Fundamental Limits for Support Recovery of Tree-Sparse Signals from Noisy Compressive Samples

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 compressive sensing (CS) have established that many high dimensional signals can be accurately recovered from a relatively small number of non-adaptive linear observations, provided that the signals possess a sparse representation in some basis. Subsequent efforts have shown that the performance of CS can be improved by exploiting structure in the locations of the nonzero signal coefficients during inference, or by utilizing some form of data-dependent adaptive measurement scheme during the sensing process. Our previous work established that an adaptive sensing strategy specifically tailored to signals that are tree-sparse can significantly outperform adaptive and non-adaptive sensing strategies that are agnostic to the underlying structure in noisy support recovery tasks. In this paper we establish corresponding fundamental performance limits for these support recovery tasks, in settings where measurements may be obtained either non-adaptively (using a randomized Gaussian measurement strategy motivated by initial CS investigations) or by any adaptive sensing strategy. Our main results here imply that the adaptive tree sensing procedure analyzed in our previous work is nearly optimal, in the sense that no other sensing and estimation strategy can perform fundamentally better for identifying the support of tree-sparse signals.**

I. Introduction

Consider the task of inferring a (perhaps very high-dimensional) vector $\mathbf{x} \in \mathbb{R}^n$. Compressive sensing (CS) prescribes collecting non-adaptive linear measurements of $\mathbf{x}$ by "projecting" it onto a collection of n-dimensional "measurement vectors." Formally, CS observations may be modeled as

$$y_j = \langle \mathbf{a}_j, \mathbf{x} \rangle + w_j = \mathbf{a}_j^T \mathbf{x} + w_j, \quad \text{for } j = 1, 2, \ldots, m, \tag{1}$$

where $\mathbf{a}_j$ is the j-th n-dimensional measurement vector and w_j describes the additive error associated with the j-th measurement, which may be due to modeling error or stochastic noise. Initial breakthrough results in CS established that sparse vectors $\mathbf{x}$ having no more than $k < n$ nonzero elements can be exactly recovered (in noise-free settings) or reliably estimated (in noisy settings) from a collection of only $m = O(k \log n)$ measurements of the form (1) using, for example, ensembles of randomly generated measurement vectors whose entries are iid realizations of certain zero-mean random variables (e.g., Gaussian) – see, for example, [1].

While many of the initial efforts in CS focused on purely randomized measurement vector designs and considered recovery of arbitrary sparse vectors, several powerful extensions to the original CS paradigm have been investigated in the literature. One such extension allows for additional flexibility in the measurement process, so that information gleaned from previous observations may be employed in the design of future measurement vectors. Formally, such *adaptive sensing* strategies are those for which the j-th measurement vector $\mathbf{a}_j$ is obtained as a (deterministic or randomized) function of previous measurement vectors and observations $\{\mathbf{a}_\ell, y_\ell\}_{\ell=1}^{j-1}$, for each $j = 2, 3, \ldots, m$. Non-adaptive sensing strategies, by contrast, are those for which each measurement vector is independent of all past (and future) observations. Adaptive sensing techniques have been shown beneficial in sparse inference tasks, enabling an improved resilience to measurement noise relative to techniques based on non-adaptive measurements (see, for example, [2]–[10] as well as the summary article [11] and the references therein).

This work was supported by DARPA/ONR Award No. N66001-11-1-4090.

Another powerful extension to the canonical CS framework corresponds to the exploitation of additional *structure* that may be present in the locations of the nonzeros of $\mathbf{x}$. To formalize this notion, we first define the support $\mathcal{S} = \mathcal{S}(\mathbf{x})$ of a vector $\mathbf{x} = [x_1 \ x_2 \ \ldots \ x_n]^T$ as $\mathcal{S}(\mathbf{x}) \triangleq \{i : x_i \neq 0\}$, and note that, in general, the support of a k-sparse n-dimensional vector corresponds to one of the $\binom{n}{k}$ distinct supports of $\{1, 2, \ldots, n\}$ of cardinality k. The term *structured sparsity* describes a restricted class of sparse signals whose supports may occur only on a (known) subset of these $\binom{n}{k}$ distinct supports. Generally speaking, knowledge of the particular structure present in the object being inferred can be incorporated into sparse inference procedures, and for certain types of structure this can result either in a reduction in the number of measurements required for accurate inference, or improved estimation error guarantees, or both (see, e.g., [12], [13], as well as the recent survey article [14]).

The authors' own previous work [15] was the first to identify and quantify the benefits of using adaptive sensing strategies that are tailored to certain types of structured sparsity, in noisy sparse inference tasks. Specifically, the work [15] established that a simple adaptive compressive sensing strategy for *tree-sparse* vectors could successfully identify the support of much weaker signals than what could be recovered using non-adaptive or adaptive sensing strategies that were agnostic to the structure present in the signal being acquired. Subsequent efforts by other authors have similarly identified benefits of adaptive sensing techniques tailored to other forms of structured sparsity in noisy sparse inference tasks [8], [9], [16]. The aim of this effort is to establish the optimality of the procedure analyzed in [15], by identifying the fundamental performance limits associated with the task of support recovery of tree-sparse signals from noisy linear measurements.

II. Adaptive CS for Tree Sparse Signals

Tree sparsity essentially describes the phenomenon where the nonzero elements of the signal being inferred exhibit clustering along paths in some known underlying tree. For the purposes of our investigation here, we formalize the notion of tree sparsity as follows. Suppose that the set $\{1, 2, \ldots, n\}$ that indexes the elements of $\mathbf{x} \in \mathbb{R}^n$ is put into a one-to-one correspondence with the nodes of a known tree of degree $d \geq 1$ having n nodes, which we refer to as the *underlying tree*. We say that a vector $\mathbf{x}$ is *k-tree sparse* (with respect to the underlying tree) when the indices of the support set $\mathcal{S}(\mathbf{x})$ correspond, collectively, to a rooted connected subtree of the underlying tree. In the sequel we restrict our attention to n-dimensional signals that are tree sparse in a known underlying *binary* tree ($d = 2$), though our approach and main results can be generalized to underlying trees having degree $d > 2$.

Tree sparsity arises naturally in the wavelet coefficients of many natural signals including, in particular, natural images, and this fact has motivated several investigations into CS inference techniques that exploit or leverage underlying tree structure in the signals being acquired [12], [13], [17]–[21]. Motivated by these efforts – in particular [21] – the essential aim of the authors own prior work [15] was to assess the performance of such approaches in noisy settings. For completeness, we summarize the main results of that work here.

Let us assume, for simplicity, that the signal $\mathbf{x}$ being acquired is tree sparse in the canonical (identity) basis, though extensions to signals that are tree sparse in any other orthonormal basis (e.g., a wavelet basis) are straightforward. Noisy observations of $\mathbf{x}$ are obtained according to (1) by projecting $\mathbf{x}$ onto a sequence of adaptively designed measurement vectors, each of which corresponds to a basis vector of the canonical basis, and we assume that each measurement vector has unit norm. Now, to simplify the description of the procedure, we introduce some slightly different notation to index the individual observations. Specifically, rather than indexing observations by the order in which they were obtained as in (1), we instead index each measurement according to the index of the basis vector onto which $\mathbf{x}$ is projected, or equivalently here, according to the location of $\mathbf{x}$ that was observed. To that end, let us denote by $y_{(j)}$ the measurement obtained by projecting $\mathbf{x}$ onto the vector $\mathbf{e}_j$ having a single nonzero in the j-th location for any $j \in \{1, 2, \ldots, n\}$.

Now, begin by specifying a threshold $\tau \geq 0$, and by initializing a support estimate $\widehat{\mathcal{S}} = \emptyset$ and a data structure $\mathcal{Q}$ (which could be a stack, queue, or simply a set) to contain the index corresponding to the root of the underlying tree. While the data structure $\mathcal{Q}$ is nonempty, remove an element ℓ from $\mathcal{Q}$, collect a noisy measurement $y_{(\ell)}$ by projecting $\mathbf{x}$ onto $\mathbf{e}_\ell$, and perform the following hypothesis test. If $|y_{(\ell)}| \geq \tau$, add the indices corresponding to the children of node ℓ in the underlying tree to the data structure $\mathcal{Q}$ and update the support estimate to include the index ℓ; on the other hand, if $|y_{(\ell)}| < \tau$, then keep $\mathcal{Q}$ and $\widehat{\mathcal{S}}$ unchanged. The procedure continues in this fashion, at each step obtaining a new measurement and performing a corresponding hypothesis test to determine whether the amplitude of the coefficient measured in that step was significant. When the overall procedure terminates it outputs its final support estimate $\widehat{\mathcal{S}}$, which essentially corresponds to the set of locations of $\mathbf{x}$ for which the corresponding measurements exceeded τ in amplitude.

The main result of [15] quantifies the performance of this type of sensing strategy when measurements are noisy. We provide a restatement of that result[1] here as a Lemma, for a generalized scenario where we assume that we obtain $r \geq 1$ measurements (each with its own independent additive noise) at each step of the procedure and these replicated measurements are averaged prior to performing the hypothesis test at each step.

Lemma II.1. *Let $\mathbf{x}$ be a k-tree sparse vector for some $k \geq 2$, and consider acquiring $\mathbf{x}$ using the adaptive tree sensing procedure described above, where $r \geq 1$ measurements are obtained in each step and averaged to reduce the effective measurement noise prior to each hypothesis test. Choose $\delta \in (0, 1)$ and sparsity parameter $k' \in \mathbb{N}$ (intended to be an upper bound on the sparsity level), and set the threshold $\tau = \sqrt{2(\sigma^2/r)\log(4k'/\delta)}$. If $\mathbf{x}$ is k-tree sparse for some $k \geq 2$, the sparsity parameter $k' \leq \beta k$ for some $\beta \geq 1$, and the amplitudes of the nonzeros of $\mathbf{x}$ satisfy*

$$|x_i| \geq \sqrt{8\left[1 + \log\left(\frac{4\beta}{\delta}\right)\right]} \cdot \sqrt{\left(\frac{\sigma^2}{r}\right)\log k}, \tag{2}$$

for every $i \in \mathcal{S}(\mathbf{x})$ then with probability at least $1 - \delta$ the following are true: the algorithm terminates after collecting $m \leq r(2k+1)$ measurements, and the support estimate $\widehat{S}$ produced by the procedure satisfies $\widehat{\mathcal{S}} = \mathcal{S}(\mathbf{x})$.

[1] We note that we have not attempted to optimize constants in our derivation of Lemma II.1, opting instead for simple expressions that better illustrate the scaling behavior with respect to the problem parameters.

Note that when $m \leq r(2k+1)$ we have that $1/r \leq 3k/m$ provided $k \geq 1$. It follows from the corollary that when the sparsity parameter k' does not overestimate the true sparsity level by more than a constant factor (i.e., $\beta \geq 1$ is a *constant*), then a sufficient condition to ensure that the support estimate produced by the repeated-measurements variant of the tree sensing procedure is correct with probability at least $1 - \delta$, is that the nonzero components of $\mathbf{x}$ satisfy

$$|x_i| \geq \sqrt{24\left[1 + \log\left(\frac{4\beta}{\delta}\right)\right]} \cdot \sqrt{\sigma^2\left(\frac{k}{m}\right)\log k}, \tag{3}$$

for all $i \in \mathcal{S}(\mathbf{x})$. Identifying whether any other procedure can accurately recover the support of tree-sparse signals having fundamentally weaker amplitudes is the motivation for our present effort.

III. Problem Statement

As stated above, our specific focus here is on establishing fundamental performance limits for the support recovery task – that of identifying the locations of the nonzeros of $\mathbf{x}$ – in settings where $\mathbf{x}$ is k-tree sparse, and when observations may be designed either non-adaptively (e.g., measurement vectors whose elements are random and iid, as in traditional CS) or adaptively based on previous observations. We formalize this problem here.

1) Signal Model: Let $\mathcal{T}_{n,k}$ denote the set of all unique supports for n-dimensional vectors that are k-tree sparse in the same underlying binary tree with n nodes[2]. Our specific focus will be on classes of k-tree sparse signals, with $2 \leq k \leq (n+1)/2$, where each k-sparse signal $\mathbf{x}$ has support $\mathcal{S}(\mathbf{x}) \in \mathcal{T}_{n,k}$, and for which the amplitudes of all nonzero signal components are greater or equal to some non-negative quantity μ. Formally, for a given underlying tree, fixed sparsity level k, and $\mathcal{T}_{n,k}$ as described above, we define the signal class

$$\mathcal{X}_{\mu;\mathcal{T}_{n,k}} \triangleq \left\{\mathbf{x} \in \mathbb{R}^n : x_i = \alpha_i \mathbf{1}_{\{i \in T\}},\ |\alpha_i| \geq \mu > 0,\ T \in \mathcal{T}_{n,k}\right\}, \tag{4}$$

where $\mathbf{1}_{\{\mathcal{B}\}}$ denotes the indicator function of the event $\mathcal{B}$. In the sequel, we choose to simplify the exposition by denoting the signal class $\mathcal{X}_{\mu;\mathcal{T}_{n,k}}$ using the shorthand notation $\mathcal{X}_{\mu,k}$.

2) Sensing Strategies: We examine the support recovery task under both adaptive and non-adaptive sensing strategies. Here, when considering performance limits of non-adaptive sensing, we consider observations obtained according to the model (1), where each $\mathbf{a}_j$, $j = 1, 2, \ldots, m$, is an independent random vector, whose elements are iid $\mathcal{N}(0, 1/n)$ random variables which ensures that each measurement vector has norm one in expectation; that is, $\mathbb{E}\left[\|\mathbf{a}_j\|_2^2\right] = 1$ for all $j = 1, 2, \ldots, m$. Our investigation of adaptive sensing strategies focuses on observations obtained according to (1), using measurement vectors satisfying $\|\mathbf{a}_j\|_2^2 = 1$, for $j = 1, 2, \ldots, m$, and for which $\mathbf{a}_j$ is allowed to explicitly depend on $\{\mathbf{a}_\ell, y_\ell\}_{\ell=1}^{j-1}$ for $j = 2, 3, \ldots, m$, as described above.

Overall, as noted in [10], we can essentially view any (non-adaptive, or adaptive) sensing strategy in terms of a collection M of *conditional distributions* of measurement vectors $\mathbf{a}_j$ given $\{\mathbf{a}_\ell, y_\ell\}_{\ell=1}^{j-1}$ for $j = 2, 3, \ldots, m$. We adopt this interpretation here, denoting by $M_{m,\text{na}}$ the specific sensing strategy based on non-adaptive Gaussian random measurements described above, and by $\mathcal{M}_m$ be the collection of all adaptive (or non-adaptive) sensing

[2] For technical reasons, we further assume that the underlying trees are *nearly complete*, meaning that all levels of the underlying tree are full with the possible exception of the last (i.e., the bottom) level, and all nodes in any partially full level are as far to the left as possible.

strategies based on m measurements, where each measurement vector is exactly norm one (with probability one).

3) Observation Noise: In each case, we model the noises associated with the linear measurements as a sequence of independent $\mathcal{N}(0, \sigma^2)$ random variables. We further assume that each noise w_j is independent of the present and all past measurement vectors $\{\mathbf{a}_\ell\}_{\ell=1}^{j}$. For the non-adaptive sensing strategies we examine here, noises will also be independent of future measurement vectors, though by design, future measurement vectors generally *will not* be independent of present noises when adaptive sensing strategies are employed.

4) The Support Estimation Task: An estimator ψ takes as its input a collection of measurement vectors and associated observations, $\{\mathbf{a}_j, y_j\}_{j=1}^{m}$, denoted by $\{\mathbf{A}_m, \mathbf{y}_m\}$ in the sequel (for shorthand), and outputs a subset of the index set $\{1, 2, \ldots, n\}$. Overall, we denote a support estimate based on observations $\mathbf{A}_m, \mathbf{y}_m$ obtained using sensing strategy M by $\psi(\mathbf{A}_m, \mathbf{y}_m; M)$.

Now, under the 0/1 loss function $d(S_1, S_2) \triangleq \mathbf{1}_{\{S_1 \neq S_2\}}$ defined on elements $S_1, S_2 \subseteq \{1, 2, \ldots, n\}$, the (maximum) risk of an estimator ψ based on sensing strategy M over the set $\mathcal{X}_{\mu,k}$ is

$$\begin{aligned} \mathcal{R}_{\mathcal{X}_{\mu,k}}(\psi, M) &\triangleq \sup_{\mathbf{x} \in \mathcal{X}_{\mu,k}} \mathbb{E}_{\mathbf{x}} \left[\, d(\psi(\mathbf{A}_m, \mathbf{y}_m; M), \ \mathcal{S}(\mathbf{x})) \, \right] \\ &= \sup_{\mathbf{x} \in \mathcal{X}_{\mu,k}} \Pr_{\mathbf{x}} \left(\psi(\mathbf{A}_m, \mathbf{y}_m; M) \neq \mathcal{S}(\mathbf{x}) \right), \quad (5) \end{aligned}$$

where $\mathbb{E}_{\mathbf{x}}$ and $\Pr_{\mathbf{x}}$ denote, respectively, expectation and probability with respect to the joint distribution $\mathbb{P}(\mathbf{A}_m, \mathbf{y}_m; \mathbf{x}) \triangleq \mathbb{P}_{\mathbf{x}}(\mathbf{A}_m, \mathbf{y}_m)$ of the quantities $\{\mathbf{A}_m, \mathbf{y}_m\}$ that is induced when $\mathbf{x}$ is the true signal being observed.

Now, we define the *minimax risk* $\mathcal{R}^*_{\mathcal{X}_{\mu,k},\mathcal{M}}$ associated with the class of distributions $\{\mathbb{P}_{\mathbf{x}} : \mathbf{x} \in \mathcal{X}_{\mu,k}\}$ induced by elements $\mathbf{x} \in \mathcal{X}_{\mu,k}$ and the class $\mathcal{M}$ of allowable sensing strategies as the infimum of the (maximum) risk over all estimators ψ and sensing strategies $M \in \mathcal{M}$; that is,

$$\begin{aligned} \mathcal{R}^*_{\mathcal{X}_{\mu,k},\mathcal{M}} &\triangleq \inf_{\psi; M \in \mathcal{M}} R_{\mathcal{X}_{\mu,k}}(\psi, M), \quad (6) \\ &= \inf_{\psi; M \in \mathcal{M}} \sup_{\mathbf{x} \in \mathcal{X}_{\mu,k}} \Pr_{\mathbf{x}} \left(\psi(\mathbf{A}_m, \mathbf{y}_m; M) \neq \mathcal{S}(\mathbf{x}) \right). \end{aligned}$$

Note that when the minimax risk is bounded away from zero, so that $\mathcal{R}^*_{\mathcal{X}_{\mu,k},\mathcal{M}} \geq \gamma$ for some $\gamma > 0$, it follows that for any particular estimator ψ and sensing strategy $M \in \mathcal{M}$ employed, there will always be at least one signal $\mathbf{x} \in \mathcal{X}_{\mu,k}$ for which $\Pr_{\mathbf{x}} \left(\psi(\mathbf{A}_m, \mathbf{y}_m; M) \neq \mathcal{S}(\mathbf{x}) \right) \geq \gamma$.

A. Summary of Our Contributions

We state the results here as theorems, and refer the readers to the full version of this paper [22] for detailed proofs. Our first main result analyzes the support recovery task for tree-sparse signals in a non-adaptive sensing scenario.

Theorem III.1. *Let $\mathcal{X}_{\mu,k}$ be the class of k-tree sparse n-dimensional signals defined in* (4) *where $2 \leq k \leq (n+1)/2$, and consider acquiring m measurements of $\mathbf{x} \in \mathcal{X}_{\mu,k}$ using the non-adaptive (random, Gaussian) sensing strategy $M_{m,\mathrm{na}}$. If*

$$\mu \leq \sqrt{\frac{1 - 2\gamma}{4}} \cdot \sqrt{\sigma^2 \left(\frac{n}{m} \right) \log(k)}, \quad (7)$$

for some $\gamma \in (0, 1/3)$ then the minimax risk $\mathcal{R}^*_{\mathcal{X}_{\mu,k},M_{m,\mathrm{na}}}$ defined in (6) *obeys the bound $\mathcal{R}^*_{\mathcal{X}_{\mu,k},M_{m,\mathrm{na}}} \geq \gamma$.*

Our second main result concerns support recovery in scenarios where adaptive sensing strategies may be employed.

Theorem III.2. *Let $\mathcal{X}_{\mu,k}$ be the class of k-tree sparse n-dimensional signals defined in* (4) *where $2 \leq k \leq (n+1)/2$, and consider acquiring m measurements of $\mathbf{x} \in \mathcal{X}_{\mu,k}$ using any sensing strategy $M \in \mathcal{M}_m$. If*

$$\mu \leq (1 - 2\gamma) \sqrt{\sigma^2 \left(\frac{k}{m} \right)}, \quad (8)$$

for some $\gamma \in (0, 1/3)$ then the minimax risk $\mathcal{R}^*_{\mathcal{X}_{\mu,k},\mathcal{M}_m}$ defined in (6) *obeys the bound $\mathcal{R}^*_{\mathcal{X}_{\mu,k},\mathcal{M}_m} \geq \gamma$.*

TABLE I: Summary of necessary conditions for exact support recovery using non-adaptive or adaptive sensing strategies that obtain m measurements of k-sparse n-dimensional signals that are either unstructured or tree sparse in an underlying binary tree.

Sampling Strategy / Sparsity Model	Non-adaptive Sensing	Adaptive Sensing
Unstructured Sparsity	$\sqrt{\sigma^2 \left(\frac{n}{m}\right) \log n}$ [24], [25]	$\sqrt{\sigma^2 \left(\frac{n}{m}\right) \log k}$ [26] (when $m > n$)
Tree Sparsity	$\sqrt{\sigma^2 \left(\frac{n}{m}\right) \log k}$ Theorem III.1	$\sqrt{\sigma^2 \left(\frac{k}{m}\right)}$ Theorem III.2

Table I depicts a summary of our main results in a broader context. Overall, we compare four distinct scenarios corresponding to a taxonomy of adaptive and non-adaptive sensing strategies for recovering k-sparse signals under assumptions of unstructured sparsity and tree sparsity. For each, we identify (up to an unstated constant) a critical value of the signal amplitude parameter, say μ^*, such that for the support recovery task the minimax risk over the class $\mathcal{X}_{\mu,k}$ will necessarily be bounded away from zero when $\mu \leq \mu^*$.

The results of Theorem III.2, summarized in the lower-right corner of Table I, address our overall question – the simple adaptive tree sensing procedure described above is indeed nearly optimal for estimating the support of k-tree sparse vectors, in the following sense: Lemma II.1 describes a technique that accurately recovers (with probability at least $1-\delta$, where δ can be made arbitrarily small) the support of any k-tree sparse signal from $m \leq r(2k+1)$ measurements, provided the amplitudes of the nonzero signal components all exceed $c_\delta \cdot \sqrt{\sigma^2 \, (k/m) \log k}$ for some constant c_δ. On the other hand, for any estimation strategy based on any adaptive or non-adaptive sensing method, support recovery will fail (with probability at least γ) to accurately recover the support of some signal or signals in a class comprised of k-tree sparse vectors whose nonzero components exceed $c_\gamma \cdot \sqrt{\sigma^2 \, (k/m)}$ in amplitude, for a constant c_γ.

IV. Experimental Evaluation and Discussion

In this section we provide some experimental results to illustrate the performance improvements that can be achieved in the support recovery task using the adaptive tree sensing procedure. We evaluate four different sensing and support estimation strategies – a non-adaptive CS strategy based on the Lasso estimator; non-adaptive CS using a group-Lasso estimator, with groups designed to enforce tree-structure; the adaptive CS procedure of [23]; and the adaptive tree sensing procedure described above – intended to be illustrative approaches for each of the four scenarios identified in Table I. We refer reader to [22] for complete details on the experimental setup.

We consider overall three different scenarios, corresponding to three different values of the problem dimension ($n = 2^8 - 1$, $n = 2^{10} - 1$, and $n = 2^{12} - 1$, chosen so that the underlying trees in each case are complete), and in each case we evaluate the

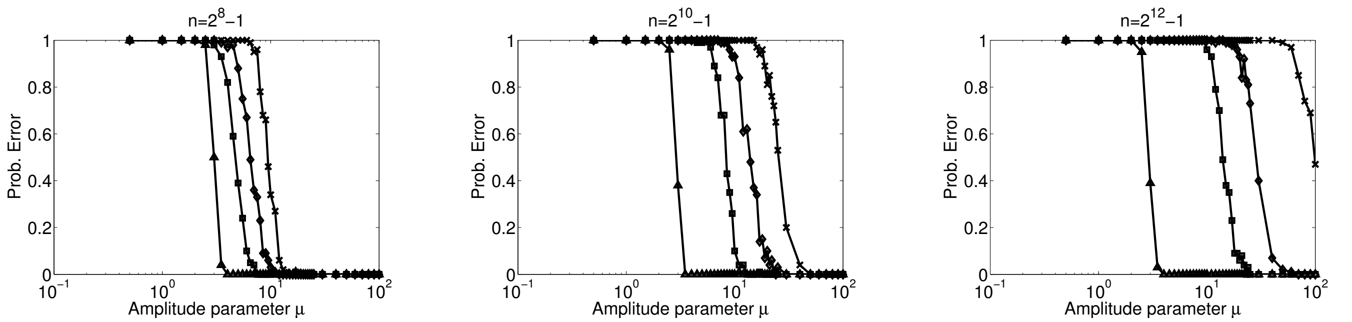

Fig. 1: Empirical probability of support recovery error as a function of signal amplitude parameter μ in three different problem dimensions n. In each case, four different sensing and support recovery approaches – the adaptive tree sensing procedure described here (△ markers); the adaptive compressive sensing approach of [23] (□ markers); a Group Lasso approach for recovering tree-sparse vectors (⋄ markers), and a Lasso approach for recovering unstructured sparse signals (× markers) – were employed to recover the support of a tree-sparse signal with 16 nonzeros of amplitude μ. The proposed tree-sensing procedure exhibits performance that is unchanged as the problem dimension increases, in agreement with the theoretical analysis here.

performance of each approach over a range of signal amplitude parameters μ, as follows. In each of 100 trials we first generate a random n-dimensional tree-sparse signal with $k = 16$ nonzero components of amplitude μ. We construct the signals here so that all nonzero components are non-negative, for simplicity, and to facilitate direct comparison with the procedure analyzed in [23]. We fix $m = 4(2k+1)$ and apply each of the procedures described above (with additive noise variance $\sigma^2 = 1$), and assess whether it correctly identifies the true support by comparing the support estimate obtained by the procedure with the true support of the tree signal. The final empirical probabilities of support recovery error for each approach (and each fixed n and μ) were obtained by averaging results over the 100 trials. Fig. 1 shows the simulation results.

A few interesting points are worth noting here. First, as expected, the adaptive tree-sensing procedure outperforms each of the other approaches in each of the three scenarios; overall, the results suggest that either utilizing adaptive sensing or exploiting tree structure (alone) can indeed result in techniques that outperform traditional CS, but even more significant improvements are possible when leveraging adaptivity and structure together, confirming our claim in the discussion in Section I. Further, it is interesting to note that the performance of the tree-sensing procedure is *unchanged* as the problem dimension increases, in agreement with the result of the result of Lemma II.1, where the sufficient condition on μ that ensures accurate support recovery does not depend on the ambient dimension n. By comparison, the performance of each of the other approaches degrades as the problem dimension increases – a "curse of dimensionality" suffered by each of these other techniques. Whether other (useful) forms of structured sparsity exhibit this favorable characteristic remains an open question.

REFERENCES

[1] E. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *The Annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, 2007.
[2] R. M. Castro, J. Haupt, R. Nowak, and G. M. Raz, "Finding needles in noisy haystacks," in *Proc. IEEE Conf. on Acoustics, Speech and Signal Processing*, 2008, pp. 5133–5136.
[3] 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.
[4] J. Haupt, R. M. Castro, and R. Nowak, "Distilled sensing: Adaptive sampling for sparse detection and estimation," *IEEE Trans. Information Theory*, vol. 57, no. 9, pp. 6222–6235, 2011.
[5] M. Malloy and R. Nowak, "Sequential testing for sparse recovery," *Submitted*, 2012, online at arxiv.org/abs/1212.1801.
[6] E. Arias-Castro, E. J. Candes, and M. Davenport, "On the fundamental limits of adaptive sensing," *Submitted*, 2011, online at arxiv.org/abs/1111.4646.
[7] J. Haupt, R. Baraniuk, R. Castro, and R. Nowak, "Sequentially designed compressed sensing," in *Proc. IEEE Statistical Signal Processing Workshop*, 2012, pp. 401–404.
[8] S. Balakrishnan, M. Kolar, A. Rinaldo, and A. Singh, "Recovering block-structured activations using compressive measurements," *Submitted*, 2012, online at arxiv.org/abs/1209.3431.
[9] A. Krishnamurthy, J. Sharpnack, and A. Singh, "Recovering graph-structured activations using adaptive compressive measurements," *Submitted*, 2013, online at arxiv.org/abs/1305.0213.
[10] R. M. Castro, "Adaptive sensing performance lower bounds for sparse signal detection and support estimation," *Submitted*, 2012, online at arxiv.org/abs/1206.0648.
[11] J. Haupt and R. Nowak, "Adaptive sensing for sparse recovery," in *Compressed Sensing: Theory and applications*, Y. Eldar and G. Kutyniok, Eds. Cambridge University Press, 2011.
[12] J. Huang, T. Zhang, and D. Metaxas, "Learning with structured sparsity," in *Proc. Intl. Conf. Machine Learning*, 2009, pp. 417–424.
[13] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inform. Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.
[14] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans Signal Proc*, vol. 59, no. 9, pp. 4053–4085, 2011.
[15] A. Soni and J. Haupt, "Efficient adaptive compressive sensing using sparse hierarchical learned dictionaries," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, 2011, pp. 1250–1254.
[16] N. Rao and R. Nowak, "Adaptive sensing with structured sparsity," in *Proc. IEEE Conf. on Acoustics, Speech, and Signal Processing*, 2013.
[17] M. F. Duarte, M. B. Wakin, and R. G. Baraniuk, "Fast reconstruction of piecewise smooth signals from incoherent projections," in *Proc. SPARS*, 2005.
[18] C. La and M. N. Do, "Signal reconstruction using sparse tree representation," in *Proc. Wavelets XI at SPIE Optics and Photonics*, 2005.
[19] L. P. Panych and F. A. Jolesz, "A dynamically adaptive imaging algorithm for wavelet-encoded MRI," *Magnetic Resonance in Medicine*, vol. 32, no. 6, pp. 738–748, 1994.
[20] M. W. Seeger and H. Nickisch, "Compressed sensing and Bayesian experimental design," in *Proc. ICML*, 2008, pp. 912–919.
[21] 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.
[22] A. Soni and J. Haupt, "On the fundamental limits of recovering tree sparse vectors from noisy linear measurements," *Submitted*, 2013, online at http://arxiv.org/abs/1306.4391.
[23] M. Malloy and R. Nowak, "Near-optimal adaptive compressive sensing," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, 2012.
[24] M. Wainwright, "Information-theoretic limitations on sparsity recovery in the high-dimensional and noisy setting," *IEEE Trans. Inform. Theory*, vol. 55, no. 12, 2009.
[25] S. Aeron, V. Saligrama, and M. Zhao, "Information theoretic bounds for compressed sensing," *IEEE Trans. Inform. Theory*, vol. 56, no. 10, pp. 5111–5130, 2010.
[26] M. Malloy and R. Nowak, "Sequential analysis in high-dimensional multiple testing and sparse recovery," in *Proc. IEEE Intl. Symp. on Information Theory*, 2011, pp. 2661–2665.