Adaptive Strategies for Target Detection and Localization in Noisy Environments

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.**

***Index Terms*—Adaptive compressed sensing, clutter noise, compressed sensing (CS), information theory, radar**

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

This paper deals with a slight variant of the standard sparse approximation problem encountered in the compressed sensing literature. To make the noise model more clear, we will begin by presenting it in a standard sparse approximation setting. In the process we will fix notation. Then, in Section I-C, we will present a closely related problem on which we focus for the remainder of the paper.

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).

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

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

This paper generalizes and improves 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].

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.

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, $1 \le n \le j-1$. 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 $C_{\min} = \min\left\{ |\langle f,\ \phi_i \rangle| \ \middle|\ i \in \mathrm{supp}(f) \right\}$ 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 $1 \le j \le m$, 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 $\mathcal{G} = \left\{ \mathcal{G}_j \ \middle|\ 1 \le j \le m \right\}$ is an ensemble of m non-adaptive random standard Gaussian noise processes independently drawn for each j. Create test functions by setting*

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

Furthermore, let $\mathcal{P}_j$ for $1 \le j \le m$ 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

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

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

Proof: See the Appendix.[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 [18] are utilized in combination with statistical binary detection and estimation techniques [19] to produce the following theorem.

Theorem 2. *Let $\mathcal{P}_j$ for $1 \le j \le m$ 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*

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

Algorithm 2 will approximate f precisely enough to reliably recover $\mathrm{supp}(f)$ *with probability $\to 1$ as $N \to \infty$.*

Proof: Apply Corollary 2 from Section III-D with $p = 1 - 1/\ln N$. □

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

[1] Since the initial appearance of Theorem 1 in [1], [2] similar results have appeared independently in [16], [17].

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

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.

Finally, it is worth noting that lower bounds for the number of adaptive measurements required in order to recover sparse signals have been developed in the "thermal noise" setting (i.e., for measurements provided by Equation 2 above) [20]. However, proving similar lower bounds for the types of adaptive measurements considered herein (i.e., see Equation 1) remains an open problem to the best of our knowledge.

C. A Simplified Problem Setup

As mentioned previously, the measurement noise model considered in this paper is motivated by target location by radar. In keeping with this motivation, below we develop an adaptive method for recovering functions of the form

$$f(x) = \sum_{j=1}^{k} C_j \cdot \delta(x - x_j) \tag{3}$$

where $\delta(x)$ is a Dirac delta function, each $C_j \in \mathbb{R}$, and $x_j \in [0,1]$, for $j \in \mathbb{Z} \cap [1,k]$. This problem is a simplified model for the problem of recovering an unknown number of ideal point targets located at positions x_j and with reflectivity C_j. The model assumes prior knowledge concerning the range of the points x_j. The given range is then normalized to the interval $[0,1]$ without loss of generality. In this form the simplified model also captures radar imaging of targets that consist of a collection of point reflectors, a target model often considered in the literature.

Selecting measurement functions, $\mathcal{M}_j$, corresponds to selecting a radar beamform and illumination pattern. For the remainder of this paper we restrict our attention to characteristic measurement functions which are constructed for sequences of adaptively refined unions of subsets of $[0,1]$. These measurement functions, or test functions, will yield noisy measurements of the signal f at each time $t \in \mathbb{R}^+$ as follows. 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 $\mathcal{I} \subseteq [0,1]$ and time $t \in \mathbb{R}^+$ we assume that we can measure

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

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

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

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 5. Let $\sigma^2_{\mathcal{I}}$ to be the variance of $m_{\mathcal{I}}(t)$ for a given $\mathcal{I} \subseteq [0,1]$. Given the assumptions above, $\sigma^2_{\mathcal{I}}$ will equal $\sigma^2_{\mathcal{J}}$ whenever $\mathcal{I}, \mathcal{J} \subseteq [0,1]$ satisfy Equation 5. Finally, denote the variance of the noise over the entire unit interval by $\sigma^2 = \sigma^2_{[0,1]}$.

To recover f we must 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

$$\frac{1}{N} < \min \{ |x_j - x_l| \;\big|\; j \in [1,k] \cap \mathbb{Z}, \; l \in \mathbb{Z} \cap [1,k] - \{j\} \}. \tag{6}$$

In other words, N gives a guaranteed separating distance between the Dirac delta functions composing f. The separating distance assumption allows us to recognize the measurements defined in Equation 4 as a slight variation of the measurement model described by Equation 1. This relationship ultimately allows us to apply the adaptive targeting methods developed below to the standard sparse approximation problem described above in Section I-A. See [2] for a detailed description of how the simplified setup presented in this section relates to the standard sparse approximation setup discussed in Section I-A.

The remainder of this paper is structured as follows. In Section II we present and analyze a simple binary search procedure for recovering 1-sparse signals in background noise. Then, in Section III, 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 sparsely representable signal. Finally, we conclude with a short discussion in Section IV.

II. Single Spike Targeting

In this section we assume that the function f consists of a single Dirac delta function (i.e., $k = 1$ in Equation 3). Given this assumption a simple adaptive binary search procedure will be employed to locate the support of f. However, before the procedure can be presented in detail we must first define the *left* and *right* subsets of a given set $\mathcal{I} \subseteq [0,1]$. Given any

$\mathcal{I} \subseteq [0,1]$ with positive measure, define $x_{\text{mid}} \in [0,1]$ to be the unique point with

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

Let the *left subset* of $\mathcal{I}$, denoted $\mathcal{I}_l$, be

$$\mathcal{I}_l = [0, x_{\text{mid}}) \cap \mathcal{I}. \tag{7}$$

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

$$\mathcal{I}_r = [x_{\text{mid}}, 1] \cap \mathcal{I}. \tag{8}$$

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 independent and identically distributed (i.i.d.) on both $[0, \frac{1}{2})$ and $[\frac{1}{2}, 1]$, the interval containing x_1 will have a larger mean than the interval not containing x_1. Thus, the 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 a 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. 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, the search assuming positivity will locate the spike with high probability. If C_1 is negative, the 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 the 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 the intervals resulting from the two binary searches, and then compare them to the mean of the arbitrary interval's measurements. Whichever binary search result differs most from the arbitrary interval in terms of measurement mean will be the correct search result with high probability. See Algorithm 1 for pseudocode.

Algorithm 1 ISOLATED DELTA

1: **Input: Initial subset $\mathcal{I} \subseteq [0,1]$, position tolerance N, magnitude tolerance α, success probability p, total measurement budget B, estimation measurement budget γ**
2: **Output: Estimate of magnitude, C_1, and position, x_1**
3: $\mathcal{I}^+ \longleftarrow \mathcal{I}$
4: $\mathcal{I}^- \longleftarrow \mathcal{I}$
5: Initialize $K \leftarrow$ constant $\in \mathbb{N}$ (See the proof of Lemma 1)
6: Initialize $B' \leftarrow (B - 3\gamma)/8K$

LOCATE x_1

7: **while** $\int \mathbb{I}_{\mathcal{I}^+}\, dx \;>\; \frac{1}{N}$ **do**
8: *Assuming C_1 is positive, find x_1...*
9: **if** $\frac{1}{K}\cdot\sum_{k=1}^K m_{\mathcal{I}_l^+}(t_k) > \frac{1}{K}\cdot\sum_{k=1}^K m_{\mathcal{I}_r^+}(t_k)$ the majority of B' trials **then**
10: $\mathcal{I}^+ \longleftarrow \mathcal{I}_l^+$
11: **else**
12: $\mathcal{I}^+ \longleftarrow \mathcal{I}_r^+$
13: **end if**
14: *Assuming C_1 is negative, find x_1...*
15: **if** $\frac{1}{K}\cdot\sum_{k=1}^K m_{\mathcal{I}_l^-}(t_k) < \frac{1}{K}\cdot\sum_{k=1}^K m_{\mathcal{I}_r^-}(t_k)$ the majority of B' trials **then**
16: $\mathcal{I}^- \longleftarrow \mathcal{I}_l^-$
17: **else**
18: $\mathcal{I}^- \longleftarrow \mathcal{I}_r^-$
19: **end if**
20: $K \longleftarrow \lceil K/2 \rceil$
21: $B' \longleftarrow \lceil B'/2 \rceil$
22: **end while**

ESTIMATE C_1

23: Choose arbitrary $\mathcal{I}^N \subset \mathcal{I} - (\mathcal{I}^- \cup \mathcal{I}^+)$ with $\int \mathbb{I}_{\mathcal{I}^N}\, dx \;= \int \mathbb{I}_{\mathcal{I}^+}\, dx$
24: $\tilde{C}^+ \longleftarrow$ Estimated mean from γ measurements of $\mathcal{I}^+$
25: $\tilde{C}^- \longleftarrow$ Estimated mean from γ measurements of $\mathcal{I}^-$
26: $\tilde{C}^N \longleftarrow$ Estimated mean from γ measurements of $\mathcal{I}^N$
27: *Decide if C_1 is positive or negative...*
28: **if** $\left|\tilde{C}^+ - \tilde{C}^N\right| > \left|\tilde{C}^- - \tilde{C}^N\right|$ **then**
29: $\tilde{C}_1 \longleftarrow \tilde{C}^+ - \tilde{C}^N$
30: $\mathcal{I}^e \longleftarrow \mathcal{I}^+$
31: **else**
32: $\tilde{C}_1 \longleftarrow \tilde{C}^- - \tilde{C}^N$
33: $\mathcal{I}^e \longleftarrow \mathcal{I}^-$
34: **end if**
35: Return $\tilde{x}_1 =$ midpoint of $\mathcal{I}^e$, $\tilde{C}_1$

Lemma 1. *$B - 3\gamma = O\left(\left(\frac{\sigma_I^2}{C_1^2} + \ln N\right)\cdot \ln\left(\frac{\ln N}{1-p}\right)\right)$ initial measurements in line 6 of Algorithm 1 are sufficient to allow lines 7 through 22 of Algorithm 1 to correctly locate x_1 within either $\mathcal{I}^+$ or $\mathcal{I}^-$ with probability at least p.*

Proof: Assume that C_1 is positive (the case for C_1 negative is analogous). For each iteration of the "while"-loop in line 7 let $\mathcal{I}_c^+$ be the left or right subset of $\mathcal{I}^+$ containing x_1, and $\mathcal{I}_w^+$ be the other subset not containing x_1. Finally, let $D_{\mathcal{I}_w^+} : \mathbb{R} \mapsto \mathbb{R}^+$ be the density function of the real random variable $m_{\mathcal{I}_w^+}(t)$. Given our assumptions about the noise, $m_{\mathcal{I}_c^+}(\tilde{t})$ will

have density $D_{\mathcal{I}_w^+}(x - C_1)$ for all $\tilde{t} \in \mathbb{R}^+$.

Define $\epsilon_{\mathcal{I}_w^+}(C_1) \in \mathbb{R}^+$ to be

$$\int_{-\infty}^{\infty} \left(\int_{y-C_1}^{y} D_{\mathcal{I}_w^+}(x)\, dx \right) D_{\mathcal{I}_w^+}(y)\, dy.$$

A straightforward calculation reveals that for each "while"-loop iteration

$$\mathbb{P}\left[m_{\mathcal{I}_c^+}(t) > m_{\mathcal{I}_w^+}(t) \right] = \frac{1}{2} + \epsilon_{\mathcal{I}_w^+}(C_1).$$

We will compare several averaged measurements from $\mathcal{I}_c^+$ to several averaged measurements from $\mathcal{I}_w^+$ in order to decide which of these intervals contains the spike. This will ensure that $\epsilon_{\mathcal{I}_w^+}(C_1)$ is effectively larger than a constant $c \in (0, \frac{1}{2})$.

Set $M_c^+ = \frac{1}{K} \cdot \sum_{k=1}^{K} m_{\mathcal{I}_c^+}(t_k)$, $M_w^+ = \frac{1}{K} \cdot \sum_{k=1}^{K} m_{\mathcal{I}_w^+}(t_k)$. In this case we can see that

$$\mathbb{P}\left[M_c^+ > M_w^+ \right] = \mathbb{P}\left[(M_c^+ - M_w^+) > 0 \right]$$

where $\mathbb{E}\left[M_c^+ - M_w^+ \right] = C_1$ and

$$\mathbb{V}\mathrm{ar}\left[M_c^+ - M_w^+ \right] = 2 \cdot \frac{\sigma^2_{\mathcal{I}_w^+}}{K} = \frac{\sigma^2_{\mathcal{I}^+}}{K}.$$

Chebyshev's Inequality then guarantees that

$$\begin{aligned} \tilde{\epsilon}_{\mathcal{I}_w^+}(C_1) &= \mathbb{P}\left[M_c^+ > M_w^+ \right] - \frac{1}{2} \\ &\geq \frac{1}{2} - \mathbb{P}\left[\left| M_c^+ - M_w^+ - C_1 \right| \geq C_1 \right] \\ &\geq \frac{1}{2} - \frac{\sigma^2_{\mathcal{I}^+}}{K \cdot C_1^2} > c \end{aligned} \tag{9}$$

whenever $K > \left(\frac{1}{2} - c \right)^{-1} \cdot \frac{\sigma^2_{\mathcal{I}^+}}{C_1^2}$. Hence, we may assume hereafter that $\epsilon_{\mathcal{I}_w^+}(C_1)$ is larger than $c = 1/4$.

Applying the Chernoff bound we can see that $\frac{\ln\left(\frac{\log_2(N)}{1-p}\right)}{2 \cdot \epsilon^2_{\mathcal{I}_w^+}(C_1)}$ comparisons of M_c^+ with M_w^+ are sufficient to correctly decide $\mathcal{I}_c^+$ with error probability at most $\frac{1-p}{\log_2(N)}$ (see, e.g., [21]). The union bound then implies that all $O(\ln(N))$ iterations of line 7's "while"-loop will succeed in locating x_1 with probability at least p. In order to bound the total number of required measurements we note that after the n^{th} iteration of line 7's loop we will have $\sigma^2_{\mathcal{I}^+} = \frac{\sigma^2_{\mathcal{I}}}{2^n}$. Hence, at the n^{th} iteration of line 7 we will require no more than

$$\begin{aligned} B' &= O\left((K+1) \cdot \ln\left(\frac{\ln(N)}{1-p} \right) \right) \\ &= O\left(\left(1 + \frac{\sigma^2_{\mathcal{I}}}{2^n \cdot C_1^2} \right) \cdot \ln\left(\frac{\ln(N)}{1-p} \right) \right) \end{aligned}$$

measurements. When utilized for all $O(\ln N)$ binary search levels to identify $\mathcal{I}^+$ and $\mathcal{I}^-$ we will require a total of

$$\begin{aligned} B - 3\gamma &= \sum_{n=0}^{O(\ln N)} O\left(\left(1 + \frac{\sigma^2_{\mathcal{I}}}{2^n \cdot C_1^2} \right) \cdot \ln\left(\frac{\ln N}{1-p} \right) \right) \\ &= O\left(\left(\frac{\sigma^2_{\mathcal{I}}}{C_1^2} + \ln N \right) \cdot \ln\left(\frac{\ln N}{1-p} \right) \right) \end{aligned}$$

measurements. The result follows. □

To finish the analysis of Algorithm 1 we address its estimation portion (lines 23 through 34). In fact, this step is also necessary to complete the location of the single spike if the sign of C_1 is unknown. The approach we use is to simply estimate the mean of measurements from 3 different intervals of the same size. One interval, $\mathcal{I}^+$, should contain the spike if C_1 is positive. Similarly, $\mathcal{I}^-$ should contain the spike if C_1 is negative. Hence we compare the estimated means of measurements from both these intervals with the estimated measurement mean from another disjoint interval, $\mathcal{I}^N$, which should contain no spike. Whichever interval, $\mathcal{I}^+$ or $\mathcal{I}^-$, has the mean least like $\mathcal{I}^N$ will contain the spike with high probability. Furthermore, the estimates of the mean allow us to estimate C_1. Following this line of reasoning we obtain the following lemma.

Lemma 2. *Let $\mathcal{I}^+, \mathcal{I}^-, \mathcal{I}^N \subset \mathcal{I}$ be pairwise disjoint unions of at most two intervals with $m = \int \mathbb{I}_{\mathcal{I}^+}\, dx = \int \mathbb{I}_{\mathcal{I}^-}\, dx = \int \mathbb{I}_{\mathcal{I}^N}\, dx \leq \frac{\int \mathbb{I}_{\mathcal{I}}\, dx}{N}$. Fix $\alpha \in (0, \frac{1}{2})$, and $p \in (0,1)$. Finally, suppose that $x_1 \in \mathcal{I}^+ \cup \mathcal{I}^-$ and $\mathbb{V}\mathrm{ar}\left[m_{\mathcal{I}}(t) \right] = \sigma^2_{\mathcal{I}}$. Then, Algorithm 1 (lines 23 through 34) 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 $\gamma = O\left(\left(1 + \frac{\sigma^2_{\mathcal{I}}}{N \cdot \alpha^2 C_1^2} \right) \cdot \ln \frac{1}{1-p} \right)$.*

Proof: Let $M_K^+ = \frac{1}{K} \cdot \sum_{k=1}^{K} m_{\mathcal{I}^+}(t_k)$. We know that both $\mathbb{E}\left[M_K^+ \right] = \mathbb{E}\left[m_{\mathcal{I}^+}(t.) \right]$ and $\mathbb{V}\mathrm{ar}\left[M_K^+ \right] = \frac{1}{K} \cdot \mathbb{V}\mathrm{ar}\left[m_{\mathcal{I}^+}(t.) \right]$ are true. Thus, if we let K be $O\left(\frac{\mathbb{V}\mathrm{ar}\left[m_{\mathcal{I}^+}(t.) \right]}{\alpha^2 \cdot C_1^2} \right)$, Chebyshev's inequality tells us that we can obtain

$$\left| M_K^+ - \mathbb{E}\left[m_{\mathcal{I}^+}(t.) \right] \right| < \frac{\alpha}{2} \cdot C_1$$

with constant probability larger than $\frac{1}{2}$ (see, e.g., [21]). Therefore, if we estimate $\mathbb{E}\left[m_{\mathcal{I}^+}(t.) \right]$ by taking the median of $O\left(\log \frac{1}{1-p} \right)$ i.i.d. M_K^+ variables, the Chernoff bound guarantees we will estimate $\mathbb{E}\left[m_{\mathcal{I}^+}(t.) \right]$ to precision $\frac{\alpha}{2} \cdot C_1$ with probability at least $\frac{2+p}{3}$. The union bound tells us that if we also estimate both $\mathbb{E}\left[m_{\mathcal{I}^-}(t.) \right]$ and $\mathbb{E}\left[m_{\mathcal{I}^N}(t.) \right]$ in a similar fashion we will locate x_1 with probability at least p.

Continuing, we note that the measurement assumptions from Section I-B imply that $\mathbb{V}\mathrm{ar}\left[m_{\mathcal{I}^+}(t.) \right]$ is $O\left(\sigma^2_{\mathcal{I}}/N \right)$. Furthermore, if the subset $\mathcal{I}^{+/-}$ containing x_1 consists of two intervals we may utilize this lemma again with fixed $\alpha \approx \frac{1}{2}$ to determine which interval actually contains the spike. The desired result follows. □

We are now able to conclude this section with a general recovery guarantee 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 $\sigma^2_{\mathcal{I}} = \mathbb{V}\mathrm{ar}\left[m_{\mathcal{I}}(t) \right]$. Fix $\alpha \in (0, \frac{1}{2})$, and $p \in (0,1)$. 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*

$$O\left(\left(\ln N + \frac{\sigma^2_{\mathcal{I}}}{C_1^2} + \frac{\sigma^2_{\mathcal{I}}}{N \cdot \alpha^2 C_1^2} \right) \cdot \ln\left(\frac{\ln N}{1-p} \right) \right).$$

Proof: Lemma 1 guarantees the location of x_1 within either $\mathcal{I}^+$ or $\mathcal{I}^-$ with probability at least $1-(1-p)/2$ using the stated number of measurements. An application of Lemma 2 using the discovered $\mathcal{I}^{+/-}$ subsets in order to approximate C_1 to the desired α-tolerance also with probability at least $1-(1-p)/2$ then finishes the proof. □

The proof of Theorem 3 follows easily from Lemmas 1 and 2. However, as one might expect, the performance of Algorithm 1 can be calculated explicitly the characteristics of the measurement noise are better known.

A. Special Case: Gaussian Measurement Noise

We now consider Lemma 1 in the context of Gaussian white measurement noise with unknown mean μ. The goal of this section is to derive explicit measurement bounds for this special case (i.e., with explicit numerical constants). As above, we assume that C_1 is positive (C_1 negative is analogous) and let $\mathcal{I}_c^+$ be the left or right subset of $\mathcal{I}^+$ containing x_1 in the current iteration of Algorithm 1 line 7's "while"-loop. Call the other subset $\mathcal{I}_w^+$ and suppose that $m_{\mathcal{I}_w^+}(t) \sim \mathcal{N}\left(\mu, \sigma^2_{\mathcal{I}_w^+}\right)$. As in the proof of Lemma 1 we will define $\epsilon_{\mathcal{I}_w^+}(C_1)$ to be the probability (in excess of 1/2) that a measurement on the half of the subset containing x_1 is larger than a measurement on the half of the subset not containing x_1. Thus, we have

$$\epsilon_{\mathcal{I}_w^+}(C_1) = \frac{1}{2} - \mathbb{P}\left[\left(m_{\mathcal{I}_c^+}(t) - m_{\mathcal{I}_w^+}(t)\right) < 0\right]. \tag{10}$$

In the Gaussian noise case we know that $\left(m_{\mathcal{I}_c^+}(t) - m_{\mathcal{I}_w^+}(t)\right) \sim \mathcal{N}\left(C_1, 2\sigma^2_{\mathcal{I}_w^+}\right)$. Therefore, we know that

$$\begin{aligned}\epsilon_{\mathcal{I}_w^+}(C_1) &= \frac{-1}{2}\cdot \operatorname{erf}\left(\frac{-C_1}{2\sigma_{\mathcal{I}_w^+}}\right) \\ &= \frac{C_1}{2\sqrt{\pi}\sigma_{\mathcal{I}_w^+}}\left(\sum_{n=0}^{\infty}\frac{(-1)^n}{n!(2n+1)}\left(\frac{C_1}{2\sigma_{\mathcal{I}_w^+}}\right)^{2n}\right).\end{aligned}$$

Continuing, we bound $\epsilon_{\mathcal{I}_w^+}(C_1)$ away from zero in the noisy setting where $C_1 < 2\sigma_{\mathcal{I}_w^+}$ (i.e., when the spike's magnitude is less than 2 standard deviations). We have

$$\begin{aligned}\epsilon_{\mathcal{I}_w^+}(C_1) &> \frac{C_1}{2\sqrt{\pi}\sigma_{\mathcal{I}_w^+}}\left(1 - \frac{1}{3}\sum_{n=1}^{\infty}\frac{\left(\frac{C_1}{2\sigma_{\mathcal{I}_w^+}}\right)^{2n}}{n!}\right) \\ &\geq \frac{C_1}{6\sqrt{\pi}\sigma_{\mathcal{I}_w^+}}\left(4 - \mathrm{e}^{\left(\frac{C_1}{2\sigma_{\mathcal{I}_w^+}}\right)^2}\right).\end{aligned}$$

Substituting this expression into the Chernoff bound we can see that $2\ln\left(\frac{1}{1-p}\right)/\operatorname{erf}^2\left(\frac{C_1}{2\sigma_{\mathcal{I}_w^+}}\right)$, which is itself less than

$$2\cdot\max\left\{\frac{36\pi\cdot\sigma^2_{\mathcal{I}_w^+}\cdot\ln\left(\frac{1}{1-p}\right)}{\left(4-\exp\left(\frac{C_1^2}{4\cdot\sigma^2_{\mathcal{I}_w^+}}\right)\right)^2\cdot C_1^2}, \frac{\ln\left(\frac{1}{1-p}\right)}{\operatorname{erf}^2(1)}\right\}, \tag{11}$$

measurements suffice to correctly decide $\mathcal{I}_c^+$ with error probability at most $1-p$ for any magnitude C_1 (see proof of Lemma 1). Finally, we note that when the sign of C_1 is unknown we must we must perform *two* binary searches (one in case C_1 is positive, and another in case C_1 is negative). Thus, we have to double the measurement bound shown in Equation 11 for all but the first iteration of Algorithm 1's line 7 – 22 loop. We can now bound the total number of measurements required by Algorithm 1 to locate x_1 within either $\mathcal{I}^+$ or $\mathcal{I}^-$ with probability at least p.

Let σ^2 be the variance of the mean μ Gaussian white measurement noise over the entire interval $[0,1]$. Then, we must utilize Equation 11 at most $\log_2 N$ times to locate x_1 within a sufficiently small $\mathcal{I}^+$ or $\mathcal{I}^-$. We are able to bound the total number of sufficient measurements by

$$\frac{2\cdot\ln\left(\frac{\log_2 2N}{1-p}\right)}{\operatorname{erf}^2\left(\frac{C_1}{2\sigma}\right)} + \sum_{n=1}^{\lceil\log_2 N\rceil-1}\frac{4\cdot\ln\left(\frac{\log_2 2N}{1-p}\right)}{\operatorname{erf}^2\left(\frac{\left(\sqrt{2}\right)^n\cdot C_1}{2\sigma}\right)}$$

which in turn is bounded by

$$A := 2A_{\sigma^2} + 4\left(\sum_{n=1}^{\lceil\log_2 N\rceil-1} A_{\left(\frac{\sigma^2}{2^n}\right)}\right) \tag{12}$$

where

$$A_{\sigma^2} := \max\left\{\frac{36\pi\cdot\sigma^2\cdot\ln\left(\frac{\log_2 2N}{1-p}\right)}{\left(4-\exp\left(\frac{C_1^2}{4\cdot\sigma^2}\right)\right)^2\cdot C_1^2}, \frac{\ln\left(\frac{\log_2 2N}{1-p}\right)}{\operatorname{erf}^2(1)}\right\}.$$

This number of measurements suffices to locate x_1 to within the given tolerance with probability at least p. We obtain the following corollary of Lemma 1.

Corollary 1. *Let σ^2 be the variance of mean μ Gaussian measurement noise over the interval $[0,1]$. 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*

$$A = O\left(\left(\frac{\sigma^2}{C_1^2} + \ln(N)\right)\cdot\ln\left(\frac{\ln(N)}{1-p}\right)\right)$$

measurements (see Equation 12).

B. Summary: Single Spike Detection

To conclude, we note that by employing a binary search for single spike recovery we are essentially transforming the spike localization problem into $O(\ln N)$ binary detection problems. Without loss of generality, at each stage of the 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 the binary search. When viewed from this perspective, the single spike recovery problem becomes equivalent to a series of statistical detection/estimation problems (e.g., see [19], [22]). 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.,

[23]) at each stage of the 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.

III. Multiple Spike Targeting

This section demonstrates how to utilize Algorithm 1 to recover signals consisting of at most k spikes (i.e., how to determine f in Equation 3). The solution 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. An example of a set of disjoint subsets of $[0,1]$ which isolates each spike from all the others is constructed in the next section.

A. 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 $|x_{j_1} - x_{j_2}| > \frac{1}{N}$, one may represent $[0,1]$ by its N subintervals,

$$s_0 = \left[0, \frac{1}{N}\right), \ \ldots, \ s_{N-1} = \left[1 - \frac{1}{N}, 1\right], \tag{13}$$

only k of which contain spikes (i.e., one 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. Several of these disjoint unions can then be used to separate each of the spikes from all the others with arbitrarily high probability.[3] We begin by describing these disjoint unions of s_j-subsets.

Let q be one of the first $2k\lfloor \log_k N \rfloor$ prime numbers larger than k. These primes are easily found via standard sieving algorithms (see [24]). For each $h \in [0, q) \cap \mathbb{Z}$ form the set

$$\mathcal{I}_{q,h} = \bigcup_{j \equiv h \bmod q} s_j \tag{14}$$

and then set

$$\mathcal{I}_q = \left\{\mathcal{I}_{q,0}, \ \mathcal{I}_{q,1}, \ \ldots, \ \mathcal{I}_{q,q-1}\right\}. \tag{15}$$

The following Lemma demonstrates that a randomly constructed $\mathcal{I}_q$ is likely to contain many subsets of $[0,1]$ that isolate a given spike from all the others.

Lemma 3. *Fix an f containing at most k spikes (see Equation 3). 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 $\{x_1, \ldots, x_j, \ldots, x_k\} \cap \mathcal{I}_{q,h} = \{x_j\}$ is true with probability at least $\frac{1}{2}$.*

Proof: We prove this result along the lines of similar work in [25]. 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 (see [24]). 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 the $2k\lfloor \log_k N \rfloor$ potential q-values must isolate x_j from the other at most $k-1$ spike supports in one of its $\mathcal{I}_q$-subsets. □

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 the 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 the overall desired success probability p. 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

$$\frac{1-p}{2 \cdot q_{\max} \cdot \log_2\left(\frac{2k}{1-p}\right) \cdot \log_2 N}. \tag{16}$$

The end result will be that we correctly locate all spikes at least once with probability at least p. We can then estimate each located spike's magnitude using Lemma 2.

To finish recovering all spikes, 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 3) 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 arbitrarily high probability p.[4] See Algorithm 2 for multiple spike recovery pseudocode.

B. Bounding the Required Measurements

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 the 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 the $O\left(q_{\max} \cdot \ln\left(\frac{2k}{1-p}\right)\right)$ $\mathcal{I}_{q,h}$-subsets with the required

[3] In fact, using $\Omega(k^2 \ln N)$ of these disjount unions guarantees separation of all k spikes from one another deterministically (i.e., with probability 1).

[4] In particular, we may allow p to depend on N. For the purposes of proving Theorem 2 we let $p = 1 - 1/\log N$.

Algorithm 2 No More Than k Deltas

1: **Input: Maximum number of spikes k, Position tolerance N, magnitude tolerance α, smallest spike magnitude of interest $C_{\min}$, success probability p, total measurement budget B**
2: **Output: Estimates of magnitudes $> \frac{1}{2}C_{\min}$, $\{C_1, \dots, C_k\}$, and their positions, $\{x_1, \dots, x_k\}$**
3: *Find all spikes at least once...*
4: $SPIKES \longleftarrow \emptyset$
5: **for** $j = 1,\ j < P = O\left(\ln\left(\frac{k}{1-p}\right)\right),\ j{+}{+}$ **do**
6: $q \longleftarrow$ Randomly select one of the $2k\lfloor \log_k N \rfloor$ primes $> k$
7: Form $\mathcal{I}_q$ (see Equation 15)
8: **for each** $\mathcal{I}_{q,h} \in \mathcal{I}_q$ **do**
9: $(\tilde{x}, \tilde{C}) \longleftarrow$ Algorithm 1 with input ($\mathcal{I}_{q,h}$, $2N$, α, $O(1 - \text{Equation 16})$, $O(B/qP)$, γ from Lemma 2)
10: **if** $|\tilde{C}| > \frac{1}{2}C_{\min}$ **then**
11: $SPIKES \longleftarrow SPIKES \cup \{(\tilde{x}, \tilde{C})\}$
12: **end if**
13: **end for**
14: **end for**
15: *Remove excess spike approximations...*
16: $\{(\tilde{x}_0, \tilde{C}_0), (\tilde{x}_1, \tilde{C}_1) \dots\} \longleftarrow$ Sort $SPIKES$ by $\tilde{x}$'s
17: **for** $n = 0,\ n < |SPIKES|,\ n{+}{+}$ **do**
18: **while** $|x_n - x_{n+1}| \leq \frac{N}{2}$ **do**
19: $SPIKES \leftarrow SPIKES - \{(\tilde{x}_{n+1}, \tilde{C}_{n+1})\}$
20: **end while**
21: **end for**
22: Return $SPIKES$

Algorithm 1 success probability guarantee from Equation 16, we can see that we will need

$$O\left(\left(q_{\max} \cdot \ln N + \frac{\sigma^2_{[0,1]}}{C^2_{\min}}\right) \cdot \ln^2\left(\frac{q_{\max} \cdot \ln^2\left(\frac{N}{1-p}\right)}{1-p}\right)\right)$$

measurements whenever $\alpha = \Omega\left(1/\sqrt{N}\right)$.

To finish, we bound $q_{\max}$ in terms of k and N. Using results from [26] it is not difficult to prove that the $2k\lfloor \log_k N \rfloor^{\text{th}}$ prime larger than k is itself at most the

$$\left[2k \log_k N \cdot \left(2 + \frac{1.2762}{2 \ln k \ln N} + \frac{1}{2k \log_k N}\right)\right]^{\text{th}}$$

prime number. Therefore, we can see that $q_{\max}$ is at most the $(5k \cdot \log_k N)^{\text{th}}$ prime for all $N \geq k \geq 3$. Appealing again to results from [26] we can see that

$$q_{\max} \leq 10k \cdot \log_k N \log_2(5k \cdot \log_k N). \tag{17}$$

In fact, this bound is fairly pessimistic (especially for large k and N). However, it is good enough to assert that we need no more than

$$O\left[\left(k \cdot \ln^2 N \cdot \log_k(k \log_k N) + \frac{\sigma^2_{[0,1]}}{C^2_{\min}} + \frac{\sigma^2_{[0,1]}}{N \cdot \alpha^2 C^2_{\min}}\right) \cdot \ln^2\left(\frac{k \cdot \ln^2\left(\frac{k \ln N}{1-p}\right) \cdot \ln^2 N}{1-p}\right)\right] \tag{18}$$

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, $C_1 \cdot \delta(x - x_1), \dots, C_k \cdot \delta(x - x_k)$, in $[0,1]$. Then Algorithm 2 will, with probability at least p, output a $(\tilde{x}_j, \tilde{C}_j)$-pair for every spike 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 18.*

C. Discussion

It is worth mentioning that the measurement bounds for Theorem 4 can be improved slightly by using adaptive group testing methods from [27] together with Algorithm 1. However, the related constructions are less straightforward and have universal separation guarantees which yield unnecessarily complicated theorems in the presence of measurement noise. Furthermore, a careful comparison of these bounds reveals that Equation 18's work is only slightly improved, if at all, in the noisy case (i.e., when $\sigma^2_{[0,1]}$ is relatively large). In the no noise case (i.e., when $\sigma^2_{[0,1]} = 0,\ p$ fixed) Equation 18 is improved by roughly a $O\left(\ln\left(\frac{k \ln N}{1-p}\right)\right)$ factor. However, we are primarily interested in sublinear-time noisy sparse recovery here. Hence, we have focused on deriving the best achievable bounds with respect to the $\frac{\sigma^2_{[0,1]}}{C^2_{\min}}$-term using the simplest possible methods.

Note that the $\mathcal{I}_q$-sets (see Equation 15) can be created non-adaptively before any measurements are taken. As presented here, Algorithm 1 requires the fast adaptive bisection of its initial input subset. Assuming that both bisecting and measuring intervals can be done at unit cost, Algorithm 2 runs in $O\left(\left(\frac{\sigma^2_{[0,1]}}{C^2_{\min}} + \frac{\sigma^2_{[0,1]}}{N \cdot \alpha^2 C^2_{\min}} + k\right) \cdot \ln^{O(1)}\left(\frac{N}{1-p}\right)\right)$-time. Hence, the required runtime is sublinear in N for reasonable noise levels.

We finish the present discussion by noting that Algorithm 2 can be improved in several respects in practice. First, if we adaptively select and create the $\mathcal{I}_q$-subsets with smaller q-values as more spikes are discovered, we should be able to reduce the measurement costs of Algorithm 2 significantly on average. Furthermore, if measurements are also fast to construct on the fly, adaptive creation of the $\mathcal{I}_q$-subsets should also decrease the runtime in practice. Lastly, Algorithm 2 as presented here is highly parallel. That is, each line 9 application of Algorithm 1 can effectively be carried out independently of all the others at the same scale. In practice far fewer measurements will be utilized overall if the subsets of $[0,1]$ that are identified as most likely containing no spikes after the i^{th} application of Algorithm 1 in line 9 are subsequently removed from consideration in all $\mathcal{I}_{q,h}$ subsets thereafter. Similarly, removing identified spike locations from consideration as they are discovered should also reduce measurement usage in practice.

D. Special Case: Gaussian Measurement Noise

In this section we derive explicit upper bounds for the number of measurements required in order to recover all k spikes in the special case of Gaussian measurement noise.

Appealing to Section II-A we can see that the number of required Algorithm 1 measurements for each $\mathcal{I}_{q,h}$-subset is always bounded above by

$$\tilde{A} := \frac{2 \cdot \ln\left(\frac{1}{1-\tilde{p}}\right)}{\text{erf}^{\,2}\left(\frac{\sqrt{k} \cdot C_{\min}}{2\sigma}\right)} + \sum_{n=1}^{\lceil \log_2(N/k) \rceil} \frac{4 \cdot \ln\left(\frac{1}{1-\tilde{p}}\right)}{\text{erf}^{\,2}\left(\frac{\left(\sqrt{2}\right)^n \cdot \sqrt{k} \cdot C_{\min}}{2\sigma_{[0,1]}}\right)} \tag{19}$$

when the q-primes are chosen as per Lemma 3. All that remains for us to do is to bound (*i*) the number of $\mathcal{I}_{q,h}$-subsets to which we will have to apply Algorithm 1, and (*ii*) the probability $\tilde{p}$ with which each Algorithm 1 binary search decision must succeed.

Let q_j be the j^{th} prime number. Thus, $q_1 = 2, q_2 = 3, q_3 = 5, \ldots$ Let M be such that $q_M = q_{\max}$. From Section III-B we know that M is bounded by $5k \cdot \log_k N$ for all $N \geq k \geq 3$. Assuming that we choose $\left\lceil \log_2\left(\frac{2k}{1-p}\right) \right\rceil$ primes uniformly without replacement as per Section III, we can see that the total number of $\mathcal{I}_{q,h}$-subsets to which we must apply Algorithm 1 is bounded by

$$\begin{aligned} S_1 &:= \sum_{j=M-\left\lfloor \log_2\left(\frac{2k}{1-p}\right) \right\rfloor}^{M} q_j \\ &< 10k \log_2\left(\frac{4k}{1-p}\right) \log_k N \log_2(5k \log_k N). \end{aligned} \tag{20}$$

Similarly, the total number of required binary search decisions over the course of these Algorithm 1 executions is bounded above by

$$\begin{aligned} S_2 &:= \sum_{j=M-\left\lfloor \log_2\left(\frac{2k}{1-p}\right) \right\rfloor}^{M} q_j \cdot \log_2(2N/q_j) \\ &< S_1 \cdot \log_2\left(\frac{2N}{k}\right). \end{aligned} \tag{21}$$

Hence, in order to guarantee that an incorrect binary search decision is made with probability at most $\frac{1-p}{2}$, it suffices to set $1 - \tilde{p} = \frac{1-p}{2 \cdot S_2}$ in Equation 19. Combining Equations 19, 20, and 21 brings us to the following Corollary to Theorem 4.

Corollary 2. *Let σ^2 be the variance of mean μ Gaussian measurement noise over the interval $[0, 1]$. Fix $\alpha \in (0, \frac{1}{2})$, $p \in (0, 1)$, and $C_{\min} \in \mathbb{R}^+$. 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, at most $S_1 \cdot \tilde{A}$ measurements are required in order to locate all spikes with magnitude at least $C_{\min}$ to within $\frac{1}{N}$-tolerance with probability at least p. Furthermore, the number of required measurements is bounded above by Equation 18.*

IV. Conclusion

The adaptive algorithm (i.e., Algorithm 2) described and analyzed throughout the majority of this paper is only one of many potential recovery methods that can be created by combining combinatorial group testing constructions (e.g., see [18], [27]) with signal estimation and detection methods (e.g., see [19], [22]). More specifically, any k-disjunct group testing matrix (see [18]) will be guaranteed to isolate the k nonzero signal components of f from one another. Thus, they can be used to segment the search space (i.e., $[0, 1]$) into smaller regions each containing only one signal component or target. Signal detection and estimation methods can then be used to search each of these smaller regions for a single isolated signal component.

As previously mentioned, this type of search and recovery scheme is very easy to parallelize since each disjoint region of the search space dictated by the group testing construction can be searched independently. This essentially follows from the fact that the group testing methods we have considered here 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. In practice the total number of utilized measurements can probably be reduced further by adaptively partitioning the search space into smaller regions.

Another potential path toward improving the search space partitioning scheme utilized herein would be to adapt the techniques recently proposed in [8] to our adaptive measurement model (i.e., see Equation 1). In particular, the method utilized in [8] to extend their 1-sparse recovery technique to a k-sparse recovery scheme might allow a modest reduction in Equation 18 (e.g., by a multiplicative logarithmic factor). This would improve the results reported herein concerning noisy recovery. However, we will leave more careful consideration of such improvements and modifications to future work.

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] E. J. Candès and M. A. Davenport, "How well can we estimate a sparse vector?" *CoRR*, vol. abs/1104.5246, 2011.

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

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

[20] E. Arias-Castro, E. J. Candes, and M. Davenport, "On the Fundamental Limits of Adaptive Sensing," *ArXiv e-prints*, Nov. 2011.

[21] R. Motwani and P. Raghavan, *Randomized Algorithms*. Cambridge University Press, 1995.

[22] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. New York, NY, USA: Cambridge University Press, 2005.

[23] 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.

[24] I. Niven, H. S. Zuckerman, and H. L. Montgomery, *An Introduction to The Theory of Numbers*. John Wiley & Sons, Inc., 1991.

[25] G. Cormode and S. Muthukrishnan, "Combinatorial Algorithms for Compressed Sensing," *Conference on Information Sciences and Systems*, March 2006.

[26] P. Dusart, "The k^{th} prime is greater than $k(\ln k + \ln\ln k - 1)$ for $k \geq 2$," *Mathematics of Computation*, vol. 68, no. 225, 1999.

[27] 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.

Appendix

It suffices to consider the discrete case where 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

$$\langle \mathcal{M}_j, f + \mathcal{P}_j \rangle = \langle \mathcal{M}_j, f \rangle + \langle \mathcal{M}_j, \mathcal{P}_j \rangle. \tag{22}$$

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

$$\mathbb{E}\left[D(\mathbb{P}_i \| \mathbb{P}_{i'})\right] \leq \left(\frac{C_{\min}^2}{\sigma^2}\right)\left(k - \left|U[i] \cap U[i']\right|\right) \cdot \frac{m}{1 - \frac{2k-2}{N}}. \tag{23}$$

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 23. 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 above 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.