

# Sparse Signal Recovery from Sparsely Corrupted Measurements

Christoph Studer, Patrick Kuppinger, Graeme Pope, and Helmut Bölcskei

Dept. of Information Technology and Electrical Engineering, ETH Zurich, 8092 Zurich, Switzerland

Email: {studerc,patricku,gpoppe,boelcskei}@nari.ee.ethz.ch

**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 the recovery of signals that are corrupted by, e.g., clipping, impulse noise, or narrowband interference. We present deterministic recovery guarantees based on a recently developed uncertainty relation and provide corresponding 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 on the support sets of signal and noise.

## 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

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

where  $\mathbf{A} \in \mathbb{C}^{M \times N_a}$  and  $\mathbf{B} \in \mathbb{C}^{M \times N_b}$  are deterministic and general (i.e., not necessarily of the same cardinality and possibly redundant or incomplete) *dictionaries*, and  $\mathbf{e} \in \mathbb{C}^{N_b}$  is 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 numerous applications, some of which are described next:

- *Clipping*: Non-linearities in analog-to-digital converters often cause signal clipping, e.g., [2]. Specifically, 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  $[-a, +a]$ . 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  $\mathbf{e} = g_a(\mathbf{y}) - \mathbf{y}$ , we see that signal clipping is contained in the model (1). The nonzero entries of  $\mathbf{e}$  are those for which clipping occurred; the vector  $\mathbf{e}$  will therefore be sparse if the clipping level is chosen high enough. 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 or recovery of audio signals impaired by click/pop noise. The model in (1) incorporates such impairments by setting  $\mathbf{B} = \mathbf{I}_M$  and identifying  $\mathbf{e}$  with the impulse-noise vector.

- *Narrowband interference*: In many applications one is interested in recovering audio, video, or communication signals that are corrupted by narrowband interference. Such impairments typically exhibit a sparse representation in the frequency domain, which amounts to setting  $\mathbf{B} = \mathbf{F}_M$  in (1), where  $\mathbf{F}_M$  denotes the  $M \times M$  discrete Fourier transform matrix.
- *Super-resolution and inpainting*: Our framework also encompasses super-resolution [4] and inpainting [5] 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. The missing entries are accounted for by choosing the vector  $\mathbf{e}$  so that the entries of  $\mathbf{z} = \mathbf{y} + \mathbf{e}$ , corresponding to the missing entries in  $\mathbf{y}$ , are set to some (arbitrary) value. The missing entries of  $\mathbf{y}$  are then filled in by recovering  $\mathbf{x}$  from  $\mathbf{z}$  followed by computing  $\mathbf{y} = \mathbf{A}\mathbf{x}$ . Note that in both applications, the dictionary  $\mathbf{A}$  is typically redundant, i.e.,  $\mathbf{A}$  has more columns than rows, which demonstrates the need for recovery results that apply to redundant dictionaries.
- *Signal separation*: The decomposition of audio or video signals into two distinct components also fits into our framework. A prominent example is the separation of texture from cartoon parts in images [5]. In this application, one chooses the dictionaries  $\mathbf{A}$  and  $\mathbf{B}$  such that texture parts can be sparsely represented in  $\mathbf{A}$  and cartoon parts are sparsely represented in  $\mathbf{B}$ . Signal separation then amounts to extracting the sparse vectors  $\mathbf{x}$  and  $\mathbf{e}$ , simultaneously, from  $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$ , where  $\mathbf{z}$  represents the image to be decomposed;  $\mathbf{A}\mathbf{x}$  corresponds to the texture part, and  $\mathbf{B}\mathbf{e}$  represents the cartoon part.

Naturally, it is of significant interest to identify fundamental limits on the recovery of  $\mathbf{x}$  (and  $\mathbf{e}$ , in the case of signal separation) from  $\mathbf{z}$  in (1). For the noiseless case  $\mathbf{z} = \mathbf{A}\mathbf{x}$  such *recovery guarantees* are known [6], [7] and typically set limits on the maximum allowed number of nonzero entries of  $\mathbf{x}$ . For the case of unstructured noise, i.e.,  $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{n}$  with no constraints imposed on  $\mathbf{n}$  apart from  $\|\mathbf{n}\|_2 < \infty$ , recovery

guarantees were derived in [8]–[10].

*Contributions:* We derive conditions that guarantee *perfect* recovery of  $\mathbf{x}$  (and  $\mathbf{e}$ ) from the sparsely corrupted observation  $\mathbf{z}$ . Specifically, based on an uncertainty relation for pairs of general dictionaries presented in [1], we find 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 recovery algorithms and derive corresponding performance guarantees.

*Notation:* Lowercase boldface letters stand for column vectors and uppercase boldface letters designate matrices. For the matrix  $\mathbf{M}$ , we denote its transpose, conjugate transpose, and pseudo-inverse by  $\mathbf{M}^T$ ,  $\mathbf{M}^H$ , and  $\mathbf{M}^\dagger$ , respectively. The  $k$ th column of  $\mathbf{M}$  is designated by  $\mathbf{m}_k$  and the  $k$ th entry of  $\mathbf{m}$  is  $[\mathbf{m}]_k$ . The space spanned by the columns of  $\mathbf{M}$  is denoted by  $\mathcal{R}(\mathbf{M})$ . Throughout the paper, we assume that the columns of the dictionaries  $\mathbf{A}$  and  $\mathbf{B}$  have unit  $\ell_2$ -norm. The  $M \times M$  identity matrix is denoted by  $\mathbf{I}_M$ , the  $M \times N$  all zeros matrix is  $\mathbf{0}_{M,N}$ , and  $\mathbf{0}_M$  stands for the all-zeros vector of dimension  $M$ . The  $\ell_2$ -norm of the vector  $\mathbf{x}$  is denoted by  $\|\mathbf{x}\|_2$ ,  $\|\mathbf{x}\|_1$  stands for the  $\ell_1$ -norm of  $\mathbf{x}$ , and  $\|\mathbf{x}\|_0$  designates the number of nonzero entries in  $\mathbf{x}$ . Sets are designated by upper-case calligraphic letters; the cardinality of the set  $\mathcal{T}$  is  $|\mathcal{T}|$ . The complement of a set  $\mathcal{S}$  is denoted by  $\mathcal{S}^c$ . For two sets  $\mathcal{S}_1$  and  $\mathcal{S}_2$ ,  $s \in (\mathcal{S}_1 + \mathcal{S}_2)$  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 set of subsets of  $\mathcal{S}$  of cardinality less than or equal to  $n$  is denoted by  $\wp_n(\mathcal{S})$ . The support set of the vector  $\mathbf{m}$ , i.e., the index set corresponding to the nonzero entries of  $\mathbf{m}$ , is designated by  $\text{supp}(\mathbf{m})$ . The matrix  $\mathbf{M}_{\mathcal{T}}$  is obtained by retaining the columns of  $\mathbf{M}$  with indices in  $\mathcal{T}$ ; the vector  $\mathbf{m}_{\mathcal{T}}$  is obtained analogously from the vector  $\mathbf{m}$ . 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  $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$  corresponds to a sparse-signal recovery problem subject to *structured* (i.e., sparse) noise. In this section, we briefly summarize existing results for sparse-signal recovery from noiseless measurements, and we review recovery results available for unstructured and structured noise.

### A. Recovery in the Noiseless Case

If we do not impose additional assumptions on  $\mathbf{x}$ , the problem of recovering  $\mathbf{x}$  from the (noiseless) observation  $\mathbf{y} = \mathbf{A}\mathbf{x}$  with  $\mathbf{A}$  being redundant (i.e.,  $M < N_a$ ) is obviously ill-posed. However, assuming that  $\mathbf{x}$  is sparse changes the situation drastically. More specifically, one can recover  $\mathbf{x}$  by solving

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

This results, however, in prohibitive computational complexity, even for small problem sizes. Two of the most popular and computationally tractable alternatives to solving (P0) are basis pursuit (BP) [6] and orthogonal matching pursuit (OMP) [7]. BP amounts to solving

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

OMP is a greedy algorithm that iteratively constructs a sparse representation of  $\mathbf{y}$ . Recovery guarantees are usually expressed in terms of the sparsity level  $n_x = \|\mathbf{x}\|_0$  and the coherence of the dictionary  $\mathbf{A}$  defined as  $\mu_a = \max_{k,\ell,k \neq \ell} |\mathbf{a}_k^H \mathbf{a}_\ell|$ . As shown in [6], [7], a sufficient condition for  $\mathbf{x}$  to be the unique solution of (P0) applied to  $\mathbf{y} = \mathbf{A}\mathbf{x}$  and for BP and OMP to deliver this solution is

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

### B. Recovery in the Presence of Unstructured Noise

Recovery guarantees in the presence of unstructured (and deterministic) noise, i.e.,  $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{n}$  with no constraints imposed on  $\mathbf{n}$  apart from  $\|\mathbf{n}\|_2 < \infty$ , were derived in, e.g., [8], [9]. The corresponding results guarantee that a suitably modified version of BP recovers an  $\hat{\mathbf{x}}$  satisfying  $\|\mathbf{x} - \hat{\mathbf{x}}\|_2 < C\|\mathbf{n}\|_2$  provided that  $n_x < (1 + \mu_a^{-1})/4$ , where  $C > 0$  depends on  $\mu_a$  and  $n_x$ . Another result in [8] states that OMP delivers the correct support set (but does not perfectly recover the nonzero entries of  $\mathbf{x}$ ) provided that

$$n_x < \frac{1}{2} \left( 1 + \mu_a^{-1} \right) - \frac{\|\mathbf{n}\|_2}{\mu_a |x_{\min}|} \quad (3)$$

where  $|x_{\min}|$  denotes the absolute value of the component of  $\mathbf{x}$  with smallest nonzero magnitude. This recovery condition yields sensible results only if  $\|\mathbf{n}\|_2/|x_{\min}|$  is small. Recovery guarantees in the case of stochastic noise  $\mathbf{n}$  can be found in [9], [10]. We finally point out that *perfect* recovery of  $\mathbf{x}$  is, in general, impossible in the presence of *unstructured* noise.

### C. Recovery Guarantees in the Presence of Structured Noise

Special cases of the general setup (1) were considered in [2], [3], [11]–[14]. Specifically, in [11] it was shown that for  $\mathbf{A} = \mathbf{F}_M$  and  $\mathbf{B} = \mathbf{I}_M$ , *perfect* recovery of the  $M$ -dimensional vector  $\mathbf{x}$  is possible if  $2n_x n_e < M$ , where  $n_e = \|\mathbf{e}\|_0$ . We emphasize that this result assumes the support set of  $\mathbf{e}$  to be known (prior to recovery), an assumption that is often difficult to meet in practice. It is interesting to observe that the condition  $2n_x n_e < M$  is—in contrast to the recovery guarantee (3)—independent of the  $\ell_2$ -norm of the noise vector, i.e.,  $\|\mathbf{B}\mathbf{e}\|_2$  may, in principle, be arbitrarily large. We note that the recovery guarantees reported in [11] are based on an uncertainty relation that puts limits on how sparse a given signal can simultaneously be in the Fourier basis and in the identity basis. The present paper is heavily inspired by this observation.

In [2], [12]–[14] probabilistic recovery results for  $\mathbf{A}$  i.i.d. zero-mean Gaussian or a randomly sub-sampled unitary matrix and  $\mathbf{B}$  an orthonormal basis (ONB) were reported.

### III. RECOVERY OF SPARSELY CORRUPTED SIGNALS

Based on the uncertainty relation in [1, Thm. 1], we next derive conditions that guarantee perfect recovery of  $\mathbf{x}$  from the sparsely corrupted measurement  $\mathbf{z}$  under different assumptions on prior knowledge about the support sets of signal and noise. Specifically, these conditions depend on the number of nonzero entries of  $\mathbf{x}$  and  $\mathbf{e}$ , and on the coherence parameters  $\mu_a$  and  $\mu_b$  of the dictionaries  $\mathbf{A}$  and  $\mathbf{B}$ , respectively, and the mutual coherence  $\mu_m = \max_{k,\ell} |\mathbf{a}_k^H \mathbf{b}_\ell|$ . To simplify notation, we define the function

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

In the remainder of the paper,  $\mathcal{X}$  denotes  $\text{supp}(\mathbf{x})$  and  $\mathcal{E}$  stands for  $\text{supp}(\mathbf{e})$ . We furthermore assume that<sup>1</sup>  $\mu_m > 0$ .

#### 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, of course. This scenario is relevant, for example, in applications requiring recovery of clipped band-limited signals with known spectral support  $\mathcal{X}$  (i.e.,  $\mathbf{A} = \mathbf{F}_M$  and  $\mathbf{B} = \mathbf{I}_M$ ).

We first rewrite the input-output relation in (1) according to  $\mathbf{z} = \mathbf{D}_{\mathcal{X},\mathcal{E}} \mathbf{s}_{\mathcal{X},\mathcal{E}}$  with the concatenated dictionary  $\mathbf{D}_{\mathcal{X},\mathcal{E}} = [\mathbf{A}_{\mathcal{X}} \ \mathbf{B}_{\mathcal{E}}]$  and the stacked vector  $\mathbf{s}_{\mathcal{X},\mathcal{E}} = [\mathbf{x}_{\mathcal{X}}^T \ \mathbf{e}_{\mathcal{E}}^T]^T$ . It is now important to realize that we can recover the stacked vector  $\mathbf{s}_{\mathcal{X},\mathcal{E}}$  perfectly and hence the nonzero entries of both  $\mathbf{x}$  and  $\mathbf{e}$ , if  $\mathbf{D}_{\mathcal{X},\mathcal{E}}$  has full column rank and the pseudo-inverse is given by  $\mathbf{D}_{\mathcal{X},\mathcal{E}}^\dagger = (\mathbf{D}_{\mathcal{X},\mathcal{E}}^H \mathbf{D}_{\mathcal{X},\mathcal{E}})^{-1} \mathbf{D}_{\mathcal{X},\mathcal{E}}^H$ . In this case, we get  $\mathbf{s}_{\mathcal{X},\mathcal{E}}$  according to

$$\mathbf{s}_{\mathcal{X},\mathcal{E}} = \mathbf{D}_{\mathcal{X},\mathcal{E}}^\dagger \mathbf{z}. \quad (4)$$

The following theorem states a sufficient condition for  $\mathbf{D}_{\mathcal{X},\mathcal{E}}$  to have full column rank, and hence for the pseudo-inverse  $\mathbf{D}_{\mathcal{X},\mathcal{E}}^\dagger$  to exist. This condition depends on the coherence parameters  $\mu_a$ ,  $\mu_b$ , and  $\mu_m$ , of the involved dictionaries  $\mathbf{A}$  and  $\mathbf{B}$  and on  $\mathcal{X}$  and  $\mathcal{E}$  only through the cardinalities  $|\mathcal{X}|$  and  $|\mathcal{E}|$ , i.e., the number of nonzero entries in  $\mathbf{x}$  and  $\mathbf{e}$ , respectively.

*Theorem 1:* Let  $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$  with  $\mathcal{X} = \text{supp}(\mathbf{x})$  and  $\mathcal{E} = \text{supp}(\mathbf{e})$ . Define  $n_x = \|\mathbf{x}\|_0$  and  $n_e = \|\mathbf{e}\|_0$ . If

$$n_x n_e < f(n_x, n_e) \quad (5)$$

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.  $\blacksquare$

Consequently, if (5) holds, one can perfectly recover  $\mathbf{x}$  (and also  $\mathbf{e}$ ) from  $\mathbf{z}$  using (4).

#### B. $\mathcal{E}$ is Known and $\mathcal{X}$ is Unknown

Next, we find recovery guarantees for the case where  $\mathcal{E}$  is known and  $\mathcal{X}$  is unknown (prior to recovery). A prominent application for this setup is the recovery of clipped band-limited signals, where the spectral support of  $\mathbf{x}$  is unknown (i.e.,  $\mathbf{A} = \mathbf{F}_M$ ,  $\mathbf{B} = \mathbf{I}_M$ , and  $\mathcal{X}$  is unknown),  $\mathcal{E}$  is known and corresponds to the set of clipped entries. The case where  $\mathcal{X}$

<sup>1</sup>If  $\mu_m = 0$ , the recovery (separation) task is straightforward as  $\mathcal{R}(\mathbf{A})$  is orthogonal to  $\mathcal{R}(\mathbf{B})$ .

is known and  $\mathcal{E}$  is unknown can be treated similarly (see [1] for details).

The setting of  $\mathcal{E}$  known and  $\mathcal{X}$  unknown was considered previously in [11] for the special case  $\mathbf{A} = \mathbf{F}_M$  and  $\mathbf{B} = \mathbf{I}_M$ . The recovery condition (8) in Theorem 2 below extends the result in [11, Thms. 5 and 9] to pairs of general dictionaries  $\mathbf{A}$  and  $\mathbf{B}$ .

*Theorem 2:* Let  $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$  where  $\mathcal{E} = \text{supp}(\mathbf{e})$  is known. Consider the problem

$$(\text{P0}, \mathcal{E}) \begin{cases} \text{minimize} & \|\tilde{\mathbf{x}}\|_0 \\ \text{subject to} & \mathbf{A}\tilde{\mathbf{x}} \in (\{\mathbf{z}\} + \mathcal{R}(\mathbf{B}_{\mathcal{E}})) \end{cases} \quad (6)$$

and the convex program

$$(\text{BP}, \mathcal{E}) \begin{cases} \text{minimize} & \|\tilde{\mathbf{x}}\|_1 \\ \text{subject to} & \mathbf{A}\tilde{\mathbf{x}} \in (\{\mathbf{z}\} + \mathcal{R}(\mathbf{B}_{\mathcal{E}})). \end{cases} \quad (7)$$

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

$$2n_x n_e < f(2n_x, n_e) \quad (8)$$

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

*Proof:* See Appendix B.  $\blacksquare$

Solving (P0,  $\mathcal{E}$ ) requires a combinatorial search, which results in prohibitive computational complexity, even for moderate problem sizes. The convex relaxation (BP,  $\mathcal{E}$ ) can, however, often be solved more efficiently.

Rather than solving (P0,  $\mathcal{E}$ ) or (BP,  $\mathcal{E}$ ), we may compute the projection onto the orthogonal complement of  $\mathcal{R}(\mathbf{B}_{\mathcal{E}})$  according to:

$$\mathbf{R}_{\mathcal{E}} \mathbf{z} = \mathbf{R}_{\mathcal{E}} (\mathbf{A}\mathbf{x} + \mathbf{B}_{\mathcal{E}} \mathbf{e}_{\mathcal{E}}) = \mathbf{R}_{\mathcal{E}} \mathbf{A}\mathbf{x} \triangleq \hat{\mathbf{z}} \quad (9)$$

where  $\mathbf{R}_{\mathcal{E}} = \mathbf{I}_M - \mathbf{B}_{\mathcal{E}} \mathbf{B}_{\mathcal{E}}^\dagger$ . This leaves us with the standard problem of recovering  $\mathbf{x}$  from the modified measurement outcome  $\hat{\mathbf{z}} = \mathbf{R}_{\mathcal{E}} \mathbf{A}\mathbf{x}$ . Since the columns of  $\mathbf{R}_{\mathcal{E}} \mathbf{A}$  will, in general, not have unit  $\ell_2$ -norm, an assumption underlying the threshold in (2), we normalize the modified dictionary  $\mathbf{R}_{\mathcal{E}} \mathbf{A}$  by rewriting (9) as

$$\hat{\mathbf{z}} = \mathbf{R}_{\mathcal{E}} \mathbf{A} \mathbf{\Delta} \hat{\mathbf{x}} \quad (10)$$

where  $\mathbf{\Delta}$  is the diagonal matrix with elements  $[\mathbf{\Delta}]_{\ell,\ell} = 1/\|\mathbf{R}_{\mathcal{E}} \mathbf{a}_\ell\|_2$ ,  $\ell = 1, \dots, N_a$ , and  $\hat{\mathbf{x}} \triangleq \mathbf{\Delta}^{-1} \mathbf{x}$ . Now,  $\mathbf{R}_{\mathcal{E}} \mathbf{A} \mathbf{\Delta}$  plays the role of the dictionary (with normalized columns) and  $\hat{\mathbf{x}}$  is the unknown sparse vector that we wish to recover. Obviously,  $\text{supp}(\hat{\mathbf{x}}) = \text{supp}(\mathbf{x})$  and  $\mathbf{x}$  can be recovered from  $\hat{\mathbf{x}}$  by noting that  $\mathbf{x} = \mathbf{\Delta} \hat{\mathbf{x}}$ . The following theorem shows that (8) is sufficient to guarantee the following: i) The columns of  $\mathbf{B}_{\mathcal{E}}$  are linearly independent (guaranteeing the existence of  $\mathbf{B}_{\mathcal{E}}^\dagger$ ), ii)  $\|\mathbf{R}_{\mathcal{E}} \mathbf{a}_\ell\|_2 > 0$  for  $\ell = 1, \dots, N_a$ , and iii) no nonzero vector  $\mathbf{x}' \in \mathbb{C}^{N_a}$  satisfying  $\|\mathbf{x}'\|_0 \leq 2n_x$  lies in the kernel of  $\mathbf{R}_{\mathcal{E}} \mathbf{A}$ . Hence, (8) guarantees perfect recovery of  $\mathbf{x}$  from (10).

*Theorem 3:* If (8) is satisfied, the unique solution of (P0) applied to  $\hat{\mathbf{z}} = \mathbf{R}_{\mathcal{E}} \mathbf{A} \mathbf{\Delta} \hat{\mathbf{x}}$  is given by  $\hat{\mathbf{x}}$ . Furthermore, BP and OMP applied to  $\hat{\mathbf{z}} = \mathbf{R}_{\mathcal{E}} \mathbf{A} \mathbf{\Delta} \hat{\mathbf{x}}$  are guaranteed to recover the unique (P0)-solution.

*Proof:* The proof can be found in [1].  $\blacksquare$

For the case of  $\mathcal{X}$  known and  $\mathcal{E}$  unknown the recovery guarantee (8) is replaced by [1]

$$2n_x n_e < f(n_x, 2n_e). \quad (11)$$

### 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). A possible corresponding application scenario 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., limited by the acquisition bandwidth). 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 4:* 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

$$(P0, n_e) \begin{cases} \text{minimize} & \|\tilde{\mathbf{x}}\|_0 \\ \text{subject to} & \mathbf{A}\tilde{\mathbf{x}} \in \left( \{\mathbf{z}\} + \bigcup_{\mathcal{E}' \in \mathcal{P}} \mathcal{R}(\mathbf{B}_{\mathcal{E}'}) \right) \end{cases} \quad (12)$$

where  $\mathcal{P} = \wp_{n_e}(\{1, \dots, N_b\})$ . The unique solution of (P0,  $n_e$ ) applied to  $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$  is given by  $\mathbf{x}$  if

$$4n_x n_e < f(2n_x, 2n_e). \quad (13)$$

*Proof:* The proof is similar to that of Theorem 2. The corresponding details can be found in [1]. ■

We emphasize that (P0,  $n_e$ ) exhibits prohibitive computational complexity, in general. Unfortunately, replacing the  $\ell_0$ -norm of  $\tilde{\mathbf{x}}$  in the minimization in (12) by the  $\ell_1$ -norm does not lead to a computationally tractable algorithm, as the constraint  $\mathbf{A}\tilde{\mathbf{x}} \in (\{\mathbf{z}\} + \bigcup_{\mathcal{E}' \in \mathcal{P}} \mathcal{R}(\mathbf{B}_{\mathcal{E}'}))$  specifies a non-convex set, in general. Nonetheless, the recovery threshold in (13) 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 (see Section IV).

### D. Case IV: No Knowledge about the Support Sets

Finally, we consider the case where no knowledge about the support sets  $\mathcal{X}$  and  $\mathcal{E}$  is available (prior to recovery). A typical application scenario for this setting is the decomposition of images into two distinct features, i.e., into a part that exhibits a sparse representation in  $\mathbf{A}$  and another part exhibiting a sparse representation in  $\mathbf{B}$ .

Recovery guarantees for this case follow from results in [15] for the concatenation of general (possibly redundant or incomplete) dictionaries. This can be seen by writing (1) as  $\mathbf{z} = \mathbf{D}\mathbf{w}$  with  $\mathbf{D} = [\mathbf{A} \ \mathbf{B}]$  and  $\mathbf{w} = [\mathbf{x}^T \ \mathbf{e}^T]^T$ . In particular, it was shown in [15, Eq. 10] that the unique solution of (P0) applied to  $\mathbf{z} = \mathbf{D}\mathbf{w}$  is given by  $\mathbf{w}$  and, furthermore, in [15, Eq. 13] that this solution is delivered by BP and OMP applied to  $\mathbf{z} = \mathbf{D}\mathbf{w}$ , if the number of nonzero entries of  $\mathbf{w}$  is less than a corresponding sparsity threshold. These thresholds (i.e., [15, Eqs. 10 and 13]) are more restrictive than those in (5), (8), (11), and (13) (cf. Section IV), which reflects the fact that additional knowledge about the support sets  $\mathcal{X}$  and  $\mathcal{E}$  can only improve the recovery guarantees.

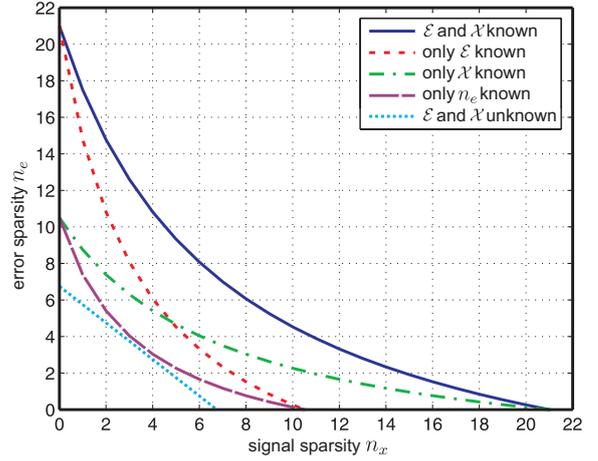


Figure 1. Recovery thresholds (5), (8), (11), (13), and [15, Eq. 13] for  $\mu_a = 0.05$ ,  $\mu_b = 0.05$ , and  $\mu_m = 0.1$ .

## IV. THE FACTOR-OF-TWO PENALTY

Comparing the recovery thresholds (5), (8), (11), and (13) (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. For example, consider the recovery thresholds (5) and (8). For given  $n_e \in [0, 1 + 1/\mu_b]$ , solving (5) 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, solving (8) 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 twice as many nonzero entries in  $\mathbf{x}$  compared to the case where  $\mathcal{X}$  is unknown. This factor-of-two penalty has the same roots as the well-known factor-of-two penalty in spectrum-blind sampling [16]–[18].

The factor-of-two penalty is illustrated in Fig. 1, where we plot the recovery thresholds (5), (8), (11), (13), and (for completeness) [15, Eq. 13] (which guarantees that  $\mathbf{w}$  is recovered through BP and OMP in Case IV without knowledge about  $\mathcal{X}$  or  $\mathcal{E}$ ) for  $\mu_a = \mu_b = 0.05$  and  $\mu_m = 0.1$ . In particular, we can see that for a fixed error sparsity level, e.g.,  $n_e = 6$ , for the case where both  $\mathcal{X}$  and  $\mathcal{E}$  are known the threshold evaluates to  $n_x < 8$ ; when only  $\mathcal{E}$  is known we have  $n_x < 4$ . For a fixed signal sparsity level, e.g.,  $n_x = 4$ , we can furthermore see that in the case where only  $n_e$  is known we can tolerate only  $n_e < 3$  errors, compared to  $n_e < 6$  in the case when  $\mathcal{E}$  is known but  $\mathcal{X}$  is unknown. Finally, for Case IV where no knowledge about the support sets is available, the recovery threshold is worse than in the case where  $n_e$  is known.

## APPENDIX A

### PROOF OF THEOREM 1

We prove the full column-rank property of  $\mathbf{D}_{\mathcal{X}, \mathcal{E}}$  by showing that if (5) is satisfied, there is a unique pair  $(\mathbf{x}, \mathbf{e})$  with  $\text{supp}(\mathbf{x}) = \mathcal{X}$  and  $\text{supp}(\mathbf{e}) = \mathcal{E}$  such that  $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$ . Assume that there exists an alternative pair  $(\mathbf{x}', \mathbf{e}')$  such that

$\mathbf{z} = \mathbf{A}\mathbf{x}' + \mathbf{B}\mathbf{e}'$  with  $\text{supp}(\mathbf{x}') \subseteq \mathcal{X}$  and  $\text{supp}(\mathbf{e}') \subseteq \mathcal{E}$  (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  $\mathbf{A}(\mathbf{x} - \mathbf{x}') = \mathbf{B}(\mathbf{e}' - \mathbf{e})$ . Since both  $\mathbf{x}$  and  $\mathbf{x}'$  are supported in  $\mathcal{X}$  it follows that  $\mathbf{x} - \mathbf{x}'$  is also supported 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  $\mathcal{P} = \text{supp}(\mathbf{x} - \mathbf{x}') \subseteq \mathcal{X}$ , and, similarly,  $\mathbf{q} = \mathbf{e}' - \mathbf{e}$  and  $\mathcal{Q} = \text{supp}(\mathbf{e}' - \mathbf{e}) \subseteq \mathcal{E}$ , we obtain the following chain of inequalities:

$$\begin{aligned} n_x n_e &\geq \|\mathbf{p}\|_0 \|\mathbf{q}\|_0 = |\mathcal{P}| |\mathcal{Q}| \\ &\geq [1 - \mu_a(|\mathcal{P}| - 1)]^+ [1 - \mu_b(|\mathcal{Q}| - 1)]^+ / \mu_m^2 \quad (14) \\ &\geq [1 - \mu_a(n_x - 1)]^+ [1 - \mu_b(n_e - 1)]^+ / \mu_m^2 \quad (15) \end{aligned}$$

where (14) follows from the uncertainty relation [1, Cor. 2] and (15) is a consequence of  $|\mathcal{P}| \leq n_x$  and  $|\mathcal{Q}| \leq n_e$ . Obviously, (15) contradicts the assumption in (5), which completes the proof.

## APPENDIX B PROOF OF THEOREM 2

We begin by proving that  $\mathbf{x}$  is the unique solution of  $(\text{P0}, \mathcal{E})$  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  $\text{supp}(\mathbf{e}') \subseteq \mathcal{E}$ , such that  $\mathbf{A}(\mathbf{x} - \mathbf{x}') = \mathbf{B}(\mathbf{e}' - \mathbf{e})$ . Since  $\text{supp}(\mathbf{e}) = \mathcal{E}$  and  $\text{supp}(\mathbf{e}') \subseteq \mathcal{E}$ , we have  $\text{supp}(\mathbf{e}' - \mathbf{e}) \subseteq \mathcal{E}$  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 (at possibly different positions), we have  $\|\mathbf{x} - \mathbf{x}'\|_0 \leq 2n_x$ . Defining  $\mathbf{p} = \mathbf{x} - \mathbf{x}'$  and  $\mathcal{P} = \text{supp}(\mathbf{x} - \mathbf{x}')$ , and, similarly,  $\mathbf{q} = \mathbf{e}' - \mathbf{e}$  and  $\mathcal{Q} = \text{supp}(\mathbf{e}' - \mathbf{e}) \subseteq \mathcal{E}$ , we obtain the following chain of inequalities

$$\begin{aligned} 2n_x n_e &\geq \|\mathbf{p}\|_0 \|\mathbf{q}\|_0 = |\mathcal{P}| |\mathcal{Q}| \\ &\geq [1 - \mu_a(|\mathcal{P}| - 1)]^+ [1 - \mu_b(|\mathcal{Q}| - 1)]^+ / \mu_m^2 \quad (16) \\ &\geq [1 - \mu_a(2n_x - 1)]^+ [1 - \mu_b(n_e - 1)]^+ / \mu_m^2 \quad (17) \end{aligned}$$

where (16) follows from the uncertainty relation [1, Cor. 2] and (17) is a consequence of  $|\mathcal{P}| \leq 2n_x$  and  $|\mathcal{Q}| \leq n_e$ . Obviously, (17) contradicts the assumption in (8), which concludes the first part of the proof.

We next prove that  $\mathbf{x}$  is also the unique solution of  $(\text{BP}, \mathcal{E})$  applied to  $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$ . Assume that there exists an alternative vector  $\mathbf{x}'$  satisfying  $\mathbf{A}\mathbf{x}' \in (\{\mathbf{z}\} + \mathcal{R}(\mathbf{B}_\mathcal{E}))$  with  $\|\mathbf{x}'\|_1 \leq \|\mathbf{x}\|_1$ ; this would imply the existence of a vector  $\mathbf{e}'$  with  $\text{supp}(\mathbf{e}') \subseteq \mathcal{E}$ , such that  $\mathbf{A}(\mathbf{x} - \mathbf{x}') = \mathbf{B}(\mathbf{e}' - \mathbf{e})$ . Defining  $\mathbf{p} = \mathbf{x} - \mathbf{x}'$  and  $\|\mathbf{P}_\mathcal{S}\mathbf{m}\|_1 = \sum_{i \in \mathcal{S}} |\mathbf{m}_i|$  for some vector  $\mathbf{m}$ , we obtain the following lower bound on the  $\ell_1$ -norm of  $\mathbf{x}'$ :

$$\begin{aligned} \|\mathbf{x}'\|_1 &= \|\mathbf{x} - \mathbf{p}\|_1 = \|\mathbf{P}_\mathcal{X}(\mathbf{x} - \mathbf{p})\|_1 + \|\mathbf{P}_{\mathcal{X}^c}\mathbf{p}\|_1 \\ &\geq \|\mathbf{P}_\mathcal{X}\mathbf{x}\|_1 - \|\mathbf{P}_\mathcal{X}\mathbf{p}\|_1 + \|\mathbf{P}_{\mathcal{X}^c}\mathbf{p}\|_1 \quad (18) \\ &= \|\mathbf{x}\|_1 - \|\mathbf{P}_\mathcal{X}\mathbf{p}\|_1 + \|\mathbf{P}_{\mathcal{X}^c}\mathbf{p}\|_1 \end{aligned}$$

where (18) is a consequence of the reverse triangle inequality. Now, the  $\ell_1$ -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  $\|\mathbf{P}_\mathcal{X}\mathbf{p}\|_1 \geq 0.5\|\mathbf{p}\|_1$ . Defining  $\mathbf{q} = \mathbf{e}' - \mathbf{e}$  and  $\mathcal{Q} =$

$\text{supp}(\mathbf{e}' - \mathbf{e})$ , and noting that  $\text{supp}(\mathbf{e}) = \mathcal{E}$  and  $\text{supp}(\mathbf{e}') \subseteq \mathcal{E}$ , it follows that  $|\mathcal{Q}| \leq n_e$ . Setting  $\mathcal{P} = \mathcal{X}$  (with  $|\mathcal{X}| = n_x$ ), we obtain the following chain of inequalities:

$$\begin{aligned} n_x n_e &\geq |\mathcal{P}| |\mathcal{Q}| \\ &\geq \frac{[(1 + \mu_a)(1 - \epsilon_\mathcal{P}) - |\mathcal{P}| \mu_a]^+ [1 - \mu_b(|\mathcal{Q}| - 1)]^+}{\mu_m^2} \quad (19) \\ &\geq [1 - \mu_a(2n_x - 1)]^+ [1 - \mu_b(n_e - 1)]^+ / (2\mu_m^2) \quad (20) \end{aligned}$$

where (19) follows from the uncertainty relation [1, Thm. 1] applied to the difference vectors  $\mathbf{p}$  and  $\mathbf{q}$ , and by using the fact that  $\|\mathbf{P}_\mathcal{X}\mathbf{p}\|_1 \geq 0.5\|\mathbf{p}\|_1$ . Furthermore, (20) is a consequence of  $|\mathcal{P}| = n_x$  and  $|\mathcal{Q}| \leq n_e$ . Rewriting (20), we obtain

$$2n_x n_e \geq [1 - \mu_a(2n_x - 1)]^+ [1 - \mu_b(n_e - 1)]^+ / \mu_m^2. \quad (21)$$

Since (21) contradicts the assumption in (8), this proves that  $\mathbf{x}$  is the unique solution of  $(\text{BP}, \mathcal{E})$  applied to  $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$ .

## REFERENCES

- [1] C. Studer, P. Kuppinger, G. Pope, and H. Bölcskei, "Recovery of sparsely corrupted signals," *IEEE Trans. Inf. Theory*, Feb. 2011, submitted.
- [2] J. N. Laska, P. Boufounos, M. A. Davenport, and R. G. Baraniuk, "Democracy in action: Quantization, saturation, and compressive sensing," *preprint*, 2009.
- [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] 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.
- [5] 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.
- [6] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via  $\ell_1$  minimization," *Proc. Natl. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.
- [7] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] J. N. Laska, M. A. Davenport, and R. G. Baraniuk, "Exact signal recovery from sparsely corrupted measurements through the pursuit of justice," *Proc. of 43rd Asilomar Conf. on Signals, Systems, and Comput., Pacific Grove, CA, USA*, pp. 1556–1560, Nov. 2009.
- [13] N. H. Nguyen and T. D. Tran, "Exact recoverability from dense corrupted observations via  $\ell_1$  minimization," Feb. 2011. [Online]. Available: <http://arxiv.org/abs/1102.1227v1>
- [14] J. Wright and Y. Ma, "Dense error correction via  $\ell_1$ -minimization," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3540–3560, Jul. 2010.
- [15] P. Kuppinger, G. Durisi, and H. Bölcskei, "Uncertainty relations and sparse signal recovery for pairs of general signal sets," *IEEE Trans. Inf. Theory*, Feb. 2011, submitted.
- [16] P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," *Proc. of IEEE Int. Conf. Acoust. Speech Sig. Proc. (ICASSP)*, vol. 3, pp. 1689–1692, Atlanta, GA, May 1996.
- [17] 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.
- [18] 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.