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 $\ell_1$-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 $\ell_1$ minimization is done over only one basis. Next, we analyze these optimal-ity 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 success-ful recovery hence overcoming the classical requirement of $\Theta(\min\{k_1, k_2\} \log(\frac{1}{\min\{k_1, k_2\}}))$ for $\ell_1$ 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 $\ell_1$ norm is minimized subject to convex constraints. Assuming $x$ has a sparse representation over the basis $U$ (i.e. $Ux$ is a sparse vector) and assuming we get to see the observations $Ax$, Basis Pursuit performs the following optimization to get back to $x$.

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

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 $x$ is sparse over $U_1, U_2$, in order to induce sparsity in both bases, we will be considering the following approach, which we call Joint Basis Pursuit (JBP).

$$\min_{\hat{x}} \|U_1\hat{x}\|_1 + \lambda \|U_2\hat{x}\|_2 \quad \text{s.t.} \quad A\hat{x} = Ax \quad (JBP)$$

For the case of a matrix $X$ that is simultaneously sparse and low rank, we may minimize the summation of $\ell_1$ norm and the matrix nuclear norm, which is denoted by $\| \cdot \|_*$ and is equal to summation of the singular values. Assuming, we observe linear measurements $A(x)$, we propose solving the following problem (JBP-Matrix) to recover $X$.

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

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 $x$ and observe $|\langle a_i, x \rangle|^2$ 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, $xx^*$. 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
by exploiting the joint sparsity of \(x\) and \(\Sigma\) and given as follows,

\[
C_1(U_1^*(A^*s_1 + s)) = \|S_1(sgn(U_1x))\|_1 < 1
\]

\[
C_2(U_2^*(A^*s_2 - s)) = \|S_2(sgn(U_2x))\|_1 < \lambda
\]

\[
A \text{ is invertible over } \bigcap_{i=1}^2 \{v|S_i(U_1v) = U_1v\}
\]

Then \(x\) is the unique optimum of (JBP).

**Proof:** Let \(f(\hat{x})\) be the cost of \((\text{JBP})\), i.e., 
\[
 f(\hat{x}) = \sum_{i=1}^2 \lambda_i \|U_i\hat{x}\|_1
\]

Then, for any \(w \in \mathcal{N}(A)\), 
\[
f(\hat{x}) - f(x) \geq \text{lower bounded by the left hand side of (2), which follows from the sub gradient of the } \ell_1 \text{ norm. Hence } f(\hat{x}) > f(x) \text{ for all } A\hat{x} = Ax, \hat{x} \neq x.
\]

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

**Lemma II.2.** Assume \(s_1, s_2 \in \mathbb{C}^n, s \in \mathbb{C}^n\) satisfying the following conditions exist:

- \(S_1(U_1^*(A^*s_1 + s)) = S_1(sgn(U_1x))\)
- \(\|S_1(U_1^*(A^*s_1 + s))\|_{\infty} < 1\)
- \(S_2(U_2^*(A^*s_2 - s)) = \lambda S_2(sgn(U_2x))\)
- \(\|S_2(U_2^*(A^*s_2 - s))\|_{\infty} < \lambda\)
- \(A\) is invertible over \(\bigcap_{i=1}^2 \{v|S_i(U_1v) = U_1v\}\)

Then \(x\) is the unique optimum of (JBP).

**Proof:** What we need to show is that if such \(s_1, s_2, s\) exist and the invertibility assumption holds then the left hand side of (2) is strictly positive for all \(w \in \mathcal{N}(A)\). Assume such \(s_1, s_2, s\) exist and let \(v_1, v_2 \in \mathbb{C}^n\) to be:

\[
v_1 = U_1^*(A^*s_1 + s) \quad \text{and} \quad v_2 = U_2^*(A^*s_2 - s)
\]

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

\[
\sum_{i=1}^2 \langle U_i w, v_i \rangle = \sum_{i=1}^2 w^* A^* s_i + w^* s - w^* s = 0
\]

To end the proof observe that \(v_1, v_2\) 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 \(S_1(U_1w)\) or \(S_2(U_2w)\) is nonzero due to invertibility assumption.

The dual certificate approach for regular BP has been used in [1], [2], [5]. Letting \(U = U_1\), compared to Lemma II.2 it requires invertibility of \(A\) over \(\{v|S_1(U_1v) = U_1v\}\) rather than the intersection and it requires \(\|S_1(U_1^* A^* s_1)\|_{\infty} < 1\), while Lemma II.2 can overcome this by making use of the extra variable \(s\). From this perspective, JBP can be viewed as a combination of two regular BP’s that are allowed to “help” each other via \(s\).

### 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:

\[i \in S \iff j \in S \text{ for all } j \text{ such that } j \equiv i \pmod{l}\]

Observe that if \(S\) is a \(l\) periodic support, \(|S|\) is divisible by \(n/l\).
Theorem III.1. Let $U_1 = I$, $U_2 = D$. $1 > \alpha \geq 0$ 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 $n = n_1n_2$.
- $|S_2| \leq |S_1| \log^\alpha(n)$.

Then, for the following scenarios, $\mathbf{x}$ can be successfully recovered via JBP with high probability (for sufficiently large $n$) when the matrix $A \in \mathbb{C}^{m \times n}$ 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 $\ell_1$ optimization over $U_1 = I$ 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 $I$ and $D$ respectively, where $n = n_1n_2$. 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 $0 \leq t < n$,

$$v_j = \begin{cases} 0 & \text{if } j \neq l \pmod{n_1} \\ W_j^{it} & \text{else} \end{cases}$$

(6)

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

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

(7)

will have $n_1$ periodic support and $D \mathbf{x}$ will have $\frac{\mathbb{Z}}{n_1}$ periodic support. The reason we say almost all is because cancellations may occur when $\mathbf{v}_i$’s are added. However, if $\alpha_j$’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 $I, D$ pair is the fact that almost all bases $U_1$ and $U_2$ do not permit signals that are sparse in both. The following lemma illustrates this.

Lemma III.1. Assume $U_1^{-1}, U_2^{-1}$ 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, [13]. In particular, if $n$ is a prime number, we still have $|S_1| + |S_2| \geq n$ requirement for a signal over $U_1 = I$ and $U_2 = D$ 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 $s_1, s_2, s$ that certifies optimality of $\mathbf{x}$.

A. Construction of $s_1, s_2, s$

For the following discussion, we’ll be using $(U_1, U_2)$ and $(I, D)$ and $(1, \lambda)$ and $(\lambda, 1)$ interchangeably. The construction of $s_1, s_2$ will follow a classical approach previously used in [2], [5], [7]. Letting $A_{S_1} \in \mathbb{C}^{m \times |S_1|}$ denote the submatrix by choosing columns corresponding to $S_1$ and $B = AD^*$, we will use the following $s_1, s_2$.

$$s_1 = A_{S_1}(A_{S_1}^* A_{S_1})^{-1} S_1(\text{sgn}(x))$$

(8)

$$s_2 = B_{S_2}(B_{S_2}^* B_{S_2})^{-1} S_2(\text{sgn}(Dx))$$

(9)

Since $I, D$ are unitary we have $U_i^{-1} = U_i$. By construction $s_1, s_2$ already satisfies,

$$S_i(U_i A^* s_i) = \lambda_i S_i(\text{sgn}(U_i x)) \quad i \in \{1, 2\}$$

(10)

However, one has to control the term $||S_i(U_i A^* s_i)||_\infty$ and will make use of $s_0$ to achieve this. Denote $U_i A^* s_i$ by $y_i$. Define the vectors $\{b_1, b_2\}$ as follows:

$$R(b_{i,j}) = \begin{cases} 0 & \text{if } j \in S_i \\ \text{sgn}(y_{i,j}) - \lambda_i / 4 & \text{else} \end{cases}$$

(11)

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

$$s = D^*(b_2 - c_2) - I(b_1 - c_1)$$

(12)

$$c_1 = D I S_2^* D b_1, \quad c_2 = D I S_1^* D b_2$$

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

Lemma IV.1. Assume $\mathbf{x}$, $\{y_1, b_1, c_1\}_{i=1}^3$ are the same as described previously. Then, one has the following:

$$S_1(y_1 + s) = S_1(\text{sgn}(s))$$

$$S_2(y_2 - Ds) = \lambda S_2(\text{sgn}(Dx))$$

$$||S_1(y_1 + s)||_\infty < \frac{1}{2} + ||S_1(c_1)||_\infty + ||S_1(D^* b_2)||_\infty$$

$$||S_2(y_2 - Ds)||_\infty < \frac{\lambda}{2} + ||S_2(c_2)||_\infty + ||S_2(D b_1)||_\infty$$

Based on Lemma [IV.1] and Lemma [II.2] JBP recovers $\mathbf{x}$ if we have, $||S_1(c_1)||_\infty + ||S_1(D^* b_2)||_\infty \leq 1/2$ and $||S_2(c_2)||_\infty + ||S_2(D b_1)||_\infty \leq \lambda/2$.
As a next step, we can analyze $\|S_1(D^*I_{S_2}D_b1)\|_\infty$ and $\|S_1(D^*b_2)\|_\infty$ 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 $A$ is i.i.d complex normal with variance $\frac{1}{m}$ and $m \geq 64 \max\{|S_1|, |S_2|\}$. This will guarantee,

$$\sigma_{min}(A_{S_1}) \geq 1/\sqrt{2} \text{ and } \sigma_{min}(B_{S_2}) \geq 1/\sqrt{2} \quad (13)$$

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

Now, conditioned on $A_{S_1}, B_{S_2}$, satisfy \[13\].

$$\|s_i\|_2^2 = s_i^*s_i = \lambda_i^2 \text{sgn}(\bar{y}_i)(A_{S_1}A_{S_1})^{-1}\text{sgn}(\bar{y}) \leq 2\lambda_i^2 |S_1|$$

does not matter. \[13\] is an i.i.d Gaussian vector whose entries have

$$\|\bar{s}_i\|_2^2 \leq \frac{1}{m}$$

Given these, we need to understand, when can we make sure,

$$\|\bar{S}_1(D^*I_{S_2}D_b1)\|_\infty \leq \frac{1}{4} \quad \text{and} \quad \|\bar{S}_1(D^*b_2)\|_\infty \leq \frac{1}{4} \quad (14)$$

From \[11\], observe that $\bar{S}_1(b_1)$ is a function of $\bar{S}_1(y_1)$ which is i.i.d random Gaussian. The next lemma, gives a characterization of $b_1$.

**Lemma IV.2.** Suppose $m \geq 64 \max\{|S_1|, |S_2|\}$. Then, the entries $(\bar{S}_1(b_1))_{i=1}^{\bar{S}_1(i)}$ of $\bar{S}_1(b_1)$ are i.i.d. random variables with the following distribution, \[11\] is \begin{enumerate}
  \item Analysis of $\|\bar{S}_1(c_1)\|_\infty$: We need to show,

$$\|\bar{S}_1(D^*I_{S_2}D_b1)\|_\infty \leq \frac{1}{4} \quad (15)$$

Calling $C = D^*I_{S_2}D$, from Lemma \[11\], each row of $C$ has energy $\|S_2\|^2$. Let $c_i$ be the $i$th column of $C^*$. Then, using Lemma \[11\] and Proposition 5.10 of \[11\], for any $i$ and an absolute constant $c > 0$,

$$\mathbb{P}(\|c_i^*b_1\| \geq \frac{1}{4}) \leq 12\exp(-\frac{m c}{2\bar{S}_1(|c_i|^2)}) \quad (16)$$

$$= 12\exp(-\frac{m c}{2\bar{S}_1(|S_2|^2)}) \quad (17)$$

Using a union bound over all $i$’s, shows \[15\] reduces to arguing $n\mathbb{P}(\|c_i^*b_1\| \geq \frac{1}{4}) \to 0$ which is equivalent to ensuring,

$$\frac{m c}{2\bar{S}_1(|S_2|^2)} - \log n \to \infty \quad \text{as} \quad n \to \infty \quad (18)$$

Using $n \geq \min\{|S_1|, |S_2|\} \log n$ in the statement of Theorem \[11\], \[18\] holds for $m \geq 2\epsilon_1c^{-1}c_0^2 \max\{|S_1|, |S_2|\} = O(\max\{|S_1|, |S_2|\})$ as desired.

2) Analysis of $\|\bar{S}_1(D^*b_2)\|_\infty$: In a similar fashion, we would like to show,

$$\|\bar{S}_1(D^*b_2)\|_\infty \leq \frac{1}{4} \quad (19)$$

holds with high probability, to conclude. Each row of $D^*$ has unit $\ell_2$ norm and nonzero entries of $b_2$ are i.i.d subgaussians from Lemma \[11\]. Letting, $p = 4\exp(-\frac{m}{\bar{S}_1(|S_2|^2)})$ and applying a Chernoff bound w.p.a.1 $1 - \exp(-np/4)$, number of non zeros in $b_2$ is at most $2np$. Considering the inner products between each row of $D^*$ and $\bar{b}_2$, and using a union bound, \[19\] holds, with probability at least,

$$1 - 12n\exp(-\frac{m c}{2c_0^2\bar{S}_1(|S_2|^2)}) - \exp(-np/4) \quad (20)$$

Assuming $m = O(|S_2|^2\log^2(n))$ 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,

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

Following the same arguments for the other basis will yield,

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

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

**V. Empirical Results**

While Theorem \[11\] 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 $x \in \mathbb{R}^n$ be the following dirac comb,

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

It is clear that $Dx = x$ 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 $k = \{2, 4, 6, \ldots, 32\}$ and for $1 \leq m \leq 30$. Interestingly, in order to achieve $50\%$ success, JBP required $\frac{k}{4} \leq m \leq \frac{k}{2}$ and $\frac{k}{4}$ slightly increased as a function of $k$. This is shown as the straight line in Figure 1. These results are quite consistent with Theorem \[11\] 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 $\ell_1$ minimization. In particular $\frac{k}{4}$ increases from 1 to 2.4 as $k$ moves from 2 to 32. While JBP outperforms BP in this setting, the
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 $x$ is a sparse vector to be recovered from observations $\{|\langle a_i, x \rangle|^2\}_{i=1}^{m}$, where $\{a_i\}_{i=1}^{m} \in \mathbb{C}^n$ are the measurement vectors. Although, these measurements are not linear in $x$, they are linear in $xx^*$ as $|\langle a_i, x \rangle|^2 = a_i^*xx^*a_i$. Using the fact that $xx^*$ is rank 1 and sparse, JBPM can be used in order to recover $X = xx^*$ 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 $X$ is low-rank and sparse and we wish to recover it from observations $A(X)$. Let us first introduce notation relevant to structure of $X \in \mathbb{C}^{n \times n}$.

- Let $S \subseteq [n] \times [n]$ be the usual support of $X$ and $S : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{|S|}$ be the projection onto $S$.
- Assuming $X$ has singular value decomposition $U\Sigma V^*$, Define the subspace $\mathcal{L} \subseteq \mathbb{C}^{n \times n}$ as,
  $$\mathcal{L} = \{ Y \in \mathbb{C}^{n \times n} \mid (I - UU^*)Y(I - VV^*) = 0 \}$$
- $\mathcal{L}$ denotes complement of $\mathcal{L}$ and projection onto $\mathcal{L}$ is denoted by $\mathcal{P}(\cdot) : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{n \times n}$.
- $A^*(\cdot) : \mathbb{C}^m \rightarrow \mathbb{C}^{n \times n}$ 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 $X$ to be unique optimizer of JBPM.

**Lemma VI.1.** Assume $S_1, S_2 \subseteq \mathbb{C}^m, S \subseteq \mathbb{C}^{n \times n}$ satisfying the following conditions exist:

- $\mathcal{L}(A^*(S_1) + S) = UV^*$.
- $\|\mathcal{L}(A^*(S_1) + S)\| < 1$.
- $S(A^*(S_2) - S) = \lambda \cdot S(\text{sgn}(A))$.
- $\|\mathcal{L}(A^*(S_2) - S)\|_{\infty} < \lambda$.
- $A(\cdot)$ is invertible over $\{ Y | \mathcal{L}(Y) = S(Y) = Y \}$.

Then $A$ 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**

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 $C_1 \in \mathbb{R}^{n \times |S_1|}$ be the matrix obtained by taking columns of $U_1^{-1}$, over $S_1$. If $z_1 = U_1x$ and $z_2 = U_2x$ are supported over $S_1, S_2$, we may write:

$$0 = U_1^{-1}z_1 - U_2^{-1}z_2 = [C_1 \ C_2][S_1(z_1)^* - S_2(z_2)^*]^*$$

By assumption, $[C_1 \ C_2] \in \mathbb{R}^{n \times (|S_1|+|S_2|)}$ 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 $(z_1, z_2)$ 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 $x$ having combined sparsities of $U_1x$ and $U_2x$ 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 $D \in \mathbb{R}^{n \times n}$ be the DFT matrix as previously. Further, let $C = D^*I_{S_2}D$. Then,

1. $C_{i,j} = 0$ for any $(i, j)$ with $i \neq j$ (mod $n_1$).
2. For any $i, i'$th row of $C$ satisfies $\|c_i\|_2 = |S_2|/n$.
3. For any $x$ that is supported on $S_1$, we have, $S_1(Cx) = 0$.
4. First three results similarly hold for $C = DI_{S_1}D^*$.

Proof: Let us start by analyzing the matrix $D^*I_{S_2}D$. Let $d_i$ be the $i$th column of $D$. Then,

$$C_{i,j} = d_i^*I_{S_2}d_j = \sum_{k \in S_2} d_i^*d_{j,k}$$

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,

$$C_{i,j} = \sum_{t \in T} \sum_{c=1}^{n_1} c_{i,t+cn_2}^*c_{j,t+cn_2}$$

Next, for any $i \neq j$ (mod $n_1$) and any $t \leq n_1$,

$$\sum_{c=0}^{n_1-1} c_{i,t+cn_2}^*c_{j,t+cn_2} = \sum_{c=0}^{n_1-1} W^{(j-i)(t+cn_2)}$$

$$= W^{(j-i)}\sum_{c=0}^{n_1-1} W^{c(j-i)n_2}$$

$$= n_1 W^{(j-i)t}\delta(i - j \text{ (mod } n_1))$$

where $\delta(k) = 1 \iff k \neq 0$. This proves the first statement. To show the second, $r_i = d_i^*I_{S_2}D$ implies:

$$\|r_i\|^2 = r_i^*r_i = d_i^*I_{S_2}DD^*I_{S_2}d_i = d_i^*I_{S_2}d_i = |S_2|/n$$

Third result will be a direct consequence of the first one: If $x \in S_1$, then

$$(Cx)_i = \sum_{j=1}^{n} C_{i,j}x_j = \sum_{j \in S_1} C_{i,j}x_j$$

When $i \in S_1, j \in S_1$, we have $i \neq j$ (mod $n_1$) by definition, which implies $C_{i,j} = C_{i,j}x_j = 0$ due to the first result. Fourth result can be shown by repeating these arguments for $DI_{S_1}D^*$.

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 separately.

Analyzing $S_1$: We may start by considering, $y_1 + s$ and write,

$$y_1 + s = y_1 + D^*(b_2 - c_2) - (b_1 - c_1)$$

$$= y_1 + D^*b_2 - IS_1D^*b_2 - b_1 + D^*IS_2Db_1$$

First, we’ll consider, $S_1(y_1 + s)$. We have the following,

$$S_1(y_1) = S_1(\text{sgn}(x)) \text{ by construction of } y_1$$

$$S_1(D^*b_2 - IS_1D^*b_2) = S_1((I - IS_1)D^*b_2) = 0$$

$$S_1(b_1) = 0 \text{ by construction of } b_1$$

$$S_1(D^*IS_2Db_1) = 0 \text{ from Lemma VII.1}$$

Hence, we find, $S_1(y_1 + s) = S_1(y_1) = S_1(\text{sgn}(x))$.

To upper bound $\|S_1(y_1 + s)\|_\infty$, we may simply use $\|S_1(y_1 - b_1)\|_\infty < 1/2$ and write,

$$\|S_1(y_1 + s)\|_\infty \leq \|S_1(y_1 - b_1)\|_\infty + \|S_1(c_1)\|_\infty + \|S_1(D^*b_2)\|_\infty$$

Analyzing $S_2$: Similarly, for $S_2(y_2 + D_2s)$, we have the following,

$$S_2(y_2) = \lambda S_2(\text{sgn}(Dx)) \text{ by construction}$$

$$S_2(\text{Db}_1 - IS_2Db_1) = S_2((I - IS_2)Db_1) = 0$$

$$S_2(b_2) = 0 \text{ by construction}$$

$$S_2(\text{DI}_{S_1}D^*b_2) = 0 \text{ from Lemma VII.1}$$

Hence, $S_2(y_2 + D_2s) = \lambda S_2(\text{sgn}(Dx))$ as desired.

To upper bound $\|S_2(y_2 + D_2s)\|_\infty$, we may use $\|S_2(y_2 - b_2)\|_\infty < \lambda/2$ and write,

$$\|S_2(y_2 - D_2s)\|_\infty < \frac{\lambda}{2} + \|S_2(c_2)\|_\infty + \|S_2(Db_1)\|_\infty$$

■
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 \(t \geq 0\)
\[
\mathbb{P}(|g| > t) \leq 2 \exp(-t^2/2) \tag{36}
\]

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

Case 1: Estimating \(\mathbb{P}(S_1(b_1)_i = 0)\)
Observe that \(\mathbf{A}_{S_1}\) and \(\mathbf{A}_{S_2}\) are independent matrices with i.i.d. Gaussian entries. Hence, for fixed \(\mathbf{A}_{S_1}\), \(\mathbf{A}_{S_2}\) is i.i.d. \(\mathcal{N}(\mathbf{0}, \sigma^2_{\mathbf{I}})\) for all integers \(\mathbf{A}_{S_1}\) and conditioned on \(\sigma_{\text{min}}(\mathbf{A}_{S_1}) \geq 1/\sqrt{2}\) for any \(i \in S_1\)
\[
\mathbb{P}(\mathcal{R}(b_1)_i = 0) = \mathbb{P}(|\mathcal{R}(y_{1,i})| < \frac{1}{4}) \geq 1 - 2 \exp(-\frac{m}{16|S_1|}) \tag{37}
\]
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 \(b_1, j\), we find,
\[
\mathbb{P}(\mathcal{R}(b_1)_i = 0) \geq 1 - 4 \exp(-\frac{m}{16|S_1|}) \tag{38}
\]

Case 2: Subgaussian norm when \(S_1(b_1)_j \neq 0\)
Let us first define a subgaussian random variable and its norm.

Definition VII.1. Let \(z \in \mathbb{R}\) be a scalar random variable. Assume for some \(K < \infty\),
\[
(E[|z|^n])^{1/n} \leq K \sqrt{n} \text{ for all integers } n \geq 1 \tag{39}
\]
Then, \(z\) is a subgaussian random variable and smallest \(K\) satisfying (39) is norm of \(z\).

Assume \(i \in S_1\). 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 \(\mathcal{R}(y_{1,i}) - \frac{1}{\sqrt{2}} \text{sgn}(\mathcal{R}(y_{1,i}))\) 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 \(1/4 \geq \sqrt{2}/8\).

Lemma VII.3. Let \(c \geq \sqrt{2}\) be a scalar, \(x\) be a standard normal random variable and,
\[
z = x - c \cdot \text{sgn}(x) \text{ conditioned on } |x| \geq c \tag{40}
\]
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,
\[
\frac{1}{\sqrt{2\pi x}}(1 - \frac{1}{x^2}) \exp(-x^2/2) \leq Q(x) < \frac{1}{\sqrt{2\pi x}} \exp(-x^2/2)
\]
Hence, using \(c \geq \sqrt{2}\), for \(t \geq 0\) we have,
\[
\mathbb{P}(|z| > t) = \frac{Q(t + c)}{Q(c)} \leq \frac{c \exp(-(t + c)^2/2)}{(t + c)(1 - c^2)^{1/2}} \leq 2 \exp(-t^2/2)
\]
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 subgaussians

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 \(z_1, \ldots, z_t\) be subgaussian random variables with subgaussian norms upper bounded by \(c_0 > 0\). Let \(a \in \mathbb{R}^t\) be an arbitrarily chosen vector. Then, for all \(t \geq 0\),
\[
\mathbb{P}(|\sum_{i=1}^t a_i z_i| \geq t) \leq 3 \exp(-\frac{ct^2}{c_0^2||a||_2^2}) \tag{41}
\]
where \(c > 0\) is an absolute constant.

Based on this, we can obtain \([16]\) as \(\mathbb{R}(\mathbf{S}_1(b_1))\) 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 \(\mathbf{C}\),
\[
\mathbb{P}(|\sum_{i} \mathcal{R}(c_{j,i}) \mathcal{R}(b_{1,i})| > \frac{1}{8\sqrt{2}}) \leq 3 \exp(-\frac{cmn}{288|S_1||S_2|}) \tag{42}
\]
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,
\[
\mathbb{P}(|\sum_{i \neq 0} \mathcal{R}(r_{j,i}) \mathcal{R}(b_{2,i})| > \frac{1}{8\sqrt{2}}) \leq 3 \exp(-\frac{cm}{2^p|S_2|\lambda^2c_0^2})
\]
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 $S_1, S_2, S$ exist then a certain null space condition will hold for $A$ which will guarantee recovery. Let us state this condition based on the sub gradients of nuclear norm and $\ell_1$ norm: For all $W \in \mathcal{N}(A)$ if the following holds then $A$ is the unique optimum of JBPM.

$$f(W) := \lambda [R(\langle \text{sgn}(A), S(W) \rangle) + \|\bar{S}(W)\|_1] + R(\langle UV^*, W \rangle) + \|\bar{L}(W)\|_* > 0$$

(43)

Now, assume such $S_1, S_2, S$ exist and consider $v_1, v_2$

where:

$$v_1 = A^*(S_1) + S \quad \text{and} \quad v_2 = A^*(S_2) - S \quad (45)$$

Observe that for any $W \in \mathcal{N}(A)$, we have $\langle v_1 + v_2, W \rangle = 0$. Now, using this:

$$0 = R(\langle v_1 + v_2, W \rangle) \quad (46)$$

$$= R(\lambda \cdot \langle \text{sgn}(A), S(W) \rangle + \langle \bar{S}(A^*(S_2) - S), W \rangle)$$

$$+ R(\langle UV^*, W \rangle + \langle \bar{L}(A^*(S_1) + S), W \rangle)$$

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

$$R(\langle \bar{L}(A^*(S_1) + S), W \rangle) < \|\bar{L}(W)\|_* \quad \text{or} \quad (47)$$

$$R(\langle \bar{S}(A^*(S_2) - S), W \rangle) < \lambda \|\bar{S}(W)\|_\infty \quad (48)$$

Overall, existence of $S_1, S_2, S$ implies the desired null space condition, i.e., $f(W) > 0$ for all $W \in \mathcal{N}(A)$. ■