Stable restoration and separation of approximately sparse signals

Christoph Studer[*], Richard G. Baraniuk[**]

Department of Electrical and Computer Engineering, Rice University, 6100 Main Street, Houston, TX 77005, USA

A B S T R A C T

This paper develops new theory and algorithms to recover signals that are approximately sparse in some general dictionary (i.e., a basis, frame, or over-/incomplete matrix) but corrupted by a combination of interference having a sparse representation in a second general dictionary and measurement noise. The algorithms and analytical recovery conditions consider varying degrees of signal and interference support-set knowledge. Particular applications covered by the proposed framework include the restoration of signals impaired by impulse noise, narrowband interference, or saturation/clipping, as well as image in-painting, super-resolution, and signal separation. Two application examples for audio and image restoration demonstrate the efficacy of the approach.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

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

$$\mathbf{z} = \mathbf{Ax} + \mathbf{Be} + \mathbf{n}, \tag{1}$$

where $\mathbf{A} \in \mathbb{C}^{M \times N_a}$ and $\mathbf{B} \in \mathbb{C}^{M \times N_b}$ are general deterministic dictionaries; examples for general dictionaries include bases, frames, or over-/incomplete matrices whose columns have unit Euclidean (or ℓ_2) norm. The vector $\mathbf{x}$ is assumed to be *approximately sparse*, i.e., its main energy (in terms of the sum of absolute values, for example) is concentrated in only a few entries. The M-dimensional signal vector is defined as $\mathbf{y} = \mathbf{Ax}$. 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, and $\mathbf{n} \in \mathbb{C}^M$ corresponds to measurement noise. Apart from the bound $\|\mathbf{n}\|_2 < \varepsilon$, the measurement noise is arbitrary. We emphasize that the interference and noise components $\mathbf{e}$ and $\mathbf{n}$ can depend on the vector $\mathbf{x}$ and/or the dictionary $\mathbf{A}$.

The setting (1) also allows us to study *signal separation*, i.e., the separation of two distinct features $\mathbf{Ax}$ and $\mathbf{Be}$ from the noisy observation $\mathbf{z}$. Here, the vector $\mathbf{e}$ in (1) is also allowed to be approximately sparse and is used to represent a second desirable feature (rather than undesired interference). 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 $\mathbf{Ax}$ and $\mathbf{Be}$.

---

* Corresponding author. Fax: +1 713 348 5685.
** Principal corresponding author. Fax: +1 713 348 5685.
E-mail addresses: studer@rice.edu (C. Studer), richb@rice.edu (R.G. Baraniuk).
URLs: http://www.ece.rice.edu/~cs32/ (C. Studer), http://web.ece.rice.edu/richb/ (R.G. Baraniuk).

1.1. Applications for the model (1)

Both the recovery and separation problems outlined above feature prominently in numerous applications (see [1–18] and the references therein), including:

- *Impulse noise:* The recovery of approximately sparse signals corrupted by impulse noise [13] corresponds to recovery of $\mathbf{x}$ from (1) by setting $\mathbf{B} = \mathbf{I}_M$ and associating the interference $\mathbf{e}$ with the impulse-noise vector. Practical examples include restoration of audio signals impaired by click/pop noise [1,2] and reading from unreliable memories [14].

- *Narrowband interference:* Audio, video, and communication signals are often corrupted by narrowband interference. A particular example is electric hum, which typically occurs in improperly designed audio or video equipment. Such impairments naturally exhibit a sparse representation in the frequency domain, which amounts to setting $\mathbf{B}$ to the inverse discrete Fourier transform matrix.

- *Saturation and clipping:* Nonlinearities in amplifiers may result in signal saturation, cf. [7,16,17]. Here, instead of the signal vector $\mathbf{y}$ of interest, one observes a saturated (or clipped) version $\mathbf{z} = \mathbf{y} + \mathbf{e} + \mathbf{n}$, where the nonzero entries of $\mathbf{e}$ correspond to the difference between the saturated signal and the original signal $\mathbf{y}$. The noise vector $\mathbf{n}$ can be used to model residual errors that are not captured by the interference component $\mathbf{Be}$.

- *Super-resolution and in-painting:* In super-resolution [3,15] and in-painting [6,8–10] applications, only a subset of the entries of the (full-resolution) signal vector $\mathbf{y} = \mathbf{Ax}$ is available. With (1), the interference vector $\mathbf{e}$ accounts for the missing parts of the signal, i.e., the locations of the nonzero entries of $\mathbf{e}$ correspond to the missing entries in $\mathbf{y}$ and are set to some arbitrary value. The missing parts of $\mathbf{y}$ are then filled in by recovering $\mathbf{x}$ from $\mathbf{z} = \mathbf{Ax} + \mathbf{e} + \mathbf{n}$ followed by computation of the (full-resolution) signal vector $\mathbf{y} = \mathbf{Ax}$.

- *Signal separation:* The framework (1) can be used to model the decomposition of signals into two distinct features. Prominent application examples are the separation of texture from cartoon parts in images [4,6,18] and the separation of neuronal calcium transients from smooth signals caused by astrocytes in calcium imaging [5]. In both applications, $\mathbf{A}$ and $\mathbf{B}$ are chosen such that each feature can be represented by approximately sparse vectors in one dictionary. Signal separation then amounts to simultaneously extracting $\mathbf{x}$ and $\mathbf{e}$ from $\mathbf{z}$, where $\mathbf{Ax}$ and $\mathbf{Be}$ represent the individual features.

In many applications outlined above, a predetermined (and possibly optimized) dictionary pair $\mathbf{A}$ and $\mathbf{B}$ is used. It is therefore of significant practical interest to identify the 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 [12,19]. The results in [12,19] rely on an uncertainty relation for pairs of general dictionaries and depend on the number of nonzero entries in $\mathbf{x}$ and $\mathbf{e}$, on the coherence parameters of the dictionaries $\mathbf{A}$ and $\mathbf{B}$, and on the amount of prior knowledge on the support of the signal and interference vector. However, the algorithms and proof techniques used in [12,19] cannot be adapted for the general (and practically more relevant) setting formulated in (1), which features approximately sparse signals and additive measurement noise.

1.2. Contributions

In this paper, we generalize the recovery guarantees of [12,19] to the framework (1) detailed above. In particular, we provide computationally efficient 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. Our results also depend on the amount of knowledge on the location of the dominant entries available prior to recovery. In particular, we investigate the following cases: (1) The locations of the dominant entries of the approximately sparse vector $\mathbf{x}$ and the support set of the perfectly sparse interference vector $\mathbf{e}$ are known (prior to recovery), (2) only the support set of the interference vector $\mathbf{e}$ is known, and (3) no support-set knowledge about $\mathbf{x}$ and $\mathbf{e}$ is available. Moreover, we present coherence-based bounds on the restricted isometry constants (RICs) for all these cases, which can be used to derive alternative recovery conditions. We provide a comparison to the recovery conditions for perfectly sparse signals and noiseless measurements presented in [12,19]. Finally, we demonstrate the efficacy of the proposed approach with two representative applications: restoration of audio signals impaired by a mixture of impulse noise and Gaussian noise, and removal of scratches from color photographs.

1.3. 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 $\mathbf{M}^\dagger = (\mathbf{M}^H \mathbf{M})^{-1} \mathbf{M}^H$, respectively. The kth entry of the vector $\mathbf{m}$ is $[\mathbf{m}]_k$, and the 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 $\|\mathbf{x}\|_2$, $\|\mathbf{x}\|_1 = \sum_k |[\mathbf{x}]_k|$ 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. $\|\mathbf{M}\|_F = \sqrt{\sum_{k,\ell} |[\mathbf{M}]_{k,\ell}|^2}$ 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 $\mathrm{supp}(\mathbf{m})$. 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 $[x]^+ = \max\{x, 0\}$.

1.4. Synopsis

The remainder of the paper is organized as follows. In Section 2, we briefly summarize the relevant prior art. Our new recovery algorithms and corresponding recovery guarantees are presented in Section 3. A set of alternative recovery guarantees obtained through the restricted isometry property (RIP) framework and a comparison to existing recovery guarantees are provided in Section 4. The application examples are shown in Section 5, and we conclude in Section 6. All proofs are relegated to Appendices A–D.

2. Relevant prior art

In this section, we review the relevant prior art in recovering sparse signals from noiseless and noisy measurements in the deterministic setting and summarize the existing guarantees for recovery of sparsely corrupted signals.

2.1. Recovery of perfectly sparse signals from noiseless measurements

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., $M < N_a$) corresponds to solving an underdetermined system of linear equations, which 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 one to uniquely recover $\mathbf{x}$ by solving

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

Unfortunately, P0 has prohibitive (combinatorial) computational complexity, even for small dimensions N_a. One of the most popular and computationally tractable alternative to solving P0 is basis pursuit (BP) [20–25], which corresponds to the convex program

(BP)   *minimize* $\|\tilde{\mathbf{x}}\|_1$ *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 $\mathbf{A}$, defined as $\mu_a = \max_{k,\ell,k \neq \ell} |\mathbf{a}_k^H \mathbf{a}_\ell|$. Specifically, a sufficient condition for $\mathbf{x}$ to be the unique solution of P0 and for BP to deliver this solution[1] is [22,23,25]

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

2.2. Recovery of approximately sparse signals from noisy measurements

For the case of bounded (otherwise arbitrary) measurement noise, i.e., $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{n}$ with $\|\mathbf{n}\|_2 \leqslant \varepsilon$, recovery guarantees based on the coherence parameter μ_a were developed in [28–32]. The corresponding recovery conditions mostly treat the case of perfectly sparse signals, i.e., where only a small fraction of the entries $\mathbf{x}$ are nonzero. Fortunately, many real-world signals exhibit the property that most of the signal's energy (e.g., in terms of the sum of absolute values) is concentrated in only a few entries. We refer to this class of signals as *approximately sparse* in the remainder of the paper. For such signals, the support set associated to the best n_x-sparse approximation is defined as

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

---

[1] The condition (2) also ensures perfect recovery using orthogonal matching pursuit (OMP) [25–27], which is, however, not further investigated in this paper.

where the set Σ_{n_x} contains all support sets of size n_x corresponding to perfectly n_x-sparse vectors having the same dimension as $\mathbf{x}$. A particular sub-class of approximately sparse signals is the set of *compressible signals*, whose approximation error decreases according to a power law [33].

The following theorem provides a sufficient condition for which a suitably modified version of BP, known as BP denoising (BPDN) [20], stably recovers an approximately sparse vector $\mathbf{x}$ from the noisy observation $\mathbf{z}$.

Theorem 1 *(BP denoising). (See [32, Theorem 2.1].) Let* $\mathbf{z} = \mathbf{Ax} + \mathbf{n}$, $\|\mathbf{n}\|_2 \leqslant \varepsilon$, *and* $\mathcal{X} = \mathrm{supp}_{n_x}(\mathbf{x})$. *If* (2) *is met, then the solution* $\hat{\mathbf{x}}$ *of the convex program*

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

with $\varepsilon \leqslant \eta$ *satisfies*

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

where both (*non-negative*) *constants* C_0 *and* C_1 *depend on* μ_a *and* n_x.

Proof. The proof in [32, Theorem 2.1] is given for perfectly sparse vectors only. Since some of the proofs presented in the remainder of the paper are developed for approximately sparse signals, we detail the relevant steps that generalize [32, Theorem 2.1] in Appendix A. $\square$

We emphasize that *perfect* recovery of $\mathbf{x}$ is, in general, impossible in the presence of bounded (but otherwise arbitrary) measurement noise $\mathbf{n}$. Hence, we 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 ground truth $\mathbf{x}$ is bounded from above by the ℓ_2-norm of the noise $\|\mathbf{n}\|_2$ and the best n_x-sparse approximation as in (3). The constants C_0 and C_1 depend on the coherence parameter μ_a and on n_x, and increase as one approaches the limits of (2). As an example, we obtain $C_0 \approx 2.59$ and $C_1 \approx 0.16$ for $\mu_a = 0.01$ and $n_x = 20$. Note that if ε is known, one should set $\varepsilon = \eta$ to minimize the error (3). We furthermore note that Theorem 1 generalizes the results for noiseless measurements and perfectly sparse signals in [22,23,25] using BP (cf. Section 2.1). Specifically, for $\|\mathbf{n}\|_2 = 0$ and $\|\mathbf{x} - \mathbf{x}_{\mathcal{X}}\|_1 = 0$, BPDN with $\varepsilon = \eta = 0$ corresponds to BP and *perfectly* recovers $\mathbf{x}$ whenever (2) is met.

2.3. Recovery guarantees from sparsely corrupted 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 input–output relation (1). Special or related cases of the general model (1) have been studied in [7,11–13,19,34–39].

2.3.1. Probabilistic recovery guarantees

Recovery guarantees for the probabilistic setting (i.e., recovery of $\mathbf{x}$ is guaranteed with high probability) for random (sub-)Gaussian matrices, which are of particular interest for applications based on compressive sensing (CS) [40,41] have been reported in [7,11,37,39]. Similar results for randomly sub-sampled unitary matrices $\mathbf{A}$ have been developed in [38]. The problem of sparse signal recovery from a nonlinear measurement process in the presence of impulse noise was considered in [13], and probabilistic results for signal detection based on ℓ_1-norm minimization in the presence of impulse noise was investigated in [36]. Another strain of probabilistic recovery guarantees has considered perfectly sparse signals from noiseless measurements with randomness on the location and values of the coefficient vectors [19,42,43]. In the remainder of the paper, we will focus on the deterministic setting exclusively.

2.3.2. Deterministic recovery guarantees

Recovery guarantees in the deterministic setting for noiseless measurements and signals being perfectly sparse, i.e., the model $\mathbf{z} = \mathbf{Ax} + \mathbf{Be}$, have been studied in [12,19,34,35,44,45]. In [34], it has been shown that when $\mathbf{A}$ is the discrete Fourier transform (DFT) 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 (whereas $\mathbf{x}$ and $\mathbf{e}$ are assumed to be perfectly sparse and for noiseless measurements) has been studied for different cases of support-set knowledge in [12,19]. There, deterministic recovery guarantees depending on the number of nonzero entries n_x and n_e in $\mathbf{x}$ 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 $\mu_m = \max_{k,\ell} |\mathbf{a}_k^H \mathbf{b}_\ell|$. A summary of the recovery guarantees presented in [12,19] (along with the novel recovery guarantees presented in the next section) is given in Table 1, where, for the sake of simplicity of exposition, we define the following function:

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

We emphasize that the results presented in [12,19] are for perfectly sparse and noiseless measurements only, and furthermore, that the algorithms and proof techniques cannot be adapted for the more general setting proposed in (1). In

Table 1

Summary of deterministic recovery guarantees for perfectly/approximately sparse signals that are corrupted by interference in the absence/presence of measurement noise.

Support-set knowledge	Recovery condition	Perfectly sparse and no noise	Approx. sparse and noise
x and **e**	$n_x n_e \mu_m^2 < f(n_x, n_e)$	[12, Theorem 3]	Theorem 2
e only	$2 n_x n_e \mu_m^2 < f(2n_x, n_e)$	[12, Theorems 4 and 5]	Theorem 3
None	[19, Eq. 12][a]	[19, Theorem 3]	–
	Eq. (13)	–	Theorem 4

[a] The recovery condition is valid for BP and OMP; a less restrictive condition for P0 is given in [19, Theorem 2].

order to gain insight into the practically more relevant case of approximately sparse signals and noisy measurements, we next develop new restoration and separation algorithms for several different cases of support-set knowledge and provide corresponding recovery guarantees. Our results complement those in [12,19] (cf. Table 1).

The case of signal separation with more than two orthonormal bases has been studied in [35]. Those results have been derived for perfectly sparse signals from noiseless measurements; a generalization of the results shown next to more than two dictionaries is left for future work. Another set of theoretical results for sparsity-based signal separation has been derived in [18,44,46]. Those results focus on the *analysis* separation problem in general (possibly infinite-dimensional) frames, which aims at minimizing the number of nonzero entries of the analysis coefficients rather than the synthesis coefficients considered here (see [18] for the details). The recovery conditions have been derived using a joint concentration measure and the so-called *cluster coherence*, which enable the derivation of recovery conditions for the analysis separation problem that explicitly exploit the structure of particular pairs of frames (e.g., wavelets and curvelets). Coherence-based results for hybrid synthesis–analysis problems for pairs of general dictionaries were developed recently in [45].

3. Main results

We now develop several computationally efficient methods for restoration or separation under the model (1) and derive corresponding recovery conditions that guarantee their stability. Our recovery guarantees depend on the ℓ_2-norm of the noise vector and on the amount of knowledge on the dominant nonzero entries of the signal and noise vectors. Specifically, we consider the following three cases: (1) *Direct restoration:* The locations of the entries corresponding to the best n_x-sparse approximation of **x** and the support set of the (perfectly sparse) interference vector **e** are known prior to recovery, (2) *BP restoration:* Only the support set of **e** is known, (3) *BP separation:* No knowledge about **x** and **e** is available, except for the fact that each vector exhibits an approximately a sparse representation in **A** and **B**, respectively.

3.1. Direct restoration: Support-set knowledge of **x** and **e**

We start by addressing the case where the locations of the dominant entries (in terms of absolute value) of the approximately sparse vector **x** and the support set $\mathcal{E}$ associated with the perfectly sparse interference vector **e** are known prior to recovery. This scenario is relevant, for example, in the restoration of old phonograph records [1,2], where one wants to recover a band-limited signal that is impaired by impulse noise, such as clicks and pops. The occupied frequency band of phonograph recordings is typically known prior to recovery. In this case, one may assume **A** to be the inverse M-dimensional discrete cosine transform (DCT) matrix and $\mathbf{B} = \mathbf{I}_M$. The locations of the clicks and pops, i.e., the support set $\mathcal{E} = \text{supp}(\mathbf{e})$, can be determined prior to recovery using the techniques described in [2], for example.

The restoration approach considered for this setup is as follows. Since $\mathcal{E}$ and $\mathcal{X} = \text{supp}_{n_x}(\mathbf{x})$ are both known prior to recovery, we start by projecting the noisy observation vector **z** onto the orthogonal complement of the range space spanned by $\mathbf{B}_\mathcal{E}$, which eliminates the sparse interference. Concretely, we consider

$$\mathbf{R}_\mathcal{E}\mathbf{z} = \mathbf{R}_\mathcal{E}(\mathbf{Ax} + \mathbf{Be}_\mathcal{E} + \mathbf{n}) = \mathbf{R}_\mathcal{E}\mathbf{Ax} + \mathbf{R}_\mathcal{E}\mathbf{n}, \tag{4}$$

where $\mathbf{R}_\mathcal{E} = \mathbf{I}_M - \mathbf{B}_\mathcal{E}\mathbf{B}_\mathcal{E}^\dagger$ is the projector onto the orthogonal complement of the range space of $\mathbf{B}_\mathcal{E}$, and we used the fact that $\mathbf{R}_\mathcal{E}\mathbf{Be}_\mathcal{E} = \mathbf{0}_{M \times 1}$. Next, one can separate (4) by exploiting the fact that $\mathcal{X}$ is known

$$\mathbf{R}_\mathcal{E}\mathbf{z} = \mathbf{R}_\mathcal{E}\mathbf{A}(\mathbf{x}_\mathcal{X} + \mathbf{x}_{\mathcal{X}^c}) + \mathbf{R}_\mathcal{E}\mathbf{n} = \mathbf{R}_\mathcal{E}\mathbf{A}_\mathcal{X}[\mathbf{x}]_\mathcal{X} + \mathbf{R}_\mathcal{E}\mathbf{Ax}_{\mathcal{X}^c} + \mathbf{R}_\mathcal{E}\mathbf{n}$$

and, by assuming $\mathbf{R}_\mathcal{E}\mathbf{A}_\mathcal{X}$ has full rank, we can isolate the dominant entries $[\mathbf{x}]_\mathcal{X}$ as

$$(\mathbf{R}_\mathcal{E}\mathbf{A}_X)^\dagger \mathbf{R}_\mathcal{E}\mathbf{z} = [\mathbf{x}]_\mathcal{X} + (\mathbf{R}_\mathcal{E}\mathbf{A}_X)^\dagger \mathbf{R}_\mathcal{E}(\mathbf{Ax}_{\mathcal{X}^c} + \mathbf{n}). \tag{5}$$

In the case where both vectors $\mathbf{x}_{\mathcal{X}^c}$ and **n** are equal to zero, we obtain

$$(\mathbf{R}_\mathcal{E}\mathbf{A}_X)^\dagger \mathbf{R}_\mathcal{E}\mathbf{z} = [\mathbf{x}]_\mathcal{X}, \tag{6}$$

and therefore the entries of **x** contained in the support set $\mathcal{X}$ are recovered perfectly by this approach. Note that conjugate gradient methods (e.g., [47]) offer an efficient way of computing (6).

The following theorem provides a sufficient condition for $(\mathbf{R}_{\mathcal{E}}\mathbf{A}_X)^{\dagger}\mathbf{R}_{\mathcal{E}}$ to exist and for which the vector $\mathbf{x}$ can be restored stably from the noisy measurement $\mathbf{z}$ using the direct restoration (DR) procedure outlined above.

Theorem 2 *(Direct restoration). Let $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e} + \mathbf{n}$ with $\|\mathbf{n}\|_2 \leqslant \varepsilon$, $\mathbf{e}$ perfectly n_e-sparse with support set $\mathcal{E}$, and $\mathcal{X} = \mathrm{supp}_{n_x}(\mathbf{x})$. Furthermore, assume that the support sets $\mathcal{E}$ and $\mathcal{X}$ are known prior to recovery. If*

$$n_x n_e \mu_m^2 < f(n_x, n_e), \tag{7}$$

then $\mathbf{R}_{\mathcal{E}}\mathbf{A}_{\mathcal{X}}$ is full rank and the vector $\hat{\mathbf{x}}$ computed according to

$$(\text{DR}) \quad [\hat{\mathbf{x}}]_{\mathcal{X}} = (\mathbf{R}_{\mathcal{E}}\mathbf{A}_{\mathcal{X}})^{\dagger}\mathbf{R}_{\mathcal{E}}\mathbf{z}, \quad [\hat{\mathbf{x}}]_{\mathcal{X}^c} = \mathbf{0}_{|\mathcal{X}^c| \times 1}$$

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

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leqslant C_3 \varepsilon + C_4 \|\mathbf{x} - \mathbf{x}_{\mathcal{X}}\|_1,$$

where the (non-negative) constants C_3 and C_4 depend on the coherence parameters μ_a, μ_b, and μ_m, and on the sparsity levels n_x and n_e.

Proof. The proof is given in Appendix B. □

Theorem 2 and in particular (7) provides a sufficient condition for which DR enables the stable recovery of $\mathbf{x}$ from $\mathbf{z}$. Specifically, (7) states that for a given number of sparse corruptions n_e, the smaller the coherence parameters μ_a, μ_b, and μ_m, the more dominant entries of $\mathbf{x}$ can be recovered stably from $\mathbf{z}$. The case that guarantees the recovery of the largest number of dominant entries in $\mathbf{x}$ is when $\mathbf{A}$ and $\mathbf{B}$ are orthonormal bases (ONBs) ($\mu_a = \mu_b = 0$) that are maximally incoherent ($\mu_m = 1/\sqrt{M}$); this is, for example, the case for the Fourier–identity pair, leading to the recovery condition $n_x n_e < M$.

The recovery guarantee in Theorem 2 generalizes that in [12, Theorem 3] to approximately sparse signals and noisy measurements. Since (7) is identical to the condition [12, Theorem 3] (cf. Table 1) we see that considering approximately sparse signals and (bounded) measurement noise does *not* result in a more restrictive recovery condition. We finally note that the recovery condition in (7) was shown in [12] to be tight for certain signals in the case where $\mathbf{A}$ is the DFT matrix and $\mathbf{B} = \mathbf{I}_M$.

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

Next, we find conditions guaranteeing the stable recovery in the setting where the support set of the interference vector $\mathbf{e}$ is known prior to recovery. A prominent application for this setting is the restoration of saturated signals [7,16]. Here, no knowledge on the locations of the dominant entries of $\mathbf{x}$ is required. The support set $\mathcal{E}$ of the sparse interference vector 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 [1,2,14], as well as sparsity-based in-painting and super-resolution [3,8,15].

The recovery procedure for this case is as follows. Since $\mathcal{E}$ is known prior to recovery, we may recover the vector $\mathbf{x}$ by projecting the noisy observation vector $\mathbf{z}$ onto the orthogonal complement of the range space spanned by $\mathbf{B}_{\mathcal{E}}$ (cf. Section 3.1). 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

$$\mathbf{R}_{\mathcal{E}}\mathbf{z} = \mathbf{R}_{\mathcal{E}}(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}_{\mathcal{E}} + \mathbf{n}) = \mathbf{R}_{\mathcal{E}}\mathbf{A}\mathbf{x} + \mathbf{R}_{\mathcal{E}}\mathbf{n}, \tag{8}$$

where $\mathbf{R}_{\mathcal{E}} = \mathbf{I}_M - \mathbf{B}_{\mathcal{E}}\mathbf{B}_{\mathcal{E}}^{\dagger}$. The following theorem provides a sufficient condition that guarantees the stable restoration of the vector $\mathbf{x}$ from (8).

Theorem 3 *(BP restoration). Let $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e} + \mathbf{n}$ with $\|\mathbf{n}\|_2 \leqslant \varepsilon$. Assume $\mathbf{e}$ to be perfectly n_e-sparse and $\mathcal{E} = \mathrm{supp}(\mathbf{e})$ to be known prior to recovery. Furthermore, let $\mathcal{X} = \mathrm{supp}_{n_x}(\mathbf{x})$. If*

$$2n_x n_e \mu_m^2 < f(2n_x, n_e), \tag{9}$$

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

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

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

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

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

Proof. The proof is given in Appendix C. □

The inequality (9) 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 condition depends on the coherence parameters μ_a, μ_b, and μ_m, and the number of sparse corruptions n_e. As for the case of DR, the situation that guarantees the recovery of the largest number n_x of dominant coefficients in $\mathbf{x}$, is when $\mathbf{A}$ and $\mathbf{B}$ are maximally incoherent ONBs. In this situation, (9) reduces to $2n_x n_e < M$, which is two times more restrictive than that for DR (see [12] for an extensive discussion on this factor-of-two penalty).

The following observations are immediate consequences of Theorem 3:

- If the vector $\mathbf{x}$ is perfectly n_x-sparse and for noiseless measurements, BP-RES using $\eta = 0$ *perfectly* recovers $\mathbf{x}$ if (9) is met. Note that two restoration procedures have been developed for this particular setting in [12, Theorems 4 and 5]. Both methods enable perfect recovery under exactly the same conditions (cf. Table 1). Hence, generalizing the recovery procedure to approximately sparse signals and measurement noise does not incur a penalty in terms of the recovery condition.
- The restoration method in [12, Theorem. 5] requires a column-normalization procedure to guarantee perfect recovery under the condition (9). Since in this special case, BP-RES (with $\eta = 0$) corresponds to BP, Theorem 3 implies that this normalization procedure is not necessary for guaranteeing perfect recovery under (9). Note, however, that this observation does not apply to OMP-based recovery (see [48] for the details).

We finally note that (9) has been shown in [12] to be tight for certain signal and interference pairs in the case where $\mathbf{A}$ is the DFT matrix and $\mathbf{B} = \mathbf{I}_M$.

3.3. BP separation: No knowledge on the support sets

We finally consider the case where no knowledge about the support sets of the approximately sparse vectors $\mathbf{x}$ and $\mathbf{e}$ is available. A typical application scenario is signal separation [4,6,18], 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 $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e} + \mathbf{n}$, followed by computation of the individual signal features according to $\mathbf{A}\mathbf{x}$ and $\mathbf{B}\mathbf{e}$. The main idea underlying this signal-separation approach involves rewriting (1) as

$$\mathbf{z} = \mathbf{D}\mathbf{w} + \mathbf{n}, \tag{10}$$

where $\mathbf{D} = [\mathbf{A}\ \mathbf{B}]$ is the concatenated dictionary of $\mathbf{A}$ and $\mathbf{B}$ and the stacked vector $\mathbf{w}^T = [\mathbf{x}^T\ \mathbf{e}^T]$. Signal separation now amounts to performing BPDN on (10) for recovery of $\mathbf{w}$ from $\mathbf{z}$.

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

$$\mu_d = \max_{i,j,i \neq j} \left| \mathbf{d}_i^H \mathbf{d}_j \right| = \max\{\mu_a, \mu_b, \mu_m\}. \tag{11}$$

One can now use BPDN to recover $\mathbf{w}$ from (10) and invoke Theorem 1 with the recovery condition in (2), resulting in

$$n_w < \frac{1}{2}\left(1 + \frac{1}{\mu_d}\right) \tag{12}$$

with $n_w = n_x + n_e$. It is, however, important to realize that (12) 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 condition (12) does not provide insight into which pairs of dictionaries $\mathbf{A}$ and $\mathbf{B}$ are most useful for signal separation. The following theorem improves upon (12) by taking into account the structure underlying $\mathbf{D}$ and enables us to gain insight into which pairs of dictionaries enable signal separation.

Theorem 4 *(BP separation). Let $\mathbf{z} = \mathbf{D}\mathbf{w} + \mathbf{n}$, with $\mathbf{D} = [\mathbf{A}\ \mathbf{B}]$, $\mathbf{w}^T = [\mathbf{x}^T\ \mathbf{e}^T]$, and $\|\mathbf{n}\|_2 \leqslant \varepsilon$. The dictionary $\mathbf{D}$ is characterized by the coherence parameters μ_a, μ_b, μ_m, and μ_d, and we assume $\mu_b \leqslant \mu_a$ without loss of generality. If*

$$n_w = n_x + n_e < \max\left\{\frac{2(1 + \mu_a)}{\mu_a + 2\mu_d + \sqrt{\mu_a^2 + \mu_m^2}}, \frac{1 + \mu_d}{2\mu_d}\right\}, \tag{13}$$

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

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

using $\varepsilon \leqslant \eta$ *satisfies*

$$\|\mathbf{w} - \hat{\mathbf{w}}\|_2 \leqslant C_7(\varepsilon + \eta) + C_8 \|\mathbf{w} - \mathbf{w}_{\mathcal{W}}\|_1, \tag{14}$$

with $\mathcal{W} = \text{supp}_{n_w}(\mathbf{w})$ *and the* (*non-negative*) *constants* C_7 *and* C_8.

Proof. The proof is given in Appendix D. □

The recovery condition (13) refines that in (12). Consider the two-ONB setting for which $\mu_a = \mu_b = 0$ and $\mu_m = \mu_d$. In this case, (11) corresponds to $n_w = n_x + n_e < (1 + 1/\mu_d)/2$, whereas the condition for BP separation (13) is given by

$$n_w = n_x + n_e < \frac{2}{3\mu_d}. \tag{15}$$

Hence, (13) guarantees the stable recovery for a larger number of dominant entries in the stacked vector $\mathbf{w}^T = [\mathbf{x}^T \ \mathbf{e}^T]$. Recovery guarantees for perfectly sparse signals and noiseless measurements in the case of two ONBs have been developed in [23,24,35,49]. The corresponding recovery condition $n_w = n_x + n_e < (\sqrt{2} - 0.5)/\mu_d$ turns out to be less restrictive than the recovery condition for approximately sparse signals and measurement noise in (15). Whether this behavior is a fundamental result of considering approximately sparse signals and noisy measurements or is an artifact of the proof technique is part of on-going work.

4. Coherence-based bounds on restricted isometry constants

Coherence-based bounds on restricted isometry constants (RICs) are useful to efficiently compute bounds on the RIC that would otherwise require a combinatorial search [50]. Moreover, such bounds enable us to develop an alternative set of recovery conditions from the restricted isometry property (RIP) framework [36,51–56]. We next show RIC bounds all three cases of support-set knowledge and provide a comparison with the recovery guarantees obtained in the previous section, those from the RIP framework, and with existing ones from [12,19] (recall Table 1).

4.1. Restricted isometry property (RIP)

An alternative route of obtaining deterministic recovery guarantees for approximately sparse signals and measurement noise, i.e., for $\mathbf{z} = \mathbf{Dx} + \mathbf{n}$, has been developed under the RIP framework [36,51–56]. There, the dictionary $\mathbf{D}$ is characterized by RICs rather than the coherence parameter μ_d.

Definition 1 (*Restricted isometry constant (RIC)*). (*See* [36].) *For each integer* $n_x \geqslant 1$, *the RIC* δ_{n_x} *of* $\mathbf{D}$ *is the smallest number such that*

$$(1 - \delta_{n_x})\|\mathbf{x}\|_2^2 \leqslant \|\mathbf{Dx}\|_2^2 \leqslant (1 + \delta_{n_x})\|\mathbf{x}\|_2^2 \tag{16}$$

holds for all perfectly n_x*-sparse vectors* $\mathbf{x}$.

Stable recovery of $\mathbf{x}$ with $\|\mathbf{n}\|_2 \leqslant \varepsilon$ using BDPN can be guaranteed if the dictionary $\mathbf{D}$ satisfies a restricted isometry property (RIP), e.g., of the form (i) $\delta_{2n_x} < \sqrt{2} - 1$ [52] or (ii) $\delta_{n_x} < 1/3$ [56]. The main issue with such recovery conditions is the fact that computation of the RIC requires a combinatorial search [50]. Nevertheless, it has been shown in [57,58] that random dictionaries $\mathbf{D}$ (e.g., with i.i.d. (sub-)Gaussian entries) satisfy such RIP conditions with high probability, which is of particular interest in CS [40,41].

To arrive at recovery conditions that are explicit in the number of nonzero entries n_x and can be computed efficiently, one may bound the RIC in (16) using the coherence parameter μ_d as $\delta_{n_x} \leqslant \mu_d(n_x - 1)$ [32,53]. Such bounds in combination with the RIP conditions in [52,56] can be used to derive alternative recovery conditions (i) $n_x < (1 + (\sqrt{2} - 1)/\mu_d)/2$ or (ii) $n_x < 1 + 1/(3\mu_d)$, which are, in general, more restrictive than (2).

4.2. Coherence-based RIC bounds for sparsely corrupted signals

We next provide coherence-based bounds on the RIC for DR, BP-RES, and BP-SEP, and derive corresponding alternative recovery guarantees using results obtained in the RIP framework [52,55].
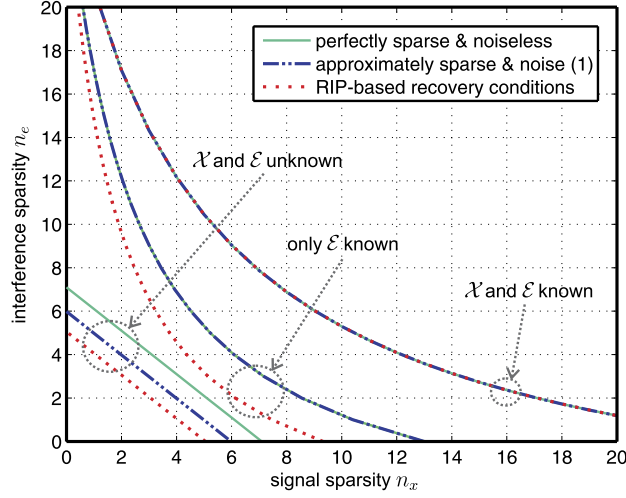
Fig. 1. Comparison of the recovery conditions for $\mu_m = \mu_d = 0.1$ and $\mu_a = \mu_b = 0.04$. For the case "$\mathcal{X}$ and $\mathcal{E}$ known," all three recovery conditions coincide; for the case "only $\mathcal{E}$ known," the conditions for perfectly and approximately sparse signals coincide.

4.2.1. RIC bound for signal restoration

As a byproduct of the proof for BP-RES detailed in Appendix C, the following coherence-based upper bound on the RIC for the matrix $\widetilde{\mathbf{A}} = \mathbf{R}_{\mathcal{E}}\mathbf{A}$ has been obtained:

Lemma 5 (*RIC bound for* $\widetilde{\mathbf{A}}$). *Let* $\widetilde{\mathbf{A}} = \mathbf{R}_{\mathcal{E}}\mathbf{A}$ *with* $\mathbf{R}_{\mathcal{E}} = \mathbf{I}_M - \mathbf{B}_{\mathcal{E}}\mathbf{B}_{\mathcal{E}}^{\dagger}$. *For each integer* $n_x \geqslant 1$, *the smallest number* δ_{n_x} *such that*

$$(1 - \delta_{n_x})\|\mathbf{x}\|_2^2 \leqslant \|\widetilde{\mathbf{A}}\mathbf{x}\|_2^2 \leqslant (1 + \delta_{n_x})\|\mathbf{x}\|_2^2$$

holds for all perfectly n_x-*sparse vectors* $\mathbf{x} \in \mathbb{C}^{N_a}$, *is bounded from above by*

$$\delta_{n_x} \leqslant \mu_a(n_x - 1) + \frac{n_x n_e \mu_m^2}{[1 - \mu_b(n_e - 1)]^+}. \tag{17}$$

For direct recovery (DR), a recovery condition from the RIP framework follows straightforwardly from Lemma 5; i.e., we need $\delta_{n_x} < 1$ to ensure that the inverse of $\widetilde{\mathbf{A}} = \mathbf{R}_{\mathcal{E}}\mathbf{A}$ exists and to enable the stable recovery of $\mathbf{x}$ using DR. For BP restoration, a recovery condition from the RIP framework is obtained by combining (17) with the RIP condition $\delta_{n_x} < 1/3$ from [56]. We note that for $\mu_d < 1/3$ this condition is more restrictive than the condition (9). Hence, for most relevant values of μ_d, the recovery condition in (9) is less restrictive than those obtained from the RIP framework.

4.2.2. RIC bound for signal separation

As a byproduct of the proof for BP-SEP detailed in Appendix D, the following coherence-based upper bound on the RIC for the concatenated dictionary $\mathbf{D} = [\mathbf{A}\,\mathbf{B}]$ has been obtained:

Lemma 6 (*RIC bound for* $\mathbf{D}$). *Let* $\mathbf{D} = [\mathbf{A}\,\mathbf{B}]$ *be characterized by* μ_a, μ_b, μ_m, *and* μ_d, *and assume* $\mu_b \leqslant \mu_a$ *without loss of generality. For each integer* $n_w \geqslant 1$ *the smallest number* δ_{n_w} *such that*

$$(1 - \delta_{n_w})\|\mathbf{w}\|_2^2 \leqslant \|\mathbf{D}\mathbf{w}\|_2^2 \leqslant (1 + \delta_{n_w})\|\mathbf{w}\|_2^2$$

holds for all perfectly n_w-*sparse vectors* $\mathbf{w} \in \mathbb{C}^{N_a + N_b}$, *is bounded by*

$$\delta_{n_w} \leqslant \min\left\{ \frac{1}{2}\left(\mu_a(n_w - 2) + n_w\sqrt{\mu_a^2 + \mu_m^2}\right), \mu_d(n_w - 1) \right\}. \tag{18}$$

As shown above, one can use the right hand side (RHS) of (18) in combination with RIP condition $\delta_{n_w} < 1/3$ of [56] to obtain an alternative recovery guarantee for signal separation. As for BP restoration, this condition is more restrictive than that in (13) in most cases. Surprisingly, in the two-ONB case (i.e., $\mu_a = \mu_b = 0$ and $\mu_m = \mu_d$), the recovery condition for signal separation obtained from the RIP framework coincides to that in (15).

4.3. Comparison of the recovery guarantees

Fig. 1 compares the recovery conditions for the general model (1) to those obtained in [12,19] for perfectly sparse signals and noiseless measurements (see also Table 1), and the conditions derived from the RIP framework. We compare the three different cases analyzed in Section 3:

- *Direct restoration:* For DR, the recovery conditions for the general model (1) detailed in (7), the condition in [12, Eq. 11] for perfectly sparse signals and noiseless measurements, and those obtained through the RIP framework coincide. Hence, the generalization to approximately sparse signals and measurement noise does not incur a degradation in terms of the recovery condition.
- *BP restoration:* The recovery conditions for the general setup considered in this paper and the condition [12, Eq. 14] for perfectly sparse signals and noiseless measurements also coincide. Again, generalizing the results does not incur a loss in terms of the recovery conditions. As expected, the recovery condition obtained trough the RIP framework turns out to be more restrictive (cf. Section 4.2).
- *BP separation:* We see that all of the recovery conditions differ. In particular, the condition [19, Eq. 13] for perfectly sparse signals and noiseless measurements is less restrictive than (1). As expected, the recovery condition from the RIP framework is most restrictive.

In summary, we see that having more knowledge on the support sets prior to recovery yields less restrictive recovery conditions. This intuitive behavior can also be observed in practice and is illustrated in Section 5.

We finally emphasize that all of the recovery conditions derived above are deterministic in nature and therefore conservative in the sense that, in practice, recovery often succeeds for sparsity levels n_x and n_e much higher than the corresponding guarantees indicate. In particular, it is well-known that coherence-based deterministic recovery guarantees are typically limited by the so-called *square-root bottleneck*, e.g., [12,19,42,43], as they are valid for *all* dictionary pairs **A** and **B** with given coherence parameters, and all signal and interference realizations with given sparsity levels n_x and n_e. Nevertheless, we next show that our recovery conditions enable us to gain considerable insights into practical applications; i.e., they are useful for identification of appropriate dictionary pairs that should be used for sparsity-based signal restoration or separation.

5. Application examples

We now develop two application examples to illustrate the main results of the paper. First, we show that direct restoration, BP restoration, and BP separation can be used for simultaneous denoising and declicking of corrupted speech signals. Then, we illustrate the impact of support-set knowledge for a sparsity-based in-painting example.

5.1. Simultaneous denoising and declicking

In this example, we attempt the recovery of a speech signal that has been corrupted by a combination of additive Gaussian noise and impulse noise. The main goal of this example is to illustrate the performance of our algorithms and not to benchmark the performance relative to existing methods; a detailed performance and restoration-complexity comparison with existing methods for simultaneous denoising and declicking is left for future work.

We corrupt a 9.5 s segment (44 100 kHz sampling rate and 16 bit precision) from the speech signal in [59] by adding zero-mean i.i.d. Gaussian noise and impulse noise. The amplitudes of the audio signal has been normalized to the range $[-1, 1]$. The variance of the additive noise is chosen such that the signal-to-noise ratio (SNR) between the L-dimensional original audio signal **y** and the noisy version $\tilde{\mathbf{y}}$, defined as

$$\mathsf{SNR} = 10 \log_{10}\big(\|\mathbf{y}\|_2^2 / \|\mathbf{y} - \tilde{\mathbf{y}}\|_2^2\big),$$

is 10.7 dB. The impulse interference (used to model the clicking artifacts in the audio signal) is generated as follows: We corrupt 10% of the samples and chose the locations of the random clicks, which are modeled by the interference vector **e**, uniformly at random. We then generate the clicks at these locations by adding i.i.d. zero-mean Gaussian random samples with variance 0.1 to the noisy signal. The resulting SNR is 0.31 dB.

5.1.1. Recovery procedure

Recovery is performed with overlapping blocks of dimension $M = 1024$. The amount of overlap between adjacent blocks is 128 samples. We set **A** to the 1024×1024 DCT matrix, $\mathbf{B} = \mathbf{I}_M$, and perform recovery based on $\mathbf{z} = \mathbf{Ax} + \mathbf{e} + \mathbf{n}$. The main reasons for using the DCT matrix in this example are (i) the speech signal is approximately sparse in the DCT basis; (ii) we have $\mu_a = \mu_b = 0$; and (iii) the mutual coherence of the DCT–identity pair is small, i.e., $\mu_m \approx 0.0442$, which leads to less restrictive recovery conditions (7), (9), and (13). For all three recovery methods, we first compute an estimate $\hat{\mathbf{x}}$ of **x** (and of **e** in the case of BP separation) followed by computing an estimate of speech signal according to $\hat{\mathbf{y}} = \mathbf{A}\hat{\mathbf{x}}$. In order to reduce undesired artifacts occurring at the boundaries between two adjacent blocks, we overlap and add the recovered blocks using a raised-cosine window function when re-synthesizing the entire speech signal.
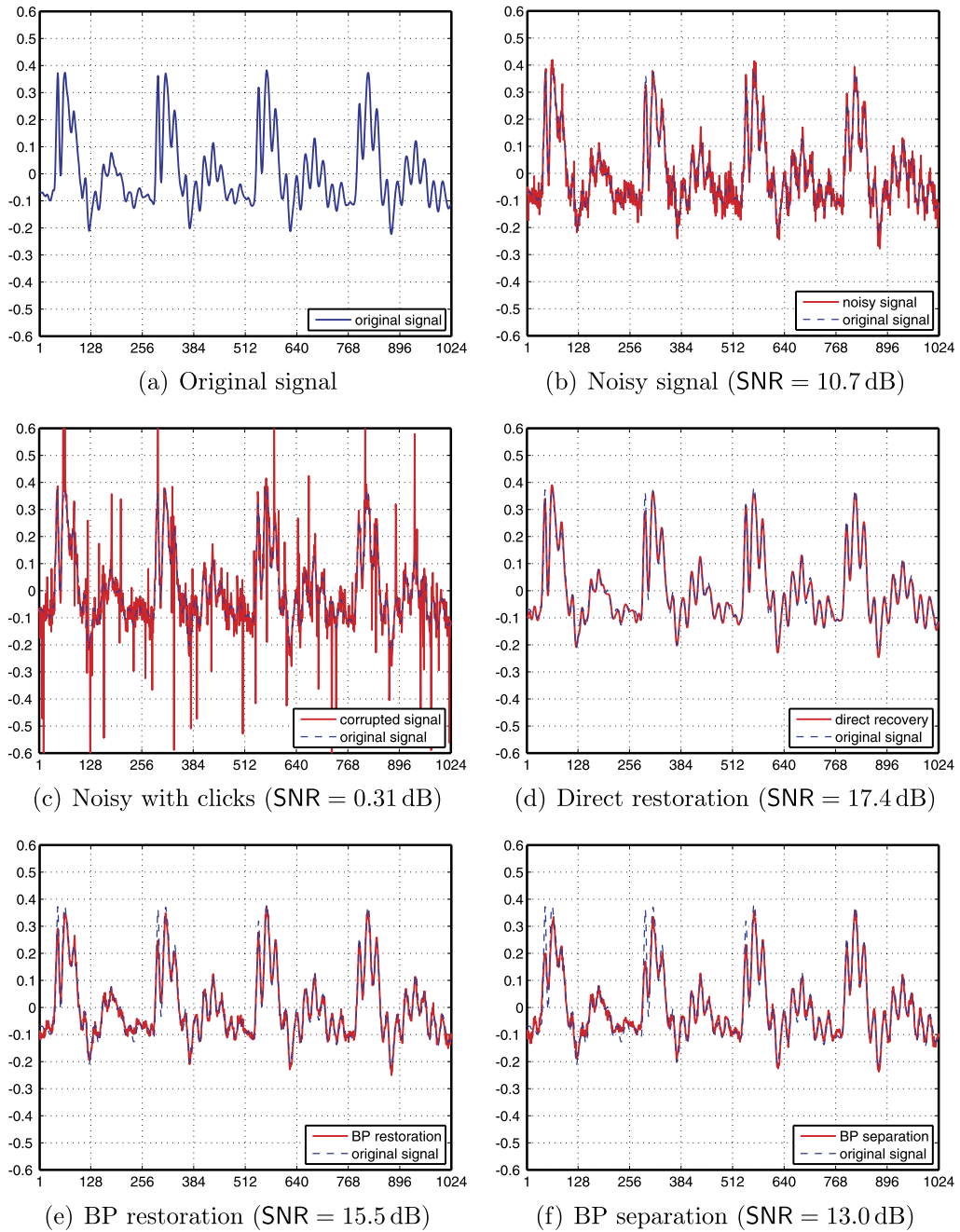
Fig. 2. Signal-to-noise ratio (SNR) results of simultaneous reduction of Gaussian noise and interference in a corrupted speech signal (the *x*-axes correspond to sample indices, the *y*-axes to magnitudes).

5.1.2. Recovery results

Fig. 2 shows snapshots of the corruption and recovery procedure and the associated SNR values. The individual results of the three recovery procedures analyzed in this paper are as follows:

- *Direct restoration:* In this case, the locations $\mathcal{E}$ of the impulse-noise realizations are assumed to be known prior to recovery. To this end, we compute the DCT coefficients of the uncorrupted signal to identify the 128 largest (in magnitude) coefficients in each block. This genie-aided support-set estimate is then used in the DR recovery stage. The SNR of the signal recovered through DR is 17.4 dB and, hence, DR is able to improve the SNR by roughly 17 dB (see Fig. 2(d)).
- *BP restoration:* In this case, the locations of the impulse-noise spikes $\mathcal{E}$ are assumed to be known prior to recovery, but nothing is known about **x**. We perform BP restoration with $\eta = 0.8$, which results in an SNR of 15.5 dB (see Fig. 2(e)).

Table 2

Coherence parameters and SNR of BP-RES for two windowing methods and redundant DCT dictionaries **A** for $M = N_b = 1024$ and a 2×128 raised-cosine window.

N_a	Windowing after BP-RES			Windowing within BP-RES		
	μ_a	μ_m	SNR	μ_a	μ_m	SNR
1024	0	0.0442	15.5	0.3824	0.0522	13.3
1152	0.0447	0.0447	14.9	0.4062	0.0523	12.9
1280	0.0658	0.0449	14.6	0.4127	0.0523	12.7
1536	0.0854	0.0452	14.3	0.3737	0.0521	12.5
2048	0.1147	0.0456	13.8	0.4185	0.0526	12.1

Note that the parameter η determines the amount of denoising (for the Gaussian noise) and is used to optimize the resulting SNR.

- *BP separation:* In this case, we assume that nothing is known about the support sets of either **x** or **e**. We perform BP-SEP with $\eta = 0.8$ and discard the recovered error component **e**; the resulting SNR corresponds to 13.0 dB (see Fig. 2(f)). BP separation achieves surprisingly good recovery performance (compared to DR and BP-RES), while being completely *blind* to the locations of the sparse interference. Hence, BP separation offers an elegant way to mitigate impulse noise in speech signals, without requiring sophisticated algorithms that detect the locations of the sparse interference (e.g., clicks and pops).

5.1.3. Alternative recovery procedure

Rectangular windowing, as used in the example above, is known to yield sub-optimal sparsification of audio signals in the DCT basis. Furthermore, over-complete dictionaries often enable sparser representations than orthonormal bases. Therefore, it is natural to ask what happens if recovery is performed *directly* on windowed audio signals using a redundant dictionary. To answer this question, we alternatively perform signal restoration using BP-RES[2] on the basis of $\mathbf{Wz} = \mathbf{W}(\mathbf{Ax} + \mathbf{Bn} + \mathbf{n})$ rather than on (1), with **W** denoting a diagonal matrix containing the windowing coefficients on the main diagonal. Moreover, we set **A** to a redundant (or over-complete) DCT matrix [60]. To perform windowing within BP-RES, we use $\tilde{\mathbf{z}} = \mathbf{Wz}$, $\widetilde{\mathbf{A}} = \mathbf{WA}$, and $\widetilde{\mathbf{B}} = \mathbf{WB}$, instead of **z**, **A**, and **B**.[3]

Table 2 shows the coherence parameters of **A** and $\widetilde{\mathbf{A}}$, the mutual coherences μ_m, and the recovery SNR for windowing after and within BP-RES, and for different redundancies. We can see that the SNR for the windowing approach within BP-RES is worse than that of the approach used above (cf. Fig. 2). The reason for this behavior is the fact that even if windowing of **A** improves sparsification of **x**, it also increases the coherence parameter μ_a. The recovery condition (9) reflects this behavior and shows that even for small coherence values, a strong sparsification is necessary. Table 2 furthermore shows that increasing the redundancy of **A** also degrades the SNR. This behavior is, once again, a result of the fact that the coherence of **A** (or the windowed version $\widetilde{\mathbf{A}}$) increases with the redundancy. We conclude that windowing within BP-RES and/or increasing the redundancy in **A** does not improve the performance in the considered audio-restoration example.

5.1.4. Discussion of the results

The results shown above show that more knowledge on the support sets $\mathcal{X}$ and/or $\mathcal{E}$ leads to improved recovery results (i.e., larger SNR). We emphasize that DR, BP restoration, and BP separation are all able to *simultaneously* reduce Gaussian noise and impulse interference as the resulting SNR values are all larger than 10.7 dB (corresponding to the SNR of the noisy signal). The recovery procedure one should use in practice depends on the amount of support-set knowledge available prior to recovery.

The use of redundant dictionaries or windowing within the restoration procedure did not show any advantage (over rectangular windowing and the use of ONBs) in the considered example. Nevertheless, we believe that specifically trained (or learned) dictionaries, e.g., using the method in [61], have the potential to further improve the recovery performance; the exploration of methods that also maintain incoherence to other dictionaries is part of on-going work.

We furthermore note that noise and clicks removal in audio signals is a well-studied topic in the literature (see, e.g., [2] and references therein). However, most of the established methods rely on Bayesian estimation techniques, e.g., [2,62,63], which do not have theoretical guarantees. Sparsity-based audio restoration has been proposed recently in [16,17]; however, no recovery guarantees are available for the associated restoration algorithms.

We finally emphasize that virtually all proposed methods require knowledge of the locations of the sparse corruptions prior to recovery, whereas our results for BP separation show that sparse errors can effectively be removed *blindly* from speech signals.

---

[2] The performance for DR and BP-SEP behaves analogously to that of BP-RES.

[3] The dictionaries $\widetilde{\mathbf{A}}$ and $\widetilde{\mathbf{B}}$ were normalized to have unit-norm columns. Note that re-normalizing the windowed identity basis $\widetilde{\mathbf{B}}$ leads to $\mathbf{I}_M$ and, hence, we have $\mu_b = 0$.

(a) Original (b) Corrupted (c) R: DCT–identity (d) S: DCT–identity

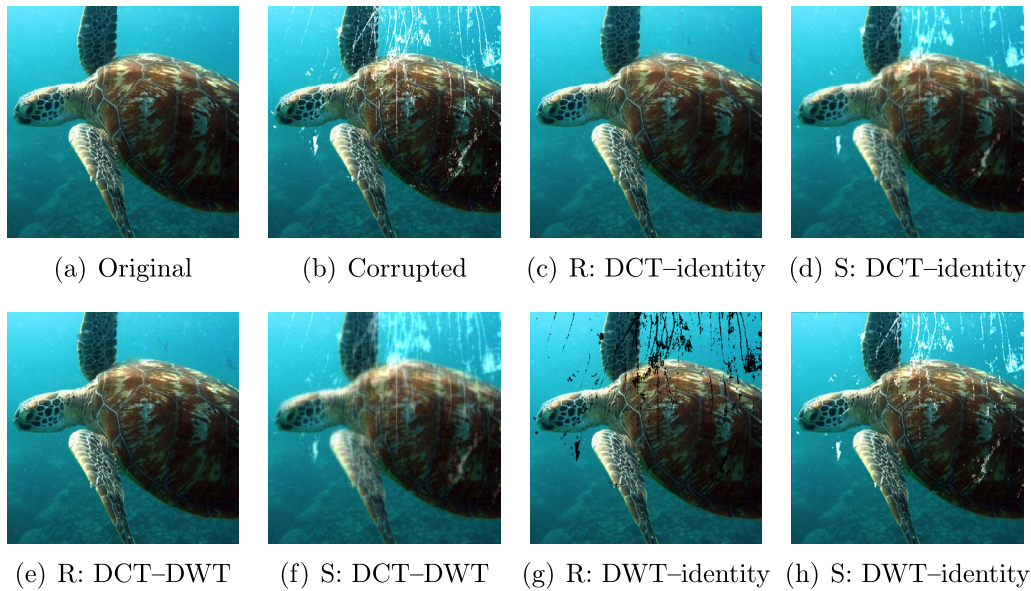(e) R: DCT–DWT (f) S: DCT–DWT (g) R: DWT–identity (h) S: DWT–identity

Fig. 3. Example of BP restoration (R) and BP separation (S) for removal of scratches from a color photograph: (a) original photo (courtesy of Graeme Pope); (b) corrupted photo (10.5 dB); (c) BP-RES (30.3 dB); (d) BP-SEP (15.6 dB); (e) BP-RES (30.3 dB); (f) BP-SEP (15.2 dB); (g) BP-RES (11.2 dB); (h) BP-SEP (12.8 dB). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.2. Removal of scratches in color photographs

We now consider a simple sparsity-based in-painting application. While a plethora of in-painting methods have been proposed in the literature (see, e.g., [6,8–10] and the references therein), our goal here is to not to benchmark our performance vs. theirs but rather to illucidate the differences between BP restoration and BP separation, i.e., to quantify the impact of support-set knowledge and of the coherence parameters on the in-painting performance.

In the following example, we seek to remove scratches from a color photograph, whose color channels are normalized to the range [0, 1]. We corrupt 15% of the pixels of a 512×512 color image by adding a mask containing artificially generated scratch patterns. We consider noiseless measurements and set $\eta = 0$. The SNR of the corrupted color photo shown in Fig. 3(b) for all color channels corresponds to 10.5 dB. In order to demonstrate the recovery performance for approximately sparse signals, the image *has not been sparsified* prior to adding the corruptions, which is in stark contrast to the in-painting example shown in [12].

5.2.1. Restoration procedure

We independently recover each color channel using BP restoration and BP separation of the full 512×512 pixel image, i.e., we have $M = 512^2$ corrupted measurements for each color channel. We consider the following pairs of bases to sparsify images and scratches: (i) a 2-dimensional DCT basis to sparsify images and the identity basis to sparsify scratches, (ii) a 2-dimensional DCT to sparsify images and a discrete wavelet transform (DWT) basis to sparsify scratches, and a DWT to sparsify images and the identity basis to sparsify scratches.[4] 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 image estimate as $\hat{\mathbf{y}} = \mathbf{A}\hat{\mathbf{x}}$. Note that DR is not considered here as information on the location of the dominant entries of $\mathbf{x}$ is difficult to acquire in practice.

5.2.2. Discussion of the results

Fig. 3 shows results of the corruption and recovery procedures along with the associated SNR values. For the DCT-identity pair, we see that the recovered image has an SNR of 30.3 dB for BP restoration (see Fig. 3(c)) and well approximates the ground truth; this is a result of the DCT and identity basis being incoherent (with $\mu_m \approx 0.004$), as reflected by the recovery condition (9). For BP separation (see Fig. 3(d)), the recovery SNR improves by 5.1 dB over the corrupted image, but in parts where large areas of the image are corrupted, blind removal of scratches fails to recover the corrupted entries. Hence, knowing the locations of the sparse corruptions leads to a significant advantage in terms of SNR and is highly desirable for sparsity-based in-painting applications.

Fig. 3 furthermore shows the recovery performance for pairs of matrices that improve sparsification of the image or interference component (compared to the DCT or identity) but have larger mutual coherence. Concretely, the results in

---

[4] We use a Daubechies 9 (DB9) wavelet decomposition on two octaves [64].

Figs. 3(e) and 3(f) assume that the scratches are sparse in the DWT basis. Since the DWT is more coherent ($\mu_m \approx 0.011$) to the DCT than the DCT–identity pair ($\mu_m \approx 0.004$), we obtain slightly worse recovery SNR. In the case of using the DWT to sparsify the image and the identity basis to sparsify the scratches, the recovery procedure fails for both BP-RES and BP-SEP. The reason is the high coherence between the DWT and identity basis, i.e., $\mu_m \approx 0.432$, as reflected by the recovery conditions (9) and (13).

Therefore, we conclude that the proposed recovery conditions (7), (9) help to identify good dictionary pairs for a variety of sparsity-based restoration and separation problems. In particular, they show that the dictionary **A** must both (i) *sparsify* the signal to be recovered and (ii) be *incoherent* with the interference dictionary **B**. Note that the second requirement is satisfied for the DCT–identity pair, whereas other transform bases typically used to sparsify images (i.e., to satisfy only the first requirement), such as DWT bases, exhibit high mutual coherence with the identity basis.

6. Conclusions

In this paper, we have generalized the results presented in [12,19] for the recovery of perfectly sparse signals that are corrupted by perfectly sparse interference to the much more practical case of approximately sparse signals and noisy measurements. We have proposed novel restoration and separation algorithms for three different cases of knowledge on the location of the dominant entries (in terms of absolute value) in the vector **x**, namely (1) direct restoration, (2) BP restoration, and (3) BP separation. Moreover, we have developed deterministic recovery guarantees for all three cases. The application examples have demonstrated that our recovery guarantees explain which dictionary pairs **A** and **B** are most suited for sparsity-based signal restoration or separation. Our comparison of the presented deterministic guarantees with similar ones obtained using the restricted isometry property (RIP) framework and to those provided in [12,19] has shown that, for BP restoration and BP separation, considering the general model does not result in more restrictive recovery conditions. For BP separation, however, the recovery conditions for the general model considered here turn out to be slightly more restrictive as it is for perfectly sparse signals and noiseless measurements.

There are many avenues for follow-on work, and we thank you in advance for the citations. Probabilistic recovery guarantees for the restoration and separation with randomness in the signal and/or interference rather than in the dictionaries have been developed recently in [43]. A generalization to approximately sparse signals and the noisy case is an interesting open research problem. Furthermore, a detailed exploration of more real-world applications using the sparsity-based restoration and separation techniques analyzed in this paper is left for future work. The development of novel dictionary learning algorithms (e.g., based on the method in [61]) that enforce incoherence to the interference dictionary **B** could further improve the performance of signal recovery from sparsely corrupted measurements in practical applications. We finally note that an integrated circuit design making use of the proposed methods for real-time audio declicking has been developed recently in [65]; this further highlights the practical relevance of our methods.

Acknowledgments

Appendix A. Proof of Theorem 1

The proof detailed next follows closely that given in [32, Theorem 2.1] and relies on techniques developed earlier in [21, 32,51].

A.1. Prerequisites

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

A.1.1. 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 [21,51,54]

$$\|\mathbf{h} - \mathbf{h}_0\|_1 \leqslant \|\mathbf{h}_0\|_1 + e_0 \tag{A.1}$$

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

$$\|\mathbf{x}\|_1 \geqslant \|\hat{\mathbf{x}}\|_1 = \|\hat{\mathbf{x}}_{\mathcal{X}}\|_1 + \|\hat{\mathbf{x}}_{\mathcal{X}^c}\|_1 = \|\mathbf{x}_{\mathcal{X}} + \mathbf{h}_0\|_1 + \|\mathbf{h} - \mathbf{h}_0 + \mathbf{x}_{\mathcal{X}^c}\|_1$$
$$\geqslant \|\mathbf{x}_{\mathcal{X}}\|_1 - \|\mathbf{h}_0\|_1 + \|\mathbf{h} - \mathbf{h}_0\|_1 - \|\mathbf{x}_{\mathcal{X}^c}\|_1.$$

Application of the reverse triangle inequality to the left-hand side term of (A.1) yields the following useful bound:

$$\|\mathbf{h}\|_1 \leqslant 2\|\mathbf{h}_0\|_1 + e_0. \tag{A.2}$$

A.1.2. Tube constraint

We furthermore have [51,54]

$$\|\mathbf{A}\mathbf{h}\|_2 = \left\|\mathbf{A}\hat{\mathbf{x}} - \mathbf{y} - (\mathbf{A}\mathbf{x} - \mathbf{y})\right\|_2$$
$$\leqslant \|\mathbf{A}\hat{\mathbf{x}} - \mathbf{y}\|_2 + \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leqslant \eta + \varepsilon. \tag{A.3}$$

A.1.3. Coherence-based restricted isometry property (RIP)

Since $\mathbf{h}_0$ is perfectly n_x-sparse, Geršgorin's disc theorem [66, Theorem 6.1.1] applied to $\|\mathbf{A}\mathbf{h}_0\|_2^2$ yields

$$\left(1 - \mu_a(n_x - 1)\right)\|\mathbf{h}_0\|_2^2 \leqslant \|\mathbf{A}\mathbf{h}_0\|_2^2 \leqslant \left(1 + \mu_a(n_x - 1)\right)\|\mathbf{h}_0\|_2^2. \tag{A.4}$$

A.2. Bounding the error $\|\mathbf{h}_0\|_2$ on the signal support

The goal of the following steps is to bound the recovery error $\|\mathbf{h}_0\|_2$ on the support set $\mathcal{X}$. We follow the steps in [32] to arrive at the following chaîne d'inégalités:

$$\left|\mathbf{h}^H \mathbf{A}^H \mathbf{A}\mathbf{h}_0\right| \geqslant \left|\mathbf{h}_0^H \mathbf{A}^H \mathbf{A}\mathbf{h}_0\right| - \left|(\mathbf{h} - \mathbf{h}_0)^H \mathbf{A}^H \mathbf{A}\mathbf{h}_0\right|$$
$$\geqslant \left(1 - \mu_a(n_x - 1)\right)\|\mathbf{h}_0\|_2^2 - \left|\sum_{k \in \mathcal{X}} \sum_{\ell \in \mathcal{X}^c} [\mathbf{h}_0^H]_k \mathbf{a}_k^H \mathbf{a}_\ell [\mathbf{h}]_\ell\right| \tag{A.5}$$
$$\geqslant \left(1 - \mu_a(n_x - 1)\right)\|\mathbf{h}_0\|_2^2 - \mu_a\|\mathbf{h}_0\|_1\|\mathbf{h} - \mathbf{h}_0\|_1 \tag{A.6}$$
$$\geqslant \left(1 - \mu_a(n_x - 1)\right)\|\mathbf{h}_0\|_2^2 - \mu_a\|\mathbf{h}_0\|_1\left(\|\mathbf{h}_0\|_1 + e_0\right) \tag{A.7}$$
$$\geqslant \left(1 - \mu_a(n_x - 1)\right)\|\mathbf{h}_0\|_2^2 - \mu_a n_x\|\mathbf{h}_0\|_2^2 - \mu_a\sqrt{n_x}\|\mathbf{h}_0\|_2 e_0 \tag{A.8}$$
$$= \left(1 - \mu_a(2n_x - 1)\right)\|\mathbf{h}_0\|_2^2 - \mu_a\sqrt{n_x}\|\mathbf{h}_0\|_2 e_0, \tag{A.9}$$

where (A.5) follows from (A.4), (A.6) is a consequence of $|\mathbf{a}_k^H \mathbf{a}_\ell| \leqslant \mu_a$, $\forall k \neq \ell$, (A.7) results from the cone constraint (A.1), and (A.8) from the Cauchy–Schwarz inequality. We emphasize that (A.9) is crucial, since it determines the recovery condition for BPDN. In particular, if the first RHS term in (A.9) satisfies $(1 - \mu_a(2n_x - 1)) > 0$ and $\mathbf{h}_0 \neq \mathbf{0}_{N_a \times 1}$, then the error $\|\mathbf{h}_0\|_2$ is bounded from above as follows:

$$\|\mathbf{h}_0\|_2 \leqslant \frac{|\mathbf{h}^H \mathbf{A}^H \mathbf{A}\mathbf{h}_0| + \mu_a\sqrt{n_x}\|\mathbf{h}_0\|_2 e_0}{(1 - \mu_a(2n_x - 1))\|\mathbf{h}_0\|_2} \tag{A.10}$$
$$\leqslant \frac{\|\mathbf{A}\mathbf{h}\|_2\|\mathbf{A}\mathbf{h}_0\|_2 + \mu_a\sqrt{n_x}\|\mathbf{h}_0\|_2 e_0}{(1 - \mu_a(2n_x - 1))\|\mathbf{h}_0\|_2} \tag{A.11}$$
$$\leqslant \frac{(\varepsilon + \eta)\sqrt{1 + \mu_a(n_x - 1)}\|\mathbf{h}_0\|_2 + \mu_a\sqrt{n_x}\|\mathbf{h}_0\|_2 e_0}{(1 - \mu_a(2n_x - 1))\|\mathbf{h}_0\|_2} \tag{A.12}$$
$$= \frac{(\varepsilon + \eta)\sqrt{1 + \mu_a(n_x - 1)} + \mu_a\sqrt{n_x} e_0}{1 - \mu_a(2n_x - 1)}. \tag{A.13}$$

Here, (A.10) is a consequence of (A.9), (A.11) follows from the Cauchy–Schwarz inequality, and (A.12) results from the tube constraint (A.3) and the RIP (A.4). The case $\mathbf{h}_0 = \mathbf{0}_{N_a \times 1}$ is trivial as it implies $\|\mathbf{h}_0\|_2 = 0$.

A.3. Bounding the recovery error $\|\mathbf{h}\|_2$

We are now ready to derive an upper bound on the recovery error $\|\mathbf{h}\|_2$. To this end, we first bound $\|\mathbf{A}\mathbf{h}\|_2^2$ from below as in [32]

$$\|\mathbf{A}\mathbf{h}\|_2^2 = \mathbf{h}^H \mathbf{A}^H \mathbf{A}\mathbf{h} = \sum_{k,\ell} [\mathbf{h}^H]_k \mathbf{a}_k^H \mathbf{a}_\ell [\mathbf{h}]_\ell$$
$$= \sum_k \|\mathbf{a}_k\|_2^2 |[\mathbf{h}]_k|^2 + \sum_{k,\ell,k \neq \ell} [\mathbf{h}^H]_k \mathbf{a}_k^H \mathbf{a}_\ell [\mathbf{h}]_\ell$$

$$\geqslant \|\mathbf{h}\|_2^2 - \mu_a \sum_{k,\ell, k \neq \ell} \big|[\mathbf{h}^H]_k[\mathbf{h}]_\ell\big| \tag{A.14}$$

$$= \|\mathbf{h}\|_2^2 + \mu_a \sum_k \big|[\mathbf{h}]_k\big|^2 - \mu_a \sum_{k,\ell} \big|[\mathbf{h}^H]_k[\mathbf{h}]_\ell\big|$$

$$= (1 + \mu_a)\|\mathbf{h}\|_2^2 - \mu_a\|\mathbf{h}\|_1^2, \tag{A.15}$$

where (A.14) follows from $\|\mathbf{a}_k\|_2 = 1$, $\forall k$, and $|\mathbf{a}_k^H \mathbf{a}_\ell| \leqslant \mu_a$, $\forall k \neq \ell$. With (A.15), the recovery error can be bounded as

$$\|\mathbf{h}\|_2^2 \leqslant \frac{\|\mathbf{Ah}\|_2^2 + \mu_a\|\mathbf{h}\|_1^2}{1 + \mu_a} \leqslant \frac{(\varepsilon + \eta)^2 + \mu_a(2\|\mathbf{h}_0\|_1 + e_0)^2}{1 + \mu_a}, \tag{A.16}$$

where (A.2) is used to arrive at (A.16). By taking the square root of (A.16) and applying the Cauchy–Schwarz inequality, we arrive at the following bound:

$$\|\mathbf{h}\|_2 \leqslant \frac{\sqrt{(\varepsilon + \eta)^2 + \mu_a(2\|\mathbf{h}_0\|_1 + e_0)^2}}{\sqrt{1 + \mu_a}}$$

$$\leqslant \frac{(\varepsilon + \eta) + \sqrt{\mu_a}(2\|\mathbf{h}_0\|_1 + e_0)}{\sqrt{1 + \mu_a}}. \tag{A.17}$$

Finally, using $\|\mathbf{h}_0\|_1 \leqslant \sqrt{n_x}\|\mathbf{h}_0\|_2$ with the bound in (A.13) followed by algebraic simplifications yields

$$\|\mathbf{h}\|_2 \leqslant \frac{(\varepsilon + \eta) + \sqrt{\mu_a}(2\sqrt{n_x}\|\mathbf{h}_0\|_2 + e_0)}{\sqrt{1 + \mu_a}}$$

$$\leqslant (\varepsilon + \eta)\frac{1 - \mu_a(2n_x - 1) + 2\sqrt{\mu_a n_x}\sqrt{1 + \mu_a(n_x - 1)}}{\sqrt{1 + \mu_a}(1 - \mu_a(2n_x - 1))}$$

$$+ e_0 \frac{\sqrt{\mu_a + \mu_a^2}}{(1 - \mu_a(2n_x - 1))} = C_0(\eta + \varepsilon) + C_1\|\mathbf{x} - \mathbf{x}_{\mathcal{X}}\|_1,$$

which concludes the proof. We note that by imposing a more restrictive condition than $n_x < (1 + 1/\mu_a)/2$ in (2), one may arrive at smaller constants C_0 and C_1 (see [54] for the details).

Appendix B. Proof of Theorem 2

The proof is accomplished by deriving an upper bound on the residual errors resulting from direct restoration. Furthermore, we show that the recovery condition (7) guarantees the existence of $\mathbf{R}_{\mathcal{E}} = \mathbf{I}_M - \mathbf{B}_{\mathcal{E}}\mathbf{B}_{\mathcal{E}}^{\dagger}$ and of the pseudo-inverse $(\mathbf{R}_{\mathcal{E}}\mathbf{A}_{\mathcal{X}})^{\dagger}$.

B.1. Bounding the recovery error

We start by bounding the recovery error of DR as

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leqslant \|\mathbf{x}_{\mathcal{X}} - \hat{\mathbf{x}}_{\mathcal{X}}\|_2 + \|\mathbf{x}_{\mathcal{X}^c} - \hat{\mathbf{x}}_{\mathcal{X}^c}\|_2 \leqslant \|\mathbf{x}_{\mathcal{X}} - \hat{\mathbf{x}}_{\mathcal{X}}\|_2 + \|\mathbf{x}_{\mathcal{X}^c}\|_1. \tag{B.1}$$

The only term in (B.1) that needs further investigation is $\|\mathbf{x}_{\mathcal{X}} - \hat{\mathbf{x}}_{\mathcal{X}}\|_2$. As shown in (5), we have

$$[\hat{\mathbf{x}}]_{\mathcal{X}} = (\mathbf{R}_{\mathcal{E}}\mathbf{A}_{\mathcal{X}})^{\dagger}\mathbf{R}_{\mathcal{E}}\mathbf{z} = \mathbf{x}_{\mathcal{X}} + (\mathbf{R}_{\mathcal{E}}\mathbf{A}_{X})^{\dagger}\mathbf{R}_{\mathcal{E}}(\mathbf{A}\mathbf{x}_{\mathcal{X}^c} + \mathbf{n})$$

and hence, it follows that

$$\|\mathbf{x}_{\mathcal{X}} - \hat{\mathbf{x}}_{\mathcal{X}}\|_2 \leqslant \big\|(\mathbf{R}_{\mathcal{E}}\mathbf{A}_{\mathcal{X}})^{\dagger}\mathbf{R}_{\mathcal{E}}\mathbf{v}\big\|_2, \tag{B.2}$$

where $\mathbf{v} = \mathbf{A}\mathbf{x}_{\mathcal{X}^c} + \mathbf{n}$ represents the residual error term. The remainder of the proof amounts to deriving an upper bound on the RHS in (B.2). We start with the definition of the pseudo-inverse

$$\big\|(\mathbf{R}_{\mathcal{E}}\mathbf{A}_{\mathcal{X}})^{\dagger}\mathbf{R}_{\mathcal{E}}\mathbf{v}\big\|_2 = \big\|(\mathbf{A}_{\mathcal{X}}^H\mathbf{R}_{\mathcal{E}}\mathbf{A}_{\mathcal{X}})^{-1}\mathbf{A}_{\mathcal{X}}^H\mathbf{R}_{\mathcal{E}}\mathbf{v}\big\|_2 \tag{B.3}$$

$$\leqslant \big\|(\mathbf{A}_{\mathcal{X}}^H\mathbf{R}_{\mathcal{E}}\mathbf{A}_{\mathcal{X}})^{-1}\big\|_2\big\|\mathbf{A}_{\mathcal{X}}^H\mathbf{R}_{\mathcal{E}}\mathbf{v}\big\|_2, \tag{B.4}$$

where (B.3) is a consequence of $\mathbf{R}_{\mathcal{E}} = \mathbf{R}_{\mathcal{E}}^H\mathbf{R}_{\mathcal{E}}$, and (B.4) follows from the Rayleigh–Ritz theorem [66, Theorem 4.2.2]. We next individually bound the RHS terms in (B.4) from above.

B.2. Bounding the ℓ_2-norm of the inverse

The bound on the norm of the inverse in (B.4) is based upon an idea developed in [25]. Specifically, we use the Neumann series $(\mathbf{I}_{|\mathcal{X}|} - \mathbf{K})^{-1} = \mathbf{I}_{|\mathcal{X}|} + \sum_{k=1}^{\infty} \mathbf{K}^k$ [67, Lem. 2.3.3] to obtain

$$
\begin{aligned}
\left\| \left(\mathbf{A}_{\mathcal{X}}^H \mathbf{R}_{\mathcal{E}} \mathbf{A}_{\mathcal{X}} \right)^{-1} \right\|_2 &= \left\| \left(\mathbf{A}_{\mathcal{X}}^H \mathbf{A}_{\mathcal{X}} - \mathbf{A}_{\mathcal{X}}^H \mathbf{B}_{\mathcal{E}} \mathbf{B}_{\mathcal{E}}^{\dagger} \mathbf{A}_{\mathcal{X}} \right)^{-1} \right\|_2 \\
&= \left\| (\mathbf{I}_{|\mathcal{X}|} - \mathbf{K})^{-1} \right\|_2 = \left\| \mathbf{I}_{|\mathcal{X}|} + \sum_{k=1}^{\infty} \mathbf{K}^k \right\|_2 \\
&\leqslant 1 + \left\| \sum_{k=1}^{\infty} \mathbf{K}^k \right\|_2 \leqslant 1 + \sum_{k=1}^{\infty} \|\mathbf{K}\|_2^k = \frac{1}{1 - \|\mathbf{K}\|_2},
\end{aligned}
\tag{B.5}
$$

which is guaranteed to exist whenever $\|\mathbf{K}\|_2 < 1$ with $\mathbf{K} = \mathbf{I}_{|\mathcal{X}|} - \mathbf{A}_{\mathcal{X}}^H \mathbf{A}_{\mathcal{X}} + \mathbf{A}_{\mathcal{X}}^H \mathbf{B}_{\mathcal{E}} \mathbf{B}_{\mathcal{E}}^{\dagger} \mathbf{A}_{\mathcal{X}}$. We next derive a sufficient condition for which $\|\mathbf{K}\|_2 < 1$ and, hence, the matrix $\mathbf{A}_{\mathcal{X}}^H \mathbf{R}_{\mathcal{E}} \mathbf{A}_{\mathcal{X}}$ is invertible. We bound $\|\mathbf{K}\|_2$ from above as

$$
\|\mathbf{K}\|_2 \leqslant \left\| \mathbf{I}_{|\mathcal{X}|} - \mathbf{A}_{\mathcal{X}}^H \mathbf{A}_{\mathcal{X}} \right\|_2 + \left\| \mathbf{A}_{\mathcal{X}}^H \mathbf{B}_{\mathcal{E}} \mathbf{B}_{\mathcal{E}}^{\dagger} \mathbf{A}_{\mathcal{X}} \right\|_2
\tag{B.6}
$$

$$
\leqslant \mu_a(n_x - 1) + \left\| \mathbf{A}_{\mathcal{X}}^H \mathbf{B}_{\mathcal{E}} \left(\mathbf{B}_{\mathcal{E}}^H \mathbf{B}_{\mathcal{E}} \right)^{-1} \mathbf{B}_{\mathcal{E}}^H \mathbf{A}_{\mathcal{X}} \right\|_2,
\tag{B.7}
$$

where (B.6) results from the triangle inequality and (B.7) is a consequence of Geršgorin's disc theorem [66, Theorem 6.1.1] applied to the ℓ_2-norm of the hollow matrix $\mathbf{I}_{|\mathcal{X}|} - \mathbf{A}_{\mathcal{X}}^H \mathbf{A}_{\mathcal{X}}$. We next bound the RHS term in (B.7) as

$$
\left\| \mathbf{A}_{\mathcal{X}} \mathbf{B}_{\mathcal{E}} \left(\mathbf{B}_{\mathcal{E}}^H \mathbf{B}_{\mathcal{E}} \right)^{-1} \mathbf{B}_{\mathcal{E}}^H \mathbf{A}_{\mathcal{X}} \right\|_2 \leqslant \left\| \mathbf{A}_{\mathcal{X}}^H \mathbf{B}_{\mathcal{E}} \right\|_2 \left\| \left(\mathbf{B}_{\mathcal{E}}^H \mathbf{B}_{\mathcal{E}} \right)^{-1} \right\|_2 \left\| \mathbf{B}_{\mathcal{E}}^H \mathbf{A}_{\mathcal{X}} \right\|_2
\tag{B.8}
$$

$$
\leqslant \left\| \mathbf{A}_{\mathcal{X}}^H \mathbf{B}_{\mathcal{E}} \right\|_F^2 \left\| \left(\mathbf{B}_{\mathcal{E}}^H \mathbf{B}_{\mathcal{E}} \right)^{-1} \right\|_2
\tag{B.9}
$$

$$
\leqslant n_x n_e \mu_m^2 \left\| \left(\mathbf{B}_{\mathcal{E}}^H \mathbf{B}_{\mathcal{E}} \right)^{-1} \right\|_2
\tag{B.10}
$$

$$
\leqslant \frac{n_x n_e \mu_m^2}{\lambda_{\min}(\mathbf{B}_{\mathcal{E}}^H \mathbf{B}_{\mathcal{E}})} \leqslant \frac{n_x n_e \mu_m^2}{[1 - \mu_b(n_e - 1)]^+},
\tag{B.11}
$$

where (B.8) follows from the ℓ_2-matrix-norm bound, (B.9) from $\|\mathbf{A}_{\mathcal{X}}^H \mathbf{B}_{\mathcal{E}}\|_2 \leqslant \|\mathbf{A}_{\mathcal{X}}^H \mathbf{B}_{\mathcal{E}}\|_F$ and $\|\mathbf{A}_{\mathcal{X}}^H \mathbf{B}_{\mathcal{E}}\|_F = \|\mathbf{B}_{\mathcal{E}}^H \mathbf{A}_{\mathcal{X}}\|_F$, and (B.10) from

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

Note that (B.11) requires $n_e < 1 + 1/\mu_b$, which provides a sufficient condition for when the pseudo-inverse $\mathbf{B}_{\mathcal{E}}^{\dagger}$ exists.

Combining (B.5), (B.7), and (B.11) yields the upper bound

$$
\left\| \left(\mathbf{A}_{\mathcal{X}}^H \mathbf{R}_{\mathcal{E}} \mathbf{A}_{\mathcal{X}} \right)^{-1} \right\|_2 \leqslant \frac{1}{1 - \mu_a(n_x - 1) - \frac{n_x n_e \mu_m^2}{[1 - \mu_b(n_e - 1)]^+}},
\tag{B.12}
$$

which requires

$$
1 - \mu_a(n_x - 1) - \frac{n_x n_e \mu_m^2}{[1 - \mu_b(n_e - 1)]^+} > 0
\tag{B.13}
$$

for the matrix $(\mathbf{A}_{\mathcal{X}}^H \mathbf{R}_{\mathcal{E}} \mathbf{A}_{\mathcal{X}})^{-1}$ to exist. We emphasize that the condition (B.13) determines the recovery condition for DR (7). In particular, if

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

then (B.13) and $n_e < 1 + 1/\mu_b$ are both satisfied and, hence, the recovery matrix $(\mathbf{R}_{\mathcal{E}} \mathbf{A}_{\mathcal{X}})^{\dagger} \mathbf{R}_{\mathcal{E}}$ required for DR exists.

B.3. Bounding the residual error term

We now derive an upper bound on the RHS residual error term in (B.4) according to

$$
\begin{aligned}
\left\| \mathbf{A}_{\mathcal{X}}^H \mathbf{R}_{\mathcal{E}} \mathbf{v} \right\|_2 &\leqslant \left\| \mathbf{A}_{\mathcal{X}}^H \mathbf{v} \right\|_2 + \left\| \mathbf{A}_{\mathcal{X}}^H \mathbf{B}_{\mathcal{E}} \mathbf{B}_{\mathcal{E}}^{\dagger} \mathbf{v} \right\|_2 \\
&\leqslant \sqrt{n_x} \|\mathbf{v}\|_2 + \left\| \mathbf{A}_{\mathcal{X}}^H \mathbf{B}_{\mathcal{E}} \mathbf{B}_{\mathcal{E}}^{\dagger} \mathbf{v} \right\|_2,
\end{aligned}
\tag{B.14}
$$

where (B.14) is a result of

$$\big\| \mathbf{A}_{\mathcal{X}}^H \mathbf{v} \big\|_2 = \sqrt{\sum_{k \in \mathcal{X}} \big| \mathbf{a}_k^H \mathbf{v} \big|^2} \leqslant \sqrt{\sum_{k \in \mathcal{X}} \| \mathbf{a}_k \|_2^2 \| \mathbf{v} \|_2^2} \leqslant \sqrt{n_x} \| \mathbf{v} \|_2. \tag{B.15}$$

The bound on the second RHS term in (B.14) is obtained by carrying out similar steps used to arrive at (B.11), i.e.,

$$\begin{aligned}
\big\| \mathbf{A}_{\mathcal{X}}^H \mathbf{B}_{\mathcal{E}} \mathbf{B}_{\mathcal{E}}^\dagger \mathbf{v} \big\|_2 &\leqslant \big\| \mathbf{A}_{\mathcal{X}}^H \mathbf{B}_{\mathcal{E}} \big\|_2 \big\| (\mathbf{B}_{\mathcal{E}}^H \mathbf{B}_{\mathcal{E}})^{-1} \big\|_2 \big\| \mathbf{B}_{\mathcal{E}}^H \mathbf{v} \big\|_2 \\
&\leqslant \frac{\sqrt{n_x n_e \mu_m^2}}{\lambda_{\min}(\mathbf{B}_{\mathcal{E}}^H \mathbf{B}_{\mathcal{E}})} \big\| \mathbf{B}_{\mathcal{E}}^H \mathbf{v} \big\|_2 \\
&\leqslant \frac{n_e \sqrt{n_x \mu_m^2}}{[1 - \mu_b(n_e - 1)]^+} \| \mathbf{v} \|_2.
\end{aligned} \tag{B.16}$$

Finally, we bound the ℓ_2-norm of the residual error term according to

$$\| \mathbf{v} \|_2 = \| \mathbf{A}\mathbf{x}_{\mathcal{X}^c} + \mathbf{n} \|_2 \leqslant \| \mathbf{A}\mathbf{x}_{\mathcal{X}^c} \|_2 + \| \mathbf{n} \|_2 \leqslant \| \mathbf{x}_{\mathcal{X}^c} \|_1 + \| \mathbf{n} \|_2 \tag{B.17}$$

since

$$\| \mathbf{A}\mathbf{x}_{\mathcal{X}^c} \|_2 = \bigg\| \sum_{k \in \mathcal{X}^c} \mathbf{a}_k [\mathbf{x}]_k \bigg\|_2 \leqslant \sum_{k \in \mathcal{X}^c} \big\| \mathbf{a}_k [\mathbf{x}]_k \big\|_2 = \| \mathbf{x}_{\mathcal{X}^c} \|_1.$$

B.4. Putting the pieces together

In order to bound the recovery error on the support set $\mathcal{X}$, we combine (B.12) with (B.14) and (B.16) to arrive at

$$\begin{aligned}
\big\| (\mathbf{R}_{\mathcal{E}} \mathbf{A}_{\mathcal{X}})^\dagger \mathbf{R}_{\mathcal{E}} \mathbf{v} \big\|_2 &\leqslant \frac{([1 - \mu_b(n_e - 1)]^+ + n_e \mu_m)\sqrt{n_x}}{(1 - \mu_a(n_x - 1))[1 - \mu_b(n_e - 1)]^+ - n_x n_e \mu_m^2} \| \mathbf{v} \|_2 \\
&= c \| \mathbf{v} \|_2.
\end{aligned} \tag{B.18}$$

Finally, using (B.17) in combination with (B.1), (B.2), and (B.18) leads to

$$\| \mathbf{x} - \hat{\mathbf{x}} \|_2 \leqslant c\varepsilon + (c + 1)\| \mathbf{x} - \mathbf{x}_{\mathcal{X}} \|_1 = C_3 \varepsilon + C_4 \| \mathbf{x} - \mathbf{x}_{\mathcal{X}} \|_1,$$

which concludes the proof.

Appendix C. Proof of Theorem 3

We first derive a set of key properties of the matrix $\widetilde{\mathbf{A}} = \mathbf{R}_{\mathcal{E}} \mathbf{A}$, which are then used to prove the main result following the steps in Appendix A.

C.1. Properties of the matrix $\widetilde{\mathbf{A}}$

BP restoration operates on the input–output relation

$$\mathbf{R}_{\mathcal{E}} \mathbf{z} = \mathbf{R}_{\mathcal{E}} (\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{e}_{\mathcal{E}} + \mathbf{n}) = \widetilde{\mathbf{A}}\mathbf{x} + \mathbf{R}_{\mathcal{E}}\mathbf{n} \tag{C.1}$$

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 restoration (9), which will be derived next, also ensures that $\mathbf{R}_{\mathcal{E}}$ exists; this is due to fact that the recovery condition for DR (7) ensures that $\mathbf{R}_{\mathcal{E}}$ exists and the condition for BP restoration (9) is met whenever (7) is satisfied.

In order to adapt the proof in Appendix A for the projected input–output relation in (C.1), the following properties of $\widetilde{\mathbf{A}}$ are required.

C.1.1. Tube constraint

Analogously to Appendix A, we obtain

$$\| \widetilde{\mathbf{A}} \mathbf{h} \|_2 \leqslant \big\| \mathbf{R}_{\mathcal{E}} (\mathbf{A}\hat{\mathbf{x}} - \mathbf{z}) \big\|_2 + \big\| \mathbf{R}_{\mathcal{E}} (\mathbf{A}\mathbf{x} - \mathbf{z}) \big\|_2 \leqslant \eta + \| \mathbf{R}_{\mathcal{E}} \mathbf{n} \|_2 \leqslant \eta + \varepsilon,$$

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

C.1.2. Coherence-based bound on the RIC

We next compute a coherence-based bound on the RIC for the matrix $\widetilde{\mathbf{A}}$. To this end, let $\mathbf{h}_0$ be perfectly n_x-sparse and

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

$$\leqslant \left(1 + \mu_a(n_x - 1) \right) \|\mathbf{h}_0\|_2^2 + \left| \mathbf{h}_0^H \mathbf{A}^H \mathbf{B}_\mathcal{E} \mathbf{B}_\mathcal{E}^\dagger \mathbf{A}\mathbf{h}_0 \right|, \tag{C.3}$$

where (C.2) follows from $\mathbf{R}_\mathcal{E}^H \mathbf{R}_\mathcal{E} = \mathbf{R}_\mathcal{E}$ and (C.3) from Geršgorin's disc theorem [66, Theorem 6.1.1]. Next, we bound the second RHS term in (C.3) as follows:

$$\left| \mathbf{h}_0^H \mathbf{A}^H \mathbf{B}_\mathcal{E} \mathbf{B}_\mathcal{E}^\dagger \mathbf{A}\mathbf{h}_0 \right| = \left| \mathbf{h}_0^H \mathbf{A}^H \mathbf{B}_\mathcal{E} \left(\mathbf{B}_\mathcal{E}^H \mathbf{B}_\mathcal{E} \right)^{-1} \mathbf{B}_\mathcal{E}^H \mathbf{A}\mathbf{h}_0 \right|$$

$$\leqslant \lambda_{\min}^{-1} \left(\mathbf{B}_\mathcal{E}^H \mathbf{B}_\mathcal{E} \right) \left\| \mathbf{B}_\mathcal{E}^H \mathbf{A}\mathbf{h}_0 \right\|_2^2 \tag{C.4}$$

$$\leqslant \lambda_{\min}^{-1} \left(\mathbf{B}_\mathcal{E}^H \mathbf{B}_\mathcal{E} \right) \left\| \mathbf{B}_\mathcal{E}^H \mathbf{A}_\mathcal{X} \right\|_2^2 \|\mathbf{h}_0\|_2^2 \tag{C.5}$$

$$\leqslant \frac{n_x n_e \mu_m^2}{[1 - \mu_b(n_e - 1)]^+} \|\mathbf{h}_0\|_2^2, \tag{C.6}$$

where (C.4) follows from [66, Theorem. 4.2.2], (C.5) from the ℓ_2-norm inequality. The inequality (C.6) results from

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

Note that (C.6) 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 in (7) is satisfied. Combining (C.3) with (C.6) results in

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

$$= (1 + \hat{\delta}) \|\mathbf{h}_0\|_2^2. \tag{C.7}$$

We next compute the lower bound as

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

$$\geqslant \left(1 - \mu_a(n_x - 1) \right) \|\mathbf{h}_0\|_2^2 - \left| \mathbf{h}_0^H \mathbf{A}^H \mathbf{B}_\mathcal{E} \mathbf{B}_\mathcal{E}^\dagger \mathbf{A}\mathbf{h}_0 \right| \tag{C.8}$$

$$\geqslant \left(1 - \mu_a(n_x - 1) \right) \|\mathbf{h}_0\|_2^2 - \frac{n_x n_e \mu_m^2}{[1 - \mu_b(n_e - 1)]^+} \|\mathbf{h}_0\|_2^2 \tag{C.9}$$

$$\geqslant \left(1 - \mu_a(n_x - 1) - \frac{n_x n_e \mu_m^2}{[1 - \mu_b(n_e - 1)]^+} \right) \|\mathbf{h}_0\|_2^2 \tag{C.10}$$

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

where (C.8) follows from [66, Theorem. 6.1.1] and (C.9) is obtained by carrying out similar steps used to arrive at (C.6). Note that (C.7) and (C.10) provide a coherence-based upper bound $\hat{\delta}$ on the RIC of the matrix $\widetilde{\mathbf{A}} = \mathbf{R}_\mathcal{E} \mathbf{A}$.

C.1.3. Upper bound on the inner products

The proof detailed in Appendix A requires an upper bound on the inner products of columns of the matrix $\widetilde{\mathbf{A}}$. For $i \neq j$, we obtain

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

$$\leqslant \mu_a + \left| \mathbf{a}_i^H \mathbf{B}_\mathcal{E} \left(\mathbf{B}_\mathcal{E}^H \mathbf{B}_\mathcal{E} \right)^{-1} \mathbf{B}_\mathcal{E}^H \mathbf{a}_j \right| \tag{C.11}$$

$$\leqslant \mu_a + \frac{|\mathbf{a}_i^H \mathbf{B}_\mathcal{E} \mathbf{B}_\mathcal{E}^H \mathbf{a}_j|}{[1 - \mu_b(n_e - 1)]^+} \tag{C.12}$$

$$\leqslant \mu_a + \frac{\|\mathbf{B}_\mathcal{E}^H \mathbf{a}_i\|_2 \|\mathbf{B}_\mathcal{E}^H \mathbf{a}_j\|_2}{[1 - \mu_b(n_e - 1)]^+}, \tag{C.13}$$

where (C.11) follows from the definition of the coherence parameter μ_a, (C.12) is a consequence of Geršgorin's disc theorem, and (C.13) from the Cauchy–Schwarz inequality. Since $\|\mathbf{B}_\mathcal{E}^H \mathbf{a}_i\|_2^2 = \sum_{k \in \mathcal{E}} |\mathbf{b}_k^H \mathbf{a}_i|^2 \leqslant n_e \mu_m^2$ for all $i = 1, \ldots, N_a$, the inner products with $i \neq j$ satisfy

$$\left| \tilde{\mathbf{a}}_i^H \tilde{\mathbf{a}}_j \right| \leqslant \mu_a + \frac{n_e \mu_m^2}{[1 - \mu_b(n_e - 1)]^+} \triangleq a. \tag{C.14}$$

C.1.4. Lower bound on the column norm

The final 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 (C.14) results in

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

C.2. Recovery guarantee

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

C.2.1. Bounding the error on the signal support

We start by bounding the error $\|\mathbf{h}_0\|_2$ as follows:

$$\begin{aligned}
\left|\mathbf{h}^H\widetilde{\mathbf{A}}^H\widetilde{\mathbf{A}}\mathbf{h}_0\right| &\geqslant \left|\mathbf{h}_0^H\widetilde{\mathbf{A}}^H\widetilde{\mathbf{A}}\mathbf{h}_0\right| - \left|(\mathbf{h} - \mathbf{h}_0)^H\widetilde{\mathbf{A}}^H\widetilde{\mathbf{A}}\mathbf{h}_0\right| \\
&\geqslant (1 - \hat\delta)\|\mathbf{h}_0\|_2^2 - an_x\|\mathbf{h}_0\|_2^2 - a\sqrt{n_x}\|\mathbf{h}_0\|_2 e_0 \\
&= c\|\mathbf{h}_0\|_2^2 - a\sqrt{n_x}\|\mathbf{h}_0\|_2 e_0
\end{aligned} \tag{C.15}$$

with

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

Note that the parameter c is crucial, since it determines the recovery condition for BP-RES (9). In particular, $c > 0$ is equivalent to (9)

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

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

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

C.2.2. Bounding the recovery error

We next compute an upper bound on $\|\mathbf{h}\|_2$. To this end, we start with a lower bound on $\|\widetilde{\mathbf{A}}\mathbf{h}\|_2^2$ as

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

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

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

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

Appendix D. Proof of Theorem 4

In order to prove the recovery guarantee in Theorem 4, 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.

D.1. Coherence-based RIC for $\mathbf{D} = [\mathbf{A}\ \mathbf{B}]$

In this section, we obtain a bound to that in Appendix A.1 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 $n_w = n_x + n_e$.

D.1.1. 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

$$
\begin{aligned}
\|\mathbf{Dh}_0\|_2^2 &= \begin{bmatrix} \mathbf{h}_x^H & \mathbf{h}_e^H \end{bmatrix} \begin{bmatrix} \mathbf{A}^H\mathbf{A} & \mathbf{A}^H\mathbf{B} \\ \mathbf{B}^H\mathbf{A} & \mathbf{B}^H\mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{h}_x \\ \mathbf{h}_e \end{bmatrix} \\
&= \mathbf{h}_0^H \begin{bmatrix} \mathbf{I}_{N_a} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N_b} \end{bmatrix} \mathbf{h}_0 + \mathbf{h}_0^H \begin{bmatrix} \mathbf{A}^H\mathbf{A} - \mathbf{I}_{N_a} & \mathbf{A}^H\mathbf{B} \\ \mathbf{B}^H\mathbf{A} & \mathbf{B}^H\mathbf{B} - \mathbf{I}_{N_b} \end{bmatrix} \mathbf{h}_0 \\
&\geqslant \|\mathbf{h}_0\|_2^2 - \left\| \begin{bmatrix} \mathbf{A}_{\mathcal{X}}^H\mathbf{A}_{\mathcal{X}} - \mathbf{I}_{|\mathcal{X}|} & \mathbf{A}_{\mathcal{X}}^H\mathbf{B}_{\mathcal{E}} \\ \mathbf{B}_{\mathcal{E}}^H\mathbf{A}_{\mathcal{X}} & \mathbf{B}_{\mathcal{E}}^H\mathbf{B}_{\mathcal{E}} - \mathbf{I}_{|\mathcal{E}|} \end{bmatrix} \right\|_2 \|\mathbf{h}_0\|_2^2,
\end{aligned}
\tag{D.1}
$$

where (D.1) 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 in (D.1) as follows:

$$
\begin{aligned}
&\left\| \begin{bmatrix} \mathbf{A}_{\mathcal{X}}^H\mathbf{A}_{\mathcal{X}} - \mathbf{I}_{|\mathcal{X}|} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{\mathcal{E}}^H\mathbf{B}_{\mathcal{E}} - \mathbf{I}_{|\mathcal{E}|} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{A}_{\mathcal{X}}^H\mathbf{B}_{\mathcal{E}} \\ \mathbf{B}_{\mathcal{E}}^H\mathbf{A}_{\mathcal{X}} & \mathbf{0} \end{bmatrix} \right\|_2 \\
&\leqslant \max\{\|\mathbf{A}_{\mathcal{X}}^H\mathbf{A}_{\mathcal{X}} - \mathbf{I}_{|\mathcal{X}|}\|_2, \|\mathbf{B}_{\mathcal{E}}^H\mathbf{B}_{\mathcal{E}} - \mathbf{I}_{|\mathcal{E}|}\|_2\} + \|\mathbf{A}_{\mathcal{X}}^H\mathbf{B}_{\mathcal{E}}\|_2,
\end{aligned}
\tag{D.2}
$$

where (D.2) is a result of 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 to the $\max\{\cdot\}$-term in (D.2) and

$$
\|\mathbf{A}_{\mathcal{X}}^H\mathbf{B}_{\mathcal{E}}\|_2 \leqslant \|\mathbf{A}_{\mathcal{X}}^H\mathbf{B}_{\mathcal{E}}\|_F \leqslant \sqrt{\sum_{k\in\mathcal{X}}\sum_{\ell\in\mathcal{E}}|\mathbf{a}_k^H\mathbf{b}_\ell|^2} \leqslant \sqrt{n_x n_e \mu_m^2}
$$

leads to

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

Hence, we arrive at the following lower bound

$$
\|\mathbf{Dh}_0\|_2^2 \geqslant \|\mathbf{h}_0\|_2^2 \left(1 - \max\{\mu_a(n_x - 1), \mu_b(n_e - 1)\} - \sqrt{n_x n_e \mu_m^2}\right).
\tag{D.3}
$$

By performing similar steps used to arrive at (D.3) we obtain the upper bound

$$
\|\mathbf{Dh}_0\|_2^2 \leqslant \|\mathbf{h}_0\|_2^2 \left(1 + \max\{\mu_a(n_x - 1), \mu_b(n_e - 1)\} + \sqrt{n_x n_e \mu_m^2}\right).
\tag{D.4}
$$

D.1.2. Bounds depending on $n_w = n_x + n_e$

Both bounds in (D.3) and (D.4) are explicit in n_x and n_e. Since the individual sparsity levels n_x and n_e are unknown prior to recovery, a coherence-based RIC bound, which depends solely on the total number $n_w = n_x + n_e$ of nonzero entries of $\mathbf{h}_0$ rather than on n_x and n_e, is required. 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

$$
\hat{g}(w) = \max_{0 \leqslant n_x \leqslant n_w} g(n_x, n_w - n_x).
\tag{D.5}
$$

Since $\hat{g}(n_w)$ only depends on $n_w = n_x + n_e$ and $g(n_x, n_e) \leqslant \hat{g}(n_w)$, we can replace $g(n_x, n_e)$ by $\hat{g}(n_w)$ in both bounds (D.3) and (D.4).

We start by computing the maximum in (D.5). Assume $\mu_a(n_x - 1) \geqslant \mu_b(n_e - 1)$ and consider the function

$$
g_a(n_x, n_w - n_x) = \mu_a(n_x - 1) + \sqrt{n_x(n_w - n_x)\mu_m^2}.
\tag{D.6}
$$

It can easily be shown that $g_a(n_x, n_w - n_x)$ is strictly concave in n_x for all $0 \leqslant n_x \leqslant n_w$ and $0 \leqslant n_w < \infty$ 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 (D.6) corresponds to

$$
\hat{g}_a(n_w) = \frac{1}{2}\left(\mu_a(n_w - 2) + n_w\sqrt{\mu_a^2 + \mu_m^2}\right).
\tag{D.7}
$$

For the case where $\mu_a(n_x - 1) < \mu_b(n_e - 1)$, we carry out similar steps used to arrive at (D.6) and exploit the symmetry of (D.5) to arrive at

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

Hence, by assuming that $\mu_b \leqslant \mu_a$, we obtain upper and lower bounds on (D.3) and (D.4) in terms of $n_w = n_x + n_e$ with the aid of (D.7) as follows:

$$\big(1 - \hat{g}_a(n_w)\big)\|\mathbf{h}_0\|_2^2 \leqslant \|\mathbf{D}\mathbf{h}_0\|_2^2 \leqslant \big(1 + \hat{g}_a(n_w)\big)\|\mathbf{h}_0\|_2^2. \tag{D.8}$$

It is important to realize that for some values of μ_a, μ_m, and n_w, the bounds in (D.8) are inferior to those obtained when ignoring the structure of the concatenated dictionary $\mathbf{D}$, i.e.,

$$\big(1 - \mu_d(n_w - 1)\big)\|\mathbf{h}_0\|_2^2 \leqslant \|\mathbf{D}\mathbf{h}_0\|_2^2 \leqslant \big(1 + \mu_d(n_w - 1)\big)\|\mathbf{h}_0\|_2^2 \tag{D.9}$$

with $\mu_d = \max\{\mu_a, \mu_b, \mu_m\}$. However, for $n_w \geqslant 2$, $\mu_m = \mu_d$, and

$$\mu_a < \mu_m + \frac{\mu_m n_w}{2}\left(\sqrt{\frac{n_w - 2}{n_w - 1}} - 1\right),$$

the RIP considering the structure of $\mathbf{D}$ in (D.8) turns out to be more tight than (D.9). For other values of n_w and/or μ_a, (D.8) turns out to be less tight than (D.9). In order to tighten the RIP in both cases, we consider

$$(1 - \hat{\delta}_{n_w})\|\mathbf{h}_0\|_2^2 \leqslant \|\mathbf{D}\mathbf{h}_0\|_2^2 \leqslant (1 + \hat{\delta}_{n_w})\|\mathbf{h}_0\|_2^2,$$

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

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

D.2. 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

$$\frac{1}{2}\Big(\mu_a(n_w - 2) + n_w\sqrt{\mu_a^2 + \mu_m^2}\Big) \leqslant \mu_d(n_w - 1).$$

The other case, i.e., where the standard RIP (D.9) is tighter than (D.8), readily follows from the proof in Appendix A, by replacing $\mathbf{A}$ by $\mathbf{D}$, μ_a by μ_d, and n_x by n_w.

D.2.1. Bounding the error on the signal support

We start by bounding the error $\|\mathbf{h}_0\|_2$. Since $\mu_m \leqslant \mu_d$, we arrive at

$$\begin{aligned}
\big|\mathbf{h}^H \mathbf{D}^H \mathbf{D}\mathbf{h}_0\big| &\geqslant \big|\mathbf{h}_0^H \mathbf{D}^H \mathbf{D}\mathbf{h}_0\big| - \big|(\mathbf{h} - \mathbf{h}_0)^H \mathbf{D}^H \mathbf{D}\mathbf{h}_0\big| \\
&\geqslant (1 - \hat{\delta}_{n_w})\|\mathbf{h}_0\|_2^2 - \mu_d w\|\mathbf{h}_0\|_2^2 - \mu_d\sqrt{n_w}\|\mathbf{h}_0\|_2 e_0 \\
&= d\|\mathbf{h}_0\|_2^2 - \mu_d\sqrt{n_w}\|\mathbf{h}_0\|_2 e_0
\end{aligned}$$

with

$$d \triangleq 1 - \delta_{n_w} - \mu_d n_w = 1 - \frac{n_w}{2}\Big(\mu_a + 2\mu_d + \sqrt{\mu_a^2 + \mu_m^2}\Big) + \mu_a.$$

It is important to note that d is crucial for the recovery guarantee as it determines the condition for which BP-SEP in (13) enables stable separation. Specifically, if $d > 0$ or, equivalently, if

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

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

$$\|\mathbf{h}_0\|_2 \leqslant \frac{(\varepsilon + \eta)\sqrt{1 + \hat{\delta}_{n_w}} + \mu_d\sqrt{n_w} e_0}{d},$$

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

D.2.2. Bounding the recovery error

Analogously to the derivation in A.3, we now compute an upper bound on $\|\mathbf{h}\|_2$, i.e.,

$$\|\mathbf{Dh}\|_2^2 \geqslant (1+\mu_d)\|\mathbf{h}\|_2^2 - \mu_d\|\mathbf{h}\|_1^2. \tag{D.10}$$

Finally, bounding $\|\mathbf{h}\|_2$ similarly to A.3 results in

$$\|\mathbf{h}\|_2 \leqslant (\varepsilon + \eta)\frac{d + 2\sqrt{\mu_d n_w}\sqrt{1 + \hat{\delta}_{n_w}}}{\sqrt{1 + \mu_d d}} + e_0 \frac{\sqrt{\mu_d}(d + 2\mu_d n_w)}{\sqrt{1 + \mu_d d}}$$
$$= C_7(\eta + \varepsilon) + C_8\|\mathbf{w} - \mathbf{w}_\mathcal{W}\|_1,$$

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

References

[1] S.V. Vaseghi, R. Frayling-Cork, Restoration of old gramophone recordings, J. Audio Eng. Soc. 40 (10) (1992) 791–801.
[2] S.J. Godsill, P.J.W. Rayner, Digital Audio Restoration – A Statistical Model-Based Approach, Springer-Verlag, London, 1998.
[3] M. Elad, Y. Hel-Or, Fast super-resolution reconstruction algorithm for pure translational motion and common space-invariant blur, IEEE Trans. Image Process. 10 (8) (2001) 1187–1193.
[4] M. Elad, J.-L. Starck, P. Querre, D.L. Donoho, Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA), Appl. Comput. Harmon. Anal. 19 (2005) 340–358.
[5] W. Göbel, F. Helmchen, In vivo calcium imaging of neural network function, Physiology 22 (6) (2007) 358–365.
[6] J.-F. Cai, S. Osher, Z. Shen, Split Bregman methods and frame based image restoration, Multiscale Model. Simul. 8 (2) (2009) 337–369.
[7] J.N. Laska, P. Boufounos, M.A. Davenport, R.G. Baraniuk, Democracy in action: Quantization, saturation, and compressive sensing, Appl. Comput. Harmon. Anal. 31 (3) (2011) 429–443.
[8] M. Bertalmio, G. Sapiro, V. Caselles, C. Ballester, Image inpainting, in: Proc. of 27th Ann. Conf. Comp. Graph. Int. Tech., 2000, pp. 417–424.
[9] A. Wong, J. Orchard, A nonlocal-means approach to exemplar-based inpainting, in: Proc. of IEEE ICIP, San Diego, CA, USA, 2008, pp. 2600–2603.
[10] J.-F. Cai, R.H. Chan, L. Shen, Z. Shen, Simultaneously inpainting in image and transformed domains, Numer. Math. 112 (4) (2009) 509–533.
[11] J.N. Laska, M.A. Davenport, 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, 2009, pp. 1556–1560.
[12] C. Studer, P. Kuppinger, G. Pope, H. Bölcskei, Recovery of sparsely corrupted signals, IEEE Trans. Inform. Theory 58 (5) (2012) 3115–3130.
[13] R.E. Carrillo, K.E. Barner, T.C. Aysal, Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise, IEEE J. Select. Top. Signal Process. 4 (2) (2010) 392–408.
[14] C. Novak, C. Studer, A. Burg, 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, 2007.
[15] S.G. Mallat, G. Yu, Super-resolution with sparse mixing estimators, IEEE Trans. Image Process. 19 (11) (2010) 2889–2900.
[16] A. Adler, V. Emiya, M.G. Jafari, M. Elad, R. Gribonval, M.D. Plumbley, A constrained matching pursuit approach to audio declipping, in: Proc. of IEEE Int. Conf. Acoustics, Speech, and Signal Processing, ICASSP, Prague, Czech Republic, 2011.
[17] A. Adler, V. Emiya, M.G. Jafari, M. Elad, R. Gribonval, M.D. Plumbley, Audio inpainting, IEEE Trans. Audio Speech Lang. Process. 20 (3) (2012) 922–932.
[18] G. Kutyniok, Data separation by sparse representations, arXiv:1102.4527v1.
[19] P. Kuppinger, G. Durisi, H. Bölcskei, Uncertainty relations and sparse signal recovery for pairs of general signal sets, IEEE Trans. Inform. Theory 58 (1) (2012) 263–277.
[20] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM J. Sci. Comput. 20 (1) (1998) 33–61.
[21] D.L. Donoho, X. Huo, Uncertainty principles and ideal atomic decomposition, IEEE Trans. Inform. Theory 47 (7) (2001) 2845–2862.
[22] D.L. Donoho, M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization, Proc. Natl. Acad. Sci. USA 100 (5) (2003) 2197–2202.
[23] R. Gribonval, M. Nielsen, Sparse representations in unions of bases, IEEE Trans. Inform. Theory 49 (12) (2003) 3320–3325.
[24] M. Elad, A.M. Bruckstein, A generalized uncertainty principle and sparse representation in pairs of bases, IEEE Trans. Inform. Theory 48 (9) (2002) 2558–2567.
[25] J.A. Tropp, Greed is good: Algorithmic results for sparse approximation, IEEE Trans. Inform. Theory 50 (10) (2004) 2231–2242.
[26] Y.C. Pati, R. Rezaifar, 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, 1993, pp. 40–44.
[27] G. Davis, S.G. Mallat, Z. Zhang, Adaptive time–frequency decompositions, Optim. Eng. 33 (7) (1994) 2183–2191.
[28] D.L. Donoho, M. Elad, V.N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, IEEE Trans. Inform. Theory 52 (1) (2006) 6–18.
[29] J.J. Fuchs, Recovery of exact sparse representations in the presence of bounded noise, IEEE Trans. Inform. Theory 51 (10) (2005) 3601–3608.
[30] J.A. Tropp, Just relax: Convex programming methods for identifying sparse signals in noise, IEEE Trans. Inform. Theory 52 (3) (2006) 1030–1051.
[31] Z. Ben-Haim, Y.C. Eldar, M. Elad, Coherence-based performance guarantees for estimating a sparse vector under random noise, IEEE Trans. Signal Process. 58 (10) (2010) 5030–5043.
[32] T.T. Cai, L. Wang, G. Xu, Stable recovery of sparse signals and an oracle inequality, IEEE Trans. Inform. Theory 56 (7) (2010) 3516–3522.
[33] R.G. Baraniuk, V. Cevher, M.F. Duarte, C. Hedge, Model-based compressive sensing, IEEE Trans. Inform. Theory 56 (4) (2010) 1982–2001.
[34] D.L. Donoho, P.B. Stark, Uncertainty principles and signal recovery, SIAM J. Appl. Math. 49 (3) (1989) 906–931.
[35] R. Gribonval, M. Nielsen, Beyond sparsity: Recovering structured representations by ℓ^1-minimization and greedy algorithms, Adv. Comput. Math. 28 (1) (2008) 23–41.
[36] E.J. Candès, T. Tao, Decoding by linear programming, IEEE Trans. Inform. Theory 51 (12) (2005) 4203–4215.
[37] J. Wright, Y. Ma, Dense error correction via ℓ_1-minimization, IEEE Trans. Inform. Theory 56 (7) (2010) 3540–3560.
[38] N.H. Nguyen, T.D. Tran, Exact recoverability from dense corrupted observations via ℓ_1 minimization, arXiv:1102.1227v1.
[39] X. Li, Compressed sensing and matrix completion with constant proportion of corruptions, arXiv:1104.1041v1.
[40] D.L. Donoho, Compressed sensing, IEEE Trans. Inform. Theory 52 (4) (2006) 1289–1306.
[41] E.J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inform. Theory 52 (2) (2006) 489–509.

[42] J.A. Tropp, On the conditioning of random subdictionaries, Appl. Comput. Harmon. Anal. 25 (2008) 1–24.
[43] G. Pope, A. Bracher, C. Studer, Probabilistic recovery guarantees for sparsely corrupted signals, IEEE Trans. Inform. Theory 59 (5) (2013) 3104–3116, arXiv:1203.6001.
[44] D.L. Donoho, G. Kutyniok, Analysis of ℓ_1 minimization in the geometric separation problem, in: 42nd Ann. Conf. on Inf. Sciences and Systems, CISS, 2008.
[45] C. Aubel, C. Studer, G. Pope, H. Bölcskei, Sparse signal separation in redundant dictionaries, in: Proc. IEEE Intl. Symp. on Inf. Theory, ISIT, Boston, MA, USA, 2012, pp. 2047–2051.
[46] D.L. Donoho, G. Kutyniok, Microlocal analysis of the geometric separation problem, arXiv:1004.3006v1.
[47] M. Hestenes, E. Stiefel, Methods of conjugate gradients for solving linear systems, J. Res. Natl. Bur. Stand. 49 (6) (1952) 409–436.
[48] Graeme Pope, Structured sparse signal recovery in general Hilbert spaces, Doctoral Dissertation, ETH Zurich, Switzerland, 2013, http://dx.doi.org/10.3929/ethz-a-009932306.
[49] A. Feuer, A. Nemirovski, On sparse representations in pairs of bases, IEEE Trans. Inform. Theory 49 (6) (2003) 1579–1581.
[50] M.E. Pfetsch, A.M. Tillmann, The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing, arXiv:1205.2081v2.
[51] E.J. Candès, J. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Comm. Pure Appl. Math. 59 (2006) 1207–1223.
[52] E.J. Candès, The restricted isometry property and its implications for compressed sensing, C. R. Acad. Sci. Paris, Ser. I 346 (2008) 589–592.
[53] T.T. Cai, G. Xu, J. Zhang, On recovery of sparse signals via ℓ_1 minimization, IEEE Trans. Inform. Theory 55 (2009) 3388–3397.
[54] T.T. Cai, L. Wang, G. Xu, Shifting inequality and recovery of sparse signals, IEEE Trans. Inform. Theory 58 (3) (2010) 1300–1308.
[55] T.T. Cai, L. Wang, G. Xu, New bounds for restricted isometry constants, IEEE Trans. Inform. Theory 56 (9) (2010) 4388–4394.
[56] T. Cai, A. Zhang, Sharp RIP bound for sparse signal and low-rank matrix recovery, Appl. Comput. Harmon. Anal. 35 (2013) 74–93.
[57] E.J. Candès, M.B. Wakin, An introduction to compressive sampling, IEEE Signal Process. Mag. 25 (2) (2008) 21–30.
[58] R.G. Baraniuk, M. Davenport, R.A. DeVore, M.B. Wakin, A simple proof of the restricted isometry property for random matrices, Constr. Approx. 28 (2008) 253–263.
[59] R.G. Baraniuk, Goodbye, textbooks, hello, open-source learning, [online accessed on July 1, 2011], http://www.ted.com/talks, Feb. 2006.
[60] M. Elad, Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing, 1st ed., Springer, 2010.
[61] M. Aharon, M. Elad, A.M. Bruckstein, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Process. 54 (11) (2006) 4311–4322.
[62] C. Févotte, S.J. Godsill, A Bayesian approach for blind separation of sparse sources, IEEE Trans. Audio Speech Lang. Process. 14 (6) (2006) 2174–2188.
[63] S.J. Godsill, A. Cemgil, C. Févotte, P.J. Wolfe, Bayesian computational methods for sparse audio and music processing, in: Proc. EUSIPCO, Poznan, Poland, 2007, pp. 345–349.
[64] S.G. Mallat, A Wavelet Tour of Signal Processing, Academic Press, San Diego, CA, 1998.
[65] P. Maechler, C. Studer, D. Bellasi, A. Maleki, A. Burg, N. Felber, H. Kaeslin, R.G. Baraniuk, VLSI design of approximate message passing for signal restoration and compressive sensing, IEEE J. Emerging Select. Top. Circuits Systems 2 (3) (2012) 579–590.
[66] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge Press, New York, NY, 1985.
[67] G.H. Golub, C.F. van Loan, Matrix Computations, 3rd ed., The Johns Hopkins Univ. Press, 1969.