Adaptive Compressed Sensing for Sparse Signals in Noise

M. A. Iwen*, A. H. Tewfik+, *Fellow, IEEE*

***Abstract*—This paper studies the problem of recovering a signal with a sparse representation in a given orthonormal basis using as few noisy observations as possible. Herein, observations are subject to the type of background clutter noise encountered in radar applications. Given this model, this paper proves for the first time that highly sparse signals contaminated with Gaussian background noise can be recovered by adaptive methods using fewer noisy linear measurements than required by any possible recovery method based on non-adaptive Gaussian measurement ensembles.**

I. Introduction

This paper considers adaptive acquisition strategies for estimating a signal, f, which admits a sparse representation in terms of a linear combination of k unknown elements from a set of N orthonormal functions. Adaptive Bayesian techniques for estimating the support of a sparse signal were proposed by Ji et al. and Castro et al. (e.g., see [3] and [4], respectively). These Bayesian methods have been demonstrated to work well empirically, often requiring fewer noisy measurements to recover sparse signals than non-adaptive competitors in practice. Similarly, (compressive) distilled sensing techniques [5], [6], [7] demonstrate that adaptive methods can improve error bounds for sparse recovery problems over nonadaptive techniques in other measurement noise models related to the one considered herein. Finally, very recently Indyk et al. demonstrated that adaptive compressed sensing methods can outperform nonadaptive methods in the standard compressed sensing context [8]. In this paper we develop additional theory supporting the further use and consideration of such adaptive methods by proving that a simple adaptive measurement procedure can reliably recover sparse signals using fewer linear measurements than any possible approach utilizing non-adaptive Gaussian measurement matrices when the measurements are contaminated with background noise.

A. The Noise Model

Let $\Phi = \{\phi_i \mid 1 \leq i \leq N\}$ be a set of real valued orthonormal functions on $[0,1]$ which span a given function space of interest. A generic observable signal (i.e., function on $[0,1]$) can have a component outside of Φ. However, sparse approximation techniques generally only consider the signal's projection, f, onto Φ and, furthermore, assume that f has a sparse representation in Φ. Suppose that

$$f = \sum_{i=1}^{N} f_i \cdot \phi_i.$$

Given a sparsity assumption for f it makes sense to define the support of f in Φ to be the positions where its coefficients, f_i, are nonzero (or otherwise larger in magnitude than an application dependent threshold). The support of f is thus

$$\text{supp}(f) = \left\{ \; j \; \middle| \; |f_j| \; > \; 0 \; \right\} \; \subseteq \; [1, N].$$

Note that in order to recover f one must identify $\text{supp}(f)$. Thus, the primarily focus of this paper – signal recovery – is integrally linked to support identification.

A solution to the sparse approximation problem necessitates the design of a set of test, or measurement, functions $\mathcal{M}_j : [0,1] \longrightarrow \mathbb{R}$, $1 \leq j \leq m$. Each test function, $\mathcal{M}_j$, is a specified linear combination of basis elements from Φ. In this paper each of these test functions, $\mathcal{M}_j$, will generate noisy observations of the form

$$y_j = \langle \mathcal{M}_j, \; f + \mathcal{P}_j \rangle = \langle \mathcal{M}_j, \; f \rangle + \langle \mathcal{M}_j, \; \mathcal{P}_j \rangle, \tag{1}$$

where $\langle f, \; g \rangle$ denotes the inner product between $f, g : [0,1] \rightarrow \mathbb{R}$. In the above equation $\mathcal{P} = \{\mathcal{P}_j \mid 1 \leq j \leq m\}$ is a sequence of identically distributed measurement noise processes. Each $\mathcal{P}_j$ is assumed to be independent of all the other $\mathcal{P}_{j'}$ processes whenever $j \neq j'$. In effect, every measurement of f is contaminated with background noise, or clutter noise, generated by a rapidly varying random background signal.

In practice, measurement noise will consist of two components. The first component will be due to environmental or physical noise outside of the processing and acquisition system (e.g., clutter in radar). This component will depend on the measurement function. For example, using a wider beam to cover a wider area in radar will increase the observed clutter noise. The second noise component is due to thermal noise in the acquisition and processing circuitry. This noise component does not depend on the measurement function, and its effect can be reduced by using more sophisticated electronics (e.g., by cooling a detector).

In many previous studies (e.g., see [9], [10], [11] and references therein) the measurements utilized for signal (support) recovery were of the form

$$y_j = \langle \mathcal{M}_j, \; f \rangle + w_j, \tag{2}$$

where $w_j \sim \mathcal{N}(0,1)$ is independent Gaussian noise for each j, f is considered as a sparse vector in $\mathbb{R}^N$ (i.e., the problem is discrete), and $M_j \sim \mathcal{N}(0, I_{N \times N})$ is a random vector independently drawn from the zero-mean isotropic Gaussian distribution for each j. The measurements provided by Equation 2 account for situations where the second (thermal) component of measurement noise discussed in the preceding paragraph dominates the first (clutter) noise component. In contrast, the noise model considered herein (see Equation 1) focuses on situations where the first noise component dominates the second noise component.

This paper expands on results first reported in [1]. The vast majority of this paper was written in early 2010 and is available as an expanded preprint [2].

*Math Department, Duke University, markiwen@math.duke.edu

+Dept. of Electrical and Computer Engineering, University of Texas at Austin, tewfik@austin.utexas.edu

The measurement model in the Equation 1 will be referred to as *non-adaptive* if the generation of the j^{th} measurement, y_j, is independent of all previous noisy observations, y_n, $1 \leq n \leq j-1$. In effect, a set of measurements $\mathcal{M} = \{\mathcal{M}_1, \ \ldots, \ \mathcal{M}_m\}$ is non-adaptive if it can be wholly instantiated before any measurements are actually taken. If, on the other hand, any single measurement may depend on the results of previous measurements, $\mathcal{M}$ will be called *adaptive*. We will always suppose that $f = \sum_{i=1}^{N} f_i \cdot \phi_i$ is k-sparse with respect to Φ. The value $C_{\min} = \min\left\{|\langle f, \ \phi_i\rangle| \ \middle| \ i \in \text{supp}(f)\right\}$ will always be the magnitude of the smallest of the k non-zero coefficients of f.

B. Results

Much of the previous work on solving sparse support identification problems has concentrated on methods utilizing non-adaptive randomly generated Gaussian measurements contaminated with zero mean Gaussian noise. The non-adaptive Gaussian measurement ensembles, $M_j \sim \mathcal{N}(0, I_{N\times N})$ for $1 \leq j \leq m$, are particularly relevant to study given their near-optimal properties with respect to non-adaptive compressive sensing measurement design (e.g., see [12], [13], [14], [15]). Here, we will momentarily focus on the following result concerning support recovery using noisy non-adaptive Gaussian measurements contaminated with Gaussian noise. The objective is to construct a lower bound on the number of measurements, m, required by any sparse recovery algorithm in order to correctly recover the support of f using the general measurement model considered herein (see Equation 1).

Theorem 1. *Suppose that $\mathcal{G} = \left\{\mathcal{G}_j \ \middle| \ 1 \leq j \leq m\right\}$ is an ensemble of m non-adaptive random standard Gaussian noise processes independently drawn for each j. Create test functions by setting*

$$\mathcal{M} = \left\{\mathcal{M}_j = \sum_{i=1}^{N} \langle \mathcal{G}_j, \phi_i\rangle \cdot \phi_i \ \middle| \ 1 \leq j \leq m\right\}.$$

Furthermore, let $\mathcal{P}_j$ for $1 \leq j \leq m$ be m independent Gaussian measurement noise processes with mean 0 so that the accumulated noise for each Equation 1 measurement, conditioned on $\mathcal{M}_j$, is $\langle \mathcal{M}_j, \mathcal{P}_j\rangle \sim \mathcal{N}(0, \|\mathcal{M}_j\|_2^2 \cdot \sigma^2/N)$. Then, there exists a constant $c \in \mathbb{R}^+$ such that any algorithm using

$$m \ < \ c \cdot \frac{\sigma^2}{C_{\min}^2} \cdot \log(N/k)$$

non-adaptive Gaussian measurements as input will asymptotically fail to reliably recover $\text{supp}(f)$ *and, therefore, f itself. That is, for N sufficiently large any algorithm will fail to recover* $\text{supp}(f)$ *with probability bounded above 0.*

Proof: See Section II below.[1] □

In effect, Theorem 1 provides a non-adaptive Gaussian measurement bound below which any recovery method must fail to be asymptotically reliable for the support identification of some sparse input vectors. In this paper ideas from group testing [17] are utilized in combination with statistical binary detection and estimation techniques [18] to produce the following theorem.

Theorem 2. *Let $\mathcal{P}_j$ for $1 \leq j \leq m$ be m independent Gaussian measurement noise processes with mean 0 so that the accumulated noise for each Equation 1 measurement, conditioned on the adaptive test function $\mathcal{M}_j$, is $\langle \mathcal{M}_j, \mathcal{P}_j\rangle \sim \mathcal{N}(0, \|\mathcal{M}_j\|_2^2 \cdot \sigma^2/N)$. Furthermore, suppose that $\sigma^2/C_{\min}^2$ is $\Omega\left(k \cdot \ln^3 N\right)$.[2] Then, there exists a constant $c \in \mathbb{R}^+$ such that whenever the number of allowed measurements, m, exceeds*

$$c \cdot \frac{\sigma^2}{C_{\min}^2} \cdot \ln^2 (k \cdot \ln N)$$

Algorithm 1 will approximate f precisely enough to reliably recover $\text{supp}(f)$ *with probability $\to 1$ as $N \to \infty$.*

Proof: Algorithm 1 is outlined in Section III. See [2] for the proof.

In order to compare Theorems 1 and 2, consider the following example. Suppose that k is $\ln^{O(1)} N$ and $\sigma^2/C_{\min}^2$ is $\Omega\left(k \cdot \ln^3 N\right)$. In this regime one can see that any asymptotically reliable non-adaptive Gaussian measurement scheme will require the use of

$$\Omega\left(\frac{\sigma^2}{C_{\min}^2} \cdot \ln N\right)$$

measurements. On the other hand, the adaptive methods developed below are asymptotically reliable using

$$O\left(\frac{\sigma^2}{C_{\min}^2} \cdot \ln^2 (\ln N)\right)$$

measurements. Hence, if f is sufficiently sparse and its measurements sufficiently noisy, the adaptive methods presented below will asymptotically outperform any sparse support recovery method utilizing non-adaptive Gaussian measurement ensembles. That is, adaptive methods can outperform nonadaptive methods for target location in noisy environments.

Intuitively, it should not be surprising that methods utilizing non-adaptive measurements are less effective under the observational model in Equation 1 than methods based on adaptive measurements. Every non-adaptive measurement must necessarily allocate significant amounts of sensing energy to a large fraction of the basis elements in Φ (i.e., to a large fraction of the entire search area). This essentially guarantees that every non-adaptive observation will be contaminated with a large fraction of the additive observational noise from the entire search area. Adaptive measurements, on the other hand, can eventually avoid observational noise from large portions of the search area by ignoring regions where signal components are unlikely to be present. The end result is that any method utilizing non-adaptive observations must ultimately deal with higher collective noise levels from their measurement ensembles than methods which adaptively focus their measurements toward regions likely to contain signal components.

[1] Since the initial appearance of Theorem 1 in [1], [2] a similar result by Aeron et al. has appeared independently in [16].

[2] Let $f, g : \mathbb{R}^+ \to \mathbb{R}^+$. Then, f is $\Omega(g)$ if and only if g is $O(f)$.

II. Proof of Theorem 1

It suffices to consider the discrete case were a k-sparse $f \in \mathbb{R}^N$ is known a priori to have $f_j = C_{\min}$ for all $j \in \text{supp}(f)$. We sketch the proof here (see appendix A of [2] for a detailed proof). The non-adaptive measurements are given by an $m \times N$ random matrix, $\mathcal{M}$, with each row, $\mathcal{M}_j$, independently drawn from the zero-mean isotropic Gaussian distribution $\mathcal{N}(0, I_{N\times N})$. We define the background noise, $\mathcal{P}$, to be an $m \times N$ random real-valued noise matrix consisting of $m \cdot N$ independently and identically distributed (i.i.d.) normal random variables. Finally, we assume that we have a single detector which, at time $t_j \in \mathbb{R}^+$, returns a noisy linear measurement (i.e., a discrete dot product along the lines of Equation 1) of the form

$$\langle \mathcal{M}_j, f + \mathcal{P}_j \rangle = \langle \mathcal{M}_j, f \rangle + \langle \mathcal{M}_j, \mathcal{P}_j \rangle. \tag{3}$$

Let $\vec{1}_k$ be the k length vector of ones and $\mathcal{I}_{N\times N}$ be the $N \times N$ identity matrix. We then define $\vec{v}$ to be $C_{\min} \cdot \vec{1}_k$. Next, for each $U \subset [1, N] \cap \mathbb{N}$ with $|U| = k$, we will define $\mathcal{M}_U$ to be the $m \times k$ matrix formed by selecting the columns of $\mathcal{M}$ indexed by U. Finally, we define the random vectors $\vec{p}, \vec{w} \in \mathbb{R}^m$ to have $p_j = \langle \mathcal{M}_j, \mathcal{P}_j \rangle \sim \mathcal{N}(0, \|\mathcal{M}_j\|_2^2 \cdot \sigma^2/N)$, conditioned on $\mathcal{M}$, and $w_j \sim \mathcal{N}(0, \sigma^2)$, respectively, for all $j \in [1, m] \cap \mathbb{N}$.

Order the k-element subsets of $[1, N] \cap \mathbb{N}$ lexicographically and then index them from 1 to $\tilde{N} = \binom{N}{k}$. For any $i \in [1, \tilde{N}] \cap \mathbb{N}$ we will let $U[i]$ denote the i^{th} subset in this ordering. Next, let $\mathcal{D}$ be an $m \times m$ diagonal matrix with $\mathcal{D}_{j,j} = \|\mathcal{M}_j\|_2^2 \cdot \sigma^2/N$ for each $j \in [1, m] \cap \mathbb{N}$. From above we know that $\mathcal{D} \to \sigma^2 \cdot \mathcal{I}_{m\times m}$ as $N \to \infty$ almost surely. It is not difficult to see that non-adaptive Gaussian measurements of f will produce a random vector of the form $\mathbb{P}_i = \mathcal{M}_{U[i]}\vec{v} + \vec{p} \sim \mathcal{N}(\mathcal{M}_{U[i]}\vec{v}, \mathcal{D})$ for some $i \in [1, \tilde{N}] \cap \mathbb{N}$. The Kullback-Leibler divergence between two such potential non-adaptive measurement distributions is

$$D(\mathbb{P}_i \| \mathbb{P}_{i'}) = \frac{1}{2}\left((\mathcal{M}_{U[i]}\vec{v} - \mathcal{M}_{U[i']}\vec{v})^{\mathrm{T}} \mathcal{D}^{-1} (\mathcal{M}_{U[i]}\vec{v} - \mathcal{M}_{U[i']}\vec{v}) \right).$$

Furthermore, this divergence is a function of the random non-adaptive measurement matrix $\mathcal{M}$. Hence, we have that

$$D(\mathbb{P}_i \| \mathbb{P}_{i'}) = \left(\frac{C_{\min}^2}{\sigma^2} \right) \cdot \left(k - \left| U[i] \cap U[i'] \right| \right) \cdot \sum_{j=1}^{m} \frac{N}{Y_j} \cdot Z_j^2$$

where $Y_j = \|\mathcal{M}_j\|_2^2 \sim \chi_N^2$ and $Z_j \sim \mathcal{N}(0, 1)$ are dependent for each $j \in [1, m] \cap \mathbb{N}$.

More carefully considering the dependence of $Y_j \sim \chi_N^2$ and $Z_j \sim \mathcal{N}(0, 1)$ for each $j \in [1, m] \cap \mathbb{N}$ we can see that it is entirely due to the at most $2k$ standard normal variables making up the entries of $\mathcal{M}_j$ indexed by $U[i] \cup U[i']$. Furthermore, the net contribution of these at most $2k$ variables to Y_j will always be nonnegative. Therefore we will have

$$\mathbb{E}\left[D(\mathbb{P}_i \| \mathbb{P}_{i'}) \right] \leq \left(\frac{C_{\min}^2}{\sigma^2} \right) \left(k - \left| U[i] \cap U[i'] \right| \right) \cdot \frac{m}{1 - \frac{2k-2}{N}}. \tag{4}$$

The remainder of the proof depends on employing the following weakened form of Fano's inequality (see Lemma 2 in [9]). That is, the average probability of error, p_{error}, in performing a hypothesis test over a family of distributions $\{\mathbb{P}_1, \ldots, \mathbb{P}_{\tilde{N}}\}$ is bounded by

$$p_{\text{error}} \geq 1 - \frac{\frac{1}{\tilde{N}^2} \cdot \sum_{i,i'=1}^{\tilde{N}} D(\mathbb{P}_i \| \mathbb{P}_{i'}) + \log 2}{\log\left(\tilde{N} - 1\right)}.$$

Considering the expected average probability of success as a function of the random non-adaptive measurement matrix we can see that

$$\mathbb{E}\left[1 - p_{\text{error}}\right] \leq \frac{k}{\log\left(\tilde{N} - 1\right)} \left(\frac{C_{\min}^2}{\sigma^2} \right) \frac{m}{1 - \frac{2k-2}{N}} + \frac{\log 2}{\log\left(\tilde{N} - 1\right)}$$

by Equation 4. Applying Markov's Inequality we have

$$\mathbb{P}\left[1 - p_{\text{error}} \geq \frac{1}{2} \right] \leq \frac{2k}{\log\left(\tilde{N} - 1\right)} \left(\frac{C_{\min}^2}{\sigma^2} \right) \frac{m}{1 - \frac{2k-2}{N}} + \frac{2 \cdot \log 2}{\log\left(\tilde{N} - 1\right)}.$$

If the right hand side of the inequality above is less than one then the probability of choosing a Gaussian measurement matrix capable of "almost always" decoding the correct support of most sparse vectors, f, will also be less than one.

Finishing, we can see that $\mathbb{P}\left[1 - p_{\text{error}} \geq \frac{1}{2}\right] < \frac{1}{2}$ whenever

$$m < \left(\frac{1 - \frac{2k-2}{N}}{8} \right) \cdot \frac{\sigma^2}{C_{\min}^2} \cdot \frac{\log\left(\tilde{N} - 1\right)}{k}$$

with $N \geq 2k \geq 32$. Theorem 1 follows.

Note that methods from [9] can be used even more directly to prove that non-adaptive Bernoulli measurement matrices, $\mathcal{M} \in \{-1, 1\}^{m\times N}$, can also only accommodate reliable sparse recovery in the presence of Gaussian background noise if m is $\Omega\left(\frac{\sigma^2}{C_{\min}^2} \cdot \log(N/k) \right)$. Similarly, we expect that more complicated modifications of this argument can also be used to prove that this scaling for m is also required for other random non-adaptive measurement ensembles utilized for sparse recovery problems (e.g., see [19]).

III. Adaptive Approach

It suffices to consider recovering functions of the form

$$f(x) = \sum_{j=1}^{k} C_j \cdot \delta(x - x_j) \tag{5}$$

where $\delta(x)$ is a Dirac delta, each $C_j \in \mathbb{R}$, and $x_j \in [0, 1]$, for $j \in \mathbb{Z} \cap [1, k]$. This is a simplified model for the problem of recovering an unknown number of ideal point targets with reflectivity C_j located at positions x_j. The model assumes prior knowledge of the range of the arbitrary positions x_j, which we normalize to the interval $[0, 1]$. The model also captures radar imaging of targets that consist of a collection of point reflectors. Selecting the measurement functions, $\mathcal{M}_j$, corresponds to selecting a radar beamform and illumination pattern.

Let $\mathcal{I}$ be a subset of $[0, 1]$ and define the indicator function for $\mathcal{I}$,

$$\mathbb{I}_{\mathcal{I}} : [0, 1] \mapsto \{0, 1\},$$

to be

$$\mathbb{I}_{\mathcal{I}}(x) = \begin{cases} 1 & \text{if } x \in \mathcal{I} \\ 0 & \text{otherwise} \end{cases}.$$

Without loss of generality, we will assume that we can measure any subset $\mathcal{I} \subseteq [0,1]$ at time $t \in \mathbb{R}^+$ via

$$\langle \mathbb{I}_{\mathcal{I}}, f \rangle = \int \mathbb{I}_{\mathcal{I}} \cdot f \; dx + \int \mathbb{I}_{\mathcal{I}} \; d\mathcal{P}_t, \tag{6}$$

where $\mathcal{P}_t(x)$ represents stochastic measurement noise (i.e., a diffusion process). We assume for each time $t \in \mathbb{R}^+$ that

$$\int \mathbb{I}_{\mathcal{I}} \; d\mathcal{P}_t \quad \text{and} \quad \int \mathbb{I}_{\mathcal{J}} \; d\mathcal{P}_t$$

are independent and identically distributed (i.i.d.) whenever $\mathcal{I} \cap \mathcal{J} = \emptyset$ and

$$\int \mathbb{I}_{\mathcal{I}} \; dx = \int \mathbb{I}_{\mathcal{J}} \; dx. \tag{7}$$

Similarly, we assume for every two times $t_1 \neq t_2$ that

$$\int \mathbb{I}_{\mathcal{I}} \; d\mathcal{P}_{t_1} \quad \text{and} \quad \int \mathbb{I}_{\mathcal{J}} \; d\mathcal{P}_{t_2}$$

are i.i.d. as long as $\mathcal{I}, \mathcal{J} \subseteq [0,1]$ satisfy Equation 7. We define $\sigma^2_{\mathcal{I}}$ to be the variance of $\langle \mathbb{I}_{\mathcal{I}}, f \rangle$ for a given $\mathcal{I} \subseteq [0,1]$. Given the assumptions above, we can see that $\sigma^2_{\mathcal{I}} = \sigma^2_{\mathcal{J}}$ whenever $\mathcal{I}, \mathcal{J} \subseteq [0,1]$ satisfy Equation 7. Finally, we will denote the variance of the noise over the unit interval by $\sigma^2 = \sigma^2_{[0,1]}$.

A. Single Spike Targeting

In this section we will assume that our function f consists of a single Dirac delta (i.e., $k = 1$ in Equation 5). We will employ a simple adaptive binary search procedure to locate the support of f. Suppose for the time being that the single Dirac delta has a positive magnitude, $C_1 > 0$. If so, we can look for the support of f (i.e., x_1) in $[0,1]$ using a binary search strategy. As long as the additive measurement noise is independent and identically distributed (i.i.d.) on both subintervals under consideration (e.g.,$[0, \frac{1}{2})$ and $[\frac{1}{2}, 1]$ to begin with), the subinterval which contains x_1 will produce measurements with larger means than the subinterval not containing x_1. Thus, our measurements of the subinterval containing x_1 will tend to be larger more often. Using this observation to our advantage, we can correctly choose the subinterval containing x_1 with high probability by choosing the subinterval that returns the largest measurements most often. Repeated application of this decision principle yields a binary search.

If C_1 is negative the binary search is analogous. We simply repeatedly choose the subinterval which returns smaller measurements more often. Finally, we deal with the fact that we don't have apriori knowledge of the sign of C_1 by performing two binary searches in parallel. One search assumes that C_1 is positive, while the other assumes it is negative. One of the two searches must succeed with high probability since C_1 is nonzero (i.e., either positive or negative). If C_1 is positive, our search assuming positivity will locate the spike with high probability. If C_1 is negative, our search assuming negativity will locate the spike with high probability. The problem is thus reduced to deciding which search result (i.e., the interval resulting from the search assuming C_1 is positive versus negative) is correct. We denote the interval resulting from the binary search assuming C_1 is positive by $\mathcal{I}^+$. Similarly, we let $\mathcal{I}^-$ denote the interval resulting from the binary search that assumes C_1 is negative. We are guaranteed to have $\mathcal{I}^+ \cap \mathcal{I}^- = \emptyset$. To finish we must decide whether $x_1 \in \mathcal{I}^+$ or $x_1 \in \mathcal{I}^-$.

To help make our final decision we arbitrarily chose an interval whose noise characteristics will be, by assumption, distributed identically to the additive noise in both the resulting positive/negative binary search intervals (i.e., $\mathcal{I}^+$ and $\mathcal{I}^-$). The resulting positive/negative interval containing x_1 should yield measurements with a mean that is different from the arbitrary interval measurements' mean. Hence, we estimate the measurement means of both intervals resulting from our binary searches, and then compare them to the mean of our arbitrary interval measurements. Whichever binary search result differs most from our arbitrary interval in terms of measurement mean will be the correct search result with high probability. See [2] for details.

To conclude, we note that by employing a binary search for single spike recovery we are essentially transforming the spike localization problem into $O(\log_2 N)$ binary detection problems. Without loss of generality, at each stage of our binary search we must decide whether measurements of the left subinterval currently under consideration were generated by (*i*) a spike in noise, or (*ii*) noise alone. The answer to this question entirely determines whether the left or right subinterval becomes the new interval of interest in the next stage of our binary search. When viewed from this perspective the single spike recovery problem becomes equivalent to a series of statistical detection/estimation problems (see [18]). We simply localize the spike by repeatedly detecting its presence in each right/left subinterval. Hence, there are as many strategies for recovering a single spike as there are strategies for detecting the presence of a signal in noise. Other possible approaches include the use of optimal sequential detection methods (e.g., [20]) at each stage of our binary search. These methods could be used to collect measurements dynamically until a decision regarding the presence/absence of a spike can be made with error probability below a user specified tolerance.

B. Recovering Multiple Spikes One at a Time

Given that any two distinct spike locations, x_{j_1} and x_{j_2}, are assumed to have $|x_{j_1} - x_{j_2}| > \frac{1}{N}$, we may represent $[0,1]$ by its N subintervals,

$$s_0 = \left[0, \frac{1}{N}\right), \; s_1 = \left[\frac{1}{N}, \frac{2}{N}\right), \; \ldots, \; s_{N-1} = \left[1 - \frac{1}{N}, 1\right], \tag{8}$$

only k of which contain spikes (i.e., we may consider $[0,1]$ to be a k-sparse array of length N). Keeping this in mind we will demonstrate how to create q disjoint unions of these s_j-subsets, each of length $O\left(\frac{1}{q}\right)$, which will isolate each spike from all the other $(k-1)$ spikes with fixed probability. We can then use several of these disjoint unions to separate each of our spikes from all the others with arbitrarily high probability. We begin by describing our disjoint unions of s_j-subsets.

Let q be one of the first $2k\lfloor \log_k N \rfloor$ prime numbers larger than k. For each $h \in [0, q) \cap \mathbb{Z}$ form the set

$$\mathcal{I}_{q,h} = \bigcup_{j \equiv h \bmod q} s_j \tag{9}$$

Algorithm 1 No More Than k Deltas

Input: Maximum number of spikes k, Position tolerance N, smallest spike magnitude of interest $C_{\min}$, success probability p
Output: Estimates of magnitudes $> \frac{1}{2}C_{\min}$, $\{C_1, \ldots, C_k\}$, and their positions, $\{x_1, \ldots, x_k\}$
Find all spikes at least once...
$SPIKES \longleftarrow \emptyset$
for $j = 1,\ j < P = O\left(\log\left(\frac{k}{1-p}\right)\right),\ j{+}{+}$ **do**
  $q \longleftarrow$ Randomly select one of $2k\lfloor \log_k N \rfloor$ primes $> k$
  Form $\mathcal{I}_q$ (see Equation 10)
  for each $\mathcal{I}_{q,h} \in \mathcal{I}_q$ **do**
    $(\tilde{x}, \tilde{C}) \longleftarrow$ Section III-A methods on $\mathcal{I}_{q,h}$
    if $|\tilde{C}| > \frac{1}{2}C_{\min}$ **then**
      $SPIKES \longleftarrow SPIKES \cup \{(\tilde{x}, \tilde{C})\}$
    end if
  end for
end for
Remove excess spike approximations...
$\{(\tilde{x}_0, \tilde{C}_0), (\tilde{x}_1, \tilde{C}_1) \ldots\} \longleftarrow$ Sort $SPIKES$ by $\tilde{x}$'s
for $n = 0,\ n < |SPIKES|,\ n{+}{+}$ **do**
  while $|x_n - x_{n+1}| \leq \frac{N}{2}$ **do**
    $SPIKES \leftarrow SPIKES - \{(\tilde{x}_{n+1}, \tilde{C}_{n+1})\}$
  end while
end for
Return $SPIKES$

and then set

$$\mathcal{I}_q = \left\{\mathcal{I}_{q,0},\ \mathcal{I}_{q,1},\ \ldots,\ \mathcal{I}_{q,q-1}\right\}. \tag{10}$$

$\mathcal{I}_q$ is our set of unions of disjoint s_j-subsets.

To recover all spikes, we simply apply the methods of Section III-A to each subset in $\mathcal{I}_q$ separately, for several randomly selected values of q. We then collect all the output spikes (allowing only one from each s_j interval) which have estimated magnitudes that are larger than half the smallest spike magnitude we care to detect. By not reporting spikes with smaller estimated magnitudes we exclude the recovery of 'fake' or 'insignificant' spikes. If we have prior knowledge of the smallest spike magnitude, $C_{\min}$, in f (see Equation 5) we can guarantee f's approximate recovery with high probability. If we have no prior knowledge of the smallest spike magnitude, then all at most k spikes with magnitude larger than any given $C_{\min}$ value will be returned. Thus, in general, we can guarantee the recovery of all sufficiently large (i.e., at least $C_{\min}$ in magnitude) spikes in f with high probability. See Algorithm 1 for multiple spike recovery pseudocode.

IV. Conclusion

The adaptive algorithm described in this paper is only one of many potential recovery methods that can be created by combining combinatorial group testing constructions [17] with signal estimation and detection methods [18]. More generally, any good binary group testing matrix which tends to isolate the members of any small number of signal components can be used to segment a search space into smaller regions likely containing only one signal component, or target. Signal detection and estimation methods can then be used to search each smaller region for a single isolated signal component.

Note that the recovery method considered herein is very easy to parallelize since each disjoint region of the search space dictated by the group testing construction in Section III-B can be searched independently. This follows from the fact that the group testing methods we used to segment the search space are themselves non-adaptive, despite the fact that each smaller resulting region is itself searched adaptively. Although this non-adaptive partitioning of the search space promotes parallelism, it may ultimately hurt performance. The number of required measurements can probably be reduced by partitioning the search space into smaller regions adaptively.

References

[1] M. A. Iwen, "Group testing strategies for recovery of sparse signals in noise," *Proc. 43rd Asilomar Conf. on Signals, Systems, and Computers*, 2009.

[2] M. Iwen and A. H. Tewfik, "Adaptive Group Testing Strategies for Target Detection and Localization in Noisy Environments," *IMA Preprint Series*, vol. 2311, June 2010.

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

[4] R. Castro, J. Haupt, R. Nowak, and G. Raz, "Finding needles in noisy haystacks," *ICASSP*, 2008.

[5] J. D. Haupt, R. G. Baraniuk, R. M. Castro, and R. D. Nowak, "Compressive distilled sensing: Sparse recovery using adaptivity in compressive measurements," *Proc. 43rd Asilomar Conf. on Signals, Systems, and Computers*, 2009.

[6] J. Haupt, R. Castro, and R. Nowak, "Improved bounds for sparse recovery from adaptive measurements," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*. IEEE, 2010, pp. 1563–1567.

[7] J. Haupt, R. Baraniuk, R. Castro, and R. Nowak, "Compressive distilled sensing: Sparse recovery using adaptivity in compressive measurements," in *Proc. 43rd Asilomar Conf. on Signals, Systems, and Computers*. IEEE, 2009.

[8] P. Indyk, E. Price, and D. Woodruff, "On the power of adaptivity in sparse recovery," *FOCS*, 2011.

[9] M. Wainwright, "Information-theoretic bounds on sparsity recovery in the high-dimensional and noisy setting," *International Symposium on Information Theory*, June 2007.

[10] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions on sparsity pattern recovery," *Preprint*, April 2008.

[11] G. Reeves and M. Gastpar, "Sampling bounds for sparse support recovery in the presence of noise," *Proc. IEEE Int. Symp. of Information Theory*, 2008.

[12] D. Donoho, "Compressed Sensing," *IEEE Trans. on Information Theory*, vol. 52, pp. 1289–1306, 2006.

[13] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, pp. 489–509, 2006.

[14] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inform. Theory*, vol. 52, no. 1, Jan 2006.

[15] M. Rudelson and R. Vershynin, "Sparse reconstruction by convex relaxation: Fourier and gaussian measurements," in *40th Annual Conference on Information Sciences and Systems (CISS)*, 2006.

[16] 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, 2010.

[17] D. Z. Du and F. K. Hwang, *Combinatorial Group Testing and Its Applications*. World Scientific, 1993.

[18] D. Middleton, *An Introduction to Statistical Communication Theory*. New York, NY, USA: McGraw-Hill Book Company, Inc., 1960.

[19] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, 2008.

[20] A. G. Tartakovsky, X. R. Li, and G. Yaralov, "Sequential detection of targets in multichannel systems," *IEEE Transactions on Information Theory*, vol. 49, no. 2, Feb. 2003.