RIP-Based Near-Oracle Performance Guarantees for SP, CoSaMP, and IHT

Raja Giryes and Michael Elad

***Abstract*—This correspondence presents an average case denoising performance analysis for SP, CoSaMP, and IHT algorithms. This analysis considers the recovery of a noisy signal, with the assumptions that it is corrupted by an additive random zero-mean white Gaussian noise and has a K-sparse representation with respect to a known dictionary $\mathbf{D}$. The proposed analysis is based on the RIP, establishing a near-oracle performance guarantee for each of these algorithms. Beyond bounds for the reconstruction error that hold with high probability, in this work we also provide a bound for the average error.**

***Index Terms*—Additive white noise, compressed sensing, Gaussian noise, signal denoising, signal reconstruction, signal representations.**

I. Introduction

The area of sparse approximation is an emerging field that has received much attention in the last decade. In one of the most basic problems posed in this field, we consider a noisy measurement vector $\mathbf{y} \in \mathbb{R}^m$ of the form $\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{e}$, where $\mathbf{x} \in \mathbb{R}^N$ is the signal's representation with respect to the dictionary $\mathbf{D} \in \mathbb{R}^{m\times N}$, where $N \geq m$. The vector $\mathbf{e} \in \mathbb{R}^m$ is an additive noise, assumed to be either an adversarial disturbance, or a random white Gaussian noise with zero mean and variance σ^2.[1] We further assume that the columns of $\mathbf{D}$ are normalized, and that the representation vector $\mathbf{x}$ is K-sparse, or nearly so.[2] Our goal is to find the K-sparse vector $\mathbf{x}$ that approximates the true representation of the measured signal $\mathbf{y}$. Solving this problem directly is quite hard and problematic [2]. For this reason, approximation algorithms were proposed—these are often referred to as pursuit algorithms. We measure the quality of the approximate solution $\hat{\mathbf{x}}$ by the mean-square error (MSE)

$$\text{MSE}(\hat{\mathbf{x}}) = E\left\|\mathbf{x} - \hat{\mathbf{x}}\right\|_2^2 \tag{I.1}$$

where the expectation is taken over the distribution of the noise ($\mathbf{x}$ is assumed to be deterministic). Naturally, we desire pursuit algorithms that are guaranteed to lead to as small as possible reconstruction error.

When analyzing the performance of pursuit algorithms, two features that characterize the dictionary $\mathbf{D}$ are often used. The first is the mutual-coherence μ of a matrix $\mathbf{D}$—the largest absolute normalized inner product between different columns from $\mathbf{D}$. The second is the RIP—it is said that $\mathbf{D}$ satisfies the K-RIP condition with parameter δ_K if it is the smallest value that satisfies

$$(1-\delta_K)\left\|\mathbf{x}\right\|_2^2 \leq \left\|\mathbf{D}\mathbf{x}\right\|_2^2 \leq (1+\delta_K)\left\|\mathbf{x}\right\|_2^2 \tag{I.2}$$

for any K-sparse vector $\mathbf{x}$ [3], [4].

Turning to pursuit algorithms, one popular approach is based on ℓ_1 relaxation and known as basis pursuit (BP) [5]. Another ℓ_1-based relaxation algorithm is the Dantzig selector (DS), as proposed in [6]. For the case of an adversarial noise these techniques satisfy bounds on the reconstruction error in the form of a constant factor $(\text{Const} \geq 2)$ multiplying the noise power,

$$\left\|\mathbf{x} - \hat{\mathbf{x}}\right\|_2^2 \leq \text{Const} \cdot \left\|\mathbf{e}\right\|_2^2. \tag{I.3}$$

One such example is the work by Candès and Tao, reported in [4], which analyzed the BP error. This work has shown that if the dictionary $\mathbf{D}$ satisfies $\delta_K + \delta_{2K} + \delta_{3K} < 1$, then the BP MSE is bounded as in (I.3). This condition has been improved several times. As far as we know, the tightest known condition is $\delta_{2K} \leq \frac{3}{4+\sqrt{6}} \approx 0.4652$ [7].

A different pursuit approach is the greedy strategy [8], leading to algorithms such as thresholding and orthogonal matching pursuit (OMP). Unlike the BP, these algorithms were shown to be more sensitive, incapable of providing a uniform guarantee for the reconstruction when m is proportional to K [9].

The last family of pursuit methods we mention here are the greedy-like algorithms. As opposed to the greedy strategies, these algorithms enable removal of elements from the detected support. Algorithms belonging to this group are regularized OMP (ROMP) [10], compressive sampling matching pursuit (CoSaMP) [11], subspace-pursuit (SP) [12], and iterative hard thresholding (IHT) [13].

Interestingly, unlike the greedy methods, the greedy-like approach was found to be closer in spirit to BP and DS, in the sense that it leads to uniform guarantees on the bounded MSE. ROMP was the first of these algorithms to be analyzed [10], leading to the more strict requirement $\delta_{2K} < \frac{0.03}{\sqrt{\log K}}$. CoSaMP [11] and SP [12], which came later, have similar RIP conditions without the $\log K$ factor. IHT was also shown to have a uniform guarantee for bounded error of the same flavor [13]. This correspondence focuses on this specific family of methods, as it poses an interesting compromise between the simplicity of the greedy methods and the proven strength of the relaxed algorithms.

Until now, our discussion dealt with the performance of the different pursuit methods in the presence of an adversarial noise. The guarantees in this case are of the form of (I.3) with a constant larger than 2. These results guarantee stability of the different algorithms but no effective denoising is to be expected. To obtain better results, one must change the perspective and consider the noise to be random, rather than adversarial.

We first consider an oracle estimator that knows the support of $\mathbf{x}$, i.e., the locations of the K nonzeros in this vector. The oracle estimator is easily given by $\hat{\mathbf{x}}_{\text{Oracle}} = \mathbf{D}_T^{\dagger}\mathbf{y}$, where T is the support of $\mathbf{x}$ and $\mathbf{D}_T$ is a submatrix of $\mathbf{D}$ that contains only the columns involved in the support T. In the case of a zero-mean white Gaussian noise with variance σ^2, the oracle's MSE is given by [6]

$$\text{MSE}(\hat{\mathbf{x}}_{\text{Oracle}}) = \text{trace}\left\{(\mathbf{D}_T^*\mathbf{D}_T)^{-1}\right\}\sigma^2 \geq \frac{K\sigma^2}{1+\delta_K}. \tag{I.4}$$

This is the smallest possible error, and it is proportional to the number of nonzeros K multiplied by σ^2. It is natural to ask how close do we get to this best error by practical pursuit methods that do not assume the knowledge of the support. The first to answer such a question were Candès and Tao in the analysis of the DS algorithm [6]. Requiring $\delta_{2K} + \delta_{3K} < 1$, the reconstruction result of DS, $\hat{\mathbf{x}}_{\text{DS}}$, was shown to obey

$$\left\|\mathbf{x} - \hat{\mathbf{x}}_{\text{DS}}\right\|_2^2 \leq 2C_{\text{DS}}^2(1+a)\log N \cdot K\sigma^2, \tag{I.5}$$

Manuscript received January 18, 2011; revised July 08, 2011 and October 09, 2011; accepted October 12, 2011. Date of publication November 07, 2011; date of current version February 10, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Philippe Ciblat. This research was supported by the European Community FP7-FET program, SMALL project, under grant agreement 225913. The work of R. Giryes is supported in part by the Azrieli Foundation through an Azrieli Fellowship.

The authors are with the Department of Computer Science, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: raja@cs.technion.ac.il; elad@cs.technion.ac.il).

Digital Object Identifier 10.1109/TSP.2011.2174985

[1]More details about the difference between adversarial and random noise can be found in [1].

[2]Has K-dominant elements.

with probability exceeding $1-(\sqrt{\pi(1+a)\log N}\cdot N^a)^{-1}$, where $C_{\mathrm{DS}}=\frac{4}{(1-2\delta_{3K})}$.[3] Up to a constant and a $\log N$ factor, this bound is the same as the oracle's one in (I.4). The $\log N$ factor in (I.5) in unavoidable, as proven in [1] and therefore this bound is optimal up to a constant factor. We note that here and in other existing work on sparsity based recovery algorithms, the derived bounds with high-probability are compared to the oracle's expected error, as given in (I.4).

A similar bound for BP was presented in [14], showing that by requiring $\delta_{2K}+3\delta_{3K}<1$, the reconstruction result of BP satisfies the bound in (I.5) with probability exceeding $1-(N^a)^{-1}$ and with a different constant C_{BP}. This result is weaker than the one obtained for DS.

Mutual-Coherence based results for DS and BP were derived in [15], [16], but we shall not dwell on those results here—for more information on how they compare to the above, see [16].

We mentioned that the greedy-like algorithms, SP, CoSaMP, and IHT, enjoy a uniform recovery bound in the adversarial noise case. In this correspondence, we extend these results and present RIP-based near-oracle performance guarantees for these algorithms that resemble the ones obtained for DS and BP. We show that the proposed guarantees are valid also for the expectation of the error and not only with high probability. In Section II we develop, based on results from [11]–[13], the RIP-based near oracle performance bounds for the SP, CoSaMP and IHT techniques. We also derive an average bound from the given probabilistic bound that guarantees near-oracle performance for the expectation of the error. In Section III we extend the above results, by considering the nearly sparse case.

II. Near Oracle Performance of the Algorithms

We begin with a short description of the notation we use. The support of $\mathbf{x}$ is denoted by $\mathrm{suppa}(\mathbf{x})$ (a set with the locations of the nonzero elements of $\mathbf{x}$) and $\mathrm{supp}(\mathbf{x},K)$ is the support of the K largest magnitude elements in $\mathbf{x}$. Similar to $\mathbf{D}_T$, $\mathbf{x}_T$ is a vector composed of the entries of the vector $\mathbf{x}$ over the set T. T^C symbolizes the complementary set of T and $T-\tilde{T}$ is the set of all elements contained in T but not in $\tilde{T}$. The projection of a vector $\mathbf{y}$ on the subspace spanned by the columns of the matrix $\mathbf{A}$ (assumed to have more rows than columns) is $\mathbf{A}\mathbf{A}^{\dagger}\mathbf{y}$. The residual is $\mathbf{y}-\mathbf{A}\mathbf{A}^{\dagger}\mathbf{y}$. T denotes the set of the nonzero places of the original signal $\mathbf{x}$; As such, $|T|\le K$ when $\mathbf{x}$ is K-sparse. T_e is the subset of columns of size K in $\mathbf{D}$ that gives the maximum correlation with the noise vector $\mathbf{e}$, namely, $T_e=\mathrm{argmax}_{T\,|\,|T|=K}\|\mathbf{D}_T^*\mathbf{e}\|_2$.

Before we turn to derive the bounds for the three techniques, we start with a short description of them. SP [12] holds a temporal solution with K nonzero entries, and in each iteration it adds an additional set of K candidate nonzeros that are most correlated with the residual, and prunes this list back to K elements by choosing the dominant ones. CoSaMP [11], in a similar way to SP, holds a temporal solution with K nonzero entries, with the difference that in each iteration it adds an additional set of $2K$ (instead of K) candidate nonzeros that are most correlated with the residual. Another difference between the two algorithms is that after the pruning step in SP we use a matrix inversion in order to calculate a new projection for the K dominant elements, while in CoSaMP we just take the biggest K elements. A detailed description of these two algorithms can be found in [11], [12], and [17].

IHT [13] uses a different strategy than SP and CoSaMP—it applies only multiplications by $\mathbf{D}$ and $\mathbf{D}^*$, and a hard thresholding operator, $[\cdot]_K$, that takes the K-largest elements. The IHT iteration is simply

$$\hat{\mathbf{x}}_{\mathrm{IHT}}^{\ell}=\left[\hat{\mathbf{x}}_{\mathrm{IHT}}^{\ell-1}+\mathbf{D}^*(\mathbf{y}-\mathbf{D}\hat{\mathbf{x}}_{\mathrm{IHT}}^{\ell-1})\right]_K. \tag{II.1}$$

[3]In [6], a slightly different constant was presented.

For all three methods, different stopping criteria can be used as described in [11]–[13].

We turn now to derive the bounds. We first present bounds for the case where $\mathbf{e}$ is an adversarial noise using the same techniques used in [11]–[13]. In these works the reconstruction error was bounded by a constant times the noise power in the same form as in (I.3). We propose a bound that is a constant times $\|\mathbf{D}_{T_\mathbf{e}}^*\mathbf{e}\|_2$. Its proof is based on the proofs in [11]–[13]. The full development appears in [17]. Armed with the new bound, we change perspective and look at the case where $\mathbf{e}$ is a zero-mean white Gaussian noise vector with variance σ^2, and derive a near-oracle performance result of the same form as in (I.5), using tools developed in [6]. The following theorem presents bounds for SP, CoSaMP and IHT for the adversarial noise case. T^ℓ stands for the support found at the ℓth iteration, and $\hat{x}^\ell$ is the ℓth iteration result.

Theorem 2.1: For a K-sparse vector $\mathbf{x}$, under the condition $\delta_{bK}\le\delta$, SP solution at the ℓth iteration satisfies

$$\left\|\mathbf{x}_{T-T^\ell}\right\|_2\le 2^{-\ell}\|\mathbf{x}\|_2+2\cdot 8.22\|\mathbf{D}_{T_\mathbf{e}}^*\mathbf{e}\|_2, \tag{II.2}$$

and CoSaMP and IHT solutions at the ℓth iteration satisfy

$$\left\|\mathbf{x}-\hat{\mathbf{x}}^\ell\right\|_2\le 2^{-\ell}\|\mathbf{x}\|_2+(C-1)\|\mathbf{D}_{T_\mathbf{e}}^*\mathbf{e}\|_2. \tag{II.3}$$

In addition, after at most $\ell^*=\left\lceil\log_2\left(\frac{\|\mathbf{x}\|_2}{\left\|\mathbf{D}_{T_\mathbf{e}}^*\mathbf{e}\right\|_2}\right)\right\rceil$ iterations, the solution $\hat{\mathbf{x}}$ leads to an accuracy

$$\left\|\mathbf{x}-\hat{\mathbf{x}}^\ell\right\|_2\le C\|\mathbf{D}_{T_\mathbf{e}}^*\mathbf{e}\|_2 \tag{II.4}$$

where $b=3$, $\delta=0.139$ and $C_{SP}=2\cdot\frac{7-9\delta_{3K}+7\delta_{3K}^2-\delta_{3K}^3}{(1-\delta_{3K})^4}\le 21.41$ for SP; $b=4$, $\delta=0.1$ and $C_{\mathrm{CoSaMP}}=\frac{29-14\delta_{4K}+\delta_{4K}^2}{(1-\delta_{4K})^2}\le 34.1$ for CoSaMP; and $b=3$, $\delta=\frac{1}{\sqrt{32}}$ and $C_{\mathrm{IHT}}=9$ for IHT.

The bounds for CoSaMP and IHT in this theorem are similar to the bounds in [11, Theorem 4.1] and [13, Theorem 5] and have almost the same constants. The difference is that we replaced $\|\mathbf{e}\|_2^2$ with $\|\mathbf{D}_{T_\mathbf{e}}^*\mathbf{e}\|_2^2$ in the bounds. The exact details of the derivation appear in an extended version of this correspondence [17].

The corresponding result for SP appears in Theorem 9 in [12]. However, we cannot use directly the constant given there as we have done for CoSaMP and IHT. The result there states that $\|\mathbf{x}-\hat{\mathbf{x}}_{SP}\|_2\le C'_{SP}\|\mathbf{e}\|_2$, where $C'_{SP}=\frac{1+\delta_{3K}+\delta_{3K}^2}{\delta_{3K}(1-\delta_{3K})}$. The occurrence of δ_{3K} in the denominator makes this constant very large, and it goes to infinity for very small values of δ_{3K}. Thus, we use a variation of [12, Theorem 10] for obtaining the following:

Theorem 2.2: For a K-sparse vector $\mathbf{x}$, the SP solution at the ℓth iteration satisfies the recurrence inequality

$$\left\|\mathbf{x}_{T-T^\ell}\right\|_2\le\frac{2\delta_{3K}(1+\delta_{3K})}{(1-\delta_{3K})^3}\left\|\mathbf{x}_{T-T^{\ell-1}}\right\|_2+\frac{6-6\delta_{3K}+4\delta_{3K}^2}{(1-\delta_{3K})^3}\|\mathbf{D}_{T_\mathbf{e}}^*\mathbf{e}\|_2. \tag{II.5}$$

For $\delta_{3K}\le 0.139$, this leads to

$$\left\|\mathbf{x}_{T-T^\ell}\right\|_2\le 0.5\left\|\mathbf{x}_{T-T^{\ell-1}}\right\|_2+8.22\|\mathbf{D}_{T_\mathbf{e}}^*\mathbf{e}\|_2. \tag{II.6}$$

Using this theorem and elements from the proof techniques presented in [11] lead to the result in Theorem 2.1. The exact steps appear in the proof of [17, Corollary 3.2].

Now that we have a bound for SP, CoSaMP, and IHT for the adversarial case, we proceed and consider a bound for the random noise case, which leads to a near-oracle performance guarantee for these techniques.

TABLE I
NEAR ORACLE PERFORMANCE GUARANTEES FOR THE DS, BP, SP, COSAMP, AND IHT ALGORITHMS

Alg.	RIP Condition	Probability of Correctness	Constant	The Obtained Bound
DS	$\delta_{2K}+\delta_{3K}\le 1$	$1-(\sqrt{\pi(1+a)\log N}\cdot N^a)^{-1}$	$\frac{4}{1-2\delta_{3K}}$	$C_{DS}^2\cdot(2(1+a)\log N)\cdot K\sigma^2$
BP	$\delta_{2K}+3\delta_{3K}\le 1$	$1-(N^a)^{-1}$	$>\frac{32}{\kappa^4},(\kappa<1)$	$C_{BP}^2\cdot(2(1+a)\log N)\cdot K\sigma^2$
SP	$\delta_{3K}\le 0.139$	$1-(\sqrt{\pi(1+a)\log N}\cdot N^a)^{-1}$	≤ 21.41	$C_{SP}^2\cdot(2(1+a)\log N)\cdot K\sigma^2$
CoSaMP	$\delta_{4K}\le 0.1$	$1-(\sqrt{\pi(1+a)\log N}\cdot N^a)^{-1}$	≤ 34.2	$C_{\text{CoSaMP}}^2\cdot(2(1+a)\log N)\cdot K\sigma^2$
IHT	$\delta_{3K}\le\frac{1}{\sqrt{32}}$	$1-(\sqrt{\pi(1+a)\log N}\cdot N^a)^{-1}$	9	$C_{IHT}^2\cdot(2(1+a)\log N)\cdot K\sigma^2$

Theorem 2.3: For a K-sparse vector $\mathbf{x}$, assuming that $\mathbf{e}$ is a zero-mean white Gaussian noise vector with variance σ^2, if the condition $\delta_{bK}\le\delta$ holds, then with probability exceeding $1-(\sqrt{\pi(1+a)\log N}\cdot N^a)^{-1}$ we obtain after at most $\ell^*=\left\lceil\log_2\left(\frac{\|\mathbf{x}\|_2}{\left\|\mathbf{D}_{T_\mathbf{e}}^*\mathbf{e}\right\|_2}\right)\right\rceil$ iterations

$$\|\mathbf{x}-\hat{\mathbf{x}}\|_2^2\le 2C^2(1+a)\log N\cdot K\sigma^2 \tag{II.7}$$

where $b=3$ and $\delta=0.139$ for SP; $b=4$ and $\delta=0.1$ for CoSaMP; and $b=3$ and $\delta=\frac{1}{\sqrt{32}}$ for IHT, and C is the appropriate constant from Theorem 2.1.

Proof: Following [6, Sec. III], it holds true that

$$\mathbf{P}\left(\sup_i|\mathbf{D}_i^*\mathbf{e}|>\sigma\cdot\sqrt{2(1+a)\log N}\right)\le\left(\sqrt{\pi(1+a)\log N}\cdot N^a\right)^{-1}.$$

Combining this with (II.4) while bearing in mind that $|T_e|=K$ gives the above. □

The results obtained for the greedy-like techniques are similar to the ones posed for DS and BP, but with different constants. In Table I, we summarize the performance guarantees of these algorithms. A comparison between the methods' constants is made in [17]. However, note that the bounds are not tight, and thus these constants cannot truly predict which method behaves better. Also, we should emphasize that the SP and CoSaMP methods are just templates, and parameters such as K or 2K could be tuned and will affect the constants in the error bounds.

In [16], similar guarantees are presented for OMP and thresholding with better constants. However, these results hold under mutual-coherence based conditions, which are more restricting. Their validity relies on the magnitude of the entries of $\mathbf{x}$ and the noise power, which is not the case for the results presented in this section for the greedy-like methods.

The importance of the obtained bounds is in showing that using the greedy-like methods we can recover signals with an effective reduction of the additive noise. In the case of an adversarial noise, such a guarantee does not exist. Furthermore, the obtained results suggest that the reconstruction results' error behaves like the oracle's error up to a $\log N$ and a constant factor.

So far we have seen that with high probability the greedy-like algorithms achieve near oracle performance. It is interesting to ask whether we can derive a similar bound on the expected error. The next theorem shows that the answer to this question is positive.[4]

Theorem 2.4: For a K-sparse vector $\mathbf{x}$, assuming that $\mathbf{e}$ is a zero-mean white Gaussian noise vector with variance σ^2 and that, $N>3$[5] if the condition $\delta_{bK}\le\delta$ holds, then after at most $\ell^*=\left\lceil\log_2\left(\frac{\|\mathbf{x}\|_2}{\left\|\mathbf{D}_{T_\mathbf{e}}^*\mathbf{e}\right\|_2}\right)\right\rceil$ iterations

$$E\|\mathbf{x}-\hat{\mathbf{x}}\|_2^2\le 4C^2(1+a)\log N\cdot K\sigma^2 \tag{II.8}$$

where $b=3$ and $\delta=0.139$ for SP; $b=4$ and $\delta=0.1$ for CoSaMP; and $b=3$ and $\delta=\frac{1}{\sqrt{32}}$ for IHT. The constant C is the one from Theorem 2.1.

Proof: Utilizing simple rules of probability theory with the result of Theorem 2.3 as a first step and of Theorem 2.1 as a second step give

$$\begin{aligned}E\|\mathbf{x}-\hat{\mathbf{x}}\|_2^2 &\le \mathbf{P}\left(\sup_i|\mathbf{D}_i^*\mathbf{e}|<\sigma\sqrt{2(1+a)\log N}\right)\\&\quad\cdot 2C^2(1+a)\log N\cdot K\sigma^2\\&\quad+E_{\sup_i|\mathbf{D}_i^*\mathbf{e}|>\sigma\sqrt{2(1+a)\log N}}\|\mathbf{x}-\hat{\mathbf{x}}\|_2^2\\&\le 2C^2(1+a)\log N\cdot K\sigma^2\\&\quad+C^2E_{\sup_i|\mathbf{D}_i^*\mathbf{e}|>\sigma\sqrt{2(1+a)\log N}}\|D_T^*\mathbf{e}\|_2^2.\end{aligned} \tag{II.9}$$

The facts that the supremum in the last inequality is over N elements and that the support of T[6] is of size K leads to

$$\begin{aligned}&E_{\sup_i|\mathbf{D}_i^*\mathbf{e}|>\sigma\sqrt{2(1+a)\log N}}\|\mathbf{D}_T^*\mathbf{e}\|_2^2\\&\quad\le N\cdot E_{\substack{|\mathbf{D}_1^*\mathbf{e}|>\sigma\sqrt{2(1+a)\log N}\\|\mathbf{D}_i^*\mathbf{e}|<|\mathbf{D}_1^*\mathbf{e}|,2\le i\le K}}\|\mathbf{D}_T^*\mathbf{e}\|_2^2\\&\quad\le NK\cdot E_{|\mathbf{D}_1^*\mathbf{e}|>\sigma\sqrt{2(1+a)\log N}}(\mathbf{D}_1^*\mathbf{e})^2.\end{aligned} \tag{II.10}$$

Since the columns of $\mathbf{D}$ are normalized and $\mathbf{e}$ is a zero-mean white Gaussian noise with variance σ^2, we have that $\mathbf{D}_1^*\mathbf{e}\sim N(0,\sigma^2)$. Using the symmetry of the Gaussian distribution, we have that

$$\begin{aligned}&E_{|\mathbf{D}_1^*\mathbf{e}|>\sigma\cdot\sqrt{2(1+a)\log N}}(\mathbf{D}_1^*\mathbf{e})^2\\&=2\int_{\sigma\cdot\sqrt{2(1+a)\log N}}^{\infty}\frac{x^2}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2}{2\sigma^2}}dx\\&\le 2\sigma\sqrt{2(1+a)\log N}e^{-\frac{\sigma^2\cdot2(1+a)\log N}{4\sigma^2}}\\&\quad\cdot\int_{\sigma\cdot\sqrt{2(1+a)\log N}}^{\infty}\frac{x}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2}{4\sigma^2}}dx\\&=2\sigma\sqrt{2(1+a)\log N}N^{-(1+a)}\int_{\frac{(1+a)\log N}{2}}^{\infty}\sqrt{\frac{2\sigma^2}{\pi}}e^{-t}dt\\&=2\sigma^2\sqrt{\frac{2}{\pi}}\sqrt{2(1+a)\log N}N^{-(1+a)}.\end{aligned} \tag{II.11}$$

The last inequality follows from the fact that $\frac{x^2}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2}{2\sigma^2}}=xe^{-\frac{x^2}{4\sigma^2}}\cdot\frac{x}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2}{4\sigma^2}}$ and that the maximum of $xe^{-\frac{x^2}{4\sigma^2}}$ in the range $[\sigma\cdot\sqrt{2(1+a)\log N},\infty)$ is achieved in the point $x=\sigma\cdot\sqrt{2(1+a)\log N}$. The equalities holds due to simple arithmetic and changing of variables in the integral with $t=\frac{x^2}{4\sigma^2}$.

By summing all the above and observing that $\frac{2}{\pi\log N}<1$ when $N>3$, we get the desired result. □

We note that though the proof is presented only for the greedy-like algorithms, as this is the scope of this correspondence, it can be easily used to extend the results of the other algorithms that guarantee near-oracle performance with high probability.

[4]Similar result for DS appears in [1].

[5]The assumption that $N>3$ is noncrucial for the proof and is used only for getting a better constant.

[6]We assume with no loss of generality that $T=\{1,\ldots,K\}$.

III. Extension to the Nonexact Sparse Case

In the case where $\mathbf{x}$ is not exactly K-sparse, our analysis has to change. Following the work reported in [11], we have the following error bounds for all algorithms (with the different RIP conditions and constants):

Theorem 3.1: If $\mathbf{e}$ is a zero-mean white Gaussian noise vector with variance σ^2, then after at most

$$\ell^* = \left\lceil \log_2 \left(\frac{\|\mathbf{x}\|_2}{\left\|\mathbf{D}_{T_\mathbf{e}}^* \mathbf{e}\right\|_2} \right) \right\rceil$$

iterations and under the appropriate RIP conditions, the reconstruction result, $\hat{\mathbf{x}}$, of SP, CoSaMP and IHT, satisfies

$$E\|\mathbf{x}-\hat{\mathbf{x}}\|_2^2 \le 4C^2(1+a)\log N \cdot K\sigma^2 + 4C^2\left(\|\mathbf{x}-\mathbf{x}_T\|_2 + \frac{1}{\sqrt{K}}\|\mathbf{x}-\mathbf{x}_T\|_1\right)^2 \quad \text{(III.1)}$$

where T denotes the support of the K largest elements in $\mathbf{x}$ and C is the constant from Theorem 2.1.

Proof: Reference [11, Proposition 3.5] provides us with the following claim:

$$\|\mathbf{D}\mathbf{x}\|_2 \le \sqrt{1+\delta_K}\left(\|\mathbf{x}\|_2 + \frac{1}{\sqrt{K}}\|\mathbf{x}\|_1\right). \quad \text{(III.2)}$$

When $\mathbf{x}$ is a nonexact K-sparse vector we get that the effective error in our results becomes $\tilde{\mathbf{e}} = \mathbf{e} + \mathbf{D}(\mathbf{x}-\mathbf{x}_T)$. Thus, using the error bounds of the algorithms (II.4) with the inequality in (III.2) and the relation $\|\mathbf{x}-\hat{\mathbf{x}}\|_2 \le \|\mathbf{x}_T - \hat{\mathbf{x}}\|_2 + \|\mathbf{x}-\mathbf{x}_T\|_2$, we have

$$\begin{aligned} E\|\mathbf{x}-\hat{\mathbf{x}}\|_2^2 &\le E\left(C\|\mathbf{D}_{T_\mathbf{e}}^*\tilde{\mathbf{e}}\|_2 + \|\mathbf{x}-\mathbf{x}_T\|_2\right)^2 \\ &\le C^2 E\left(\|\mathbf{D}_{T_\mathbf{e}}^*\mathbf{e}\|_2 + \sqrt{1+\delta_K}\|\mathbf{D}(\mathbf{x}-\mathbf{x}_T)\|_2 + \frac{1}{C}\|\mathbf{x}-\mathbf{x}_T\|_2\right)^2 \\ &\le C^2 E\Bigg(\|\mathbf{D}_{T_\mathbf{e}}^*\mathbf{e}\|_2 + (1+\delta_K)\|\mathbf{x}-\mathbf{x}_T\|_2 \\ &\quad + \frac{1+\delta_K}{\sqrt{K}}\|\mathbf{x}-\mathbf{x}_T\|_1 + \frac{1}{C}\|\mathbf{x}-\mathbf{x}_T\|_2\Bigg)^2. \end{aligned} \quad \text{(III.3)}$$

Using the fact that $E\mathbf{e} = 0$ and similar steps to those taken in Theorem 2.4 gives

$$E\|\mathbf{x}-\hat{\mathbf{x}}\|_2^2 \le 4C^2(1+a)\log N \cdot K\sigma^2 + C^2\left(\left(\frac{1}{C}+1+\delta_K\right)\|\mathbf{x}-\mathbf{x}_T\|_2 + \frac{1+\delta_K}{\sqrt{K}}\|\mathbf{x}-\mathbf{x}_T\|_1\right)^2. \quad \text{(III.4)}$$

Since the RIP condition for all the algorithms satisfies $\delta_K \le 0.5$ and $C \ge 2$, plugging this into (III.4) gives (III.1), which concludes the proof. □

Embarking from (III.3) and using (III.4) for the first term, we obtain also the inequality

$$E\|\mathbf{x}-\hat{\mathbf{x}}\|_2^2 \le 4C(1+a)\log N \cdot K\sigma^2 + \left(\|\mathbf{x}-\mathbf{x}_T\|_2 + C\left\|\mathbf{D}_{T_\mathbf{e}}^\dagger \mathbf{D}(\mathbf{x}-\mathbf{x}_T)\right\|_2\right)^2. \quad \text{(III.5)}$$

Remark 3.2: For a K-sparse vector $\mathbf{x}$, by applying the SP, CoSaMP, and IHT algorithms with $K = \sum_i I(|\mathbf{x}_i| > \sigma)$ (I is the indicator function and $\mathbf{x}_i$ is the ith element in $\mathbf{x}$), one can easily get from (III.5) a bound of the form $E\|\mathbf{x}-\hat{\mathbf{x}}\|_2^2 \le 4(1+a)C^2\log N \cdot \sum \min(\mathbf{x}_i^2, \sigma^2)$. This bound is proportional to a better oracle that for small elements of $\mathbf{x}$ estimates 0. Unlike the regular oracle that uses the support of the original vector $\mathbf{x}$, this oracle uses the support that minimizes the MSE. Its MSE is lower bounded by $0.5\sum \min(\mathbf{x}_i^2, \sigma^2)$ [1], [6].

IV. Conclusion

In this correspondence, we have presented near-oracle performance guarantees for three greedy-like algorithms—subspace pursuit, CoSaMP, and iterative hard-thresholding. The approach taken in our analysis is an RIP-based (as opposed to mutual-coherence) and uses the existing worst case guarantees of these algorithms. Our study leads to uniform guarantees for the three algorithms explored, i.e., the near-oracle error bounds are dependent only on the dictionary properties (RIP constant) and the sparsity level of the sought solution. In addition, those bounds hold also for the MSE of the reconstruction and not only with high probability for the squared error, as was done in previous works for other algorithms. We have also presented a simple extension of our results to the case where the representation is only approximately sparse.

Acknowledgment

The authors would like to thank Z. Ben-Haim for fruitful discussions and relevant references to existing literature, which helped in shaping this work. The authors are also thankful to the reviewers, who helped improve this correspondence to its current form. In particular, the bound that the authors have obtained for the expected-error is a direct consequence of comments the reviewers made.

References

[1] E. J. Candès, "Modern statistical estimation via oracle inequalities," *Acta Numerica*, vol. 15, pp. 257–325, 2006.
[2] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, 2009.
[3] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
[4] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
[5] 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.
[6] E. J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Statist.*, vol. 35, p. 2313, 2007.
[7] S. Foucart, "A note on guaranteed sparse recovery via l1-minimization," *Appl. Comput. Harmon. Anal.*, vol. 29, no. 1, pp. 97–103, 2010.
[8] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, 1993.
[9] H. Rauhut, "On the impossibility of uniform sparse reconstruction using greedy methods," *Sampl. Theory Signal Image Process.*, vol. 7, no. 2, pp. 197–215, 2008.
[10] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 310–316, Apr. 2010.
[11] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.
[12] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.
[13] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, 2009.
[14] P. J. Bickel, Y. Ritov, and A. B. Tsybakov, "Simultaneous analysis of Lasso and Dantzig selector," *Ann. Statist.*, vol. 37, no. 4, pp. 1705–1732, 2009.
[15] T. Cai, L. Wang, and G. Xu, "Stable recovery of sparse signals and an oracle inequality," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3516–3522, Jul. 2010.
[16] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5030–5043, Oct. 2010.
[17] R. Giryes and M. Elad, "RIP-Based near-oracle performance guarantees for Subspace-Pursuit, CoSaMP, and Iterative Hard-Thresholding," Technion—Israel Institute of Technology, Haifa, Israel, Technion CS-2010-23, 2010.