values of Φ. Since Φ has a unit spectral norm then

$$|||\Phi||| = |||V\Sigma U^T||| = 1$$

but $V$and $U^T$are preserve norm because V and U^T are row orthonormal , i.e.

$\|Vx\|_{l_2} = \|x\|_{l_2}$ for all $x \in \mathbb{R}^{\mathrm{M}}$. Hence

$$|||\Phi||| = |||V\Sigma U^T||| = |||\Sigma||| = 1$$

and for any vector x and any index set Ω , we have

$$\|\Phi x_\Omega\|_{l_2}^2 = \|V\Sigma U^T x_\Omega\|_{l_2}^2 = \|\Sigma U^T x_\Omega\|_{l_2}^2 \leq \|U^T x_\Omega\|_{l_2}^2.$$

Therefore

$$\frac{\|\Phi x_\Omega\|_{l_2}^2}{\|x_\Omega\|_{l_2}^2} \leq \frac{\|U^T x_\Omega\|_{l_2}^2}{\|x_\Omega\|_{l_2}^2}.$$

Now, since Φ is l_1-failing for some k-sparse vector, then there exists some failing vector $\mathrm{z} \in \mathcal{N}(\Phi)$ and an index set Ω_{k} of size k such that

$$\left\|\mathrm{z}_{\Omega_k}\right\|_{l_1} \geq \left\|\mathrm{z}_{\Omega_k^{\mathrm{c}}}\right\|_{l_1} \tag{3.8}$$

Now let $\mathrm{W} \in \mathbb{R}^{\mathrm{N}\times(\mathrm{N}-\mathrm{M}-1)}$ be a matrix such that the rows of W forms an orthonormal basis of the orthogonal complement to $\{z, U\}$ (note that the rows of W forms an orthonormal basis to the orthogonal complement of rows of U).

Since $\mathrm{z} \in \mathcal{N}(\Phi)$ then $\Phi\mathrm{z} = V\Sigma U^T z = 0$. Also, since $V\Sigma \neq 0$ because $V\Sigma$ is an invertible matrix. Therefore, the only way that $\Phi\mathrm{z} = 0$ is that $U^T z = 0$, i.e. z is orthonormal to the rows of U. Therefore, the columns of $\{z, U, \mathrm{W}\}$ forms an orthonormal basis over $\mathbb{R}^{\mathrm{N}}$ whereby we can write any $x_\Omega \in \mathbb{R}^{\mathrm{N}}$as:

$$x_\Omega = az + Ub + \mathrm{Wc}$$

for some $a \in \mathbb{R}$, $b \in \mathbb{R}^{\mathrm{M}}$ and $\mathrm{c} \in \mathbb{R}^{\mathrm{N}-\mathrm{M}-1}$. Next, define the minimally redundant row orthonormal matrix $\Phi^* := [U, \mathrm{W}]^{\mathrm{T}} \in \mathbb{R}^{(\mathrm{N}-1)\times\mathrm{N}}$. Since U and W are row orthonormal and the spectral norm of both of them is one, then spectral norm of Φ^* is one too, i.e.

$$|||\Phi^*||| = 1$$

and

$$\|\Phi^* x_\Omega\|_{l_2} = \|[U, \mathrm{W}]^{\mathrm{T}} az\|_{l_2} + \|[U, \mathrm{W}]^{\mathrm{T}} Ub\|_{l_2} + \|[U, \mathrm{W}]^{\mathrm{T}} \mathrm{Wc}\|_{l_2}. \tag{3.9}$$

Moreover, for any x_Ω we have

$$\frac{\|\Phi x_\Omega\|_{l_2}^2}{\|x_\Omega\|_{l_2}^2} \leq \frac{\|U^{\mathrm{T}} x_\Omega\|_{l_2}^2}{\|x_\Omega\|_{l_2}^2}. \tag{3.10}$$

So we compute the right hand side of the above inequality:

$$\|U^{\mathrm{T}} x_\Omega\|_{l_2}^2 = \|U^{\mathrm{T}} az + U^{\mathrm{T}} Ub + U^{\mathrm{T}} \mathrm{Wc}\|_{l_2}^2.$$

We previously showed that $U^T z$ vanishes, and $U^T U = I_{M\times M}$. Also, since every row of W is orthogonal to $\{z, u_1, \dots, u_2\}$ (by orthogonal complement property) then $\langle \mathrm{W}_i, z\rangle = \mathrm{W}_i^T z = 0$. Similarly, $\langle \mathrm{W}_i, U\rangle = \mathrm{W}_i^T U = 0$. Then

$$\|U^{\mathrm{T}} x_\Omega\|_2^2 = \|U^{\mathrm{T}} az + U^{\mathrm{T}} Ub + U^{\mathrm{T}} \mathrm{Wc}\|_2^2 = \|0 + Ib + 0\|_2^2 = \|b\|_2^2$$

Now we compute the denominator of (4.10), i.e.

$$\|x_\Omega\|_{l_2}^2 = \|az\|_{l_2}^2 + \|Ub\|_{l_2}^2 + \|\mathrm{Wc}\|_{l_2}^2 .$$

Since z is orthogonal vector, then $\|az\|_{l_2}^2 = a^2$ and U is row orthonormal such that $\|Ub\|_{l_2}^2 = \|b\|_{l_2}^2$ where $b \in \mathbb{R}^{\mathrm{M}}$. Also, since W is row orthonormal then $\|\mathrm{Wc}\|_{l_2}^2 = \|\mathrm{c}\|_{l_2}^2$. Thus we can write the value of $\|x_\Omega\|_{l_2}^2$ as

$$\|x_\Omega\|_{l_2}^2 = a^2 + \|b\|_{l_2}^2 + \|\mathrm{c}\|_{l_2}^2.$$

Then the inequality (3.10) becomes

$$\frac{\|\Phi x_\Omega\|_{l_2}^2}{\|x_\Omega\|_{l_2}^2} \leq \frac{\|U^{\mathrm{T}} x_\Omega\|_{l_2}^2}{\|x_\Omega\|_{l_2}^2} = \frac{\|b\|_{l_2}^2}{a^2 + \|b\|_{l_2}^2 + \|\mathrm{c}\|_{l_2}^2}$$

By adding $\|\mathrm{c}\|_{l_2}^2$ to the right hand side numerator in the above inequality we get

$$\frac{\|\Phi x_\Omega\|_{l_2}^2}{\|x_\Omega\|_{l_2}^2} \leq \frac{\|U^{\mathrm{T}} x_\Omega\|_{l_2}^2}{\|x_\Omega\|_{l_2}^2} = \frac{\|b\|_{l_2}^2}{a^2 + \|b\|_{l_2}^2 + \|\mathrm{c}\|_{l_2}^2} \leq \frac{\|b\|_{l_2}^2 + \|\mathrm{c}\|_{l_2}^2}{a^2 + \|b\|_{l_2}^2 + \|\mathrm{c}\|_{l_2}^2}$$

Since $\mathrm{W}^{\mathrm{T}}\mathrm{W} = I_{(N-M-1)\times(N-M-1)}$, then (3.9) becomes

$$\|\Phi^* x_\Omega\|_{l_2}^2 = \|b\|_{l_2}^2 + \|\mathrm{c}\|_{l_2}^2$$

then we obtain

$$\frac{\|\Phi x_\Omega\|_{l_2}^2}{\|x_\Omega\|_{l_2}^2} \le \frac{\|U^{\mathrm{T}} x_\Omega\|_{l_2}^2}{\|x_\Omega\|_{l_2}^2} = \frac{\|b\|_{l_2}^2}{a^2 + \|b\|_{l_2}^2 + \|\mathrm{c}\|_{l_2}^2} \le \frac{\|b\|_{l_2}^2 + \|\mathrm{c}\|_{l_2}^2}{a^2 + \|b\|_{l_2}^2 + \|\mathrm{c}\|_{l_2}^2} = \frac{\|\Phi^* x_\Omega\|_{l_2}^2}{\|x_\Omega\|_{l_2}^2}$$

Hence

$$\sigma_m^2(\Phi) \le \sigma_m^2(\Phi^*)$$

Note that we concluded that $\Phi^* z = 0$, hence z is in the null space of Φ^* which is l_1-failing for at least one k-sparse vector.

■

Next, by the proposition below we show that for the minimally redundant row orthonormal matrices, the asymmetric RIC $\sigma_k^2(\Phi)$ is completely determined by the unit vector z which spans the null space $\mathcal{N}(\Phi)$.

Proposition (3.2) [Davies and Gribonval [11], Proposition 2]:

Let $\Phi \in \mathbb{R}^{(\mathrm{N}-1)\times \mathrm{N}}$ be a minimally redundant row orthonormal matrix, and let $z \in \mathbb{R}^{\mathrm{N}}$ with $\|\mathrm{z}\|_{l_2} = 1$ be a vector which spans $\mathcal{N}(\Phi)$. Denoting Ω_m the set indexing the m largest components of z we have for every m

$$\sigma_m^2(\Phi) = 1 - \left\|\mathrm{z}_{\Omega_m}\right\|_{l_2}^2$$

Proof: Since z is in the null space of Φ then $\Phi z = 0$ and Φ is row orthonormal whereby the columns of $\{z, \Phi^{\mathrm{T}}\}$ forms an orthonormal basis in $\mathbb{R}^{\mathrm{N}}$. Therefore, we can write any vector x as :

$$x = az + \Phi^{\mathrm{T}} b \tag{3.11}$$

where $a \in \mathbb{R}$ and $b \in \mathbb{R}^{(\mathrm{N}-1)}$ and $\Phi x = \Phi z a + \Phi\Phi^{\mathrm{T}} b$ but $\Phi z = 0$ and $\Phi\Phi^{\mathrm{T}} = I$ then

$$\Phi x = b$$

and

$$\|\Phi x\|_{l_2}^2 = \|b\|_{l_2}^2. \tag{3.12}$$

If x has a unit norm then

$$1 = \|x\|_{l_2}^2 = a^2 + \|b\|_{l_2}^2. \tag{3.13}$$

But

$$a^2 = |a|^2 = |a\, z^{\mathrm{T}}\mathrm{z} + b^T \Phi z|^2 = |(a\, z^{\mathrm{T}} + b^T \Phi)\mathrm{z}|^2 = |x^{\mathrm{T}}\mathrm{z}|^2 = |\mathrm{z}x|^2 = |\langle z, x\rangle|^2$$

where x has the form of (3.11). By using (3.12) together with the value of a^2 we can write (3.13) as

$$1 = \|x\|_{l_2}^2 = a^2 + \|b\|_{l_2}^2 = |\langle z, x\rangle|^2 + \|\Phi y\|_{l_2}^2$$

hence

$$\|\Phi x\|_{l_2}^2 = 1 - |\langle z, x\rangle|^2. \tag{3.14}$$

Finally, to find the asymmetric RIC of Φ_Ω, we need to solve the problem

$$\sigma_m^2(\Phi) = \min_{\Omega, |\Omega| \le m} \min_{\substack{x_\Omega \\ \|x_\Omega\|=1}} \|\Phi x_\Omega\|_{l_2}^2 \ .$$

By using (3.14) we obtain

$$\min_{\Omega, |\Omega| \le m} \min_{\substack{x_\Omega \\ \|x_\Omega\|=1}} \|\Phi x_\Omega\|_{l_2}^2 = 1 - \max_{\Omega, |\Omega| \le m} \max_{\substack{x_\Omega \\ \|x_\Omega\|=1}} |\langle \mathrm{z}, x_\Omega\rangle|^2$$

then

$$\sigma_m^2(\Phi) = \quad 1 - \max_{\Omega, |\Omega| \le m} \max_{\substack{x_\Omega \\ \|x_\Omega\|=1}} |\langle z, x_\Omega\rangle|^2 \ .$$

In other words, we need to find the unit vector x_Ω^* that is maximally correlated with z. For a given Ω that satisfies with $x_\Omega^* = \frac{\mathrm{z}_\Omega}{\|\mathrm{z}_\Omega\|_{l_2}}$, whereby $|\langle \mathrm{z}, x_\Omega\rangle|^2 = \|\mathrm{z}_\Omega\|_{l_2}^2$. Hence, the best Ω is the one that captures the k largest components of z because asymmetric RIC $\sigma_k^2(\Phi)$ is the minimum of $\|\Phi x_\Omega\|_{l_2}^2$which obtains by subtracting the result of maximum inner product between z and x_Ω from one.

■

Next, we introduce a technical lemma which we need in order to explain the structure of optimal z in the null space.

Lemma (3.1) [Davies and Gribonval [11], lemma 2]: Let $\mathrm{u}_1 > \mathrm{v}_1 \geq \mathrm{v}_2 > \mathrm{u}_2 \geq 0$ such that $\mathrm{u}_1 + \mathrm{u}_2 = \mathrm{v}_1 + \mathrm{v}_2$. Then

$$\mathrm{u}_1{}^2 + \mathrm{u}_2{}^2 > \mathrm{v}_1{}^2 + \mathrm{v}_2{}^2.$$

Proof: Let $\mathrm{J}(\mathrm{u}_1, \mathrm{u}_2) = \mathrm{u}_1{}^2 + \mathrm{u}_2{}^2$, and we have $\mathrm{u}_1 + \mathrm{u}_2 = \mathrm{v}_1 + \mathrm{v}_2$ which implies

$$\mathrm{u}_2 = \mathrm{v}_1 + \mathrm{v}_2 - \mathrm{u}_1.$$

Substituting the last equation in J we get

$$\mathrm{J}\big(\mathrm{u}_1, \mathrm{u}_2(\mathrm{u}_1)\big) = \mathrm{u}_1{}^2 + (\mathrm{v}_1 + \mathrm{v}_2 - \mathrm{u}_1)^2.$$

This implies

$$\mathrm{J}\big(\mathrm{u}_1, \mathrm{u}_2(\mathrm{u}_1)\big) = 2(\mathrm{u}_1{}^2 - \mathrm{u}_1(\mathrm{v}_1 + \mathrm{v}_2) + \mathrm{v}_1\mathrm{v}_2) + (\mathrm{v}_1{}^2 + \mathrm{v}_2{}^2).$$

Hence

$$\mathrm{J}\big(\mathrm{u}_1, \mathrm{u}_2(\mathrm{u}_1)\big) = 2(\mathrm{u}_1 - \mathrm{v}_1)(\mathrm{u}_1 - \mathrm{v}_2) + (\mathrm{v}_1{}^2 + \mathrm{v}_2{}^2)$$

which is strictly positive by the assumption $\mathrm{u}_1 > \mathrm{v}_1 \geq \mathrm{v}_2 \geq 0$.

■

Keep in mind, our original problem which was to select a minimally redundant row orthonormal matrices with largest σ_{2k}^2. Then by proposition (3.2), minimally redundant row orthonormal matrices with largest σ_{2k}^2 is completely determined by a unit vector z which spans the null space of Φ. Characterizing the unit vector z with largest σ_{2k}^2 can be modified in to an optimization problem where we wish to determine the form of the optimal vector z.

Up to column permutation of Φ and sign changes we can assume that $z_i \geq z_{i+1} \geq 0$ and the l_1-failing assumption is that $\left\|\mathrm{z}_{\Omega_{2k}}\right\|_{l_1}^2 \geq \left\|\mathrm{z}_{\Omega_{2k}^c}\right\|_{l_1}^2$. When z is l_1-failing for some Ω_k , then it is l_1-failing for k-largest entry of z_{Ω_k}.

In other words, consider a matrix Φ and a vector z that solves Φ, i.e. $\Phi\mathrm{z} = 0$, then by swapping any 2 appropriate columns of Φ and changing signs then $z_i \geq z_{i+1} \geq 0$. We take an example to illustrate what we said about swapping the columns of Φ more clearly.

Consider $\Phi = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 1 & 2 \end{bmatrix}$ then $= \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} \in \mathcal{N}(\Phi)$., i.e. $\Phi z = 0$. Now, swap 2 appropriate columns of Φ (in this example we must swap columns 2 and 3), then $\Phi^{'} = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 2 & 1 \end{bmatrix}$ and then $z^{'} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} \in \mathcal{N}(\Phi^{'})$. Multiply the second column of $\Phi^{'}$ by (-1) then $\Phi^{''} = \begin{bmatrix} 1 & -2 & 0 \\ 1 & -2 & 1 \end{bmatrix}$. Thus, $z^{'} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$ solves $\Phi^{''}$ and $z_1 = 2, z_2 = 1$ and $z_3 = 0$ then we obtain that $z_i \geq z_{i+1} \geq 0$ as we claimed. Note that $\Phi^{''}$ is minimally redundant because the number of its rows is one less than the number of its columns and it is rows are orthonormal because swapping columns the inner product is still zero. Also when Φ has unit spectral norm, then swapping columns does not change the unit spectral norm property of $\Phi^{''}$.

By defining $\Lambda_0 = \{1, \dots, m\}, \Lambda_1 = \{m+1, \dots, k\}$ and $\Lambda_2 = \{k+1, \dots, N\}$ we can turn our problem of finding the largest $\sigma_{2k}^2(\Phi)$ into an optimisation problem, given by lemma (3.2), whereby we wish to select a unit vector z that allows l_1-failing reconstruction for k-sparse vector while minimizing $\|z_{\Omega_k}\|_{l_2}^2$. The next lemma shows that one only has to consider a simple form for the optimal vector z. Furthermore, this optimization problem below is equivalent to the problem that we are interested in, i.e. the problem of finding the largest asymmetric RIC σ_{2k}^2. By proposition (3.2) finding the largest asymmetric RIC σ_{2k}^2 of the matrix Φ can be determined by $(1 - \|z_{\Omega_m}\|_{l_2}^2)$ where $\|z\|_{l_2}^2 = 1$ which is constraint (3.17) in lemma (3.2). We can write the unit vector z in terms of Λ_0, Λ_1 and Λ_2 as follow:

$$\|z\|_{l_2}^2 = \|z_{\Lambda_0}\|_{l_2}^2 + \|z_{\Lambda_1}\|_{l_2}^2 + \|z_{\Lambda_2}\|_{l_2}^2 = 1.$$

This implies

$$\|z_{\Lambda_0}\|_{l_2}^2 + \|z_{\Lambda_1}\|_{l_2}^2 = 1 - \|z_{\Lambda_2}\|_{l_2}^2.$$

By proposition (3.2), minimizing the function, J(z), (defined below in (3.15)) is equivalent with minimizing $\|z_{\Lambda_0}\|_{l_2}^2 + \|z_{\Lambda_1}\|_{l_2}^2 = \|z_{\Omega_k}\|_{l_2}^2$. The l_1-failing inequality can be interpreted in terms of Λ_0, Λ_1 and Λ_2 as

$$\|z_{\Lambda_0}\|_{l_1} \geq \|z_{\Lambda_0^c}\|_{l_1} = \|z_{\Lambda_1 \cup \Lambda_2}\|_{l_1}$$

and by lemma (2.8) this implies

$$\left\|z_{\Lambda_0}\right\|_{l_1} \geq \left\|z_{\Lambda_1}\right\|_{l_1} + \left\|z_{\Lambda_2}\right\|_{l_1}.$$

Hence

$$\frac{\left\|z_{\Lambda_1}\right\|_{l_1} + \left\|z_{\Lambda_2}\right\|_{l_1}}{\left\|z_{\Lambda_0}\right\|_{l_1}} \leq 1$$

which is constraint (3.16) in the lemma below. Finally, as we previously discussed the form of unit vector z can be assumed as $z_i \geq z_{i+1} \geq 0$ which is constraint (3.18) in the lemma below. By previous analysis the problem of selecting l_1-failing minimally redundant row orthonormal unit spectral norm matrix Φ becomes an optimization problem as shown below.

Lemma (3.2) [Davies and Gribonval [11], lemma 3]:

Consider $k \geq 2m$ and let $z^* \in \mathbb{R}^N$ be a solution to the following optimization problem:

$$\text{Minimise:} \quad J(z) := \frac{\left\|z_{\Lambda_0}\right\|_{l_2}^2 + \left\|z_{\Lambda_1}\right\|_{l_2}^2}{\left\|z_{\Lambda_2}\right\|_{l_2}^2} \tag{3.15}$$

$$\text{Subject to:} \quad \frac{\left\|z_{\Lambda_1}\right\|_{l_1} + \left\|z_{\Lambda_2}\right\|_{l_1}}{\left\|z_{\Lambda_0}\right\|_{l_1}} \leq 1 \tag{3.16}$$

$$\|z\|_{l_2}^2 = 1 \tag{3.17}$$

$$\text{and} \qquad z_i \geq z_{i+1} \geq 0 \tag{3.18}$$

Then z^* is piecewise flat (by flat we mean that the elements have the same value), and has the form:

$$z^* = [\underbrace{\alpha, \dots, \alpha}_{m}, \underbrace{\beta, \dots, \beta}_{L}, \gamma, 0, \dots, 0]^T \tag{3.19}$$

for some constants $\alpha > \beta > \gamma \geq 0$ and some L such that $k + 1 \leq m + L \leq N$.

Proof: Firstly, due to the continuity of $J(z)$ and the compactness of the constraint set, an optimum z^* is guaranteed to exist. This is because a continuous function on a compact set always attains its minimum and maximum value. Then by the contradiction we prove that the shape of z^* is flat.

We split the proof into three parts, firstly we show that $z^*_{\Lambda_0}$ is flat, then we show that $z^*_{\Lambda_1}$ is flat as well and finally by proving $z^*_{\Lambda_2}$ is flat we will conclude that z^* has the claimed form in (3.19). By using the lower bound of Lemma (2.5), we have

$$\left\|z^*_{\Lambda_0}\right\|_{l_2} \geq m^{-1/2}\left\|z^*_{\Lambda_0}\right\|_{l_1} \tag{3.20}$$

where m is the order of sparsity. In terms of Lemma (2.5), the equality holds when x and y are linearly dependent and we obtain

$$|\langle x, y\rangle| = \|x\|_{l_2}\|y\|_{l_2}$$

and then by using the above formula for proving the lower bound of Lemma (2.5) we get the equality in (3.20). In other words, the equality holds only if $z^*_i = m^{-1}\left\|z^*_{\Lambda_0}\right\|_{l_1}$ for all $i \in \Lambda_0$, which means that $z^*_{\Lambda_0}$ is flat. Next, we define $z^{'}$ by $z^{'}_{\Lambda_1\cup\Lambda_2} = z^*_{\Lambda_1\cup\Lambda_2}$ and $z^{'}_i = m^{-1}\left\|z^*_{\Lambda_0}\right\|_{l_1}$ we then rescale the vector $z^{'}$ to make it feasible, i.e. let $z^{''} = \frac{z^{'}}{\|z^{'}\|_{l_2}}$ and later we must show that $z^{''}$ is satisfies the constraints (3.16), (3.17) and (3.18). Firstly, (3.17) holds as $z^{''}$ is a unit vector. Secondly, we shall prove that $z^{''}$ is satisfies (3.16), in other words if we could show that $z^{'}$ is satisfying (3.16) then $z^{''}$ is satisfies (3.16) too because (3.16) is a scale invariant constraint, now by using the assumption $z^{'}_{\Lambda_1\cup\Lambda_2} = z^*_{\Lambda_1\cup\Lambda_2}$ we obtain

$$\left\|z^{'}_{\Lambda_1}\right\|_{l_1} = \left\|z^*_{\Lambda_1}\right\|_{l_1}$$

and

$$\left\|z^{'}_{\Lambda_2}\right\|_{l_1} = \left\|z^*_{\Lambda_2}\right\|_{l_1}.$$

Let $z^{'}_{\Lambda_0} = (\alpha, \dots, \alpha)$ where we have m number of α's in $z^{'}_{\Lambda_0}$, and by computing the l_1-norm of $z^{'}_{\Lambda_0}$ we get:

$$\left\|z^{'}_{\Lambda_0}\right\|_{l_1} = m\alpha = m\left(m^{-1}\left\|z^*_{\Lambda_0}\right\|_{l_1}\right).$$

By the assumption $z^*_i = m^{-1}\left\|z^*_{\Lambda_0}\right\|_{l_1}$ where $i \in \Lambda_0$. We end up with

$$\left\|z^{'}_{\Lambda_0}\right\|_{l_1} = \left\|z^*_{\Lambda_0}\right\|_{l_1}.$$

Hence $z^{'}$ is satisfies (3.16), i.e.

$$\frac{\|z'_{\Lambda_1}\|_{l_1} + \|z'_{\Lambda_2}\|_{l_1}}{\|z'_{\Lambda_0}\|_{l_1}} = \frac{\|z^*_{\Lambda_1}\|_{l_1} + \|z^*_{\Lambda_2}\|_{l_1}}{\|z^*_{\Lambda_0}\|_{l_1}} \leq 1$$

As z^* is feasible. Finally, to show that $z''_i \geq z''_{i+1} \geq 0$ we need to show that $z'_i \geq z'_{i+1} \geq 0$ as $z''_i \geq z''_{i+1} \geq 0$ if and only if $z'_i \geq z'_{i+1} \geq 0$ as the constraint (3.18) is scale invariant.

In order to show that $z'_i \geq z'_{i+1}$, we shall take all possibilities that i might take, i.e. when $i < m$, $i = m$ and $i > m$. In the case of $i < m$, $z'_i = z^*_{i+1}$ because we flattened z' by dividing it by its length. When $i = m$, $z'_i \geq z^*_m$ by the fact that flatten vector is greater than (or equal) to the smallest entry (element) of the same vector before flattening it, i.e. let x be a set of numbers, then $\min(x) \leq \text{average}(x)$. This implies that

$$z'_m \geq z^*_m \geq z^*_{m+1}$$

But $z'_{m+1} = z^*_{m+1}$, hence $z'_m \geq z'_{m+1}$. Lastly, in the case of $i > m$, we previously assumed that $z'_{\Lambda_1 \cup \Lambda_2} = z^*_{\Lambda_1 \cup \Lambda_2}$ whereby we concluded that z' is feasible.

Next, it will be shown that $z^*_{\Lambda_1}$ is flat with all entries equal to $z^*_{k+1} = \|z^*_{\Lambda_2}\|_{l_\infty}$. By contradiction, assume that $z^*_i \neq z^*_{k+1}$ for some $i \in \Lambda_1$ and then construct z' by letting $z'_{\Lambda_0 \cup \Lambda_2} = z^*_{\Lambda_0 \cup \Lambda_2}$ and $z'_i = z^*_{k+1}$ for all $i \in \Lambda_1$. By rescaling, $z'' = \frac{z'}{\|z'\|_{l_2}}$ then we must show that z'' is feasible, i.e. z'' must satisfy constraints (3.16),(3.17) and (3.18). It is obvious that $\|z''\|^2_{l_2} = 1$ because z'' is a unit vector and thus (3.17) holds. By the assumption

$$z'_{\Lambda_0 \cup \Lambda_2} = z^*_{\Lambda_0 \cup \Lambda_2}$$

we obtain

$$\|z'_{\Lambda_0}\|_{l_1} = \|z^*_{\Lambda_0}\|_{l_1} \text{ and } \|z'_{\Lambda_2}\|_{l_1} = \|z^*_{\Lambda_2}\|_{l_1}$$

Also since $z'_i = z^*_{k+1}\ \forall\, i \in \Lambda_1$, then $\|z'_{\Lambda_1}\|_{l_1} = \|max(z^*_{\Lambda_2})\|_{l_1}$ and we conclude

$$\frac{\|z'_{\Lambda_1}\|_{l_1} + \|z'_{\Lambda_2}\|_{l_1}}{\|z'_{\Lambda_0}\|_{l_1}} \leq 1.$$

hence

$$\frac{\|z''_{\Lambda_1}\|_{l_1} + \|z''_{\Lambda_2}\|_{l_1}}{\|z''_{\Lambda_0}\|_{l_1}} \leq 1.$$

Thus constraint (3.16) holds for z''. As we mentioned previously $z''_i \geq z''_{i+1} \geq 0$ if and only if $z'_i \geq z'_{i+1} \geq 0$ which means that in order for z'' to satisfy (3.18) we shall prove that z' satisfies (3.18).To do that we investigate all possible cases that i might have as follow:

When $i < m$, then $z'_{\Lambda_0} = z^*_{\Lambda_0}$ because we assumed that $z'_{\Lambda_0 \cup \Lambda_2} = z^*_{\Lambda_0 \cup \Lambda_2}$ which implies that $z'_{\Lambda_0} = z^*_{\Lambda_0}$. If $i = m$, then $z'_m = z^*_m$ and since $k + 1 > m$ we obtain

$$z'_m = z^*_{m+1}.$$

Whereas if $i \leq k$ then $z'_i = z^*_{k+1}$ since all $i \in \Lambda_1$. Finally, in the case of $i \geq k + 1$, by the assumption $z'_{\Lambda_0 \cup \Lambda_2} = z^*_{\Lambda_0 \cup \Lambda_2}$we obtain $z'_i = z^*_i$.Therefore, by the above analysis we conclude $z'_i \geq z'_{i+1} \geq 0$ which leads to $z''_i \geq z''_{i+1} \geq 0$.

We are now going to describe the shape of $z^*_{\Lambda_2}$. Let j be the smallest index such that $z^*_j < z^*_k$ and l be the largest index such that $z^*_l > 0$. Suppose that $j \neq l$, otherwise we already have the form in (3.19). Next, construct z' with non-increasing, positive entries such that $z'_i = z^*_i$ for all $i \neq \{j, l\}$ as follow:

C1: if $z^*_j + z^*_l \leq z^*_k$, we set:

$$z'_l = 0 \text{ and } z'_j = z^*_j + z^*_l$$

C2: if $z^*_j + z^*_l > z^*_k$, we set:

$$z'_j = z^*_k \text{ and } z'_l = z^*_j + z^*_l - z^*_k .$$

At the beginning, by ***C1*** and ***C2*** we obtain that $\|z'\|_{l_1} = \|z^*\|_{l_1}$ as follow:

$$\|z'\|_{l_1} = \sum_{i=1}^{N} |z'_i| = \sum_{i=1}^{N} z'_i = \sum_{\zeta} z'_\zeta + z'_j + z'_l$$

where $= \{1, \dots, N\} - \{j, l\}$. If we are in the ***C1*** case;

$$\sum_{\zeta} z'_\zeta + z'_j + z'_l = \sum_{\zeta} z'_\zeta + (z^*_j + z^*_l) + 0 = \sum_{\zeta} z'_\zeta + z^*_j + z^*_l = \|z^*\|_{l_1}.$$

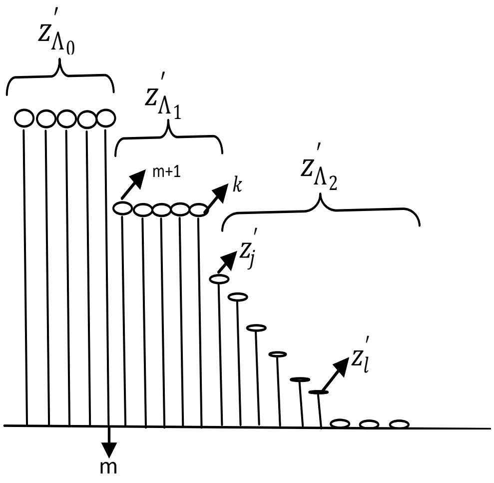

Figure 4: We flattened z'_{Λ_0} and z'_{Λ_1}, and we want to show that the vector z'_{Λ_2} contains one element which is not equal to z^*_k or 0.

In ***C2***, Since $z'_l > 0$ then

$$\sum_{\zeta}(z'_{\zeta} + z'_j + z'_l) = \sum_{i=1}^{N} z^*_i = \|z^*\|_{l_1}$$

hence

$$\|z'\|_{l_1} = \|z^*\|_{l_1}. \tag{3.21}$$

Furthermore, lemma (3.1) implies that $\left\|z'_{\Lambda_2}\right\|_{l_2} > \left\|z^*_{\Lambda_2}\right\|_{l_2}$ as follow:

Let $u_1 = z'_j, u_2 = z'_l, v_1 = z^*_j$ and $v_2 = z^*_l$ then we need to show that

$$(z'_j)^2 + (z'_l)^2 > (z^*_j)^2 + (z^*_l)^2.$$

In other words, we shall show that $z'_j > z^*_j \geq z^*_l > z'_l \geq 0$ in terms of both ***C1*** and ***C2***. We start by showing that $z'_j > z^*_j$ as follow:

In ***C1***, since $z^*_l > 0$ then $z'_j > z^*_j$ because $z'_j = z^*_j + z^*_l$. Whereas in ***C2***, since $z'_j = z^*_k$ and by the assumption $z^*_j < z^*_k$ we end up with $z'_j > z^*_j$.

Furthermore, $z^*_j \geq z^*_l$ since $z^*_i \geq z^*_{i+1}$. Finally, $z^*_l > z'_l$ as follow:

In ***C1***, since $z_l^* > 0$ and $z_l^{'} = 0$ thus $z_l^* > z_l^{'}$. In ***C2***, it is given that $z_l^{'} = z_l^* + z_j^* - z_k^*$, so $z_j^* - z_k^*$ must be less than zero in order to obtain $z_l^{'} < z_l^*$. But $z_j^* < z_k^*$ then $z_j^* - z_k^*$ is less than zero and thus $z_l^* > z_l^{'}$ as claimed. Finally, $z_l^{'} \geq 0$ as follow:

In ***C1***, we already assumed that $z_l^{'} = 0$ whereas in ***C2*** the constructed $z_l^{'} > 0$. Therefore,

$z_j^{'} > z_j^* \geq z_l^* > z_l^{'} \geq 0$ and then by lemma (3.1) we obtain

$$(z_j^{'})^2 + (z_l^{'})^2 > (z_j^*)^2 + (z_l^*)^2.$$

Hence

$$\left\|z_{\Lambda_2}^{'}\right\|_{l_2} > \left\|z_{\Lambda_2}^*\right\|_{l_2}.$$

This implies that $J(z^{'}) < J(z^*)$ (by the fact that we divide $\left\|z_{\Lambda_0}\right\|_{l_2}^2 + \left\|z_{\Lambda_1}\right\|_{l_2}^2$ by a greater quantity which is $\left\|z_{\Lambda_2}^{'}\right\|_{l_2}$ in comparison with dividing $\left\|z_{\Lambda_0}^*\right\|_{l_2}^2 + \left\|z_{\Lambda_1}^*\right\|_{l_2}^2$ by a smaller quantity which is $\left\|z_{\Lambda_2}^*\right\|_{l_2}$). Again we rescale to make the vector $z^{'}$ feasible, i.e. let $z^{''} = \frac{z^{'}}{\|z^{'}\|_{l_2}}$ which satisfies the constraints (3.16),(3.17) and (3.18) as follow:

By using (3.21) we have

$$\frac{\left\|z_{\Lambda_1}^{'}\right\|_{l_1} + \left\|z_{\Lambda_2}^{'}\right\|_{l_1}}{\left\|z_{\Lambda_0}^{'}\right\|_{l_1}} = \frac{\left\|z_{\Lambda_1}^*\right\|_{l_1} + \left\|z_{\Lambda_2}^*\right\|_{l_1}}{\left\|z_{\Lambda_0}^*\right\|_{l_1}} \leq 1.$$

Thus constraint (3.16) holds. It is obvious that constraint (3.17) holds as $z^{''}$ is a unit vector. There is only one constraint left that $z^{'}$ must satisfy which is constraint (3.18). It has been assumed that

$$z_i^{'} = z_i^* \quad \forall\, i \neq \{j, l\}$$

So we need to investigate the cases below in order to show $z^{'}$ has the form of (3.18):

$$z_{j-1}^* \geq z_j^{'} \tag{3.22}$$

$$z_j^{'} \geq z_{j+1}^* \tag{3.23}$$

$$z_{l-1}^* \geq z_l^{'} \tag{3.24}$$

$$z_l^{'} \geq z_{l+1}^* \tag{3.25}$$

Furthermore, for all cases above from (3.22 - 3.25) we need to consider each case in both ***C1*** and ***C2***. Starting with (3.22) in ***C1***, we already have $z_{j-1}^* \geq z_j^*, z_j^* < z_k^*$ and $z_l^* > 0$. This implies that

$$z_{j-1}^* \geq z_k^* \geq z_j^{'}. \tag{3.26}$$

This is because we assumed ***j*** to be the smallest index and thus (3.22) holds in ***C1***. Moving on to verify (3.22) in ***C2***, since $z_j^{'} = z_k^*$ and $z_{j-1}^* \geq z_j^*$ then

$$z_{j-1}^* \geq z_k^* = z_j^{'} > z_j^*.$$

This implies $z_{j-1}^* \geq z_j^{'}$ and then (3.22) holds in ***C2*** as well. Next, (3.23) will be verified in both ***C1*** and ***C2*** as follow:

In general, we have

$$z_{j-1}^* \geq z_j^* \geq z_{j+1}^* \geq z_l^* \geq 0$$

and the assumption $z_k^* \geq z_j^* + z_l^*$ in ***C1*** implies

$$z_{j-1}^* \geq z_j^* \geq z_j^* + z_l^* \geq z_{j+1}^* \geq z_l^* \geq 0.$$

But $z_j^* + z_l^* = z_j^{'}$ then we obtain

$$z_{j-1}^* \geq z_j^* \geq z_j^{'} \geq z_{j+1}^* \geq z_l^* \geq 0.$$

Hence $z_j^{'} \geq z_{j+1}^*$, i.e. (3.23) holds in ***C1***. On the other hand, in ***C2***, $z_j^{'} = z_k^*$ and it is given that

$$z_k^* > z_j^* \geq z_{j+1}^*$$

thus $z_j^{'} \geq z_{j+1}^*$ and (3.23) holds in ***C2***. To confirm (3.24) in both ***C1*** and ***C2***. In ***C1***, $z_l^{'} = 0$ and since $z_{j-1}^* \geq z_l^* \geq 0$ then it is clear that $z_l^{'} \leq z_{j-1}^*$. Thus (3.24) holds in ***C1***. Also, in ***C2*** we have $z_l^{'} = z_l^* + z_j^* - z_k^*$ but $z_j^* - z_k^*$ is negative since $z_j^* < z_k^*$, then $z_l^{'} \leq z_l^*$ and it is given that $z_l^* \leq z_{l-1}^*$ thus

$$z_l^{'} < z_l^* \leq z_{l-1}^*.$$

This implies $z_l^{'} \leq z_{l-1}^*$ then (3.24) holds in ***C2***. Finally, (3.25) holds in both ***C1*** and ***C2*** as follow:

In ***C1***, $z_l^{'} = 0$ and since $z_l^* > 0 \geq z_{l+1}^*$ then $z_l^* > 0 = z_l^{'} \geq z_{l+1}^*$. Thus $z_l^{'} \geq z_{l+1}^*$,i.e. (3.25) holds in ***C1***. Furthermore, in ***C2*** $\mathrm{z}_l^{'} < \mathrm{z}_l^*$ and we have $z_l^* \geq z_{l+1}^*$, therefore $z_l^* > z_l^{'} \geq z_{l+1}^*$ whereby (4.25) holds in ***C2***.

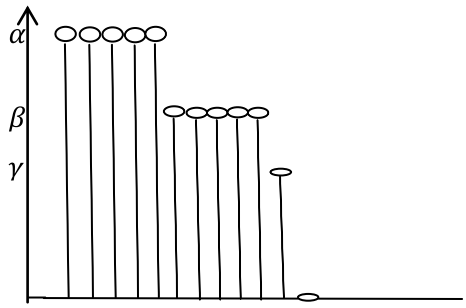

Figure 5: Shape of the optimal vector *z*.

To sum up, by the above analysis we can conclude that $z_{\Lambda_2}^*$ can only have one element not equal to z_k^* or 0. This concludes the proof that z^* must have the claimed form in (3.19) with parameters $\alpha \geq \beta > \gamma \geq 0$ and $k + 1 \leq m + L \leq N$.

■

Next, we calculate the largest asymmetric restricted isometry constant σ_{2k}^2. To do so, we need to evaluate optimal α, β, γ, m and L. Next lemma will be used later to proof theorem (3.2) .

Lemma (3.3) [Davies and Gribonval [11], lemma 4]:

Consider $2k < N$, and $\eta_1 = \sqrt{2} - 1$. Let $z \in \mathbb{R}^N$ be of the form (3.19) with $\alpha > \beta > \gamma \geq 0$ and $k + 1 \leq L \leq N - k$, and assume that z satisfies (3.17) with

$$\frac{\|z_{\Lambda_1}\|_{l_1} + \|z_{\Lambda_2}\|_{l_1}}{\|z_{\Lambda_0}\|_{l_1}} = 1. \tag{3.27}$$

Then

$$\|z_{\Omega_{2k}}\|_{l_2}^2 \geq 2\eta_1.$$

Proof: We start the proof by defining both $\mathrm{L}^{'}$ and η as follow

$$L^{'} := L + \left(\frac{\gamma}{\beta}\right)$$

$$\eta := \frac{k}{L'} \tag{3.28}$$

Since $\beta > \gamma$ then $L \leq L^{'} < L + 1$ and the l_1-failing equality constraint (3.27) reads

$$m\alpha = (k + 1 - m)\beta + \gamma$$

and $k + 1 - m = L$, then

$$m\alpha = L\beta + \gamma$$

Dividing the last equation by β , we obtain

$$\beta = \frac{m}{L'}\ \alpha.$$

By (3.28) we obtain

$$\beta = \eta\alpha.$$

Since $\|z\|_{l_2}^2 = 1$, then it implies

$$\left\|z_{\Lambda_0}\right\|_{l_2}^2 + \left\|z_{\Lambda_1}\right\|_{l_2}^2 + \left\|z_{\Lambda_2}\right\|_{l_2}^2 = 1.$$

Since $\left\|z_{\Lambda_0}\right\|_{l_2}^2 = m\alpha^2$, $\left\|z_{\Lambda_1}\right\|_{l_2}^2 = L\beta^2$ and $\left\|z_{\Lambda_2}\right\|_{l_2}^2 = \gamma^2$, then

$$m\alpha^2 + L\beta^2 + \gamma^2 = 1.$$

We can re-write the above last equation as follow

$$m\alpha^2 + L\beta^2 + \left(\frac{\gamma}{\beta}\right)\beta^2 - \left(\frac{\gamma}{\beta}\right)\beta^2 + \gamma^2 = 1.$$

This implies

$$m\alpha^2 + (L + \frac{\gamma}{\beta})\beta^2 - \left(\frac{\gamma}{\beta}\right)\beta^2 + \gamma^2 = 1.$$

Since $L + \frac{\gamma}{\beta} = L'$ then we obtain

$$m\alpha^2 + L^{'}\beta^2 = 1 + \left(\frac{\gamma}{\beta}\right)\beta^2 - \gamma^2 \geq 1 \tag{3.29}$$

with equality when $\gamma = 0$. Substituting $\beta = \eta\alpha$ and $L^{'} = \frac{m}{\eta}$ in (3.29) we get:

$$m\alpha^2 + \frac{m}{\eta}(\eta\alpha)^2 \geq 1$$

then

$$m\alpha^2 \geq (1+\eta)^{-1} \tag{3.30}$$

and it follows that

$$\left\|z_{\Omega_{2k}}\right\|_{l_2}^2 = m\alpha^2 + m\beta^2 = m\alpha^2 + m(\eta^2\alpha^2) = m\alpha^2(1+\eta^2).$$

By (3.30), we get

$$\left\|z_{\Omega_{2k}}\right\|_{l_2}^2 \geq \frac{(1+\eta^2)}{(1+\eta)} \tag{3.31}$$

Differentiating the right hand side and equating to zero, i.e. let $\varpi = \frac{(1+\eta_1^2)}{(1+\eta_1)}$, then

$$\frac{d\varpi}{d\eta} = \frac{\eta_1^2 + 2\eta_1 - 1}{(1+\eta_1)^2}.$$

We observe that the zero of the derivative indeed yields a minimum, therefore

$$\eta_1^2 + 2\eta_1 - 1 = 0$$

and then $\eta_1 = \sqrt{2} - 1$. Substituting the value of η_1 in (3.31) gives:

$$\left\|z_{\Omega_{2k}}\right\|_{l_2}^2 \geq 2\sqrt{2} - 2 = 2\eta_1.$$

This implies

$$\left\|z_{\Omega_{2k}}\right\|_{l_2}^2 \geq 2\eta_1$$

■

By previous lemmas and propositions we construct minimally redundant row orthonarmal matrices whereby there exists a sparse vector which cannot be recovered by solving (3.3). In their paper particularly theorem 3 in [11], Davies and Gribonval proved a more general result whereby they showed how one can construct successful candidate matrix for l_1-recovery by assuming above lemmas and propositions.

Theorem (3.2) [Davies and Gribonval [11], Theorem 3]:

Consider $0 < \eta_1 < 1$ be the unique positive solution of

$$\eta_1^2 + 2\eta_1 - 1 = 0$$

i.e.$\eta_1 = \sqrt{2} - 1$. Then for every $\varepsilon > 0$, there exist integers $k \geq 1, N \geq 2k + 1$ and a minimally redundant row orthonormal matrix $\Phi \in \mathbb{R}^{(N-1)\times N}$ with:

$$\sigma_{2k}^2(\Phi) \geq 1 - 2\eta_1 - \varepsilon$$

for which there exist a k-sparse vector which cannot be uniquely recovered by solving $\|z_\Omega\|_{l_1} < \|z_{\Omega^c}\|_{l_1}$.

Proof: Consider a unit spectral norm matrix Φ. Assume that Φ is l_1-failing for some k-sparse vector. Then, by proposition (3.1), $\exists\ \Phi^* \in \mathbb{R}^{(N-1)\times N}$ where Φ^* is a minimally redundant row orthonormal (unit spectral norm) matrix which is l_1-failing for the same k-sparse vector such that

$$\sigma_{2k}^2(\Phi) \leq \sigma_{2k}^2(\Phi^*)$$

Also, by proposition (3.2),

$$\sigma_{2k}^2(\Phi^*) \leq 1 - \left\|z_{\Omega_{2k}}\right\|_{l_2}^2$$

Where z is a unit vector which spans the null space $\mathcal{N}(\Phi^*)$. Since Φ^* is l_1-failing, then after proper re-indexing and taking the absolute value, z satisfies the constraints (3.16),(3.17) and (3.18). Therefore by lemma (3.2) z has the form of (3.19) and by lemma (3.3)

$$\left\|z_{\Omega_{2k}}\right\|_{l_2}^2 \geq 2\eta_1 \,. \tag{3.32}$$

Note that η_1 is the largest lower bound. Now, for every $\varepsilon > 0$, $2\eta_1 + \varepsilon$ will be the new lower bound. By contradiction, suppose that there is not such a bound then (3.32) becomes

$$\left\|z_{\Omega_{2k}}\right\|_{l_2}^2 \leq 2\eta_1 + \varepsilon.$$

This produces a minimally redundant row orthonormal unit spectral norm matrix Φ_1^* with

$$\sigma_{2k}^2(\Phi_1^*) \geq 1 - (2\eta_1 + \varepsilon)$$

which is l_1-failing for some k-sparse vector.

■

Finally, we show that when RIC $\delta_{2k} \approx 0.7071$ then the l_1-recovery fails to recover all k-sparse vectors. This indicates that there is a limited room over the best known result of for which all k-sparse vectors can be recovered by l_1-recovery when $\delta_{2k} \approx 0.4531$.

Proof (of Theorem 4.1): In order to find l_1-failing matrices with smallest RIC δ_{2k}, we will be looking for the l_1-failing matrices with the largest asymmetric RIC σ_{2k}^2. By theorem (3.2), we have $\sigma_{2k}^2 \geq 3 - 2\sqrt{2} - \varepsilon$. Using lemma (3.1) for the rescaled matrix, Ψ_{2k} , we have

$$\Psi_{2k} := (\frac{2}{1+\sigma_{2k}^2(\Phi)})^{\frac{1}{2}}\Phi$$

whereby the rescaled RIC, $\delta_{2k}(\Psi_{2k})$, is

$$\delta_{2k}(\Psi_{2k}) \leq \frac{1-\sigma_{2k}^2(\Phi)}{1+\sigma_{2k}^2(\Phi)}.$$

Assume that $\sigma_{2k}^2 = 3 - 2\sqrt{2}$, then substituting this value of σ_{2k}^2 in the above inequality we obtain

$$\delta_{2k}(\Psi_{2k}) \leq \frac{1-3+2\sqrt{2}}{1+3-2\sqrt{2}} = \frac{-2+2\sqrt{2}}{4-2\sqrt{2}} = \frac{\sqrt{2}-1}{2-\sqrt{2}} = \frac{1}{\sqrt{2}}.$$

Then

$$\delta_{2k} \leq \frac{1}{\sqrt{2}}. \tag{3.33}$$

But subtracting a positive quantity from σ_{2k}^2 which is ε, means we add a positive quantity (i.e. ε) to the right hand side of (3.33) whereby δ_{2k} will not be equal to $\frac{1}{\sqrt{2}}$. Hence

$$\delta_{2k} < \frac{1}{\sqrt{2}} + \varepsilon$$

such that for every $\varepsilon > 0$ there exist matrices Ψ_{2k} where l_1-recovery can fail.

■

After we theoretically showed that one can construct l_1-failing minimally redundant row orthonormal unit spectral norm matrices, we next try to construct an example of a matrix that has these properties. Before going to discuss the procedure of building up l_1-failing matrices, we need a theorem known as Gram-Schmidt theorem. Below we state Gram-Schmidt then by using this theorem we go through the procedure of constructing l_1-failing matrices.

Theorem (3.4) (Gram-Schmidt): Consider a vector v in a finite dimensional inner product space V. Then there exist a set of orthonormal basis that contain vector v.

3.1 Constructing l_1-failing matrices

In this sub-section we try to give a general procedure for constructing l_1-failing matrices. Building up l_1-failing matrices from the unit vector z have not been discussed deeply in [11]. Whereas here we somehow show that one can construct these matrices (i.e. l_1-failing matrices) in a general form. To do that we need to use what is known as Gram-Schmidt theorem, given by theorem (3.4), for constructing a set of orthonormal basis from the unit vector z.

Proposition (3.3): Let $z \in \mathbb{R}^N$ be a vector which has the form of (3.19). Then we can construct a minimally redundant row orthonormal unit spectral norm matrix Φ that contain z such that $\Phi \in \mathbb{R}^{(N-1)\times N}$.

Proof: Select vector z in a way that has the form of (3.19). Then use the Gram-Schmidt procedure to create a set of orthonormal basis, say $\{u_1, u_2, \dots, u_{N-1}\}$, such that $z = u_1$ is the first orthonormal vector in that set. Then select the rows of matrix Φ to be the transpose of the orthonormal basis vectors above, i.e.

$$\Phi = \begin{bmatrix} z^T \\ u_2^T \\ \vdots \\ u_{N-1}^T \end{bmatrix}_{(N-1)\times N}.$$

Firstly, Φ is minimally redundant as the number of its rows is equal to the number of its columns minus 1. Secondly, Φ is row orthonormal as we constructed the rows of Φ by Gram-Schmidt process. Finally, multiplying Φ by one of its rows, then the result of taking Euclidean norm of this multiplication and dividing it by the norm the row will be 1. In other words, in order to show that Φ is a unit spectral norm then Φ must satisfy definition (3.1), i.e.

$$|||\Phi||| = \sup_{x\neq 0} \frac{\|\Phi x\|_{l_2}}{\|x\|_{l_2}} = 1$$

where x is the transpose of one of the rows of Φ. It is clear that $\|x\|_{l_2} = 1$ as it is created by Gram-Schmidt procedure. Also, because we selected a row which is orthonormal with all of the rows of Φ to multiply Φ by, then the result of taking norm of this multiplication will be 1. Finally we end up with a matrix Φ which has the property of minimally redundant row orthonormal unit spectral norm whereby l_1-recovery can fail.

■

To make this procedure clearer, we take an example and we go through each step that we claimed above.

Let $\Phi \in \mathbb{R}^{3\times 4}$ be our given matrix. Then we select the vector z such that it has the following form

$$\mathrm{z} = \begin{bmatrix} \sqrt{3/4} \\ 1/\sqrt{12} \\ 1/\sqrt{12} \\ 1/\sqrt{12} \end{bmatrix} \in \mathbb{R}^4$$

where z has the form of (3.19). In other words, $L = 2$ and $m = 1$ where $k \geq 2m = 2$ then $k + 1 = 3 \leq m + L = 3 \leq N = 4$.

Now, z is a unit norm and it has the form of (3.19). Next apply the Gram-Schmidt procedure.

Let $v_1 = z,$ and $v_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, v_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ and $v_4 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$ are three standard basis of $\mathbb{R}^4$. According to the Gram-Schmidt theorem, we assume that the first orthonormal vector u_1 is the same as the first vector in the inner product space which is z, i.e.

$$u_1 = v_1 = z = \begin{bmatrix} \sqrt{3/4} \\ 1/\sqrt{12} \\ 1/\sqrt{12} \\ 1/\sqrt{12} \end{bmatrix}.$$

Also , the second orthonormal vector u_2 can be obtained by

$$u_2 = v_2 - \frac{\langle u_1, v_2 \rangle}{\langle u_1, u_1 \rangle} u_1.$$

Since $\langle u_1, u_1 \rangle = 1$ and $\langle u_1, v_2 \rangle = \sqrt{3/4}$, then the above formula can be written as

$$u_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} {}^{-3}/{}_{-4} \\ -1/4 \\ -1/4 \\ -1/4 \end{bmatrix} = \begin{bmatrix} 1/4 \\ -1/4 \\ -1/4 \\ -1/4 \end{bmatrix}.$$

Then we normalize u_2, i.e. $u_2 = \frac{u_2}{\|u_2\|}$ such that

$$u_2 = 2 \times \begin{bmatrix} 1/4 \\ -1/4 \\ -1/4 \\ -1/4 \end{bmatrix} = \begin{bmatrix} 1/2 \\ -1/2 \\ -1/2 \\ -1/2 \end{bmatrix}.$$

This implies that $\|u_2\| = 1$. Next we find the third orthonormal vector u_3 as follow:

$$u_3 = v_3 - \frac{\langle u_1, v_3 \rangle}{\langle u_1, u_1 \rangle} u_1 - \frac{\langle u_2, v_3 \rangle}{\langle u_2, u_2 \rangle} u_2.$$

Since $\langle u_1, u_1 \rangle = 1$, $\langle u_2, u_2 \rangle = 1$ and by substituting the values of v_3, u_2 and u_1 in the above formula we obtain:

$$u_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -1/4 \\ -1/12 \\ -1/12 \\ -1/12 \end{bmatrix} + \begin{bmatrix} 1/4 \\ -1/4 \\ -1/4 \\ -1/4 \end{bmatrix} = \begin{bmatrix} 0 \\ 2/3 \\ -1/3 \\ -1/3 \end{bmatrix}.$$

Normalizing u_3, i.e. $u_3 = \frac{u_3}{\|u_3\|}$ where $\|u_3\| = \frac{\sqrt{2}}{\sqrt{3}}$ will give a normalized vector u_3 as follow:

$$u_3 = \begin{bmatrix} 0 \\ \sqrt{2}/\sqrt{3} \\ -1/\sqrt{6} \\ -1/\sqrt{6} \end{bmatrix}$$

which has unit norm. Finally, we calculate the last orthonormal vector as follow:

$$u_4 = v_4 - \frac{\langle u_1, v_4 \rangle}{\langle u_1, u_1 \rangle} u_1 - \frac{\langle u_2, v_4 \rangle}{\langle u_2, u_2 \rangle} u_2 - \frac{\langle u_3, v_4 \rangle}{\langle u_3, u_3 \rangle} u_3.$$

By substituting the values of v_4, u_1, u_2 and u_3 we get:

$$u_4 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -1/2\sqrt{2} \\ -1/12 \\ -1/12 \\ -1/12 \end{bmatrix} + \begin{bmatrix} 1/4 \\ -1/4 \\ -1/4 \\ -1/4 \end{bmatrix} + \begin{bmatrix} 0 \\ 1/3 \\ -1/6 \\ -1/6 \end{bmatrix} = \begin{bmatrix} 1 - \sqrt{2}/4 \\ 0 \\ 1/2 \\ -1/2 \end{bmatrix}.$$

By normalizing u_4 we obtain:

$$u_4 = \begin{bmatrix} \frac{1-\sqrt{2}}{\sqrt{11-2\sqrt{2}}} \\ 0 \\ \frac{2}{\sqrt{11-2\sqrt{2}}} \\ \frac{-2}{\sqrt{11-2\sqrt{2}}} \end{bmatrix}.$$

Finally, the rows of matrix Φ is the transpose of u_1, u_2 and u_3 as follow:

$$\Phi = \begin{bmatrix} \frac{1}{2} & \frac{-1}{2} & \frac{-1}{2} & \frac{-1}{2} \\ 0 & \frac{\sqrt{2}}{\sqrt{3}} & \frac{-1}{\sqrt{6}} & \frac{-1}{\sqrt{6}} \\ \frac{1-\sqrt{2}}{\sqrt{11-2\sqrt{2}}} & 0 & \frac{2}{\sqrt{11-2\sqrt{2}}} & \frac{-2}{\sqrt{11-2\sqrt{2}}} \end{bmatrix}.$$

It is clear that Φ is minimally redundant as $\Phi \in \mathbb{R}^{3\times 4} = \mathbb{R}^{(N-1)\times N}$ where $N = 4$. Also, the rows of Φ are orthonormal as they constructed by Gram-Schmidt process. Hence Φ is a minimally redundant row orthonormal matrix. To show that Φ has unit spectral norm, let $x = u_3$ which is the second row in matrix Φ. Since

$$\|\Phi x\|_{l_2} = 1 \text{ and } \|x\|_{l_2} = 1$$

then

$$|||\Phi||| = 1.$$

Finally, we concluded that Φ is a minimally redundant row orthonormal matrix which has unit spectral norm.

3.2 Future work

After we showed that in order to search for the l_1-failing matrices we need to search for the minimally redundant row orthonormal matrices. Then this might help for investigating whether it is sensible to consider that l_1-failing matrices have the minimally redundant property (i.e. to throw away the unit spectral norm property). Furthermore, by theorem (3.1) we stated that δ_{2k} cannot be bigger than 0.7071 in order to recover all sparse vectors and it was already known that when $\delta_{2k} < 0.472$ then we recover all sparse vectors. Then there is a gap between these two results and the question that arises here is: what if we have a minimally redundant row orthonormal matrix such that δ_{2k} has a value in the gap between these two results above, i.e. between 0.472 and 0.7071 ?

References

[1] ACM SIGKDD and Netflix, "Proceedings of KDD Cup and Workshop, " 2007. [Online]. [Cited: August 27, 2012] http://www.cs.uic.edu/~liub/KDD-cup-2007/proceedings.html.

[2] Baranuik, R. G., "Compressive Sensing." s.l. : IEEE Signal Processing Magazine, 2007, Issue 4, Vol. 24, pp. 118-121.

[3] Cai, T. T., Wang, L. and Xu, G., "Shifting Inequality and Recovery of Sparse Signals." s.l. : IEEE Transactions on Signal Processing, March 2010, Issue 3, Vol. 58, pp. 1300-1308.

[4] Candes, E. J. and Tao, T., "Decoding by linear programming." s.l. : IEEE Transactions on Information Theory, December 2005, Issue 12, Vol. 51, pp. 4203-4215.

[5] Candes, E. J., Romberg, J. and Tao, T., "Robust uncertainty priciples: exact signal reconstruction from highly incomplete frequency information." s.l. : IEEE Transactions on Information Theory, February 2006, Issue 2, Vol. 52, pp. 489-509.

[6] Candes, E. J. and Wakin, M. B., "An Introduction to Compressive Sampling." s.l. : IEEE Signal Processing Magazine, 2008, Issue 2, Vol. 25, pp. 21-30.

[7] Candes, E. J., "The restricted isometry propery and its impications for compressed sensing." Paris : Compte Rendus de I'Academie des Science, 2008, Issue 346, Vol. series I, pp. 589-592.

[8] Candes, E. J., "Issac Newton Inestitute for Mathematical Science". www.newton.ac.uk. [Online] University Of Cambridge, March 21, 2011. [Cited: April 4, 2012.] http://www.newton.ac.uk/programmes/INI/seminars/032114001.html.

[9] Coyne, K., "MRI: A Guided Tour," [Online]. [Cited: August 11, 2012] http://www.magnet.fsu.edu/education/tutorials/magnetacademy/mri/.

[10] Devanport, M. A., et al., "Introduction to Compressed Sensing." www.dfg-spp1324.de. [Online] May (2011). [Cited: July 17, 2012.] http://www.dfg-spp1324.de/download/preprints/preprint093.pdf

[11] Davies, M. E. and Gribonval, R., "Restricted Isometry Constants Where l_p sparse recovery can fail for $0 < p \leq 1$." s.l. : IEEE Transactions on Information Theory, May 2009, Issue 5, Vol. 55, pp. 2203-2214.

[12] Donoho, D. L., "Compressed Sensing." s.l. : IEEE Transactions on Information Theory, April 2006, Issue 4, Vol. 52, pp. 1289-1306.

[13] Foucart, S. and Lai, M., "Sparsest Solutions of Underdetermined linear systems via l_q-minimization for $0 < q < 1$." s.l. : Applied and Computational Harmonic Analysis, 2009, Vol. 26, pp. 395-407.

[14] Lay, D. C., "The Singular Value Decomposition," in *Linear Algebra and its Applications*, Boston, Pearson education, Inc., 2012, pp. 414-432.