On the Limits of Sequential Testing in High Dimensions

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 presents results pertaining to sequential methods for support recovery of sparse signals in noise. Specifically, we show that *any* sequential measurement procedure fails provided the average number of measurements per dimension grows less then $D(f_0||f_1)^{-1}\log s$ where s is the level of sparsity, and $D(f_0||f_1)$ the Kullback-Leibler divergence between the underlying distributions. Moreover, we show that a simple procedure termed *sequential thresholding* guarantees exact support recovery provided the average number of measurements per dimension grows faster than $2D(f_0||f_1)^{-1}\log s$, a mere factor of 2 more than the lower bound. Lastly, we show any *non-sequential* procedure fails provided the number of measurements grows at a rate less than $D(f_1||f_0)^{-1}\log n$, where n is the total dimension of the problem.**

I. Introduction

High-dimensional signal support recovery is a fundamental problem arising in many aspects of science and engineering. The goal of the basic problem is to determine, based on noisy observations, a sparse set of elements that somehow differ from the others.

In this paper we study the following problem. Consider a support set $\mathcal{S} \subset \{1, ..., n\}$ and a random variable $y_{i,j}$ such that

$$y_{i,j} \sim \begin{cases} f_0(\cdot) & i \notin \mathcal{S} \\ f_1(\cdot) & i \in \mathcal{S} \end{cases} \tag{1}$$

where $f_0(\cdot)$ and $f_1(\cdot)$ are probability measures on $\mathcal{Y}$, and j indexes multiple measurements of any component $i \in \{1, ..., n\}$. The dimension of the problem, n, is large – perhaps thousands or millions or more – but the support set $\mathcal{S}$ is sparse in the sense that the number of elements following f_1 is much less than the dimension, i.e., $|\mathcal{S}| = s \ll n$. The goal of the sparse recovery problem is to identify the set $\mathcal{S}$.

In a *non-sequential* setting $m \geq 1$ independent observations of each component are made ($y_{i,1}, ..., y_{i,m}$ are observed for each i) and the fundamental limits for reliable recovery are readily characterized in terms of Kullback-Leibler divergence and dimension.

Sequential approaches to the high dimensional support recovery problem have been given much attention recently (see [1], [2], [3], etc). In the *sequential* setting, the decision to observe $y_{i,j}$ is based on prior observations, i.e. $y_{i,1}, ..., y_{i,j-1}$. Herein lies the advantage of sequential methods: if prior measurements indicate a particular component belongs (or doesn't belong) to $\mathcal{S}$ with sufficient certainty, measurement of that component can cease, and resources can be diverted to a more uncertain element.

The results presented in this paper are in terms of asymptotic rate at which the average number of measurements per dimension, denoted m, must increase with n to ensure exact recovery of $\mathcal{S}$ for any fixed distributions f_0 and f_1. The main contributions are *1)* to present a necessary condition for success of any sequential procedure in the sparse setting in terms of the average number of observations per dimension based on the known optimal *sequential probability ratio test*, *2)* show success of a simple sequential procedure first presented in [2] is guaranteed provided the average number of measurements per dimension is within a factor of 2 of the necessary condition for any sequential procedure, and *3)* lastly, compare these results to the performance of any non-sequential procedure. Table I summarizes these results.

TABLE I
AVERAGE NUMBER OF MEASUREMENTS PER DIMENSION FOR EXACT RECOVERY

Non-sequential	$m \geq \frac{\log n}{D(f_1\|\|f_0)}$	necessary
Sequential	$m \geq \frac{\log s}{D(f_0\|\|f_1)}$	necessary
Sequential Thresholding	$m > \frac{2\log s}{D(f_0\|\|f_1)}$	sufficient if $\log n < s \ll n$

Our results are striking primarily for two reasons. First, *sequential* procedures succeed when the number of measurements per dimension increases at a rate logarithmic in the level of *sparsity*, i.e. $\log s$. In contrast, *non-sequential* procedures require the average number of measurements per dimension to increase at a rate logarithmic in the *dimension*, i.e. $\log n$. For signals where sparsity is sublinear in dimension, the gains of sequential methods are polynomial; in scenarios where the sparsity grows logarithmically, the gains are exponential.

Secondly, and perhaps equally as surprising, a simple procedure dubbed sequential thresholding achieves nearly optimal performance provided minor constraints on the level of sparsity are met (specifically, that s is sublinear in n). In terms of the average number of measurements per dimension, the procedure comes within a mere factor of two of the lower bound of any sequential procedure.

II. Problem Formulation

Let $\mathcal{S}$ be a sparse subset of $\{1, ..., n\}$ with cardinality $s = |\mathcal{S}|$. For any index $i \in \{1, ..., n\}$, the random variable $y_{i,j}$ is independent, identically distributed according to (1). That is, for all j, $y_{i,j}$ follows distribution $f_1(\cdot)$ if i belongs to $\mathcal{S}$, and follows $f_0(\cdot)$ otherwise.

In this paper, we limit our analysis to exact recovery of $\mathcal{S}$ using coordinate wise methods. Letting $\hat{\mathcal{S}}$ be an estimate of $\mathcal{S}$, the family wise error rate is defined as:

$$\mathbb{P}(\mathcal{E}) = \mathbb{P}(\hat{\mathcal{S}} \neq \mathcal{S}) = \mathbb{P}\left(\bigcup_{i \notin \mathcal{S}} \mathcal{E}_i \cup \bigcup_{i \in \mathcal{S}} \mathcal{E}_i\right)$$

where $\mathcal{E}_i$, $i \notin \mathcal{S}$ is a false positive error event and $\mathcal{E}_i$, $i \in \mathcal{S}$ a false negative error event. To simplify notation, we define the false positive and false negative probabilities: $\alpha = \mathbb{P}(\mathcal{E}_i | i \notin \mathcal{S})$, and $\beta = \mathbb{P}(\mathcal{E}_i | i \in \mathcal{S})$.

The test to decide if component i belongs to $\mathcal{S}$ is based on the log-likelihood ratio. For y_j distributed i.i.d. f_0 or f_1,

$$t^{(m)} := \sum_{j=1}^{m} \log \frac{f_1(y_j)}{f_0(y_j)}$$

which is a function of $(y_1, ..., y_m) \in \mathcal{Y}^m$. The superscript m explicitly indicates the number of measurements used to form the likelihood ratio and is suppressed when unambiguous. The log-likelihood ratio is compared against a scalar threshold τ to hypothesize if a component follows f_0 or f_1:

$$t \gtrless_{f_0}^{f_1} \tau.$$

Additionally, the Kullback-Liebler divergence of distribution f_0 from f_1 is defined as:

$$D(f_1 || f_0) = \mathbb{E}_1\left[\log \frac{f_1(y)}{f_0(y)}\right]$$

where $\mathbb{E}_1[\cdot]$ is expectation with respect to distribution f_1.

A. Measurement procedures

To be precise in characterizing a measurement procedure, we continue with three definitions.

Definition II.1. Measurement Procedure. *A procedure, denoted π, used to determine if $y_{i,j}$ is measured. If $\pi_{i,j} = 1$, then $y_{i,j}$ is measured. Conversely, if $\pi_{i,j} = 0$, then $y_{i,j}$ is not measured.*

Definition II.2. Non-sequential measurement procedure. *Any measurement procedure π such that $\pi_{i,j}$ is not a function of $y_{i,j'}$ for any j'.*

Definition II.3. Sequential measurement procedure. *Any measurement procedure π where $\pi_{i,j}$ depends on prior measurements, specifically, $\pi_{i,j} : \{y_{i,1}, ..., y_{i,j-1}\} \mapsto \{0, 1\}$.*

B. Measurement Budget

In order to make fair comparison between measurement schemes, we limit the total number of observations of $y_{i,j}$ in expectation. For any procedure π, we require

$$\mathbb{E}\left[\sum_{i,j} \pi_{i,j}\right] \leq nm \tag{2}$$

for some integer m. This implies, on average, we use m or fewer observations per dimension.

III. Sequential Thresholding

Sequential thresholding, first presented in [2], relies on simple bisection idea. The procedure consists of a series of K measurement passes, where each pass eliminates from consideration approximately half of the components measured on the prior pass. After the last pass the procedure terminates and the remaining components are taken as the estimate of $\mathcal{S}$. Sequential thresholding is described in the following algorithm.

Sequential Thresholding

input: $K \approx \log n$ steps, threshold $\mu := \text{median}(t_i | i \notin \mathcal{S})$
initialize: $\boldsymbol{\pi}_{i,1} = \mathbf{1}$ for all i
for $k = 1, \ldots, K$ **do**
  for $\{i : \boldsymbol{\pi}_{i,k} = 1\}$ **do**
    measure: $t_i^{(m/2)}$
    threshold: $\boldsymbol{\pi}_{i,k+1} = \begin{cases} \mathbf{1} & t_i > \mu \\ \mathbf{0} & \text{else} \end{cases}$
  end for
end for
output: $\hat{\mathcal{S}} = \{i : \boldsymbol{\pi}_{i,K+1} = \mathbf{1}\}$

A. Example of Sequential Thresholding

Sequential thresholding is perhaps best illustrated by example. Consider a simple case where $m = 2$, $f_0 \sim \mathcal{N}(0, \sigma^2)$ and $f_1 \sim \mathcal{N}(\theta, \sigma^2)$ for some $\theta > 0$.

On the first pass, the procedures measures $y_{i,1}$ for all i, using n measurements (half of the total budget as $mn = 2n$). On subsequent passes, the procedure observes $y_{i,k}$ if $\pi_{i,k} = 1$. To set $\pi_{i,k+1}$ the procedure thresholds observations that fall below 0, the *median* of distribution f_0:

$$\pi_{i,k+1} = \begin{cases} 1 & y_{i,k} > 0 \\ 0 & y_{i,k} \leq 0. \end{cases}$$

That is, if a measurement of component i falls below the threshold on any pass, that component is *not* measured for the remainder of the procedure, and not included in the estimate of $\mathcal{S}$. After $K \approx \log n$ passes, the procedure terminates, and estimates $\mathcal{S}$ as the set of indices that have not been eliminated from consideration: $\hat{\mathcal{S}} = \{i : \pi_{i,K+1} = 1\}$.

B. Details of Sequential Thresholding

Sequential thresholding operates by making a series of K measurement passes. In general, on each pass, $m/2$ measurements of a subset of components are made, and the log-likelihood ratio $t_i^{(m/2)}$ is formed for each component. As measurements are made in blocks of size $m/2$, we use *boldface* $\boldsymbol{\pi}_{i,k}$ to indicate a block of measurements are taken of component i on the kth measurement pass. $\boldsymbol{\pi}_{i,k}$ can be interpreted as a vector:

$$\boldsymbol{\pi}_{i,k} = (\pi_{i,(k-1)m/2+1}, ..., \pi_{i,km/2}).$$

Sequential thresholding aims to eliminate approximately half of the components remaining on each pass from further consideration. To this end, we define the threshold μ to be the median of the log-likelihood ratio under the null distribution; i.e., μ satisfies

$$\mathbb{P}(t_i^{(m/2)} \geq \mu | i \notin \mathcal{S}) = \frac{1}{2}.$$

With μ and $K \approx \log n$ as inputs, sequential thresholding operates as follows. First, the procedure initializes, setting $\boldsymbol{\pi}_{i,1} = \mathbf{1}$. For passes $k = 1, ..., K$ the procedure measures $t_i^{(m/2)}$ if $\boldsymbol{\pi}_{i,k} = \mathbf{1}$. To set $\boldsymbol{\pi}_{i,k+1}$, the procedure tests the corresponding log-likelihood ratio against the threshold μ:

$$\boldsymbol{\pi}_{i,k+1} \quad = \quad \begin{cases} \mathbf{1} & \text{if} \quad t_i^{(m/2)} > \mu \\ \mathbf{0} & \text{else.} \end{cases}$$

That is, if $t_i^{(m/2)}$ is below μ, no further measurements of component i are taken. Otherwise, component i is measured on the subsequent pass. By definition of μ, approximately half of the components that follow f_0 will be eliminated on each pass; if $s \ll n$, thresholding eliminates approximately half of all remaining components.

After pass K, the procedure terminates and estimates $\mathcal{S}$ as the indices still under consideration: $\hat{\mathcal{S}} = \{i : \boldsymbol{\pi}_{i,K+1} = \mathbf{1}\}$.

C. Measurement Budget

Sequential thresholding satisfies the measurement budget in (2) provided s grows sublinearly with n. For brevity, we argue the procedure comes arbitrarily close to satisfying the measurement budget for large n:

$$\begin{aligned} \mathbb{E}\left[\sum_{i,j} \pi_{i,j}\right] \quad &\leq \quad \sum_{k=0}^{K-1} \left(\frac{(n-s)m/2}{2^k} + sm/2 \right) \\ &\leq \quad m(n-s) + msK. \end{aligned}$$

Letting $K = \log n$, in the limit, the procedure comes arbitrarily close to satisfying the constraint. To be rigorous in showing the procedure satisfies (2), m can be replaced by $m-1$, and the analysis throughout holds.

D. Ability of Sequential Thresholding

We present the first of the three main theorems of the paper to quantify the performance of sequential thresholding.

Theorem III.1. Ability of sequential thresholding. *Provided*

$$m \geq \frac{2\log(s\log n)}{D(f_0||f_1)} \tag{3}$$

sequential thresholding recovers $\mathcal{S}$ with high probability. More precisely, if

$$\lim_{n\to\infty} \frac{m}{\log(s\log n)} > \frac{2}{D(f_0||f_1)}$$

then, $\mathbb{P}(\mathcal{E}) \to 0$.

Proof: From a union bound on the family wise error rate, we have

$$\mathbb{P}(\mathcal{E}) \quad \leq \quad (n-s)\alpha + s\beta. \tag{4}$$

Employing sequential thresholding, from the definition of μ, $\alpha = \frac{1}{2^K}$ and

$$\begin{aligned} \beta \quad &= \quad \mathbb{P}\left(\bigcup_{k=1}^{K} t_i^{(m/2)} < \mu | i \in \mathcal{S} \right) \\ &\leq \quad Ks\mathbb{P}\left(t_i^{(m/2)} < \mu | i \in \mathcal{S} \right) \end{aligned}$$

where the inequality follows from a union bound.

We can further bound the false negative error event using Stein's Lemma [4], p. 384. Consider a test with a fixed probability of false positive at $\alpha = \frac{1}{2}$. By Stein's Lemma, the false negative probability is then given as

$$\mathbb{P}\left(t_i^{(m/2)} < \mu \mid i \in \mathcal{S} \right) \doteq e^{-\frac{m}{2} D(f_0||f_1)}$$

where $a \doteq e^{-mD}$ is equivalent to

$$\lim_{m\to\infty} \frac{1}{m} \log a = -D.$$

This implies, for any $\epsilon_1 > 0$, for sufficiently large m,

$$\mathbb{P}\left(t_i^{(m/2)} < \mu \mid i \in \mathcal{S} \right) \leq e^{-\frac{m}{2}(D(f_0||f_1)-\epsilon_1)}.$$

Letting $K = (1+\epsilon_2)\log_2 n$, for sufficiently large n and m, (4) becomes

$$\mathbb{P}(\mathcal{E}) \quad \leq \quad \frac{(n-s)}{n^{(1+\epsilon_2)}} + s(1+\epsilon_2)\log_2 n e^{-\frac{m}{2}(D(f_0||f_1)-\epsilon_1)}.$$

Hence, $\mathbb{P}(\mathcal{E})$ goes to zero provided

$$m \geq \frac{2\log((1+\epsilon_1)s\log_2 n)}{(D(f_0||f_1)-\epsilon_2)}$$

which, as ϵ_1 and ϵ_2 can be made arbitrarily small, directly gives the theorem:

$$m > \frac{2\log(s\log_2 n)}{D(f_0||f_1)}.$$

Moreover, provided $s > \log_2 n$, we have

$$m > \frac{2\log s}{D(f_0||f_1)}.$$

If s is sub-logarithmic in n, i.e., $s < \log_2 n$ we have

$$m > \frac{2 \log \log n}{D(f_0||f_1)}.$$

■

IV. Lower Bound on Sequential Performance: the SPRT

In this section we derive a lower bound on the rate at which m must grow with n for any sequential procedure based on the well studied sequential probability ratio test (SPRT).

A. The SPRT

A sequential probability ratio test operates by continuing to measure a component if the corresponding likelihood ratio is within an upper and lower boundary, and terminating measurement otherwise. Specifically, for two scalars A and B

$$\pi_{i,j+1} = \begin{cases} 1 & \text{if} \quad A < t_i^{(j)}(y_{i,1}, ..., y_{i,j}) < B \\ 0 & \text{else} \end{cases}$$

where $t_i^{(j)}$ is the log likelihood ratio comprised of *all* prior measurements (unlike sequential thresholding, in which the likelihood ratio is only formed using measurements from a single pass). If $t_i^{(j)} < A$, the SPRT labels index i as *not* belonging to $\mathcal{S}$, and if $t_i^{(j)} > B$, index i is assigned to $\mathcal{S}$. For a thorough discussion of the SPRT, see [5].

Sequential probability ratio tests are optimal for binary hypothesis tests in terms of minimum expected number of measurements for any error probabilities α and β (shown originally in [6]). To make the connection to the high dimensional problem, consider a single component i, and the corresponding binary hypothesis test. To simplify notation, define:

$$\mathbb{E}_0[N] = \mathbb{E}\left[\sum_j \pi_{i,j} | i \notin \mathcal{S}\right] \quad \mathbb{E}_1[N] = \mathbb{E}\left[\sum_j \pi_{i,j} | i \in \mathcal{S}\right]$$

that is, $\mathbb{E}_0[N]$ and $\mathbb{E}_1[N]$ are the expected number of measurements under f_0 and f_1 respectively. To be thorough, we restate the optimal property of the SPRT in the following lemma.

Lemma IV.1. Optimality of the SPRT for binary tests [7] (p.63). *Consider an SPRT with expected number of measurements $\mathbb{E}_0[N]$ and $\mathbb{E}_1[N]$, and corresponding error probabilities α and β. Any other sequential test with expected number of measurements $\mathbb{E}_0[N]'$ and $\mathbb{E}_1[N]'$ and error probabilities $\alpha' \leq \alpha$ and $\beta' \leq \beta$ will also have $\mathbb{E}_0[N]' \geq \mathbb{E}_0[N]$ and $\mathbb{E}_0[N]' \geq \mathbb{E}_0[N]$.*

In short, no procedure with the same error probabilities can have fewer measurements in expectation than the SPRT. To translate the optimality of the SPRT to the high dimensional case, we introduce the following lemma.

Lemma IV.2. Optimality of the SPRT. *Consider n component-wise sequential probability ratio tests used to estimate $\mathcal{S}$ with family wise error rate $\mathbb{P}(\mathcal{E})$ and expected number of measurements $\mathbb{E}[\sum_{i,j} \pi_{i,j}]$. Any other component wise test with family wise error rate $\mathbb{P}(\mathcal{E})' \leq \mathbb{P}(\mathcal{E})$ will also have expected number of measurements $\mathbb{E}[\sum_{i,j} \pi_{i,j}]' \geq \mathbb{E}[\sum_{i,j} \pi_{i,j}]$.*

Proof: First note that $\mathbb{P}(\mathcal{E})$ (see (6)) is monotonically increasing in both α and β. Moreover, we can write the total expected number of measurements as:

$$\mathbb{E}\left[\sum_{i,j} \pi_{i,j}\right] = (n-s)\mathbb{E}_0[N] + s\mathbb{E}_1[N]$$

which is monotonically increasing in both $\mathbb{E}_0[N]$ and $\mathbb{E}_1[N]$. Together with IV.1, this implies the lemma. ■

B. Limitation of any sequential procedure

Armed with the two lemmas above we are ready to state the theorem regarding the performance of any sequential measurement procedure.

Theorem IV.3. *Consider any sequential measurement procedure. Provided*

$$m < \frac{\log s}{D(f_0||f_1)}$$

the family wise error rate tends to one. More precisely, if

$$\lim_{n\to\infty} \frac{m}{\log s} < \frac{1}{D(f_0||f_1)} \tag{5}$$

then $\mathbb{P}(\mathcal{E}) \to 1$.

Proof: First, we show conditions under which the family wise error rate goes to one:

$$\begin{aligned} \mathbb{P}(\mathcal{E}) &= \mathbb{P}\left(\bigcup_{i\notin\mathcal{S}} \mathcal{E}_i \cup \bigcup_{i\in\mathcal{S}} \mathcal{E}_i\right) \\ &= 1 - \mathbb{P}\left(\bigcap_{i\notin\mathcal{S}} \mathcal{E}_i^c \cap \bigcap_{i\in\mathcal{S}} \mathcal{E}_i^c\right) \\ &= 1 - (1-\beta)^s(1-\alpha)^{n-s} \\ &\geq 1 - e^{-\beta s} e^{-\alpha(n-s)} \end{aligned} \tag{6}$$

which goes to one provided either

$$\alpha > \frac{1}{n-s} \qquad \beta > \frac{1}{s}. \tag{7}$$

Second, assume an SPRT with false positive and false negative probabilities α and β. From [5] p.21, we have

$$E_0[N] \geq \frac{1}{D(f_0||f_1)}\left(\alpha \log \frac{\alpha}{1-\beta} + (1-\alpha)\log\frac{1-\alpha}{\beta}\right)$$

which is derived from a simple argument using Jensen's inequality. The total expected number of measurements, constrained by the measurement budget, is

$$mn \geq E\left[\sum_{i,j} \pi_{i,j}\right] = (n-s)E_0[N] + sE_1[N] \tag{8}$$

By Lemma IV.2, no sequential test with the a smaller $\mathbb{P}(\mathcal{E})$ can have fewer measurements and satisfy (8). Thus, dropping the $sE_1[N]$ term from (8), if

$$mn \geq \frac{n-s}{D(f_0||f_1)}\left(\alpha \log \frac{\alpha}{1-\beta} + (1-\alpha)\log\frac{1-\alpha}{\beta}\right)$$

implies $\mathbb{P}(\mathcal{E}) \to 1$, any sequential procedures will also drive $\mathbb{P}(\mathcal{E}) \to 1$. Dividing by $n \log s$, the inequality becomes

$$\frac{m}{\log s} \geq \frac{n-s}{D(f_0||f_1) n \log s}\left(\alpha \log \frac{\alpha}{1-\beta} + (1-\alpha)\log\frac{1-\alpha}{\beta}\right).$$

Imposing the condition in (5) and cancelling $D(f_0||f_1)$ from both sides, we have

$$\lim_{n\to\infty} \frac{n-s}{n\log s}\left(\alpha \log \frac{\alpha}{1-\beta} + (1-\alpha)\log\frac{1-\alpha}{\beta}\right) < 1. \quad (9)$$

It is sufficient to show that (9) implies either $\alpha > \frac{1}{n-s}$ or $\beta > \frac{1}{s}$ in the high dimensional limit.

With this in mind, let $\beta = \frac{1-\epsilon_1}{s}$, and $\alpha = \frac{1-\epsilon_2}{n-s}$ for some $\epsilon_1, \epsilon_2 \in [0,1)$. Taking the limit as $n \to \infty$ in (9) and reducing terms we have:

$$\lim_{n\to\infty} (\cdot) = 1 \quad (10)$$

which contradicts (9), and negates our assumption that both $\beta = \frac{1-\epsilon_1}{s}$ and $\alpha = \frac{1-\epsilon_2}{n-s}$ for $\epsilon_1, \epsilon_2 \in [0,1)$. Hence, by (7), the family wise error rate must go to one, completing the proof. ∎

C. Comparison of the SPRT to Sequential Thresholding

Although a fully rigorous proof is quite involved, using standard approximations for the sequential probability ratio test (again, see [5]) it is relatively straightforward to show the SPRT does achieve the lower bound presented above.

Sequential thresholding is similar in spirit to the SPRT. In many scenarios, however, implementing the SPRT can be substantially more complicated, if not infeasible, when compared to sequential thresholding. To set the stopping boundaries, an SPRT requires knowledge of the underlying distributions as well as the level of sparsity s. Even when these are available, only approximations relating error probabilities to the stopping boundaries can be derived in closed-form.

On the contrary, sequential thresholding does not require knowledge of s. Since its sample requirements are within a factor of 2 of the lower bound, sequential thresholding is automatically *adaptive* to unknown levels of sparsity. Moreover, in practice, sequential thresholding needs only approximate knowledge of the distributions to operate (such that a substantial number of components that follow f_0 can be eliminated on each pass).

V. Limitation of Non-Sequential Methods

Our analysis would not be complete with out comparison of sequential thresholding and the lower bound based on the SPRT to non-sequential methods. To do so, we analyze the performance of any non-sequential method using Stein's lemma.

Theorem V.1. Limitation of non-sequential testing. *Consider any non-sequential thresholding procedure. Provided*

$$m < \frac{\log n}{D(f_1||f_0)} \quad (11)$$

the family wise error rate goes to 1. To be precise, (11) is equivalent to

$$\lim_{n\to\infty} \frac{m}{\log n} < \frac{1}{D(f_1||f_0)}.$$

Proof: From [4], p. 386, (Stein's Lemma) and by (7) we require

$$\alpha \doteq e^{-mD(f_\lambda||f_0)} > \frac{1}{n-s}$$

and

$$\beta \doteq e^{-mD(f_\lambda||f_1)} > \frac{1}{s}$$

where

$$f_\lambda = \frac{f_0^\lambda f_1^{1-\lambda}}{\int_\Omega f_0^\lambda f_1^{1-\lambda} dy}$$

for $\lambda \in [0,1]$. Hence, any sequential procedure fails provided

$$m \quad < \quad \min_{\lambda\in[0,1]} \max\left(\frac{\log(n-s)}{D(f_\lambda||f_0)}, \frac{\log s}{D(f_\lambda||f_1)}\right)$$

which is implied if

$$m \quad < \quad \frac{\log n}{D(f_1||f_0)}.$$

∎

VI. Conclusion

This paper showed sequential methods for support recovery of high dimensional sparse signals in noise can succeed using far fewer measurements than non-sequential methods. Specifically, non-sequential methods require the number of measurements to grow logarithmically with the dimension, while sequential methods succeed if the number of measurements grows logarithmically with the level of sparsity. Additionally, a simple procedure termed sequential thresholding comes within a factor of two of optimal in terms of number of measurements per dimension.

References

[1] J. Haupt, R. Castro, and R. Nowak, "Distilled sensing: Selective sampling for sparse signal recovery," *http://arxiv.org/abs/1001.5311*.

[2] M. Malloy and R. Nowak, "Sequential analysis in high dimensional multiple testing and sparse recovery," *Accepted to ISIT 2011, http://arxiv.org/abs/1103.5991*, March 2011.

[3] V. Srivastava, K. Plarre, and F. Bullo, "Randomized sensor selection in sequential hypothesis testing," *Signal Processing, IEEE Transactions on*, vol. 59, no. 5, pp. 2342 –2354, may 2011.

[4] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. New York, NY, USA: Wiley-Interscience, 2005.

[5] D. Siegmund, *Sequential Analysis*. New York, NY, USA: Springer-Verlag, 2010.

[6] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *The Annals of Mathematical Statistics*, vol. 19, no. 3, pp. 326–339, 1948.

[7] H. Chernoff, *Sequential Analysis and Optimal Design*. Bristol, England: Society for Industrial and Applied Mathematics, 1979.