Probabilistic Recovery Guarantees for Sparsely Corrupted Signals

Graeme Pope, *Student Member, IEEE*, Annina Bracher, and Christoph Studer, *Member, IEEE*

Abstract

We consider the recovery of sparse signals subject to sparse interference, as introduced in Studer *et al.*, IEEE Trans. IT, 2012. We present novel probabilistic recovery guarantees for this framework, covering varying degrees of knowledge of the signal and interference support, which are relevant for a large number of practical applications. Our results assume that the sparsifying dictionaries are solely characterized by coherence parameters and we require randomness only in the signal and/or interference. The obtained recovery guarantees show that one can recover sparsely corrupted signals with overwhelming probability, even if the sparsity of both the signal and interference scale (near) linearly with the number of measurements.

Index Terms

Sparse signal recovery, probabilistic recovery guarantees, coherence, basis pursuit, signal restoration, signal separation, compressed sensing.

Part of this paper has been submitted to the IEEE Information Theory Workshop (ITW), Lausanne, Switzerland, 2012.

G. Pope and A. Bracher are with the Dept. of Information Technology and Electrical Engineering, ETH Zurich, 8092 Zurich, Switzerland (e-mail: gpope@nari.ee.ethz.ch, brachera@student.ethz.ch).

C. Studer is with the Dept. of Electrical and Computer Engineering, Rice University, Houston, TX 77005, USA (e-mail: studer@rice.edu).

The work of C. Studer was supported by the Swiss National Science Foundation (SNSF) under Grant PA00P2-134155.

The authors would like to thank C. Aubel, R. G. Baraniuk, H. Bölcskei, I. Koch, P. Kuppinger, and A. Pope for inspiring discussions.

I. INTRODUCTION

We consider the problem of recovering the sparse signal vector $\mathbf{x} \in \mathbb{C}^{n_a}$ with support set $\mathcal{X}$ (containing the locations of the non-zero entries of $\mathbf{x}$) from m linear measurements [1]

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}. \tag{1}$$

Here, $\mathbf{A} \in \mathbb{C}^{m \times n_a}$ and $\mathbf{B} \in \mathbb{C}^{m \times n_b}$ are given and known dictionaries, i.e., matrices that are possibly over-complete and whose columns have unit Euclidean norm. The vector $\mathbf{e} \in \mathbb{C}^{n_b}$ with support set $\mathcal{E}$ represents the sparse interference. We investigate the following models for the sparse signal vector $\mathbf{x}$ and sparse interference vector $\mathbf{e}$, and their support sets $\mathcal{X}$ and $\mathcal{E}$:

- The interference support set $\mathcal{E}$ is *arbitrary*, i.e., $\mathcal{E} \subseteq \{1, \ldots, n_b\}$ can be any subset of cardinality n_e. In particular, $\mathcal{E}$ may depend upon the sparse signal vector $\mathbf{x}$ and/or the dictionary $\mathbf{A}$, and hence, may also be chosen adversarially. The support set $\mathcal{X}$ of $\mathbf{x}$ is chosen at *random*, i.e., $\mathcal{X}$ is chosen uniformly at random from all subsets of $\{1, \ldots, n_a\}$ with cardinality n_x.
- The support set $\mathcal{E}$ of the sparse interference vector $\mathbf{e}$ is chosen at random, i.e., $\mathcal{E}$ is chosen uniformly at random from all subsets of $\{1, \ldots, n_b\}$ with cardinality n_e. The support set $\mathcal{X}$ is assumed to be arbitrary and of size n_x.
- Both $\mathcal{X}$ and $\mathcal{E}$, the support sets of the signal and of the interference with size n_x and n_e, respectively, are chosen uniformly at random.

In addition, for each model on the support sets $\mathcal{X}$ and $\mathcal{E}$ we may or may not know either of the support sets prior to recovery.

As discussed in [1], recovery of the sparse signal vector $\mathbf{x}$ from the sparsely corrupted observation $\mathbf{z}$ in (1) is relevant in a large number of practical applications. In particular, restoration of saturated signals [2]–[4], signals impaired by impulse noise [5]–[7], or removal of narrowband interference is captured by the input-output relation (1). Furthermore, the setting (1) enables us to investigate sparsity-based super-resolution and in-painting [8], [9], as well as signal separation [10], [11]. Hence, identifying the fundamental limits on the recovery of the vector $\mathbf{x}$ from the sparsely corrupted observation $\mathbf{z}$ is of significant practical interest.

Recovery guarantees for sparsely corrupted signals have been partially studied in [1], [2], [12]–[17]. In particular, [1], [12] investigated coherence-based recovery guarantees for arbitrary support sets $\mathcal{X}$ and $\mathcal{E}$ and for varying levels of support-set knowledge; [13] analyzed the special case where both support sets are unknown, but one is chosen arbitrarily and the other at random. The recovery guarantees in [14] require that the measurement matrix $\mathbf{A}$ is chosen at random and those in [2], [15]–[17] characterize $\mathbf{A}$ by the restricted isometry property (RIP), which is, in general, difficult to verify in practice. The recovery guarantees [2], [14], [15] require $\mathbf{B}$ to be unitary, whereas [16], [17] only consider a *single* dictionary $\mathbf{A}$ and partial support-set knowledge within $\mathbf{A}$. The specific models and assumptions underlying the results in [2], [14]–[17] reduce their utility for the applications outlined above.

A. Generality of the signal and interference model

In this paper, we will exclusively focus on probabilistic results where the randomness is in the signal and/or the interference but *not* in the dictionary. Furthermore, the dictionaries $\mathbf{A}$ and $\mathbf{B}$ will be characterized only by their coherence parameters and their dimensions. Such results enable us to operate with a given (and arbitrary) pair of sparsifying dictionaries $\mathbf{A}$ and $\mathbf{B}$, rather than hoping that the signal will be sparse in a randomly generated dictionary or that $\mathbf{A}$ satisfies the RIP. The following two application examples illustrate the generality of our results.

1) Restoration of saturated signals: In this example, a signal $\mathbf{y} = \mathbf{A}\mathbf{x}$ is subject to saturation [1]. This impairment is captured by setting $\mathbf{z} = g_a(\mathbf{y})$ in (1), where $g_a(\cdot)$ implements element-wise saturation to $[-a, a]$ with a being the saturation level. By writing $\mathbf{z} = \mathbf{y} + \mathbf{e}$ with $\mathbf{e} = g_a(\mathbf{y}) - \mathbf{y}$, where $\mathbf{e}$ is non-zero only for the entries where the saturation in $\mathbf{z}$ occurs, we see that for moderate saturation levels a, the vector $\mathbf{e}$ will be sparse. The reconstruction of the (uncorrupted) signal $\mathbf{y}$ from the saturated measurement $\mathbf{z}$, amounts to recovering $\mathbf{x}$ from $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{e}$, followed by computing $\mathbf{y} = \mathbf{A}\mathbf{x}$.

We assume that the signal $\mathbf{y} = \mathbf{A}\mathbf{x}$ is drawn from a stochastic model where $\mathcal{X}$ has a support set chosen uniformly at random. Since the saturation artifacts modeled by $\mathbf{e}$ are dependent on $\mathbf{y}$, we want to guarantee recovery for arbitrary $\mathcal{E}$. Furthermore, we can identify the locations where the saturation occurs (e.g., by comparing the entries of $\mathbf{z}$ to the saturation level a) and hence, we can assume that $\mathcal{E}$ is known prior to recovery. The recovery guarantees developed in this paper include this particular combination of support-set knowledge and randomness as a special case, whereas the recovery guarantees in [1], [13], [18] are unable to consider all aspects of this model and turn out to be more restrictive.

2) Removal of impulse noise: Consider a signal $\mathbf{y} = \mathbf{A}\mathbf{x}$ that is subject to impulse noise. Specifically, we observe $\mathbf{z} = \mathbf{y} + \mathbf{e}$, where $\mathbf{e}$ is the impulse noise vector. For a sufficiently low impulse-noise rate, $\mathbf{e}$ will be sparse in the identity basis, i.e., $\mathbf{B} = \mathbf{I}$. As before, consider the setting where $\mathbf{y} = \mathbf{A}\mathbf{x}$ is generated from a stochastic model with unknown support set $\mathcal{X}$. Since impulse noise does not, in general, depend on the signal $\mathbf{y}$, we may chose $\mathcal{E}$ at random. In addition, the locations $\mathcal{E}$ of the impulse noise are normally unknown.

Recovery guarantees for this setting are partially covered by [1], [13], [18]. However, as for the saturation example above, the recovery guarantees in [1], [13], [18] are unable to exploit all aspects of support-set knowledge and randomness. The results developed here cover this particular setting as a special case and hence, lead to less restrictive recovery guarantees.

B. Contributions

In this paper, we present probabilistic recovery guarantees that improve upon the ones in [1], [13], [18] and cover novel cases for varying degrees of knowledge of the signal and interference support sets. Our results depend on the coherence parameters of the two dictionaries $\mathbf{A}$ and $\mathbf{B}$ and their dimensions. In particular, we present novel recovery guarantees for the situations where the support sets $\mathcal{X}$ and/or $\mathcal{E}$ are chosen at random, and for the cases where knowledge of neither, one, or both support sets $\mathcal{X}$ and $\mathcal{E}$ is available prior to recovery. For the case where one support set is random and the other arbitrary, but no knowledge of $\mathcal{X}$ and $\mathcal{E}$ is available, we present an improved (i.e., less restrictive) recovery guarantee than the existing one in [13, Thm. 6]. Finally, we show that ℓ_1-norm minimization is able to recover the vectors $\mathbf{x}$ and $\mathbf{e}$ with overwhelming probability, even if the number of non-zero components in both scales (near) linearly with the number of measurements.

A summary of all the cases studied in this paper is given in Table I; the theorems highlighted in dark gray indicate novel recovery guarantees, light gray indicates improved ones. We will only prove the boldface theorems; the corresponding symmetric cases are shown in italics and the associated recovery guarantees can be obtained by interchanging the roles of $\mathbf{x}$ and $\mathbf{e}$.

C. Notation

Lowercase and uppercase boldface letters stand for column vectors and matrices, respectively. For the matrix $\mathbf{M}$, we denote its transpose, adjoint, and (Moore–Penrose) pseudo-inverse by

TABLE I

SUMMARY OF ALL RECOVERY GUARANTEES FOR SPARSELY CORRUPTED SIGNALS

	$\mathcal{X}$, $\mathcal{E}$ arbitrary	$\mathcal{X}$ random, $\mathcal{E}$ arbitrary	$\mathcal{X}$ arbitrary, $\mathcal{E}$ random	$\mathcal{X}$, $\mathcal{E}$ random
$\mathcal{X}$, $\mathcal{E}$ known	Case 1a [1, Thm. 3]	Case 1b **Theorem 1**	*Case 1b* Theorem 1	Case 1c **Theorem 1**
$\mathcal{E}$ known	Case 2a [1, Thm. 4]	Case 2b **Theorem 2**	*Case 2c* Theorem 4	Case 2d **Theorem 3**
$\mathcal{X}$ known	*Case 2a* [1, Cor. 6]	Case 2c **Theorem 4**	*Case 2b* Theorem 2	*Case 2d* Theorem 3
neither known	Case 3a [13, Thms. 2 and 3]	Case 3b **Theorem 5 and [13, Thm. 6]**	*Case 3b* Theorem 5 and [13, Thm. 6]	Case 3c **Theorem 6**

$\mathbf{M}^T$, $\mathbf{M}^H$, and $\mathbf{M}^\dagger$, respectively. The jth column and the entry in the ith row and jth column of the matrix $\mathbf{M}$ is designated by $\mathbf{m}_j$ and $[\mathbf{M}]_{i,j}$, respectively. The minimum and maximum singular value of $\mathbf{M}$ are given by $\sigma_{\min}(\mathbf{M})$ and $\sigma_{\max}(\mathbf{M})$, respectively; the spectral norm is $\|\mathbf{M}\|_{2,2} = \sigma_{\max}(\mathbf{M})$. The ℓ_1-norm of the vector $\mathbf{v}$ is denoted by $\|\mathbf{v}\|_1$ and $\|\mathbf{v}\|_0$ stands for the number of nonzero entries in $\mathbf{v}$. Sets are designated by upper-case calligraphic letters; the cardinality of the set $\mathcal{S}$ is $|\mathcal{S}|$. The support set of $\mathbf{v}$, i.e., the indices of the nonzero entries, is given by $\mathrm{supp}(\mathbf{v})$. The matrix $\mathbf{M}_\mathcal{S}$ is obtained from $\mathbf{M}$ by retaining the columns of $\mathbf{M}$ with indices in $\mathcal{S}$; the vector $\mathbf{v}_\mathcal{S}$ is obtained analogously from the vector $\mathbf{v}$. The $\mathrm{sign}(\cdot)$ function applied to a vector returns a vector consisting of the phases of each entry. The $N \times N$ restriction matrix $\mathbf{R}_\mathcal{S}$ for the set $\mathcal{S} \subseteq \{1, \ldots, N\}$ has $[\mathbf{R}_\mathcal{S}]_{k,k} = 1$ if $k \in \mathcal{S}$ and is zero otherwise. For random variables X and Y, we define $\mathbb{E}^q[X] = \mathbb{E}[|X|^q]^{1/q}$ to be the qth moment and $\mathbb{E}^q_X[f(X,Y)]$ to be the qth moment with respect to X. We define $\mathbb{1}[\mu \neq 0]$ to be equal to 1 if the condition $\mu \neq 0$ holds and 0 otherwise.

Throughout the paper, $\mathcal{X} = \mathrm{supp}(\mathbf{x})$ is assumed to be of cardinality n_x and $\mathcal{E} = \mathrm{supp}(\mathbf{e})$ of cardinality n_e. We define $\mathbf{D} = [\,\mathbf{A}\ \mathbf{B}\,]$ and $\mathbf{D}_{\mathcal{X},\mathcal{E}} = [\,\mathbf{A}_\mathcal{X}\ \mathbf{B}_\mathcal{E}\,]$ to be the sub-dictionary of $\mathbf{D}$ associated with the non-zero entries of $\mathbf{x}$ and $\mathbf{e}$. Similarly, we define the vector $\mathbf{s}_{\mathcal{X},\mathcal{E}} = [\,\mathbf{x}_\mathcal{X}^T\ \mathbf{e}_\mathcal{E}^T\,]^T$ which consists of the non-zero components of $\mathbf{s} = [\,\mathbf{x}^T\ \mathbf{e}^T\,]^T$.

D. Outline of the paper

The remainder of the paper is organized as follows. Relevant prior work is summarized in Section II. The main theorems are presented in Section III and a corresponding discussion is given in Section IV. We conclude in Section V. All proofs are relegated to the Appendices.

II. Relevant Prior Work

We next summarize relevant prior work on sparse signal recovery and sparsely corrupted signals, and we put our results into perspective.

A. Coherence-based recovery guarantees

During the last decade, numerous deterministic and probabilistic guarantees for the recovery of sparse signals from linear (and non-adaptive) measurements have been developed [18]–[26]. These results give sufficient conditions for when one can reconstruct the sparse signal vector $\mathbf{x}$ from the (interference-less) observation $\mathbf{y} = \mathbf{A}\mathbf{x}$ by solving

$$(\mathrm{P0}) \quad \underset{\hat{\mathbf{x}}}{\text{minimize}} \, \|\hat{\mathbf{x}}\|_0 \quad \text{subject to } \mathbf{y} = \mathbf{A}\hat{\mathbf{x}},$$

or its convex relaxation, known as basis pursuit, defined as

$$(\mathrm{BP}) \quad \underset{\hat{\mathbf{x}}}{\text{minimize}} \, \|\hat{\mathbf{x}}\|_1 \quad \text{subject to } \mathbf{y} = \mathbf{A}\hat{\mathbf{x}}.$$

In particular, in [19]–[21] it is shown that if $\|\mathbf{x}\|_0 \leqslant n_x$ for some $n_x < (1 + 1/\mu_a)/2$ with the coherence parameter

$$\mu_a = \max_{i,j,i \neq j} |\langle \mathbf{a}_i, \mathbf{a}_j \rangle| \,, \tag{2}$$

then (P0) and (BP) are able to perfectly recover the sparse signal vector $\mathbf{x}$. Such coherence-based recovery guarantees are, however, subject to the infamous "square-root bottleneck", which only guarantees the recovery of $\mathbf{x}$ for sparsity levels on the order of $n_x \sim \sqrt{m}$ [18]. This behavior is an immediate consequence of the Welch bound [27] and dictates that the number of measurements must grow at least quadratically in the sparsity level of $\mathbf{x}$ to guarantee recovery. In order to overcome this square-root bottleneck, one must either resort to a RIP-based analysis, e.g., [22]–[25], which typically requires randomness in the dictionary $\mathbf{A}$, or a *probabilistic* analysis that only considers randomness in the vector $\mathbf{x}$, whereas $\mathbf{A}$ is constant and solely characterized by its coherence parameter [18]. In this paper, we are interested in the latter type of results. Such probabilistic and coherence-based recovery guarantees that overcome the square-root bottleneck have been derived for (P0) and (BP) in [18]. The corresponding results, however, do not exploit the structure of the problem (1), i.e., the fact that we are dealing with two dictionaries and that knowledge of $\mathcal{X}$ and/or $\mathcal{E}$ may be available prior to recovery.

B. Recovery guarantees for sparsely corrupted signals

Guarantees for the recovery of sparsely corrupted signals as modeled by (1) have been developed recently in [1], [12], [13]. The reference [1] considers deterministic (and coherence-based) results for several cases[1] which arise in different applications: 1) $\mathcal{X} = \mathrm{supp}(\mathbf{x})$ and $\mathcal{E} = \mathrm{supp}(\mathbf{e})$ are known prior to recovery, 2) only one of $\mathcal{X}$ and $\mathcal{E}$ is known, and 3) neither $\mathcal{X}$ nor $\mathcal{E}$ are known. For case 1), the non-zero entries of both the signal and interference vectors can be recovered by [1]

$$\mathbf{s}_{\mathcal{X},\mathcal{E}} = \mathbf{D}^{\dagger}_{\mathcal{X},\mathcal{E}}\mathbf{z}, \tag{3}$$

if the recovery guarantee in [1, Thm. 2] is satisfied. For case 2), recovery is performed by using modified versions of (P0) and (BP); the associated recovery guarantees can be found in [1, Thm. 4 and Cor. 6]. For case 3), recovery guarantees for the standard (P0) or (BP) algorithms are given in [13, Thms. 2 and 3]. However, all these recovery guarantees suffer from the square-root bottleneck, as they guarantee recovery for *all* signal and *all* interference vectors satisfying the given sparsity constraints. A notable exception for case 3) was discussed in [13, Thm. 6]. There, $\mathbf{e}$ is assumed to be random, but $\mathbf{x}$ is assumed to be arbitrary. This model overcomes the square-root bottleneck and is able to significantly improve upon the corresponding deterministic recovery guarantees in [13, Thms. 2 and 3].

Another strain of recovery guarantees for sparsely corrupted signals that are able to overcome the square-root bottleneck have been developed in [2], [14]–[17]. The work of [14] considers the case where $\mathbf{A}$ is random, whereas [2], [15]–[17] consider matrices $\mathbf{A}$ that are characterized by the RIP, which is, in general, difficult to verify for a given (deterministic) $\mathbf{A}$. Moreover, the recovery guarantees in [2], [14], [15] require that $\mathbf{B}$ is an orthogonal matrix and, hence, these

[1]Note that no efficient recovery algorithm with corresponding guarantees is known for the case studied in [1], where only the cardinality of $\mathcal{X}$ or $\mathcal{E}$ is known. Thus, we do not consider this case in the remainder of the paper.

Model 1 $\mathcal{M}$(P0)

- Let $\mathbf{A} \in \mathbb{C}^{m \times n_a}$ and $\mathbf{B} \in \mathbb{C}^{m \times n_b}$ be dictionaries with coherence μ_a, μ_b, and mutual coherence μ_m.
- Let $\mathbf{x} \in \mathbb{C}^{n_a}$ and $\mathbf{e} \in \mathbb{C}^{n_b}$ have support set $\mathcal{X}$ and $\mathcal{E}$, respectively, of which at least one is chosen at random. If a support set is chosen at random, then assume that the corresponding non-zero entries of the associated vector are drawn from a continuous distribution.
- The observation $\mathbf{z}$ is given by $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$.

Model 2 $\mathcal{M}$(BP)

- The conditions of $\mathcal{M}$(P0) hold.
- If $\mathcal{X}$ or $\mathcal{E}$ is chosen at random, then assume that the corresponding non-zero entries of the associated vector(s) are drawn from a continuous distribution, where the phases of the individual components are independent and uniformly distributed on $[0, 2\pi)$.

results do not allow for *arbitrary* pairs of dictionaries $\mathbf{A}$ and $\mathbf{B}$. The results in [16], [17] only consider a *single* dictionary with partial support-set knowledge and, thus, are unable to exploit the fact that the signal and interference exhibit sparse representations in two *different* dictionaries. In addition, [14], [15] do not study the impact of support-set knowledge on the recovery guarantees. While all these assumptions are valid for applications based on compressive sensing (see, e.g., [28], [29]), they are not suitable for the application scenarios outlined in Section I.

To overcome the square-root bottleneck for *arbitrary* pairs of dictionaries $\mathbf{A}$ and $\mathbf{B}$, we next investigate a generalization of the probabilistic models developed in [13], [18] for the cases 1), 2), and 3) outlined above. In particular, we impose a random model on the signal and/or interference vectors rather than on the dictionaries, and we allow for varying degrees of knowledge of the support sets $\mathcal{X}$ and $\mathcal{E}$. An overview of the coherence-based recovery guarantees developed next is given in Table I.

III. Main Results

The recovery guarantees developed next rely upon the models $\mathcal{M}$(P0) and $\mathcal{M}$(BP) summarized in Model 1 and Model 2, respectively. Both models use the coherence parameters of the dictionaries $\mathbf{A}$ and $\mathbf{B}$, i.e., the coherence μ_a of $\mathbf{A}$ in (2), the coherence μ_b of $\mathbf{B}$ given by

$$\mu_b = \max_{i,j,i\neq j} \left|\langle \mathbf{b}_i, \mathbf{b}_j \rangle\right|,$$

and the mutual coherence μ_m between $\mathbf{A}$ and $\mathbf{B}$, defined as

$$\mu_m = \max_{i,j} \left|\langle \mathbf{a}_i, \mathbf{b}_j \rangle\right|.$$

Our main results for the cases highlighted in Table I are detailed next.

A. *Cases 1b and 1c: $\mathcal{X}$ and $\mathcal{E}$ known*

We start with the case where both support sets $\mathcal{X}$ and $\mathcal{E}$ are known prior to recovery. The following theorem guarantees recovery of $\mathbf{x}$ and $\mathbf{e}$ from $\mathbf{z}$, with high probability, using (3).

Theorem 1 (Cases 1b and 1c): Let $\mathbf{x}$ and $\mathbf{e}$ be signals satisfying the conditions of $\mathcal{M}$(P0), assume that both $\mathcal{X}$ and $\mathcal{E}$ are known, and choose $\beta \geqslant \log(n_x)$. If $\mathcal{X}$ is chosen uniformly at random, $\mathcal{E}$ is arbitrary, and if

$$\begin{aligned}\delta e^{1/4} \geqslant \|\mathbf{A}\|_{2,2} \|\mathbf{B}\|_{2,2} \sqrt{\frac{n_x}{n_a}} + 12\mu_a\sqrt{\beta n_x} + (n_e-1)\mu_b \\ + \mathbb{1}[\mu_a \neq 0]\frac{2n_x}{n_a}\|\mathbf{A}\|_{2,2}^2 + 3\mu_m\sqrt{2\beta n_e},\end{aligned} \tag{4}$$

holds with $\delta = 1$, then we can recover $\mathbf{x}$ and $\mathbf{e}$ using (3) with probability at least $1 - e^{-\beta}$.

If both $\mathcal{X}$ and $\mathcal{E}$ are chosen at random and if

$$\begin{aligned}\delta e^{1/4} \geqslant 12\sqrt{\beta}\left(\mu_a\sqrt{n_x} + \mu_b\sqrt{n_e}\right) + \mathbb{1}[\mu_a \neq 0]\frac{2n_x}{n_a}\|\mathbf{A}\|_{2,2}^2 + \mathbb{1}[\mu_b \neq 0]\frac{2n_e}{n_b}\|\mathbf{B}\|_{2,2}^2 \\ + \min\left\{3\mu_m\sqrt{2\beta n_x} + \sqrt{\frac{n_e}{n_b}}\left\|\mathbf{A}^H\mathbf{B}\right\|_{2,2}, 3\mu_m\sqrt{2\beta n_e} + \sqrt{\frac{n_x}{n_a}}\left\|\mathbf{A}^H\mathbf{B}\right\|_{2,2}\right\}\end{aligned} \tag{5}$$

holds with $\delta = 1$ and $\beta \geqslant \max\{\log(n_x), \log(n_e)\}$, then we can recover $\mathbf{x}$ and $\mathbf{e}$ using (3) with probability at least $1 - e^{-\beta}$.

Proof: See Appendix B. ∎

A discussion of the recovery conditions (4) and (5) is relegated to Section IV.

B. *Cases 2b and 2d: $\mathcal{E}$ known*

Consider the case where only the support set $\mathcal{E}$ of $\mathbf{e}$ is known prior to recovery. In this case, recovery of $\mathbf{x}$ (and the non-zero entries of $\mathbf{e}$) from $\mathbf{z}$ can be achieved by solving [1]

$$(\text{P0}, \mathcal{E}) \quad \begin{cases} \underset{\hat{\mathbf{x}}, \hat{\mathbf{e}}_{\mathcal{E}}}{\text{minimize}} & \|\hat{\mathbf{x}}\|_0 \\ \text{subject to} & \mathbf{z} = \mathbf{A}\hat{\mathbf{x}} + \mathbf{B}_{\mathcal{E}}\hat{\mathbf{e}}_{\mathcal{E}} \end{cases} \tag{6}$$

or its convex relaxation

$$(\text{BP}, \mathcal{E}) \quad \begin{cases} \underset{\hat{\mathbf{x}}, \hat{\mathbf{e}}_{\mathcal{E}}}{\text{minimize}} & \|\hat{\mathbf{x}}\|_1 \\ \text{subject to} & \mathbf{z} = \mathbf{A}\hat{\mathbf{x}} + \mathbf{B}_{\mathcal{E}}\hat{\mathbf{e}}_{\mathcal{E}}. \end{cases} \tag{7}$$

The following theorems guarantee the recovery of $\mathbf{x}$ and $\mathbf{e}$ from $\mathbf{z}$, using $(\text{P0}, \mathcal{E})$ or $(\text{BP}, \mathcal{E})$, with high probability.

Theorem 2 (Case 2b): Let $\mathbf{x}$ and $\mathbf{e}$ be signals satisfying the conditions of $\mathcal{M}(\text{P0})$, assume that $\mathcal{E}$ is known prior to recovery and chosen arbitrarily, and assume that $\mathcal{X}$ is unknown and drawn uniformly at random. Choose $\beta \geqslant \log(n_x)$. If (4) holds for some δ where $0 < \delta < 1$ and if

$$n_x \mu_a^2 + n_e \mu_m^2 < 1 - \delta, \tag{8}$$

then we can recover $\mathbf{x}$ and $\mathbf{e}$ using $(\text{P0}, \mathcal{E})$ with probability at least $1 - e^{-\beta}$.

Moreover, if $\mathbf{x}$ and $\mathbf{e}$ are signals satisfying the conditions of $\mathcal{M}(\text{BP})$, and, in addition to (4) and (8), if either $\mathbf{A}$ is unitary or

$$n_x \mu_a^2 + n_e \mu_m^2 < \frac{(1-\delta)^2}{2(\log(n_a) + \beta)} \tag{9}$$

holds, then we can recover $\mathbf{x}$ and $\mathbf{e}$ using $(\text{BP}, \mathcal{E})$ with probability at least $1 - 3e^{-\beta}$.

Proof: See Appendices C and D. ■

Note that since we allow for $0 < \delta < 1$, we arrive at less restrictive recovery conditions than those in [13], [18], where a fixed value of δ is chosen, i.e., $\delta = 1 - \lambda^2$ with $\lambda = 1/\sqrt{2}$. Moreover, by combining (4), (8), and possibly (9) into a *single* recovery condition, by effectively removing δ, we can easily calculate the largest values of n_x and n_e for which successful recovery with high probability is guaranteed (see Section IV-C for a corresponding discussion).

Theorem 3 (Case 2d): Let $\mathbf{x}$ and $\mathbf{e}$ be signals satisfying the conditions of $\mathcal{M}(\text{P0})$, assume that $\mathcal{E}$ is known but $\mathcal{X}$ is unknown prior to recovery, and assume that both $\mathcal{X}$ and $\mathcal{E}$ are drawn uniformly at random. If (5) and (8) hold for some $0 < \delta < 1$ and $\beta \geqslant \max\{\log(n_x), \log(n_e)\}$, then we can recover $\mathbf{x}$ and $\mathbf{e}$ using $(\text{P0}, \mathcal{E})$ with probability at least $1 - e^{-\beta}$.

Moreover, if $\mathbf{x}$ and $\mathbf{e}$ are signals satisfying the conditions of $\mathcal{M}(\text{BP})$ and if (9) holds in addition to (5) and (8), then we can recover $\mathbf{x}$ and $\mathbf{e}$ using $(\text{BP}, \mathcal{E})$ with probability at least $1 - 3e^{-\beta}$.

Proof: See Appendices C and D. ■

A discussion of both theorems is relegated to Section IV.

C. Case 2c: $\mathcal{X}$ known

The case where $\mathcal{X}$ is random and known, and $\mathcal{E}$ is unknown and arbitrary, differs slightly to the case where $\mathcal{X}$ is random and unknown, and $\mathcal{E}$ is arbitrary and known (covered by Theorem 2). Hence, we need to consider both cases separately. The recovery problems $(\text{P0}, \mathcal{X})$ and $(\text{BP}, \mathcal{X})$ required here are defined analogously to $(\text{P0}, \mathcal{E})$ and $(\text{BP}, \mathcal{E})$.

Theorem 4 (Case 2c): Let $\mathbf{x}$ and $\mathbf{e}$ be signals satisfying the conditions of $\mathcal{M}(\text{P0})$, assume that the support set $\mathcal{X}$ is known and chosen uniformly at random, and assume that $\mathcal{E}$ is unknown and arbitrary. If

$$\begin{aligned} \delta e^{1/4} \geqslant \|\mathbf{A}\|_{2,2} \|\mathbf{B}\|_{2,2} \sqrt{\frac{n_e}{n_b}} + 12\mu_b\sqrt{\beta n_e} + (n_x - 1)\mu_a \\ + \mathbb{1}[\mu_b \neq 0]\frac{2n_e}{n_b} \|\mathbf{B}\|_{2,2}^2 + 3\mu_m\sqrt{2\beta n_x} \end{aligned} \tag{10}$$

holds for some $0 < \delta < 1$ and $\beta \geqslant \log(n_e)$, and if

$$n_x\mu_m^2 + n_e\mu_b^2 < 1 - \delta, \tag{11}$$

then we can recover $\mathbf{x}$ and $\mathbf{e}$ using $(\text{P0}, \mathcal{X})$ with probability at least $1 - e^{-\beta}$.

Moreover, if $\mathbf{x}$ and $\mathbf{e}$ are signals satisfying the conditions of $\mathcal{M}(\text{BP})$, and, in addition to (10) and (11), if either $\mathbf{B}$ is unitary or if

$$n_x\mu_m^2 + n_e\mu_b^2 < \frac{(1-\delta)^2}{2(\log(n_b) + \beta)} \tag{12}$$

holds, then we can recover $\mathbf{x}$ and $\mathbf{e}$ using $(\text{BP}, \mathcal{X})$ with probability at least $1 - 3e^{-\beta}$.

Proof: See Appendices C and D. ■

A discussion of this theorem is relegated to Section IV.

D. Cases 3b and 3c: No support-set knowledge

Recovery guarantees for the case of no support-set knowledge, but where one support set is chosen at random and the other arbitrarily can be found in [13, Thm. 6]. The theorem shown next slightly improves upon [13, Thm. 6]. The improvements are due to the following facts: i) We allow for arbitrary $0 < \delta < 1$, whereas $\delta = 1/2$ in [13, Thm. 6], and ii) we do not use a global coherence parameter $\mu = \max\{\mu_a, \mu_b, \mu_m\}$, but rather we further exploit the individual coherence parameters μ_a, μ_b, and μ_m of $\mathbf{A}$ and $\mathbf{B}$. See Section IV-A for a corresponding discussion.

Theorem 5 (Case 3b): Let $\mathbf{x}$ and $\mathbf{e}$ be signals satisfying the conditions of $\mathcal{M}(\text{P0})$, assume that $\mathcal{X}$ is chosen uniformly at random, and assume that $\mathcal{E}$ is arbitrary. If (4), (8), and (11) hold for some $0 < \delta < 1$ and $\beta \geqslant \log(n_x)$, then

$$(\text{P0}^*) \quad \underset{\hat{\mathbf{x}},\hat{\mathbf{e}}}{\text{minimize}} \, \|\hat{\mathbf{x}}\|_0 + \|\hat{\mathbf{e}}\|_0 \quad \text{subject to } \mathbf{z} = \mathbf{A}\hat{\mathbf{x}} + \mathbf{B}\hat{\mathbf{e}}$$

recovers $\mathbf{x}$ and $\mathbf{e}$ with probability at least $1 - e^{-\beta}$.

Moreover, if $\mathbf{x}$ and $\mathbf{e}$ are signals satisfying the conditions of $\mathcal{M}(\text{BP})$ and if (9) and (12) hold in addition to (4), (8), and (11), then

$$(\text{BP}^*) \quad \underset{\hat{\mathbf{x}},\hat{\mathbf{e}}}{\text{minimize}} \, \|\hat{\mathbf{x}}\|_1 + \|\hat{\mathbf{e}}\|_1 \quad \text{subject to } \mathbf{z} = \mathbf{A}\hat{\mathbf{x}} + \mathbf{B}\hat{\mathbf{e}}$$

recovers $\mathbf{x}$ and $\mathbf{e}$ with probability at least $1 - 3e^{-\beta}$.

Proof: See Appendices C and D ■

Theorem 6 (Case 3c): Let $\mathbf{x}$ and $\mathbf{e}$ be signals satisfying the conditions of $\mathcal{M}(\text{P0})$ and assume that $\mathcal{X}$ and $\mathcal{E}$ are both unknown and chosen uniformly at random. If (5), (8), and (11) hold for some $0 < \delta < 1$ and $\beta \geqslant \max\{\log(n_x), \log(n_e)\}$, then (P0^*) recovers $\mathbf{x}$ and $\mathbf{e}$ with probability at least $1 - e^{-\beta}$.

Moreover, if $\mathbf{x}$ and $\mathbf{e}$ are signals from $\mathcal{M}(\text{BP})$ and if (9) and (12) hold in addition to (5), (8), and (11), then (BP^*) recovers $\mathbf{x}$ and $\mathbf{e}$ with probability at least $1 - 3e^{-\beta}$.

Proof: See Appendices C and D. ■

A discussion of both theorems is given below.

IV. Discussion of the Recovery Guarantees

We now discuss the theorems presented in Section III. In particular, we study the impact of support-set knowledge on the recovery guarantees and characterize the asymptotic behavior of the corresponding recovery conditions, i.e., the threshold for which recovery is guaranteed with high probability.

In the ensuing discussion, we assume $\mathbf{A}$ and $\mathbf{B}$ are unitary, i.e., $n_a = n_b = m$ and $\mu_a = \mu_b = 0$, and maximally incoherent, i.e., $\mu_m = 1/\sqrt{m}$. For example, $\mathbf{A}$ could be the discrete Fourier transform (or Hadamard) matrix with appropriately normalized columns and $\mathbf{B}$ the identity matrix. We furthermore set $\beta = \log(m)$, so that recovery is guaranteed with probability at least $1 - 1/m$ and $1 - 3/m$ when solving the ℓ_0-norm and ℓ_1-norm-based recovery problems, respectively.

In order to plot the recovery conditions, we note that for a pair of unitary matrices and a given n_e, the recovery conditions of the theorems are quadratic equations in $\sqrt{n_x}$; this enables us to calculate the maximum n_x guaranteeing the successful recovery of $\mathbf{x}$ and $\mathbf{e}$ in closed form.

A. *Recovery guarantees*

1) $\mathcal{X}$ and $\mathcal{E}$ known: Figure 1 shows the recovery conditions for the cases when both support sets $\mathcal{X}$ and $\mathcal{E}$ are assumed to be known. For small problem dimensions, i.e., $m = 10^4$, the recovery conditions where both support sets are assumed to be arbitrary turn out to be less restrictive than for the case where both support sets are chosen at random. For large problem dimensions, i.e., $m = 10^8$, we see, however, that the probabilistic results of Theorem 1 guarantee the recovery (with high probability) for larger n_x and n_e than the deterministic results of [1] considering arbitrary support sets. Hence, the probabilistic recovery conditions presented here require a sufficiently large problem size in order to outperform the corresponding deterministic results. We furthermore see from Figure 1 that one can guarantee the recovery of signals having a larger number of non-zero entries if both support sets are chosen at random compared to the situation where $\mathcal{X}$ is random but $\mathcal{E}$ is arbitrary.

2) Only $\mathcal{E}$ known: Figure 2 shows the recovery conditions from Theorems 2 and 3 for the cases where only $\mathcal{E}$ is known prior to recovery (the case of only $\mathcal{X}$ known behaves analogously). We see that for a random $\mathcal{X}$ and random $\mathcal{E}$ successful recovery at high probability is guaranteed for significantly larger n_x and n_e compared to the case where one or both support sets are assumed to be arbitrary. Hence, having more randomness in the support sets leads to less restrictive recovery guarantees. We furthermore see from Figure 2 that there is no difference between the conditions for $(\text{P0}, \mathcal{E})$ and $(\text{BP}, \mathcal{E})$; this is due to the fact that $\mathbf{A}$ and $\mathbf{B}$ are both unitary. For

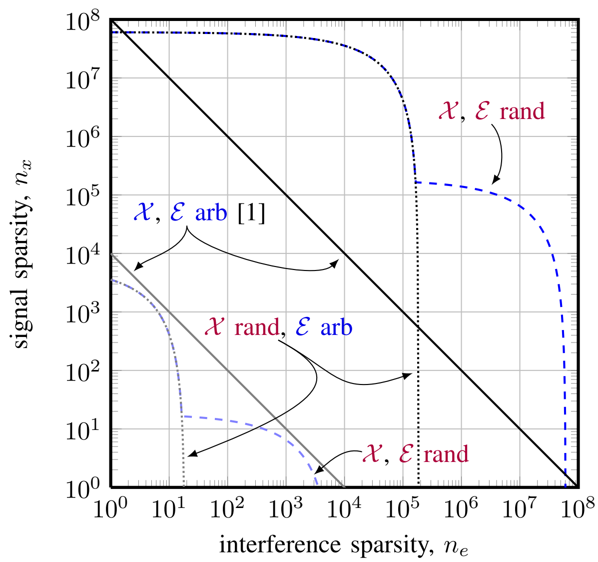

Fig. 1. Comparison of the recovery conditions for the case where $\mathcal{X}$ and $\mathcal{E}$ are known prior to recovery. $\mathbf{A}$ and $\mathbf{B}$ are unitary with $m = n_a = n_b$ and $\mu_m = 1/\sqrt{m}$; the darker curves in the upper-right are for $m = 10^8$ and the lighter curves in the lower-left are for $m = 10^4$.

arbitrary dictionary pairs $\mathbf{A}$ and $\mathbf{B}$, however, the recovery conditions for $(\text{P0}, \mathcal{E})$ are slightly less restrictive than those for $(\text{BP}, \mathcal{E})$.

3) No support-set knowledge: Finally, Figure 3 shows the recovery conditions for (BP^*) for the case of no support-set knowledge. We see that for random $\mathcal{X}$ and $\mathcal{E}$, successful recovery is guaranteed for significantly larger n_x and n_e compared to the case where one or both support sets are assumed to be arbitrary. As a comparison, we also show the recovery conditions derived in [13, Thm. 6] and the conditions from [18], the latter of which does not take into account the structure of the problem (1). We see that the recovery conditions derived in Theorems 5 and 6 are less restrictive, i.e., they guarantee the successful recovery (with high probability) for a larger number of nonzero coefficients in both the sparse signal vector $\mathbf{x}$ and the sparse interference $\mathbf{e}$.

B. Impact of support-set knowledge

As detailed in [1], having knowledge of the support set of $\mathbf{x}$ or $\mathbf{e}$ implies that one can guarantee the recovery of $\mathbf{x}$ and $\mathbf{e}$ having up to twice as many non-zero entries (compared to the case of no support-set knowledge). A similar behavior is also apparent in the probabilistic results presented here. Specifically, the recovery conditions in Figure 4 for (P0) and $(\text{P0}, \mathcal{E})$ show a

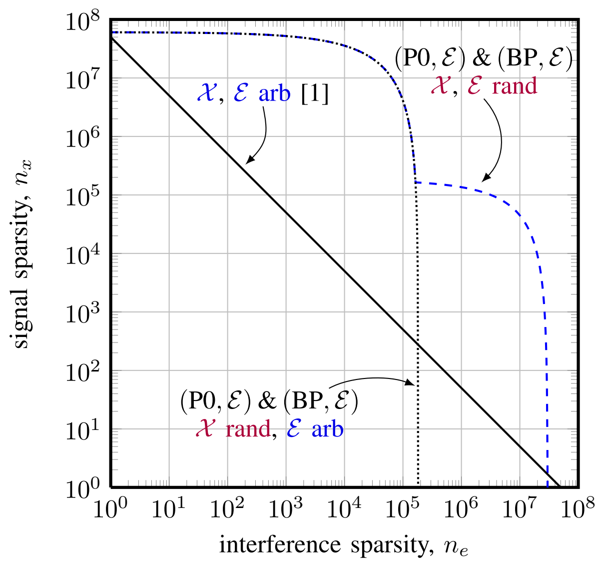

Fig. 2. Comparison of the recovery conditions for the case where only $\mathcal{E}$ is known prior to recovery. $\mathbf{A}$ and $\mathbf{B}$ are unitary with $m = n_a = n_b = 10^8$ and $\mu_m = 1/\sqrt{m}$.

similar factor-of-two gain in the case where both $\mathcal{X}$ and $\mathcal{E}$ are chosen at random. For example, knowledge of $\mathcal{X}$ enables one to recover a signal $\mathbf{x}$ with approximately twice as many non-zero components compared to the case of not knowing $\mathcal{X}$. We note that a similar gain is apparent for $\mathcal{X}$ arbitrary and $\mathcal{E}$ random, as well as for using (BP) and $(\text{BP}, \mathcal{E})$ instead of (P0) and $(\text{P0}, \mathcal{E})$.

C. Asymptotic behavior of the recovery conditions

We now compare the asymptotic behavior of probabilistic and deterministic recovery conditions, i.e., we study the scaling behavior of n_x and n_e. To this end, we are interested in the largest n_x for which recovery of $\mathbf{x}$ (and $\mathbf{e}$) from $\mathbf{z}$ can be guaranteed with high probability. In particular, we consider the following models for the sparse interference vector $\mathbf{e}$: i) Constant sparsity, i.e., $n_e = 10^3$, ii) sparsity proportional to the square root of the problem size, i.e., $n_e = \sqrt{m}$, and iii) sparsity proportional to the problem size, i.e., $n_e = m/10^5$.

Figure 5 shows the largest n_x for which recovery can be guaranteed using $(\text{BP}, \mathcal{E})$. Here, $\mathcal{E}$ is assumed to be known and arbitrary and $\mathcal{X}$ is unknown and chosen at random. Note that the other cases of support-set knowledge and arbitrary/random exhibit the same scaling behavior. We see from Figure 5 that for a constant interference sparsity (i.e., $n_e = 10^3$), the probabilistic and deterministic results show the same scaling behavior. For the cases where n_e scales with

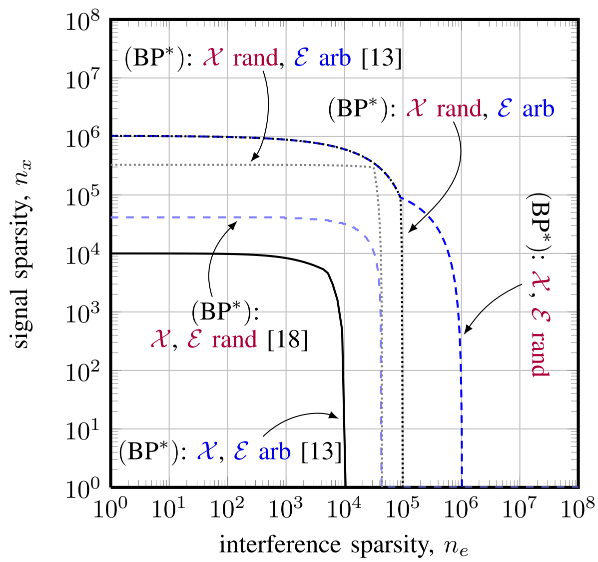

Fig. 3. Comparison of the recovery conditions for the case of no support-set knowledge. $\mathbf{A}$ and $\mathbf{B}$ are unitary with $m = n_a = n_b = 10^8$ and $\mu_m = 1/\sqrt{m}$.

$\sqrt{m}$ or m, however, the deterministic thresholds developed in [1] result in worse scaling, while the behavior of the probabilistic guarantees derived in this paper remain unaffected.

We now investigate the scaling behavior observed in Figure 5 analytically. Again, we only consider the case where $\mathcal{X}$ is unknown and chosen at random and $\mathcal{E}$ is known and chosen arbitrarily; an analysis of the other cases yields similar results. From Theorem 2, recovery of $\mathbf{x}$ with probability at least $1 - 3/n_a$ (i.e., for $\beta = \log(n_a)$) is guaranteed if

$$1 - e^{-1/4}\sqrt{\frac{n_x}{n_a}} - 3e^{-1/4}\mu_m\sqrt{2n_e \log n_a} \geqslant n_e \mu_m^2.$$

For the special case $\mu_a = \mu_b = 0$, $n_a = n_b = m$, and $\mu_m = 1/\sqrt{m}$, we get the following sufficient condition for recovery:

$$e^{1/4}\sqrt{m} \geqslant \sqrt{n_x} + 3\sqrt{2}\sqrt{n_e \log m} + e^{1/4}\frac{n_e}{\sqrt{m}}. \tag{13}$$

Hence, if $n_x \sim m$ and $n_e \sim m/\log m$, the condition (13) can be satisfied. Consequently, recovery of $\mathbf{x}$ (and of $\mathbf{e}$) is guaranteed with probability at least $1 - 3/m$ even if n_x scales *linearly* in the number of (corrupted) measurements m *and* n_e scales near-linearly (i.e., with $m/\log m$) in m.

We finally note that the recovery guarantees in [14] also allow for the sparsity of the interference vector to scale near-linearly in the number of measurements. The results in [14], however,

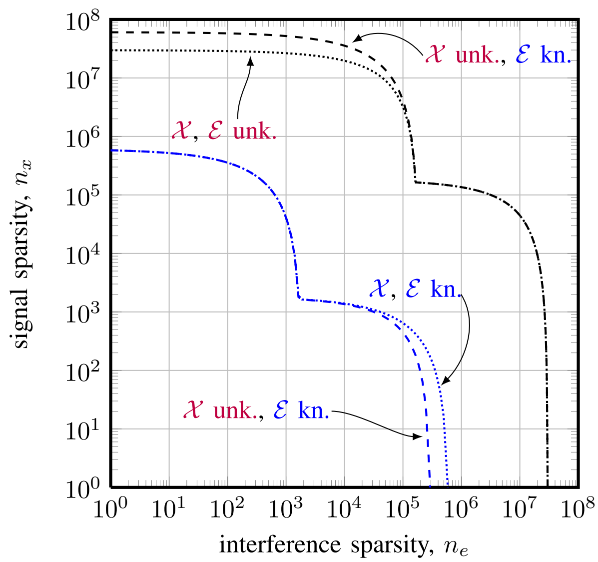

Fig. 4. Impact of support-set knowledge on the recovery conditions for (P0) and (P0, $\mathcal{E}$) in the case where $\mathcal{X}$ and $\mathcal{E}$ are both random. $\mathbf{A}$ and $\mathbf{B}$ are unitary with $m = n_a = n_b = 10^6$ (lower-left curves) and $m = n_a = n_b = 10^8$ (upper-right curves) and $\mu_m = 1/\sqrt{m}$.

require the matrix $\mathbf{A}$ to be random and $\mathbf{B}$ to be orthogonal, whereas the recovery guarantees shown here are for *arbitrary* pairs of dictionaries $\mathbf{A}$ and $\mathbf{B}$ (characterized by the coherence parameters) and for varying degrees of support-set knowledge.

V. Conclusions

In this paper, we have presented novel coherence-based recovery guarantees for sparsely corrupted signals in the probabilistic setting. In particular, we have studied the case where the sparse signal and/or sparse interference vectors are modeled as random and the dictionaries $\mathbf{A}$ and $\mathbf{B}$ are solely characterized by their coherence parameters. Our recovery guarantees complete all missing cases of support-set knowledge and improve upon the results in [1], [13]. Furthermore, we have shown that the reconstruction of sparse signals is guaranteed with high probability, even if the number of non-zero entries in both the sparse signal and sparse interference are allowed to scale (near) linearly with the number of (corrupted) measurements.

There are many avenues for follow-on work. The derivation of probabilistic recovery guarantees for the more general setting studied in [12], i.e., $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e} + \mathbf{n}$ with $\mathbf{n}$ being additive noise and $\mathbf{x}$ and $\mathbf{e}$ being approximately sparse (rather than perfectly sparse), is left for future

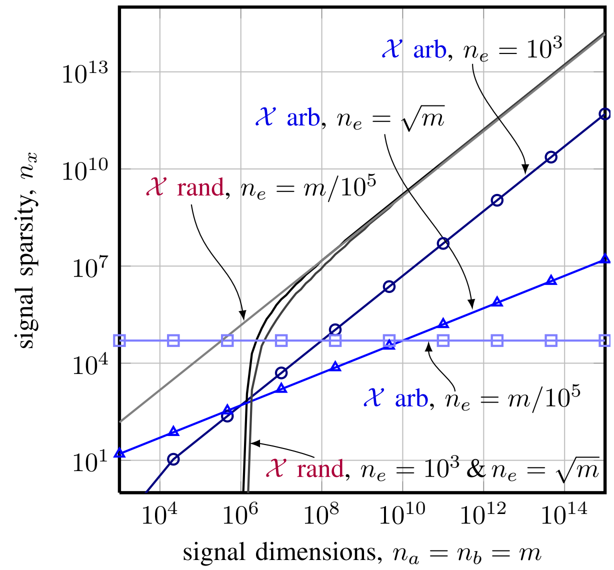

Fig. 5. Maximum signal sparsity n_x that ensures recovery of $\mathbf{x}$ for $\mathcal{E}$ known and arbitrary. We assume $n_e = 10^3$, $n_e = \sqrt{m}$, and $n_e = m/10^5$. The probability of successful recovery is set to 10^{-15}.

work. Finally, the derivation of probabilistic uncertainty relations for pairs of general dictionaries is an interesting open problem and would complete the deterministic results in [1], [13].

APPENDIX A
BOUNDS ON $\sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}})$

We now derive probabilistic bounds on $\sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}})$, which are key in showing when the recovery from sparsely corrupted signals succeeds. We extend [13, Lemma 7] to the case where both supports $\mathcal{X}$ and $\mathcal{E}$ are chosen at random and give improved results for the case where only one support set is random. First, we require the following two results from [18].

Theorem 7 (Thm. 8 of [18]): Let $\mathbf{M} \in \mathbb{C}^{m \times n}$ be a matrix. Let $\mathcal{S} \subseteq \{1, 2, \ldots, n\}$ be a set of size s drawn uniformly at random and define $\mathbf{R}_\mathcal{S}$ to be the restriction operator to the set of columns specified by $\mathcal{S}$. Fix $q \geqslant 1$, then for each $p \geqslant \max\{2, 2\log(\operatorname{rank}(\mathbf{M}\mathbf{R}_\mathcal{S}^H)), q/2\}$ we have

$$\mathbb{E}^q\Big[\|\mathbf{M}\mathbf{R}_\mathcal{S}\|_{2,2}\Big] \leqslant 3\sqrt{p}\,\|\mathbf{M}\|_{1,2} + \sqrt{\frac{s}{n}}\,\|\mathbf{M}\|_{2,2}\,,$$

where $\|\mathbf{M}\|_{1,2} = \sup_{\mathbf{v}\in\mathbb{C}^n} \|\mathbf{M}\mathbf{v}\|_2 \,/\, \|\mathbf{v}\|_1$ and is the maximum ℓ_2-norm of the columns of $\mathbf{M}$.

Lemma 8 (Eq. 6.1 of [18]): Let $\mathbf{M} \in \mathbb{C}^{m\times n}$ be a matrix with coherence μ and let $\mathcal{S} \subseteq \{1,2,\ldots,n\}$ be a set of size s chosen uniformly at random. Then, for $\beta \geqslant \log(s)$ and $q = 4\beta$

$$\mathbb{E}^q\Big[\left\|\mathbf{M}_\mathcal{S}^H\mathbf{M}_\mathcal{S} - \mathbf{I}\right\|_{2,2}\Big] \leqslant 12\mu\sqrt{\beta s} + \mathbb{1}[\mu \neq 0]\frac{2s}{n}\left\|\mathbf{M}\right\|_{2,2}^2.$$

Note that the result in [18, Eq. 6.1] does not include the indicator function $\mathbb{1}[\mu \neq 0]$. It is, however, straightforward to verify that if $\mathbf{M}$ is orthonormal, then $\mu = 0$ and hence, $\left\|\mathbf{M}_\mathcal{S}^H\mathbf{M}_\mathcal{S} - \mathbf{I}\right\|_{2,2} = 0$ for all sets $\mathcal{S}$.

We now state the main result for $\sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}})$.

Theorem 9: Choose $\beta \geqslant \log(n_x)$, $q = 4\beta$ and assume that $\mathbf{A}$ and $\mathbf{B}$ are characterized by the coherence parameters μ_a, μ_b, and μ_m. If i) $\mathcal{X}$ is chosen uniformly at random with cardinality n_x, $\mathcal{E}$ is arbitrary, and (4) holds, or ii) $\mathcal{E}$ is chosen uniformly at random with cardinality n_e, $\mathcal{X}$ is arbitrary, and (10) holds, or iii) both $\mathcal{X}$ and $\mathcal{E}$ are chosen uniformly at random with cardinalities n_x and n_e respectively, and (5) holds, then

$$\mathbb{P}\Big\{\left\|\mathbf{D}_{\mathcal{X},\mathcal{E}}^H\mathbf{D}_{\mathcal{X},\mathcal{E}} - \mathbf{I}\right\|_{2,2} \geqslant \delta\Big\} \leqslant e^{-\beta} \tag{14}$$

and if (4), (5) or (10) hold with $\delta = 1$, then

$$\mathbb{P}\{\sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}}) = 0\} \leqslant e^{-\beta}. \tag{15}$$

Proof: The proof follows that of [13, Lemma 7]. We start by defining the hollow Gram matrix

$$\mathbf{H} = \mathbf{D}_{\mathcal{X},\mathcal{E}}^H\mathbf{D}_{\mathcal{X},\mathcal{E}} - \mathbf{I} = \begin{bmatrix} \mathbf{A}_\mathcal{X}^H\mathbf{A}_\mathcal{X} - \mathbf{I} & \mathbf{A}_\mathcal{X}^H\mathbf{B}_\mathcal{E} \\ \mathbf{B}_\mathcal{E}^H\mathbf{A}_\mathcal{X} & \mathbf{B}_\mathcal{E}^H\mathbf{B}_\mathcal{E} - \mathbf{I} \end{bmatrix}.$$

Splitting $\mathbf{H}$ into diagonal and off-diagonal blocks and applying the triangle inequality leads to

$$\begin{aligned} \left\|\mathbf{H}\right\|_{2,2} &\leqslant \left\|\begin{bmatrix} \mathbf{A}_\mathcal{X}^H\mathbf{A}_\mathcal{X} - \mathbf{I} & 0 \\ 0 & \mathbf{B}_\mathcal{E}^H\mathbf{B}_\mathcal{E} - \mathbf{I} \end{bmatrix}\right\|_{2,2} + \left\|\begin{bmatrix} 0 & \mathbf{A}_\mathcal{X}^H\mathbf{B}_\mathcal{E} \\ \mathbf{B}_\mathcal{E}^H\mathbf{A}_\mathcal{X} & 0 \end{bmatrix}\right\|_{2,2} \\ &\leqslant \max\Big\{\left\|\mathbf{A}_\mathcal{X}^H\mathbf{A}_\mathcal{X} - \mathbf{I}\right\|_{2,2}, \left\|\mathbf{B}_\mathcal{E}^H\mathbf{B}_\mathcal{E} - \mathbf{I}\right\|_{2,2}\Big\} + \left\|\mathbf{B}_\mathcal{E}^H\mathbf{A}_\mathcal{X}\right\|_{2,2} \\ &\leqslant \left\|\mathbf{A}_\mathcal{X}^H\mathbf{A}_\mathcal{X} - \mathbf{I}\right\|_{2,2} + \left\|\mathbf{B}_\mathcal{E}^H\mathbf{B}_\mathcal{E} - \mathbf{I}\right\|_{2,2} + \left\|\mathbf{B}_\mathcal{E}^H\mathbf{A}_\mathcal{X}\right\|_{2,2}. \end{aligned}$$

Since the qth moment effectively defines an ℓ_q-norm, it satisfies the triangle inequality, namely, $\mathbb{E}^q[|X+Y|] \leqslant \mathbb{E}^q[|X|] + \mathbb{E}^q[|Y|]$. Hence, it follows that

$$\mathbb{E}^q\Big[\left\|\mathbf{H}\right\|_{2,2}\Big] \leqslant \mathbb{E}^q\Big[\left\|\mathbf{A}_\mathcal{X}^H\mathbf{A}_\mathcal{X} - \mathbf{I}\right\|_{2,2}\Big] + \mathbb{E}^q\Big[\left\|\mathbf{B}_\mathcal{E}^H\mathbf{B}_\mathcal{E} - \mathbf{I}\right\|_{2,2}\Big] + \mathbb{E}^q\Big[\left\|\mathbf{B}_\mathcal{E}^H\mathbf{A}_\mathcal{X}\right\|_{2,2}\Big]. \tag{16}$$

We now separately bound each of the terms in (16) and we do this for each case where $\mathcal{X}$ and $\mathcal{E}$ is either chosen at random or arbitrarily. If $\mathcal{X}$ is chosen uniformly at random, then it follows from Lemma 8 that

$$\mathbb{E}^q\left[\left\|\mathbf{A}_{\mathcal{X}}^H\mathbf{A}_{\mathcal{X}}-\mathbf{I}\right\|_{2,2}\right] \leqslant 12\mu_a\sqrt{\beta n_x}+\mathbb{1}[\mu_a \neq 0]\frac{2n_x}{n_a}\left\|\mathbf{A}\right\|_{2,2} \tag{17}$$

for any $4\beta = q \geqslant 4\log(n_x)$. If $\mathcal{X}$ is allowed to be arbitrary, then for all $\mathcal{X}$ we have

$$\left\|\mathbf{A}_{\mathcal{X}}^H\mathbf{A}_{\mathcal{X}}-\mathbf{I}\right\|_{2,2} \leqslant \max_k \sum_{j\neq k}\left|[\mathbf{A}_{\mathcal{X}}^H\mathbf{A}_{\mathcal{X}}]_{j,k}\right| \leqslant (n_x-1)\mu_a, \tag{18}$$

where the first inequality follows from the Geršgorin disc theorem [30, Thm. 6.1.1] and the second inequality is a consequence of the definition of μ_a. By reversing the role of $\mathbf{A}$ and $\mathbf{B}$, we get the analogous bounds for the right-hand side (RHS) term $\mathbb{E}^q\left[\left\|\mathbf{B}_{\mathcal{E}}^H\mathbf{B}_{\mathcal{E}}-\mathbf{I}\right\|_{2,2}\right]$ in (16).

For the third summand appearing in the RHS of (16), let us first consider the case where $\mathcal{E}$ is chosen arbitrarily and $\mathcal{X}$ uniformly at random. We then want to apply Theorem 7 to $\mathbf{M}=\mathbf{B}_{\mathcal{E}}^H\mathbf{A}$ with $\mathbf{R}_{\mathcal{X}}$. Since $\mathbf{M}\mathbf{R}_{\mathcal{X}}$ is an $n_e \times n_x$ matrix, $\operatorname{rank}(\mathbf{M}\mathbf{R}_{\mathcal{X}}) \leqslant \min\{n_x, n_e\}$ and thus we can apply Theorem 7 with $q = 2p = 4\beta$ where $q \geqslant 4\min\{\log(n_x),\log(n_e)\} \geqslant 4\log(\operatorname{rank}(\mathbf{M}\mathbf{R}_{\mathcal{X}}))$ to get

$$\begin{aligned}\mathbb{E}^q\left[\left\|\mathbf{B}_{\mathcal{E}}^H\mathbf{A}_{\mathcal{X}}\right\|_{2,2}\right] &\leqslant 3\sqrt{p}\left\|\mathbf{B}_{\mathcal{E}}^H\mathbf{A}\right\|_{1,2}+\sqrt{\frac{n_x}{n_a}}\left\|\mathbf{B}_{\mathcal{E}}^H\mathbf{A}\right\|_{2,2}\\ &\leqslant 3\mu_m\sqrt{2\beta n_e}+\sqrt{\frac{n_x}{n_a}}\left\|\mathbf{B}^H\mathbf{A}\right\|_{2,2},\end{aligned} \tag{19}$$

where the entries of $\mathbf{B}_{\mathcal{E}}^H\mathbf{A}$ are bounded by the mutual coherence μ_m. The case where $\mathcal{E}$ is random and $\mathcal{X}$ is arbitrary follows by reversing the roles of $\mathbf{A}$ and $\mathbf{B}$.

Now consider the case where both $\mathcal{E}$ and $\mathcal{X}$ are random. Set $\mathbf{M}=\mathbf{B}^H\mathbf{A}$ and then, as in [18, Sec. 6], we can write

$$\begin{aligned}\mathbb{E}^q\left[\left\|\mathbf{B}_{\mathcal{E}}^H\mathbf{A}_{\mathcal{X}}\right\|_{2,2}\right] &= \mathbb{E}^q\left[\left\|\mathbf{R}_{\mathcal{E}}\mathbf{M}\mathbf{R}_{\mathcal{X}}\right\|_{2,2}\right]\\ &= \mathbb{E}^q_{\mathbf{R}_{\mathcal{E}}}\left[\mathbb{E}^q_{\mathbf{R}_{\mathcal{X}}}\left[\left\|\mathbf{R}_{\mathcal{E}}\mathbf{M}\mathbf{R}_{\mathcal{X}}\right\|_{2,2}\right]\right].\end{aligned} \tag{20}$$

Let us first consider the inner expectation term of (20). We set $\mathbf{M}'=\mathbf{R}_{\mathcal{E}}\mathbf{M}$ and apply Theorem 7 to get

$$\begin{aligned}\mathbb{E}^q_{\mathbf{R}_{\mathcal{X}}}\left[\left\|\mathbf{M}'\mathbf{R}_{\mathcal{X}}\right\|_{2,2}\right] &\leqslant 3\sqrt{p}\left\|\mathbf{M}'\right\|_{1,2}+\sqrt{\frac{n_x}{n_a}}\left\|\mathbf{M}'\right\|_{2,2}\\ &= 3\sqrt{p}\left\|\mathbf{B}_{\mathcal{E}}^H\mathbf{A}\right\|_{1,2}+\sqrt{\frac{n_x}{n_a}}\left\|\mathbf{B}_{\mathcal{E}}^H\mathbf{A}\right\|_{2,2},\end{aligned}$$

where we take $q = 2p = 4\beta \geqslant 4\min\{\log(n_x), \log(n_e)\} \geqslant 4\log(\operatorname{rank}(\mathbf{M}'\mathbf{R}_\mathcal{X}))$ as $\mathbf{M}'\mathbf{R}_\mathcal{X}$ is an $n_e \times n_x$ matrix. Thus, (20) results in

$$\mathbb{E}^q\left[\left\|\mathbf{B}_\mathcal{E}^H\mathbf{A}_\mathcal{X}\right\|_{2,2}\right] \leqslant 3\sqrt{p}\,\mathbb{E}^q_{\mathbf{R}_\mathcal{E}}\left[\left\|\mathbf{R}_\mathcal{E}\mathbf{B}^H\mathbf{A}\right\|_{1,2}\right] + \sqrt{\frac{n_x}{n_a}}\,\mathbb{E}^q_{\mathbf{R}_\mathcal{E}}\left[\left\|\mathbf{R}_\mathcal{E}\mathbf{B}^H\mathbf{A}\right\|_{2,2}\right].$$

Since one may choose the dictionaries $\mathbf{A}$ and $\mathbf{B}$ so that $\left\|\mathbf{R}_\mathcal{E}\mathbf{B}^H\mathbf{A}\right\|_{1,2} = \left\|\mathbf{R}_{\mathcal{E}'}\mathbf{B}^H\mathbf{A}\right\|_{1,2}$ for all $\mathcal{E}, \mathcal{E}'$ of the same cardinality (e.g., by setting $\mathbf{A}$ to the identity and $\mathbf{B}$ to the DFT matrix), we bound this term by $\mu_m\sqrt{n_e}$ and the corresponding expectation term disappears.[2] Hence, we have

$$\mathbb{E}^q\left[\left\|\mathbf{B}_\mathcal{E}^H\mathbf{A}_\mathcal{X}\right\|_{2,2}\right] \leqslant 3\mu_m\sqrt{2\beta n_e} + \sqrt{\frac{n_x}{n_a}}\left\|\mathbf{A}^H\mathbf{B}\right\|_{2,2}.$$

By performing the same steps with $\mathbf{M} = \mathbf{A}^H\mathbf{B}$, we arrive at

$$\begin{aligned}\mathbb{E}^q\left[\left\|\mathbf{B}_\mathcal{E}^H\mathbf{A}_\mathcal{X}\right\|_{2,2}\right] \leqslant \min\Bigg\{&3\mu_m\sqrt{2\beta n_x} + \sqrt{\frac{n_e}{n_b}}\left\|\mathbf{A}^H\mathbf{B}\right\|_{2,2}, \\ &3\mu_m\sqrt{2\beta n_e} + \sqrt{\frac{n_x}{n_a}}\left\|\mathbf{A}^H\mathbf{B}\right\|_{2,2}\Bigg\},\end{aligned} \tag{21}$$

for any $\beta \geqslant \min\{\log(n_x), \log(n_e)\}$. Combining (17), (18), (19), and (21) with the analogous results for $\mathbf{B}$ and $\mathcal{E}$ leads to the conditions (4), (10), and (5).

Due to (17) and the analogous result for $\mathbf{B}_\mathcal{E}$, if $\mathcal{X}$ is chosen at random, we require $\beta \geqslant \log(n_x)$, if $\mathcal{E}$ is chosen at random we need $\beta \geqslant \log(n_e)$, and if both $\mathcal{X}$ and $\mathcal{E}$ are chosen at random, both of these conditions need to be satisfied, namely that $\beta \geqslant \max\{\log(n_x), \log(n_e)\}$.

We now show that the conditions (4), (10), and (5) are sufficient to show that (14) holds. Chebyshev's Inequality [31, Sec. 1.3] states that for a random variable X and a function $f\colon \mathbb{R} \to \mathbb{R}^+$

$$\mathbb{P}\{X \in \mathcal{A}\} \leqslant \frac{\mathbb{E}[f(X)]}{\inf\{f(x)\colon x \in \mathcal{A}\}}. \tag{22}$$

Application of (22) with $f(x) = x^q$ and the random variable $X = \left\|\mathbf{D}_\mathcal{S}^H\mathbf{D}_\mathcal{S} - \mathbf{I}\right\|_{2,2}$ gives

$$\mathbb{P}\{X \geqslant \delta\} \leqslant \frac{\mathbb{E}[X^q]}{\inf\{x^q\colon x \geqslant \delta\}} \leqslant \frac{\left(\delta e^{-1/4}\right)^q}{\delta^q} = e^{-q/4}, \tag{23}$$

provided that $(\delta e^{-1/4})^q \geqslant \mathbb{E}[X^q]$. But this is guaranteed by the assumptions in (4), (5), or (10), depending on the signal and interference model. Therefore, we have

$$\mathbb{P}\Big\{\|\mathbf{H}\|_{2,2} \geqslant \delta\Big\} \leqslant e^{-\beta},$$

[2]Note that using Theorem 7 twice to bound $\mathbb{E}^q_{\mathbf{R}_\mathcal{E}}\left[\left\|\mathbf{R}_\mathcal{E}\mathbf{B}^H\mathbf{A}\right\|_{2,2}\right]$ from above leads to a worse result than bounding $\left\|\mathbf{R}_\mathcal{E}\mathbf{B}^H\mathbf{A}\right\|_{2,2}$ straightforwardly with $\left\|\mathbf{B}^H\mathbf{A}\right\|_{2,2}$.

since $q = 4\beta$. The second part of the theorem, (15), is a result of the fact that $\sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}}) = 0$ implies that $\|\mathbf{H}\|_{2,2} \geqslant 1$ and hence, $\mathbb{P}\{\sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}}) = 0\} \leqslant \mathbb{P}\left\{\|\mathbf{H}\|_{2,2} \geqslant 1\right\}$. ■

APPENDIX B
BOTH SUPPORTS KNOWN

Proof of Theorem 1: It suffices to show that $\mathbf{D}_{\mathcal{X},\mathcal{E}}$ is invertible, which is equivalent to the condition that $\sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}}) > 0$. By assumption, the conditions of Theorem 9 hold, which implies $\mathbb{P}\{\sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}}) = 0\} \leqslant e^{-\beta}$. Hence, recovery of $\mathbf{x}$ and $\mathbf{e}$ using (3) succeeds with probability at least $1 - e^{-\beta}$. ■

APPENDIX C
(P0) WITH LIMITED SUPPORT KNOWLEDGE

We now prove the recovery guarantees for (P0*), (P0, $\mathcal{E}$), and (P0, $\mathcal{X}$) for partial (or no) support-set knowledge of $\mathcal{E}$ and $\mathcal{X}$. We follow the proof of [18] and present the three cases 1) $\mathcal{X}$ known, 2) $\mathcal{E}$ known, and 3) no support-set knowledge, all together, since the corresponding proofs are similar. Note that $\mathcal{R}(\mathbf{D})$ denotes the space spanned by the columns of $\mathbf{D}$.

We begin by generalizing [18, Thm. 13] to the case of pairs of dictionaries $\mathbf{A}$ and $\mathbf{B}$ where we know the support set of $\mathbf{e}$. The result gives us a sufficient condition for when there is a unique minimizer of (P0*), (P0, $\mathcal{E}$), or (P0, $\mathcal{X}$).

Lemma 10 (Based on Thm. 13 of [18]): Let $\tilde{\mathbf{A}} \in \mathbb{C}^{n_a \times m}$ and $\tilde{\mathbf{B}} \in \mathbb{C}^{n_b \times m}$ be two dictionaries and suppose that we observe the signal $\mathbf{z} = \tilde{\mathbf{A}}\mathbf{x} + \tilde{\mathbf{B}}\mathbf{e}$ where $\mathcal{X} = \operatorname{supp}(\mathbf{x})$ and $\mathcal{E} = \operatorname{supp}(\mathbf{e})$ and the non-zero entries of $\mathbf{x}$ are drawn from a continuous distribution. Furthermore, suppose that $\mathcal{E}$ is known. If

$$\dim\left(\mathcal{R}\left(\tilde{\mathbf{D}}_{\mathcal{X},\mathcal{E}}\right) \cap \mathcal{R}\left(\tilde{\mathbf{D}}_{\mathcal{X}',\mathcal{E}}\right)\right) < |\mathcal{X}| + |\mathcal{E}|\,, \tag{24}$$

for all sets $\mathcal{X}'$ where $|\mathcal{X}| = |\mathcal{X}'|$, then, almost surely, (P0, $\mathcal{E}$) recovers the vectors $\mathbf{x}$ and $\mathbf{e}$. This result also provides a sufficient condition for (P0*), if we set $\tilde{\mathbf{A}} = \mathbf{D}$ and take $\tilde{\mathbf{B}}$ to be the empty matrix, or for (P0, $\mathcal{X}$), if we set $\tilde{\mathbf{A}} = \mathbf{B}$ and $\tilde{\mathbf{B}} = \mathbf{A}$.

Proof: We follow the proof of [18, Thm. 13]. We begin by defining the set $\mathcal{A}^{\mathcal{E}}_{\mathcal{X},\mathcal{X}'}$ as follows:

$$\mathcal{A}^{\mathcal{E}}_{\mathcal{X},\mathcal{X}'} = \left\{\mathbf{z} \colon \mathbf{z} = \tilde{\mathbf{A}}_{\mathcal{X}}\mathbf{x}_{\mathcal{X}} + \tilde{\mathbf{B}}_{\mathcal{E}}\mathbf{e}_{\mathcal{E}} = \tilde{\mathbf{A}}_{\mathcal{X}'}\mathbf{x}'_{\mathcal{X}'} + \tilde{\mathbf{B}}_{\mathcal{E}}\mathbf{e}'_{\mathcal{E}}\right\},$$

so that $\mathcal{A}^{\mathcal{E}}_{\mathcal{X},\mathcal{X}'}$ is the set of observations that can be written in terms of two pairs of signals $(\mathbf{x},\mathbf{e})$ and $(\mathbf{x}',\mathbf{e}')$ where $\mathcal{X}=\text{supp}(\mathbf{x})$, $\mathcal{X}'=\text{supp}(\mathbf{x}')$, and $\mathcal{E}=\text{supp}(\mathbf{e})=\text{supp}(\mathbf{e}')$.

For any $\mathcal{X}'$ of size $|\mathcal{X}|$ and $\mathcal{X}'\neq\mathcal{X}$, we have

$$\mathcal{A}^{\mathcal{E}}_{\mathcal{X},\mathcal{X}'}\subseteq\mathcal{R}\Big(\tilde{\mathbf{D}}_{\mathcal{X},\mathcal{E}}\Big)\cap\mathcal{R}\Big(\tilde{\mathbf{D}}_{\mathcal{X}',\mathcal{E}}\Big).$$

Now assume that (24) holds for $\mathcal{X}$, $\mathcal{X}'$, and $\mathcal{E}$, then $\dim(\mathcal{A}^{\mathcal{E}}_{\mathcal{X},\mathcal{X}'})<|\mathcal{X}|+|\mathcal{E}|$. Thus, the smallest subspace containing $\mathcal{A}^{\mathcal{E}}_{\mathcal{X},\mathcal{X}'}$ is a strict subspace of $\mathcal{R}(\tilde{\mathbf{D}}_{\mathcal{X},\mathcal{E}})$ and hence, has zero measure with respect to any nonatomic measure. Since $\mathbf{x}$ and hence $\mathbf{z}$, have non-zero entries drawn from a continuous distribution

$$\mathbb{P}\Big\{\tilde{\mathbf{A}}\mathbf{x}+\tilde{\mathbf{B}}\mathbf{e}=\mathbf{z}\in\mathcal{A}^{\mathcal{E}}_{\mathcal{X},\mathcal{X}'}\Big\}=0.$$

Thus, with probability zero, there exists no alternative pair $(\mathbf{x}',\mathbf{e}')$ with supports $\mathcal{X}'$ and $\mathcal{E}$, respectively, otherwise $\mathbf{z}$ would lie in $\mathcal{A}^{\mathcal{E}}_{\mathcal{X},\mathcal{X}'}$. If (24) holds for all $\mathcal{X}'$, then, almost surely, given $\mathbf{z}=\tilde{\mathbf{A}}\mathbf{x}+\tilde{\mathbf{B}}\mathbf{e}$, $(\text{P0},\mathcal{E})$ returns the vectors $\mathbf{x}$ and $\mathbf{e}$. ■

We can use Lemma 10 to prove the first part of Theorems 2, 3, 4, 5, and 6 by showing that (24) holds with high probability. To show that (24) holds for all $\mathcal{X}'$ we show that for every column $\tilde{\mathbf{a}}_\gamma$ of $\tilde{\mathbf{A}}$ not in $\tilde{\mathbf{A}}_\mathcal{X}$ (i.e., for all $\gamma\notin\mathcal{X}$) that $\tilde{\mathbf{a}}_\gamma\notin\mathcal{R}(\tilde{\mathbf{D}}_{\mathcal{X},\mathcal{E}})$, which is equivalent to showing that

$$\left\|\tilde{\mathbf{P}}_{\mathcal{X},\mathcal{E}}\tilde{\mathbf{a}}_\gamma\right\|_2<\|\tilde{\mathbf{a}}_\gamma\|_2=1, \tag{25}$$

for all $\gamma\notin\mathcal{X}$ and where $\tilde{\mathbf{P}}_{\mathcal{X},\mathcal{E}}=\tilde{\mathbf{D}}^{\dagger}_{\mathcal{X},\mathcal{E}}\tilde{\mathbf{D}}^{H}_{\mathcal{X},\mathcal{E}}$ is the projection onto the range space of $\tilde{\mathbf{D}}_{\mathcal{X},\mathcal{E}}$. We will now bound the probability that (25) holds for the following three situations: 1) only $\mathcal{E}$ known, 2) only $\mathcal{X}$ known, and 3) both support sets unknown.

1) Only $\mathcal{E}$ known: Consider the setting where $\mathcal{E}$ is known, but $\mathcal{X}$ is unknown; this case fits the setting of Lemma 10 with $\tilde{\mathbf{A}}=\mathbf{A}$ and $\tilde{\mathbf{B}}=\mathbf{B}$. Hence, the condition (25) is equivalent to $\|\mathbf{P}_{\mathcal{X},\mathcal{E}}\mathbf{a}_\gamma\|_2<\|\mathbf{a}_\gamma\|_2=1$. We have

$$\begin{aligned}\|\mathbf{P}_{\mathcal{X},\mathcal{E}}\mathbf{a}_\gamma\|_2&\leqslant\left\|\mathbf{D}^{\dagger}_{\mathcal{X},\mathcal{E}}\right\|_{2,2}\left\|\mathbf{D}^{H}_{\mathcal{X},\mathcal{E}}\mathbf{a}_\gamma\right\|_2\\&\leqslant\sigma^{-1}_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}})\sqrt{\|\mathbf{A}^H_{\mathcal{X}}\mathbf{a}_\gamma\|_2^2+\|\mathbf{B}^H_{\mathcal{E}}\mathbf{a}_\gamma\|_2^2}.\end{aligned}$$

From the definitions of the coherence parameters[3]

$$\left\|\mathbf{D}^H_{\mathcal{X},\mathcal{E}}\mathbf{a}_\gamma\right\|_2 \leqslant \xi_{\mathcal{E}} = \sqrt{\mu_a^2 n_x + \mu_m^2 n_e}. \tag{26}$$

Thus, in order to guarantee $\left\|\mathbf{P}_{\mathcal{X},\mathcal{E}}\mathbf{a}_\gamma\right\|_2 < 1$ it suffices to have

$$\xi_{\mathcal{E}} < \sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}}). \tag{27}$$

2) Only $\mathcal{X}$ known: For the setting where only $\mathcal{X}$ is known, we apply Lemma 10 with $\tilde{\mathbf{A}} = \mathbf{B}$ and $\tilde{\mathbf{B}} = \mathbf{A}$, thus the condition of (24) becomes

$$\dim(\mathcal{R}(\mathbf{D}_{\mathcal{X},\mathcal{E}}) \cap \mathcal{R}(\mathbf{D}_{\mathcal{X},\mathcal{E}'})) < |\mathcal{X}| + |\mathcal{E}|\,,$$

and so we only want to show that $\left\|\mathbf{P}_{\mathcal{X},\mathcal{E}}\mathbf{b}_\gamma\right\|_2 < \left\|\mathbf{b}_\gamma\right\|_2$ for all $\gamma \notin \mathcal{E}$. Proceeding as before, it follows that

$$\begin{aligned}\left\|\mathbf{P}_{\mathcal{X},\mathcal{E}}\mathbf{b}_\gamma\right\|_2 &\leqslant \left\|\mathbf{D}^\dagger_{\mathcal{X},\mathcal{E}}\right\|_{2,2}\left\|\mathbf{D}^H_{\mathcal{X},\mathcal{E}}\mathbf{b}_\gamma\right\|_2 \\ &\leqslant \sigma^{-1}_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}})\left\|\mathbf{D}^H_{\mathcal{X},\mathcal{E}}\mathbf{b}_\gamma\right\|_2 \\ &\leqslant \sigma^{-1}_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}})\,\xi_{\mathcal{X}},\end{aligned} \tag{28}$$

where $\xi_{\mathcal{X}} = \sqrt{\mu_m^2 n_x + \mu_b^2 n_e}$. Hence, it suffices to show that

$$\xi_{\mathcal{X}} < \sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}}). \tag{29}$$

3) No support-set knowledge: Finally, we consider the setting where neither $\mathcal{X}$ nor $\mathcal{E}$ is known, so we apply Lemma 10 with $\tilde{\mathbf{A}} = [\,\mathbf{A}\ \mathbf{B}\,]$ and $\tilde{\mathbf{B}}$ being the empty matrix, thus this is exactly the condition of [18, Thm. 13]. Then, we show that $\left\|\mathbf{P}_{\mathcal{X},\mathcal{E}}\mathbf{d}_\gamma\right\|_2 < \left\|\mathbf{d}_\gamma\right\|_2$ for any column $\mathbf{d}_\gamma$ of $\mathbf{D}$ not in $\mathbf{D}_{\mathcal{X},\mathcal{E}}$. In other words, we want both (26) and (28) to hold as $\mathbf{d}_\gamma$ can be a column of either $\mathbf{A}$ or $\mathbf{B}$. So it suffices to show

$$\left\|\mathbf{P}_{\mathcal{X},\mathcal{E}}\mathbf{d}_\gamma\right\|_2 \leqslant \sigma^{-1}_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}})\,\xi_+ < 1, \tag{30}$$

where $\xi_+ = \max\{\xi_{\mathcal{X}}, \xi_{\mathcal{E}}\}$.

Finally, to show that the (P0) based problems succeed, we want to bound the probability that (27), (29), or (30) holds (depending on which, if any, support sets we know). In each of the

[3]Note that we use bounds that hold for *all* $\mathcal{X}$, rather than a bound that holds with high probability. The underlying reason is the fact that if $\mathbf{A}$ is an equiangular tight frame, the associated inequalities hold with equality and hence, we cannot do any better by using probabilistic bounds, unless we take advantage of a property of $\mathbf{A}$ other than the coherence μ_a.

cases, we know that (P0*), $(\text{P0}, \mathcal{E})$, or $(\text{P0}, \mathcal{X})$ returns the correct solution if $\xi < \sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}})$, where $\xi \in (0, 1)$ is equal to $\xi_{\mathcal{E}}$, $\xi_{\mathcal{X}}$, or ξ_+ (as appropriate to the case). Hence, we can bound the probability of error as follows

$$\begin{aligned}\mathbb{P}\{\text{error}\} &\leqslant \mathbb{P}\{\xi \geqslant \sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}})\}\\ &\leqslant \mathbb{P}\left\{\left\|\mathbf{D}_{\mathcal{X},\mathcal{E}}^H \mathbf{D}_{\mathcal{X},\mathcal{E}} - \mathbf{I}\right\|_{2,2} \geqslant 1 - \xi^2\right\} \leqslant e^{-\beta},\end{aligned}$$

where we use Theorem 9 with $\delta = 1 - \xi^2$. Therefore, with probability exceeding $1 - e^{-\beta}$, the pair $(\mathbf{x}, \mathbf{e})$ is the unique minimizer.

APPENDIX D
(BP) WITH LIMITED SUPPORT KNOWLEDGE

We now prove the recovery results for the (BP) based algorithms. To do this, we extend the sufficient recovery conditions of [32] to (BP*), $(\text{BP}, \mathcal{E})$, and $(\text{BP}, \mathcal{X})$, however we first require the following useful lemma.

Lemma 11 (Lem. 6 of [32]): Suppose that $\mathbf{u}$ is a vector whose components are all nonzero and that $\mathbf{v}$ is a vector whose entries do not have identical moduli. Then

$$|\langle \mathbf{u}, \mathbf{v} \rangle| < \|\mathbf{u}\|_1 \|\mathbf{v}\|_\infty .$$

We now give a sufficient condition for when (BP*), $(\text{BP}, \mathcal{E})$ or $(\text{BP}, \mathcal{X})$ can recover the solution we want. We then show that we can satisfy this condition with high probability, thus proving our recovery theorems.

Lemma 12: Let $\tilde{\mathbf{A}} \in \mathbb{C}^{n_a \times m}$ and $\tilde{\mathbf{B}} \in \mathbb{C}^{n_b \times m}$ be two dictionaries and suppose that we observe the signal $\mathbf{z} = \tilde{\mathbf{A}}\mathbf{x} + \tilde{\mathbf{B}}\mathbf{e}$ where $\mathcal{X} = \text{supp}(\mathbf{x})$ and $\mathcal{E} = \text{supp}(\mathbf{e})$. Write $\mathbf{s} = [\, \mathbf{x}^T \; \mathbf{e}^T \,]^T$, $\tilde{\mathbf{D}} = [\, \tilde{\mathbf{A}} \; \tilde{\mathbf{B}} \,]$, and assume that $\tilde{\mathbf{D}}_{\mathcal{X},\mathcal{E}} = [\, \tilde{\mathbf{A}}_{\mathcal{X}} \; \tilde{\mathbf{B}}_{\mathcal{E}} \,]$ is full rank. Suppose that both $\mathcal{X}$ and $\mathcal{E}$ are unknown. If there exists a vector $\mathbf{h} \in \mathbb{C}^m$ such that

$$\tilde{\mathbf{D}}_{\mathcal{X},\mathcal{E}}^H \mathbf{h} = \text{sign}(\mathbf{s}_{\mathcal{X},\mathcal{E}}), \text{ and} \tag{31a}$$

$$|\langle \mathbf{h}, \tilde{\mathbf{d}}_\gamma \rangle| < 1 \text{ for all columns } \tilde{\mathbf{d}}_\gamma \text{ of } \tilde{\mathbf{D}} \text{ not in } \tilde{\mathbf{D}}_{\mathcal{X},\mathcal{E}}, \tag{31b}$$

then $(\mathbf{x}, \mathbf{e})$ is the unique minimizer of (BP*).

Now suppose that $\mathcal{E}$ is known. If there exists a vector $\mathbf{h} \in \mathbb{C}^m$ such that

$$\tilde{\mathbf{A}}_{\mathcal{X}}^H \mathbf{h} = \operatorname{sign}(\mathbf{x}_{\mathcal{X}}), \text{ and} \tag{32a}$$

$$|\langle \mathbf{h}, \tilde{\mathbf{a}}_\gamma \rangle| < 1 \text{ for all columns } \tilde{\mathbf{a}}_\gamma \text{ of } \tilde{\mathbf{A}} \text{ not in } \tilde{\mathbf{A}}_{\mathcal{X}}, \tag{32b}$$

then $(\mathbf{x}, \mathbf{e}_{\mathcal{E}})$ is the unique minimizer of $(\mathrm{BP}, \mathcal{E})$.
Note that conditions for recovery when using $(\mathrm{BP}, \mathcal{X})$ are identical to those of $(\mathrm{BP}, \mathcal{E})$, except we exchange the roles of both dictionaries and both vectors $\mathbf{x}$ and $\mathbf{e}$.

Proof: Follows Thm. 5 of [32]: First note that (31a) and (31b) are exactly the conditions of [32, Thm. 5] only rewritten in terms of the concatenated dictionary $\tilde{\mathbf{D}} = [\, \tilde{\mathbf{A}} \ \tilde{\mathbf{B}} \,]$, so we only need to prove the sufficient condition for $(\mathrm{BP}, \mathcal{E})$.

Suppose that $\mathbf{x}$ and $\mathbf{e}$ give the sparsest representation of the signal and suppose that $\mathbf{x}'$ and $\mathbf{e}'$ give an alternative representation with supports $\mathcal{X}'$ and $\mathcal{E}' = \mathcal{E}$, respectively. We want to show that $\|\mathbf{x}\|_1 < \|\mathbf{x}'\|_1$ and hence, that the pair $(\mathbf{x}, \mathbf{e})$ is the unique $(\mathrm{BP}, \mathcal{E})$ minimizer. First, observe that

$$\|\mathbf{x}\|_1 = \left|\operatorname{sign}(\mathbf{x})^H \mathbf{x}\right| = \left|\mathbf{h}^H \tilde{\mathbf{A}}_{\mathcal{X}} \mathbf{x}\right| = \left|\mathbf{h}^H \tilde{\mathbf{A}}_{\mathcal{X}'} \mathbf{x}'\right| = \left\langle \mathbf{x}', \tilde{\mathbf{A}}_{\mathcal{X}'}^H \mathbf{h} \right\rangle.$$

If the entries of $\tilde{\mathbf{A}}_{\mathcal{X}'} \mathbf{h}$ do not have constant moduli, then we can apply Lemma 11 to get

$$\begin{aligned} \|\mathbf{x}\|_1 = \left\langle \mathbf{x}', \tilde{\mathbf{A}}_{\mathcal{X}'}^H \mathbf{h} \right\rangle &< \|\mathbf{x}'\|_1 \left\| \tilde{\mathbf{A}}_{\mathcal{X}'}^H \mathbf{h} \right\|_\infty \\ &= \|\mathbf{x}'\|_1 \max_{\gamma \in \mathcal{X}'} \left\langle \mathbf{h}, \mathbf{a}'_\gamma \right\rangle \leqslant \|\mathbf{x}'\|_1 , \end{aligned}$$

since by assumption $\left|\left\langle \mathbf{h}, \mathbf{a}'_\gamma \right\rangle\right| < 1$ for any $\gamma \notin \mathcal{X}$ (by (32b)) and $\left|\left\langle \mathbf{h}, \mathbf{a}'_\gamma \right\rangle\right| = 1$ for any $\gamma \in \mathcal{X}$ (by (32a)). Therefore, $\|\mathbf{x}\|_1 < \|\mathbf{x}'\|_1$.

Now suppose that the entries of $\tilde{\mathbf{A}}_{\mathcal{X}'} \mathbf{h}$ have constant moduli, that is, $\max_{\gamma \in \mathcal{X}'} \langle \mathbf{h}, \mathbf{d}_\gamma \rangle = |\langle \mathbf{h}, \mathbf{d}_\gamma \rangle|$ for all $\gamma \in \mathcal{X}'$. If $\mathbf{x}'$ has support $\mathcal{X}' \subset \mathcal{X}$, then $[\, \tilde{\mathbf{A}}_{\mathcal{X}} \ \tilde{\mathbf{B}}_{\mathcal{E}} \,]$ cannot be full rank (as both $\mathbf{e}$ and $\mathbf{e}'$ have support $\mathcal{E}$). Hence there must exist some $\gamma \in \mathcal{X}' \setminus \mathcal{X}$. Then, since by assumption $|\langle \mathbf{h}, \mathbf{d}_\gamma \rangle| < 1$, we have

$$\begin{aligned} \|\mathbf{x}\|_1 = \left\langle \mathbf{x}', \tilde{\mathbf{A}}_{\mathcal{X}'}^H \mathbf{h} \right\rangle &= \|\mathbf{x}'\|_1 \left\| \tilde{\mathbf{A}}_{\mathcal{X}'}^H \mathbf{h} \right\|_\infty \\ &= \|\mathbf{x}'\|_1 \max_{\gamma \in \mathcal{X}'} \langle \mathbf{h}, \mathbf{d}_\gamma \rangle = \|\mathbf{x}'\|_1 |\langle \mathbf{h}, \mathbf{d}_\gamma \rangle| < \|\mathbf{x}'\|_1 . \end{aligned}$$

In both cases we have shown that $\|\mathbf{x}\|_1 < \|\mathbf{x}'\|_1$ and hence, the pair $(\mathbf{x}, \mathbf{e})$ is the unique $(\mathrm{BP}, \mathcal{E})$ minimizer. ∎

Finally, before we can prove the probabilistic recovery guarantees for the ℓ_1-norm-based algorithms of Theorems 2, 3, 4, 5, and 6, we require the following lemma.

Lemma 13 (Bernstein's Inequality, Prop. 16 of [18]): Let $\mathbf{v} \in \mathbb{C}^n$ and let $\boldsymbol{\varepsilon} \in \mathbb{C}^n$ be a Steinhaus sequence. Then, for $u \geqslant 0$ we have

$$\mathbb{P}\left\{\left|\sum_{i=1}^{n} \varepsilon_i v_i\right| \geqslant u \left\|\mathbf{v}\right\|_2\right\} \leqslant 2\exp\left(-\frac{u^2}{2}\right). \tag{33}$$

A Steinhaus sequence is a (countable) collection of independent complex-valued random variables, whose entries are uniformly distributed on the unit circle [18].

We now prove the second part of Theorems 2, 3, 4, 5, and 6. To show that recovery with (BP*), (BP, $\mathcal{E}$), or (BP, $\mathcal{X}$) succeeds, we demonstrate that the vector $\mathbf{h}$, as in Lemma 12, exists with high probability. We now consider the following three settings in turn: 1) only $\mathcal{E}$ known, 2) only $\mathcal{X}$ known, and 3) both support sets unknown. But first, let us assume that in each case $\mathbf{D}_{\mathcal{X},\mathcal{E}}$ is full rank.

1) Only $\mathcal{E}$ known: Consider the case where $\mathcal{E}$ is known but $\mathcal{X}$ is unknown, so we want to use Lemma 12 with $\tilde{\mathbf{A}} = \mathbf{A}$ and $\tilde{\mathbf{B}} = \mathbf{B}$ and we show that a vector $\mathbf{h}$ exists that satisfies (32a) and (32b) with high probability. To this end, set $\mathbf{h} = \mathbf{A}_{\mathcal{X}}\left(\mathbf{A}_{\mathcal{X}}^H\mathbf{A}_{\mathcal{X}}\right)^{-1}\operatorname{sign}(\mathbf{x}_{\mathcal{X}})$, so that (32a) is satisfied. Then, for any column $\mathbf{a}_\gamma$ of $\mathbf{A}$ where $\gamma \notin \mathcal{X}$,

$$\begin{aligned}
|\langle \mathbf{h}, \mathbf{a}_\gamma\rangle| &= \left|\left\langle \mathbf{A}_{\mathcal{X}}^{\dagger}\operatorname{sign}(\mathbf{x}_{\mathcal{X}}), \mathbf{a}_\gamma\right\rangle\right| \\
&= \left|\left\langle \operatorname{sign}(\mathbf{x}_{\mathcal{X},}), \left(\mathbf{A}_{\mathcal{X}}^H\mathbf{A}_{\mathcal{X}}\right)^{-1}\mathbf{A}_{\mathcal{X}}^H\mathbf{a}_\gamma\right\rangle\right| = \left|\sum_{j=1}^{n_x}\varepsilon_j v_j^\gamma\right|,
\end{aligned}$$

with $\boldsymbol{\varepsilon} = \operatorname{sign}(\mathbf{x}_{\mathcal{X}})$ and $\mathbf{v}^\gamma = \left(\mathbf{A}_{\mathcal{X}}^H\mathbf{A}_{\mathcal{X}}\right)^{-1}\mathbf{A}_{\mathcal{X}}^H\mathbf{a}_\gamma$. If $\mathbf{A}$ is unitary, then $\mathbf{v}^\gamma = \mathbf{0}$ and (32b) is immediately satisfied. However if $\mathbf{A}$ is not unitary, we need to proceed as follows. Since $\boldsymbol{\varepsilon}$ is a Steinhaus sequence by assumption, we can apply Lemma 13 with $u = \|\mathbf{v}^\gamma\|_2^{-1}$ to arrive at

$$\mathbb{P}\left\{\left|\sum_{j=1}^{n_x}\varepsilon_j v_j^\gamma\right| \geqslant 1\right\} \leqslant 2\exp\left(-\frac{1}{2\left\|\mathbf{v}^\gamma\right\|_2^2}\right). \tag{34}$$

But we have that

$$\begin{aligned}
\|\mathbf{v}^\gamma\|_2^2 &= \left\|\left(\mathbf{A}_{\mathcal{X}}^H\mathbf{A}_{\mathcal{X}}\right)^{-1}\mathbf{A}_{\mathcal{X}}^H\mathbf{a}_\gamma\right\|_2^2 \\
&\leqslant \left\|\left(\mathbf{A}_{\mathcal{X}}^H\mathbf{A}_{\mathcal{X}}\right)^{-1}\right\|_{2,2}^2 \left\|\mathbf{A}_{\mathcal{X}}^H\mathbf{a}_\gamma\right\|_2^2 \leqslant \sigma_{\min}^{-4}(\mathbf{A}_{\mathcal{X}})\,\xi_{\mathcal{E}}^2,
\end{aligned}$$

where $\xi_{\mathcal{E}}^2 = n_x \mu_a^2$. Hence, (34) results in

$$\mathbb{P}\left\{\left|\sum_{j=1}^{n_x} \varepsilon_j v_j^\gamma\right| \geqslant 1\right\} \leqslant 2\exp\left(-\frac{\sigma_{\min}^4(\mathbf{A}_{\mathcal{X}})}{2\xi_{\mathcal{E}}^2}\right).$$

Now we want (31b) to hold for all $\gamma \notin \mathcal{X}$. Hence, applying the union bound to the result above leads to

$$\mathbb{P}\left\{\max_{\gamma\notin\mathcal{X}}\left|\sum_{j=1}^{n_x} \varepsilon_j v_j^\gamma\right| \geqslant 1\right\} \leqslant 2n_a\exp\left(-\frac{\sigma_{\min}^4(\mathbf{A}_{\mathcal{X}})}{2\xi_{\mathcal{E}}^2}\right). \tag{35}$$

To make this result consistent with the case when $\mathbf{A}$ is unitary we will multiply the right hand side of (35) with the term $\mathbb{1}[\mu_a \neq 0]$ so that the right hand side of (35) is 0 if $\mathbf{A}$ is unitary.

2) Only $\mathcal{X}$ known: Consider the setting where $\mathcal{X}$ is known, but $\mathcal{E}$ is unknown. Again we use Lemma 12 but with $\tilde{\mathbf{A}} = \mathbf{B}$ and $\tilde{\mathbf{B}} = \mathbf{A}$. Proceeding as before, we show that

$$|\langle \mathbf{h}, \mathbf{b}_\gamma\rangle| = \left|\sum_{j=1}^{n_e} \varepsilon_j v_j^\gamma\right| < 1,$$

for all $\gamma \notin \mathcal{E}$ and where $\mathbf{v}^\gamma = \mathbf{B}_{\mathcal{E}}^\dagger \mathbf{b}_\gamma$. If $\mathbf{B}$ is not unitary, we have

$$\|\mathbf{v}^\gamma\|_2^2 \leqslant \left\|\left(\mathbf{B}_{\mathcal{E}}^H\mathbf{B}_{\mathcal{E}}\right)^{-1}\right\|_{2,2}^2 \left\|\mathbf{B}_{\mathcal{E}}^H\mathbf{b}_\gamma\right\|_2^2 \leqslant \sigma_{\min}^{-4}(\mathbf{B}_{\mathcal{E}})\,\xi_{\mathcal{X}}^2,$$

where $\xi_{\mathcal{X}}^2 = n_e\mu_b^2$. Hence, we obtain

$$\mathbb{P}\left\{\left|\sum_{j=1}^{n_x+n_e} \varepsilon_j v_j^\gamma\right| \geqslant 1\right\} \leqslant 2\exp\left(-\frac{\sigma_{\min}^4(\mathbf{B}_{\mathcal{E}})}{2\xi_{\mathcal{X}}^2}\right).$$

Finally, we want (31b) to hold for all $\gamma \notin \mathcal{E}$. Therefore, applying the union bound to the above leads to

$$\mathbb{P}\left\{\max_{\gamma\notin\mathcal{E}}\left|\sum_{j=1}^{n_e} \varepsilon_j v_j^\gamma\right| \geqslant 1\right\} \leqslant 2n_b\exp\left(-\frac{\sigma_{\min}^4(\mathbf{B}_{\mathcal{E}})}{2\xi_{\mathcal{X}}^2}\right). \tag{36}$$

3) No support-set knowledge: Finally, we consider the third setting where neither $\mathcal{X}$ nor $\mathcal{E}$ are known. In particular, we want to show that in Lemma 12, we can satisfy (31a) and (31b) with high probability. For any column $\mathbf{d}_\gamma$ of $\mathbf{D}$ not in $\mathbf{D}_{\mathcal{X},\mathcal{E}}$, set $\mathbf{v}^\gamma = \mathbf{D}_{\mathcal{X},\mathcal{E}}^\dagger \mathbf{d}_\gamma$. In this case, we have

$$\begin{aligned}\|\mathbf{v}^\gamma\|_2^2 &\leqslant \left\|\left(\mathbf{D}_{\mathcal{X},\mathcal{E}}^H\mathbf{D}_{\mathcal{X},\mathcal{E}}\right)^{-1}\right\|_{2,2}^2 \left\|\mathbf{D}_{\mathcal{X},\mathcal{E}}^H\mathbf{d}_\gamma\right\|_2^2 \\ &\leqslant \sigma_{\min}^{-4}(\mathbf{D}_{\mathcal{X},\mathcal{E}})\,\xi_+^2,\end{aligned}$$

where $\xi_+^2 = \max\{n_x\mu_a^2 + n_e\mu_m^2, n_x\mu_m^2 + n_e\mu_b^2\}$ and hence,

$$\mathbb{P}\left\{\left|\sum_{j=1}^{n_x+n_e} \varepsilon_j v_j^\gamma\right| \geqslant 1\right\} \leqslant 2\exp\left(-\frac{\sigma_{\min}^4(\mathbf{D}_{\mathcal{X},\mathcal{E}})}{2\xi_+^2}\right).$$

Finally, we want (31b) to hold for all $\mathbf{d}_\gamma$. Therefore, applying the union bound to the result above leads to

$$\mathbb{P}\left\{\max_{\gamma\notin\mathcal{X}\cup\mathcal{E}}\left|\sum_{j=1}^{n_x+n_e} \varepsilon_j v_j^\gamma\right| \geqslant 1\right\} \leqslant 2(n_a+n_b)\exp\left(-\frac{\sigma_{\min}^4(\mathbf{D}_{\mathcal{X},\mathcal{E}})}{2\xi_+^2}\right). \tag{37}$$

We now want to derive an upper bound on the right hand sides of (35), (36), and (37). First, we use the facts that $\sigma_{\min}(\mathbf{A}_\mathcal{X}) \geqslant \sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}})$ and $\sigma_{\min}(\mathbf{B}_\mathcal{E}) \geqslant \sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}})$ to give an upper bound for (35) and (36). Then, we calculate the probability conditioned on $\sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}}) > \lambda \in (0,1)$. Note that if $\lambda > 0$, then $\sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}}) > \lambda > 0$ and we satisfy the remaining assumption of Lemma 12, namely that $\mathbf{D}_{\mathcal{X},\mathcal{E}}$ is full rank.

For convenience, in the case where $\mathcal{E}$ is known, let us set $N = n_a$, $\xi = \xi_\mathcal{E}$ and $\zeta = \mathbb{1}[\mu_a \neq 0]$. In the case where $\mathcal{X}$ is known, set $N = n_b$, $\xi = \xi_\mathcal{X}$ and $\zeta = \mathbb{1}[\mu_b \neq 0]$ and finally, in the case where neither $\mathcal{X}$ nor $\mathcal{E}$ are known, set $N = n_a + n_b$, $\xi = \xi_+$ and $\zeta = 1$. Thus, we have

$$\mathbb{P}\left\{\max_{\gamma\notin\mathcal{S}}\left|\sum_{j=1}^{N} \varepsilon_j v_j^\gamma\right| \geqslant 1 \middle| \sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}}) > \lambda\right\} \leqslant 2N\zeta\exp\left(-\frac{\lambda^4}{2\xi^2}\right) \leqslant 2\zeta e^{-\beta}, \tag{38}$$

for some $\beta \leqslant \lambda^4/(2\xi^2) - \log N$.

For our particular choice of $\mathbf{h}$, (32a) (in the case where $\mathcal{X}$ or $\mathcal{E}$ is known) or (31a) (in the case where both supports are unknown) will always be satisfied. So let $\mathfrak{E}$ be the event that (32b) (in the case where one support is unknown) or (31b) (in the case where both supports are known) is not fulfilled with our choice of $\mathbf{h}$ and let $\mathfrak{R}$ be the event that $\mathbf{D}_{\mathcal{X},\mathcal{E}}$ is not full rank. As $\mathfrak{E}\cup\mathfrak{R}$ is a necessary condition for the (BP) based algorithms not to be able to recover the vectors $\mathbf{x}$ and $\mathbf{e}$, $\mathbb{P}\{\mathfrak{E}\cup\mathfrak{R}\}$ is an upper bound on the probability of error. Then, since $\sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}}) > \lambda > 0$ implies that $\mathfrak{R}$ cannot occur, and hence that $\mathbb{P}\{\mathfrak{E}\cup\mathfrak{R} | \sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}}) > \lambda\} = \mathbb{P}\{\mathfrak{E} | \sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}}) > \lambda\}$, we have that for any $\lambda > 0$

$$\begin{aligned}\mathbb{P}\{\mathfrak{E}\cup\mathfrak{R}\} &= \mathbb{P}\{\mathfrak{E}\cup\mathfrak{R} | \sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}}) > \lambda\}\,\mathbb{P}\{\sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}}) > \lambda\} \\ &\quad + \mathbb{P}\{\mathfrak{E}\cup\mathfrak{R} | \sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}}) \leqslant \lambda\}\,\mathbb{P}\{\sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}}) \leqslant \lambda\} \\ &\leqslant \mathbb{P}\{\mathfrak{E} | \sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}}) > \lambda\} + \mathbb{P}\{\sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}}) \leqslant \lambda\}.\end{aligned} \tag{39}$$

We can bound the first summand in (39) using (38) under the assumption that $\beta \leqslant \lambda^4/(2\xi^2) - \log N$. The second term we can bound using Theorem 9 with $\delta = 1 - \lambda^2 \in (0, 1)$, which, provided that $\beta \geqslant N'$ where N' is the size of the supports chosen at random, says that

$$\mathbb{P}\{\sigma_{\min}(\mathbf{D}_{\mathcal{X},\mathcal{E}}) \leqslant \lambda\} \leqslant e^{-\beta}.$$

Therefore, we have

$$\mathbb{P}\{\mathfrak{E} \cup \mathfrak{R}\} \leqslant (2\zeta + 1)e^{-\beta} \leqslant 3e^{-\beta}, \tag{40}$$

and hence, we can recover $\mathbf{x}$ and $\mathbf{e}$ with probability at least $1 - 3e^{-\beta}$.

References

[1] C. Studer, P. Kuppinger, G. Pope, and H. Bölcskei, "Recovery of sparsely corrupted signals," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, May 2012.

[2] J. N. Laska, P. T. Boufounos, M. A. Davenport, and R. G. Baraniuk, "Democracy in action: Quantization, saturation, and compressive sensing," *App. Comp. Harm. Anal.*, vol. 31, no. 3, pp. 429–443, 2011.

[3] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "A constrained matching pursuit approach to audio declipping," in *Proc. of IEEE Int. Conf. Acoustics, Speech, and Sig. Proc.*, Prague, Czech Republic, May 2011, pp. 329–332.

[4] ——, "Audio inpainting," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 20, no. 3, pp. 922–932, Mar. 2012.

[5] S. V. Vaseghi and R. Frayling-Cork, "Restoration of old gramophone recordings," *Journal of Audio Engineering*, vol. 40, pp. 791–801, Oct. 1992.

[6] S. J. Godsill and P. J. W. Rayner, *Digital audio restoration: a statistical model based approach*. Springer-Verlag, 1998.

[7] C. Novak, C. Studer, A. Burg, and G. Matz, "The effect of unreliable LLR storage on the performance of MIMO-BICM," in *Proc. of 44th Asilomar Conf. on Signals, Systems, and Comput.*, Pacific Grove, CA, USA, Nov. 2010, pp. 736–740.

[8] M. Elad and Y. Hel-Or, "A fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur," *IEEE Trans. Image Process*, vol. 10, no. 8, pp. 1187–1193, Aug. 2001.

[9] S. G. Mallat and G. Yu, "Super-resolution with sparse mixing estimators," *IEEE Trans. Image Process*, vol. 19, no. 11, pp. 2889–2900, Nov. 2010.

[10] M. Elad, J.-L. Starck, P. Querre, and D. L. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *App. Comp. Harm. Anal.*, vol. 19, pp. 340–358, Dec. 2005.

[11] J.-F. Cai, S. Osher, and Z. Shen, "Split bregman methods and frame based image restoration," *Multiscale Model. Simul*, Dec. 2009.

[12] C. Studer and R. G. Baraniuk, "Stable restoration and separation of approximately sparse signals," *submitted*.

[13] P. Kuppinger, G. Durisi, and H. Bölcskei, "Uncertainty relations and sparse signal recovery for pairs of general signal sets," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 263–277, Jan. 2012.

[14] X. Li, "Compressed sensing and matrix completion with constant proportion of corruptions," *arXiv:1104.1041v1*, Apr. 2011.

[15] J. N. Laska, M. A. Davenport, and R. G. Baraniuk, “Exact signal recovery from sparsely corrupted measurements through the pursuit of justice,” *preprint*, Dec. 2010.

[16] N. Vaswani and W. Lu, “Modified-CS: Modifying compressive sensing for problems with partially known support,” *IEEE Trans. Sig. Proc.*, vol. 58, no. 9, pp. 4595–4607, Sep. 2010.

[17] L. Jacques, “A short note on compressed sensing with partially known signal support,” *Signal Processing*, vol. 90, pp. 3308–3312, 2010.

[18] J. A. Tropp, “On the conditioning of random subdictionaries,” *App. Comp. Harm. Anal.*, vol. 25, pp. 1–24, July 2008.

[19] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization,” *Proc. of Natl. Acad. Sci*, vol. 100, no. 5, pp. 2197–2202, Dec. 2003.

[20] J. A. Tropp, “Greed is good: Algorithmic results for sparse approximation,” *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[21] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[22] E. J. Candès and J. Romberg, “Sparsity and incoherence in compressive sampling,” *Inverse Problems*, vol. 23, no. 3, pp. 969–985, 2007.

[23] E. J. Candès and Y. Plan, “A probabilistic and RIPless theory of compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7235–7254., Nov. 2010.

[24] E. J. Candès, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Comm. Pure and Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[25] E. J. Candès, “The restricted isometry property and its implications for compressed sensing,” *C. R. Acad. Sci. Paris*, vol. 346, no. 9-10, pp. 589–592, 2008.

[26] T. Cai, L. Wany, and G. Xu, “Stable recovery of sparse signals and an oracle inequality,” *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3516–3522, Jul. 2010.

[27] L. R. Welch, “Lower bounds on the maximum cross correlation of signals,” *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 397–399, May 1974.

[28] E. J. Candès and T. Tao, “Decoding by linear programming,” *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[29] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[30] R. Horn and C. Johnson, *Matrix analysis*. Cambridge Univ. Press, 1990.

[31] R. Durrett, *Probability: Theory and Examples*. Cambridge Univ. Press, 2010.

[32] J. A. Tropp, “Recovery of short, complex linear combinations via ℓ_1 minimization,” *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1568–1570, Apr. 2005.