ADAPTIVE GROUP TESTING STRATEGIES FOR TARGET DETECTION AND LOCALIZATION IN NOISY ENVIRONMENTS

By

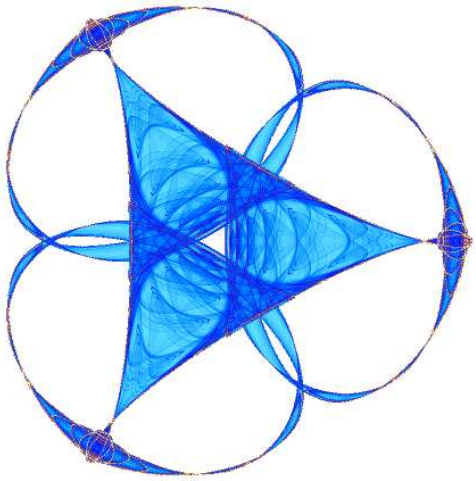
M.A. Iwen

and

A.H. Tewfik

IMA Preprint Series # 2311

(June 2010)



INSTITUTE FOR MATHEMATICS AND ITS APPLICATIONS

UNIVERSITY OF MINNESOTA

400 Lind Hall
207 Church Street S.E.
Minneapolis, Minnesota 55455–0436

Phone: 612-624-6066 Fax: 612-626-7370
URL: http://www.ima.umn.edu

Adaptive Group Testing Strategies for Target Detection and Localization in Noisy Environments*

M. A. Iwen
Institute for Mathematics and its Applications (IMA)
University of Minnesota
iwen@ima.umn.edu

A. H. Tewfik
Dept. of Electrical and Computer Engineering
University of Minnesota
tewfik@umn.edu

June 4, 2010

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. As opposed to previous studies, this paper models observations which are subject to the type of 'clutter noise' encountered in radar applications (i.e., the measurements used influence the observed noise). Given this model, the paper develops bounds on the number of measurements required to reconstruct the support of the signal and the signal itself up to any given accuracy level when the measurement noise is Gaussian using non-adaptive and adaptive measurement strategies. Further, the paper demonstrates that group testing measurement constructions may be combined with statistical binary detection and estimation methods to produce practical and computationally efficient adaptive algorithms for sparse signal approximation and support recovery.

In particular, the paper proves that a wide class of sparse signals can be recovered by adaptive methods using fewer noisy linear measurements than required by any recovery method based on non-adaptive Gaussian measurement ensembles. This result demonstrates an improvement over previous non-adaptive methods in the compressed sensing literature for sparse support pattern recovery in the sublinear-sparse support regime under the measurement model considered herein.

1 Introduction

In this paper, we study the problem of recovering a signal with a sparse representation of unknown structure from noisy observations. Specifically, we study *adaptive* acquisition strategies for estimating a signal f that admits a sparse representation in terms of a linear combination of k unknown elements from a set of N orthonormal functions ϕ_i, $1 \leq i \leq N$. The problem has received renewed attention in the emerging field of compressed sensing. It has also been the subject of many investigations in related areas, including radar imaging, medical imaging, model selection and signal classification in statistics and machine learning, and adaptive signal denoising. For example, a radar equipped surveillance plane might be required to locate a target within a specified search area to a desired accuracy level in a limited amount of time. Similarly, a smart missile may be required to locate a target under energy and time constraints. In testing, the target is a fault which is required to be localized to within a module or a component in a limited amount of time and using a pre-specified set of tests. In imaging applications, the goal is to reconstruct a high resolution image,

*This paper generalizes and improves on results first reported in (1).

e.g., a 3D magnetic resonance or coherent tomography image, in as short time as possible using a minimal amount of radiation. Solving such problems is equivalent to designing an optimal search strategy that may need to satisfy certain constraints.

More specifically, we are interested here in reconstructing the projection of an observable signal, f, onto a finite dimensional function space. Let $\Phi = \{\phi_i \mid 1 \leq i \leq N\}$ be a set of real valued orthonormal functions on $[0, 1]$ that span the given function space of interest. We will denote f as

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

Note that the true observable signal may have a component outside of Φ. However, we are only interested in its projection, f, onto Φ and assume that f has a sparse representation in Φ. Given this sparsity assumption it makes sense to define the support of f to be the positions where the coefficients $\vec{f_i}$ are nonzero (or otherwise larger in magnitude than a pre-specified application dependent threshold δ_*). The support of f, or $\vec{f} \in \mathbb{R}^N$, is thus denoted

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

Note that in order to recover f we must identify $\text{supp}(f)$. Furthermore, once we have identified the support of f it is straightforward to approximate its nonzero coefficients. Thus, the primarily focus of this paper – signal recovery – is integrally linked to support recovery.

The problem of allocating search effort has been studied as early as World War II (e.g., see (2),(3)). The problem was revisited in the context of fault diagnostics, e.g., (4), and radar imaging, e.g., (5), (6), (7), (8), where the emphasis has been on rapidly locating a single target using an optimal beamforming and target illumination strategy, designing adaptive waveform for optimal target detection and identification, or achieving the best resolution in time constrained radar imaging. In the compressed sensing theory, several papers addressed the problem of reconstructing a signal f with a sparse support as defined above using quadratic programming, e.g., (9), (10), (11), (12), and in particular several papers investigate the asymptotic performance of the lasso approach (e.g., see (13), (14)). Information theoretic bounds have also been developed for the asymptotic performance of support recovery and partial support recovery algorithms (e.g. see (15), (16) and references therein). Bayesian approaches for estimating the support of a signal have also been proposed for compressive sensing with adaptive measurements (e.g., see (17), (18)). These Bayesian methods have been demonstrated to work well empirically, often requiring fewer noisy measurements to recover sparse signals than their non-adaptive competitors in practice. Note also that the "compressive distilled sensing" techniques (19) demonstrate that adaptive methods can improve error bounds for sparse recovery problems under a different measurement model. In contrast, we discuss and develop a more general class of adaptive measurement procedures for a different measurement model and then rigorously prove that they indeed reliably recover the support of sparse signals using fewer noisy linear measurements than any possible approach utilizing non-adaptive Gaussian measurement matrices. We accomplish this goal by designing adaptive sparse support recovery procedures with sufficient measurement conditions that are asymptotically smaller than the necessary conditions developed for non-adaptive approaches utilizing Gaussian measurement matrices below.

The most general form of the support recovery problem that we consider here can be formulated as follows. We are allowed to design and adaptively select 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. Corresponding to each such test function $\mathcal{M}_j$, we can generate noisy observations of the form

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

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 $j \neq j'$. We will refer to the measurement model in the Equation 1 as *non-adaptive* if the generation of the j^{th} measurement, y_j, is independent of all previous noisy observations, y_n, $1 \leq n \leq j - 1$. In effect, a set of measurements $\mathcal{M} = \{\mathcal{M}_1,\, \ldots,\, \mathcal{M}_m\}$ is non-adaptive if it can be wholly instantiated before any measurements are actually taken. If, on the other hand, any single measurement may depend on the results of previous measurements 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 compressed sensing support recovery methods in the noisy sublinear-sparse support regime. In this context the work herein is a generalization of previous work which considered group testing methods for sparse signal recovery using noiseless measurements (see Section 5 of (20) and references therein).

Our goal is to determine the minimal number m of measurements we can take to recover $f : [0,1] \longrightarrow \mathbb{R}$ up to an accuracy level at least precise enough to guarantee we recover its correct support in Φ. To accomplish this goal we construct an algorithm for designing and adaptively selecting the m measurement functions $\mathcal{M}_j$, $1 \leq j \leq m$. To simplify the presentation of the results we will, without loss of generality, restrict our attention to a specific instance of the problem that arises in multi-target localization in radar. We conclude the paper with a discussion that explains how the results and algorithms developed here also apply to the more general problem.

Specifically, we will derive our results in the context of the following problem. 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 $\delta(x)$ 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 of the range of the arbitrary positions x_j, which we normalized without loss of generality to the interval $[0,1]$. The model also captures radar imaging of targets that consist of a collection of point reflectors, a target model that is often used in the literature. Selecting the measurement functions $\mathcal{M}_j$ corresponds to selecting a radar beamform and illumination pattern.

Without loss of generality as we explain below, we restrict our attention in the sequel to a specific class of test functions. The test functions we consider are box or indicator functions that yield a measurement model consisting of noisy integral measurements of the signal f at each time $t \in \mathbb{R}^+$. 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 can measure

$$m_{\mathcal{I}}(t) = \int \mathbb{I}_{\mathcal{I}} \cdot f \, dx + \int \mathbb{I}_{\mathcal{I}} \, d\mathcal{P}(t), \tag{3}$$

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

$$\int \mathbb{I}_{\mathcal{I}} \, d\mathcal{P}(t) \text{ and } \int \mathbb{I}_{\mathcal{J}} \, d\mathcal{P}(t)$$

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

$$\int \mathbb{I}_{\mathcal{I}} \, dx = \int \mathbb{I}_{\mathcal{J}} \, dx. \tag{4}$$

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

$$\int \mathbb{I}_{\mathcal{I}} \, d\mathcal{P}(t_1) \text{ and } \int \mathbb{I}_{\mathcal{J}} \, d\mathcal{P}(t_2)$$

are i.i.d. as long as $\mathcal{I}, \mathcal{J} \subseteq [0,1]$ satisfy Equation 4. We define $\sigma^2_{\mathcal{I}}$ to be the variance of $m_{\mathcal{I}}(t)$ for a given $\mathcal{I} \subseteq [0,1]$. Given the assumptions above, we can see that $\sigma^2_{\mathcal{I}} = \sigma^2_{\mathcal{J}}$ whenever $\mathcal{I}, \mathcal{J} \subseteq [0,1]$ satisfy Equation 4. Finally, we will denote the variance of the noise over the entire unit interval by $\sigma^2 = \sigma^2_{[0,1]}$.

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

$$\frac{1}{N} < \min\left\{|x_j - x_l| \;\middle|\; j \in [1,k] \cap \mathbb{Z},\ l \in \mathbb{Z} \cap [1,k] - \{j\}\right\}. \tag{5}$$

In other words, N gives a guaranteed separating distance between the Dirac delta functions composing f. The separating distance assumption ultimately allows us to recognize the measurements defined in Equation 3 as a specialized form of the more general measurement model described by Equation 1. For example, consider the following two related problems:

1. **Discrete Setting:** Given that we will be satisfied to locate each x_j to within a $\frac{1}{N}$-tolerance, we can simply model f as a k-sparse vector $\vec{f} \in \mathbb{R}^N$ whose n^{th} element is given by

$$\vec{f}_n = \begin{cases} C_j & \text{if } x_j \in [(n-1)/N,\ n/N) \text{ for } n \in [1, N-1] \cap \mathbb{N} \\ C_j & \text{if } x_j \in [(N-1)/N,\ 1] \\ 0 & \text{otherwise} \end{cases}. \tag{6}$$

   In this setting each indicator function based measurement in Equation 3 used by our adaptive approach will directly correspond to a discrete inner product measurement of a binary measurement vector, $\mathcal{M}_j \in \{0,1\}^N$, with $\vec{f}$.[1] Choose the measurement noise process in Equation 1, $\mathcal{P}_j$, to be a random vector in $\mathbb{R}^N$ with i.i.d. entries each having variance σ^2/N. The total variance of a measurement over the entire search space will be the given σ^2 value so that stochastic measurement noise behaves as expected above. Under these conditions, the discrete Equation 1 measurements of the vector $\vec{f}$ from Equation 6 will be essentially indistinguishable from the corresponding indicator function based measurements in Equation 3 of the function f from Equation 2. Hence, similar adaptive recovery algorithms may be utilized in either case.

   In the paragraph above we made the assumption that the $\vec{f} \in \mathbb{R}^N$ defined in Equation 6 was sparse in the standard basis. However, standard methods allow us to complete recovery if $\vec{f}$ is instead sparse under a different orthonormal basis. Suppose, in this discrete setting, that Φ is a known $N \times N$ orthonormal matrix for which $\Phi\vec{f}$ is k-sparse. Instead of gathering the discrete Equation 1 measurements discussed in the preceding paragraph, we will slightly alter them by gathering

$$y_j = \langle \Phi^{\mathrm{T}} \mathcal{M}_j,\ \vec{f} + \mathcal{P}_j \rangle = \langle \mathcal{M}_j,\ \Phi\vec{f} \rangle + \langle \mathcal{M}_j,\ \Phi\mathcal{P}_j \rangle.$$

   Note that the only substantive change to our measurements of the sparse vector of interest – now $\Phi\vec{f}$ instead of f – is that the additive noise term is now determined by $\Phi\mathcal{P}_j \in \mathbb{R}^N$ instead of just $\mathcal{P}_j \in \mathbb{R}^N$. However, if each random entry of $\mathcal{P}_j$ has mean 0, or more generally has a known mean, then this is not problematic. The fact that each entry of $\mathcal{P}_j$ is i.i.d. with mean 0 and variance σ^2/N will conspire with the orthonormality of Φ to ensure that all the entries of $\Phi\mathcal{P}_j$ are still uncorrelated with mean 0 and variance σ^2/N. These noise characteristics are sufficient for the proofs of all the adaptive results below.

2. **The Characteristic Function Basis:** For the case where $f : [0,1] \to \mathbb{R}$ is a true function, the measurements defined in Equation 3 correspond most naturally with Equation 1 type measurements performed over functions approximated with respect to the orthonormal characteristic functions $\Phi = \{\phi_i \mid 1 \le i \le N\}$ on $[0,1]$ defined by

$$\phi_i = \sqrt{N} \cdot \mathbb{I}_{[(i-1)/N,\ i/N)}$$

   for $i \in [1,N) \cap \mathbb{N}$, and $\phi_N = \sqrt{N} \cdot \mathbb{I}_{[(N-1)/N,\ 1]}$. Thus, Φ spans the space of piecewise constant functions over the N subintervals of equal length in $[0,1]$. With respect to this Φ, f from Equation 2 is represented by

$$f(x) = \sum_{i=1}^{N} \vec{f}_i \cdot \phi_i(x),$$

[1] In fact, it may also be necessary to occasionally measure a single additional $\vec{f}$ entry in order to guarantee that the discrete analogs of the left/right subsets defined in Section 3 below both always have an equal number of entries. However, this does not alter the statement of any results proven herein.

where $\vec{f}$ is defined as in Equation 6. In this setting each indicator function based measurement in Equation 3 must be reweighted by a factor of $\sqrt{N}$ when utilized with the more general measurement definition of Equation 1 (i.e., we are working with very tall and thin characteristic functions which 'approximate delta functions'). Hence, each Equation 1 type measurement of this piecewise constant f will use a test function of the form $\mathcal{M}_j(x) = \sqrt{N} \cdot \mathbb{I}_{\mathcal{I}}(x)$, where each set $\mathcal{I}$ is a subset of subintervals dictated by our adaptive recovery methods developed below.[2] Finally, the stochastic measurement noise will behave as assumed in Equation 3 if we set the noise process from Equation 1, $\mathcal{P}_j$, to be the rescaled noise term from Equation 3, $d\mathcal{P}(j)/\sqrt{N}$. Using these measurement definitions it is not difficult to see that both Equations 1 and 3 will again produce essentially indistinguishable measurement results for their corresponding functions.

Of course, $f : [0,1] \longrightarrow \mathbb{R}$ need not be sparse with respect to a basis of orthonormal characteristic functions in order to be recoverable. Suppose f is sparse with respect to an arbitrary orthonormal set of N functions, $\Phi = \{\phi_i \mid 1 \leq i \leq N\}$, on $[0,1]$. In this general case we can still pursue recovery along the same lines as in the previous discrete example. Let $\mathcal{M}_j \in \{0,1\}^N$ be the j^{th} binary measurement vector resulting from the application of our adaptive recovery techniques in the discrete setting (i.e., see the first paragraph of the first example above). We simply use this binary measurement vector to define our j^{th} test function for Equation 1, $\mathcal{M}'_j : [0,1] \longrightarrow \mathbb{R}$, as

$$\mathcal{M}'_j(x) = \sum_{i=1}^{N} \mathcal{M}_{j,i} \cdot \phi_i(x).$$

This will produce the required Equation 1 measurement results anytime the measurement noise process, $\mathcal{P}_j$, has both of the following properties: (i) $\langle \phi_i, \mathcal{P}_j \rangle$ has σ^2/N variance and a fixed mean for all $1 \leq i \leq N$, and (ii) $\langle \phi_i, \mathcal{P}_j \rangle$ and $\langle \phi_{i'}, \mathcal{P}_j \rangle$ are uncorrelated for all $i \neq i'$. This will certainly be the case if, for example, $\mathcal{P}_j$ is defined to be an independent incarnation of weighted Gaussian white noise, $\mathcal{P}_j = \frac{\sigma}{\sqrt{N}} \cdot \frac{dW}{dt}$, for each j^{th} measurement (see (21) for a rigorous introductory treatment of Gaussian white noise, dW/dt).

Using this model and goals, we make several contributions to the problem of recovering a signal with a sparse representation of unknown structure from noisy observations. First, we demonstrate that adaptive group testing measurement constructions may be combined with statistical binary detection and estimation methods to produce efficient adaptive sequential algorithms for sparse signal support recovery. More specifically, we construct new results contrasting the number of noisy non-adaptive Gaussian measurements versus noisy adaptive measurements required to recover a sparsely representable signal. In doing so we prove that a sufficiently sparse signal, f, can be recovered up to any desired accuracy level using fewer noisy adaptive linear measurements than required by any recovery method based on non-adaptive Gaussian measurement ensembles. These results ultimately improve on previous sufficient conditions for sparse support recovery in the noisy sublinear-sparse support regime. Finally, we also present explicit algorithms and techniques for achieving our adaptive recovery results, as well as discuss modifications likely to further improve the performance of such adaptive methods in practice.

The remainder of this paper is structured as follows. In Section 2 we present new results contrasting the recovery of the support of a signal f using noisy non-adaptive and adaptive Gaussian measurements contaminated with Gaussian noise. These results anchor the developments in the remaining portions of the paper and extend prior results in the literature. Then, in Section 3, we present and analyze a simple binary search procedure for recovering 1-sparse signals in noise. In Section 4 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 5.

[2] Similar to the last example, it may occasionally be necessary to adjust a measurement interval, $\mathcal{I}$, to avoid bisecting one of the fundamental $[(i-1)/N, i/N)$-intervals. When this happens a few additional measurements of a single fundamental interval may be necessary (i.e., to measure the contribution of a single basis function to f). However, this does not alter the statement of any results proven herein.

2 Results and Discussion

In many previous studies (e.g., see (13), (16), (15) and references therein) the measurements utilized for signal (support) recovery were of the form

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

where $w_j \sim \mathcal{N}(0,1)$ is independent Gaussian noise for each j, $f \in \mathbb{R}^N$ is a sparse vector (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. In practice, the 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, in radar, using a wider beam to cover a wider area will increase the observed clutter noise. This first noise component is captured by our model in equations (1) and (3) and is neglected in the measurement model above used in prior work.

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. Its effect can be reduced by using more sophisticated electronics, e.g., by cooling a detector. This noise component is neglected in our model and is the only component captured by prior works. Adding this noise component to our model does not change the conclusions drawn below, or proof techniques, but adds to the complexity of the derivations.

As mentioned above, 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, e.g., (22), (23), (9), measurement design (e.g., see (24)). Here, we will focus on the following result concerning support recovery using noisy non-adaptive Gaussian measurements contaminated with Gaussian noise. Our 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 our general measurement model (see Equation 1).

Theorem 1. *Suppose f is k-sparse when represented in terms of a known orthonormal basis $\Phi = \{\phi_i \mid 1 \le i \le N\}$. That is, $f = \sum_{i=1}^{N} \vec{f_i} \cdot \phi_i$ where $\vec{f} \in \mathbb{R}^N$ is k-sparse. Let $C_{\min} = \min\left\{ \left| \langle f,\ \phi_i \rangle \right| \ \middle|\ i \in \mathrm{supp}(f) \right\}$ be the magnitude of the smallest of the k non-zero coefficients of $\vec{f}$. Next, 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*

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

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 $\langle \mathcal{M}_j, \mathcal{P}_j \rangle \sim \mathcal{N}(0, \|\mathcal{M}_j\|_2^2 \cdot \sigma^2/N)$. 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 $\mathrm{supp}(f)$ 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 Appendix A. □

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 we utilize adaptive combinatorial group testing (25) methods to guarantee the recovery of highly sparse signals with asymptotically 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 (26) we obtain the following corollary of our main theorem for the special case of support identification of sparse vectors with Gaussian measurement noise (see Corollary 3).

Theorem 2. *Suppose f is k-sparse when represented in terms of a known orthonormal basis $\Phi = \{\phi_i \mid 1 \leq i \leq N\}$. That is, $f = \sum_{i=1}^{N} \vec{f_i} \cdot \phi_i$ where $\vec{f} \in \mathbb{R}^N$ is k-sparse. Let $C_{\min} = \min\left\{ \left| \langle f,\ \phi_i \rangle \right| \ \middle|\ i \in \operatorname{supp}(f) \right\}$ be the magnitude of the smallest of the k non-zero coefficients of $\vec{f}$. Furthermore, let $\mathcal{P}_j$ for $1 \leq j \leq m$ be m independent Gaussian measurement noise processes with mean 0 so that the accumulated noise for each Equation 1 measurement, conditioned on the adaptive test function $\mathcal{M}_j$, is $\langle \mathcal{M}_j, \mathcal{P}_j \rangle \ \sim\ \mathcal{N}(0, \|\mathcal{M}_j\|_2^2 \cdot \sigma^2/N)$. Finally, suppose that $\sigma^2/C_{\min}^2$ is $\Omega\left(k \cdot \ln^3 N\right)$.[3] 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 will recover f to a precision guaranteed to correctly determine $\operatorname{supp}(f)$ *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 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.

Intuitively, it should not be surprising that methods utilizing non-adaptive measurements are less effective under the observational model in Equation 1 than methods based on adaptive measurements. Every non-adaptive measurement must necessarily allocate significant amounts of sensing energy to a large fraction of the basis elements in Φ (i.e., to a large fraction of the entire search area). This essentially guarantees that every non-adaptive observation will be contaminated with a large fraction of the additive observational noise from the entire search area. Adaptive measurements, on the other hand, can eventually avoid additive 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 sensing energy toward regions likely to contain signal components.

3 Single Spike Targeting

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

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

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

Algorithm 1 ISOLATED DELTA

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
Output: Estimate of magnitude, C_1, and position, x_1
$\mathcal{I}^+ \longleftarrow \mathcal{I}$
$\mathcal{I}^- \longleftarrow \mathcal{I}$
Initialize $B' \leftarrow (B - 3\gamma)/4$

LOCATE x_1

while $\int \mathbb{I}_{\mathcal{I}^+}\, dx > \frac{1}{N}$ **do**
  Assuming C_1 is positive, find x_1...
  if $m_{\mathcal{I}_l^+}(t) > m_{\mathcal{I}_r^+}(t)$ the majority of B' trails (possibly averaged) **then**
    $\mathcal{I}^+ \longleftarrow \mathcal{I}_l^+$
  else
    $\mathcal{I}^+ \longleftarrow \mathcal{I}_r^+$
  end if
  Assuming C_1 is negative, find x_1...
  if $m_{\mathcal{I}_l^-}(t) < m_{\mathcal{I}_r^-}(t)$ the majority of B' trails (possibly averaged) **then**
    $\mathcal{I}^- \longleftarrow \mathcal{I}_l^-$
  else
    $\mathcal{I}^- \longleftarrow \mathcal{I}_r^-$
  end if
  $B' \leftarrow B'/2$
end while

ESTIMATE C_1

Choose arbitrary $\mathcal{I}^N \subset \mathcal{I} - (\mathcal{I}^- \cup \mathcal{I}^+)$ with $\int \mathbb{I}_{\mathcal{I}^N}\, dx = \int \mathbb{I}_{\mathcal{I}^+}\, dx$
$\tilde{C}^+ \longleftarrow$ Estimated mean from γ $m_{\mathcal{I}^+}(t.)$ measurements (α-precise)
$\tilde{C}^- \longleftarrow$ Estimated mean from γ $m_{\mathcal{I}^-}(t.)$ measurements (α-precise)
$\tilde{C}^N \longleftarrow$ Estimated mean from γ $m_{\mathcal{I}^N}(t.)$ measurements (α-precise)
Decide if C_1 is positive or negative...
if $\left|\tilde{C}^+ - \tilde{C}^N\right| > \left|\tilde{C}^- - \tilde{C}^N\right|$ **then**
  $\tilde{C}_1 \longleftarrow \tilde{C}^+ - \tilde{C}^N$
  $\mathcal{I}^e \longleftarrow \mathcal{I}^+$
else
  $\tilde{C}_1 \longleftarrow \tilde{C}^- - \tilde{C}^N$
  $\mathcal{I}^e \longleftarrow \mathcal{I}^-$
end if
Return $\tilde{x}_1 =$ midpoint of $\mathcal{I}^e$, $\tilde{C}_1$

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

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

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

To help make our final decision we arbitrarily chose an interval whose noise characteristics will be, by assumption, distributed identically to the additive noise in both the resulting positive/negative binary search intervals (i.e., $\mathcal{I}^+$ and $\mathcal{I}^-$). The resulting positive/negative interval containing x_1 should yield measurements with a mean that is different from the arbitrary interval measurements' mean. Hence, we estimate the measurement means of both intervals resulting from our binary searches, and then compare them to the mean of our arbitrary interval measurements. Whichever binary search result differs most from our arbitrary interval in terms of measurement mean will be the correct search result with high probability. See Algorithm 1 for pseudocode.

Lemma 1. *Using* $B - 3\gamma = O\left(\left(\frac{\sigma_{\mathcal{I}}^2}{C_1^2} + \log N\right) \cdot \log\left(\frac{\log N}{1-p}\right)\right)$ *total initial measurements in line 5 of Algorithm 1 is sufficient to allow lines 6 through 20 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 line 6's "while"-loop 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 $\tilde{t} \in \mathbb{R}^+$.

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

$$\mathbb{P}\left[m_{\mathcal{I}_c^+}(t) > m_{\mathcal{I}_w^+}(t)\right] = \frac{1}{2} + \int_{-\infty}^{\infty}\left(\int_{y-C_1}^{y} D_{\mathcal{I}_w^+}(x)\,dx\right) D_{\mathcal{I}_w^+}(y)\,dy = \frac{1}{2} + \epsilon_{\mathcal{I}_w^+}(C_1),$$

where $\epsilon_{\mathcal{I}_w^+}(C_1)$ is greater than 0. However, instead of comparing single measurements from each subset of $\mathcal{I}^+$, we will instead consider comparing several averaged measurements from $\mathcal{I}_c^+$ to several averaged measurements from $\mathcal{I}_c^+$. This will help us ensure that $\epsilon_{\mathcal{I}_w^+}(C_1)$ is effectively larger than any desired constant $c \in (0, \frac{1}{2})$.

Let $M_c^+ = \frac{1}{K} \cdot \sum_{k=1}^{K} m_{\mathcal{I}_c^+}(t_k)$ and $M_w^+ = \frac{1}{K} \cdot \sum_{k=1}^{K} m_{\mathcal{I}_w^+}(t_k)$. 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 $\mathbb{V}\mathrm{ar}\left[M_c^+ - M_w^+\right] = 2 \cdot \frac{\sigma_{\mathcal{I}_w^+}^2}{K} = \frac{\sigma_{\mathcal{I}^+}^2}{K}$. Chebyshev's Inequality then guarantees that

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

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

Applying the Chernoff bound we can see that $\frac{\ln\left(\frac{\log_2(N)}{1-p}\right)}{2 \cdot \epsilon_{\mathcal{I}_w^+}^2(C_1)}$ (averaged) measurement comparisons are sufficient to correctly decide $\mathcal{I}_c^+$ with error probability at most $\frac{1-p}{\log_2(N)}$ (see (27)). The union bound then implies that all $O(\log(N))$ iterations of line 6'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 6's loop we will have $\sigma^2_{\mathcal{I}^+} = \frac{\sigma^2_{\mathcal{I}}}{2^n}$. Hence, at the n^{th} iteration of line 6 we will require no more than

$$B' = O\left((K+1)\cdot\log\left(\frac{\log_2(N)}{1-p}\right)\right) = O\left(\left(1+\frac{\sigma^2_{\mathcal{I}}}{2^n\cdot C_1^2}\right)\cdot\log\left(\frac{\log_2(N)}{1-p}\right)\right)$$

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

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

measurements. The result follows. □

To finish the analysis of Algorithm 1 we address its estimation portion (lines 21 through 32). 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 3 different intervals of the same size. 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. 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, our mean estimates 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 21 through 32) 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 $\gamma = O\left(\left(1+\frac{\sigma^2_{\mathcal{I}}}{N\cdot\alpha^2 C_1^2}\right)\cdot\log\frac{1}{1-p}\right)$.*

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 (27)). 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 correctly locate x_1 with probability at least p.

Continuing, we relate $\mathbb{V}\mathrm{ar}\left[m_{\mathcal{I}^+}(t.)\right]$ to $\sigma^2_{\mathcal{I}}$. We have that

$$\sigma^2_{\mathcal{I}} = \mathbb{V}\mathrm{ar}\left[\sum_{j=1}^N \int_{x_j}^{x_{j+1}} \mathbb{I}_{\mathcal{I}}\,dN(t) + \int_{x_{N+1}}^1 \mathbb{I}_{\mathcal{I}}\,dN(t)\right],$$

where the points $x_1 < x_2 < \cdots < x_{N+1} \in [0,1]$ are chosen so that $\int_{x_j}^{x_{j+1}} \mathbb{I}_{\mathcal{I}}\,dx = m$ for all $j \in [1,N]\cap\mathbb{Z}$. Using our assumptions from Section 2 we have that

$$\sigma^2_{\mathcal{I}} \geq \sum_{j=1}^N \mathbb{V}\mathrm{ar}\left[\int \mathbb{I}_{\mathcal{I}\cap[x_j,x_{j+1})}\,dN(t.)\right] \geq (N-4)\mathbb{V}\mathrm{ar}\left[m_{\mathcal{I}^+}(t.)\right].$$

Finally, we note that 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 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 $\sigma_{\mathcal{I}}^2 = \mathbb{V}\mathrm{ar}[m_{\mathcal{I}}(t)]$. 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 $O\left(\left(\log N + \frac{\sigma_{\mathcal{I}}^2}{C_1^2} + \frac{\sigma_{\mathcal{I}}^2}{N \cdot a^2 C_1^2}\right) \cdot \log\left(\frac{\log N}{1-p}\right)\right)$.*

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. Alternatively, it may be proven for a modified variant of Algorithm 1 by repeatedly applying Lemma 2 with fixed $\alpha \approx \frac{1}{2}$ during the binary search process (as opposed to using Lemma 1). If we take the majority of averages of $O\left(\sigma_{\mathcal{I}_{l/r}^+}^2 \Big/ C_1^2\right)$-measurements in Algorithm 1's lines 8 and 14 instead of simply taking the majority of individual measurements, we can see that the number of measurements required to decide on the correct subinterval containing x_1 will scale linearly in $\frac{\sigma_{\mathcal{I}_{l/r}^+}^2}{C_1^2}$ (see the proof of Lemma 1). The end result is that Theorem 3 holds in the presence of arbitrary (e.g., adversarial) measurement noise. However, as one might expect, the performance of Algorithm 1 can be improved when the characteristics of the measurement noise are better known.

3.1 Gaussian White Measurement Noise

We now consider Lemma 1 in the context of Gaussian white measurement noise with *unknown* mean μ. Our goal in this section is to derive concrete measurement bounds for this special case. 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 6's "while"-loop. Call the other subset $\mathcal{I}_w^+$ and suppose that $m_{\mathcal{I}_w^+}(t) \sim \mathcal{N}\left(\mu, \sigma_{\mathcal{I}_w^+}^2\right)$. 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

$$\epsilon_{\mathcal{I}_w^+}(C_1) = \mathbb{P}\left[m_{\mathcal{I}_c^+}(t) > m_{\mathcal{I}_w^+}(t)\right] - \frac{1}{2} = \frac{1}{2} - \mathbb{P}\left[(m_{\mathcal{I}_c^+}(t) - m_{\mathcal{I}_w^+}(t)) < 0\right]. \tag{9}$$

In the Gaussian white noise case we know that $(m_{\mathcal{I}_c^+}(t) - m_{\mathcal{I}_w^+}(t)) \sim \mathcal{N}\left(C_1, 2\sigma_{\mathcal{I}_w^+}^2\right)$ (see (28)). Therefore, we know that

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

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 our spike magnitude is less than 2 standard deviations). We have

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

Substituting this expression into the Chernoff bound we can see that

$$\frac{\ln\left(\frac{1}{1-p}\right)}{\text{erf}^{\,2}\left(\frac{C_1}{2\sigma_{\mathcal{I}_w^+}}\right)} < \max\left\{\frac{36\pi\cdot\sigma^2_{\mathcal{I}_w^+}\cdot\ln\left(\frac{1}{1-p}\right)}{\left(4-\exp\left(\frac{C_1^2}{4\cdot\sigma^2_{\mathcal{I}_w^+}}\right)\right)^2\cdot C_1^2}, \frac{\ln\left(\frac{1}{1-p}\right)}{\text{erf}^{\,2}(1)}\right\} \tag{10}$$

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 10 for all but the first iteration of Algorithm 1's line 6 – 20 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 our mean μ Gaussian white measurement noise over the entire interval $[0,1]$. Then, we must utilize Equation 10 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

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

which in turn is bounded by

$$\max\left\{\frac{36\pi\cdot\sigma^2\cdot\ln\left(\frac{\log_2 2N}{1-p}\right)}{\left(4-\exp\left(\frac{C_1^2}{4\cdot\sigma^2}\right)\right)^2\cdot C_1^2}, \frac{\ln\left(\frac{\log_2 2N}{1-p}\right)}{\text{erf}^{\,2}(1)}\right\} + \sum_{n=1}^{\lceil\log_2 N\rceil-1} 2\cdot\max\left\{\frac{36\pi\cdot\sigma^2\cdot\ln\left(\frac{\log_2 2N}{1-p}\right)}{2^n\left(4-\exp\left(\min\left\{1,\frac{2^n\cdot C_1^2}{4\cdot\sigma^2}\right\}\right)\right)^2\cdot C_1^2}, \frac{\ln\left(\frac{\log_2 2N}{1-p}\right)}{\text{erf}^{\,2}(1)}\right\}.$$

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

Corollary 1. *Let σ^2 be the variance of our 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*

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

measurements. As $\sigma^2\to\infty$ the number of required measurements is bounded above by $12\pi\cdot\frac{\sigma^2}{C_1^2}\cdot\ln\left(\frac{\log_2 2N}{1-p}\right)$.

Proof:

The explicit measurement bounds are developed above. As $\sigma^2\to\infty$ we simply note that the number of required measurements will be less than

$$\frac{4\pi\cdot\sigma^2\cdot\ln\left(\frac{\log_2 2N}{1-p}\right)}{C_1^2} + \sum_{n=1}^{\lceil\log_2 N\rceil-1}\frac{4\pi\cdot\sigma^2\cdot\ln\left(\frac{\log_2 2N}{1-p}\right)}{2^{n-1}\cdot C_1^2} < \frac{4\pi\cdot\sigma^2\cdot\ln\left(\frac{\log_2 2N}{1-p}\right)}{C_1^2}\cdot\left(1+\sum_{n=0}^{\infty}\frac{1}{2^n}\right) = 12\pi\cdot\frac{\sigma^2}{C_1^2}\cdot\ln\left(\frac{\log_2 2N}{1-p}\right).$$

The result follows. □

It is worth noting that Corollary 1's number of required measurements can be calculated much more precisely if desired. In our work above the use of the Chernoff bound may always be replaced by the exact combinatorial sum expressing the probability of the simultaneous occurrence of more than half of a given number T independent Bernoulli events, each occurring with probability $\mathbb{P}\left[m_{\mathcal{I}_c^+}(t) > m_{\mathcal{I}_w^+}(t)\right]$. In the Gaussian white noise case we can

calculate $\mathbb{P}\left[m_{\mathcal{I}_c^+}(t) > m_{\mathcal{I}_w^+}(t)\right]$ to arbitrarily high precision at each stage of our binary search(es). Then, at each stage, the smallest T for the combinatorial sum yielding an acceptably high success rate provides the required number of measurements. To finish, we simply sum these $\log_2 N$ T-values to obtain a better upper bound on the total required number of measurements. We next consider an alternate bounding strategy for our measurements which does not utilize the Chernoff bound.

3.2 Gaussian Noise: Alternative Measurement Bounds

In Section 3.1 above we bounded the number of measurements required to employ Algorithm 1 in the special case of Gaussian noise. In doing so we appealed to the Chernoff bound to obtain Corollary 1. In this section we consider replacing lines 8 and 14 of Algorithm 1 with standard methods for signal detection and estimation in additive mean μ Gaussian noise (see, e.g., (26), (29)). By doing so we demonstrate that Algorithm 1 uses at most a constant factor more measurements than standard methods to locate a single spike in the noisy Gaussian case.

Let $m(p)$ be the smallest number of measurements required in order for Algorithm 1 to correctly decide $\mathcal{I}_c^+$, the right/left subset of $\mathcal{I}^+$ containing x_1, with probability at least $p \in (1/2, 1)$ as $\sigma_{\mathcal{I}_w^+} \to \infty$. Equation 10, derived using the Chernoff bound, tells us that

$$m(p) < 4\pi \cdot \frac{\sigma^2_{\mathcal{I}_w^+}}{C_1^2} \cdot \ln\left(\frac{1}{1-p}\right). \tag{11}$$

Alternatively, we can decide which subinterval of $\mathcal{I}^+$ contains x_1 by directly comparing the average of several measurements from the right subinterval of $\mathcal{I}^+$ to the average of several measurements from the left subinterval of $\mathcal{I}^+$. In effect, we perform a generalized likelihood ratio test by testing the maximum likelihood estimates of each subinterval's mean against one another. Assuming, without loss of generality, that we know the spike amplitude C_1 is positive, we simply choose the larger of the two maximum likelihood mean estimates.

Let $\tilde{m}(p)/2$ be the number of measurements averaged from each subinterval for our mean estimates. Then we will correctly decide $\mathcal{I}_c^+$ with probability

$$p = \mathbb{P}\left[\frac{2}{\tilde{m}(p)} \cdot \sum_{j=1}^{\tilde{m}(p)/2} m_{\mathcal{I}_c^+}(t_j) > \frac{2}{\tilde{m}(p)} \cdot \sum_{j=1}^{\tilde{m}(p)/2} m_{\mathcal{I}_w^+}(t_j)\right]. \tag{12}$$

Because our additive noise is Gaussian, we can see that

$$p = \frac{1}{2} \cdot \left(1 + \operatorname{erf}\left(-\frac{\sqrt{\tilde{m}(p)} \cdot C_1}{2\sqrt{2}\sigma_{\mathcal{I}_w^+}}\right)\right).$$

Hence, we have that

$$\tilde{m}(p) \leq 2\left\lceil 4 \cdot \frac{\sigma^2_{\mathcal{I}_w^+}}{C_1^2} \cdot \left(\operatorname{erf}^{-1}(2p-1)\right)^2 \right\rceil \leq 8 \cdot \frac{\sigma^2_{\mathcal{I}_w^+}}{C_1^2} \cdot \left(\operatorname{erf}^{-1}(2p-1)\right)^2 + 2. \tag{13}$$

See Figure 1 for a comparison of the bounds in Equations 11 and 13 in the noisy setting (i.e., $\sigma_{\mathcal{I}_w^+} \to \infty$).

Looking at Figure 1 we can see that the Equation 11 bound obtained for Algorithm 1 using Chernoff is within a relatively small constant factor of the Equation 13 bound obtained for Algorithm 1 modified to use maximum likelihood based methods. Indeed, for all desired success probabilities larger than 0.9 both bounds are within a factor of 5 from one another. Furthermore, one must remember that the Chernoff bound is not tight. Algorithm 1 actually requires less measurements per decision than the $m(p)$-bound graphed in Figure 1. *We may conclude that Algorithm 1 does not suffer significantly in the Gaussian case despite its ability to handle more general measurement noise.*

When Algorithm 1 is modified as described above, Equation 13 may be used to obtain the following alternate version of Corollary 1.

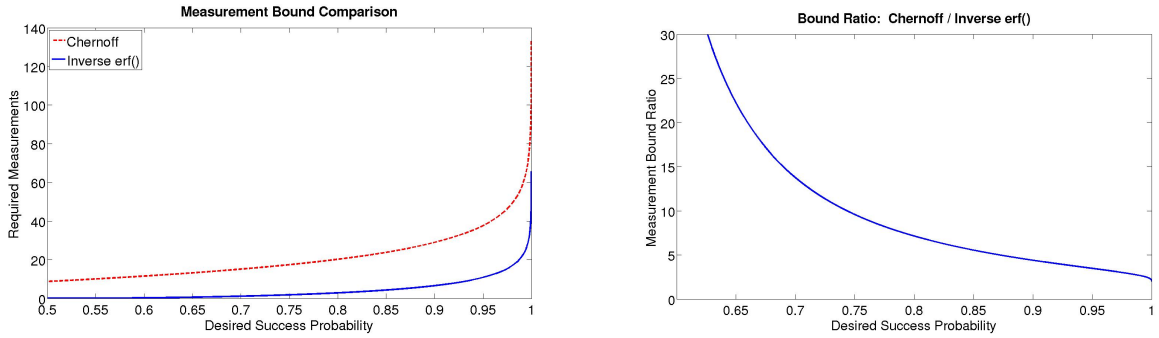


Figure 1: **Left:** Comparison of the bounds in Equations 11 and 13, each divided by $\frac{\sigma^2_{\mathcal{I}^+_w}}{C_1^2}$, as $\sigma_{\mathcal{I}^+_w} \to \infty$. **Right:** The ratio of the two bounds as $\sigma_{\mathcal{I}^+_w} \to \infty$. In effect, the graph of $\frac{4\pi \cdot \ln\left(\frac{1}{1-p}\right)}{8 \cdot \left(\text{erf}^{-1}(2p-1)\right)^2}$.

Corollary 2. *Let σ^2 be the variance of our Gaussian measurement noise over the interval $[0,1]$. Fix $p \in (0,1)$. Then, a modified variant of Algorithm 1 can correctly locate a C_1 magnitude spike within either $\mathcal{I}^+$ or $\mathcal{I}^-$ with probability at least p using no more than*

$$\sum_{n=0}^{\lceil \log_2 N \rceil - 1} 2^{2-\mathbb{I}_{\{0\}}(n)} \left\lfloor 4 \cdot \frac{\sigma^2}{2^n \cdot C_1^2} \cdot \left(\text{erf}^{-1}\left(2p^{\frac{1}{\lceil \log_2 N \rceil}} - 1\right)\right)^2 \right\rfloor \; < \; 24 \cdot \frac{\sigma^2}{C_1^2} \cdot \left(\text{erf}^{-1}\left(2p^{\frac{1}{\lceil \log_2 N \rceil}} - 1\right)\right)^2 + 4 \cdot \lceil \log_2 N \rceil - 2$$

measurements.

3.3 Gaussian Noise with Known Mean

Note that if we have prior information regarding the magnitude of C_1, Corollary 2's measurement bound may be improved by a constant factor of roughly $\frac{2}{3}$ for large desired success probabilities (e.g., p larger than 0.99) when the mean, μ, is known. To do so we modify our subinterval decision scheme (i.e., the use of mean estimates in Equation 12) as follows. First, if we know $|C_1|$ we need only consider one subinterval of $\mathcal{I}$ (note we have dropped the superscript '+' to indicate we are no longer making assumptions regarding the sign of C_1). Without loss of generality, let us always consider the right subinterval of our current test interval $\mathcal{I}$. Then, if

$$\left| \frac{1}{\tilde{m}(p)} \sum_{j=1}^{\tilde{m}(p)} m_{\mathcal{I}_r}(t_j) - \mu \right| > \frac{|C_1|}{2} \tag{14}$$

we conclude that the right subinterval contains our spike. Otherwise, we conclude the left subinterval contains our spike. No assumptions about the sign of C_1 are required. Thus, we no longer need to perform *two* binary searches (one assuming C_1 positive, the other assuming C_1 negative). The end result is an approximately $\frac{2}{3}$ constant factor reduction in the measurements necessary to locate our single spike for large desired success probabilities.

Given that our Gaussian measurement noise mean, μ, is known, the absolute value of the averaged measurements in Equation 14 follows a folded normal distribution (30). If we would like to use measurements along the lines of Equation 14, but do not want the resulting Algorithm 1 variant to depend explicitly on an assumed $|C_1|$ lower bound (despite the fact that its performance still will), we can always compare the two folded normal variables

$$\left| \frac{2}{\tilde{m}(p)} \sum_{j=1}^{\tilde{m}(p)/2} m_{\mathcal{I}_r}(t_j) - \mu \right| \text{ and } \left| \frac{2}{\tilde{m}(p)} \sum_{j=1}^{\tilde{m}(p)/2} m_{\mathcal{I}_l}(t_j) - \mu \right|$$

along the lines of Equation 12. This, again, allows us to avoid having to perform two binary searches, one for each possible sign of C_1. Many such variants of Algorithm 1 exist. For example, we could compare maximum likelihood estimates of folded normal variables (31) for both the right and left subintervals of $\mathcal{I}$, or we could directly utilize the ratio of these two folded normal variables (32) to derive measurement bounds. However, we neglect further consideration of such Algorithm 1 variants.

3.4 Summary

To conclude, we note that by employing a binary search for single spike recovery we are essentially transforming the spike localization problem into $O(\log_2 N)$ binary detection problems. Without loss of generality, at each stage of our binary search we must decide whether measurements of the left subinterval currently under consideration were generated by (*i*) a spike in noise, or (*ii*) noise alone. The answer to this question entirely determines whether the left or right subinterval becomes the new interval of interest in the next stage of our binary search. When viewed from this perspective the single spike recovery problem becomes equivalent to a series of statistical detection/estimation problems (see (26)). 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., (33)) at each stage of our binary search. These methods could be used to collect measurements dynamically until a decision regarding the presence/absence of a spike can be made with error probability below a user specified tolerance.

4 Multiple Spike Targeting

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 our disjoint subsets of $[0,1]$.

4.1 Recovering Multiple Spikes One at a Time

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

$$s_0 = \left[0, \frac{1}{N}\right),\ s_1 = \left[\frac{1}{N}, \frac{2}{N}\right),\ \ldots,\ s_{N-1} = \left[1 - \frac{1}{N}, 1\right], \tag{15}$$

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

Let q be one of the first $2k\lfloor \log_k N \rfloor$ prime numbers larger than k. These primes are easily found via standard sieving algorithms (see (34)). For each $h \in [0,q) \cap \mathbb{Z}$ form the set

$$\mathcal{I}_{q,h} = \bigcup_{j \equiv h \bmod q} s_j \tag{16}$$

and then set

$$\mathcal{I}_q = \left\{\mathcal{I}_{q,0},\ \mathcal{I}_{q,1},\ \ldots,\ \mathcal{I}_{q,q-1}\right\}. \tag{17}$$

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

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 $\{x_1, \ldots, x_j, \ldots, x_k\} \cap \mathcal{I}_{q,h} = \{x_j\}$ is true with probability at least $\frac{1}{2}$.*

Proof:

We prove this result along the lines of similar work in (35). 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 (34)). 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 $2k\lfloor \log_k N \rfloor$ 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 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 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

$$\frac{1-p}{2 \cdot q_{\max} \cdot \log_2\left(\frac{2k}{1-p}\right) \cdot \log_2 N}. \tag{18}$$

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.

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 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 any given $C_{\min}$ value will be returned. Thus, in general, we can guarantee the recovery of all sufficiently large (i.e., at least $C_{\min}$ in magnitude) spikes in f with high probability. See Algorithm 2 for multiple spike recovery pseudocode.

4.2 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 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 from Equation 18, we can see that we will need a total of

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

measurements.

To finish, we bound $q_{\max}$ in terms of k and N. Using results from (36) 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

$$\left[2k \log_k N \cdot \left(1 + \frac{1}{2 \ln N} + \frac{1.2762}{2 \ln k \ln N} + \frac{1}{2k \log_k N}\right)\right]^{\text{th}} \tag{19}$$

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, total measurement budget B
Output: Estimates of magnitudes $> \frac{1}{2}C_{\min}$, $\{C_1, \ldots, C_k\}$, and their positions, $\{x_1, \ldots, x_k\}$
Find all spikes at least once...
$SPIKES \longleftarrow \emptyset$
for $j = 1,\ j < P = O\left(\log\left(\frac{k}{1-p}\right)\right),\ j++$ **do**
  $q \longleftarrow$ Randomly select one of the $2k\lfloor \log_k N \rfloor$ primes $> k$
  Form $\mathcal{I}_q$ (see Equation 17)
  for each $\mathcal{I}_{q,h} \in \mathcal{I}_q$ **do**
    $(\tilde{x}, \tilde{C}) \longleftarrow$ Algorithm 1 with input ($\mathcal{I}_{q,h}$, $2N$, α, $O(1 -$ Equation 18$)$, $O(B/qP)$, γ from Lemma 2)
    if $|\tilde{C}| > \frac{1}{2}C_{\min}$ **then**
      $SPIKES \longleftarrow SPIKES \cup \{(\tilde{x}, \tilde{C})\}$
    end if
  end for
end for
Remove excess spike approximations...
$\{(\tilde{x}_0, \tilde{C}_0), (\tilde{x}_1, \tilde{C}_1) \ldots\} \longleftarrow$ Sort $SPIKES$ by $\tilde{x}$'s
for $n = 0,\ n < |SPIKES|,\ n++$ **do**
  while $|x_n - x_{n+1}| \leq \frac{N}{2}$ **do**
    $SPIKES \leftarrow SPIKES - \{(\tilde{x}_{n+1}, \tilde{C}_{n+1})\}$
  end while
end for
Return $SPIKES$

prime number. Therefore, we can see that $q_{\max}$ is at most the $\left(5k \cdot \log_k N\right)^{\text{th}}$ prime for all $N \geq k \geq 3$. Appealing again to results from (36) we can see that

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

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

$$O\left(\left(k \cdot \log^2 N \cdot \log_k(k \log_k N) + \frac{\sigma^2_{[0,1]}}{C^2_{\min}} + \frac{\sigma^2_{[0,1]}}{N \cdot \alpha^2 C^2_{\min}}\right) \cdot \log^2\left(\frac{k \cdot \log^2\left(\frac{k \log N}{1-p}\right) \cdot \log^2 N}{1-p}\right)\right) \tag{20}$$

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 20.*

4.3 Discussion

It is worth mentioning that the measurement bounds for Theorem 4 can be improved slightly to

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

by using adaptive group testing methods from (20) 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 20'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 20 is improved by roughly a $O\left(\log\left(\frac{k \log N}{1-p}\right)\right)$ factor. Similarly, in the noiseless positive spikes case (i.e., $\sigma^2_{[0,1]} = 0$, all $C_j \geq 0$) the measurement bounds for Theorem 4 can be improved to within a small constant factor of the $O(k \cdot \log(N/k))$ information theoretic lower measurement bound by using the linear-time two-stage identification process from (37). 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 our $\mathcal{I}_q$-subsets (see Equation 17) can be created non-adaptively before any measurements are taken. In fact, when used in combination with a bit testing matrix (see (20)) Theorem 4 can be achieved using entirely non-adaptive measurements. However, non-adaptive measurements incur an additional $O(\log N)$-factor on Theorem 4's noisy measurement bounds. Hence, we have focused on adaptive methods. 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 $O\left(\left(\frac{\sigma^2_{[0,1]}}{C^2_{\min}} + \frac{\sigma^2_{[0,1]}}{N \cdot a^2 C^2_{\min}} + k\right) \cdot \log^{O(1)}\left(\frac{N}{1-p}\right)\right)$-time. Hence, for modest noise levels the required runtime is sublinear in N.

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 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 on average. Furthermore, 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. 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. 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.

4.4 A Special Case: Gaussian White Measurement Noise

We now consider multiple spike support recovery in the case of Gaussian white measurement noise with unknown mean μ. Our goal in this section is to bound the number of measurements required to locate a set of approximate spike locations (i.e., $\mathcal{I}^{+/-}$-subsets) which contains good approximate locations of all k spike supports. Note that we only require that we locate a set containing the spike supports. We may also recover additional 'fake' or 'insignificant' spike supports. However, once a set containing the true supports is located, Lemma 2 can easily be applied to each potential spike at relatively small cost in order to determine whether it is significant. Thus, we concentrate on bounding the number of measurements required to locate some small superset of f's support.

We begin by considering the application of Corollary 1 to a slightly modified $\mathcal{I}_{q,h}$-subset (see Equation 16). We will assume for the purposes of this section that $\int \mathbb{I}_{\mathcal{I}_{q,h_1}}\, dx = \int \mathbb{I}_{\mathcal{I}_{q,h_2}}\, dx$ for all $h_1, h_2 \in [0, q) \cap \mathbb{Z}$. This is easily enforced by creating each $\mathcal{I}_{q,h}$ subset as before, except for the last $N \bmod q$ s_j-intervals (see Equation 15) which can be equally divided by length between the q $\mathcal{I}_{q,h}$-subsets. Thus, if $\sigma^2_{[0,1]}$ is the variance of $m_{[0,1]}(t)$, each $m_{\mathcal{I}_{q,h}}(t)$ will have variance $\frac{\sigma^2_{[0,1]}}{q}$. Using this modification along with Corollary 1, we have that the number of required Algorithm 1 measurements for each $\mathcal{I}_{q,h}$-subset is bounded above by

$$\frac{\ln\left(\frac{1}{1-\tilde{p}}\right)}{\operatorname{erf}^2\left(\frac{\sqrt{\bar{q}} \cdot C_{\min}}{2\sigma}\right)} + \sum_{n=1}^{\lceil \log_2(N/q) \rceil} \frac{2 \cdot \ln\left(\frac{1}{1-\tilde{p}}\right)}{\operatorname{erf}^2\left(\frac{(\sqrt{2})^n \cdot \sqrt{\bar{q}} \cdot C_{\min}}{2\sigma_{[0,1]}}\right)} = O\left(\left(\frac{\sigma^2_{[0,1]}}{q \cdot C^2_{\min}} + \log_2(N/q)\right) \cdot \ln\left(\frac{1}{1-\tilde{p}}\right)\right). \tag{21}$$

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 $q_M = q_{\max}$. Equation 19 tells us 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 4, we can see that the total number of $\mathcal{I}_{q,h}$-subsets to which we must apply Algorithm 1 is bounded by

$$S_1 = \sum_{j=M-\left\lfloor \log_2\left(\frac{2k}{1-p}\right)\right\rfloor}^{M} q_j < 10k \cdot \log_2\left(\frac{4k}{1-p}\right) \cdot \log_k N \cdot \log_2(5k \cdot \log_k N). \tag{22}$$

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

$$S_2 = \sum_{j=M-\left\lfloor \log_2\left(\frac{2k}{1-p}\right)\right\rfloor}^{M} q_j \cdot \log_2(2N/q_j) < S_1 \cdot \log_2(2N/k). \tag{23}$$

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 21. Combining Equations 21, 22, and 23 brings us to the following Corollary to Theorem 4.

Corollary 3. *Let σ^2 be the variance of our 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 $[0, 1]$. Let S_2 be defined as in Equation 23. Using less than*

$$\sum_{j=M-\left\lfloor \log_2\left(\frac{2k}{1-p}\right)\right\rfloor}^{M} q_j \left(\frac{\ln\left(\frac{2 \cdot S_2}{1-p}\right)}{erf^2\left(\frac{\sqrt{q_j} \cdot C_{\min}}{2\sigma}\right)} + \sum_{n=1}^{\lceil \log_2(N/q_j)\rceil} \frac{2 \cdot \ln\left(\frac{2 \cdot S_2}{1-p}\right)}{erf^2\left(\frac{\left(\sqrt{2}\right)^n \cdot \sqrt{q_j} \cdot C_{\min}}{2\sigma}\right)} \right)$$

measurements, we can correctly locate all spikes with magnitude at least $C_{\min}$ to within $\frac{1}{N}$-tolerance with probability at least p. Thus, the required number of measurements is

$$O\left(\left(\left(\frac{\sigma^2}{C_{\min}^2}\right) + k \cdot \frac{\ln N \cdot \ln(N/k) \cdot \ln(k \ln N)}{\ln k}\right) \cdot \ln\left(\frac{k}{1-p}\right) \cdot \ln\left(\frac{k \cdot \ln N \cdot \ln(N/k) \cdot \ln(k \ln N) \cdot \ln\left(\frac{k}{1-p}\right)}{\ln k \cdot (1-p)}\right)\right).$$

As $\sigma^2 \to \infty$ the number of required measurements is bounded above by

$$12\pi \cdot \frac{\sigma^2}{C_{\min}^2} \cdot \log_2\left(\frac{4k}{1-p}\right) \cdot \ln\left(\frac{20k \cdot \ln N \cdot \log_2(5k \cdot \log_k N) \cdot \log_2\left(\frac{4k}{1-p}\right) \cdot \log_2(2N/k)}{\ln k \cdot (1-p)}\right).$$

If, in addition, $N \to \infty$ in such a way that $\frac{N}{\sigma^2} \to 0$ this bound improves slightly to

$$12\pi \cdot \frac{\sigma^2}{C_{\min}^2} \cdot \log_2\left(\frac{4k}{1-p}\right) \cdot \ln\left(\frac{4k \ln N \cdot \ln(2k \log_k N) \cdot \log_2\left(\frac{4k}{1-p}\right) \cdot \log_2(2N/k)}{\ln k \cdot (1-p)}\right).$$

Proof:

We begin by bounding S_2 (see Equation 23). We have

$$S_2 < q_M \cdot \log_2\left(\frac{4k}{1-p}\right) \cdot \log_2(2N/k) < 10k \cdot \log_k N \cdot \log_2(5k \cdot \log_k N) \cdot \log_2\left(\frac{4k}{1-p}\right) \cdot \log_2(2N/k).$$

It is well know that $q_M \sim M \ln M$ (see (36)). Combining this fact with Equation 19 we can see that

$$q_M \sim 2k \log_k N \cdot \ln(2k \log_k N).$$

Hence, as $N \to \infty$, S_2 will be bounded above by

$$2k \log_k N \cdot \ln(2k \log_k N) \cdot \log_2\left(\frac{4k}{1-p}\right) \cdot \log_2(2N/k).$$

Combining Equations 21, 22, and 23 shows us that the number of required measurements is bounded above by

$$\sum_{j=M-\left\lfloor \log_2\left(\frac{2k}{1-p}\right)\right\rfloor}^{M} q_j \left(\frac{2 \cdot \ln\left(\frac{2 \cdot S_2}{1-p}\right)}{\operatorname{erf}^2\left(\frac{\sqrt{q_j} \cdot C_{\min}}{2\sigma}\right)} + \sum_{n=1}^{\lceil \log_2(N/q_j) \rceil} \frac{4 \cdot \ln\left(\frac{2 \cdot S_2}{1-p}\right)}{\operatorname{erf}^2\left(\frac{(\sqrt{2})^n \cdot \sqrt{q_j} \cdot C_{\min}}{2\sigma}\right)} \right) \tag{24}$$

which is

$$O\left(\sum_{j=M-\left\lfloor \log_2\left(\frac{2k}{1-p}\right)\right\rfloor}^{M} q_j \cdot \left(\left(\frac{\sigma^2}{q_j \cdot C_{\min}^2} + \log_2(N/q_j) \right) \cdot \ln\left(\frac{S_2}{1-p}\right) \right) \right) = O\left(\left(\frac{\sigma^2}{C_{\min}^2} \cdot \log\left(\frac{k}{1-p}\right) + S_2 \right) \cdot \ln\left(\frac{S_2}{1-p}\right) \right).$$

Substituting the S_2 bounds into the rightmost equation directly above gives us our stated result.

Finally, we consider the limit as $\sigma^2 \to \infty$. Utilizing Corollary 1 as we can see that as σ^2 increases Equation 24 becomes bounded by

$$\sum_{j=M-\left\lfloor \log_2\left(\frac{2k}{1-p}\right)\right\rfloor}^{M} q_j \cdot \left(12\pi \cdot \frac{\sigma^2}{q_j \cdot C_{\min}^2} \cdot \ln\left(\frac{2S_2}{1-p}\right) \right) \leq 12\pi \cdot \frac{\sigma^2}{C_{\min}^2} \cdot \log_2\left(\frac{4k}{1-p}\right) \cdot \ln\left(\frac{2S_2}{1-p}\right).$$

As before, combining this last inequality with our S_2 results finishes the proof. □

We conclude by noting that alternate bounds for multiple spike recovery can be developed along the lines of Section 3.2 for the Gaussian case. However, the overall improvements obtained are on the order of the modest constant factor decreases in measurements seen previously (e.g., see Figure 1). Hence, we leave the careful analysis of such alternate Algorithm 2 variants for future work.

5 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 (37)) with signal estimation and detection methods (e.g., see (26), (29)). More generally, any good binary group testing matrix which tends to isolate the members of any small number of signal components can be used to segment a search space into smaller regions likely containing only one signal component or target. Signal detection and estimation methods can then be used to search each smaller region for a single isolated signal component.

As a one final observation, we point out that 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. However, we leave further considerations along these lines to future work.

Acknowledgments

We would like to thank Vasileios Maroulas, Charles Doering, and Tsvetanka Sendova for helpful discussions and advice. Mark Iwen would also like to thank the IMA for support during the completion of the work chronicled herein.

References

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

[2] KM Mjelde, *Methods of the allocation of limited resources*, John Wiley & Sons, 1983.

[3] S.J. Benkoski, M.G. Monticino, and J.R. Weisinger, "A survey of the search theory literature," *Naval Research Logistics*, vol. 38, no. 4, pp. 469–494, 2006.

[4] JA Nachlas, SR Loney, and BA Binney, "Diagnostic-strategy selection for series systems," *IEEE Transactions on Reliability*, vol. 39, no. 3, pp. 273–280, 1990.

[5] AA Abdel-Samad and AH Tewfik, "Hierarchical radar target localization," in *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings*, 2000, vol. 5.

[6] SM Sowelam and AH Tewfik, "Waveform selection in radar target classification," *IEEE Transactions on Information Theory*, vol. 46, no. 3, pp. 1014–1029, 2000.

[7] M.R. Bell and Chieh-Fu Chang, "Frequency coded waveforms for adaptive waveform radar," in *Information Sciences and Systems, 2006 40th Annual Conference on*, 22-24 2006, pp. 508 –511.

[8] AA Abdel-Samad and AH Tewfik, "Sequential techniques in hierarchical radar target localization," in *Image Processing, 2000. Proceedings. 2000 International Conference on*, 2000, vol. 1.

[9] 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.

[10] JA Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.

[11] JJ Fuchs, "Recovery of exact sparse representations in the presence of bounded noise," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3601–3608, 2005.

[12] E.J. Candès, J.K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207, 2006.

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

[14] N. Meinshausen and P. B
"uhlmann, "High-dimensional graphs and variable selection with the lasso," *The Annals of Statistics*, pp. 1436–1462, 2006.

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

[16] Alyson K. Fletcher, Sundeep Rangan, and Vivek K. Goyal, "Necessary and sufficient conditions on sparsity pattern recovery," *Preprint*, April 2008.

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

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

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

[20] Anna C. Gilbert, M. A. Iwen, and Martin J. Strauss, "Group testing and sparse signal recovery," *42nd Asilomar Conference on Signals, Systems, and Computers*, 2008.

[21] Charles R. Doering, "Modeling complex systems: Stochastic processes, stochastic differential equations, and fokker-plank equations," *1990 Lectures In Complex Systems*, vol. 3, 1990.

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

[23] 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.

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

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

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

[27] Rajeev Motwani and Prabhakar Raghavan, *Randomized Algorithms*, Cambridge University Press, 1995.

[28] Geoffrey Grimmett and David Stirzaker, *Probability and Random Processes*, Oxford University Press, 2001.

[29] David Tse and Pramod Viswanath, *Fundamentals of wireless communication*, Cambridge University Press, New York, NY, USA, 2005.

[30] F.C. Leone, L. S. Nelson, and R. B. Nottingham, "The folded normal distribution," *Technometrics*, vol. 3, no. 4, 1961.

[31] N. L. Johnson, "The folded normal distribution: Accuracy of estimation by maximum likelihood," *Technometrics*, vol. 4, no. 2, 1962.

[32] Hea-Jung Kim, "On the ratio of two folded normal distributions," *Communications in Statistics - Theory and Methods*, vol. 35, no. 6, 2006.

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

[34] Ivan Niven, Herbert S. Zuckerman, and Hugh L. Montgomery, *An Introduction to The Theory of Numbers*, John Wiley & Sons, Inc., 1991.

[35] M. A. Iwen, "Combinatorial sublinear-time fourier algorithms," *Foundations of Computational Mathematics*, vol. 10, 2010.

[36] Pierre 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.

[37] David Eppstein, Michael T. Goodrich, and Daniel S. Hirschberg, "Improved combinatorial group testing algorithms for real-world problem sizes," May 2005.

[38] Marco F. Duarte, Mark A. Davenport, Dharmpal Takhar, Jason N. Laska, Ting Sun, Kevin F. Kelly, and Richard G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 83, March 2008.

[39] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, 2008.

A Proof of Theorem 1

Suppose f is k-sparse when represented in terms of a known orthonormal basis $\Phi = \{\phi_i \mid 1 \leq i \leq N\}$ of either (i) $\mathbb{R}^N$ in the discrete case, or (ii) a subspace of periodic functions on $[0,1]$ in the continuous setting. That is, $f = \sum_{i=1}^{N} \vec{f_i} \cdot \phi_i$ where $\vec{f} \in \mathbb{R}^N$ is k-sparse. Let $C_{\min} = \min\left\{\left|\vec{f_i} = \langle f,\ \phi_i\rangle\right| \ \middle|\ i \in \text{supp}(f)\right\}$ be the magnitude of the smallest of the k non-zero coefficients of $\vec{f} = \Phi[f] \in \mathbb{R}^N$.

In the discrete case we directly define $\mathcal{G} = \left\{\mathcal{G}_j \ \middle|\ 1 \leq j \leq m\right\}$ to be an $m \times N$ real-valued random matrix. In this case $\mathcal{G}_j$ denotes the j^{th}-row of $\mathcal{G}$, each of which is independently drawn from the zero-mean isotropic Gaussian distribution $\mathcal{N}(0, I_{N\times N})$. Each measurement test function, $\mathcal{M}_j = \Phi\mathcal{G}_j$, will then also have the zero-mean isotropic Gaussian distribution $\mathcal{N}(0, I_{N\times N})$ since Φ is orthonormal. In the continuous case we assume that each $\mathcal{G}_j \in \mathcal{G}$ is an independent Gaussian white noise process (i.e., $\mathcal{G}_j = dW/dt$ for $1 \leq j \leq m$). Thus, each measurement will have

$$\mathcal{M}_j = \sum_{i=1}^{N} \langle \mathcal{G}_j, \phi_i\rangle \cdot \phi_i = \sum_{i=1}^{N} \mathfrak{g}_{j,i} \cdot \phi_i$$

where the $\mathfrak{g}_{j,i} \sim \mathcal{N}(0,1)$ are uncorrelated standard normal random variables. Hence, in both the discrete and continuous cases we ultimately end up with measurements of the form

$$\langle \mathcal{M}_j, f + \mathcal{P}_j\rangle = \langle \mathcal{M}_j, f\rangle + \langle \mathcal{M}_j, \mathcal{P}_j\rangle = \sum_{i=1}^{N} \mathfrak{g}_{j,i} \cdot \vec{f_i} + \sum_{i=1}^{N} \mathfrak{g}_{j,i} \cdot \langle \phi_i, \mathcal{P}_j\rangle. \tag{25}$$

Therefore, we see that our test functions, $\mathcal{M}_j$, will always produce measurements equivalent to the inner product of $\vec{f} \in \mathbb{R}^N$ with a random Gaussian vector $\mathfrak{g}_j \sim \mathcal{N}(0, I_{N\times N})$, plus additive noise. We will next consider the additive measurement noise generated by $\mathcal{P}_j$.

In the discrete setting we will simply define an $m \times N$ noise matrix, $\mathcal{P}$, having independent and identically distributed (i.i.d.) normal entries with mean 0 and variance σ^2/N. In this case $\mathcal{P}_j \sim \mathcal{N}(0, \frac{\sigma^2}{N} \cdot I_{N\times N})$ will denote the j^{th} row of $\mathcal{P}$. In the continuous setting we assume that each $\mathcal{P}_j$ is an independent weighted Gaussian white noise process (i.e., $\mathcal{P}_j = \frac{\sigma}{\sqrt{N}} \cdot dW/dt$ for $1 \leq j \leq m$). In either case it is not difficult to see that the accumulated additive measurement noise from Equation 25, conditioned on the test function $\mathcal{M}_j$, is $\langle \mathcal{M}_j, \mathcal{P}_j\rangle \ \sim\ \mathcal{N}(0, \|\mathcal{M}_j\|_2^2 \cdot \sigma^2/N)$.

Based on the two previous paragraphs we can see that it suffices to prove Theorem 1 in the simplified discrete setting where $f = \vec{f} \in \mathbb{R}^N$ is itself a k-sparse vector. Our non-adaptive measurements are then 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 $\mathcal{N}(0, I_{N\times N})$. Furthermore, we may define $\mathcal{P}$ 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

$$\langle \mathcal{M}_j, f + \mathcal{P}_j\rangle = \langle \mathcal{M}_j, f\rangle + \langle \mathcal{M}_j, \mathcal{P}_j\rangle. \tag{26}$$

Examples of such single detector systems include the famous single pixel camera (38). Given such a simplified setup we want to know how many measurements we must take in order to recover the support of f. Equivalently, we want to know how large we must make m in order to be able to entirely recover $\text{supp}(f)$ using only the measurements $\mathcal{M}f$. We will answer this question by adapting the proof of Theorem 2 from (13) to our simplified Equation 26 measurement model.

We begin setting up the proof by fixing some notation. 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 $\vec{p}_j = \langle \mathcal{M}_j, \mathcal{P}_j\rangle \ \sim\ \mathcal{N}(0, \|\mathcal{M}_j\|_2^2 \cdot \sigma^2/N)$ and $\vec{w}_j \sim \mathcal{N}(0, \sigma^2)$, respectively, for all $j \in [1, m] \cap \mathbb{N}$.

Consider the related restricted problem of recovering a k-sparse vector, $\tilde{f} \in \mathbb{R}^N$, which is known a priori to have both (i) $\text{supp}(\tilde{f}) = \text{supp}(f)$, and (ii) $\tilde{f}_j = C_{\min}$ for all $j \in \text{supp}(\tilde{f})$. We can solve this restricted problem by searching

all $\tilde{N} = \binom{N}{k}$ subsets of size k from $[1, N] \cap \mathbb{N}$ seeking to minimize

$$\begin{aligned} g(U) &= \left\| \left(\mathcal{M}\tilde{f} + \vec{p} \right) - \mathcal{M}_U \vec{v} \right\|_2^2 \\ &= \left\| \left(\mathcal{M}_{\text{supp}(\tilde{f})} - \mathcal{M}_U \right) \vec{v} + \vec{p} \right\|_2^2. \end{aligned} \tag{27}$$

Note that $\|\mathcal{M}_j\|_2^2 \cdot \sigma^2/N$ follows the Chi- square distribution with mean σ^2 and variance $2\sigma^4/N$. Thus, Chebyshev's Inequality tells us that

$$\mathbb{P}\left[\left| \|\mathcal{M}_j\|_2^2 \cdot \sigma^2/N - \sigma^2 \right| \geq \epsilon \right] \leq \frac{2\sigma^4}{N \cdot \epsilon^2}$$

for $\epsilon \in (0, 1)$ sufficiently small. Setting, for example, $\epsilon = \sqrt{2}\sigma^2/\sqrt{\log N}$ the union bound tells us that $\vec{p} \xrightarrow{d} \vec{w}$ as $N \to \infty$ for all $m = o(N/\log N)$. Furthermore, we can see that $g(U)$ in Equation 27 provides an asymptotically optimal decoder for the restricted problem.

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 $\tilde{f}$ 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 our random non-adaptive measurement matrix $\mathcal{M}$. Hence, we have that

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

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. Let

$$Y_j' = Y_j - \sum_{n \in U[i] \cup U[i']} \mathcal{M}_{j,n}^2 \quad = \sum_{n \in [1,N] \cap \mathbb{N} - (U[i] \cup U[i'])} \mathcal{M}_{j,n}^2.$$

Then

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

where $Y_j' \sim \chi^2_{N - |U[i] \cup U[i']|}$ and $Z_j \sim \mathcal{N}(0, 1)$ are independent for each $j \in [1, m] \cap \mathbb{N}$. Therefore we will have

$$\begin{aligned} \mathbb{E}\left[D(\mathbb{P}_i \| \mathbb{P}_{i'}) \right] &\leq \left(\frac{C_{\min}^2}{\sigma^2} \right) \cdot (k - |U[i] \cap U[i']|) \cdot \sum_{j=1}^{m} \mathbb{E}\left[\frac{N}{Y_j'} \right] \cdot \mathbb{E}\left[Z_j^2 \right] \\ &\leq \left(\frac{C_{\min}^2}{\sigma^2} \right) \cdot (k - |U[i] \cap U[i']|) \cdot \frac{m}{1 - \frac{2k-2}{N}}. \end{aligned} \tag{28}$$

Equipped with this upper bound for the expectation of each divergence as a function of our random measurement matrix we may now consider the number of measurements necessary to reliably recover the support of $\tilde{f}$.

The remainder of the proof depends on employing the following weakened form of Fano's inequality (see Lemma 2 in (13)). 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

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

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

$$\begin{aligned}
\mathbb{E}\left[1 - p_{\text{error}}\right] &\leq \frac{1}{\log\left(\tilde{N} - 1\right)} \cdot \mathbb{E}\left[\frac{1}{\tilde{N}^2} \cdot \sum_{i,i'=1}^{\tilde{N}} D\left(\mathbb{P}_i \| \mathbb{P}_{i'}\right)\right] + \frac{\log 2}{\log\left(\tilde{N} - 1\right)} \\
&\leq \frac{1}{\log\left(\tilde{N} - 1\right)} \cdot \left(\frac{C_{\min}^2}{\sigma^2}\right) \cdot k \cdot \frac{m}{1 - \frac{2k-2}{N}} + \frac{\log 2}{\log\left(\tilde{N} - 1\right)}
\end{aligned}$$

by Equation 28. Applying Markov's Inequality we can see that the average probability of successfully decoding the support of an input $\tilde{f}$ as a function of the randomly chosen measurement matrix is bounded by

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

Therefore, 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, $\tilde{f}$, 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

$$\begin{aligned}
m &< \left(\frac{1 - \frac{2k-2}{N}}{16}\right) \cdot \frac{\sigma^2}{C_{\min}^2} \cdot \log\left(N/k\right) \\
&\leq \left(\frac{1 - \frac{2k-2}{N}}{8}\right) \cdot \frac{\sigma^2}{C_{\min}^2} \cdot \frac{\log\left(\tilde{N} - 1\right)}{k}
\end{aligned}$$

with $N \geq 2k \geq 32$. Theorem 1 follows. Mainly, we see that we can not expect non-adaptive Gaussian measurements to allow the recovery of all k-sparse signals with high probability unless m is $\Omega\left(\frac{\sigma^2}{C_{\min}^2} \cdot \log\left(N/k\right)\right)$.

As one final comment we note that methods from (13) can be used much 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 $\Omega\left(\frac{\sigma^2}{C_{\min}^2} \cdot \log\left(N/k\right)\right)$. Similarly, we expect that more complicated modifications of this argument can also be used to demonstrate that this scaling for m is in fact required for other random non-adaptive measurement ensembles commonly utilized for sparse recovery problems (e.g., see (39)).