ADAPTIVE SENSING FOR SPARSE SIGNAL RECOVERY

Jarvis Haupt, Robert Nowak[*]

University of Wisconsin—Madison
Dept. of Electrical and Computer Engineering

Rui Castro

Columbia University
Department of Electrical Engineering

ABSTRACT

The theory of compressed sensing shows that sparse signals in high-dimensional spaces can be recovered from a relatively small number of samples in the form of random projections. However, in severely resource-constrained settings even CS techniques may fail, and thus, a less aggressive goal of partial signal recovery is reasonable. This paper describes a simple *data-adaptive* procedure that efficiently utilizes information from previous observations to focus subsequent measurements into subspaces that are increasingly likely to contain true signal components. The procedure is analyzed in a simple setting, and more generally, shown experimentally to be more effective than methods based on traditional (non-adaptive) random projections for partial signal recovery.

Index Terms— sparsity, compressed sensing, signal detection and estimation, adaptive sampling

1. INTRODUCTION

The theory of *compressed sensing* (CS) has had a profound effect on our understanding of signal acquisition and sampling, and is currently at the forefront of signal processing research. Consider a signal (vector) that has a sparse representation (i.e., can be written as a linear combination of a small number of basis vectors). The basic idea of compressed sensing is that if one takes *non-adaptive* samples (samples that do not depend on the actual signal being observed) in the form of projections of the signal onto a set of test vectors, and if the test vectors are *incoherent* with the basis vectors comprising the basis in which the signal is sparse, then the signal representation can be recovered exactly from a relatively small number of such samples, roughly proportional to the number of components in the sparse representation [1, 2]. In addition, compressed sensing remains stable in the presence of random noise; i.e., the recovery degrades gracefully, but markedly, as the noise level is increased [3, 4].

Despite its recent popularity and notable success in noiseless settings, traditional CS methods do suffer several drawbacks. First, in noisy settings, using incoherent projections leads to an inherent reduction in measurement signal-to-noise ratio (SNR) relative to what could be attained using directly focused samples (such as traditional point samples). Incoherence essentially ensures that each projection observation provides some information about the unknown signal, and this is accomplished by allocating an equal fraction of "sensing power" to every dimension of the signal space. While this approach is sensible in noise-free settings, the spreading of sensing power necessarily results in a reduction in the SNR *per dimension* in noisy settings.

A second, more subtle, limitation of CS arises in settings where the acquisition process is severely resource-constrained. Despite the tremendous reduction in the number of observations required for provable success of CS methods, there are some application domains in which even the CS requirements are too restrictive. For example, in settings where the signal of interest is time-varying, there may not be sufficient time to collect a enough CS observations to obtain a useful estimate of the entire signal. Another scenario where this restriction could be evident is in rapid-detection settings, where the goal is to identify a subset of the true signal support using as few observations as possible.

In this paper we explore a method that addresses both of these limitations simultaneously through the use of adaptivity in sampling (making subsequent observations using knowledge of the specific signal of interest gleaned from prior observations). In general, it is easy to envision a sampling procedure that begins with incoherent sampling, and utilizes some feedback information from each observation to focus subsequent observations more directly into the signal subspace. Indeed, as more information is obtained from the observations, subsequent samples could be focused more directly onto the relevant signal components, providing a significant improvement in measurement SNR. In addition, if the feedback process quickly focuses in on the actual signal support, only a few observations need be collected.

The potential advantages of adaptive projection shaping schemes for sparse signal recovery were examined in several recent works. The first published approach was the fully-Bayesian procedure proposed in [5] (and later extended in [6]). While shown to be effective in practice, the computational complexity of these projection shaping procedures limit their application to problems of relatively small dimension. A more computationally efficient Bayesian algorithm was pro-

[*]This work was partially supported by NSF grants ECS-0529381 and CCF-0353079.

posed in [7]. Absent from these prior works was any theoretical evaluation of the performance of the algorithms, which is prohibitive due to the complicated statistical dependencies among observations.

Here we propose a simple adaptive projection shaping procedure that is amenable to direct analysis in simple settings, and which leverages favorable implementation features of each of the previously proposed adaptive methods. Like the adaptive procedure described in [6], the method proposed here does not require explicit knowledge of the observation noise power (though the performance will, of course, depend on it). In addition, as with the procedure in [7], the proposed procedure enjoys the benefit of lower implementation complexity than the methods of [5, 6].

The remainder of the paper is organized as follows. Following a brief formalization of the observation model in Section 2, we propose our adaptive procedure in Section 3, and we provide a theoretical analysis of its performance in a simple setting. Section 4 provides a more comprehensive (experimental) validation of the proposed procedure, demonstrating its improvement relative to methods that utilize non-adaptive random projections, for the task of partial signal support recovery. Conclusions are discussed in Section 5.

2. OBSERVATION MODEL

Let $\boldsymbol{x} \in \mathbb{R}^n$ denote the unknown signal of interest, which will be our object of interest throughout the paper. We assume the signal $\boldsymbol{x}$ is sparse, meaning that most of its entries are zero. Observations of the signal are modeled as noisy projections onto a set of user-specified test vectors. Formally, the observations are given by

$$y_i = \boldsymbol{a}_i^T \boldsymbol{x} + \eta_i, \;\; i = 1, \ldots, k \tag{1}$$

where $\{\boldsymbol{a}_i\}_{i=1}^k$ are the test vectors, and η_i are independent and identically distributed (i.i.d.) normal random variables with zero mean and variance σ^2. In addition, we assume that each test vector satisfies $\|\boldsymbol{a}_i\|_2^2 = \sum_{j=1}^n a_i^2(j) = 1$, essentially limiting the amount of "sensing energy" available for each projection observation. At times, we will find it useful to express the observation model in matrix form

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\eta}, \tag{2}$$

where $\boldsymbol{y}$ is $k \times 1$, $\boldsymbol{A}$ is a $k \times n$ matrix whose rows are the test vectors $\{\boldsymbol{a}_i\}_{i=1}^k$, and $\boldsymbol{\eta}$ is the $k \times 1$ vector whose entries are the noises $\{\eta_i\}_{i=1}^k$.

Traditional CS methods utilize non-adaptive collections of test vectors $\{\boldsymbol{a}_i\}_{i=1}^k$ whose entries are incoherent with the basis vectors of the basis in which the signal is sparse. Early works in CS established that incoherence is not difficult to attain in practice—many random vectors, such as vectors whose entries are independent and identically distributed (iid) Gaussian or symmetric Bernoulli random variables are, with high probability, incoherent with any fixed orthonormal basis [2, 8, 9, 10]. In contrast, the sequential procedure proposed here provides more flexibility, and allows the test vectors to be chosen adaptively based on prior test vectors and observations. In other words, we allow $\boldsymbol{a}_i$ to depend on $\{\boldsymbol{a}_j, y_j\}_{j<i}$.

3. SEQUENTIAL PROJECTION FOCUSING

Our proposed algorithm is a simple iterative refocusing scheme. The observations at each step of a given iteration are reminiscent of traditional CS observations, but focused only on a subset of indices of the original signal. A simple testing procedure reduces the size of the subset of interest at each step, allowing sensing energy to be more directly focused in subsequent steps. A pseudocode description of the algorithm appears as Algorithm 1.

Algorithm 1: Sequential Projection Focusing.

Input:
Polarity $\rho \in \{-1, +1\}$;
Block size $b \in \mathbb{N}$;

Initialize:
Index set $I^{(1)} \longleftarrow \{1, 2, \ldots, n\}$;

for $\ell = 1$ **to** $\lceil \log_2 n \rceil$ **do**

Generate $b \times n$ sign matrix Υ;

$$A^{(\ell)}(i,j) = \begin{cases} \sqrt{\frac{1}{|I^{(\ell)}|}}\Upsilon(i,j) & i = 1, \ldots, b, \; j \in I^{(\ell)} \\ 0, & i = 1, \ldots, b, \; j \in \left(I^{(\ell)}\right)^c \end{cases};$$

$$\boldsymbol{y}^{(\ell)} = \boldsymbol{A}^{(\ell)}\boldsymbol{x} + \boldsymbol{\eta}^{(\ell)};$$

$$\widehat{\boldsymbol{x}}^{(\ell)} = \left(\boldsymbol{A}^{(\ell)}\right)^T \boldsymbol{y}^{(\ell)};$$

$$I^{(new)} \longleftarrow \{j : \rho \cdot \widehat{x}^{(\ell)}(j) > 0\};$$

if $|I^{(new)}| > 0$ **then**
$I^{(\ell+1)} \longleftarrow I^{(new)}$;
else
$I^{(\ell+1)} \longleftarrow I^{(\ell)}$;
end

end

Output:
Support estimate $\widehat{I} = \max_j \;\; \rho \cdot \widehat{x}^{(\lceil \log_2 n \rceil)}(j)$

In words, the algorithm proceeds as follows. Begin by specifying the polarity $\rho \in \{-1, +1\}$ of the signal com-

ponents to search for ($\rho = -1$ implies searching for negative signal entries, and $\rho = +1$ implies searching for positive entries), and an integer $b \in \mathbb{N}$ (the block size) which is the number of observations to collect for each step. In the first step, obtain a collection of b observations, each of which corresponds to projecting the unknown signal onto a vector whose amplitude is $1/n$ in each of the n signal dimensions and whose sign is given by the entries of the sign matrix Υ generated for that step.

Next, compute the back-projection $\widehat{\boldsymbol{x}}^{(1)} \in \mathbb{R}^n$, which is a vector whose entries quantify the correlation between the current observation vector $\boldsymbol{y}^{(1)}$ and the corresponding column of the current observation matrix $\boldsymbol{A}^{(1)}$ (strong positive correlations suggest a positive signal entry at that vector index, and likewise a strong negative correlation suggests a negative entry). Refine the set of indices of interest by retaining only the indices of $\widehat{\boldsymbol{x}}^{(1)}$ having correlations that are sign-consistent with the specified polarity.

In the following step, collect another b observations, but this time, only observe the dimensions corresponding to the indices of interest from the previous step. Eliminating some of the indices allows for an increase in the amount of energy that can be allotted to the remaining indices of interest. This is quantified by the scaling of the nonzero entries of $\boldsymbol{A}^{(\ell)}$—as the size of the set of indices of interest decreases, the sensing energy per index of interest increases. Continue in this fashion until the size of the set of indices of interest reduces to one. Because the specification of the sign matrix Υ is rather general, the 'if' statement is required to guarantee that the set of indices of interest remains non-empty. Finally, the output of the procedure is the index of the strongest sign-consistent correlation present in the final back projection vector $\widehat{\boldsymbol{x}}^{(\lceil \log_2 n \rceil)}$. In the case where the maximum is not unique, the tie can be broken deterministically.

At a high level, note that the procedure essentially amounts to applying a naïve reconstruction approach to the collection of observations at each step, then using that coarse signal estimate to direct subsequent measurements. The key feature is that rather than attempt a full reconstruction at each step (by retaining a small set of indices that are most likely to correspond to signal components), the simple thresholding test instead *rejects* a relatively large subset of indices at which signal components are least likely to be present. In other words, rather than requiring the entries of the back projection at indices where signal components are present to exhibit the *strongest* correlations, we only require that they are not among the *weakest*. Thus, each refinement is relatively forgiving, and the true components can be more easily identified in subsequent steps because of the increasing SNR per dimension.

Notice that the procedure as stated returns a single index corresponding to a likely location of a nonzero signal component. To recover multiple entries of the signal, the procedure could be performed multiple times (perhaps with differing values of ρ, to identify signal components with both positive and negative signs). If, after each run, the previously identified indices are removed from consideration, this multi-step procedure corresponds to a greedy method of identifying several likely indices corresponding to nonzero signal entries.

3.1. Complexity Analysis

The complexity of the proposed procedure is governed by the complexity of the matrix creation and matrix-vector multiplications at each step, both of which are $O(bn)$ operations. As a result, one implementation of the proposed procedure in this setting requires $O(bn \log n)$ operations, and the multi-step process requires $O(mbn \log n)$ operations to identify m components. Since the total number of observations is $k = mb \log n$, the general complexity of the proposed approach is $O(kn)$.

For comparison, we note that the adaptive procedure proposed in [6] requires $O(n^3)$ operations, while a "fast version" of the same algorithm requires $O(k^2 n)$ operations. On the other hand, the adaptive procedure proposed in [7] requires n simple parameter updates following each observation, resulting in an overall complexity of $O(kn)$ operations. Thus, like the procedure proposed in [7], the procedure described here enjoys the benefits of low computational complexity, making it a viable option in high-dimensional settings.

3.2. Performance Analysis

Our adaptive procedure is amenable to direct analysis in a simple setting. Consider a signal $\boldsymbol{x} \in \mathbb{R}^n$ consisting of a single nonzero entry of amplitude 1, assume that the signal dimension n is a power of 2, and let $b = 1$. In this case, the sign matrix can be chosen so that the set of indices of interest is *exactly* bisected at each step. For example, this could be accomplished by choosing the signs of the nonzero entries of the test vector at each step to be the entries of any vector with half of the entries equal to $+1$ and half equal to -1. For this simplified procedure, we have the following result.

Theorem 1. *When the adaptive procedure (using the specially-chosen sign pattern described above) is applied to a sparse signal $\boldsymbol{x} \in \mathbb{R}^n$ (where n is a power of 2) having a single nonzero entry of amplitude 1, then the index returned corresponds to the true location of the nonzero entry with probability*

$$P^a = \prod_{\ell=1}^{\log_2 n} \Phi\left(\frac{\sqrt{2^{\ell-1}}}{\sigma\sqrt{n}}\right), \tag{3}$$

where $\Phi(z)$ is the standard normal CDF.

Proof. To establish the claim, we let j^* denote the location of the single nonzero entry of $\boldsymbol{x}$. Model each iteration of the procedure as a Bernoulli test with success probability P_ℓ. Because the noise is independent for each observation, the prob-

ability of identifying the true index after $\log_2 n$ steps is simply given by the product of the probabilities of successfully retaining the location of the nonzero entry at each refinement step, conditioned on the event that the true location is still in the set of indices of interest at that step.

Given successful refinements at steps $1, 2, \ldots, \ell-1$, at step ℓ the observation is distributed as $y^{(\ell)} \sim \mathcal{N}\left(a^{(\ell)}(j^*), \sigma^2\right)$, where $|a^{(\ell)}| = 2^{\ell-1}/n$. Regardless the actual sign of $a^{(\ell)}(j^*)$, the probability that the observation is sign-consistent with the entry of the projection vector at index j^* is equivalent to the probability that $y^{(\ell)}$ exceeds zero, which is given by

$$\begin{aligned}\Pr\left(\text{Refine correctly at step } \ell\right) &= P_\ell(n,\sigma) \\ &= \int_{-\infty}^{\sqrt{2^{\ell-1}/n}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-z^2}{2\sigma^2}\right) dz \\ &= \Phi\left(\frac{\sqrt{2^{\ell-1}}}{\sigma\sqrt{n}}\right).\end{aligned}$$

Overall, we have that the probability of finding the spike in one iteration of the procedure is

$$P^a = \prod_{\ell=1}^{\log_2 n} P_\ell(n,\sigma) = \prod_{\ell=1}^{\log_2 n} \Phi\left(\frac{\sqrt{2^{\ell-1}}}{\sigma\sqrt{n}}\right).$$

□

The same technique can be applied using passive observations, where the set of indices of interest is refined as above but the observation energy is not focused. Applying the same reasoning as above, it is easy to see that the probability of success in this case is given by

$$P^p = \prod_{\ell=1}^{\log_2 n} \Phi\left(\frac{1}{\sigma\sqrt{n}}\right).$$

Notice that the relative benefit of focusing observation energy is evident in the numerator of the argument in the Φ function, and it is clear that the success probability for the adaptive procedure is always greater than the success of the same procedure using passive observations.

This simple result can be extended to a slightly more general setting ($b > 1$) simply by repeating the bisecting measurement for each observation in the block. The net effect is a reduction in the noise variance. We give the result here as a corollary.

Corollary 1. *If for each block of observations, the bisecting sign pattern observations (described above) are repeated for each of $b \geq 1$ measurements, the proposed adaptive procedure correctly identifies the location of the single positive nonzero entry with probability*

$$P^a(b) = \prod_{\ell=1}^{\log_2 n} \Phi\left(\frac{\sqrt{2^{\ell-1}b}}{\sigma\sqrt{n}}\right), \tag{4}$$

where again, $\Phi(z)$ is the standard normal CDF.

Proof. The analysis is similar to the proof of Theorem 1, but now, the noise level is reduced by a factor of b because of the averaging of identical measurements. The result follows. □

The presence of additional nonzero components complicates the relatively simple theoretical analysis presented in the proof of Theorem 1. To compare the performance of the proposed procedure to traditional CS approaches for partial support recovery in these more general settings, we provide a detailed experimental evaluation in the next section.

4. EXPERIMENTAL EVALUATION

To illustrate the effectiveness of the proposed adaptive procedure relative to traditional CS methods, we compare the performance of each in a variety of settings corresponding to a range of algorithm parameters, noise levels, and signal types. In the following simulations, we use a randomly generated Υ for the adaptive procedure, in which each entry of the sign matrix is independently ± 1 with probability $1/2$.

4.1. Varying Block Size

We begin by considering the effect of varying the block size b. Our test signals $\boldsymbol{x} \in \mathbb{R}^n$ ($n = 2^{10}$) are constructed to have 15 nonzero (positive) entries of equal amplitude (at randomly selected locations), and we normalize so that $\|\boldsymbol{x}\|_2^2 = 1$. For several fixed values of SNR ($S = 1/n\sigma^2$), we compare the proposed procedure with one step of Orthogonal Matching Pursuit (OMP) [11] applied to a collection of $b\log_2 n$ traditional CS measurements using test vectors whose entries are (appropriately scaled) random symmetric Bernoulli random variables. Recall that the index selection step of OMP identifies the index j at which the absolute value of the back-projection $\widehat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{x}$ is maximized.

The empirical probabilities of successfully identifying one of the true support entries (averaged over 200 trials) for four different values of S ($S = 0.01$, $S = 0.1$, $S = 0.5$, and $S = 1$) are shown in Figure 1. Notice that, while OMP (dotted line) performs better than random guessing (which succeeds with probability $1/2^{10}$), the proposed active procedure (solid line) consistently and significantly outperforms OMP across the entire range of block sizes examined.

4.2. Varying Noise Level

Next we consider the effect of varying noise level. As above, the test signals $\boldsymbol{x} \in \mathbb{R}^n$ ($n = 2^{10}$) are constructed with 15 nonzero (positive) entries of equal amplitude (at randomly selected locations), and we normalize so that $\|\boldsymbol{x}\|_2^2 = 1$. For several fixed values of block size b, we compare the proposed procedure with one step of (OMP) applied to a collection of $b\log_2 n$ traditional CS measurements.

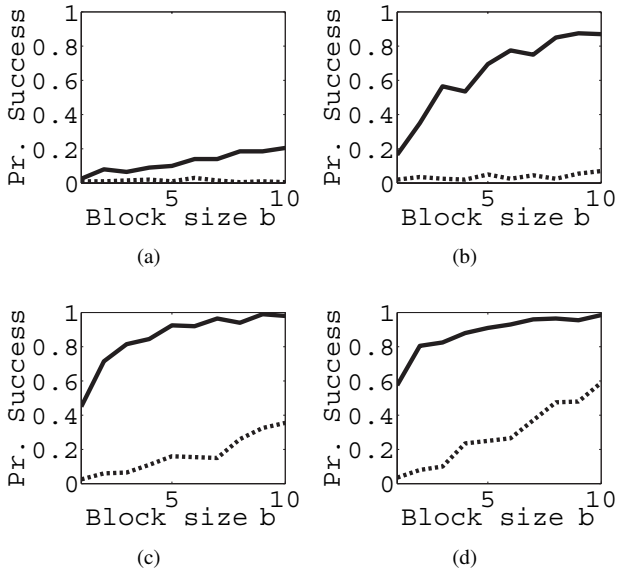

Fig. 1. Empirical probabilities of successfully identifying one entry of the signal support for the proposed adaptive procedure (solid line) and OMP (dotted line), as a function of the block size b. Each panel corresponds to a fixed SNR ($S = 0.01$, $S = 0.1$, $S = 0.5$, and $S = 1$ in panels (a)-(d), respectively).

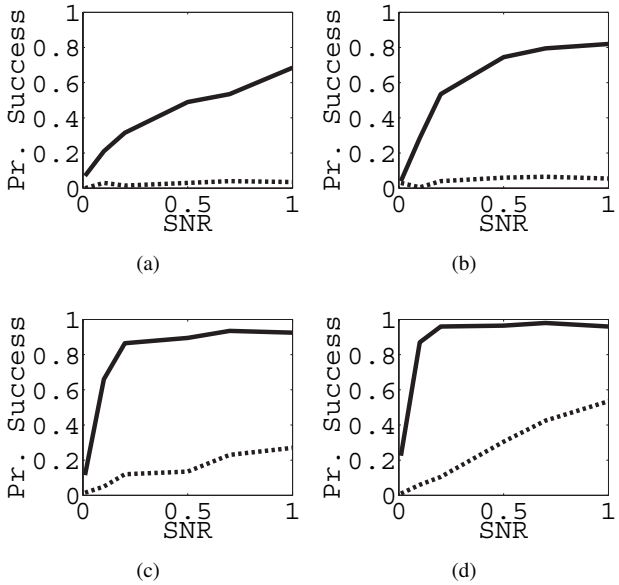

Fig. 2. Empirical probabilities of successfully identifying one entry of the signal support for the proposed adaptive procedure (solid line) and OMP (dotted line), as a function of the SNR. Each panel corresponds to a fixed block size b ($b = 1$, $b = 2$, $b = 5$, and $b = 10$ in panels (a)-(d), respectively).

The empirical probabilities of success as a function of S for each method, again averaged over 200 trials, are shown in Figure 2 for four different values of b ($b = 1$, $b = 2$, $b = 5$, and $b = 10$). Again, the proposed procedure consistently outperforms the traditional CS method, and as expected, increasing the block size makes the procedure more effective.

4.3. Varying Sparsity Level

Here we evaluate the performance of the proposed procedure as a function of the sparsity level of the signal, for the nominal value $b = 1$, for four different SNR levels. For the same signal model as above, the empirical probabilities of success as a function of sparsity level for each method, again averaged over 200 trials, are shown in Figure 3(a)-(d) for $S = 0.01$, $S = 0.1$, $S = 0.5$, and $S = 1$, respectively.

As expected the performance improves with increasing SNR, and the adaptive procedure generally outperforms the nonadaptive method for any sparsity level. However, when the signal becomes very non-sparse, the problem of finding one entry of the signal becomes relatively easy and both methods perform well in this case.

4.4. Signals with Positive and Negative Entries

As a final test, we apply the proposed procedure to recover entries of a signal having positive and negative entries, and for which the nonzero entries may differ in amplitude. For each of 200 trials, we execute the proposed procedure twice (once each for $\rho = \pm 1$), identifying two potential signal support entries. We term each trial a success if at least one of the signal entries identified corresponds to the location of a nonzero signal entry. For comparison, we consider a non-adaptive approach comprising two steps of the OMP recovery procedure (for the same number of observations as two executions of the adaptive procedure, $2b \log_2 n$).

As above, the signals of interest $\boldsymbol{x} \in \mathbb{R}^n$ ($n = 2^{10}$) have 15 nonzero entries and satisfy $\|\boldsymbol{x}\|_2^2 = 1$, but the sign of each entry is selected randomly and the amplitudes exhibit a polynomial decay (the amplitudes of the nonzero entries are proportional to $p^{-1/2}$, for $p = 1, 2, \ldots, 15$). We consider varying block sizes b, and compare the performance of each method for four different values of SNR. The results are shown in Figure 4. It is clear that the adaptive procedure still outperforms OMP, and remains an effective approach for partial support recovery in this more general setting.

5. CONCLUSIONS

We have demonstrated that the adaptive procedure proposed in this paper is consistently more effective than traditional CS methods for partial support recovery, making it a more viable option than CS in severely resource-constrained applications. We also note that the simplicity of the feedback adaptivity in the proposed procedure alleviates the difficulty that often accompanies the analysis of adaptive procedures—the pres-

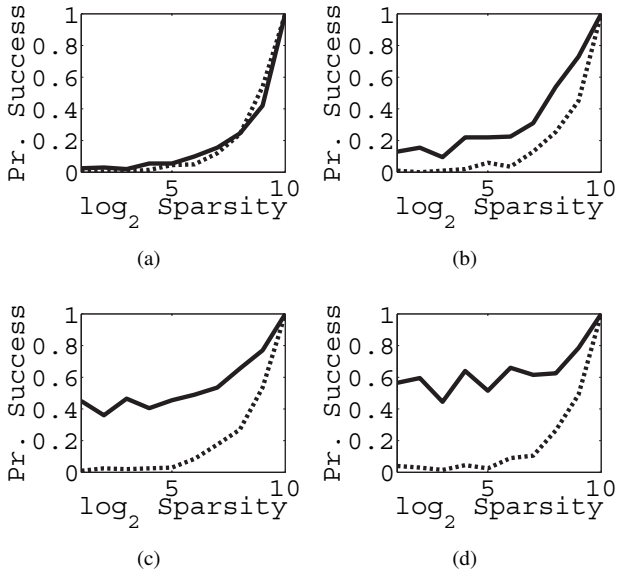

Fig. 3. Empirical probabilities of successfully identifying one entry of the signal support for the proposed adaptive procedure (solid line) and OMP (dotted line), as a function of the sparsity level. Each panel corresponds to a fixed SNR ($S = 0.01$, $S = 0.1$, $S = 0.5$, and $S = 1$ in panels (a)-(d), respectively).

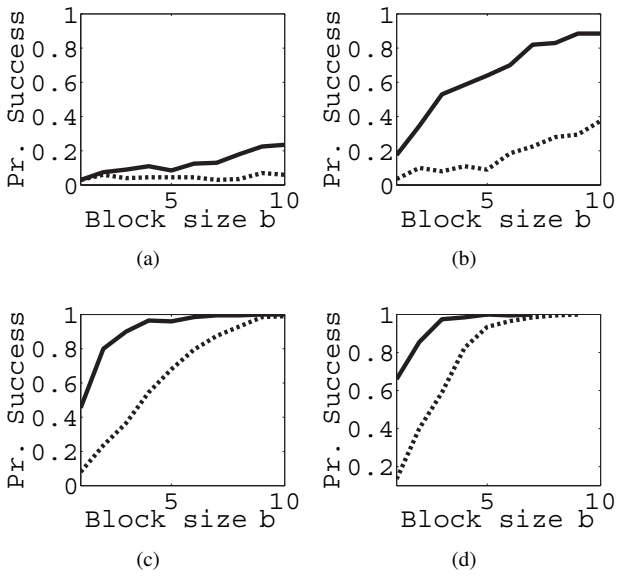

Fig. 4. Empirical probabilities of successfully identifying one entry of the signal support for the proposed adaptive procedure (solid line) and OMP (dotted line), as a function of the block size, for signals having both positive and negative entries. Each panel corresponds to a fixed SNR ($S = 0.01$, $S = 0.1$, $S = 0.5$, and $S = 1$ in panels (a)-(d), respectively).

ence of complicated statistical dependence among observations. This makes our algorithm amenable to direct analysis in simple settings, and possibly sheds new light on the potential for very simple adaptive sampling algorithms to provide significant provable performance improvements.

6. REFERENCES

[1] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 4036–4048, Sept. 2006.

[4] E. Candès and T. Tao, "The Dantzig selector: statistical estimation when p is much larger than n," *Annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, Dec. 2007.

[5] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Proc.*, vol. 56, no. 6, pp. 2346–2356, June 2008.

[6] S. Ji, D. Dunson, and L. Carin, "Multi-task compressive sensing," *IEEE Trans. Signal Proc.*, submitted.

[7] R. Castro, J. Haupt, R. Nowak, and G. Raz, "Finding needles in noisy haystacks," in *Proc. IEEE Intl. Conf. Acoustics, Speech and Sig. Proc.*, Las Vegas, NV, Apr. 2008, pp. 5133–5136.

[8] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[9] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[10] R. Baraniuk, M. Davenport, R. A. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," in *Constructive Approximation*. Springer, New York, 2008.

[11] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. on Signals, Systems, and Computers*, Nov. 1993.