Sequential Analysis in High Dimensional Multiple Testing and Sparse Recovery

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

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

Abstract—This paper studies the problem of high-dimensional multiple testing and sparse recovery from the perspective of sequential analysis. In this setting, the probability of error is a function of the dimension of the problem. A simple sequential testing procedure for this problem is proposed. We derive necessary conditions for reliable recovery in the non-sequential setting and contrast them with sufficient conditions for reliable recovery using the proposed sequential testing procedure. Applications of the main results to several commonly encountered models show that sequential testing can be exponentially more sensitive to the difference between the null and alternative distributions (in terms of the dependence on dimension), implying that subtle cases can be much more reliably determined using sequential methods.

I. INTRODUCTION

High dimensional testing and sparse recovery problems arise in a broad range of scientific and engineering applications. The basic problem is summarized as follows. Let $\boldsymbol{\theta} \in \mathbb{R}^n$ denote a parameter vector. The dimension n may be very large (thousands or millions or more), but $\boldsymbol{\theta}$ is sparse in the sense that most its elements are equal to a baseline/null value denoted by θ_0 (e.g., $\theta_0 = 0$). The support of the sparse subset of elements that deviate from the baseline is denoted by $\mathcal{S}$.

The parameter $\boldsymbol{\theta}$ is observed stochastically according to

$$y_i \sim \begin{cases} f(y_i|\theta_0) & i \notin \mathcal{S} \\ f(y_i|\theta_1) & i \in \mathcal{S} \end{cases}, \quad (1)$$

where $f(\cdot|\theta)$ is a parametric family of densities indexed by a scalar parameter $\theta \in \mathbb{R}$. The goal of the high-dimensional testing and sparse recovery problem is to identify $\mathcal{S}$ from observations of this form. This problem has attracted attention lately due to its importance in the biological sciences. It is also relevant in communications problems including spectrum sensing in cognitive radio, one of the motivations for our work.

The conventional theoretical treatment of this problem assumes that a set of observations are collected prior to data analysis. Typically, in what we refer to as the non-sequential setting, each of the n components is measured (one or more times) according to the model above *and then* component-wise tests are performed to estimate $\mathcal{S}$.

This papers investigates the high-dimensional testing problem from the perspective of sequential analysis. In this setting, observations are gathered sequentially and adaptively, based on information gleaned from previous observations. This allows the observation process to focus sensing resources on certain components at the expense of ignoring others. For example, the process might first measure each component once, then focus on a reduced subset of 'interesting' components in a second pass.

To compare sequential and non-sequential methods we impose a budget on the total number of measurements that can be made. The main results show sequential methods can be dramatically more sensitive to small differences between the baseline/null θ_0 and the alternative value of θ_1. Our approach is similar to the so-called *distilled sensing* method proposed in [1] [2], however there are two main distinctions. First, the results in this paper are applicable to a large class of problems characterized by one-sided tests; the distilled sensing approach is specific to the Gaussian setting. Second, here we are concerned with the probability of error in identifying $\mathcal{S}$, whereas the distilled sensing controls the false discovery and non-discovery rates which is less demanding than the error rate control. The probability of error is more natural and appropriate in applications such as spectrum sensing.

To give a sense of the main results, consider the case in which $f(\cdot|\theta)$ is a Gaussian with mean θ and variance 1. If $\theta_0 = 0$ and the alternative is $\theta_1 > 0$, then reliable detection (probability of error tending to zero as $n \to \infty$) is possible using non-sequential methods if and only if $\theta_1 > \sqrt{2\log n}$. In contrast, a sequential method that we will demonstrate is reliable as long as $\theta_1 > \sqrt{4\log|\mathcal{S}|}$, where $|\mathcal{S}|$ is the cardinality of the support set. This shows that the sequential method is more sensitive whenever $|\mathcal{S}| < n^{1/2}$; i.e., the sparse setting. The improvement is especially remarkable when $\boldsymbol{\theta}$ is very sparse; e.g., if $|\mathcal{S}| \sim \log n$, then sequential methods succeed as long as θ_1 is larger than a constant mutiple of $\sqrt{\log\log n}$. The gains provided by the sequential method are even more pronounced for certain one-sided distributions. Under the Gamma distribution model (which arises in spectrum sensing), for constants c and C, if $\theta_0 \geq C \log(|\mathcal{S}|\log_2 n)$ then the sequential method is reliable, but any non-sequential thresholding procedure is unreliable if $\theta_0 \leq c\, n^{\frac{1}{2}}$. To dramatize this result, if $|\mathcal{S}| \sim \log n$, then the gap between these conditions is *doubly* exponential in n.

II. PROBLEM STATEMENT

We begin by stating a main assumption about the family $f(\cdot|\theta)$. Let $y_1, \ldots, y_m$ be i.i.d. random variables with common

distribution $f(\cdot|\theta)$, for some $\theta \in \mathbb{R}$. Let $y = (y_1, \ldots, y_m)$ and define the likelihood ratio

$$\Gamma(y) := \prod_{j=1}^{m} \frac{f(y_j|\theta_1)}{f(y_j|\theta_0)} \ .$$

Assumption A1. $\Gamma(y)$ *is a monotone non-decreasing function for* $\theta_1 \geq \theta_0$.

We will state our main results with this monotonicity assumption. However, in certain applications we consider it is more natural to consider $\theta_1 \leq \theta_0$ and assume that the likelihood ratio is a monotone non-increasing function. The main results carry over to this setting with appropriate modification. Define T_m as the (log) likelihood ratio test statistic, which is a function of y. The test statistic depends on the number of independent observations, and so this is indicated by the subscript m. If **A1** holds, then the test at threshold $\tau \in \mathbb{R}$

$$T_m \overset{\theta_1}{\underset{\theta_0}{\gtrless}} \tau \ ,$$

is the uniformly most powerful test (UMP) of $\theta \leq \theta_0$ versus $\theta > \theta_0$. The monotonicity of the likelihood ratio is satisfied by a large number of distributions in the exponential family (including Gaussian, Poisson and exponential distributions).

A. Measurement Budget

To compare sequential and non-sequential methods we impose a budget on the total number of measurements. The total number of measurements $N \leq 2mn$, where $m \geq 1$ is an integer and n is the dimension of $\boldsymbol{\theta}$.

B. Non-Sequential Testing

The non-sequential approach distributes the measurement budget uniformly over the n components, making $2m$ i.i.d. observations of each. Let $y_{i,1}, \ldots, y_{i,2m}$ denote the m observations of component i, and let $T_{i,2m}$ denote the corresponding test statistic. The UMP test takes the form

$$T_{i,2m} \overset{\theta_1}{\underset{\theta_0}{\gtrless}} \tau \ . \tag{2}$$

The estimated support set at threshold τ is defined to be

$$\mathcal{S}_\tau := \{i \ : \ T_{i,2m} > \tau\} \ .$$

This estimator is optimal among all (non-sequential) component-wise procedures because each test is UMP.

C. Sequential Thresholding

The sequential method we proposed is based on the following simple bisection idea. Instead of aiming to identify the components in $\mathcal{S}$ (those with $\theta = \theta_1$), at each step of the sequential procedure we aim to eliminate about $1/2$ of the components that follow the null, $\theta = \theta_0$, from further consideration. The components that remain after K such steps is our estimate of the set $\mathcal{S}$.

Suppose we begin by using half of our measurement budget to collect m observations of each component. The test statistic

for each is $T_{i,m} := T(y_{i,1}, \ldots, y_{i,m})$. Assume θ_0 is known and let $T_m|\theta_0$ denote the random variable whose distribution is that of the test statistic under the null, $\theta = \theta_0$. Consider the threshold test

$$T_{i,m} \ > \ \mathrm{median}(T_m|\theta_0) \ .$$

For $i \notin \mathcal{S}$, the test statistic $T_{i,m}$ falls below $\mathrm{median}(T_m|\theta_0)$ with probability $1/2$. The threshold test above thus eliminates approximately $1/2$ of the components that follow the null. We can next use a portion of our remaining budget of mn to repeat the same measurement and thresholding procedure on the remaining components. Since approximately $n/2$ components remain this will require $mn/2$ of the remaining budget. Repeating this process for sufficiently many iterations will remove, with high probability, all of the null components. We call this process *sequential thresholding* and give a formal algorithm below. The output of the procedure, $\mathcal{S}_K$, is the estimated support set. Notice that sequential thresholding does not require prior knowledge of the size of the support set.

Sequential Thresholding

input: $K > 0$ steps, $\gamma_0 := \mathrm{median}(T_m|\theta_0)$
initialize: $\mathcal{S}_0 = \{1, \ldots, n\}$
for $k = 1, \ldots, K$ **do**
  for $i \in \mathcal{S}_{k-1}$ **do**
    measure: $\{y_{i,j}^{(k)}\}_{j=1}^{m} \overset{iid}{\sim} \begin{cases} \prod_{j=1}^{m} f(y_{i,j}^{(k)}|\theta_0) & i \notin \mathcal{S} \\ \prod_{j=1}^{m} f(y_{i,j}^{(k)}|\theta_1) & i \in \mathcal{S} \end{cases}$
    threshold: $\mathcal{S}_k := \{i \in \mathcal{S}_{k-1} \ : T_{i,m}^{(k)} > \gamma_0\}$
  end for
end for
output: $\mathcal{S}_K$

D. Sequential Thresholding Satisfies Budget

The number of measurements used by sequential thresholding satisfies the overall measurement budget $N \leq 2mn$ in expectation. Let $s = |\mathcal{S}|$, the cardinality of the support set. The expected number of measurements is

$$\mathbb{E}\left[\sum_{k=0}^{K-1} |\mathcal{S}_k|\right] \leq \sum_{k=0}^{K-1} \left(\frac{m(n-s)}{2^k} + ms\right)$$
$$\leq 2m(n-s) + msK \ .$$

Our interest is in high-dimensional limits of n and s (and possibly K). Suppose that sK grows sublinearly with n. Then for any $\epsilon > 0$ there exists an N_ϵ such that $\mathbb{E}\left[\sum_{k=0}^{K-1} |\mathcal{S}_k|\right] \leq 2(1+\epsilon)mn$ for every $n > N_\epsilon$. For ease of exposition, we suppress the factor $1+\epsilon$ as we proceed; it does not effect the main results and conclusions of the paper as allowing the non-sequential method $2(1+\epsilon)mn$ observations is inconsequential.

E. Implementations

There are two possible implementations of sequential thresholding which we refer to as *parallel* and *scanning*.

parallel: The parallel implementation measures and tests all n components in parallel according to the procedure.

scanning: The scanning implementation measures and tests the n components in a sequence (which can be arbitrary). For example, the scanning implementation can begin with component $i = 1$ and repeatedly measure and threshold the observations up to K times. If an observation falls below the threshold at any point, then the scanning procedure immediately moves on to the next component. If K observations are made without an observation falling below the threshold, then the component is added to the set S_K. The expected number of observations obeys the same bound as derived above.

The two implementations are equivalent from a theoretical perspective. The parallel implementation may be more natural for large-scale experimental designs (e.g., in the biological sciences), whereas the scanning implementation is more appropriate in communications applications such as spectrum sensing. The latter also reveals natural connections between sequential thresholding and sequential probability ratio tests.

F. Connection to Sequential Probability Ratio Tests

As we will show in the following section, in the high-dimensional limit ($n \to \infty$) sequential thresholding can drive the probability of error to zero if the divergence between the null and alternative distributions is $\log |\mathcal{S}|$ times a small constant. This specializes in the Gaussian setting to the requirement that the difference between the means is at least $\sqrt{4 \log |\mathcal{S}|}$, which compares favorably to the requirement that the difference exceeds $\sqrt{2 \log n}$ for non-sequential methods.

In fact, the $\log |\mathcal{S}|$ dependence of sequential thresholding is optimal, up to constant factors. This follows from well-known results in sequential testing. Let $\widehat{\mathcal{S}}$ denote the result of any testing procedure based on n local (component-wise) tests of the form $H_0 : i \notin \mathcal{S}$ against $H_1 : i \in \mathcal{S}$. Each test is based on the sequential observations $y_{i,1}, y_{i,2}, \ldots, y_{i,N}$, and the *stopping time* of the test is the value of N (possibly random) when the decision is made.

Suppose that each individual test has false-positive and false-negative error probabilities less than $\alpha := \epsilon/(n - |\mathcal{S}|)$ and $\beta := \epsilon/|\mathcal{S}|$, respectively. Then the expected total number of errors is $\mathbb{E}|\widehat{S} \cap \mathcal{S}^C| + \mathbb{E}|\widehat{S}^C \cap \mathcal{S}| \leq 2\epsilon$. It is necessary that this expected number tend to zero in order for the probability of error, $\mathbb{P}(\widehat{S} \neq \mathcal{S})$, to tend zero. With the above specifications for the two types of error, it is possible to design a sequential probability ratio test (SPRT) for each component.

The SPRT computes a sequence of likelihood ratios, where $\ell_{i,n}$ is the likelihood ratio of $y_{i,1}, \ldots, y_{i,n}, n \geq 1$. The SPRT terminates when $\ell_{i,n} \geq B$ or $\ell_{i,n} \leq A$, where the thresholds A and B are determined by the equations $\alpha = B^{-1}(1 - \beta)$ and $\beta = A(1 - \alpha)$ (see [3] p. 11). Note that, unlike sequential thresholding, the SPRT requires knowlege of both distributions as well as the level of sparsity. Since such information is usually unavailable in applications, we advocate the use of sequential thresholding instead; it requires only crude knowledge of the null and nothing about the alternative or sparsity

level.

From the Wald equation, the expected stopping time of the SPRT per index is (approximately) [3]

$$\mathbb{E}_0[N'] \cong \mu_0^{-1}\left\{\alpha \log\left(\frac{1-\beta}{\alpha}\right) + (1-\alpha)\log\left(\frac{\beta}{1-\alpha}\right)\right\}$$

$$\mathbb{E}_1[N'] \cong \mu_1^{-1}\left\{(1-\beta)\log\left(\frac{1-\beta}{\alpha}\right) + \beta \log\left(\frac{\beta}{1-\alpha}\right)\right\},$$

where $\mathbb{E}_i$ denotes the expectation under $f(\cdot|\theta_i)$ and $\mu_i := \mathbb{E}_i\left[\log \frac{f(y|\theta_1)}{f(y|\theta_0)}\right]$, $i = 0, 1$. In our case $\alpha = \epsilon/(n - |\mathcal{S}|)$ and $\beta = \epsilon/|\mathcal{S}|$, and as $\epsilon \to 0$ we have

$$\mathbb{E}_0[N'] \cong \mu_0^{-1} \log \frac{\epsilon}{|\mathcal{S}|}$$

$$\mathbb{E}_1[N'] \cong \mu_1^{-1} \log \frac{n - |\mathcal{S}|}{\epsilon} .$$

If $|\mathcal{S}| \ll n$, then the expected total number of measurements of made by all n SPRTs is

$$E[N] = (n - |\mathcal{S}|)\mathbb{E}_0[N'] + |\mathcal{S}|\mathbb{E}_1[N'] \cong \frac{n}{\mu_0} \log \frac{\epsilon}{|\mathcal{S}|} .$$

Note that $\mu_0 = -D_0 := -D(f(\cdot|\theta_0)\|f(\cdot|\theta_1))$, the KL divergence of $f(\cdot|\theta_1)$ from $f(\cdot|\theta_0)$, so expected total number of observations made by the n SPRTs is

$$E[N] \cong \frac{n}{D_0} \log \frac{|\mathcal{S}|}{\epsilon} .$$

It follows from the optimality of the SPRT that no other component-wise testing procedure with ϵ error-rate requires fewer observations. Now let us constrain this expected total to be less than or equal to $2mn$. This yields a necessary condition for controlling the probability of error of any sequential test:

$$D(f(\cdot|\theta_0)\|f(\cdot|\theta_1)) \gtrsim \frac{1}{2m} \log \frac{|\mathcal{S}|}{\epsilon} .$$

III. MAIN RESULTS

The main results rely on the extremal properties of the test statistic. We say that a testing procedure is reliable if it drives the probability of error to zero in the high-dimensional limit. More formally, consider a sequence of multiple testing problems indexed by dimension n. Let $\mathcal{S}(n)$ denote the true support set and let $\widehat{\mathcal{S}}(n) = \mathcal{S}_\tau$ (non-sequential procedure) or $\widehat{\mathcal{S}}(n) = \mathcal{S}_K$ (sequential procedure). We define a notion of reliability as follows.

Definition III.1. (Reliability) *Let $\mathcal{E}$ denote the error event $\{\widehat{\mathcal{S}}(n) \neq \mathcal{S}(n)\}$. We say that the support set estimator $\widehat{\mathcal{S}}(n)$ is reliable if $\lim_{n\to\infty} \mathbb{P}(\mathcal{E}) = 0$; i.e., if the probability of error tends to zero as n grows.*

To keep the notation simple, in what follows we will not explicitly indicate the dependence of the statistics on n. We show that the non-sequential testing procedure in (2) is *unreliable* at every threshold level τ if

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\max_{i\notin\mathcal{S}} T_{i,2m}}{\text{median}(T_{2m}|\theta_1)} \geq 1\right) = 1 . \tag{3}$$

Sequential testing according to sequential thresholding is reliable if

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\min_{k=1}^{K} \min_{i\in\mathcal{S}} T_{i,m}^{(k)}}{\text{median}\,(T_m|\theta_0)} \le 1 \right) = 0 \;, \tag{4}$$

and $K = (1+\epsilon)\log_2 n$, for any $\epsilon > 0$. We are interested in ranges of $\theta_1 > \theta_0$ that satisfy the conditions above. In many cases of interest, (3) and (4) hold simultaneously for a wide range parameter values. This implies that there are many regimes in which sequential methods are reliable, but non-sequential methods are not.

For example, we show in Section IV-A that if the underlying component distributions are unit variance Gaussian with means $\theta_0 = 0$ and $\theta_1 > 0$, then the non-sequential procedure (2) is unreliable if $\theta_1 < \sqrt{\frac{1}{m}\log n}$ whereas sequential thresholding is reliable if $\theta_1 \ge \sqrt{\frac{2}{m}\log(|\mathcal{S}|\log_2 n)}$. The size of the sparse support $|\mathcal{S}|$ is typically much smaller than the overall dimension n, and so there are many cases in which the sequential method is reliable but the non-sequential method is not. The gap between the two conditions can be exponentially large in terms of the dimension n. As a specific example, if the support is $|\mathcal{S}| \sim \log n$, then there are constants $C, c > 0$ such that the sequential method is reliable if $\theta_1 \ge C\sqrt{\log\log n}$ and the non-sequential method is not if $\theta_1 \le c\sqrt{\log n}$.

A. Limitation of Non-Sequential Testing

Theorem III.2. *If (3) holds, then the non-sequential procedure in (2) is unreliable. Specifically, if $\mathcal{E}_\tau$ is the error event $\{\mathcal{S}_\tau \ne \mathcal{S}\}$, then for every τ*

$$\lim_{n\to\infty} \mathbb{P}\,(\mathcal{E}_\tau) \;\ge\; \frac{1}{2}.$$

Proof: The non-sequential testing procedure accepts the null hypothesis if the test statistic $T_{i,2m}$ is less than some threshold, τ, and conversely, rejects the null hypothesis if $T_{i,2m} \ge \tau$. The probability of error at threshold level τ is

$$\mathbb{P}\,(\mathcal{E}_\tau) \;=\; \mathbb{P}\left(\bigcup_{i\notin\mathcal{S}}\{T_{i,2m} \ge \tau\} \bigcup_{i\in\mathcal{S}}\{T_{i,2m} < \tau\} \right),$$

and the *minimum* probability of error is $\min_\tau \mathbb{P}\,(\mathcal{E}_\tau)$. Now suppose we take $\tau = \text{median}(T_{2m}|\theta_1)$, the median value of the test statistic under the alternative. At this threshold level, the false-negative rate would be $1/2$, and so the overall probability of error would be at least $1/2$. It follows that the minimum probability of error can be bounded from below by

$$\min_\tau \mathbb{P}\,(\mathcal{E}_\tau) \ge \min\left(1/2\,,\, \mathbb{P}(\cup_{i\notin\mathcal{S}}\{T_{i,2m} \ge \text{median}(T_{2m}|\theta_1)\}\right).$$

According to (3) the second argument above tends to 1 as $n\to\infty$, which completes the proof. ∎

B. Capability of Sequential Thresholding

Theorem III.3. *If (4) holds, then sequential thresholding is reliable if $K = (1+\epsilon)\log_2 n$, for $\epsilon > 0$. Specifically, if $\mathcal{E}_\epsilon$ is the error event $\{\mathcal{S}_K \ne \mathcal{S}\}$, then for any $\epsilon > 0$*

$$\lim_{n\to\infty} \mathbb{P}\,(\mathcal{E}_\epsilon) \;=\; 0.$$

Proof: The probability of error is

$$
\begin{aligned}
\mathbb{P}(\mathcal{E}_\epsilon) &:= \mathbb{P}\,(\mathcal{S}_K \ne \mathcal{S}) \\
&= \mathbb{P}\,(\{\mathcal{S}\cap\mathcal{S}_K^c \ne \emptyset\} \cup \{\mathcal{S}^c\cap\mathcal{S}_K \ne \emptyset\}) \\
&\le \mathbb{P}\,(\mathcal{S}\cap\mathcal{S}_K^c \ne \emptyset) + \mathbb{P}\,(\mathcal{S}^c\cap\mathcal{S}_K \ne \emptyset) \;, \tag{5}
\end{aligned}
$$

where the superscript c denotes the complementation of the set. The upper bound on the probability of error consists of two terms, the false-negative and false-positive probabilities. The false positive probability (second term in (5)) can be bounded as follows. Because we threshold at the median of the null distribution, approximately half of the null components survive each step.

$$
\begin{aligned}
&\mathbb{P}\,(\mathcal{S}^c\cap\mathcal{S}_K \ne \emptyset) \\
&= \mathbb{P}\left(\bigcup_{i\notin\mathcal{S}}\bigcap_{k=1}^{K}\left\{ T_{i,m}^{(k)} \ge \text{median}\,(T_m|\theta_0)\right\} \right) \\
&\le \sum_{i\notin\mathcal{S}}\left(\mathbb{P}\left(T_{i,m}^{(1)} \ge \text{median}\,(T_m|\theta_0)\right)\right)^K \\
&= \frac{n - |\mathcal{S}|}{2^K}
\end{aligned}
$$

Since $K = (1+\epsilon)\log_2 n$, with $\epsilon > 0$, we have

$$\lim_{n\to\infty} \mathbb{P}\,(\mathcal{S}^c\cap\mathcal{S}_K \ne \emptyset) = 0$$

Bounding the false-negative probability (first term in (5)) depends on the distribution of the test statistic under the alternative θ_1:

$$
\begin{aligned}
&\mathbb{P}\,(S\cap S_K^c \ne \emptyset) \\
&= \mathbb{P}\left(\bigcup_{k=1}^{K}\bigcup_{i\in\mathcal{S}}\left\{ T_{i,m}^{(k)} \le \text{median}\,(T_m|\theta_0)\right\} \right) \\
&= \mathbb{P}\left(\min_{k=1}^{K}\min_{i\in\mathcal{S}} T_{i,m}^{(k)} \le \text{median}\,(T_m|\theta_0) \right)
\end{aligned}
$$

which, from (4), goes to zero in the limit, which completes the proof. ∎

IV. APPLICATIONS

To illustrate the main results we consider three canonical settings arising in high-dimensional multiple testing. We again have in mind a sequence of problems and consider behavior in the high-dimensional limit. Thus, when we write $\theta \le g(n)$ (or $\theta \ge g(n)$) we mean that the parameter θ may (must) grow with dimension n no faster (slower) than the function $g(n)$. Throughout this section we let $s := |\mathcal{S}|$, the cardinality of the support set (which may also be considered to be a function of n).

A. Gaussian Model

Gaussian noise models are commonly assumed in multiple testing problems arising in the biological sciences (e.g., testing which of many genes or proteins are involved in a certain process or function). For example, a multistage testing procedure similar in spirit to sequential thresholding was used to determine genes important for virus replication in [4]. Consider a high dimensional hypothesis test in additive Gaussian noise where the parameter θ represents the mean of the distribution. We assume the null hypothesis follows zero mean ($\theta_0 = 0$), unit variance gaussian statistics; the alternate hypothesis, mean $\theta_1 > 0$, unit variance:

$$y_i \overset{iid}{\sim} \begin{cases} \mathcal{N}(0,1), & i \notin \mathcal{S} \\ \mathcal{N}(\theta_1,1), & i \in \mathcal{S}. \end{cases}$$

1) Non-Sequential Testing: We make $2m$ measurements of each element of θ. The test statistic again follows a normal distribution:

$$T_{i,2m} = \frac{1}{2m}\sum_{j=1}^{2m} y_{i,j} \overset{iid}{\sim} \begin{cases} \mathcal{N}(0,\frac{1}{2m}), & i \notin \mathcal{S} \\ \mathcal{N}(\theta_1,\frac{1}{2m}), & i \in \mathcal{S}. \end{cases} \quad (6)$$

Corollary IV.1. *If $\theta_1 < \sqrt{\frac{\log(n-s)}{m}}$, then the non-sequential testing procedure in (2) is unreliable, i.e., $\min_\tau \mathbb{P}(\mathcal{E}_\tau) \geq 1/2$.*

Proof: For the test statistic in equation (6), we satisfy (3) provided $\text{median}(T_{i,2m}|\theta_1) \leq \sqrt{\frac{\log(n-s)}{m}}$ (see, for example [5]). By Theorem III.2 and since $\text{median}(T_{i,2m}|\theta_1) = \theta_1$, if

$$\theta_1 \leq \sqrt{\frac{\log(n-s)}{m}}$$

then non-sequential thresholding is unreliable. ∎

2) Sequential Testing: Sequential thresholding makes m measurements of each component in the set $\mathcal{S}_k$ at each step. The test statistic follows a normal distribution:

$$T_{i,m}^{(k)} = \frac{1}{m}\sum_{j=1}^{m} y_{i,j} \overset{iid}{\sim} \begin{cases} \mathcal{N}(0,\frac{1}{m}) & i \notin \mathcal{S} \\ \mathcal{N}(\theta_1,\frac{1}{m}) & i \in \mathcal{S}. \end{cases} \quad (7)$$

Corollary IV.2. *If $\theta_1 > \sqrt{\frac{2}{m}\log(s\log_2 n)}$, then sequential thresholding will reliably recover $\mathcal{S}$.*

Proof: In this case, equation (4) is satisfied provided $\text{median}(T_m|\theta_0) \leq \theta_1 - \sqrt{\frac{2\log Ks}{m}}$ (see for example [5]). Since $\text{median}(T_m|\theta_0) = 0$, Theorem III.3 tells us that provided

$$\theta_1 \geq \sqrt{\frac{2}{m}\log Ks} \quad (8)$$

with $K = (1+\epsilon)\log_2 n$, we reliably recover $\mathcal{S}$. ∎

B. Gamma Model: Spectrum Sensing

Often termed *hole detection*, the objective of spectrum sensing is to identify unoccupied communication bands in the electromagnetic spectrum. Most of the bands will be occupied by *primary* users, but these users may come and go, leaving certain bands momentarily open and available for *secondary* users. Recent work in spectrum sensing has given considerable attention to such scenarios, including some work employing adaptive sensing methods (see, for example [6], [7]).

Following the notation throughout this paper, channel occupation is parameterized by θ, with θ_0 denoting the *signal plus noise* power in the occupied bands, and θ_1 representing the *noise only* power in the un-occupied bands. Without loss of generality, we let $\theta_1 = 1$. The statistics of a single measurement follow a complex Gaussian distribution – $y_i \overset{iid}{\sim} \mathcal{CN}(0,\theta)$. From Urkowitz's seminal work [8], making m measurements of each index, the likelihood ratio test statistic follows a Gamma distribution:

$$T_{i,m}^{(k)} = \sum_{j=1}^{m} |y_{i,\ell}|^2 \overset{iid}{\sim} \begin{cases} \text{Gamma}(m,\theta_0) & i \notin \mathcal{S} \\ \text{Gamma}(m,1) & i \in \mathcal{S}. \end{cases} \quad (9)$$

Remarkably, for this problem there exist constants $C, c > 0$ such that the sequential testing procedure is reliable if $\theta_0 \geq C\log(s\log_2 n)$, but the non-sequential testing procedure is unreliable if $\theta_0 \leq c(n-s)^{\frac{1}{2m}}$. To highlight this effect, if $s \sim \log n$, then the gap between these conditions is *doubly exponential* in n.

Since we are interested in detecting the sparse set of vacancies in the spectrum, our hypothesis test is reversed. We reject the null hypothesis (occupied component) if the test statistic falls *below* (rather than above) a certain threshold. In this case, the likelihood ratio is monotone non-increasing for $\theta_1 \leq \theta_0$, and so the inequalities in the key conditions (3) and (4) are reversed: specifically, the non-sequential thresholding procedure is *unreliable* if

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\min_{i\notin\mathcal{S}} T_{i,2m}}{\text{median}(T_{2m}|\theta_1)} \leq 1\right) = 1 \quad (10)$$

and sequential thresholding is *reliable* if

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\max_{k=1}^{K}\max_{i\in\mathcal{S}} T_{i,m}^{(k)}}{\text{median}(T_m|\theta_0)} \geq 1\right) = 0. \quad (11)$$

1) Non-Sequential Testing: In the non-sequential procedure (2), we make $2m$ measurements per index. The distribution of the test statistic follows a gamma distribution with shape parameter $2m$.

Corollary IV.3. *If $\theta_0 < 2(m-1)(n-s)^{\frac{1}{2m}}$, then the non-sequential procedure in (2) is unreliable.*

Proof: In this case, because the hypothesis test is reversed, we aim to satisfy (10). Since $\text{median}(T_{2m}|\theta_1) \geq 2(m-1)$, we have

$$\mathbb{P}\left(\frac{\min_{i\notin\mathcal{S}} T_{i,2m}}{\text{median}(T_{2m}|\theta_1)} \leq 1\right) \geq \mathbb{P}\left(\frac{\min_{i\notin\mathcal{S}} T_{i,2m}}{2(m-1)} \leq 1\right).$$

If $2(m-1) > \frac{\theta_0}{(n-s)^{\frac{1}{2m}}}$, we show in Appendix A that the right hand side above goes to 1 as n grows large. Together with Theorem III.2 this implies that if $\theta_0 < 2(m-1)(n-s)^{\frac{1}{2m}}$ then the non-sequential procedure is unreliable. ∎

2) Sequential Testing: Sequential thresholding makes m measurements of each component in the set $\mathcal{S}_k$ at each step. The test statistic follows the Gamma distributions in (9).

Corollary IV.4. *If $\theta_0 > \frac{\log(s \log_2 n)}{m}$, then sequential thresholding is reliable.*

Proof: It suffices to show (11) is satisfied. For all m and θ_0, we have $\mathrm{median}(T_m|\theta_0) \geq \theta_0(m-1)$. We upper bound (11) by

$$\lim_{n \to \infty} \mathbb{P}\left(\frac{\max_{k=1}^{K} \max_{i \in \mathcal{S}} T_{i,m}^{(k)}}{\theta_0(m-1)} \geq 1 \right)$$

which goes to zero in the limit provided $\theta_0(m-1) > \log Ks$ (see appendix B) . Together with Theorem III.3 if

$$\theta_0 > \frac{\log Ks}{m-1} \tag{12}$$

with $K = (1 + \epsilon)\log_2 n$, then sequential thresholding is reliable. ∎

C. Poisson Model: Photon-based Detection

Lastly we consider a situation in which the component distributions are Poisson. This model arises naturally in testing problems involving photon counting (e.g., optical communications or biological applications using fluorescent markers). We let the (sparse) alternative follow a Poisson with *fixed* rate θ_1, and the null hypothesis a rate θ_0, $\theta_0 > \theta_1$:

$$y_i \overset{iid}{\sim} \begin{cases} \mathrm{Poisson}(\theta_0) & i \notin \mathcal{S} \\ \mathrm{Poisson}(\theta_1) & i \in \mathcal{S} \end{cases},$$

Note that as $\theta_0 > \theta_1$, our hypothesis test is reversed as in the spectrum sensing example (and equations (10) and (11)).

The sufficient statistic for the likelihood ratio test is a sum of the individual measurements, again following a Poisson distribution. In this setting, the gap between sequential and non-sequential testing is similar to that of the Gaussian case. Proofs are left to Appendices C and D.

Corollary IV.5. *For any fixed θ_1, if $\theta_0 < \frac{\log(n-s)}{2m}$, non-sequential thresholding is unreliable.*

Corollary IV.6. *For any fixed θ_1, if $\theta_0 > \frac{\log(s \log_2 n)+1}{m}$, sequential thresholding is reliable.*

V. CONCLUSION

This paper studied the problem of high-dimensional testing and sparse recovery from the perspective of sequential analysis. The gap between the null parameter θ_0 and the alternative θ_1 plays a crucial role in this problem. We derived necessary conditions for reliable recovery in the non-sequential setting and contrasted them with sufficient conditions for reliable recovery using the proposed sequential testing procedure. Applications of the main results to several commonly encountered

models show that sequential testing can be exponentially (in dimension n) more sensitive to the difference between the null and alternative distributions, implying that subtle cases can be much more reliably determined using sequential methods.

REFERENCES

[1] J. Haupt, R. Castro, and R. Nowak, "Distilled sensing: Selective sampling for sparse signal recovery," *http://arxiv.org/abs/1001.5311*.
[2] ——, "Improved bounds for sparse recovery from adaptive measurements," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, 2010, pp. 1563 –1567.
[3] D. Siegmund, *Sequential Analysis*. New York, NY, USA: Springer-Verlag, 2010.
[4] L. Hao, A. Sakurai, T. Watanabe, E. Sorensen, C. Nidom, M. Newton, P. Ahlquist, and Y. Kawaoka, "Drosophila rnai screen identifies host genes important for influenza virus replication," *Nature*, pp. 890–3, 2008.
[5] M. R. Leadbetter, G. Lindgren, and H. Rootzen, *Extremes and Related Properties of Random Sequences and Processes*. Berlin: Springer, 1983.
[6] A. Tajer, R. Castro, and X. Wang, "Adaptive spectrum sensing for agile cognitive radios," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, 2010, pp. 2966 –2969.
[7] W. Zhang, A. Sadek, C. Shen, and S. Shellhammer, "Adaptive spectrum sensing," in *Information Theory and Applications Workshop (ITA), 2010*, 31 2010.
[8] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523 – 531, 1967.
[9] J. A. Gubner, *Probability and Random Processes for Electrical and Computer Engineers*. New York, NY, USA: Cambridge University Press, 2006.

A. Gamma Non-Sequential

The cumulative distribution function of $\mathrm{Gamma}(2m, \theta_0)$ is given as

$$F(\gamma) = 1 - e^{-\frac{\gamma}{\theta_0}} \sum_{\ell=0}^{2m-1} \left(\frac{\gamma}{\theta_0}\right)^{\ell} \frac{1}{\ell!}$$

hence,

$$\mathbb{P}\left(\frac{\min_{i \notin \mathcal{S}} T_{i,2m}}{\gamma} \leq 1\right) = 1 - \left(e^{-\frac{\gamma}{\theta_0}} \sum_{\ell=0}^{2m-1} \left(\frac{\gamma}{\theta_0}\right)^{\ell} \frac{1}{\ell!}\right)^{n-s}$$

Letting $\gamma = \frac{\theta_0}{(n-s)^{\frac{1}{2m}}}$ and taking the limit, it can be shown

$$\lim_{n \to \infty} 1 - \left(e^{-\left((n-s)^{-\frac{1}{2m}}\right)} \sum_{\ell=0}^{2m-1} \frac{(n-s)^{\frac{-\ell}{2m}}}{\ell!}\right)^{n-s}$$

$$= 1 - e^{-\frac{1}{(2m)!}}.$$

If $\gamma > \frac{\theta_0}{(n-s)^{\frac{1}{2m}}}$, then

$$\mathbb{P}\left(\frac{\min_{i \notin \mathcal{S}} T_{i,2m}}{\gamma} \leq 1\right) = 1.$$

B. Gamma Sequential

The cumulative distribution function of $\mathrm{Gamma}(m, 1)$ is given as

$$F(\gamma) = 1 - e^{-\gamma} \sum_{\ell=0}^{m-1} \frac{\gamma^{\ell}}{\ell!}$$

hence,

$$\mathbb{P}\left(\max_{k=1}^{K} \max_{i \in \mathcal{S}} T_{i,m}^{(k)} \geq \gamma\right) = 1 - \left(1 - e^{-\gamma} \sum_{\ell=0}^{m-1} \frac{\gamma^{\ell}}{\ell!}\right)^{Ks}.$$

Letting $\gamma = (1 + \epsilon) \log Ks$, for some $\epsilon > 0$, we have

$$\lim_{n \to \infty} 1 - \left(1 - \frac{1}{(Ks)^{1+\epsilon}} \sum_{\ell=0}^{m-1} \frac{((1+\epsilon) \log Ks)^{\ell}}{\ell!}\right)^{Ks} = 0 \ .$$

C. Poisson Non-Sequential

The likelihood ratio statistic is distributed as

$$T_{i,2m} = \sum_{j=1}^{2m} y_{i,j} \overset{iid}{\sim} \begin{cases} \mathrm{Poisson}(2m\theta_0) & i \notin \mathcal{S} \\ \mathrm{Poisson}(2m\theta_1) & i \in \mathcal{S}. \end{cases}$$

It suffices to show

$$\lim_{n \to \infty} \mathbb{P}\left(\frac{\min_{i \notin \mathcal{S}} T_{i,2m}}{\mathrm{median}(T_{2m} | \theta_1)} \leq 1\right) = 1.$$

for any $\theta_0 < \frac{\log(n-s)}{2m}$. The bound we derive is loose, but sufficient to show the adaptive scheme is superior. First, we assume that $\mathrm{median}(T_{2m} | \theta_1) > 0$. Next we have

$$\mathbb{P}\left(\min_{i \notin \mathcal{S}} T_{i,2m} \leq \mathrm{median}(T_{2m} | \theta_1)\right)$$

$$\geq \ \mathbb{P}\left(\min_{i \notin \mathcal{S}} T_{i,2m} = 0\right)$$

$$= 1 - \left(1 - e^{-2m\theta_0}\right)^{n-s} \ .$$

If $2m\theta_0 < \log(n-s)$, then

$$\lim_{n \to \infty} 1 - \left(1 - e^{-2m\theta_0}\right)^{n-s} = 1$$

which is also true provided $\theta_0 < \frac{\log(n-s)}{2m}$ and concludes the proof.

D. Poisson Sequential

In sequential thresholding, for each $i \in S_k$

$$T_{i,m}^{(k)} = \sum_{j=1}^{m} y_{i,j} \overset{iid}{\sim} \begin{cases} \mathrm{Poisson}(m\theta_0) & i \notin \mathcal{S} \\ \mathrm{Poisson}(m\theta_1) & i \in \mathcal{S}. \end{cases}$$

We need to show, for the test statistic above,

$$\lim_{n \to \infty} \mathbb{P}\left(\frac{\max_{k=1}^{K} \max_{i \in \mathcal{S}} T_{i,m}}{\mathrm{median}(T_m | \theta_0)} \geq 1\right) = 0.$$

First, we note $\mathrm{median}(T_m | \theta_0) \geq m\theta_0 - 1$. Hence,

$$\mathbb{P}\left(\max_{k=1}^{K} \max_{i \in \mathcal{S}} T_{i,m} \geq \mathrm{median}(T_m | \theta_0)\right)$$

$$\leq \mathbb{P}\left(\max_{k=1}^{K} \max_{i \in \mathcal{S}} T_{i,m} \geq m\theta_0 - 1\right).$$

We can bound the probability of a single event by Chernoff's bound [9], p.166. For $T_{i,m} \sim \mathrm{Possion}(m\theta_1)$ we have:

$$\mathbb{P}\left(T_{i,m} \geq \gamma\right) \ \leq \ e^{-m\theta_1 - \gamma\left(\log\left(\frac{\gamma}{m\theta_1}\right) - 1\right)}$$

$$\leq \ e^{-\gamma\left(\log\left(\frac{\gamma}{m\theta_1}\right) - 1\right)}.$$

which implies

$$\mathbb{P}\left(\max_{k=1}^{K} \max_{i \in \mathcal{S}} T_{i,m} \geq \gamma\right) \leq 1 - \left(1 - e^{-\gamma\left(\log\left(\frac{\gamma}{m\theta_1}\right) - 1\right)}\right)^{Ks}$$

Letting $\gamma = \log Ks$ and taking the limit as $n \to \infty$ of the expression above for any fixed θ_1, we conclude

$$\lim_{n \to \infty} \mathbb{P}\left(\frac{\max_{k=1}^{K} \max_{i \in \mathcal{S}} T_{i,m}}{\log Ks} \geq 1\right) = 0$$

Thus, if $\log Ks \leq m\theta_0 - 1$, or equivalently

$$\theta_0 \geq \frac{\log Ks + 1}{m},$$

sequential thresholding is reliable.