Recovering Jointly Sparse Signals via Joint Basis Pursuit

Samet Oymak Babak Hassibi

Department of Electrical Engineering
California Institute of Technology
Pasadena, CA 91125

Abstract

This work considers recovery of signals that are sparse over two bases. For instance, a signal might be sparse in both time and frequency, or a matrix can be low rank and sparse simultaneously. To facilitate recovery, we consider minimizing the sum of the ℓ_1-norms that correspond to each basis, which is a tractable convex approach. We find novel optimality conditions which indicates a gain over traditional approaches where ℓ_1 minimization is done over only one basis. Next, we analyze these optimality conditions for the particular case of time-frequency bases. Denoting sparsity in the first and second bases by k_1, k_2 respectively, we show that, for a general class of signals, using this approach, one requires as small as $O(\max\{k_1,k_2\}\log\log n)$ measurements for successful recovery hence overcoming the classical requirement of $\Theta(\min\{k_1,k_2\}\log(\frac{n}{\min\{k_1,k_2\}}))$ for ℓ_1 minimization when $k_1 \approx k_2$. Extensive simulations show that, our analysis is approximately tight.

***Index Terms*—basis pursuit, compressed sensing, phase retrieval, duality, convex optimization**

I. Introduction

Compressed sensing is concerned with the recovery of sparse vectors and has recently been the subject of immense interest. One of the main methods is Basis Pursuit (BP) where the ℓ_1 norm is minimized subject to convex constraints. Assuming $\mathbf{x}$ has a sparse representation over the basis $\mathbf{U}$ (i.e. $\mathbf{Ux}$ is a sparse vector) and assuming we get to see the observations $\mathbf{Ax}$, Basis Pursuit performs the following optimization to get back to $\mathbf{x}$.

$$\min_{\hat{\mathbf{x}}} \|\mathbf{U}\hat{\mathbf{x}}\|_1 \quad \text{subject to} \quad \mathbf{Ax} = \mathbf{A}\hat{\mathbf{x}} \qquad \text{(BP)}$$

In this work, we'll be investigating recovery of vectors that can be sparsely represented over two bases. For example, a vector such as a Dirac comb can be sparse in time and frequency. Similarly, we can consider a low rank matrix which is supported over an unknown submatrix and zero elsewhere and hence sparse. Assuming $\mathbf{x}$ is sparse over $\mathbf{U}_1, \mathbf{U}_2$, in order to induce sparsity in both bases, we will be considering the following approach, which we call Joint Basis Pursuit (JBP).

$$\min_{\hat{\mathbf{x}}} \|\mathbf{U}_1\hat{\mathbf{x}}\|_1 + \lambda\|\mathbf{U}_2\hat{\mathbf{x}}\|_2 \quad \text{s.t.} \quad \mathbf{A}\hat{\mathbf{x}} = \mathbf{Ax} \qquad \text{(JBP)}$$

For the case of a matrix $\mathbf{X}$ that is simultaneously sparse and low rank, we may minimize the summation of ℓ_1 norm and the matrix nuclear norm, which is denoted by $\|\cdot\|_\star$ and is equal to summation of the singular values. Assuming, we observe linear measurements $\mathcal{A}(\mathbf{X})$, we propose solving the following problem (JBP-Matrix) to recover $\mathbf{X}$.

$$\min_{\hat{\mathbf{X}}} \|\hat{\mathbf{X}}\|_\star + \lambda\|\hat{\mathbf{X}}\|_1 \quad \text{s.t.} \quad \mathcal{A}(\hat{\mathbf{X}}) = \mathcal{A}(\mathbf{X}) \qquad \text{(JBPM)}$$

While it is possible to come up with relevant problems, this paper will focus on JBP and JBPM. Our motivations are,

- Investigating whether JBP can outperform regular BP.
- The sparse phase retrieval problem, in which one has measurements of a sparse vector $\mathbf{x}$ and observe $|\langle \mathbf{a}_i, \mathbf{x}\rangle|^2$ as measurements [7], [8]. While it is not possible to cast this as a regular compressed sensing problem, it can be cast as JBPM where we wish to recover sparse and low rank matrix, $\mathbf{xx}^*$. This problem is known to have applications to X-Ray crystallography [6] and has recently attracted interest [7]–[10].

Background: It should be emphasized that, recently, there has been significant interest in using a combination of different norms to exploit the structure of a signal. While this paper deals with signals having sparse representations in both bases, [3]–[5] considers the problem of separating the signals that are combinations of sparsely representable incoherent pieces.

Contributions: In this work we provide sharp recovery conditions that guarantees success of JBP and JBPM. Next, we cast these conditions in a *dual certificate* framework to facilitate analysis. For the case of time-frequency bases, we analyze the dual certificate construction to find that for the class of "periodic signals", one needs at most $O(\max\{k_1,k_2\}\log\log n)$ measurements where

This work was supported in part by the National Science Foundation under grants CCF-0729203, CNS-0932428 and CCF-1018927, by the Office of Naval Research under the MURI grant N00014-08-1-0747, and by Caltech's Lee Center for Advanced Networking.

k_1, k_2 represents the sparsity in $\mathbf{U}_1, \mathbf{U}_2$. This shows that JBP can indeed outperform regular BP which requires $\Theta(k \log \frac{n}{k})$ measurements for recovery of a k sparse vector [12], [13]. Finally, simulation results indicate that our results are sharp. We believe that, the result of this paper can be seen as negative in nature. While, JBP provides an improvement, it is not a significant improvement when we consider the fact that signals that are simultaneously sparse are *few* in number.

II. Problem Setup

We begin by considering the (JBP) problem and assume $\mathbf{x} \in \mathbb{C}^n$ is a signal that is sparse over two complete bases, $\mathbf{U}_1, \mathbf{U}_2$. Later on we will briefly extend our approach to (JBPM) and the recovery of matrices that are simultaneously sparse and low rank.

The basic question we would like to answer is whether one can do better in recovering $\mathbf{x}$ from measurements $\mathbf{A}\mathbf{x}$ by exploiting the joint sparsity of $\mathbf{x}$.

Before, going into technical details, we'll introduce the relevant notation. Denote the set $\{1, 2, \dots, n\}$ by $[n]$. Let $S_1, S_2 \subseteq [n]$ denote the supports of $\mathbf{x}$ in the bases $\mathbf{U}_1$ and $\mathbf{U}_2$, i.e., locations of nonzero entries of $\mathbf{U}_1\mathbf{x}$ and $\mathbf{U}_2\mathbf{x}$ respectively. Further, let $\mathcal{S}_1(\cdot) : \mathbb{C}^n \to \mathbb{C}^{|S_1|}, \mathcal{S}_2(\cdot) : \mathbb{C}^n \to \mathbb{C}^{|S_2|}$ denote the operators that collapse a vector onto S_1, S_2 respectively. $\text{sgn}(\cdot) : \mathbb{C}^n \to \mathbb{C}^n$ is the function that returns entry wise signs of a vector, i.e., 0 is mapped to 0 and $a \neq 0$ is mapped to $\frac{a}{|a|}$. $\mathbf{I}$ will be the identity matrix of the appropriate size. Null space of a linear operator $\mathbf{A}$ is denoted by $\mathcal{N}(\mathbf{A})$. $\mathcal{R}(\cdot), \mathcal{I}(\cdot) : \mathbb{C}^n \to \mathbb{R}^n$ are the functions that returns entry-wise real and imaginary parts of a vector. Denote $\sqrt{-1}$ by $\mathbf{i}$. $\mathbf{D}$ is the Discrete Fourier Transform (DFT) matrix of the appropriate size and given as follows,

$$\mathbf{D}_{i,j} = \frac{W^{(i-1)(j-1)}}{\sqrt{n}} \quad 1 \leq i, j \leq n \tag{1}$$

where W is always $\exp(-\frac{2\pi\mathbf{i}}{n})$. We will use λ_1, λ_2 and $1, \lambda$ alternatively.

Remark: Proofs that are omitted can be found in the appendix.

A. Recovery Conditions for JBP

We will start with explaining our approach. Let $\mathbf{A} \in \mathbb{C}^{m \times n}$ where m is the number of measurements. The following lemma gives a condition that guarantees $\mathbf{x}$ to be the unique optimum of (JBP).

Lemma II.1 (Null Space Condition)**.** *Assume, for all* $\mathbf{w} \in \mathcal{N}(\mathbf{A})$*, the following holds,*

$$\sum_{i=1}^{2} \lambda_i (\mathcal{R}(\langle \textit{sgn}(\mathbf{U}_i\mathbf{x}), \mathbf{U}_i\mathbf{w}\rangle) + |\bar{\mathcal{S}}_i(\mathbf{U}_i\mathbf{w})|_1) > 0 \tag{2}$$

Then, $\mathbf{x}$ *is the unique optimizer of (JBP).*

Proof: Let $f(\hat{\mathbf{x}})$ be the cost of (JBP) , i.e., $f(\hat{\mathbf{x}}) = \sum_{i=1}^2 \lambda_i \|\mathbf{U}_i\hat{\mathbf{x}}\|_1$. Then, for any $\mathbf{w} \in \mathcal{N}(\mathbf{A})$, $f(\mathbf{x}+\mathbf{w}) - f(\mathbf{x})$ is lower bounded by the left hand side of (2), which follows from the sub gradient of the ℓ_1 norm. Hence $f(\hat{\mathbf{x}}) > f(\mathbf{x})$ for all $\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}\mathbf{x}$, $\hat{\mathbf{x}} \neq \mathbf{x}$. ■

Based on (2), the following lemma connects success of (JBP) to the existence of dual certificates.

Lemma II.2. *Assume* $\mathbf{s}_1, \mathbf{s}_2 \in \mathbb{C}^m, \mathbf{s} \in \mathbb{C}^n$ *satisfying the following conditions exist:*

- $\mathcal{S}_1(\mathbf{U}_1^{-*}(\mathbf{A}^*\mathbf{s}_1 + \mathbf{s})) = \mathcal{S}_1(\textit{sgn}(\mathbf{U}_1\mathbf{x}))$
- $\|\bar{\mathcal{S}}_1(\mathbf{U}_1^{-*}(\mathbf{A}^*\mathbf{s}_1 + \mathbf{s}))\|_\infty < 1$
- $\mathcal{S}_2(\mathbf{U}_2^{-*}(\mathbf{A}^*\mathbf{s}_2 - \mathbf{s})) = \lambda\mathcal{S}_2(\textit{sgn}(\mathbf{U}_2\mathbf{x}))$
- $\|\bar{\mathcal{S}}_2(\mathbf{U}_2^{-*}(\mathbf{A}^*\mathbf{s}_2 - \mathbf{s}))\|_\infty < \lambda$
- $\mathbf{A}$ *is invertible over* $\bigcap_{i=1}^2 \{\mathbf{v} \big| \mathcal{S}_i(\mathbf{U}_i\mathbf{v}) = \mathbf{U}_i\mathbf{v}\}$.

Then $\mathbf{x}$ *is the unique optimum of (JBP).*

Proof: What we need to show is that if such $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}$ exist and the invertibility assumption holds then the left hand side of (2) is strictly positive for all $\mathbf{w} \in \mathcal{N}(\mathbf{A})$. Assume such $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}$ exist and let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{C}^n$ to be:

$$\mathbf{v}_1 = \mathbf{U}_1^{-*}(\mathbf{A}^*\mathbf{s}_1 + \mathbf{s}) \quad \text{and} \quad \mathbf{v}_2 = \mathbf{U}_2^{-*}(\mathbf{A}^*\mathbf{s}_2 - \mathbf{s}^*) \tag{3}$$

Observe that for any $\mathbf{w} \in \mathcal{N}(\mathbf{A})$, using $\mathbf{A}\mathbf{w} = 0$,

$$\sum_{i=1}^{2} \langle \mathbf{U}_i\mathbf{w}, \mathbf{v}_i \rangle = \sum_{i=1}^{2} \mathbf{w}^*\mathbf{A}^*\mathbf{s}_i + \mathbf{w}^*\mathbf{s} - \mathbf{w}^*\mathbf{s} = 0 \tag{4}$$

To end the proof observe that $\mathbf{v}_1, \mathbf{v}_2$ satisfies the conditions listed in Lemma II.2 which implies that the LHS of (2) is strictly positive when combined with (4). This follows from the fact that either $\bar{\mathcal{S}}_1(\mathbf{U}_1\mathbf{w})$ or $\bar{\mathcal{S}}_2(\mathbf{U}_2\mathbf{w})$ is nonzero due to invertibility assumption. ■

The dual certificate approach for regular BP has been used in [1], [2], [5]. Letting $\mathbf{U} = \mathbf{U}_1$, compared to Lemma II.2, it requires invertibility of $\mathbf{A}$ over $\{\mathbf{v} \big| \mathcal{S}_1(\mathbf{U}_1\mathbf{v}) = \mathbf{U}_1\mathbf{v}\}$ rather than the intersection and it requires $\|\bar{\mathcal{S}}_1(\mathbf{U}_1^{-*}\mathbf{A}^*\mathbf{s}_1)\|_\infty < 1$, while Lemma II.2 can overcome this by making use of the extra variable $\mathbf{s}$. From this perspective, JBP can be viewed as a combination of two regular BP's that are allowed to "help" each other via $\mathbf{s}$.

III. Main Results

Our main result is concerned with the time-frequency bases, i.e., Identity and the DFT matrices. Before stating the main result, let us first describe the setting for which it holds.

Definition III.1. S *is a* l *periodic subset of* $[n]$ *if* n *is divisible by* l *and for any* $i \in [n]$*, we have,*

$$i \in S \iff j \in S \textit{ for all } j \textit{ such that } j \equiv i \textit{ (mod } l) \tag{5}$$

Observe that if S *is a* l *periodic support,* $|S|$ *is divisible by* n/l.

Theorem III.1. *Let $\mathbf{U}_1 = \mathbf{I}$, $\mathbf{U}_2 = \mathbf{D}$, $1 > \alpha \geq 0$ be an arbitrary constant and without loss of generality assume $|S_1| \leq |S_2|$. Further, assume the followings hold,*

- $|S_1| \leq \frac{n}{\log n}$.
- *S_1, S_2 are n_1, n_2 periodic supports, where $n = n_1 n_2$.*
- $|S_2| \leq |S_1| \log^{\alpha}(n)$.

Then, for the following scenarios, $\mathbf{x}$ can be successfully recovered via JBP with high probability (for sufficiently large n) when the matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$ is generated with i.i.d complex Gaussian entries.

- *If $|S_2| \leq |S_1| \log\log n$ setting $\lambda = 1$ and using $m = O(|S_2| \log\log n)$ measurements.*
- *If $|S_2| \geq |S_1| \log\log n$, setting $\lambda = \log^{-1}(n)$ and using $m = O(|S_2|)$ measurements.*

Remark: Our proof approach will inherently require $m \geq \max\{|S_1|, |S_2|\}$. Consequently, if $|S_2| \geq |S_1| \log(n)$, then one can already perform the regular ℓ_1 optimization over $\mathbf{U}_1 = \mathbf{I}$ to ensure recovery with $m = O(|S_2|)$ measurements. Hence, $|S_2| \leq |S_1| \log^{\alpha}(n)$ is a reasonable assumption.

A. Signals with Periodic Supports

Theorem III.1 holds for signals whose supports are periodic with n_1, n_2 over $\mathbf{I}$ and $\mathbf{D}$ respectively, where $n = n_1 n_2$. Here, we give a family of such signals that satisfy this requirement. Let T be the set of signals $\mathbf{v} \in \mathbb{C}^n$ such that for some $l \leq n_1$ and $0 \leq t < n$,

$$v_j = \begin{cases} 0 \text{ if } j \not\equiv l \pmod{n_1} \\ W^{jt} \text{ else} \end{cases} \tag{6}$$

Basically, T is the set of Dirac combs with period n_1 and hence for any $\mathbf{v} \in T$, $\mathbf{D}\mathbf{v}$ will have $\frac{n}{n_1}$ periodic support. In general, almost all $\mathbf{x}$ of the form,

$$\mathbf{x} = \sum_{\mathbf{v}_i \in T} \alpha_i \mathbf{v}_i \tag{7}$$

will have n_1 periodic support and $\mathbf{D}\mathbf{x}$ will have $\frac{n}{n_1}$ periodic support. The reason we say almost all is because cancellations may occur when $\mathbf{v}_i$'s are added. However, if α_j's are chosen from a continuous distribution, the chance of cancellation is 0.

B. Converse Results

We should emphasize that, the main reason we have considered the $\mathbf{I}, \mathbf{D}$ pair is the fact that almost all bases $\mathbf{U}_1$ and $\mathbf{U}_2$ do not permit signals that are sparse in both. The following lemma illustrates this.

Lemma III.1. *Assume $\mathbf{U}_1^{-1}, \mathbf{U}_2^{-1}$ have i.i.d entries chosen from a continuous distribution. Then, with probability 1, there exists no nonzero vector $\mathbf{x}$ satisfying $|S_1| + |S_2| \leq n$.*

An interesting work by Tao shows that, such results are true even for highly structured bases, [14]. In particular, if n is a prime number, we still have $|S_1| + |S_2| > n$ requirement for a signal over $\mathbf{U}_1 = \mathbf{I}$ and $\mathbf{U}_2 = \mathbf{D}$ bases.

IV. Proof of Theorem III.1

This section will be dedicated to the analysis of Lemma II.2 to prove Theorem III.1. We start by proposing a construction for $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}$ that certifies optimality of $\mathbf{x}$.

A. Construction of $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}$

For the following discussion, we'll be using $(\mathbf{U}_1, \mathbf{U}_2)$ and $(\mathbf{I}, \mathbf{D})$ and $(1, \lambda)$ and (λ_1, λ_2) interchangeably. The construction of $\mathbf{s}_1, \mathbf{s}_2$ will follow a classical approach previously used in [2], [5], [7]. Letting $\mathbf{A}_{S_1} \in \mathbb{C}^{m\times|S_1|}$ denote the submatrix by choosing columns corresponding to S_1 and $\mathbf{B} = \mathbf{A}\mathbf{D}^*$, we will use the following $\mathbf{s}_1, \mathbf{s}_2$.

$$\mathbf{s}_1 = \mathbf{A}_{S_1}(\mathbf{A}_{S_1}^*\mathbf{A}_{S_1})^{-1}\mathcal{S}_1(\text{sgn}(\mathbf{x})) \tag{8}$$

$$\mathbf{s}_2 = \mathbf{B}_{S_2}(\mathbf{B}_{S_2}^*\mathbf{B}_{S_2})^{-1}\lambda\mathcal{S}_2(\text{sgn}(\mathbf{D}\mathbf{x})) \tag{9}$$

Since $\mathbf{I}, \mathbf{D}$ are unitary we have $\mathbf{U}_i^{-*} = \mathbf{U}_i$. By construction $\mathbf{s}_1, \mathbf{s}_2$ already satisfies,

$$\mathcal{S}_i(\mathbf{U}_i\mathbf{A}^*\mathbf{s}_i) = \lambda_i\mathcal{S}_i(\text{sgn}(\mathbf{U}_i\mathbf{x})) \quad i \in \{1, 2\} \tag{10}$$

However, one has to control the term $\|\bar{\mathcal{S}}_i(\mathbf{U}_i\mathbf{A}^*\mathbf{s}_i)\|_\infty$ and we will make use of $\mathbf{s}$ to achieve this. Denote $\mathbf{U}_i\mathbf{A}^*\mathbf{s}_i$ by $\mathbf{y}_i$. Define the vectors $\{\mathbf{b}_1, \mathbf{b}_2\}$ as follows:

$$\mathcal{R}(b_{i,j}) = \begin{cases} 0 \text{ if } j \in S_i \\ 0 \text{ if } j \in \bar{S}_i \text{ and } |\mathcal{R}(y_{i,j})| \leq \lambda_i/4 \\ \mathcal{R}(y_{i,j}) - \lambda_i \text{sgn}(\mathcal{R}(y_{i,j}))/4 \text{ else} \end{cases} \tag{11}$$

and imaginary part $\mathcal{I}(b_{i,j})$ is obtained from $\mathcal{I}(y_{i,j})$ in the same way. Observe that, $\|\bar{\mathcal{S}}_i(\mathbf{y}_i - \mathbf{b}_i)\|_\infty < \lambda_i/2$. Based on $\{\mathbf{b}_i\}_{i=1}^2$ construct $\mathbf{s}$ as follows,

$$\begin{aligned} &\mathbf{s} = \mathbf{D}^*(\mathbf{b}_2 - \mathbf{c}_2) - \mathbf{I}(\mathbf{b}_1 - \mathbf{c}_1) \quad \text{where} \\ &\mathbf{c}_1 = \mathbf{D}^*\mathbf{I}_{S_2}\mathbf{D}\mathbf{b}_1,\ \mathbf{c}_2 = \mathbf{D}\mathbf{I}_{S_1}\mathbf{D}^*\mathbf{b}_2 \end{aligned} \tag{12}$$

Here, $\mathbf{I}_{S_1}$, $\mathbf{I}_{S_2}$ are diagonal matrices whose diagonal entries corresponding to S_1, S_2 are 1 and the rest are zero.

Lemma IV.1. *Assume $\mathbf{x}, \{\mathbf{y}_i, \mathbf{b}_i, \mathbf{c}_i\}_{i=1}^2$ are the same as described previously. Then, one has the following:*

$$\mathcal{S}_1(\mathbf{y}_1 + \mathbf{s}) = \mathcal{S}_1(sgn(\mathbf{x}))$$
$$\mathcal{S}_2(\mathbf{y}_2 - \mathbf{D}\mathbf{s}) = \lambda\mathcal{S}_2(sgn(\mathbf{D}\mathbf{x}))$$
$$\|\bar{\mathcal{S}}_1(\mathbf{y}_1 + \mathbf{s})\|_\infty < \frac{1}{2} + \|\bar{\mathcal{S}}_1(\mathbf{c}_1)\|_\infty + \|\bar{\mathcal{S}}_1(\mathbf{D}^*\mathbf{b}_2)\|_\infty$$
$$\|\bar{\mathcal{S}}_2(\mathbf{y}_2 - \mathbf{D}\mathbf{s})\|_\infty < \frac{\lambda}{2} + \|\bar{\mathcal{S}}_2(\mathbf{c}_2)\|_\infty + \|\bar{\mathcal{S}}_2(\mathbf{D}\mathbf{b}_1)\|_\infty$$

Based on Lemma IV.1 and Lemma II.2, JBP recovers $\mathbf{x}$ if we have, $\|\bar{\mathcal{S}}_1(\mathbf{c}_1)\|_\infty + \|\bar{\mathcal{S}}_1(\mathbf{D}^*\mathbf{b}_2)\|_\infty \leq 1/2$ and $\|\bar{\mathcal{S}}_2(\mathbf{c}_2)\|_\infty + \|\bar{\mathcal{S}}_2(\mathbf{D}\mathbf{b}_1)\|_\infty \leq \lambda/2$.

As a next step, we can analyze $\|\bar{\mathcal{S}}_1(\mathbf{D}^*\mathbf{I}_{S_2}\mathbf{D}\mathbf{b}_1)\|_\infty$ and $\|\bar{\mathcal{S}}_1(\mathbf{D}^*\mathbf{b}_2)\|_\infty$ and find the conditions that guarantees their sum to be small. The analysis for S_2 will be identical to S_1 and hence is omitted.

B. Probabilistic Analysis

Assume $\mathbf{A}$ is i.i.d complex normal with variance $\frac{1}{m}$ and $m \geq 64\max\{|S_1|,|S_2|\}$. This will guarantee,

$$\sigma_{min}(\mathbf{A}_{S_1}) \geq 1/\sqrt{2} \text{ and } \sigma_{min}(\mathbf{B}_{S_2}) \geq 1/\sqrt{2} \quad (13)$$

with probability $1-\exp(-\Omega(m))$, [11].

Now, conditioned on $\mathbf{A}_{S_1}, \mathbf{B}_{S_2}$ satisfy (13),

$$\|\mathbf{s}_i\|_2^2 = \mathbf{s}_i^*\mathbf{s}_i = \lambda_i^2 \text{sgn}(\mathbf{x})^*(\mathbf{A}_{S_1}^*\mathbf{A}_{S_1})^{-1}\text{sgn}(\mathbf{x}) \leq 2\lambda_i^2|S_i|$$

and $\bar{\mathcal{S}}_i(\mathbf{y}_i)$ is an i.i.d Gaussian vector whose entries having variance $\frac{\|\mathbf{s}_i\|_2^2}{m}$. Given these, we need to understand, when can we make sure,

$$\|\bar{\mathcal{S}}_1(\mathbf{D}^*\mathbf{I}_{S_2}\mathbf{D}\mathbf{b}_1)\|_\infty \leq \frac{1}{4} \text{ and } \|\bar{\mathcal{S}}_1(\mathbf{D}^*\mathbf{b}_2)\|_\infty \leq \frac{1}{4}$$

From (11), observe that $\bar{\mathcal{S}}_i(\mathbf{b}_i)$ is a function of $\bar{\mathcal{S}}_i(\mathbf{y}_i)$ which is i.i.d. random Gaussian. The next lemma, gives a characterization of $\mathbf{b}_i$.

Lemma IV.2. *Assume $m \geq 64\max\{|S_1|,|S_2|\}$. Then, the entries $\{\bar{\mathcal{S}}_i(\mathbf{b}_i)_j\}_{j=1}^{|\bar{S}_i|}$ of $\bar{\mathcal{S}}_i(\mathbf{b}_i)$ are i.i.d. random variables with the following distribution,*

$$\bar{\mathcal{S}}_i(\mathbf{b}_i)_j \text{ is } \begin{cases} 0 \text{ with probability at least } 1-4\exp(-\frac{m}{16|S_i|}) \\ \text{otherwise distributed as } \mathbf{z} \end{cases} \quad (14)$$

where $\mathbf{z}$ is 0 mean and subgaussian norms (see [11]) of $\mathcal{R}(\mathbf{z}), \mathcal{I}(\mathbf{z})$ are upper bounded by $c_0\lambda_i\sqrt{\frac{|S_i|}{m}}$ for an absolute constant $c_0 > 0$.

1) Analysis of $\|\bar{\mathcal{S}}_1(\mathbf{c}_1)\|_\infty$: We need to show,

$$\|\bar{\mathcal{S}}_1(\mathbf{D}^*\mathbf{I}_{S_2}\mathbf{D}\mathbf{b}_1)\|_\infty \leq \frac{1}{4} \quad (15)$$

Calling $\mathbf{C} = \mathbf{D}^*\mathbf{I}_{S_2}\mathbf{D}$, from Lemma VII.1, each row of $\mathbf{C}$ has energy $\frac{|S_2|}{n}$. Let $\mathbf{c}_i$ be the i'th column of $\mathbf{C}^*$. Then, using Lemma IV.2 and Proposition 5.10 of [11], for any i and an absolute constant $c > 0$,

$$\mathbb{P}(|\mathbf{c}_i^*\mathbf{b}_1| \geq \frac{1}{4}) \leq 12\exp(-\frac{mc}{2^7c_0^2|S_1|\|\mathbf{c}_{i,T}\|_2^2}) \quad (16)$$

$$= 12\exp(-\frac{mnc}{2^7c_0^2|S_1||S_2|}) \quad (17)$$

Using a union bound over all i's, shows (15) reduces to arguing $n\mathbb{P}(|\mathbf{c}_i^*\mathbf{b}_1| \geq \frac{1}{4}) \to 0$ which is equivalent to ensuring,

$$\frac{mnc}{2^7c_0^2|S_1||S_2|} - \log n \to \infty \quad \text{as} \quad n \to \infty \quad (18)$$

Using $n \geq \min\{|S_1|,|S_2|\}\log n$ in the statement of Theorem III.1, (18) holds for $m \geq 2^8c^{-1}c_0^2\max\{|S_1|,|S_2|\} = O(\max\{|S_1|,|S_2|\})$ as desired.

2) Analysis of $\|\bar{\mathcal{S}}_1(\mathbf{D}^*\mathbf{b}_2)\|_\infty$: In a similar fashion, we would like to show,

$$\|\bar{\mathcal{S}}_1(\mathbf{D}^*\mathbf{b}_2)\|_\infty \leq \frac{1}{4} \quad (19)$$

holds with high probability, to conclude. Each row of $\mathbf{D}^*$ has unit ℓ_2 norm and nonzero entries of $\mathbf{b}_2$ are i.i.d subgaussians from Lemma IV.2. Letting, $p = 4\exp(-\frac{m}{16|S_2|})$ and applying a Chernoff bound w.p.a.l $1-\exp(-np/4)$, number of non zeros in $\mathbf{b}_2$ is at most $2np$. Considering the inner products between each row of $\mathbf{D}^*$ and $\mathbf{b}_2$, and using a union bound, (19) holds, with probability at least,

$$1 - 12n\exp(-\frac{mc}{2^8c_0^2\lambda^2|S_2|p}) - \exp(-np/4) \quad (20)$$

Assuming $m = O(|S_2|\log^\alpha(n))$ for some $\alpha < 1$, we have $\exp(-np/4) \to 0$. Finally, to show the second term in (20) approaches 0, for some absolute constants $c_1, c_2 > 0$, we need to argue,

$$\frac{m}{c_1\lambda^2|S_2|}\exp(\frac{m}{c_2|S_2|}) - \log n \to \infty \quad \text{as} \quad n \to \infty \quad (21)$$

Following the same arguments for the other basis will yield,

$$\frac{m\lambda^2}{c_1|S_1|}\exp(\frac{m}{c_2|S_1|}) - \log n \to \infty \quad \text{as} \quad n \to \infty \quad (22)$$

By choosing $m = O(\max\{|S_1|,|S_2|\}\log\log n)$ and $\lambda = 1$ one can always satisfy these. In case $|S_2| \geq |S_1|\log\log n$, choose $\lambda = \log^{-1}(n)$ and m sufficiently large but $O(|S_2|)$ to still satisfy both.

V. Empirical Results

While Theorem III.1 shows that JBP can indeed outperform BP it is important to understand how good it actually is. We considered the following basic setup: Let k be a positive integer and $n = k^2$. Then, let $\mathbf{x} \in \mathbb{R}^n$ be the following dirac comb,

$$x_i = 1 \text{ if } i \equiv 1 \pmod{k} \text{ and } 0 \text{ else} \quad (23)$$

It is clear that $\mathbf{D}\mathbf{x} = \mathbf{x}$ hence the signal is only $\sqrt{n}$ sparse in both domain and the optimal weight in JBP is $\lambda = 1$ by symmetry. Simulation for JBP is performed for $k = \{2,4,6,\dots,32\}$ and for $1 \leq m \leq 30$. Interestingly, in order to achieve 50% success, JBP required $\frac{k}{2} \leq m \leq \frac{3k}{5}$ and $\frac{m}{k}$ slightly increased as a function of k. This is shown as the straight line in Figure 1. These results are quite consistent with Theorem III.1 from which we expect to have $m = O(k\log\log k)$ measurements.

On the other hand, 50% success curve for BP is shown as the dashed line in Figure 1 and obeys $m = O(k\log k)$ as expected from classical results on ℓ_1 minimization. In particular $\frac{m}{k}$ increases from 1 to 2.4 as k moves from 2 to 32. While JBP outperforms BP in this setting, the

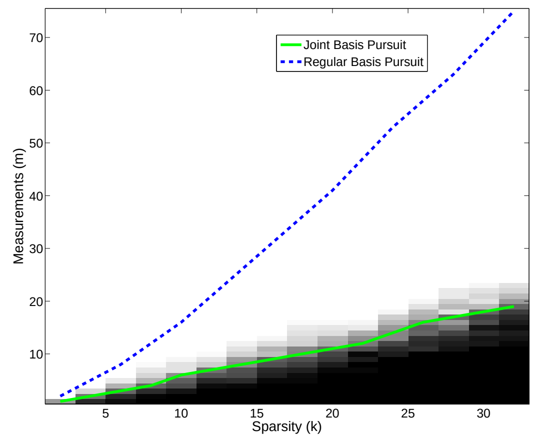

Fig. 1. Phase transitions of JBP vs BP where sparsity k varies between 2 to 32 and $n = k^2$. Dark region indicates failure for JBP while light region corresponds success. Straight and dashed lines are 50% success curves for JBP and BP respectively. While JBP outperforms BP, it still requires $\Omega(k)$ measurements.

fact that it requires $\Omega(k)$ samples to recover a highly structured signal is disappointing. It would be interesting to see whether a greedy algorithm can be developed to attack this problem.

VI. Extension to Matrices

As it has been discussed in the introduction, similar to jointly sparse signals one might as well consider matrices that are sparse and low rank. The motivation is the sparse phase retrieval problem where $\mathbf{x}$ is a sparse vector to be recovered from observations $\{|\langle \mathbf{a}_i, \mathbf{x}\rangle|^2\}_{i=1}^m$ where $\{\mathbf{a}_i\}_{i=1}^m \in \mathbb{C}^n$ are the measurement vectors. Although, these measurements are not linear in $\mathbf{x}$, they are linear in $\mathbf{x}\mathbf{x}^*$ as $|\langle \mathbf{a}_i, \mathbf{x}\rangle|^2 = \mathbf{a}_i^*\mathbf{x}\mathbf{x}^*\mathbf{a}_i$. Using the fact that $\mathbf{x}\mathbf{x}^*$ is rank 1 and sparse, JBPM can be used in order to recover $\mathbf{X} = \mathbf{x}\mathbf{x}^*$ as it will enforce a low-rank and sparse solution.

Although, this work will not deal with the analysis of this problem, we'll point out that our framework for JBP can be used for JBPM as well. In general, assume matrix $\mathbf{X}$ is low-rank and sparse and we wish to recover it from observations $\mathcal{A}(\mathbf{X})$. Let us first introduce notation relevant to structure of $\mathbf{X} \in \mathbb{C}^{n\times n}$.

- Let $S \in [n] \times [n]$ be the usual support of $\mathbf{X}$ and $\mathcal{S} : \mathbb{C}^{n\times n} \to \mathbb{C}^{|S|}$ be the projection onto S.
- Assuming $\mathbf{X}$ has singular value decomposition $\mathbf{U}\Sigma\mathbf{V}^*$, Define the subspace $\mathcal{L} \in \mathbb{C}^{n\times n}$ as,
$$\mathcal{L} = \{\mathbf{Y} \in \mathbb{C}^{n\times n} \bigm| (\mathbf{I} - \mathbf{U}\mathbf{U}^*)\mathbf{Y}(\mathbf{I} - \mathbf{V}\mathbf{V}^*) = 0\}$$
- $\bar{\mathcal{L}}$ denotes complement of $\mathcal{L}$ and projection onto $\mathcal{L}$ is denoted by $\mathcal{L}(\cdot) : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$.
- $\mathcal{A}^*(\cdot) : \mathbb{C}^m \to \mathbb{C}^{n\times n}$ denotes the adjoint operator. Operator norm is denoted by $\|\cdot\|$.

The following lemma is effectively equivalent to Lemma II.2 and characterizes a simple condition for $\mathbf{X}$ to be unique optimizer of JBPM.

Lemma VI.1. *Assume $\mathbf{S}_1, \mathbf{S}_2 \in \mathbb{C}^m, \mathbf{S} \in \mathbb{C}^{n\times n}$ satisfying the following conditions exist:*

- $\mathcal{L}(\mathcal{A}^*(\mathbf{S}_1) + \mathbf{S}) = \mathbf{U}\mathbf{V}^*$.
- $\|\bar{\mathcal{L}}(\mathcal{A}^*(\mathbf{S}_1) + \mathbf{S})\| < 1$.
- $\mathcal{S}(\mathcal{A}^*(\mathbf{S}_2) - \mathbf{S}) = \lambda \cdot \mathcal{S}(sgn(\mathbf{A}))$.
- $\|\bar{\mathcal{S}}(\mathcal{A}^*(\mathbf{S}_2) - \mathbf{S})\|_\infty < \lambda$.
- $\mathcal{A}(\cdot)$ *is invertible over* $\{\mathbf{Y} \big| \mathcal{L}(\mathbf{Y}) = \mathcal{S}(\mathbf{Y}) = \mathbf{Y}\}$.

Then $\mathbf{A}$ is the unique optimum of (JBPM).

Finally, it would be interesting to see whether similar or better improvements can be shown for JBPM over regular BP or regular nuclear norm minimization algorithms.

References

[1] E. J. Candès and B. Recht, "Simple Bounds for Low-complexity Model Reconstruction", arXiv:1106.1474v1.

[2] J. A. Tropp, "On the Conditioning of Random Subdictionaries", Appl. Comput. Harmon. Anal., vol. 25, pp. 1-24, 2008.

[3] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust Principal Component Analysis?", Journal of ACM 58(1), 1-37.

[4] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and Alan S. Willsky, "Rank-Sparsity Incoherence for Matrix Decomposition", SIAM Journal on Optimization, Vol. 21, Issue 2, pp. 572-596, 2011.

[5] E. J. Candès and J. Romberg, "Quantitative Robust Uncertainty Principles and Optimally Sparse Decompositions", Found. of Comput. Math., 6 227-254.

[6] R. P. Millane, "Phase retrieval in crystallography and optics," J. Opt. Soc. Am. A, 1990.

[7] E. J. Candès, T. Strohmer, and V. Voroninski, "PhaseLift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming", To appear in Communications on Pure and Applied Mathematics.

[8] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski "Phase Retrieval via Matrix Completion", arXiv:1109.0573v2

[9] H. Ohlsson, A. Y. Yang, R. Dong, and S. S. Sastry, "Compressive Phase Retrieval From Squared Output Measurements Via Semidefinite Programming", arXiv:1111.6323v2

[10] Y. M. Lu and M. Vetterli, "Sparse spectral factorization: unicity and reconstruction algorithms," in IEEE Trans. Acoust., Speech, and Signal Process., 2011.

[11] R. Vershynin "Introduction to the non-asymptotic analysis of random matrices", available at arXiv:1011.3027v7

[12] D. L. Donoho, "High-dimensional centrally symmetric polytopes with neighborliness pro- portional to dimension", Discrete Comput. Geometry 35 (2006), 617-652.

[13] K. D. Ba, P. Indyk, E. Price, and D. P. Woodruff, "Lower Bounds for Sparse Recovery", SODA 2010.

[14] T. Tao, "An uncertainty principle for cyclic groups of prime order", Mathematical Research Letters 12, 121-127 (2005).

VII. Appendix

We will start by proving Lemma III.1 using a classical argument.

Proof of Lemma III.1: Let us first fix S_1, S_2 and consider these particular supports. Let $\mathbf{C}_i \in \mathcal{R}^{n\times|S_i|}$ be the matrix obtained by taking columns of $\mathbf{U}_i^{-1}$, over S_i. If $\mathbf{z}_1 = \mathbf{U}_1\mathbf{x}$ and $\mathbf{z}_2 = \mathbf{U}_2\mathbf{x}$ are supported over S_1, S_2, we may write:

$$0 = \mathbf{U}_1^{-1}\mathbf{z}_1 - \mathbf{U}_2^{-1}\mathbf{z}_2 = [\mathbf{C}_1\ \ \mathbf{C}_2][\mathcal{S}_1(\mathbf{z}_1)^* \ \ -\mathcal{S}_2(\mathbf{z}_2)^*]^* \tag{24}$$

By assumption, $[\mathbf{C}_1\ \ \mathbf{C}_2] \in \mathcal{R}^{n\times(|S_1|+|S_2|)}$ has i.i.d. entries from a continuous distribution and hence full column rank with probability 1 whenever $|S_1|+|S_2| \leq n$. It follows that only $(\mathbf{z}_1, \mathbf{z}_2)$ satisfying (24) is $(0,0)$. There are finitely many S_1, S_2 pairs satisfying $|S_1|+|S_2| \leq n$ hence a union bound will still give, with probability 1, there exists no nonzero vector $\mathbf{x}$ having combined sparsities of $\mathbf{U}_1\mathbf{x}$ and $\mathbf{U}_2\mathbf{x}$ at most n. ■

Following lemma gives a simple but useful property of the DFT matrix.

Lemma VII.1. *Let $n = n_1n_2$ and S_1, S_2 be n_1 and n_2 periodic supports. Let $\mathbf{D} \in \mathbb{R}^{n\times n}$ be the DFT matrix as previously. Further, let $\mathbf{C} = \mathbf{D}^*\mathbf{I}_{S_2}\mathbf{D}$. Then,*

1) $\mathbf{C}_{i,j} = 0$ *for any (i,j) with $i \not\equiv j$ (mod n_1).*
2) *For any i, i'th row $\mathbf{r}_i$ of $\mathbf{C}$ satisfies $\|\mathbf{c}_i\|_2^2 = \frac{|S_2|}{n}$.*
3) *For any $\mathbf{x}$ that is supported on $\bar{S}_1$, we have, $\mathcal{S}_1(\mathbf{C}\mathbf{x}) = 0$.*
4) *First three results similarly hold for $\mathbf{C} = \mathbf{D}\mathbf{I}_{S_2}\mathbf{D}^*$.*

Proof: Let us start by analyzing the matrix $\mathbf{D}^*\mathbf{I}_{S_2}\mathbf{D}$. Let $\mathbf{d}_i$ be the i'th column of $\mathbf{D}$. Then,

$$\mathbf{C}_{i,j} = \mathbf{d}_i^*\mathbf{I}_{S_2}\mathbf{d}_j = \sum_{k\in S_2}\mathbf{d}_{i,k}^*\mathbf{d}_{j,k} \tag{25}$$

Using S_2 is n_2 periodic, for some set $T \in [n_1]$ (which is simply S_2 (mod n_1)), we may write,

$$\mathbf{C}_{i,j} = \sum_{t\in T}\sum_{c=1}^{n_1}\mathbf{c}_{i,t+cn_2}^*\mathbf{c}_{j,t+cn_2} \tag{26}$$

Next, for any $i \not\equiv j$ (mod n_1) and any $t \leq n_1$,

$$\begin{aligned}\sum_{c=0}^{n_1-1}\mathbf{c}_{i,t+cn_2}^*\mathbf{c}_{j,t+cn_2} &= \sum_{c=0}^{n_1-1}W^{(j-i)(t+cn_2)}\\ &= W^{(j-i)t}\sum_{c=0}^{n_1-1}W^{c(j-i)n_2}\\ &= n_1W^{(j-i)t}\delta(i-j \text{ (mod } n_1))\end{aligned}$$

where $\delta(k) = 1 \iff k \neq 0$. This proves the first statement. To show the second, $\mathbf{r}_i = \mathbf{d}_i^*\mathbf{I}_{S_2}\mathbf{D}$ implies:

$$\|\mathbf{r}_i\|_2^2 = \mathbf{r}_i^*\mathbf{r}_i = \mathbf{d}_i^*\mathbf{I}_{S_2}\mathbf{D}\mathbf{D}^*\mathbf{I}_{S_2}\mathbf{d}_i = \mathbf{d}_i^*\mathbf{I}_{S_2}\mathbf{d}_i = \frac{|S_2|}{n}$$

Third result will be a direct consequence of the first one: If $\mathbf{x} \in \bar{S}_1$, then

$$(\mathbf{C}\mathbf{x})_i = \sum_{j=1}^{n}\mathbf{C}_{i,j}\mathbf{x}_j = \sum_{j\in\bar{S}_1}\mathbf{C}_{i,j}\mathbf{x}_j \tag{27}$$

When $i \in S_1, j \in \bar{S}_1$, we have $i \not\equiv j$(mod n_1) by definition, which implies $\mathbf{C}_{i,j} = \mathbf{C}_{i,j}\mathbf{x}_j = 0$ due to the first result. Fourth result can be shown by repeating these arguments for $\mathbf{D}\mathbf{I}_{S_2}\mathbf{D}^*$. ■

Using Lemma VII.1, we'll now proceed with the proof of Lemma IV.1.

A. Proof of Lemma IV.1

Proof: S_1 and S_2 components will be analyzed seperately.

Analyzing S_1: We may start by considering, $\mathbf{y}_1 + \mathbf{s}$ and write,

$$\begin{aligned}\mathbf{y}_1 + \mathbf{s} &= \mathbf{y}_1 + \mathbf{D}^*(\mathbf{b}_2 - \mathbf{c}_2) - (\mathbf{b}_1 - \mathbf{c}_1)\\ &= \mathbf{y}_1 + \mathbf{D}^*\mathbf{b}_2 - \mathbf{I}_{S_1}\mathbf{D}^*\mathbf{b}_2 - \mathbf{b}_1 + \mathbf{D}^*\mathbf{I}_{S_2}\mathbf{D}\mathbf{b}_1\end{aligned}$$

First, we'll consider, $\mathcal{S}_1(\mathbf{y}_1 + \mathbf{s})$. We have the following,

$$\mathcal{S}_1(\mathbf{y}_1) = \mathcal{S}_1(\text{sgn}(\mathbf{x})) \text{ by construction of } \mathbf{y}_1 \tag{28}$$

$$\mathcal{S}_1(\mathbf{D}^*\mathbf{b}_2 - \mathbf{I}_{S_1}\mathbf{D}^*\mathbf{b}_2) = \mathcal{S}_1((\mathbf{I} - \mathbf{I}_{S_1})\mathbf{D}^*\mathbf{b}_2) = 0 \tag{29}$$

$$\mathcal{S}_1(\mathbf{b}_1) = 0 \text{ by construction of } \mathbf{b}_1 \tag{30}$$

$$\mathcal{S}_1(\mathbf{D}^*\mathbf{I}_{S_2}\mathbf{D}\mathbf{b}_1) = 0 \text{ from Lemma VII.1.} \tag{31}$$

Hence, we find, $\mathcal{S}_1(\mathbf{y}_1 + \mathbf{s}) = \mathcal{S}_1(\mathbf{y}_1) = \mathcal{S}_1(\text{sgn}(\mathbf{x}))$.
To upper bound $\|\bar{\mathcal{S}}_1(\mathbf{y}_1 + \mathbf{s})\|_\infty$, we may simply use $\|\bar{\mathcal{S}}_1(\mathbf{y}_1 - \mathbf{b}_1)\|_\infty < 1/2$ and write,

$$\|\bar{\mathcal{S}}_1(\mathbf{y}_1+\mathbf{s})\|_\infty \leq \|\bar{\mathcal{S}}_1(\mathbf{y}_1-\mathbf{b}_1)\|_\infty + \|\bar{\mathcal{S}}_1(\mathbf{c}_1)\|_\infty + \|\bar{\mathcal{S}}_1(\mathbf{D}^*\mathbf{b}_2)\|_\infty$$

Analyzing S_2: Similarly, for $\mathcal{S}_2(\mathbf{y}_2 + \mathbf{D}\mathbf{s})$, we have the following,

$$\mathcal{S}_2(\mathbf{y}_2) = \lambda\mathcal{S}_2(\text{sgn}(\mathbf{D}\mathbf{x})) \text{ by construction} \tag{32}$$

$$\mathcal{S}_2(\mathbf{D}\mathbf{b}_1 - \mathbf{I}_{S_2}\mathbf{D}\mathbf{b}_1) = \mathcal{S}_2((\mathbf{I} - \mathbf{I}_{S_2})\mathbf{D}\mathbf{b}_1) = 0 \tag{33}$$

$$\mathcal{S}_2(\mathbf{b}_2) = 0 \text{ by construction} \tag{34}$$

$$\mathcal{S}_2(\mathbf{D}\mathbf{I}_{S_1}\mathbf{D}^*\mathbf{b}_2) = 0 \text{ from Lemma VII.1.} \tag{35}$$

Hence, $\mathcal{S}_2(\mathbf{y}_2 - \mathbf{D}\mathbf{s}) = \lambda\mathcal{S}_2(\text{sgn}(\mathbf{D}\mathbf{x}))$ as desired.
To upper bound $\|\bar{\mathcal{S}}_2(\mathbf{y}_2 - \mathbf{D}\mathbf{s})\|_\infty$, we may use $\|\bar{\mathcal{S}}_2(\mathbf{y}_2 - \mathbf{b}_2)\|_\infty < \lambda/2$ and write,

$$\|\bar{\mathcal{S}}_2(\mathbf{y}_2 - \mathbf{D}\mathbf{s})\|_\infty < \frac{\lambda}{2} + \|\bar{\mathcal{S}}_2(\mathbf{c}_2)\|_\infty + \|\bar{\mathcal{S}}_2(\mathbf{D}\mathbf{b}_1)\|_\infty$$

■

B. Proof of Lemma IV.2

Proof: We start by stating a useful lemma on Gaussian variables, [11].

Lemma VII.2. *Let g be a real standard normal random variable. Then, for any $t \geq 0$*

$$\mathbb{P}(|g| > t) \leq 2\exp(-t^2/2) \quad (36)$$

Our discussion will be for S_1 only. Proof for S_2 is identical.

Case 1: Estimating $\mathbb{P}(\bar{\mathcal{S}}_1(b_1)_i = 0)$
Observe that $\mathbf{A}_{S_1}$ and $\mathbf{A}_{\bar{S}_1}$ are independent matrices with i.i.d. Gaussian entries. Hence, for fixed $\mathbf{A}_{S_1}$, $\mathbf{A}_{\bar{S}_1}$ is i.i.d. $\bar{\mathcal{S}}_1(\mathbf{y}_1)$ is a vector with i.i.d. complex Gaussian entries with variance $\frac{\|\mathbf{s}_1\|_2^2}{m}$. Next, from (11) it can be seen that $\bar{\mathcal{S}}_1(\mathbf{b}_1)$ is an entry wise function of $\bar{\mathcal{S}}_1(\mathbf{y}_1)$ and hence i.i.d. Using Lemma VII.2 and conditioned on $\sigma_{min}(\mathbf{A}_{S_1}) \geq 1/\sqrt{2}$ for any $i \in \bar{S}_1$

$$\mathbb{P}(\mathcal{R}(b_{1,i}) = 0) = \mathbb{P}(|\mathcal{R}(y_{1,i})| < \frac{1}{4}) \geq 1-2\exp(-\frac{m}{16|S_1|}) \quad (37)$$

as variance of $\mathcal{R}(y_{1,i})$ is at most $\frac{|S_1|}{m}$. Using a union bound over real and imaginary parts of $b_{i,j}$, we find,

$$\mathbb{P}(\mathcal{R}(b_{1,i}) = 0) \geq 1 - 4\exp(-\frac{m}{16|S_1|}) \quad (38)$$

Case 2: Subgaussian norm when $\bar{\mathcal{S}}_1(\mathbf{b}_1)_j \neq 0$
Let us first define a subgaussian random variable and its norm.

Definition VII.1. *Let $\mathbf{z} \in \mathbb{R}$ be a scalar random variable. Assume for some $K < \infty$,*

$$(\mathbb{E}[|\mathbf{z}|^n])^{1/n} \leq K\sqrt{n} \text{ for all integers } n \geq 1 \quad (39)$$

Then, $\mathbf{z}$ is a subgaussian random variable and smallest K satisfying (39) is norm of $\mathbf{z}$.

Assume $i \in \bar{S}_1$. This time, we consider the case where $|y_{1,i}| > 0$. Clearly real and imaginary components of $b_{1,i}$ are independent as it is the case for $y_{1,i}$. Without loss of generality consider the real part. Observe that, if $\mathcal{R}(b_{1,i}) \neq 0$ then it is $\mathcal{R}(y_{1,i}) - \frac{1}{4}\text{sgn}(\mathcal{R}(y_{1,i}))$ where $\text{var}(\mathcal{R}(y_{1,i})) \leq \frac{|S_1|}{m} \leq \frac{1}{64}$ by assumption. Hence, using following lemma we can conclude that subgaussian norm of $b_{1,j}$ is upper bounded by $c_0\sqrt{|S_1|/m}$ as $1/4 > \sqrt{2}/8$.

Lemma VII.3. *Let $c \geq \sqrt{2}$ be a scalar, x be a standard normal random variable and,*

$$z = x - c \cdot \text{sgn}(x) \text{ conditioned on } |x| \geq c \quad (40)$$

Then, z has subgaussian norm at most c_0 for some absolute constant c_0.

Proof: Following inequality is true for tail of Gaussian p.d.f,

$$\frac{1}{\sqrt{2\pi}x}(1-\frac{1}{x^2})\exp(-x^2/2) < Q(x) < \frac{1}{\sqrt{2\pi}x}\exp(-x^2/2)$$

Hence, using $c \geq \sqrt{2}$, for $t \geq 0$ we have,

$$\begin{aligned}\mathbb{P}(|z| > t) &= \frac{Q(t+c)}{Q(c)} \leq \frac{c\exp(-(t+c)^2/2)}{(t+c)(1-c^{-2})\exp(-c^2/2)} \\ &\leq 2\exp(-t^2/2)\end{aligned}$$

Result immediately follows from Lemma 5.5 of [11] and from the bound on $\mathbb{P}(|z| > t)$. ■

Finally, $b_{1,i}$ is zero mean as $y_{1,i}$ is distributed symmetrically around 0 and construction of $b_{1,i}$ preserves the symmetry. ■

C. Proposition 5.10 and sums of sub-gaussians

Next, we state Proposition $5,10$ of [11] for completeness, which gives a bound on weighted sum of subgaussians.

Theorem VII.1 (Proposition 5.10 of [11])**.** *Let $\mathbf{z}_1, \dots, \mathbf{z}_l$ be subgaussian random variables with subgaussian norms upper bounded by $c_0 > 0$. Let $\mathbf{a} \in \mathcal{R}^l$ be an arbitrarily chosen vector. Then, for all $t \geq 0$,*

$$\mathbb{P}(|\sum_{i=1}^{l} a_i\mathbf{z}_i| \geq t) \leq 3\exp(-\frac{ct^2}{c^2\|\mathbf{a}\|_2^2}) \quad (41)$$

where $c > 0$ is an absolute constant.

Based on this, we can obtain (16) as $\mathcal{R}(\bar{\mathcal{S}}_1(\mathbf{b}_1))$ is i.i.d. subgaussian with norm at most $c_0\sqrt{|S_1|/m}$ and we need to argue both contributions from real and imaginary parts are at most $\frac{1}{4\sqrt{2}}$ with high probability. In particular for j'th row of $\mathbf{C}$,

$$\mathbb{P}(|\sum_i \mathcal{R}(c_{j,i})\mathcal{R}(b_{1,i})| > \frac{1}{8\sqrt{2}}) \leq 3\exp(-\frac{cmn}{128|S_1||S_2|}) \quad (42)$$

Writing similar bounds for $|\sum_i \mathcal{I}(c_{j,i})\mathcal{R}(b_{1,i})|$, $|\sum_i \mathcal{R}(c_{j,i})\mathcal{I}(b_{1,i})|$, $|\sum_i \mathcal{I}(c_{j,i})\mathcal{I}(b_{1,i})|$ we can conclude in (16). Similarly, to obtain, (20), we again use bounds on real and imaginary parts. This time we consider only the nonzero entries which are at most $2np$ with high probability. Then, denoting, for j'th row of $\mathbf{D}^*$ we can write,

$$\mathbb{P}(|\sum_{i\big|_{b_{2,i}\neq 0}} \mathcal{R}(r_{j,i})\mathcal{R}(b_{2,i})| > \frac{1}{8\sqrt{2}}) \leq 3\exp(-\frac{cm}{2^8p|S_2|\lambda^2c_0^2})$$

Doing this for all components and union bounding similarly yields (20).

D. Proof of Lemma VI.1

Finally, we give the proof of Lemma VI.1 which is quite similar to the proof of Lemma II.2.

Proof: Following the notation introduced for the matrix case, we need to show if such $\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}$ exist then a certain null space condition will hold for $\mathcal{A}$ which will guarantee recovery. Let us state this condition based on the sub gradients of nuclear norm and ℓ_1 norm: For all $\mathbf{W} \in \mathcal{N}(\mathcal{A})$ if the following holds then $\mathbf{A}$ is the unique optimum of JBPM.

$$f(\mathbf{W}) := \lambda[\mathcal{R}(\langle \text{sgn}(\mathbf{A}), \mathcal{S}(\mathbf{W})\rangle) + \|\bar{\mathcal{S}}(\mathbf{W})\|_1] \tag{43}$$
$$+ \mathcal{R}(\langle \mathbf{U}\mathbf{V}^*, \mathbf{W}\rangle) + \|\bar{\mathcal{L}}(\mathbf{W})\|_\star > 0 \tag{44}$$

Now, assume such $\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}$ exist and consider $\mathbf{v}_1, \mathbf{v}_2$ where:

$$\mathbf{v}_1 = \mathcal{A}^\star(\mathbf{S}_1) + \mathbf{S} \quad \text{and} \quad \mathbf{v}_2 = \mathcal{A}^\star(\mathbf{S}_2) - \mathbf{S} \tag{45}$$

Observe that for any $\mathbf{W} \in \mathcal{N}(\mathcal{A})$, we have $\langle \mathbf{v}_1 + \mathbf{v}_2, \mathbf{W}\rangle = 0$. Now, using this:

$$\begin{aligned} 0 &= \mathcal{R}(\langle \mathbf{v}_1 + \mathbf{v}_2, \mathbf{W}\rangle) \qquad (46)\\ &= \mathcal{R}(\lambda \cdot \langle \text{sgn}(\mathbf{A}), \mathcal{S}(\mathbf{W})\rangle + \left\langle \bar{\mathcal{S}}(\mathcal{A}^\star(\mathbf{S}_2) - \mathbf{S}), \mathbf{W}\right\rangle)\\ &\quad + \mathcal{R}(\langle \mathbf{U}\mathbf{V}^*, \mathbf{W}\rangle + \left\langle \bar{\mathcal{L}}(\mathcal{A}^*(\mathbf{S}_1) + \mathbf{S}), \mathbf{W}\right\rangle) \end{aligned}$$

To end the proof, using invertibility of $\mathcal{A}(\cdot)$ on $\mathcal{L} \cap \mathcal{S}$ we can conclude $\bar{\mathcal{L}}(\mathbf{W}) \neq 0$ or $\bar{\mathcal{S}}(\mathbf{W}) \neq 0$ hence:

$$\mathcal{R}(\left\langle \bar{\mathcal{L}}(\mathcal{A}^\star(\mathbf{S}_1) + \mathbf{S}), \mathbf{W}\right\rangle) < \|\bar{\mathcal{L}}(\mathbf{W})\|_\star \quad \text{or} \tag{47}$$
$$\mathcal{R}(\left\langle \bar{\mathcal{S}}(\mathcal{A}^\star(\mathbf{S}_2) - \mathbf{S}), \mathbf{W}\right\rangle) < \lambda \|\bar{\mathcal{S}}(\mathbf{W})\|_\infty \tag{48}$$

Overall, existence of $\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}$ implies the desired null space condition, i.e., $f(\mathbf{W}) > 0$ for all $\mathbf{W} \in \mathcal{N}(\mathcal{A})$. ■