DENOISING WITH GREEDY-LIKE PURSUIT ALGORITHMS

Raja Giryes and Michael Elad

Computer Science Department,
Technion - Israel Institute of Technology

ABSTRACT

This paper provides theoretical guarantees for denoising performance of greedy-like methods. Those include Compressive Sampling Matching Pursuit (CoSaMP), Subspace Pursuit (SP), and Iterative Hard Thresholding (IHT). Our results show that the denoising obtained with these algorithms is a constant and a log-factor away from the oracle's performance, if the signal's representation is sufficiently sparse. Turning to practice, we show how to convert these algorithms to work without knowing the target cardinality, and instead constrain the solution to an error-budget. Denoising tests on synthetic data and image patches show the potential in this stagewise technique as a replacement of the classical OMP.

1. INTRODUCTION

Signal denoising is a long-studied problem: We are given a signal $\mathbf{y} \in \mathbb{R}^m$, which is a result of contamination of an unknown clean signal $\mathbf{y}_0$ with additive noise $\mathbf{e}$, i.e. $\mathbf{y} = \mathbf{y}_0 + \mathbf{e}$. The most popular denoising problem assumes the case where the noise is i.i.d., white and Gaussian with known variance σ^2. The task is to recover $\mathbf{y}_0$ from $\mathbf{y}$.

In order to be able to denoise the signal, a model for the ideal data should be added. We shall assume that the ideal signal is created as $\mathbf{y}_0 = \mathbf{D}\mathbf{x}$, where $\mathbf{D} \in \mathbb{R}^{m\times n}$ is a redundant dictionary (a matrix with $m \leq n$), and $\mathbf{x}$ is the signal's representation, known to have K dominant elements (almost K-sparse). Thus,

$$\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{e}. \quad (1.1)$$

We shall further assume that the columns of $\mathbf{D}$ are ℓ_2-normalized, in order to simplify the analysis that follows.

Denoising of $\mathbf{y}$ requires finding a sparse vector that could explain the measurements. Put differently, we need to find (or, more practically, approximate) the representation vector, obtaining $\hat{\mathbf{x}}$. There are many pursuit methods that aim to do just that, and in this paper we shall concentrate on three such algorithms, the CoSaMP [1], the SP [2], and the IHT [3]. These are greedy-like methods that estimate $\hat{\mathbf{x}}$ by a series of iterations that detect/cancel likely non-zeros in $\mathbf{x}$.

In our recent work, [4], we have analyzed the performance of these algorithms, focusing on the error obtained in estimating the representation $\hat{\mathbf{x}}$ compared to its true value $\mathbf{x}$, i.e., studying the error[1] $\|\hat{\mathbf{x}} - \mathbf{x}\|$. However, as described above, our goal is signal denoising, and thus, we should instead focus on the error these methods yield in the signal (and not the representation) domain, $\|\mathbf{D}(\hat{\mathbf{x}} - \mathbf{x})\|$. In this paper we extend our earlier results and obtain theoretical guarantees for denoising performance of these greedy-like methods. Our results show that the denoising obtained with these algorithms is a constant and a log-factor away from the oracle's performance, if the signal's representation is sufficiently sparse.

Turning to practice, the CoSaMP, SP, and IHT algorithms suffer from a shortcoming that limits their usability. All three assume that the ideal representation's cardinality, K, is known, which is rarely the case. Instead, since σ^2 is assumed known, it is more natural to use pursuit techniques that aim to find $\hat{\mathbf{x}}$ such that the representation error of the found signal is below the noise energy. In this paper we propose a way to convert these three algorithms to work without knowing the target cardinality, instead constraining the solution to an error-budget. We refer to these as stagewise algorithms, as their support cardinality varies from one iteration to the next.

In order to demonstrate the stagewise variation in practice, we introduce denoising tests on synthetic data and image patches, both showing the potential in this technique as a replacement of the Orthogonal Matching Pursuit (OMP) [5].

The paper is organized as follows: In Section 2 we provide background theoretical results on the CoSaMP, SP, and IHT. Section 3 presents a guarantees obtained for the reconstruction of the signal using the three greedy-like methods. In Section 4 we present a variation of the greedy-like techniques that do not require the knowledge of the cardinality K. In Section 5 we present some simulations results, and in Section 6 we conclude the paper.

2. BACKGROUND

The analysis in this work uses the Restricted Isometry Property (RIP) [6]. We say that a matrix $\mathbf{D}$ satisfies the RIP with parameter δ_K if for every K-sparse (have at most K non-zero entries) vector $\mathbf{v}$

$$(1-\delta_K)\|\mathbf{v}\|^2 \leq \|\mathbf{D}\mathbf{v}\|^2 \leq (1+\delta_K)\|\mathbf{v}\|^2, \quad (2.1)$$

where $\|\cdot\|$ is the ℓ_2 norm.

When the support, T (of cardinality K), of the most dominant elements in $\mathbf{x}$ is known beforehand, the oracle estimator is given by $\hat{\mathbf{x}}_{\text{Oracle}} = \mathbf{D}_T^{\dagger}\mathbf{y}$, where $\mathbf{D}_T$ is a sub-matrix of $\mathbf{D}$, containing the columns related to the support T. The oracle approximation satisfies the upper bound [3, 4]

$$\begin{aligned} E\|\mathbf{x} - \hat{\mathbf{x}}_{\text{Oracle}}\|_2^2 \quad &\leq \quad \frac{K}{1-\delta_K}\sigma^2 \qquad (2.2) \\ + \left(\left(1+\frac{\sqrt{1+\delta_K}}{\sqrt{1-\delta_K}}\right)\|\mathbf{x}-\mathbf{x}_T\| \right. &+ \left. \frac{\sqrt{1+\delta_K}}{\sqrt{1-\delta_K}}\frac{\|\mathbf{x}-\mathbf{x}_T\|_1}{\sqrt{K}}\right)^2, \end{aligned}$$

This research was supported by the the European Community's FP7-FET program SMALL project (grant agreement no. 225913) and the Israel Science Foundation (ISF) grant number 1031/08. RG is also grateful to the Azrieli Foundation for the award of an Azrieli Fellowship.

[1]Here and elsewhere in the paper, such norm corresponds to ℓ_2.

and the lower bound [4]

$$E\|\mathbf{x}-\hat{\mathbf{x}}_{\text{Oracle}}\|_2^2 \geq \frac{K\sigma^2}{1+\delta_K} + \left\|\mathbf{x}-\mathbf{x}_T - \mathbf{D}_T^{\dagger}\mathbf{D}(\mathbf{x}-\mathbf{x}_T)\right\|_2^2,$$

where the expectation is relative to the noise and $\mathbf{x}_T$ is the true representation vector, containing its leading K entries, and zeroing the rest.[2] Thus, for a truly K-sparse representation, the second terms vanish and we get $\frac{K\sigma^2}{1+\delta_K} \leq E\|\mathbf{x}-\hat{\mathbf{x}}_{\text{Oracle}}\|^2 \leq \frac{K\sigma^2}{1-\delta_K}$.

Since the oracle is an impossible tool in practice, it is interesting to ask how practical methods do in this case. Candes and Tao have shown that, when $\mathbf{D}$ satisfies the RIP with $\delta_{2K}+\delta_{3K}<1$, the Dantzig Selector's (DS) performance is similar to the oracle's up to a constant and $\log(n)$ factor with high probability [7]. These factors are unavoidable according to [8]. Similar results were presented in [9] for the Basis Pursuit (BP). Mutual-Coherence based results for these algorithms were presented in [10, 11]. The work in [10] also presented parallel and somewhat weaker results for the OMP and a thresholding algorithm.

The relaxation based techniques are high complexity algorithms. On the other hand, classical greedy methods, as the OMP are known to be much simpler. Unfortunately, Mutual-Coherence based bounds for the OMP and the Thresholding algorithms in [10] show a dependency on the values of the entries of $\mathbf{x}$, implying weaker performance guarantees.

The CoSaMP, SP, and IHT stand as a midway between the simpler and weaker greedy methods and the more complex but better performing relaxation techniques. CoSaMP and SP are described in Algorithm 1. The operator $\text{supp}(\cdot,K)$, used in these algorithms, gives the support of the K-th largest elements in a given vector. IHT is simpler, using the following iterative formula,

$$\hat{\mathbf{x}}_{IHT}^{\ell} = \left[\hat{\mathbf{x}}_{IHT}^{\ell-1} + \mathbf{D}^T(\mathbf{y}-\mathbf{D}\hat{\mathbf{x}}_{IHT}^{\ell-1})\right]_K \tag{2.3}$$

where $[\cdot]_K$ is a hard thresholding operator that takes the K largest elements and zeros the rest. Different stopping criteria can be sought for these algorithms as described in [1, 2, 3]. In this work, we look on the residual's norm and stop when it goes under the noise power.

In our earlier work, [4], RIP-based performance results were presented for these greedy-like methods. While these algorithms' complexity is comparable to that of the greedy techniques, their performance is similar to the one obtained by relaxation based methods (BP and DS). The main theorem in [4] paper states that:

Theorem 2.1 (Theorem 3.1 in [4]) *If the conditions* $\delta_{3K} \leq 0.139$, $\delta_{4K} \leq 0.1$ *and* $\delta_{3K} \leq 1/\sqrt{32}$ *hold for SP, CoSaMP and IHT respectively, then with probability exceeding* $1-(\sqrt{\pi(1+a)\log n}\cdot n^a)^{-1}$ *we obtain*

$$\begin{aligned}\|\mathbf{x}-\hat{\mathbf{x}}\|^2 \quad &\leq \quad 2\cdot C^2\Big(\sqrt{(1+a)\log n\cdot K}\cdot\sigma \\ &+\|\mathbf{x}-\mathbf{x}_T\| + \frac{1}{\sqrt{K}}\|\mathbf{x}-\mathbf{x}_T\|_1\Big)^2.\end{aligned} \tag{2.4}$$

where $\hat{\mathbf{x}}$ *is the reconstruction result and* C *is a constant that depends only on the RIP (and differs from one algorithm to another – see [4]).*

Algorithm 1 Subspace Pursuit (SP) and CoSaMP

Require: $K,\mathbf{D},\mathbf{y}$ where $\mathbf{y}=\mathbf{D}\mathbf{x}+\mathbf{e}$, K is the cardinality of $\mathbf{x}$ and $\mathbf{e}$ is the additive noise. $\alpha=1$ (SP), $\alpha=2$ (CoSaMP).

Ensure: $\hat{\mathbf{x}}_{\text{CoSaMP}}$ or $\hat{\mathbf{x}}_{SP}$: K-sparse approximation of $\mathbf{x}$.

Initialize the support and the residual: $T^0=\emptyset$, $\mathbf{y}_r^0=\mathbf{y}$.

Set $\ell=0$.

while halting criterion is not satisfied **do**

  $\ell=\ell+1$.

  Find new support elements: $T_\Delta = \text{supp}(\mathbf{D}^*\mathbf{y}_r^{\ell-1},\alpha K)$.

  Update the support: $\tilde{T}^\ell = T^{\ell-1}\cup T_\Delta$.

  Compute the representation: $\mathbf{x}_p = \mathbf{D}_{\tilde{T}^\ell}^{\dagger}\mathbf{y}$.

  Prune small entries: $T^\ell = \text{supp}(\mathbf{x}_p,K)$.

  Update the residual: $\mathbf{y}_r^\ell = \mathbf{y}-\mathbf{D}_{T^\ell}(\mathbf{x}_p)_{T^\ell}$ for CoSaMP, and $\mathbf{y}_r^\ell = \mathbf{y}-\mathbf{D}_{T^\ell}\mathbf{D}_{T^\ell}^{\dagger}\mathbf{y}$ for SP.

end while

Form the final solution: $\hat{\mathbf{x}}_{\text{CoSaMP},T^\ell} = (\mathbf{x}_p)_{T^\ell}$ for CoSaMP and $\hat{\mathbf{x}}_{SP,T^\ell} = \mathbf{D}_{T^\ell}^{\dagger}\mathbf{y}$ for SP.

This theorem shows that the reconstruction result of the greedy-like algorithms achieves the performance of the oracle up to a constant and a $\log(n)$ factor. We now turn to derive a similar result for the error of the reconstructed signal $\mathbf{D}\hat{\mathbf{x}}$.

3. DENOISING GUARANTEES

Since the results we obtain are to be compared to the oracle ones, we start by deriving the signal approximation-error expected by the oracle. The oracle estimator, $\mathbf{D}\hat{\mathbf{x}}_{\text{Oracle}}$, satisfies

$$\begin{aligned} E\|\mathbf{D}(\mathbf{x}-\hat{\mathbf{x}}_{\text{Oracle}})\|^2 \quad &= \\ &= E\left\|\mathbf{D}\left(\mathbf{x}-\mathbf{x}_T-\mathbf{D}_T^{\dagger}\mathbf{D}(\mathbf{x}-\mathbf{x}_T)-\mathbf{D}_T^{\dagger}\mathbf{e}\right)\right\|^2 \\ &= E\left\|\left(\mathbf{I}-\mathbf{D}_T\mathbf{D}_T^{\dagger}\right)\mathbf{D}(\mathbf{x}-\mathbf{x}_T)\right\|^2 + E\left\|\mathbf{D}_T\mathbf{D}_T^{\dagger}\mathbf{e}\right\|^2 \\ &= E\left\|\left(\mathbf{I}-\mathbf{D}_T\mathbf{D}_T^{\dagger}\right)\mathbf{D}(\mathbf{x}-\mathbf{x}_T)\right\|^2 + K\sigma^2. \end{aligned} \tag{3.1}$$

In the above derivation we have used the definition of $\hat{\mathbf{x}}_{\text{Oracle}}$ in the first step, the fact that $E\mathbf{e}=0$ in the second step and the property $E\mathbf{e}\mathbf{e}^T=\sigma^2\mathbf{I}$ in the last one. By observing that $\left\|\mathbf{I}-\mathbf{D}_T\mathbf{D}_T^{\dagger}\right\|=1$ we have the upper bound

$$E\|\mathbf{D}(\mathbf{x}-\hat{\mathbf{x}}_{\text{Oracle}})\|^2 \leq E\|\mathbf{D}(\mathbf{x}-\mathbf{x}_T)\|^2 + K\sigma^2. \tag{3.2}$$

For a truly K-sparse representation, the oracle error is simply $K\sigma^2$, as the first term drops.

We now turn to derive a similar bound for the greedy-like methods. Theorem 2.1 have shown that these algorithms have near oracle performance in the representation domain. The following theorem shows that the same holds true for the signal estimation:

Theorem 3.1 *If the conditions* $\delta_{3K} \leq 0.139$, $\delta_{4K} \leq 0.1$ *and* $\delta_{3K} \leq 1/\sqrt{32}$ *hold for SP, CoSaMP and IHT respectively,*

[2] Abusing notation, $\mathbf{D}_T^{\dagger}\mathbf{y}$ and $\mathbf{x}_T$ may refer to vectors of length K or vectors with K non-zero elements padded with zeros. The meaning will be clear from the context.

then with probability exceeding $1-(\sqrt{\pi(1+a)\log n}\cdot n^a)^{-1}$ *we obtain*

$$\|\mathbf{D}(\mathbf{x}-\hat{\mathbf{x}})\| \le \|\mathbf{D}(\mathbf{x}-\mathbf{x}_T)\| + \sqrt{2(1+\delta_{2K})}C\cdot \left(\sqrt{(1+a)\log n}\sqrt{K}\sigma + \|\mathbf{D}(\mathbf{x}-\mathbf{x}_T)\|\right). \tag{3.3}$$

where $\hat{\mathbf{x}}$ *is the reconstruction result and* C *is a constant that depends only on the RIP.*

Proof: Using the triangle inequality and the fact that $\mathbf{x}_T - \hat{\mathbf{x}}$ is $2K$-sparse at most we have that

$$\|\mathbf{D}(\mathbf{x}-\hat{\mathbf{x}})\| \le \sqrt{1+\delta_{2K}}\,\|(\mathbf{x}_T-\hat{\mathbf{x}})\| + \|\mathbf{D}(\mathbf{x}-\mathbf{x}_T)\|. \tag{3.4}$$

Using the same technique used for the proof of Theorem 5.1 in [4], we get that the first term in the above inequality can be bounded by

$$\begin{aligned}\|\mathbf{x}_T-\hat{\mathbf{x}}\| \quad &\le \quad C\sqrt{2(1+a)\log n}\sqrt{K}\sigma \\ &+C\sqrt{1+\delta_K}\,\|\mathbf{D}(\mathbf{x}-\mathbf{x}_T)\|.\end{aligned} \tag{3.5}$$

Plugging (3.5) into (3.4) and using the fact that the RIP condition for all the algorithms satisfies $\delta_K \le 0.5$ gives us the desired result. □

In some cases the noise power is stronger than some of the signal representation elements. In this case, the zero estimation is more favorable and leads to a lower error. The oracle estimator that suites this case is the one that chooses the support that minimizes the MSE, instead of using the original support. Denoting by $\hat{\mathbf{x}}^T = \mathbf{D}_T^{\dagger}\mathbf{y}$, this oracle estimator is:

$$\mathbf{x}^*_{\text{Oracle}} = arg\min E\left\|\hat{\mathbf{x}}^T - \mathbf{x}\right\|^2. \tag{3.6}$$

In the exact K-sparse case the representation error is bounded from below by $\frac{1}{2}\sum\min(\mathbf{x}_i^2,\sigma^2)$, where $\mathbf{x}_i$ is the i-th element in $\mathbf{x}$. Using the RIP, the bound for the signal is

$$E\|\mathbf{D}(\mathbf{x}-\mathbf{x}_{\text{Oracle}})\|^2 \ge \frac{1-\delta_{2K}}{2}\sum\min(\mathbf{x}_i^2,\sigma^2). \tag{3.7}$$

We are apt to ask whether the greedy-like techniques' performance are also proportional to this oracle. The following theorem shows that by applying these algorithms with $K' = \sum_i I(|\mathbf{x}_i| > \sigma) \le K$ the reconstruction error is like the better oracle up to a constant and a $\log(n)$ factor as before. An equivalent result for the representation error appears in Remark 3.2 in [4] without a proof.

Theorem 3.2 *By applying SP, CoSaMP and IHT with* $K' = \sum_i I(|\mathbf{x}_i| > \sqrt{\log n}\sigma)$ *in the exact* K*-sparse case, if the conditions* $\delta_{3K} \le 0.139$, $\delta_{4K} \le 0.1$ *and* $\delta_{3K} \le 1/\sqrt{32}$ *hold for SP, CoSaMP and IHT respectively, then with probability exceeding* $1-(\sqrt{\pi(1+a)\log n}\cdot n^a)^{-1}$ *we obtain*

$$\begin{aligned}\|\mathbf{D}(\mathbf{x}-\hat{\mathbf{x}})\|_2^2 \quad &\le \quad 2(a+3)(1+\delta_{2K})C^2\cdot \\ &\sum\min(\mathbf{x}_i^2,\log n\sigma^2).\end{aligned} \tag{3.8}$$

where $\hat{\mathbf{x}}$ *is the reconstruction result and* C *is a constant that depends only on the RIP.*

Proof: We denote by T' the support of the K' largest elements in $\mathbf{x}$. By looking at $\mathbf{x}$ as a nearly K'-sparse vector and using the result of Theorem 3.1 we have

$$\begin{aligned}\|\mathbf{D}(\mathbf{x}-\hat{\mathbf{x}})\| &\le ((1+\delta_{2K})C+1)\,\|\mathbf{D}(\mathbf{x}-\mathbf{x}_{T'})\| \\ &\quad + C\sqrt{2(1+\delta_{2K})(1+a)\log n}\sqrt{K'}\sigma \\ &\le C\sqrt{2(1+\delta_{2K})}\big(\sqrt{2}\,\|\mathbf{x}-\mathbf{x}_{T'}\| \\ &\quad + \sqrt{(1+a)\log n}\sqrt{K'}\sigma\big).\end{aligned} \tag{3.9}$$

The last step relies on the observation that $\mathbf{x}-\mathbf{x}_T$ is $2K$-sparse and thus $\|\mathbf{D}(\mathbf{x}-\hat{\mathbf{x}})\| \le \sqrt{1+\delta_{2K}}\,\|\mathbf{x}-\hat{\mathbf{x}}\|$, and the observation that since $C \ge 2$ and $\delta_{2K} \le 0.5$ we have that $(1+\delta_{2K})C+1 \le 2C$. Taking square on both sides of (3.9) we have

$$\begin{aligned}&\|\mathbf{D}(\mathbf{x}-\hat{\mathbf{x}})\|^2 \\ &\le 2(1+\delta_{2K})C^2\Bigg(2\,\|\mathbf{x}-\mathbf{x}_{T'}\|^2 + (1+a)\log nK'\sigma^2 \\ &\quad +2\sqrt{(1+a)\log n}\sqrt{K'}\sigma\sqrt{2}\,\|\mathbf{x}-\mathbf{x}_{T'}\|\Bigg) \\ &\le 2(a+3)(1+\delta_{2K})C^2\Bigg(\|\mathbf{x}-\mathbf{x}_{T'}\|^2 + \log nK'\sigma^2\Bigg).\end{aligned} \tag{3.10}$$

The last step follows from the fact that any three scalars a,b,c satisfies $ab \le \frac{ca^2}{2} + \frac{b^2}{2c}$. In our case $a = \sqrt{(1+a)\log n}\sqrt{K'}\sigma$, $b = \sqrt{2}\,\|\mathbf{x}-\mathbf{x}_{T'}\|$ and $c = \frac{2}{1+a}$. Having the result in (3.10), we get easily (3.8) by using the definition of K'. □

4. STAGEWISE GREEDY-LIKE ALGORITHMS

Theorems 3.1 and 3.2 seem promising for signal denoising using greedy-like techniques. However, the problem with the discussed algorithms is that they require the cardinalities K or K' to be known. This restriction limits their use in denoising problems, where generally they are not known beforehand. Thus we introduce a variation on the algorithms that removes this restriction, replacing it with the knowledge of σ. In this variation, the effective cardinality grows in each iteration by a fixed value β. The stopping criterion used is $\|\mathbf{y}-\mathbf{D}\hat{\mathbf{x}}\| \le c\sqrt{n}\sigma$ where c is a pre-chosen constant (typically $c = 1.1$). The stagewise variation on the CoSaMP and SP is shown in Algorithm 2.

An open question at this stage is a theoretical guarantee for the denoising performance for these modified algorithms, and this is part of our future work. Under the assumption that these algorithms stop after $O(K)$ iterations, a property that holds for CoSaMP and SP for a large family of signals [1, 2], their complexity is $O(K(K^2m+mn))$, like CoSaMP and SP. Generally $K^2 \ll n$ and, therefore, the complexity is effectively $O(Kmn)$. For high dimensions, one can calculate the pseudo-inverse in the algorithms using an iterative method for reducing their complexity, as was done in [1].

In a similar way, stagewise IHT would be

$$\hat{\mathbf{x}}^{\ell}_{IHT} = \left[\hat{\mathbf{x}}^{\ell-1}_{IHT} + \mathbf{D}^T(\mathbf{y}-\mathbf{D}\hat{\mathbf{x}}^{\ell-1}_{IHT})\right]_{\beta\ell}, \tag{4.1}$$

As before, by assuming that the algorithm stops after $O(K)$ iterations, the complexity is $O(Kmn)$ and of the same order of the original IHT.

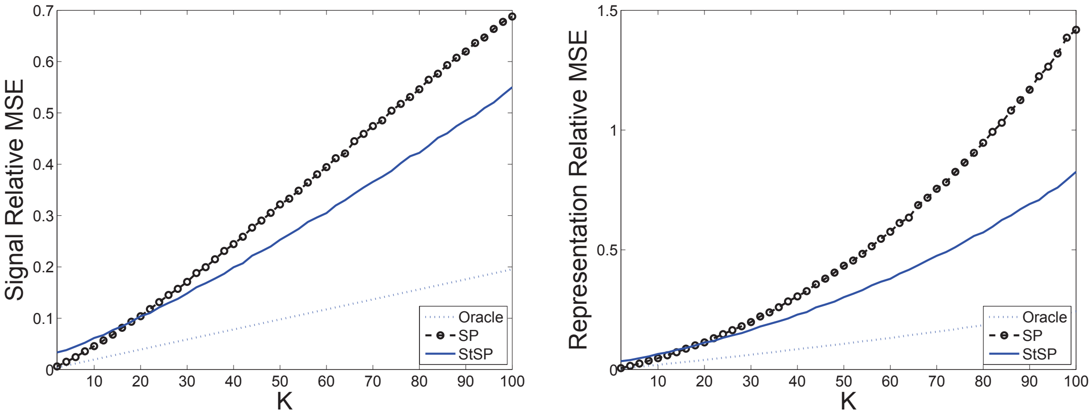

Figure 1: The signal relative error (left) and representation error (right) as achieved by SP and StSP as a function of the cardinality for SNR=4. The graphs also show the oracle performance.

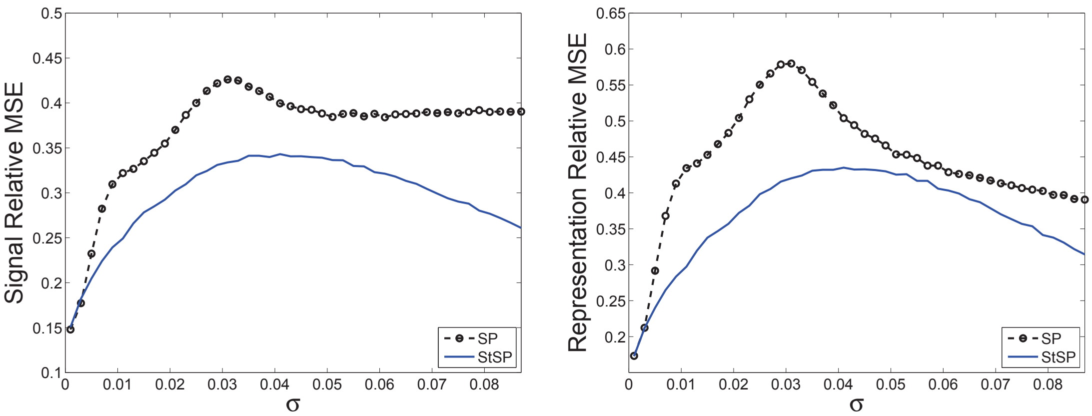

Figure 2: The signal relative error (left) and representation error (right) as achieved by SP and StSP as a function of the noise variance for $K = 50$.

Algorithm 2 Stagewise CoSaMP (StCoSaMP) and Stagewise SP (StSP) Algorithms

---

Require: $K, \mathbf{D}, \mathbf{y}, \beta$ where $\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{e}$, K is the cardinality of $\mathbf{x}$, $\mathbf{e}$ is the additive noise and β is the increase rate of the cardinality. $\alpha = 1$ (StSP), $\alpha = 2$ (StCoSaMP).

Ensure: $\hat{\mathbf{x}}_{\text{StCoSaMP}}$ or $\hat{\mathbf{x}}_{\text{StSP}}$: K-sparse approximation of $\mathbf{x}$.

Initialize the support and the residual: $T^0 = \emptyset$, $\mathbf{y}_r^0 = \mathbf{y}$.

Set $\ell = 0$.

while halting criterion is not satisfied **do**

  $\ell = \ell + 1$.

  Find new support elements: $T_\Delta = \text{supp}(\mathbf{D}^*\mathbf{y}_r^{\ell-1}, \alpha\beta\ell)$.

  Update the support: $\tilde{T}^\ell = T^{\ell-1} \cup T_\Delta$.

  Compute the representation: $\mathbf{x}_p = \mathbf{D}_{\tilde{T}^\ell}^\dagger \mathbf{y}$.

  Prune small entries: $T^\ell = \text{supp}(\mathbf{x}_p, \beta\ell)$.

  Update the residual: $\mathbf{y}_r^\ell = \mathbf{y} - \mathbf{D}_{T^\ell}(\mathbf{x}_p)_{T^\ell}$ for StCoSaMP, and $\mathbf{y}_r^\ell = \mathbf{y} - \mathbf{D}_{T^\ell}\mathbf{D}_{T^\ell}^\dagger \mathbf{y}$ for StSP.

end while

Form the final solution: $\hat{\mathbf{x}}_{\text{StCoSaMP},T^\ell} = (\mathbf{x}_p)_{T^\ell}$ for StCoSaMP and $\hat{\mathbf{x}}_{\text{StSP},T^\ell} = \mathbf{D}_{T^\ell}^\dagger \mathbf{y}$ for StSP.

Before moving to the next section we just note that a thresholding operation can be performed on the result of each of the algorithms using the rule for choosing K' in Theorem 3.2. All the entries in the representation that are smaller than $\sqrt{\log n}\sigma$ can be pruned.

5. EXPERIMENTS

In this section we check the denoising results of the greedy-like methods. Due to lack of space we present results only for SP and StSP (with $\beta = 1$) as representatives of the greedy-like techniques.

The first experiment compares the reconstruction results of SP and StSP with the oracle error. In the experiment a random dictionary with entries drawn from the canonic normal distribution is used. The columns of the dictionary are normalized and the dimensions are $m = 512$ and $n = 1024$. The vector $\mathbf{x}$ is an exact K-sparse vector and is normalized. Its non-zero entries are chosen from a white Gaussian distribution and its support is selected uniformly at random. The support and the non-zero values are statistically independent. We repeat each experiment 2000 times and average.

Fig. 1 presents the relative (with respect to the initial

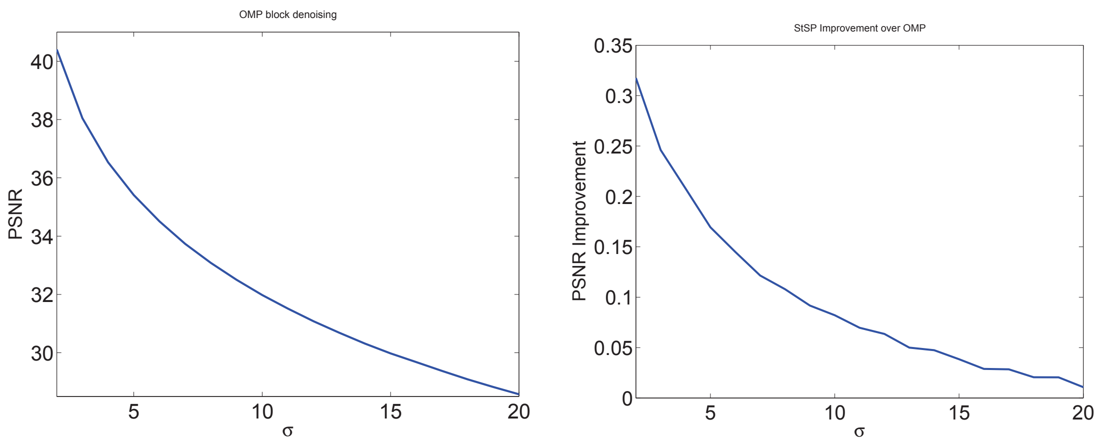

Figure 3: The PSNR of OMP for image patches as a function of σ (left), and the PSNR improvement (PSNR(StSP) - PSNR(OMP)) obtained using the StSP (right).

noise) MSE of each of the methods as a function of the cardinality where the noise power is set to satisfy an SNR (signal to noise ratio) of 4. It can be seen that SP and StSP raise linearly with the cardinality as predicted theoretically for SP. For high cardinalities, StSP behaves better than SP despite the fact that the last has an information about the support size. This happens since StSP, unlike SP, is not restricted to the support size and can throw elements smaller than the noise power. For low cardinalities the significance of the number of elements raises and thus the prior knowledge about the support size grants SP a better performance results.

Fig. 2 presents the relative MSE of each of the methods as a function of the noise power. The cardinality is set to $K = 50$ and σ ranges from 0.001 (high SNR) to $\frac{2}{\sqrt{m}}$ (SNR of 0.5). For high SNR, SP and StSP have similar performance. For lower SNR, StSP becomes better since it has the freedom, as observed before, to select a smaller support size leaving out elements that are smaller than the noise power.

The second experiment uses overlapping image patches of size 8×8, taken from the image *Lenna*. A denoising for each patch is performed using error-driven OMP and StSP, with the redundant DCT dictionary. Fig. 3 (left) shows the average PSNR of all the patches for denoising using OMP as a function of σ. On the right, this figure presents the improvement obtained by replacing OMP by StSP. We can see that StSP performs better, and especially so for weak noise.

6. CONCLUSION

In this paper we derive a theoretical guarantee for the denoising performance of three greedy-like methods. This guarantee, posed in the signal domain, suggests that CoSaMP, SP, and IHT, have near-oracle denoising performance for sufficiently sparse signals. On the practical side, we show that a stagewise version of these algorithm can be posed, and used in applications where the cardinality is unknown. More work is required to close the gap and suggest a theoretical oracle-like denoising guarantee for the stagewise version of these algorithms.

REFERENCES

[1] D. Needell and J.A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301 – 321, May 2009.

[2] Wei Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory.*, vol. 55, no. 5, pp. 2230 –2249, May 2009.

[3] T. Blumensath and M.E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265 – 274, 2009.

[4] R. Giryes and M. Elad, "RIP-based near-oracle performance guarantees for SP, CoSaMP, and IHT," *submitted to IEEE Trans. Signal Process.*, 2011.

[5] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, pp. 3397–3415, 1993.

[6] 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.

[7] E.J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Annals Of Statistics*, vol. 35, pp. 2313, 2007.

[8] E.J. Candès, "Modern statistical estimation via oracle inequalities," *Acta Numerica*, vol. 15, pp. 257–325, 2006.

[9] P.J. Bickel, Y. Ritov, and A.B. Tsybakov, "Simultaneous analysis of lasso and dantzig selector," *Annals of Statistics*, vol. 37, no. 4, pp. 1705–1732, 2009.

[10] 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.

[11] 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.