Recovery Guarantees for Restoration and Separation of Approximately Sparse Signals

Christoph Studer and Richard G. Baraniuk

Dept. Electrical and Computer Engineering, Rice University
Houston, TX, USA; e-mail: {studer,richb}@rice.edu

***Abstract*—In this paper, we present performance guarantees for the recovery and separation of signals that are approximately sparse in some general (i.e., basis, frame, over-complete, or in-complete) dictionary but corrupted by a combination of measurement noise and interference that is sparse in a second general dictionary. Applications covered by this framework include the restoration of signals impaired by impulse noise, narrowband interference, or saturation, as well as image in-painting, super-resolution, and signal separation. We develop computationally efficient algorithms for signal restoration and signal separation and present deterministic conditions that guarantee their stability. A simple in-painting example demonstrates the efficacy of our approach.**

I. Introduction

We investigate the *restoration* of the vector $\mathbf{x} \in \mathbb{C}^{N_a}$ from the corrupted M-dimensional observations

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

where $\mathbf{A} \in \mathbb{C}^{M\times N_a}$ and $\mathbf{B} \in \mathbb{C}^{M\times N_b}$ are general (basis, frame, over-complete, or in-complete) deterministic dictionaries, i.e., matrices whose columns have unit Euclidean norm. The vector $\mathbf{x}$ is assumed to be *approximately sparse*, i.e., its main energy (in terms of the sum of absolute values) is concentrated in only a few entries; the M-dimensional signal vector is defined as $\mathbf{y} = \mathbf{A}\mathbf{x}$. The vector $\mathbf{e} \in \mathbb{C}^{N_a}$ represents interference and is assumed to be *perfectly sparse*, i.e., only a few entries are nonzero. Furthermore, the locations of the nonzero coefficients are assumed to be known prior to recovery. The vector $\mathbf{n} \in \mathbb{C}^{M}$ is used to model measurement noise, which is, apart from $\|\mathbf{n}\|_2 < \varepsilon$, allowed to be arbitrary.

In addition to restoration, the setting (1) also allows us to study *signal separation*, i.e., the separation of two distinct features $\mathbf{A}\mathbf{x}$ and $\mathbf{B}\mathbf{e}$ from the noisy observation $\mathbf{z}$. Here, the vector $\mathbf{e}$ in (1) is used to represent a second desirable feature (rather than undesired interference) and is also allowed to be approximately sparse. Furthermore, no knowledge on the locations of the dominant coefficients in $\mathbf{e}$ is required in this scenario. Signal separation amounts to simultaneously recovering the vectors $\mathbf{x}$ and $\mathbf{e}$ from the noisy measurement $\mathbf{z}$ followed by computation of the individual signal features $\mathbf{A}\mathbf{x}$ and $\mathbf{B}\mathbf{e}$.

Both the recovery and separation problems outlined above feature prominently in numerous applications (see [1], [2] and the references therein), including the restoration of signals from impulse noise [3], narrowband interference, and saturation [4], as well as super-resolution [5], in-painting [6]–[8], and signal separation [9], [10]. In almost all these applications, a predetermined (and possibly optimized) dictionary pair $\mathbf{A}$ and $\mathbf{B}$ is used. It is therefore of significant practical interest to identify fundamental limits on the performance of restoration or separation from the model (1) for the *deterministic setting*, i.e., assuming no randomness in the dictionaries, the signal, interference, or the noise vector. Deterministic recovery guarantees for the special case of *perfectly* sparse vectors $\mathbf{x}$ and $\mathbf{e}$ and *no* measurement noise have been studied in [2], [11]. The corresponding algorithms and proof techniques, however, cannot be adapted for the general (and practically more relevant) setting formulated in (1), which features approximately sparse signals and additive measurement noise.

A. Contributions

In this paper, we generalize the recovery guarantees of [2], [11] to the framework detailed above. In particular, we provide novel, computationally efficient signal restoration and separation algorithms, and derive corresponding recovery guarantees for the deterministic setting. Our guarantees depend in a natural way on the number of dominant nonzero entries of $\mathbf{x}$ and $\mathbf{e}$, on the *coherence parameters* of the dictionaries $\mathbf{A}$ and $\mathbf{B}$, and on the Euclidean norm of the measurement noise. We provide a comparison to the recovery conditions for perfectly sparse signals and noiseless measurements presented in [2], [11] and, finally, we demonstrate the efficacy of our approach for simultaneous attenuation of Gaussian noise and removal of scratches from old photographs.

B. Notation

Lowercase and uppercase boldface letters stand for column vectors and matrices, respectively. The transpose, conjugate transpose, and (Moore–Penrose) pseudo-inverse of the matrix $\mathbf{M}$ are denoted by $\mathbf{M}^T$, $\mathbf{M}^H$, and $\mathbf{M}^\dagger = \left(\mathbf{M}^H\mathbf{M}\right)^{-1}\mathbf{M}^H$, respectively. The kth entry of the vector $\mathbf{m}$ is $[\mathbf{m}]_k$, and the

An extended version of this paper was submitted to *Applied and Computational Harmonic Analysis* [1]. The authors would like to thank C. Aubel, H. Bölcskei, P. Kuppinger, and G. Pope for inspiring discussions. This work was supported by the Swiss National Science Foundation (SNSF) under Grant PA00P2-134155 and by the Grants NSF CCF-0431150, CCF-0728867, CCF-0926127, DARPA/ONR N66001-08-1-2065, N66001-11-1-4090, N66001-11-C-4092, ONR N00014-08-1-1112, N00014-10-1-0989, AFOSR FA9550-09-1-0432, ARO MURIs W911NF-07-1-0185 and W911NF-09-1-0383, and by the Texas Instruments Leadership University Program.

kth column of $\mathbf{M}$ is $\mathbf{m}_k$ and the entry in the kth row and ℓth column is designated by $[\mathbf{M}]_{k,\ell}$. The $M \times M$ identity matrix is denoted by $\mathbf{I}_M$ and the $M \times N$ all zeros matrix by $\mathbf{0}_{M\times N}$. 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 of $\mathbf{x}$. The spectral norm of the matrix $\mathbf{M}$ is $\|\mathbf{M}\|_2 = \sqrt{\lambda_{\max}(\mathbf{M}^H\mathbf{M})}$, where the minimum and maximum eigenvalue of a positive-semidefinite matrix $\mathbf{M}$ are denoted by $\lambda_{\min}(\mathbf{M})$ and $\lambda_{\max}(\mathbf{M})$, respectively. $\|\mathbf{M}\|_F$ stands for the Frobenius matrix norm. Sets are designated by upper-case calligraphic letters. The cardinality of the set $\mathcal{T}$ is $|\mathcal{T}|$ and the complement of a set $\mathcal{S}$ in some superset $\mathcal{T}$ is denoted by $\mathcal{S}^c$. The support set of the vector $\mathbf{m}$, i.e., the index set corresponding to the nonzero entries of $\mathbf{m}$, is designated by $\operatorname{supp}(\mathbf{m})$. We define the $M \times M$ diagonal (projection) matrix $\mathbf{P}_\mathcal{S}$ for the set $\mathcal{S} \subseteq \{1, \ldots, M\}$ as follows:

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

and $\mathbf{m}_\mathcal{T} = \mathbf{P}_\mathcal{T}\mathbf{m}$. The matrix $\mathbf{M}_\mathcal{T}$ is obtained from $\mathbf{M}$ by retaining the columns of $\mathbf{M}$ with indices in $\mathcal{T}$ and the $|\mathcal{T}|$-dimensional vector $[\mathbf{m}]_\mathcal{T}$ is obtained analogously. For $x \in \mathbb{R}$, we set $[x]^+ = \max\{x, 0\}$.

C. Outline of the paper

The remainder of the paper is organized as follows. In Section II, we summarize the relevant prior art. Our restoration and separation algorithms, the corresponding recovery guarantees, and a comparison to the results in [2], [11] are presented in Section III. We shown an application example in Section IV and conclude in Section V.

II. Prior Art

Recovery of a vector $\mathbf{x} \in \mathbb{C}^{N_a}$ from the noiseless observations $\mathbf{y} = \mathbf{A}\mathbf{x}$ with $\mathbf{A}$ over-complete (i.e., $M < N_a$) is well-known to be ill-posed. However, assuming that $\mathbf{x}$ is perfectly sparse (i.e., that only small number of its entries are nonzero) enables us to uniquely recover $\mathbf{x}$ by solving

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

which exhibits a prohibitive computational complexity, even for small dimensions N_a. One of the most popular and computationally tractable alternative to solving P0 is basis pursuit (BP) [12]–[17], which corresponds to the convex program

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

Recovery guarantees for P0 and BP are usually expressed in terms of the sparsity level $n_x = \|\mathbf{x}\|_0$ and the coherence parameter of the dictionary $\mathbf{A}$, which is defined as

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

A sufficient condition for $\mathbf{x}$ to be the unique solution of P0 and for BP to deliver this solution is [14], [15], [17]

$$n_x < 1/2\,(1 + 1/\mu_a)\,. \tag{2}$$

A. Recovery guarantees for approximately sparse signals from noisy observations

For the case of bounded (otherwise arbitrary) measurement noise, i.e., $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{n}$ with $\|\mathbf{n}\|_2 \leq \varepsilon$, recovery guarantees based on the coherence parameter μ_a were developed in [18]–[22]. In practice, most signals $\mathbf{x}$ are not perfectly-sparse (only a small fraction of the entries $\mathbf{x}$ are nonzero), but rather *approximately sparse*, i.e., most of the signal's energy (in terms of the sum of absolute values) is concentrated in only a few entries. For such signals, the support set associated to the best n_x-sparse approximation (in ℓ_1-norm) corresponds to

$$\widehat{\mathcal{X}} = \operatorname{supp}_{n_x}(\mathbf{x}) = \arg\min_{\widetilde{\mathcal{X}} \in \Sigma_{n_x}} \left\|\mathbf{x} - \mathbf{x}_{\widetilde{\mathcal{X}}}\right\|_1,$$

where Σ_{n_x} contains all support sets of size n_x corresponding to perfectly n_x-sparse vectors having the dimension of $\mathbf{x}$.

The following theorem provides a sufficient condition for which a suitably modified version of BP, known as BP denoising (BPDN) [12], [22], stably recovers an approximately sparse vector $\mathbf{x}$ from the noisy observation $\mathbf{z}$.

Theorem 1 ([22, Thm. 2.1]): Let $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{n}$, $\|\mathbf{n}\|_2 \leq \varepsilon$, and $\mathcal{X} = \operatorname{supp}_{n_x}(\mathbf{x})$. If (2) is met, then the solution $\hat{\mathbf{x}}$ of

$$\text{(BPDN)} \quad \text{minimize } \|\tilde{\mathbf{x}}\|_1 \quad \text{subject to } \|\mathbf{z} - \mathbf{A}\tilde{\mathbf{x}}\|_2 \leq \eta$$

with $\varepsilon \leq \eta$ satisfies

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq C_0(\varepsilon + \eta) + C_1\|\mathbf{x} - \mathbf{x}_\mathcal{X}\|_1\,, \tag{3}$$

where both constants $C_0, C_1 > 0$ depend on μ_a and n_x.

Proof: The proof in [22] is detailed for perfectly sparse vectors only. The general case for approximately sparse signals and measurement noise can be found in [1]. ■

Theorem 1 generalizes the results for noiseless measurements and perfectly sparse signals in [14], [15], [17] using BP. Specifically, for $\|\mathbf{n}\|_2 = 0$ and $\|\mathbf{x} - \mathbf{x}_\mathcal{X}\|_1 = 0$, BPDN with $\eta = 0$ corresponds to BP and (3) results in $\|\mathbf{x} - \hat{\mathbf{x}}\|_2 = 0$, which ensures perfect recovery of the vector $\mathbf{x}$ whenever (2) is met. We emphasize that *perfect* recovery of $\mathbf{x}$ in the presence of bounded (but otherwise arbitrary) measurement noise $\mathbf{n}$, is impossible, in general. We therefore consider *stable recovery* instead, i.e., in a sense that the ℓ_2-norm of the difference between the estimate $\hat{\mathbf{x}}$ and the true $\mathbf{x}$ is bounded from above by η and the best n_x-sparse approximation as in (3).

B. Recovery guarantees for perfectly sparse signals from sparsely corrupted and noiseless measurements

A large number of restoration and separation problems occurring in practice can be formulated as sparse signal recovery from sparsely corrupted signals using the model (1).

1) Probabilistic recovery guarantees: Recovery guarantees for the probabilistic setting (i.e., recovery of $\mathbf{x}$ is guaranteed with high probability) for random Gaussian matrices, which are of particular interest for applications based on compressive sensing (CS), were reported in, e.g., [23]–[26] and the references therein. In the remainder of the paper, however, we will consider the deterministic setting exclusively.

2) Deterministic recovery guarantees: Recovery guarantees in the deterministic setting for noiseless measurements and signals being perfectly sparse were studied in [2], [11], [27]. In [27], it was shown that when $\mathbf{A}$ is the Fourier matrix, $\mathbf{B} = \mathbf{I}_M$ and when the support set of the interference $\mathbf{e}$ is known, perfect recovery of $\mathbf{x}$ is possible if $2n_x n_e < M$, where $n_e = \|\mathbf{e}\|_0$. The case of $\mathbf{A}$ and $\mathbf{B}$ being arbitrary dictionaries was studied for different cases of support-set knowledge in [2], [11]. The presented recovery guarantees depend upon the number of nonzero entries n_x and n_e in the perfectly sparse vectors $\mathbf{x}$ and $\mathbf{e}$, respectively, and on the coherence parameters μ_a and μ_b of $\mathbf{A}$ and $\mathbf{B}$, as well as on the mutual coherence between the dictionaries $\mathbf{A}$ and $\mathbf{B}$, which is defined as

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

For the case of signal restoration, i.e., where the support set of $\mathbf{e}$ is known prior to recovery, the recovery guarantee in [2, Thms. 4 and 5] states that if

$$2n_x n_e \mu_m^2 < f(2n_x, n_e) \tag{4}$$

with the definition

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

is satisfied, then perfect recovery of $\mathbf{x}$ from $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}$ is guaranteed. An equivalent recovery condition for the case of signal separation can be found in [11, Thm. 3]. We emphasize that the results presented in [2], [11] are for perfectly sparse and noiseless measurements only, and furthermore, the algorithms and proof techniques cannot be adapted for the more general setting proposed in (1). In order to gain insight into the practically more relevant case of approximately sparse signals and noisy measurements, we next develop and analyze novel restoration and separation algorithms.

III. Recovery Guarantees

We next develop two computationally efficient methods for restoration or separation under the model (1) and derive corresponding recovery conditions that guarantee their stability.

A. BP restoration: Support-set knowledge of $\mathbf{e}$ only

A prominent application for the setting where the support set of $\mathbf{e}$ is known prior to recovery, is the restoration of saturated signals [4]. Here, $\mathbf{A}$ is chosen to sparsify the audio signal $\mathbf{y}$ (e.g., using the Fourier matrix) and $\mathbf{B} = \mathbf{I}_M$, and no knowledge on the locations of the dominant entries of $\mathbf{x}$ is required. The support set $\mathcal{E}$ of $\mathbf{e}$ can, however, be easily identified by comparing the measured signal entries $[\mathbf{z}]_i$, $i = 1, \ldots, M$, to a saturation threshold. Further application examples for this setting include the removal of impulse noise [3], in-painting [6]–[8], and super-resolution [5] of signals admitting an approximately sparse representation in some carefully-chosen dictionary $\mathbf{A}$.

The recovery procedure for this case is as follows. Since the support set $\mathcal{E}$ of the interference vector $\mathbf{e}$ is known prior to recovery, we may recover the vector $\mathbf{x}$ by projecting the noisy observation $\mathbf{z}$ onto the orthogonal complement of the range space spanned by $\mathbf{B}_\mathcal{E}$. This projection eliminates the sparse noise and leaves us with a sparse signal recovery problem similar to that in Theorem 1. In particular, we consider recovery from

$$\mathbf{R}_\mathcal{E}\mathbf{z} = \mathbf{R}_\mathcal{E}(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}_\mathcal{E} + \mathbf{n}) = \mathbf{R}_\mathcal{E}\mathbf{A}\mathbf{x} + \mathbf{R}_\mathcal{E}\mathbf{n}, \tag{5}$$

where $\mathbf{R}_\mathcal{E} = \mathbf{I}_M - \mathbf{B}_\mathcal{E}\mathbf{B}_\mathcal{E}^\dagger$, and we used the fact that $\mathbf{R}_\mathcal{E}\mathbf{B}\mathbf{e}_\mathcal{E} = \mathbf{0}_{M\times 1}$. The following theorem provides a sufficient condition that guarantees the stable restoration of the vector $\mathbf{x}$ from (5).

Theorem 2 (BP restoration): Let $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e} + \mathbf{n}$ with $\|\mathbf{n}\|_2 \le \varepsilon$. Assume $\mathbf{e}$ to be perfectly n_e-sparse and $\mathcal{E} = \operatorname{supp}(\mathbf{e})$ to be known prior to recovery. Furthermore, let $\mathcal{X} = \operatorname{supp}_{n_x}(\mathbf{x})$. If

$$2n_x n_e \mu_m^2 < f(2n_x, n_e), \tag{6}$$

then the result $\hat{\mathbf{x}}$ of BP restoration

$$\text{(BP-RES)} \quad \begin{cases} \text{minimize} & \|\tilde{\mathbf{x}}\|_1 \\ \text{subject to} & \|\mathbf{R}_\mathcal{E}(\mathbf{z} - \mathbf{A}\tilde{\mathbf{x}})\|_2 \le \eta \end{cases}$$

with $\mathbf{R}_\mathcal{E} = \mathbf{I}_M - \mathbf{B}_\mathcal{E}\mathbf{B}_\mathcal{E}^\dagger$ and $\varepsilon \le \eta$ satisfies

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \le C_5(\varepsilon + \eta) + C_6\|\mathbf{x} - \mathbf{x}_\mathcal{X}\|_1,$$

where the (non-negative) constants C_5 and C_6 depend on μ_a, μ_b, μ_m, n_x and n_e.

Proof: The proof is given in Appendix A. ■

We note that (6) provides a sufficient condition on the number n_x of dominant entries of $\mathbf{x}$, for which BP-RES can stably recover $\mathbf{x}$ from $\mathbf{z}$. The situation guaranteeing that the largest number n_x of dominant coefficients in $\mathbf{x}$ will be recovered stably using BP-RES, is when $\mathbf{A}$ and $\mathbf{B}$ are maximally incoherent orthonormal bases (ONBs). In this case, the recovery condition reduces to $2n_x n_e < M$. Furthermore, (6) turns out to be equivalent to the condition (4) for perfectly-sparse signals and noiseless measurement provided in [2, Thms. 4 and 5]. Hence, generalizing the recovery procedure to approximately sparse signals and measurement noise does not incur a penalty in terms of the recovery condition.

B. BP separation: No knowledge of the support sets

A typical application scenario for no knowledge on the support sets is signal separation [9], [10], e.g., the decomposition of audio, image, or video signals into two or more distinct features, i.e., in a part that exhibits an approximately sparse representation in the dictionary $\mathbf{A}$ and another part that exhibits an approximately sparse representation in $\mathbf{B}$. Decomposition then amounts to performing simultaneous recovery of $\mathbf{x}$ and $\mathbf{e}$ from (1), followed by computation of the individual signal features according to $\mathbf{A}\mathbf{x}$ and $\mathbf{B}\mathbf{e}$. The idea underlying the signal-separation approach studied here is to rewrite (1) as

$$\mathbf{z} = \mathbf{D}\mathbf{w} + \mathbf{n} \tag{7}$$

where $\mathbf{D} = [\,\mathbf{A}\ \mathbf{B}\,]$ is the concatenated dictionary of $\mathbf{A}$ and $\mathbf{B}$ and the stacked vector $\mathbf{w} = [\,\mathbf{x}^T\ \mathbf{e}^T\,]^T$. Signal separation now amounts to performing BPDN on (7) for recovery of $\mathbf{w}$ from $\mathbf{z}$, which is also known as the synthesis separation problem (see, e.g., [10] and the references therein).

A straightforward way to arrive at a corresponding recovery guarantee for this problem is to consider $\mathbf{D}$ as the new dictionary with the *dictionary coherence* defined as

$$\mu_d = \max_{i,j,i\neq j} \left|\mathbf{d}_i^H \mathbf{d}_j\right| = \max\left\{\mu_a, \mu_b, \mu_m\right\}. \tag{8}$$

One can now use BPDN to recover $\mathbf{w}$ from (7) and invoke Theorem 1 with the recovery condition in (2), resulting in

$$w = n_x + n_e < 1/2\,(1 + 1/\mu_d)\,. \tag{9}$$

It is, however, important to realize that (9) ignores the structure underlying the dictionary $\mathbf{D}$, i.e., it does not take into account the fact that $\mathbf{D}$ is a concatenation of two dictionaries that are characterized by the coherence parameters μ_a, μ_b, and μ_m. Hence, the recovery guarantee (9) does not provide insight into which pairs of dictionaries $\mathbf{A}$ and $\mathbf{B}$ are most useful for signal separation. The following theorem takes into account this structure, enabling us to gain insight into which pairs of dictionaries $\mathbf{A}$ and $\mathbf{B}$ support signal separation.

Theorem 3 (BP separation): Let $\mathbf{z} = \mathbf{D}\mathbf{w} + \mathbf{n}$, with $\|\mathbf{n}\|_2 \leq \varepsilon$, $\mathbf{D} = [\,\mathbf{A}\ \mathbf{B}\,]$, and $\mathbf{w} = [\,\mathbf{a}^T\ \mathbf{e}^T\,]^T$. The dictionary $\mathbf{D}$ is characterized by the coherence parameters μ_a, μ_b, μ_m, and μ_d, and we assume $\mu_b \leq \mu_a$ without loss of generality. Furthermore, let $\mathcal{W} = \mathrm{supp}_w(\mathbf{w})$. If

$$w < \max\left\{\frac{2(1+\mu_a)}{\mu_a + 2\mu_d + \sqrt{\mu_a^2 + \mu_m^2}}, \frac{1+\mu_d}{2\mu_d}\right\}, \tag{10}$$

then the solution $\hat{\mathbf{w}}$ of BP separation

$$\text{(BP-SEP)} \quad \begin{cases} \text{minimize} & \|\tilde{\mathbf{w}}\|_1 \\ \text{subject to} & \|\mathbf{z} - \mathbf{D}\tilde{\mathbf{w}}\|_2 \leq \eta \end{cases}$$

using $\varepsilon \leq \eta$ satisfies

$$\|\mathbf{w} - \hat{\mathbf{w}}\|_2 \leq C_7(\varepsilon + \eta) + C_8\|\mathbf{w} - \mathbf{w}_{\mathcal{W}}\|_1\,, \tag{11}$$

with the (non-negative) constants C_7 and C_8.

Proof: The proof is given in Appendix B. ■

The recovery condition (10) refines that in (9); in particular, considering the two-ONB setting for which $\mu_a = \mu_b = 0$ and $\mu_m = \mu_d$. In this case, the straightforward recovery condition corresponds to (9), whereas the one for BP separation (10) is $w < 2/(3\mu_d)$. Hence, (10) guarantees the stable recovery for a larger number of dominant entries w. The recovery condition for perfectly sparse signals and noiseless measurements in the two-ONB setting corresponds to $w < (\sqrt{2} - 0.5)/\mu_d$ (see [16], [17], [28]) and turns out to be slightly less restrictive than the recovery condition for approximately sparse signals and measurement noise in (10). Whether this behavior is a fundamental result of considering approximately sparse signals or is an artifact of the proof remains an open research problem.

C. Comparison of the recovery conditions

Figure 1 compares the recovery conditions for the general model (1) to those obtained in [2], [11] for perfectly sparse signals and noiseless measurements. We set $\mu_d = \mu_m = 0.1$ and $\mu_a = \mu_b = 0.04$, and compare the recovery conditions of BP restoration and BP separation analyzed in Section III. The following observations can be made:

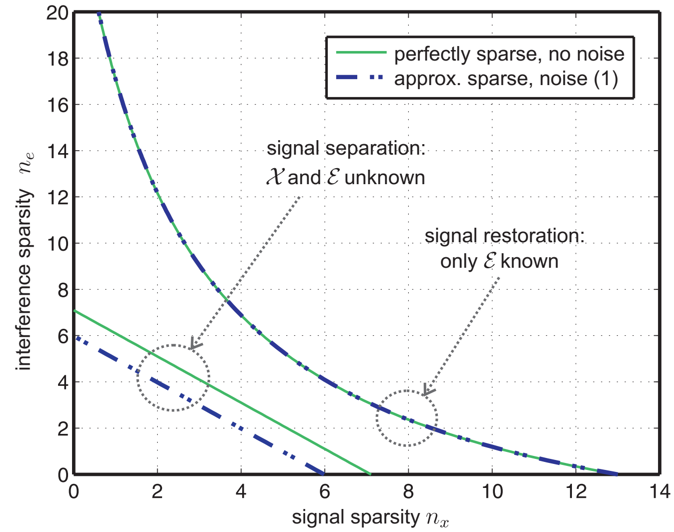


Fig. 1. Comparison of the recovery conditions using the coherence parameters $\mu_m = \mu_d = 0.1$ and $\mu_a = \mu_b = 0.04$. In the case of only $\mathcal{E}$ known, the curves for both models overlap.

- *BP restoration:* In this case, the recovery conditions for the general setup considered in this paper and the condition [2, Eq. 14] for perfectly sparse signals and noiseless measurements coincide. Hence, generalizing the results does not incur a loss in terms of the recovery conditions. The asymmetry of the curve results from the fact that only knowledge of the support set of $\mathbf{e}$ is available (see [2] for a detailed discussion).
- *BP separation:* We see that the recovery conditions for the general case (1) and the case of perfectly sparse signals with noiseless measurements differ, i.e., the condition [11, Eq. 13] is slightly less restrictive. Note that both recovery conditions are a linear function in n_x and n_e as they guarantee the recovery of $w = n_x + n_e$ nonzero (or dominant) entries in $\mathbf{w}$.

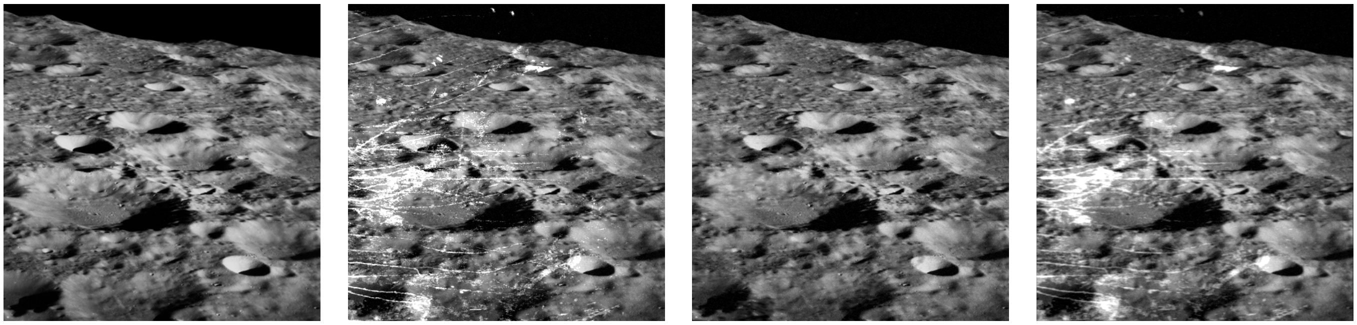
In summary, we see that having more knowledge on the support sets of $\mathbf{x}$ and/or $\mathbf{e}$ prior to recovery yields less restrictive recovery conditions. This intuitive behavior can also be observed in practice and is illustrated next.

IV. Application Example

We now present a simple sparsity-based in-painting example, where we simultaneously attenuate Gaussian noise and remove scratches from old photographs. A plethora of in-painting methods have been proposed in the literature (see, e.g., [6]–[8] and the references therein). Our goal here is not to benchmark our performance vs. theirs, but rather to quantify the impact of support-set knowledge on the recovery performance in a practical application.

A. Corruption and restoration procedure

We corrupt a 512×512 greyscale image (the pixel intensities are scaled to be within 0 and 1) by adding a mask containing artificially generated scratch patterns, which destroys 15% of the image. We furthermore corrupt each pixel with additive



(a) Original (courtesy of NASA [29]) (b) Corrupted ($\mathsf{MSE} = -16.6\,\text{dB}$) (c) BP-RES ($\mathsf{MSE} = -28.1\,\text{dB}$) (d) BP-SEP ($\mathsf{MSE} = -19.3\,\text{dB}$)

Fig. 2. Example of using BP restoration (BP-RES) and BP separation (BP-SEP) for simultaneous denoising and scratch removal for old photographs.

i.i.d. zero-mean Gaussian noise and variance 0.04. The mean-squared error (MSE) between the original in Fig. 2(a) and the corrupted version in Fig. 2(b) is $-16.6\,\text{dB}$.

Scratch removal proceeds as follows. We assume that the image admits an approximately sparse representation in the two-dimensional DCT basis $\mathbf{A}$, whereas the interference is assumed to be sparse in the identity basis $\mathbf{B} = \mathbf{I}_M$. Recovery is performed on the basis of the full 512×512 pixel image, i.e., we have $M = 512^2$ corrupted measurements. For BP restoration, we assume that the locations of the scratches are known prior to recovery, whereas no such knowledge is required for BP separation. For BP restoration we recover $\hat{\mathbf{x}}$ (for BP separation we additionally recover $\hat{\mathbf{e}}$) and then compute an estimate of the (uncorrupted) image as $\hat{\mathbf{y}} = \mathbf{A}\hat{\mathbf{x}}$. To arrive at a low recovery MSE, we set $\eta = 0.5$ for both algorithms.

B. Discussion of the results

Figure 2 shows results of the corruption and recovery procedure along with the associated MSE values. For BP restoration, we see that the recovered image has an MSE of $-29.2\,\text{dB}$ and well approximates the ground truth. For BP separation, the MSE improves over the corrupted image, but in parts where large areas of the image are corrupted, blind removal of scratches fails. Hence, knowing the locations of the sparse corruptions leads to a significant MSE advantage and is therefore highly desirable for sparsity-based in-painting methods using the proposed algorithms.

We finally emphasize that the recovery conditions (6) and (10) turn out to be useful in practice as they show that the dictionary $\mathbf{A}$ must both i) *sparsify* the signal to be recovered and ii) be *incoherent* with the interference dictionary $\mathbf{B}$ [2]. Note that the second requirement is satisfied for the DCT–Identity pair used here, whereas other transform bases typically used to sparsify images (i.e., to satisfy the first requirement), such as wavelet bases exhibit high mutual coherence with the identity basis. Hence, our recovery conditions also help to identify suitable dictionary pairs for a variety of sparsity-based restoration and separation problems.

V. Conclusions

In this paper, we have generalized the recovery guarantees presented in [2], [11] for the restoration and separation of perfectly sparse signals to the much more practical case of approximately sparse signals and noisy measurements. We proposed computationally efficient restoration and separation algorithms which build upon basis-pursit (BP) denoising and derived corresponding deterministic recovery guarantees. The provided recovery conditions put limits on the number of dominant coefficients that guarantee stability of recovery and separation, and additionally explain which dictionary pairs $\mathbf{A}$ and $\mathbf{B}$ are most suited for signal restoration or separation using the proposed algorithms.

Appendix A
Proof of Theorem 2

The proof follows that for Theorem 1 detailed in [22, Thm. 2.1] and relies on techniques developed earlier in [13], [22], [30]. We first derive a set of key properties, which are then used to prove the main result.

A. Prerequisites

We start with the following definitions. Let $\mathbf{h} = \hat{\mathbf{x}} - \mathbf{x}$, where $\hat{\mathbf{x}}$ denotes the solution of BP-RES and $\mathbf{x}$ is the vector to be recovered. Furthermore, define $\mathbf{h}_0 = \mathbf{P}_{\mathcal{X}}\mathbf{h}$ with the set $\mathcal{X} = \text{supp}_{n_x}(\mathbf{x})$. The proof relies on the following facts.

1) Cone constraint: Let $e_0 = 2\|\mathbf{x} - \mathbf{x}_{\mathcal{X}}\|_1$ with $\mathbf{x}_{\mathcal{X}} = \mathbf{P}_{\mathcal{X}}\mathbf{x}$; then [13], [30]

$$\|\mathbf{h} - \mathbf{h}_0\|_1 \leq \|\mathbf{h}_0\|_1 + e_0 \tag{12}$$

which follows from the fact that BPDN delivers a feasible solution $\hat{\mathbf{x}}$ satisfying $\|\mathbf{x}\|_1 \geq \|\hat{\mathbf{x}}\|_1$ and from

$$\begin{aligned}\|\mathbf{x}\|_1 &\geq \|\mathbf{x}_{\mathcal{X}} + \mathbf{h}_0\|_1 + \|\mathbf{h} - \mathbf{h}_0 + \mathbf{x}_{\mathcal{X}^c}\|_1 \\ &\geq \|\mathbf{x}_{\mathcal{X}}\|_1 - \|\mathbf{h}_0\|_1 + \|\mathbf{h} - \mathbf{h}_0\|_1 - \|\mathbf{x}_{\mathcal{X}^c}\|_1 .\end{aligned}$$

Application of the reverse triangle inequality to the left-hand side term of (12) yields the following useful bound:

$$\|\mathbf{h}\|_1 \leq 2\|\mathbf{h}_0\|_1 + e_0. \tag{13}$$

2) Tube constraint: Analogously to [13], [22], [30], we arrive at the following chain of ineqialities:

$$\begin{aligned}\left\|\widetilde{\mathbf{A}}\mathbf{h}\right\|_2 &\leq \|\mathbf{R}_{\mathcal{E}}(\mathbf{A}\hat{\mathbf{x}} - \mathbf{z})\|_2 + \|\mathbf{R}_{\mathcal{E}}(\mathbf{A}\mathbf{x} - \mathbf{z})\|_2 \\ &\leq \eta + \|\mathbf{R}_{\mathcal{E}}\mathbf{n}\|_2 \leq \eta + \varepsilon,\end{aligned}$$

where the last inequality follows from the fact that $\mathbf{R}_\mathcal{E}$ is a projection matrix and, hence, $\|\mathbf{R}_\mathcal{E}\mathbf{n}\|_2 \leq \|\mathbf{n}\|_2 \leq \varepsilon$.

B. Properties of the matrix $\widetilde{\mathbf{A}}$

It is important to realize that BP restoration operates on the input-output relation (5) using with $\mathbf{R}_\mathcal{E} = \mathbf{I}_M - \mathbf{B}_\mathcal{E}\mathbf{B}_\mathcal{E}^\dagger$ and $\widetilde{\mathbf{A}} = \mathbf{R}_\mathcal{E}\mathbf{A}$. The recovery condition for BP-RES (6), which will be derived next, also ensures that $\mathbf{R}_\mathcal{E}$ exists. Specifically, in [2] it was shown that the same condition as in (6) ensures the existence of $\mathbf{R}_\mathcal{E}$. In order to adapt the proof in [22] for the projected input-output relation (5), the following properties of $\widetilde{\mathbf{A}}$ are required.

1) Coherence-based bound on the RIC: We next compute a coherence-based bound on the restricted isometry constant (RIC) for the matrix $\widetilde{\mathbf{A}}$. To this end, let $\mathbf{h}_0$ be perfectly n_x-sparse and bound $\|\widetilde{\mathbf{A}}\mathbf{h}_0\|^2$ as

$$\left\|\widetilde{\mathbf{A}}\mathbf{h}_0\right\|_2^2 = \left|\mathbf{h}_0^H\mathbf{A}^H\mathbf{A}\mathbf{h}_0 - \mathbf{h}_0^H\mathbf{A}^H\mathbf{B}_\mathcal{E}\mathbf{B}_\mathcal{E}^\dagger\mathbf{A}\mathbf{h}_0\right| \tag{14}$$

$$\leq (1+\mu_a(n_x-1))\|\mathbf{h}_0\|_2^2 + \left|\mathbf{h}_0^H\mathbf{A}^H\mathbf{B}_\mathcal{E}\mathbf{B}_\mathcal{E}^\dagger\mathbf{A}\mathbf{h}_0\right|, \tag{15}$$

where (14) follows from $\mathbf{R}_\mathcal{E}^H\mathbf{R}_\mathcal{E} = \mathbf{R}_\mathcal{E}$ and (15) from Geršgorin's disc theorem [31, Thm. 6.1.1]. Next, we bound the second RHS term in (15) as follows:

$$\left|\mathbf{h}_0^H\mathbf{A}^H\mathbf{B}_\mathcal{E}\mathbf{B}_\mathcal{E}^\dagger\mathbf{A}\mathbf{h}_0\right| \leq \lambda_{\min}^{-1}(\mathbf{B}_\mathcal{E}^H\mathbf{B}_\mathcal{E})\left\|\mathbf{B}_\mathcal{E}^H\mathbf{A}\mathbf{h}_0\right\|_2^2 \tag{16}$$

$$\leq \lambda_{\min}^{-1}(\mathbf{B}_\mathcal{E}^H\mathbf{B}_\mathcal{E})\left\|\mathbf{B}_\mathcal{E}^H\mathbf{A}_\mathcal{X}\right\|_2^2\|\mathbf{h}_0\|_2^2 \tag{17}$$

$$\leq \frac{n_x n_e \mu_m^2}{[1-\mu_b(n_e-1)]^+}\|\mathbf{h}_0\|_2^2 \tag{18}$$

where (16) follows from [31, Thm. 4.2.2], (17) from the ℓ_2-norm inequality. The inequality (18) results from

$$\left\|\mathbf{B}_\mathcal{E}^H\mathbf{A}_\mathcal{X}\right\|_2^2 \leq \left\|\mathbf{B}_\mathcal{E}^H\mathbf{A}_\mathcal{X}\right\|_F^2 = \sum_{\ell\in\mathcal{E}}\sum_{k\in\mathcal{X}}\left|\mathbf{b}_\ell^H\mathbf{a}_k\right|^2 \leq n_x n_e \mu_m^2.$$

Note that (18) requires $n_e < 1 + 1/\mu_b$, which is a sufficient condition for $(\mathbf{B}_\mathcal{E}^H\mathbf{B}_\mathcal{E})^{-1}$ to exist. Note that $n_e < 1 + 1/\mu_b$ holds whenever the recovery condition for BP-RES is satisfied. Combining (15) with (18) results in

$$\|\mathbf{R}_\mathcal{E}\mathbf{A}\mathbf{h}_0\|_2^2 \leq \left(1+\mu_a(n_x-1)+\frac{n_x n_e \mu_m^2}{[1-\mu_b(n_e-1)]^+}\right)\|\mathbf{h}_0\|_2^2 \tag{19}$$

$$= (1+\hat{\delta})\|\mathbf{h}_0\|_2^2.$$

We next compute the lower bound on the RIC as

$$\begin{aligned}\left\|\widetilde{\mathbf{A}}\mathbf{h}_0\right\|_2^2 &\geq (1-\mu_a(n_x-1))\|\mathbf{h}_0\|_2^2 - \left|\mathbf{h}_0^H\mathbf{A}^H\mathbf{B}_\mathcal{E}\mathbf{B}_\mathcal{E}^\dagger\mathbf{A}\mathbf{h}_0\right| \\ &\geq \left(1-\mu_a(n_x-1)-\frac{n_x n_e\mu_m^2}{[1-\mu_b(n_e-1)]^+}\right)\|\mathbf{h}_0\|_2^2 \\ &= (1-\hat{\delta})\|\mathbf{h}_0\|_2^2,\end{aligned} \tag{20}$$

which is obtained by carrying out similar steps used to arrive at (18). Note that (19) and (20) provide a coherence-based upper bound $\hat{\delta}$ on the RIC of the projected matrix $\widetilde{\mathbf{A}} = \mathbf{R}_\mathcal{E}\mathbf{A}$.

2) Upper bound on the inner products: To carry out the proof in [22], we need an upper bound on the inner products of columns of the matrix $\widetilde{\mathbf{A}}$. For $i \neq j$, we obtain

$$\left|\tilde{\mathbf{a}}_i^H\tilde{\mathbf{a}}_j\right| = \left|\mathbf{a}_i^H\mathbf{R}_\mathcal{E}\mathbf{a}_j\right| \leq \left|\mathbf{a}_i^H\mathbf{a}_j\right| + \left|\mathbf{a}_i^H\mathbf{B}_\mathcal{E}\mathbf{B}_\mathcal{E}^\dagger\mathbf{a}_j\right|$$

$$\leq \mu_a + \frac{\left|\mathbf{a}_i^H\mathbf{B}_\mathcal{E}\mathbf{B}_\mathcal{E}^H\mathbf{a}_j\right|}{[1-\mu_b(n_e-1)]^+} \tag{21}$$

$$\leq \mu_a + \frac{\left\|\mathbf{B}_\mathcal{E}^H\mathbf{a}_i\right\|_2\left\|\mathbf{B}_\mathcal{E}^H\mathbf{a}_j\right\|_2}{[1-\mu_b(n_e-1)]^+}, \tag{22}$$

where (21) is a consequence of Geršgorin's disc theorem, and (22) of the Cauchy-Schwarz inequality. Since

$$\left\|\mathbf{B}_\mathcal{E}^H\mathbf{a}_i\right\|_2 = \sqrt{\sum_{k\in\mathcal{E}}\left|\mathbf{b}_k^H\mathbf{a}_i\right|^2} \leq \sqrt{n_e\mu_m^2}$$

for all $i = 1, \ldots, N_a$, the inner products with $i \neq j$ satisfy

$$\left|\tilde{\mathbf{a}}_i^H\tilde{\mathbf{a}}_j\right| \leq \mu_a + \frac{n_e\mu_m^2}{[1-\mu_b(n_e-1)]^+} \triangleq a. \tag{23}$$

3) Lower bound on the column norm: The last prerequisite for the proof is a lower bound on the column-norms of $\widetilde{\mathbf{A}}$. Application of the reverse triangle inequality, using the fact that $\|\mathbf{a}_i\|_2 = 1$, $\forall i$, and carrying out the similar steps used to arrive at (23) results in

$$\begin{aligned}\|\tilde{\mathbf{a}}_i\|_2^2 = \|\mathbf{R}_\mathcal{E}\mathbf{a}_i\|_2^2 &\geq \left|\mathbf{a}_i^H\mathbf{a}_i\right| - \left|\mathbf{a}_i^H\mathbf{B}_\mathcal{E}\mathbf{B}_\mathcal{E}^\dagger\mathbf{a}_i\right| \\ &\geq 1 - \frac{n_e\mu_m^2}{[1-\mu_b(n_e-1)]^+} \triangleq b.\end{aligned}$$

C. The recovery guarantee

We now derive the recovery condition and bound the corresponding error $\|\mathbf{h}\|_2$. The proof follows that of [22, Thm. 2.1]. For the sake of simplicity of exposition, we make use of the previously defined quantities $\hat{\delta}$, a, and b.

1) Bounding the error on the signal support: We start by bounding the error $\|\mathbf{h}_0\|_2$ as follows:

$$\begin{aligned}\left|\mathbf{h}^H\widetilde{\mathbf{A}}^H\widetilde{\mathbf{A}}\mathbf{h}_0\right| &\geq \left|\mathbf{h}_0^H\widetilde{\mathbf{A}}^H\widetilde{\mathbf{A}}\mathbf{h}_0\right| - \left|(\mathbf{h}-\mathbf{h}_0)^H\widetilde{\mathbf{A}}^H\widetilde{\mathbf{A}}\mathbf{h}_0\right| \\ &\geq (1-\hat{\delta})\|\mathbf{h}_0\|_2^2 - a n_x\|\mathbf{h}_0\|_2^2 - a\sqrt{n_x}\|\mathbf{h}_0\|_2 e_0 \\ &= c\|\mathbf{h}_0\|_2^2 - a\sqrt{n_x}\|\mathbf{h}_0\|_2 e_0\end{aligned}$$

with

$$c \triangleq 1 - \hat{\delta} - a n_x = 1 - \mu_a(2n_x-1) - \frac{2n_x n_e\mu_m^2}{[1-\mu_b(n_e-1)]^+}.$$

Note that the parameter c is crucial, since it determines the recovery condition for BP-RES (6). In particular, $c > 0$ is equivalent to (6)

$$[1-\mu_a(2n_x-1)]^+[1-\mu_b(n_e-1)]^+ > 2n_x n_e\mu_m^2.$$

If this condition is satisfied, then we can bound $\|\mathbf{h}_0\|_2$ from above as follows:

$$\|\mathbf{h}_0\|_2 \leq \frac{(\varepsilon+\eta)\sqrt{1+\hat{\delta}} + a\sqrt{n_x}e_0}{c}.$$

2) Bounding the recovery error: We next compute an upper bound on $\|\mathbf{h}\|_2$. To this end, we start with the lower bound

$$\left\|\widetilde{\mathbf{A}}\mathbf{h}\right\|_2^2 \geq (b+a)\|\mathbf{h}\|_2^2 - a\|\mathbf{h}\|_1^2 = (1+\mu_a)\|\mathbf{h}\|_2^2 - a\|\mathbf{h}\|_1^2,$$

since $b + a = 1 + \mu_a$. Finally, we bound $\|\mathbf{h}\|_2$ as follows:

$$\begin{aligned}\|\mathbf{h}\|_2 &\leq (\varepsilon+\eta)\frac{c + 2\sqrt{an_x}\sqrt{1+\hat{\delta}}}{\sqrt{1+\mu_a}c} + e_0\frac{\sqrt{a}\sqrt{1+\mu_a}}{c} \\ &= C_5(\eta+\varepsilon) + C_6\|\mathbf{x}-\mathbf{x}_{\mathcal{X}}\|_1,\end{aligned}$$

where the constants C_5 and C_6 depend on μ_a, μ_b, n_x, and n_e, which concludes the proof.

APPENDIX B
PROOF OF THEOREM 3

We start by deriving a coherence-based bound on the RIC of the concatenated matrix $\mathbf{D} = [\,\mathbf{A}\ \mathbf{B}\,]$, which is then used to prove the main result.

A. Coherence-based RIC for $\mathbf{D} = [\,\mathbf{A}\ \mathbf{B}\,]$

In this section, we obtain an equivalent bound to that in Section A-B1 for the dictionary $\mathbf{D}$ that depends only on the coherence parameters μ_a, μ_b, μ_m, and μ_d, and the total number of nonzero entries denoted by $w = n_x + n_e$.

1) Bounds that are explicit in n_x and n_e: Let $\mathbf{h}_0 = [\,\mathbf{h}_x^T\ \mathbf{h}_e^T\,]^T$ where $\mathbf{h}_x = \mathbf{P}_{\mathcal{X}}(\hat{\mathbf{x}} - \mathbf{x})$ and $\mathbf{h}_e = \mathbf{P}_{\mathcal{E}}(\hat{\mathbf{e}} - \mathbf{e})$ are perfectly n_x and n_e sparse, respectively. We start by the lower bound on the squared ℓ_2-norm according to

$$\begin{aligned}\|\mathbf{D}\mathbf{h}_0\|_2^2 &= \begin{bmatrix} \mathbf{h}_x^H & \mathbf{h}_e^H \end{bmatrix} \begin{bmatrix} \mathbf{A}^H\mathbf{A} & \mathbf{A}^H\mathbf{B} \\ \mathbf{B}^H\mathbf{A} & \mathbf{B}^H\mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{h}_x \\ \mathbf{h}_e \end{bmatrix} \\ &= \mathbf{h}_0^H \begin{bmatrix} \mathbf{I}_{N_a} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N_b} \end{bmatrix} \mathbf{h}_0 + \mathbf{h}_0^H \begin{bmatrix} \mathbf{A}^H\mathbf{A} - \mathbf{I}_{N_a} & \mathbf{A}^H\mathbf{B} \\ \mathbf{B}^H\mathbf{A} & \mathbf{B}^H\mathbf{B} - \mathbf{I}_{N_b} \end{bmatrix} \mathbf{h}_0 \\ &\geq \|\mathbf{h}_0\|_2^2 - \left\|\begin{bmatrix} \mathbf{A}_{\mathcal{X}}^H\mathbf{A}_{\mathcal{X}} - \mathbf{I}_{|\mathcal{X}|} & \mathbf{A}_{\mathcal{X}}^H\mathbf{B}_{\mathcal{E}} \\ \mathbf{B}_{\mathcal{E}}^H\mathbf{A}_{\mathcal{X}} & \mathbf{B}_{\mathcal{E}}^H\mathbf{B}_{\mathcal{E}} - \mathbf{I}_{|\mathcal{E}|} \end{bmatrix}\right\|_2 \|\mathbf{h}_0\|_2^2,\end{aligned}$$

which follows from the reverse triangle inequality and elementary properties of the ℓ_2 matrix norm. We next compute an upper bound on the matrix norm as follows:

$$\begin{aligned}&\left\|\begin{bmatrix} \mathbf{A}_{\mathcal{X}}^H\mathbf{A}_{\mathcal{X}} - \mathbf{I}_{|\mathcal{X}|} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{\mathcal{E}}^H\mathbf{B}_{\mathcal{E}} - \mathbf{I}_{|\mathcal{E}|} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{A}_{\mathcal{X}}^H\mathbf{B}_{\mathcal{E}} \\ \mathbf{B}_{\mathcal{E}}^H\mathbf{A}_{\mathcal{X}} & \mathbf{0} \end{bmatrix}\right\|_2 \\ &\leq \max\left\{\left\|\mathbf{A}_{\mathcal{X}}^H\mathbf{A}_{\mathcal{X}} - \mathbf{I}_{|\mathcal{X}|}\right\|_2, \left\|\mathbf{B}_{\mathcal{E}}^H\mathbf{B}_{\mathcal{E}} - \mathbf{I}_{|\mathcal{E}|}\right\|_2\right\} + \left\|\mathbf{A}_{\mathcal{X}}^H\mathbf{B}_{\mathcal{E}}\right\|_2,\end{aligned}$$

which results from the triangle inequality for matrix norms and the facts that the spectral norm of both a block-diagonal matrix and an anti-block-diagonal matrix is given by the largest among the spectral norms of the individual nonzero blocks. The application of Geršgorin's disc theorem and

$$\left\|\mathbf{A}_{\mathcal{X}}^H\mathbf{B}_{\mathcal{E}}\right\|_2 \leq \left\|\mathbf{A}_{\mathcal{X}}^H\mathbf{B}_{\mathcal{E}}\right\|_F \leq \sqrt{n_x n_e \mu_m^2}$$

leads to

$$\begin{aligned}\max&\left\{\left\|\mathbf{A}_{\mathcal{X}}^H\mathbf{A}_{\mathcal{X}} - \mathbf{I}_{|\mathcal{X}|}\right\|_2, \left\|\mathbf{B}_{\mathcal{E}}^H\mathbf{B}_{\mathcal{E}} - \mathbf{I}_{|\mathcal{E}|}\right\|_2\right\} + \left\|\mathbf{A}_{\mathcal{X}}^H\mathbf{B}_{\mathcal{E}}\right\|_2 \\ &\leq \max\left\{\mu_a(n_x-1), \mu_b(n_e-1)\right\} + \sqrt{n_x n_e \mu_m^2}.\end{aligned}$$

Hence, we arrive at the following lower bound

$$\|\mathbf{D}\mathbf{h}_0\|_2^2 \geq \|\mathbf{h}_0\|_2^2\Big(1 - m(n_x,n_e) - \sqrt{n_x n_e \mu_m^2}\Big) \qquad (24)$$

with $m(n_x, n_e) = \max\{\mu_a(n_x-1), \mu_b(n_e-1)\}$. Performing similar steps as for (24) results in the upper bound

$$\|\mathbf{D}\mathbf{h}_0\|_2^2 \leq \|\mathbf{h}_0\|_2^2\Big(1 + m(n_x,n_e) + \sqrt{n_x n_e \mu_m^2}\Big). \qquad (25)$$

2) Bounds depending on $w = n_x + n_e$: Both bounds in (24) and (25) are explicit in n_x and n_e. Since the individual sparsity levels n_x and n_e are unknown prior to recovery, we require a coherence-based bound on the RIC that depends solely on the total number $w = n_x + n_e$ of nonzero entries of $\mathbf{h}_0$ rather than on n_x and n_e. To this end, we define the function

$$g(n_x, n_e) = \max\{\mu_a(n_x-1), \mu_b(n_e-1)\} + \sqrt{n_x n_e \mu_m^2}$$

and find the maximum

$$\hat{g}(w) = \max_{0 \leq n_x \leq w} g(n_x, w - n_x). \qquad (26)$$

Since $\hat{g}(w)$ only depends on $w = n_x + n_e$ and $g(n_x, n_e) \leq \hat{g}(w)$, we can replace $g(n_x, n_e)$ by $\hat{g}(w)$ in (24) and (25).

We start by computing the maximum in (26). Assume $\mu_a(n_x - 1) \geq \mu_b(n_e - 1)$ and consider the function

$$g_a(n_x, w - n_x) = \mu_a(n_x - 1) + \sqrt{n_x(w-n_x)\mu_m^2}. \qquad (27)$$

It can easily be shown that $g_a(n_x, w - n_x)$ is strictly concave in n_x for all $0 \leq n_x \leq w$ and $0 \leq w < \infty$ and, therefore, the maximum is either achieved at a stationary point or a boundary point. Standard arithmetic manipulations show that the (global) maximum of the function in (27) corresponds to

$$\hat{g}_a(w) = \frac{1}{2}\left(\mu_a(w-2) + w\sqrt{\mu_a^2 + \mu_m^2}\right). \qquad (28)$$

For the case where $\mu_a(n_x - 1) < \mu_b(n_e - 1)$, we carry out similar steps used to arrive at (27) and exploit the symmetry of (26) to arrive at

$$\hat{g}_b(w) = \frac{1}{2}\left(\mu_b(w-2) + w\sqrt{\mu_b^2 + \mu_m^2}\right).$$

Hence, by assuming that $\mu_b \leq \mu_a$, we obtain upper and lower bounds on (24) and (25) in terms of $w = n_x + n_e$ with the aid of (28) as follows:

$$(1 - \hat{g}_a(w))\|\mathbf{h}_0\|_2^2 \leq \|\mathbf{D}\mathbf{h}_0\|_2^2 \leq (1 + \hat{g}_a(w))\|\mathbf{h}_0\|_2^2. \qquad (29)$$

It is important to realize that for some values of μ_a, μ_m, and w, the bounds in (29) are inferior to those obtained when ignoring the structure of the concatenated dictionary $\mathbf{D}$, i.e.,

$$\begin{aligned}(1 - \mu_d(w-1))\|\mathbf{h}_0\|_2^2 &\leq \|\mathbf{D}\mathbf{h}_0\|_2^2 \\ &\leq (1 + \mu_d(w-1))\|\mathbf{h}_0\|_2^2 \end{aligned} \qquad (30)$$

with $\mu_d = \max\{\mu_a, \mu_b, \mu_m\}$. In order to tighten the RIC in both cases, we consider

$$\left(1 - \hat{\delta}_w\right)\|\mathbf{h}_0\|_2^2 \leq \|\mathbf{D}\mathbf{h}_0\|_2^2 \leq \left(1 + \hat{\delta}_w\right)\|\mathbf{h}_0\|_2^2,$$

where the coherence-based upper bound on the RIC of the concatenated dictionary $\mathbf{D} = [\,\mathbf{A}\ \mathbf{B}\,]$ corresponds to

$$\hat{\delta}_w = \min\left\{1/2\Big(\mu_a(w-2) + w\sqrt{\mu_a^2 + \mu_m^2}\Big), \mu_d(w-1)\right\}.$$

B. Recovery guarantee

We now bound the error $\|\mathbf{h}\|_2$ and derive the recovery guarantee by following the proof in Appendix A. In the following, we only show the case where

$$1/2\left(\mu_a(w-2)+w\sqrt{\mu_a^2+\mu_m^2}\right)\leq\mu_d(w-1).$$

The other case, i.e., where the standard RIC bound (30) is tighter than (29), readily follows from the proof for Theorem 1 in [22], by replacing $\mathbf{A}$ by $\mathbf{D}$, μ_a by μ_d, and n_x by w.

1) Bounding the error on the signal support: We start by bounding the error $\|\mathbf{h}_0\|_2$. Since $\mu_m\leq\mu_d$, we arrive at

$$\begin{aligned}\left|\mathbf{h}^H\mathbf{D}^H\mathbf{D}\mathbf{h}_0\right| &\geq\left|\mathbf{h}_0^H\mathbf{D}^H\mathbf{D}\mathbf{h}_0\right|-\left|(\mathbf{h}-\mathbf{h}_0)^H\mathbf{D}^H\mathbf{D}\mathbf{h}_0\right|\\ &\geq(1-\hat{\delta}_w)\|\mathbf{h}_0\|_2^2-\mu_d w\|\mathbf{h}_0\|_2^2-\mu_d\sqrt{w}\|\mathbf{h}_0\|_2\,e_0\\ &=d\|\mathbf{h}_0\|_2^2-\mu_d\sqrt{w}\|\mathbf{h}_0\|_2\,e_0\end{aligned}$$

with

$$d\triangleq 1-\delta_w-\mu_d w=1-\frac{w}{2}\left(\mu_a+2\mu_d+\sqrt{\mu_a^2+\mu_m^2}\right)+\mu_a.$$

It is important to note that d is crucial for the recovery guarantee as it determines the condition for which BP-SEP in (10) enables stable separation. Specifically, if $d>0$ or, equivalently, if

$$w<\frac{2(1+\mu_a)}{\mu_a+2\mu_d+\sqrt{\mu_a^2+\mu_m^2}}$$

then the error on the signal support $\|\mathbf{h}_0\|_2$ is bounded from above as

$$\|\mathbf{h}_0\|_2\leq\frac{(\varepsilon+\eta)\sqrt{1+\hat{\delta}_w}+\mu_d\sqrt{w}e_0}{d}.$$

where $e_0=2\|\mathbf{w}-\mathbf{w}_\mathcal{W}\|_1$ with $\mathcal{W}=\operatorname{supp}_w(\mathbf{w})$.

2) Bounding the recovery error: We now compute an upper bound on $\|\mathbf{h}\|_2$, i.e.,

$$\|\mathbf{D}\mathbf{h}\|_2^2\geq(1+\mu_d)\|\mathbf{h}\|_2^2-\mu_d\|\mathbf{h}\|_1^2\,. \tag{31}$$

Finally, bounding $\|\mathbf{h}\|_2$ similarly to Appendix A results in

$$\begin{aligned}\|\mathbf{h}\|_2&\leq(\varepsilon+\eta)\frac{d+2\sqrt{\mu_d w}\sqrt{1+\hat{\delta}_w}}{\sqrt{1+\mu_d}d}+e_0\frac{\sqrt{\mu_d}(d+2\mu_d w)}{\sqrt{1+\mu_d}d}\\ &=C_7(\eta+\varepsilon)+C_8\|\mathbf{w}-\mathbf{w}_\mathcal{W}\|_1\,.\end{aligned}$$

where the constants C_7 and C_8 depend on the parameters μ_a, μ_b, μ_m, μ_d, and $w=n_x+n_e$, which concludes the proof.

References

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

[2] C. Studer, P. Kuppinger, G. Pope, and H. Bölcskei, "Recovery of sparsely corrupted signals," *submitted to IEEE Trans. Inf. Theory*, Feb. 2011.

[3] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration – A Statistical Model-Based Approach*. Springer-Verlag London, 1998.

[4] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, "Audio inpainting," *to appear in IEEE Trans. on Audio, Speech, and Language Processing*, 2011.

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

[7] J.-F. Cai, R. H. Chan, L. Shen, and Z. Shen, "Simultaneously inpainting in image and transformed domains," *Num. Mathematik*, vol. 112, no. 4, pp. 509–533, May 2009.

[8] J.-F. Cai, S. Osher, and Z. Shen, "Split Bregman methods and frame based image restoration," *Multiscale Model. Simul.*, vol. 8, no. 2, pp. 337–369, Dec. 2009.

[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] G. Kutyniok, "Data separation by sparse representations," 02 2011. [Online]. Available: http://arxiv.org/abs/1102.4527v1

[11] P. Kuppinger, G. Durisi, and H. Bölcskei, "Uncertainty relations and sparse signal recovery for pairs of general signal sets," *to appear in IEEE Trans. Inf. Theory*, 2011.

[12] 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.

[13] 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.

[14] 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.

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

[16] 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.

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

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] J. N. Laska, P. Boufounos, M. A. Davenport, and R. G. Baraniuk, "Democracy in action: Quantization, saturation, and compressive sensing," *to appear in Appl. Comput. Harmon. Anal.*, 2011.

[24] 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.

[25] X. Li, "Compressed sensing and matrix completion with constant proportion of corruptions," Apr. 2011. [Online]. Available: http://arxiv.org/abs/1104.1041v1

[26] 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/1102.1227v1

[27] 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.

[28] 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.

[29] National Aeronautics and Space Administration, "Apollo 11: Orbital view, west of daedalus crater," [accessed on Jun 6, 2011:] URL `http://www.apolloarchive.com`, July 1969.

[30] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, pp. 1207–1223, 2006.

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