A Weighted l_1-Minimization for Distributed Compressive Sensing

by

Xiaowei Li

B.Sc., Peking University, 2013

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Mathematics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

September 2015

Abstract

Distributed Compressive Sensing (DCS) studies the recovery of jointly sparse signals. Compared to separate recovery, the joint recovery algorithms in DCS are usually more effective as they make use of the joint sparsity. In this thesis, we study a weighted l_1-minimization algorithm for the joint sparsity model JSM-1 proposed by Baron et al. Our analysis gives a sufficient null space property for the joint sparse recovery. Furthermore, this property can be extended to stable and robust settings. We also presents some numerical experiments for this algorithm.

Preface

This thesis is original, unpublished, independent work by the author, X. Li under the supervision of Dr. Ö. Yılmaz.

Table of Contents

List of Figures

Acknowledgements

I would like to thank Dr. Özgür Yılmaz for his guidance and support. I would also want to thank my parents for their love throughout the years.

Chapter 1

Introduction

1.1 Compressive Sensing with l_1-Minimization

Compressive Sensing (CS) is a modern technique for recovering sparse signals from a seemingly incomplete set of linear measurements. It exploits the sparsity assumption which leads to recovery algorithms that require far fewer samples than the ambient dimension of the signal to be recovered. One such algorithm is Basis Pursuit, which relies on l_1-minimization.

Mathematically, a vector (signal) x is sparse if most of its entries are zero and we say it is s-sparse if $|\{i : x(i) \neq 0\}| \leq s$. The set of all s-sparse vectors in the ambient dimension is denoted as Σ_s.

In practice, most signals we encounter are not necessarily sparse but can be well approximated by sparse signals. Such signals are called compressible signals and are usually modeled as signals whose k-th largest coefficient $x_{(k)}$ obeys a power decay such as

$$|x_{(k)}| \leq Ck^{-1/q}. \tag{1.1}$$

To measure how close a signal can be approximated by sparse vectors, we can define the best s-term approximation error to x in l_p by

$$\sigma_s(x)_p = \inf_{z \in \Sigma_s} \|x - z\|_p, \quad p > 0.$$

Note that $\sigma_s(x)_p = 0$ if x is s-sparse. For compressible signals as in (1.1), we can show that for $p > q > 0$,

$$\sigma_s(x)_p \leq C \left(\frac{p}{q} - 1\right)^{-1/p} s^{1/p-1/q}.$$

Given a compressible signal $x \in \mathbb{R}^N$, sparse recovery aims to recover a sparse approximation to x with only m measurements where $m \ll N$.

Note that this problem is ill-posed if sparsity is not assumed. Ideally, each measurement is a linear function of x taking value in $\mathbb{R}$, but in reality, noise is also added each time we take a measurement. Writing this in matrix form, we have

$$y = Ax + e,$$

where $A \in \mathbb{R}^{m \times N}$ is the sampling matrix, $e \in \mathbb{R}^m$ is additive noise and $y \in \mathbb{R}^m$ is the final measurements. Each row in this equation represents one measurement and we typically assume that $\|e\|_2 \leq \varepsilon$.

The following l_0-minimization is perhaps the most straightforward algorithm to find the sparest solution in this setting:

$$\min_{x \in \mathbb{R}^N} \|x\|_0 \text{ s.t. } \|y - Ax\|_2 \leq \varepsilon. \tag{1.2}$$

In the noiseless case (i.e. $\varepsilon = 0$), (1.2) recovers every s-sparse vector if and only if $spark(A) > 2s$ [15].[1] This means $m \geq 2s$ is sufficient for perfect recovery (provided $spark(A) > 2s$). However, (1.2) is intractable in practice as it is NP-hard for any given $\varepsilon \geq 0$ [18] and furthermore, it is neither stable nor robust [15].

As the convex relaxation of l_0-minimization, consider the l_1-minimization

$$\min_{x \in \mathbb{R}^N} \|x\|_1 \text{ s.t. } \|y - Ax\|_2 \leq \varepsilon. \tag{1.3}$$

This is also referred to as Basis Pursuit DeNoise (BPDN) and simply Basis Pursuit (BP) when $\varepsilon = 0$. It is shown (e.g. [7],[11]) that BPDN recovers x stably and robustly for some properly chosen sampling matrix A with $m \geq Cs\log(N/s)$. We will now discuss this in more detail.

One important concept in CS is the restricted isometry property, which is first introduced by Candès and Tao [10].

Definition 1.1.1. A matrix A satisfies the *restricted isometry property* (RIP) of order s with *restricted isometry constant* (RIC) δ_s if

$$(1 - \delta_s)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_s)\|x\|_2^2$$

[1] The spark of a matrix A is the smallest number of columns of A that are linearly dependent.

holds for every s-sparse vector x.

Remark 1.1.1. Roughly speaking, a small δ_s implies every s columns of A are nearly orthogonal. In fact, let A_S be the matrix made from some s columns in A, then by this definition, A_S preserves the square of Euclidean distance up to a factor of δ_s. Hence a small δ_s means A_S almost preserves the Euclidean distance and is therefore nearly orthogonal.

RIP provides uniform guarantees for sparse recovery, namely a measurement matrix with a small RIC can stably and robustly recover every sparse vector. The following theorem is a precise statement of this:

Theorem 1.1.1 ([8]). *Let $A \in \mathbb{R}^{m \times N}$ be the measurement matrix satisfying $\delta_{2s} < \sqrt{2} - 1$. Then for any vector $x_* \in \mathbb{R}^N$ and $y = Ax_* + e \in \mathbb{R}^m$ with $\|e\|_2 \leq \varepsilon$, let $x_\#$ be the solution of* (1.3)*, we have*

$$\|x_\# - x_*\|_2 \leq \frac{C}{\sqrt{s}} \sigma_s(x_*)_1 + D\varepsilon, \tag{1.4}$$

where C, D are well behaved constants depending only on δ_{2s}.

Remark 1.1.2. In the sparse and noiseless case, by putting $\sigma_s(x_*)_1 = 0$ and $\varepsilon = 0$, we know that BP recovers every s-sparse vector when $\delta_{2s} < \sqrt{2} - 1$.

Remark 1.1.3. It is also true that (see e.g., [15])

$$\|x_\# - x_*\|_1 \leq C\sigma_s(x_*)_1 + D\sqrt{s}\varepsilon. \tag{1.5}$$

Remark 1.1.4. The bound $\delta_{2s} < \sqrt{2} - 1$ is not optimal. In fact, along with $\delta_{3s} + 3\delta_{4s} < 2$ [9], it is one of the earliest RIP-based guarantee for stable and robust s-sparse recovery. Various conditions and improvements have been made since (e.g. [3, 4, 17]). For the optimal results, Cai and Zhang [6] recently showed that for any given $t \geq 4/3$, $\delta_{ts} < \sqrt{(t-1)/t}$ is a sharp bound for stable and robust s-sparse recovery (sharp means for any $\epsilon > 0$, there exists a matrix A with $\delta_{ts} < \sqrt{(t-1)/t} + \epsilon$ who cannot stably recover every s-sparse vector). Consequently the sharp bound for δ_{2s} is $\delta_{2s} < 1/\sqrt{2} \approx 0.7071$. They also proved that $\delta_s < 1/3$ is another sharp bound [5].

Despite the simple result of Theorem 1.1.1, verifying whether a given matrix satisfies the RIP condition is still NP-hard [1]. But it turns out, with exponentially high probability, various types of random matrices including Gaussian and Bernoulli satisfy RIP with small RIC. This is why random matrices play an important role in CS.

One of the most common and well-studied random ensembles is the Gaussian ensemble. A matrix $A \in \mathbb{R}^{m\times N}$ is Gaussian if its entries A_{ij} are independent Gaussian variables, all with distribution $N(0, \frac{1}{m})$. For Gaussian matrices, RIP conditions may be satisfied when $m \ll N$.

Theorem 1.1.2 ([15])**.** *Let $A \in \mathbb{R}^{m\times N}$ be a Gaussian matrix, then there exists a universal constant $C > 0$ such that A satisfies $\delta_s \le \delta$ with probability at least $1 - 2\exp(-\delta^2 m/C)$ provided that*

$$m \ge C\delta^{-2} s \log(eN/s). \tag{1.6}$$

Remark 1.1.5. When (1.6) is satisfied, the failure probability is at most

$$2\exp(-\delta^2 m/C) \le 2(s/eN)^s = 2e^{-\gamma N},$$

where $\gamma = \frac{s}{N} \log \frac{eN}{s}$ dependents only on the ratio s/N. Therefore, with $\frac{s}{N} \ge \alpha > 0$ for some constant α, we have $\gamma \ge \alpha \log \frac{e}{\alpha} > 0$ so that the failure probability decreases exponentially as N increases. Theorem 1.1.2 also holds for certain classes of subgaussian matrices [15].

Remark 1.1.6. Putting Theorem 1.1.1 and Theorem 1.1.2 together, we know that Gaussian matrices guarantee stable and robust s-sparse recovery (with high probability) provided (1.6) is satisfied. On the other hand, [12] shows that if we want (1.5) or (1.4) to hold with $x_\#$ being the solution of any recovery algorithm, then we must have $m = O(s \log(eN/s))$. So (1.6) is optimal up to a constant factor.

Remark 1.1.7. If we take an orthonormal basis Φ in $\mathbb{R}^N$, it is easy to see that $A\Phi$ is still Gaussian, so Theorem 1.1.2 holds for $A\Phi$. Furthermore, Theorem 1.1.2 also holds for $A\Phi$ when A is subgaussian. This is known as the universality of subgaussian matrices [15].

We end this section by noting the case when a signal x is sparse (or compressible) in some other domain Φ, namely $x = \Phi u$ where Φ is a orthonormal basis and u a sparse (or compressible) vector. Now consider $A^{'} = A\Phi$, by Theorem 1.1.1 and Remark 1.1.7 we know that BPDN still offers stable and robust sparse recovery. This is a very useful fact as it extends the basis for representing signals form the canonical one to any orthonormal one.

1.2 Distributed Compressive Sensing

In compressive sensing, sometimes we may have a collection of signals that are similar in certain aspects. To recover all these signals, one way is to apply the method and theories in section 1.1 separately on each signal and its measurements. However, this is often unnecessary and by considering their similarities, much better methods may be designed.

As a toy example, in the noiseless case, suppose that we have n s-sparse signals $x_1, \ldots, x_n$ all with the same support. Consider the following recovery method: first recover x_1 through BP and second, recover each of x_i for $i \geq 2$ by solving a s by s linear system with $\text{supp}(x_i) = \text{supp}(x_1)$. Obviously, this method only requires s measurements each for $x_2, \ldots, x_n$ rather than $O(s \log(eN/s))$ measurements when we recover each signal separately.

Distributed Compressive Sensing (DCS) arises in similar circumstances where multiple (sparse) signals share common information that we can exploit, and studies the recovery of these signals. The theory of DCS is introduced in [2], in which the authors use the Joint Sparsity Models (JSM) to model the intra- and inter-signal correlations. Other types of signal models have also been studied in [22], [21], [20]. It is worth noting that signal model is a prerequisite for DCS and different models usually lead to different algorithms and analysis. In this paper we are mainly concerned with JSM-1, the first signal model in [2].

JSM-1 assumes that each signal is made of two sparse components: a common part and an innovation part. The common part, modeling the shared information, is the same across all signals while the innovation part, modeling additional individual information for each signal, varies across signals. So, for such a model with J signals $x_1, \ldots, x_J \in \mathbb{R}^N$, we write

$$x_i = z_0 + z_i, \quad i \in [J] := \{1, \ldots, J\} \tag{1.7}$$

where z_0 denotes the common part and z_i the innovation part.

Here we say $\{z_i\}_{i=0}^J$ is a JSM representation of $\{x_i\}_{i=1}^J$. Notice that from (1.7), this representation is not unique. For example, given $x_1(n), \ldots, x_J(n)$, we cannot uniquely determine $z_0(n), \ldots, z_J(n)$ since there are $J+1$ variables and only J equations. However, if we know one of them is zero, the rest of $z_i(n)$ can then be determined. We call this the reduced JSM representation.

Definition 1.2.1. $\{z_i\}_{i=0}^J$ is a *reduced JSM representation* of $\{x_i\}_{i=1}^J$ if

$$\#\{0 \le i \le J : z_i(n) \ne 0\} \le J, \quad \forall n \in [N].$$

It is easy to see that every JSM representation can be rewritten as a reduced one.

We assume each signal is sampled by $y_i = A_i x_i$ where $A_i \in \mathbb{R}^{m_i \times N}$ is the measurement matrix, then to recover the sparse solutions, we can consider the following l_0-minimization problem.

$$\min \sum_{k=0}^{J} \|z_k\|_0 \text{ s.t. } y_i = A_i(z_0 + z_i),\ i = 1, \ldots, J. \tag{1.8}$$

Baron et al. give conditions for recovering $\{x_i\}$ through (1.8) in [2]. To state this, they first define the overlap size $K_C(\Gamma, \{z_i\})$ and conditional sparsity $K_{\text{cond}}(\Gamma, \{z_i\})$.

Definition 1.2.2 ([2])**.** Given a JSM representation $\{z_i\}_{i=0}^J$, the *overlap size* for a set of signals $\Gamma \subseteq [J]$, denoted $K_C(\Gamma, \{z_i\})$, is the number of indices in which there is overlap between the common and the innovation part supports at all signals $j \notin \Gamma$:

$$K_C(\Gamma, \{z_i\}) = |\{n \in [N] : z_0(n) \ne 0 \textit{ and } \forall j \notin \Gamma, z_j(n) \ne 0\}|.$$

Also define $K_C([J], \{z_i\}) = K_C(\{z_i\})$ and $K_C(\emptyset, \{z_i\}) = 0$. The *conditional sparsity* for a set of signals $\Gamma \subseteq [J]$ is defined as

$$K_{\text{cond}}(\Gamma, \{z_i\}) = \left(\sum_{j \in \Gamma} |\text{supp}(z_j)| \right) + K_C(\Gamma, \{z_i\}).$$

The conditional sparsity is the number of entries form $\{z_i\}_{i\in\Gamma}$ that must be recovered by measurements y_j $(j \in \Gamma)$. Roughly speaking, the support of z_j $(j \in \Gamma)$ must be recovered by y_j $(j \in \Gamma)$ since other y_j provides no information about them. All entries in z_0 with $z_0(n) \neq 0$ and $z_j(n) \neq 0$ $(\forall j \notin \Gamma)$ must also be recovered by y_j $(j \in \Gamma)$ since outside of Γ, $z_0(n)z_j(n) \neq 0$ for all j, so we cannot uniquely determine $z_0(n)$ from $j \notin \Gamma$. For the number of required measurements, Baron et al. showed the following theorem:

Theorem 1.2.1 ([2])**.** *Let $\{x_i\}_{i\in[J]}$ be signals in $\mathbb{R}^N$ from JSM-1 and each measured by a Gaussian matrix $A_i \in \mathbb{R}^{m_i\times N}$ yielding $y_i = A_i x_i$. Suppose there exists a reduced JSM representation $\{z_i\}$ of $\{x_i\}$ such that*

$$\sum_{j\in\Gamma} m_j \geq K_{cond}(\Gamma, \{z_i\}) + |\Gamma| \tag{1.9}$$

holds for all $\Gamma \subseteq [J]$, then $\{x_i\}$ can be uniquely recovered from (1.8) *with probability one over $\{A_i\}_{i\in[J]}$.*
Conversely, if there exists a reduced JSM representation $\{z_i'\}$ such that

$$\sum_{j\in\Gamma} m_j < K_{cond}(\Gamma, \{z_i'\}) \tag{1.10}$$

holds for some $\Gamma \subseteq [J]$, then $\{x_i\}$ cannot be uniquely recovered.

Remark 1.2.1. The gap $|\Gamma|$ between (1.9) and (1.10) is due to identifying the support of z_i. If all support locations of $\{z_i\}$ are prior known, this gap can be removed [2].

Aside from JSM-1, another popular signal model in DCS is the Multiple Measurement Vector (MMV), which is also the JSM-2 in [2]. MMV assumes we have measurements of multiple (sparse) signals with exactly the same support. This model is much more well-studied than JSM-1 as its support structure is more simple and direct. Many algorithms have been proposed for MMV, e.g., Simultaneous Orthogonal Matching Pursuit (SOMP [23]), Reduce MMV and Boost (ReMBo [16]) and the mixed-norm optimization ([13],[14]), etc.

Even with Theorem 1.2.1, the l_0-minimization as in (1.8) is again NP-hard, just like (1.2). This prompts us to look at the convexly relaxed weighted l_1 minimization (also called the *γ-weighted l_1-norm formulation* in [2]):

$$\min \sum_{k=0}^{J} \gamma_k \|z_k\|_1 \text{ s.t. } y_i = A_i(z_0 + z_i),\ i = 1, \dots, J$$

where each wight $\gamma_k > 0$ is introduced to balance the magnitudes of z_k. By further assuming that $\{z_i\}_{i>0}$ have approximately the same magnitudes, we then arrive at a relatively simplified algorithm which only imposes a weight on the common part z_0:

$$\mathcal{P}_Z(\gamma): \quad \min \gamma\|z_0\|_1 + \sum_{k=1}^{J} \|z_k\|_1 \text{ s.t. } y_i = A_i(z_0 + z_i),\ i = 1, \dots, J$$

where $\gamma > 0$ is a given parameter. This is the algorithm we will study in the rest of this thesis.

1.3 Notations

Here we clarify the notations for this thesis.

All vectors are column vectors unless stated otherwise. For simplicity, we will write

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_J \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_J \end{bmatrix}, \quad Z = \begin{bmatrix} \gamma z_0 \\ z_1 \\ \vdots \\ z_J \end{bmatrix}, \tag{1.11}$$

and

$$A = \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_J \end{bmatrix}, \tag{1.12}$$

where $X \in \mathbb{R}^{JN}$, $Y \in \mathbb{R}^{\sum m_i}$, $Z \in \mathbb{R}^{(J+1)N}$ and $A_i \in \mathbb{R}^{m_i \times N}$. Also write

$$\Psi = \begin{bmatrix} \gamma^{-1} I_N & I_N & & & \\ \gamma^{-1} I_N & & I_N & & \\ \vdots & & & \ddots & \\ \gamma^{-1} I_N & & & & I_N \end{bmatrix} \tag{1.13}$$

so that we have $X = \Psi Z$ and $\mathcal{P}_Z(\gamma)$ becomes

$$\mathcal{P}_Z(\gamma): \quad \min_{Z \in \mathbb{R}^{(J+1)N}} \|Z\|_1 \text{ s.t. } Y = A\Psi Z.$$

Note that a parameter γ is implied whenever we write Ψ or Z.

Moreover, we will write $[M] = \{1, 2, \ldots, M\}$ for any positive integer M; $|S|$ or $\#S$ for the cardinality of a set S; $\overline{S}$ for the complement of a set S; $w(i)$ for the i-th coordinate of a vector w; $\text{supp}(w)$ for the support of a vector w; $\mathbb{N}$ for natural numbers and $\mathbb{N}_+$ for positive integers; $\psi = \ker(\Psi)$ for the kernel of Ψ. Note that ψ is a N dimensional subspace and

$$\psi = \{(\gamma u^T, -u^T, \ldots, -u^T)^T : u \in \mathbb{R}^N\}.$$

We denote by w_S the restriction of a vector w to indices in S. Particularly, for $X \in \mathbb{R}^{JN}$ and $\mathcal{S} \subseteq [JN]$, we have $X_{\mathcal{S}}$ defined as

$$X_{\mathcal{S}}(i) = X(i)\mathbf{1}_{\{i \in \mathcal{S}\}}, \quad i \in [JN]$$

where $\mathbf{1}_{\{i \in \mathcal{S}\}}$ is the indicator function of whether $i \in \mathcal{S}$. We will also use the notation $X|_T$ for $X \in \mathbb{R}^{JN}$ and $T \subseteq [N]$, which means the restriction to T of every x_j in X, i.e.,

$$X|_T = \begin{bmatrix} (x_1)_T \\ \vdots \\ (x_j)_T \end{bmatrix}.$$

Especially, we write $X|_i = X|_{\{i\}}$ for $i \in [N]$.

1.4 Main Contributions

The main contribution of this thesis is a null space property that guarantees the recovery of JSM-1 signals through $\mathcal{P}_Z(\gamma)$. Furthermore, this property can be made to be stable and robust. To the best of the author's knowledge, this is the first sufficient condition of such kind. Aside form these recovery guarantees, we also give a geometric interpretation for understanding the effectiveness of $\mathcal{P}_Z(\gamma)$.

Chapter 2 gives a detailed discussion about these contributions. The key point of observation is to write $\mathcal{P}_Z(\gamma)$ into an equivalent form. Our null space property is also derived by studying this equivalent problem. Chapter 3 is a case study for $\mathcal{P}_Z(\gamma)$ with $J = 2$, namely the recovery of two signals. Chapter 4 presents several numerical experiments on $\mathcal{P}_Z(\gamma)$ with $J = 2$ and chapter 5 summarizes the thesis.

Chapter 2

A Weighted l_1-Minimization

Recall that under the Joint Sparsity Model (JSM) we assume every signal x_i is the sum of a common part z_0 and an innovation part z_i. We also consider a weighted l_1-minimization for recovering J such signals:

$$\mathcal{P}_Z(\gamma): \quad \min \gamma \|z_0\|_1 + \sum_{k=1}^{J} \|z_k\|_1 \text{ s.t. } y_i = A_i(z_0 + z_i), \;\; i = 1, \ldots, J$$

where $\gamma > 0$ is a weight imposed on the common part. Using the matrix notations as in section 1.3, we can write $\mathcal{P}_Z(\gamma)$ in a more concise form

$$\mathcal{P}_Z(\gamma): \quad \min_{Z \in \mathbb{R}^{(J+1)N}} \|Z\|_1 \text{ s.t. } Y = A\Psi Z.$$

The signals are then recovered by $X = \Psi Z$. To study this algorithm, we first write it into a different but equivalent form.

2.1 An Equivalent Problem

As to be shown in Theorem 2.1.3, the optimization problem $\mathcal{P}_Z(\gamma)$ is equivalent to

$$\mathcal{P}_X(\gamma): \quad \min_{X \in \mathbb{R}^{JN}} \|X\|_\Psi \text{ s.t. } Y = AX.$$

Here the Ψ-norm $\|\cdot\|_\Psi$ is defined as follows (with $\psi = \text{Ker}(\Psi)$)

$$\|X\|_\Psi = \inf_{U \in \psi} \| \begin{bmatrix} 0 \\ X \end{bmatrix} + U\|_1. \tag{2.1}$$

In other words, the Ψ-norm of a $X \in \mathbb{R}^{JN}$ is the smallest l_1-norm of $Z \in \mathbb{R}^{(J+1)N}$ such that $X = \Psi Z$. So another way of writing this is

$$\|X\|_\Psi = \min_{\mathbb{R}^{(J+1)N}} \|Z\|_1 \text{ s.t. } X = \Psi Z.$$

We first give an intuitive (but not rigorous) argument for the equivalence. Looking at $\mathcal{P}_Z(\gamma)$, by adding any $U \in \psi$ to Z, we can rewrite it as

$$\min_{U\in\psi, Z} \|Z+U\|_1 \text{ s.t. } Y = A\Psi(Z+U) = A\Psi Z.$$

Note the above minimization is taken over both U and Z. Now if we minimize over U first, it becomes

$$\min_{Z} \min_{U\in\psi} \|Z+U\|_1 \text{ s.t. } Y = A\Psi Z.$$

By definition, the Ψ-norm of ΨZ is the smallest l_1-norm of $\hat{Z}$ such that $\Psi\hat{Z} = \Psi Z$, so $\|\Psi Z\|_\Psi$ is the same as $\min_{U\in\psi}\|Z+U\|_1$ (this can also be shown as in Lemma 2.1.2). Hence we have

$$\min_{Z} \|\Psi Z\|_\Psi \text{ s.t. } Y = A\Psi Z.$$

Lastly, writing this into a minimization problem of X using $X = \Psi Z$, we arrive at $\mathcal{P}_X(\gamma)$.

Although it is not a proof, the above argument does highlight a direct connection between $\mathcal{P}_Z(\gamma)$ and $\mathcal{P}_X(\gamma)$, and we now give a detailed proof of their equivalence below. First need to show that $\|\cdot\|_\Psi$ is indeed a norm.

Lemma 2.1.1. *$\|\cdot\|_\Psi$ is a norm on $\mathbb{R}^{JN}$.*

Proof. Notice that vector spaces $\mathbb{R}^{JN}$ and $\mathbb{R}^{(J+1)N}/\psi$ are isomorphic with isomorphism

$$\begin{aligned} f: \quad \mathbb{R}^{JN} &\to \mathbb{R}^{(J+1)N}/\psi \\ X &\mapsto \begin{bmatrix} 0 \\ X \end{bmatrix} + \psi. \end{aligned}$$

The claim then follows from the fact that $\inf_{U\in\psi}\|\begin{bmatrix} 0 \\ X \end{bmatrix} + U\|_1$ is the canonical quotient norm on $\mathbb{R}^{(J+1)N}/\psi$. □

The following lemma is a helpful result in proving the equivalence of $\mathcal{P}_Z(\gamma)$ and $\mathcal{P}_X(\gamma)$.

Lemma 2.1.2. *For $Z \in \mathbb{R}^{(J+1)N}$ as defined in* (1.11) *and* Ψ *in* (1.13)*, we have*

$$\|\Psi Z\|_\Psi = \inf_{U \in \psi} \|Z + U\|_1.$$

Proof. Let

$$\hat{Z} = \begin{bmatrix} \gamma z_0 \\ -z_0 \\ \vdots \\ -z_0 \end{bmatrix},$$

then $\hat{Z} \in \psi$ (as $\Psi \hat{Z} = 0$) and

$$\begin{bmatrix} 0 \\ \Psi Z \end{bmatrix} = \begin{bmatrix} 0 \\ z_0 + z_1 \\ \vdots \\ z_0 + z_J \end{bmatrix} + \hat{Z} - \hat{Z} = Z - \hat{Z}.$$

By definition,

$$\|\Psi Z\|_\Psi = \inf_{U \in \psi} \| \begin{bmatrix} 0 \\ \Psi Z \end{bmatrix} + U\|_1 = \inf_{U \in \psi} \|Z + U - \hat{Z}\|_1 = \inf_{U \in \psi} \|Z + U\|_1.$$

□

Now the main theorem for this section:

Theorem 2.1.3. *The optimization problems $\mathcal{P}_Z(\gamma)$ and $\mathcal{P}_X(\gamma)$ are equivalent, in the sense that, with the same Y and A, if $\mathfrak{Z}$ and $\mathfrak{X}$ are the solution sets for $\mathcal{P}_Z(\gamma)$ and $\mathcal{P}_X(\gamma)$ respectively, then*

$$\Psi\mathfrak{Z} := \{\Psi Z : Z \in \mathfrak{Z}\} = \mathfrak{X}.$$

Furthermore, for every $Z \in \mathfrak{Z}$ and $X \in \mathfrak{X}$,

$$\|Z\|_1 = \|X\|_\Psi. \tag{2.2}$$

Proof. If Z_* is a minimizer for $\mathcal{P}_Z(\gamma)$, it is easy to see that

$$\|Z_*\|_1 = \inf_{V\in\ker(A), U\in\psi} \|Z_* + \begin{bmatrix} 0 \\ V \end{bmatrix} + U\|_1. \tag{2.3}$$

We also claim that

$$\|Z_*\|_1 = \inf_{U\in\psi} \|Z_* + U\|_1. \tag{2.4}$$

This is because if not, let

$$\hat{U} = \arg\min_{U\in\psi} \|Z_* + U\|_1, \quad \hat{Z} = Z_* + \hat{U},$$

then $\hat{Z}$ is also a solution for $Y = A\Psi Z$ and $\|\hat{Z}\|_1 < \|Z_*\|_1$, which contradicts to Z_* being a minimizer.

Now for $V \in \ker(A)$, by Lemma 2.1.2,

$$\|\Psi Z_* + V\|_\Psi = \|\Psi \left(Z_* + \begin{bmatrix} 0 \\ V \end{bmatrix} \right) \|_\Psi = \inf_{U\in\psi} \|Z_* + \begin{bmatrix} 0 \\ V \end{bmatrix} + U\|_1, \tag{2.5}$$

then by (2.3), (2.4) and (2.5) we have

$$\|\Psi Z_* + V\|_\Psi \geq \|Z_*\|_1 = \inf_{U\in\psi} \|Z_* + U\|_1.$$

Using Lemma 2.1.2 again, this becomes

$$\|\Psi Z_* + V\|_\Psi \geq \|\Psi Z_*\|_\Psi,$$

since the above holds for any $V \in \ker(A)$, we conclude that ΨZ_* is a minimizer for $\mathcal{P}_X$, i.e., $\Psi Z_* \in \mathfrak{X}$.

If X_* is a minimizer for $\mathcal{P}_X$, then

$$\|X_*\|_\Psi \leq \|X_* + V\|_\Psi, \quad \forall V \in \ker(A).$$

Take $W_* \in \mathbb{R}^{(J+1)N}$ such that

$$X_* = \Psi W_*, \quad \|X_*\|_\Psi = \|W_*\|_1.$$

For any $V \in \ker(A)$, we have

$$\|W_*\|_1 = \|X_*\|_\Psi \leq \|X_* + V\|_\Psi.$$

Also notice that

$$\|X_* + V\|_\Psi = \|\Psi\left(W_* + \begin{bmatrix} 0 \\ V \end{bmatrix}\right)\|_\Psi = \inf_{U\in\psi} \|W_* + \begin{bmatrix} 0 \\ V \end{bmatrix} + U\|_1.$$

So we conclude that

$$\|W_*\|_1 \le \inf_{U\in\psi} \|W_* + \begin{bmatrix} 0 \\ V \end{bmatrix} + U\|_1$$

holds for every $V \in \ker(A)$. Therefore W_* is a minimizer for $\mathcal{P}_Z$, i.e., $W_* \in \mathfrak{Z}$.

Finally, (2.2) is true because, as shown above, $\|X_*\|_\Psi = \|W_*\|_1$ and $X_* \in \mathfrak{X}$, $W_* \in \mathfrak{Z}$. □

Remark 2.1.1. From Theorem 2.1.3, Ψ maps the solution set $\mathfrak{Z}$ onto $\mathfrak{X}$. Here we point out that this mapping is not necessarily one to one. For example, take $N = 1$, $J = 3$, $\gamma = 1$ and assume that $X_0 = [x_1 \quad x_2 \quad x_3]^T = [1 \quad 1 \quad 0]^T$ is in $\mathfrak{X}$. Notice that $\|X_0\|_\Psi = 2$ and

$$X_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}.$$

So we have $Z_0 = [0 \quad 1 \quad 1 \quad 0]^T$ and $Z_1 = [1 \quad 0 \quad 0 \quad -1]^T$ both in $\mathfrak{Z}$ with $\Psi Z_0 = \Psi Z_1 = X_0$. (In fact, $Z_\alpha = [\alpha \quad 1-\alpha \quad 1-\alpha \quad -\alpha]^T \in \mathfrak{Z}$ for all $\alpha \in [0,1]$ and $\Psi Z_\alpha = X_0$).

However, for most choices of γ, the mapping Ψ is indeed one to one. This is stated in the following proposition.

Proposition 2.1.4. *Let $\mathfrak{Z}$ and $\mathfrak{X}$ be the same as in Theorem 2.1.3. If*

$$\gamma \notin \{s - t : s, t \in \mathbb{N}_+, s + t = J\}, \tag{2.6}$$

then Ψ is a bijection between $\mathfrak{Z}$ and $\mathfrak{X}$.

Proof. For $X \in \mathfrak{X}$, consider

$$\mathfrak{Z}_X = \{Z \in \mathfrak{Z} : \Psi Z = X\}.$$

We need to show $|\mathfrak{Z}_X| = 1$ so that Ψ is injective. First notice that for every $Z \in \mathfrak{Z}_X$, if its first N coordinates γz_0 are given, then we can uniquely determine $z_1, \ldots, z_J$ through $\Psi Z = X$. In other words, $Z \in \mathfrak{Z}_X$ can be uniquely determined by γz_0, or simply z_0.

Now look at z_0. As a result from Theorem 2.1.3 we have $\|Z\|_1 = \|X\|_\Psi$. Writing this out we can see $z_0(i) \in \mathfrak{z}_i$ where

$$\mathfrak{z}_i = \arg\min_{u \in \mathbb{R}} \gamma|u| + \sum_{j \in [J]} |x_j(i) - u|, \quad i \in [N].$$

To study the cardinality of $\mathfrak{z}_i$, let

$$f_i(u) = \gamma|u| + \sum_{j \in [J]} |x_j(i) - u|,$$

then $f_i(u)$ is a convex piecewise linear function and we can calculate its right derivative

$$f'_{i+}(u) = \begin{cases} \gamma - s + t, & u \geq 0 \\ -\gamma - s + t, & u < 0 \end{cases}$$

where $s = \{j : x_j(i) > u\}$ and $t = \{j : x_j(i) \leq u\}$. According to condition (2.6), $f'_{i+}(u) \neq 0$ for every $u \in \mathbb{R}$.

Thus we conclude that $f_i(u)$ has a unique minimum, i.e., $|\mathfrak{z}_i| = 1$. This is because if not, take u_1 and u_2 to be two minimizers with $u_1 < u_2$, then by convexity, f_i is a constant on $[u_1, u_2]$, which implies $f'_{i+}(u_1) = 0$.

Since every $\mathfrak{z}_i$ ($i \in [N]$) has exactly one element, z_0 can have only one possible value. So we have $|\mathfrak{Z}_X| = 1$.

Ψ is surjective by Theorem 2.1.3. □

Theorem 2.1.3 suggests that $\mathcal{P}_Z(\gamma)$ and $\mathcal{P}_X(\gamma)$ are essentially the same, so we will not distinguish them from now on. In the rest of the thesis, a solution X_* of $\mathcal{P}_Z(\gamma)$ simply means that X_* is a solution to the equivalent problem $\mathcal{P}_X(\gamma)$. Similarly, a solution Z_* of $\mathcal{P}_X(\gamma)$ means Z_* is a solution to the equivalent problem $\mathcal{P}_Z(\gamma)$.

In practice, we are more interested in the case where $\mathfrak{X}$ contains only one (sparse) element. The reason for this is that most algorithms for l_1-minimization will only return with one solution. Besides, even if we are able to find more solutions, we may not be able to identify which one is the true signal(s). As a result from Theorem 2.1.3, the following corollary provides a characterization for the unique minimizer of $\mathcal{P}_Z(\gamma)$.

Corollary 2.1.5. *X_* is the unique minimizer of $\mathcal{P}_Z(\gamma)$ if and only if*

$$\| \begin{bmatrix} 0 \\ X_* \end{bmatrix} + \begin{bmatrix} 0 \\ V \end{bmatrix} + U\|_1 > \|X_*\|_\Psi \tag{2.7}$$

holds for every $V \in \ker(A)\backslash\{0\}$ and every $U \in \psi$.

Proof. This follows directly from the fact that X_* is the unique minimizer if and only if

$$\|X_* + V\|_\Psi > \|X\|_\Psi, \quad \forall V \in \ker(A)\backslash\{0\}.$$

□

Corollary 2.1.5 may be used to further derive guarantees for sparse recovery. For now, we turn our eyes to a study of the unit ball of the Ψ-norm as it presents us a geometric interpretation of how and why $\mathcal{P}_Z(\gamma)$ works.

2.2 Geometric Intuition: The Unit Ball for Ψ-Norm

The main result (Theorem 2.2.3) of this section is a description of $B_{\|\cdot\|_\Psi}$, the unit ball for Ψ-norm. Roughly speaking, $B_{\|\cdot\|_\Psi}$ is the convex hull of the l_1 unit ball $B_{\|\cdot\|_1}$ and a few "other points". So, in some degree, $B_{\|\cdot\|_\Psi}$ and $B_{\|\cdot\|_1}$ are similar to each other. This can be used to understand why $\mathcal{P}_Z(\gamma)$ promotes sparsity. Meanwhile, for the different parts of these two unit balls, the existence of those "other points" favors signals with (large) common part to be the solution for $\mathcal{P}_Z(\gamma)$. This can help us understand why $\mathcal{P}_Z(\gamma)$ performs well for the joint model. The weight γ also plays an role in determining the shape of $B_{\|\cdot\|_\Psi}$.

We start by looking at an easier scenario: each signal x_j is only a single value in $\mathbb{R}$, i.e., $N = 1$ and $x_j = x_j(1)$. Now we have the following lemma.

Lemma 2.2.1. *For X and Ψ with $N = 1$, we have*

$$\{X \in R^J : \|X\|_\Psi \leq 1\} = Conv\{\pm e_i^J, \pm\frac{1}{\gamma}\mathbf{1}^J : i \in [J]\},$$

where $e_i^J \in \mathbb{R}^J$ is the vector with i-th coordinate one and all others zero, $\mathbf{1}^J \in \mathbb{R}^J$ is the all one vector.

Proof. "$\subseteq$": By definition,

$$\|X\|_\Psi = \inf_u |\gamma u| + \sum_{j \in [J]} |x_j(1) - u|.$$

Let

$$u_* = \arg\min_u |\gamma u| + \sum_{j \in [J]} |x_j(1) - u|,$$

we can then write X as

$$X = \gamma u_* \frac{1}{\gamma}\mathbf{1}^J + \sum_{j \in [J]} (x_j(1) - u_*) e_j^J.$$

For $\|X\|_\Psi = 1$, notice that

$$|\gamma u_*| + \sum_{j \in [J]} |x_j(1) - u_*| = 1,$$

so $X \in \text{Conv}\{\pm e_i^J, \pm\gamma^{-1}\mathbf{1}^J : i \in [J]\}$.

"$\supseteq$": Notice that $\|e_i^J\|_\Psi = 1$, $\|\gamma^{-1}\mathbf{1}^J\|_\Psi \leq 1$ and $\{X \in R^J : \|X\|_\Psi \leq 1\}$ is convex. □

One result of Lemma 2.2.1 is the following corollary, which is from the convexity of the unit ball.

Corollary 2.2.2. *For X and Ψ with $N = 1$, if $\|X\|_\Psi = 1$, then there exists λ_k such that*

$$X = \sum_k \lambda_k p_k,$$

$$\lambda_k \geq 0, \quad \sum_k \lambda_k = 1,$$

where $p_k \in \{\pm e_i^J, \pm\gamma^{-1}\mathbf{1}^J : i \in [J]\}$.

With the preparations above, we can now describe the unit ball for Ψ-norm. The key observation here is to write X as a sum of projections where Lemma 2.2.1 and Corollary 2.2.2 can be applied to each projection.

Theorem 2.2.3. *Let e_i^n be the vector in $\mathbb{R}^n$ with i-th coordinate one and all other coordinates zero. Denote $B_{\|\cdot\|}^{JN}$ the unit ball for any norm $\|\cdot\|$ in $\mathbb{R}^{JN}$, i.e., $B_{\|\cdot\|}^{JN} := \{X \in \mathbb{R}^{JN} : \|X\| \leq 1\}$, then*

$$\begin{aligned} B_{\|\cdot\|_\Psi}^{JN} &= \mathit{Conv}\{\pm e_i, \pm d_j : i \in [JN], j \in [N]\} \\ &= \mathit{Conv}\{B_{\|\cdot\|_1}^{JN}, \pm d_j : j \in [N]\} \end{aligned}$$

where $e_i = e_i^{JN}$ *and* $d_j = \frac{1}{\gamma}\begin{bmatrix} e_j^N \\ \vdots \\ e_j^N \end{bmatrix} \in \mathbb{R}^{JN}$.

Proof. "$\subseteq$": $\forall X \in \partial B_{\|\cdot\|_\Psi}^{JN}$,[2] let

$$X|_i = \sum_{j=1}^{J} x_j(i) e_{i+(j-1)N}, \quad i = 1, \cdots, N$$

be the projection of every x_j to its i-th coordinate, then $X = \sum_{i \in [N]} X|_i$ and by definition,

$$\|X\|_\Psi = \sum_{i \in [N]} \left(\inf_{u_i} |\gamma u_i| + \sum_{j \in [J]} |x_j(i) - u_i| \right) = \sum_{i \in [N]} \|X|_i\|_\Psi.$$

Let $\alpha_i = \|X|_i\|_\Psi$, by Corollary 2.2.2, $\exists \lambda_{i_k}$ and p_{i_k} such that

$$X|_i = \alpha_i \sum_k \lambda_{i_k} p_{i_k},$$

$$\lambda_{i_k} \geq 0, \quad \sum_k \lambda_{i_k} = 1,$$

$$p_{i_k} \in \{\pm d_i, \pm e_i, \ldots, \pm e_{i+(J-1)N}\}.$$

So

$$X = \sum_{i \in [N]} X|_i = \sum_{i \in [N]} \sum_k \alpha_i \lambda_{i_k} p_{i_k},$$

[2] ∂S denotes the boundary of a set S. Here $\partial B_{\|\cdot\|_\Psi}^{JN} = \{X \in \mathbb{R}^{JN} : \|X\|_\Psi = 1\}$.

and we also have

$$\sum_{i\in[N]}\sum_{k}\alpha_i\lambda_{i_k} = \sum_{i\in[N]}\alpha_i = \|X\|_\Psi = 1.$$

This shows $X \in \text{Conv}\{\pm e_i, \pm d_j : i \in [JN], j \in [N]\}$.

"$\supseteq$": Notice that $\|e_i\|_\Psi = 1$, $\|d_j\|_\Psi \leq 1$ and $B^{JN}_{\|\cdot\|_\Psi}$ is convex. □

Remark 2.2.1. $B^{JN}_{\|\cdot\|_\Psi}$ is the convex hull of $B^{JN}_{\|\cdot\|_1}$ and $\{\pm d_j : j \in [N]\}$. Despite $\pm d_j$, the two unit balls $B^{JN}_{\|\cdot\|_\Psi}$ and $B^{JN}_{\|\cdot\|_1}$ are similar in shape. Since l_1-norm can effectively promote sparsity, it should be no surprise that Ψ-norm is also capable of finding sparse solutions. On the other hand, look at d_j and divide it into sequential blocks of length N, every block has only the j-th coordinate to be nonzero with value $1/\gamma$. This structure corresponds to the common part in JSM-1. For $\gamma < J$, $\pm d_j$ are outside of $B^{JN}_{\|\cdot\|_1}$, namely $B^{JN}_{\|\cdot\|_1} \subset B^{JN}_{\|\cdot\|_\Psi}$. In this case, consider the minimization of X over both Ψ-norm and l_1-norm under the same linear constraints, and let X_* and X^* be their unique minimizer respectively. If $\|X_*\|_\Psi = \|X^*\|_\Psi$, then $X_* = X^*$; otherwise we have $\|X_*\|_\Psi < \|X^*\|_\Psi$ and $\|X_*\|_1 > \|X^*\|_1$, which implies that X_* has a larger common part than X^*. Thus Ψ-norm has a higher probability of finding solutions with larger common parts than l_1-norm. In all, we can see that $\mathcal{P}_Z(\gamma)$ promotes both sparsity and the common part when $\gamma < J$. (Note that when $\gamma \geq J$, Ψ-norm becomes l_1-norm).

Remark 2.2.2. The value of γ can no doubt affect the performance of the algorithm due to its impact on $B_{\|\cdot\|_\Psi}$.

1. For small γ (close to 0), the scale of d_j, which is $1/\gamma$, becomes very large, so the algorithm will favor solutions with large shared components. This is only suitable for signals with predominantly a common part.

2. As γ increases but remains less than J, d_j moves towards $\frac{\gamma}{J}d_j \in \partial B_{\|\cdot\|_1}$. So $B_{\|\cdot\|_\Psi}$ shrinks in all $\pm d_j$ directions and gets more and more similar to $B_{\|\cdot\|_1}$. A good choice for γ in practice is usually somewhere in here, one that can balance the common part and the innovation part of the signals.

3. Once $\gamma \geq J$, Ψ-norm becomes l_1-norm as $\pm d_j \in B_{\|\cdot\|_1}$ for all j. This can also be seen from the definition of Ψ-norm. Recall that

$$\|X\|_\Psi = \sum_{i\in[N]} \left(\min_u \gamma|u| + \sum_{j\in[J]} |x_j(i) - u| \right).$$

When $\gamma \geq J$, for every $i \in [N]$, the minimum is achieved at $u = 0$ with value $\sum_{j\in[J]} |x_j(i)|$, and the right hand side above becomes $\|X\|_1$. There is no need for $\mathcal{P}_Z(\gamma)$ in this scenario since it is computationally heavier than the l_1-minimization on X.

Corollary 2.2.4. *For $X \in \mathbb{R}^{JN}$ and $p \geq 1$,*

$$\min\left\{\frac{\gamma}{J^{1/p}}, 1\right\} \|X\|_p \leq \|X\|_\Psi \leq (JN)^{1-1/p}\|X\|_p.$$

Proof. Since $B^{JN}_{\|\cdot\|_1} \subseteq B^{JN}_{\|\cdot\|_\Psi}$, we have

$$\|X\|_\Psi \leq \|X\|_1 \leq (JN)^{1-1/p}\|X\|_p. \tag{2.8}$$

For the other side of the inequality, consider the e_i and d_j as in Theorem 2.2.3, we have

$$\|e_i\|_\Psi = \|e_i\|_p = 1, \quad \|d_j\|_\Psi = 1, \quad \|d_j\|_p = J^{1/p}/\gamma.$$

If $0 < \gamma < J$, this suggests that $B^{JN}_{\|\cdot\|_\Psi} \subseteq (J^{1/p}/\gamma) \cdot B^{JN}_{\|\cdot\|_1}$. Otherwise $\gamma > J$ and this suggests $B^{JN}_{\|\cdot\|_\Psi} \subseteq B^{JN}_{\|\cdot\|_1}$. In all, the first inequality holds. □

Remark 2.2.3. By setting $p = 1$ we get

$$\min\left\{\frac{\gamma}{J}, 1\right\} \|X\|_1 \leq \|X\|_\Psi \leq \|X\|_1. \tag{2.9}$$

The equalities in (2.9) can be achieved at e_i or d_j.
For $p > 1$, the left equality in Corollary 2.2.4 can always be achieved (at e_i or d_j), but not always for the right. Consequently $(JN)^{1-1/p}$ above is not exactly sharp. However, we note that it is sharp when $\gamma \geq 1$ or J is even. To show this, consider $\tilde{X} = [\mathbf{1}_J \quad -\mathbf{1}_J \quad \cdots \quad (-1)^{J+1}\mathbf{1}_J]^T$ where $\mathbf{1}_J$ is the all-one row vector in $\mathbb{R}^J$. With $X = \tilde{X}$, (2.8) then becomes equations. As for when $\gamma \in (0, 1)$ and J is odd, $(JN)^{1-1/p}$ is not sharp because every nonzero X that achieve the equality on the right in (2.8) (i.e. $|X(i)| = |X(1)| > 0$ for all $i \in [JN]$) will have a strict inequality on the left.

2.3 Stability and Robustness

In section 2.1, we gave a condition (Corollary 2.1.5) for the exact recovery of X through algorithm $\mathcal{P}_Z(\gamma)$. Here we will extend this condition so that $\mathcal{P}_Z(\gamma)$ have guaranteed stability and robustness.

Stability

The stability for $\mathcal{P}_Z(\gamma)$ will enable us the recovery of not only the exact spare signals under JSM-1, but also the compressible ones. It is important for practice as most real-life signals are among the latter. To study this, we first need to model the compressible signals under JSM-1.

We now write the signals to be recovered as $\hat{X} = [\hat{x}_1^T, \ldots, \hat{x}_J^T]^T$ (where $\hat{x}_j \in \mathbb{R}^N$) and the measurements $\hat{Y} = A\hat{X}$. Other notations remain the same as before. We also say that $\hat{X}$ is *JSM-1 compressible* if it can be written as

$$\hat{X} = X + E, \tag{2.10}$$

where X is a vector under JSM-1 and E is small compared to X. Note that this definition is rather loose and allows some freedom by choosing X and E. Unlike the definitions for compressible signals and best s-term approximation in compressive sensing, each x_j in (2.10) need not to be a best $\|x_j\|_0$-term approximation for $\hat{x}_j$. This is because if forced to be, the resulting X may no longer be in JSM-1.

Under this new setting, $\mathcal{P}_Z(\gamma)$ and $\mathcal{P}_X(\gamma)$ become

$$\hat{\mathcal{P}}_Z(\gamma): \quad \min_{Z\in\mathbb{R}^{(J+1)N}} \|Z\|_1 \text{ s.t. } \hat{Y} = A\Psi Z.$$

$$\hat{\mathcal{P}}_X(\gamma): \quad \min_{X\in\mathbb{R}^{JN}} \|X\|_\Psi \text{ s.t. } \hat{Y} = AX.$$

To establish stability, one natural idea is to apply the stable null space property in compressive sensing to $\hat{\mathcal{P}}_X(\gamma)$, but with Ψ-norm. A similar argument like that of l_1-norm [15] then yields the desired result. However, the stable condition derived such way is too strong and impractical. A brief description of this is given as below.

Recall the definition of the stable null space property in compressive sensing.

Definition 2.3.1. A matrix $A \in \mathbb{R}^{m \times N}$ satisfies the *stable null space property* with constant $0 < \rho < 1$ relative to S (denoted as (S, ρ)-NSP) if

$$\|v_S\|_1 \leq \rho \|v_{\overline{S}}\|_1, \quad \forall v \in \ker(A) \backslash \{0\}.$$

It is said to satisfy the *stable null space property of order s* (denoted as (s, ρ)-NSP) if the above holds for every $S \subseteq [N]$ with $|S| \leq s$.

If matrix Φ is (S, ρ)-NSP with $S = \text{supp}(x)$, then l_1-minimization is stable due to the following result [15]:

$$\|x - \Delta_{l_1}(x)\|_1 \leq \frac{2(1+\rho)}{1-\rho} \sigma_{|S|}(x)_1,$$

where

$$\Delta_{l_1}(x) = \arg\min_w \|w\|_1 \text{ s.t. } \Phi w = \Phi x.$$

Note that in showing this, the only property required form l_1-norm is that $\|w_S\|_1 + \|w_{\overline{S}}\|_1 = \|w\|_1$ for all w.

To apply this approach to $\hat{\mathcal{P}}_X(\gamma)$, first we need to choose the support set $T = \cup_{j \in [J]} \text{supp}(x_j)$ so that property $\|W|_T\|_\Psi + \|W|_{\overline{T}}\|_\Psi = \|W\|_\Psi$ holds for all $W \in \mathbb{R}^{JN}$. A stable null space property with Ψ-norm then comes as

$$\|V|_T\|_\Psi \leq \rho \|V|_{\overline{T}}\|_\Psi, \quad \forall V \in \ker(A) \backslash \{0\}. \tag{2.11}$$

Unfortunately, (2.11) is a little too strong. In fact, take any $V = \begin{bmatrix} v_1 \\ \mathbf{0} \end{bmatrix}$, where $v_1 \in \ker(A_1) \backslash \{0\}$, then $V \in \psi \backslash \{0\}$ and by (2.9) (2.11) we have

$$\|(v_1)_T\|_1 = \|V|_T\|_\Psi \leq \rho \|V|_{\overline{T}}\|_\Psi \leq \rho \|V|_{\overline{T}}\|_1 = \rho \|(v_1)_{\overline{T}}\|_1.$$

Thus we arrive at an even stronger condition than $(\text{supp}(x_1), \rho)$-NSP on A_1, which already implies the stable recovery of x_1 through l_1-minimization. Similar results can also be said for $x_2, \ldots, x_J$. Therefore in the case of (2.11), there is no need to even consider JSM and $\hat{\mathcal{P}}_Z(\gamma)$ as we can simply recover through separate l_1-minimizations.

For $\hat{X} = X + E \in \mathbb{R}^{JN}$, we will use the following (2.12) to provide stability for $\hat{\mathcal{P}}_Z(\gamma)$. Here $\alpha > 0$ is some constant.

$$\|X + V\|_\Psi \geq \|X\|_\Psi + \alpha \|V\|_\Psi, \quad \forall V \in \ker(A). \tag{2.12}$$

Note that (2.12) depends on X rather than $\hat{X}$ and, compared with the necessary and sufficient condition in Corollary 2.1.5, it is slightly stronger than (2.7) in the existence of term $\alpha\|V\|_\Psi$. Theorem 2.3.1 gives a stability result based on (2.12).

Theorem 2.3.1. *Given JSM-1 compressible signals $\hat{X} \in \mathbb{R}^{JN}$ as in* (2.10) *and let $X_\#$ be a solution of $\hat{\mathcal{P}}_X(\gamma)$ with $\hat{Y} = A\hat{X}$. If* (2.12) *holds for some $\alpha > 0$, then*

$$\|\hat{X} - X_\#\|_\Psi \leq (2/\alpha + 2)\|E\|_\Psi \tag{2.13}$$

Proof. Let

$$V = X_\# - \hat{X} = X_\# - X - E,$$

we will prove (2.13) by contradiction.

If (2.13) does not hold, then

$$\|X_\# - X\|_\Psi \geq \|V\|_\Psi - \|E\|_\Psi > (2/\alpha + 1)\|E\|_\Psi.$$

Together with (2.12), we have

$$\begin{aligned} \|X_\# - E\|_\Psi - \|X\|_\Psi &= \|X + V\|_\Psi - \|X\|_\Psi \\ &\geq \alpha\|X_\# - X - E\|_\Psi \\ &\geq \alpha\left(\|X_\# - X\|_\Psi - \|E\|_\Psi\right) \\ &> 2\|E\|_\Psi. \end{aligned}$$

On the other hand, since $X_\#$ is a solution of $\hat{\mathcal{P}}_X(\gamma)$,

$$\|X_\# - E\|_\Psi \leq \|X_\#\|_\Psi + \|E\|_\Psi \leq \|\hat{X}\|_\Psi + \|E\|_\Psi \leq \|X\|_\Psi + 2\|E\|_\Psi$$

This gives us a contradiction. Therefore (2.13) must hold. ☐

Remark 2.3.1. We may think (2.12) as roughly another form of the stable null space property. In fact, when considering the l_1-minimization in compressive sensing, it is equivalent to the stable null space property as stated in the following proposition.

Proposition 2.3.2. *A matrix $\Phi \in \mathbb{R}^{m\times N}$ is (s,ρ)-NSP if and only if*

$$\|x+v\|_1 \geq \|x\|_1 + \alpha\|v\|_1 \tag{2.14}$$

holds for all $v \in \ker(\Phi)$ and $x \in \mathbb{R}^N$ with $\|x\|_0 \leq s$. Here the corresponding relationship between ρ and α is

$$\alpha = \frac{1-\rho}{1+\rho}, \quad \rho = \frac{1-\alpha}{1+\alpha}.$$

Proof. *"if"*: For $S \subseteq [N]$ with $|S| \leq s$ and $v \in \ker(\Phi)$, by putting $x = -v_S$ in (2.14), we have

$$\|v_{\overline{S}}\|_1 \geq \|v_S\|_1 + \alpha\|v\|_1$$

Simplify this gives $\|v_S\|_1 \leq \rho\|v_{\overline{S}}\|_1$ with $\rho = \frac{1-\alpha}{1+\alpha}$.

"only if": For $v \in \ker(\Phi)$ and $x \in \mathbb{R}^N$ with $\|x\|_0 \leq s$, let $S = \mathrm{supp}(x)$. From (s,ρ)-NSP we have

$$\|v\|_1 = \|v_{\overline{S}}\|_1 + \|v_S\|_1 \leq (1+\rho)\|v_{\overline{S}}\|_1.$$

Hence

$$\|v_{\overline{S}}\|_1 = \rho\|v_{\overline{S}}\|_1 + (1-\rho)\|v_{\overline{S}}\|_1 \geq \|v_S\|_1 + \frac{1-\rho}{1+\rho}\|v\|_1.$$

Using $\mathrm{supp}(x) = S$, (2.14) then follows by

$$\|x+v\|_1 \geq \|x\|_1 - \|v_S\|_1 + \|v_{\overline{S}}\|_1 \geq \|x\|_1 + \frac{1-\rho}{1+\rho}\|v\|_1.$$

□

Remark 2.3.2. A similar argument to the *"only if"* part in Proposition 2.3.2 with Ψ-norm can show that (2.11) implies (2.12). However, a similar *"if"* argument does not work because we cannot simply put $X = V|_T$ and invoke (2.12) as $V|_T$ may well not be in JSM-1. This tells us that condition (2.12) is no stronger than (2.11). In fact, it is actually weaker as demonstrated in the following example.

Consider the case when $J = 2$, $N = 2$ and $\gamma \geq 1$, let

$$A_1 = [1 \quad 0], \quad A_2 = [1 \quad 2],$$

and

$$x_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Then we have $z_0 = \mathbf{0}$, $z_1 = x_1$, $z_2 = x_2$, $X = [1 \quad 0 \quad 0 \quad 0]^T$ and

$$V = [0 \quad a \quad -2b \quad b]^T, \quad a, b \in \mathbb{R}.$$

Notice that for $\gamma \geq 1$ and $W \in \mathbb{R}^{JN}$,

$$\max\{|W(1)|, |W(3)|\} + \max\{|W(2)|, |W(4)|\} \leq \|W\|_\Psi \leq \|W\|_1.$$

Hence

$$\|X + V\|_\Psi \geq 1 + \max\{|a|, |b|\},$$

and

$$\|X\|_\Psi + \alpha\|V\|_\Psi \leq 1 + \alpha(|a| + 3|b|) \leq 1 + 3\alpha \cdot \max\{|a|, |b|\}.$$

Therefore (2.12) holds with $\alpha = 1/3$.

On the other hand, if we take $V_* = [0 \quad 1 \quad -2 \quad 1]^T \in \ker(A)$ and $T = \cup \text{supp}(x_j) = \{1\}$, then we have

$$\|V_*|_T\|_\Psi = 2 > 1 = \|V_*|_{\overline{T}}\|_\Psi.$$

So (2.11) does not hold for any $\rho \in (0, 1)$.

Robustness

The robustness for $\mathcal{P}_Z(\gamma)$ lets us deal with noisy samplings. Here we assume the measurements $Y^{'}$ are corrupted with noise and write

$$Y^{'} = A\hat{X} + \epsilon, \quad \|\epsilon\|_2 \leq \varepsilon, \tag{2.15}$$

where $\epsilon = [\epsilon_1^T, \ldots, \epsilon_J^T]^T$ and $\epsilon_j \in \mathbb{R}^{m_j}$ is the noise when measuring x_j. Problems $\mathcal{P}_Z(\gamma)$ and $\mathcal{P}_X(\gamma)$ then become

$$\mathcal{P}_Z^{'}(\gamma): \qquad \min_{Z \in \mathbb{R}^{(J+1)N}} \|Z\|_1 \text{ s.t. } \|Y^{'} - A\Psi Z\|_2 \leq \varepsilon.$$

$$\mathcal{P}_X^{'}(\gamma): \qquad \min_{X \in \mathbb{R}^{JN}} \|X\|_\Psi \text{ s.t. } \|Y^{'} - AX\|_2 \leq \varepsilon.$$

The main result for robustness is the following theorem.

Theorem 2.3.3. *Given JSM-1 compressible signals* $\hat{X} \in \mathbb{R}^{JN}$ *as in* (2.10) *and its measurement* $Y^{'}$ *as in* (2.15). *Let* $X_{\#}$ *be a solution of* $\mathcal{P}_X^{'}(\gamma)$. *If*

$$\|X + V\|_\Psi + \tau \|AV\|_2 \geq \|X\|_\Psi + \alpha \|V\|_\Psi, \quad \forall V \in \mathbb{R}^{JN}. \tag{2.16}$$

holds for some $\alpha > 0$ *and* $\tau \geq 0$, *then*

$$\|\hat{X} - X_{\#}\|_\Psi \leq (2/\alpha + 2)\|E\|_\Psi + (2\tau/\alpha)\varepsilon. \tag{2.17}$$

Proof. This proof is similar to that of Theorem 2.3.1. Let

$$V = X_{\#} - \hat{X} = X_{\#} - X - E,$$

we have

$$\|AV\|_2 \leq \|AX_{\#} - Y^{'}\|_2 + \|Y^{'} - A\hat{X}\|_2 \leq 2\varepsilon.$$

If (2.17) does not hold, then

$$\|X_{\#} - X\|_\Psi \geq \|V\|_\Psi - \|E\|_\Psi > (2/\alpha + 1)\|E\|_\Psi + (2\tau/\alpha)\varepsilon.$$

Hence

$$\begin{aligned}
\|X_{\#} - E\|_\Psi - \|X\|_\Psi &= \|X + V\|_\Psi - \|X\|_\Psi \\
&\geq \alpha \|X_{\#} - X - E\|_\Psi - \tau \|AV\|_2 \\
&\geq \alpha \left(\|X_{\#} - X\|_\Psi - \|E\|_\Psi\right) - 2\tau\varepsilon \\
&> 2\|E\|_\Psi + 2\tau\varepsilon - 2\tau\varepsilon \\
&= 2\|E\|_\Psi.
\end{aligned}$$

On the other hand, $X_{\#}$ is a solution of $\mathcal{P}_X^{'}(\gamma)$ implies

$$\|X_{\#} - E\|_\Psi \leq \|X_{\#}\|_\Psi + \|E\|_\Psi \leq \|\hat{X}\|_\Psi + \|E\|_\Psi \leq \|X\|_\Psi + 2\|E\|_\Psi.$$

This gives a contradiction. □

Remark 2.3.3. The idea for condition (2.16) comes from the robust null space property in compressive sensing. In fact, for $\Phi \in \mathbb{R}^{m \times N}$ and a positive integer s, if we assume the robust null space property with constants $\rho \in (0, 1)$ and $\eta > 0$, i.e.,

$$\|v_S\|_1 \leq \rho \|v_{\overline{S}}\|_1 + \eta \|\Phi v\|_2, \quad \forall v \in \mathbb{R}^N, S \subseteq [N], |S| \leq s.$$

then similar to the argument in Proposition 2.3.2, we have

$$\|x+v\|_1 + \frac{2\eta}{1+\rho}\|\Phi v\|_2 \geq \|x\|_1 + \frac{1-\rho}{1+\rho}\|v\|_1$$

for all $v \in \mathbb{R}^N$ and $x \in \mathbb{R}^N$ with $\|x\|_0 \leq s$.

2.4 A Null Space Property

So far we have derived a condition for the exact recovery of jointly sparse signals (Corollary 2.1.5), as well as its stable and robust variants (2.12) and (2.16). However, these conditions are still primitive as they include X, the signals to be recovered. In this section, we will try to drop X and find some kind of null space property for A. First we study the recovery of exact jointly sparse signals with noiseless sampling.

Definition 2.4.1. A matrix $A \in \mathbb{R}^{M\times JN}$ is said to satisfy the *null space property for JSM* with constant $0 \leq \rho \leq 1$ relative to $(T, \mathcal{S})$ if

$$\|V|_{\overline{T}}\|_\Psi + \rho\|\tilde{V}_{\mathcal{S}}\|_\Psi > \|V_{\mathcal{S}}\|_\Psi, \quad \forall V \in \ker(A)\backslash\{0\} \tag{2.18}$$

where $T \subseteq [N]$, $\mathcal{S} \subseteq [JN]$ and

$$\tilde{V}_{\mathcal{S}} = V|_T - V_{\mathcal{S}}.$$

With this null space property, we have the result for exact recovery as Theorem 2.4.1 below.

Theorem 2.4.1. *For jointly sparse signals $X \in \mathbb{R}^{JN}$, let $Z \in \mathbb{R}^{(J+1)N}$ such that $\Phi Z = X$ and $\|Z\|_1 = \|X\|_\Psi$. Define sets*

$$T = \cup supp(z_j) = \{i \in [N] : \exists 0 \leq j \leq J \text{ such that } z_j(i) \neq 0\},$$

$$\mathcal{S} = \{i + (j-1)N \in [JN] : z_0(i) \neq 0 \text{ or } z_j(i) \neq 0, i \in [N], j \in [J]\}.$$

If we further have

$$|\{j \in [J] : i \in supp(z_j)\}| \leq t, \quad \forall i \notin supp(z_0) \tag{2.19}$$

for some $t \leq \gamma$, then the null space property for JSM with constant $\frac{\gamma-t}{\gamma+t}$ relative to $(T, \mathcal{S})$ guarantees the exact recovery of X through $\mathcal{P}_Z(\gamma)$.

Proof. By Corollary 2.1.5, it is enough to show

$$\|X+V\|_\Psi > \|X\|_\Psi, \quad \forall V \in \ker(A)\backslash\{0\}.$$

Notice that $\|X+V\|_\Psi = \sum_{i\in[N]} \|(X+V)|_i\|_\Psi$, so we will first estimate $\|(X+V)|_i\|_\Psi$ for all i. There are three cases:

(a) $i \in \text{supp}(z_0)$. By triangular inequality,

$$\|(X+V)|_i\|_\Psi \geq \|X|_i\|_\Psi - \|V|_i\|_\Psi.$$

(b) $i \in \overline{T}$. Notice that $X|_i = \mathbf{0}$, so

$$\|(X+V)|_i\|_\Psi = \|V|_i\|_\Psi.$$

(c) $i \in T\backslash\text{supp}(z_0)$. Let $\mathbf{1}_{ij} = \mathbf{1}_{\{i\in\text{supp}(z_j)\}}$ be the indicator of whether $i \in \text{supp}(z_j)$ and let $\bar{\mathbf{1}}_{ij}$ be its negation. By definition,

$$\begin{aligned}\|(X+V)|_i\|_\Psi &= \inf_u \gamma|u| + \sum_{j\in[J]} |z_j(i)+v_j(i)-u|\mathbf{1}_{ij} + |v_j(i)-u|\bar{\mathbf{1}}_{ij} \\ &\geq \inf_u \gamma|u| + \sum_{j\in[J]} |z_j(i)|\mathbf{1}_{ij} - |v_j(i)|\mathbf{1}_{ij} - |u|\mathbf{1}_{ij} \\ &\quad +|v_j(i)-u|\bar{\mathbf{1}}_{ij}.\end{aligned}$$

Notice that

$$\|X|_i\|_\Psi = \gamma|z_0(i)| + \sum_{j\in[J]} |z_j(i)| = \sum_{j\in[J]} |z_j(i)|\mathbf{1}_{ij},$$

$$\|V_{\mathcal{S}}|_i\|_\Psi = \inf_u \gamma|u| + \sum_{j\in[J]} |v_j(i)-u|\mathbf{1}_{ij} + |u|\bar{\mathbf{1}}_{ij} = \sum_{j\in[J]} |v_j(i)|\mathbf{1}_{ij}.$$

(The last equality is true because $\gamma - \sum_j \mathbf{1}_{ij} \geq \gamma - t \geq 0$ and

$$\|V_{\mathcal{S}}|_i\|_\Psi \geq \sum_{j\in[J]} |v_j(i)|\mathbf{1}_{ij} + \inf_u \left(\gamma - \sum_{j\in[J]} \mathbf{1}_{ij} + \sum_{j\in[J]} \bar{\mathbf{1}}_{ij}\right)|u|.$$

Together they imply $\|V_{\mathcal{S}}|_i\|_\Psi \geq \|V_{\mathcal{S}}|_i\|_1$. On the other hand, $\|V_{\mathcal{S}}|_i\|_\Psi \leq \|V_{\mathcal{S}}|_i\|_1 = \sum_j |v_j(i)|\mathbf{1}_{ij}$.)

Substitute $\|X|_i\|_\Psi$, $\|V_{\mathcal{S}}|_i\|_\Psi$ into above and use the assumption that $\sum_j \mathbf{1}_{ij} \le t$, we have

$$\|(X+V)|_i\|_\Psi \ge \|X|_i\|_\Psi - \|V_{\mathcal{S}}|_i\|_\Psi + \inf_u h_i(u),$$

where

$$h_i(u) = (\gamma - t)|u| + \sum_{j\in[J]} |v_j(i) - u|\bar{\mathbf{1}}_{ij}.$$

Next we show $\inf_u h_i(u) \ge \frac{\gamma-t}{\gamma+t}\|\tilde{V}_{\mathcal{S}}|_i\|_\Psi$ where $\tilde{V}_{\mathcal{S}} = V|_T - V_{\mathcal{S}}$. Let

$$g_i(u) = \gamma|u| + \sum_{j\in[J]} |u|\mathbf{1}_{ij} + |v_j(i) - u|\bar{\mathbf{1}}_{ij}$$

and by definition, $\|\tilde{V}_{\mathcal{S}}|_i\|_\Psi = \inf_u g_i(u)$. Applying $\sum_j \mathbf{1}_{ij} \le t \le \gamma$ we get

$$\frac{\gamma-t}{\gamma+t} g_i(u) \le (\gamma-t)|u| + \frac{\gamma-t}{\gamma+t}\sum_{j\in[J]} |v_j(i) - u|\bar{\mathbf{1}}_{ij} \le h_i(u).$$

The inequality $\inf_u h_i(u) \ge \frac{\gamma-t}{\gamma+t}\|\tilde{V}_{\mathcal{S}}|_i\|_\Psi$ then follows by taking the infimum of u on both sides, and we have the final estimation

$$\|(X+V)|_i\|_\Psi \ge \|X|_i\|_\Psi - \|V_{\mathcal{S}}|_i\|_\Psi + \frac{\gamma-t}{\gamma+t}\|\tilde{V}_{\mathcal{S}}|_i\|_\Psi.$$

Putting (a) (b) (c) all together and by the Null Space Property for JSM,

$$\|X+V\|_\Psi \ge \|X\|_\Psi - \|V|_{\mathcal{S}}\|_\Psi + \|V|_{\overline{T}}\|_\Psi + \frac{\gamma-t}{\gamma+t}\|\tilde{V}_{\mathcal{S}}\|_\Psi > \|X\|_\Psi.$$

□

Remark 2.4.1. Although not always satisfiable, condition (2.19) in Theorem 2.4.1 is reasonable, especially when J is large. It requires that, on a coordinate whose common part is zero, only a few (no more than the weight γ) signals can have a nonzero entry. This may be seen as an interpretation of JSM-1 because if a coordinate is shared by many signals, then it is more likely that a nonzero common part exists there. We also note that by increasing the value of γ, (2.19) can be easily satisfied. However, as studied before, Ψ-norm changes towards l_1-norm as γ increases to J, so a large γ may not be able to effectively promote the common part.

Remark 2.4.2. T can also be equivalently defined in terms of X by $T = \cup\text{supp}(x_j)$. However, we define $\mathcal{S}$ only in terms of Z, rather than using $\{i+(j-1)N \in [JN] : x_j(i) \neq 0\}$, because in some rare cases like $0 = x_j(i) = z_0(i) + z_j(i)$ with $z_0(i) \neq 0$, we would still want to include $i + (j-1)N$ in the set $\mathcal{S}$.

Finally, we give the null space properties for stable and robust recovery. Based on the discussion in section 2.3, it is easy to see that we can similarly define the *stable null space property for JSM* by replacing (2.18) in Definition 2.4.1 with

$$\|V|_{\overline{T}}\|_\Psi + \rho\|\tilde{V}_{\mathcal{S}}\|_\Psi \geq \|V_{\mathcal{S}}\|_\Psi + \alpha\|V\|_\Psi, \quad \forall V \in \ker(A) \tag{2.20}$$

and similarly define the *robust null space property for JSM* by replacing (2.18) with

$$\|V|_{\overline{T}}\|_\Psi + \rho\|\tilde{V}_{\mathcal{S}}\|_\Psi + \tau\|AV\|_2 \geq \|V_{\mathcal{S}}\|_\Psi + \alpha\|V\|_\Psi, \quad \forall V \in \mathbb{R}^{JN} \tag{2.21}$$

As their counterparts to Theorem 2.4.1, we have the following theorem.

Theorem 2.4.2. *For $\hat{X} = X + E \in \mathbb{R}^{JN}$, let $Z \in \mathbb{R}^{(J+1)N}$ such that $\Phi Z = X$ and $\|Z\|_1 = \|X\|_\Psi$. Define sets T and $\mathcal{S}$ the same as in Theorem 2.4.1. Also assume that*

$$|\{j \in [N] : i \in supp(z_j)\}| \leq t, \quad \forall i \notin supp(z_0)$$

for some $t \leq \gamma$.

(i) If (2.20) holds for some $\alpha > 0$ and $\rho = \frac{\gamma-t}{\gamma+t}$, then

$$\|\hat{X} - X_\#\|_\Psi \leq (2/\alpha + 2)\|E\|_\Psi,$$

where $X_\#$ is any solution to problem:

$$\min_{W\in\mathbb{R}^{JN}} \|W\|_\Psi \ s.t.\ A\hat{X} = AW.$$

(ii) If (2.21) holds for some $\alpha > 0$, $\tau \geq 0$ and $\rho = \frac{\gamma-t}{\gamma+t}$, then

$$\|\hat{X} - X_\#\|_\Psi \leq (2/\alpha + 2)\|E\|_\Psi + (2\tau/\alpha)\varepsilon,$$

where $X_{\#}$ *is any solution to problem:*

$$\min_{W \in \mathbb{R}^{JN}} \|W\|_{\Psi} \; s.t. \; \|A\hat{X} - AW\|_2 \leq \varepsilon.$$

Proof. This is readily seen from the proof of Theorem 2.4.1 and Theorem 2.3.1 (for (i)) or Theorem 2.3.3 (for (ii)). □

At the end of this chapter we note that the (stable/robust) null space property for JSM above is a non-uniform guarantee since it relies on the support structure of the signals (i.e. $\mathcal{S}$ and T). However, by considering all possible support structures for the set of desired jointly sparse signals, we can write it into a uniform guarantee. In this case, (2.18) (or (2.20), (2.21)) holds for all possible pairs of $(T, \mathcal{S})$.

On the other hand, we do not have any condition on the matrices A_i (such as an RIP-based sufficient condition) that will ensure that such a uniform null space property holds. This is an important open problem we intend to work on as a follow up to this thesis.

Chapter 3

The Recovery of Two Signals

As a case study, in this chapter we focus on the recovery of two signals (i.e. $J = 2$). For the notations, here we have

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad Z = \begin{bmatrix} \gamma z_0 \\ z_1 \\ z_2 \end{bmatrix},$$

and

$$A = \begin{bmatrix} A_1 & \\ & A_2 \end{bmatrix}, \quad \Psi = \begin{bmatrix} \gamma^{-1} I_N & I_N & \\ \gamma^{-1} I_N & & I_N \end{bmatrix},$$

where $X \in \mathbb{R}^{2N}$, $Y \in \mathbb{R}^{m_1+m_2}$, $Z \in \mathbb{R}^{3N}$ and $A_i \in \mathbb{R}^{m_i \times N}$. We shall consider $\mathcal{P}_Z(\gamma)$ and its equivalent form $\mathcal{P}_X(\gamma)$ for recovery.

$$\mathcal{P}_Z(\gamma): \quad \min_{Z \in \mathbb{R}^{3N}} \|Z\|_1 \text{ s.t. } Y = A\Psi Z.$$

$$\mathcal{P}_X(\gamma): \quad \min_{X \in \mathbb{R}^{2N}} \|X\|_\Psi \text{ s.t. } Y = AX.$$

3.1 Ψ-Norm and Unit Ball Revisited

Recall the definition of Ψ-norm, now we have

$$\begin{aligned} \|X\|_\Psi &= \inf_{U \in \psi} \| \begin{bmatrix} 0 \\ X \end{bmatrix} + U\|_1 \\ &= \inf_{u \in \mathbb{R}^N} \|\gamma u\|_1 + \|x_1 - u\|_1 + \|x_2 - u\|_1. \end{aligned}$$

Also by Proposition 2.1.4, for every $X \in \mathbb{R}^{2N}$ and any choice of $\gamma > 0$, there exists one and only one $Z \in \mathbb{R}^{3N}$ such that $X = \Psi Z$ and $\|Z\|_1 = \|X\|_\Psi$. So when considering $\mathcal{P}_X(\gamma)$ to have a unique solution, it is the same as when $\mathcal{P}_Z(\gamma)$ has a unique solution.

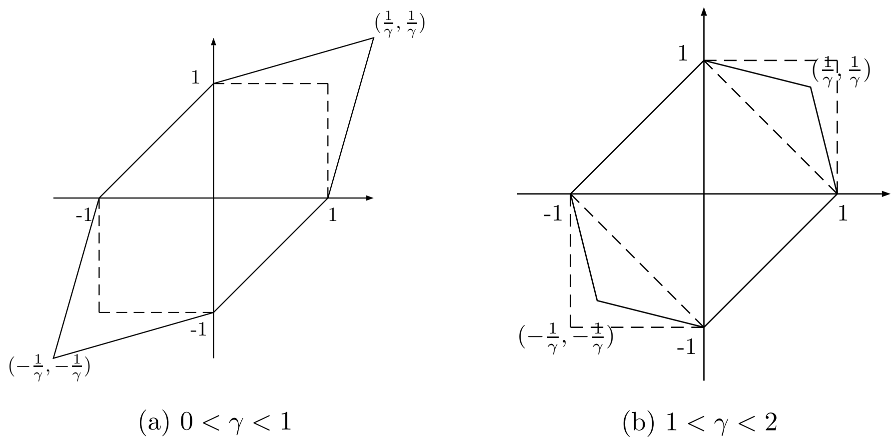


Figure 3.1: Unit ball for Ψ-norm when $J = 2$ and $N = 1$

For the unit ball of Ψ-norm, we can easily apply the results in section 2.2. Figure 3.1 shows the unit ball with $N = 1$ and γ changing from 0 to 2 (the left one is an example when $0 < \gamma < 1$ and the right one an example when $1 < \gamma < 2$). As γ increases to 2, the ball "shrinks" to l_1 ball in directions $\pm(1/\gamma, 1/\gamma)$. Particularly, it is the "mixture" of the l_1 and l_∞ ball when $\gamma = 1$. Ψ-norm also becomes l_1-norm once $\gamma \geq 2$. For an explicit formula, have the following proposition.

Proposition 3.1.1. *If $N = 1$, for $x = (a, b)^T$*

$$\|x\|_\Psi = \begin{cases} |a| + |b| & ab \leq 0 \\ |a| + |b| & ab > 0, \gamma \geq 2 \\ M + (\gamma - 1)m & ab > 0, \gamma \in (0, 2) \end{cases}$$

where $M = \max\{|a|, |b|\}$, $m = \min\{|a|, |b|\}$.

This proposition can be shown by dropping the absolute value sign of $|\gamma u| + |a - u| + |b - u|$ and finding its minimum for all a, b since

$$\|x\|_\Psi = \inf_u |\gamma u| + |a - u| + |b - u|.$$

Based on Proposition 3.1.1, we have the inequalities below.

Proposition 3.1.2. *For $\gamma > 0$ and any $u, a, b \in \mathbb{R}$, let $\theta = \min\{\gamma - 1, 1\}$, then*

$$|\gamma u| + |a - u| + |b - u| \geq \begin{cases} |a| + \theta|b| \\ |b| + \theta|a| \end{cases}.$$

Proof. First, notice that

$$|\gamma u| + |a - u| + |b - u| \geq \|(a, b)^T\|_\Psi.$$

Then by Proposition 3.1.1, let $M = \max\{|a|, |b|\}$ and $m = \min\{|a|, |b|\}$ we have

$$\|(a, b)^T\|_\Psi \geq M + \theta m \geq m + \theta M.$$

□

3.2 A Null Space Property

In chapter 2, Theorem 2.4.1 states that the null space property for JSM can guarantee the exact recovery of jointly sparse X through $\mathcal{P}_Z(\gamma)$. However, it is based on condition

$$|\{j \in [J] : i \in \operatorname{supp}(z_j)\}| \leq \gamma, \quad \forall i \notin \operatorname{supp}(z_0),$$

and in the case of $J = 2$, this condition may be too strong. In fact, if we have a $i \in [N]$ with $x_1(i)x_2(i) < 0$, then it is easy to see that $z_0(i) = 0$ and $z_j(i) = x_j(i)$ for $j = 1, 2$. Hence we must have $\gamma \geq 2$ for the above condition to hold. On the other hand, for $\mathcal{P}_Z(\gamma)$ to be useful (Ψ-norm to be different form l_1-norm), we should only focus on $0 < \gamma < 2$.

In this section we will derive a null space property just for $J = 2$. First we consider jointly sparse signals X and define its support structure.

Definition 3.2.1. For $X \in \mathbb{R}^{2N}$, $\mathcal{T}_X = \mathcal{T}_X(\overline{T}, T_0, T_1, \ldots, T_5)$ is the support model of X if

$$\begin{aligned} T_3 &= \text{supp}(z_0) \cap \text{supp}(z_1) \\ T_4 &= \text{supp}(z_0) \cap \text{supp}(z_2) \\ T_5 &= \text{supp}(z_1) \cap \text{supp}(z_2) \\ T_0 &= \text{supp}(z_0) \backslash (T_3 \cup T_4) \\ T_1 &= \text{supp}(z_1) \backslash (T_3 \cup T_5) \\ T_2 &= \text{supp}(z_2) \backslash (T_5 \cup T_4) \\ \overline{T} &= [N] \backslash (\cup_{i=0}^5 T_i) \end{aligned}$$

where $Z = (\gamma z_0^T, z_1^T, z_2^T)^T$, $\Psi Z = X$ and $\|Z\|_1 = \|X\|_\Psi$.

Figure 3.2 illustrates the supports $\overline{T}, T_0, T_1, \ldots, T_5$. We also note that there is one and only one Z satisfying the conditions above for any $X \in \mathbb{R}^{2N}$ and $\gamma > 0$, so $\mathcal{T}_X$ is well-defined. In such case we have

$$\begin{aligned} &\cap_{i=0}^2 \text{supp}(z_i) = \emptyset \\ &z_0(i) z_1(i) \geq 0, \quad i \in T_3 \\ &z_0(i) z_2(i) \geq 0, \quad i \in T_4 \\ &z_1(i) z_2(i) \leq 0, \quad i \in T_5 \end{aligned}$$

Now we can state a null space property and its result. For simplicity, consider the special case when $\gamma = 1$.

Theorem 3.2.1. *For $\gamma = 1$ and X_* with support model $\mathcal{T}_{X_*}(\overline{T}, T_0, \ldots, T_5)$. If for every $V = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in \ker(A) \backslash \{0\}$ (here $v_1, v_2 \in \mathbb{R}^N$),*

$$\sum_{\overline{T}} \max_{j \in [2]} |v_j(i)| > \sum_{T_0} \min_{j \in [2]} |v_j(i)| + \sum_{T_1 \cup T_3 \cup T_5} |v_1(i)| + \sum_{T_2 \cup T_4 \cup T_5} |v_2(i)|,$$

then X_* is the unique minimizer of $\mathcal{P}_Z(\gamma)$ with $Y = AX_*$.

Proof. By Corollary 2.1.5, we only need to show that (2.7) is satisfied. Let $X_* = \Psi Z$ such that $\|Z\|_1 = \|X\|_\Psi$. For $U \in \psi$, we have

$$U = \begin{bmatrix} u \\ -u \\ -u \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix},$$

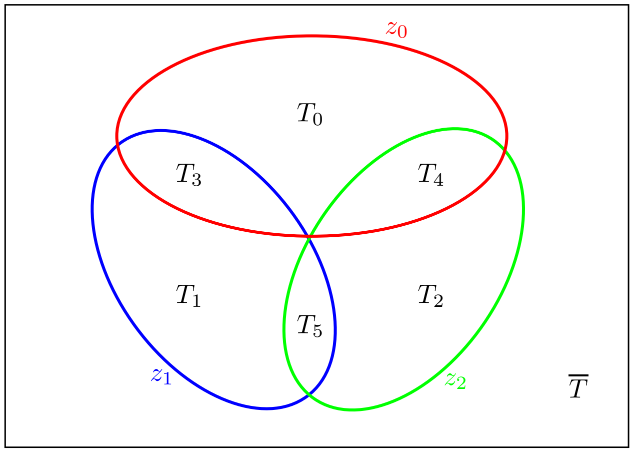


Figure 3.2: Support model illustration

so for $V \in \ker(A)\backslash\{0\}$,

$$\begin{aligned} \| \begin{bmatrix} 0 \\ X_* \end{bmatrix} + \begin{bmatrix} 0 \\ V \end{bmatrix} + U\|_1 &= \sum_{i=1}^{N} |u_i| + \sum_{j=1}^{2}\sum_{i=1}^{N} |z_0(i) + z_j(i) + v_j(i) + u_i| \\ &= \sum_{i\in[N]} f_i(U,V,Z) \end{aligned}$$

where

$$f_i(U,V,Z) = f_i = |u_i| + \sum_{j=1}^{2} |z_0(i) + z_j(i) + v_j(i) + u_i|.$$

Next we estimate f_i.

(a) $i \in \overline{T}$. We have $z_0(i) = z_1(i) = z_2(i) = 0$,

$$\begin{aligned} f_i &= |u_i| + |v_1(i) + u_i| + |v_2(i) + u_i| \\ &\geq \max_{j\in[2]} |v_j(i)| \end{aligned}$$

The inequality here follows from Proposition 3.1.2.

(b) $i \in T_0$. We have $z_1(i) = z_2(i) = 0$,

$$\begin{aligned} f_i &= |u_i| + |z_0(i) + v_1(i) + u_i| + |z_0(i) + v_2(i) + u_i| \\ &\geq |u_i| + |z_0(i) + u_i| + |z_0(i) + v_2(i) + u_i| - |v_1(i)| \\ &\geq |z_0(i)| - |v_1(i)| \end{aligned}$$

Similarly we can show that $f_i \geq |z_0(i)| - |v_2(i)|$, therefore

$$f_i \geq |z_0(i)| - \min_{j\in[2]} |v_j(i)|$$

(c) $i \in T_1$. We have $z_0(i) = z_2(i) = 0$,

$$\begin{aligned} f_i &= |u_i| + |z_1(i) + v_1(i) + u_i| + |v_2(i) + u_i| \\ &\geq |u_i| + |z_1(i) + u_i| + |v_2(i) + u_i| - |v_1(i)| \\ &\geq |z_1(i)| - |v_1(i)| \end{aligned}$$

(d) $i \in T_2$. Similar to (c), we have

$$f_i \geq |z_2(i)| - |v_2(i)|$$

(e) $i \in T_3$. We have $z_2(i) = 0$, $z_0(i)z_1(i) \geq 0$,

$$\begin{aligned} f_i &= |u_i| + |z_0(i) + z_1(i) + v_1(i) + u_i| + |z_0(i) + v_2(i) + u_i| \\ &\geq |u_i| + |z_0(i) + z_1(i) + u_i| + |z_0(i) + v_2(i) + u_i| - |v_1(i)| \\ &\geq |z_0(i) + z_1(i)| - |v_1(i)| \\ &= |z_0(i)| + |z_1(i)| - |v_1(i)| \end{aligned}$$

(f) $i \in T_4$. Similar to (e), we have

$$f_i \geq |z_0(i)| + |z_2(i)| - |v_2(i)|$$

(g) $i \in T_5$. We have $z_0(i) = 0$, $z_1(i)z_2(i) \leq 0$,

$$\begin{aligned} f_i &= |u_i| + |z_1(i) + v_1(i) + u_i| + |z_2(i) + v_2(i) + u_i| \\ &\geq |u_i| + |z_1(i) + u_i| + |z_2(i) + u_i| - |v_1(i)| - |v_2(i)| \\ &\geq |z_1(i)| + |z_2(i)| - |v_1(i)| - |v_2(i)| \end{aligned}$$

Now sum up all above, we have

$$\begin{aligned} \sum_{i\in[N]} f_i &\geq \|Z\|_1 + \sum_{\overline{T}} \max_{j\in[2]} |v_j(i)| \\ &\quad - \left(\sum_{T_0} \min_{j\in[2]} |v_j(i)| + \sum_{T_1\cup T_3\cup T_5} |v_1(i)| + \sum_{T_2\cup T_4\cup T_5} |v_2(i)| \right). \end{aligned}$$

So we can conclude that

$$\| \begin{bmatrix} 0 \\ X_* \end{bmatrix} + \begin{bmatrix} 0 \\ V \end{bmatrix} + U\|_1 > \|Z\|_1 = \|X\|_\Psi .$$

□

Lastly, we comment that although Theorem 3.2.1 is the special case for $\gamma = 1$, general case for $0 < \gamma \leq 2$ can be similarly considered using Proposition 3.1.1 and 3.1.2. For stability and robustness, the method is same as that in section 2.4.

Chapter 4

Numerical Experiments

In this chapter, we present some numerical experiments of the main algorithm studied in this thesis:

$$\mathcal{P}_Z(\gamma): \quad \min \gamma\|z_0\|_1 + \sum_{k=1}^{J} \|z_k\|_1 \text{ s.t. } y_i = A_i(z_0 + z_i),$$

where

$$A_i \in \mathbb{R}^{m_i \times N}, \quad y_i = A_i x_i, \quad i = 1, \ldots, J.$$

Continued from chapter 3, we mainly focus on the two-signal case. Similar experiments can be found in [2] [19] et al.

4.1 Sparse Signals

For the simple experiments here, we choose the number of signals $J = 2$, signal dimension $N = 50$ and the number of measurements for both signals to be the same ($m_1 = m_2$). We will calculate the recovery rate as m_i changes form 0 to N. The recovery rate is measured by the number of successful instances out of a thousand trials where each trial is conducted by first generating signals x_i and Gaussian matrices A_i, and then running joint recovery algorithm $\mathcal{P}_Z(\gamma)$. A trial is considered successful if

$$\frac{\|X_\# - X\|_2}{\|X\|_2} \leq \theta,$$

where $X_\#$ is the numerical solution of recovery algorithm and θ is the desired accuracy.

Consider the recovery for sparse signals. We will compare $\mathcal{P}_Z(\gamma)$ with the separate recovery:

$$\min \|w_i\|_1 \text{ s.t. } y_i = A_i w_i, \quad i = 1, 2.$$

Note that $\mathcal{P}_Z(\gamma)$ is useful in practice only if it outperforms the separate recovery. We would also like to see how γ can affect the joint recovery, particularly if close values of γ achieves similar recovery rates (the stability of γ).

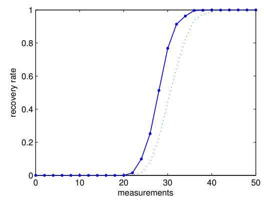
Here we take sparse vectors $z_i \in \mathbb{R}^N$ for $i = 0, 1, 2$ where each one has sparsity $\|z_i\|_0 = k_i$ and the nonzero entries are assigned to be Gaussian. The sparse signals are $x_j = z_0 + z_j$ $(j = 1, 2)$. We also take $\theta = 10^{-3}$.

Figure 4.1 is a comparison of joint (solid lines) and separate (dotted lines) sparse recovery under various settings. Here we fix $k_1 = k_2$ and $k_0 + k_1 = 10$ while changing the values of k_0 and γ. We make the following observations form this set of experiments.

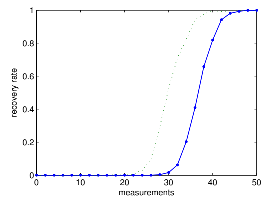
First, depending on the signals and γ, joint recovery may perform better, worse or similar to separate recovery. For example, (d) (e) (f) together shows that with fixed γ, the value of k_0 can influence whether joint recovery outperforms separate recovery; while (b) (e) (h) together shows that γ can have the same influence with fixed k_0. Second, generally speaking, the larger the common part z_0 (or k_0), the better the performance for joint recovery with $\gamma < 2$. So we should use joint recovery only when there is a relative large common part, otherwise it may be better to do separate recovery. Third, looking at (j) (k) (l), the recovery rates for both methods are almost the same regardless of k_0. This verifies our claim earlier that Ψ-norm becomes l_1-norm when $\gamma = 2$.

To further see the stability of γ. Figure 4.2 shows the joint recovery under different γ when we fix $k_0 = 8$ and $k_1 = k_2 = 2$. The dotted lines represent results for corresponding separate recoveries. These results suggest that γ is stable in this case, i.e., close values of γ give similar recovery rates.

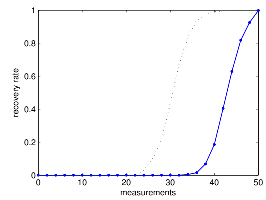
Figure 4.3 studies whether choosing Gaussian measurement matrices A_1 and A_2 independently or the same will affect the performance of $\mathcal{P}_Z(\gamma)$. This problem is more thoroughly studied in [19], which considers recoveries with the same, partly and completely independent A_1, A_2. From the figures we can see that it is better to choose A_i independently. This is consistent with the results form [19].



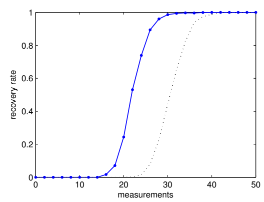
(a) $k_0 = 8, \gamma = 0.5$



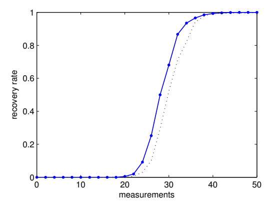
(b) $k_0 = 5, \gamma = 0.5$



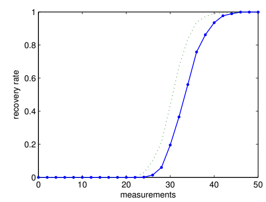
(c) $k_0 = 2, \gamma = 0.5$



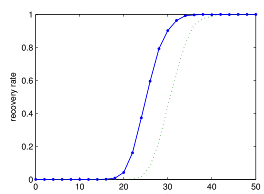
(d) $k_0 = 8, \gamma = 1$



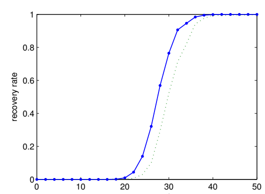
(e) $k_0 = 5, \gamma = 1$



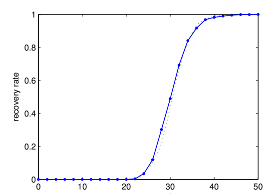
(f) $k_0 = 2, \gamma = 1$



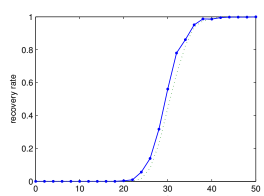
(g) $k_0 = 8, \gamma = 1.5$



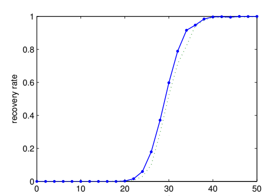
(h) $k_0 = 5, \gamma = 1.5$



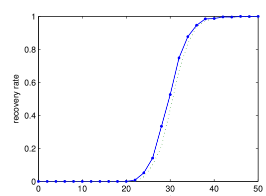
(i) $k_0 = 2, \gamma = 1.5$



(j) $k_0 = 8, \gamma = 2$

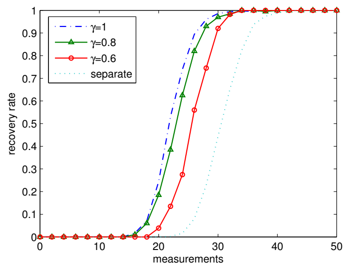


(k) $k_0 = 5, \gamma = 2$

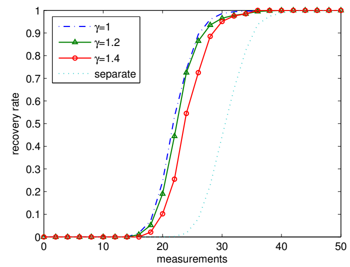


(l) $k_0 = 2, \gamma = 2$

Figure 4.1: Comparison of joint and separate sparse recovery when $J = 2$ and $N = 50$. Here we fix $k_1 = k_2$ and $k_0 + k_1 = 10$. The horizontal axis is the value of m_1 (or m_2) and the vertical axis is recovery rate. Solid lines in the figures represent joint recovery $\mathcal{P}_Z(\gamma)$ and the dotted lines represent separate recovery.

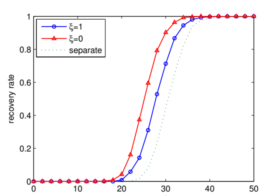


(a) $\gamma = 0.6,\, 0.8,\, 1$

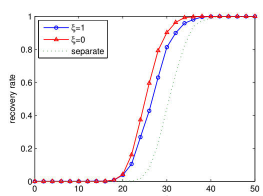


(b) $\gamma = 1,\, 1.2,\, 1.4$

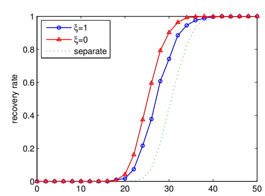
Figure 4.2: Joint sparse recovery under different γ when $J = 2$ and $N = 50$. Here we fix $k_0 = 8$ and $k_1 = k_2 = 2$. The dotted lines represent results for separate recovery. The horizontal axis is the value of m_1 (or m_2) and the vertical axis is recovery rate.



(a) $\gamma = 0.5$



(b) $\gamma = 1$



(c) $\gamma = 1.5$

Figure 4.3: Joint sparse recovery with $A_1 = A_2$ under different γ when $J = 2$ and $N = 50$. Here we fix $k_0 = 8$ and $k_1 = k_2 = 2$. ξ denotes the overlap between A_1 and A_2 such that $\xi = 1$ means $A_1 = A_2$ while $\xi = 0$ means that they are independent. The dotted lines represent results for separate recovery. The horizontal axis is the value of m_1 (or m_2) and the vertical axis is recovery rate.

4.2 Compressible Signals

We present a few experiments below to demonstrate that $\mathcal{P}_Z(\gamma)$ also works for compressible signals. For experiments with noise as well, we refer to e.g.,[19]. Here we assume the signals are

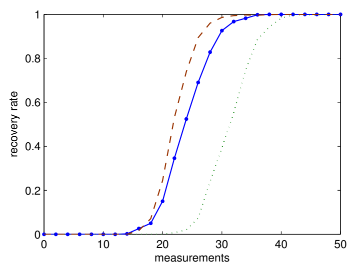
$$x_j = z_0 + z_j + e_j, \quad j = 1, 2$$

where z_i $(0 \le i \le 2)$ are the same as the previous sparse case and e_j has coefficients drawn from a zero-mean normal distribution. We also define

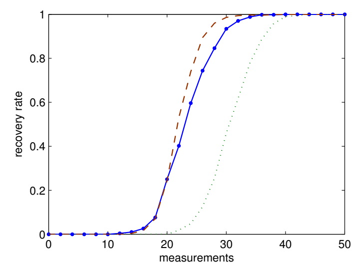
$$\eta = \max_{j \in [J]} \frac{\|e_j\|_2}{\|z_0 + z_j\|_2}$$

to quantify compressibility and take accuracy $\theta = \eta$.

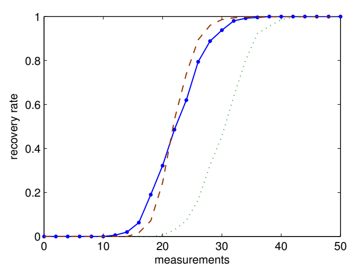
Figure 4.4 shows the joint recovery for compressible signals (joint stable recovery) under different η when $J = 2$ and $N = 50$. Here we fix $k_0 = 8$, $k_1 = k_2 = 2$ and $\gamma = 1$. The dotted lines represent separate stable recovery under the same settings and the dashed lines represent joint sparse recovery ($\eta = 0$ and $\theta = 10^{-3}$) as in Figure 4.1 (a). From these figures we can see that the recovery rates for compressible signals are similar to that of sparse signals, which verifies the stability of $\mathcal{P}_Z(\gamma)$.



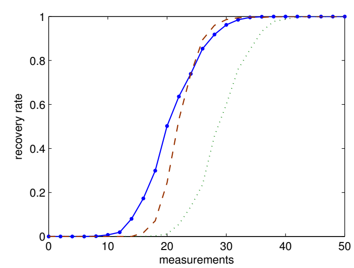
(a) $\eta = 1\%$



(b) $\eta = 3\%$



(c) $\eta = 5\%$



(d) $\eta = 10\%$

Figure 4.4: Joint stable recovery (solid lines) under different η when $J = 2$ and $N = 50$. Here we fix $k_0 = 8$, $k_1 = k_2 = 2$ and $\gamma = 1$. The dotted lines are the results for separate stable recovery and the dashed lines are for joint sparse recovery ($\eta = 0$ and $\theta = 10^{-3}$) as in Figure 4.1 (a). The horizontal axis is the value of m_1 (or m_2) and the vertical axis is recovery rate.

Chapter 5

Summary

In this thesis we studied the weighted l_1-minimization problem $\mathcal{P}_Z(\gamma)$ for the recovery of jointly sparse signals in DCS. This algorithm is first put forth in [2] by Baron et al., and then followed by a probabilistic study of the two-signal case. Here, by recognizing and defining the Ψ-norm, we adopt another approach, paralleled to the study of l_1-minimization from CS framework, to analyze $\mathcal{P}_Z(\gamma)$. This analysis eventually leads to a null space guarantee for recovery.

Chapter 2 contains the main arguments of this thesis. In section 2.1 we introduce Ψ-norm, write $\mathcal{P}_Z(\gamma)$ into $\mathcal{P}_X(\gamma)$ and prove their equivalence. This is the base of our deduction as we can then analyze $\mathcal{P}_X(\gamma)$ in a similar fashion as the l_1-minimization in CS framework. Section 2.2 takes a look at the unit ball for Ψ-norm, whose structure promotes both sparsity and similarity (all signals sharing a common part). This can be seen as a geometric interpretation for understanding the effectiveness of $\mathcal{P}_Z(\gamma)$. Motivated by the stable/robust null space property in CS, section 2.3 derives the stable and robust versions for Corollary 2.1.5 (a null space characterization for exact recovery). Section 2.4 further simplify these conditions to sufficient null space properties.

The following chapter 3 serves as a case study for $J = 2$. We first review Ψ-norm for this special case in section 3.1 and then give a sufficient null space property in section 3.2. Note that this property is not the same as the one in section 2.4 as the latter requires assumption (2.19), which may not be feasible in the two-signal case.

Chapter 4 presents some numerical experiments for $\mathcal{P}_Z(\gamma)$ and compares their results with that of separate recovery. From the results we know that $\mathcal{P}_Z(\gamma)$ is only efficient when our jointly sparse signals have a large enough common part, otherwise it may perform worse than separate recovery. We also test different values of weight γ to see its impact on the performance of $\mathcal{P}_Z(\gamma)$. The results show that γ can greatly influence $\mathcal{P}_Z(\gamma)$, but close values of γ should not lead to drastic performance change.

At the end of this thesis, we discuss several open problems and possible future works.

First problem is to prove that for Gaussian matrices A_i with enough measurements, $\mathcal{P}_Z(\gamma)$ recovers jointly sparse signals with overwhelming probability. To show this, we probably need to find another recovery guarantee (e.g. RIP-based guarantee). Moreover, this guarantee should not be stronger when compared to the ones for separate recovery, because we would also like to explain why $\mathcal{P}_Z(\gamma)$ can work better than the other.

Second problem is to determine the optimal choice of γ for recovery. Of course, as numerical results form chapter 4 suggest, this should base on some prior information on the joint sparsity model (e.g. $\{\|z_i\|_0 : 0 \leq i \leq J\}$). Further, if the optimum is impossible or to hard to find, the goal can be lowered to a suboptimal one.

Third problem is to explain why less repetition in measurement matrices A_i leads to better performance for $\mathcal{P}_Z(\gamma)$, at least when $J = 2$, as observed in chapter 4 and [19]. Moreover, an interesting phenomenon seen in [19] is also worth studying: in the two-signal case, when computing the difference between two innovation parts by subtracting the x_1, x_2 recovered from $\mathcal{P}_Z(\gamma)$, contrary to the previous result, more repetition in A_i now leads to better performance. The null space property derived in this thesis may be a way to study these questions.

Last problem is to generalize our analysis to some other sparsity models in DCS. The main obstacle here is that our analysis relies heavily on Ψ and its structure, which is determined by the joint sparsity model.

Bibliography

[1] Afonso S Bandeira, Edgar Dobriban, Dustin G Mixon, and William F Sawin. Certifying the restricted isometry property is hard. *arXiv preprint arXiv:1204.1580*, 2012.

[2] Dror Baron, Michael B Wakin, Marco F Duarte, Shriram Sarvotham, and Richard G Baraniuk. Distributed compressed sensing, 2005.

[3] T Tony Cai, Lie Wang, and Guangwu Xu. New bounds for restricted isometry constants. *Information Theory, IEEE Transactions on*, 56(9):4388–4394, 2010.

[4] T Tony Cai, Lie Wang, and Guangwu Xu. Shifting inequality and recovery of sparse signals. *Signal Processing, IEEE Transactions on*, 58(3):1300–1308, 2010.

[5] T Tony Cai and Anru Zhang. Sharp rip bound for sparse signal and low-rank matrix recovery. *Applied and Computational Harmonic Analysis*, 35(1):74–93, 2013.

[6] T Tony Cai and Anru Zhang. Sparse representation of a polytope and recovery of sparse signals and low-rank matrices. *Information Theory, IEEE Transactions on*, 60(1):122–132, 2014.

[7] Emmanuel J Candès. Compressive sampling. In *Proceedings oh the International Congress of Mathematicians: Madrid, August 22-30, 2006: invited lectures*, pages 1433–1452, 2006.

[8] Emmanuel J Candes. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9):589–592, 2008.

[9] Emmanuel J Candes, Justin K Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, 2006.

[10] Emmanuel J Candes and Terence Tao. Decoding by linear programming. *Information Theory, IEEE Transactions on*, 51(12):4203–4215, 2005.

[11] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21–30, 2008.

[12] Albert Cohen, Wolfgang Dahmen, and Ronald DeVore. Compressed sensing and best -term approximation. *Journal of the American mathematical society*, 22(1):211–231, 2009.

[13] Shane F Cotter, Bhaskar D Rao, Kjersti Engan, and Kenneth Kreutz-Delgado. Sparse solutions to linear inverse problems with multiple measurement vectors. *Signal Processing, IEEE Transactions on*, 53(7):2477–2488, 2005.

[14] Yonina C Eldar and Moshe Mishali. Robust recovery of signals from a structured union of subspaces. *Information Theory, IEEE Transactions on*, 55(11):5302–5316, 2009.

[15] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*. Springer, 2013.

[16] Moshe Mishali and Yonina C Eldar. Reduce and boost: Recovering arbitrary sets of jointly sparse vectors. *Signal Processing, IEEE Transactions on*, 56(10):4692–4702, 2008.

[17] Qun Mo and Song Li. New bounds on the restricted isometry constant δ_{2k}. *Applied and Computational Harmonic Analysis*, 31(3):460–468, 2011.

[18] Balas Kausik Natarajan. Sparse approximate solutions to linear systems. *SIAM journal on computing*, 24(2):227–234, 1995.

[19] Felix Oghenekohwo, Ernie Esser, and FJ Herrmann. Time-lapse seismic without repetition-reaping the benefits from randomized sampling and joint recovery. In *76th EAGE Conference and Exhibition 2014*, 2014.

[20] Jeonghun Park, Seunggye Hwang, Janghoon Yang, and Dongku Kim. Generalized distributed compressive sensing. *arXiv preprint arXiv:1211.6522*, 2012.

[21] Dennis Sundman, Saikat Chatterjee, and Mikael Skoglund. Methods for distributed compressed sensing. *Journal of Sensor and Actuator Networks*, 3(1):1–25, 2013.

[22] Dennis Sundman, Saikat Chatterjee, and Mikael Skoglund. Distributed greedy pursuit algorithms. *Signal Processing*, 105:298–315, 2014.

[23] Joel A Tropp, Anna C Gilbert, and Martin J Strauss. Simultaneous sparse approximation via greedy pursuit. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, volume 5, pages v–721. IEEE, 2005.