Sparse Signal Separation in Redundant Dictionaries

Céline Aubel*, Christoph Studer†, Graeme Pope*, and Helmut Bölcskei*

*Dept. of IT & EE, ETH Zurich, 8092 Zurich, Switzerland
†Dept. of ECE, Rice University, Houston, TX, USA
Email: {aubelc, gpope, boelcskei}@nari.ee.ethz.ch, studer@rice.edu

***Abstract*—We formulate a unified framework for the separation of signals that are sparse in "morphologically" different redundant dictionaries. This formulation incorporates the so-called "analysis" and "synthesis" approaches as special cases and contains novel hybrid setups. We find corresponding coherence-based recovery guarantees for an ℓ_1-norm based separation algorithm. Our results recover those reported in Studer and Baraniuk, ACHA, *submitted*, for the synthesis setting, provide new recovery guarantees for the analysis setting, and form a basis for comparing performance in the analysis and synthesis settings. As an aside our findings complement the D-RIP recovery results reported in Candès *et al.*, ACHA, 2011, for the "analysis" signal recovery problem**

$$\underset{\widetilde{\mathbf{x}}}{\text{minimize}}\ \|\mathbf{\Psi}\widetilde{\mathbf{x}}\|_1 \quad \text{subject to}\ \|\mathbf{y}-\mathbf{A}\widetilde{\mathbf{x}}\|_2 \le \varepsilon$$

by delivering corresponding coherence-based recovery results.

I. Introduction

We consider the problem of splitting the signal $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ into its constituents $\mathbf{x}_1 \in \mathbb{C}^d$ and $\mathbf{x}_2 \in \mathbb{C}^d$—assumed to be sparse in "morphologically" different (redundant) dictionaries [1]—based on m linear, nonadaptive, and noisy measurements $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$. Here, $\mathbf{A} \in \mathbb{C}^{m\times d}$, $m \le d$, is the measurement matrix, assumed to be known, and $\mathbf{e} \in \mathbb{C}^m$ is a noise vector, assumed to be unknown and to satisfy $\|\mathbf{e}\|_2 \le \varepsilon$, with ε known.

Redundant dictionaries [2], [3] often lead to sparser representations than nonredundant ones, such as, e.g., orthonormal bases, and have therefore become pervasive in the sparse signal recovery literature [3]. In the context of signal separation, redundant dictionaries lead to an interesting dichotomy [1], [4], [5]:

- In the so-called "synthesis" setting, it is assumed that, for $\ell = 1, 2$, $\mathbf{x}_\ell = \mathbf{D}_\ell \mathbf{s}_\ell$, where $\mathbf{D}_\ell \in \mathbb{C}^{d\times n}$ $(d < n)$ is a redundant dictionary (of full rank) and the coefficient vector $\mathbf{s}_\ell \in \mathbb{C}^n$ is sparse (or approximately sparse in the sense of [6]). Given the vector $\mathbf{y} \in \mathbb{C}^m$, the problem of finding the constituents $\mathbf{x}_1$ and $\mathbf{x}_2$ is formalized as [7]:

$$\text{(PS)}\begin{cases} \underset{\tilde{\mathbf{s}}_1,\tilde{\mathbf{s}}_2}{\text{minimize}} & \|\tilde{\mathbf{s}}_1\|_1 + \|\tilde{\mathbf{s}}_2\|_1 \\ \text{subject to} & \|\mathbf{y} - \mathbf{A}(\mathbf{D}_1\tilde{\mathbf{s}}_1 + \mathbf{D}_2\tilde{\mathbf{s}}_2)\|_2 \le \varepsilon. \end{cases}$$

- In the so-called "analysis" setting, it is assumed that, for $\ell = 1, 2$, there exists a matrix $\mathbf{\Psi}_\ell \in \mathbb{C}^{n\times d}$ such that $\mathbf{\Psi}_\ell \mathbf{x}_\ell$ is sparse (or approximately sparse). The problem of recovering $\mathbf{x}_1$ and $\mathbf{x}_2$ from $\mathbf{y}$ is formalized as [5]:

$$\text{(PA)}\begin{cases} \underset{\widetilde{\mathbf{x}}_1,\widetilde{\mathbf{x}}_2}{\text{minimize}} & \|\mathbf{\Psi}_1\widetilde{\mathbf{x}}_1\|_1 + \|\mathbf{\Psi}_2\widetilde{\mathbf{x}}_2\|_1 \\ \text{subject to} & \|\mathbf{y} - \mathbf{A}(\widetilde{\mathbf{x}}_1 + \widetilde{\mathbf{x}}_2)\|_2 \le \varepsilon. \end{cases}$$

Throughout the paper, we exclusively consider redundant dictionaries as for $\mathbf{D}_\ell$, $\ell = 1, 2$, square, the synthesis setting can be recovered from the analysis setting by taking $\mathbf{\Psi}_\ell = \mathbf{D}_\ell^{-1}$.

The problems (PS) and (PA) arise in numerous applications including denoising [8], super-resolution [8], inpainting [9]–[11], deblurring [11], and recovery of sparsely corrupted signals [12]. Coherence-based recovery guarantees for (PS) were reported in [7]. The problem (PA) was mentioned in [5]. In the noiseless case, recovery guarantees for (PA), expressed in terms of a concentration inequality, are given in [13] for $\mathbf{A} = \mathbf{I}_d$ and $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_2$ both Parseval frames [2].

Contributions: We consider the general problem

$$\text{(P)}\begin{cases} \underset{\widetilde{\mathbf{x}}_1,\widetilde{\mathbf{x}}_2}{\text{minimize}} & \|\mathbf{\Psi}_1\widetilde{\mathbf{x}}_1\|_1 + \|\mathbf{\Psi}_2\widetilde{\mathbf{x}}_2\|_1 \\ \text{subject to} & \|\mathbf{y} - \mathbf{A}_1\widetilde{\mathbf{x}}_1 - \mathbf{A}_2\widetilde{\mathbf{x}}_2\|_2 \le \varepsilon, \end{cases}$$

which encompasses (PS) and (PA). To recover (PS) from (P), one sets $\mathbf{A}_\ell = \mathbf{A}\mathbf{D}_\ell$ and $\mathbf{\Psi}_\ell = [\mathbf{I}_d\ \mathbf{0}_{d,n-d}]^T$, for $\ell = 1, 2$. (PA) is obtained by choosing $\mathbf{A}_\ell = \mathbf{A}$, for $\ell = 1, 2$. Our main contribution is a coherence-based recovery guarantee for the general problem (P). This result recovers [7, Th. 4], which deals with (PS), provides new recovery guarantees for (PA), and constitutes a basis for comparing performance in the analysis and synthesis settings. As an aside, it also complements the D-RIP recovery guarantee in [5, Th. 1.2] for the problem

$$\text{(P}^*\text{)}\ \underset{\widetilde{\mathbf{x}}}{\text{minimize}}\ \|\mathbf{\Psi}\widetilde{\mathbf{x}}\|_1 \quad \text{subject to}\ \|\mathbf{y}-\mathbf{A}\widetilde{\mathbf{x}}\|_2 \le \varepsilon$$

by delivering a corresponding coherence-based recovery guarantee. Moreover, the general formulation (P) encompasses novel hybrid problems of the form

$$\begin{array}{ll} \underset{\tilde{\mathbf{s}}_1,\widetilde{\mathbf{x}}_2}{\text{minimize}} & \|\tilde{\mathbf{s}}_1\|_1 + \|\mathbf{\Psi}_2\widetilde{\mathbf{x}}_2\|_1 \\ \text{subject to} & \|\mathbf{y} - \mathbf{A}(\mathbf{D}_1\tilde{\mathbf{s}}_1 - \widetilde{\mathbf{x}}_2)\|_2 \le \varepsilon. \end{array}$$

Notation: Lowercase boldface letters stand for column vectors and uppercase boldface letters denote matrices. The transpose, conjugate transpose, and Moore-Penrose inverse of the matrix $\mathbf{M}$ are designated as $\mathbf{M}^T$, $\mathbf{M}^H$, and $\mathbf{M}^\dagger$, respectively. The jth column of $\mathbf{M}$ is written $[\mathbf{M}]_j$, and

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

(a) Original cartoon image (b) Corrupted image (c) Restored cartoon image

Fig. 1: Image separation in the presence of Gaussian noise (SNR $= 20\,\text{dB}$).

the entry in the ith row and jth column of $\mathbf{M}$ is $[\mathbf{M}]_{i,j}$. We let $\sigma_{\min}(\mathbf{M})$ denote the smallest singular value of $\mathbf{M}$, use $\mathbf{I}_n$ for the $n \times n$ identity matrix, and let $\mathbf{0}_{k\times m}$ be the $k \times m$ all zeros matrix. For matrices $\mathbf{M}$ and $\mathbf{N}$, we let $\omega_{\min}(\mathbf{M}) \triangleq \min_j \|[\mathbf{M}]_j\|_2$, $\omega_{\max}(\mathbf{M}) \triangleq \max_j \|[\mathbf{M}]_j\|_2$, $\omega_{\min}(\mathbf{M},\mathbf{N}) \triangleq \min\{\omega_{\min}(\mathbf{M}), \omega_{\min}(\mathbf{N})\}$, and $\omega_{\max}(\mathbf{M},\mathbf{N}) \triangleq \max\{\omega_{\max}(\mathbf{M}), \omega_{\max}(\mathbf{N})\}$. The kth entry of the vector $\mathbf{x}$ is written $[\mathbf{x}]_k$, and $\|\mathbf{x}\|_1 \triangleq \sum_k |[\mathbf{x}]_k|$ stands for its ℓ_1-norm. We take $\text{supp}_k(\mathbf{x})$ to be the set of indices corresponding to the k largest (in magnitude) coefficients of $\mathbf{x}$. Sets are designated by uppercase calligraphic letters; the cardinality of the set $\mathcal{S}$ is $|\mathcal{S}|$ and the complement of $\mathcal{S}$ (in some given set) is denoted by $\mathcal{S}^c$. For a set $\mathcal{S}$ of integers and $n \in \mathbb{Z}$, we let $n + \mathcal{S} \triangleq \{n + p : p \in \mathcal{S}\}$. The $n \times n$ diagonal projection matrix $\mathbf{P}_\mathcal{S}$ for the set $\mathcal{S} \subset \{1, \ldots, n\}$ is defined as follows:

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

and we set $\mathbf{M}_\mathcal{S} \triangleq \mathbf{P}_\mathcal{S}\mathbf{M}$. We define $\sigma_k(\mathbf{x})$ to be the ℓ_1-norm approximation error of the best k-sparse approximation of $\mathbf{x}$, i.e., $\sigma_k(\mathbf{x}) \triangleq \|\mathbf{x}-\mathbf{x}_\mathcal{S}\|_1$ where $\mathcal{S} = \text{supp}_k(\mathbf{x})$ and $\mathbf{x}_\mathcal{S} \triangleq \mathbf{P}_\mathcal{S}\mathbf{x}$.

II. Recovery Guarantees

Coherence definitions in the sparse signal recovery literature [3] usually apply to dictionaries with normalized columns. Here, we will need coherence notions valid for general (unnormalized) dictionaries $\mathbf{M}$ and $\mathbf{N}$, assumed, for simplicity of exposition, to consist of nonzero columns only.

Definition 1 (Coherence): The coherence of the dictionary $\mathbf{M}$ is defined as

$$\hat{\mu}(\mathbf{M}) = \max_{i,j,i\neq j} \frac{|[\mathbf{M}^H\mathbf{M}]_{i,j}|}{\omega^2_{\min}(\mathbf{M})}. \tag{1}$$

Definition 2 (Mutual coherence): The mutual coherence of the dictionaries $\mathbf{M}$ and $\mathbf{N}$ is defined as

$$\hat{\mu}_m(\mathbf{M},\mathbf{N}) = \max_{i,j} \frac{|[\mathbf{M}^H\mathbf{N}]_{i,j}|}{\omega^2_{\min}(\mathbf{M},\mathbf{N})}. \tag{2}$$

The main contribution of this paper is the following recovery guarantee for (P).

Theorem 1: Let $\mathbf{y} = \mathbf{A}_1\mathbf{x}_1 + \mathbf{A}_2\mathbf{x}_2 + \mathbf{e}$ with $\|\mathbf{e}\|_2 \leq \varepsilon$ and let $\mathbf{\Psi}_1 \in \mathbb{C}^{n_1\times p_1}$ and $\mathbf{\Psi}_2 \in \mathbb{C}^{n_2\times p_2}$ be full-rank matrices. Let $\mathbf{x} = [\mathbf{x}_1^T\ \mathbf{x}_2^T]^T$, $\hat{\mu}_1 = \hat{\mu}(\mathbf{A}_1\mathbf{\Psi}_1^\dagger)$, $\hat{\mu}_2 = \hat{\mu}(\mathbf{A}_2\mathbf{\Psi}_2^\dagger)$, $\hat{\mu}_m = \hat{\mu}_m(\mathbf{A}_1\mathbf{\Psi}_1^\dagger, \mathbf{A}_2\mathbf{\Psi}_2^\dagger)$, and $\hat{\mu}_{\max} = \max\{\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_m\}$. Without loss of generality, we assume that $\hat{\mu}_1 \leq \hat{\mu}_2$. Let k_1 and k_2 be nonnegative integers such that

$$k_1 + k_2 < \max\left\{\frac{2(1+\hat{\mu}_2)}{\hat{\mu}_2 + 2\hat{\mu}_{\max} + \sqrt{\hat{\mu}_2^2 + \hat{\mu}_m^2}}, \frac{1+\hat{\mu}_{\max}}{2\hat{\mu}_{\max}}\right\}. \tag{3}$$

Then, the solution $(\mathbf{x}_1^*, \mathbf{x}_2^*)$ to the convex program (P) satisfies

$$\|\mathbf{x}^* - \mathbf{x}\|_2 \leq C_0\,\varepsilon + C_1(\sigma_{k_1}(\mathbf{\Psi}_1\mathbf{x}_1) + \sigma_{k_2}(\mathbf{\Psi}_2\mathbf{x}_2)), \tag{4}$$

where $C_0, C_1 \geq 0$ are constants that do not depend on $\mathbf{x}_1$ and $\mathbf{x}_2$ and where $\mathbf{x}^* = [\mathbf{x}_1^{*T}\ \mathbf{x}_2^{*T}]^T$.

Note that the quantities $\hat{\mu}_1$, $\hat{\mu}_2$, and $\hat{\mu}_m$ characterize the interplay between the measurement matrix $\mathbf{A}$ and the sparsifying transforms $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_2$.

As a corollary to our main result, we get the following statement for the problem (P*) considered in [5].

Corollary 2: Let $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ with $\|\mathbf{e}\|_2 \leq \varepsilon$ and let $\mathbf{\Psi} \in \mathbb{C}^{n\times p}$ be a full-rank matrix. Let k be a nonnegative integer such that

$$k < \frac{1}{2}\left(1 + \frac{1}{\hat{\mu}(\mathbf{A}\mathbf{\Psi}^\dagger)}\right). \tag{5}$$

Then, the solution $\mathbf{x}^*$ to the convex program (P*) satisfies

$$\|\mathbf{x}^* - \mathbf{x}\|_2 \leq C_0\,\varepsilon + C_1\sigma_k(\mathbf{\Psi}\mathbf{x}), \tag{6}$$

where $C_0, C_1 \geq 0$ are constants[1] that do not depend on $\mathbf{x}$.

The proofs of Theorem 1 and Corollary 2 can be found in the Appendix.

We conclude by noting that D-RIP recovery guarantees for (P*) were provided in [5]. As is common in RIP-based

[1]Note that the constants C_0 and C_1 may take on different values at each occurrence.

recovery guarantees the restricted isometry constants are, in general, hard to compute. Moreover, the results in [5] hinge on the assumption that $\mathbf{\Psi}$ forms a Parseval frame, i.e., $\mathbf{\Psi}^H\mathbf{\Psi} = \mathbf{I}_d$; a corresponding extension to general $\mathbf{\Psi}$ was provided in [14]. We finally note that it does not seem possible to infer the coherence-based threshold (5) from the D-RIP recovery guarantees in [5], [14].

III. Numerical Results

We analyze an image-separation problem where we remove a fingerprint from a cartoon image. We corrupt the 512×512 greyscale cartoon image depicted in Fig. 1(a) by adding a fingerprint[2] and i.i.d. zero-mean Gaussian noise.

Cartoon images are constant apart from (a small number of) discontinuities and are thus sparse under the finite difference operator $\mathbf{\Delta}$ defined in [15]. Fingerprints are sparse under the application of a wave atom transform, $\mathbf{W}$, such as the redundancy 2 transform available in the WaveAtom toolbox[3] [16]. It is therefore sensible to perform separation by solving the problem (PA) with $\mathbf{\Psi}_1 = \mathbf{\Delta}$, $\mathbf{\Psi}_2 = \mathbf{W}$, and $\mathbf{A} = \mathbf{I}_d$. For our simulation, we use a regularized version of $\mathbf{\Delta}$ and we employ the TFOCS solver[4] from [17].

Fig. 1(c) shows the corresponding recovered image. We can see that the restoration procedure gives visually satisfactory results.

Appendix A
Proofs

For simplicity of exposition, we first present the proof of Corollary 2 and then describe the proof of Theorem 1.

A. Proof of Corollary 2

We define the vector $\mathbf{h} = \mathbf{x}^* - \mathbf{x}$, where $\mathbf{x}^*$ is the solution to (P*) and $\mathbf{x}$ is the vector to be recovered. We furthermore set $\mathcal{S} = \text{supp}_k(\mathbf{\Psi}\mathbf{x})$.

1) Prerequisites: Our proof relies partly on two important results developed earlier in [5], [6] and summarized, for completeness, next.

Lemma 3 (Cone constraint [5], [6]): The vector $\mathbf{\Psi}\mathbf{h}$ obeys

$$\|\mathbf{\Psi}_{\mathcal{S}^c}\mathbf{h}\|_1 \leq \|\mathbf{\Psi}_{\mathcal{S}}\mathbf{h}\|_1 + 2\|\mathbf{\Psi}_{\mathcal{S}^c}\mathbf{x}\|_1, \tag{7}$$

where $\mathcal{S} = \text{supp}_k(\mathbf{\Psi}\mathbf{x})$.

Proof: Since $\mathbf{x}^*$ is the minimizer of (P*), the inequality $\|\mathbf{\Psi}\mathbf{x}\|_1 \geq \|\mathbf{\Psi}\mathbf{x}^*\|_1$ holds. Using $\mathbf{\Psi} = \mathbf{\Psi}_{\mathcal{S}} + \mathbf{\Psi}_{\mathcal{S}^c}$ and $\mathbf{x}^* = \mathbf{x} + \mathbf{h}$, we obtain

$$\begin{aligned}
\|\mathbf{\Psi}_{\mathcal{S}}\mathbf{x}\|_1 + \|\mathbf{\Psi}_{\mathcal{S}^c}\mathbf{x}\|_1 &= \|\mathbf{\Psi}\mathbf{x}\|_1 \\
&\geq \|\mathbf{\Psi}\mathbf{x}^*\|_1 = \|\mathbf{\Psi}_{\mathcal{S}}\mathbf{x} + \mathbf{\Psi}_{\mathcal{S}}\mathbf{h}\|_1 + \|\mathbf{\Psi}_{\mathcal{S}^c}\mathbf{x} + \mathbf{\Psi}_{\mathcal{S}^c}\mathbf{h}\|_1 \\
&\geq \|\mathbf{\Psi}_{\mathcal{S}}\mathbf{x}\|_1 - \|\mathbf{\Psi}_{\mathcal{S}}\mathbf{h}\|_1 + \|\mathbf{\Psi}_{\mathcal{S}^c}\mathbf{h}\|_1 - \|\mathbf{\Psi}_{\mathcal{S}^c}\mathbf{x}\|_1.
\end{aligned}$$

We retrieve (7) by simple rearrangement of terms. ■

Lemma 4 (Tube constraint [5], [6]): The vector $\mathbf{A}\mathbf{h}$ satisfies $\|\mathbf{A}\mathbf{h}\|_2 \leq 2\varepsilon$.

Proof: Since both $\mathbf{x}^*$ and $\mathbf{x}$ are feasible (we recall that $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ with $\|\mathbf{e}\|_2 \leq \varepsilon$), we have the following

$$\begin{aligned}
\|\mathbf{A}\mathbf{h}\|_2 &= \|\mathbf{A}(\mathbf{x}^* - \mathbf{x})\|_2 \\
&\leq \|\mathbf{A}\mathbf{x}^* - \mathbf{y}\|_2 + \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq 2\varepsilon,
\end{aligned}$$

thus establishing the lemma. ■

2) Bounding the recovery error: We want to bound $\|\mathbf{h}\|_2$ from above. Since $\sigma_{\min}(\mathbf{\Psi}) > 0$ by assumption ($\mathbf{\Psi}$ is assumed to be full-rank), it follows from the Rayleigh-Ritz theorem [18, Th. 4.2.2] that

$$\|\mathbf{h}\|_2 \leq \frac{1}{\sigma_{\min}(\mathbf{\Psi})}\|\mathbf{\Psi}\mathbf{h}\|_2. \tag{9}$$

We now set $\mathcal{Q} = \text{supp}_k(\mathbf{\Psi}\mathbf{h})$. Clearly, we have for $i \in \mathcal{Q}^c$,

$$|[\mathbf{\Psi}\mathbf{h}]_i| \leq \frac{\|\mathbf{\Psi}_{\mathcal{Q}}\mathbf{h}\|_1}{k}.$$

Using the same argument as in [19, Th. 3.1], we obtain

$$\begin{aligned}
\|\mathbf{\Psi}_{\mathcal{Q}^c}\mathbf{h}\|_2^2 = \sum_{i\in\mathcal{Q}^c} |[\mathbf{\Psi}\mathbf{h}]_i|^2 &\leq \sum_{i\in\mathcal{Q}^c} |[\mathbf{\Psi}\mathbf{h}]_i| \frac{\|\mathbf{\Psi}_{\mathcal{Q}}\mathbf{h}\|_1}{k} \\
&= \|\mathbf{\Psi}_{\mathcal{Q}^c}\mathbf{h}\|_1 \frac{\|\mathbf{\Psi}_{\mathcal{Q}}\mathbf{h}\|_1}{k}. \qquad (10)
\end{aligned}$$

Since $\mathcal{Q}$ is the set of indices of the k largest (in magnitude) coefficients of $\mathbf{\Psi}\mathbf{h}$ and since $\mathcal{Q}$ and $\mathcal{S}$ both contain k elements, we have $\|\mathbf{\Psi}_{\mathcal{S}}\mathbf{h}\|_1 \leq \|\mathbf{\Psi}_{\mathcal{Q}}\mathbf{h}\|_1$ and $\|\mathbf{\Psi}_{\mathcal{Q}^c}\mathbf{h}\|_1 \leq \|\mathbf{\Psi}_{\mathcal{S}^c}\mathbf{h}\|_1$, which, combined with the cone constraint in Lemma 3, yields

$$\|\mathbf{\Psi}_{\mathcal{Q}^c}\mathbf{h}\|_1 \leq \|\mathbf{\Psi}_{\mathcal{Q}}\mathbf{h}\|_1 + 2\|\mathbf{\Psi}_{\mathcal{S}^c}\mathbf{x}\|_1. \tag{11}$$

The inequality in (10) then becomes

$$\begin{aligned}
\|\mathbf{\Psi}_{\mathcal{Q}^c}\mathbf{h}\|_2^2 &\leq \frac{\|\mathbf{\Psi}_{\mathcal{Q}}\mathbf{h}\|_1^2}{k} + 2\|\mathbf{\Psi}_{\mathcal{S}^c}\mathbf{x}\|_1 \frac{\|\mathbf{\Psi}_{\mathcal{Q}}\mathbf{h}\|_1}{k} \\
&\leq \|\mathbf{\Psi}_{\mathcal{Q}}\mathbf{h}\|_2^2 + 2\|\mathbf{\Psi}_{\mathcal{S}^c}\mathbf{x}\|_1 \frac{\|\mathbf{\Psi}_{\mathcal{Q}}\mathbf{h}\|_2}{\sqrt{k}} \qquad (12a) \\
&\leq 2\|\mathbf{\Psi}_{\mathcal{Q}}\mathbf{h}\|_2^2 + \frac{\|\mathbf{\Psi}_{\mathcal{S}^c}\mathbf{x}\|_1^2}{k}, \qquad (12b)
\end{aligned}$$

where (12a) follows from $\|\mathbf{u}\|_1 \leq \sqrt{k}\|\mathbf{u}\|_2$ for k-sparse[5] $\mathbf{u}$ and (12b) is a consequence of $2xy \leq x^2 + y^2$, for $x, y \in \mathbb{R}$.

It now follows that

$$\begin{aligned}
\|\mathbf{\Psi}\mathbf{h}\|_2 &= \sqrt{\|\mathbf{\Psi}_{\mathcal{Q}}\mathbf{h}\|_2^2 + \|\mathbf{\Psi}_{\mathcal{Q}^c}\mathbf{h}\|_2^2} \\
&\leq \sqrt{3\|\mathbf{\Psi}_{\mathcal{Q}}\mathbf{h}\|_2^2 + \frac{\|\mathbf{\Psi}_{\mathcal{S}^c}\mathbf{x}\|_1^2}{k}} \qquad (13a) \\
&\leq \sqrt{3}\|\mathbf{\Psi}_{\mathcal{Q}}\mathbf{h}\|_2 + \frac{\|\mathbf{\Psi}_{\mathcal{S}^c}\mathbf{x}\|_1}{\sqrt{k}}, \qquad (13b)
\end{aligned}$$

where (13a) is a consequence of (12b) and (13b) results from $\sqrt{x^2 + y^2} \leq x + y$, for $x, y \geq 0$.

Combining (9) and (13b) leads to

$$\|\mathbf{h}\|_2 \leq \frac{1}{\sigma_{\min}(\mathbf{\Psi})}\left(\sqrt{3}\|\mathbf{\Psi}_{\mathcal{Q}}\mathbf{h}\|_2 + \frac{\|\mathbf{\Psi}_{\mathcal{S}^c}\mathbf{x}\|_1}{\sqrt{k}}\right). \tag{14}$$

[2] The fingerprint image is taken from http://commons.wikimedia.org/

[3] We used the WaveAtom toolbox from http://www.waveatom.org/

[4] We used TFOCS from http://tfocs.stanford.edu/

[5] A vector is k-sparse if it has at most k nonzero entries.

3) Bounding the term $\|\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_2$ in (14): In the last step of the proof, we bound the term $\|\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_2$ in (14). To this end, we first bound $\|\mathbf{A}\mathbf{\Psi}^\dagger\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_2^2$, with $\mathbf{\Psi}^\dagger = (\mathbf{\Psi}^H\mathbf{\Psi})^{-1}\mathbf{\Psi}^H$, using Geršgorin's disc theorem [18, Th. 6.2.2]:

$$\theta_{\text{min}}\|\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_2^2 \leq \|\mathbf{A}\mathbf{\Psi}^\dagger\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_2^2 \leq \theta_{\text{max}}\|\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_2^2 \quad (15)$$

where $\theta_{\text{min}} \triangleq \omega_{\text{min}}^2 - \mu(k-1)$ and $\theta_{\text{max}} \triangleq \omega_{\text{max}}^2 + \mu(k-1)$ with

$$\mu = \max_{i,j,i\neq j} |[(\mathbf{A}\mathbf{\Psi}^\dagger)^H\mathbf{A}\mathbf{\Psi}^\dagger]_{i,j}| \quad (16)$$

and $\omega_{\text{min}} \triangleq \omega_{\text{min}}(\mathbf{A}\mathbf{\Psi}^\dagger)$ and $\omega_{\text{max}} \triangleq \omega_{\text{max}}(\mathbf{A}\mathbf{\Psi}^\dagger)$.

Using Lemma 4 and (15) and following the same steps as in [20, Th. 2.1] and [7, Th. 1], we arrive at the following chain of inequalities:

$$\begin{aligned}
&\theta_{\text{min}}\|\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_2^2 \leq \|\mathbf{A}\mathbf{\Psi}^\dagger\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_2^2 = (\mathbf{A}\mathbf{\Psi}^\dagger\mathbf{\Psi}_\mathcal{Q}\mathbf{h})^H\mathbf{A}\mathbf{\Psi}^\dagger\mathbf{\Psi}_\mathcal{Q}\mathbf{h} \\
&= (\mathbf{A}\mathbf{h})^H\mathbf{A}\mathbf{\Psi}^\dagger\mathbf{\Psi}_\mathcal{Q}\mathbf{h} - (\mathbf{A}\mathbf{\Psi}^\dagger\mathbf{\Psi}_{\mathcal{Q}^c}\mathbf{h})^H\mathbf{A}\mathbf{\Psi}^\dagger\mathbf{\Psi}_\mathcal{Q}\mathbf{h} \quad (17a) \\
&\leq |(\mathbf{A}\mathbf{h})^H\mathbf{A}\mathbf{\Psi}^\dagger\mathbf{\Psi}_\mathcal{Q}\mathbf{h}| + |(\mathbf{\Psi}_{\mathcal{Q}^c}\mathbf{h})^H(\mathbf{A}\mathbf{\Psi}^\dagger)^H\mathbf{A}\mathbf{\Psi}^\dagger(\mathbf{\Psi}_\mathcal{Q}\mathbf{h})| \\
&\leq \|\mathbf{A}\mathbf{h}\|_2\|\mathbf{A}\mathbf{\Psi}^\dagger\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_2 \\
&\qquad + \sum_{i\in\mathcal{Q}^c, j\in\mathcal{Q}} |[(\mathbf{A}\mathbf{\Psi}^\dagger)^H\mathbf{A}\mathbf{\Psi}^\dagger]_{i,j}||[\mathbf{\Psi}\mathbf{h}]_i||[\mathbf{\Psi}\mathbf{h}]_j| \quad (17b) \\
&\leq 2\varepsilon\sqrt{\theta_{\text{max}}}\|\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_2 + \mu\|\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_1\|\mathbf{\Psi}_{\mathcal{Q}^c}\mathbf{h}\|_1 \quad (17c) \\
&\leq 2\varepsilon\sqrt{\theta_{\text{max}}}\|\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_2 + \mu\|\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_1\left(\|\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_1 + 2\|\mathbf{\Psi}_{\mathcal{S}^c}\mathbf{x}\|_1\right) \quad (17d) \\
&\leq 2\varepsilon\sqrt{\theta_{\text{max}}}\|\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_2 + \mu k\|\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_2^2 \\
&\qquad + 2\mu\sqrt{k}\|\mathbf{\Psi}_{\mathcal{S}^c}\mathbf{x}\|_1\|\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_2, \quad (17e)
\end{aligned}$$

where (17a) follows from $\mathbf{\Psi}_\mathcal{Q}\mathbf{h} = \mathbf{\Psi}\mathbf{h} - \mathbf{\Psi}_{\mathcal{Q}^c}\mathbf{h}$ and $\mathbf{\Psi}^\dagger\mathbf{\Psi} = \mathbf{I}_d$, (17b) is a consequence of the Cauchy-Schwarz inequality, (17c) is obtained from (15), Lemma 4, and the definition of μ in (16), (17d) results from (11), and (17e) comes from $\|\mathbf{u}\|_1 \leq \sqrt{k}\|\mathbf{u}\|_2$, for k-sparse $\mathbf{u}$.

If $\mathbf{h} \neq \mathbf{0}$, then $\|\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_2 \neq 0$, since $\mathbf{\Psi}$ is assumed to be full-rank and $\mathcal{Q}$ is the set of indices of the k largest (in magnitude) coefficients of $\mathbf{\Psi}\mathbf{h}$, and therefore, the inequality between $\theta_{\text{min}}\|\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_2^2$ and (17e) simplifies to

$$(\omega_{\text{min}}^2 - \mu(2k-1))\|\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_2 \leq 2\varepsilon\sqrt{\theta_{\text{max}}} + 2\mu\sqrt{k}\|\mathbf{\Psi}_{\mathcal{S}^c}\mathbf{x}\|_1.$$

This finally yields

$$\|\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_2 \leq \frac{2\varepsilon\sqrt{\theta_{\text{max}}} + 2\mu\sqrt{k}\|\mathbf{\Psi}_{\mathcal{S}^c}\mathbf{x}\|_1}{\omega_{\text{min}}^2 - \mu(2k-1)} \quad (18)$$

provided that

$$\omega_{\text{min}}^2 - \mu(2k-1) > 0.$$

4) Recovery guarantee: Using Definition 1, we get $\hat{\mu} = \hat{\mu}(\mathbf{A}\mathbf{\Psi}^\dagger) = \mu/\omega_{\text{min}}^2$. Combining (14) and (18), we therefore conclude that for

$$k < \frac{1}{2}\left(1 + \frac{1}{\hat{\mu}}\right) \quad (19)$$

we have

$$\|\mathbf{x}^* - \mathbf{x}\|_2 = \|\mathbf{h}\|_2 \leq C_0\,\varepsilon + C_1\|\mathbf{\Psi}_{\mathcal{S}^c}\mathbf{x}\|_1$$

with

$$C_0 = \frac{2\sqrt{3}}{\sigma_{\text{min}}(\mathbf{\Psi})\omega_{\text{min}}}\frac{\sqrt{\frac{\omega_{\text{max}}^2}{\omega_{\text{min}}^2}(1+\hat{\mu}(k-1))}}{1-\hat{\mu}(2k-1)}$$

$$C_1 = \frac{1}{\sigma_{\text{min}}(\mathbf{\Psi})}\left(\frac{2\hat{\mu}\sqrt{3k}}{1-\hat{\mu}(2k-1)} + \frac{1}{\sqrt{k}}\right).$$

B. Proof of Theorem 1

We start by transforming (P) into the equivalent problem

$$(\text{P}^*)\ \underset{\widetilde{\mathbf{x}}}{\text{minimize}}\ \ \|\mathbf{\Psi}\widetilde{\mathbf{x}}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{A}\widetilde{\mathbf{x}}\|_2 \leq \varepsilon$$

by amalgamating $\mathbf{\Psi}_1, \mathbf{\Psi}_2$ and $\mathbf{A}_1, \mathbf{A}_2$ into the matrices $\mathbf{\Psi}$ and $\mathbf{A}$ as follows:

$$\mathbf{A} = \begin{bmatrix}\mathbf{A}_1 & \mathbf{A}_2\end{bmatrix} \in \mathbb{C}^{m\times p} \quad (20)$$

$$\mathbf{\Psi} = \begin{bmatrix}\mathbf{\Psi}_1 & \mathbf{0}_{n\times d} \\ \mathbf{0}_{n\times d} & \mathbf{\Psi}_2\end{bmatrix} \in \mathbb{C}^{2n\times 2d}, \quad (21)$$

where $p = 2d$ in the analysis setting, $p = 2n$ in the synthesis setting, and $p = d + n$ in hybrid settings. The corresponding measurement vector is $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$, where we set $\mathbf{x} = [\mathbf{x}_1^T\ \mathbf{x}_2^T]^T$.

A recovery condition for (P) could now be obtained by simply inserting $\mathbf{A}$ and $\mathbf{\Psi}$ in (20), (21) above into (5). In certain cases, we can, however, get a better (i.e., less restrictive) threshold following ideas similar to those reported in [7] and detailed next.

We define the vectors $\mathbf{h}_1 = \mathbf{x}_1^* - \mathbf{x}_1$, $\mathbf{h}_2 = \mathbf{x}_2^* - \mathbf{x}_2$, the sets $\mathcal{Q}_1 \triangleq \text{supp}_{k_1}(\mathbf{\Psi}_1\mathbf{h}_1)$, $\mathcal{Q}_2 \triangleq n + \text{supp}_{k_2}(\mathbf{\Psi}_2\mathbf{h}_2)$, and $\mathbf{h} = [\mathbf{h}_1^T\ \mathbf{h}_2^T]^T$, $\mathcal{Q} = \mathcal{Q}_1 \cup \mathcal{Q}_2$, and set $k = k_1 + k_2$.

We furthermore let, for $\ell = 1, 2$,

$$\mu_\ell = \max_{i,j,i\neq j} |[(\mathbf{A}_\ell\mathbf{\Psi}_\ell^\dagger)^H\mathbf{A}_\ell\mathbf{\Psi}_\ell^\dagger]_{i,j}|$$

$$\mu_m = \max_{i,j} |[(\mathbf{A}_1\mathbf{\Psi}_1^\dagger)^H\mathbf{A}_2\mathbf{\Psi}_2^\dagger]_{i,j}|.$$

With the definitions of $\mathcal{Q}_1$ and $\mathcal{Q}_2$, we have from (15)

$$\begin{aligned}\|\mathbf{A}\mathbf{\Psi}^\dagger\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_2^2 &= \|\mathbf{A}\mathbf{\Psi}^\dagger\mathbf{\Psi}_{\mathcal{Q}_1}\mathbf{h}\|_2^2 + \|\mathbf{A}\mathbf{\Psi}^\dagger\mathbf{\Psi}_{\mathcal{Q}_2}\mathbf{h}\|_2^2 \\ &\quad + 2(\mathbf{A}\mathbf{\Psi}^\dagger\mathbf{\Psi}_{\mathcal{Q}_1}\mathbf{h})^H\mathbf{A}\mathbf{\Psi}^\dagger\mathbf{\Psi}_{\mathcal{Q}_2}\mathbf{h}. \quad (22)\end{aligned}$$

The application of Geršgorin's disc theorem [18] gives

$$\theta_{\text{min},1}\|\mathbf{\Psi}_{\mathcal{Q}_1}\mathbf{h}\|_2^2 \leq \|\mathbf{A}\mathbf{\Psi}^\dagger\mathbf{\Psi}_{\mathcal{Q}_1}\mathbf{h}\|_2^2 \leq \theta_{\text{max},1}\|\mathbf{\Psi}_{\mathcal{Q}_1}\mathbf{h}\|_2^2 \quad (23)$$

$$\theta_{\text{min},2}\|\mathbf{\Psi}_{\mathcal{Q}_2}\mathbf{h}\|_2^2 \leq \|\mathbf{A}\mathbf{\Psi}^\dagger\mathbf{\Psi}_{\mathcal{Q}_2}\mathbf{h}\|_2^2 \leq \theta_{\text{max},2}\|\mathbf{\Psi}_{\mathcal{Q}_2}\mathbf{h}\|_2^2 \quad (24)$$

with $\theta_{\text{min},\ell} \triangleq \omega_{\text{min}}^2(\mathbf{A}_\ell\mathbf{\Psi}_\ell^\dagger) - \mu_\ell(k_\ell - 1)$ and $\theta_{\text{max},\ell} \triangleq \omega_{\text{max}}^2(\mathbf{A}_\ell\mathbf{\Psi}_\ell^\dagger) + \mu_\ell(k_\ell - 1)$, for $\ell = 1, 2$.

In addition, the last term in (22) can be bounded as

$$\begin{aligned}
&|(\mathbf{A}\mathbf{\Psi}^\dagger\mathbf{\Psi}_{\mathcal{Q}_1}\mathbf{h})^H\mathbf{A}\mathbf{\Psi}^\dagger\mathbf{\Psi}_{\mathcal{Q}_2}\mathbf{h}| \\
&\quad \leq \sum_{i\in\mathcal{Q}_1, j\in\mathcal{Q}_2} |[(\mathbf{A}\mathbf{\Psi}^\dagger)^H\mathbf{A}\mathbf{\Psi}^\dagger]_{i,j}||[\mathbf{\Psi}\mathbf{h}]_i||[\mathbf{\Psi}\mathbf{h}]_j| \\
&\quad \leq \mu_m\|\mathbf{\Psi}_{\mathcal{Q}_1}\mathbf{h}\|_1\|\mathbf{\Psi}_{\mathcal{Q}_2}\mathbf{h}\|_1 \quad (25a) \\
&\quad \leq \mu_m\sqrt{k_1k_2}\|\mathbf{\Psi}_{\mathcal{Q}_1}\mathbf{h}\|_2\|\mathbf{\Psi}_{\mathcal{Q}_2}\mathbf{h}\|_2 \quad (25b) \\
&\quad \leq \frac{\mu_m}{2}\sqrt{k_1k_2}\left(\|\mathbf{\Psi}_{\mathcal{Q}_1}\mathbf{h}\|_2^2 + \|\mathbf{\Psi}_{\mathcal{Q}_2}\mathbf{h}\|_2^2\right) \quad (25c) \\
&\quad \leq \frac{\mu_m}{2}\sqrt{k_1k_2}\|\mathbf{\Psi}_\mathcal{Q}\mathbf{h}\|_2^2, \quad (25d)
\end{aligned}$$

where (25a) follows from the definition of μ_m, (25b) results from $\|\mathbf{u}\|_1 \leq \sqrt{k}\|\mathbf{u}\|_2$, for k-sparse $\mathbf{u}$, and (25c) is a consequence of the arithmetic-mean geometric-mean inequality.

Combining (23), (24), and (25d) gives

$$\theta_{\text{min}}\|\mathbf{\Psi}_{\mathcal{Q}}\mathbf{h}\|_2^2 \leq \|\mathbf{A}\mathbf{\Psi}^{\dagger}\mathbf{\Psi}_{\mathcal{Q}}\mathbf{h}\|_2^2 \leq \theta_{\text{max}}\|\mathbf{\Psi}_{\mathcal{Q}}\mathbf{h}\|_2^2,$$

where $\theta_{\text{min}} \triangleq \omega_{\text{min}}^2 - f(k_1,k_2)$, $\theta_{\text{max}} \triangleq \omega_{\text{max}}^2 + f(k_1,k_2)$, $\omega_{\text{min}} \triangleq \omega_{\text{min}}(\mathbf{A}_1\mathbf{\Psi}_1^{\dagger}, \mathbf{A}_2\mathbf{\Psi}_2^{\dagger})$, $\omega_{\text{max}} \triangleq \omega_{\text{max}}(\mathbf{A}_1\mathbf{\Psi}_1^{\dagger}, \mathbf{A}_2\mathbf{\Psi}_2^{\dagger})$, and

$$f(k_1,k_2) \triangleq \max\{\mu_1(k_1-1), \mu_2(k_2-1)\} + \mu_m\sqrt{k_1k_2}.$$

Using the same steps as in (17a)-(17e), we get

$$g(k_1,k_2)\|\mathbf{\Psi}_{\mathcal{Q}}\mathbf{h}\|_2 \leq 2\varepsilon\sqrt{\theta_{\text{max}}} + 2\mu\sqrt{k}\|\mathbf{\Psi}_{\mathcal{S}^c}\mathbf{x}\|_1,$$

where $g(k_1,k_2) \triangleq \omega_{\text{min}}^2 - f(k_1,k_2) - \mu k$.

Next, we bound $g(k_1,k_2)$ from below by a function of $k = k_1+k_2$. This can be done, e.g., by looking for the minimum [7]

$$\hat{g}(k) \triangleq \min_{k_1:\ 0\leq k_1\leq k} g(k_1, k-k_1) \tag{26}$$

or equivalently

$$\hat{g}(k) \triangleq \min_{k_2:\ 0\leq k_2\leq k} g(k-k_2, k_2). \tag{27}$$

To find $\hat{g}(k)$ in (26) or in (27), we need to distinguish between two cases:

- Case 1: $\mu_1(k_1-1) \leq \mu_2(k_2-1)$

In this case, we get

$$g(k-k_2,k_2) = \omega_{\text{min}}^2 - \mu_2(k_2-1) - \mu_m\sqrt{k_2(k-k_2)} - \mu k.$$

A straightforward calculation reveals that the minimum of g is achieved at

$$k_2 = \frac{k}{2}\left(1 + \frac{\mu_2}{\sqrt{\mu_2^2+\mu_m^2}}\right),$$

resulting in

$$\hat{g}(k) = \omega_{\text{min}}^2 - \frac{1}{2}\left(\mu_2(k-2) + k\sqrt{\mu_2^2+\mu_m^2}\right) - \mu k.$$

If $\hat{g}(k) > 0$, then we have

$$\|\mathbf{x}^* - \mathbf{x}\|_2 = \|\mathbf{h}\|_2 \leq C_0\,\varepsilon + C_1\|\mathbf{\Psi}_{\mathcal{S}^c}\mathbf{x}\|_1 \tag{28}$$

where

$$C_0 = \frac{2\sqrt{3}}{\sigma_{\text{min}}(\mathbf{\Psi})\hat{g}(k)}$$

and

$$C_1 = \frac{1}{\sigma_{\text{min}}(\mathbf{\Psi})}\left(\frac{2\mu\sqrt{3k}}{\hat{g}(k)} + \frac{1}{\sqrt{k}}\right).$$

Setting $\hat{g}(k) > 0$ amounts to imposing

$$k < \frac{2\,(1+\hat{\mu}_2)}{\hat{\mu}_2 + 2\hat{\mu}_{\text{max}} + \sqrt{\hat{\mu}_2^2+\hat{\mu}_m^2}}, \tag{29}$$

where we used Definitions 1 and 2 to get a threshold depending on the coherence parameters only.

- Case 2: $\mu_2(k_2-1) \leq \mu_1(k_1-1)$

Similarly to Case 1, we get

$$\hat{g}(k) = \omega_{\text{min}}^2 - \frac{1}{2}\left(\mu_1(k-2) + k\sqrt{\mu_1^2+\mu_m^2}\right) - \mu k.$$

If $\hat{g}(k) > 0$, we must have

$$k < \frac{2\,(1+\hat{\mu}_1)}{\hat{\mu}_1 + 2\hat{\mu}_{\text{max}} + \sqrt{\hat{\mu}_1^2+\hat{\mu}_m^2}}. \tag{30}$$

Since $\hat{\mu}_1 \leq \hat{\mu}_2$, by assumption, the inequality in (30) is tighter than the one in (29). We complete the proof by combining the thresholds in (19) and (29) to get (3).

References

[1] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045–1057, Apr. 2010.

[2] O. Christensen, *An Introduction to Frames and Riesz Bases*, ser. Applied and Numerical Harmonic Analysis, J. J. Benedetto, Ed. Boston, MA, USA: Birkhäuser, 2002.

[3] M. Elad, *Sparse and Redundant Representations – From Theory to Applications in Signal and Image Processing*. New York, NY, USA: Springer, 2010.

[4] P. Milanfar and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, pp. 947–968, Jan. 2007.

[5] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmon. Anal.*, vol. 31, no. 1, pp. 59–73, Sep. 2011.

[6] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 2, pp. 1207–1223, Mar. 2005.

[7] C. Studer and R. Baraniuk, "Stable restoration and separation of approximately sparse signals," *Appl. Comput. Harmon. Anal.*, submitted. [Online]. Available: http://arxiv.org/pdf/1107.0420v1.pdf

[8] S. G. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*. Burlington, MA, USA: Academic Press, 2009.

[9] M. Elad, J.-L. Starck, P. Querre, and D. L. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *Appl. Comput. Harmon. Anal.*, vol. 19, no. 3, pp. 340–358, Jan. 2005.

[10] J. Fadili, J.-L. Starck, M. Elad, and D. L. Donoho, "MCALab: Reproducible research in signal and image decomposition and inpainting," *Computing in Science & Engineering*, vol. 12, no. 1, pp. 44–63, Feb. 2010.

[11] J.-F. Cai, S. Osher, and Z. Shen, "Split Bregman methods and frame based image restoration," *Multiscale Modeling & Simulation*, vol. 8, no. 2, pp. 337–369, Jan. 2010.

[12] C. Studer, P. Kuppinger, G. Pope, and H. Bölcskei, "Recovery of sparsely corrupted signals," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3115–3130, May 2012.

[13] G. Kutyniok, "Data separation by sparse representations," in *Compressed Sensing: Theory and Applications*, Y. C. Eldar and G. Kutyniok, Eds., New York, NY, USA: Cambridge University Press, 2012.

[14] Y. Liu, T. Mi, and S. Li, "Compressed sensing with general frames via optimal-dual-based ℓ_1-analysis," *IEEE Trans. Inf. Theory*, submitted. [Online]. Available: http://arxiv.org/pdf/1111.4345.pdf

[15] S. Nam, M. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," INRIA, Tech. Rep., Jun. 2011. [Online]. Available: http://arxiv.org/pdf/1106.4987v1.pdf

[16] L. Demanet and L. Ying, "Wave atoms and sparsity of oscillatory patterns," *Appl. Comput. Harmon. Anal.*, vol. 23, no. 3, pp. 368–387, Jan. 2007.

[17] S. Becker, E. J. Candès, and M. Grant, "Templates for convex cone problems with applications to sparse signal recovery," in *Mathematical Programming Computation*, W. J. Cook, Ed., 2012, vol. 3, no. 3, pp. 165–218.

[18] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY, USA: Cambridge University Press, 1991.

[19] T. T. Cai, L. Wang, and G. Xu, "New bounds for restricted isometry constants," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 1–7, Aug. 2010.

[20] ——, "Stable recovery of sparse signals and an oracle inequality," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3516–3522, Jul. 2010.