Recovery of Sparsely Corrupted Signals

Christoph Studer, *Member, IEEE*, Patrick Kuppinger, *Student Member, IEEE*, Graeme Pope, *Student Member, IEEE*, and Helmut Bölcskei, *Fellow, IEEE*

***Abstract*—We investigate the recovery of signals exhibiting a sparse representation in a general (i.e., possibly redundant or incomplete) dictionary that are corrupted by additive noise admitting a sparse representation in another general dictionary. This setup covers a wide range of applications, such as image inpainting, super-resolution, signal separation, and recovery of signals that are impaired by, e.g., clipping, impulse noise, or narrowband interference. We present deterministic recovery guarantees based on a novel uncertainty relation for pairs of general dictionaries and we provide corresponding practicable recovery algorithms. The recovery guarantees we find depend on the signal and noise sparsity levels, on the coherence parameters of the involved dictionaries, and on the amount of prior knowledge about the signal and noise support sets.**

***Index Terms*—Uncertainty relations, signal restoration, signal separation, coherence-based recovery guarantees, ℓ_1-norm minimization, greedy algorithms.**

I. Introduction

We consider the problem of identifying the sparse vector $\mathbf{x} \in \mathbb{C}^{N_a}$ from M linear and non-adaptive measurements collected in the vector

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e} \qquad (1)$$

where $\mathbf{A} \in \mathbb{C}^{M \times N_a}$ and $\mathbf{B} \in \mathbb{C}^{M \times N_b}$ are known deterministic and general (i.e., not necessarily of the same cardinality, and possibly redundant or incomplete) *dictionaries*, and $\mathbf{e} \in \mathbb{C}^{N_b}$ represents a sparse noise vector. The support set of $\mathbf{e}$ and the corresponding nonzero entries can be *arbitrary*; in particular, $\mathbf{e}$ may also depend on $\mathbf{x}$ and/or the dictionary $\mathbf{A}$.

This recovery problem occurs in many applications, some of which are described next:

- *Clipping:* Non-linearities in (power-)amplifiers or in analog-to-digital converters often cause signal clipping or saturation [2]. This impairment can be cast into the signal model (1) by setting $\mathbf{B} = \mathbf{I}_M$, where $\mathbf{I}_M$ denotes the $M \times M$ identity matrix, and rewriting (1) as $\mathbf{z} = \mathbf{y} + \mathbf{e}$ with $\mathbf{e} = g_a(\mathbf{y}) - \mathbf{y}$. Concretely, instead of the M-dimensional signal vector $\mathbf{y} = \mathbf{A}\mathbf{x}$ of interest, the device in question delivers $g_a(\mathbf{y})$, where the function $g_a(\mathbf{y})$ realizes entry-wise signal clipping to the interval $[-a, +a]$. The vector $\mathbf{e}$ will be sparse, provided the clipping level is high enough. Furthermore, in this case the support set of $\mathbf{e}$ can be identified prior to recovery, by simply comparing the absolute values of the entries of $\mathbf{y}$ to the clipping threshold a. Finally, we note that here it is essential that the noise vector $\mathbf{e}$ be allowed to depend on the vector $\mathbf{x}$ and/or the dictionary $\mathbf{A}$.
- *Impulse noise:* In numerous applications, one has to deal with the recovery of signals corrupted by impulse noise [3]. Specific applications include, e.g., reading out from unreliable memory [4] or recovery of audio signals impaired by click/pop noise, which typically occurs during playback of old phonograph records. The model in (1) is easily seen to incorporate such impairments. Just set $\mathbf{B} = \mathbf{I}_M$ and let $\mathbf{e}$ be the impulse-noise vector. We would like to emphasize the generality of (1) which allows impulse noise that is sparse in general dictionaries $\mathbf{B}$.
- *Narrowband interference:* In many applications one is interested in recovering audio, video, or communication signals that are corrupted by narrowband interference. Electric hum, as it may occur in improperly designed audio or video equipment, is a typical example of such an impairment. Electric hum typically exhibits a sparse representation in the Fourier basis as it (mainly) consists of a tone at some base-frequency and a series of corresponding harmonics, which is captured by setting $\mathbf{B} = \mathbf{F}_M$ in (1), where $\mathbf{F}_M$ is the M-dimensional discrete Fourier transform (DFT) matrix defined below in (2).
- *Super-resolution and inpainting:* Our framework also encompasses super-resolution [5], [6] and inpainting [7] for images, audio, and video signals. In both applications, only a subset of the entries of the (full-resolution) signal vector $\mathbf{y} = \mathbf{A}\mathbf{x}$ is available and the task is to fill in the missing entries of the signal vector such that $\mathbf{y} = \mathbf{A}\mathbf{x}$. The missing entries are accounted for by choosing the vector $\mathbf{e}$ such that the entries of $\mathbf{z} = \mathbf{y} + \mathbf{e}$ corresponding to the missing entries in $\mathbf{y}$ are set to some (arbitrary) value, e.g., 0. The missing entries of $\mathbf{y}$ are then filled in by first recovering $\mathbf{x}$ from $\mathbf{z}$ and then computing $\mathbf{y} = \mathbf{A}\mathbf{x}$. Note that in both applications the support set $\mathcal{E}$ is known (i.e., the locations of the missing entries can easily be identified) and the dictionary $\mathbf{A}$ is typically redundant (see, e.g., [8] for a corresponding discussion), i.e., $\mathbf{A}$ has more dictionary elements (columns) than rows, which

Part of this paper was presented at the IEEE International Symposium on Information Theory (ISIT), Saint-Petersburg, Russia, July 2011 [1]. This work was supported in part by the Swiss National Science Foundation (SNSF) under Grant PA00P2-134155.

C. Studer was with the Dept. of Information Technology and Electrical Engineering, ETH Zurich, Switzerland, and is now with the Dept. of Electrical and Computer Engineering, Rice University, Houston, TX, USA (e-mail: studer@rice.edu).

P. Kuppinger was with the Dept. of Information Technology and Electrical Engineering, ETH Zurich, Switzerland, and is now with UBS, Zurich, Switzerland (e-mail: patrick.kuppinger@gmail.com).

G. Pope and H. Bölcskei are with the Dept. of Information Technology and Electrical Engineering, ETH Zurich, Switzerland (e-mail: gpope@nari.ee.ethz.ch; boelcskei@nari.ee.ethz.ch).

demonstrates the need for recovery results that apply to general (i.e., possibly redundant) dictionaries.

- *Signal separation:* Separation of (audio or video) signals into two distinct components also fits into our framework. A prominent example for this task is the separation of texture from cartoon parts in images (see [9], [10] and references therein). In the language of our setup, the dictionaries $\mathbf{A}$ and $\mathbf{B}$ are chosen such that they allow for sparse representation of the two distinct features; $\mathbf{x}$ and $\mathbf{e}$ are the corresponding coefficients describing these features (sparsely). Note that here the vector $\mathbf{e}$ no longer plays the role of (undesired) noise. Signal separation then amounts to simultaneously extracting the sparse vectors $\mathbf{x}$ and $\mathbf{e}$ from the observation (e.g., the image) $\mathbf{z} = \mathbf{Ax} + \mathbf{Be}$.

Naturally, it is of significant practical interest to identify fundamental limits on the recovery of $\mathbf{x}$ (and $\mathbf{e}$, if appropriate) from $\mathbf{z}$ in (1). For the noiseless case $\mathbf{z} = \mathbf{Ax}$ such *recovery guarantees* are known [11]–[13] and typically set limits on the maximum allowed number of nonzero entries of $\mathbf{x}$ or—more colloquially—on the "sparsity" level of $\mathbf{x}$. These recovery guarantees are usually expressed in terms of restricted isometry constants (RICs) [14], [15] or in terms of the coherence parameter [11]–[13], [16] of the dictionary $\mathbf{A}$. In contrast to coherence parameters, RICs can, in general, not be computed efficiently. In this paper, we focus exclusively on coherence-based recovery guarantees. For the case of unstructured noise, i.e., $\mathbf{z} = \mathbf{Ax} + \mathbf{n}$ with no constraints imposed on $\mathbf{n}$ apart from $\|\mathbf{n}\|_2 < \infty$, coherence-based recovery guarantees were derived in [16]–[20]. The corresponding results, however, do not guarantee perfect recovery of $\mathbf{x}$, but only ensure that either the recovery error is bounded above by a function of $\|\mathbf{n}\|_2$ or only guarantee perfect recovery of the support set of $\mathbf{x}$. Such results are to be expected, as a consequence of the generality of the setup in terms of the assumptions on the noise vector $\mathbf{n}$.

A. Contributions

In this paper, we consider the following questions: 1) Under which conditions can the vector $\mathbf{x}$ (and the vector $\mathbf{e}$, if appropriate) be recovered *perfectly* from the (sparsely corrupted) observation $\mathbf{z} = \mathbf{Ax} + \mathbf{Be}$, and 2) can we formulate practical recovery algorithms with corresponding (analytical) performance guarantees? Sparsity of the signal vector $\mathbf{x}$ and the error vector $\mathbf{e}$ will turn out to be key in answering these questions. More specifically, based on an uncertainty relation for pairs of general dictionaries, we establish recovery guarantees that depend on the number of nonzero entries in $\mathbf{x}$ and $\mathbf{e}$, and on the coherence parameters of the dictionaries $\mathbf{A}$ and $\mathbf{B}$. These recovery guarantees are obtained for the following different cases: I) The support sets of both $\mathbf{x}$ and $\mathbf{e}$ are known (prior to recovery), II) the support set of only $\mathbf{x}$ or only $\mathbf{e}$ is known, III) the number of nonzero entries of only $\mathbf{x}$ or only $\mathbf{e}$ is known, and IV) nothing is known about $\mathbf{x}$ and $\mathbf{e}$. We formulate efficient recovery algorithms and derive corresponding performance guarantees. Finally, we compare our analytical recovery thresholds to numerical results and we demonstrate the application of our algorithms and recovery guarantees to an image inpainting example.

B. Outline of the paper

The remainder of the paper is organized as follows. In Section II, we briefly review relevant previous results. In Section III, we derive a novel uncertainty relation that lays the foundation for the recovery guarantees reported in Section IV. A discussion of our results is provided in Section V and numerical results are presented in Section VI. We conclude in Section VII.

C. Notation

Lowercase boldface letters stand for column vectors and uppercase boldface letters designate matrices. For the matrix $\mathbf{M}$, we denote its transpose and conjugate transpose by $\mathbf{M}^T$ and $\mathbf{M}^H$, respectively, its (Moore–Penrose) pseudo-inverse by $\mathbf{M}^\dagger = \left(\mathbf{M}^H\mathbf{M}\right)^{-1}\mathbf{M}^H$, its kth column by $\mathbf{m}_k$, and the entry in the kth row and ℓth column by $[\mathbf{M}]_{k,\ell}$. The kth entry of the vector $\mathbf{m}$ is $[\mathbf{m}]_k$. The space spanned by the columns of $\mathbf{M}$ is denoted by $\mathcal{R}(\mathbf{M})$. The $M \times M$ identity matrix is denoted by $\mathbf{I}_M$, the $M \times N$ all zeros matrix by $\mathbf{0}_{M,N}$, and the all-zeros vector of dimension M by $\mathbf{0}_M$. The $M \times M$ discrete Fourier transform matrix $\mathbf{F}_M$ is defined as

$$[\mathbf{F}_M]_{k,\ell} = \frac{1}{\sqrt{M}} \exp\left(-\frac{2\pi i(k-1)(\ell-1)}{M}\right), \ k,\ell = 1,\ldots,M \tag{2}$$

where $i^2 = -1$. The Euclidean (or ℓ_2) norm of the vector $\mathbf{x}$ is denoted by $\|\mathbf{x}\|_2$, $\|\mathbf{x}\|_1$ stands for the ℓ_1-norm of $\mathbf{x}$, and $\|\mathbf{x}\|_0$ designates the number of nonzero entries in $\mathbf{x}$. Throughout the paper, we assume that the columns of the dictionaries $\mathbf{A}$ and $\mathbf{B}$ have unit ℓ_2-norm. The minimum and maximum eigenvalue of the positive-semidefinite matrix $\mathbf{M}$ is denoted by $\lambda_{\text{min}}(\mathbf{M})$ and $\lambda_{\text{max}}(\mathbf{M})$, respectively. The spectral norm of the matrix $\mathbf{M}$ is $\|\mathbf{M}\| = \sqrt{\lambda_{\text{max}}(\mathbf{M}^H\mathbf{M})}$. Sets are designated by uppercase calligraphic letters; the cardinality of the set $\mathcal{T}$ is $|\mathcal{T}|$. The complement of a set $\mathcal{S}$ (in some superset $\mathcal{T}$) is denoted by $\mathcal{S}^c$. For two sets $\mathcal{S}_1$ and $\mathcal{S}_2$, $s \in \left(\mathcal{S}_1 + \mathcal{S}_2\right)$ means that s is of the form $s = s_1 + s_2$, where $s_1 \in \mathcal{S}_1$ and $s_2 \in \mathcal{S}_2$. The support set of the vector $\mathbf{m}$ is designated by $\text{supp}(\mathbf{m})$. The matrix $\mathbf{M}_\mathcal{T}$ is obtained from $\mathbf{M}$ by retaining the columns of $\mathbf{M}$ with indices in $\mathcal{T}$; the vector $\mathbf{m}_\mathcal{T}$ is obtained analogously. We define the $N \times N$ diagonal (projection) matrix $\mathbf{P}_\mathcal{S}$ for the set $\mathcal{S} \subseteq \{1,\ldots,N\}$ as follows:

$$[\mathbf{P}_\mathcal{S}]_{k,\ell} = \begin{cases} 1, & k = \ell \text{ and } k \in \mathcal{S} \\ 0, & \text{otherwise.} \end{cases}$$

For $x \in \mathbb{R}$, we set $[x]^+ = \max\{x, 0\}$.

II. Review of Relevant Previous Results

Recovery of the vector $\mathbf{x}$ from the sparsely corrupted measurement $\mathbf{z} = \mathbf{Ax} + \mathbf{Be}$ corresponds to a sparse-signal recovery problem subject to *structured* (i.e., sparse) noise. In this section, we briefly review relevant existing results for sparse-signal recovery from noiseless measurements, and we summarize the results available for recovery in the presence of unstructured and structured noise.

A. Recovery in the noiseless case

Recovery of $\mathbf{x}$ from $\mathbf{z} = \mathbf{A}\mathbf{x}$ where $\mathbf{A}$ is redundant (i.e., $M < N_a$) amounts to solving an underdetermined linear system of equations. Hence, there are infinitely many solutions $\mathbf{x}$, in general. However, under the assumption of $\mathbf{x}$ being sparse, the situation changes drastically. More specifically, one can recover $\mathbf{x}$ from the observation $\mathbf{z} = \mathbf{A}\mathbf{x}$ by solving

$$\text{(P0)} \quad \text{minimize } \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathbf{z} = \mathbf{A}\mathbf{x}.$$

This approach results, however, in prohibitive computational complexity, even for small problem sizes. Two of the most popular and computationally tractable alternatives to solving (P0) by an exhaustive search are basis pursuit (BP) [11]–[13], [21]–[23] and orthogonal matching pursuit (OMP) [13], [24], [25]. BP is essentially a convex relaxation of (P0) and amounts to solving

$$\text{(BP)} \quad \text{minimize } \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{z} = \mathbf{A}\mathbf{x}.$$

OMP is a greedy algorithm that recovers the vector $\mathbf{x}$ by iteratively selecting the column of $\mathbf{A}$ that is most "correlated" with the difference between $\mathbf{z}$ and its current best (in ℓ_2-norm sense) approximation.

The questions that arise naturally are: Under which conditions does (P0) have a unique solution and when do BP and/or OMP deliver this solution? To formulate the answer to these questions, define $n_x = \|\mathbf{x}\|_0$ and the coherence of the dictionary $\mathbf{A}$ as

$$\mu_a = \max_{k,\ell,k\neq\ell} \left|\mathbf{a}_k^H \mathbf{a}_\ell\right|. \tag{3}$$

As shown in [11]–[13], a sufficient condition for $\mathbf{x}$ to be the unique solution of (P0) applied to $\mathbf{z} = \mathbf{A}\mathbf{x}$ and for BP and OMP to deliver this solution is

$$n_x < \frac{1}{2}\left(1 + \mu_a^{-1}\right). \tag{4}$$

B. Recovery in the presence of unstructured noise

Coherence-based recovery guarantees in the presence of unstructured (and deterministic) noise, i.e., for $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{n}$, with no constraints imposed on $\mathbf{n}$ apart from $\|\mathbf{n}\|_2 < \infty$, were derived in [16]–[20] and the references therein. Specifically, it was shown in [16] that a suitably modified version of BP, referred to as BP denoising (BPDN), recovers an estimate $\hat{\mathbf{x}}$ satisfying $\|\mathbf{x} - \hat{\mathbf{x}}\|_2 < C\|\mathbf{n}\|_2$ provided that (4) is met. Here, $C > 0$ depends on the coherence μ_a and on the sparsity level n_x of $\mathbf{x}$. Note that the support set of the estimate $\hat{\mathbf{x}}$ may differ from that of $\mathbf{x}$. Another result, reported in [17], states that OMP delivers the correct support set (but does not perfectly recover the nonzero entries of $\mathbf{x}$) provided that

$$n_x < \frac{1}{2}\left(1 + \mu_a^{-1}\right) - \frac{\|\mathbf{n}\|_2}{\mu_a |x_{\min}|} \tag{5}$$

where $|x_{\min}|$ denotes the absolute value of the component of $\mathbf{x}$ with smallest nonzero magnitude. The recovery condition (5) yields sensible results only if $\|\mathbf{n}\|_2/|x_{\min}|$ is small. Results similar to those reported in [17] were obtained in [18], [19]. Recovery guarantees in the case of stochastic noise $\mathbf{n}$ can be found in [19], [20]. We finally point out that *perfect* recovery of $\mathbf{x}$ is, in general, impossible in the presence of *unstructured* noise. In contrast, as we shall see below, *perfect* recovery is possible under structured noise according to (1).

C. Recovery guarantees in the presence of structured noise

As outlined in the introduction, many practically relevant signal recovery problems can be formulated as (sparse) signal recovery from sparsely corrupted measurements, a problem that seems to have received comparatively little attention in the literature so far and does not appear to have been developed systematically.

A straightforward way leading to recovery guarantees in the presence of structured noise, as in (1), follows from rewriting (1) as

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e} = \mathbf{D}\mathbf{w} \tag{6}$$

with the concatenated dictionary $\mathbf{D} = [\,\mathbf{A}\ \mathbf{B}\,]$ and the stacked vector $\mathbf{w} = [\,\mathbf{x}^T\ \mathbf{e}^T\,]^T$. This formulation allows us to invoke the recovery guarantee in (4) for the concatenated dictionary $\mathbf{D}$, which delivers a sufficient condition for $\mathbf{w}$ (and hence, $\mathbf{x}$ and $\mathbf{e}$) to be the unique solution of (P0) applied to $\mathbf{z} = \mathbf{D}\mathbf{w}$ and for BP and OMP to deliver this solution [11], [12]. However, the so obtained recovery condition

$$n_w = n_x + n_e < \frac{1}{2}\left(1 + \mu_d^{-1}\right) \tag{7}$$

with the dictionary coherence μ_d defined as

$$\mu_d = \max_{k,\ell,k\neq\ell} \left|\mathbf{d}_k^H \mathbf{d}_\ell\right| \tag{8}$$

ignores the structure of the recovery problem at hand, i.e., is agnostic to i) the fact that $\mathbf{D}$ consists of the dictionaries $\mathbf{A}$ and $\mathbf{B}$ with known coherence parameters μ_a and μ_b, respectively, and ii) knowledge about the support sets of $\mathbf{x}$ and/or $\mathbf{e}$ that may be available prior to recovery. As shown in Section IV, exploiting these two structural aspects of the recovery problem yields superior (i.e., less restrictive) recovery thresholds. Note that condition (7) guarantees perfect recovery of $\mathbf{x}$ (and $\mathbf{e}$) independent of the ℓ_2-norm of the noise vector, i.e., $\|\mathbf{B}\mathbf{e}\|_2$ may be arbitrarily large. This is in stark contrast to the recovery guarantees for noisy measurements in [16] and (5) (originally reported in [17]).

Special cases of the general setup (1), explicitly taking into account certain structural aspects of the recovery problem were considered in [3], [14], [26]–[30]. Specifically, in [26] it was shown that for $\mathbf{A} = \mathbf{F}_M$, $\mathbf{B} = \mathbf{I}_M$, and knowledge of the support set of $\mathbf{e}$, *perfect* recovery of the M-dimensional vector $\mathbf{x}$ is possible if

$$2n_x n_e < M \tag{9}$$

where $n_e = \|\mathbf{e}\|_0$. In [27], [28], recovery guarantees based on the RIC of the matrix $\mathbf{A}$ for the case where $\mathbf{B}$ is an orthonormal basis (ONB), and where the support set of $\mathbf{e}$ is either known or unknown, were reported; these recovery guarantees are particularly handy when $\mathbf{A}$ is, for example, i.i.d. Gaussian [31], [32]. However, results for the case of $\mathbf{A}$ and $\mathbf{B}$ both *general* (and deterministic) dictionaries taking into account prior knowledge about the support sets of $\mathbf{x}$ and

$$|\mathcal{P}|\,|\mathcal{Q}| \geq \frac{\left[(1+\mu_a)(1-\epsilon_{\mathcal{P}}) - |\mathcal{P}|\,\mu_a\right]^+ \left[(1+\mu_b)(1-\epsilon_{\mathcal{Q}}) - |\mathcal{Q}|\,\mu_b\right]^+}{\mu_m^2}. \tag{10}$$

$\mathbf{e}$ seem to be missing in the literature. Recovery guarantees for $\mathbf{A}$ i.i.d. non-zero mean Gaussian, $\mathbf{B} = \mathbf{I}_M$, and the support sets of $\mathbf{x}$ and $\mathbf{e}$ unknown were reported in [29]. In [30] recovery guarantees under a probabilistic model on both $\mathbf{x}$ and $\mathbf{e}$ and for unitary $\mathbf{A}$ and $\mathbf{B} = \mathbf{I}_M$ were reported showing that $\mathbf{x}$ can be recovered perfectly with high probability (and independently of the ℓ_2-norm of $\mathbf{x}$ and $\mathbf{e}$). The problem of sparse-signal recovery in the presence of impulse noise (i.e., $\mathbf{B} = \mathbf{I}_M$) was considered in [3], where a particular nonlinear measurement process combined with a non-convex program for signal recovery was proposed. In [14], signal recovery in the presence of impulse noise based on ℓ_1-norm minimization was investigated. The setup in [14], however, differs considerably from the one considered in this paper as $\mathbf{A}$ in [14] needs to be tall (i.e., $M > N_a$) and the vector $\mathbf{x}$ to be recovered is not necessarily sparse.

We conclude this literature overview by noting that the present paper is inspired by [26]. Specifically, we note that the recovery guarantee (9) reported in [26] is obtained from an uncertainty relation that puts limits on how sparse a given signal can simultaneously be in the Fourier basis and in the identity basis. Inspired by this observation, we start our discussion by presenting an uncertainty relation for pairs of general dictionaries, which forms the basis for the recovery guarantees reported later in this paper.

III. A General Uncertainty Relation for ϵ-Concentrated Vectors

We next present a novel uncertainty relation, which extends the uncertainty relation in [33, Lem. 1] for pairs of general dictionaries to vectors that are *ϵ-concentrated* rather than perfectly sparse. As shown in Section IV, this extension constitutes the basis for the derivation of recovery guarantees for BP.

A. The uncertainty relation

Define the mutual coherence between the dictionaries $\mathbf{A}$ and $\mathbf{B}$ as

$$\mu_m = \max_{k,\ell} \left|\mathbf{a}_k^H \mathbf{b}_\ell\right|.$$

Furthermore, we will need the following definition, which appeared previously in [26].

Definition 1: A vector $\mathbf{r} \in \mathbb{C}^{N_r}$ is said to be *$\epsilon_{\mathcal{R}}$-concentrated* to the set $\mathcal{R} \subseteq \{1,\ldots,N_r\}$ if $\|\mathbf{P}_{\mathcal{R}}\mathbf{r}\|_1 \geq (1-\epsilon_{\mathcal{R}})\|\mathbf{r}\|_1$, where $\epsilon_{\mathcal{R}} \in [0,1]$. We say that the vector $\mathbf{r}$ is *perfectly concentrated* to the set $\mathcal{R}$ and, hence, $|\mathcal{R}|$-sparse if $\mathbf{P}_{\mathcal{R}}\mathbf{r} = \mathbf{r}$, i.e., if $\epsilon_{\mathcal{R}} = 0$.

We can now state the following uncertainty relation for pairs of general dictionaries and for ϵ-concentrated vectors.

Theorem 1: Let $\mathbf{A} \in \mathbb{C}^{M\times N_a}$ be a dictionary with coherence μ_a, $\mathbf{B} \in \mathbb{C}^{M\times N_b}$ a dictionary with coherence μ_b, and denote the mutual coherence between $\mathbf{A}$ and $\mathbf{B}$ by μ_m. Let $\mathbf{s}$ be a vector in $\mathbb{C}^M$ that can be represented as a linear combination of columns of $\mathbf{A}$ and, similarly, as a linear combination of columns of $\mathbf{B}$. Concretely, there exists a pair of vectors $\mathbf{p} \in \mathbb{C}^{N_a}$ and $\mathbf{q} \in \mathbb{C}^{N_b}$ such that $\mathbf{s} = \mathbf{A}\mathbf{p} = \mathbf{B}\mathbf{q}$ (we exclude the trivial case where $\mathbf{p} = \mathbf{0}_{N_a}$ and $\mathbf{q} = \mathbf{0}_{N_b}$).[1] If $\mathbf{p}$ is $\epsilon_{\mathcal{P}}$-concentrated to $\mathcal{P}$ and $\mathbf{q}$ is $\epsilon_{\mathcal{Q}}$-concentrated to $\mathcal{Q}$, then (10) holds.

Proof: The proof follows closely that of [33, Lem. 1], which applies to perfectly concentrated vectors $\mathbf{p}$ and $\mathbf{q}$. We therefore only summarize the modifications to the proof of [33, Lem. 1]. Instead of using $\sum_{p\in\mathcal{P}} |[\mathbf{p}]_p| = \|\mathbf{p}\|_1$ to arrive at [33, Eq. 29]

$$\left[(1+\mu_a) - |\mathcal{P}|\,\mu_a\right]^+ \|\mathbf{p}\|_1 \leq |\mathcal{P}|\,\mu_m \|\mathbf{q}\|_1$$

we invoke $\sum_{p\in\mathcal{P}} |[\mathbf{p}]_p| \geq (1-\epsilon_{\mathcal{P}})\|\mathbf{p}\|_1$ to arrive at the following inequality valid for $\epsilon_{\mathcal{P}}$-concentrated vectors $\mathbf{p}$:

$$\left[(1+\mu_a)(1-\epsilon_{\mathcal{P}}) - |\mathcal{P}|\,\mu_a\right]^+ \|\mathbf{p}\|_1 \leq |\mathcal{P}|\,\mu_m \|\mathbf{q}\|_1. \tag{11}$$

Similarly, $\epsilon_{\mathcal{Q}}$-concentration, i.e., $\sum_{\mathcal{Q}} |[\mathbf{q}]_q| \geq (1-\epsilon_{\mathcal{Q}})\|\mathbf{q}\|_1$, is used to replace [33, Eq. 30] by

$$\left[(1+\mu_b)(1-\epsilon_{\mathcal{Q}}) - |\mathcal{Q}|\,\mu_b\right]^+ \|\mathbf{q}\|_1 \leq |\mathcal{Q}|\,\mu_m \|\mathbf{p}\|_1. \tag{12}$$

The uncertainty relation (10) is then obtained by multiplying (11) and (12) and dividing the resulting inequality by $\|\mathbf{p}\|_1\,\|\mathbf{q}\|_1$. ■

In the case where both $\mathbf{p}$ and $\mathbf{q}$ are *perfectly concentrated*, i.e., $\epsilon_{\mathcal{P}} = \epsilon_{\mathcal{Q}} = 0$, Theorem 1 reduces to the uncertainty relation reported in [33, Lem. 1], which we restate next for the sake of completeness.

Corollary 2 ([33, Lem. 1]): If $\mathcal{P} = \mathrm{supp}(\mathbf{p})$ and $\mathcal{Q} = \mathrm{supp}(\mathbf{q})$, the following holds:

$$|\mathcal{P}|\,|\mathcal{Q}| \geq \frac{\left[1-\mu_a(|\mathcal{P}|-1)\right]^+ \left[1-\mu_b(|\mathcal{Q}|-1)\right]^+}{\mu_m^2}. \tag{13}$$

As detailed in [33], [34], the uncertainty relation in Corollary 2 generalizes the uncertainty relation for two orthonormal bases (ONBs) found in [23]. Furthermore, it extends the uncertainty relations provided in [35] for pairs of square dictionaries (having the same number of rows and columns) to pairs of general dictionaries $\mathbf{A}$ and $\mathbf{B}$.

B. Tightness of the uncertainty relation

In certain special cases it is possible to find signals that satisfy the uncertainty relation (10) with equality. As in [26], consider $\mathbf{A} = \mathbf{F}_M$ and $\mathbf{B} = \mathbf{I}_M$, so that $\mu_m = 1/\sqrt{M}$, and define the comb signal containing equidistant spikes of unit height as

$$[\boldsymbol{\delta}_t]_\ell = \begin{cases} 1, & \text{if } (\ell-1) \bmod t = 0 \\ 0, & \text{otherwise} \end{cases}$$

[1] The uncertainty relation continues to hold if either $\mathbf{p} = \mathbf{0}_{N_a}$ or $\mathbf{q} = \mathbf{0}_{N_b}$, but does not apply to the trivial case $\mathbf{p} = \mathbf{0}_{N_a}$ *and* $\mathbf{q} = \mathbf{0}_{N_b}$. In all three cases we have $\mathbf{s} = \mathbf{0}_M$.

where we shall assume that t divides M. It can be shown that the vectors $\mathbf{p} = \boldsymbol{\delta}_{\sqrt{M}}$ and $\mathbf{q} = \boldsymbol{\delta}_{\sqrt{M}}$, both having $\sqrt{M}$ nonzero entries, satisfy $\mathbf{F}_M\mathbf{p} = \mathbf{I}_M\mathbf{q}$. If $\mathcal{P} = \mathrm{supp}(\mathbf{p})$ and $\mathcal{Q} = \mathrm{supp}(\mathbf{q})$, the vectors $\mathbf{p}$ and $\mathbf{q}$ are perfectly concentrated to $\mathcal{P}$ and $\mathcal{Q}$, respectively, i.e., $\epsilon_\mathcal{P} = \epsilon_\mathcal{Q} = 0$. Since $|\mathcal{P}| = |\mathcal{Q}| = \sqrt{M}$ and $\mu_m = 1/\sqrt{M}$ it follows that $|\mathcal{P}|\,|\mathcal{Q}| = 1/\mu_m^2 = M$ and, hence, $\mathbf{p} = \mathbf{q} = \boldsymbol{\delta}_{\sqrt{M}}$ satisfies (10) with equality.

We will next show that for pairs of general dictionaries $\mathbf{A}$ and $\mathbf{B}$, finding signals that satisfy the uncertainty relation (10) with equality is NP-hard. For the sake of simplicity, we restrict ourselves to the case $\mathcal{P} = \mathrm{supp}(\mathbf{p})$ and $\mathcal{Q} = \mathrm{supp}(\mathbf{q})$, which implies $|\mathcal{P}| = \|\mathbf{p}\|_0$ and $|\mathcal{Q}| = \|\mathbf{q}\|_0$. Next, consider the problem

$$\text{(U0)} \quad \begin{cases} \text{minimize} & \|\mathbf{p}\|_0\,\|\mathbf{q}\|_0 \\ \text{subject to} & \mathbf{A}\mathbf{p} = \mathbf{B}\mathbf{q},\ \|\mathbf{p}\|_0 \geq 1, \|\mathbf{q}\|_0 \geq 1. \end{cases}$$

Since we are interested in the minimum of $\|\mathbf{p}\|_0\,\|\mathbf{q}\|_0$ for nonzero vectors $\mathbf{p}$ and $\mathbf{q}$, we imposed the constraints $\|\mathbf{p}\|_0 \geq 1$ and $\|\mathbf{q}\|_0 \geq 1$ to exclude the case where $\mathbf{p} = \mathbf{0}_{N_a}$ and/or $\mathbf{q} = \mathbf{0}_{N_b}$. Now, it follows that for the particular choice $\mathbf{B} = \mathbf{z} \in \mathbb{C}^M$ and hence $\mathbf{q} = q \in \mathbb{C}\setminus\{0\}$ (note that we exclude the case $q = 0$ as a consequence of the requirement $\|\mathbf{q}\|_0 \geq 1$) the problem (U0) reduces to

$$\text{(U0}^*) \quad \text{minimize } \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{z}$$

where $\mathbf{x} = \mathbf{p}/q$. However, as (U0*) is equivalent to (P0), which is NP-hard [36], in general, we can conclude that finding a pair $\mathbf{p}$ and $\mathbf{q}$ satisfying the uncertainty relation (10) with equality is NP-hard.

IV. Recovery of Sparsely Corrupted Signals

Based on the uncertainty relation in Theorem 1, we next derive conditions that guarantee perfect recovery of $\mathbf{x}$ (and of $\mathbf{e}$, if appropriate) from the (sparsely corrupted) measurement $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$. These conditions will be seen to depend on the number of nonzero entries of $\mathbf{x}$ and $\mathbf{e}$, and on the coherence parameters μ_a, μ_b, and μ_m. Moreover, in contrast to (5), the recovery conditions we find will not depend on the ℓ_2-norm of the noise vector $\|\mathbf{B}\mathbf{e}\|_2$, which is hence allowed to be arbitrarily large. We consider the following cases: I) The support sets of both $\mathbf{x}$ and $\mathbf{e}$ are known (prior to recovery), II) the support set of only $\mathbf{x}$ or only $\mathbf{e}$ is known, III) the number of nonzero entries of only $\mathbf{x}$ or only $\mathbf{e}$ is known, and IV) nothing is known about $\mathbf{x}$ and $\mathbf{e}$. The uncertainty relation in Theorem 1 is the basis for the recovery guarantees in all four cases considered. To simplify notation, motivated by the form of the right-hand side (RHS) of (13), we define the function

$$f(u,v) = \frac{[1-\mu_a(u-1)]^+\,[1-\mu_b(v-1)]^+}{\mu_m^2}.$$

In the remainder of the paper, $\mathcal{X}$ denotes $\mathrm{supp}(\mathbf{x})$ and $\mathcal{E}$ stands for $\mathrm{supp}(\mathbf{e})$. We furthermore assume that the dictionaries $\mathbf{A}$ and $\mathbf{B}$ are known perfectly to the recovery algorithms. Moreover, we assume that[2] $\mu_m > 0$.

[2] If $\mu_m = 0$, the space spanned by the columns of $\mathbf{A}$ is orthogonal to the space spanned by the columns of $\mathbf{B}$. This makes the separation of the components $\mathbf{A}\mathbf{x}$ and $\mathbf{B}\mathbf{e}$ given $\mathbf{z}$ straightforward. Once this separation is accomplished, $\mathbf{x}$ can be recovered from $\mathbf{A}\mathbf{x}$ using (P0), BP, or OMP, if (4) is satisfied.

A. Case I: Knowledge of $\mathcal{X}$ and $\mathcal{E}$

We start with the case where both $\mathcal{X}$ and $\mathcal{E}$ are known prior to recovery. The values of the nonzero entries of $\mathbf{x}$ and $\mathbf{e}$ are unknown. This scenario is relevant, for example, in applications requiring recovery of clipped band-limited signals with known spectral support $\mathcal{X}$. Here, we would have $\mathbf{A} = \mathbf{F}_M$, $\mathbf{B} = \mathbf{I}_M$, and $\mathcal{E}$ can be determined as follows: Compare the measurements $[\mathbf{z}]_i$, $i = 1,\ldots,M$, to the clipping threshold a; if $|[\mathbf{z}]_i| = a$ add the corresponding index i to $\mathcal{E}$.

Recovery of $\mathbf{x}$ from $\mathbf{z}$ is then performed as follows. We first rewrite the input-output relation in (1) as

$$\mathbf{z} = \mathbf{A}_\mathcal{X}\mathbf{x}_\mathcal{X} + \mathbf{B}_\mathcal{E}\mathbf{e}_\mathcal{E} = \mathbf{D}_{\mathcal{X},\mathcal{E}}\mathbf{s}_{\mathcal{X},\mathcal{E}}$$

with the concatenated dictionary $\mathbf{D}_{\mathcal{X},\mathcal{E}} = [\,\mathbf{A}_\mathcal{X}\ \mathbf{B}_\mathcal{E}\,]$ and the stacked vector $\mathbf{s}_{\mathcal{X},\mathcal{E}} = \left[\,\mathbf{x}_\mathcal{X}^T\ \mathbf{e}_\mathcal{E}^T\,\right]^T$. Since $\mathcal{X}$ and $\mathcal{E}$ are known, we can recover the stacked vector $\mathbf{s}_{\mathcal{X},\mathcal{E}} = \left[\,\mathbf{x}_\mathcal{X}^T\ \mathbf{e}_\mathcal{E}^T\,\right]^T$, perfectly and, hence, the nonzero entries of both $\mathbf{x}$ and $\mathbf{e}$, if the pseudo-inverse $\mathbf{D}_{\mathcal{X},\mathcal{E}}^\dagger$ exists. In this case, we can obtain $\mathbf{s}_{\mathcal{X},\mathcal{E}}$, as

$$\mathbf{s}_{\mathcal{X},\mathcal{E}} = \mathbf{D}_{\mathcal{X},\mathcal{E}}^\dagger\mathbf{z}. \tag{14}$$

The following theorem states a sufficient condition for $\mathbf{D}_{\mathcal{X},\mathcal{E}}$ to have full (column) rank, which implies existence of the pseudo-inverse $\mathbf{D}_{\mathcal{X},\mathcal{E}}^\dagger$. This condition depends on the coherence parameters μ_a, μ_b, and μ_m, of the involved dictionaries $\mathbf{A}$ and $\mathbf{B}$ and on $\mathcal{X}$ and $\mathcal{E}$ through the cardinalities $|\mathcal{X}|$ and $|\mathcal{E}|$, i.e., the *number of nonzero entries* in $\mathbf{x}$ and $\mathbf{e}$, respectively.

Theorem 3: Let $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$ with $\mathcal{X} = \mathrm{supp}(\mathbf{x})$ and $\mathcal{E} = \mathrm{supp}(\mathbf{e})$. Define $n_x = \|\mathbf{x}\|_0$ and $n_e = \|\mathbf{e}\|_0$. If

$$n_x n_e < f(n_x, n_e), \tag{15}$$

then the concatenated dictionary $\mathbf{D}_{\mathcal{X},\mathcal{E}} = [\,\mathbf{A}_\mathcal{X}\ \mathbf{B}_\mathcal{E}\,]$ has full (column) rank.

Proof: See Appendix A. ■

For the special case $\mathbf{A} = \mathbf{F}_M$ and $\mathbf{B} = \mathbf{I}_M$ (so that $\mu_a = \mu_b = 0$ and $\mu_m = 1/\sqrt{M}$) the recovery condition (15) reduces to $n_x n_e < M$, a result obtained previously in [26]. Tightness of (15) can be established by noting that the pairs $\mathbf{x} = \lambda\boldsymbol{\delta}_{\sqrt{M}}$, $\mathbf{e} = (1-\lambda)\boldsymbol{\delta}_{\sqrt{M}}$ with $\lambda \in (0,1)$ and $\mathbf{x}' = \lambda'\boldsymbol{\delta}_{\sqrt{M}}$, $\mathbf{e}' = (1-\lambda')\boldsymbol{\delta}_{\sqrt{M}}$ with $\lambda' \neq \lambda$ and $\lambda' \in (0,1)$ both satisfy (15) with equality and lead to the same measurement outcome $\mathbf{z} = \mathbf{F}_M\mathbf{x} + \mathbf{e} = \mathbf{F}_M\mathbf{x}' + \mathbf{e}'$ [34].

It is interesting to observe that Theorem 3 yields a sufficient condition on n_x and n_e for any $(M-n_e)\times n_x$-submatrix of $\mathbf{A}$ to have full (column) rank. To see this, consider the special case $\mathbf{B} = \mathbf{I}_M$ and hence, $\mathbf{D}_{\mathcal{X},\mathcal{E}} = [\,\mathbf{A}_\mathcal{X}\ \mathbf{I}_\mathcal{E}\,]$. Condition (15) characterizes pairs (n_x, n_e), for which all matrices $\mathbf{D}_{\mathcal{X},\mathcal{E}}$ with $n_x = |\mathcal{X}|$ and $n_e = |\mathcal{E}|$ are guaranteed to have full (column) rank. Hence, the sub-matrix consisting of all rows of $\mathbf{A}_\mathcal{X}$ with row index in $\mathcal{E}^c$ must have full (column) rank as well. Since the result holds for all support sets $\mathcal{X}$ and $\mathcal{E}$ with $|\mathcal{X}| = n_x$ and $|\mathcal{E}| = n_e$, all possible $(M-n_e)\times n_x$-submatrices of $\mathbf{A}$ must have full (column) rank.

B. Case II: Only $\mathcal{X}$ or only $\mathcal{E}$ is known

Next, we find recovery guarantees for the case where either only $\mathcal{X}$ or only $\mathcal{E}$ is known prior to recovery.

1) Recovery when $\mathcal{E}$ is known and $\mathcal{X}$ is unknown: A prominent application for this setup is the recovery of clipped band-limited signals [27], [37], where the signal's spectral support, i.e., $\mathcal{X}$, is unknown. The support set $\mathcal{E}$ can be identified as detailed previously in Section IV-A. Further application examples for this setup include inpainting and super-resolution [5]–[7] of signals that admit a sparse representation in $\mathbf{A}$ (but with unknown support set $\mathcal{X}$). The locations of the missing elements in $\mathbf{y} = \mathbf{A}\mathbf{x}$ are known (and correspond, e.g., to missing paint elements in frescos), i.e., the set $\mathcal{E}$ can be determined prior to recovery. Inpainting and super-resolution then amount to reconstructing the vector $\mathbf{x}$ from the sparsely corrupted measurement $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{e}$ and computing $\mathbf{y} = \mathbf{A}\mathbf{x}$.

The setting of $\mathcal{E}$ known and $\mathcal{X}$ unknown was considered previously in [26] for the special case $\mathbf{A} = \mathbf{F}_M$ and $\mathbf{B} = \mathbf{I}_M$. The recovery condition (18) in Theorem 4 below extends the result in [26, Thms. 5 and 9] to pairs of general dictionaries $\mathbf{A}$ and $\mathbf{B}$.

Theorem 4: Let $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$ where $\mathcal{E} = \operatorname{supp}(\mathbf{e})$ is known. Consider the problem

$$(\text{P0}, \mathcal{E}) \quad \begin{cases} \text{minimize} & \|\tilde{\mathbf{x}}\|_0 \\ \text{subject to} & \mathbf{A}\tilde{\mathbf{x}} \in (\{\mathbf{z}\} + \mathcal{R}(\mathbf{B}_\mathcal{E})) \end{cases} \tag{16}$$

and the convex program

$$(\text{BP}, \mathcal{E}) \quad \begin{cases} \text{minimize} & \|\tilde{\mathbf{x}}\|_1 \\ \text{subject to} & \mathbf{A}\tilde{\mathbf{x}} \in (\{\mathbf{z}\} + \mathcal{R}(\mathbf{B}_\mathcal{E})) \,. \end{cases} \tag{17}$$

If $n_x = \|\mathbf{x}\|_0$ and $n_e = \|\mathbf{e}\|_0$ satisfy

$$2 n_x n_e < f(2n_x, n_e), \tag{18}$$

then the unique solution of $(\text{P0}, \mathcal{E})$ applied to $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$ is given by $\mathbf{x}$ and $(\text{BP}, \mathcal{E})$ will deliver this solution.

Proof: See Appendix B. ■

Solving $(\text{P0}, \mathcal{E})$ requires a combinatorial search, which results in prohibitive computational complexity even for moderate problem sizes. The convex relaxation $(\text{BP}, \mathcal{E})$ can, however, be solved more efficiently. Note that the constraint $\mathbf{A}\tilde{\mathbf{x}} \in (\{\mathbf{z}\} + \mathcal{R}(\mathbf{B}_\mathcal{E}))$ reflects the fact that any error component $\mathbf{B}_\mathcal{E}\mathbf{e}_\mathcal{E}$ yields consistency on account of $\mathcal{E}$ known (by assumption). For $n_e = 0$ (i.e., the noiseless case) the recovery threshold (18) reduces to $n_x < (1 + 1/\mu_a)/2$, which is the well-known recovery threshold (4) guaranteeing recovery of the sparse vector $\mathbf{x}$ through (P0) and BP applied to $\mathbf{z} = \mathbf{A}\mathbf{x}$. We finally note that RIC-based guarantees for recovering $\mathbf{x}$ from $\mathbf{z} = \mathbf{A}\mathbf{x}$ (i.e., recovery in the absence of (sparse) corruptions) that take into account partial knowledge of the signal support set $\mathcal{X}$ were developed in [38], [39].

Tightness of (18) can be established by setting $\mathbf{A} = \mathbf{F}_M$ and $\mathbf{B} = \mathbf{I}_M$. Specifically, the pairs $\mathbf{x} = \boldsymbol{\delta}_{2\sqrt{M}} - \boldsymbol{\delta}_{\sqrt{M}}$, $\mathbf{e} = \boldsymbol{\delta}_{\sqrt{M}}$ and $\mathbf{x}' = \boldsymbol{\delta}_{2\sqrt{M}}$, $\mathbf{e}' = \mathbf{e}$ both satisfy (18) with equality. One can furthermore verify that $\mathbf{x}$ and $\mathbf{x}'$ are both in the admissible set specified by the constraints in $(\text{P0}, \mathcal{E})$ and $(\text{BP}, \mathcal{E})$ and $\|\mathbf{x}'\|_0 = \|\mathbf{x}\|_0$, $\|\mathbf{x}'\|_1 = \|\mathbf{x}\|_1$. Hence, $(\text{P0}, \mathcal{E})$ and $(\text{BP}, \mathcal{E})$ both cannot distinguish between $\mathbf{x}$ and $\mathbf{x}'$ based on the measurement outcome $\mathbf{z}$. For a detailed discussion of this example we refer to [34].

Rather than solving $(\text{P0}, \mathcal{E})$ or $(\text{BP}, \mathcal{E})$, we may attempt to recover the vector $\mathbf{x}$ by exploiting more directly the fact that $\mathcal{R}(\mathbf{B}_\mathcal{E})$ is known (since $\mathbf{B}$ and $\mathcal{E}$ are assumed to be known) and projecting the measurement outcome $\mathbf{z}$ onto the orthogonal complement of $\mathcal{R}(\mathbf{B}_\mathcal{E})$. This approach would eliminate the (sparse) noise component and leave us with a standard sparse-signal recovery problem for the vector $\mathbf{x}$. We next show that this ansatz is guaranteed to recover the sparse vector $\mathbf{x}$ provided that condition (18) is satisfied. Let us detail the procedure. If the columns of $\mathbf{B}_\mathcal{E}$ are linearly independent, the pseudo-inverse $\mathbf{B}_\mathcal{E}^\dagger$ exists, and the projector onto the orthogonal complement of $\mathcal{R}(\mathbf{B}_\mathcal{E})$ is given by

$$\mathbf{R}_\mathcal{E} = \mathbf{I}_M - \mathbf{B}_\mathcal{E}\mathbf{B}_\mathcal{E}^\dagger. \tag{19}$$

Applying $\mathbf{R}_\mathcal{E}$ to the measurement outcome $\mathbf{z}$ yields

$$\mathbf{R}_\mathcal{E}\mathbf{z} = \mathbf{R}_\mathcal{E}(\mathbf{A}\mathbf{x} + \mathbf{B}_\mathcal{E}\mathbf{e}_\mathcal{E}) = \mathbf{R}_\mathcal{E}\mathbf{A}\mathbf{x} \triangleq \hat{\mathbf{z}} \tag{20}$$

where we used the fact that $\mathbf{R}_\mathcal{E}\mathbf{B}_\mathcal{E} = \mathbf{0}_{M,n_e}$. We are now left with the standard problem of recovering $\mathbf{x}$ from the modified measurement outcome $\hat{\mathbf{z}} = \mathbf{R}_\mathcal{E}\mathbf{A}\mathbf{x}$. What comes to mind first is that computing the standard recovery threshold (4) for the modified dictionary $\mathbf{R}_\mathcal{E}\mathbf{A}$ should provide us with a recovery threshold for the problem of extracting $\mathbf{x}$ from $\hat{\mathbf{z}} = \mathbf{R}_\mathcal{E}\mathbf{A}\mathbf{x}$. It turns out, however, that the columns of $\mathbf{R}_\mathcal{E}\mathbf{A}$ will, in general, not have unit ℓ_2-norm, an assumption underlying (4). What comes to our rescue is that under condition (18) we have (as shown in Theorem 5 below) $\|\mathbf{R}_\mathcal{E}\mathbf{a}_\ell\|_2 > 0$ for $\ell = 1, \ldots, N_a$. We can, therefore, normalize the modified dictionary $\mathbf{R}_\mathcal{E}\mathbf{A}$ by rewriting (20) as

$$\hat{\mathbf{z}} = \mathbf{R}_\mathcal{E}\mathbf{A}\boldsymbol{\Delta}\hat{\mathbf{x}} \tag{21}$$

where $\boldsymbol{\Delta}$ is the diagonal matrix with elements

$$[\boldsymbol{\Delta}]_{\ell,\ell} = \frac{1}{\|\mathbf{R}_\mathcal{E}\mathbf{a}_\ell\|_2}, \quad \ell = 1, \ldots, N_a,$$

and $\hat{\mathbf{x}} \triangleq \boldsymbol{\Delta}^{-1}\mathbf{x}$. Now, $\mathbf{R}_\mathcal{E}\mathbf{A}\boldsymbol{\Delta}$ plays the role of the dictionary (with normalized columns) and $\hat{\mathbf{x}}$ is the unknown sparse vector that we wish to recover. Obviously, $\operatorname{supp}(\hat{\mathbf{x}}) = \operatorname{supp}(\mathbf{x})$ and $\mathbf{x}$ can be recovered from $\hat{\mathbf{x}}$ according to[3] $\mathbf{x} = \boldsymbol{\Delta}\hat{\mathbf{x}}$. The following theorem shows that (18) is sufficient to guarantee the following: i) The columns of $\mathbf{B}_\mathcal{E}$ are linearly independent, which guarantees the existence of $\mathbf{B}_\mathcal{E}^\dagger$, ii) $\|\mathbf{R}_\mathcal{E}\mathbf{a}_\ell\|_2 > 0$ for $\ell = 1, \ldots, N_a$, and iii) no vector $\mathbf{x}' \in \mathbb{C}^{N_a}$ with $\|\mathbf{x}'\|_0 \leq 2n_x$ lies in the kernel of $\mathbf{R}_\mathcal{E}\mathbf{A}$. Hence, (18) enables perfect recovery of $\mathbf{x}$ from (21).

Theorem 5: If (18) is satisfied, the unique solution of (P0) applied to $\hat{\mathbf{z}} = \mathbf{R}_\mathcal{E}\mathbf{A}\boldsymbol{\Delta}\hat{\mathbf{x}}$ is given by $\hat{\mathbf{x}}$. Furthermore, BP and OMP applied to $\hat{\mathbf{z}} = \mathbf{R}_\mathcal{E}\mathbf{A}\boldsymbol{\Delta}\hat{\mathbf{x}}$ are guaranteed to recover the unique (P0)-solution.

Proof: See Appendix C. ■

Since condition (18) ensures that $[\boldsymbol{\Delta}]_{\ell,\ell} > 0$, $\ell = 1, \ldots, N_a$, the vector $\mathbf{x}$ can be obtained from $\hat{\mathbf{x}}$ according to $\mathbf{x} = \boldsymbol{\Delta}\hat{\mathbf{x}}$. Furthermore, (18) guarantees the existence of $\mathbf{B}_\mathcal{E}^\dagger$ and hence the nonzero entries of $\mathbf{e}$ can be obtained from $\mathbf{x}$ as follows:

$$\mathbf{e}_\mathcal{E} = \mathbf{B}_\mathcal{E}^\dagger(\mathbf{z} - \mathbf{A}\mathbf{x}) \,.$$

[3]If $\|\mathbf{R}_\mathcal{E}\mathbf{a}_\ell\|_2 > 0$ for $\ell = 1, \ldots, N_a$, then the matrix $\boldsymbol{\Delta}$ corresponds to a one-to-one mapping.

Theorem 5 generalizes the results in [26, Thms. 5 and 9] obtained for the special case $\mathbf{A} = \mathbf{F}_M$ and $\mathbf{B} = \mathbf{I}_M$ to pairs of general dictionaries and additionally shows that OMP delivers the correct solution provided that (18) is satisfied.

It follows from (21) that other sparse-signal recovery algorithms, such as iterative thresholding-based algorithms [40], CoSaMP [41], or subspace pursuit [42] can be applied to recover $\mathbf{x}$.[4] Finally, we note that the idea of projecting the measurement outcome onto the orthogonal complement of the space spanned by the active columns of $\mathbf{B}$ and investigating the effect on the RICs, instead of the coherence parameter μ_a (as was done in Appendix C-C) was put forward in [27], [43] along with RIC-based recovery guarantees that apply to random matrices $\mathbf{A}$ and guarantee the recovery of $\mathbf{x}$ with high probability (with respect to $\mathbf{A}$ and irrespective of the locations of the sparse corruptions).

2) Recovery when $\mathcal{X}$ is known and $\mathcal{E}$ is unknown: A possible application scenario for this situation is the recovery of spectrally sparse signals with known spectral support that are impaired by impulse noise with unknown impulse locations.

It is evident that this setup is formally equivalent to that discussed in Section IV-B1, with the roles of $\mathbf{x}$ and $\mathbf{e}$ interchanged. In particular, we may apply the projection matrix $\mathbf{R}_\mathcal{X} = \mathbf{I}_M - \mathbf{A}_\mathcal{X}\mathbf{A}_\mathcal{X}^\dagger$ to the corrupted measurement outcome $\mathbf{z}$ to obtain the standard recovery problem $\hat{\mathbf{z}}' = \mathbf{R}_\mathcal{X}\mathbf{B}\boldsymbol{\Delta}'\hat{\mathbf{e}}$, where $\boldsymbol{\Delta}'$ is a diagonal matrix with diagonal elements $[\boldsymbol{\Delta}']_{\ell,\ell} = 1/\|\mathbf{R}_\mathcal{X}\mathbf{b}_\ell\|_2$. The corresponding unknown vector is given by $\hat{\mathbf{e}} \triangleq (\boldsymbol{\Delta}')^{-1}\,\mathbf{e}$. The following corollary is a direct consequence of Theorem 5.

Corollary 6: Let $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$ where $\mathcal{X} = \text{supp}(\mathbf{x})$ is known. If the number of nonzero entries in $\mathbf{x}$ and $\mathbf{e}$, i.e., $n_x = \|\mathbf{x}\|_0$ and $n_e = \|\mathbf{e}\|_0$, satisfy

$$2n_x n_e < f(n_x, 2n_e) \tag{22}$$

then the unique solution of (P0) applied to $\hat{\mathbf{z}}' = \mathbf{R}_\mathcal{X}\mathbf{B}\boldsymbol{\Delta}'\hat{\mathbf{e}}$ is given by $\hat{\mathbf{e}} = (\boldsymbol{\Delta}')^{-1}\,\mathbf{e}$. Furthermore, BP and OMP applied to $\hat{\mathbf{z}}' = \mathbf{R}_\mathcal{X}\mathbf{B}\boldsymbol{\Delta}'\hat{\mathbf{e}}$ recover the unique (P0)-solution.

Once we have $\hat{\mathbf{e}}$, the vector $\mathbf{e}$ can be obtained easily, since $\mathbf{e} = \boldsymbol{\Delta}'\hat{\mathbf{e}}$ and the nonzero entries of $\mathbf{x}$ are given by

$$\mathbf{x}_\mathcal{X} = \mathbf{A}_\mathcal{X}^\dagger(\mathbf{z} - \mathbf{B}\mathbf{e}).$$

Since (22) ensures that the columns of $\mathbf{A}_\mathcal{X}$ are linearly independent, the pseudo-inverse $\mathbf{A}_\mathcal{X}^\dagger$ is guaranteed to exist. Note that tightness of the recovery condition (22) can be established analogously to the case of $\mathcal{E}$ known and $\mathcal{X}$ unknown (discussed in Section IV-B1).

C. Case III: Cardinality of $\mathcal{E}$ or $\mathcal{X}$ known

We next consider the case where neither $\mathcal{X}$ nor $\mathcal{E}$ are known, but knowledge of either $\|\mathbf{x}\|_0$ or $\|\mathbf{e}\|_0$ is available (prior to recovery). An application scenario for $\|\mathbf{x}\|_0$ unknown and $\|\mathbf{e}\|_0$ known would be the recovery of a sparse pulse-stream with unknown pulse-locations from measurements that are corrupted by electric hum with unknown base-frequency but known number of harmonics (e.g., determined by the base frequency of the hum and the acquisition bandwidth of the system under consideration). We state our main result for the case $n_e = \|\mathbf{e}\|_0$ known and $n_x = \|\mathbf{x}\|_0$ unknown. The case where n_x is known and n_e is unknown can be treated similarly.

Theorem 7: Let $\mathbf{z} = \mathbf{A}\mathbf{x}+\mathbf{B}\mathbf{e}$, define $n_x = \|\mathbf{x}\|_0$ and $n_e = \|\mathbf{e}\|_0$, and assume that n_e is known. Consider the problem

$$(\text{P0}, n_e) \quad \begin{cases} \text{minimize} & \|\tilde{\mathbf{x}}\|_0 \\ \text{subject to} & \mathbf{A}\tilde{\mathbf{x}} \in \left(\{\mathbf{z}\} + \bigcup_{\mathcal{E}' \in \mathscr{P}} \mathcal{R}(\mathbf{B}_{\mathcal{E}'})\right) \end{cases} \tag{23}$$

where $\mathscr{P} = \wp_{n_e}(\{1,\ldots,N_b\})$ denotes the set of subsets of $\{1,\ldots,N_b\}$ of cardinality less than or equal to n_e. The unique solution of $(\text{P0}, n_e)$ applied to $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$ is given by $\mathbf{x}$ if

$$4n_x n_e < f(2n_x, 2n_e). \tag{24}$$

Proof: See Appendix D. ∎

We emphasize that the problem $(\text{P0}, n_e)$ exhibits prohibitive (concretely, combinatorial) computational complexity, in general. Unfortunately, replacing the ℓ_0-norm of $\tilde{\mathbf{x}}$ in the minimization in (23) by the ℓ_1-norm does not lead to a computationally tractable alternative either, as the constraint $\mathbf{A}\tilde{\mathbf{x}} \in (\{\mathbf{z}\}+\bigcup_{\mathcal{E}' \in \mathscr{P}} \mathcal{R}(\mathbf{B}_{\mathcal{E}'}))$ specifies a non-convex set, in general. Nevertheless, the recovery threshold in (24) is interesting as it completes the picture on the impact of knowledge about the support sets of $\mathbf{x}$ and $\mathbf{e}$ on the recovery thresholds. We refer to Section V-A for a detailed discussion of this matter. Note, though, that greedy recovery algorithms, such as OMP [13], [24], [25], CoSaMP [41], or subspace pursuit [42], can be modified to incorporate prior knowledge of the *individual* sparsity levels of $\mathbf{x}$ and/or $\mathbf{e}$. Analytical recovery guarantees corresponding to the resulting modified algorithms do not seem to be available.

We finally note that tightness of (24) can be established for $\mathbf{A} = \mathbf{F}_M$ and $\mathbf{B} = \mathbf{I}_M$. Specifically, consider the pair $\mathbf{x} = \boldsymbol{\delta}_{2\sqrt{M}}$, $\mathbf{e} = -\boldsymbol{\delta}_{2\sqrt{M}}$ and the alternative pair $\mathbf{x}' = \boldsymbol{\delta}_{2\sqrt{M}} - \boldsymbol{\delta}_{\sqrt{M}}$, $\mathbf{e}' = -\boldsymbol{\delta}_{2\sqrt{M}} + \boldsymbol{\delta}_{\sqrt{M}}$. It can be shown that both $\mathbf{x}$ and $\mathbf{x}'$ are in the admissible set of (P0, n_e) in (23), satisfy $\|\mathbf{x}'\|_0 = \|\mathbf{x}\|_0$, and lead to the same measurement outcome $\mathbf{z}$. Therefore, (P0, n_e) cannot distinguish between $\mathbf{x}$ and $\mathbf{x}'$ (we refer to [34] for details).

D. Case IV: No knowledge about the support sets

Finally, we consider the case of no knowledge (prior to recovery) about the support sets $\mathcal{X}$ and $\mathcal{E}$. A corresponding application scenario would be the restoration of an audio signal (whose spectrum is sparse with unknown support set) that is corrupted by impulse noise, e.g., click or pop noise occurring at unknown locations. Another typical application can be found in the realm of signal separation; e.g., the decomposition of images into two distinct features, i.e., into a part that exhibits a sparse representation in the dictionary $\mathbf{A}$ and another part that exhibits a sparse representation in $\mathbf{B}$. Decomposition of the image $\mathbf{z}$ then amounts to performing sparse-signal recovery based on $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$ with no knowledge about the support sets $\mathcal{X}$ and $\mathcal{E}$ available prior to recovery. The individual image features are given by $\mathbf{A}\mathbf{x}$ and $\mathbf{B}\mathbf{e}$.

[4]Finding analytical recovery guarantees for these algorithms remains an interesting open problem.

Recovery guarantees for this case follow from the results in [33]. Specifically, by rewriting (1) as $\mathbf{z} = \mathbf{D}\mathbf{w}$ as in (6), we can employ the recovery guarantees in [33], which are explicit in the coherence parameters μ_a and μ_b, and the dictionary coherence μ_d of $\mathbf{D}$. For the sake of completeness, we restate the following result from [33].

Theorem 8 ([33, Thm. 2]): Let $\mathbf{z} = \mathbf{D}\mathbf{w}$ with $\mathbf{w} = [\,\mathbf{x}^T\ \mathbf{e}^T\,]^T$ and $\mathbf{D} = [\,\mathbf{A}\ \mathbf{B}\,]$ with the coherence parameters $\mu_a \le \mu_b$ and the dictionary coherence μ_d as defined in (8). A sufficient condition for the vector $\mathbf{w}$ to be the unique solution of (P0) applied to $\mathbf{z} = \mathbf{D}\mathbf{w}$ is

$$n_x + n_e = n_w < \frac{f(\hat{x}) + \hat{x}}{2} \tag{25}$$

where

$$f(x) = \frac{(1+\mu_a)(1+\mu_b) - x\mu_b(1+\mu_a)}{x(\mu_d^2 - \mu_a\mu_b) + \mu_a(1+\mu_b)}$$

and $\hat{x} = \min\{x_b, x_s\}$. Furthermore, $x_b = (1+\mu_b)/(\mu_b + \mu_d^2)$ and

$$x_s = \begin{cases} 1/\mu_d, & \text{if } \mu_a = \mu_b = \mu_d, \\[1ex] \dfrac{\mu_d\sqrt{(1+\mu_a)(1+\mu_b)} - \mu_a - \mu_a\mu_b}{\mu_d^2 - \mu_a\mu_b}, & \text{otherwise.} \end{cases}$$

Obviously, once the vector $\mathbf{w}$ has been recovered, we can extract $\mathbf{x}$ and $\mathbf{e}$. The following theorem, originally stated in [33], guarantees that BP and OMP deliver the unique solution of (P0) applied to $\mathbf{z} = \mathbf{D}\mathbf{w}$ and the associated recovery threshold, as shown in [33], is only slightly more restrictive than that for (P0) in (25).

Theorem 9 ([33, Cor. 4]): A sufficient condition for BP and OMP to deliver the unique solution of (P0) applied to $\mathbf{z} = \mathbf{D}\mathbf{w}$ is given by

$$n_w < \begin{cases} \dfrac{\delta\big(\epsilon - (\mu_d + 3\mu_b)\big)}{2(\mu_d^2 - \mu_b^2)}, & \text{if } \mu_b < \mu_d \text{ and } \kappa(\mu_d, \mu_b) > 1, \\[2ex] \dfrac{1 + 2\mu_d^2 + 3\mu_b - \mu_d\delta}{2(\mu_d^2 + \mu_b)}, & \text{otherwise} \end{cases} \tag{26}$$

with $n_w = n_x + n_e$ and

$$\kappa(\mu_d, \mu_b) = \frac{\delta\sqrt{2\mu_d\,(\mu_b + 3\mu_d + \epsilon)} - 2\mu_d - 2\mu_b(\delta + \mu_d)}{2(\mu_d^2 - \mu_b^2)}$$

where $\delta = 1 + \mu_b$ and $\epsilon = 2\sqrt{2}\sqrt{\mu_d(\mu_b + \mu_d)}$.

We emphasize that both thresholds (25) and (26) are more restrictive than those in (15), (18), (22), and (24) (see also Section V-A), which is consistent with the intuition that additional knowledge about the support sets $\mathcal{X}$ and $\mathcal{E}$ should lead to higher recovery thresholds. Note that tightness of (25) and (26) was established before in [44] and [33], respectively.

V. Discussion of the Recovery Guarantees

The aim of this section is to provide an interpretation of the recovery guarantees found in Section IV. Specifically, we discuss the impact of support-set knowledge on the recovery thresholds we found, and we point out limitations of our results.

A. Factor of two in the recovery thresholds

Comparing the recovery thresholds (15), (18), (22), and (24) (Cases I–III), we observe that the price to be paid for not knowing the support set $\mathcal{X}$ or $\mathcal{E}$ is a reduction of the recovery threshold by a factor of two (note that in Case III, both $\mathcal{X}$ and $\mathcal{E}$ are unknown, but the cardinality of either $\mathcal{X}$ or $\mathcal{E}$ is known). For example, consider the recovery thresholds (15) and (18). For given $n_e \in [0, 1 + 1/\mu_b]$, solving (15) for n_x yields

$$n_x < \frac{(1+\mu_a)(1 - \mu_b(n_e - 1))}{n_e(\mu_m^2 - \mu_a\mu_b) + \mu_a(1+\mu_b)}.$$

Similarly, still assuming $n_e \in [0, 1 + 1/\mu_b]$ and solving (18) for n_x, we get

$$n_x < \frac{1}{2}\,\frac{(1+\mu_a)(1 - \mu_b(n_e - 1))}{n_e(\mu_m^2 - \mu_a\mu_b) + \mu_a(1+\mu_b)}.$$

Hence, knowledge of $\mathcal{X}$ prior to recovery allows for the recovery of a signal with twice as many nonzero entries in $\mathbf{x}$ compared to the case where $\mathcal{X}$ is not known. This factor-of-two penalty has the same roots as the well-known factor-of-two penalty in spectrum-blind sampling [45]–[47]. Note that the same factor-of-two penalty can be inferred from the RIC-based recovery guarantees in [15], [39], when comparing the recovery threshold specified in [39, Thm. 1] for signals where partial support-set knowledge is available (prior to recovery) to that given in [15, Thm. 1.1] which does not assume prior support-set knowledge.

We illustrate the factor-of-two penalty in Figs. 1 and 2, where the recovery thresholds (15), (18), (22), (24), and (26) are shown. In Fig. 1, we consider the case $\mu_a = \mu_b = 0$ and $\mu_m = 1/\sqrt{64}$. We can see that for $\mathcal{X}$ and $\mathcal{E}$ known the threshold evaluates to $n_x n_e < 64$. When only $\mathcal{X}$ or $\mathcal{E}$ is known we have $n_x n_e < 32$, and finally in the case where only n_e is known we get $n_x n_e < 16$. Note furthermore that in Case IV, where no knowledge about the support sets is available, the recovery threshold is more restrictive than in the case where n_e is known.

In Fig. 2, we show the recovery thresholds for $\mu_a = 0.1258$, $\mu_b = 0.1319$, and $\mu_m = 0.1321$. We see that all threshold curves are straight lines. This behavior can be explained by noting that (in contrast to the assumptions underlying Fig. 1) the dictionaries $\mathbf{A}$ and $\mathbf{B}$ have $\mu_a, \mu_b > 0$ and the corresponding recovery thresholds are essentially dominated by the numerator of the RHS expressions in (15), (18), (22), and (24), which depends on both n_x and n_e. More concretely, if $\mu_a = \mu_b = \mu_m = \mu_d > 0$, then the recovery threshold for Case II (where the support set $\mathcal{E}$ is known) becomes

$$2n_x + n_e < \left(1 + \mu_d^{-1}\right) \tag{27}$$

which reflects the behavior observed in Fig. 2.

B. The square-root bottleneck

The recovery thresholds presented in Section IV hold for *all signal and noise realizations* $\mathbf{x}$ and $\mathbf{e}$ and for *all* dictionary pairs (with given coherence parameters). However, as is well-known in the sparse-signal recovery literature, coherence-based recovery guarantees are—in contrast to RIC-based recovery guarantees—fundamentally limited by the so-called

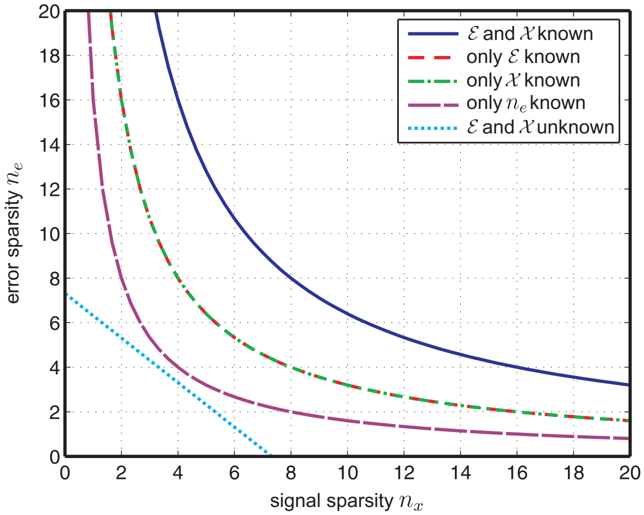

Fig. 1. Recovery thresholds (15), (18), (22), (24), and (26) for $\mu_a = \mu_b = 0$, and $\mu_m = 1/\sqrt{64}$. Note that the curves for "only $\mathcal{E}$ known" and "only $\mathcal{X}$ known" are on top of each other.

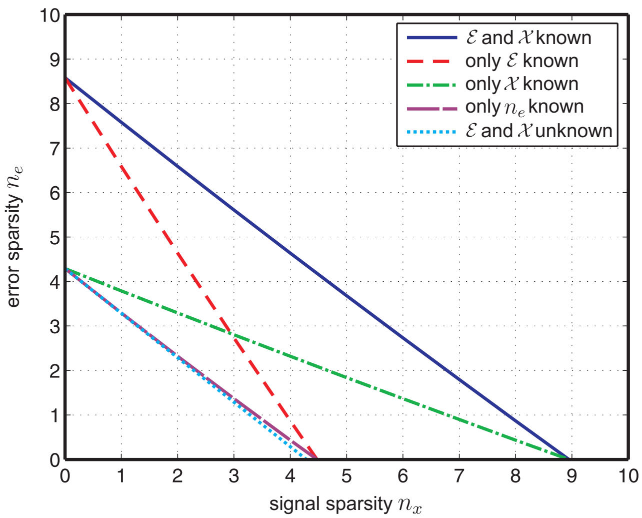

Fig. 2. Recovery thresholds (15), (18), (22), (24), and (26) for $\mu_a = 0.1258$, $\mu_b = 0.1319$, and $\mu_m = 0.1321$.

square-root bottleneck [48]. More specifically, in the noiseless case (i.e., for $\mathbf{e} = \mathbf{0}_{N_b}$), the threshold (4) states that recovery can be guaranteed only for up to $\sqrt{M}$ nonzero entries in $\mathbf{x}$. Put differently, for a fixed number of nonzero entries n_x in $\mathbf{x}$, i.e., for a fixed sparsity level, the number of measurements M required to recover $\mathbf{x}$ through (P0), BP, or OMP is on the order of n_x^2.

As in the classical sparse-signal recovery literature, the square-root bottleneck can be broken by performing a probabilistic analysis [48]. This line of work—albeit interesting—is outside the scope of the present paper and is further investigated in [33], [49], [50].

C. Trade-off between n_x and n_e

We next illustrate a trade-off between the sparsity levels of $\mathbf{x}$ and $\mathbf{e}$. Following the procedure outlined in [12], [51], we construct a dictionary $\mathbf{A}$ consisting of A ONBs and a dictionary $\mathbf{B}$ consisting of B ONBs such that $\mu_a = \mu_b = \mu_m = 1/\sqrt{M}$, where $A + B \leq M + 1$ with $M = p^k$, p prime, and $k \in \mathbb{N}^+$. Now, let us assume that the error sparsity level scales according to $n_e = \alpha\sqrt{M}$ for some $0 \leq \alpha \leq 1$. For the case where only $\mathcal{E}$ is known but $\mathcal{X}$ is unknown (Case II), we find from (18) that any signal $\mathbf{x}$ with (order-wise) $(1-\alpha)\sqrt{M}/2$ non-zero entries (ignoring terms of order less than $\sqrt{M}$) can be reconstructed. Hence, there is a trade-off between the sparsity levels of $\mathbf{x}$ and $\mathbf{e}$ (here quantified through the parameter α), and both sparsity levels scale with $\sqrt{M}$.

VI. Numerical Results

We first report simulation results and compare them to the corresponding analytical results in the paper. We will find that even though the analytical thresholds are pessimistic in general, they do reflect the numerically observed recovery behavior correctly. In particular, we will see that the factor-of-two penalty discussed in Section V-A can also be observed in the numerical results. We then demonstrate, through a simple inpainting example, that perfect signal recovery in the presence of sparse errors is possible even if the corruptions are significant (in terms of the ℓ_2-norm of the sparse noise vector $\mathbf{e}$). In all numerical results, OMP is performed with a predetermined number of iterations [13], [24], [25], i.e., for Case II and Case IV, we set the number of iterations to n_x and $n_x + n_e$, respectively. To implement BP, we employ SPGL1 [52], [53].

A. Impact of support-set knowledge on recovery thresholds

We first compare simulation results to the recovery thresholds (15), (18), (22), and (26). For a given pair of dictionaries $\mathbf{A}$ and $\mathbf{B}$ we generate signal vectors $\mathbf{x}$ and error vectors $\mathbf{e}$ as follows: We first fix n_x and n_e, then the support sets of the n_x-sparse vector $\mathbf{x}$ and the n_e-sparse vector $\mathbf{e}$ are chosen uniformly at random among all possible support sets of cardinality n_x and n_e, respectively. Once the support sets have been chosen, we generate the nonzero entries of $\mathbf{x}$ and $\mathbf{e}$ by drawing from i.i.d. zero mean, unit variance Gaussian random variables. For each pair of support-set cardinalities n_x and n_e, we perform 10 000 Monte-Carlo trials and declare success of recovery whenever the recovered vector $\hat{\mathbf{x}}$ satisfies

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 < 10^{-3}\|\mathbf{x}\|_2\,. \tag{28}$$

We plot the 50% success-rate contour, i.e., the border between the region of pairs (n_x, n_e) for which (28) is satisfied in at least 50% of the trials and the region where (28) is satisfied in less than 50% of the trials. The recovered vector $\hat{\mathbf{x}}$ is obtained as follows:

- *Case I:* When $\mathcal{X}$ and $\mathcal{E}$ are both known, we perform recovery according to (14).
- *Case II:* When either only $\mathcal{E}$ or only $\mathcal{X}$ is known, we apply BP and OMP using the modified dictionary as detailed in Theorem 5 and Corollary 6, respectively.
- *Case IV:* When neither $\mathcal{X}$ nor $\mathcal{E}$ is known, we apply BP and OMP to the concatenated dictionary $\mathbf{D} = [\,\mathbf{A}\ \mathbf{B}\,]$ as described in Theorem 9.

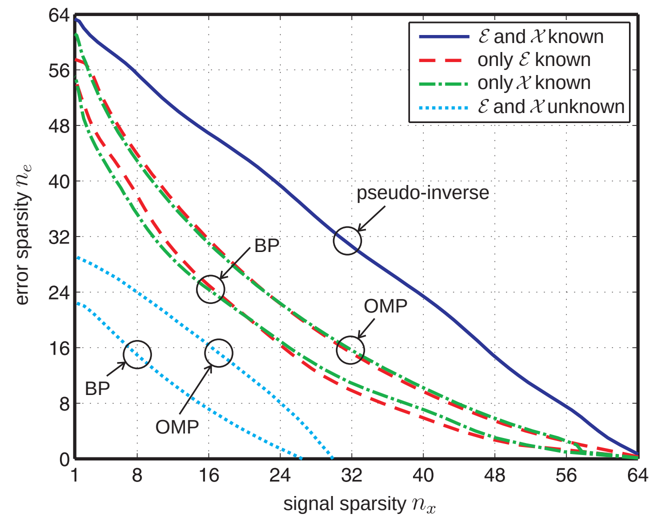

Fig. 3. Impact of support-set knowledge on the 50% success-rate contour of OMP and BP for the Hadamard–identity pair.

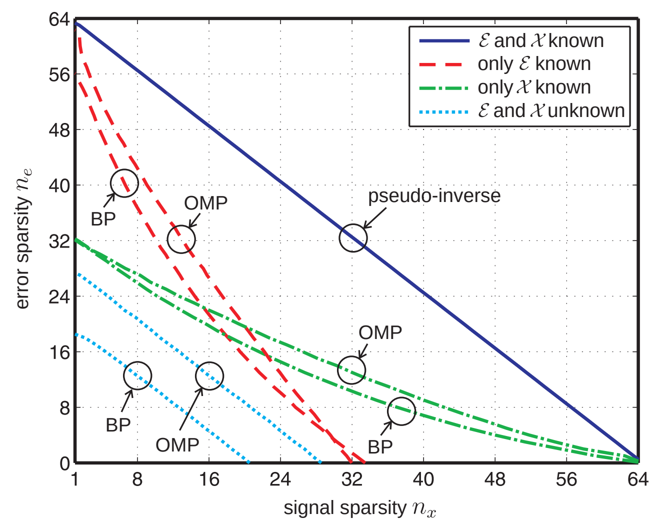

Fig. 4. Impact of support-set knowledge on the 50% success-rate contour of OMP and BP performing recovery in pairs of approximate ETFs each of dimension 64×80.

Note that for *Case III*, i.e., the case where the cardinality n_e of the support set $\mathcal{E}$ is known—as pointed out in Section IV-C—we only have uniqueness results but no analytical recovery guarantees, neither for BP nor for greedy recovery algorithms that make use of the *separate* knowledge of n_x or n_e (whereas, e.g., standard OMP makes use of knowledge of $n_x + n_e$, rather than knowledge of n_x and n_e individually). This case is, therefore, not considered in the simulation results below.

1) Recovery performance for the Hadamard–identity pair using BP and OMP: We take $M = 64$, let $\mathbf{A}$ be the Hadamard ONB [54] and set $\mathbf{B} = \mathbf{I}_M$, which results in $\mu_a = \mu_b = 0$ and $\mu_m = 1/\sqrt{M}$. Fig. 3 shows 50% success-rate contours, under different assumptions of support-set knowledge. For perfect knowledge of $\mathcal{X}$ and $\mathcal{E}$, we observe that the 50% success-rate contour is at about $n_x + n_e \approx M$, which is significantly better than the sufficient condition $n_x n_e < M$ (guaranteeing perfect recovery) provided in (15).[5] When either only $\mathcal{X}$ or only $\mathcal{E}$ is known, the recovery performance is essentially independent of whether $\mathcal{X}$ or $\mathcal{E}$ is known. This is also reflected by the analytical thresholds (18) and (22) when evaluated for $\mu_a = \mu_b = 0$ (see also Fig. 1). Furthermore, OMP is seen to outperform BP. When neither $\mathcal{X}$ nor $\mathcal{E}$ is known, OMP again outperforms BP.

It is interesting to see that the factor-of-two penalty discussed in Section V-A is reflected in Fig. 3 (for $n_x = n_e$) between Cases I and II. Specifically, we can observe that for full support-set knowledge (Case I) the 50% success-rate is achieved at $n_x = n_e \approx 31$. If either $\mathcal{X}$ or $\mathcal{E}$ only is known (Case II), OMP achieves 50% success-rate at $n_x = n_e \approx 23$, demonstrating a factor-of-two penalty since $31 \cdot 31 \approx 23 \cdot 23 \cdot 2$. Note that the results from BP in Fig. 3 do not seem to reflect the factor-of-two penalty. For lack of an efficient recovery algorithm (making use of knowledge of n_e) we do not show numerical results for Case III.

2) Impact of $\mu_a, \mu_b > 0$: We take $M = 64$ and generate the dictionaries $\mathbf{A}$ and $\mathbf{B}$ as follows. Using the alternating projection method described in [56], we generate an approximate equiangular tight frame (ETF) for $\mathbb{R}^M$ consisting of 160 columns. We split this frame into two sets of 80 elements (columns) each and organize them in the matrices $\mathbf{A}$ and $\mathbf{B}$ such that the corresponding coherence parameters are given by $\mu_a \approx 0.1258$, $\mu_b \approx 0.1319$, and $\mu_m \approx 0.1321$. Fig. 4 shows the 50% success-rate contour under four different assumptions of support-set knowledge. In the case where either only $\mathcal{X}$ or only $\mathcal{E}$ is known and in the case where $\mathcal{X}$ and $\mathcal{E}$ are unknown, we use OMP and BP for recovery. It is interesting to note that the graphs for the cases where only $\mathcal{X}$ or only $\mathcal{E}$ are known, are symmetric with respect to the line $n_x = n_e$. This symmetry is also reflected in the analytical thresholds (18) and (22) (see also Fig. 2 and the discussion in Section V-A).

We finally note that in all cases considered above, the numerical results show that recovery is possible for significantly higher sparsity levels n_x and n_e than indicated by the corresponding analytical thresholds (15), (18), (22), and (26) (see also Figs. 1 and 2). The underlying reasons are i) the deterministic nature of the results, i.e., the recovery guarantees in (15), (18), (22), and (26) are valid for *all* dictionary pairs (with given coherence parameters) and *all* signal and noise realizations (with given sparsity level), and ii) we plot the 50% success-rate contour, whereas the analytical results guarantee *perfect* recovery in 100% of the cases.

B. Inpainting example

In transform coding one is typically interested in maximally sparse representations of a given signal to be encoded [57]. In our setting, this would mean that the dictionary $\mathbf{A}$ should be chosen so that it leads to maximally sparse representations of a given family of signals. We next demonstrate, however, that in the presence of structured noise, the signal dictionary $\mathbf{A}$ should *additionally* be incoherent to the noise dictionary $\mathbf{B}$.

[5] For $\mathbf{A} = \mathbf{F}_M$ and $\mathbf{B} = \mathbf{I}_M$ it was proven in [55] that a set of columns chosen randomly from both $\mathbf{A}$ and $\mathbf{B}$ is linearly independent (with high probability) given that the total number of chosen columns, i.e., $n_x + n_e$ here, does not exceed a constant proportion of M.

(a) Corrupted image (MSE $= -11.2\,$dB)

(b) Recovery when $\mathcal{X}$ and $\mathcal{E}$ are known (MSE $= -184.6\,$dB)

(c) Recovery when only $\mathcal{E}$ is known (MSE $= -113.3\,$dB)

(d) Recovery for $\mathcal{X}$ and $\mathcal{E}$ unknown (MSE $= -13.0\,$dB)

Fig. 5. Recovery results using the DCT basis for the signal dictionary and the identity basis for the noise dictionary, for the cases where (b) $\mathcal{X}$ and $\mathcal{E}$ are known, (c) only $\mathcal{E}$ is known, and (d) no support-set knowledge is available. (Picture origin: ETH Zürich/Esther Ramseier).

(a) Corrupted image (MSE $= -11.2\,$dB)

(b) Recovery when $\mathcal{X}$ and $\mathcal{E}$ are known (MSE $= -27.1\,$dB)

(c) Recovery when only $\mathcal{E}$ is known (MSE $= -27.1\,$dB)

(d) Recovery for $\mathcal{X}$ and $\mathcal{E}$ unknown (MSE $= -12.0\,$dB)

Fig. 6. Recovery results for the signal dictionary given by the Haar wavelet basis and the noise dictionary given by the identity basis, for the cases where (b) both $\mathcal{X}$ and $\mathcal{E}$ are known, (c) only $\mathcal{E}$ is known, and (d) no support-set knowledge is available. (Picture origin: ETH Zürich/Esther Ramseier).

This extra requirement can lead to very different criteria for designing transform bases (frames).

To illustrate this point, and to show that perfect recovery can be guaranteed even when the ℓ_2-norm of the noise term $\mathbf{Be}$ is large, we consider the recovery of a sparsely corrupted 512×512-pixel grayscale image of the main building of ETH Zurich. The dictionary $\mathbf{A}$ is taken to be either the two-dimensional discrete cosine transform (DCT) or the Haar wavelet decomposed on three octaves [58]. We first "sparsify" the image by retaining the 15% largest entries of the image's representation $\mathbf{x}$ in $\mathbf{A}$. We then corrupt (by overwriting with text) 18.8% of the pixels in the sparsified image by setting them to the brightest grayscale value; this means that the errors are sparse in $\mathbf{B} = \mathbf{I}_M$ and that the noise is structured but may have large ℓ_2-norm. Image recovery is performed according to (14) if $\mathcal{X}$ and $\mathcal{E}$ are known. (Note, however, that knowledge of $\mathcal{X}$ is usually not available in inpainting applications.) We use BP when only $\mathcal{E}$ is known and when neither $\mathcal{X}$ nor $\mathcal{E}$ are known. The recovery results are evaluated by computing the mean-square error (MSE) between the sparsified image and its recovered version.

Figs. 5 and 6 show the corresponding results. As expected, the MSE increases as the amount of knowledge about the support sets decreases. More interestingly, we note that even though Haar wavelets often yield a smaller approximation error in classical transform coding compared to the DCT, here the wavelet transform performs worse than the DCT. This behavior is due to the fact that sparsity is not the only factor determining the performance of a transform coding basis (or frame) in the presence of structured noise. Rather the mutual coherence between the dictionary used to represent the signal and that used to represent the structured noise becomes highly relevant. Specifically, in the example at hand, we have $\mu_m = 1/2$ for the Haar-wavelet and the identity basis, and $\mu_m \approx 0.004$ for the DCT and the identity basis. The dependence of the analytical thresholds (15), (18), (22), (24), and (26) on the mutual coherence μ_m explains the performance difference between the Haar wavelet basis and the DCT basis. An intuitive explanation for this behavior is as follows: The Haar-wavelet basis contains only four non-zero entries in the columns associated to fine scales, which is reflected in the high mutual coherence (i.e., $\mu_m = 1/2$) between the Haar-wavelet basis and the identity basis. Thus, when projecting onto the orthogonal complement of $(\mathbf{I}_M)_\mathcal{E}$, it is likely that all non-zero entries of such columns are deleted, resulting in columns of all zeros. Recovery of the corresponding non-zero entries of $\mathbf{x}$ is thus not possible. In summary, we see that the choice of the transform basis (frame) for a sparsely corrupted signal should not only aim at sparsifying the signal as much as possible but should also take into account the mutual coherence between the transform basis (frame) and the noise sparsity basis (frame).

VII. Conclusion

The setup considered in this paper, in its generality, appears to be new and a number of interesting extensions are possible. In particular, developing (coherence-based) recovery guarantees for greedy algorithms such as CoSaMP [41] or subspace pursuit [42] for all cases studied in the paper are interesting open problems. Note that probabilistic recovery guarantees for the case where nothing is known about the signal and noise support sets (i.e., Case IV) readily follow from the results in [33]. Probabilistic recovery guarantees for the other cases studied in this paper are in preparation [50]. Furthermore, an extension of the results in this paper that accounts for measurement noise (in addition to sparse noise) and applies to approximately sparse signals can be found in [59].

Acknowledgments

The authors would like to thank C. Aubel, A. Bracher, and G. Durisi for interesting discussions, and the anonymous reviewers for their valuable comments which helped to improve the exposition.

Appendix A
Proof of Theorem 3

We prove the full (column-)rank property of $\mathbf{D}_{\mathcal{X},\mathcal{E}}$ by showing that under (15) there is a unique pair $(\mathbf{x},\mathbf{e})$ with $\operatorname{supp}(\mathbf{x}) = \mathcal{X}$ and $\operatorname{supp}(\mathbf{e}) = \mathcal{E}$ satisfying $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$. Assume that there exists an alternative pair $(\mathbf{x}',\mathbf{e}')$ such that $\mathbf{z} = \mathbf{A}\mathbf{x}'+\mathbf{B}\mathbf{e}'$ with $\operatorname{supp}(\mathbf{x}') \subseteq \mathcal{X}$ and $\operatorname{supp}(\mathbf{e}') \subseteq \mathcal{E}$ (i.e., the support sets of $\mathbf{x}'$ and $\mathbf{e}'$ are contained in $\mathcal{X}$ and $\mathcal{E}$, respectively). This would then imply that

$$\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e} = \mathbf{A}\mathbf{x}' + \mathbf{B}\mathbf{e}'$$

and thus

$$\mathbf{A}(\mathbf{x} - \mathbf{x}') = \mathbf{B}(\mathbf{e}' - \mathbf{e}).$$

Since both $\mathbf{x}$ and $\mathbf{x}'$ have support in $\mathcal{X}$ it follows that $\mathbf{x} - \mathbf{x}'$ also has support in $\mathcal{X}$, which implies $\|\mathbf{x} - \mathbf{x}'\|_0 \leq n_x$. Similarly, we get $\|\mathbf{e}' - \mathbf{e}\|_0 \leq n_e$. Defining $\mathbf{p} = \mathbf{x} - \mathbf{x}'$ and $\mathcal{P} = \operatorname{supp}(\mathbf{x} - \mathbf{x}') \subseteq \mathcal{X}$, and, similarly, $\mathbf{q} = \mathbf{e}' - \mathbf{e}$ and $\mathcal{Q} = \operatorname{supp}(\mathbf{e}' - \mathbf{e}) \subseteq \mathcal{E}$, we obtain the following chain of inequalities:

$$\begin{aligned} n_x n_e &\geq \|\mathbf{p}\|_0 \|\mathbf{q}\|_0 = |\mathcal{P}|\,|\mathcal{Q}| \\ &\geq \frac{[1-\mu_a(|\mathcal{P}|-1)]^+ [1-\mu_b(|\mathcal{Q}|-1)]^+}{\mu_m^2} \qquad (29) \\ &\geq \frac{[1-\mu_a(n_x-1)]^+ [1-\mu_b(n_e-1)]^+}{\mu_m^2} = f(n_x,n_e) \qquad (30) \end{aligned}$$

where (29) follows by applying the uncertainty relation in Theorem 1 (with $\epsilon_\mathcal{P} = \epsilon_\mathcal{Q} = 0$ since both $\mathbf{p}$ and $\mathbf{q}$ are perfectly concentrated to $\mathcal{P}$ and $\mathcal{Q}$, respectively) and (30) is a consequence of $|\mathcal{P}| \leq n_x$ and $|\mathcal{Q}| \leq n_e$. Obviously, (30) contradicts the assumption in (15), which completes the proof.

Appendix B
Proof of Theorem 4

We begin by proving that $\mathbf{x}$ is the unique solution of $(\mathrm{P0},\mathcal{E})$ applied to $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$. Assume that there exists an alternative vector $\mathbf{x}'$ that satisfies $\mathbf{A}\mathbf{x}' \in (\{\mathbf{z}\} + \mathcal{R}(\mathbf{B}_\mathcal{E}))$ with $\|\mathbf{x}'\|_0 \leq n_x$. This would imply the existence of a vector $\mathbf{e}'$ with $\operatorname{supp}(\mathbf{e}') \subseteq \mathcal{E}$, such that

$$\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e} = \mathbf{A}\mathbf{x}' + \mathbf{B}\mathbf{e}'$$

and hence

$$\mathbf{A}(\mathbf{x} - \mathbf{x}') = \mathbf{B}(\mathbf{e}' - \mathbf{e}).$$

Since $\operatorname{supp}(\mathbf{e}) = \mathcal{E}$ and $\operatorname{supp}(\mathbf{e}') \subseteq \mathcal{E}$, we have $\operatorname{supp}(\mathbf{e}'-\mathbf{e}) \subseteq \mathcal{E}$ and hence $\|\mathbf{e}' - \mathbf{e}\|_0 \leq n_e$. Furthermore, since both $\mathbf{x}$ and $\mathbf{x}'$ have at most n_x nonzero entries, we have $\|\mathbf{x} - \mathbf{x}'\|_0 \leq 2n_x$. Defining $\mathbf{p} = \mathbf{x} - \mathbf{x}'$ and $\mathcal{P} = \operatorname{supp}(\mathbf{x} - \mathbf{x}')$, and, similarly, $\mathbf{q} = \mathbf{e}'-\mathbf{e}$ and $\mathcal{Q} = \operatorname{supp}(\mathbf{e}'-\mathbf{e}) \subseteq \mathcal{E}$, we obtain the following chain of inequalities

$$\begin{aligned} 2n_x n_e &\geq \|\mathbf{p}\|_0 \|\mathbf{q}\|_0 = |\mathcal{P}|\,|\mathcal{Q}| \\ &\geq \frac{[1-\mu_a(|\mathcal{P}|-1)]^+ [1-\mu_b(|\mathcal{Q}|-1)]^+}{\mu_m^2} \qquad (31) \\ &\geq \frac{[1-\mu_a(2n_x-1)]^+ [1-\mu_b(n_e-1)]^+}{\mu_m^2} = f(2n_x,n_e) \qquad (32) \end{aligned}$$

where (31) follows by applying the uncertainty relation in Theorem 1 (with $\epsilon_\mathcal{P} = \epsilon_\mathcal{Q} = 0$ since both $\mathbf{p}$ and $\mathbf{q}$ are perfectly concentrated to $\mathcal{P}$ and $\mathcal{Q}$, respectively) and (32) is a consequence of $|\mathcal{P}| \leq 2n_x$ and $|\mathcal{Q}| \leq n_e$. Obviously, (32) contradicts the assumption in (18), which concludes the first part of the proof.

We next prove that $\mathbf{x}$ is also the unique solution of $(\mathrm{BP},\mathcal{E})$ applied to $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$. Assume that there exists an alternative vector $\mathbf{x}'$ that satisfies $\mathbf{A}\mathbf{x}' \in (\{\mathbf{z}\} + \mathcal{R}(\mathbf{B}_\mathcal{E}))$ with $\|\mathbf{x}'\|_1 \leq \|\mathbf{x}\|_1$. This would imply the existence of a vector $\mathbf{e}'$ with $\operatorname{supp}(\mathbf{e}') \subseteq \mathcal{E}$, such that

$$\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e} = \mathbf{A}\mathbf{x}' + \mathbf{B}\mathbf{e}'$$

and hence

$$\mathbf{A}(\mathbf{x} - \mathbf{x}') = \mathbf{B}(\mathbf{e}' - \mathbf{e}).$$

Defining $\mathbf{p} = \mathbf{x} - \mathbf{x}'$, we obtain the following lower bound for the ℓ_1-norm of $\mathbf{x}'$

$$\begin{aligned} \|\mathbf{x}'\|_1 = \|\mathbf{x} - \mathbf{p}\|_1 &= \|\mathbf{P}_\mathcal{X}(\mathbf{x} - \mathbf{p})\|_1 + \|\mathbf{P}_{\mathcal{X}^c}\mathbf{p}\|_1 \\ &\geq \|\mathbf{P}_\mathcal{X}\mathbf{x}\|_1 - \|\mathbf{P}_\mathcal{X}\mathbf{p}\|_1 + \|\mathbf{P}_{\mathcal{X}^c}\mathbf{p}\|_1 \qquad (33) \\ &= \|\mathbf{x}\|_1 - \|\mathbf{P}_\mathcal{X}\mathbf{p}\|_1 + \|\mathbf{P}_{\mathcal{X}^c}\mathbf{p}\|_1 \end{aligned}$$

where (33) is a consequence of the reverse triangle inequality. Now, the ℓ_1-norm of $\mathbf{x}'$ can be smaller than or equal to that of $\mathbf{x}$ only if $\|\mathbf{P}_\mathcal{X}\mathbf{p}\|_1 \geq \|\mathbf{P}_{\mathcal{X}^c}\mathbf{p}\|_1$. This would then imply that the difference vector $\mathbf{p}$ needs to be at least 50%-concentrated to the set $\mathcal{P} = \mathcal{X}$ (of cardinality n_x), i.e., we require that $\epsilon_\mathcal{P} \leq 0.5$. Defining $\mathbf{q} = \mathbf{e}' - \mathbf{e}$ and $\mathcal{Q} = \operatorname{supp}(\mathbf{e}' - \mathbf{e})$, and noting that $\text{supp}(\mathbf{e}) = \mathcal{E}$ and $\text{supp}(\mathbf{e}') \subseteq \mathcal{E}$, it follows that $|\mathcal{Q}| \leq n_e$. This leads to the following chain of inequalities:

$$n_x\, n_e \geq |\mathcal{P}|\,|\mathcal{Q}| \geq \frac{[(1+\mu_a)(1-\epsilon_{\mathcal{P}}) - |\mathcal{P}|\,\mu_a]^+\,[1-\mu_b\,(|\mathcal{Q}|-1)]^+}{\mu_m^2} \tag{34}$$

$$\geq \frac{1}{2}\,\frac{[1-\mu_a(2n_x-1)]^+\,[1-\mu_b(n_e-1)]^+}{\mu_m^2} \tag{35}$$

where (34) follows from the uncertainty relation in Theorem 1 applied to the difference vectors $\mathbf{p}$ and $\mathbf{q}$ (with $\epsilon_{\mathcal{P}} \leq 0.5$ since $\mathbf{p}$ is at least 50%-concentrated to $\mathcal{P}$ and $\epsilon_{\mathcal{Q}} = 0$ since $\mathbf{q}$ is perfectly concentrated to $\mathcal{Q}$) and (35) is a consequence of $|\mathcal{P}| = n_x$ and $|\mathcal{Q}| \leq n_e$. Rewriting (35), we obtain

$$2n_xn_e \geq \frac{[1-\mu_a(2n_x-1)]^+\,[1-\mu_b(n_e-1)]^+}{\mu_m^2} = f(2n_x, n_e). \tag{36}$$

Since (36) contradicts the assumption in (18), this proves that $\mathbf{x}$ is the unique solution of $(\text{BP}, \mathcal{E})$ applied to $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$.

APPENDIX C
PROOF OF THEOREM 5

We first show that condition (18) ensures that the columns of $\mathbf{B}_{\mathcal{E}}$ are linearly independent. Then, we establish that $\|\mathbf{R}_{\mathcal{E}}\mathbf{a}_\ell\|_2 > 0$ for $\ell = 1, \ldots, N_a$. Finally, we show that the unique solution of (P0), BP, and OMP applied to $\hat{\mathbf{z}} = \mathbf{R}_{\mathcal{E}}\mathbf{A}\boldsymbol{\Delta}\hat{\mathbf{x}}$ is given by $\hat{\mathbf{x}} = \boldsymbol{\Delta}^{-1}\mathbf{x}$.

A. The columns of $\mathbf{B}_{\mathcal{E}}$ *are linearly independent*

Condition (18) can only be satisfied if $[1-\mu_b(n_e-1)]^+ > 0$, which implies that $n_e < 1 + 1/\mu_b$. It was shown in [11]–[13] that for a dictionary $\mathbf{B}$ with coherence μ_b no fewer than $1 + 1/\mu_b$ columns of $\mathbf{B}$ can be linearly dependent. Hence, the n_e columns of $\mathbf{B}_{\mathcal{E}}$ must be linearly independent.

B. $\|\mathbf{R}_{\mathcal{E}}\mathbf{a}_\ell\|_2 > 0$ *for* $\ell = 1, \ldots, N_a$

We have to verify that condition (18) implies $\|\mathbf{R}_{\mathcal{E}}\mathbf{a}_\ell\|_2 > 0$ for $\ell = 1, \ldots, N_a$. Since $\mathbf{R}_{\mathcal{E}}$ is a projector and, therefore, Hermitian and idempotent, it follows that

$$\|\mathbf{R}_{\mathcal{E}}\mathbf{a}_\ell\|_2^2 = \mathbf{a}_\ell^H\mathbf{R}_{\mathcal{E}}\mathbf{a}_\ell \tag{37}$$

$$= \left|\mathbf{a}_\ell^H\mathbf{R}_{\mathcal{E}}\mathbf{a}_\ell\right|$$

$$\geq 1 - \underbrace{\left|\mathbf{a}_\ell^H\mathbf{B}_{\mathcal{E}}\left(\mathbf{B}_{\mathcal{E}}^H\mathbf{B}_{\mathcal{E}}\right)^{-1}\mathbf{B}_{\mathcal{E}}^H\mathbf{a}_\ell\right|}_{\triangleq C_1} \tag{38}$$

where (37) is a consequence of $\mathbf{R}_{\mathcal{E}}^H\mathbf{R}_{\mathcal{E}} = \mathbf{R}_{\mathcal{E}}$, and (38) follows from the reverse triangle inequality and $\|\mathbf{a}_\ell\|_2 = 1$, $\ell = 1, \ldots, N_a$. Next, we derive an upper bound on C_1 according to

$$C_1 \leq \lambda_{\max}\Big(\left(\mathbf{B}_{\mathcal{E}}^H\mathbf{B}_{\mathcal{E}}\right)^{-1}\Big)\left\|\mathbf{B}_{\mathcal{E}}^H\mathbf{a}_\ell\right\|_2^2 \tag{39}$$

$$\leq \lambda_{\min}^{-1}\big(\mathbf{B}_{\mathcal{E}}^H\mathbf{B}_{\mathcal{E}}\big)\, n_e\,\mu_m^2 \tag{40}$$

where (39) follows from the Rayleigh-Ritz theorem [60, Thm. 4.2.2] and (40) results from

$$\left\|\mathbf{B}_{\mathcal{E}}^H\mathbf{a}_\ell\right\|_2^2 = \sum_{i\in\mathcal{E}}\left|\mathbf{b}_i^H\mathbf{a}_\ell\right|^2 \leq n_e\,\mu_m^2.$$

Next, applying Geršgorin's disc theorem [60, Theorem 6.1.1], we arrive at

$$\lambda_{\min}\big(\mathbf{B}_{\mathcal{E}}^H\mathbf{B}_{\mathcal{E}}\big) \geq [1-\mu_b(n_e-1)]^+. \tag{41}$$

Combining (38), (40), and (41) leads to the following lower bound on $\|\mathbf{R}_{\mathcal{E}}\mathbf{a}_\ell\|_2^2$:

$$\|\mathbf{R}_{\mathcal{E}}\mathbf{a}_\ell\|_2^2 \geq 1 - \frac{n_e\,\mu_m^2}{[1-\mu_b(n_e-1)]^+}. \tag{42}$$

Note that if condition (18) holds for[6] $n_x \geq 1$, it follows that $n_e\,\mu_m^2 < [1-\mu_b(n_e-1)]^+$ and hence the RHS of (42) is strictly positive. This ensures that $\boldsymbol{\Delta}$ defines a one-to-one mapping. We next show that, moreover, condition (18) ensures that for every vector $\mathbf{x}' \in \mathbb{C}^{N_a}$ satisfying $\|\mathbf{x}'\|_0 \leq 2n_x$, $\mathbf{A}\mathbf{x}'$ has a nonzero component that is orthogonal to $\mathcal{R}(\mathbf{B}_{\mathcal{E}})$.

C. Unique recovery through (P0), BP, and OMP

We now need to verify that (P0), BP, and OMP (applied to $\hat{\mathbf{z}} = \mathbf{R}_{\mathcal{E}}\mathbf{A}\boldsymbol{\Delta}\hat{\mathbf{x}}$) recover the vector $\hat{\mathbf{x}} = \boldsymbol{\Delta}^{-1}\mathbf{x}$ provided that (18) is satisfied. This will be accomplished by deriving an upper bound on the coherence $\mu(\mathbf{R}_{\mathcal{E}}\mathbf{A}\boldsymbol{\Delta})$ of the modified dictionary $\mathbf{R}_{\mathcal{E}}\mathbf{A}\boldsymbol{\Delta}$, which, via the well-known coherence-based recovery guarantee [11]–[13]

$$n_x < \frac{1}{2}\Big(1 + \mu_d(\mathbf{R}_{\mathcal{E}}\mathbf{A}\boldsymbol{\Delta})^{-1}\Big) \tag{43}$$

leads to a recovery threshold guaranteeing perfect recovery of $\hat{\mathbf{x}}$. This threshold is then shown to coincide with (18). More specifically, the well-known sparsity threshold in (4) guarantees that the unique solution of (P0) applied to $\hat{\mathbf{z}} = \mathbf{R}_{\mathcal{E}}\mathbf{A}\boldsymbol{\Delta}\hat{\mathbf{x}}$ is given by $\hat{\mathbf{x}}$, and, furthermore, that this unique solution can be obtained through BP and OMP if (43) holds. It is important to note that $\|\hat{\mathbf{x}}\|_0 = \|\mathbf{x}\|_0 = n_x$. With

$$[\boldsymbol{\Delta}]_{\ell,\ell} = \frac{1}{\|\mathbf{R}_{\mathcal{E}}\mathbf{a}_\ell\|_2}, \quad \ell = 1, \ldots, N_a$$

we obtain

$$\mu(\mathbf{R}_{\mathcal{E}}\mathbf{A}\boldsymbol{\Delta}) = \max_{r,\ell,\ell\neq r}\frac{\left|\mathbf{a}_r^H\mathbf{R}_{\mathcal{E}}^H\mathbf{R}_{\mathcal{E}}\mathbf{a}_\ell\right|}{\|\mathbf{R}_{\mathcal{E}}\mathbf{a}_r\|_2\,\|\mathbf{R}_{\mathcal{E}}\mathbf{a}_\ell\|_2}. \tag{44}$$

Next, we upper-bound the RHS of (44) by upper-bounding its numerator and lower-bounding its denominator. For the numerator we have

$$\left|\mathbf{a}_r^H\mathbf{R}_{\mathcal{E}}^H\mathbf{R}_{\mathcal{E}}\mathbf{a}_\ell\right| = \left|\mathbf{a}_r^H\mathbf{R}_{\mathcal{E}}\mathbf{a}_\ell\right| \tag{45}$$

$$\leq \left|\mathbf{a}_r^H\mathbf{a}_\ell\right| + \left|\mathbf{a}_r^H\mathbf{B}_{\mathcal{E}}\mathbf{B}_{\mathcal{E}}^\dagger\mathbf{a}_\ell\right| \tag{46}$$

$$\leq \mu_a + \underbrace{\left|\mathbf{a}_r^H\mathbf{B}_{\mathcal{E}}\left(\mathbf{B}_{\mathcal{E}}^H\mathbf{B}_{\mathcal{E}}\right)^{-1}\mathbf{B}_{\mathcal{E}}^H\mathbf{a}_\ell\right|}_{\triangleq C_2} \tag{47}$$

[6] The case $n_x = 0$ is not interesting, as $n_x = 0$ corresponds to $\mathbf{x} = \mathbf{0}_{N_a}$ and hence recovery of $\mathbf{x} = \mathbf{0}_{N_a}$ only could be guaranteed.

where (45) follows from $\mathbf{R}_{\mathcal{E}}^H\mathbf{R}_{\mathcal{E}} = \mathbf{R}_{\mathcal{E}}$, (46) is obtained through the triangle inequality, and (47) follows from $\left|\mathbf{a}_r^H\mathbf{a}_\ell\right| \le \mu_a$. Next, we derive an upper bound on C_2 according to

$$C_2 \le \left\|\mathbf{B}_{\mathcal{E}}^H\mathbf{a}_r\right\|_2 \left\|\left(\mathbf{B}_{\mathcal{E}}^H\mathbf{B}_{\mathcal{E}}\right)^{-1}\mathbf{B}_{\mathcal{E}}^H\mathbf{a}_\ell\right\|_2 \tag{48}$$

$$\le \left\|\mathbf{B}_{\mathcal{E}}^H\mathbf{a}_r\right\|_2 \left\|\left(\mathbf{B}_{\mathcal{E}}^H\mathbf{B}_{\mathcal{E}}\right)^{-1}\right\| \left\|\mathbf{B}_{\mathcal{E}}^H\mathbf{a}_\ell\right\|_2 \tag{49}$$

where (48) follows from the Cauchy-Schwarz inequality and (49) from the Rayleigh-Ritz theorem [60, Thm. 4.2.2]. Defining $i = \arg\max_r \left\|\mathbf{B}_{\mathcal{E}}^H\mathbf{a}_r\right\|_2$, we further have

$$\begin{aligned} C_2 &\le \left\|\left(\mathbf{B}_{\mathcal{E}}^H\mathbf{B}_{\mathcal{E}}\right)^{-1}\right\| \left\|\mathbf{B}_{\mathcal{E}}^H\mathbf{a}_i\right\|_2^2 \\ &= \lambda_{\max}\!\left(\left(\mathbf{B}_{\mathcal{E}}^H\mathbf{B}_{\mathcal{E}}\right)^{-1}\right) \left\|\mathbf{B}_{\mathcal{E}}^H\mathbf{a}_i\right\|_2^2. \end{aligned}$$

We obtain an upper bound on C_2 using the same steps that were used to bound C_1 in (39) – (41):

$$C_2 \le \frac{n_e\,\mu_m^2}{C_b} \tag{50}$$

where $C_b = [1-\mu_b(n_e-1)]^+$. Combining (47) and (50) leads to the following upper bound

$$\left|\mathbf{a}_r^H\mathbf{R}_{\mathcal{E}}^H\mathbf{R}_{\mathcal{E}}\mathbf{a}_\ell\right| \le \mu_a + \frac{n_e\,\mu_m^2}{C_b}. \tag{51}$$

Next, we derive a lower bound on the denominator on the RHS of (44). To this end, we set $j = \arg\min_r \left\|\mathbf{R}_{\mathcal{E}}\mathbf{a}_r\right\|_2$ and note that

$$\begin{aligned} \left\|\mathbf{R}_{\mathcal{E}}\mathbf{a}_r\right\|_2 \left\|\mathbf{R}_{\mathcal{E}}\mathbf{a}_\ell\right\|_2 &\ge \left\|\mathbf{R}_{\mathcal{E}}\mathbf{a}_j\right\|_2^2 \\ &\ge 1 - \frac{n_e\,\mu_m^2}{C_b} \end{aligned} \tag{52}$$

where (52) follows from (42). Finally, combining (51) and (52) we arrive at

$$\mu(\mathbf{R}_{\mathcal{E}}\mathbf{A}\boldsymbol{\Delta}) \le \frac{\mu_a C_b + n_e\,\mu_m^2}{C_b - n_e\,\mu_m^2}. \tag{53}$$

Inserting (53) into the recovery threshold in (43), we obtain the following threshold guaranteeing recovery of $\hat{\mathbf{x}}$ from $\hat{\mathbf{z}} = \mathbf{R}_{\mathcal{E}}\mathbf{A}\boldsymbol{\Delta}\hat{\mathbf{x}}$ through (P0), BP, and OMP:

$$n_x < \frac{1}{2}\left(\frac{C_b(1+\mu_a)}{\mu_a C_b + n_e\,\mu_m^2}\right). \tag{54}$$

Since $2n_x n_e\,\mu_m^2 \ge 0$, we can transform (54) into

$$\begin{aligned} 2n_x n_e\,\mu_m^2 &< C_b[1-\mu_a(2n_x-1)]^+ \\ &= [1-\mu_b(n_e-1)]^+\,[1-\mu_a(2n_x-1)]^+. \end{aligned} \tag{55}$$

Rearranging terms in (55) finally yields

$$2n_x n_e < f(2n_x, n_e)$$

which proves that (18) guarantees recovery of the vector $\hat{\mathbf{x}}$ (and thus also of $\mathbf{x} = \boldsymbol{\Delta}\hat{\mathbf{x}}$) through (P0), BP, and OMP.

Appendix D
Proof of Theorem 7

Assume that there exists an alternative vector $\mathbf{x}'$ that satisfies $\mathbf{A}\mathbf{x}' \in (\{\mathbf{z}\} + \bigcup_{\mathcal{E}\in\mathscr{P}}\mathcal{R}(\mathbf{B}_{\mathcal{E}}))$ (with $\mathscr{P} = \wp_{n_e}(\{1,\ldots,N_b\})$) with $\|\mathbf{x}'\|_0 \le n_x$. This implies the existence of a vector $\mathbf{e}'$ with $\|\mathbf{e}'\|_0 \le n_e$ such that

$$\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e} = \mathbf{A}\mathbf{x}' + \mathbf{B}\mathbf{e}'$$

and therefore

$$\mathbf{A}(\mathbf{x}-\mathbf{x}') = \mathbf{B}(\mathbf{e}'-\mathbf{e}).$$

From $\|\mathbf{x}\|_0 = n_x$ and $\|\mathbf{x}'\|_0 \le n_x$ it follows that $\|\mathbf{x}-\mathbf{x}'\|_0 \le 2n_x$. Similarly, $\|\mathbf{e}\|_0 = n_e$ and $\|\mathbf{e}'\|_0 \le n_e$ imply $\|\mathbf{e}'-\mathbf{e}\|_0 \le 2n_e$. Defining $\mathbf{p} = \mathbf{x}-\mathbf{x}'$ and $\mathcal{P} = \operatorname{supp}(\mathbf{x}-\mathbf{x}')$, and, similarly, $\mathbf{q} = \mathbf{e}'-\mathbf{e}$ and $\mathcal{Q} = \operatorname{supp}(\mathbf{e}'-\mathbf{e})$, we arrive at

$$\begin{aligned} 4n_x n_e &\ge \|\mathbf{p}\|_0\,\|\mathbf{q}\|_0 = |\mathcal{P}|\,|\mathcal{Q}| \\ &\ge \frac{[1-\mu_a(|\mathcal{P}|-1)]^+\,[1-\mu_b(|\mathcal{Q}|-1)]^+}{\mu_m^2} \qquad (56) \\ &\ge \frac{[1-\mu_a(2n_x-1)]^+\,[1-\mu_b(2n_e-1)]^+}{\mu_m^2} = f(2n_x, 2n_e) \qquad (57) \end{aligned}$$

where (56) follows from the uncertainty relation in Theorem 1 applied to the difference vectors $\mathbf{p}$ and $\mathbf{q}$ (with $\epsilon_{\mathcal{P}} = \epsilon_{\mathcal{Q}} = 0$ since both $\mathbf{p}$ and $\mathbf{q}$ are perfectly concentrated to $\mathcal{P}$ and $\mathcal{Q}$, respectively) and (57) is a consequence of $|\mathcal{P}| \le 2n_x$ and $|\mathcal{Q}| \le 2n_e$. Obviously, (57) is in contradiction to (24), which concludes the proof.

References

[1] C. Studer, P. Kuppinger, G. Pope, and H. Bölcskei, “Sparse signal recovery from sparsely corrupted measurements,” in *Proc. of IEEE Int. Symp. on Inf. Theory (ISIT)*, St. Petersburg, Russia, Aug. 2011, pp. 1422–1426.

[2] J. S. Abel and J. O. Smith III, “Restoring a clipped signal,” in *Proc. of IEEE Int. Conf. Acoust. Speech Sig. Proc.*, vol. 3, May 1991, pp. 1745–1748.

[3] R. E. Carrillo, K. E. Barner, and T. C. Aysal, “Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise,” *IEEE J. Sel. Topics Sig. Proc.*, vol. 4, no. 2, pp. 392–408, 2010.

[4] 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.

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

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

[7] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, “Image inpainting,” in *Proc. of 27th Ann. Conf. Comp. Graph. Int. Tech.*, 2000, pp. 417–424.

[8] M. Aharon, M. Elad, and A. M. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Trans. Sig. Proc.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[9] M. Elad, J.-L. Starck, P. Querre, and D. L. Donoho, “Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA),” *Appl. Comput. Harmon. Anal.*, vol. 19, pp. 340–358, Nov. 2005.

[10] D. L. Donoho and G. Kutyniok, “Microlocal analysis of the geometric separation problem,” Feb. 2010. [Online]. Available: http://arxiv.org/abs/1004.3006v1

[11] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proc. Natl. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.

[12] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

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

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

[15] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *C. R. Acad. Sci. Paris, Ser. I*, vol. 346, pp. 589–592, 2008.

[16] T. T. Cai, L. Wang, 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.

[17] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[18] J. J. Fuchs, "Recovery of exact sparse representations in the presence of bounded noise," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3601–2608, Oct. 2005.

[19] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.

[20] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Trans. Sig. Proc.*, vol. 58, no. 10, pp. 5030–5043, Oct. 2010.

[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] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[23] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2558–2567, Sep. 2002.

[24] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. of 27th Asilomar Conf. on Signals, Systems, and Comput., Pacific Grove, CA, USA*, Nov. 1993, pp. 40–44.

[25] G. Davis, S. G. Mallat, and Z. Zhang, "Adaptive time-frequency decompositions," *Opt. Eng.*, vol. 33, no. 7, pp. 2183–2191, Jul. 1994.

[26] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM J. Appl. Math*, vol. 49, no. 3, pp. 906–931, Jun. 1989.

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

[28] J. N. Laska, M. A. Davenport, and R. G. Baraniuk, "Exact signal recovery from sparsely corrupted measurements through the pursuit of justice," in *Proc. of 43rd Asilomar Conf. on Signals, Systems, and Comput.*, Pacific Grove, CA, USA, Nov. 2009, pp. 1556–1560.

[29] J. Wright and Y. Ma, "Dense error correction via ℓ_1-minimization," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3540–3560, Jul. 2010.

[30] N. H. Nguyen and T. D. Tran, "Exact recoverability from dense corrupted observations via ℓ_1 minimization," Feb. 2011. [Online]. Available: http://arxiv.org/abs/1102.1227v3

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

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

[33] P. Kuppinger, G. Durisi, and H. Bölcskei, "Uncertainty relations and sparse signal recovery for pairs of general signal sets," *IEEE Trans. Inf. Theory*, Feb. 2012, to appear.

[34] P. Kuppinger, "General uncertainty relations and sparse signal recovery," Ph.D. dissertation, Series in Communication Theory, vol. 8, ed. H. Bölcskei, Hartung-Gorre Verlag, 2011.

[35] S. Ghobber and P. Jaming, "On uncertainty principles in the finite dimensional setting," 2010. [Online]. Available: http://arxiv.org/abs/0903.2923

[36] G. Davis, S. G. Mallat, and M. Avellaneda, "Adaptive greedy algorithms," *Constr. Approx.*, vol. 13, pp. 57–98, 1997.

[37] 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.

[38] 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.

[39] L. Jacques, "A short note on compressed sensing with partially known signal support," *EURASIP J. Sig. Proc.*, vol. 90, no. 12, pp. 3308–3312, Dec. 2010.

[40] A. Maleki and D. L. Donoho, "Optimally tuned iterative reconstruction algorithms for compressed sensing," *IEEE J. Sel. Topics Sig. Proc.*, vol. 4, no. 2, pp. 330–341, Apr. 2010.

[41] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, May 2009.

[42] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[43] M. A. Davenport, P. T. Boufonos, and R. G. Baraniuk, "Compressive domain interference cancellation," in *Proc. of Workshop on Sig. Proc. with Adaptive Sparse Structured Representations (SPARS)*, Saint-Malo, France, Apr. 2009.

[44] A. Feuer and A. Nemirovski, "On sparse representations in pairs of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 6, pp. 1579–1581, Jun. 2003.

[45] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," in *Proc. of IEEE Int. Conf. Acoust. Speech Sig. Proc. (ICASSP)*, vol. 3, Atlanta, GA, May 1996, pp. 1689–1692.

[46] Y. Bresler, "Spectrum-blind sampling and compressive sensing for continuous-index signals," *Proc. of Information Theory and Applications Workshop (ITA), San Diego, CA*, pp. 547–554, Jan. 2008.

[47] M. Mishali and Y. C. Eldar, "Blind multi-band signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Sig. Proc.*, vol. 57, no. 3, pp. 993–1009, Mar. 2009.

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

[49] P. Kuppinger, G. Durisi, and H. Bölcskei, "Where is randomness needed to break the square-root bottleneck?" in *Proc. of IEEE Int. Symp. on Inf. Theory (ISIT)*, Jun. 2010, pp. 1578–1582.

[50] G. Pope, A. Bracher, and C. Studer, "Probabilistic recovery guarantees for sparsely corrupted signals," *in preparation*.

[51] A. R. Calderbank, P. J. Cameron, W. M. Kantor, and J. J. Seidel, "Z_4-Kerdock codes, orthogonal spreads, and extremal Euclidean line-sets," *Proc. London Math. Soc. (3)*, vol. 75, no. 2, pp. 436–480, 1997.

[52] E. van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM J. Sci. Comput.*, vol. 31, no. 2, pp. 890–912, 2008. [Online]. Available: http://link.aip.org/link/?SCE/31/890

[53] ——, "SPGL1: A solver for large-scale sparse reconstruction," June 2007, http://www.cs.ubc.ca/labs/scl/spgl1.

[54] S. S. Agaian, *Hadamard matrices and their applications*, ser. Lecture notes in mathematics. Springer, 1985, vol. 1168.

[55] J. A. Tropp, "On the linear independence of spikes and sines," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 838–858, 2008.

[56] J. A. Tropp, I. S. Dhillon, R. W. Heath Jr., and T. Strohmer, "Designing structured tight frames via an alternating projection method," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 188–209, Jan. 2005.

[57] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Prentice Hall, 1984.

[58] S. G. Mallat, *A wavelet tour of signal processing*. San Diego, CA: Academic Press, 1998.

[59] C. Studer and R. G. Baraniuk, "Stable restoration and separation of approximately sparse signals," *submitted to Appl. Comput. Harm. Anal.*, July 2011.

[60] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY: Cambridge Press, 1985.