ONLINE PERFORMANCE GUARANTEES FOR SPARSE RECOVERY

Raja Giryes[1] and Volkan Cevher[2*]*

[1]Computer Science Department
[2]Laboratory for Information and Inference Systems

ABSTRACT

A K^*-sparse vector $x^* \in \mathbb{R}^N$ produces measurements via linear dimensionality reduction as $u = \Phi x^* + n$, where $\Phi \in \mathbb{R}^{M \times N}$ ($M < N$), and $n \in \mathbb{R}^M$ consists of independent and identically distributed, zero mean Gaussian entries with variance σ^2. An algorithm, *after its execution*, determines a vector $\hat{x}$ that has K-nonzero entries, and satisfies $\|u - \Phi \hat{x}\| \leq \epsilon$. How far can $\hat{x}$ be from x^*? When the measurement matrix Φ provides stable embedding to $2K$-sparse signals (the so-called restricted isometry property), they must be very close. This paper therefore establishes worst-case bounds to characterize the distance $\|\hat{x} - x^*\|$ based on the online meta-information. These bounds improve the pre-run algorithmic recovery guarantees, and are quite useful in exploring various data error and solution sparsity trade-offs. We also evaluate the performance of some sparse recovery algorithms in the context of our bound.

***Index Terms**—* near-oracle performance guarantees, restricted isometry property, compressive sensing

1. INTRODUCTION

Many problems in data mining, learning, and compressive sensing (CS) feature a linear dimensionality reduction where an unknown signal $x^* \in \mathbb{R}^N$ is related to its observations $u \in \mathbb{R}^M$ via

$$u = \Phi x^* + n, \qquad (1)$$

where the vector $n \in \mathbb{R}^M$ typically consists of independent and identically distributed (iid), zero mean Gaussian noise with a known variance σ^2. In the sequel, we assume that the columns of Φ are normalized. Because of the broad applicability of (1), there is now a significant interest to obtain improved compression rates (i.e., smaller M); more stable recovery schemes; faster sampling and recovery times; and improved analytical recovery bounds.

Since dimensionality reduction loses information in general (i.e., Φ has a non-trivial null space), we need prior information to correctly identify the true vector x among–possibly–infinitely many other vectors that can produce the same measurements u. It is now well-known that sparse representations provide sufficient prior information for algorithms to stably and provably recover signals when the measurement matrix provide stable embedding to the set of sparse signals. A signal $x \in \mathbb{R}^N$ has a sparse representation as $x = \Psi \mathbf{v}$ in a basis $\Psi \in \mathbb{R}^{N \times N}$, when $K \ll N$ coefficients of $\mathbf{v}$ can exactly represent the signal x. By stable embedding, we mean that the matrix Φ satisfies the restricted isometry property (RIP) with isometry constants $\tilde{\mu}_K$ and $\tilde{L}_K$:

$$\tilde{\mu}_K \|\mathbf{v}\|^2 \leq \|\Phi \mathbf{v}\|^2 \leq \tilde{L}_K \|\mathbf{v}\|^2, \forall \mathbf{v} \in \Sigma_K; \qquad (2)$$

where Σ_K is the set of all K-sparse vectors.[1]

Random matrices satisfy RIP with high probability [1, 2]. Moreover, their isometry constants are well-characterized in many cases. For example, when Φ is generated iid Gaussian and is column normalized to unit norm, we have $\tilde{L}_K \leq (1 + \sqrt{K/M} + \sqrt{2t/M})^2$, and $\tilde{\mu}_K \geq (1 - \sqrt{K/M} - \sqrt{2t/M})^2$ with probability $1 - e^{-t}$. Explicit calculations for sub-sampled Fourier matrices can be found in Theorem 3 in [3]. Other calculations for sparse matrices, based on expander graphs can be found in [4].

Over the last decade, quite a few sparsity seeking recovery algorithms has been proposed, each of which boasts different recovery guarantees and speed, along with different parameters for turning, such as relaxation parameter and desired sparsity level. Surprisingly, when we execute the available software packages for sparse recovery problems, we typically obtain *different* solutions with varying levels of sparsity. Then, the typical rule-of-thumb, e.g., in machine learning, stipulates that we pick the sparsest solution instead of the one that provides the least data error, i.e., $f(x) = \|u - \Phi x\|^2$. A more rigorous understanding of this trade-off could help practitioners to decide which algorithm is more useful for their own problem.

In this paper, we analytically explore such trade-offs in the presence of iid Gaussian noise. We assume the measurement matrix Φ provides stable embedding with known or bounded isometry constants $\tilde{\mu}$ and $\tilde{L}$. We then show that if an algorithm obtains a K-sparse solution $\hat{x}$, where $f(\hat{x}) \leq \tilde{\epsilon}^2$, then

$$\|\hat{x} - x^*\| \leq \frac{\tilde{\epsilon} + \Delta_{x^*, K}}{\sqrt{\tilde{\mu}_{2K}}} + 4\sqrt{\frac{(1+a)K \log N}{\tilde{\mu}_{2K}}} \sigma, \qquad (3)$$

with probability exceeding $1 - (\sqrt{\pi(1+a)\log N} \cdot N^a)^{-1}$, where $\Delta_{x^*, K} = \frac{\sqrt{\tilde{L}_K}}{\sqrt{\tilde{\mu}_{2K}}} \nu_{x^*} + \|x^* - x_K^*\|$ is the irrecoverable energy, x_K^* is the best K-term approximation of the true solution x^* and $\nu_{x^*} = \|x^* - x_K^*\|_2 + \frac{1}{\sqrt{K}} \|x^* - x_K^*\|_1$. ν is known as the irrecoverable energy in sparse approximation [5]. This bound is quite useful to compare the recovery performance of different algorithms *after their execution* when we have estimates of the decay profile and the total energy of typical signals.

The paper is organized as follows. Section 2 presents an oracle performance bound on the problem in the presence of iid Gaussian noise and some preliminary definitions. Section 3 establishes our main result in (3) via a series of theorems. Section 4 compares the performance of three algorithms in the context of our bound. Section 5 provides a concluding discussion.

---

[1]In the sequel, we assume that the signals are canonically sparse without loss of generality (e.g., $\Psi = I$).

2. PRELIMINARIES

We first set the stage by establishing a lower-bound on the minimum error that we can hope to achieve, and then introduce various algorithms that we use in Section 4.

The least-squares oracle estimator is defined as $x_{\text{oracle}} = \Phi^{\dagger}_{x^*_K} \mathbf{u}$, which has the full knowledge of the support of x^*_K (e.g., the locations of the sparse coefficients). We denote $\Phi_{x^*_K}$ as the submatrix of Φ that contains the columns in the support of x^*_K. According to [6, 7], the mean squared error (MSE) of the oracle for the iid white Gaussian noise model is bounded by

$$E \|\hat{x}_{\text{oracle}} - x^*\|^2 \leq \tilde{\Delta}^2_{x^*, K} + \frac{K\sigma^2}{\tilde{\mu}_K},$$

where $\tilde{\Delta}_{x^*, K} \triangleq \|x^* - x^*_K\| + \sqrt{\frac{\tilde{L}_K}{\tilde{\mu}_K}} \nu_{x^*}$ and $\nu_{x^*} = \|x^* - x^*_K\|_2 + \frac{1}{\sqrt{K}} \|x^* - x^*_K\|_1$ [5]. In a similar way, using Jensen's inequality, we can obtain a bound on the mean of the error:

$$E \|\hat{x}_{\text{oracle}} - x^*\| \leq \tilde{\Delta}_{x^*, K} + \frac{\sqrt{K}\sigma}{\sqrt{\tilde{\mu}_K}}.$$

While the oracle estimator seems impractical as we do not know the support of x^* *a priori*, certain algorithms can provide solutions, which are provably a (nearly dimension independent) constant factor off the oracle error. The Dantzig Selector (DS) estimation bound, as characterized by Candes and Tao in [8], is one such example. In [8], the authors show that the DS produces a solution, which is close to the oracle distance up to $\sqrt{\log N}$ and an absolute constant by using the RIP assumption on the measurement matrix Φ. Similar results were shown for basis pursuit denoising (BPDN) in [9] and for compressive sampling matching pursuit (CoSaMP), subspace pursuit (SP) and iterative hard thresholding (IHT) methods in [7]. In [10, 11] mutual coherence based results, which is a weaker descriptor of Φ, were presented for DS, BPDN, orthogonal matching pursuit (OMP) and thresholding methods.

In this work, we develop a recipe to obtain online algorithmic bounds relying on the meta-information we have after signal recovery, such as signal sparsity, data error, etc. We show that when the noise is Gaussian, these bounds are asymptotically close to the oracle bound, hence may provide a much tighter range for our signal estimation error.

We demonstrate our approach on two algorithms. The first, which is a variation of IHT, is fast Lipschitz iterative hard thresholding (FLIHT) proposed in [12]. This algorithm minimizes the ℓ_2 data error by taking two inputs. The first input is the desired sparsity K of the solution, and the second input is a step-size for gradient updates. For instance, the FLIHT iteration is of the form $(a_{i+1} = 0.5(1 + \sqrt{1 + 4a_i^2}))$

$$
\begin{aligned}
y_{i+1} &= x_i + \frac{a_i - 1}{a_{i+1}}(x_i - x_{i-1}) \\
x_{i+1} &= \left[y_i - \frac{1}{2\tilde{L}_{3K}} \nabla f(y_i) \right]_K,
\end{aligned}
\tag{4}
$$

where $[\cdot]_K$ is a hard thresholding operator that keeps the largest K elements in a vector and zeros the rest.

The second algorithm is the the fast iterative shrinkage-thresholding algorithm (FISTA) [13], which solves the BPDN formulation:

$$\hat{x}_{BPDN} = \operatorname*{argmin}_x \|\Phi x - u\|^2 + \gamma \|x\|_1. \tag{5}$$

The FISTA iterations are of the same form as FLIHT, but uses soft thresholding instead of hard thresholding and $L = 2\lambda_{\max}(\Phi^T \Phi)$, the Lipschitz constant of the gradient ∇f, instead of $2\tilde{L}_{3K}$.

Finally, we use the following key lemma from [8] used for establishing the online bounds in our paper based on the iid Gaussian noise:

Lemma 1. *Given a white Gaussian noise vector n with variance σ^2, and a signal support set Ω, it holds that $\left\|\Phi_\Omega^T n\right\|^2 \leq (2(1 + a)\log N)|\Omega|\sigma^2$ with probability exceeding $1 - (\sqrt{\pi(1+a)\log N} \cdot N^a)^{-1}$.*

Proof. The proof is straightforward, and relies on the following concentration-of-measure inequality from Section 3 in [8]:

$$
\mathbf{P}\left(\sup_i |\Phi_i^* n| > \sigma \cdot \sqrt{2(1+a)\log N} \right) \leq
$$
$$
1 - (\sqrt{\pi(1+a)\log N} \cdot N^a)^{-1}.
$$

$\square$

3. ONLINE GUARANTEES

In this section, we present our online bounds for reconstruction algorithms. We assume that we can have access to the following information after the execution of an algorithm: $\hat{K}$, the sparsity of the reconstructed signal, and the data error $\tilde{\epsilon}^2$. This additional information will help us to sharpen the algorithmic guarantees and see the sparsity versus data error trade-offs.

Based on the meta-information, there are two ways of obtaining an improved online estimation bound. The first one uses the known sparsity of the original signal. The second one relies on the sparsity of the reconstruction of the algorithm. The former setting implies the following theorem:
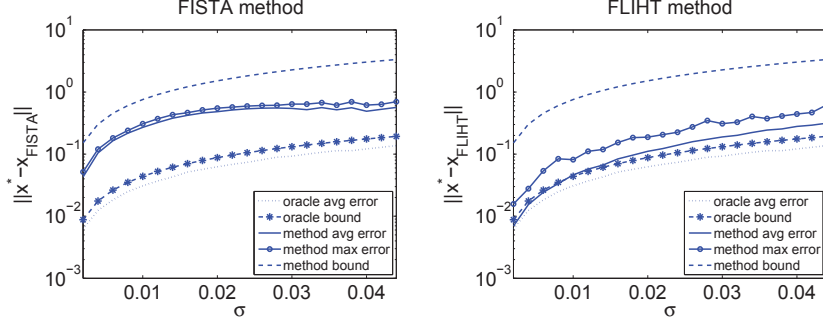
Theorem 2. *Suppose that the measurement matrix Φ has RIP, and that we have a reconstruction algorithm that returns $\hat{x}$ as the estimate of x^* with $f(\hat{x}) - f(x^*) \leq \epsilon^2$. For the case where n is a general adversarial noise, the reconstruction result of the algorithm satisfies $\|x^* - \hat{x}\| \leq \frac{\epsilon}{\sqrt{\tilde{\mu}_{2K}}} + \Delta_{x^*, K} + \Delta_{\hat{x}, K} + \frac{2}{\tilde{\mu}_{2K}} \left\|\Phi_{x^* - \hat{x}}^T n\right\|$.*

The latter setting leads us to the next theorem:

Theorem 3. *Suppose that the measurement matrix Φ has RIP, and that we have a reconstruction algorithm that returns $\hat{K}$-sparse reconstruction $\hat{x}$ as the estimate of x^* and satisfies $f(\hat{x}) - f(x^*) \leq \epsilon^2$. For the case where n is a general adversarial noise, the reconstruction result of the algorithm satisfies $\|x^* - \hat{x}\| \leq \frac{\epsilon}{\sqrt{\tilde{\mu}_{2\hat{K}}}} + \Delta_{x^*, \hat{K}} + \frac{2}{\tilde{\mu}_{2\hat{K}}} \left\|\Phi_{x^* - \hat{x}}^T n\right\|$.*

The proofs are presented in Appendix A. ϵ in the theorems can be approximated by $\tilde{\epsilon}$. In addition, we note that if the sparsity of the reconstruction result is smaller or equal to the one of the original representation than the result of Theorem 3 coincides with that of Theorem 2.

Theorems 2 and 3 above did not assume any distribution on the noise n. In many cases, $\left\|\Phi_{x^* - \hat{x}}^T n\right\|$ can be equal or even greater than the full noise norm $\|n\|$. Adding to that the fact that $\frac{2}{\tilde{\mu}_{2\hat{K}}} > 1$ shows that we might get amplification of the noise power instead of attenuation. Leveraging the noise statistics, we can improve these theorems via Lemma 1 and re-state them in terms of σ:

Fig. 1. Maximum and average error of the FISTA (left) and FLIHT (right) method versus σ compared to the oracle error and the theoretical upper bounds. The FISTA results are not de-biased.

Theorem 4. *Let us assume that the matrix Φ satisfies the RIP condition and that we have a reconstruction algorithm that returns $\hat{x}$ as the estimate of x^* and satisfies $f(\hat{x}) - f(x^*) \le \epsilon^2$. For the case where n is white Gaussian noise with a known variance σ^2, the reconstruction result of the algorithm satisfies $\|x^* - \hat{x}\| \le \frac{\epsilon}{\sqrt{\tilde{\mu}_{2K}}} +$*

$\Delta_{x^*,K} + \Delta_{\hat{x},K} + \frac{4\sqrt{(1+a)K\log N}}{\tilde{\mu}_{2K}}\sigma$, *with probability exceeding $1 - (\sqrt{\pi(1+a)\log N} \cdot N^a)^{-1}$.*

Theorem 5. *Let us assume that the matrix Φ satisfies the RIP condition and that we have a reconstruction algorithm that returns $\hat{K}$-sparse reconstruction $\hat{x}$ as the estimate of x^* and satisfies $f(\hat{x}) - f(x^*) \le \epsilon^2$. For the case where n is white Gaussian noise with a known variance σ^2, the reconstruction result of the algorithm satisfies $\|x^* - \hat{x}\| \le \frac{\epsilon}{\sqrt{\tilde{\mu}_{2\hat{K}}}} + \Delta_{x^*,\hat{K}} + \frac{4\sqrt{(1+a)\hat{K}\log N}}{\tilde{\mu}_{2\hat{K}}}\sigma$, with probability exceeding $1 - (\sqrt{\pi(1+a)\log N} \cdot N^a)^{-1}$.*

The claims in Theorems 4 and 5 are a straightforward result of combining Lemma 1 with Theorems 2 and 3. Knowing ϵ and $\hat{K}$ for an algorithm, these theorems provide us with improved guarantees on its performance. By noting that $f(x^*) = \|n\|^2$ and using the concentration-of-measure, in a similar way to Lemma 1, one can better estimate ϵ by $\tilde{\epsilon}$ minus a lower bound on $\|n\|^2$.

We demonstrate the use of Theorems 4 and 5 for FLIHT and BPDN minimization techniques in the next section.

4. EXPERIMENTS

Exact sparsity example. Having a closer look on the guarantees achieved for FLIHT and BPDN minimization, it is interesting to check whether the constants in the bounds are tight or not. For this task we have simulated FLIHT and FISTA (which solves BPDN).

Figure 1 presents the error (averaging over 250 instances of the experiment) as a function of σ for FISTA (left) and FLIHT (right). The representation size is $N = 1024$ and the measured signal size is $m = 512$. The columns of Φ and the vector x^* are normalized and drawn from the canonical Gaussian distribution, where the support of x^* is selected uniformly at random with $K = 10$. The σ values range from high signal-to-noise ratio (SNR) to an SNR of 0dB. The theoretical bounds in the curves are calculated using Theorem 5. Since we are looking on the exact K-sparse case the value of Δ is zero for FLIHT. In the current experiment, FISTA's result is also K sparse and thus Δ is zero also for it. In addition, ϵ is negligible compared to the noise term for both techniques. Thus, we present only the noise term in the bound using $a = 0$. Approximations for $\tilde{\mu}_K$, $\tilde{\mu}_{2K}$ and $\tilde{L}_K$ are calculated using equations (3.1) and (3.2) from [14] as explained in Section 1.

In both figures it can be seen that the bounds are meaningful only for high SNR (small values of σ) since the test signals have unit energy. As the error bound goes above 1, a better estimator is the zero estimator. In addition, we observe that the FLIHT method error closely follows the theoretical and oracle bounds, while being significantly close to the latter. However, for FISTA, when the SNR approaches to 0dB the error curve stops behaving in a linear way. In FLIHT, because we have exploited the correct knowledge of K we do not observe this kind of behavior.

Compressible signal example. In the previous experiment the values of ϵ and Δ were negligible and we did not observe their effect on the bound. We turn to look on the case of a compressible signal. In this case we cannot neglect these terms. x^* is not an exact K-sparse signal but obey a certain decay law on its elements. Having additive noise on the measurement, the reconstruction algorithm should aim at reconstructing the K dominant elements in x^* and throw the rest since they will be lost by the noise. Choosing the right value of K is not an easy task. In the next experiment we see that using the upper bound, one can estimate the right value.

In this experiment the elements in x^* are generated using the generalized Pareto distribution [15]. For each element x_i^* we have

$$x_i^* = \epsilon_{\pm} r \left(U^{-\frac{1}{r}} - 1 \right), \tag{6}$$

where U is uniformly distributed $U[0,1]$ and $\epsilon_{\pm}$ is ± 1 with probability 0.5. We choose $r = 0.8$ and Φ is selected in the same way as in the previous experiment. The noise power is $\sigma = 100$ (the SNR varies from 6dB to 20dB depending on the signal strength) and FLIHT is applied with various values of K ranging from 1 to 100. Figure 2 presents the average ℓ_2 error of FLIHT and the oracle together with their theoretical upper bounds. The upper bound for FLIHT is calculated using Theorem 5 and is composed of three elements: ϵ, Δ and the noise term. The last is calculated in the same way as in the first experiment. ϵ is predicted using $\tilde{\epsilon}$ and Δ is estimated using the statistics of the signal x by

$$\|x^* - x_K^*\|_p \approx N^{1/r} \left(\frac{r}{p-r} \right)^{1/p} K^{-1/r+1/p}. \tag{7}$$

By looking on the figure we can have several observations. First, FLIHT has a similar behavior to the oracle when applied with small values of K. For larger values of K, it tends to miss more elements from the right support and thus get higher error. The theoretical upper bound of the oracle in the non-exact case is not tight as in the exact K-sparse case but still has a behavior that is similar to the oracle error. The last observation and the most important one is that FLIHT's upper bound has the same behavior as the real error providing a good prediction for selecting the right value of K for
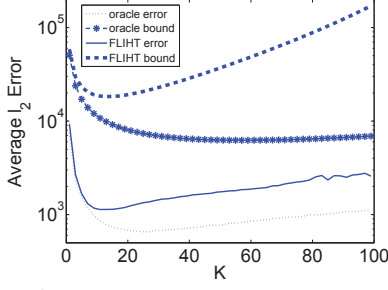
Fig. 2. Average ℓ_2 error of FLIHT and the oracle versus K compared to the theoretical bounds. The behavior of the FLIHT method bound predicts the value of K that minimizes the error in an accurate way.

FLIHT. This shows that the given bound can be used as a parameter selection tool when using a certain algorithm if we have a prior on the distribution of the unknown signal.

Note that our upper bounds are a constant factor off in predicting the actual behavior. A more focused analysis might improve these bounds and make them tighter. In particular, when we take a closer look at the 1-sparse solution of the compressible signal, the upper bound multiplies the actual error by approximately 3. This factor is due to the irrecoverable energy bound, which is obtained by a worse case analysis.

5. CONCLUSIONS

We presented a scheme for obtaining online bounds for reconstruction techniques. Our scheme is called online because the final guarantees depend on the data reconstruction errors. We demonstrate our bounds for two algorithms: FLIHT and BPDN. Simulation results show that our bounds are currently not tight for typical, small-scale signal realizations, while they are much tighter than the recovery guarantees established for the algorithms prior to execution. However, in practice the given bounds can help in parameter selection for a given algorithm.

A. PROOF OF THEOREM 2 AND 3

We begin with the proof of Theorem 2. Using the triangle inequality and Proposition 3.5 from [5] we have:

$$
\begin{aligned}
f(\hat{x}) - f(x^*) &= \|u - \Phi\hat{x}\|^2 - \|u - \Phi x^*\|^2 \\
&= \|\Phi x^* + n - \Phi\hat{x}\|^2 - \|n\|^2 \\
&= \|\Phi(x_K^* - \hat{x}_K + x^* - x_K^* - \hat{x} + \hat{x}_K) + n\|^2 - \|n\|^2 \\
&\geq \|\Phi(x_K^* - \hat{x}_K) + n\|^2 - \|n\|^2 - \|\Phi(\hat{x} - \hat{x}_K)\|^2 \\
&- \|\Phi(x^* - x_K^*)\|^2 \geq \|\Phi(x_K^* - \hat{x}_K)\|^2 \\
&\qquad - n^T \Phi(x_K^* - \hat{x}_K) - \tilde{L}_K \nu_{x^*}^2 - \tilde{L}_K \nu_{\hat{x}}^2. \quad (8)
\end{aligned}
$$

Combining the above result with $f(\hat{x}) - f(x^*) \leq \epsilon^2$ leads to:

$$
\|\Phi(x_K^* - \hat{x}_K)\|^2 \leq \epsilon^2 + \tilde{L}_K \nu_{x^*}^2 + \tilde{L}_K \nu_{\hat{x}}^2 + 2n^T \Phi(x_K^* - \hat{x}_K). \quad (9)
$$

Using the RIP $\tilde{\mu}_{2K}\|x_K^* - \hat{x}_K\|^2 \leq \|\Phi(x_K^* - \hat{x}_K)\|^2$. Since $\Phi(x_K^* - \hat{x}_K) = \Phi_{x_K^* - \hat{x}_K}(x_K^* - \hat{x}_K)$ we get from (9)

$$
\begin{aligned}
\|x_K^* - \hat{x}_K\|^2 &\leq \frac{\epsilon^2 + \tilde{L}_K(\nu_{x^*}^2 + \nu_{\hat{x}}^2)}{\tilde{\mu}_{2K}} \\
&+ \frac{2}{\tilde{\mu}_{2K}}\left\|\Phi_{x_K^* - \hat{x}_K}^T n\right\| \|x_K^* - \hat{x}_K\|. \quad (10)
\end{aligned}
$$

Based on the observation that $a^2 \leq c^2 + 2ba$ yields $a \leq c + 2b$, and that $\sqrt{a^2 + b^2} < a + b$ we have

$$
\|x_K^* - \hat{x}_K\| \leq \frac{\epsilon + \sqrt{\tilde{L}_K}(\nu_{x^*} + \nu_{\hat{x}})}{\sqrt{\tilde{\mu}_{2K}}} + \frac{2}{\tilde{\mu}_{2K}}\left\|\Phi_{x_K^* - \hat{x}_K}^T n\right\|. \quad (11)
$$

The desired result is achieved by the fact that $\|x^* - \hat{x}\| \leq \|x_K^* - \hat{x}_K\| + \|x^* - x_K^*\| + \|\hat{x} - \hat{x}_K\|$ and the definition of $\Delta_{x,K}$. $\qquad\square$

The proof of Theorem 3 is straightforward by using the result of Theorem 2 and looking at x^* as a $\hat{K}$-sparse vector.

B. REFERENCES

[1] 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.

[2] M. Rudelson and R. Vershynin, "Sparse reconstruction by convex relaxation: Fourier and gaussian measurements," in *Information Sciences and Systems, 2006 40th Annual Conference on*, 22-24 2006, pp. 207 –212.

[3] H. Rauhut, "Compressive sensing and structured random matrices," *Theoretical Foundations and Numerical Methods for Sparse Recovery*, 2010.

[4] A. Wigderson and D. Zuckerman, "Expanders that beat the eigenvalue bound: Explicit construction and applications," *Combinatorica*, vol. 19, no. 1, pp. 125–138, 1999.

[5] 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, 2009.

[6] 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.

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

[8] 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.

[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.

[12] V. Cevher, "An ALPS view of sparse recovery," in *ICASSP*, 2011.

[13] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Img. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[14] E.J. Candès and T. Tao, "Decoding by linear programming," vol. 51, no. 12, pp. 4203 – 4215, dec. 2005.

[15] V. Cevher, "Learning with compressible priors," in *NIPS*, Vancouver, B.C., Canada, 7–9 December 2009.