DETERMINISTIC PHASE GUARANTEES FOR ROBUST RECOVERY IN INCOHERENT DICTIONARIES

Cheuk Ting Li

The Chinese University of Hong Kong
Department of Information Engineering
Shatin, NT, Hong Kong

Samet Oymak, Babak Hassibi

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

ABSTRACT

This paper presents a relaxation of an assumption usually imposed in the recovery of sparse vectors with random support in pairs of orthonormal bases or incoherent dictionaries by basis pursuit. The assumption requires the phases of the entries of the sparse vector to be chosen randomly in $[0, 2\pi)$. This paper provides probabilistic recovery guarantees for deterministic phases. We prove that, if a phase pattern is fixed, then a sparse vector with random support and corresponding phases can be recovered with high probability. As a result, the phases can take any distribution and can be dependent, as long as they are independent of the support. Furthermore, this improvement does not come at the expense of the maximum recoverable sparsity.

***Index Terms*—** uncertainty principle, basis pursuit, sparsity, duality in optimization, incoherent dictionary

1. INTRODUCTION

In recent years, sparse representation problems have received extensive attention, where we aim to find the sparse signal $\mathbf{x}$ that underlies the, possibly underdetermined, observations $\mathbf{y} = \mathbf{\Phi x}$. One popular way of recovering $\mathbf{x}$ is the following ℓ_1-minimization program.

$$\min \|\tilde{\mathbf{x}}\|_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{\Phi}\tilde{\mathbf{x}} \qquad \text{(P1)}$$

The program (P1) is called *basis pursuit* ([1]) and it is well known that certain conditions can provably guarantee that the output of (P1) is equal to $\mathbf{x}$ ([2]).

This work will be concerned with the problem where $\mathbf{\Phi} \in \mathbb{R}^{d\times N}$ is an *overcomplete dictionary*, i.e. a deterministic matrix whose columns have unit ℓ_2 norm, $N > d$ and the inner products between any two distinct columns are bounded by the coherence μ.

Background: The program (P1) was first proposed in a seminal work by Chen et al. ([1]) and was analyzed in [3]. Sharp deterministic guarantees are provided in [3], in particular, when $\mathbf{\Phi}$ is the concatenation of identity and DFT matrices of size d, (P1) can recover any underlying signal $\mathbf{x}$ with sparsity at most $\frac{1}{2}\sqrt{d}$ from underdetermined observations.

A more recent work by Candès and Romberg [4] showed that, by introducing randomness in the support and phases of the nonzero components of the signal, (P1) can tolerate significantly more nonzero entries for the same dictionary, as much as $O(\frac{d}{log(d)})$. Basically this means, (P1) works for most "not-so-sparse" signals and we'll call this the "robust sparsity threshold (RST)". The paper [5] yields results comparable to [4]; however its results apply to general dictionaries rather than only pairs of orthonormal matrices.

Contributions: Works related to RST ([6, 7, 8]) generally assumes uniformly i.i.d. random phase for nonzero entries and uniformly random support for the underlying signal. In this paper, it is shown that the assumption on the phases can be relaxed with no additional cost at all. Our results require, orderwise, the same amount of maximum tolerable sparsity as in [5] which provides the best available results.

In our model, we first fix an arbitrary phase pattern $\boldsymbol{\beta} \in \mathbb{C}^N$, $|\beta_i| = 1$ for $i = 1, ..., N$ and choose the support T of the signal uniformly at random from $\{1, ..., N\}$ and then use the elements of $\boldsymbol{\beta}$ that corresponds to T as phases of the nonzero entries. In Theorem 6.1, we prove that, for a general dictionary, if $\mathbf{x}$ is sufficiently sparse and obeys this model, it can be recovered via (P1) with high probability (w.h.p.). In Theorem 6.2, we study the case of a pair of orthonormal basis. It is shown that, if the support on the first basis is fixed arbitrarily and the phases are uniformly i.i.d., while the support on the second basis is random and the phases are chosen in the same way (from $\boldsymbol{\beta}$), (P1) will recover the sparse signal $\mathbf{x}$ w.h.p.

Our results are based on a simple but novel dual certificating method that guarantees success of (P1). We compensate for the loss of random phases by introducing an additional random support during the dual vector construction.

2. BACKGROUND AND NOTATION

It is always assumed $\mathbf{\Phi} \in \mathbb{C}^{d\times N}$. The spectral norm is denoted by $\|\cdot\|_2$. Let $T \subseteq \{1, 2, \ldots, N\}$. $\mathbf{\Phi}_T$ denotes the

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. This work was performed while the first author was an exchange student at the California Institute of Technology, supported by Professor Charles K. Kao Research Exchange Scholarship of The Chinese University of Hong Kong.

matrix obtained from $\mathbf{\Phi}$ by selecting only the columns in T. Similarly, the vector $\mathbf{x}_T$ is obtained by selecting the entries of $\mathbf{x} \in \mathbb{C}^N$ in T. The selection of columns or entries is applied before transposition or inversion (i.e. $\mathbf{\Phi}_T^* = (\mathbf{\Phi}_T)^*$). Denote the support (set of non-zero positions) of a vector $\mathbf{x}$ by $\mathrm{supp}(\mathbf{x})$. For any scalar $x \in \mathbb{C}$, let $\mathrm{sgn}(x) = x/|x|$ when $x \neq 0$ and $\mathrm{sgn}(x) = 0$ when $x = 0$.

Definition 2.1. *A* dictionary *is a matrix $\mathbf{\Phi} \in \mathbb{C}^{d\times N}$ in which columns have unit ℓ_2 norm. The* coherence *of a dictionary $\mathbf{\Phi} = [\boldsymbol{\varphi}_1, ..., \boldsymbol{\varphi}_N]$ is defined as*

$$\mu = \max_{i\neq j} \left|\langle \boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j \rangle\right|. \tag{1}$$

In particular, we may construct a dictionary $\mathbf{\Phi} = [\mathbf{\Phi}_1\ \mathbf{\Phi}_2]$ by a pair of orthonormal bases $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$. Obviously, (P1) is interesting when $N > d$, i.e. when $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$ is underdetermined.

3. PREVIOUS WORK

This section presents a comparison between results of [5] and this work. The following theorem is a result of Theorem B and Theorem 14 of [5].

Theorem 3.1 (Tropp)**.** *Let $\mathbf{\Phi} \in \mathbb{C}^{d\times N}$ be a dictionary with coherence μ and $s \geq 1$. If*

$$\sqrt{\mu^2 m \cdot s \log N} + \frac{m}{N} \|\mathbf{\Phi}\|_2^2 \leq C,$$

where $C > 0$ is a constant, then with probability $1 - N^{-s}$, a vector $\mathbf{x} \in \mathbb{C}^N$ with $\mathrm{supp}(\mathbf{x})$ of size m randomly chosen in $\{1, ..., N\}$ and $\mathrm{sgn}(x_i)$ i.i.d. uniform on the unit circle, is the unique solution of (P1) *with $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$.*

Compared to this result, Theorem 6.1 relaxes the random phase assumption, while keeping the condition on sparsity the same (except for a smaller C).

The following theorem is a result of Theorem D and Theorem 14 of [5].

Theorem 3.2 (Tropp)**.** *Let $\mathbf{\Phi} \in \mathbb{C}^{d\times 2d}$ be a pair of orthonormal bases with coherence μ and $s \geq 1$. Assume*

$$\mu^2(m_1 + m_2)s \log d \leq C,$$

where $C > 0$ is a constant. Then with probability $1 - d^{-s}$, a vector $\mathbf{x} = [\mathbf{v}_1^T\ \mathbf{v}_2^T]^T$, where $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{C}^d$, with $\mathrm{supp}(\mathbf{v}_1)$ of size m_1 arbitrary in $\{1, ..., d\}$ and $\mathrm{supp}(\mathbf{v}_2)$ of size m_2 randomly chosen in $\{1, ..., d\}$, and $\mathrm{sgn}(x_i)$ i.i.d. uniform on the unit circle, is the unique solution of (P1) *with $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$.*

Compared to this result, Theorem 6.2 relaxes the random phase assumption on the second basis, while keeping the condition on sparsity the same (except for a smaller C).

Next, we state Theorem B of [5], which will be useful in the proof of Theorem 6.1.

Theorem 3.3 (Tropp)**.** *Let $\mathbf{\Phi} \in \mathbb{C}^{d\times N}$ be a dictionary with coherence μ, and let T be a random subset of $\{1, ..., N\}$ with size m. Let $s \geq 1$. If*

$$\sqrt{\mu^2 m \log(m+1) \cdot s} + \frac{m}{N} \|\mathbf{\Phi}\|_2^2 \leq c\delta,$$

where $c > 0$ is a constant, then we have

$$\mathbb{P}\left\{\|\mathbf{\Phi}_T^* \mathbf{\Phi}_T - I\|_2 \geq \delta\right\} \leq m^{-s}.$$

4. CONCENTRATION INEQUALITIES

The following inequality can be obtained by applying the classical Hoeffding inequality [9] on the real and imaginary parts of $\{X_i\}_{i=1}^N$ and then using a union bound.

Lemma 4.1 (Complex Hoeffding Inequality)**.** *Let $\mathbf{a} \in \mathbb{R}^N$. Let $X_1, ..., X_N$ be independent zero-mean complex-valued random variables with $|X_i| \leq a_i$ almost surely. Then for $\delta \geq 0$,*

$$\mathbb{P}\left\{\left|\sum_{i=1}^N X_i\right| \geq \delta\right\} \leq 4 \cdot \exp\left(-\frac{\delta^2}{4\|\mathbf{a}\|_2^2}\right).$$

Now, we introduce a useful inequality on random subsets.

Lemma 4.2. *Fix $N = 3m$ and $\mathbf{x} \in \mathbb{C}^N$. Let T be a random subset of $\{1, ..., N\}$ with size m. Let $\mathbf{c} \in \mathbb{R}^N$ with*

$$c_i = \begin{cases} 1 & \text{if } i \in T \\ -1/2 & \text{if } i \notin T \end{cases},$$

then we have

$$\mathbb{P}\left\{|\langle \mathbf{x}, \mathbf{c}\rangle| \geq \delta\right\} \leq 16 \cdot \exp\left(-\frac{\delta^2}{16\|\mathbf{x}\|_2^2}\right).$$

Proof. To avoid ambiguity between the modulus sign and conditional probability, we write the magnitude of a scalar t as $\|t\|$. Here we use a trick similar to Lemma 18 of [5]. Let $\boldsymbol{\sigma}$ be a random permutation of $\{1, ..., N\}$. Assume the first m entries of $\boldsymbol{\sigma}$ correspond to the elements of T, then we have

$$\begin{aligned} p &\overset{def}{=} \mathbb{P}\left\{\left\|\sum_{i=1}^N x_i c_i\right\| \geq \delta\right\} \\ &= \mathbb{P}\left\{\left\|\sum_{i=1}^m \left(x_{\sigma_i} - x_{\sigma_{i+m}}/2 - x_{\sigma_{i+2m}}/2\right)\right\| \geq \delta\right\}. \end{aligned}$$

Draw a random vector $\mathbf{v} \in \{0, 1, 2, 3\}^N$ uniformly and independent of $\boldsymbol{\sigma}$. Let $w_{i,j} = 1$ if $v_i = j$ and 0 otherwise. Further, let

$$f_{\boldsymbol{\sigma}}(\mathbf{w}) = \sum_{i=1}^m \left(\left(w_{i,0} + w_{i,1}\right) x_{\sigma_i} - w_{i,2} x_{\sigma_{i+m}} - w_{i,3} x_{\sigma_{i+2m}}\right). \tag{2}$$

Using the fact $\mathbb{E}[w_{i,j}] = 1/4$, we have

$$p = \mathbb{P}\{\|\mathbb{E}[f_{\boldsymbol{\sigma}}(\mathbf{w})|\boldsymbol{\sigma}]\| \geq \delta/2\}.$$

We now fix $\boldsymbol{\sigma}$ and consider the event $\|\mathbb{E}[f_{\boldsymbol{\sigma}}(\mathbf{w})]\| \geq \delta/2$. We write $r(\mathbf{w}, n) = \mathbf{w}'$ where $w'_{i,j} = w_{i,(j+n) \bmod 4}$. Note that

$$\begin{aligned}\sum_{i=0}^{3} \|f_{\boldsymbol{\sigma}}(r(\mathbf{w}, i))\| &\geq \left\| \sum_{i=0}^{3} f_{\boldsymbol{\sigma}}(r(\mathbf{w}, i)) \right\| \\ &= \left\| \sum_{i=1}^{m} (w_{i,0} + ... + w_{i,3})(2x_{\sigma_i} - x_{\sigma_{i+m}} - x_{\sigma_{i+2m}}) \right\| \\ &= 4 \left\| \mathbb{E}\left[f_{\boldsymbol{\sigma}}(\mathbf{w})\right] \right\| \geq 2\delta.\end{aligned}$$

Therefore, at least one of $\mathbf{w}$, $r(\mathbf{w}, 1)$, $r(\mathbf{w}, 2)$ and $r(\mathbf{w}, 3)$ will have $\|f_{\boldsymbol{\sigma}}\left(r(\mathbf{w}, n)\right)\| \geq \delta/2$. As a result, whenever $\boldsymbol{\sigma}$ satisfies $\|\mathbb{E}\left[f_{\boldsymbol{\sigma}}(\mathbf{w})|\boldsymbol{\sigma}\right]\| \geq \delta/2$,

$$\mathbb{P}\left\{ \|f_{\boldsymbol{\sigma}}(\mathbf{w})\| \geq \delta/2 \middle| \boldsymbol{\sigma} \right\} \geq 1/4.$$

Taking average on $\boldsymbol{\sigma}$ with $\|\mathbb{E}\left[f_{\boldsymbol{\sigma}}(\mathbf{w})|\,\boldsymbol{\sigma}\right]\| \geq \delta/2$,

$$\mathbb{P}\left\{ \|f_{\boldsymbol{\sigma}}(\mathbf{w})\| \geq \delta/2 \middle| \left\|\mathbb{E}\left[f_{\boldsymbol{\sigma}}(\mathbf{w})|\,\boldsymbol{\sigma}\right]\right\| \geq \delta/2 \right\} \geq 1/4.$$

Consequently, we have

$$p = \mathbb{P}\left\{\left\|\mathbb{E}\left[f_{\boldsymbol{\sigma}}(\mathbf{w})|\,\boldsymbol{\sigma}\right]\right\| \geq \delta/2\right\} \leq 4 \cdot \mathbb{P}\left\{\|f_{\boldsymbol{\sigma}}(\mathbf{w})\| \geq \delta/2\right\}.$$

Now, let $g(n) = 1$ when $n = 0$ or 1, $g(n) = -1$ otherwise. Due to (2), we simply have

$$f_{\boldsymbol{\sigma}}(\mathbf{w}) = \sum_{i=1}^{m} g(v_i) x_{\sigma_{i+\max(0, v_i - 1)\cdot m}}.$$

Let $\sigma'_i = \sigma_{i+\max(0,v_i-1)\cdot m}$ for $i = 1...m$. Note that if v_i is fixed, then $\boldsymbol{\sigma}'$ is a random permutation of a random m-element subset of $\{1, ..., N\}$ regardless of the values of v_i, and thus $\boldsymbol{\sigma}'$ is independent of v_i. For any fixed $\boldsymbol{\sigma}'$, as $g(v_i)$ are i.i.d. with $\mathbb{P}\left\{g(v_i) = \pm 1\right\} = 1/2$, Lemma 4.1 gives

$$\begin{aligned}\mathbb{P}\left\{ \left\| \sum_{i=1}^{m} g(v_i) x_{\sigma'_i} \right\| \geq \delta/2 \right\} &\leq 4 \cdot \exp\left(-\frac{(\delta/2)^2}{4 \sum_{i=1}^{m} \left| x_{\sigma'_i} \right|^2} \right) \\ &\leq 4 \cdot \exp\left(-\frac{\delta^2}{16 \left\|\mathbf{x}\right\|_2^2} \right).\end{aligned}$$

Consequently, we obtain the desired result as

$$p \leq 16 \cdot \exp\left(-\frac{\delta^2}{16 \left\|\mathbf{x}\right\|_2^2} \right).$$

□

5. ℓ_1-DUALITY

We state a condition for recovery by ℓ_1 minimization presented in [10], which generalizes [11]. See also [2].

Lemma 5.1 (ℓ_1-Duality)**.** *Let $\mathbf{\Phi} \in \mathbb{C}^{d \times N}$, $\mathbf{x} \in \mathbb{C}^N$ and $T = \mathrm{supp}(\mathbf{x})$ and assume $\mathbf{\Phi}_T$ has rank $|T|$. Then, if there exists a vector $\mathbf{h} \in \mathbb{C}^d$ which satisfies*

$$\begin{aligned} (\mathbf{\Phi}^*\mathbf{h})_i &= \mathrm{sgn}(x_i) && \text{for all } i \in T \\ |(\mathbf{\Phi}^*\mathbf{h})_i| &< 1 && \text{for all } i \notin T \end{aligned}$$

then $\mathbf{x}$ is the unique solution of the program (P1) with $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$.

We call the vector $\mathbf{h}$ in Lemma 5.1 a *dual vector* of $\mathbf{x}$. Next, we present a method to construct the dual vector.

Lemma 5.2. *Let $\mathbf{\Phi} \in \mathbb{C}^{d \times N}$, $\mathbf{x} \in \mathbb{C}^N$ and $T = \mathrm{supp}(\mathbf{x})$. Let Γ be a superset of T and assume $\mathbf{\Phi}_\Gamma$ has rank $|\Gamma|$. If there exists a vector $\mathbf{v} \in \mathbb{C}^N$ satisfying*

$$\begin{aligned} v_i &= \mathrm{sgn}(x_i) && \text{for all } i \in T, \\ |v_i| &< 1 && \text{for all } i \in \Gamma \backslash T, \text{ and} \\ \left| \left\langle \mathbf{\Phi}_\Gamma^\dagger \boldsymbol{\varphi}_i, \mathbf{v}_\Gamma \right\rangle \right| &< 1 && \text{for all } i \notin \Gamma \end{aligned}$$

then $\mathbf{\Phi}_\Gamma^{\dagger*}\mathbf{v}_\Gamma$ is a dual vector of $\mathbf{x}$. Here, $\boldsymbol{\varphi}_i$ denotes the i'th column of $\mathbf{\Phi}$ and $\mathbf{\Phi}_\Gamma^\dagger = \left(\mathbf{\Phi}_\Gamma^*\mathbf{\Phi}_\Gamma\right)^{-1} \mathbf{\Phi}_\Gamma^*$ is the Moore–Penrose pseudoinverse.

Proof. Note that $(\mathbf{\Phi}^*\mathbf{\Phi}_\Gamma^{\dagger*}\mathbf{v}_\Gamma)_i = v_i$ for any $i \in \Gamma$. It remains to check $(\mathbf{\Phi}^*\mathbf{\Phi}_\Gamma^{\dagger*}\mathbf{v}_\Gamma)_i$ for $i \notin \Gamma$, which is bounded by the third assumption. □

6. MAIN RESULTS

This section is dedicated to two theorems, which are the main contributions of this paper.

Theorem 6.1. *Let $N \geq 32$ and $\mathbf{\Phi} \in \mathbb{C}^{d \times N}$ be a dictionary with coherence μ. Fix $\boldsymbol{\beta} \in \mathbb{C}^N$ with $|\beta_i| = 1$. Let T be a random subset of $\{1, ..., N\}$ with size $m \geq 2$. Let $s \geq 1$. If*

$$\sqrt{\mu^2 m \cdot s \log N} + \frac{m}{N} \left\|\mathbf{\Phi}\right\|_2^2 \leq C,$$

where C is a constant, then a vector $\mathbf{x} \in \mathbb{C}^N$ with $\mathrm{supp}(\mathbf{x}) = T$ and $\mathrm{sgn}(x_i) = \beta_i$ is the unique solution of (P1) *with $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$ with probability $1 - N^{-s}$.*

Before going into the proof, it should be noted that, if a particular $\mathbf{x}$ with some support and phase pattern can be recovered via the program (P1) , then all vectors with same support and phase pattern can be recovered as well.

Proof of Theorem 6.1. We take $C = \min(c/6, 1/24)$, where c is the constant in Theorem 3.3. Let $\mathbf{\Phi} = [\boldsymbol{\varphi}_1, ..., \boldsymbol{\varphi}_N]$. Note that we can multiply $\boldsymbol{\beta}$ into each column of $\mathbf{\Phi}$ to obtain a new dictionary $\mathbf{\Phi}_2 = [\beta_1\boldsymbol{\varphi}_1, ..., \beta_N\boldsymbol{\varphi}_N]$ with the same norm and incoherence, and then we can take the signs of $\mathbf{x}$ to be 1. Therefore, without loss of generality, assume $\beta_i = 1$.

After picking T, we draw a random $2m$-element subset T_2 from $\{1, ..., N\} \backslash T$. The union $\Gamma = T \cup T_2$ is a random $3m$ subset of $\{1, ..., N\}$.

Let $s' = s\log(N)/\log(m)$, then

$$\begin{aligned}\sqrt{\mu^2 \cdot 3ms'\log(3m+1)} + \frac{3m}{N}\|\mathbf{\Phi}\|_2^2 \\ \leq 3\left(\sqrt{\mu^2 m \cdot s\log N} + \frac{m}{N}\|\mathbf{\Phi}\|_2^2\right) \\ \leq 3C \leq c/2\end{aligned}$$

as $3\log(m) \geq \log(3m+1)$ when $m \geq 2$. Invoking Theorem 3.3 on Γ and using the fact that $s' \geq 1$, we have

$$\mathbb{P}\left\{\|\mathbf{\Phi}_\Gamma^*\mathbf{\Phi}_\Gamma - I\|_2 \geq \frac{1}{2}\right\} \leq (3m)^{-s'} \leq \frac{1}{2}N^{-s}. \quad (3)$$

From now on, we fix Γ and assume $\|\mathbf{\Phi}_\Gamma^*\mathbf{\Phi}_\Gamma - I\|_2 < 1/2$. Let $\mathbf{v} \in \mathbb{C}^N$ be as follows,

$$v_i = \begin{cases} 1 & \text{if } i \in T \\ -\frac{1}{2} & \text{if } i \in T_2 \\ 0 & otherwise \end{cases}.$$

We want to show that $(\mathbf{\Phi}_\Gamma)^{\dagger *}\mathbf{v}_\Gamma$ is a dual vector for $\mathbf{x}$ with $\mathrm{supp}(\mathbf{x}) = T$ and $\mathrm{sgn}(x_i) = 1$ for $i \in T$. Note that $\|\mathbf{\Phi}_\Gamma^*\mathbf{\Phi}_\Gamma - I\|_2 < 1/2$ implies rank$(\mathbf{\Phi}_\Gamma) = 3m$. By Lemma 5.2, it remains to show the following, with high probability

$$\left|\left\langle \mathbf{\Phi}_\Gamma^\dagger \boldsymbol{\varphi}_i, \mathbf{v}_\Gamma\right\rangle\right| < 1 \text{ for } i \notin \Gamma.$$

Consider $\mathbf{\Phi}_\Gamma^\dagger \boldsymbol{\varphi}_i$. We have

$$\|\mathbf{\Phi}_\Gamma^\dagger \boldsymbol{\varphi}_i\|_2 \leq \|\left(\mathbf{\Phi}_\Gamma^*\mathbf{\Phi}_\Gamma\right)^{-1}\|_2 \|\mathbf{\Phi}_\Gamma^* \boldsymbol{\varphi}_i\|_2 .$$

Since $\|\mathbf{\Phi}_\Gamma^*\mathbf{\Phi}_\Gamma - I\|_2 < 1/2$, $\lambda_{min}(\mathbf{\Phi}_\Gamma^*\mathbf{\Phi}_\Gamma) > 1/2$, and thus $\left\|(\mathbf{\Phi}_\Gamma^*\mathbf{\Phi}_\Gamma)^{-1}\right\|_2 \leq 2$. By coherence of $\mathbf{\Phi}$, $\|\mathbf{\Phi}_\Gamma^*\boldsymbol{\varphi}_i\|_2 \leq \mu\sqrt{3m}$. Therefore, we have $\|\mathbf{\Phi}_\Gamma^\dagger\boldsymbol{\varphi}_i\|_2 \leq 2\mu\sqrt{3m}$.

Note that, when Γ is fixed, T is a random subset of Γ. By Lemma 4.2, for each $i \notin \Gamma$,

$$\mathbb{P}\left\{\left|\left\langle \mathbf{\Phi}_\Gamma^\dagger \boldsymbol{\varphi}_i, \mathbf{v}_\Gamma\right\rangle\right| \geq 1\right\} \leq 16 \cdot \exp\left(-\frac{1}{192\mu^2 m}\right).$$

Now, from the initial assumption,

$$\sqrt{\mu^2 m \cdot s\log N} \leq C \leq 1/24 \implies \mu^2 m \leq (24^2 \cdot s\log N)^{-1}.$$

Hence we obtain,

$$\begin{aligned}\mathbb{P}\left\{\left|\left\langle \mathbf{\Phi}_\Gamma^\dagger \boldsymbol{\varphi}_i, \mathbf{v}_\Gamma\right\rangle\right| \geq 1\right\} &\leq 16 \cdot \exp\left(-\frac{24^2 \cdot s\log N}{192}\right) \\ &\leq 16N^{-3s} \leq \frac{1}{2}N^{-3s+1} \leq \frac{1}{2}N^{-s-1}.\end{aligned}$$

Taking union bound,

$$\mathbb{P}\left\{\left|\left\langle \mathbf{\Phi}_\Gamma^\dagger \boldsymbol{\varphi}_i, \mathbf{v}_\Gamma\right\rangle\right| < 1 \text{ for } i \notin \Gamma\right\} \geq 1 - \frac{1}{2}N^{-s}.$$

By Lemma 5.2, for any given Γ with $\|\mathbf{\Phi}_\Gamma^*\mathbf{\Phi}_\Gamma - I\|_2 < 1/2$, with probability $1 - (1/2)N^{-s}$, every vector supported on T with signs 1 can be recovered using ℓ_1 minimization. The result follows from (3). □

Theorem 6.2. *Let $d \geq 80$. Let $\mathbf{\Phi} \in \mathbb{C}^{d\times 2d}$ be a pair of orthonormal bases with coherence μ. Let $\boldsymbol{\zeta} \in \mathbb{C}^d$ where ζ_i are i.i.d. uniform on the unit circle. Fix $\boldsymbol{\eta} \in \mathbb{C}^d$ with $|\eta_i| = 1$. Fix $T_1 \subseteq \{1, ..., d\}$ with $|T_1| = m_1 \geq 2$. Let T_2 be a random subset of $\{d+1, ..., 2d\}$ with size $m_2 \geq 2$. Let $T = T_1 \cup T_2$ and $\boldsymbol{\beta} = [\boldsymbol{\zeta}\ \boldsymbol{\eta}]$. Let $s \geq 1$. If*

$$\mu^2(m_1 + m_2)s\log d \leq C,$$

where $C > 0$ is a constant, then a vector $\mathbf{x} \in \mathbb{C}^N$ with $\mathrm{supp}(x) = T$ *and* $\mathrm{sgn}(x_i) = \beta_i$ *for all $i \in T$ is the unique solution of* (P1) *with $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$, with probability $1 - d^{-s}$.*

Proof. We take $C = \min(c/6, 1/2304)$, where c is the constant in Theorem D of [5]. The proof is similar to that of Theorem 6.1. □

7. REFERENCES

[1] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, pp. 33–61, 1999.

[2] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[3] David L. Donoho and Xiaoming Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

[4] E. J. Candès and J. Romberg, "Quantitative robust uncertainty principles and optimally sparse decomposition," *Found. Comput. Math.*, vol. 6, no. 2, pp. 227–254, 2006.

[5] J. A. Tropp, "On the conditioning of random subdictionaries," *Appl. Comput. Harmon. Anal.*, vol. 25, pp. 1–24, 2008.

[6] E. J. Cands and Y. Plan, "Near-ideal model selection by ℓ_1 minimization," *Ann. Statist.*, vol. 37, no. 5A, pp. 2145–2177, 2009.

[7] P. Kuppinger, G. Durisi, and H. Bölcskei, "Where is Randomness Needed to Break the Square-Root Bottleneck?" *in Proceedings of ISIT 2010*, pp. 1578–1582.

[8] P. Kuppinger, G. Durisi, and H. Bölcskei, "Uncertainty Relations and Sparse Signal Recovery for Pairs of General Signal Sets," submitted to *IEEE Trans. Inf. Theory*.

[9] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Am. Stat. Assoc.*, vol. 58, no. 301, pp. 13–30, 1963.

[10] J. A. Tropp, "Recovery of short, complex linear combinations via ℓ_1 minimization," *IEEE Trans. Inform. Theory*, vol. 51, pp. 1568–2242, 2005.

[11] J. J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Inform. Theory*, vol. 50, pp. 1341–1344, 2004.