

Near-Optimal Adaptive Compressed Sensing

Matthew L. Malloy

Electrical and Computer Engineering
University of Wisconsin-Madison
Email: mmalloy@wisc.edu

Robert D. Nowak

Electrical and Computer Engineering
University of Wisconsin-Madison
Email: nowak@ece.wisc.edu

Abstract—This paper proposes a simple adaptive sensing and group testing algorithm for sparse signal recovery. The algorithm, termed Compressive Adaptive Sense and Search (CASS), is shown to be near-optimal in that it succeeds at the lowest possible signal-to-noise-ratio (SNR) levels. Like traditional compressed sensing based on random non-adaptive design matrices, the CASS algorithm requires only $k \log n$ measurements to recover a k -sparse signal of dimension n . However, CASS succeeds at SNR levels that are a factor $\log n$ less than required by standard compressed sensing. From the point of view of constructing and implementing the sensing operation as well as computing the reconstruction, the proposed algorithm is substantially less computationally intensive than standard compressed sensing. CASS is also demonstrated to perform considerably better in practice through simulation. To the best of our knowledge, this is the first demonstration of an adaptive compressed sensing algorithm with near-optimal theoretical guarantees *and* excellent practical performance. This paper also shows that methods like compressed sensing, group testing, and pooling have an advantage beyond simply reducing the number of measurements or tests – adaptive versions of such methods can also improve detection and estimation performance when compared to non-adaptive direct (uncompressed) sensing.

I. INTRODUCTION

Compressed sensing (CS) has had a tremendous impact on signal processing, machine learning, and statistics, fundamentally changing the way we think about sensing and data acquisition. Beyond the standard compressed sensing methods that gave birth to the field, the compressive framework naturally suggests an ability to make measurements in an on-line and adaptive manner. Adaptive sensing uses previously collected measurements to guide the design and selection of the next measurement in order to optimize the gain of new information.

There is now a reasonably complete understanding of the potential advantages of adaptive sensing over non-adaptive sensing, the main one being that adaptive sensing can reliably recover sparse signals at lower SNRs than non-adaptive sensing. Roughly speaking, to recover a k -sparse signal of length n , standard (non-adaptive) sensing requires the SNR to grow like $\log n$. Adaptive sensing, on the other hand, succeeds as long as the SNR scales like $\log k$. This is a significant improvement, especially in high-dimensional regimes. In terms of the number of measurements, both standard compressed sensing and adaptive compressed sensing require about $k \log n$ measurements.

This paper makes two main contributions in adaptive sensing. First, we propose a simple adaptive sensing algorithm, termed Compressive Adaptive Sense and Search (CASS) that is proved to be near-optimal in that it succeeds if the SNR scales like $\log k$. From the point of view of constructing and implementing the sensing operation as well as computing the reconstruction, the CASS algorithm is comparatively less computationally intensive than standard compressed sensing. The algorithm could be easily realized in a number of existing compressive systems, including those based on digital micromirror devices [2], [3]. Second, CASS is demonstrated in simulation to perform considerably better than 1) compressed sensing based on random Gaussian sensing matrices and 2) non-adaptive direct sensing (which requires $m = n$ measurements). To the best of our knowledge, this is the first demonstration of an adaptive sensing algorithm with near-optimal theoretical guarantees *and* excellent practical performance.

The results presented in this paper answer the following important question. Can sensing systems that make adaptive measurements significantly outperform non-adaptive systems in practice? Perhaps not surprisingly, the answer is *yes*, contradictory to the theme of [4]. This is for two reasons; first, while the required SNR for success of adaptive procedures is at best a logarithmic factor in dimension smaller than non-adaptive procedures, this factor *is* significant for problems of even modest size. On test signals (natural and synthetic images, see Figs. 4 and 5) the CASS procedure consistently outperforms standard compressed sensing by 2-8 dB. Second, in terms of computational complexity, adaptive sensing algorithms such as CASS require little or no computation after the measurement stage of the procedure is completed; in general, non-adaptive algorithms involve computation and memory intense optimization routines after measurements are gathered.

This work also sheds light on another relevant question. Can compressive sensing systems with an SNR budget ever hope to outperform direct non-compressive systems which measure each element of an unknown vector directly? The answer to this question is somewhat surprising. For the task of support recovery, when the measurements are collected non-adaptively, compressive techniques are always less reliable than direct sensing with n measurements. However, *adaptive* compressed sensing with just $k \log n$ measurements can be more reliable than non-adaptive direct sensing, provided the signal is sufficiently sparse. This means that methods like compressed sensing, group testing, and pooling may have an

This work was presented in part at the Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA [1]

advantage beyond simply reducing the number of measurements – adaptive versions of such methods can also improve detection and estimation performance.

A. Problem Setup and Background

Sparse support recovery in compressed sensing refers to the following problem. Let $\mathbf{x} \in \mathbb{R}^n$ be an unknown signal with $k \ll n$ non-zero entries. The unknown signal is measured through m linear projections of the form:

$$y_i = \langle \mathbf{a}_i, \mathbf{x} \rangle + z_i \quad i = 1, \dots, m \quad (1)$$

where z_i are i.i.d. $\mathcal{N}(0, 1)$, and $\mathbf{a}_i \in \mathbb{R}^n$ are sensing vectors with a total sensing energy constraint

$$\sum_{i=1}^m \|\mathbf{a}_i\|_2^2 \leq M. \quad (2)$$

The goal of the support recovery problem is to identify the locations of the non-zero entries of \mathbf{x} . Roughly speaking, the paramount result of compressed sensing states this is possible if 1) m is greater than a constant times $k \log n$ (which can be much less than n), and 2) the amplitude of the non-zero entries of \mathbf{x} , and the total sensing energy M , aren't too small.

This paper is concerned with relaxing the second requirement through adaptivity. If the sensing vectors $\mathbf{a}_1, \dots, \mathbf{a}_m$ are fixed prior to making any measurements (the *non-adaptive* setting), then a necessary condition for exact support recovery is that the smallest non-zero entry of \mathbf{x} be greater than a constant times $\sqrt{\frac{n}{M} \log n}$ [5], [6]. The factor of M/n is best interpreted as the sensing energy per dimension, and $\log n$ is required to control the error rate across the n dimensions. Standard compressed sensing approaches based on random sensing matrices are known to achieve this lower bound while requiring on the order of $k \log n$ measurements [7].

In adaptive sensing, \mathbf{a}_i can be a function of y_1, \dots, y_{i-1} . The sensing vectors $\mathbf{a}_1, \dots, \mathbf{a}_m$ are *not* fixed prior to making observations, but instead depend on previous measurements. In this adaptive scenario, a necessary condition for sparse recovery is that the smallest non-zero entry exceed a constant times $\sqrt{\frac{n}{M} \log k}$ [8]. To the best of our knowledge, no proposed algorithms achieve this lower bound while also using only order $k \log n$ measurements.

A handful of adaptive sensing procedures have been proposed, some coming close to meeting the lower bound while requiring only order $k \log n$ measurements. Most recently, in [9], the authors propose an algorithm that guarantees exact recovery provided the smallest non-zero entry exceeds an unspecified constant times $\sqrt{\frac{n}{M} (\log k + \log \log_2 \log n)}$, coming within a triply logarithmic factor of the lower bound. While this triply logarithmic factor is not of practical concern, removing the suboptimal dependence on n is of theoretical interest. Reducing the leading constant, on the other hand, is of great practical concern.

¹Sufficient for exact support recovery with Alg. 1 if $x_i \geq 0, i = 1, \dots, n$ or if $k \leq \log n$; sufficient to recover any fixed fraction of support for any \mathbf{x} , k . See Theorems 1 and 2.

TABLE I
ASYMPTOTIC REQUIREMENTS FOR SUPPORT RECOVERY
SUBLINEAR SPARSITY ($\frac{k}{n} \rightarrow 0$)

	<i>non-adaptive</i>	<i>adaptive</i>
<i>direct</i>	1) $m \geq n$ 2) $x_{\min} > \sqrt{2 \frac{n}{M} \log n}$ necessary and sufficient traditional detection problem	1) $m > n$ 2) $x_{\min} > \sqrt{2 \frac{n}{M} \log k}$ necessary [10] sufficient [11]
<i>compressive</i>	1) $m \geq k \log n$ 2) $x_{\min} > \sqrt{C \frac{n}{M} \log n}$ necessary [5], [6] sufficient [7]	1) $m \geq k \log n$ 2) $x_{\min} > \sqrt{C \frac{n}{M} \log k}$ necessary [4], [8] sufficient using CASS ¹

Table I summarizes the necessary and sufficient conditions for exact support recovery in the compressive, direct, non-adaptive and adaptive settings. *Direct* or *un-compressed* refers to the setting where individual measurement are made of each component – specifically, each sensing vector is supported on only one index. Here, support recovery is a traditional detection problem, and the non-adaptive SNR requirement becomes apparent by considering the following. If we measure each element of \mathbf{x} once using sensing vectors that form an identity matrix when stacked together, then $M = n$, and the requirement implies the smallest non-zero entry be greater than $\sqrt{2 \log n}$. It is well known that the maximum of n i.i.d. Gaussians grows as $\sqrt{2 \log n}$; the smallest signal must exceed this value.

Adaptive direct sensing has recently received a fair amount of attention in the context of sparse recovery [11]–[16], and is closely related to traditional work in sequential statistical analysis. In this non-compressive setting, the number of measurements must be at least on the order of the dimension – i.e., $m \geq n$, as testing procedures must at a minimum measure each index once. In this setting, a number of works have shown that scaling of the SNR as $\log k$ is necessary and sufficient for exact recovery [11], [15].

To summarize, Table I highlights the potential gains from compressive and adaptive measurements. Allowing for compressive measurements can reduce the total number of measurements from n to $k \log n$, but does not relax the SNR requirement. Allowing for adaptive measurements does not reduce the total number of measurements required; instead, it can reduce the required SNR scaling from $\log n$ to $\log k$. Prior to this work, in the adaptive compressive case, it was unknown if $\log k$ scaling of the SNR was sufficient for support recovery.

B. Main Results and Contributions

The main theoretical contribution of this work is to complete Table I by introducing a simple adaptive compressed sensing procedure that 1) requires only order $k \log n$ measurements, and 2) succeeds provided the minimum non-zero entry is greater than a constant times $\sqrt{\frac{n}{M} \log k}$, showing this scaling is sufficient. To the best of our knowledge, this is the first

demonstration of a procedure that is optimal in terms of dependence on SNR and dimension.

Specifically, we propose a procedure termed Compressive Adaptive Sense and Search (CASS). For recovery of non-negative signals, the procedure succeeds in exact support recovery with probability greater than $1 - \delta$ provided the minimum non-zero entry of \mathbf{x} is greater than

$$\sqrt{20 \frac{n}{M} \left(\log k + \log \left(\frac{8}{\delta} \right) \right)},$$

requires exactly $m = 2k \log_2(n/k)$ measurements, and has total sensing energy $\sum_{i=1}^m \|\mathbf{a}_i\|_2^2 = M$. This result implies that as n grows large, if the minimum non-zero entry exceeds $\sqrt{20 \frac{n}{M} \log k}$, the procedure succeeds in exact support recovery.

When the signal contains both positive and negative entries, the CASS procedure guarantees the following. On average, the procedure succeeds in recovery of at least a fraction $1 - \epsilon$ of the non-zero components provided their magnitude is at least

$$\sqrt{20 \frac{n}{M} \left(\log k + 2 \log \left(\frac{8}{\epsilon} \right) \right)}.$$

In this setting the procedure requires approximately $2k \log_2(n/k) + 4k/\epsilon$ measurements and has total sensing energy $\sum_{i=1}^m \|\mathbf{a}_i\|_2^2 = M$.

In addition to these theoretical developments, we present a series of numerical results that show CASS outperforms standard compressed sensing for problems of even modest size. The numerical experiments reinforce the theoretical results – as the dimension of the problem grows, performance of the CASS algorithm does not deteriorate, confirming the logarithmic dependence on dimension has been removed. We also show side-by-side comparisons between standard compressed sensing, direct sensing, and CASS on approximately sparse signals.

C. Prior Work in Adaptive Compressed Sensing

A number of adaptive compressed sensing procedures have been proposed (see [9], [17]–[22] and references therein), some coming close to meeting the lower bound while requiring only order $k \log n$ measurements.

Most closely related to the CASS procedure proposed here, from an algorithmic perspective, are the procedures of Iwen [17], Iwen and Tewfik [19], [21], and the Compressive Binary Search of Davenport and Arias-Castro [20]. Much like CASS, these procedures rely on repeated bisection of the support of the signal. The first proposal of a bisecting search procedure applied to the adaptive compressed sensing problem, to the best of our knowledge, was [17]. The theoretical guarantees of this work when translated to the setting studied here ensure exact support recovery provided the non-zero entries of \mathbf{x} are greater than a constant times $\sqrt{\frac{n}{M} (\log^2 k + \log^2 \log n)}$ (see [21, Theorem 2] and more succinctly, [19, Theorem 2]).

Most recently, in [9], the authors propose an algorithm termed Sequential Compressed Sensing. The procedure is

based on random sensing vectors with sequentially masked support, allowing the procedure to focus sensing energy on the suspected non-zero components. The procedure guarantees exact recovery provided the smallest non-zero entry exceeds an unspecified constant times $\sqrt{\frac{n}{M} (\log k + \log \log_2 \log n)}$, coming within in a triply logarithmic factor of the theoretical bound lower bound.

While the bisection approach used by CASS is suggested in a number of other works, the distinguishing feature of the procedure is the allocation of the sensing energy. Roughly speaking, both procedures in [17], and later in [20], suggest an allocation of sensing energy across the bisecting steps of the procedure that ensure the probability a mistake is made at any step is the same. The CASS procedure instead uses less sensing energy for initial steps, and more sensing energy at later steps, reducing the the probability the procedure makes an error from step to step. This allocation of the sensing energy removes any sub-optimal dependence on dimension from the SNR requirement; the CASS procedure succeeds provided the minimum non-zero entry exceeds $\sqrt{20 \frac{n}{M} \log k}$.

D. Notation

Notation adopted throughout, in general, follows convention. Bold face letters, \mathbf{x} , denote vectors (both random and deterministic), matrices are denoted with capitol letters, for example, A , while calligraphic font, such as \mathcal{S} , denotes a set or an event. The indicator function $\mathbb{I}_{\{\mathcal{E}\}}$ is equal to one if the event \mathcal{E} occurs, and equals zero otherwise. The vector $\mathbf{1}_{\{\mathcal{S}\}}$, for any support set \mathcal{S} , is a vector of ones over \mathcal{S} , and 0 elsewhere. Expectation is denoted $\mathbb{E}[\cdot]$ and the probability of an event $\mathbb{P}(\cdot)$.

II. PROBLEM DETAILS

Consider the canonical compressed sensing setup in which a sparse signal $\mathbf{x} \in \mathbb{R}^n$ supported on $\mathcal{S} \subset \{1, \dots, n\}$, $|\mathcal{S}| = k$, is measured through m linear projections as in (1). In matrix-vector form,

$$\mathbf{y} = A\mathbf{x} + \mathbf{z}$$

where $\mathbf{z} \in \mathbb{R}^m \sim \mathcal{N}(0, I)$ and $A \in \mathbb{R}^{m \times n}$ is the sensing matrix with total sensing energy constraint

$$\|A\|_{\text{fro}}^2 \leq M. \quad (3)$$

The total sensing energy constraint reflects a limit on the total amount of energy or time available to the sensing device, as in [9], [20]. Importantly, it results in a decoupling between the energy constraint and the number of measurements, allowing for fair comparison between procedures that take a different number of measurements. In adaptive sensing the rows of A , denoted \mathbf{a}_i^T , are designed sequentially and adaptively based on prior observations, y_1, \dots, y_{i-1} , as opposed to the non-adaptive setting in which \mathbf{a}_i , $i = 1, \dots, m$ are fixed a-priori.

The error metrics considered are the probability of exact support recovery, or the family wise error rate,

$$\mathbb{P}(\hat{\mathcal{S}} \neq \mathcal{S})$$

and the symmetric set difference

$$d(\hat{\mathcal{S}}, \mathcal{S}) = \left| \{\mathcal{S} \setminus \hat{\mathcal{S}}\} \cup \{\hat{\mathcal{S}} \setminus \mathcal{S}\} \right|$$

where $\hat{\mathcal{S}}$ is an estimate of \mathcal{S} . Both depend on the procedure, n , k , M and the amplitude of the non-zero entries of \mathbf{x} . In particular, we quantify performance in terms of the minimum non-zero amplitude of \mathbf{x} , denoted x_{\min} , defined as

$$x_{\min} = \min_{i \in \mathcal{S}} |x_i|.$$

In general, no assumption is made on the support set \mathcal{S} other than $|\mathcal{S}| = k$. Theorem 2 requires the sparse support be chosen uniformly at random over all possible cardinality k support sets, which can be avoided by randomly permuting the support before running the algorithm.

Lastly, throughout the paper we assume that both n and k are powers of two for simplicity of presentation. This can of course be relaxed, at the cost of increased constants.

III. COMPRESSIVE ADAPTIVE SENSE AND SEARCH

Conceptually, the CASS procedure operates by initially dividing the signal into a number of partitions, and then using compressive measurements to test for the presence of one or more non-zero elements in each partition. The procedure continues its search by bisecting the most promising partitions, with the goal of returning the k largest components of the vector. This compressive sensing and bisecting search approach inspired the name of the algorithm.

A. The CASS Procedure

The CASS procedure is detailed in Alg. 1. The procedure requires three inputs: 1) k , the number of non-zero coefficients to be returned by the procedure, 2) M , the total budget, and 3) an initial scale parameter ϵ , which is in most practical scenarios is set to $\epsilon = 1$.

Since the procedure is based on repeated bisections of the signal, it is helpful to define the dyadic subintervals of $\{1, \dots, n\}$. The dyadic subintervals, $\mathcal{J}_{j,\ell} \subset \{1, \dots, n\}$, are given by

$$\mathcal{J}_{j,\ell} = \left\{ \left\lfloor \frac{(\ell-1)n}{2^j} \right\rfloor + 1, \dots, \left\lfloor \frac{\ell n}{2^j} \right\rfloor \right\},$$

$$j = 0, 1, \dots, \log_2 n, \quad \ell = 1, \dots, 2^j$$

where j indicates the scale and ℓ the location of the dyadic partition.

The procedure consists of a series of s_0 steps. Conceptually, on the first step, $s = 1$, the procedure begins by dividing the signal into ℓ_0 partitions, where ℓ_0 is the smaller of n or the next power of two greater than $4k/\epsilon$. These initial partitions are given by the dyadic sub-intervals of $\{1, \dots, n\}$ at scale $\log_2(\ell_0)$:

$$\mathcal{J}_{\log_2(\ell_0),\ell} \quad \ell = 1, \dots, \ell_0.$$

The procedure measures the signal with ℓ_0 sensing vectors, each with support over a single dyadic sub-interval. The procedure then selects the largest k measurements in absolute

value. These k measurements are used to define the support of the sensing vectors used on step two of the procedure. More specifically, the supports of the sensing vectors corresponding to the k largest measurements are bisected, giving $2k$ support sets. These $2k$ support sets define the support of the sensing vectors on step $s = 2$.

The procedure continues in the fashion, taking $2k$ measurements on each step (after the initial step, which requires ℓ_0 measurements) and bisecting the support of the k largest measurements to define the support of the sensing vectors on the next step. On the final step, $s = s_0$, where $s_0 = \log_2 \frac{n}{\ell_0} + 1$, the support of the sensing vectors consists of a single index; the procedure returns the support of the sensing vectors corresponding to the k largest measurements as the estimate of \mathcal{S} and terminates. Additionally, estimates of the values of the non-zero coefficients are returned based on the measurements made on the last step.

One of the key aspects of the CASS algorithm is the allocation of the measurement budget across the steps of the procedure. On step s , the amplitude of the non-zero entries of the sensing vectors is given as

$$\sqrt{\frac{Ms}{\gamma n}} \quad (4)$$

where γ is an internal parameter of the procedure used to ensure the total sensing energy constraint in (3) is satisfied with equality. Specifically,

$$\gamma = 1 + \frac{4k}{\ell_0} \sum_{s=1}^{s_0} s 2^{-s}.$$

This particular allocation of the sensing energy gives the CASS procedure its optimal theoretical guarantees. While the amplitude of each sensing vector grows in a polynomial manner with each step, the support of the sensing vectors decreases geometrically; this results in a decrease in the sensing energy across steps, but an increase in reliability.

Details of the CASS procedure are found in Alg. 1, and performance is quantified in the following section.

B. Theoretical Guarantees

The theoretical guarantees of CASS are presented in Theorems 1 and 2. Theorem 1 presents theoretical guarantees for non-negative signals, while Theorem 2 quantifies performance when the signal is both positive and negative.

Theorem 1. Assume $x_i \geq 0$ for all $i = 1, \dots, n$. Set $\epsilon = 1$. For $\delta \in (0, 1]$, Alg. 1 has $\mathbb{P}(\hat{\mathcal{S}} \neq \mathcal{S}) \leq \delta$ provided

$$x_{\min} \geq \sqrt{20 \frac{n}{M} \left(\log k + \log \left(\frac{8}{\delta} \right) \right)},$$

has total sensing energy $\|\mathbf{A}\|_{\text{fro}}^2 = M$ and uses $m = 2k \log_2 \frac{n}{k}$ measurements.

Proof: See Appendix A. ■

The theoretical guarantees of Theorem 1 do not apply to recovery of signals with positive and negative entries, as

Algorithm 1 CASS

input: number of terms in approximation k ,
budget M ,
initial scale parameter $\epsilon \in (0, 1]$ – (default $\epsilon = 1$)
initialize: $\ell_0 = \min \{4 \cdot 2^{\lceil \log_2(k/\epsilon) \rceil}, n\}$ partitions
 $s_0 = \log_2 \frac{n}{\ell_0} + 1$ steps
 $\gamma = 1 + 4k/\ell_0 \sum_{s=2}^{s_0} s2^{-s}$
 $\mathcal{L} = \{1, 2, \dots, \ell_0\}$
for $s = 1, \dots, s_0$ **do**
 for $\ell \in \mathcal{L}$ **do**
 $\mathbf{a}_{s,\ell} = \sqrt{\frac{Ms}{\gamma n}} \mathbf{1}_{\{\mathcal{J}_{\log_2(\ell_0)-1+s,\ell}\}}$
 measure: $y_{s,\ell} = \langle \mathbf{a}_{s,\ell}, \mathbf{x} \rangle + z_{s,\ell}$ $z_{s,\ell} \stackrel{iid}{\sim} \mathcal{N}(0, 1)$
 end for
 let $\ell_1, \ell_2, \dots, \ell_{|\mathcal{L}|}$ be such that
 $|y_{s,\ell_1}| \geq |y_{s,\ell_2}| \geq \dots \geq |y_{s,\ell_{|\mathcal{L}|}}|$
 if $s \neq s_0$ **then**
 $\mathcal{L} = \{2\ell_1 - 1, 2\ell_1, 2\ell_2 - 1, 2\ell_2, \dots, 2\ell_k - 1, 2\ell_k\}$
 else
 $\hat{\mathcal{S}} = \{\ell_1, \ell_2, \dots, \ell_k\}$
 $\hat{x}_i = y_{s,i} \cdot \sqrt{\frac{\gamma n}{Ms}}$ for all $i \in \hat{\mathcal{S}}$
 end if
end for
output: $\hat{\mathcal{S}}$ (k indices), estimates \hat{x}_i for $i \in \hat{\mathcal{S}}$

scenarios can arise where two non-zero components can cancel when measured by the same sensing vector. In order to avoid this effect, the *initial scale parameter*, ϵ , can be set to a value less than one. Doing so increases the number of partitions that are created on the first step, but reduces the occurrence of positive and negative signal cancellation. Theorem 2 bounds the expected fraction of the components that are recovered when the signal contains positive and negative entries.

Theorem 2. Assume \mathcal{S} is chosen uniformly at random. For any $\epsilon \in (0, 1]$ Alg. 1 has $\mathbb{E}[d(\hat{\mathcal{S}}, \mathcal{S})] \leq k\epsilon$ provided

$$x_{\min} \geq \sqrt{20 \frac{n}{M} \left(\log k + 2 \log \left(\frac{8}{\epsilon} \right) \right)},$$

has total sensing energy $\|A\|_F^2 = M$ and has

$$m \leq \frac{8k}{\epsilon} + 2k \log_2 \left(\frac{n}{k} \right)$$

measurements.

Proof: See Appendix B. ■

C. Discussion

Theorem 2 implies conditions under which a fraction $1 - \epsilon$ of the support can be recovered in expectation. Provided $\epsilon \geq 1/\log n$, the procedure requires order $k \log n$ measurements. In scenarios where one wishes to control $\mathbb{P}(\hat{\mathcal{S}} \neq \mathcal{S})$, ϵ can be set to satisfy $\epsilon < 1/k$. In this case, while the SNR requirement remains essentially unchanged, the procedure would require on the order of k^2 measurements. In this sense, the compressive case with positive and negative non-zero values is more

difficult, a phenomena noted in other work [23]. Note that when $k \geq \log n$, the theoretical guarantee of Theorem 2 adapted to control the family wise error rate requires more than order $k \log n$ measurements.

The minimax lower bound of [8, Proposition 4.2] states the following (over a slightly larger class of signals with sparsity k , $k - 1$, and $k + 1$, with $k < n/2$): if $\mathbb{E}[d(\hat{\mathcal{S}}, \mathcal{S})] \leq \epsilon$ then necessarily

$$x_{\min} \geq \sqrt{2 \frac{n}{M} \left(\log k + \log \frac{1}{2\epsilon} - 1 \right)}.$$

Since the condition $\mathbb{P}(\hat{\mathcal{S}} \neq \mathcal{S}) \leq \delta$ in Theorem 1 implies $\mathbb{E}[d(\hat{\mathcal{S}}, \mathcal{S})] \leq 2k\delta$, the upper and lower bounds can be compared directly by setting $\delta = \epsilon/2k$. In the same manner, Theorem 2 can be directly compared. Doing so shows the theorems are tight to within constant factors.

The splitting or bisecting approach in CASS is a common technique for a number search problems, including adaptive sensing. As discussed in Sec. I-C, to the best of our knowledge, the first work to propose a bisecting search for the compressed sensing problem was [17]. When translated to the framework studied here, the authors in essence allocate the sensing vectors on each bisecting step to be proportional to $\sqrt{M/n}$, as opposed to increasing with the index of the step as in (4). Likewise, the *compressive binary search* (CBS) algorithm [20] also employs sensing vectors that are proportional to $\sqrt{M/n}$. The CBS procedure finds a single non-zero entry with vanishing probability of error as n gets large provided $x_{\min} \geq \sqrt{8 \frac{n}{M} \log \log_2 n}$, where x_{\min} is the amplitude of the non-zero entry. The authors of [20] rightly question whether the $\log \log_2 n$ term is needed. Thm. 1 answers this question, removing the doubly logarithmic term and coming within a constant factor of the lower bound in [20]. A more direct comparison to the work in [20] is found in [24].

Increasing the number of partitions to further isolate sparse entries is an idea suggested in [17], [21]. In particular, the authors suggest a procedure that creates a large random partitions based on primes, essentially decreasing the support of the sensing vector such that each non-zero element is isolated with arbitrarily high probability.

IV. NUMERICAL EXPERIMENTS

This section presents a series of numerical experiments aimed at highlighting the performance gains of adaptive sensing. In all experiments, the CASS procedure was implemented according to Alg. 1 with inputs as specified. The signal to noise ratio is defined as

$$\text{SNR (dB)} = 10 \log_{10} \left(\frac{x_{(k)}^2 M}{n} \right)$$

where $x_{(k)}$ is the amplitude of the k th largest element of \mathbf{x} . This definition reflects the SNR of measuring the k th largest element using a fraction $1/n$ of the total sensing energy. In all experiments, Gaussian noise of unit variance is added to each measurement, as specified in (1).

Fig. 1 shows empirical performance of CASS for 1-sparse recovery, and, for comparison, 1) traditional compressed sensing with orthogonal matching pursuit (OMP) using a Gaussian ensemble with normalized columns and 2) direct sensing. As the dimension of the problem is increased, notice that performance of CASS remains constant, while the error probability of both OMP and direct sensing increase. Note for non-adaptive methods, in the one sparse case OMP is equivalent to the maximum likelihood support estimator, and is thus optimal amongst non-adaptive recovery algorithms. The total sensing energy is the same in all cases, allowing for fair comparison. For sufficiently large n , CASS outperforms direct sensing.

Figs. 2 and 3 show performance of CASS for k -sparse signals with equal magnitude positive and negative non-zero entries. Performance is in terms of empirical value of $d(\hat{\mathcal{S}}, \mathcal{S})$ averaged over the trials. The support of the test signals was chosen uniformly at random from the set of k -sparse signals, and the non-zero entries were assigned an amplitude of $+x_{\min}$ or $-x_{\min}$ at random. The cancellation effect results in an error floor of around $d(\hat{\mathcal{S}}, \mathcal{S})/(2k) = 0.10$; this is greatly reduced when the scale parameter is set to $\epsilon = 1/8$, clearly visible in both Fig. 2 and Fig. 3. The procedure is compared against traditional compressed sensing using a random Gaussian ensemble, recovered with LASSO and a regularizer tuned to return a k -sparse signal (using SpaRSA, [25]). Performance of traditional compressed sensing with $m = 2k \log(n/k)$, fairly compared against CASS with $\epsilon = 1$, is shown in both figures. Traditional compressed sensing with $m = 4k/\epsilon + 2k \log_2(n\epsilon/k)$ is also shown, which is more fairly compared to CASS with $\epsilon = 1/8$.

In Fig. 3 notice that performance of CASS remains constant as n becomes large. For sufficiently large n , CASS outperforms direct sensing, highlighting a major advantage of adaptive sensing: for support recovery, while standard compressed sensing does not outperform direct sensing in terms of dependence on SNR, for sufficiently sparse problems, CASS does. As solving LASSO with a dense sensing matrix for large problem sizes becomes computationally prohibitive, performance was only evaluated up to $n = 2^{18}$. For all dimensions evaluated, compressed sensing using LASSO was inferior to CASS and direct sensing.

Fig. 4 shows recovery of a test image from [26] for three SNRs. The image is *approximately* sparse in the Daubechies wavelet domain ('db4' in MATLAB). Alg. 1 was run with $k = 256$ as input and compared against LASSO with a random Gaussian ensemble, and the regularizer was tuned to return $k = 256$ elements (again using SpaRSA, [25]). Using traditional compressed sensing, at an SNR of -8 dB, the reconstructed image is essentially un-recognizable; at the same SNR, the image recovered with CASS is still very much recognizable. The peak-signal-to-noise ratio (PSNR) for images on $[0, 1]$ is defined as

$$\text{PSNR (dB)} = -10 \log_{10} \left(\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2}{n} \right)$$

where $\hat{\mathbf{x}}$ is an estimate of the image.

TABLE II
HANDWRITTEN EIGHT, 10 TRIALS, $n = 256^2$, $k = 256$, $\epsilon = 1$

	CASS (Alg. 1)	CS (LASSO)	Direct
SNR = -2 dB			
$d(\hat{\mathcal{S}}, \mathcal{S})$	120.0	153.2	131.0
PSNR (dB)	16.88	8.43	12.71
runtime (s)	0.11	118.3	0.04
SNR = -8 dB			
$d(\hat{\mathcal{S}}, \mathcal{S})$	152.0	202.0	163.8
PSNR (dB)	12.91	4.40	6.15
runtime (s)	0.11	65.0	0.04
SNR = -14 dB			
$d(\hat{\mathcal{S}}, \mathcal{S})$	236.4	379.6	301.6
PSNR (dB)	6.69	2.52	-1.76
runtime (s)	0.11	54.2	0.03

The experiments of Fig. 4 were repeated 10 times, and the empirical results averaged over the 10 trials are shown in Table II. In terms of PSNR, CASS outperforms compressed sensing by 4-8 dB. The runtime of the simulation, which consists of the entire simulation time, including both measurement and recovery, is included. While the compressed sensing experiments were not optimized for speed, it seems unlikely any recovery method using dense Gaussian matrices would approach the runtime of CASS – standard compressed sensing runtime was on the order of a minute, while runtime of CASS was approximately 0.1 seconds.

Fig. 5 shows side-by-side performance of CASS, traditional compressed sensing with a dense Gaussian sensing matrix and LASSO tuned to return $k = 2048$ non-zero components, and direct sensing on a natural image. In terms of PSNR, CASS outperforms compressed sensing by 2-4 dB. In terms of symmetric set difference, CASS substantially outperforms compressed sensing on all SNRs evaluated.

Fig. 6 shows results of running CASS on the 20 image classes in the Microsoft Research Object Class Recognition database [27]. Each class consists of approximately 30 images of size 128×128 . Results were averaged over all images in a particular class. Direct sensing, standard compressed sensing using a Gaussian Ensemble and the group LASSO (G-LASSO) of [28], [29] which exploits signal structure, are included. G-LASSO is known to be one of the best non-adaptive methods for recovery of natural images, making it a natural benchmark for comparison (see [30, Fig. 5]). The total sensing energy was the same in all cases, and all methods used $m = 5120$ measurements (except direct sensing which requires $m = 16,384$ measurements). Each algorithm was tuned to output a 512 coefficient approximation. CASS universally outperforms both G-LASSO and standard CS on all image sets and SNRs evaluated.

A. Discussion

The leading constant in the upper bound – a factor of 20 required by the CASS procedure – is significant in practice. Roughly speaking, traditional direct sensing requires $x_{\min} \geq \sqrt{2 \frac{n}{M} \log n}$, while CASS requires $x_{\min} \geq \sqrt{20 \frac{n}{M} \log k}$. One would then conclude that anytime the signal is sufficiently sparse, specifically, if $k \leq n^{1/10}$, the CASS procedure should

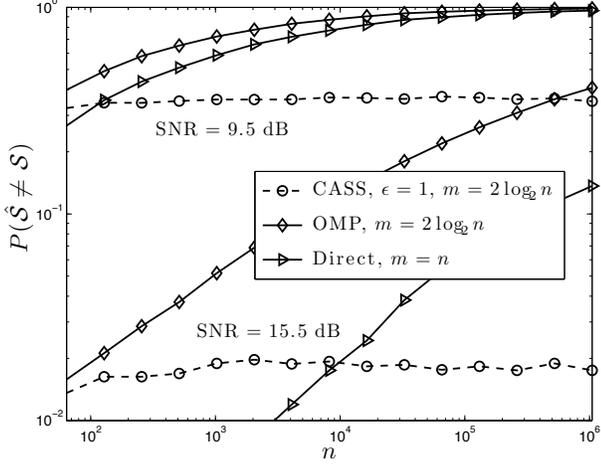


Fig. 1. Recovery of one sparse signal ($k = 1$). Empirical performance of CASS (Alg. 1), traditional compressed sensing with orthogonal matching pursuit (OMP), and direct sensing as a function of n , for $\text{SNR}(\text{dB}) = 10 \log_{10}(x_{\min}^2 M/n) = 9.5$ dB and $\text{SNR} = 15.5$ dB. 10,000 trials for each n .

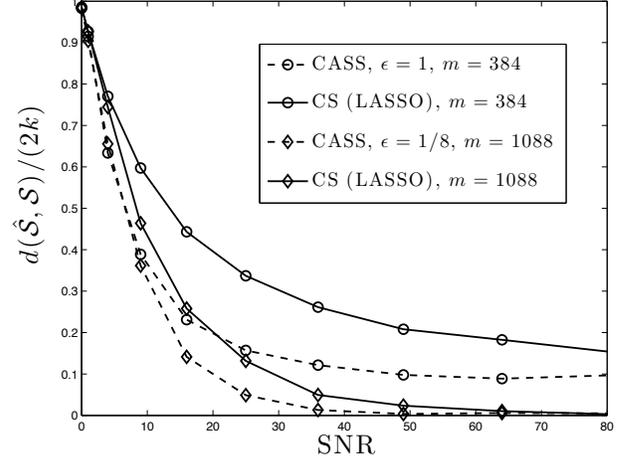


Fig. 2. Empirical performance of CASS (Alg. 1) with non-zero entries of \mathbf{x} positive and negative of equal amplitude as a function of SNR compared against traditional compressed sensing with Gaussian ensemble and recovery using LASSO, $k = 32$, $n = 2048$. $\text{SNR} = x_{\min}^2 M/n$. 100 trials for each SNR (linear units).

outperform direct sensing. Notice that in simulation CASS outperforms direct sensing on signals where $k \geq n^{1/10}$, implying the leading constant of 20 is loose.

V. CONCLUSION

This work presented an adaptive compressed sensing and group testing procedure for recovery of sparse signals in additive Gaussian noise, termed CASS. The procedure was shown to be near-optimal in that it comes within a constant factor of the best possible dependence on SNR, while requiring on the order of $k \log n$ measurements. Simulation showed that the theoretical developments hold in practice, including scenarios in which the signals is only approximately sparse.

While many of the questions regarding the gains of adaptivity for sparse support recovery have been settled, small questions remain. First, it would be of practical interest to further reduce the leading constant in the SNR requirement of CASS. The question of whether or not *exact* support recovery is possible in full generality (positive and negative non-zero entries, any level of sparsity), with SNR scaling as $\log k$, and the number of measurements scaling as $k \log n$, is still outstanding.

APPENDIX A

Proof of Theorem 1: The total number of measurements is given as

$$m = \ell_0 + 2k(s_0 - 1) = 2k \log_2 \left(\frac{n}{k} \right)$$

since the first step requires ℓ_0 measurements, and the remaining $s_0 - 1$ steps each require $2k$ measurements. The equality

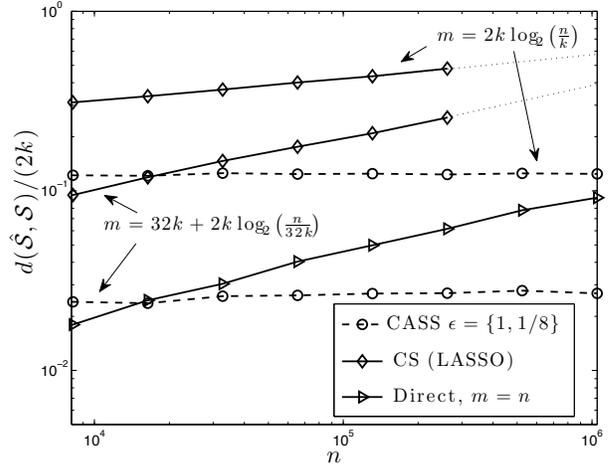


Fig. 3. Non-zero entries of equal magnitude, positive and negative, as a function of n . $k = 32$. CASS Alg. 1 with $\epsilon = 1$, $m = 2k \log_2(n/k)$, and $\epsilon = 1/8$, $m = 32k + 2k \log_2(\frac{n}{32k})$. LASSO using Gaussian ensemble, $m = 2k \log_2(n/k)$, $m = 32k + 2k \log_2(\frac{n}{32k})$. $\text{SNR} = 15.5$ dB. 1000 trials for each n . LASSO was not evaluated for $n \geq 2^{19}$ because of computational intensity.

follows as $\ell_0 = 4k$. The constraint in (3) is satisfied:

$$\begin{aligned} \|A\|_{\text{fro}}^2 &= \sum_{\ell=1}^{\ell_0} \|a_{1,\ell}\|^2 + \sum_{s=2}^{s_0} \sum_{\ell=1}^{2k} \|a_{s,\ell}\|^2 \\ &= \ell_0 \left(\frac{n}{\ell_0} \right) \left(\frac{M}{\gamma n} \right) + 2k \sum_{s=2}^{s_0} \left(\frac{n}{\ell_0 2^{s-1}} \right) \left(\frac{Ms}{\gamma n} \right) \\ &= \frac{M}{\gamma} \left(1 + \frac{4k}{\ell_0} \sum_{s=2}^{s_0} s 2^{-s} \right) = M \end{aligned} \quad (5)$$

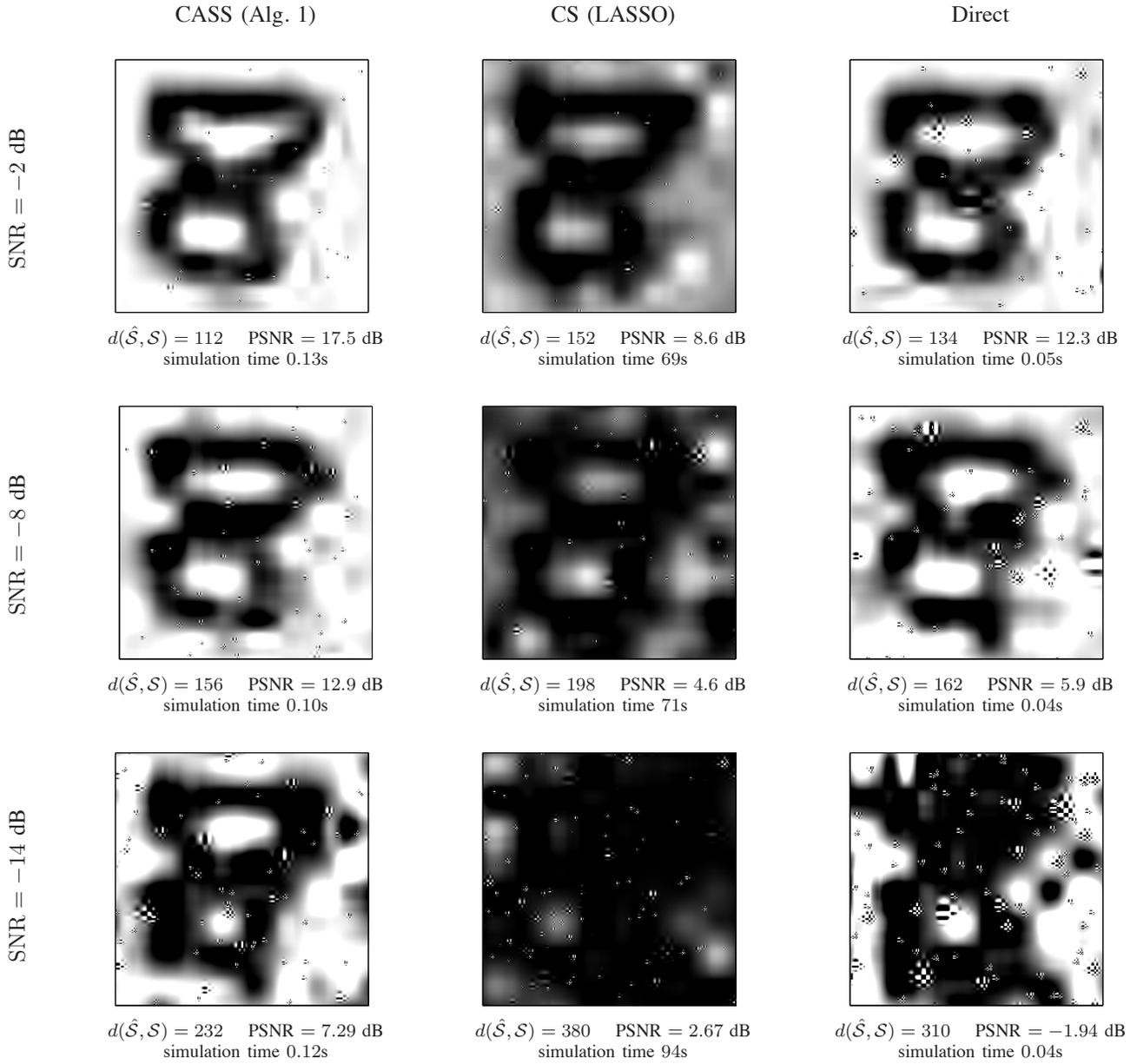


Fig. 4. Handwritten eight [26], approximately sparse in ‘dB4’ wavelet basis, $n = 256^2$. Recovery with CASS (Alg. 1), $k = 256$, $\epsilon = 1$, as input, $m = 4608$ (column 1). Traditional compressed sensing with Gaussian ensemble and LASSO (using SpaRSA [25]), tuned to return $k = 2048$ components, $m = 4608$ (column 2). Direct sensing, $m = n$, using best $k = 256$ term approximation (column 3). Each method evaluated for three SNRs, $\text{SNR}(\text{dB}) = 10 \log_{10}(x_{(k)}^2 M/n)$, where $x_{(k)}$ is the amplitude of the k th largest component.

where the second equality follows as each sensing vector has support $n2^{-(s-1)}/\ell_0$ and amplitude $Ms/(\gamma n)$ on step s .

A sufficient conditions for exact recovery of the support set is that the procedure never incorrectly eliminate a measurement corresponding to a non-zero entry on any of the s_0 steps. Let $y_{s,1}, \dots, y_{s,t_s}$ be measurements corresponding to sensing vectors with one or more non-zero entries in their support, and y_{s,t_s+1}, \dots be measurements corresponding to the zero entries on step s . Consider a series of thresholds denoted τ_s , $s = 1, \dots, s_0$. If $|y_{s,1}|, \dots, |y_{s,t_s}|$ are all greater than τ_s , and $|y_{s,t_s+1}|, \dots$ are all less than τ_s , for all s , the

sufficient condition is implied. Taking the complement of this event, and from a union bound,

$$\mathbb{P}(\hat{\mathcal{S}} \neq \mathcal{S}) \leq \mathbb{P} \left(\bigcup_{j=1}^{t_1} \{|y_{1,j}| \leq \tau_1\} \cup \bigcup_{j=t_1+1}^{\ell_0} \{|y_{1,j}| \geq \tau_1\} \right) + \sum_{s=2}^{s_0} \mathbb{P} \left(\bigcup_{j=1}^{t_s} \{|y_{s,j}| \leq \tau_s\} \cup \bigcup_{j=t_s+1}^{2k} \{|y_{s,j}| \geq \tau_s\} \right)$$

for any series of thresholds τ_s . As all non-zero elements are positive by assumption, and the signals combine constructively, it is straightforward to see $y_{s,1}, \dots, y_{s,t_s}$ are normally



Fig. 5. Cameraman, approximately sparse in Haar basis, $n = 256^2$. Recovery with CASS, Alg. 1, $k = 2048$, $\epsilon = 1$, as input, $m = 20480$ (column 1). Traditional compressed sensing with Gaussian ensemble and recovery with LASSO (using SpaRSA [25]), regularizer tuned to return $k = 2048$ components, $m = 20480$ (column 2). Direct sensing, $m = n$, using best $k = 2048$ term approximation (column 3). Each method evaluated for three SNRs, $\text{SNR}(\text{dB}) = 10 \log_{10}(x_{(k)}^2 M/n)$, where $x_{(k)}$ is the amplitude of the k th largest component.

distributed with mean greater than $x_{\min} a_s$ and unit variance. On the other hand, y_{s, t_s+1}, \dots are normally distributed with mean zero and unit variance. Setting the thresholds at $\tau_s = x_{\min} a_s / 2$, where a_s is the amplitude of the non-zero entries of the sensing vector on step s , gives the series of equations in (6) - (8). Here, (6) follows from the union bound. Equation (7) follows from a standard Gaussian tail bound: $1 - F_{\mathcal{N}}(x) \leq \frac{1}{2} \exp(-x^2/2)$ for $x \geq 0$, where $F_{\mathcal{N}}(\cdot)$ is the Gaussian cumulative density function. Lastly, (8) follows as γ

is always less than $5/2$:

$$\gamma = 1 + 4k/\ell_0 \sum_{s=2}^{s_0} s 2^{-s} \leq \frac{5}{2} \quad (8)$$

as the sum is arithmetico-geometric, and $\ell = 4k$. Setting

$$x_{\min} \geq \sqrt{20 \frac{n}{M} \left(\log k + \log \left(\frac{8}{\delta} \right) \right)}$$

gives $\mathbb{P}(\hat{\mathcal{S}} \neq \mathcal{S}) \leq \sum_{s=1}^{s_0} (2/\delta)^{-s} \leq \delta$, completing the proof.

$$\begin{aligned} \mathbb{P}(\hat{\mathcal{S}} \neq \mathcal{S}) &\leq \sum_{j=1}^{t_1} \mathbb{P}\left(|y_{1,j}| \leq \frac{x_{\min} a_1}{2}\right) + \sum_{j=t_1+1}^{4k} \mathbb{P}\left(|y_{1,j}| \geq \frac{x_{\min} a_1}{2}\right) \\ &\quad + \sum_{s=2}^{s_0} \left(\sum_{j=1}^{t_s} \mathbb{P}\left(|y_{s,j}| \leq \frac{x_{\min} a_s}{2}\right) + \sum_{j=t_s+1}^{2k} \mathbb{P}\left(|y_{s,j}| \geq \frac{x_{\min} a_s}{2}\right) \right) \end{aligned} \quad (6)$$

$$\leq 4k \exp\left(-\frac{x_{\min}^2 a_1^2}{8}\right) + \sum_{s=2}^{s_0} 2k \exp\left(-\frac{x_{\min}^2 a_s^2}{8}\right) \quad (7)$$

$$\leq \sum_{s=1}^{s_0} \exp\left(-\frac{x_{\min}^2 M s}{8\gamma n} + \log 4k\right) \leq \sum_{s=1}^{s_0} \exp\left(-\frac{x_{\min}^2 M s}{20n} + \log 4k\right) \quad (8)$$

APPENDIX B

Proof of Theorem 2.

First, we confirm that the sensing budget is satisfied. In the same manner as before

$$\begin{aligned} \|A\|_{\text{fro}}^2 &= \sum_{\ell=1}^{\ell_0} \|\mathbf{a}_{1,\ell}\|^2 + \sum_{s=2}^{s_0} \sum_{\ell=1}^{2k} \|\mathbf{a}_{s,\ell}\|^2 \\ &= \frac{M}{\gamma} \left(1 + \frac{4k}{\ell_0} \sum_{s=2}^{s_0} s 2^{-s} \right) = M. \end{aligned}$$

The total number of measurement satisfies

$$m = \ell_0 + 2k(s_0 - 1) \leq 8k/\epsilon + 2k \log_2(n/k)$$

where the inequality follows as $\min\{n, 4k/\epsilon\} \leq \ell_0 \leq 8k/\epsilon$.

We proceed by bounding the expected number of non-zero components that are isolated when the signal is partitioned, and then show these indices are returned in $\hat{\mathcal{S}}$ with high probability. We assume the support of the signal is chosen uniformly at random from all possible k -sparse supports sets (note this is equivalent to randomly permuting the labels of all indices). Conceptually, the algorithm divides the signal support into ℓ_0 disjoint support sets, where $\ell_0 = \min\{n, \text{next power of } 2 \text{ greater than } 4k/\epsilon\}$. Let the set $\mathcal{I} \subset \{1, \dots, \ell_0\}$ index the disjoint support sets that contain exactly one index corresponding to a non-zero entry:

$$\mathcal{I} := \{\ell \in \{1, \dots, \ell_0\} : |\mathcal{I}_{\log_2 \ell_0, \ell} \cap \mathcal{S}| = 1\}.$$

Then

$$\begin{aligned} \mathbb{E}[|\mathcal{I}|] &= \mathbb{E}\left[\sum_{\ell=1}^{\ell_0} \mathbb{I}_{\{\ell \in \mathcal{I}\}}\right] = \sum_{\ell=1}^{\ell_0} \mathbb{E}[\mathbb{I}_{\{\ell \in \mathcal{I}\}}] \\ &= \sum_{\ell=1}^{\ell_0} \mathbb{P}(\ell \in \mathcal{I}) = \sum_{\ell=1}^{\ell_0} \frac{k}{\ell_0} \left(\frac{\ell_0 - 1}{\ell_0}\right)^{k-1} \\ &= k \left(\frac{\ell_0 - 1}{\ell_0}\right)^{k-1} \geq k \frac{\ell_0 - k}{\ell_0} \end{aligned}$$

where $\mathbb{I}_{\{\cdot\}}$ is the indicator function. Since $\ell_0 \geq \min\{4k/\epsilon, n\}$ we have

$$\mathbb{E}[|\mathcal{I}|] \geq k(1 - \epsilon/4). \quad (9)$$

Let \mathcal{E} be the event that the procedure fails to find at least as many true non-zero elements as the number that were isolated on the first partitioning, specifically,

$$\mathcal{E} = \{|\hat{\mathcal{S}} \cap \mathcal{S}| < |\mathcal{I}|\}.$$

Additionally, notice that since the procedure returns exactly k indices,

$$\mathcal{E}^c \Rightarrow d(\hat{\mathcal{S}}, \mathcal{S}) \leq 2(k - |\mathcal{I}|). \quad (10)$$

We can bound the probability of the event \mathcal{E} in a fashion similar to the previous proof. Consider a threshold at step s given as $\tau_s = x_{\min} a_s/2$. Assume that $y_{s,1}, y_{s,2}, \dots, y_{s,t'_s}$ are measurements corresponding to isolated components, $y_{s,t'_s+1}, \dots, y_{s,t_s}$ correspond to measurements that contain more than one non-zero index, and y_{s,t_s+1}, \dots correspond to noise only measurements. If all noise only measurements (y_{s,t_s+1}, \dots) are below the threshold in absolute value, and all isolated non-zero components are above the threshold ($y_{s,1}, \dots, y_{s,t'_s}$), for all s , this implies \mathcal{E} does not occur, regardless of $y_{s,t'_s+1}, \dots, y_{s,t_s}$. As before, after applying a union bound,

$$\begin{aligned} \mathbb{P}(\mathcal{E}) &\leq \sum_{j=1}^{t'_1} \mathbb{P}(|y_{1,j}| \leq \tau_1) + \sum_{j=t_1+1}^{\ell_0} \mathbb{P}(|y_{1,j}| \geq \tau_1) \\ &\quad + \sum_{s=2}^{s_0} \left(\sum_{j=1}^{t'_s} \mathbb{P}(|y_{s,j}| \leq \tau_s) + \sum_{j=t_s+1}^{2k} \mathbb{P}(|y_{s,j}| \geq \tau_s) \right) \\ &\leq \ell_0 \exp\left(-\frac{x_{\min}^2 a_1^2}{8}\right) + \sum_{s=2}^{s_0} 2k \exp\left(-\frac{x_{\min}^2 a_s^2}{8}\right) \\ &\leq \sum_{s=1}^{s_0} \exp\left(-\frac{x_{\min}^2 a_s^2}{8} + \log\left(\frac{8k}{\epsilon}\right)\right) \end{aligned}$$

where the inequality follows from a standard Gaussian tail bound, and the last line follows as $\ell_0 \leq 8k/\epsilon$. Setting $a_s = \sqrt{Ms/\gamma n}$ gives

$$\begin{aligned} \mathbb{P}(\mathcal{E}) &\leq \sum_{s=1}^{s_0} \exp\left(-\frac{x_{\min}^2 M s}{8\gamma n} + \log\left(\frac{8k}{\epsilon}\right)\right) \\ &\leq \sum_{s=1}^{s_0} \exp\left(-\frac{x_{\min}^2 M s}{20n} + \log\left(\frac{8k}{\epsilon}\right)\right) \end{aligned}$$

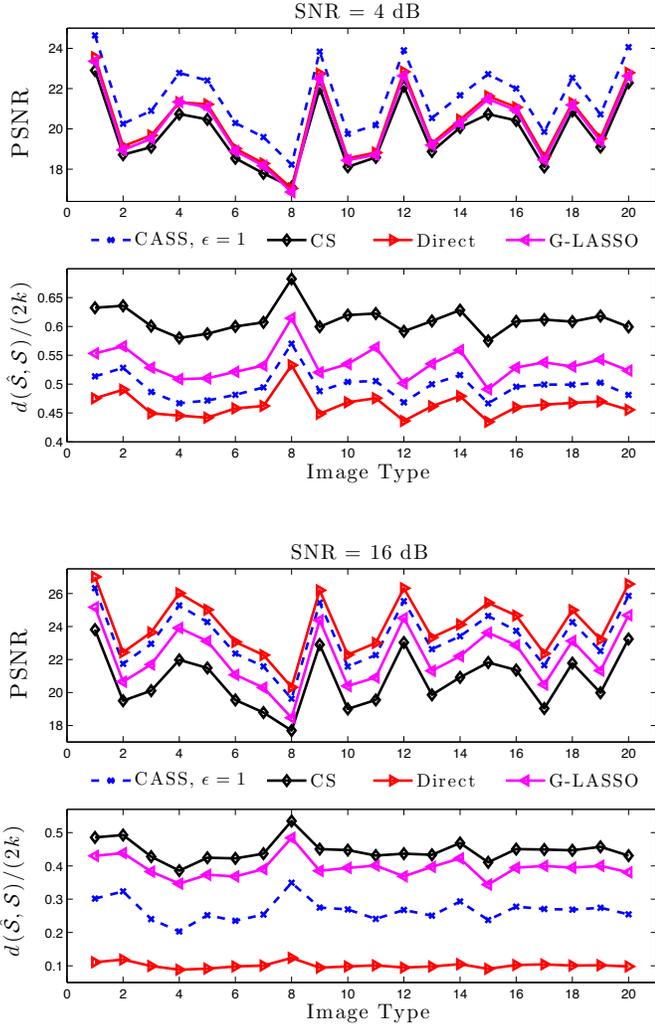


Fig. 6. Performance of CASS, standard compressed sensing with LASSO and Gaussian ensemble, Group LASSO of [28] (G-LASSO) and Gaussian ensemble, which exploits signal structure, and direct sensing on the Microsoft Research Object Class Recognition database [27] for two SNRs. The database consists of 20 image classes, with approximately 30 images per class. The images are approximately sparse in the Haar basis. The total sensing energy was the same for all methods. $n = 128^2$. The results are averaged over each class. $k = 512$, $\epsilon = 1$.

since, from (8), $\gamma \leq 5/2$. Next, imposing the condition of the Theorem,

$$\begin{aligned} x_{\min} &\geq \sqrt{20 \frac{n}{M} \left(\log k + 2 \log \left(\frac{8}{\epsilon} \right) \right)} \\ &= \sqrt{20 \frac{n}{M} \left(\log \left(\frac{8}{\epsilon} \right) + \log \left(\frac{8k}{\epsilon} \right) \right)}, \end{aligned}$$

then

$$\mathbb{P}(\mathcal{E}) \leq \sum_{s=1}^{s_0} \left(\frac{8}{\epsilon} \right)^{-s} \leq \epsilon/4. \quad (11)$$

By combining (9), (10), and (11), since $\mathbb{E}[d(\hat{\mathcal{S}}, \mathcal{S})|\mathcal{E}] \leq 2k$,

$$\begin{aligned} \mathbb{E}[d(\hat{\mathcal{S}}, \mathcal{S})] &= \mathbb{P}(\mathcal{E})\mathbb{E}[d(\hat{\mathcal{S}}, \mathcal{S})|\mathcal{E}] + \mathbb{P}(\mathcal{E}^c)\mathbb{E}[d(\hat{\mathcal{S}}, \mathcal{S})|\mathcal{E}^c] \\ &\leq k\epsilon, \end{aligned}$$

completing the proof.

REFERENCES

- [1] M. L. Malloy and R. D. Nowak, "Near-optimal adaptive compressed sensing," in *Signals, Systems and Computers (ASILOMAR), 2012 Conference Record of the Forty Sixth Asilomar Conference on*. IEEE, 2012, pp. 1124–1130.
- [2] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk, "Single-pixel imaging via compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 83–91, 2008.
- [3] V. Studer, J. Bobin, M. Chahid, H. Mousavi, E. Candes, and M. Dahan, "Compressive fluorescence microscopy for biological and hyperspectral imaging," *Proceedings of the National Academy of Sciences*, vol. 109, no. 26, pp. E1679–E1687, 2012.
- [4] E. Arias-Castro, E. J. Candes, and M. Davenport, "On the Fundamental Limits of Adaptive Sensing," *ArXiv e-prints*, Nov. 2011.
- [5] M. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *Information Theory, IEEE Transactions on*, vol. 55, no. 12, pp. 5728–5741, dec. 2009.
- [6] S. Aeron, V. Saligrama, and M. Zhao, "Information theoretic bounds for compressed sensing," *Information Theory, IEEE Transactions on*, vol. 56, no. 10, pp. 5111–5130, oct. 2010.
- [7] M. J. Wainwright, "Sharp thresholds for high-dimensional and noisy recovery of sparsity," *arXiv preprint math/0605740*, 2006.
- [8] R. M. Castro, "Adaptive Sensing Performance Lower Bounds for Sparse Signal Estimation and Testing," *ArXiv e-prints*, Jun. 2012.
- [9] J. Haupt, R. Baraniuk, R. Castro, and R. Nowak, "Sequentially designed compressed sensing," in *Statistical Signal Processing Workshop (SSP), 2012 IEEE*. IEEE, 2012, pp. 401–404.
- [10] M. L. Malloy and R. D. Nowak, "On the limits of sequential testing in high dimensions," in *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*. IEEE, 2011, pp. 1245–1249.
- [11] M. L. Malloy and R. Nowak, "Sequential testing for sparse recovery," *arXiv preprint arXiv:1212.1801*, 2012.
- [12] D. Wei and A. O. Hero, "Multistage adaptive estimation of sparse signals," in *Statistical Signal Processing Workshop (SSP), 2012 IEEE*. IEEE, 2012, pp. 153–156.
- [13] G. Newstadt, E. Bashan, and A. Hero, "Adaptive search for sparse targets with informative priors," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 3542–3545.
- [14] J. Haupt, R. M. Castro, and R. Nowak, "Distilled sensing: Adaptive sampling for sparse detection and estimation," *Information Theory, IEEE Transactions on*, vol. 57, no. 9, pp. 6222–6235, 2011.
- [15] M. L. Malloy and R. D. Nowak, "Sequential analysis in high-dimensional multiple testing and sparse recovery," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*. IEEE, 2011.
- [16] M. L. Malloy, G. Tang, and R. D. Nowak, "The sample complexity of search over multiple populations," *arXiv preprint arXiv:1209.1380*, 2012.
- [17] M. Iwen, "Group testing strategies for recovery of sparse signals in noise," in *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on*. IEEE, 2009, pp. 1561–1565.
- [18] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *Signal Processing, IEEE Transactions on*, vol. 56, no. 6, pp. 2346–2356, 2008.
- [19] M. Iwen and A. Tewfik, "Adaptive compressed sensing for sparse signals in noise," in *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*. IEEE, 2011.
- [20] M. A. Davenport and E. Arias-Castro, "Compressive binary search," *ArXiv e-prints*, Feb. 2012. [Online]. Available: <http://arxiv.org/abs/1202.0937>
- [21] M. Iwen and A. Tewfik, "Adaptive strategies for target detection and localization in noisy environments," *Signal Processing, IEEE Transactions on*, vol. 60, no. 5, pp. 2344–2353, 2012.

- [22] J. Haupt, R. Baraniuk, R. Castro, and R. Nowak, "Compressive distilled sensing: Sparse recovery using adaptivity in compressive measurements," in *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on*, nov. 2009, pp. 1551–1555.
- [23] E. Arias-Castro, "Detecting a vector based on linear measurements," *Electronic Journal of Statistics*, vol. 6, pp. 547–558, 2012.
- [24] M. L. Malloy and R. D. Nowak, "Near-Optimal Compressive Binary Search," *ArXiv e-prints*, Mar. 2012. [Online]. Available: <http://arxiv.org/abs/1203.1804>
- [25] S. J. Wright, R. D. Nowak, and M. A. Figueiredo, "Sparse reconstruction by separable approximation," *Signal Processing, IEEE Transactions on*, vol. 57, no. 7, pp. 2479–2493, 2009.
- [26] A. Frank and A. Asuncion, "UCI machine learning repository," 2010. [Online]. Available: <http://archive.ics.uci.edu/ml>
- [27] "Microsoft Research Cambridge Object Recognition Image Database, <http://research.microsoft.com/en-us/projects/objectclassrecognition/>."
- [28] N. S. Rao, R. D. Nowak, S. J. Wright, and N. G. Kingsbury, "Convex approaches to model wavelet sparsity patterns," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE, 2011, pp. 1917–1920.
- [29] L. Jacob, G. Obozinski, and J.-P. Vert, "Group lasso with overlap and graph lasso," in *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 2009, pp. 433–440.
- [30] N. S. Rao and R. D. Nowak, "Correlated Gaussian designs for compressive imaging," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*. IEEE, 2012, pp. 921–924.