

Efficient Adaptive Compressive Sensing Using Sparse Hierarchical Learned Dictionaries

Akshay Soni and Jarvis Haupt
 University of Minnesota, Twin Cities
 Department of Electrical and Computer Engineering
 Minneapolis, Minnesota USA 55455
 e-mail: {sonix022, jdhaupt}@umn.edu

Abstract—Recent breakthrough results in compressed sensing (CS) have established that many high dimensional objects can be accurately recovered from a relatively small number of non-adaptive linear projection observations, provided that the objects possess a sparse representation in some basis. Subsequent efforts have shown that the performance of CS can be improved by exploiting the structure in the location of the non-zero signal coefficients (structured sparsity) or using some form of online measurement focusing (adaptivity) in the sensing process. In this paper we examine a powerful hybrid of these two techniques. First, we describe a simple adaptive sensing procedure and show that it is a provably effective method for acquiring sparse signals that exhibit structured sparsity characterized by tree-based coefficient dependencies. Next, employing techniques from sparse hierarchical dictionary learning, we show that representations exhibiting the appropriate form of structured sparsity can be learned from collections of training data. The combination of these techniques results in an effective and efficient adaptive compressive acquisition procedure.

I. INTRODUCTION

Motivated in large part by breakthrough results in compressed sensing (CS), significant attention has been focused in recent years on the development and analysis of sampling and inference methods that make efficient use of measurement resources. The essential idea underlying many directions of research in this area is that signals of interest often possess a parsimonious representation in some basis or frame. For example, let $x \in \mathbb{C}^n$ be a (perhaps very high dimensional) vector which denotes our signal of interest. Suppose that for some fixed (known) matrix D whose columns are n -dimensional vectors d_i , x may be expressed as a linear combination of the columns of D , as

$$x = \sum_i \alpha_i d_i, \quad (1)$$

where the α_i are the coefficients corresponding to the relative weight of the contribution of each of the d_i in the representation. The dictionary D may, for example, consist of all of the columns of an orthonormal matrix (eg., a discrete wavelet or Fourier transform matrix), though other representations may be possible (eg., D may be a frame). In any case, we define the *support set* \mathcal{S} to be the set of indices corresponding to the nonzero values of α_i in the representation of x . When $|\mathcal{S}|$ is small relative to the ambient dimension n , we say that the

signal x is *sparse* in the dictionary D , and we call the vector α , whose entries are the coefficients α_i , the sparse representation of x in the dictionary D .

The most general CS observation model prescribes collecting (noisy) linear measurements of x in the form of projections of x onto a set of $m (< n)$ “test vectors” ϕ_i . Formally, these measurements can be expressed as

$$y_i = \phi_i^T x + w_i, \quad i = 1, 2, \dots, m, \quad (2)$$

where w_i denotes the additive measurement uncertainty associated with the i th measurement. In “classic” CS settings, the measurements are *non-adaptive* in nature, meaning that the $\{\phi_i\}$ are specified independently of $\{y_i\}$ (eg., the test vectors can be specified before any measurements are obtained). Initial breakthrough results in CS establish that for certain choices of the test vectors, or equivalently the matrix Φ whose rows are the test vectors, sparse vectors x can be exactly recovered (or accurately approximated) from $m \ll n$ measurements. For example, if x has no more than k nonzero entries, and the entries of the test vectors/matrix are chosen as iid realizations of zero-mean random variables having sub-Gaussian distributions, then only $m = O(k \log(n/k))$ measurements of the form (2) suffice to exactly recover (if noise free) or accurately estimate (when $w_i \neq 0$) the unknown vector x , with high probability [1], [2].

Several extensions to the traditional CS paradigm have been investigated recently in the literature. One such extension corresponds to exploiting additional *structure* that may be present in the sparse representation of x , which can be quantified as follows. Suppose that $\alpha \in \mathbb{R}^p$, the sparse representation of x in an $n \times p$ orthonormal dictionary D , has k nonzero entries. Then, there are generally $\binom{p}{k}$ possible subspaces on which x could be supported, and the space of all k -sparse vectors can be understood as a *union of (k -dimensional) linear subspaces* [3]. *Structured sparsity* refers to sparse representations that are drawn from a *restricted* union of subspaces (where only a subset of the $\binom{p}{k}$ subspaces are allowable). Recent works exploiting structured sparsity in CS reconstruction include [4], [5]. One particular example of structured sparsity, which will be our primary focus here, is *tree-sparsity*. Let $\mathcal{T}_{p,d}$ denote a balanced rooted connected tree of degree d with p nodes. Suppose that the components of a sparse representation $\alpha \in \mathbb{R}^p$ can be put into a one-to-one correspondence with the nodes of the tree $\mathcal{T}_{p,d}$. We say that the vector $\alpha \in \mathbb{R}^p$ is *k -tree-sparse* in

the tree $\mathcal{T}_{p,d}$ if its nonzero components correspond to a rooted connected subtree of $\mathcal{T}_{p,d}$. This type of tree structure arises, for example, in the wavelet coefficients of many natural images [6].

Another extension to the “classic” CS observation model is to allow additional flexibility in the *measurement* process in the form of feedback. Sequential *adaptive sensing* strategies are those for which subsequent test vectors $\{\phi_i\}_{i \geq j}$ may explicitly depend on (or be a function of) past measurements and test vectors $\{\phi_l, y_l\}_{l < j}$. Adaptive CS procedures have been shown to provide an improved resilience to noise relative to traditional CS – see, for example, [7]–[9], as well as the summary article [10] and the references therein. The essential idea of these sequential procedures is to gradually “steer” measurements towards the subspace in which signal x resides, in an effort to increase the signal to noise ratio (SNR) of each measurement.

In this paper we examine a hybrid technique to exploit structured sparsity and adaptivity in the context of noisy compressed sensing. Adaptive sensing techniques that exploit the hierarchical tree-structured dependencies present in wavelet representations of images have been examined in the context of non-Fourier encoding in magnetic resonance imaging [11], and more recently in the context of compressed sensing for imaging [12]. Our first contribution here is to quantify the performance of such procedures when measurements are corrupted by zero-mean additive white Gaussian measurement noise. Our main theoretical results establish sufficient conditions (in terms of the number of measurements required, and the minimum amplitude of the nonzero components) under which the support of tree-sparse vectors may be exactly recovered (with high probability) using these adaptive sensing techniques. Our results stand in stark contrast with existing results for support recovery for (generally unstructured) sparse vectors, highlighting the significant improvements that can be achieved by the intelligent exploitation of structure throughout the measurement process.

Further, we demonstrate that tree-based adaptive compressed sensing strategies can be applied with representations learned from a collection of *training data* using recent techniques in hierarchical dictionary learning. This procedure of learning structured sparse representations gives rise to a powerful general-purpose sensing and reconstruction method, which we refer to as **Learning Adaptive Sensing Representations**, or **LASeR**. We demonstrate the performance improvements that may be achieved via this approach, relative to other compressed sensing methods.

The remainder of this paper is organized as follows. Section II provides a discussion of the top down adaptive compressed sensing procedure motivated by the approaches in [11], [12], and contains our main theoretical results which quantify the performance of such approaches in noisy settings. In Section III we discuss the LASeR approach for extending this adaptive compressed sensing idea to general compressed sensing applications using recent techniques in dictionary learning. The performance of the LASeR procedure is evaluated in Section IV, and conclusions and directions for future work are discussed in Section V. Finally, a sketch of

the proof of our main result is provided in Section VI.

II. ADAPTIVE CS FOR TREE SPARSE SIGNALS

Our analysis here pertains to a simple adaptive compressed sensing procedure for tree sparse signals, similar to the techniques proposed in [11], [12]. As above, let $\alpha \in \mathbb{R}^p$ denote the tree-sparse representation of an unknown signal $x \in \mathbb{R}^n$ in a known $n \times p$ dictionary D having orthonormal columns. We assume sequential measurements of the form specified in (2) where the additive noises w_i are taken to be iid $\mathcal{N}(0, 1)$.

Rather than projecting onto randomly generated test vectors, here we will obtain measurements of x by projecting onto selectively chosen, scaled versions of columns of the dictionary D , as follows. Without loss of generality suppose that the index 1 corresponds to the root of the tree $\mathcal{T}_{p,d}$. Begin by initializing a data structure (a stack or queue) with the index 1, and collect a (noisy) measurement of the coefficient α_1 according to (2) by selecting $\phi_1 = \beta d_1$, where $\beta > 0$ is a fixed scaling parameter. That is, obtain a measurement

$$y = \beta d_1^T x + w. \quad (3)$$

Note that our assumptions on the additive noise imply that $y \sim \mathcal{N}(\beta \alpha_1, 1)$. Now, perform a significance test to determine whether the amplitude of the measured value y exceeds a specified threshold $\tau > 0$. If the measurement is deemed significant (ie, $|y| \geq \tau$), then add the locations of the d children of index 1 in the tree $\mathcal{T}_{p,d}$ to the stack (or queue). If the measurement is not deemed significant, then obtain the next index from the data structure (if the structure is nonempty) to determine which column of D should comprise the next test vector, and proceed as above. If the data structure is empty, the procedure stops. Notice that using a stack as the data structure results in *depth-first* traversal of the tree, while using a queue results in *breadth-first* traversal. The aforementioned algorithm is adaptive in the sense that the decision on which locations of α to measure depends on outcomes of the statistical tests corresponding to the previous measurements.

The performance of this procedure is quantified by the following result, which comprises the main theoretical contribution of this work. A sketch of the proof of the theorem is given in Sec. VI.

Theorem 1: Let α be k -tree-sparse in the tree $\mathcal{T}_{p,d}$ with support set \mathcal{S} , and suppose $k < p/d$. For any $c_1 > 0$ and $c_2 \in (0, 1)$, there exists a constant $c_3 > 0$ such that if

$$\alpha_{\min} = \min_{i \in \mathcal{S}} |\alpha_i| \geq \sqrt{c_3 \frac{\log k}{\beta^2}} \quad (4)$$

and $\tau = c_2 \beta \alpha_{\min}$, the following hold with probability at least $1 - k^{-c_1}$: the total number of measurements obtained $m = dk + 1$, and the support estimate $\hat{\mathcal{S}}$ comprised of all the measured locations for which corresponding measured value exceeds τ in amplitude is equal to \mathcal{S} .

A brief discussion is in order here to put the results of this theorem in context. Note that in practical settings, physical constraints (eg., power or time limitations) effectively impose a limit on the precision of the measurements that may

be obtained. This can be modeled by introducing a global constraint of the form

$$\sum_i \|\phi_i\|_2^2 \leq R, \quad (5)$$

on the model (2) in order to limit the ‘‘sensing energy’’ that may be expended throughout entire measurement process. In the context of Thm. 1, this corresponds to a constraint of the form $\sum_{i=1}^m \beta^2 \leq R$. In this case for the choice

$$\beta = \sqrt{\frac{R}{(d+1)k}}, \quad (6)$$

Thm. 1 guarantees exact support recovery with high probability from $O(k)$ measurements provided that α_{\min} exceeds a constant times $\sqrt{(d+1)(k/R)\log k}$. To assess the benefits of exploiting structure via adaptive sensing, it is illustrative to compare the result of Thm. 1 with results obtained in several recent works that examined support recovery for *unstructured* sparse signals under a Gaussian noise model. The consistent theme identified in these works is that exact support recovery is impossible unless the minimum signal amplitude α_{\min} exceeds a constant times $\sqrt{(n/R)\log n}$ for non-adaptive measurement strategies [13], [14], or $\sqrt{(n/R)\log k}$ for adaptive sensing strategies [15]. Clearly, when the signal being acquired is sparse ($k \ll n$), the procedure analyzed in this work succeeds in recovering much weaker signals.

Our proof of Thm. 1 can be extended to obtain guarantees on the accuracy of an estimate obtained via a related adaptive sensing procedure.

Corollary 1: There exists a two-stage (support recovery, then estimation) adaptive compressed sensing procedure for k -tree sparse signals that produces an estimate from $m = O(k)$ measurements that (with high probability) satisfies

$$\|\hat{\alpha} - \alpha\|_2^2 = O\left(k\left(\frac{k}{R}\right)\right) \quad (7)$$

provided α_{\min} exceeds a constant times $\sqrt{(k/R)\log k}$.

By comparison, non-adaptive CS estimation techniques that do not assume any structure in the sparse representation can achieve estimation error

$$\|\alpha - \hat{\alpha}\|_2^2 = O\left(k\left(\frac{n}{R}\right)\log n\right), \quad (8)$$

from $m = O(k\log(n/k))$ measurements [16]. Exploiting structure in non-adaptive CS, as in [5], results in an estimation procedure that achieves error

$$\|\alpha - \hat{\alpha}\|_2^2 = O\left(k\left(\frac{n}{R}\right)\right) \quad (9)$$

from $m = O(k)$ measurements. Again, we see that the results of the corollary to Thm. 1 provide a significant improvement over these existing error bounds, especially in the case when $k \ll n$.

III. LEARNING ADAPTIVE SENSING REPRESENTATIONS

The approach outlined above can be applied in general settings, by employing techniques from *dictionary learning* [17], [18]. Let X denote an $n \times q$ matrix whose n -dimensional columns x_i comprise a collection of training data, and suppose

we can find a factorization of X of the form $X \approx DA$, where D is an $n \times p$ dictionary with orthonormal columns, and A is a $p \times q$ matrix whose columns $a_i \in \mathbb{R}^p$ each exhibit tree-sparsity in some tree $\mathcal{T}_{p,d}$. The task of finding the dictionary D and associated coefficient matrix A with tree-sparse columns can be accomplished by solving an optimization of the form

$$\{D, A\} = \arg \min_{D \in \mathbb{R}^{n \times p}, \{a_i\} \in \mathbb{R}^q} \sum_{i=1}^p \|x_i - Da_i\|_2^2 + \lambda \Omega(a_i), \quad (10)$$

subject to the constraint $D^T D = I$. Here, the regularization term is given by

$$\Omega(a_i) = \sum_{g \in \mathcal{G}} \omega_g \|(a_i)_g\|, \quad (11)$$

where \mathcal{G} is the set of p groups, each comprised of a node with all of its descendants in the tree $\mathcal{T}_{p,d}$, the notation $(a_i)_g$ refers to the subvector of a_i restricted to the indices in the set $g \in \mathcal{G}$, the ω_g are non-negative weights, and the norm can be either the ℓ_2 or ℓ_∞ norm. Efficient software packages have been developed (eg., [19]) for solving the optimizations of the form (10) via alternating minimization over D and A . Enforcing the additional constraint of orthogonality of the columns of D can be achieved in a straightforward manner. In the context of the procedure outlined in Sec. II, we refer to solving this form of constrained structured dictionary learning task as **Learning Adaptive Sensing Representations**, or **LASeR**. The performance of LASeR is evaluated in the next section.

IV. EXPERIMENTAL RESULTS

We performed experiments on the Psychological Image Collection at Stirling [20] which contains a set of 72 man-made and 91 natural images. The files are in JPG and TIFF format respectively, with each image of size 256×256 (here, each of the images was rescaled to 128×128 to reduce computational demands on the dictionary learning procedure). The training data were then each reshaped to a 16384×1 vector and stacked together to form the training matrix $X \in \mathbb{R}^{16384 \times 163}$. After centering the training data by subtracting the column mean of the training matrix from each of the training vectors, we learned a balanced binary tree structured orthonormal dictionary with 7 levels (comprising 127 orthogonal dictionary elements).

The LASeR sensing procedure was then applied with rows of dictionary scaled to meet the total sensing budget R for two test signals (chosen from the original training set). Since, during the dictionary learning process we specify the sparsity level of the signal in the learned dictionary, allocation of sensing energy to each measurement can be done beforehand (specifically β is defined as in (6)). We evaluated the performance of the procedure for various values of τ (the threshold for determining significance of a measured coefficient) in a noisy setting corrupted by zero-mean additive white Gaussian measurement noise. The reconstruction from the LASeR procedure is obtained as the column mean plus a weighted sum of the atoms of the dictionary used to obtain the projections, where the weights are taken to be the actual observation values obtained by projecting onto the corresponding atom. When assessing the performance of the procedure in noisy

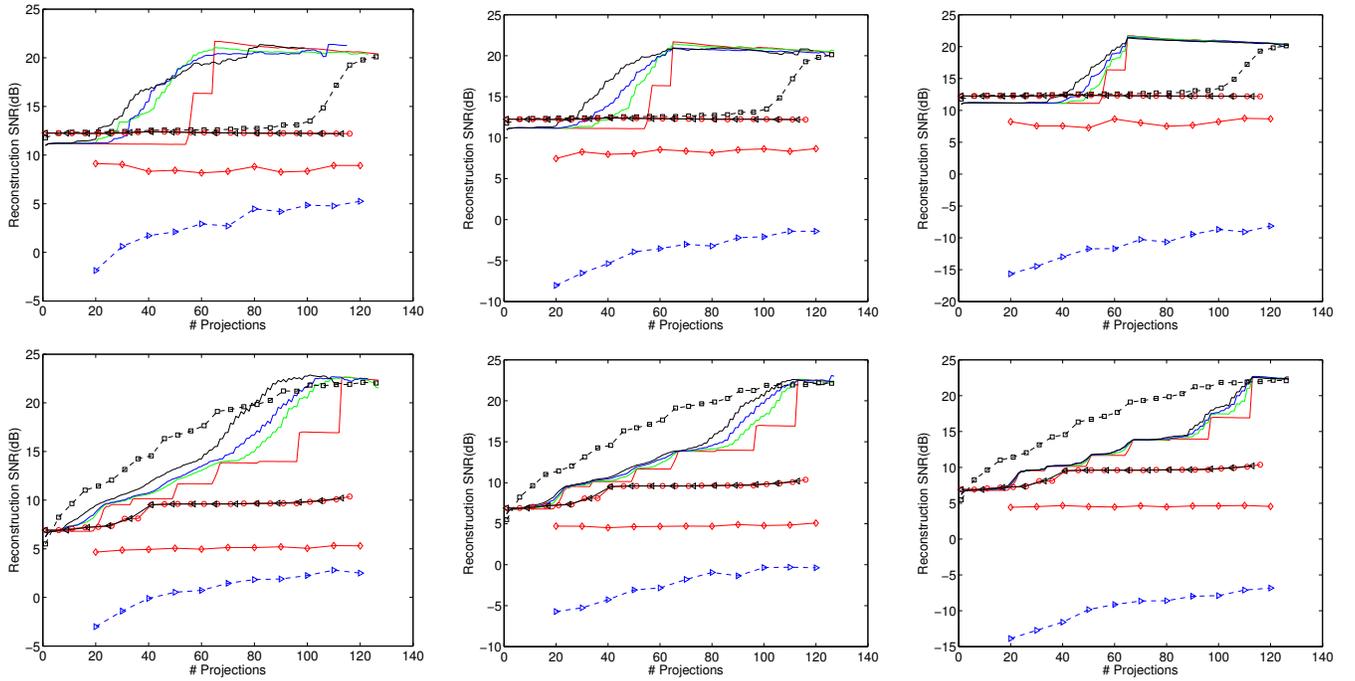


Fig. 1: Reconstruction SNR vs. Number of measurements plots (best viewed in color) with different sensing energy R and fixed noise level $\sigma^2 = 1$ for different schemes (LASeR, PCA, direct wavelet sensing, model-based CS and Lasso). Results in each row corresponds to a different test image. Column 1: $R = 128 \times 128$, Column 2: $R = (128 \times 128)/8$, Column 3: $R = (128 \times 128)/32$. Here, \square is PCA, \diamond is model-based CS, \triangleright is CS-Lasso, \triangleleft and \circ are for direct wavelet sensing with $\tau = 0$ and $\tau = 0.5$ respectively. Colored solid lines are for LASeR with red for $\tau = 0$, green for $\tau = 0.04$, blue for $\tau = 0.06$ and black for $\tau = 0.1$.

settings, we averaged performance over a total of 500 trials corresponding to different realizations of the random noise.

Reconstruction performance is quantified by the reconstruction signal to noise ratio (SNR), given by

$$\text{SNR} = 10 \log_{10} \left(\frac{\|\mathbf{x}\|_2^2}{\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2} \right). \quad (12)$$

where \mathbf{x} and $\hat{\mathbf{x}}$ are the original test and reconstructed signal respectively.

To provide a performance comparison for LASeR, we also evaluate the reconstruction performance of the direct wavelet sensing algorithm described in [12], as well as Principal component analysis (PCA) based reconstruction. For PCA, the reconstruction is obtained by taking projections of the test signal onto the principal components and adding back the subtracted column mean to the reconstruction. We also compare with “traditional” compressed sensing and model-based compressed sensing [5], where measurements are obtained by projecting onto random vectors (in this case, vectors whose entries are i.i.d. zero-mean Gaussian distributed) and reconstruction is obtained via the Lasso and CoSaMP respectively. In order to make a fair comparison among all of the different strategies, we scale so that the constraint on the total sensing energy is met.

Reconstruction SNR values vs. number of measurements for two of the test images is shown in Fig. 1. The results in the top row (for the first test image) show that for a range of threshold values τ one can get a good reconstruction

SNR by taking only 60 – 65 measurements using LASeR with very limited sensing budget R . On the other hand, reconstruction SNR for Lasso and model-based CS degrade as we decrease the sensing energy R . The results in the bottom row (corresponding to the second test image) demonstrate a case where the performance of LASeR is on par with PCA. In this case too, the SNR for Lasso and model-based CS decrease significantly as we decrease R . The advantage of LASeR is in the low measurement (high threshold) and low sensing budget scenario where we can get a good reconstruction from few measurements.

V. DISCUSSION/CONCLUSION

In this paper, we presented a novel sensing and reconstruction procedure called LASeR, which uses dictionaries learned from training data, in conjunction with adaptive sensing, to perform compressed sensing. Bounds on minimum feature strength in the presence of measurement noise were explicitly proven for LASeR. Simulations demonstrate that the proposed procedure can provide significant improvements over traditional compressed sensing (based on random projection measurements), as well as other established methods such as PCA.

Future work in this direction will entail obtaining a complete characterization of the performance of the LASeR procedure for different dictionaries, and for different learned tree structures (we restricted attention here to binary trees, though higher degrees can also be obtained via the same procedure).

VI. PROOF OF MAIN RESULT

Before proceeding with the proof of the main theorem, we state an intermediate result concerning the number of measurements that are obtained via the procedure described in Sec. II when sensing a k -tree-sparse vector. We state the result here as a lemma. The proof is by induction on k , and is straightforward, so we omit it here due to space constraints.

Lemma 1: Let $\mathcal{T}_{p,d}$ denote a completed rooted connected tree of degree d with p nodes, and let $\alpha \in \mathbb{R}^p$ be k -tree-sparse in $\mathcal{T}_{p,d}$ with $k \leq q/d$. If the procedure described in Sec. II is used to acquire α , and the outcome of the statistical test is correct at each step, then the procedure halts when $m = dk + 1$ measurements have been collected.

In other words, suppose that for a k -tree sparse α with support \mathcal{S} , the outcomes of each of the statistical tests of the procedure described in Sec. II are correct. Then, the set of locations that are measured is of the form $\mathcal{S} \cup \tilde{\mathcal{S}}^c$, where \mathcal{S} and $\tilde{\mathcal{S}}^c$ are disjoint, and $|\tilde{\mathcal{S}}^c| = (d-1)k + 1$.

A. Sketch of Proof of Theorem 1

An error can occur in two different ways, corresponding to missing a truly significant signal component (a *miss hit*) and determining a component to be significant when it is not (a *false alarm*). Let y_{d_j} correspond to the measurement obtained according to the noisy linear model (2) by projecting onto the column d_j . A false alarm corresponds to the event $|y_{d_j}| \geq \tau$ for some $j \in \tilde{\mathcal{S}}^c$. Since in this case, we have $y_{d_j} \sim \mathcal{N}(0, 1)$, using a standard Gaussian tail bound for zero mean and unit variance random variables, the probability of false alarm can be upper bounded as

$$\Pr(\text{false alarm}) \leq e^{-\tau^2/2}. \quad (13)$$

Likewise, for $j \in \mathcal{S}$, a miss hit corresponds to $|y_{d_j}| < \tau$. Letting $\alpha_{\min} = \min_{j \in \mathcal{S}} |\alpha_j|$, we have

$$\Pr(\text{miss hit}) \leq e^{-(\beta\alpha_{\min}-\tau)^2/2} \quad (14)$$

for $\tau < \beta\alpha_{\min}$.

Now, the probability of exact support recovery corresponds to the probability of the event that each of the $|\mathcal{S}| = k$ statistical tests corresponding to measurements of nonzero signal components is correct, as are each of the $|\tilde{\mathcal{S}}^c| = m - k = (d-1)k + 1$ tests corresponding to measurements obtained at locations where the signal has a zero component. Thus, the probability of the failure event can be obtained via the union bound, as

$$\begin{aligned} \Pr(\text{failure}) &\leq |\tilde{\mathcal{S}}^c| \Pr(\text{false alarm}) + |\mathcal{S}| \Pr(\text{miss hit}) \\ &\leq (m-k)e^{-\tau^2/2} + ke^{-(\beta\alpha_{\min}-\tau)^2/2} \end{aligned} \quad (15)$$

Let $\tau = a(\beta\alpha_{\min})$, where $a \in (0, 1)$. If, for some $c_1 > 0$, each of the terms in the bound above is less than $k^{-c_1}/2$, then the overall failure probability is upper bounded by k^{-c_1} .

Consider the first term on the right hand side of (15), the condition $(m-k)e^{-\tau^2/2} \leq k^{-c_1}/2$ implies that (for $m = dk + 1$),

$$\alpha_{\min} \geq \sqrt{\frac{2 \log((d-1)k+1) + 2c_1 \log k + 2 \log 2}{\beta^2 a^2}}. \quad (16)$$

Similarly, the condition $ke^{-(\beta\alpha_{\min}-\tau)^2/2} \leq k^{-c_1}/2$ implies

$$\alpha_{\min} \geq \sqrt{\frac{2(1+c_1) \log k + 2 \log 2}{\beta^2(1-a)^2}}. \quad (17)$$

There exists a constant c_3 (depending on d and a) such that when $\alpha_{\min} \geq \sqrt{c_3 \log(k)}/\beta^2$, both (16) and (17) are satisfied.

REFERENCES

- [1] 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.
- [2] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [3] Y. Lu and M. Do, "A theory for sampling signals from a union of subspaces," *IEEE Trans. Signal Proc.*, vol. 56, no. 6, pp. 2334–2345, 2008.
- [4] J. Huang, T. Zhang, and D. Metaxas, "Learning with structured sparsity," *Technical Report*, 2009, Online: arxiv.org/pdf/0903.3002v2.
- [5] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.
- [6] M. Crouse, R. Nowak, and R. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Proc.*, vol. 46, pp. 886–902, Apr. 1998.
- [7] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. on Sig. Proc.*, vol. 56, no. 6, pp. 2346–2356, 2008.
- [8] R. Castro, J. Haupt, R. Nowak, and G. Raz, "Finding needles in noisy haystacks," in *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Proc.*, 2008, pp. 5133–5136.
- [9] J. Haupt, R. Baraniuk, R. Castro, and R. Nowak, "Compressive distilled sensing: Sparse recovery using adaptivity in compressive measurements," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, 2009, pp. 1551–1555.
- [10] J. Haupt and R. Nowak, "Adaptive sensing for sparse recovery," To appear in *Compressed Sensing: Theory and applications*, Cambridge University Press, 2011, http://www.ece.umn.edu/~jdhaupt/publications/cs10_adaptive_sensing.pdf.
- [11] L. P. Panych and F. A. Jolesz, "A dynamically adaptive imaging algorithm for wavelet-encoded MRI," *Magnetic Resonance in Medicine*, vol. 32, pp. 738–748, 1994.
- [12] S. Deutsch, A. Averbuch, and S. Dekel, "Adaptive compressed image sensing based on wavelet modelling and direct sampling," in *8th International Conference on Sampling, Theory and Applications. Marseille, France*, 2009.
- [13] D. Donoho and J. Jin, "Higher criticism for detecting sparse heterogenous mixtures," *Ann. Statist.*, vol. 32, no. 3, pp. 962–994, 2004.
- [14] C. Genovese, J. Jin, and L. Wasserman, "Revisiting marginal regression," *preprint*, 2009.
- [15] M. Malloy and R. Nowak, "On the limits of sequential testing in high dimensions," *preprint*, 2011.
- [16] E. J. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n ," *Ann. Statist.*, vol. 35, no. 6, pp. 2313–2351, 2007.
- [17] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?," *Vision Research*, vol. 37, pp. 3311–3325, 1997.
- [18] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Proc.*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [19] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach, "Proximal methods for sparse hierarchical dictionary learning," in *Proc. ICML*, 2010.
- [20] "Psychological image collection at stirling," <http://www.pics.stir.ac.uk/>.