Further results on stable recovery of sparse overcomplete representations in the presence of noise

Paul Tseng

***Abstract*—Sparse overcomplete representations have attracted much interest recently for their applications to signal processing. In a recent work, Donoho, Elad, and Temlyakov [12] showed that, assuming sufficient sparsity of the ideal underlying signal and approximate orthogonality of the overcomplete dictionary, the sparsest representation can be found, at least approximately if not exactly, by either an orthogonal greedy algorithm or by ℓ_1-norm minimization subject to a noise tolerance constraint. In this paper, we sharpen the approximation bounds under more relaxed conditions. We also derive analogous results for a stepwise projection algorithm.**

***Index Terms*—Basis pursuit, greedy algorithm, ℓ_1-norm minimization, matching pursuit, overcomplete representation, mutual coherence, sparse representation.**

I. Introduction

A fundamental problem in signal processing is that of finding a "good" representation of a given (possibly noisy) signal $y \in \Re^n$. An approach that has been gaining popularity is to choose an overcomplete set of elementary signals $\phi_1, \ldots, \phi_m \in \Re^n$, normalized so that $\|\phi_j\|_2 = 1$ for $j = 1, \ldots, m$, and use optimization or other means to find a sparse representation of y from this set. This is guided by Occam's Razor principle that "simplest is best." For example, letting $\Phi := [\phi_1 \cdots \phi_m] \in \Re^{n\times m}$, we might seek a representation $\Phi\alpha = y$ for which $\alpha \in \Re^m$ has the fewest nonzeros. However, this problem is known to be intractable (NP-hard) and its solution is highly sensitive to noise in y. A less noise-sensitive problem formulation is

$$\min_{\alpha\in\Re^m} \|\alpha\|_0 \quad \text{subject to} \quad \|\Phi\alpha - y\|_2^2 \le \delta^2,$$

where $\delta \ge 0$ is a user-chosen tolerance. Here and throughout, $\|\cdot\|_p$ denotes the ℓ_p-norm ($1 \le p \le \infty$) and $\|\alpha\|_0 := |\text{supp}(\alpha)|$, where $\text{supp}(\alpha) := \{j \mid \alpha_j \ne 0\}$ and $|\mathcal{J}|$ denotes the cardinality of a finite set $\mathcal{J}$. Thus, we seek a sparsest (i.e., fewest nonzeros) representation with noise tolerance δ. However, this problem is still intractable in general.

To make the problem tractable, it was proposed in [12], [18], [31] to approximate the nonconvex discontinuous counting function $\|\cdot\|_0$ by $\|\cdot\|_1$, and thus solve the convex optimization problem

$$\min_{\alpha\in\Re^m} \|\alpha\|_1 \quad \text{subject to} \quad \|\Phi\alpha - y\|_2^2 \le \delta^2. \tag{1}$$

The author is with the Department of Mathematics, University of Washington, Seattle, WA, 98195, USA e-mail: (tseng@math.washington.edu). This research is supported by the National Science Foundation, Grant No. DMS-0511283.

Manuscript received February 26, 2007; revised June 12, 2008.

This problem can be efficiently solved by various numerical methods and is closely related to least square with ℓ_1-regularization used in basis pursuit denoising and LASSO [6], [26], [29]:

$$\min_{\alpha\in\Re^m} \|\Phi\alpha - y\|_2^2 + \gamma\|\alpha\|_1, \qquad (\gamma > 0) \tag{2}$$

where $1/\gamma$ may be interpreted as a Lagrange multiplier for (1) for some suitable $\delta \ge 0$ [16]. In some studies [2], [21], $\|\cdot\|_1$ is replaced by $\|\cdot\|_p$ for some $0 < p < 1$. However, the corresponding optimization problems (1) and (2) are no longer convex and have spurious local minima, e.g., $\alpha = 0$. Thus, a global minimum is not guaranteed to be found by a local descent method.

Another practical strategy for finding a sparse representation is to select the members of the representation one-at-a-time in a greedy manner, until the least square residual is below a user-chosen threshold $\delta^{\text{OGA}} \ge 0$. One such greedy algorithm is the orthogonal greedy algorithm (OGA) [7], [24], which is a modification of the matching pursuit algorithm of Mallat and Zhang [23]; see [1], [8], [12], [28] and Section IV for further discussions of OGA.

Let α^δ be any solution of (1), and let α^{OGA} be coefficients generated by the OGA. (Note that $\alpha^\delta = 0$ if and only if $\delta \ge \|y\|_2$.) How good approximations are α^δ and α^{OGA} of a sparsest solution? This question has been extensively studied in [11], [12], [13], [17], [18], [19], [22], [30], [31]. In the studies [11], [12], [13], [17], [22], [30], it is assumed that the original signal y^o lies in the span of $\phi_1, \ldots, \phi_m$, i.e., $y^o \in \text{Span}\Phi$, and α^δ, α^{OGA} are compared with a solution α^o of

$$\min_{\alpha\in\Re^m} \|\alpha\|_0 \quad \text{subject to} \quad \Phi\alpha = y^o. \tag{3}$$

Note that $\alpha^o = 0$ if and only if $y^o = 0$. In what follows, let

$$\begin{array}{rclcrcl} \mathcal{J}^o & := & \text{supp}(\alpha^o), & \qquad & N & := & |\mathcal{J}^o|, \\ \alpha^o_{\min} & := & \displaystyle\min_{j\in\mathcal{J}^o} |\alpha^o_j|, & & \epsilon & := & \|y - y^o\|_2. \end{array}$$

It is not difficult to see that $\alpha^\delta = \alpha^{\text{OGA}} = \alpha^o$ whenever $\phi_1, \ldots, \phi_m$ are pairwise orthogonal and $\epsilon = \delta = 0$. This suggests that α^δ and α^{OGA} are good approximations of α^o whenever $\phi_1, \ldots, \phi_m$ are approximately orthogonal and $\epsilon \approx \delta \approx 0$. Accordingly, central to the analysis in [11], [12], [13], [17], [18], [19], [22], [23], [30], [31] is the following measure of approximate orthogonality, called "mutual coherence" and introduced by Mallat and Zhang [23] in their initial study of matching pursuit,

$$\mu := \max_{1\le i\neq j\le m} \left|\phi_i^T \phi_j\right|. \tag{4}$$

As is noted in [12, page 7], there exist overcomplete sets with $m \approx n^2$ and $\mu \approx 1/\sqrt{n}$; see [27, pages 265-266] and references therein.

In the noiseless case of $\epsilon = 0$, Donoho and Elad [11, Theorem 7], Gribonval and Nielsen [22, Theorem 1], Fuchs [17], Tropp [30, Theorems A and B] independently showed that $\alpha^\delta = \alpha^o$ is the unique solution of (1) and (3) whenever $\delta = 0$ and $N < \frac{1}{2}(\mu^{-1}+1)$. If Φ is the concatenation of two square orthogonal matrices, then the latter condition can be relaxed to $N < (\sqrt{2}-\frac{1}{2})\mu^{-1}$ [11], [15]. Tropp [30, Theorems A and B] showed that $\alpha^{\text{OGA}} = \alpha^o$ whenever $\delta^{\text{OGA}} = 0$ and $N < \frac{1}{2}(\mu^{-1}+1)$. In the general noisy case, Donoho, Elad, and Temlyakov showed that

$$\|\alpha^\delta - \alpha^o\|_2^2 \le \frac{(\epsilon+\delta)^2}{1-\mu(4N-1)} \tag{5}$$

whenever $\delta \ge \epsilon$ and $N < \frac{1}{4}(\mu^{-1}+1)$ [12, Theorem 3.1]. They also showed that $\text{supp}(\alpha^\delta) \subseteq \mathcal{J}^\circ$ whenever $N < \frac{1}{2}\mu^{-1}$ and

$$\delta > \frac{1-\mu N + \sqrt{1-\mu N}\sqrt{N}}{1-2\mu N}\,\epsilon \tag{6}$$

[12, Theorem 4.1]. (A related result for (2) is derived in [18, Theorem 4], showing that $\text{supp}(\alpha^{\text{bp}}) \subseteq \mathcal{J}^\circ$ whenever $N \le \frac{1}{2}\mu^{-1}$ and $\gamma/2 > 1 + \frac{\mu N}{1-\mu(2N-1)}\epsilon$, where α^{bp} is any solution of (2).) It is further shown in [12, Theorem 5.1(a)] that $\text{supp}(\alpha^{\text{OGA}}) = \mathcal{J}^\circ$ whenever $\delta^{\text{OGA}} = \epsilon$, $N < \frac{1}{2}(\mu^{-1}+1)$, and

$$\alpha^o_{\min} \ge \frac{2\epsilon}{1-\mu(2N-1)}. \tag{7}$$

In this paper, we derive new stable recovery results for ℓ_1 minimization (1) and OGA in the noisy case. In particular, we improve on the main results in [12] by sharpening the aforementioned approximation bounds under more relaxed conditions. For example, we sharpen the bound (5) and extend it to hold whenever $N < (\frac{1}{2} - O(\mu))\mu^{-1}+1$; see Theorem 1 and (22). We further extend this bound to the case of $\delta < \epsilon$, which had not been studied previously. The bound involves an additional quantity L_k (see (12)), which is the 2,1-norm of the pseudo-inverse of at most $k = \|\alpha^\delta\|_0 + N$ linearly independent columns of Φ. If α^δ is also sparse so that $\mu k < 1$, then $L_k \le 1/\sqrt{k^{-1}-\mu}$; see (13). The sufficient conditions (6) and (7) for $\text{supp}(\alpha^\delta) \subseteq \mathcal{J}^\circ$ and $\text{supp}(\alpha^{\text{OGA}}) = \mathcal{J}^\circ$ are similarly relaxed; see Theorems 2, 3 and Corollary 1. In particular, the condition $\delta^{\text{OGA}} = \epsilon$ and (7) are relaxed to $\delta^{\text{OGA}} \ge \epsilon$ and either $N = 1$ or

$$N \ge 2 \text{ and } \alpha^o_{\min} \ge \max\left\{\frac{2\epsilon}{1-\mu(2N-1)}, \frac{\epsilon+\delta^{\text{OGA}}}{1-\mu(N-1)}\right\}. \tag{8}$$

In addition, we show that another greedy algorithm, called stepwise projection algorithm [1], has similar sparsest support identification properties as OGA; see Theorem 4 and Corollary 2.

The preceding analyses are worst-case in the sense that the results hold for all Φ subject to conditions on N, μ, ϵ, δ, etc. In particular, for $\mu \approx 1/\sqrt{n}$, the results require $N = O(\sqrt{n})$. There has been much recent work showing that, for Φ randomly generated from certain classes of distributions (e.g., Gaussian), stable recovery by ℓ_1 minimization (1) or OGA is likely even when N is nearly $O(n)$; see [3], [5], [10], [14], [32] and references therein for the noiseless case and [4], [9] for the noisy case. These results are based on approximate isometry properties of submatrices of Φ. The approximation bounds in [4], [9] for (1) require $\delta \ge \epsilon$, as well as $m = O(n)$ in [9].

II. Matrix Notations and Norms

For any $A \in \Re^{p\times q}$, define

$$\begin{aligned}\|A\|_1 &:= \max_{\|d\|_1=1}\|Ad\|_1,\\ \|A\|_{2,1} &:= \max_{\|d\|_2=1}\|Ad\|_1,\\ \|A\|_2 &:= \max_{\|d\|_2=1}\|Ad\|_2.\end{aligned}$$

Then, for any $d \in \Re^q$, we have $\|d\|_1 \le \sqrt{q}\|d\|_2 \le q\|d\|_\infty$ and $\|Ad\|_1 \le \|A\|_1\|d\|_1$, $\|Ad\|_1 \le \|A\|_{2,1}\|d\|_2$, $\|Ad\|_2 \le \|A\|_2\|d\|_2$. Also, $\|A\|_1 = \max_j \sum_i |A_{ij}|$ [20, Section 2.2], $\|A\|_{2,1} \le \sqrt{p}\|A\|_2$, and $\|A\|_{2,1} \le \sum_i \sqrt{\sum_j |A_{ij}|^2}$.

For any $\alpha \in \Re^m$ and nonempty set $\mathcal{J} \subseteq \{1,\ldots,m\}$, $\mathcal{J}_c := \{1,\ldots,m\}\setminus\mathcal{J}$, $\alpha_{\mathcal{J}}$ denotes the subvector of α comprising those α_j with $j \in \mathcal{J}$, and $\Phi_{\mathcal{J}}$ denotes the submatrix of Φ comprising those columns ϕ_j with $j \in \mathcal{J}$. The following lemma derives bounds on the singular values of $\Phi_{\mathcal{J}}$ and on the norms of its pseudo-inverse [20, page 139] and $(\Phi_{\mathcal{J}}^T\Phi_{\mathcal{J}})^{-1}$. It will be used to prove Theorems 1–4. Lemma 1(a) and its proof are similar to [12, Lemma 2.2] and its proof.

Lemma 1: For any nonempty $\mathcal{J} \subseteq \{1,\ldots,m\}$, the following results hold with $M := |\mathcal{J}|$.

(a) $\|\Phi_{\mathcal{J}}\|_2^2 \le 1 + \mu(M-1)$ and

$$\begin{aligned}\|\Phi_{\mathcal{J}}d\|_2^2 &\ge (1+\mu)\|d\|_2^2 - \mu\|d\|_1^2\\ &\ge (1-\mu(M-1))\|d\|_2^2 \qquad \forall d \in \Re^M.\end{aligned}$$

(b) If $\mu(M-1) < 1$, then $\Phi_{\mathcal{J}}^T\Phi_{\mathcal{J}}$ is invertible and $E := (\Phi_{\mathcal{J}}^T\Phi_{\mathcal{J}})^{-1}$ satisfies

$$|E_{ij}| \le \begin{cases} 1 + \frac{\mu^2(M-1)}{1-\mu(M-1)} & \text{if } i = j \in \mathcal{J}\\ \frac{\mu}{1-\mu(M-1)} & \text{if } i \ne j \in \mathcal{J}\end{cases} \tag{9}$$

$$\|E\|_2 \le \frac{1}{1-\mu(M-1)}, \tag{10}$$

$$\|E\Phi_{\mathcal{J}}^T\|_2 \le \frac{1}{\sqrt{1-\mu(M-1)}}. \tag{11}$$

Proof: By (4) and $\|\phi_j\|_2 = 1$ for all j, we have $\Phi_{\mathcal{J}}^T\Phi_{\mathcal{J}} = I - B$, where $B_{ii} = 0$ and $|B_{ij}| \le \mu$ for $i \ne j \in \mathcal{J}$.

(a). For any $d = (d_j)_{j\in\mathcal{J}} \in \Re^M$, we have

$$\|\Phi_{\mathcal{J}}d\|_2^2 = d^T\Phi_{\mathcal{J}}^T\Phi_{\mathcal{J}}d = d^T(I-B)d = \|d\|_2^2 - \sum_{i\ne j} d_iB_{ij}d_j.$$

Since $|B_{ij}| \leq \mu$, $|\sum_{i\neq j} d_i B_{ij} d_j| \leq \sum_{i\neq j} \mu |d_i||d_j| = \mu\|d\|_1^2 - \mu\|d\|_2^2$. Thus

$$(1+\mu)\|d\|_2^2 - \mu\|d\|_1^2 \leq \|\Phi_{\mathcal{J}} d\|_2^2 \leq (1-\mu)\|d\|_2^2 + \mu\|d\|_1^2.$$

Using $\|d\|_1^2 \leq M\|d\|_2^2$ and the definition of $\|\Phi_{\mathcal{J}}\|_2$ completes the proof.

(b). For $l = 2, 3, \ldots$,

$$\begin{aligned} |(B^l)_{ij}| &= \left|\sum_{k\neq j} (B^{l-1})_{ik} B_{kj}\right| \\ &\leq \mu \sum_{k\neq j} |(B^{l-1})_{ik}| \\ &\leq \mu(M-1)\max_{k\neq j} |(B^{l-1})_{ik}|, \quad \forall i,j \in \mathcal{J}, \end{aligned}$$

from which it follows by induction on l that

$$|(B^l)_{ij}| \leq \mu^l (M-1)^{l-1}, \quad \forall i,j \in \mathcal{J},\ l = 2,3,\ldots.$$

Since $\mu(M-1) < 1$, it follows that $I - B$ is invertible and $E = (I-B)^{-1} = I + B + B^2 + \cdots$, so that

$$\begin{aligned} |E_{ij}| &= |(I + B + B^2 + \cdots)_{ij}| \\ &= \begin{cases} |1 + (B^2)_{ii} + (B^3)_{ii} + \cdots| & \text{if } i = j \\ |B_{ij} + (B^2)_{ij} + \cdots| & \text{if } i \neq j \end{cases} \\ &\leq \begin{cases} 1 + \mu^2(M-1) + \mu^3(M-1)^2 + \cdots & \text{if } i = j \\ \mu + \mu^2(M-1) + \mu^3(M-1)^2 + \cdots & \text{if } i \neq j \end{cases} \\ &\leq \begin{cases} 1 + \frac{\mu^2(M-1)}{1-\mu(M-1)} & \text{if } i = j \\ \frac{\mu}{1-\mu(M-1)} & \text{if } i \neq j. \end{cases} \end{aligned}$$

This proves (9). Lastly, (a) shows that the singular values of $\Phi_{\mathcal{J}}$ have magnitude of at least $\sqrt{1-\mu(M-1)} > 0$. Using the singular value decomposition of $\Phi_{\mathcal{J}}$, it is straightforward to verify (10) and (11). ■

III. Sparse Representation from ℓ_1-Regularization

For $k = 1, 2, \ldots$, let

$$L_k := \max_{\operatorname{rank}\Phi_{\mathcal{K}} = |\mathcal{K}| \leq k} \|(\Phi_{\mathcal{K}}^T \Phi_{\mathcal{K}})^{-1} \Phi_{\mathcal{K}}^T\|_{2,1}, \tag{12}$$

i.e., the maximum is taken over all $\mathcal{K} \subseteq \{1, \ldots, m\}$ such that $|\mathcal{K}| \leq k$ and $\Phi_{\mathcal{K}}$ has linearly independent columns. When $\mu(k-1) < 1$, it follows from (11) that

$$L_k \leq \frac{\sqrt{k}}{\sqrt{1-\mu(k-1)}}. \tag{13}$$

When $|\phi_i^T \phi_j| \in \{0, 1/\sqrt{n}\}$ for $i \neq j$ [27, page 266], L_k may be estimated using Cramer's rule for any k; see [33, Lemma 5]. In general, L_k may be difficult to estimate. However, it plays a key role in the following extension of [12, Theorem 3.1] when $\delta < \epsilon$.

Theorem 1: If $N \geq 1$, $\mu(N-1) < 1$, and $\rho\mu N < 1$, then

$$\|\alpha^\delta - \alpha^o\|_1 \leq \left(\sqrt{N}\vartheta + \frac{1+\rho\mu N}{1-\rho\mu N}\left(\lambda_\delta + \sqrt{N}\vartheta\right)\right)(\epsilon + \delta), \tag{14}$$

$$\|\alpha^\delta - \alpha^o\|_2^2 \leq \left(1 + \mu\left(\sqrt{N}\vartheta + \frac{1+\rho\mu N}{1-\rho\mu N}\left(\lambda_\delta + \sqrt{N}\vartheta\right)\right)^2\right)\frac{(\epsilon+\delta)^2}{1+\mu}, \tag{15}$$

where

$$\rho := \frac{1+\mu^2(N-1)}{1-\mu(N-1)}, \qquad \vartheta := \frac{1}{\sqrt{1-\mu(N-1)}},$$

$$\lambda_\delta := \begin{cases} L_{\|\alpha^\delta\|_0 + N} & \text{if } \delta < \epsilon; \\ 0 & \text{else.} \end{cases}$$

Proof: Since $\mathcal{J}^\circ = \operatorname{supp}(\alpha^o)$, α^o solves (3), and α^δ solves (1), we have

$$\begin{aligned} \Phi_{\mathcal{J}^o}\alpha^o_{\mathcal{J}^o} &= y^o, & (16) \\ \Phi_{\mathcal{J}^o}\alpha^\delta_{\mathcal{J}^o} + \Phi_{\mathcal{J}^o_c}\alpha^\delta_{\mathcal{J}^o_c} &= y + w & (17) \end{aligned}$$

for some $w \in \Re^n$ with $\|w\|_2 \leq \delta$. Moreover, either $\|w\|_2 = \delta$ or $\alpha^\delta = 0$.

Assume that $N \geq 1$, $\mu(N-1) < 1$ (so that $\rho > 1$), and $\rho\mu N < 1$. If $\delta \geq \epsilon$, then α^o is feasible for (1) and we let $\hat{\alpha} := \alpha^o$. Otherwise $0 \leq \delta < \epsilon$, and α^o is not feasible for (1). In this case, we construct below a feasible point $\hat{\alpha}$ of (1) that is near α^o. Let

$$\Delta := y - y^o + w, \qquad \bar{\mathcal{K}} := \operatorname{supp}(\alpha^\delta - \alpha^o). \tag{18}$$

Then $\Phi_{\bar{\mathcal{K}}}(\alpha^\delta - \alpha^o)_{\bar{\mathcal{K}}} = \Delta$, i.e., $\Delta \in \operatorname{Span}(\Phi_{\bar{\mathcal{K}}})$. Moreover, $\Delta \neq 0$ (since either $\|w\|_2 = \delta < \epsilon = \|y - y^o\|_2$ or $\alpha^\delta = 0$, implying $y + w = 0 \neq y^o$). By dropping linearly dependent columns if necessary, we have that $\Delta \in \operatorname{Span}(\Phi_{\mathcal{K}})$ for some nonempty $\mathcal{K} \subseteq \bar{\mathcal{K}}$ such that the columns of $\Phi_{\mathcal{K}}$ are linearly independent. Then there exists a unique $\beta \in \Re^m$ satisfying

$$\Phi_{\mathcal{K}}\beta_{\mathcal{K}} = \Delta, \qquad \beta_{\mathcal{K}_c} = 0.$$

Multiplying both sides of the first equation by $\Phi_{\mathcal{K}}^T$ on the left yields

$$\beta_{\mathcal{K}} = (\Phi_{\mathcal{K}}^T\Phi_{\mathcal{K}})^{-1}\Phi_{\mathcal{K}}^T\Delta.$$

Let

$$\hat{\alpha} := \alpha^o + \beta.$$

Then $\Phi\hat{\alpha} = y + w$ so $\hat{\alpha}$ is feasible for (1).

Since $\rho\mu N < 1$ and $\rho > 1$, we have $\mu(N-1) < 1$ and Lemma 1(b) implies $\Phi_{\mathcal{J}^o}^T\Phi_{\mathcal{J}^o}$ is invertible. Let $E := (\Phi_{\mathcal{J}^o}^T\Phi_{\mathcal{J}^o})^{-1}$. Multiplying (16) and (17) by $E\Phi_{\mathcal{J}^o}^T$ on the left yields

$$\alpha^\delta_{\mathcal{J}^o} + EC\alpha^\delta_{\mathcal{J}^o_c} = E\Phi_{\mathcal{J}^o}^T(y^o + \Delta) = \alpha^o_{\mathcal{J}^o} + E\Phi_{\mathcal{J}^o}^T\Delta, \tag{19}$$

where $C := \Phi_{\mathcal{J}^o}^T\Phi_{\mathcal{J}^o_c} \in \Re^{N\times(m-N)}$. This and the triangle inequaliy yield

$$\begin{aligned} & \|\hat{\alpha}_{\mathcal{J}^o}\|_1 - \|\alpha^\delta_{\mathcal{J}^o}\|_1 \\ \leq\ & \|\hat{\alpha}_{\mathcal{J}^o} - \alpha^\delta_{\mathcal{J}^o}\|_1 \\ =\ & \|\hat{\alpha}_{\mathcal{J}^o} - \alpha^o_{\mathcal{J}^o} - E\Phi_{\mathcal{J}^o}^T\Delta + EC\alpha^\delta_{\mathcal{J}^o_c}\|_1 \\ \leq\ & \|\hat{\alpha}_{\mathcal{J}^o} - \alpha^o_{\mathcal{J}^o}\|_1 + \|E\Phi_{\mathcal{J}^o}^T\Delta\|_1 + \|EC\|_1\|\alpha^\delta_{\mathcal{J}^o_c}\|_1. \end{aligned}$$

This in turn yields

$$\begin{aligned} \|\hat{\alpha}\|_1 &= \|\hat{\alpha}_{\mathcal{J}^o}\|_1 + \|\hat{\alpha}_{\mathcal{J}^o_c}\|_1 \\ &\leq \|\alpha^\delta_{\mathcal{J}^o}\|_1 + \|\hat{\alpha}_{\mathcal{J}^o_c}\|_1 + \|\hat{\alpha}_{\mathcal{J}^o} - \alpha^o_{\mathcal{J}^o}\|_1 \\ &\quad + \|E\Phi_{\mathcal{J}^o}^T\Delta\|_1 + \|EC\|_1\|\alpha^\delta_{\mathcal{J}^o_c}\|_1 \\ &= \|\alpha^\delta\|_1 - (1 - \|EC\|_1)\|\alpha^\delta_{\mathcal{J}^o_c}\|_1 \\ &\quad + \|\hat{\alpha} - \alpha^o\|_1 + \|E\Phi_{\mathcal{J}^o}^T\Delta\|_1. \end{aligned}$$

Since $\hat{\alpha}$ is feasible for (1) and α^δ is a solution of (1), the left-hand side is greater than or equal to $\|\alpha^\delta\|_1$. Thus

$$(1-\|EC\|_1)\|\alpha^\delta_{\mathcal{J}^o_c}\|_1 \le \|\hat{\alpha}-\alpha^o\|_1 + \|E\Phi^T_{\mathcal{J}^o}\Delta\|_1. \quad (20)$$

By (4) and Lemma 1(b), for $i \in \mathcal{J}^o$ and $j \in \mathcal{J}^o_c$,

$$\begin{aligned} |(EC)_{ij}| &= \left|\sum_{k\in\mathcal{J}^o} E_{ik}C_{kj}\right| \le \sum_{k\in\mathcal{J}^o}|E_{ik}|\mu \\ &\le \left(1+\frac{(\mu^2+\mu)(N-1)}{1-\mu(N-1)}\right)\mu \\ &= \frac{1+\mu^2(N-1)}{1-\mu(N-1)}\mu \;=\; \rho\mu, \end{aligned}$$

so that

$$\|EC\|_1 = \max_{j\in\mathcal{J}^o_c}\sum_{i\in\mathcal{J}^o}|(EC)_{ij}| \le \rho\mu N < 1.$$

This together with (20) yields

$$\|\alpha^\delta_{\mathcal{J}^o_c}\|_1 \le \frac{1}{1-\rho\mu N}\left(\|\hat{\alpha}-\alpha^o\|_1 + \|E\Phi^T_{\mathcal{J}^o}\Delta\|_1\right).$$

We now bound the right-hand side. We have $\|\Delta\|_2 = \|y - y^o + w\|_2 \le \|y-y^o\|_2 + \|w\|_2 = \epsilon+\delta$. Thus properties of $\|\cdot\|_1, \|\cdot\|_2$ and (11) in Lemma 1(b) yield

$$\begin{aligned} \|E\Phi^T_{\mathcal{J}^o}\Delta\|_1 &\le \sqrt{N}\|E\Phi^T_{\mathcal{J}^o}\Delta\|_2 \\ &\le \sqrt{N}\|E\Phi^T_{\mathcal{J}^o}\|_2\|\Delta\|_2 \\ &\le \frac{\sqrt{N}(\epsilon+\delta)}{\sqrt{1-\mu(N-1)}} \\ &= \sqrt{N}\vartheta(\epsilon+\delta). \end{aligned}$$

If $\delta \ge \epsilon$, then $\hat{\alpha}-\alpha^o = 0$; otherwise

$$\begin{aligned} \|\hat{\alpha}-\alpha^o\|_1 &= \|\beta\|_1 \\ &= \|(\Phi^T_{\mathcal{K}}\Phi_{\mathcal{K}})^{-1}\Phi^T_{\mathcal{K}}\Delta\|_1 \\ &\le \|(\Phi^T_{\mathcal{K}}\Phi_{\mathcal{K}})^{-1}\Phi^T_{\mathcal{K}}\|_{2,1}\|\Delta\|_2 \\ &\le L_{|\mathcal{K}|}(\epsilon+\delta) \\ &\le L_{\|\alpha^\delta\|_0+N}(\epsilon+\delta), \end{aligned}$$

where the last inequality uses (12) and $|\mathcal{K}| \le |\bar{\mathcal{K}}| \le \|\alpha^\delta\|_0 + N$ (see (18)). The above three inequalities together with the definition of λ_δ imply

$$\|\alpha^\delta_{\mathcal{J}^o_c}\|_1 \le \frac{1}{1-\rho\mu N}\left(\lambda_\delta + \sqrt{N}\vartheta\right)(\epsilon+\delta).$$

Also, (19) yields

$$\begin{aligned} \|\alpha^\delta_{\mathcal{J}^o} - \alpha^o_{\mathcal{J}^o}\|_1 &= \|E\Phi^T_{\mathcal{J}^o}\Delta - EC\alpha^\delta_{\mathcal{J}^o_c}\|_1 \\ &\le \|E\Phi^T_{\mathcal{J}^o}\Delta\|_1 + \|EC\|_1\|\alpha^\delta_{\mathcal{J}^o_c}\|_1 \\ &\le \sqrt{N}\vartheta(\epsilon+\delta) + \rho\mu N\|\alpha^\delta_{\mathcal{J}^o_c}\|_1. \end{aligned}$$

Combining the above two inequalities with $\|\alpha^\delta - \alpha^o\|_1 \le \|\alpha^\delta_{\mathcal{J}^o} - \alpha^o_{\mathcal{J}^o}\|_1 + \|\alpha^\delta_{\mathcal{J}^o_c}\|_1$ yields (14).

Finally, using Lemma 1(a) with $\mathcal{J} = \{1,\ldots,m\}$ and $d = \alpha^\delta - \alpha^o$, we have

$$\begin{aligned} &(1+\mu)\|\alpha^\delta-\alpha^o\|_2^2 \\ \le\;& \|\Phi(\alpha^\delta-\alpha^o)\|_2^2 + \mu\|\alpha^\delta-\alpha^o\|_1^2 \\ =\;& \|\Delta\|_2^2 + \mu\|\alpha^\delta-\alpha^o\|_1^2 \\ \le\;& (\epsilon+\delta)^2 + \mu\left(\sqrt{N}\vartheta + \frac{1+\rho\mu N}{1-\rho\mu N}\cdot \right. \\ & \left.\left(\lambda_\delta + \sqrt{N}\vartheta\right)\right)^2(\epsilon+\delta)^2, \end{aligned}$$

where the last inequality uses (14). This proves (15). ■

Below we compare Theorem 1 with existing results.

1. By letting $R = \mu(N-1)$ we can write $\rho\mu N < 1$ equivalently as $(1+\mu R)(R+\mu) < 1-R$, which is a quadratic inequality in R. Solving this gives

$$\mu(N-1) < \frac{1-\mu}{\sqrt{1+\mu+\frac{1}{4}\mu^4}+1+\frac{1}{2}\mu^2}. \quad (21)$$

   If $N=1$, this is clearly satisfied whenever $\mu<1$. It can be verified numerically that if $N=2$, this is satisfied whenever $\mu \le .31$; and if $N=3$, this is satisfied whenever $\mu \le .19$. As $\mu \to 0$, the right-hand side of (21) increases towards $\frac{1}{2}$. Thus $\mu(N-1) < \frac{1}{2} - O(\mu)$ or, equivalently, $N < (\frac{1}{2}-O(\mu))\mu^{-1}+1$.
2. For the noiseless case of $\delta=\epsilon=0$ considered in [11], [17], [22], [30], Theorem 1 says that $\alpha^\delta = \alpha^o$ whenever (21) holds. As $\mu\to 0$, (21) approaches the sufficient condition $\mu(N-\frac{1}{2}) \le \frac{1}{2}$ derived in [11], [17], [22], [30].
3. For the noisy case of $\delta \ge \epsilon$ considered in [12, Theorem 3.1], Theorem 1 yields

$$\|\alpha^\delta-\alpha^o\|_1 \le \left(1+\frac{1+\rho\mu N}{1-\rho\mu N}\right)\sqrt{N}\vartheta(\epsilon+\delta),$$

$$\begin{aligned} &\|\alpha^\delta-\alpha^o\|_2^2 \\ &\le \left(1+\left(1+\frac{1+\rho\mu N}{1-\rho\mu N}\right)^2\mu N\vartheta^2\right)\frac{(\epsilon+\delta)^2}{1+\mu} \end{aligned} \quad (22)$$

   whenever (21) holds. In contrast, the bound (5) requires the stricter condition $\mu < (4N-1)^{-1}$. Suppose ϵ,δ,N are fixed. As $\mu \to (4N-1)^{-1}$, the right-hand side of (5) tends to ∞ while the right-hand side of (22) remains uniformly bounded. As $\mu \to 0$, the right-hand side of (5) has the form $(1+(4N-1)\mu+o(\mu))(\epsilon+\delta)^2$ while $\rho\to 1$ and $\vartheta \to 1$, so the right-hand side of (22) has the form $\frac{(1+4N\mu+o(\mu))}{1+\mu}(\epsilon+\delta)^2$. Since $\frac{1+4N\mu}{1+\mu} = 1+(4N-1)\mu+o(\mu)$, we see that (5) and (22) are equally sharp up to first-order in μ.
4. For the noisy case of $\delta<\epsilon$, we have from (13) that $\lambda_\delta \le 1/\sqrt{\frac{1+\mu}{\|\alpha^\delta\|_0+N}-\mu}$ whenever $\|\alpha^\delta\|_0 + N < \mu^{-1}+1$. If $\|\alpha^\delta\|_0 + N \le \mu^{-1}/\kappa$, with $\kappa>1$, then $\lambda_\delta \le 1/\sqrt{\mu(\kappa-1)}$ and (15) yields $\|\alpha^\delta-\alpha^o\|_2 \le \kappa'(\epsilon+\delta)$, where κ' depends on κ.

The following example shows that, even in the noiseless case of $\delta=\epsilon=0$, the sparsest solution and the least ℓ_1-norm solution can be arbitrarily far apart when the mutual coherence μ approaches 1. In this example, all except one column of Φ are pairwise orthogonal.

Example 1: Suppose that $m = n+1$, $y = y^o$, $\delta = 0$, and

$$\begin{aligned} \phi_j &= j\text{th unit coordinate vector}, \ j = 1, \cdots, n, \\ \phi_m &= \frac{ty^o - \phi_1}{\theta}, \end{aligned}$$

where $t > 0$ and $\theta := \|ty^o - \phi_1\|_2 > 0$. Then $\|\phi_j\|_2 = 1$ for $j = 1, \ldots, m$, $\phi_i^T \phi_j = 0$ for all $1 \le i \ne j \le n$. (Note that $\theta \to 1$ and $\phi_m \to -\phi_1$ as $t \to 0$.)

For any feasible point α of (1), i.e., $\Phi\alpha = y^o$, we can solve for $\alpha_1, \ldots, \alpha_n$ in terms of α_m to obtain

$$\begin{aligned} \alpha_1 &= y_1^o + \left(\frac{1 - y_1^o t}{\theta} \right) \alpha_m, \\ \alpha_j &= y_j^o \left(1 - \frac{t}{\theta} \alpha_m \right), \quad j = 2, \ldots, n. \end{aligned}$$

Thus

$$\|\alpha\|_1 = \left| y_1^o + \left(\frac{1 - y_1^o t}{\theta} \right) \alpha_m \right| + \left(\sum_{j=2}^n |y_j^o| \right) \left| 1 - \frac{t}{\theta} \alpha_m \right| + |\alpha_m|.$$

The right-hand side is a piecewise-linear function of α_m, with three breakpoints. To simplify the analysis, we further choose y^o and t so that

$$0 < y_1^o < \frac{1}{t}, \qquad -1 < \frac{1 - \|y^o\|_1 t}{\theta} < 1,$$
$$y_i^o \ne 0, \ i = 2, \cdots, n.$$

(For example, choose $y^o > 0$ and set $t = 1/\|y^o\|_1$.) Then the three breakpoints can be ordered as

$$\frac{-\theta y_1^o}{1 - y_1^o t} < 0 < \frac{\theta}{t}.$$

The derivative of the right-hand side with respect to α_m on the interval $\frac{-\theta y_1^o}{1 - y_1^o t} < \alpha_m < 0$ is

$$\left(\frac{1 - y_1^o t}{\theta} \right) - \left(\sum_{j=2}^n |y_j^o| \right) \frac{t}{\theta} - 1 = \frac{1 - \|y^o\|_1 t}{\theta} - 1 < 0.$$

The derivative of the right-hand side with respect to α_m on the interval $0 < \alpha_m < \frac{\theta}{t}$ is

$$\left(\frac{1 - y_1^o t}{\theta} \right) - \left(\sum_{j=2}^n |y_j^o| \right) \frac{t}{\theta} + 1 = \frac{1 - \|y^o\|_1 t}{\theta} + 1 > 0.$$

Thus

$$\alpha_j^\delta = y_j^o, \ j = 1, \ldots, n, \qquad \alpha_m^\delta = 0,$$

is a stationary point of (1) and hence a solution of (1) (since (1) is a convex program).

On the other hand,

$$\alpha_1^o = \frac{1}{t}, \quad \alpha_2^o = \cdots = \alpha_n^o = 0, \quad \alpha_m^o = \frac{\theta}{t}$$

is feasible for (3) and has 2 nonzeros. Since y^o is not a scalar multiple of any of $\phi_1, \ldots, \phi_m$ so that (3) cannot have a feasible point with fewer nonzeros, this is a solution of (3). As $t \to 0$, α^δ and α^o become arbitrarily far apart. Note that $\|\alpha^\delta\|_0 = n$ while $\|\alpha^o\|_0 = 2$.

Not surprisingly, $|\phi_1^T \phi_m| \to 1$ as $t \to 0$, so that $\mu \to 1$. This example shows that the sparsest solution and the least ℓ_1-norm solution can be arbitrarily far apart when the mutual coherence μ approaches 1.

By orthogonal subspace decomposition [25, page 5], we can uniquely express

$$y = y^\perp + y^{\mathrm{I}} + y^{\mathrm{II}}$$

for some $y^\perp \in \mathrm{Null}(\Phi^T)$, $y^{\mathrm{I}} \in \mathrm{Span}(\Phi) \cap \mathrm{Null}(\Phi_{\mathcal{J}^o}^T)$, and $y^{\mathrm{II}} \in \mathrm{Span}(\Phi_{\mathcal{J}^o})$. Then $\|\Phi\alpha - y\|_2^2 = \|\Phi\alpha - y^{\mathrm{I}} - y^{\mathrm{II}}\|_2^2 + \|y^\perp\|_2^2$ and, by replacing δ^2 with $\delta^2 - \|y^\perp\|_2^2$, we can without loss of generality assume that $y^\perp = 0$, i.e., $y \in \mathrm{Span}(\Phi)$. The following theorem refines [12, Theorem 4.1] and shows that α^δ has a support identification property whenever δ is sufficiently large relative to $\sqrt{N}\|y^{\mathrm{I}}\|_2$.

Theorem 2: Assume $y \in \mathrm{Span}(\Phi)$. If $N \ge 1$, $\mu(2N-1) < 1$, and

$$\delta > \epsilon, \quad \delta^2 > \|y^{\mathrm{I}}\|_2^2 + \frac{1 - \mu(N-1)}{(1 - \mu(2N-1))^2} N \|y^{\mathrm{I}}\|_2^2, \tag{23}$$

then $\mathrm{supp}(\alpha^\delta) \subseteq \mathcal{J}^o$.

Proof: For notational simplicity, let

$$Q(\alpha) := \frac{1}{2} \left(\|\Phi\alpha - y\|_2^2 - \delta^2 \right).$$

Thus α is feasible for (1) if and only if $Q(\alpha) \le 0$. Consider (1) with $\mathrm{supp}(\alpha)$ restricted to $\mathcal{J}^o$:

$$\min_{\alpha \in \Re^m} \|\alpha\|_1 \quad \text{subject to} \quad Q(\alpha) \le 0, \ \alpha_{\mathcal{J}_c^o} = 0. \tag{24}$$

Since $\delta > \epsilon$, α^o is feasible for (24) and $Q(\alpha^o) < 0$. Let $\bar\alpha$ be any solution of (24). If $Q(\bar\alpha) < 0$, then it must be that $\bar\alpha = 0$ and hence $\alpha^\delta = 0$, implying $\mathrm{supp}(\alpha^\delta) \subseteq \mathcal{J}^o$. It remains to consider the case of $Q(\bar\alpha) = 0$.

First, we show that $\bar\alpha$ is a solution of (1). Since (24) is a convex program and $Q(\alpha^o) < 0$, this implies (24) has a Lagrange multiplier [25, Theorem 28.2], i.e., there exists a $\pi \in [0, \infty)$ satisfying the following optimality condition for (24):

$$s_{\mathcal{J}^o} + g_{\mathcal{J}^o} \pi = 0, \tag{25}$$

where $s_{\mathcal{J}^o} \in [-1, 1]^N$ (with $s_j = \mathrm{sign}(\bar\alpha_j)$ if $\bar\alpha_j \ne 0$) and we let $g := \nabla Q(\bar\alpha)$. Since $\bar\alpha \ne 0$, we have $s_{\mathcal{J}^o} \ne 0$ and hence $\pi > 0$. We will use (25) to show that

$$s_{\mathcal{J}_c^o} + g_{\mathcal{J}_c^o} \pi = 0$$

for some $s_{\mathcal{J}_c^o} \in [-1, 1]^{m-N}$, or, equivalently,

$$\|g_{\mathcal{J}_c^o} \pi\|_\infty \le 1. \tag{26}$$

Since $\bar\alpha_{\mathcal{J}_c^o} = 0$, this will show that $\bar\alpha$ satisfies the optimality condition for the convex program (1) and hence is a solution of (1).

Since $g = \Phi^T(\Phi\bar\alpha - y)$, (25) yields

$$\Phi_{\mathcal{J}^o}^T (\Phi_{\mathcal{J}^o} \bar\alpha_{\mathcal{J}^o} - y)\pi = -s_{\mathcal{J}^o}.$$

Since $\mu(2N-1) < 1$, Lemma 1(b) implies $E := (\Phi_{\mathcal{J}^o}^T\Phi_{\mathcal{J}^o})^{-1}$ is well defined, so we can solve for $\bar{\alpha}_{\mathcal{J}^o}\pi$:

$$\bar{\alpha}_{\mathcal{J}^o}\pi = E(\Phi_{\mathcal{J}^o}^T y\pi - s_{\mathcal{J}^o}) \tag{27}$$

and plug it into the left-hand side of (26) to obtain

$$\begin{aligned} & \|g_{\mathcal{J}_c^o}\pi\|_\infty \\ =\ & \|\Phi_{\mathcal{J}_c^o}^T(\Phi_{\mathcal{J}^o}E(\Phi_{\mathcal{J}^o}^T y\pi - s_{\mathcal{J}^o}) - y\pi)\|_\infty \\ \le\ & \|\Phi_{\mathcal{J}_c^o}^T\Phi_{\mathcal{J}^o}Es_{\mathcal{J}^o}\|_\infty + \|\Phi_{\mathcal{J}_c^o}^T(\Phi_{\mathcal{J}^o}E\Phi_{\mathcal{J}^o}^T - I)y\pi\|_\infty \\ =\ & \max_{j\in\mathcal{J}_c^o}|\phi_j^T\Phi_{\mathcal{J}^o}Es_{\mathcal{J}^o}| + \max_{j\in\mathcal{J}_c^o}|\phi_j^T y^{\mathrm{I}}|\pi \\ \le\ & \mu\|Es_{\mathcal{J}^o}\|_1 + \max_{j\in\mathcal{J}_c^o}\|\phi_j\|_2\|y^{\mathrm{I}}\|_2\pi \\ \le\ & \mu\sqrt{N}\|Es_{\mathcal{J}^o}\|_2 + \|y^{\mathrm{I}}\|_2\pi \\ \le\ & \mu\sqrt{N}\|E\|_2\|s_{\mathcal{J}^o}\|_2 + \|y^{\mathrm{I}}\|_2\pi \\ \le\ & \frac{\mu N}{1-\mu(N-1)} + \|y^{\mathrm{I}}\|_2\pi, \end{aligned} \tag{28}$$

where the last inequality also uses (10) in Lemma 1(b).

Since $Q(\bar{\alpha}) = 0$, we have

$$\begin{aligned}\delta^2 &= \|\Phi\bar{\alpha} - y\|_2^2 = \|\Phi_{\mathcal{J}^o}\bar{\alpha}_{\mathcal{J}^o} - y^{\mathrm{I}} - y^{\mathrm{II}}\|_2^2 \\ &= \|\Phi_{\mathcal{J}^o}\bar{\alpha}_{\mathcal{J}^o} - y^{\mathrm{II}}\|_2^2 + \|y^{\mathrm{I}}\|_2^2.\end{aligned}$$

Also, using (27), $\Phi_{\mathcal{J}^o}^T y = \Phi_{\mathcal{J}^o}^T y^{\mathrm{II}}$ and (11), we have

$$\begin{aligned}\|\Phi_{\mathcal{J}^o}\bar{\alpha}_{\mathcal{J}^o} - y^{\mathrm{II}}\|_2\pi &= \|\Phi_{\mathcal{J}^o}E(\Phi_{\mathcal{J}^o}^T y^{\mathrm{II}}\pi - s_{\mathcal{J}^o}) - y^{\mathrm{II}}\pi\|_2 \\ &= \|\Phi_{\mathcal{J}^o}Es_{\mathcal{J}^o}\|_2 \\ &\le \|\Phi_{\mathcal{J}^o}E\|_2\|s_{\mathcal{J}^o}\|_2 \\ &\le \|E\Phi_{\mathcal{J}^o}^T\|_2\sqrt{N}\|s_{\mathcal{J}^o}\|_\infty \\ &\le \frac{\sqrt{N}}{\sqrt{1-\mu(N-1)}}.\end{aligned}$$

The above two relations yield

$$\pi \le \frac{\sqrt{N}\vartheta}{\|\Phi_{\mathcal{J}^o}\bar{\alpha}_{\mathcal{J}^o} - y^{\mathrm{II}}\|_2} = \frac{\sqrt{N}\vartheta}{\sqrt{\delta^2 - \|y^{\mathrm{I}}\|_2^2}},$$

where we let $\vartheta := 1/\sqrt{1-\mu(N-1)}$. This together with (28) yields

$$\|g_{\mathcal{J}_c^o}\pi\|_\infty \le \frac{\mu N}{1-\mu(N-1)} + \frac{\|y^{\mathrm{I}}\|_2\sqrt{N}\vartheta}{\sqrt{\delta^2 - \|y^{\mathrm{I}}\|_2^2}} < 1, \tag{29}$$

where the strict inequality is equivalent to (23). To see the equivalence, substract $\frac{\mu N}{1-\mu(N-1)}$ from both sides of the strict inequality and then square both sides and simplify to obtain

$$\frac{\|y^{\mathrm{I}}\|_2^2 N}{\delta^2 - \|y^{\mathrm{I}}\|_2^2} < \frac{(1-\mu(2N-1))^2}{1-\mu(N-1)}.$$

Rearranging terms yields (23). This proves (26).

Finally, we show that $\alpha_{\mathcal{J}_c^o}^\delta = 0$ for any solution α^δ of (1). Since $\bar{\alpha}$ and α^δ are both solutions of (1), we have

$$\begin{aligned} 0 &= \frac{\|\bar{\alpha} + t(\alpha^\delta - \bar{\alpha})\|_1 - \|\bar{\alpha}\|_1}{t} \\ &= (\alpha_{\bar{\mathcal{J}}}^\delta - \bar{\alpha}_{\bar{\mathcal{J}}})^T\mathrm{sign}(\bar{\alpha}_{\bar{\mathcal{J}}}) + \|\alpha_{\bar{\mathcal{J}}_c}^\delta\|_1 \end{aligned} \tag{30}$$

for all $0 < t \le 1$ sufficiently small, where $\bar{\mathcal{J}} := \mathrm{supp}(\bar{\alpha}) \subseteq \mathcal{J}^o$. Also, since $Q(\bar{\alpha}) = 0$ and $Q(\alpha^\delta) \le 0$, we have

$$0 \ge (\alpha^\delta - \bar{\alpha})^T g.$$

Multiplying this by π and adding to (30) yields

$$\begin{aligned} 0 &\ge (\alpha_{\bar{\mathcal{J}}}^\delta - \bar{\alpha}_{\bar{\mathcal{J}}})^T\mathrm{sign}(\bar{\alpha}_{\bar{\mathcal{J}}}) + \|\alpha_{\bar{\mathcal{J}}_c}^\delta\|_1 + (\alpha^\delta - \bar{\alpha})^T g\pi \\ &= \|\alpha_{\bar{\mathcal{J}}_c}^\delta\|_1 + (\alpha_{\bar{\mathcal{J}}_c}^\delta)^T g_{\bar{\mathcal{J}}_c}\pi \\ &= \|\alpha_{\mathcal{J}^o\setminus\bar{\mathcal{J}}}^\delta\|_1 + \|\alpha_{\mathcal{J}_c^o}^\delta\|_1 + (\alpha_{\mathcal{J}^o\setminus\bar{\mathcal{J}}}^\delta)^T g_{\mathcal{J}^o\setminus\bar{\mathcal{J}}}\pi \\ &\quad + (\alpha_{\mathcal{J}_c^o}^\delta)^T g_{\mathcal{J}_c^o}\pi \\ &\ge \|\alpha_{\mathcal{J}^o\setminus\bar{\mathcal{J}}}^\delta\|_1 + \|\alpha_{\mathcal{J}_c^o}^\delta\|_1 - \|\alpha_{\mathcal{J}^o\setminus\bar{\mathcal{J}}}^\delta\|_1\|g_{\mathcal{J}^o\setminus\bar{\mathcal{J}}}\pi\|_\infty \\ &\quad - \|\alpha_{\mathcal{J}_c^o}^\delta\|_1\|g_{\mathcal{J}_c^o}\pi\|_\infty, \end{aligned}$$

where the first equality uses $\bar{\alpha}_{\bar{\mathcal{J}}_c} = 0$ and $-g_{\bar{\mathcal{J}}}\pi = \mathrm{sign}(\bar{\alpha}_{\bar{\mathcal{J}}})$ (see (25)). By (25), $\|g_{\mathcal{J}^o\setminus\bar{\mathcal{J}}}\pi\|_\infty \le 1$ while, by (29), $\|g_{\mathcal{J}_c^o}\pi\|_\infty < 1$. Thus, $\alpha_{\mathcal{J}_c^o}^\delta = 0$. ■

How does condition (23) compare with (6), assuming $1 \le N < \frac{1}{2}\mu^{-1}$? Since

$$\epsilon^2 = \|y - y^o\|_2^2 \ge \min_{w\in\mathrm{Span}(\Phi_{\mathcal{J}^o})}\|y - w\|_2^2 = \|y - y^{\mathrm{II}}\|_2^2 = \|y^{\mathrm{I}}\|_2^2,$$

we see that (6) implies

$$\begin{aligned}\delta &> \frac{1-\mu N + \sqrt{1-\mu N}\sqrt{N}}{1-2\mu N}\,\epsilon \\ &= \left(1 + \frac{\mu N + \sqrt{(1-\mu N)N}}{1-2\mu N}\right)\epsilon \\ &> \left(1 + \frac{\sqrt{\mu N + (1-\mu N)N}}{1-2\mu N+\mu}\right)\|y^{\mathrm{I}}\|_2 \\ &> \sqrt{1 + \frac{\mu N + (1-\mu N)N}{(1-2\mu N+\mu)^2}}\,\|y^{\mathrm{I}}\|_2,\end{aligned}$$

where the second inequality also uses $a + \sqrt{b} \ge \sqrt{a+b}$ for $a \ge 0, b \ge \frac{1}{4}$, and the last inequality uses $1 + a \ge \sqrt{1+a^2}$ for $a \ge 0$. Since the right-hand side equals the square root of the right-hand side of (23), this shows that (6) implies (23). Notice that (23) relaxes (6) significantly only when N is small or $\|y^{\mathrm{I}}\|_2$ is small relative to ϵ (i.e., the noise is concentrated in $\mathrm{Span}(\Phi_{\mathcal{J}^o})$).

IV. Sparse Representation from Orthogonal Greedy

In [12], Donoho, Elad, and Temlyakov proposed the following noise-aware version of an orthogonal greedy algorithm (OGA) [7], [8], [24], [28] to find a sparse α with residual $\|\Phi\alpha - y\|_2$ below a prescribed tolerance.

OGA:

0. Input $\delta^{\mathrm{OGA}} \ge 0$. Initialize $\mathcal{J} \leftarrow \emptyset$. Go to Step 1.
1. Let $\alpha_{\mathcal{J}}$ be any solution of

$$\min_{\alpha_{\mathcal{J}}\in\Re^{|\mathcal{J}|}}\|\Phi_{\mathcal{J}}\alpha_{\mathcal{J}} - y\|_2, \tag{31}$$

Let $r = \Phi_{\mathcal{J}}\alpha_{\mathcal{J}} - y$. If $\|r\|_2 \le \delta^{\mathrm{OGA}}$, then output $\mathcal{J}$ and $\alpha_{\mathcal{J}}$; otherwise choose

$$\bar{j} \in \arg\max_{j\in\mathcal{J}_c}|\phi_j^T r|,$$

update $\mathcal{J} \leftarrow \mathcal{J} \cup \{\bar{j}\}$, and return to Step 1.

In the noiseless case of $\epsilon = 0$, Tropp [30, Theorems A and B] showed that $\alpha^{\text{OGA}} = \alpha^o$ (equivalently, $\text{supp}(\alpha^{\text{OGA}}) = \mathcal{J}^\circ$) whenever $\delta^{\text{OGA}} = 0$ and $\mu(2N-1) < 1$. In the general case, Donoho, Elad, and Temylakov [12, Theorem 5.1(a)] showed that $\text{supp}(\alpha^{\text{OGA}}) = \mathcal{J}^\circ$ whenever $\delta^{\text{OGA}} = \epsilon$, $\mu(2N-1) < 1$, and (7) holds. When $\epsilon = 0$, this recovers Tropp's result (since (7) holds trivially). The following theorem extends these two results to allow $\delta^{\text{OGA}} \geq \epsilon$. This shows that OGA can still identify the sparsest support when using an overestimate of the noise level ϵ.

Theorem 3: Assume $\mu(N-1) < 1$. The following results hold.

(a) For any nonempty $\mathcal{J} \subseteq \mathcal{J}^\circ$ and any solution $\alpha_{\mathcal{J}}$ of (31),

$$\|\alpha_{\mathcal{J}} - \alpha^o_{\mathcal{J}}\|_2 \leq \frac{\mu\sqrt{|\mathcal{J}|}\|\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}\|_1}{1-\mu(|\mathcal{J}|-1)} + \frac{\epsilon}{\sqrt{1-\mu(|\mathcal{J}|-1)}}.$$

(b) If $\delta^{\text{OGA}} \geq \epsilon$ and

$$\|\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}\|_\infty > 2\left(\epsilon + \frac{\mu+\mu^2}{1-\mu(|\mathcal{J}|-1)}\|\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}\|_1\right) \quad \forall \emptyset \neq \mathcal{J} \subset \mathcal{J}^\circ, \tag{32}$$

then $\mathcal{J}^{\text{OGA}} \subseteq \mathcal{J}^\circ$, where $\mathcal{J}^{\text{OGA}}$ is an output of the OGA. If in addition

$$\|\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}\|_2 > \frac{\frac{\mu\sqrt{|\mathcal{J}|}}{\sqrt{1-\mu(|\mathcal{J}|-1)}}\|\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}\|_1 + \epsilon + \delta^{\text{OGA}}}{\sqrt{1-\mu(|\mathcal{J}^\circ\setminus\mathcal{J}|-1)}} \quad \forall \emptyset \neq \mathcal{J} \subset \mathcal{J}^\circ, \tag{33}$$

then $\mathcal{J}^{\text{OGA}} = \mathcal{J}^\circ$.

Proof: For any nonempty $\mathcal{J} \subseteq \mathcal{J}^\circ$, we have $\mu(|\mathcal{J}|-1) \leq \mu(N-1) < 1$, so Lemma 1 implies $E := (\Phi_{\mathcal{J}}^T\Phi_{\mathcal{J}})^{-1}$ is well defined. Since $\alpha_{\mathcal{J}}$ is a solution of (31), we have

$$\alpha_{\mathcal{J}} = E\Phi_{\mathcal{J}}^T y. \tag{34}$$

Also, $\Phi_{\mathcal{J}}\alpha^o_{\mathcal{J}} + \Phi_{\mathcal{J}^\circ\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}} = y^o$, so that

$$\alpha^o_{\mathcal{J}} + E\Phi_{\mathcal{J}}^T\Phi_{\mathcal{J}^\circ\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}} = E\Phi_{\mathcal{J}}^T y^o. \tag{35}$$

(a) For any nonempty $\mathcal{J} \subset \mathcal{J}^\circ$, we have from (34) and (35) that

$$\begin{aligned}
&\|\alpha_{\mathcal{J}} - \alpha^o_{\mathcal{J}}\|_2 \\
= \ & \|E\Phi_{\mathcal{J}}^T\Phi_{\mathcal{J}^\circ\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}} + E\Phi_{\mathcal{J}}^T(y-y^o)\|_2 \\
\leq \ & \|E\|_2\|\Phi_{\mathcal{J}}^T\Phi_{\mathcal{J}^\circ\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}\|_2 + \|E\Phi_{\mathcal{J}}^T\|_2\|y-y^o\|_2 \\
= \ & \|E\|_2\sqrt{\sum_{j\in\mathcal{J}}|\phi_j^T\Phi_{\mathcal{J}^\circ\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}|^2} + \|E\Phi_{\mathcal{J}}^T\|_2\epsilon \\
\leq \ & \frac{\sqrt{|\mathcal{J}|}\mu\|\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}\|_1}{1-\mu(|\mathcal{J}|-1)} + \frac{\epsilon}{\sqrt{1-\mu(|\mathcal{J}|-1)}},
\end{aligned}$$

where the last inequality uses (10) and (11) in Lemma 1, as well as $|\phi_j^T\Phi_{\mathcal{J}^\circ\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}| \leq \mu\|\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}\|_1$ for $j \in \mathcal{J}$.

(b) Assume $\delta^{\text{OGA}} \geq \epsilon$ and (32) holds. First we prove by induction that $\mathcal{J} \subseteq \mathcal{J}^\circ$ at Step 1 of the OGA. This is true when initially $\mathcal{J} = \emptyset$. Suppose this is true at the beginning of Step 1. We show that it remains true at the end of Step 1. There are two cases to consider: (i) $\mathcal{J} \subset \mathcal{J}^\circ$ and (ii) $\mathcal{J} = \mathcal{J}^\circ$ at the beginning of Step 1.

In case (i), by (34) and (35), for any $j \in \mathcal{J}_c$,

$$\begin{aligned}
\phi_j^T r &= \phi_j^T(\Phi_{\mathcal{J}}\alpha_{\mathcal{J}} - y) \\
&= \phi_j^T\left(\Phi_{\mathcal{J}}(\alpha_{\mathcal{J}} - \alpha^o_{\mathcal{J}}) - \Phi_{\mathcal{J}^\circ\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}} - (y-y^o)\right) \\
&= \phi_j^T\left((\Phi_{\mathcal{J}}E\Phi_{\mathcal{J}}^T - I)(y-y^o)\right. \\
&\quad \left. + \Phi_{\mathcal{J}}E\Phi_{\mathcal{J}}^T\Phi_{\mathcal{J}^\circ\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}} - \Phi_{\mathcal{J}^\circ\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}\right)
\end{aligned} \tag{36}$$

We now bound the right-hand terms of (36). Since $I - \Phi_{\mathcal{J}}E\Phi_{\mathcal{J}}^T$ is a projection matrix and $\|\phi_j\|_2 = 1$, the Cauchy-Schwarz inequality implies that

$$\begin{aligned}
&|\phi_j^T(\Phi_{\mathcal{J}}E\Phi_{\mathcal{J}}^T - I)(y-y^o)| \\
\leq \ & \|\phi_j\|_2\|\Phi_{\mathcal{J}}E\Phi_{\mathcal{J}}^T - I\|_2\|y-y^o\|_2 = \epsilon.
\end{aligned}$$

Also, by (10) in Lemma 1,

$$\begin{aligned}
&|\phi_j^T\Phi_{\mathcal{J}}E\Phi_{\mathcal{J}}^T\Phi_{\mathcal{J}^\circ\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}| \\
\leq \ & \|\Phi_{\mathcal{J}}^T\phi_j\|_2\|E\|_2\|\Phi_{\mathcal{J}}^T\Phi_{\mathcal{J}^\circ\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}\|_2 \\
\leq \ & \|\Phi_{\mathcal{J}}^T\phi_j\|_2\frac{1}{1-\mu(|\mathcal{J}|-1)}\sum_{j'\in\mathcal{J}^\circ\setminus\mathcal{J}}\|\Phi_{\mathcal{J}}^T\phi_{j'}\|_2|\alpha^o_{j'}| \\
\leq \ & \frac{\mu^2|\mathcal{J}|}{1-\mu(|\mathcal{J}|-1)}\|\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}\|_1,
\end{aligned}$$

where the last inequality uses $\|\Phi_{\mathcal{J}}^T\phi_{j'}\|_2 \leq \mu\sqrt{|\mathcal{J}|}$ for all $j' \in \mathcal{J}_c$. If $j \in \mathcal{J}_c^\circ$, then

$$\begin{aligned}
|\phi_j^T\Phi_{\mathcal{J}^\circ\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}| &= \left|\sum_{j'\in\mathcal{J}^\circ\setminus\mathcal{J}}\phi_j^T\phi_{j'}\alpha^o_{j'}\right| \\
&\leq \sum_{j'\in\mathcal{J}^\circ\setminus\mathcal{J}}|\phi_j^T\phi_{j'}||\alpha^o_{j'}| \\
&\leq \mu\|\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}\|_1.
\end{aligned}$$

Using this and the previous two inequalities to bound the right-hand side of (36) yields

$$|\phi_j^T r| \leq \mu\|\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}\|_1 + \epsilon + \frac{\mu^2|\mathcal{J}|}{1-\mu(|\mathcal{J}|-1)}\|\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}\|_1. \tag{37}$$

If $j \in \mathcal{J}^\circ \setminus \mathcal{J}$, then

$$\begin{aligned}
|\phi_j^T\Phi_{\mathcal{J}^\circ\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}| &= |\alpha^o_j + \phi_j^T\Phi_{\mathcal{J}^\circ\setminus(\mathcal{J}\cup\{j\})}\alpha^o_{\mathcal{J}^\circ\setminus(\mathcal{J}\cup\{j\})}| \\
&\geq |\alpha^o_j| - |\phi_j^T\Phi_{\mathcal{J}^\circ\setminus(\mathcal{J}\cup\{j\})}\alpha^o_{\mathcal{J}^\circ\setminus(\mathcal{J}\cup\{j\})}| \\
&\geq |\alpha^o_j| - \mu\|\alpha^o_{\mathcal{J}^\circ\setminus(\mathcal{J}\cup\{j\})}\|_1 \\
&\geq |\alpha^o_j| - \mu\|\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}\|_1.
\end{aligned} \tag{38}$$

Using this and the earlier two inequalities to bound the right-hand side of (36) yields

$$|\phi_j^T r| \geq |\alpha^o_j| - \mu\|\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}\|_1 - \epsilon - \frac{\mu^2|\mathcal{J}|}{1-\mu(|\mathcal{J}|-1)}\|\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}\|_1. \tag{39}$$

Since, by (32),

$$\max_{j\in\mathcal{J}^\circ\setminus\mathcal{J}}|\alpha^o_j| > 2\left(\epsilon + \left(\mu + \frac{\mu^2|\mathcal{J}|}{1-\mu(|\mathcal{J}|-1)}\right)\|\alpha^o_{\mathcal{J}^\circ\setminus\mathcal{J}}\|_1\right),$$

(37) and (39) together imply that

$$\max_{j\in\mathcal{J}^o\setminus\mathcal{J}} |\phi_j^T r| > \max_{j\in\mathcal{J}_c^o} |\phi_j^T r|.$$

Hence in Step 1 we choose $\bar{j} \in \mathcal{J}^o \setminus \mathcal{J}$, so that $\mathcal{J} \cup \{\bar{j}\} \subseteq \mathcal{J}^o$.

In case (ii), since $\mathcal{J} = \mathcal{J}^o$, we have $\|r\|_2 \le \|\Phi_{\mathcal{J}}\alpha^o_{\mathcal{J}} - y\|_2 = \|y^o - y\|_2 = \epsilon$. Since $\delta^{\rm OGA} \ge \epsilon$, then we have $\|r\|_2 \le \delta^{\rm OGA}$, so Step 1 would output $\mathcal{J}$ and $\alpha_{\mathcal{J}}$.

Suppose that $\mathcal{J}$ is a nonempty proper subset of $\mathcal{J}^o$. We have from (34) and (35) that

$$\begin{aligned}
&\|\Phi_{\mathcal{J}}(\alpha_{\mathcal{J}} - \alpha^o_{\mathcal{J}}) + y^o - y\|_2 \\
=\ & \|\Phi_{\mathcal{J}} E \Phi_{\mathcal{J}}^T \left(\Phi_{\mathcal{J}^o\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}} + y - y^o\right) + y^o - y\|_2 \\
=\ & \|\Phi_{\mathcal{J}} E \Phi_{\mathcal{J}}^T \Phi_{\mathcal{J}^o\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}} + (I - \Phi_{\mathcal{J}} E \Phi_{\mathcal{J}}^T)(y^o - y)\|_2 \\
\le\ & \|\Phi_{\mathcal{J}} E\|_2 \|\Phi_{\mathcal{J}}^T \Phi_{\mathcal{J}^o\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_2 \\
& + \|I - \Phi_{\mathcal{J}} E \Phi_{\mathcal{J}}^T\|_2 \|y^o - y\|_2 \\
=\ & \|E\Phi_{\mathcal{J}}^T\|_2 \sqrt{\sum_{j\in\mathcal{J}} |\phi_j^T \Phi_{\mathcal{J}^o\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}|^2} + \epsilon \\
\le\ & \frac{\sqrt{|\mathcal{J}|}\mu\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_1}{\sqrt{1-\mu(|\mathcal{J}|-1)}} + \epsilon,
\end{aligned}$$

where the second inequality uses (11) in Lemma 1(b), as well as $|\phi_j^T \Phi_{\mathcal{J}^o\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}| \le \mu\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_1$ for $j \in \mathcal{J}$. Thus

$$\begin{aligned}
&\|\Phi_{\mathcal{J}}\alpha_{\mathcal{J}} - y\|_2 \\
=\ & \|\Phi_{\mathcal{J}}(\alpha_{\mathcal{J}} - \alpha^o_{\mathcal{J}}) + \Phi_{\mathcal{J}}\alpha^o_{\mathcal{J}} - y^o + y^o - y\|_2 \\
\ge\ & \|\Phi_{\mathcal{J}}\alpha^o_{\mathcal{J}} - y^o\|_2 - \|\Phi_{\mathcal{J}}(\alpha_{\mathcal{J}} - \alpha^o_{\mathcal{J}}) + y^o - y\|_2 \\
\ge\ & \|\Phi_{\mathcal{J}}\alpha^o_{\mathcal{J}} - y^o\|_2 - \frac{\sqrt{|\mathcal{J}|}\mu\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_1}{\sqrt{1-\mu(|\mathcal{J}|-1)}} - \epsilon.
\end{aligned}$$

Lastly, by Lemma 1(a),

$$\begin{aligned}
\|\Phi_{\mathcal{J}}\alpha^o_{\mathcal{J}} - y^o\|_2^2 &= \|\Phi_{\mathcal{J}^o\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_2^2 \\
&\ge (1 - \mu(|\mathcal{J}^o \setminus \mathcal{J}| - 1))\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_2^2.
\end{aligned}$$

Thus, $\|\Phi_{\mathcal{J}}\alpha_{\mathcal{J}} - y\|_2 > \delta^{\rm OGA}$ whenever

$$\sqrt{1-\mu(|\mathcal{J}^o \setminus \mathcal{J}| - 1)}\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_2 - \frac{\sqrt{|\mathcal{J}|}\mu\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_1}{\sqrt{1-\mu(|\mathcal{J}|-1)}} - \epsilon > \delta^{\rm OGA},$$

which is equivalent to (33). ∎

Theorem 3(a) is a generalization of [12, Theorem 5.1(b)] in the special case of $\mathcal{J} = \mathcal{J}^o$. The proof of Theorem 3(b) amounts to showing that $|\phi_j^T r|$ is small for all $j \in \mathcal{J}_c^o$ (see (37)) and is large for some $j \in \mathcal{J}^o \setminus \mathcal{J}$ (see (39)). In [12, Theorem 5.1(a)], it is shown that $\mathcal{J}^{\rm OGA} = \mathcal{J}^o$ whenever $\delta^{\rm OGA} = \epsilon$, $\mu(2N-1) < 1$, and (7) holds. The following lemma shows that this result can be improved by using Theorem 3(b).

Lemma 2: Assume $2(1+\mu)\mu(N-1) < 1$. Then $\mu(N-1) < 1$ and (32) is implied by either $N = 1$ or

$$N \ge 2 \quad \text{and} \quad \alpha^o_{\min} > \frac{2\epsilon}{1 - 2\mu(1+\mu)(N-1)}, \tag{40}$$

and (33) is implied by either $N = 1$ or

$$N \ge 2 \quad \text{and} \quad \alpha^o_{\min} > \frac{1-\mu(N-2)+\mu\sqrt{N-1}\sqrt{1-\mu(N-2)}}{(1+\mu)(1+\mu-\mu N)}(\epsilon + \delta^{\rm OGA}). \tag{41}$$

Proof: Clearly $\mu(N-1) \le 2(1+\mu)\mu(N-1) < 1$. If $N = 1$, then (32) and (33) hold trivially. Suppose $N \ge 2$.

Since $\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_1 \le (N - |\mathcal{J}|)\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_\infty$ for all nonempty $\mathcal{J} \subset \mathcal{J}^o$, (32) is implied by

$$\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_\infty > 2\epsilon + 2(\mu+\mu^2)\frac{N-|\mathcal{J}|}{1-\mu(|\mathcal{J}|-1)}\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_\infty$$

for all nonempty $\mathcal{J} \subset \mathcal{J}^o$. This in turn is implied by

$$\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_\infty > 2\epsilon + 2(\mu+\mu^2)\left(\max_{1\le t\le N-1}\frac{N-t}{1-\mu(t-1)}\right)\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_\infty.$$

It is straightforward to verify that $\frac{d}{dt}\frac{N-t}{1-\mu(t-1)} = \frac{\mu(N-1)-1}{(1-\mu(t-1))^2} < 0$ for $t \in [1, N-1]$, so that the maximum is attained at $t = 1$. Rearranging terms and using $2(\mu+\mu^2)(N-1) < 1$ yield

$$\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_\infty > \frac{2\epsilon}{1-2(\mu+\mu^2)(N-1)}.$$

Since $\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_\infty \ge \alpha^o_{\min}$, this is implied by (40).

Since $\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_1 \le \sqrt{N-|\mathcal{J}|}\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_2$ for all nonempty $\mathcal{J} \subset \mathcal{J}^o$, (33) is implied by

$$\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_2 > \frac{\frac{\mu\sqrt{|\mathcal{J}|}}{\sqrt{1-\mu(|\mathcal{J}|-1)}}\sqrt{N-|\mathcal{J}|}\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_2 + \epsilon + \delta^{\rm OGA}}{\sqrt{1-\mu(N-|\mathcal{J}|-1)}} \quad \forall \emptyset \ne \mathcal{J} \subset \mathcal{J}^o.$$

Rearranging terms and using $\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_2 \ge \sqrt{N-|\mathcal{J}|}\alpha^o_{\min}$, this in turn is implied by

$$\alpha^o_{\min}\min_{1\le t\le N-1}\sqrt{N-t}\left(\sqrt{1-\mu(N-t-1)} - \frac{\mu\sqrt{t}\sqrt{N-t}}{\sqrt{1-\mu(t-1)}}\right) > \epsilon + \delta^{\rm OGA}.$$

We now simplify the expression inside the minimization. By making the substitution $\tau = N - t$ and bringing it under a common denominator, this expression can be written as

$$\sqrt{\tau}\frac{\sqrt{1+\mu-\mu\tau}\sqrt{1+\mu-\mu N+\mu\tau} - \mu\sqrt{N-\tau}\sqrt{\tau}}{\sqrt{1+\mu-\mu N+\mu\tau}},$$

with $1 \le \tau \le N-1$. Multiplying top and bottom by the conjugate of the numerator and simplifying, we obtain

$$\frac{\sqrt{\tau}(1+\mu)(1+\mu-\mu N)}{\sqrt{1+\mu-\mu\tau}\sqrt{1+\mu-\mu N+\mu\tau} + \mu\sqrt{N-\tau}\sqrt{\tau}} \cdot \frac{1}{\sqrt{1+\mu-\mu N+\mu\tau}},$$

which can be written as

$$\frac{(1+\mu)(1+\mu-\mu N)}{\sqrt{C_1(\tau)} + \mu\sqrt{C_2(\tau)}}, \tag{42}$$

where

$$C_1(\tau) := \left(\frac{1+\mu}{\tau} - \mu\right)(1+\mu-\mu N+\mu\tau)^2,$$
$$C_2(\tau) := (N-\tau)(1+\mu-\mu N+\mu\tau).$$

We have

$$C_1'(\tau) = -\left(\frac{1+\mu}{\tau^2}(1+\mu-\mu N-\mu\tau)+2\mu^2\right)(1+\mu-\mu N+\mu\tau),$$

which is negative for $\tau \le N-1$ (since $1+\mu-\mu N-\mu\tau \ge 1-\mu(2N-2) > 0$). Similarly,

$$C_2'(\tau) = -(1-\mu(2N-1-2\tau)),$$

which is negative for $\tau \ge 1$. Thus both $C_1(\tau)$ and $C_2(\tau)$ are decreasing with τ on $1 \le \tau \le N-1$, so $\sqrt{C_1(\tau)} + \mu\sqrt{C_2(\tau)}$ attains its maximum at $\tau = 1$, so its maximum value is $(1-\mu(N-2)) + \mu\sqrt{(N-1)(1-\mu(N-2))}$. Plugging this into (42) yields (41). ■

When $\mu(2N-1) < 1$, we have $2(1+\mu)(N-1) = 2N-1-(1-\mu(2N-2)) < 2N-1$, so the condition (40) is a relaxation of (7). Also, in the special case of $\delta^{\text{OGA}} = \epsilon$, the condition (41) is implied by (40). This is because $-(N-2)+\sqrt{N-1} \le 1$ for all $N \ge 2$, so the right-hand side of (41) is at most $2\epsilon/(1+\mu-\mu N)$, which in turn is less than the right-hand side of (40). This yields the following corollary of Theorem 3 and Lemma 2, which relaxes the condition $\delta^{\text{OGA}} = \epsilon$ and sharpens [12, Theorem 5.1(a)] in the case of $\delta^{\text{OGA}} = \epsilon$.

Corollary 1: Assume $2(1+\mu)\mu(N-1) < 1$ and $\delta^{\text{OGA}} \ge \epsilon$. If either $N = 1$ or (8) holds, then $\mathcal{J}^{\text{OGA}} = \mathcal{J}^{\circ}$, where $\mathcal{J}^{\text{OGA}}$ is an output of the OGA.

V. Sparse Representation from Stepwise Projection

The OGA uses the residual correlation term $|\phi_j^T r|$ to estimate the reduction in the residual when ϕ_j is added to the representation and chooses j with the largest estimated reduction. We can instead look ahead to add a ϕ_j that yields the largest actual reduction in the residual. This is more costly, but promises to yield a larger residual reduction. The resulting algorithm is known as the stepwise projection algorithm (SPA); see [1].

SPA:

0. Input $\delta^{\text{SPA}} > 0$. Initialize $\mathcal{J} \leftarrow \emptyset$. Go to Step 1.
1. Compute

$$v_j = \min_{\alpha_{\mathcal{J}} \in \Re^{|\mathcal{J}|}, \theta \in \Re} \|\Phi_{\mathcal{J}}\alpha_{\mathcal{J}} + \phi_j\theta - y\|_2^2 \quad \forall j \in \mathcal{J}_c, \tag{43}$$

choose

$$\bar{j} \in \arg\min_{j \in \mathcal{J}_c} v_j,$$

and update $\mathcal{J} \leftarrow \mathcal{J} \cup \{\bar{j}\}$. If $\sqrt{v_{\bar{j}}} \le \delta^{\text{SPA}}$, then output $\mathcal{J}$ and the corresponding solution $\alpha_{\mathcal{J}}$ of (31); otherwise return to Step 1.

Notice that during the first step when $\mathcal{J} = \emptyset$, we have $v_j = \|\phi_j\phi_j^T y - y\|_2^2 = \|y\|_2^2 - |\phi_j^T y|^2$ for all j. Thus, the SPA would initially choose j to maximize $|\phi_j^T y|$, the same as the OGA. However, in subsequent steps, the choices could differ. The work per step increases with $|\mathcal{J}|$ in the SPA. A possibly more efficient variant would be to switch back to OGA steps when $|\mathcal{J}|$ exceeds some threshold.

The following theorem shows that the SPA has similar sparsest support identification properties as the OGA. Its proof differs from the proof of Theorem 3 in that we work directly with the least square function instead of its gradient.

Theorem 4: Assume $\mu(N-1) < 1$. If $\delta^{\text{SPA}} \ge \epsilon$ and

$$\|\alpha^o_{\mathcal{J}^o \setminus \mathcal{J}}\|_\infty > 2\left(\mu\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_1 + \mu D_{\mathcal{J}} + \epsilon\right) \quad \forall \emptyset \ne \mathcal{J} \subset \mathcal{J}^{\circ}, \tag{44}$$

then $\mathcal{J}^{\text{SPA}} \subseteq \mathcal{J}^{\circ}$, where $\mathcal{J}^{\text{SPA}}$ is an output of the SPA and

$$D_{\mathcal{J}} := \sqrt{|\mathcal{J}|}\left(\frac{\mu\sqrt{|\mathcal{J}|+1}\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_1}{1-\mu|\mathcal{J}|} + \frac{\epsilon}{\sqrt{1-\mu|\mathcal{J}|}}\right).$$

If in addition (33) holds with δ^{OGA} replaced by δ^{SPA}, then $\mathcal{J}^{\text{SPA}} = \mathcal{J}^{\circ}$.

Proof: Assume $\delta^{\text{SPA}} \ge \epsilon$ and (44) holds. First we prove by induction that $\mathcal{J} \subseteq \mathcal{J}^{\circ}$ at Step 1 of the SPA. This is true when initially $\mathcal{J} = \emptyset$. Suppose this is true at the beginning of Step 1 with $\mathcal{J} \subset \mathcal{J}^{\circ}$. We show that it remains true at the end of Step 1.

Fix any $j \in \mathcal{J}_c$. Let $\hat{\mathcal{J}} = \mathcal{J} \cup \{j\}$, $\hat{\mathcal{J}}^{\circ} = \mathcal{J}^{\circ} \cup \{j\}$, and let $\alpha_{\hat{\mathcal{J}}}$ be any solution of

$$\min_{\alpha_{\hat{\mathcal{J}}} \in \Re^{|\hat{\mathcal{J}}|}} \|\Phi_{\hat{\mathcal{J}}}\alpha_{\hat{\mathcal{J}}} - y\|_2.$$

Then $\mu(|\hat{\mathcal{J}}|-1) \le \mu(N-1) < 1$ and $\Phi_{\hat{\mathcal{J}}^o}\alpha^o_{\hat{\mathcal{J}}^o} = y^o$. Thus, arguing as in the proof of Theorem 3(a), with $\hat{\mathcal{J}}, \hat{\mathcal{J}}^{\circ}$ replacing $\mathcal{J}, \mathcal{J}^{\circ}$, we obtain that

$$\|\alpha_{\hat{\mathcal{J}}} - \alpha^o_{\hat{\mathcal{J}}}\|_2 \le \frac{\mu\sqrt{|\hat{\mathcal{J}}|}\|\alpha^o_{\hat{\mathcal{J}}^0\setminus\hat{\mathcal{J}}}\|_1}{1-\mu(|\hat{\mathcal{J}}|-1)} + \frac{\epsilon}{\sqrt{1-\mu(|\hat{\mathcal{J}}|-1)}}.$$

Since $\|\alpha_{\hat{\mathcal{J}}} - \alpha^o_{\hat{\mathcal{J}}}\|_2 \ge \|\alpha_{\mathcal{J}} - \alpha^o_{\mathcal{J}}\|_2 \ge \frac{1}{\sqrt{|\mathcal{J}|}}\|\alpha_{\mathcal{J}} - \alpha^o_{\mathcal{J}}\|_1$, $|\hat{\mathcal{J}}| = |\mathcal{J}|+1$, and $\hat{\mathcal{J}}^0 \setminus \hat{\mathcal{J}} \subseteq \mathcal{J}^0 \setminus \mathcal{J}$, this and the definition of $D_{\mathcal{J}}$ implies

$$\|\alpha_{\mathcal{J}} - \alpha^o_{\mathcal{J}}\|_1 \le D_{\mathcal{J}}. \tag{45}$$

Thus, the minimization (43) is unchanged if we add (45) as a constraint, i.e.,

$$v_j = \min_{\substack{(\alpha_{\mathcal{J}},\theta) \in \Re^{|\mathcal{J}|+1} \\ \|\alpha_{\mathcal{J}} - \alpha^o_{\mathcal{J}}\|_1 \le D_{\mathcal{J}}}} R_j(\alpha_{\mathcal{J}}, \theta) \quad \forall j \in \mathcal{J}_c, \tag{46}$$

where for simplicity we let

$$R_j(\alpha_{\mathcal{J}}, \theta) := \|\Phi_{\mathcal{J}}\alpha_{\mathcal{J}} + \phi_j\theta - y\|_2^2.$$

Let $z := y - y^o$. For any $j \in \mathcal{J}_c$, we have from $\Phi_{\mathcal{J}^o}\alpha^o_{\mathcal{J}^o} = y^o$ that

$$\begin{aligned} & R_j(\alpha_{\mathcal{J}}, \theta) \\ = \; & \|\phi_j\theta + \Phi_{\mathcal{J}}\alpha_{\mathcal{J}} - \Phi_{\mathcal{J}^o}\alpha^o_{\mathcal{J}^o} - z\|_2^2 \\ = \; & \|\phi_j\|_2^2\theta^2 + 2\theta\phi_j^T\left(\Phi_{\mathcal{J}}\alpha_{\mathcal{J}} - \Phi_{\mathcal{J}^o}\alpha^o_{\mathcal{J}^o} - z\right) + R(\alpha_{\mathcal{J}}) \\ = \; & \theta^2 + 2\theta C_j(\alpha_{\mathcal{J}}) + R(\alpha_{\mathcal{J}}), \end{aligned}$$

where $R(\alpha_{\mathcal{J}}) := \|\Phi_{\mathcal{J}}\alpha_{\mathcal{J}} - \Phi_{\mathcal{J}^o}\alpha^o_{\mathcal{J}^o} - z\|_2^2$ is a function of $\alpha_{\mathcal{J}}$ only, independent of j, and

$$\begin{aligned} C_j(\alpha_{\mathcal{J}}) &:= \phi_j^T\left(\Phi_{\mathcal{J}}\alpha_{\mathcal{J}} - \Phi_{\mathcal{J}^o}\alpha^o_{\mathcal{J}^o} - z\right) \\ &= \phi_j^T\left(\Phi_{\mathcal{J}}(\alpha_{\mathcal{J}} - \alpha^o_{\mathcal{J}}) - \Phi_{\mathcal{J}^o\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}} - z\right). \end{aligned}$$

Suppose (45) holds. Then

$$\begin{aligned} \left|\phi_j^T\left(\Phi_{\mathcal{J}}(\alpha_{\mathcal{J}} - \alpha^o_{\mathcal{J}}) - z\right)\right| &\leq |\phi_j^T\Phi_{\mathcal{J}}(\alpha_{\mathcal{J}} - \alpha^o_{\mathcal{J}})| + |\phi_j^T z| \\ &\leq \mu\|\alpha_{\mathcal{J}} - \alpha^o_{\mathcal{J}}\|_1 + \|\phi_j\|_2\|z\|_2 \\ &\leq \mu D_{\mathcal{J}} + \epsilon. \end{aligned}$$

If $j \in \mathcal{J}_c^\circ$, then

$$|\phi_j^T\Phi_{\mathcal{J}^o\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}| \leq \mu\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_1,$$

so the previous two relations yield

$$|C_j(\alpha_{\mathcal{J}})| \leq \mu\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_1 + \mu D_{\mathcal{J}} + \epsilon.$$

If $j \in \mathcal{J}^\circ \setminus \mathcal{J}$, then (38) implies

$$|\phi_j^T\Phi_{\mathcal{J}^o\setminus\mathcal{J}}\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}| \geq |\alpha^o_j| - \mu\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_1,$$

so the earlier two relations yield

$$|C_j(\alpha_{\mathcal{J}})| \geq |\alpha^o_j| - \mu\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_1 - \mu D_{\mathcal{J}} - \epsilon.$$

Thus, if (45) holds and $\theta \neq 0$, then for any $j \in \mathcal{J}^\circ \setminus \mathcal{J}$ satisfying $|\alpha^o_j| = \|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_\infty$ and for every $j' \in \mathcal{J}_c^\circ$, we have

$$\begin{aligned} &\min\{R_j(\alpha_{\mathcal{J}},\theta), R_j(\alpha_{\mathcal{J}},-\theta)\} \\ =\;& \theta^2 - 2|\theta||C_j(\alpha_{\mathcal{J}})| + R(\alpha_{\mathcal{J}}) \\ \leq\;& \theta^2 - 2|\theta|\left(|\alpha^o_j| - \mu\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_1 - \mu D_{\mathcal{J}} - \epsilon\right) + R(\alpha_{\mathcal{J}}) \\ <\;& \theta^2 - 2|\theta|\left(\mu\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_1 + \mu D_{\mathcal{J}} + \epsilon\right) + R(\alpha_{\mathcal{J}}) \\ \leq\;& \theta^2 - 2|\theta||C_{j'}(\alpha_{\mathcal{J}})| + R(\alpha_{\mathcal{J}}) \\ \leq\;& \theta^2 + 2\theta C_{j'}(\alpha_{\mathcal{J}}) + R(\alpha_{\mathcal{J}}) = R_{j'}(\alpha_{\mathcal{J}},\theta), \end{aligned}$$

where the strict inequality uses (44) and $\theta \neq 0$. This and (46) imply that

$$\min_{j\in\mathcal{J}^\circ\setminus\mathcal{J}} \upsilon_j < \min_{j\in\mathcal{J}_c^\circ} \upsilon_j.$$

Hence in Step 1 we choose $\bar{j} \in \mathcal{J}^\circ\setminus\mathcal{J}$, so that $\mathcal{J}\cup\{\bar{j}\} \subseteq \mathcal{J}^\circ$.

In case of $|\mathcal{J}| = N - 1$, we have $\mathcal{J}\cup\{\bar{j}\} = \mathcal{J}^\circ$ so that $\sqrt{\upsilon_{\bar{j}}} \leq \|\Phi_{\mathcal{J}}\alpha^o_{\mathcal{J}} - y\|_2 = \|y^o - y\|_2 = \epsilon$. Since $\delta^{\text{SPA}} \geq \epsilon$, this implies $\sqrt{\upsilon_{\bar{j}}} \leq \delta^{\text{SPA}}$, so Step 1 would output $\mathcal{J}^\circ$ and $\alpha_{\mathcal{J}^\circ}$.

If in addition (33) holds with δ^{OGA} replaced by δ^{SPA}, then the same argument as in the proof of Theorem 3 shows that $\mathcal{J}^{\text{SPA}} = \mathcal{J}^\circ$ always. ■

Condition (44) is comparable to (32) and can be similarly simplified as in Lemma 2 to obtain the following identification result analogous to Corollary 1.

Corollary 2: Assume $\frac{2\mu+\mu^2}{1-\mu}(N-1) < 1$ and $\delta^{\text{SPA}} \geq \epsilon$. If either $N = 1$ or

$$\begin{aligned} &N \geq 2 \quad \text{and} \quad \alpha^o_{\min} \geq \\ &\max\left\{\frac{2\epsilon\left(1+\mu\sqrt{\frac{N-1}{1-\mu(N-1)}}\right)}{1-\frac{2\mu+\mu^2}{1-\mu}(N-1)}, \frac{\epsilon+\delta^{\text{OGA}}}{1-\mu(N-1)}\right\}, \end{aligned} \quad (47)$$

then $\mathcal{J}^{\text{SPA}} = \mathcal{J}^\circ$, where $\mathcal{J}^{\text{SPA}}$ is an output of the SPA.

Proof: By following the proof of (40), we see that (44) is implied by

$$\begin{aligned} &\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_\infty \\ >\;& 2\epsilon\left(1+\mu\max_{1\leq t\leq N-1}\frac{\sqrt{t}}{\sqrt{1-\mu t}}\right) \\ +\;& 2\mu\max_{1\leq t\leq N-1}\left(1+\frac{\mu\sqrt{t}\sqrt{t+1}}{1-\mu t}\right)(N-t)\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_\infty. \end{aligned}$$

The argument of the first maximization is increasing with t, so the maximum is attained at $t = N-1$. By using $\sqrt{t}\sqrt{t+1} - t = \sqrt{t}/(\sqrt{t+1}+\sqrt{t}) < \frac{1}{2}$ to bound the argument of the second maximization, we see that this in turn is implied by

$$\begin{aligned} \|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_\infty \geq\;& 2\epsilon\left(1+\frac{\mu\sqrt{N-1}}{\sqrt{1-\mu(N-1)}}\right) \\ &+(2\mu+\mu^2)\max_{1\leq t\leq N-1}\frac{N-t}{1-\mu t}\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_\infty. \end{aligned}$$

Since $\frac{N-t}{1-\mu t}$ is decreasing over $t \in [1, N-1]$, the maximum is attained at $t = 1$. Rearranging terms and using $\frac{2\mu+\mu^2}{1-\mu}(N-1) < 1$, $\|\alpha^o_{\mathcal{J}^o\setminus\mathcal{J}}\|_\infty \geq \alpha^o_{\min}$, we see that this is implied by (47). Also, as is shown in Lemma 2 and the ensuing discussion, (47) implies (33) with δ^{OGA} replaced by δ^{SPA}. ■

The sufficient condition for identifying $\mathcal{J}^\circ$ by SPA is slightly more stringent than that for OGA (compare (8) and (47)). This is because the minimizing $\alpha_{\mathcal{J}}$ in (43) changes with j, and we have only a uniform bound $D_{\mathcal{J}}$ on this change (see (45)) which adds to our error estimate.

ACKNOWLEDGMENT

The author thanks the reviewers for their insightful comments and, in particular, to one reviewer for detailed suggestions and pointing out the references [1], [27].

REFERENCES

[1] A. R. Barron, A. Cohen, W. Dahmen, and R. A. DeVore, "Approximation and learning by greedy algorithms," *Ann. Statist.*, vol. 36, no. 1, pp. 64-94, 2008.
[2] A. Antoniadis and J. Fan, "Regularization of Wavelet approximations," *J. Amer. Statist. Assoc.*, vol. 96, no. 455, pp. 939-967, 2001.
[3] E. J. Candès, J. Romberg and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Inf. Theory*, vol. 52, no. 2, pp. 489-509, Feb. 2006.
[4] E. J. Candès, J. Romberg and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, pp. 1207-1223, 2006.
[5] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Inf. Theory*, vol. 51, no. 12, pp. 4203-4215, Dec. 2005.
[6] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33-61, 1998.
[7] G. Davis, S. Mallat, and Z. Zhang, "Adaptive time-frequency decompositions," *Opt. Eng.*, vol. 33, no. 7, pp. 2183-2191, 1994.
[8] R. A. DeVore and V. N. Temlyakov, "Some remarks on greedy algorithms," *Adv. Comput. Math.*, vol. 5, pp. 173-187, 1996.
[9] D. L. Donoho, "For most large underdetermined systems of linear equations, the minimal ℓ^1-norm near-solution approximates the sparsest near-solution," Department of Statistics, Stanford University, Stanford, CA, Tech. Rep., 2006.
[10] D. L. Donoho, "For most large underdetermined systems of linear equations, the minimal ℓ^1-norm solution is also the sparsest solution," *Comm. Pure Appl. Math.*, vol. 59, no. 6, pp. 797-929, 2006.
[11] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, pp. 2197-2202, 2003.

[12] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6-18, Jan. 2006.
[13] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845-2862, Nov. 2001.
[14] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," Department of Statistics, Stanford University, Stanford, CA, Tech. Rep., 2006.
[15] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2558-2567, Sep. 2002.
[16] M. P. Friedlander and P. Tseng, "Exact regularization of convex programs," *SIAM J. Optim.*, vol. 18, no. 4, pp. 1326-1350, Oct. 2007.
[17] J.-J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1341-1344, Jun. 2004.
[18] J.-J. Fuchs, "Recovery of exact sparse representations in the presence of bounded noise," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3601-3608, Oct. 2005.
[19] A. C. Gilbert, S. Muthukrishnan, S., and M. J. Strauss, "Approximation of functions over redundant dictionaries using coherence," in *Proc. 14th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2003, pp. 243-252.
[20] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: The Johns Hopkins University Press, 1985.
[21] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600-616, Mar. 1997.
[22] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320-3325, Dec. 2003.
[23] S. Mallat and Z. Zhang, "Matching pursuit in a time-frequency dictionary," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397-3415, Dec. 1993.
[24] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Ann. Asilomar Conf. Signals, Systems and Computers*, Nov. 1993, pp. 40-44.
[25] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton University Press, 1970.
[26] S. Sardy, A. Bruce, and P. Tseng, "Block coordinate relaxation methods for nonparametric Wavelet denoising," *J. Comput. Graph. Statist.*, vol. 9, no. 2, pp. 361-379, 2000.
[27] T. Strohmer, and R. Heath Jr., "Grassmannian frames with applications to coding and communications," *Appl. Comp. Harm. Anal.*, vol. 14, no. 3, pp. 257-275, May 2003.
[28] V. N. Temlyakov, "Weak greedy algorithms," *Adv. Comput. Math.*, vol. 12, pp. 213-227, 2000.
[29] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. Ser. B*, vol. 58, no. 1, pp. 267-288, 1996.
[30] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231-2242, Oct. 2004.
[31] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030-1051, Mar. 2006.
[32] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.
[33] P. Tseng and Z.-Q. Luo, "On the convergence of the affine-scaling algorithm," *Math. Program.*, vol. 56, no. 3, pp. 301-319, 1992.

P Tseng received his B.Sc. from Queens University in 1981 and his Ph.D. from MIT in 1986. After spending one year at the Management Science Division in the University of British Columbia and three years at LIDS in MIT, he joined the University of Washington in 1990, where he remained since. His research area is mainly in optimization. He likes to optimize his travels by bicycle or kayak.