Group Testing Strategies for Recovery of Sparse Signals in Noise

Mark A. Iwen
Institute for Mathematics and its Applications
University of Minnesota, Twin Cities
Email: iwen@ima.umn.edu

***Abstract*—We consider the recovery of sparse signals, $f \in \mathbb{R}^N$, containing at most $k \ll N$ nonzero entries using linear measurements contaminated with i.i.d. Gaussian background noise. Given this measurement model, we present and analyze a computationally efficient group testing strategy for recovering the support of f and approximating its nonzero entries. In particular, we demonstrate that group testing measurement matrix constructions may be combined with statistical binary detection and estimation methods to produce efficient adaptive sequential algorithms for sparse signal support recovery. Furthermore, when f exhibits sufficient sparsity, we show that these adaptive group testing methods allow the recovery of sparse signals using fewer noisy linear measurements than possible with non-adaptive methods based on Gaussian measurement ensembles. As a result we improve on previous sufficient conditions for sparsity pattern recovery in the noisy sublinear-sparsity regime.**

I. Introduction

Consider a vector $f \in \mathbb{R}^N$ containing at most $k \ll N$ nonzero entries. We want to recover f using (potentially adaptive) noisy linear measurements. Define the support of f to be the positions of f's nonzero entries,

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

Note that in order to recover f we must identify $\mathrm{supp}(f)$. Furthermore, once we have identified the support of f it is straightforward to approximate its nonzero entries. Thus, we focus primarily on support recovery. Related application areas include compressive sensing, sparse image recovery, multi-target localization, and signal denoising.

Let $\mathcal{M}$ be a $m \times N$ real-valued measurement matrix, and denote the j^{th}-row of $\mathcal{M}$ by $\mathcal{M}_j$. Next, define $\mathcal{G}$ to be an $m \times N$ random real-valued Gaussian noise matrix consisting of $m \cdot N$ independently and identically distributed (i.i.d.) mean zero Gaussian random variables. Finally, for simplicity, assume we have one detector which, at time $t_j \in \mathbb{R}^+$, returns a noisy linear measurement (i.e., dot product) of the form

$$\langle \mathcal{M}_j, f + \mathcal{G}_j \rangle = \langle \mathcal{M}_j, f \rangle + \langle \mathcal{M}_j, \mathcal{G}_j \rangle. \tag{1}$$

Examples of such single detector systems include the famous single pixel camera [5]. Given such a system we want to know how few measurements we can take and still recover the support of f. Equivalently, we want to know how small we can make m and still be able to entirely recover $\mathrm{supp}(f)$ using $\mathcal{M}$.

We will consider a measurement matrix $\mathcal{M}$ in the Equation 1 model above to be *non-adaptive* if the generation of each row, $\mathcal{M}_j$, is independent of the results of $\langle \mathcal{M}_{j-1}, f + \mathcal{G}_{j-1} \rangle, \ldots, \langle \mathcal{M}_1, f + \mathcal{G}_1 \rangle$. In effect, $\mathcal{M}$ is non-adaptive if it can be wholly instantiated before any measurements are actually taken. If, on the other hand, any single measurement row may depend on the results of previous measurements we will refer to $\mathcal{M}$ as *adaptive*. In this paper we present adaptive group testing methods for the sparse support recovery problem capable of outperforming a wide class of non-adaptive methods in the noisy sublinear-sparsity regime. The work herein is a continuation of previous work which considered group testing methods for sparse signal recovery using noiseless measurements (see Section 5 of [7] and references therein).

A. Results and Related Work

Recently, several Bayesian approaches have been proposed for compressive sensing with adaptive measurements (e.g., see [8], [1]). Furthermore, they have been shown to work well in practice, often requiring fewer noisy measurements to recover sparse signals than their non-adaptive competitors. In this paper we seek to prove theoretically that adaptive measurement procedures can indeed reliably recover sparse signals using fewer noisy linear measurements than non-adaptive approaches under the above background noise model (see Equation 1). To accomplish this goal we will study sparse support recovery.

A large body of work on solving sparse support identification problems has concentrated on measurement matrices whose entries are all i.i.d. mean zero Gaussian random variables (e.g., see [12],[6], and references therein). Such measurement matrices are particularly relevant to study given their near-optimal properties with respect to non-adaptive compressive sensing [2], [3] measurement design (e.g., see [10]). In particular, we will focus on the following result concerning support recovery using noisy non-adaptive Gaussian measurements.

Theorem 1: Let $C_{\min} = \min\left\{ |f_j| \; \middle| \; j \in \mathrm{supp}(f) \right\}$ be the magnitude of the smallest of the k non-zero entries in $f \in \mathbb{R}^N$. Next, suppose that $\mathcal{M}$ is an $m \times N$ non-adaptive random matrix with each row drawn from the zero-mean isotropic Gaussian distribution $\mathcal{N}(0, I_{N \times N})$. Furthermore, assume the noise matrix, $\mathcal{G}$, has i.i.d. $\mathcal{N}(0, \sigma^2/N)$ entries so that accumulated per-row measurement noise from Equation 1 is

$\langle \mathcal{M}_j, \mathcal{G}_j \rangle \sim \mathcal{N}(0, \|\mathcal{M}_j\|_2^2 \cdot \sigma^2/N)$. In this case 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 $\mathrm{supp}(f)$. That is, as $N \to \infty$ any algorithm will fail to recover $\mathrm{supp}(f)$ with probability bounded above 0.

Proof Sketch: The proof follows by adapting Wainwright's necessary conditions for perfect sparse support identification via non-adaptive Gaussian measurements (see [12]) to our background noise measurement model. Informally, adapting the proof depends on utilizing the fact that $\|\mathcal{M}_j\|_2^2$ quickly converges to N with probability rapidly approaching 1 as $N \to \infty$. $\square$

In effect, Theorem 1 provides a non-adaptive Gaussian measurement bound below which any recovery method must fail to be asymptotically reliable for support identification. In this paper we utilize adaptive combinatorial group testing [4] methods to recover highly sparse signals with fewer noisy measurements (see Equation 1) than such non-adaptive methods require. In particular, using group testing methods in combination with statistical binary detection and estimation techniques [9] we prove the following result for the special case of Gaussian measurement noise.

Theorem 2: Let $C_{\min} = \min\left\{|f_j| \;\middle|\; j \in \mathrm{supp}(f)\right\}$ be the magnitude of the smallest non-zero entry in $f \in \mathbb{R}^N$, and suppose the noise matrix, $\mathcal{G}$, has i.i.d. $\mathcal{N}(0, \sigma^2/N)$ entries. Furthermore, suppose that $\sigma^2/C_{\min}^2$ is $\Omega\left(k \cdot \ln^3 N\right)$[1]. Then, there exists a constant $c \in \mathbb{R}^+$ such that if the number of allowed measurements, m, exceeds

$$c \cdot \frac{\sigma^2}{C_{\min}^2} \cdot \ln^2\left(k \cdot \ln^3 N \cdot \ln^2(k \ln^2 N)\right)$$

our adaptive group testing methods can recover $\mathrm{supp}(f)$ with probability $\to 1$ as $N \to \infty$.

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 we 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, our non-adaptive group testing methods 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 our measurements sufficiently noisy, our adaptive methods will asymptotically outperform any sparse support recovery method utilizing non-adaptive Gaussian measurement ensembles.

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

The remainder of this paper is structured as follows. In Section II we describe our measurement model and fix notation. Then, in Section III, we present and analyze a simple binary search procedure for recovering 1-sparse signals in noise. In Section IV we describe a method for reducing general sparse support recovery problems to a collection of 1-sparse support recovery problems. This allows us to use multiple binary search procedures to recover the support of any sparse signal. Finally, we conclude with a short discussion in Section V.

II. Preliminaries

For ease of presentation we will phrase the sparse approximation problem in terms of recovering functions consisting of k weighted Dirac delta functions on the unit interval $[0, 1]$. This will allow us to bisect intervals (i.e., sub-arrays) at will without considering odd array sizes. Thus, we are interested in recovering functions of the form

$$f(x) = \sum_{j=1}^{k} C_j \cdot \delta(x - x_j) \tag{2}$$

where each $C_j \in \mathbb{R}$, and $x_j \in [0, 1]$, for $j \in \mathbb{Z} \cap [1, k]$. In recovering f we want to approximate both C_j and x_j for all $j \in \mathbb{Z} \cap [1, k]$. In approximating each x_j we will be satisfied to locate x_j to within $\frac{1}{N}-$tolerance for a given $N \in \mathbb{Z}^+$ which is guaranteed to have

$$\frac{1}{N} < \min\left\{|x_j - x_l| \;\middle|\; j \in [1, k] \cap \mathbb{Z},\ l \in \mathbb{Z} \cap [1, k] - \{j\}\right\}. \tag{3}$$

In other words, N gives a guaranteed separating distance between the Dirac delta functions composing f.

We are allowed to take integral measurements at each time $t \in \mathbb{R}^+$ to determine f. 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

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

For any subset $\mathcal{I} \subseteq [0, 1]$ we can measure

$$m_{\mathcal{I}}(t) = \int \mathbb{I}_{\mathcal{I}} \cdot f \; dx + \int \mathbb{I}_{\mathcal{I}} \; dW_t, \tag{4}$$

where $W_t(x)$ is a Wiener process that generates Gaussian measurement noise over $[0, 1]$. Furthermore, we assume that $W_t(x)$ is regenerated identically and independently at each time $t \in \mathbb{R}^+$. Thus, for each time $t \in \mathbb{R}^+$ we have that

$$\int \mathbb{I}_{\mathcal{I}} \; dW_t \quad \text{and} \quad \int \mathbb{I}_{\mathcal{J}} \; dW_t$$

are i.i.d. Gaussian random variables whenever $\mathcal{I} \cap \mathcal{J} = \emptyset$ and

$$\int \mathbb{I}_{\mathcal{I}} \; dx = \int \mathbb{I}_{\mathcal{J}} \; dx. \tag{5}$$

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

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

are i.i.d. Gaussian random variables as long as $\mathcal{I}, \mathcal{J} \subseteq [0,1]$ satisfy Equation 5. We finish by defining $\sigma_{\mathcal{I}}^2$ to be the variance of $m_{\mathcal{I}}(t)$ for a given $\mathcal{I} \subseteq [0,1]$. Given the assumptions above we can see that $\sigma_{\mathcal{I}}^2 = \sigma_{\mathcal{J}}^2$ whenever $\mathcal{I}, \mathcal{J} \subseteq [0,1]$ satisfy Equation 5. Finally, we will denote the noise over the entire unit interval by $\sigma^2 = \sigma_{[0,1]}^2$.

III. Single Spike Recovery

In this section we will assume that our function f consists of a single Dirac delta function (i.e., $k = 1$ in Equation 2). We will employ a simple adaptive binary search procedure to locate the support of f. However, before we can present the procedure in detail and prove that it succeeds we must first define the *left* and *right* subsets of any particular set $\mathcal{I} \subseteq [0,1]$. Given $\mathcal{I} \subseteq [0,1]$ with positive measure, define $x_{\text{mid}} \in [0,1]$ to be the unique point with

$$\int_0^{x_{\text{mid}}} \mathbb{I}_{\mathcal{I}}\, dx = \int_{x_{\text{mid}}}^1 \mathbb{I}_{\mathcal{I}}\, dx = \frac{1}{2}\int \mathbb{I}_{\mathcal{I}}\, dx.$$

We then define the *left subset* of $\mathcal{I}$, denoted $\mathcal{I}_l$, to be

$$\mathcal{I}_l = [0, x_{\text{mid}}) \cap \mathcal{I}. \tag{6}$$

Similarly, we define the *right subset* of $\mathcal{I}$, denoted $\mathcal{I}_r$, to be

$$\mathcal{I}_r = [x_{\text{mid}}, 1] \cap \mathcal{I}. \tag{7}$$

Given this definition, we are ready to discuss Algorithm 1 for locating a single Dirac delta function.

Assume that C_1 is positive for the time being. If so, we can begin looking 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 i.i.d. on both $[0, \frac{1}{2})$ and $[\frac{1}{2}, 1]$, the interval which contains x_1 will have a larger mean than the interval not containing x_1. Thus, our measurements for the interval 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 our binary search. If C_1 is negative the binary search is analogous. We simply repeatedly choose the subinterval which returns the smaller value more often. See Algorithm 1 for pseudocode.

Lemma 1: Fix $p \in (0,1)$. Then Algorithm 1 can correctly locate a C_1 magnitude spike within either $\mathcal{I}^+$ or $\mathcal{I}^-$ with probability at least p using less than

$$\left(54\pi \cdot \frac{\sigma_{\mathcal{I}}^2}{C_1^2} + \log_2(2N)\right) \cdot \ln\left(\frac{\log_2 2N}{1-p}\right)$$

measurements.

Proof Sketch: Suppose $C_1 > 0$. Define $\mathcal{I}_c^+$ to be the "correct" left/right subset of $\mathcal{I}^+$ in line 7 of Algorithm 1 containing the spike. Let $\mathcal{I}_w^+$ be the other "wrong" subset. Bound $\mathbb{P}\left[m_{\mathcal{I}_c^+}(t) > m_{\mathcal{I}_w^+}(t.)\right] - \frac{1}{2}$ away from zero in the noisy setting where $C_1 < 2\sigma_{\mathcal{I}_w^+}$ and then apply the Chernoff bound. The small noise case is obvious. □

Finally, we deal with the fact that we don't have a priori 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. 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 $x_1 \in \mathcal{I}^-$.

We now address the estimation portion of Algorithm 1 (lines 14 through 23). In fact, this step is also necessary to complete the location of our spike if the sign of C_1 is unknown. The approach we use is to simply estimate the mean of measurements from $\mathcal{I}^+$ and $\mathcal{I}^-$. One interval, $\mathcal{I}^+$, should contain our spike if C_1 is positive. Similarly, $\mathcal{I}^-$ should contain our spike if C_1 is negative. In either case, the interval not containing x_1 will have a smaller absolute mean value with high probability (i.e., if we assume mean 0 noise).

Lemma 2: Fix $\alpha \in (0, \frac{1}{2})$, and $p \in (\frac{1}{2}, 1)$. Finally, suppose

Algorithm 1 Isolated Delta

1: **Input: Initial subset** $\mathcal{I} \subseteq [0,1]$**, position tolerance** N**, magnitude tolerance** α**, success probability** p
2: **Output: Estimate of magnitude,** C_1**, and position,** x_1
3: $\mathcal{I}^+ \longleftarrow \mathcal{I}$
4: $\mathcal{I}^- \longleftarrow \mathcal{I}$

LOCATE x_1

5: **while** $\int \mathbb{I}_{\mathcal{I}^+}\, dx > \frac{1}{N}$ **do**
6: *Assuming C_1 is positive, find x_1...*
7: **if** $m_{\mathcal{I}_l^+}(t) > m_{\mathcal{I}_r^+}(t)$ the majority of $T(N,p)$ trails **then**
8: $\mathcal{I}^+ \longleftarrow \mathcal{I}_l^+$
9: **else**
10: $\mathcal{I}^+ \longleftarrow \mathcal{I}_r^+$
11: **end if**
12: **end while**
13: *Repeat lines 5 – 12 assuming C_1 is negative to get $\mathcal{I}^-$.*

ESTIMATE C_1

14: $\tilde{C}^+ \longleftarrow$ Estimated mean from $m_{\mathcal{I}^+}(t.)$ (α-precise)
15: $\tilde{C}^- \longleftarrow$ Estimated mean from $m_{\mathcal{I}^-}(t.)$ (α-precise)
16: *Decide if C_1 is positive or negative...*
17: **if** $\left|\tilde{C}^+\right| > \left|\tilde{C}^-\right|$ **then**
18: $\tilde{C}_1 \longleftarrow \tilde{C}^+$
19: $\mathcal{I}^e \longleftarrow \mathcal{I}^+$
20: **else**
21: $\tilde{C}_1 \longleftarrow \tilde{C}^-$
22: $\mathcal{I}^e \longleftarrow \mathcal{I}^-$
23: **end if**
24: Return $\tilde{x}_1 =$ midpoint of $\mathcal{I}^e$, $\tilde{C}_1$

that $x_1 \in \mathcal{I}^+ \cup \mathcal{I}^-$ and $\mathbb{V}\mathrm{ar}\left[m_{\mathcal{I}}(t)\right] = \sigma_{\mathcal{I}}^2$. Then, Algorithm 1 (lines 14 through 23) can both determine which set x_1 belongs to (i.e., either $\mathcal{I}^+$ or $\mathcal{I}^-$) and estimate C_1 to precision $\alpha \cdot C_1$ with probability at least p. The number of required measurements is less than $2 + 2 \cdot \frac{\sigma_{\mathcal{I}}^2}{N \cdot \alpha^2 C_1^2} \cdot \log\left(\frac{1}{1-p}\right)$.

Proof Sketch: Follows by bounding the probability that a maximum likelihood estimator for the mean of $m_{\mathcal{I}^{-/+}}(t.)$ measurements deviates by $> \alpha C_1$ from the true mean. □

We conclude with a recovery theorem for Algorithm 1.

Theorem 3: Suppose there is a single spike $C_1 \cdot \delta(x - x_1)$ in $\mathcal{I} \subseteq [0,1]$. Let $\sigma_{\mathcal{I}}^2 = \mathbb{V}\mathrm{ar}\left[m_{\mathcal{I}}(t)\right]$. Fix $\alpha \in (0, \frac{1}{2})$, and $p \in (\frac{1}{2}, 1)$. Then, a variant of Algorithm 1 can output $\tilde{x}_1, \tilde{C}_1$ for which both $\left|\tilde{x}_1 - x_1\right| \leq \frac{1}{2N}$ and $\left|\tilde{C}_1 - C_1\right| \leq \alpha \cdot C_1$ are true with probability at least p. The number of required measurements is $O\left(\log N + \left(\frac{\sigma_{\mathcal{I}}^2}{C_1^2} + \frac{\sigma_{\mathcal{I}}^2}{N \cdot \alpha^2 C_1^2}\right) \cdot \log\left(\frac{\log N}{1-p}\right)\right)$.

The proof of Theorem 3 follows from Lemmas 1 and 2. To finish, we quickly note that the constant factors provided in Lemmas 1 and 2 are crude. However, they suffice given our interest in asymptotic behavior. Finally, we note that many different variants of Algorithm 1 would also satisfy Theorem 3 while simultaneously yielding better results in practice. In particular, each binary search decision could be made using optimal statistical binary detection methods (e.g., [9], [11]).

IV. Multiple Spike Recovery

In this section we will demonstrate how to utilize Algorithm 1 to recover signals consisting of k spikes (i.e., how to determine f in Equation 2). Our approach will be to partition $[0,1]$ into several smaller subsets of near-equal length, so that each spike is isolated by itself in at least one of the subsets. We then apply Algorithm 1 to each subset. Algorithm 1 will recover each spike isolated in a subset by Theorem 3. On subsets which don't isolate a spike we will, at worst, recover a "fake spike" with a magnitude small enough to ignore. Thus, as long as Algorithm 1 succeeds with high enough probability on each subset, we will recover good estimates of all k spikes and nothing extra. We now construct isolating subsets of $[0,1]$.

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,

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

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 others with fixed probability. We can then use several of these disjoint unions to separate each of our spikes with arbitrarily high probability. We begin by describing our disjoint unions.

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

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

and then set

$$\mathcal{I}_q = \left\{ \mathcal{I}_{q,h} \;\middle|\; h \in [0, q) \cap \mathbb{Z} \right\} \tag{10}$$

$\mathcal{I}_q$ is our set of disjoint s_j-unions. The following Lemma demonstrates that a randomly constructed $\mathcal{I}_q$ is likely to contain many subsets of $[0,1]$ with a single spike.

Lemma 3: Fix an f containing at most k spikes (see Equation 2). Choose one of the first $2k\lfloor \log_k N \rfloor$ prime numbers larger than k uniformly at random. Then each x_j, with probability at least $\frac{1}{2}$, is isolated in its associated $\mathcal{I}_{q,h} \in \mathcal{I}_q$. In other words, for each x_j there exists an $\mathcal{I}_{q,h} \in \mathcal{I}_q$ so that $\{x_1, \ldots, x_j, \ldots, x_k\} \cap \mathcal{I}_{q,h} = \{x_j\}$ with probability $> \frac{1}{2}$.

Proof: Each x_j may collide with one of the other at most $(k-1)$ spikes in a $\mathcal{I}_{q,h}$-subset for at most $\lfloor \log_k N \rfloor$ values of q by the Chinese Remainder Theorem. Thus, x_j may collide with *any* of the other $\leq (k-1)$ spikes for at most $(k-1) \cdot \lfloor \log_k N \rfloor$ values of q. Hence, more than half of our potential q-values must isolate x_j from the other at most $k-1$ spike supports. □

Looking at Lemma 3 we can see that if we select $\log_2\left(\frac{k}{1-p}\right)$ q-primes independently and uniformly at random, and then form their related $\mathcal{I}_q$-subsets, we will isolate all of f's spikes at least once with probability at least p. Hence, we can utilize $\log_2\left(\frac{2k}{1-p}\right)$ q-primes in order to guarantee that we fail in isolating all spikes with probability at most $\frac{1-p}{2}$. Let $q_{\max}$ be the largest of our randomly selected primes. If we also guarantee that Algorithm 1 will fail (in the presence of an isolated spike) on any of these at most $q_{\max} \cdot \log_2\left(\frac{2k}{1-p}\right)$ total $\mathcal{I}_{q,h}$-subsets with probability at most $\frac{1-p}{2}$, we will assure our overall desired success probability. This can be accomplished by using Algorithm 1 with enough measurements to ensure that it fails in correctly locating an isolated spike at each binary search stage with probability at most

$$\frac{1-p}{2 \cdot q_{\max} \cdot \log_2\left(\frac{2k}{1-p}\right) \cdot \log_2 N}. \tag{11}$$

The end result will be that we correctly locate each spike at least once with probability at least p. We can then estimate each located spike's magnitude using Lemma 2. Note that $q_{\max} \leq 10k \cdot \log_k N \log_2(5k \cdot \log_k N)$.

To finish recovery, we simply return all the spikes Algorithm 1 outputs (allowing only one $\tilde{x}_j$ 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 2) 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 the supplied $C_{\min}$ 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 2 for multiple spike recovery pseudocode.

Algorithm 2 NO MORE THAN K DELTAS

Input: Maximum number of spikes k, Position tolerance N, magnitude tolerance α, smallest spike magnitude of interest $C_{\min}$, success probability p
Output: Estimates of magnitudes $> \frac{1}{2}C_{\min}$, $\{\tilde{C}_1, \ldots, \tilde{C}_k\}$, and their positions, $\{\tilde{x}_1, \ldots, \tilde{x}_k\}$
Find all spikes at least once...
$SPIKES \longleftarrow \emptyset$
for $j = 1,\ j < O\left(\log\left(\frac{k}{1-p}\right)\right),\ j++$ **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**
    $(\tilde{x}, \tilde{C}) \longleftarrow$ Algorithm 1($\mathcal{I}_{q,h}$,$2N$,α,1 - Equation 11)
    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 $|\tilde{x}_n - \tilde{x}_{n+1}| \leq \frac{N}{2}$ **do**
    $SPIKES \leftarrow SPIKES - \{(\tilde{x}_{n+1}, \tilde{C}_{n+1})\}$
  end while
end for
Return $SPIKES$

We are now ready to consider the measurements required to locate all k spikes and estimate their magnitudes. Let $\sigma^2_{[0,1]}$ be the variance of our measurement noise over $[0,1]$. Then we can see that $\mathbb{V}\mathrm{ar}\left[m_{\mathcal{I}_{q,h}}(t)\right]$ will be $O\left(\frac{\sigma^2_{[0,1]}}{q}\right)$. Applying Theorem 3 to each of our $O\left(q_{\max} \cdot \log_2\left(\frac{2k}{1-p}\right)\right)$ $\mathcal{I}_{q,h}$-subsets with the required Algorithm 1 success probability guarantee (see Equation 11) we can see that we need no more than

$$O\left(\frac{\sigma^2_{[0,1]}}{C^2_{\min}}\left(1 + \frac{1}{N\alpha^2}\right)\log^2\left(\frac{k \cdot \log^2\left(\frac{k\log N}{1-p}\right) \cdot \log^2 N}{1-p}\right) + k \cdot \log^2\left(\frac{k\log N}{1-p}\right) \cdot \frac{\log^2 N}{\log k}\right) \quad (12)$$

measurements to find and estimate all k spikes with probability at least p. We obtain the following theorem.

Theorem 4: Fix $\alpha \in (0, \frac{1}{2})$, $p \in (0,1)$, and $C_{\min} \in \mathbb{R}^+$. Let σ^2 be the variance of $m_{[0,1]}(t)$. Finally, suppose that there are at most k spikes, $C_1 \cdot \delta(x - x_1), \ldots, C_k \cdot \delta(x - x_k)$, in $[0,1]$. Then Algorithm 2 can, with probability at least p, output $\tilde{x}_j, \tilde{C}_j$ for all spikes with $|C_j| \geq C_{\min}$ such that both $\left|\tilde{x}_j - x_j\right| \leq \frac{1}{2N}$ and $\left|\tilde{C}_j - C_j\right| \leq \alpha \cdot C_{\min}$ are true. The number of required measurements is bounded above by Equation 12.

Theorem 2 is a corollary of Theorem 4. Discretizing Theorem 4 is simply a matter of carefully splitting sub-arrays during each binary search (i.e., each invocation of Algorithm 1). Splitting an even length sub-array is done in the obvious way. Splitting an odd length sub-array can be done by ignoring the last array element during the binary search and individually testing it later for a spike with Lemma 2. In order to send the probability of success, p, to 1 as N goes to infinity we can simply set $p = 1 - \frac{1}{\log N}$ in Equation 12.

V. CONCLUSION

As presented here Algorithm 1 is adaptive, requiring the fast bisection of its initial input subset. Assuming that both bisecting and measuring intervals can be done at unit cost, Algorithm 2 runs in $O\left(\left(\frac{\sigma^2_{[0,1]}}{C^2_{\min}} + \frac{\sigma^2_{[0,1]}}{N \cdot \alpha^2 C^2_{\min}} + k\right) \cdot \log^{O(1)}\left(\frac{N}{1-p}\right)\right)$-time. Hence, for modest noise levels the required runtime is sublinear in N. Finally, if we adaptively create our $\mathcal{I}_q$-subsets with smaller q-values as more spikes are discovered, we should be able to reduce the measurement costs of Algorithm 2 significantly in practice. If our measurements are also fast to construct on the fly, adaptive creation of our $\mathcal{I}_q$-subsets should also decrease the runtime in practice.

Finally, we point out that the simplicity of Algorithm 1 allows us to obtain similar theoretical results for much more general types of measurement noise. For example, it is sufficient that the measurement noise matrix, $\mathcal{G}$, have i.i.d. entries with arbitrary mean and bounded variance. Furthermore, noise matrix entries may simply be uncorrelated.

ACKNOWLEDGMENTS

I would like to thank Vasileios Maroulas and Ahmed Tewfik for helpful discussions and advice. I would also like to thank Ahmed Tewfik for commenting on an early draft of this paper.

REFERENCES

[1] R. Castro, J. Haupt, R. Nowak, and G. Raz. Finding needles in noisy haystacks. *ICASSP*, 2008.
[2] D. Donoho. Compressed Sensing. *IEEE Trans. on Information Theory*, 52:1289–1306, 2006.
[3] D. L. Donoho, M. Elad, and V. N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inform. Theory*, 52(1), Jan 2006.
[4] D. Z. Du and F. K. Hwang. *Combinatorial Group Testing and Its Applications*. World Scientific, 1993.
[5] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 83, March 2008.
[6] A. K. Fletcher, S. Rangan, and V. K. Goyal. Necessary and sufficient conditions on sparsity pattern recovery. *Preprint*, April 2008.
[7] A. C. Gilbert, M. A. Iwen, and M. J. Strauss. Group testing and sparse signal recovery. *42nd Asilomar Conference on Signals, Systems, and Computers*, 2008.
[8] S. Ji, Y. Xue, and L. Carin. Bayesian compressive sensing. *IEEE Trans. Signal Processing*, 56(6):2346–2356, June 2008.
[9] D. Middleton. *An Introduction to Statistical Communication Theory*. McGraw-Hill Book Company, Inc., New York, NY, USA, 1960.
[10] 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.
[11] A. G. Tartakovsky, X. R. Li, and G. Yaralov. Sequential detection of targets in multichannel systems. *IEEE Transactions on Information Theory*, 49(2), Feb. 2003.
[12] M. Wainwright. Information-theoretic bounds on sparsity recovery in the high-dimensional and noisy setting. *International Symposium on Information Theory*, June 2007.