Error Bounds for Compressed Sensing Algorithms With Group Sparsity

Shashank Ranjan

A Thesis Submitted to
Indian Institute of Technology Hyderabad
In Partial Fulfillment of the Requirements for
The Degree of Master of Technology

Department of Electrical Engineering

June 2011

Declaration

I declare that this written submission represents my ideas in my own words, and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from the sources that have thus not been properly cited, or from whom proper permission has not been taken when needed.

Shashank Ranjan

(Signature)

Shashank Ranjan

(– Student Name –)

EE14MTECH11038

(Roll No)

Approval Sheet

This Thesis entitled Error Bounds for Compressed Sensing Algorithms With Group Sparsity by Shashank Ranjan is approved for the degree of Master of Technology from IIT Hyderabad

---

(Dr. Phanindra Jampana) Examiner
Dept. of Chem Eng
IITH

(Dr. Sumohana Channappayya) Examiner
Dept.of Electrical Eng
IITH

(Dr. Ketan P. Detroja) Adviser
Dept. of Electrical Eng
IITH

(Prof. M Vidyasagar) Co-Adviser
Dept. of Electrical Eng
IITH

(Dr. K. Sri Rama Murty) Chairman
Dept. of Electrical Eng
IITH

Abstract

In compressed sensing, to recover a sparse signal or nearly sparse signal from noisy measurements, most popular method is ℓ_1-norm minimization [1]. The signals in this context are actually some vectors in $\mathbb{R}^n$. For conventionally sparse signal, latest approch to derive upper bound for the ℓ_2, ℓ_1-norm of error between estimated signal and original signal is shown in [2] and [3]. For "group sparse" signals, upper bound for the norm of error is given in [4]. In the present work, we focused on group sparse signals, and presented a unified approach to eatablish upper bound on the norm of error. For group sparse signal recovery, we also introduced a new bound on RIC constant which is diiferent from the one proposed in [4]. A key technical tool, which represents a vector in polytope set by convex combination of sparse vectors is discussed in [2]. We modified this key technical tool for group sparse signals to establish the above discussed unified approach.

Contents

Chapter 1

Introduction to Compressed Sensing

1.1 Motivations

Compressive sensing, also referred to as compressive sampling or compressed sensing, is an signal processing technique which has attracted a lot of attention recently. Compressed sensing deals with efficiently acquiring a signal and reconstructing it, by finding the solution to undetermined linear system. It exploits the fact that many of the natural signals are sparse when represented in proper basis. For example, it is a well known fact that a signal can't be time limited as well as band limited simultaneously, hence it is always sparse in one of its domain. This sparsity property enables us to recover the original signal from far fewer samples than required by the Shannon-Nyquist sampling theorem. Shannon-Nyquist sampling theorem states that in order to exactly recover a signal we must sample it at the or above the Nyquist rate which is twice the maximum frequency component present in the original signal. This requirement increases the computational cost. However, in applications like imaging, sensor networks, astronomy, high-speed analog-to-digital compression and biological systems, the signals of interest are often sparse over a certain basis. For example, if we take a typical image consisting of a million pixels, it is found to be very sparse or compressible over the wavelet basis. Only a small fraction of wavelet coefficients, about one hundred or one thousand out of a million wavelet coefficients, are significant in recovering the original image, while the rest of wavelet coefficients are discarded in the process of compression.

This process of sampling at full rate and then compression can prove to be costlier, as the cost of sensors, sampling resources or the cost of running them, are often high. In many applications we have the situations where the resources such as sensors, energy, and observation time etc. are limited, so sampling at Nyquist rate is not a feasible option. At this point, the compressed sensing comes into play which promises to recover high dimensional signal exactly or accurately by using far fewer, non-adaptive linear samples or measurements of the original signal. In general, signals in this context are represented by vectors from linear spaces, and in many applications they may represent an image or any other quantity. As a result of very few linear measurement of the signal we have a condition where number of equation is less than the number of unknowns. Basic principles of linear algebra says it is not possible to obtain the exact signal from incomplete set of measurements. However, as we discussed before, many of these signals such as image or audio are sparse when represented in proper basis. This enables us to get back the original signal from incomplete measurement by using some demodulation algorithm. One of the most efficient, effective and popular demodulation algorithm for compressed sensing is the basis pursuit algorithm [5], also known as ℓ_1 norm minimization algorithm.

1.2 Mathematical Formulation of Compressed Sensing

In this section, first of all we define the meaning of sparsity? Sparse means, something which is less dense. So, if someone says that a given vector is sparse, that means most of its elements are zero. More specifically, if a vector $x \in \mathbb{R}^n$ is k-sparse, then it will have maximum k non-zero elements in it. The set of all k-sparse vectors is denoted as Σ_k. Mathematically we can write,

$$\|x\|_0 = |\text{supp}(x)| \leq k \ll n, \; \forall x \in \Sigma_k.$$

Where, supp(x) is the support set of vector x, and it is defined as,

$$\text{supp}(x) := \{i \in [n] : (x)_i \neq 0\}.$$

In compressed sensing we only take few linear measurement of a sparse or linearly sparse signal. Let us assume that we have m measurements, where $m \ll n$. Each of the measurements can be seen as the inner-product between the original signal $x \in \mathbb{R}^n$ and measurement vector $a_i \in \mathbb{R}^n$, where $i = 1, 2, ..., m$. If we consider the m measurements as a vector $y \in \mathbb{R}^m$, we may then consider a measurement matrix $A \in \mathbb{R}^{m \times n}$ whose rows are the vectors a_i. So, the sparse signal recovery problem can be seen as the recovery of a signal $x \in \mathbb{R}^n$ from its measurement $y = A$. Mathematically we can write

$$y = Ax. \tag{1.1}$$

Now the question that arises at this point of time is, knowing that x is k-sparse, how to get back x from its measurement y ? Theoretically, one may attempt to recover the signal x by solving ℓ_0-minimization problem,

$$\hat{x} := \underset{z}{\text{argmin}} \, \|z\|_0 \quad \text{s.t. } Az = y. \tag{1.2}$$

Now, we define the null space of A as,

$$\mathcal{N}(A) = \{z \in \mathbb{R}^n : Az = 0\}.$$

Here we make a statement that in case of noise-less measurement, if $\mathcal{N}(A)$ does not contain any vector in Σ_{2k} except a **null vector**, or in short $\mathcal{N}(A) \cap \Sigma_{2k} = \{0\}$, then the solution to the equation (1.2) is exactly equal to x. This can can be proved very easily as, $A\hat{x} = Ax = y$, which implies $A(\hat{x} - x) = 0$ and hence, $(\hat{x} - x) \in \mathcal{N}(A)$. Further, it is not difficult to see that, $(\hat{x} - x) \in \Sigma_{2k}$. But we have assume that $\mathcal{N}(A)$ does not contain any nonzero vector in Σ_{2k}, thus the only vector in $\mathcal{N}(A)$ which is $2k$ sparse is $z = 0$, which leads to $\hat{x} = x$. Based on the above fact we conclude that, ℓ_0-minimization problem works perfectly theoretically. However, in practical it is a NP-Hard problem[6]. Fortunately in compressed sensing we have many alternative to ℓ_0-minimization algorithm, which are computationally efficient.

1.2.1 ℓ_1-minimization

ℓ_1-minimization algorithm also referred to as "basis pursuit", is an alternative to ℓ_0-minimization algorithm, which relaxes the NP-Hard problem associated with ℓ_0-norm minimization. In case of noise-free measurement we recover the x as:

$$\hat{x} := \underset{z}{\operatorname{argmin}} \|z\|_1 \text{ s.t. } Az = y. \tag{1.3}$$

The ℓ_1-norm minimization approach is sometimes referred to as the LASSO formulation, due its similarity to the LASSO formulation for sparse regression [7]. It is interesting to note that, this algorithm often recovers x exactly when x is sparse and accurately when x is nearly sparse, even though the ℓ_1-norm is different from ℓ_0-norm which is a quasi-norm. We should note that the measurement matrix A remains fixed and it is independent of the signal. The ℓ_1 minimization algorithm will succeed [8],[9] whenever signal x is sufficiently sparse and matrix A satisfies some conditions which we will discuss later. Surprisingly, ℓ_1-norm minimization gives sparsest solution. According to the paper [10] sparsity promoting nature of ℓ_1-norm was first noticed in 1960 by Logan [11] where he proved probably the first ℓ_1-uncertainty principle. The ℓ_1 minimization also finds its applications in seismology, where sparse reflection function from band limited data indicates meaningful changes between surface layers [12]. A paper [7] proposed LAASO algorithm for sparse model selection in statistics, after which search for the application areas of ℓ_1 minimization began to broden.

The basis pursuit was proposed [5] for obtaining sparse representation of a given signal from a given over complete dictionary, and a similar approach known as variation minimization was proposed in [13]. But the major breakthrough was achieved in [8],[9] and [14], therein it was shown that Basis Pursuit is able recover sparse signals from its few linear measurements. However in order to recover sparse signal, a stronger condition on the measurement matrix A is required. For example, restricted isometry property (RIP) condition which guarantees that ℓ_1 minimization accurately recovers a sparse signals. The number of rows for the measurement which satisfies these properties ig given by $m = k\ log(n)^{O(1)}$ [14].

1.2.2 Greedy Algorithms

Under the the certain conditions on measurement matrix, ℓ_1 minimization guarantees exact sparse recovery in noiseless case and robust sparse recovery in case of noisy measurements. But solving a ℓ_1 minimization problem is found to be highly complex. For example, in linear programming, the complexity grows cubic in the problem dimension n. In applications where the signal x is of very large dimension, ℓ_1 minimization takes too much time. In such cases we use greedy algorithms which are comparatively fast in both theory and practice. There are several greedy algorithm exists such as Orthogonal Matching Pursuit [15], [16] Regularized Orthogonal Matching Pursuit [17], Stagewise Orthogonal Matching Pursuit [18], Iterative Thresholding [19], Matching Pursuit [20]. Greedy algorithm approximates the sparse signal iteratively and after each iteration they come up with a better approximation. Among all the Matching Pursuit algorithm is the simplest one. It projects the measurement y on each of the column of measurement matrix $A \in \mathbb{R}^{m \times n}$. The projection which possesses largest euclidean distance gets subtracted from the measurement y. After the subtraction, the obtained vector is know as residue (r). If the euclidean distance of r is not less than some predefined value (known as stopping criteria), then we again repeat the same procedure with r in place of y. This algorithm will continue until the stopping criteria is satisfied. Even though MP takes an infinite number of iterations in general, it converges exponentially. On the other hand, in case of Orthogonal Matching Pursuit (OMP), though it selects atoms (columns) iteratively from A, but it uses the current approximate as an orthogonal projection onto the Space Spanned by the column selected so far. The main consequence of this modification is that the OMP is guaranteed to converge in n (or fewer) iterations. We can interpret the algorithm as building up a subspace in which the approximation lives and each iteration adds a dimension to this subspace, and so at most n iterations are possible.

Though the greedy algorithm are comparatively fast, but most of them do not deliver upto the expectations when compared to ℓ_1 minimization, in addition to that most of the greedy algorithms don't provides provable guarantees for robust sparse recovery.

1.3 Conditions On Measurement Matrix

As we know that, the measurement of a signal $x \in \mathbb{R}^n$, in noiseless case can be seen as $y = Ax$, where $A \in \mathbb{R}^{m\times n}$ is the measurement matrix. Since $m \ll n$, we are getting a compressed form of the signal x. But, we should must ensure that the measurement y is preserving significant amount information contained in x. In this section we are concerned with different conditions to which the sensing matrix A should satisfy such that y preserves sufficient amount of information.

1.3.1 Null Space-Based Condition

We have already seen in section 1.2 that in case of noise-less measurement, if the signal x is k-sparse and A satisfies the condition $\mathcal{N}(A) \cap \Sigma_{2k} = \{0\}$, then theoretically the ℓ_0-norm minimization algorithm in (1.2) will exactly recover x. However in practical situations, signals are not exactly sparse, and measurements are always associated with some amount of random noise. In addition, solving ℓ_0-minimization is not possible. Hence, we need to setup some appropriate condition on A as well as we have modify our recovery algorithm. In compressed sensing recovery algorithm is often termed as "demodulation map", denoted by $\Delta : \mathbb{R}^m \to \mathbb{R}^n$. For different noise model we will define different demodulation map (Δ).

Let us consider measurement as

$$y = Ax + \eta,$$

where, $x \in \mathbb{R}^n$ is our original signal and $\eta \in \mathbb{R}^m$ is some random noise. If the noise η follows the characteristic $\|\eta\|_2 \leq \epsilon$, then we generally use demodulation map (Δ) of the type,

$$\Delta(y) = \hat{x} := \underset{z}{\operatorname{argmin}} \|z\|_1 \text{ s.t. } \|Az - y\|_2 \leq \epsilon. \tag{1.4}$$

In case, if the noise η follows the characteristic $\|A^t\eta\|_\infty \leq \epsilon$, then we usually use demodulation map (Δ) of the type

$$\Delta(y) = \hat{x} := \underset{z}{\operatorname{argmin}} \|z\|_1 \text{ s.t. } \|A^t(Az - y)\|_\infty \leq \epsilon. \tag{1.5}$$

Before going into the further discussion we introduce couple of definitions which we will use frequently in this work.

Definition* 1.1*: Suppose positive integer n and $k < n$, and a norm $\|\cdot\|$ on $\mathbb{R}^n$ are specified. Let $x \in \mathbb{R}^n$ be arbitrary. Then the quantity

$$\sigma_k(x, \|\cdot\|) := \min_{z \in \Sigma_k} \|x - z\|$$

is called the **k-sparsity index** of vector $x \in \mathbb{R}^n$ with respect to the norm $\|\cdot\|$. It is obvious that $x \in \Sigma_k$ if and only if $\sigma_k(x, \|\cdot\|) = 0$.

Definition* 1.2*: Suppose we are given positive integers n and $k < n$, a measurement matrix $A \in \mathbb{R}^{m \times n}$, and a demodulation map $\Delta : \mathbb{R}^m \to \mathbb{R}^n$. The various property of pair (A, Δ) are defined as follows:

- The pair (A, Δ) is said to achieve **exact sparse recovery** of order k if

$$\Delta(Ax) = x, \ \forall x \in \Sigma_k.$$

- The pair (A, Δ) is said to achieve **stable sparse recovery** of order k if there exists a constant C_1 such that for some $p \geq 1$, it is the case that

$$\|\Delta(Ax) - x\|_p \leq \mathrm{C}_1 \sigma_k(x, \|\cdot\|_1) \ \forall x \in \mathbb{R}^n.$$

- The pair (A, Δ) is said to achieve **robust sparse recovery** of order k if there exist constants $\mathrm{C}_1, \mathrm{C}_2$ and some $p \geq 1$, such that for all $\eta \in \mathbb{R}^m$ with $\|\eta\|_2 \leq \epsilon$, it is the case that

$$\|\Delta(Ax + \eta) - x\|_p \leq \mathrm{C}_1 \sigma_k(x, \|\cdot\|_1) + \mathrm{C}_2 \epsilon \ \forall x \in \mathbb{R}^n.$$

Though there are several null space-based conditions exists, but we will talk about only two of them those are more relevant in this work.

Definition* 1.3*: A matrix $A \in \mathbb{R}^{m \times n}$ is said to satisfy the **exact null space (ENS) property** of order k if,

$$\|v\|_1 < 2\sigma_{k,\mathcal{G}}(x, \|\cdot\|_1), \forall v \in \mathcal{N}(A) \backslash \{0\}. \tag{1.6}$$

Following this definition we introduce a theorem.

Theorem* 1.1*: If the measurement matrix $A \in \mathbb{R}^{m \times n}$ satisfies **exact null space proerty (ENSP)** of order k, then in noiseless case, recovery algorithm defined in (1.3) will exactly recover k-sparse signals.

Proof: In order to prove this we will show that x is the unique minimizer to (1.3). Let an arbitrary vector $z \in \mathbb{R}^n$ such that $y = Az = Ax$ and $z \neq x$, where $x \in \Sigma_k$. We denote supp(x) by S. It is quite obvious that $h = (x - z) \in \mathcal{N}(A) \backslash \{0\}$. Using the ENSP, we have that

$$\|h\|_1 < 2\|h_{S^c}\|_1,$$

where S^c is the complement of S and it can be expressed as $= \{1, 2, ...n\} - S$. We should note that h_S denotes a vector which retains elements of vector h corresponding to index set S and the remaining elements are set to zero. On simplifying the above inequality, we get

$$\|h_S\|_1 < \|h_{S^c}\|_1. \tag{1.7}$$

Using the triangle inequality and (1.7), we have that

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

The a above fact establishes the fact that $\|x\|_1$ is the unique minimizer to (1.3) and hence it proves our theorem.

***Definition* 1.4:** A matrix $A \in \mathbb{R}^{m\times n}$ is said to satisfy the ℓ_2 **robust null space property (ℓ_2-RNSP)** of order k and norm $\|\cdot\|_2$, with constants $\rho \in (0, 1), \tau \in \mathbb{R}_+$, if, for all $h \in \mathbb{R}^n$ and all sets $S \subset [n]$ s.t. $|S| \leq k$, it is true that

$$\|h_S\|_2 \leq \frac{\rho}{\sqrt{k}}\|h_{S^c}\|_1 + \frac{\tau}{\sqrt{k}}\|Ah\|_2. \tag{1.8}$$

With this definition we introduce a theorem for "robust recovery" of signals.

***Theorem* 1.2:** Suppose $A \in \mathbb{R}^{m\times n},\ x \in \mathbb{R}^n$, and $y = Ax + \eta$ where $\|\eta\|_2 \leq \epsilon$. Define $\hat{x} = \Delta(y)$ as in (1.4). Suppose that A satisfies ℓ_2 robust null space property of order k and norm $\|\cdot\|_2$, with constants $\rho \in (0, 1), \tau \in \mathbb{R}_+$. Then, for all $p \in [1, 2]$, we have that,

$$\|\hat{x} - x\|_p \leq \frac{1}{k^{1-1/p}}.\frac{2}{1-\rho}[(1 + 2\rho)\sigma_k(x, \|\cdot\|_1) + 3\tau\epsilon]. \tag{1.9}$$

Proof of this theorem is omitted here, but the interested readers are directed to [3]. We should notice that the upper bound on the norm of error $e = \hat{x} - x$ is finite, see inequality (1.9). In case of noise-free measurements, if $x \in \Sigma_k$, then from inequality (1.9) we get $\hat{x} = x$. Hence we can say that A is preserving significant amount of information contained in x.

So far in this section we have seen different properties to which the measurement matrix A should satisfy in order to achieve sparse recovery.

1.3.2 Restricted Isometry Property

We mentioned in subsection 1.3.1 that if A satisfies ℓ_2-robust null space property (ℓ_2-RNSP) as in definition 1.4, then the demodulation map Δ defined in (1.4) will exactly recovers the signal x in case of sparse signal recovery problem. However, satisfying ℓ_2-RNSP is just an idea and it is not clear how one may go about choosing matrix A to have this property. In the present subsection we introduce a property known as Restricted Isometry Property (RIP), and it is shown in [3] that RIP implies ℓ_2-RNSP.

***Definition* 1.5:** A matrix $A \in \mathbb{R}^{m\times n}$ is said to satisfy the **restricted isometry property (RIP)** of order k with constant $\delta_k \in (0,1)$ if

$$(1-\delta_k)\|u\|_2^2 \leq \|Au\|_2^2 \leq (1+\delta_k)\|u\|_2^2, \quad \forall u \in \Sigma_k. \tag{1.10}$$

Where δ_{tk} is known as **restricted isometry constant (RIC)**. Another equivalent way of expressing (1.10) is the following: For every subset $J \subset [n]$ with $|J| \leq k$, the singular values of the matrix $A_J^t A_J$ lies in the interval $[1-\delta_k, 1+\delta_k]$.

***Theorem* 1.3:** Suppose $A \in \mathbb{R}^{m\times n}$ satisfies the RIP of order $2k$ with constant $\delta_{2k} < \sqrt{2}-1$, and that $y = Ax + \eta$ for some $x \in \mathbb{R}^n$ and $\eta \in \mathbb{R}^m$ with $\eta 2 \leq \epsilon$, then the demodulation map Δ defined in (1.4) leads to the bound:

$$\|\hat{x} - x\|_2 \leq C_1 \sigma_k(x, \|\cdot\|_1) + C_2\epsilon, \tag{1.11}$$

Where,

$$C_1 = \frac{2}{\sqrt{k}} \frac{1+(\sqrt{2}-1)\delta_{2k}}{1-(\sqrt{2}+1)\delta_{2k}} \tag{1.12}$$

$$C_2 = 4\frac{\sqrt{1+\delta_{2k}}}{1-(\sqrt{2}+1)\delta_{2k}}. \tag{1.13}$$

For the proof of this theorem readers are directed to [21].

1.4 Group Sparsity

Until now we have been studying what might be called "pure" sparsity or conventional sparsity, in which the quantity of interest is the number of nonzero components of a vector. However, in applications such as cancer biology, the quantity of interest is the number of "pathways" in which a vector have non-zero components, and not the absolute number of non-zero components. This kind of application leads naturally to the notion of group sparsity.

1.4.1 Mathematical Structure of Group Sparsity

Throughout this work, the symbol $[n]$ denotes the set $\{1, \ldots, n\}$ whenever n is an integer. For some $\Lambda \subset [n]$ and $h \in \mathbb{R}^n$, the symbol h_Λ denotes a vector which retains elements of vector h corresponding to index set Λ and the remaining elements are set to zero. The symbol Λ^c denotes a set which is defined by $\Lambda^c = [n] - \Lambda$.

Now we introduce the following definition,

***Definition* 1.6:** Let $\mathcal{G} = \{G_1, \ldots, G_g\}$ be a partition of $[n]$ such that $|G_i| \leq k$ for all i. If $S \subseteq \{1, \ldots, g\}$, define $G_S := \cup_{i\in S} G_i$. A subset $\Lambda \subseteq [n]$ is said to be **group k-sparse** if there exists a subset $S \subseteq \{1, \ldots, g\}$ such that $\Lambda = G_S$, and in addition, $|\Lambda| \leq k$. The collection of all group k-sparse subsets of $[n]$ is denoted by GkS. A vector $u \in \mathbb{R}^n$ is said to be **group k-sparse**, if its support set supp(u) is contained in a group k-sparse set. Set of all group k-sparse vector corresponding to set GkS is denoted by Σ_{GkS}.

In particular, every group k-sparse set is k-sparse i.e. has cardinality no larger than k, but the converse is not true.

One may ask why to put condition like that $\Lambda = G_S$? What will happen if this condition replaced by $\Lambda \subseteq G_S$? Answer to this is not very simple. If the stated condition is replaced by $\Lambda \subseteq G_S$, then the every k-sparse vector will become group sparse vector and hence, there will be no meaning of introducing the notion of group sparsity. Let us take an example to make the notion of group sparsity more clear. Suppose, $n = 6$, $k = 3$, $g = 4$, and that

$$G_1 = \{1\},\ G_2 = \{2\},\ G_3 = \{3\},\ G_4 = \{4, 5, 6\}.$$

Denote $\Lambda_1 = \{1, 2\},\ \Lambda_2 = \{1, 3\},\ \Lambda_3 = \{2, 3\},\ \Lambda_4 = \{1, 2, 3\},\ \Lambda_5 = \{4, 5, 6\},\ \Lambda_6 = G_1,\ \Lambda_7 = G_2,\ \Lambda_8 = G_3,\ \Lambda_9 = G_4$, then the set GkS is comprised of

$$\text{GkS} = \{\Lambda_1,\ \Lambda_2,\ \Lambda_3,\ \Lambda_4,\ \Lambda_5,\ \Lambda_6,\ \Lambda_7,\ \Lambda_8,\ \Lambda_9\}.$$

Suppose we are given vectors $v_1 = [0\ 0\ 0\ 5\ 0\ 3]$, $v_2 = [0\ 0\ 3\ 1\ 0\ 8]$. Observe that $\text{supp}(v_1) \subset \Lambda_5$, hence v_1 is group k-sparse as well as k-sparse. On the other hand,

$$\text{supp}(v_2) \nsubseteq \Lambda_i\ \forall i \in [9]$$

the above fact suggest that v_2 is not group k-sparse, but it can be easily seen that v_2 is k-sparse. This supports the fact that we stated above. For more details readers are directed to [4].

1.4.2 Group Restricted Isometry Property (GRIP)

Just like we defined Restricted Isometry Property (RIP) in case of conventional sparsity, a property called **Group Restricted Isometry Property (GRIP)** is introduced in this subsection which is meant for group sparsity model.

***Definition* 1.7:** A matrix $A \in \mathbb{R}^{m\times n}$ is said to satisfy the **group restricted isometry property (GRIP)** of order k with constant $\delta_k \in (0, 1)$ if

$$1 - \delta_k \leq \min_{\Lambda\in\text{GkS}} \min_{\text{supp}(z)\subseteq\Lambda} \frac{\|Az\|_2^2}{\|z\|_2^2} \leq \max_{\Lambda\in\text{GkS}} \max_{\text{supp}(z)\subseteq\Lambda} \frac{\|Az\|_2^2}{\|z\|_2^2} \leq 1 + \delta_k.$$

The set of group k-sparse vectors can be strictly smaller than the set of k-sparse vectors. Consequently, in general, the GRIP constant of order k can be smaller than the RIP constant of order k. When probabilistic methods are used to construct the measurement matrix A, often we require fewer measurements to achieve group sparse recovery than sparse recovery. See for example [4, Section 6]. This is why we study group sparsity.

Theorems in context to group sparse recovery is given in chapter 2.

1.4.3 Conventional Sparsity as a Special Case of Group Sparsity

Let $\mathcal{G} = \{G_1, G_2, ..., G_n\}$ is the partition set of index set $[n] = \{1, 2, 3, ..., n\}$ such that $|G_i| = 1$ for all $i \in [n]$. Then the set GkS contains all $\Lambda \subset [n]$ such that, $|\Lambda| \leq k$. Note that, for all $x \in \Sigma_k$,

$\Lambda = \text{supp}(x) \subset [n]$ with $|\text{supp}(x)| \leq k$, which implies $\Lambda \in$ GkS. So, with the above fact it is clear that conventional sparsity is a special case of group sparsity.

1.5 Application

Compressed sensing is present not only in theory but it has widespread applications, both in past, for example, in geophysical science (seismology) and promisingly in future. It has already contributed to a greater extent in the fields, space based imaging, radar design, surface metrology, genotyping, medical imaging, high-speed analog-to-digital conversion and ground-penetrating radar imaging in civil engineering, oil-exploration, geophysics, landmine detection, forensics, archeology etc. Several examples of its applications are discussed in details in the upcoming subsections.

1.5.1 Compressive Imaging

Acquiring images efficiently is one of the prominent application of compressed sensing. Most of the images we are interested in, are often sparse over some suitable basis like wavelet basis. So they meet the requirement of compressed sensing. Cameras that we use captures images with one sensor for every pixel and they acquire all the pixel before compressing the acquired data. Todays digital cameras have the mega pixel range which uses semiconductors as sensors.

One may ask why to acquire this many data as we have to throw many of them immediately? With the advances in compressive imaging, the sensors can directly acquire random linear measurements of an image while reducing the number of sensors required. With the knowledge of compressive sensing we can propose a guideline framework for implementing such an idea, including designing the measurement methods and the decoding algorithms. In fact the researchers have worked on designing such a system, for example, single pixel camera that the rice university has made [22]. This single pixel camera is consisting of a digital micro-mirror device (DMD), two lenses, a single photon detector and an analog-to-digital(A/D) converter. A DMD chip is made of 10 million oscillating mirrors or even more. We can command the action of each and everyone of these tiny (15 micrometer by 15 micrometer) mirrors. In other words, with the use of a proper set-up, every milliseconds, we can let each of these mirrors to shine on the detector. One of those lens focuses onto the DMD. Each of the mirror on DMD is for a pixel in the image, and we can make them to tilt toward or away from the second lens. Tilt towards the lens, is denoted by $+1$ and away the lens is denoted by -1. We can tell the set of mirrors on DMD to display a set of random tilting, that way a random set of mirrors are shining on second lens. This operation is analogous to creating inner products between the image (in vector form) and a vector containing elements 1 and -1. This light is then collected by the lens and focused onto the photon detector where the measurement is computed. If we do this once we obtain first compressed sensing measurement. We have to continue this process m times. The optical computer in the camera computes the random linear measurements of the image and passes those to a digital computers where it reconstructs the image. So, the single pixel Camera is very different from a tradition camera as it requires only one photon detector. One of the advantage of single pixel camera is that they can operate over a much broader range of light spectrum than a traditional camera. Sensors can be very expensive over some over some light spectrum, for example a tradition camera made for capturing infrared images would be very expensive and complicated. For more details about the working of single pixel camera readers are directed to [23].

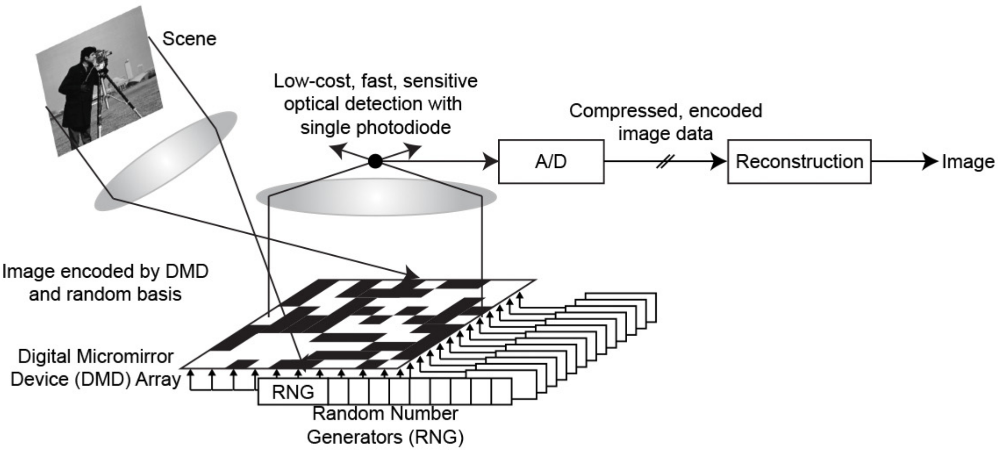

Figure 1.1: Diagram showing working of a single pixel camera.

Compressed sensing finds another important application in medical imaging, in particular in magnetic resonance imaging (MRI) which samples Fourier coefficients of an image. MR images are are found to be sparse. Some of them such as angiograms are sparse in their actual pixel representation, whereas some other more complicated MR images are sparse over some other basis, such as the wavelet basis or Fourier basis. As we all are aware about the fact that MRI are too much time costly, as we have to acquire a huge amount of data but the physical and physiological constraints don't allows us to do so in a short time window. Thus the our prime concern is to reduce the number of measurements without compromising on the quality of MR image or in other way, we can say that our concern is to increase the recovered image quality with the same number of measurements. Compressed sensing in MRI is a very hot topic now days and it has attracted a large number of researchers all over the world. For MRI application many compressive sampling algorithm have been designed [24].

1.5.2 Radar Signal Processing

Compressed sensing has got deep root in "Radar Signal processing. In a traditional radar system, radar transmits a kind of pulse form, after receiving it at the other end a matched filter is used to correlate the signal received with that pulse. Pulse compression system is used by the receiver together with a high-rate analog-to-digital (A/D) converter to process the signal. However, this approach is not only complicated and expensive, but also the resolution of the radar system is limited by radar uncertainty principle. Compressive Radar Imaging discretized the time-frequency plane into a grid. Each possible target scene is treated as a matrix. The occupations of the grids will be sparse if the number of targets is small enough, and then, compressive sensing techniques can be used to recover the target scene [25].

1.5.3 Biology

Compressive sensing can also be used efficiently in the area of biological applications. In fact, the idea of Group Testing is closely related to compressive sensing. Group testing was used in World War II to test soldiers for syphilis. Because the test for syphilis antigen in a blood sample was expensive, pathologist managed to group people and test the entire pool of blood samples for this group. If syphilis antigen was found in a pool of samples, the group was further divided into the subgroups, and then the test was done on each of the individual subgroup. Now a days, comparative DNA micro-array is a more modern example of compressive sensing idea in biology [26],[27]. Micro-arrays (DNA, protein, etc.) are colossal parallel affinity-based bio-sensors capable of detecting and computing a large number of different genomic particles simultaneously. Tens of thousands of probe spots found DNA micro-arrays are being used to test a number of targets in a single experiment. In micro-arrays, a large number of copies of a single probe is contained in a spot, designed to capture a single target, and hence collects only a single data point. However, only a chunk of the total number of genes represented by the reference sample and the test sample is differentially expressed in comparative DNA micro-array experiments. So we can use the compressed sensing ideas to create the compressed micro-arrays. In compressed micro-arrays, each spot encompass copies of different probes and the total number of spots is much smaller than the number of targets being tested. Application of compressed sensing can also be found in gene expression studies. For example, it will be a big achievement if one would be able to deduce the gene expression level of thousands of genes from only a limited number of observations [28].

1.5.4 Error Correcting

Compressive sensing also finds its applications in coding theory and practices. Error correction problem is an active research area in coding theory: In communication applications, when signals are received at the receiver, they usually get corrupted due to some random noise or disturbance. For example, in digital communications, onboard computations performed by circuits are real-valued. circuits usually experience disturbance caused by the numerous factors in the outside world. This is only one of the example of difficult real-world problem of error correction. So, It is a challenging problem that how to design system and decoding algorithms to correct the errors. As the errors usually occur in few places, compressed sensing tools can be applied to reconstruct the original signal from the corrupted data [8]. The mathematical formulation for the error correction problem is as follows:

Let us consider we are given a word "CRICKET". Assume that the code for the given word is in the form of a vector $x \in \mathbb{R}^m$, and we are asked to transmit it reliably to a remote receiver. Note that, We are not going to transmit x directly. We first encode the $x \in \mathbb{R}^m$ into a n-dimensional code y using a coding matrix B (also known as linear code). We can write y as

$$y = Bx, \text{ where } B \in \mathbb{R}^{n \times m}.$$

It is clear that in noiseless case, if B is a full rank matrix, we can get back x. But in real-world scenario, most often, we get the corrupted version of y due to the addition of some random sparse noise. So the receiver receives

$$y' = Bx + \eta, \tag{1.14}$$

where $\eta \in \mathbb{R}^n$ is some random noise. We then wish to recover the x from the corrupted received codeword $y = Bf + \eta$. To realize this as a compressed sensing problem, consider a matrix a matrix $A \in \mathbb{R}^{m\times n}$ whose null-space lies in the column space of B. On applying A to both sides of the equation (1.14), we get $Ay' = A\eta$. Set $Ay' = z$, and it becomes the problem of recovering a sparse vector η from its measurement $z = Ay'$. After obtaining the error vector η we subtract it from $y = Bx + \eta$, which gives us Bx. Since B is full rank matrix, we can get back original signal x. For more details readers are directed to [8].

Chapter 2

Literature Review

In the past, for sparse signal recovery, several condition on RIC constants have been proposed. For example, some sufficient conditions for sparse signal recovery in noiseless case are: $\delta_{2k} < \sqrt{2} - 1$ in [21], $\delta_{2k} < 0.497$ in [29], $\delta_{2k} < 0.472$ in[30], $\delta_k < 0.307$ in [31], $\delta_{2k} < 1/2$ and $\delta_k < 1/3$ in [32]. Some conditions that involves RIC of different orders have been introduced, e.g. $\delta_{3k} + 3\dagger_{4k} < 2$,$\delta_k + \dagger_{2k} < 1$ in [33],$\delta_{2k} < 0.5746$ together with $\delta_{2k} < 0.5746$, $\delta_{3k} < 0.7731$ together with $\delta_{16k} < 1$ in [34] and $\delta_{2k} < 4/\sqrt{41}$ in [35]. However, as proposed in [2], $\delta_{tk} < \sqrt{(t-1)/t}$ for some $t \geq 4/3$ is the latest bound on RIC available till date. In this chapter, first we will derive some of the bounds introduced in [2] and then we will discuss about some sufficient conditions introduced in [4] for group sparsity.

2.1 Sparse Representation of a Polytope and Recovery of Sparse Signals

In order to understand how the latest bound on RIP have been derived in [2], it is necessary to have knowledge of the key technical tool [2, Lemma 1.1], which represents a vector in polytope set by convex combination of sparse vectors. Proof of this key lemma is given in next subsection. Later we will modify this lemma for group sparsity case which will play a key role in obtaining the goal of this thesis.

2.1.1 Sparse Representation of a Polytope

For a positive number α and a positive integer k, define a polytope $T(\alpha, k) subset \mathbb{R}^n$, such that

$$T(\alpha, k) = \{v \in \mathbb{R}^n : \|v\|_\infty \leq \alpha,\ \|v\|_1 \leq \alpha k\}.$$

***Lemma* 2.1:** For any $v \in T(\alpha, k)$ define a set of sparse vector $U(\alpha, k, v) \subset \mathbb{R}^n$ such that,

$$\mathrm{U}(\alpha, k, v) = \{u \in \mathbb{R}^n : \mathrm{supp}(u) \subseteq \mathrm{supp}(v),\ \|u\|_1 = \|v\|_1,\ \|u\|_\infty \leq \alpha,\ \|u\|_0 \leq k\}\ ,$$

there exist an integer N such that v can be represented as

$$v = \sum_{i=1}^{N} \lambda_i \ u_i,$$

where

$$0 \le \lambda_i \le 1, \ \sum_{i=1}^{N} \lambda_i = 1, \ u_i \in \mathrm{U}(\alpha, k, v), \ \forall i \in [N].$$

Proof: We will prove this based on induction principle. Let l is a positive integer such that $l-1 \ge k$. Here we will prove that, if this lemma is true for all vectors $v \in T(\alpha, k)$ such that $|\mathrm{supp}(v)| = l-1$, then it also true for all l sparse vector present in $T(\alpha, k)$.

First, our assumption is that this lemma is true of all $(l-1)$-sparse vector in $T(\alpha, k)$. Let a vector $v \in T(\alpha, k)$ s.t. v is l-sparse but not $(l-1)$-sparse, otherwise there will be nothing to prove. We can express v as $v = \sum_{i=1}^{l} a_i \ e_i$ where $a_1 \ge a_2 \ge a_3 \ge ..... \ge a_l > 0$: e_i's are different unit vectors with only one non-zero entry of ± 1.

Define a set D as

$$D = \{1 \le i \le l-1 : a_i + a_{i+1} + a_{i+2} + ... + a_l \le (l-i)\alpha\}.$$

It is obvious that $1 \in D$ as, $a_1 + ... + a_l \le \alpha k \le (l-1)\alpha$. Let j is the largest element in D, which implies

$$a_j + a_{j+1} + ... + a_l \le (l-j)\alpha,$$

$$a_{j+1} + a_{j+2} + ... + a_l > (l-j-1)\alpha. \tag{2.1}$$

Define

$$b_w = \frac{\sum_{i=j}^{l} a_i}{l-j} - a_w, \ j \le w \le l. \tag{2.2}$$

It is not difficult to see that $(l-j)\sum_{i=j}^{l} b_i = \sum_{i=j}^{l} a_i$. By using (2.1), for all $j \le w \le l$,

$$\begin{aligned} b_w \ge b_j = \frac{\sum_{i=j}^{l} a_i}{l-j} - a_j =& \frac{\sum_{i=j+1}^{l} a_i}{l-j} - \frac{(l-j-1)a_j}{l-j} \\ =& \frac{\sum_{i=j+1}^{l} a_i - (l-j-1)a_j}{l-j} \\ >& \ 0. \end{aligned} \tag{2.3}$$

In addition, we define

$$\begin{aligned} v_w =& \sum_{i=1}^{j-1} a_i \ e_i + \sum_{i=j}^{l} b_i \sum_{i=j, i \ne w}^{l} e_i \ , \\ \lambda_w =& \frac{b_w}{\sum_{i=j}^{l} b_i}. \end{aligned} \tag{2.4}$$

Observe the following points:

- $v = \sum_{i=j}^{l} \lambda_w \ v_w$.

- $0 \leq \lambda_w \leq 1$.
- $\sum_{i=j}^{l} \lambda_i = 1$.
- $\text{supp}(v_w) \subset \text{supp}(v)$.
- $|\text{supp}(v_w)| = l - 1$.

We also have

$$\|v_w\|_1 = \sum_{i=1}^{j-1} a_i + (l-j) \sum_{w=j}^{l} b_w = \sum_{i=1}^{j-1} a_i + \sum_{i=j}^{l} a_i = \|v\|_1,$$

$$\begin{aligned}\|v_w\|_\infty &= \max\{a_1, ..., a_{j-1}, \sum_{i=j}^{l} b_i\} \\ &\leq \max\{\alpha, \frac{\sum_{i=j}^{l} a_i}{l-j}\} \leq \alpha.\end{aligned}$$

The last inequality is from the first part of (2.1). From the above fact it is quite clear that v_w is $(l-1)$-sparse and it belongs to the set $T(\alpha, k)$. We made an assumption that this lemma is true for all $(l-1)$-sparse vector, therefore we can find $\{u_{i,w} \in \mathbb{R}^n, \lambda_{i,w} \in \mathbb{R} \ : 1 \leq i \leq N_w, \ j \leq w \leq l\}$ such that

$$u_{i,w} \text{ is } k\text{-sparse}, \ \text{supp}(u_{i,w}) \subseteq \text{supp}(v_w) \subseteq \text{supp}(v), \ \|u_{i,w}\|_1 = \|v_w\|_1 = \|v\|_1, \ \|u_{i,w}\|_\infty \leq \alpha.$$

In addition, $v_w = \sum_{i=1}^{N_w} \lambda_{i,w} u_{i,w}$, so v can be represented as

$$v = \sum_{w=j}^{l} \sum_{i=1}^{N_w} \lambda_w \lambda_{i,w} u_{i,w}. \tag{2.5}$$

The result in (2.5) tells us that this lemma is true for all $v \in T(\alpha, k)$ that satisfies $|\text{supp}(v)| = l$. Based on induction principle we can state that this lemma is true for all $v \in T(\alpha, k)$. □

2.1.2 Exact Sparse Recovery

As we have already seen in subsection 1.3.1 that, ENSP enables "exact sparse signal recovery". In this subsection we will establish the fact that, if the measurement matrix $A \in \mathbb{R}^{m \times n}$ satisfies RIP of order tk with constant $\delta_{2k} < \sqrt{(t-1)/t}$, for some $t \geq 4/3$, then it will also satisfy **exact null space property (ENSP)** of order k. First we will begin with the following lemma,

Lemma 2.2: The number μ defined as $\mu = \sqrt{t(t-1)} - (t-1)$, for some $t > 1$, satisfies $\mu \in (0, \ 0.5)$.

Proof: As t is greater than 1, it is obvious that $\mu > 0$. In order to prove this lemma, we use obvious inequality that $\sqrt{1+q} < 1 + q/2$ for all $q > 0$. We have,

$$\begin{aligned}\mu =&\sqrt{t(t-1)} - (t-1) \\ =&\sqrt{(t-1+1)(t-1)} - (t-1) \\ =&(t-1)\Big[\sqrt{1 + \frac{1}{(t-1)}} - 1\Big]\end{aligned}$$

$$< (t-1)\frac{1}{2(t-1)} = 0.5\ .$$

Now we are going to introduce an important theorem which establishes the fact that, $\delta_{tk} < \sqrt{(t-1)/t}$ is a sufficient condition that should be satisfied by measurement matrix A in order to achieve exact sparse recovery of order k.

Theorem* 2.1*: Suppose $A \in \mathbb{R}^{m\times n}$, $x \in \Sigma_k \subset \mathbb{R}^n$, and $y = Ax$. Define recovery algorithm as in (1.3). Suppose A satisfies the RIP of order tk with constant $\delta_{tk} < \sqrt{(t-1)/t}$, for some $t \geq 4/3$ such that tk is an integer. Then the recovery algorithm in (1.3) will exactly recover x.

Proof: We only need to prove that, $\forall h \in \mathcal{N}(A)\backslash\{0\}$, it is true that $\|h_S\|_1 < \|h_{S^c}\|_1$, where S is the index set corresponding to k largest magnitude element of h. Proving this will prove the exact null space property (ENSP). However we will prove this by contradiction.

Assume there exist a vector $h \in \mathcal{N}(A)\backslash\{0\}$ such that, $\|h_S\|_1 \geq \|h_{S^c}\|_1$. Define,

$$\alpha := \|h_{S^c}\|_1/k.$$

Define index set S_1 and S_2 as,

$$S_1 := \{i \in S^c : |(h)_i| > \alpha/(t-1)\},$$

$$S_2 := \{i \in S^c : |(h)_i| \leq \alpha/(t-1)\}.$$

For brevity, denote $h_S,\ h_{S_1},\ h_{S_2}$ by $h_0,\ h_1\ h_2$ respectively. It is obvious that $\|h_1\|_1 \leq \|h_{S^c}\|_1 = \alpha k$. Let $\|h_1\|_0 = m$, then it follows that

$$\alpha k \geq \|h_1\|_1 > m\ \alpha/(t-1),\ \text{ or } m < k(t-1).$$

In addition we have

$$\begin{aligned} \|h_2\|_1 &= \|h_{S^c}\|_1 - \|h_1\|_1 \\ &\leq \alpha k - m\frac{\alpha}{(t-1)} \\ &= [k(t-1) - m]\ \frac{\alpha}{t-1} \qquad (2.6) \\ \|h_2\|_\infty &\leq \frac{\alpha}{(t-1)}. \qquad (2.7) \end{aligned}$$

Let $\alpha^* = \alpha/(t-1)$ and $p = k(t-1) - m$. On observing (2.6) and (2.7), we find that $h_2 \in T(\alpha^*, p)$. Using lemma 2.1, for some positive integer N, h_2 can be represented as

$$h_2 = \sum_{i=1}^{N} \lambda_i u_i.$$

Since, u_i are $(k(t-1) - m)$-sparse, by the known inequality

$$\|u_i\|_2 \leq \sqrt{\|u_i\|_0}\|u_i\|_\infty$$

$$\leq \sqrt{k(t-1)-m}\ \frac{\alpha}{(t-1)}$$
$$\leq \sqrt{k(t-1)}\ \frac{\alpha}{(t-1)}$$
$$\leq \sqrt{\frac{k}{(t-1)}}\ \alpha. \tag{2.8}$$

Let $\mu \geq 0$, $c = 1/2$ are two constants. Denote $\beta_i = h_0 + h_1 + \mu u_i$, then

$$\sum_{j=1}^{N} \lambda_j \beta_j - \frac{1}{2}\beta_i = h_0 + h_1 + \mu h_2 - \frac{1}{2}\beta_i$$
$$= (\frac{1}{2} - \mu)(h_0 + h_1) - \frac{1}{2}\mu u_i + \mu h$$
$$\sum_{j=1}^{N} \lambda_j \beta_j - \frac{1}{2}\beta_i - \mu h = (\frac{1}{2} - \mu)(h_0 + h_1) - \frac{1}{2}\mu u_i. \tag{2.9}$$

It is not difficult to see that vectors in equation (2.9) are tk-sparse.

We can check the following equation in l_2 norm,

$$\sum_{i=1}^{N} \lambda_i \|A(\sum_{j=1}^{N} \lambda_j \beta_j - \frac{1}{2}\beta_i)\|_2^2 \ - \ \sum_{i=1}^{N} \frac{\lambda_i}{4} \|A(\beta_i)\|_2^2 = 0. \tag{2.10}$$

Set $\mu = \sqrt{t(t-1)} - (t-1)$. Since A follows RIP, it follows that

$$\sum_{i=1}^{N} \lambda_i \|A(\sum_{j=1}^{N} \lambda_j \beta_j - \frac{1}{2}\beta_i)\|_2^2 \leq (1+\delta_{tk}) \sum_{i=i}^{N} \lambda_i \Big((\frac{1}{2} - \mu)^2 \|h_0 + h_1\|_2^2 + \frac{\mu^2}{4} \|u_i\|_2^2 \Big). \tag{2.11}$$

Using (2.9),(2.10), (2.11) we come to the conclusion that

$$0 \leq (1+\delta_{2k}) \sum_{i=i}^{N} \Big(\lambda_i (\frac{1}{2} - \mu)^2 \|h_0 + h_1\|_2^2 + \frac{\mu^2}{4} \|u_i\|_2^2 \Big)$$
$$- (1-\delta_{2k}) \sum_{i=1}^{N} \frac{\lambda_i}{4} \Big(\|h_0 + h_1\|_2^2 + \mu^2 \|u_i\|_2^2 \Big)$$
$$= \sum_{i=1}^{N} \lambda_i \Big[\Big((1+\delta_{2k})(\frac{1}{2} - \mu)^2 - (1-\delta_{2k})\frac{1}{4} \Big) \|h_0 + h_1\|_2^2 + \frac{1}{2}\delta_{tk}\mu^2 \|u_i\|_2^2 \Big]. \tag{2.12}$$

Using (2.7), we estabilish the fact that

$$\|u_i\|_2 \leq \sqrt{k/(t-1)}\ \alpha \leq \frac{\|h_0\|_2}{\sqrt{(t-1)}} \leq \frac{\|h_0 + h_1\|_2}{\sqrt{(t-1)}}. \tag{2.13}$$

On substituting (2.13) in (2.12), we get

$$\leq \sum_{i=1}^{N} \lambda_i \|h_0 + h_1\|_2^2 \Big[(\mu^2 - \mu)\delta_{tk} + \Big(\frac{1}{2} - \mu + (1 + \frac{1}{2(t-1)})\mu^2 \Big) \Big]$$

$$\leq \|h_0+h_1\|_2^2 \cdot \Big[\delta_{2k}\Big((2t-1)t - 2t\sqrt{t(t-1)}\Big) - \Big((2t-1)\sqrt{t(t-1)} - 2t(t-1)\Big)\Big]$$
$$< 0. \tag{2.14}$$

Above, we used fact that

$$\begin{aligned}\delta_{tk} &< \sqrt{(t-1)/t},\\ \|u_i\|_2 &\leq \sqrt{k/(t-1)}\alpha\\ &\leq \frac{\|h_0\|_2}{\sqrt{(t-1)}} \leq \frac{\|h_0+h_1\|_2}{\sqrt{(t-1)}}.\end{aligned} \tag{2.15}$$

Inequalities (2.12) and (2.14) are contradicting each other. Which implies

$$\|h_S\|_1 \not\geq \|h_{S^c}\|_1.$$

So we conclude that matrix A satisfies **exact null Space property**. Hence the recovery algorithm defined in (1.3) exactly recovers the k-sparse signal x. □

2.1.3 Robust Recovery of Signal

In the subsection 2.1.2 we proved that measurement matrix A in theorem 2.1 facilitates exact sparse recovery. However, in real life time most of the signals are nearly sparse instead of exactly sparse. In addition, there is always some amount of noise associated with the measurements. In this subsection we will prove that, matrix A in theorem 2.1 also facilitates robust sparse recovery.

Theorem 2.2: Suppose that, for some number $t \geq 4/3$ such that tk is an integer, the matrix $A \in \mathbb{R}^{m\times n}$ satisfies the RIP of order tk with constants $\delta_{tk} < \sqrt{(t-1)/t}$. Define recovery algorithm as in (1.4). Suppose, $x \in \mathbb{R}^n$ and that $y = Ax + \eta$ where $\|\eta\|_2 \leq \epsilon$. Then the recovery algorithm in (1.4) will lead to the bound

$$\|\hat{x}-x\|_2 \leq 2\frac{\sqrt{2(1+\delta_{tk})}}{1-\sqrt{t/(t-1)}\delta_{tk}}\epsilon + \left(\frac{\sqrt{2}\delta_{tk} + \sqrt{t(\sqrt{(t-1)/t}-\delta_{tk})\delta_{tk}}}{t(\sqrt{(t-1)/t}-\delta_{tk})} + 1\right)\frac{2\sigma_k(x, \|\cdot\|_1)}{\sqrt{k}}. \tag{2.16}$$

Proof: Suppose $h = \hat{x} - x$, first we will derive a widely known result (see, e.g.,[36],[37],[38],[39]),

$$\|h_{\Lambda_0^c}\|_1 \leq \|h_{\Lambda_0}\|_1 + 2\|x_{S_0^c}\|_1.$$

Where $S_0,\ \Lambda_0$ are the index set of k largest components by magnitude of $x,\ h$ respectively.

Optimality condition of $\hat{x}$ implies that

$$\|\hat{x}\|_1 = \|x+h\|_1 \leq \|x\|_1.$$

Using the decomposability of $\|\cdot\|_1$, we get

$$\|x_{S_0^c} + h_{S_0^c}\|_1 + \|x_{S_0} + h_{S_0}\|_1 \leq \|x_{S_0^c}\|_1 + \|x_{S_0}\|_1.$$

Applying triangle inequality twice to the left hand side of the above inequality. we get,

$$\|x_{S_0}\|_1 - \|h_{S_0}\|_1 - \|x_{S_0^c}\|_1 + \|h_{S_0^c}\|_1 \leq \|x_{S_0^c}\|_1 + \|x_{S_0}\|_1.$$

On canceling the common term and rearranging the above inequality, we get

$$\|h_{S_0^c}\|_1 \leq \|h_{S_0}\|_1 + 2\|x_{S_0^c}\|_1. \tag{2.17}$$

Observe that

$$\|h_{\Lambda_0}\|_1 \geq \ \|h_{S_0}\|_1, \text{ and } \|h_{\Lambda_0^c}\|_1 \leq \|h_{S_0^c}\|_1.$$

Using the above facts and (2.17), we get

$$\|h_{\Lambda_0^c}\|_1 \leq \|h_{\Lambda_0}\|_1 + 2\|h_{S_0^c}\|_1. \tag{2.18}$$

Bsides,

$$\|Ah\|_2 \leq \|A\hat{x} - y\|_2 + \|Ax - y\|_2 \leq 2\epsilon. \tag{2.19}$$

Define

$$\alpha = (\|h_{\Lambda_0}\|_1 + 2\|x_{S_0^c}\|_1)/k.$$

Now partition Λ_0^c as disjoint union $S_1 \cup S_2$, where

$$S_1 := \{i \in \Lambda_0^c : |(h)_i| > \alpha/(t-1)\},$$

$$S_2 := \{i \in \Lambda_0^c : |(h)_i| \leq \alpha/(t-1)\}.$$

For brevity, denote $h_{\Lambda_0},\ h_{S_1},\ h_{S_2}$ by $h_0,\ h_1,\ h_2$ respectively. Using (2.18) we can easily obtain

$$\|h_1\|_1 \leq \|h_{\Lambda_0^c}\|_1 \leq \alpha k.$$

Let $\|h_1\|_0 = m$. Then it follows that

$$\alpha k \geq \|h_1\|_1 > m\alpha/(t-1).$$

Which implies,

$$m < k(t-1).$$

Observe that, $\|h_0 + h_1\|_0 = k + m \leq tk$, it means $h_0 + h_1$ is tk-sparse. Then it follows that

$$\begin{aligned} \langle A(h_{\Lambda_0} + h_1), Ah\rangle &\leq \|A(h_{\Lambda_0} + h_1)\|_2 \|Ah\|_2 \\ &\leq 2\sqrt{1+\delta_{tk}}\|h_{\Lambda_0} + h_1\|_2\ \epsilon. \end{aligned} \tag{2.20}$$

It is not difficult to see that

$$\begin{aligned} \|h_2\|_1 &= \|h_{\Lambda_0^c}\|_1 - \|h_1\|_1 \\ &\leq \alpha k - m\frac{\alpha}{(t-1)} \end{aligned}$$

$$= [k(t-1)-m]\ \frac{\alpha}{t-1}, \tag{2.21}$$

$$\|h_2\|_\infty \le \frac{\alpha}{(t-1)}. \tag{2.22}$$

Let $\alpha^* = \alpha/(t-1)$ and $p = k(t-1)-m$. On observing (2.21) and (2.22), we find that $h_2 \in T(\alpha^*, p)$. Using lemma 2.1, for some positive integer N, h_2 can be represented as

$$h_2 = \sum_{i=1}^{N} \lambda_i u_i.$$

Since, u_i are $(k(t-1)-m)$-sparse, by the known inequality

$$\begin{aligned}
\|u_i\|_2 &\le \sqrt{\|u_i\|_0}\|u_i\|_\infty \\
&\le \sqrt{k(t-1)-m}\ \frac{\alpha}{(t-1)} \\
&\le \sqrt{k(t-1)}\ \frac{\alpha}{(t-1)} \\
&\le \sqrt{\frac{k}{(t-1)}}\ \alpha.
\end{aligned} \tag{2.23}$$

Let $1 \ge \mu \ge 0,\ c = 1/2$ and denote $\beta_i = h_0 + h_1 + \mu u_i$. We have already seen in subsection 2.1.2 that, $\sum_{j=1}^{N} \lambda_j \beta_j - \frac{1}{2}\beta_i - \mu h = (\frac{1}{2}-\mu)(h_0+h_1) - \frac{\mu}{2}u_i$ are tk sparse vectors.

Let $s = \|h_0 + h_1\|_2,\ P = 2\|x_{S_0^c}\|_1/\sqrt{k}$, then

$$\begin{aligned}
\|u_i\|_2 &\le \sqrt{k/(t-1)}\ \alpha \\
&\le \frac{\|h_{\Lambda_0} + h_1\|_2}{\sqrt{(t-1)}} + 2\frac{\|x_{\Lambda_{0*}^c}\|_1}{\sqrt{k(t-1)}} \\
&\le \frac{s+P}{\sqrt{t-1}}.
\end{aligned} \tag{2.24}$$

Set $\mu = \sqrt{t(t-1)} - (t-1)$. Using the identity (2.10)

$$\begin{aligned}
0 &= \sum_{i=1}^{N} \lambda_i \Big\| A\Big((\frac{1}{2}-\mu)(h_0+h_1) - \frac{\mu}{2}u_i + \mu h\Big)\Big\|_2^2 - \sum_{i=1}^{N} \frac{\lambda_i}{4}\|A(h_0+h_1+\mu u_i)\|_2^2 \\
&= \sum_{i=1}^{N} \lambda_i \Big\| A\Big((\frac{1}{2}-\mu)(h_0+h_1) - \frac{\mu}{2}u_i\Big)\Big\|_2^2 + 2\Big\langle A\Big((\frac{1}{2}-\mu)(h_0+h_1) - \frac{\mu}{2}h_2\Big),\ \mu A h\Big\rangle \\
&\quad + \mu^2\|Ah\|_2^2 - \sum_{i=1}^{N} \frac{\lambda_i}{4}\|A(h_0+h_1+\mu u_i)\|_2^2 \\
&= \sum_{i=1}^{N} \lambda_i \Big\| A\Big((\frac{1}{2}-\mu)(h_0+h_1) - \frac{\mu}{2}u_i\Big)\Big\|_2^2 + \mu(1-\mu)\langle A(h_0+h_1),\ Ah\rangle \\
&\quad - \frac{\lambda_i}{4}\|A(h_0+h_1+\mu u_i)\|_2^2.
\end{aligned}$$

Now observe that, $(\frac{1}{2}-\mu)(h_0+h_1) - \frac{\mu}{2}u_i,\ h_0+h_1+\mu u_i$ are tk sparse vector. Hence, by using RIP property and (2.20), we have that

$$
\begin{aligned}
0 &\le (1+\delta_{tk})\sum_{i=i}^{N}\lambda_i\Big((\frac{1}{2}-\mu)^2\|h_0+h_1\|_2^2+\frac{\mu^2}{4}\|u_i\|_2^2\Big)+\mu(1-\mu)2\sqrt{1+\delta_{tk}}\|h_0+h_1\|_2\ \epsilon\\
&\quad-(1-\delta_{tk})\sum_{i=1}^{N}\frac{\lambda_i}{4}\Big(\|(h_0+h_1)\|_2^2+\mu^2\|u_i\|_2^2\Big)\\
&=\sum_{i=1}^{N}\lambda_i\Big[\Big((1+\delta_{tk})(\frac{1}{2}-\mu)^2-(1-\delta_{tk})\frac{1}{4}\Big)\|h_0+h_1\|_2^2+\frac{1}{2}\delta_{tk}\mu^2\|u_i\|_2^2\Big]\\
&\quad+2\mu(1-\mu)\sqrt{1+\delta_{tk}}\|h_0+h_1\|_2\ \epsilon\\
&\le\sum_{i=1}^{N}\lambda_i\|h_0+h_1\|_2^2\Big[(\mu^2-\mu)\delta_{tk}+\Big(\frac{1}{2}-\mu+(1+\frac{1}{2(t-1)})\mu^2\Big)\Big]\ s^2\\
&\quad+\Big[2\mu(1-\mu)\sqrt{1+\delta_{tk}}\ \epsilon+\frac{\delta_{tk}\mu^2P}{t-1}\Big]s+\frac{\delta_{tk}\mu^2P^2}{2(t-1)}\\
&=-t\Big((2t-1)-2\sqrt{t(t-1)}\Big)\Big(\sqrt{\frac{(t-1)}{t}}-\delta_{2k}\Big)\ s^2+\Big[2\mu^2\sqrt{\frac{t-1}{t}}\cdot\sqrt{1+\delta_{tk}}\epsilon+\frac{\delta_{tk}\mu^2P}{t-1}\Big]\ s\\
&\quad+\frac{\delta_{tk}\mu^2P^2}{2(t-1)}\\
&=\frac{\mu^2}{t-1}\Big[-t\Big(\sqrt{\frac{t-1}{t}}-\delta_{tk}\Big)s^2+\Big(2\sqrt{t(t-1)(1+\delta_{tk})}\epsilon+\delta_{tk}P\Big)s+\frac{\delta_{tk}P^2}{2}\Big].
\end{aligned}
\tag{2.25}
$$

The above one is the second-order inequality for s. On solving (2.25) for s, we get

$$
\begin{aligned}
s &\le\Big\{\Big(2\sqrt{t(t-1)(1+\delta_{tk})}\epsilon+\delta_{tk}P\Big)+\Big[\Big(2\sqrt{t(t-1)(1+\delta_{tk})}\epsilon+\delta_{tk}P\Big)^2\\
&\quad+2t(\sqrt{(t-1)/t}-\delta_{tk})\delta_{tk}P^2\Big]^{1/2}\Big\}\cdot\Big(2t(\sqrt{(t-1)/t}-\delta_{tk})\Big)^{-1}\\
&\le 2\frac{\sqrt{t(t-1)(1+\delta_{tk})}}{t(\sqrt{(t-1)/t}-\delta_{tk})}\epsilon+\frac{\sqrt{2}\delta_{tk}+\sqrt{t(\sqrt{(t-1)/t}-\delta_{tk})\delta_{tk}}}{2t(\sqrt{(t-1)/t}-\delta_{tk})}\cdot P.
\end{aligned}
$$

Finally, note that $\|h_{\Lambda_0^c}\|_1\le\|h_0\|_1+P\sqrt{k}$, by [40, Lemma 5.3], we obtain $\|h_{\Lambda_0^c}\|_2\le\|h_0\|_1+P$, so

$$
\begin{aligned}
\|h\|_2 &=\sqrt{\|h_0\|_2^2+\|h_{\Lambda_0^c}\|_2^2}\\
&\le\sqrt{\|h_0\|_2^2+(\|h_0\|_2+P)^2}\\
&\le\sqrt{2\|h_0\|_2^2}+P\\
&\le\sqrt{2}\ s+P\\
&\le 2\frac{\sqrt{2t(t-1)(1+\delta_{tk})}}{t(\sqrt{(t-1)/t}-\delta_{tk})}\epsilon+\Big(\frac{\sqrt{2}\delta_{tk}+\sqrt{t(\sqrt{(t-1)/t}-\delta_{tk})\delta_{tk}}}{t(\sqrt{(t-1)/t}-\delta_{tk})}+1\Big)\frac{2\sigma_k(x,\|\cdot\|_1)}{\sqrt{k}}.
\end{aligned}
$$

As $P=\|x_{S_0^c}\|_1/\sqrt{k}=\sigma_k(x,\|\cdot\|_1)/\sqrt{k}$, we get

$$\|h\|_2 \le 2\frac{\sqrt{2(1+\delta_{tk})}}{1-\sqrt{t/(t-1)}\delta_{tk}}\epsilon + \left(\frac{\sqrt{2}\delta_{tk} + \sqrt{t(\sqrt{(t-1)/t} - \delta_{tk})\delta_{tk}}}{t(\sqrt{(t-1)/t} - \delta_{tk})} + 1\right)\frac{2\sigma_k(x, \|\cdot\|_1)}{\sqrt{k}}. \tag{2.26}$$

Which proves our theorem. □

Remarks: We made an assumption that tk is an integer. But in case, if tk is not an integer, take $t' = \lceil tk \rceil / k$. Note that $t' > t$, $t'k$ is an integer. Then we can choose $\delta_{t'k}$ as

$$\delta_{t'k} = \delta_{tk} < \sqrt{\frac{t-1}{t}} < \sqrt{\frac{t'-1}{t'}}.$$

Now we introduce a theorem which tells us that the bound $\delta_{tk} < \sqrt{(t-1)/t}$ is indeed a sharp bound. We are not going to give proof of it here, but the interested readers are directed [2].

Theorem* 2.3*: (See [2, Theorem 2.2]) Let $t \ge 4/3$. For all $\gamma > 0$ and all $k \ge 5/\gamma$, there exists a matrix A satisfying the RIP of order tk with constant $\delta_{tk} \le \sqrt{(t-1)/t} + \gamma$ such that the recovery procedure in (1.4) fails for some k-sparse vector.

In this section we came to know that $\delta_{tk} < \sqrt{(t-1)/t}$ for some $t \ge 4/3$ is a sufficient condition which should be satisfied by measurement matrix A in order to achieve both sparse recovery as well as robust recovery of order k through the constrained ℓ_1-minimization.

Chapter 3

My Contribution

The paper [4] provides upper bound on restricted isometry constant (δ_{2k}) in addition to error estimates for the group sparse recovery. It also includes several algorithms that have been proposed in the literature for both "conventional" as well as group sparsity. In order to achieve robust sparse recovery when the results proved in [4] are specialized to the case of conventional sparsity and ℓ_1-norm minimization, we get the bound $\delta_{2k} < \sqrt{2} - 1$. However, by using the *Theorem* 2.2, in order to achieve robust sparse recovery, $\delta_{2k} < 1/\sqrt{2}$ is found to be a sufficient condition. It means *Theorem* 2.2 gives a tight bound. This suggests that the method of proof adopted in [4] can be improved. That is precisely the purpose of the present work. Therefore, our objective is to establish bounds on restricted isometry constant and error estimates for group sparse recovery through constrained ℓ_1-norm minimization, which include the bounds of [2] as a special case in the case of conventional sparsity. Moreover, in the process of proving bounds for group sparse recovery, we also improve upon the error estimates given in [2].

3.1 Preliminaries

Before moving to the main results of this chapter we introduce few definitions.

***Definition* 3.1:** Given a vector $x \in \mathbb{R}^n$ and a group k-sparse set GkS. Set $\mathrm{GkS}_0 = \mathrm{GkS}$, and define the set Λ_0 as

$$\Lambda_0 = \operatorname*{argmin}_{\Lambda \in \mathrm{GkS}_0} \|x - x_\Lambda\|_1.$$

Now, for $i \geq 1$, we define an iterative algorithm as:

$$\mathrm{GkS}_i := \{\Lambda : \Lambda \cap \Lambda_{i-1} = \phi \ , \forall \Lambda \in \mathrm{GkS}_{i-1}\},$$

$$\Lambda_i := \operatorname*{argmin}_{\Lambda \in \mathrm{GkS}_i} \|x - \sum_{j=0}^{i-1} x_{\Lambda_j} - x_\Lambda\|_1.$$

Iterate the above algorithm for s number of times such that $\mathrm{GkS}_{s+1} = \phi$. Finally, we obtain the set $\{\Lambda_0,\ \Lambda_1, ...,\ \Lambda_s\}$. Then the set of vector $\{x_{\Lambda_0}, x_{\Lambda_1}, ...x_{\Lambda_s}\}$ is known as **optimal group $k-$sparse decomposition** of x.

***Definition* 3.2:** Given an integer k, let GkS denote the collection of all group k-sparse subsets of $[n]$, and define

$$\sigma_{k,\mathcal{G}}(x, \|\cdot\|) := \min_{\Lambda \in \text{GkS}} \|x - x_\Lambda\| = \min_{\Lambda \in \text{GkS}} \|x_{\Lambda_0^c}\| \tag{3.1}$$

to be the **group k-sparsity index** of the vector x with respect to the norm $\|\cdot\|$ and the group structure $\mathcal{G}$. In group sparsity, group k-sparsity index is analogous to k-sparsity index defined in conventional sparsity.

It is obvious that if $g = n$, and each group G_i is the singleton set $\{i\}$, then group sparsity and group sparsity index reduce respectively to k-sparsity and k-sparsity index. Note that, because GkS is in general a strict subset of the set of all k-sparse sets, it follows that

$$\sigma_k(x, \|\cdot\|) \leq \sigma_{k,\mathcal{G}}(x, \|\cdot\|). \tag{3.2}$$

***Definition* 3.3:** Given a vector $v \in \mathbb{R}^n$, we define the **group support set of** v, denoted by $\text{Gsupp}(v)$, as

$$\text{Gsupp}(v) := \{j \in [g] : v_{G_j} \neq 0\}. \tag{3.3}$$

Thus $\text{Gsupp}(v)$ denotes the subset of the groups on which v has a nonzero support.

***Definition* 3.4:** Suppose $A \in \mathbb{R}^{m \times n}$, known as the measurement map, and $\Delta : \mathbb{R}^m \to \mathbb{R}^n$, known as the demodulation map. Then the pair (A, Δ) is said to achieve **robust group sparse recovery** of order k if there exist constants D_1, D_2 such that, for all $\eta \in \mathbb{R}^m$ with $\|\eta\|_2 \leq \epsilon$, we have that

$$\|\Delta(Ax + \eta) - x\|_2 \leq D_1 \sigma_{k,\mathcal{G}}(x, \|\cdot\|_1) + D_2 \epsilon. \tag{3.4}$$

3.2 Polytope Decomposition Lemma

The key to the results in section 2.1 is *Lemma* 2.1, which represents a vector in polytope set by convex combination of sparse vectors. In this subsection we generalize this lemma to the case of group sparsity. Before presenting the lemma, we introduce a couple of terms.

We define $m_{\max}$ and $m_{\min}$ as

$$m_{\max} := \max_{j \in [g]} |G_j|, m_{\min} := \min_{j \in [g]} |G_j|. \tag{3.5}$$

Recall that, by assumption group sparsity, $m_{\max} \leq k$.

***Lemma* 3.1:** Given a vector $v \in \mathbb{R}^n$ and some $\alpha \in \mathbb{R}_+$ such that,

$$\|v_{G_j}\|_1 \leq \alpha, \ \forall j \in [g], \text{ and } \|v\|_1 \leq s\alpha \tag{3.6}$$

for some integer s, there an integer N and exist vectors $u_i, i \in [N]$ such that

- $\text{supp}(u_i) \subseteq \text{supp}(v), \ \forall i \in [N]$.
- $\|u_i\|_1 = \|v\|_1, \ \forall i \in [N]$.
- $\|(u_i)_{G_j}\|_1 \leq \alpha, \ \forall i \in [N], \text{ and } \ \forall j \in [g]$

- u_i is group $sm_{\max}$-sparse for each i, and finally
- v is a convex combination of $u_i, i \in [N]$.

Remarks: In the case of conventional sparsity, $m_{\max} = 1$, in which case all vectors u_i are s-sparse, which is precisely *Lemma* 2.1.

Proof: We will prove this *Lemma* by using the induction principle. Define a set of vectors

$$X := \{v \in \mathbb{R}^n : \|v_{G_j}\|_1 \leq \alpha \; \forall j \in [g], \|v\|_1 \leq s\alpha\}.$$

To begin the inductive process, suppose $|\mathrm{Gsupp}(v)| \leq s$. Then v is itself $sm_{\max}$-sparse. So we can take $N = 1$ and $u_1 = v$. Now suppose that the lemma is true for all $v \in X$ such that $|\mathrm{Gsupp}(v)| = r - 1$ where $r - 1 \geq s$. It is shown that the lemma is also true for all $v \in X$ satisfying $|\mathrm{Gsupp}(v)| = r$.

Let a vector $v \in X$ such that, $|\mathrm{Gsupp}(v)| = r$. suppose $Q \subseteq [g]$ denote the index set $\{j \in [g] : v_{G_j} \neq 0\}$, and observe that $|Q| = |\mathrm{Gsupp}(v)| = r$ by assumption. Then v can be expressed as

$$v = \sum_{j \in Q} v_{G_j}.$$

Now arrange the vectors v_{G_j} in decreasing order of their ℓ_1-norm. Denote the permuted vectors as p_1 through p_r.

Define,

$$a_i := \|p_i\|_1, \quad \text{and } \hat{p}_i = (1/a_i)p_i.$$

Then each $\hat{p}_i$ has unit ℓ_1-norm. Moreover $a_i \geq a_{i+1}$ for all i, and $v = \sum_{i=1}^r p_i = \sum_{i=1}^r a_i \hat{p}_i$. Also, because the ℓ_1-norm is decomposable and the p_i have non overlapping support sets, it follows that $\|v\|_1 = \sum_{i=1}^r a_i$.

Now define a set

$$D := \{\beta \in [r-1] : \sum_{i=\beta}^{r} a_\beta \leq (r - \beta)\alpha\}.$$

Then $1 \in D$ because

$$\sum_{i=1}^{r} a_i = \|v\|_1 \leq s\alpha \leq (r-1)\alpha.$$

Therefore D is nonempty. Now, by a slight abuse of notation, let β again denote the largest element of the set D. This implies that

$$a_\beta + a_{\beta+1} + a_{\beta+2} + ... + a_r \leq (r - \beta)\alpha,$$

$$a_{\beta+1} + a_{\beta+2} + ... + a_r > (r - \beta - 1)\alpha. \tag{3.7}$$

Define the constants,

$$b_w = \frac{\sum_{i=\beta}^r a_i}{r - \beta} - a_w, \qquad \beta \leq w \leq r,$$

which satisfies

$$\sum_{i=\beta}^{r} a_i = (r - \beta) \sum_{i=\beta}^{r} b_i. \tag{3.8}$$

For all $\beta \leq w \leq r$,

$$\begin{aligned} b_w \geq b_\beta = & \frac{\sum_{i=\beta+1}^{r} a_i}{r-\beta} - \frac{r-\beta-1}{r-\beta} a_\beta \\ \geq & \frac{\sum_{i=\beta+1}^{r} a_i - (r-\beta-1)\alpha}{r-\beta} > 0, \end{aligned}$$

where the last two steps follow from $a_i \leq \alpha$ for all i, and from the second inequality in (3.7).

Next, for $w = \beta, ..., r$, define

$$v^{(w)} := \sum_{i=1}^{\beta-1} a_i \hat{p}_i + (\sum_{i=\beta}^{r} b_i) \sum_{i=\beta, i\neq w}^{r} \hat{p}_i, \quad \lambda_w := \frac{b_w}{\sum_{i=\beta}^{r} b_w}. \tag{3.9}$$

Now observe that

$$0 < \lambda_w < 1, \qquad \sum_{w=\beta}^{r} \lambda_w = 1.$$

Next, $\text{supp}(v^{(w)}) \subseteq \text{supp}(v)$ for all w. Moreover, $|\text{Gsupp}(v^{(w)})| \leq r-1$ for all w, because the corresponding term $\hat{p}_w$ is missing from the summation in (3.9). Also, note that each $\hat{p}_i$ has unit ℓ_1-norm. Therefore, for each w between β and r, we have that

$$\begin{aligned} \|v^{(w)}\|_1 &= \sum_{i=1}^{\beta-1} a_i + (r-\beta) \sum_{i=\beta}^{r} b_i \\ &= \sum_{i=1}^{\beta-1} a_i + \sum_{i=\beta}^{r} a_i = \sum_{i=1}^{r} a_i = \|v\|_1. \end{aligned} \tag{3.10}$$

Through (3.9) it can be easily seen that $v^{(w)}$ is composed of some linear combination of $\hat{p}_i$. Since, each $\hat{p}_i$ is supported over a particular G_j and $\hat{p}_i$ have non overlapping support set, it follows that

$$\begin{aligned} \|v_{G_j}^{(w)}\|_1 &\leq \max\{a_1, a_2, ..., a_{\beta-1}, \sum_{i=\beta}^{r} b_w\} \\ &\leq \max\{\alpha, \frac{\sum_{i=\beta}^{r} a_i}{r-\beta}\} \leq \alpha. \end{aligned} \tag{3.11}$$

So, the results in (3.10), (3.11) suggest that, each $v^{(w)} \in X$. By the inductive assumption, each $v^{(w)}$ has a convex decomposition as in the statement of the lemma. It follows that v is also a convex combination as in the statement of the lemma. This completes the inductive step.

We have already established the fact that this lemma is true for all vector $v \in X$ satisfying $|\text{Gsupp}(v)| \leq s$, therefore it is also true for all $v \in X$ satisfying $|\text{Gsupp}(v)| \leq s+1$ and so on, for all $v \in X$ satisfying $|\text{Gsupp}(v)| \leq g$. It is obvious that, $\forall v \in X$, $0 \leq |\text{Gsupp}(v)| \leq g$. Hence this lemma is true for all $v \in X$. This was the one way to understand the proof of this lemma. But if, it is still not clear then forget about the induction principle. There is no need to assume that this lemma is true for all $v \in X$ satisfying $|\text{Gsupp}(v)| = r-1$. Take any vector $v \in X$ satisfying $|\text{Gsupp}(v)| = r > s$. We can split it into $v^{(w)}$ as in (3.9). We have proved that $v^{(w)} \in X$. We also showed that $|\text{Gsupp}(v^{(w)})| = r-1$. If $r-1 = s$, it means we have achieved our goal. Because if this is the case, then $v^{(w)}$ will be group $sm_{\max}-$sparse. But if, $r-1$ is still greater than s, then we will be able to apply the same procedure to each $v^{(w)}$. Split $v^{(w)}$ it into $v^{(w_t)}$, where $t = 1, ..., N_1$. Note that $v^{(w_t)}$ will satisfy $|\text{Gsupp}(v^{(w_t)})| = r-2$ and $v^{(w_t)} \in X$. We shall repeat the same procedure unless we get vectors u_i, such that $|\text{Gsupp}(u_i)| = s$. □

Now we introduce a lemma which gives a bound on the euclidean distance of u_i for all $i \in [N]$.

***Lemma* 3.2:** Let $u_i, i \in [N]$ be the vectors in the convex combination of Lemma 3.1. Then

$$\|u_i\|_2^2 \leq \frac{sm_{\max}}{m_{\min}}\alpha^2, \ \forall i \in [N]. \tag{3.12}$$

Proof: Fix the index $i \in [N]$. Define the index set

$$B_i := \{j \in [g] : (u_i)_{G_j} \neq 0\}.$$

Let $c_i = |B_i|$. Because u_i is $(sm_{\max})$-sparse, it follows that $c_i \leq \frac{sm_{\max}}{m_{\min}}$. Moreover, by using *Lemma* 3.1 for each index $j \in B_i$, we have that

$$\|(u_i)_{G_j}\|_2 \leq \|(u_i)_{G_j}\|_1 \leq \alpha.$$

Now observe that

$$u_i = \sum_{j \in B_i} (u_i)_{G_j}.$$

Please note that, there are at most c_i terms in the above summation, and each term has Euclidean norm no larger than α. Hence, using this fact and knowing that $(u_i)_{G_j}$ have non overlapping support set, we have that

$$\begin{aligned} \|u_i\|_2^2 &= \sum_{\forall j \in B} ||(u_i)_{G_j}||_2^2 \\ &\leq \frac{sm_{\max}}{m_{\min}}\alpha^2, \end{aligned}$$

which completes the proof. □

As of now we have stated and proved some important *Lemma*, now we shall move on to main results.

3.3 Main Results

Suppose for some $t > 1$, $k(t-1)/m_{\max}$ is an integer. Suppose that A satisfies the GRIP of order tk with constant $\delta_{tk} = \delta$. To facilitate the statement and proof of the main theorem, we define a few constants.

$$\mu := \sqrt{(t-1)t} - (t-1), \tag{3.13}$$

$$a := [\mu(1-\mu) - \delta(0.5 - \mu + \mu^2)]^{1/2}, \tag{3.14}$$

$$b := \mu(1-\mu)\sqrt{1+\delta}, c := \left[\frac{\delta\mu^2 m_{\max}^2}{2(t-1)m_{\min}}\right]^{1/2}, \tag{3.15}$$

$$d := \mu(1-\mu)\sqrt{k(kt - m_{\max})}, \tag{3.16}$$

$$\rho := c/a, \tau := b\sqrt{k}/a^2. \tag{3.17}$$

Note that the alternate expression for a is

$$a = \frac{[(1-\delta) - (1-2\mu)^2(1+\delta)]^{1/2}}{2}.$$

Next we extend the so-called robust null space property (see *Definition* 1.4) to group sparsity.

***Definition* 3.5:** A matrix $A \in \mathbb{R}^{m\times n}$ is said to satisfy the ℓ_2 **group robust null space property (GRNSP)** with constants $\rho \in (0,1), \tau \in \mathbb{R}_+$, if, for all $h \in \mathbb{R}^n$ and all sets $S \in$ GkS, it is true that

$$\|h_S\|_2 \leq \frac{\rho}{\sqrt{k}}\|h_{S^c}\|_1 + \frac{\tau}{\sqrt{k}}\|Ah\|_2, \tag{3.18}$$

Now, using (3.18) we obtain one more important inequality:
By Schwarz' inequality,

$$\|h_S\|_1 \leq \sqrt{k}\|h_{S^c}\|_1. \tag{3.19}$$

On substituting (3.19) into (3.18), we get

$$\|h_S\|_1 \leq \rho\|h_{S^c}\|_2 + \tau\|Ah\|_2. \tag{3.20}$$

Now we come to the key result that allows us to establish robust group k-sparse recovery. Note that, even in the case of conventional sparsity, the following result is new.

***Theorem* 3.1:** Suppose that the matrix A satisfies the GRIP of order tk with constant $\delta_{tk} = \delta$. Where $t > 1$ such that $k(t-1)/m_{\max}$ is an integer. If δ satisfies

$$\delta < \mu(1-\mu)\left(\frac{\mu^2 m_{\max}^2}{2(t-1)m_{\min}} + 0.5 - \mu + \mu^2\right)^{-1}, \tag{3.21}$$

then A satisfies the ℓ_2 GRNSP with constants ρ, τ defined in (3.18).

Proof: Let $h_{\Lambda_0}, h_{\Lambda_1}, h_{\Lambda_2}, \ldots, h_{\Lambda_s}$ be an optimal group k-sparse decomposition of h. Now denote $h_{\Lambda_0^c} = h^*$. Define sets S_1 and S_2 as follows:

$$S_1 = \left\{ j : \|h^*_{G_j}\|_1 > m_{\max}\frac{\|h_{\Lambda_0^c}\|_1}{k(t-1)},\ \forall j \in [g] \right\},$$

$$S_2 = \left\{ j : \|h^*_{G_j}\|_1 \leq m_{\max}\frac{\|h_{\Lambda_0^c}\|_1}{k(t-1)},\ \forall j \in [g] \right\}.$$

Let $GS_1 = \cup_{j\in S_1} G_j$ and $GS_2 = \cup_{j \in S_2} G_j$. Now define

$$h^{(0)} = h_{\Lambda_0},\ h^{(1)} = h^*_{GS_1},\ h^{(2)} = h^*_{GS_2}.$$

Then we have

$$h_{\Lambda_0^c} = h^* = h^*_{GS_1} + h^*_{GS_2} = h^{(1)} + h^{(2)}.$$

Assume that $|S_1| = r$. Now we will establish upper bound on $\|h^{(2)}\|_1$ and $\|h^{(2)}_{G_j}\|_1$. Because of the definition of set S_1, it follows that

$$\|h^{(1)}\|_1 \geq rm_{\max}\frac{\|h_{\Lambda_0^c}\|_1}{k(t-1)}. \tag{3.22}$$

Moreover

$$\|h^{(2)}\|_1 = \|h_{\Lambda_0^c}\|_1 - \|h^{(1)}\|_1.$$

Using (3.22) we get

$$\begin{aligned}\|h^{(2)}\|_1 &\leq \|h_{\Lambda_0^c}\|_1 - rm_{\max}\frac{\|h_{\Lambda_0^c}\|_1}{k(t-1)} \\ &= \left[\frac{k(t-1)}{m_{\max}} - r\right] m_{\max}\frac{\|h_{\Lambda_0^c}\|_1}{k(t-1)}. \end{aligned} \tag{3.23}$$

By the definition of set S_2

$$\|h^{(2)}_{G_j}\|_1 \leq m_{\max}\frac{\|h_{\Lambda_0^c}\|_1}{k(t-1)}, \ \forall j \in [g]. \tag{3.24}$$

Remarks: We should note that $r \ngtr \frac{k(t-1)}{m_{\max}} - 1$. Because, if $r = \frac{k(t-1)}{m_{\max}}$, then $\|h^{(1)}\|_1 = \|h_{\Lambda_0^c}\|_1 + \Omega$ where, $\Omega \in (0, \infty)$. This is not possible because $h_{\Lambda_0^c} = h^{(2)} + h^{(2)}$, and in addition to that $h^{(1)}$, $h^{(2)}$ have non overlapping support set. This implies that $h^{(0)} + h^{(1)}$ is group $(kt - m_{\max})$-sparse, and hence they are also group kt-sparse.

From (3.23) and (3.24), we see that the vector $h^{(2)}$ satisfies the hypotheses of *Lemma* 3.1 with

$$\alpha = m_{\max}\frac{\|h_{\Lambda_0^c}\|_1}{k(t-1)}, s = \left[\frac{k(t-1)}{m_{\max}} - r\right].$$

Therefore we can apply *Lemma* 3.1 to $h^{(2)}$. So, $h^{(2)}$ can be represented as

$$h^{(2)} = \lambda_1 \ u_1 \ + \lambda_2 \ u_2 + \ldots + \lambda_N u_N,$$

where each u_i is group $(k(t-1) - rm_{\max})$-sparse, $h^{(1)}$ is group $(rm_{\max})$-sparse, and $h^{(0)}$ is group k-sparse. Therefore $u_i + h^{(1)} + \ h^{(0)}$, is group tk-sparse for each $i \in [N]$. Now let, for all $i \in [N]$,

$$x_i = \frac{1}{2}\left(h^{(0)} + h^{(1)}\right) + \ \frac{\mu}{2}u_i,$$

$$z_i = \frac{1-2\mu}{2}\left(h^{(0)} + h^{(1)}\right) - \ \frac{\mu}{2}u_i,$$

$$\gamma = x_i + z_i = (1-\mu)\left(h^{(0)} + h^{(1)}\right),$$

$$\beta_i = x_i - z_i = \mu\left(h^{(0)} + h^{(1)} + \ u_i\right).$$

Then

$$\begin{aligned}\sum_{i=1}^{N}\lambda_i\langle A\gamma, A\beta_i\rangle &= \left\langle A\gamma, A\sum_{i=1}^{N}\lambda_i\beta_i\right\rangle \\ &= \mu(1-\mu)\langle A(h^{(0)} + h^{(1)}), Ah\rangle. \end{aligned} \tag{3.25}$$

However,for each index set i, we have that

$$\langle A\gamma, A\beta_i\rangle = \langle Ax_i + Az_i, Ax_i - Az_i\rangle = \|Ax_i\|_2^2 - \|Az_i\|_2^2.$$

Therefore it follows that

$$\sum_{i=1}^{N} \lambda_i(\|Ax_i\|_2^2 - \|Az_i\|_2^2) = \mu(1-\mu)\left\langle A(h^{(0)} + h^{(1)}), Ah\right\rangle,$$

on rearranging

$$\begin{aligned}\sum_{i=1}^{N} \lambda_i\|Ax_i\|_2^2 &= \sum_{i=1}^{N} \lambda_i\|Az_i\|_2^2 \\ &+ \mu(1-\mu)\Big\langle A(h^{(0)} + h^{(1)})\ ,\ Ah\Big\rangle.\end{aligned}$$

Since $x_i,\ z_i,\ (h^{(0)} + h^{(1)})$ are all group tk-sparse, it follows from the GRIP that,

$$\begin{aligned}(1-\delta)\sum_{i=1}^{N} \lambda_i\|x_i\|_2^2 &\leq (1+\delta)\sum_{i=1}^{N} \lambda_i\|z_i\|_2^2 \\ &+ \mu(1-\mu)\left\langle A(h^{(0)} + h^{(1)}), Ah\right\rangle.\end{aligned}$$

Since $h^{(0)},\ h^{(1)}$ and u_i have disjoint support sets, it follows that, for all $i \in [N]$, we have

$$\|x_i\|_2^2 = 0.25\left(\|(h^{(0)} + h^{(1)})\|_2^2 + \mu^2\|u_i\|_2^2\right),$$

$$\|z_i\|_2^2 = 0.25\left[(1-2\mu)^2\|(h^{(0)} + h^{(1)})\|_2^2 + \mu^2\|u_i\|_2^2\right].$$

Substituting these relationships, multiplying both sides by 4, and noting that $\sum_{i=1}^{N} \lambda_i = 1$, leads to

$$\begin{aligned}(1-\delta)\left[\|(h^{(0)} + h^{(1)})\|_2^2 + \mu^2\sum_{i=1}^{N} \lambda_i\|u_i\|_2^2\right] \leq &(1+\delta)\left[(1-2\mu)^2\|(h^{(0)} + h^{(1)})\|_2^2 + \mu^2\sum_{i=1}^{N} \lambda_i\|u_i\|_2^2\right] \\ &+ 4\mu(1-\mu)\left\langle A(h^{(0)} + h^{(1)}), Ah\right\rangle,\end{aligned}$$

or upon rearranging,

$$\begin{aligned}\|(h^{(0)} + h^{(1)})\|_2^2[(1-\delta) - (1+\delta)(1-2\mu)^2] &\leq 2\delta\mu^2\sum_{i=1}^{N} \lambda_i\|u_i\|_2^2 \\ &+ 4\mu(1-\mu)\left\langle A(h^{(0)} + h^{(1)}), Ah\right\rangle. \qquad (3.26)\end{aligned}$$

Recall that

$$\alpha = m_{\max}\frac{\|h_{\Lambda_0^c}\|_1}{k(t-1)}, s = \left[\frac{k(t-1)}{m_{\max}} - r\right].$$

Substituting these values into (3.12), we get that

$$\begin{aligned} \|u_i\|_2^2 &\leq \frac{k(t-1) - rm_{\max}}{m_{\min}} m_{\max}^2 \frac{\|h_{\Lambda_0^c}\|_1^2}{k^2(t-1)^2} \\ &\leq \frac{k(t-1)}{m_{\min}} m_{\max}^2 \frac{\|h_{\Lambda_0^c}\|_1^2}{k^2(t-1)^2} \\ &= \frac{m_{\max}^2}{m_{\min}} \frac{\|h_{\Lambda_0^c}\|_1^2}{k(t-1)}. \end{aligned} \tag{3.27}$$

Substituting the above bound which is independent of i, into (3.26), we get

$$\begin{aligned} \|(h^{(0)} + h^{(1)})\|_2^2[(1-\delta) - (1+\delta)(1-2\mu)^2] &\leq \frac{2\delta\mu^2 m_{\max}^2}{m_{\min}} \frac{\|h_{\Lambda_0^c}\|_1^2}{k(t-1)} \\ &+ 4\mu(1-\mu)\left\langle A(h^{(0)} + h^{(1)}), Ah \right\rangle. \end{aligned} \tag{3.28}$$

Using the Schwartz's inequality and the fact that $h^{(0)} + h^{(1)}$ is group tk-sparse, the above inequality becomes

$$\begin{aligned} \|(h^{(0)} + h^{(1)})\|_2^2[(1-\delta) - (1+\delta)(1-2\mu)^2] &\leq \frac{2\delta\mu^2 m_{\max}^2}{m_{\min}} \frac{\|h_{\Lambda_0^c}\|_1^2}{k(t-1)} \\ &+ 4\mu(1-\mu)\sqrt{1+\delta_k}\|(h^{(0)} + h^{(1)})\|_2 \cdot \|Ah\|_2. \end{aligned}$$

Denote $\|(h^{(0)} + h^{(1)})\|_2$ by f and invoke the definition of the constants a, b, c from (3.14) and (3.15). This gives

$$4f^2a^2 \leq 4c^2 \frac{\|h_{\Lambda_0^c}\|_1^2}{k} + 4bf\|Ah\|_2,$$

or after dividing both the sides by 4 and rearranging,

$$f^2a^2 - bf\|Ah\|_2 \leq c^2 \frac{\|h_{\Lambda_0^c}\|_1^2}{k}.$$

The next step is to complete the square on left side of the above inequality.

$$f^2a^2 - bf\|Ah\|_2 + \frac{b^2}{4a^2}\|Ah\|_2^2 \leq \frac{b^2}{4a^2}\|Ah\|_2^2 + c^2 \frac{\|h_{\Lambda_0^c}\|_1^2}{k},$$

or equivalently,

$$\left[af - \frac{b}{2a}\|Ah\|_2\right]^2 \leq \frac{b^2}{4a^2}\|Ah\|_2^2 + c^2 \frac{\|h_{\Lambda_0^c}\|_1^2}{k}.$$

Taking the square root on both sides, and using the obvious inequality that $\sqrt{x^2 + y^2} \leq x + y$ whenever $x, y \geq 0$, leads to

$$af - (b/2a)\|Ah\|_2 \leq (b/2a)\|Ah\|_2 + c\frac{\|h_{\Lambda_0^c}\|_1}{\sqrt{k}},$$

$\sqrt{x^2 + y^2} \leq x + y$ whenever $x, y \geq 0$, leads to

$$af - (b/2a)\|Ah\|_2 \leq (b/2a)\|Ah\|_2 + c\frac{\|h_{\Lambda_0^c}\|_1}{\sqrt{k}},$$

or upon rearranging and replacing f by $\|(h^{(0)}+h^{(1)})\|_2$,

$$a\|(h^{(0)}+h^{(1)})\|_2 \leq (b/a)\|Ah\|_2 + c\frac{\|h_{\Lambda_0^c}\|_1}{\sqrt{k}}.$$

Dividing both the sides by a and observing that $h_{\Lambda_0} = h^{(0)}$ and

$$\|h^{(0)}\|_2 \leq \|(h^{(0)}+h^{(1)})\|_2,$$

we get

$$\begin{aligned}
\|h_{\Lambda_0}\|_2 &\leq \|(h^{(0)}+h^{(1)})\|_2 \leq \frac{b}{a^2}\|Ah\|_2 + \frac{c}{a}\frac{\|h_{\Lambda_0^c}\|_1}{\sqrt{k}} \\
&= \frac{b\sqrt{k}}{a^2\sqrt{k}}\|Ah\|_2 + \frac{c}{a}\frac{\|h_{\Lambda_0^c}\|_1}{\sqrt{k}}.
\end{aligned}$$

This inequality is of the form (3.18) with ρ, τ given as in (3.3). The only thing left to prove is that, if δ satisfies the bound (3.21), then $\rho < 1$. This is equivalent to $c/a < 1$, so it is enough to show that $c^2 < a^2$, that is

$$\frac{\delta\mu^2\ m_{\max}^2}{2(t-1)m_{\min}} < \mu(1-\mu) - \delta(0.5-\mu+\mu^2),$$

on rearranging

$$\delta\Big(\frac{\mu^2\ m_{\max}^2}{2(t-1)m_{\min}}\ +\ 0.5-\mu+\mu^2\Big) < \mu(1-\mu),$$

which implies

$$\delta < \Big(\frac{\mu^2\ m_{\max}}{2(t-1)m_{\min}}\ +\ 0.5-\mu+\mu^2\Big)^{-1}\ \mu(1-\mu).$$

Hence we finished the proof. □

3.3.1 Robust Group Sparse Recovery

Suppose we have the measurement of the type

$$y = Ax + \eta$$

where η is some bounded noise, with $\|\eta\|_2 \leq \epsilon$. Define the recovery algorithm (or, demodulation map Δ) as,

$$\Delta(y) := \hat{x} = \text{argmin}\,\|z\|_1 \ \text{ s.t. } \ \|y - Az\|_2 \leq \epsilon \tag{3.29}$$

Now, we introduce a theorem in context to the error estimate obtained by the above algorithm.

***Theorem* 3.2:** Suppose $A \in \mathbb{R}^{m\times n}$ satisfies GRIP of order tk with $\delta_{tk} = \delta$, where $t > 1$ such that $k(t-1)/m_{\max}$ is an integer. Then, for $p \in [1,2]$, the demodulation map (Δ) defined in (3.29) leads to the bound

$$\|\hat{x}-x\|_1 \leq \frac{2}{1-\rho}[(1+\rho)\sigma_{k,G} + \ 2\tau\epsilon] \tag{3.30}$$

$$\|\hat{x}-x\|_p \leq \frac{2}{1-\rho}\Big\{\Big[\frac{\rho}{k^{1-1/p}} + (1+\rho)\Big]\sigma_{k,G} + \Big(\frac{1}{k^{1-1/p}}+2\Big)\tau\epsilon\Big\}. \tag{3.31}$$

Proof: Let $\hat{x} = x + h$. Optimality condition implies that

$$\|\hat{x}\|_1 = \|x + h\|_1 \leq \|x\|_1.$$

Let $x_{S_0}, x_{S_1}, \ldots, x_{S_b}$ be an optimal group k-sparse decomposition of x. Then

$$\|x_{S_0^c} + h_{S_0^c}\|_1 + \|x_{S_0} + h_{S_0}\|_1 \leq \|x_{S_0^c}\|_1 + \|x_{S_0}\|_1.$$

Applying triangle inequality twice to the left hand side of the above inequality. we get,

$$\|x_{S_0}\|_1 - \|h_{S_0}\|_1 - \|x_{S_0^c}\|_1 + \|h_{S_0^c}\|_1 \leq \|x_{S_0^c}\|_1 + \|x_{S_0}\|_1.$$

Canceling the common term and denoting $\|x_{S_0^c}\|$ by $\sigma_{k,G}(x, \|\cdot\|_1) = \sigma_{k,G}$, we get

$$\|h_{S_0^c}\|_1 - \|h_{S_0}\|_1 \leq 2\sigma_{k,G}. \tag{3.32}$$

Let $h_{\Lambda_0}, h_{\Lambda_1}, \ldots, h_{\Lambda_s}$ be an optimal group k-sparse decomposition of h. Then

$$\|h_{\Lambda_0}\|_1 \geq \|h_{S_0}\|_1, \text{ and } \|h_{\Lambda_0^c}\|_1 \leq \|h_{S_0^c}\|_1.$$

Using the above fact and (3.32), we get

$$\|h_{\Lambda_0^c}\|_1 - \|h_{\Lambda_0}\|_1 \leq 2\sigma_{k,G}. \tag{3.33}$$

Now,

$$\|Ah\|_2 = \|(A\hat{x} - y) - (Ax - y)\|_2 \leq 2\epsilon.$$

Using the inequality (3.20) and the above fact, we have that

$$\|h_{\Lambda_0}\|_1 \leq \rho\|h_{\Lambda_0^c}\|_1 \; + \; 2\tau\epsilon. \tag{3.34}$$

Using (3.34) and (3.33), we get

$$\begin{bmatrix} 1 & -1 \\ -\rho & 1 \end{bmatrix} \begin{bmatrix} \|h_{\Lambda_0^c}\|_1 \\ \|h_{\Lambda_0}\|_1 \end{bmatrix} \leq \begin{bmatrix} 2\sigma_{k,G} \\ 2\tau\epsilon \end{bmatrix}.$$

Let the M denotes the coefficient matrix on the left hand side

$$M^{-1} = \begin{bmatrix} 1 & -1 \\ -\rho & 1 \end{bmatrix}^{-1} = \frac{1}{1-\rho} \begin{bmatrix} 1 & 1 \\ \rho & 1 \end{bmatrix}.$$

Since $1 - \rho > 0$, all the elements of M^{-1} are positive. Therefore we can multiply both the sides by M^{-1}. This gives

$$\begin{bmatrix} \|h_{\Lambda_0^c}\|_1 \\ \|h_{\Lambda_0}\|_1 \end{bmatrix} \leq \frac{1}{1-\rho} \begin{bmatrix} 1 & 1 \\ \rho & 1 \end{bmatrix} \begin{bmatrix} 2\sigma_{k,G} \\ 2\tau\epsilon \end{bmatrix}$$

$$\leq \frac{1}{1-\rho}\begin{bmatrix} 2(\sigma_{k,G}+\tau\epsilon) \\ 2(\rho\sigma_{k,G}+\tau\epsilon) \end{bmatrix}. \tag{3.35}$$

Finally Using the triangle inequality, we get

$$\begin{aligned} \|h\|_1 &\leq \|h_{\Lambda_0^c}\|_1 + \ \|h_{\Lambda_0^c}\|_1 \\ &= \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \|h_{\Lambda_0^c}\|_1 \\ \|h_{\Lambda_0}\|_1 \end{bmatrix} \\ &\leq \frac{2}{1-\rho}[(1+\rho)\sigma_{k,G} + \ 2\tau\epsilon] \end{aligned}$$

$$\|h\|_1 \leq \frac{2}{1-\rho}[(1+\rho)\sigma_{k,G} + \ 2\tau\epsilon]. \tag{3.36}$$

The above inequality is same as (3.30). Now we move on to prove the inequality (3.31). From the triangle inequality,

$$\|h\|_p \leq \|h_{\Lambda_0}\|_p + \ \|h_{\Lambda_0^c}\|_p. \tag{3.37}$$

Now we will obtain the upper bound for both of the terms in right hand side of (3.37). Using the triangle inequality it is not difficult to see that,

$$\|h_{\Lambda_0^c}\|_p \leq \|h_{\Lambda_0^c}\|_1 \leq \|h\|_1.$$

Using the above fact and (3.36), we get

$$\|h_{\Lambda_0^c}\|_p \leq \frac{2}{1-\rho}[(1+\rho)\sigma_{k,G} + \ 2\tau\epsilon]. \tag{3.38}$$

It is the ready consequence of $Holder's$ inequality that

$$\|h_{\Lambda_0}\|_p \leq k^{1/p-1/2} \ \|h_{\Lambda_0}\|_2. \tag{3.39}$$

Using the inequalities (3.39), (3.18), we get,

$$\|h_{\Lambda_0}\|_p \leq k^{1/p-1} \ (\rho\|h_{\Lambda_0^c}\|_1 + 2\tau\epsilon). \tag{3.40}$$

From inequality (3.35)

$$\|h_{\Lambda_0^c}\|_1 \leq \frac{2}{1-\rho}(\sigma_{k,G}+\tau\epsilon). \tag{3.41}$$

On substituting (3.41) into (3.40), we get

$$\|h_{\Lambda_0}\|_p \leq \frac{1}{k^{1-1/p}}\Big[\frac{2\rho}{1-\rho}(\sigma_{k,G}+\tau\epsilon) + \ 2\frac{1-\rho}{1-\rho}\tau\epsilon\Big],$$

on rearranging

$$\|h_{\Lambda_0}\|_p \leq \frac{1}{k^{1-1/p}}\frac{2}{(1-\rho)}\Big[\rho\sigma_{k,G}+\tau\epsilon\Big]. \tag{3.42}$$

Using inequalities (3.37),(3.42),(3.38), we get

$$\|h\|_p \leq \frac{2}{1-\rho}\Big\{\Big[\frac{\rho}{k^{1-1/p}} + (1+\rho)\Big]\sigma_{k,G} + \Big(\frac{1}{k^{1-1/p}} + 2\Big)\tau\epsilon\Big\}, \tag{3.43}$$

which completes the proof. □

Now we are going to obtain error bound for the case of dantzig selector noise.

Theorem 3.3: Suppose that, for some $t > 1$ such that $k(t-1)/m_{\max}$ is an integer, matrix $A \in \mathbb{R}^{m\times n}$ satisfies $GRIP$ of order tk. We have the measurement of the type $y = Ax + \eta$, where $\|A^t\eta\|_\infty \leq \zeta$. Define the demodulation map as

$$\Delta(y) = \hat{x}_{\text{DS}} := \operatorname*{argmin}_z \|z\|_1 \text{ s.t. } \|A^t(Az - y)\|_\infty \leq \zeta. \tag{3.44}$$

Define constants $a,\ c,\ d,\ \rho$ as in (3.14) to (3.17). Then the demodulation map defined in (3.44) leads to the bound

$$\|h\|_1 \leq \|h_{\Lambda_0^c}\|_1 + \|h_{\Lambda_0}\|_1 \leq \frac{2}{(1-\rho)}\Big[(1+\rho)\sigma_{k,G} + \frac{2d}{a^2}\zeta\Big]. \tag{3.45}$$

Proof: With the $\hat{x}_{\text{DS}}$ defined in (3.44), define $h := \hat{x}_{\text{DS}} - x$. Then the computations in the proof of *Theorem*3.1 continue to apply until (3.28). We have already shown that that $h^{(0)} + h^{(1)}$ is group $(kt - m_{\max})$-sparse. Also, both $\hat{x}_{\text{DS}}$ and x feasible for the optimization problem in (3.44). Therefore it follows that

$$\|A^tAh\|_\infty = \|A^t(y - Ax) - A^t(y - A\hat{x}_{\text{DS}})\|_\infty \leq 2\zeta.$$

Now we can write

$$\begin{aligned}\langle A(h^{(0)} + h^{(1)}), Ah\rangle &= \langle h^{(0)} + h^{(1)}, A^tAh\rangle \\ &\leq \|h^{(0)} + h^{(1)}\|_1\|A^tAh\|_\infty \\ &\leq 2\|h^{(0)} + h^{(1)}\|_2\sqrt{kt - m_{\max}}\cdot\zeta. \end{aligned} \tag{3.46}$$

Substituting from (3.46) into (3.28) gives

$$\begin{aligned}\|(h^{(0)} + h^{(1)})\|_2^2[(1-\delta) - (1+\delta)(1-2\mu)^2] \leq{}& \frac{2\delta\mu^2 m_{\max}^2}{m_{\min}}\frac{\|h_{\Lambda_0^c}\|_1^2}{k(t-1)} \\ &+ 8\mu(1-\mu)\zeta\sqrt{kt - m_{\max}}\cdot\|h^{(0)} + h^{(1)}\|_2,\end{aligned}$$

or upon rewriting

$$\begin{aligned}\|(h^{(0)} + h^{(1)})\|_2^2[(1-\delta) - (1+\delta)(1-2\mu)^2] \leq{}& \frac{2\delta\mu^2 m_{\max}^2}{m_{\min}}\frac{\|h_{\Lambda_0^c}\|_1^2}{k(t-1)} \\ &+ \frac{8\mu(1-\mu)\zeta\sqrt{k(kt - m_{\max})}}{\sqrt{k}}\cdot\|h^{(0)} + h^{(1)}\|_2.\end{aligned} \tag{3.47}$$

Recall now the definitions of the constants a from (3.14), c from (3.15), d from (3.16), and ρ from (3.17). Then, after dividing both the sides by 4 and denoting $\|h^{(0)}+h^{(1)}\|_2$ by f, the above inequality becomes

$$a^2 f^2 - \frac{2d\zeta}{\sqrt{k}} f \le c^2 \frac{\|h_{\Lambda_0^c}\|_1^2}{k}.$$

Completing the square and taking the square root on both sides leads to

$$\begin{aligned}
\Big(af - \frac{d\zeta}{a\sqrt{k}}\Big)^2 &\le \frac{d^2\zeta^2}{a^2 k} + c^2 \frac{\|h_{\Lambda_0^c}\|_1^2}{k}, \\
af - \frac{d\zeta}{a\sqrt{k}} &\le \Big[\frac{d^2\zeta^2}{a^2 k} + c^2 \frac{\|h_{\Lambda_0^c}\|_1^2}{k}\Big]^{1/2} \\
&\le \frac{d\zeta}{a\sqrt{k}} + c\frac{\|h_{\Lambda_0^c}\|_1}{\sqrt{k}},
\end{aligned} \tag{3.48}$$

and finally

$$\|h^{(0)} + h^{(1)}\|_2 = f \le \frac{2d}{a^2\sqrt{k}}\zeta + \frac{c}{a}\frac{\|h_{\Lambda_0^c}\|_1}{\sqrt{k}} = \frac{2d}{a^2\sqrt{k}}\zeta + \rho\frac{\|h_{\Lambda_0^c}\|_1}{\sqrt{k}}.$$

Next, by Schwarz' inequality, we get

$$\|h^{(0)}\|_1 \le \sqrt{k}\|h^{(0)}\|_2 \le \sqrt{k}\|h^{(0)} + h^{(1)}\|_2 \le \frac{2d}{a^2}\zeta + \rho\|h_{\Lambda_0^c}\|_1. \tag{3.49}$$

We should note that inequality (3.33) is still valid in this case. Then we have

$$\|h_{\Lambda_0^c}\|_1 - \|h_{\Lambda_0}\|_1 \le 2\sigma_{k,G}, \tag{3.50}$$

where it is to be noted that $h^{(0)}$ is same as h_{Λ_0}, and $\sigma_{k,G}$ is the shorthand for $\sigma_{k,G}(x, \|\cdot\|_1)$. Inequalities (3.49) and (3.50) can be written as

$$\begin{bmatrix} 1 & -1 \\ -\rho & 1 \end{bmatrix} \begin{bmatrix} \|h_{\Lambda_0^c}\|_1 \\ \|h_{\Lambda_0}\|_1 \end{bmatrix} \le \begin{bmatrix} 2\sigma_{k,G} \\ 2d\zeta/a^2 \end{bmatrix}.$$

Because $\rho < 1$, this leads to

$$\begin{bmatrix} \|h_{\Lambda_0^c}\|_1 \\ \|h_{\Lambda_0}\|_1 \end{bmatrix} \le \frac{1}{1-\rho} \begin{bmatrix} 1 & 1 \\ \rho & 1 \end{bmatrix} \begin{bmatrix} 2\sigma_{k,G} \\ 2d\zeta/a^2 \end{bmatrix}.$$

Combining these two shows that

$$\|h\|_1 \le \|h_{\Lambda_0^c}\|_1 + \|h_{\Lambda_0}\|_1 \le \frac{2}{(1-\rho)}\Big[(1+\rho)\sigma_{k,G} + \frac{2d}{a^2}\zeta\Big]. \tag{3.51}$$

□

Remarks: In the case of conventional sparsity, $m_{\max} = m_{\min} = 1$. If we put these values in the bound obtained for δ in (3.21), we get

$$\delta < \mu(1-\mu)\left(\frac{\mu^2}{2(t-1)} + 0.5 - \mu + \mu^2\right)^{-1}. \tag{3.52}$$

Now tedious but simple computation shows that

$$\mu(1-\mu) = (2t-1)\sqrt{t(t-1)} - 2t(t-1),$$

and

$$\frac{\mu^2}{2(t-1)} + 0.5 - \mu + \mu^2 = (2t-1)t - 2t\sqrt{t(t-1)} = \mu(1-\mu)\sqrt{\frac{t}{t-1}}.$$

Therefore,

$$\mu(1-\mu)\Big[\frac{\mu^2}{2(t-1)} + 0.5 - \mu + \mu^2\Big]^{-1} = \sqrt{\frac{t-1}{t}},$$

which implies that

$$\delta < \sqrt{\frac{t-1}{t}} \tag{3.53}$$

It is interesting to note that result obtained in (3.53) is same as proposed in [2]. We have seen in *Theorem* 2.3 that the bound, $\delta < \sqrt{(t-1)/t}$ is a sharp bound. So, this fact supports our claim that the approach followed in this work (for group sparsity) provides a tight bound.

References

[1] S. Foucart and H. Rauhut. A mathematical introduction to compressive sensing. Birkhäuser, 2013.

[2] T. T. Cai and A. Zhang. Sparse representation of a polytope and recovery of sparse signals and low-rank matrices. *Information Theory, IEEE Transactions on* 60, (2014) 122–132.

[3] M. Vidyasagar. An Introduction to Compressed Sensing. `http://www.iith.ac.in/~m_vidyasagar/CS-Notes.pdf` (2015).

[4] M. E. Ahsen and M. Vidyasagar. Error bounds for compressed sensing algorithms with group sparsity: A unified approach. *Applied and Computational Harmonic Analysis* –.

[5] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM review* 43, (2001) 129–159.

[6] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM journal on computing* 24, (1995) 227–234.

[7] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 267–288.

[8] E. J. Candes and T. Tao. Decoding by linear programming. *Information Theory, IEEE Transactions on* 51, (2005) 4203–4215.

[9] L. D. Donoho. High-Dimensional Centrally Symmetric Polytopes with Neighborliness Proportional to Dimension. *Discrete & Computational Geometry* 35, (2006) 617–652.

[10] R. M. Gray. IEEE Information Theory Society Newsletter .

[11] B. F. Logan. Properties of high-pass signals .

[12] H. L. Taylor, S. C. Banks, and J. F. McCoy. Deconvolution with the 1 norm. *Geophysics* 44, (1979) 39–52.

[13] A. C. Gilbert, S. Guha, P. Indyk, S. Muthukrishnan, and M. Strauss. Near-optimal sparse Fourier representations via sampling. In Proceedings of the thiry-fourth annual ACM symposium on Theory of computing. ACM, 2002 152–161.

[14] E. J. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *Information Theory, IEEE Transactions on* 52, (2006) 5406–5425.

[15] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on* 41, (1993) 3397–3415.

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

[17] D. Needell and R. Vershynin. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. *Foundations of computational mathematics* 9, (2009) 317–334.

[18] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck. Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *Information Theory, IEEE Transactions on* 58, (2012) 1094–1121.

[19] M. Fornasier and H. Rauhut. Iterative thresholding algorithms. *Applied and Computational Harmonic Analysis* 25, (2008) 187–208.

[20] D. Needell and J. A. Tropp. CoSaMP:,Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis* 26, (2009) 301–321.

[21] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique* 346, (2008) 589–592.

[22] D. Takhar, J. N. Laska, M. B. Wakin, M. F. Duarte, D. Baron, S. Sarvotham, K. F. Kelly, and R. G. Baraniuk. A new compressive imaging camera architecture using optical-domain compression. In Electronic Imaging 2006. International Society for Optics and Photonics, 2006 606,509–606,509.

[23] N. blanche. Working of single pixel camera. `http://nuit-blanche.blogspot.in/2007/07/how-does-rice-one-pixel-camera-work.html` (2007).

[24] M. Lustig, D. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic resonance in medicine* 58, (2007) 1182–1195.

[25] M. A. Herman and T. Strohmer. High-resolution radar via compressed sensing. *Signal Processing, IEEE Transactions on* 57, (2009) 2275–2284.

[26] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi. Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays. *Selected Topics in Signal Processing, IEEE Journal of* 2, (2008) 275–285.

[27] M. A. Sheikh, O. Milenkovic, and R. G. Baraniuk. Designing compressive sensing DNA microarrays. In Computational Advances in Multi-Sensor Adaptive Processing, 2007. CAMPSAP 2007. 2nd IEEE International Workshop on. IEEE, 2007 141–144.

[28] E. J. Candès et al. Compressive sampling. In Proceedings of the international congress of mathematicians, volume 3. Madrid, Spain, 2006 1433–1452.

[29] Q. Mo and S. Li. New bounds on the restricted isometry constant δ2k. *Applied and Computational Harmonic Analysis* 31, (2011) 460–468.

[30] T. T. Cai, L. Wang, and G. Xu. Shifting inequality and recovery of sparse signals. *Signal Processing, IEEE Transactions on* 58, (2010) 1300–1308.

[31] T. T. Cai, L. Wang, and G. Xu. New bounds for restricted isometry constants. *Information Theory, IEEE Transactions on* 56, (2010) 4388–4394.

[32] T. T. Cai and A. Zhang. Sharp RIP bound for sparse signal and low-rank matrix recovery. *Applied and Computational Harmonic Analysis* 35, (2013) 74–93.

[33] T. T. Cai and A. Zhang. Compressed sensing and affine rank minimization under restricted isometry. *Signal Processing, IEEE Transactions on* 61, (2013) 3279–3290.

[34] S. Zhou, L. Kong, and N. Xiu. New bounds for RIC in compressed sensing. *Journal of the Operations Research Society of China* 1, (2013) 227–237.

[35] J. Andersson and J.-O. Stromberg. On the theorem of uniform recovery of random sampling matrices. *Information Theory, IEEE Transactions on* 60, (2014) 1700–1710.

[36] T. T. Cai, G. Xu, and J. Zhang. On recovery of sparse signals via minimization. *Information Theory, IEEE Transactions on* 55, (2009) 3388–3397.

[37] E. J. Candes, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics* 59, (2006) 1207–1223.

[38] E. Candes and T. Tao. The Dantzig selector: statistical estimation when p is much larger than n. *The Annals of Statistics* 2313–2351.

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

[40] R. Baraniuk and P. Steeghs. Compressive radar imaging. In Radar Conference, 2007 IEEE. IEEE, 2007 128–133.