Adaptive Discovery of Sparse Signals in Noise

Jarvis Haupt
Dept. of Elec. and Comp. Engr.
University of Wisconsin – Madison
Madison, WI 53706
Email: jdhaupt@wisc.edu

Rui Castro
Department of Electrical Engineering
Columbia University
New York, NY 10027
Email: rmcastro@ee.columbia.edu

Robert Nowak
Dept. of Elec. and Comp. Engr.
University of Wisconsin – Madison
Madison, WI 53706
Email: nowak@ece.wisc.edu

***Abstract*—A multi-step adaptive resampling procedure is proposed, and shown to be an effective approach when detecting high-dimensional sparse signals in noise. Each step of the proposed procedure refines an estimate of the true signal subspace, allowing sensing energy to be focused more directly into the subspace of interest and significantly improving the performance of the final detection test. Large-sample analysis shows that for the sparse signal detection problems considered, the proposed adaptive sensing procedure outperforms the best possible detection methods based on non-adaptive sensing, allowing for the detection of signals that are exponentially weaker than what can be detected using non-adaptive samples.**

I. Introduction

In many high-dimensional inference problems, the task is to detect the presence of a signal from noisy measurements. For example, in sensor network monitoring applications, a large number of sensors might be deployed over a geographical region with the goal of determining whether a signal (perhaps a chemical or biological agent) is present anywhere in the region. Other applications in microarray analysis, astronomy, and covert communications are also easy to envision. Often, signals of interest in these applications are sparse (i.e., they exist in a very low-dimensional, but unknown, subspace).

Natural questions emerge regarding the performance of any detection method in such settings. For example, one is often concerned with the trade-offs among various problem parameters—namely, the sparsity level (signal subspace dimension), the amplitudes of each nonzero component, and the observation noise power. Consider the following n-dimensional signal plus noise observation model:

$$Y_i = x_i + z_i = \mu\, s_i + Z_i\,, \quad i = 1, \ldots, n, \qquad (1)$$

where $Z_i \sim \mathcal{N}(0,1)$ denotes a Gaussian distribution with mean 0 and variance 1, the scalar μ denotes the signal strength and $s = [s_1, \ldots, s_n]$ is a binary vector whose nonzero entries indicate the signal subspace. In this setting, the detection problem amounts to determining whether $\mu \neq 0$. A simple detection scheme for this problem is based on linear data fusion. The data are averaged to obtain the test statistic $T = \frac{1}{n}\sum_{i=1}^{n} Y_i$, which is then compared to a threshold. Note that $\sqrt{n}T \sim \mathcal{N}(\mu d_n/\sqrt{n}, 1)$, where $d_n = \sum_{i=1}^{n} s_i$ (the signal subspace dimension) describes the relative the sparsity level. If $d_n/\sqrt{n}$ is bounded away from zero when n tends to infinity (and μ is nonvanishing), then both the false-alarm and miss probabilities can be driven to zero using an appropriate threshold.

Note that this simple linear fusion approach can break down if the signal subspace dimension grows sublinearly in n. If, as $n \to \infty$, $d_n/\sqrt{n} \to 0$, then for any fixed μ, $\sqrt{n}T \sim \mathcal{N}(\mu d_n/\sqrt{n}, 1) \to \mathcal{N}(0,1)$, implying that for any threshold the probability of miss does not tend to 0. An alternate approach is to test whether $\max_i Y_i$ exceeds a threshold. If $\mu = 0$, then in the large-sample limit $\max_i Y_i \leq \sqrt{2\log n}$ with probability 1, implying signals can be reliably detected when $\mu > \sqrt{2\log n}$. The limits of detectability using observations collected under the model (1) were established in [1], [2], where it was shown that more subtle testing procedures based on order statistics succeed at detecting slightly weaker signals in this setting ($\mu > \sqrt{2r\log n}$ for some $r < 1$).

The limits of detectability established in [1], [2] were based on the assumption that the observations of the signal are collected *non-adaptively*. This implies sensing/sampling resources must be equally allocated over the entire observation space (since one does not know the signal subspace a priori). In contrast, here we consider an adaptive procedure that utilizes feedback in sensing. Formally, adaptive sampling approaches allow the sampling procedure to be adjusted using previously collected samples. In the sparse signal detection problems considered here, adaptivity could allow subsequent samples to be focused more directly into subspaces where the signal of interest, if present, is most likely to exist, increasing the measurement signal to noise ratio (SNR) and ultimately improving the performance of any detection method.

In this paper we propose a multi-step adaptive resampling procedure that delivers such an improvement. In one setting we show that our adaptive sampling procedure (using a fixed number of resampling steps) effectively improves the range of problem parameters for which sparse signals are detectable. By letting the number of resampling steps tend to infinity (slowly as a function of the signal dimension n) we also establish that the proposed procedure succeeds in detecting signals that are *exponentially weaker* than any signal that can be detected using non-adaptive observations.

The remainder of this paper is organized as follows. In Section II we formalize the sparse signal and observation models that will be employed throughout. Known results on detectability using non-adaptive observations are summarized in Section III. We propose our adaptive sampling procedure

in Section IV, and describe its performance benefit relative to non-adaptive sampling in Section V. A few brief conclusions are drawn in Section VI. Proofs of the main results are relegated to the Appendix.

II. Signal and Observation Models

Throughout this paper we will be concerned with the detection of signals $x \in \mathbb{R}^n$ which exhibit sublinear sparsity, and for simplicity we restrict our attention to signals in which each nonzero entry has the same amplitude. We will assume that the signals of interest are generated randomly, according to the following model. For a fixed $\beta \in (1/2, 1)$, we let $\epsilon(n) = n^{-\beta}$ and

$$s_i = \begin{cases} 1 & \text{with probability } \epsilon(n) \\ 0 & \text{with probability } 1-\epsilon(n) \end{cases}, \; i = 1, 2, \ldots, n, \tag{2}$$

and we let $x_i = \mu s_i$ for some scalar amplitude μ. The random signal model is chosen primarily to facilitate comparison between our adaptive procedure and prior work that uses non-adaptive observation models. Note, however, that for large n, signals so generated can be identified as having $m(n) \approx n^{1-\beta}$ nonzero entries with very high probability. This claim is easily formalized using standard concentration inequalities.

We assume that observations of x come from multiple "looks," indexed by j, of the form

$$Y^{(j)} = \phi^{(j)} \cdot x \; + \; Z^{(j)} \; , \tag{3}$$

where each $\phi^{(j)} \in \mathbb{R}^n$ is a *sensing vector* with non-negative entries, $Z^{(j)} \overset{iid}{\sim} \mathcal{N}(0, 1_{n\times n})$, and the operation $\phi^{(j)} \cdot x$ denotes the Hadamard (or element-wise) product of the vectors. In addition, we impose the restriction $\sum_j \|\phi^{(j)}\|_2^2 = n$, limiting the total energy of the sensing vectors.

Using this model, an example of non-adaptive sampling arises when only one look at the signal is obtained. In this case, (3) can be written as

$$Y^{(1)} = Y = \phi \cdot x \; + \; Z \; , \tag{4}$$

and to satisfy the sensing energy condition, we can choose $\phi_i^{(1)} = \phi_i = 1$ for $i = 1, \ldots, n$. Another possibility is to make multiple iid observations, but each with only a fraction of the total sensing energy budget. For example, let $p \in \mathbb{N}$ denote the total number of looks, and let $\phi_i^{(j)} = 1/\sqrt{p}$, $i = 1, \ldots, n$, for $j = 1, \ldots, p$. Because of the independence of the noises, $\sum_{j=1}^p Y^{(j)}$ is statistically equivalent to Y in the single-observation model (4) as well. There are obviously many other choices of $\{\phi^{(j)}\}_j$ that yield the same result. Furthermore, no non-adaptive sensing scheme exists that can produce better results than those obtained using observations from the standard model (4). Therefore, we are interested here in adaptive designs of $\{\phi^{(j)}\}_j$ that tend to focus in on the non-zero components of x. Specifically, in what follows we will allow $\phi^{(j)}$ to depend explicitly on $\{\phi^{(\ell)}, Y^{(\ell)}\}_{\ell<j}$, and we will show that information gleaned from previous observations can be used to effectively guide future sampling, yielding significant performance improvements.

III. Large-Sample Detection Thresholds from Non-Adaptive Observations

In [1], a fundamental large sample detectability threshold for sparse signals using non-adaptive observations was established. Let the fraction of nonzero entries be denoted by $\epsilon(n)$, and the amplitude of each be $\mu(n)$. When samples are drawn according to (4), the detection problem amounts to a hypothesis test between a joint null distribution and an alternative mixture distribution:

$$\begin{array}{ll} H_0: & Y_i \overset{iid}{\sim} \mathcal{N}(0,1) \\ H_1: & Y_i \overset{iid}{\sim} (1-\epsilon(n))\mathcal{N}(0,1) + \epsilon(n)\mathcal{N}(\mu(n),1) \end{array} ,$$

for $i = 1, \ldots, n$. Carefully specifying how $\epsilon(n)$ and $\mu(n)$ evolve with n allows for a concise statement about the detectability threshold behavior. Specifically, for a fixed $\beta \in (1/2, 1)$ let the fraction of nonzero entries be given by $\epsilon(n) = n^{-\beta}$. In addition, suppose that the amplitude of each nonzero entry is $\mu(n) = \sqrt{2r \log n}$ for some $r \in (0, 1)$. The following result holds [1].

Theorem 3.1: Let

$$\rho^*(\beta) = \begin{cases} \beta - \frac{1}{2}, & \frac{1}{2} < \beta \leq \frac{3}{4} \\ \left(1 - \sqrt{1-\beta}\right)^2, & \frac{3}{4} < \beta < 1 \end{cases} .$$

When $r > \rho^*(\beta)$, then there exists a test for which the sum of Type I and Type II errors (the false alarm and miss errors, respectively) tends to zero as $n \to \infty$. Conversely, when $r < \rho^*(\beta)$, then the sum of Type I and Type II errors for any *test tends to one as $n \to \infty$.*

Of course, the likelihood ratio test is one such test that achieves the performance described above. However, successfully implementing the test requires knowledge of the problem parameters r and β. In [1] the authors proposed an alternative procedure called *Higher Criticism* that utilizes properties of the order statistics of the observed data to achieve the same performance as the likelihood ratio test but without requiring knowledge of the problem parameters.

IV. Adaptive Testing and Resampling

In order to improve upon the threshold behavior specified in Theorem 3.1, we consider sampling methods that allow feedback. The key to our adaptive approach is a multi-step testing and resampling procedure, described as Algorithm 1, which proceeds as follows. Initially, all locations of x are measured, but with only a fraction of the total sensing energy budget arising from an equal allocation of sensing energy among all observation steps. Each observation takes the form (3), where the sensing energy allocated to that step is distributed equally among a subset of promising entries of x. Following each of the first k observation steps, a refinement test is performed, which identifies the subset of the promising locations where the current observation is positive. The rationale is that it is highly improbable that the signal (which is assumed to be positive) is present at locations where the current observation is negative. The algorithm terminates after the final observation, and the output consists of the final collection of observations

and a set of possible signal component locations. A pseudocode description of the procedure appears as Algorithm 1.

Algorithm 1: Adaptive Testing and Resampling.

Input:
Number of refinements k;

Initialize:
Initial index set $I^{(1)} \longleftarrow \{1, 2, \ldots, n\}$;
Energy per observation $\mathcal{E} = n/(k+1)$;

Refinement:
for $j = 1$ **to** k **do**

$$Y_i^{(j)} = \left\{ \begin{array}{ll} \sqrt{\frac{\mathcal{E}}{|I^{(j)}|}} x_i + Z_i^{(j)}, & i \in I^{(j)} \\ Z_i^{(j)}, & i \in I^{(1)} \setminus I^{(j)} \end{array} \right\};$$

$I^{(j+1)} \longleftarrow \{i \in I^{(j)} : Y_i^{(j)} > 0\}$;
end

Final Observations:

$$Y_i^{(k+1)} = \left\{ \begin{array}{ll} \sqrt{\frac{\mathcal{E}}{|I^{(k+1)}|}} x_i + Z_i^{(k+1)}, & i \in I^{(k+1)} \\ Z_i^{(k+1)}, & i \in I^{(1)} \setminus I^{(k+1)} \end{array} \right\};$$

Output:
Final index set $I^{(k+1)}$;
Refined obs. $Y^{(k+1)} := \{Y_i^{(k+1)} : i \in I^{(k+1)}\}$;

To quantify the performance of this algorithm, we will show that the refinement test at each step retains *most* (in fact all, in the limit as n tends to infinity) of the indices corresponding to nonzero signal components, but only about *half* of the indices corresponding to zero entries. When the signal is sparse, this implies that the effective dimension is roughly halved at each step, allowing about twice as much sensing energy to be allocated to the signal subspace in the next step. Applying Higher Criticism testing to the output observations results in significantly lower large-sample thresholds for consistent detectability, as described in the next section.

V. Main Results

The large-sample threshold for consistent sparse signal detection based on the proposed adaptive resampling procedure is examined in two settings; one where the number of observations, $k+1$, is finite, and another where k is allowed to tend to infinity slowly as a function of the dimension n. The results are stated in this section. The proofs are relegated to the Appendix.

Our first main result quantifies the performance of adaptive resampling followed by Higher Criticism testing when the number of refinement steps is fixed. The result is an expanded region of large-sample detectability in the (β, r) plane.

Theorem 5.1: For a fixed $\beta \in (1/2, 1)$, consider sparse signals generated according to the model (2)*, and suppose that the amplitude of each nonzero entry is given by $\mu(n) = \sqrt{2r \log n}$ for some $r \in (0, 1)$. There exists an improved threshold function,*

$$\begin{aligned} \rho_a^*(\beta, k) &= \left(\frac{k+1}{2^k}\right) \rho^*(\beta) \\ &= \left\{ \begin{array}{ll} \frac{k+1}{2^k}\left(\beta - \frac{1}{2}\right), & \frac{1}{2} < \beta \leq \frac{3}{4} \\ \frac{k+1}{2^k}\left(1 - \sqrt{1-\beta}\right)^2, & \frac{3}{4} < \beta < 1 \end{array} \right. , \end{aligned}$$

such that when $r > \rho_a^*(\beta, k)$, Higher Criticism testing applied to $Y^{(k+1)}$ (the output of the adaptive procedure with k refinement testing steps) succeeds in detecting the sparse signal in the sense that the sum of Type I and Type II errors tends to zero as $n \to \infty$.

This result shows that even a small number of refinement steps leads to a significant improvement in terms of the large-sample threshold for consistent recovery (since $(k+1)/2^k \ll 1$ for even modest values of k). Figure 1 shows the large-sample threshold of Theorem 3.1 (solid line) along with several of the improved thresholds for various values of k (the threshold for $k = 3$ is shown as a dotted line, and the threshold for $k = 7$ is the dashed line).

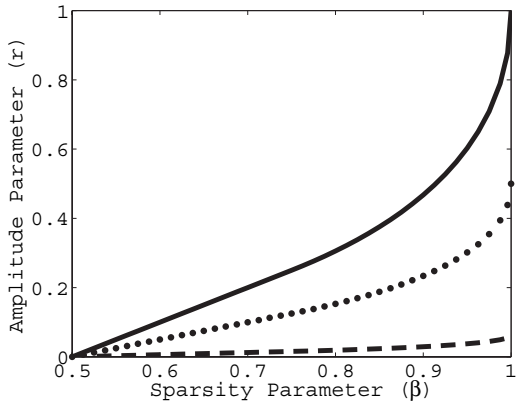

Fig. 1. Large-Sample Thresholds for Consistent Detection. Signals whose sparsity and amplitude parameters are above the thresholds can be reliably detected. The solid line shows the threshold of Theorem 3.1 obtained using observations from the standard model (4), while the thresholds obtained using the adaptive procedure (for $k = 3$ and $k = 7$ refinement steps) are shown as dotted and dashed lines, respectively.

Note that the result of Theorem 5.1 suggests increasing the number of refinement steps results in increasingly improved recovery thresholds. A natural question arises as to whether the number of observation steps can be large enough to detect signals whose amplitudes grow more slowly (as a function of the dimension n) than $\mu(n) = \sqrt{2r \log n}$. We address this question here by letting the number of observation steps tend to infinity slowly as a function of n. The result is the following theorem.

Theorem 5.2: For a fixed $\beta \in (1/2, 1)$, consider sparse signals generated according to the model (2)*, and suppose that for a fixed $\delta > 0$, the amplitude of each nonzero entry is*

given by $\mu(n) = \sqrt{2r\left(\log_2 \log n\right)^{1+\delta}}$ *for some* $r \in (0,1)$. *Let the number of refinement steps be given by* $k(n) = \log_2 \log n$. *When* $r > \rho^*(\beta)$ *(as defined in Theorem 3.1), Higher Criticism testing applied to* $Y^{(k+1)}$ *succeeds in detecting the sparse signal in the sense that the sum of Type I and Type II errors tends to zero as* $n \to \infty$.

In other words, the adaptive procedure followed by Higher Criticism testing succeeds at detecting sparse signals with non-zero amplitudes as small as $\sqrt{2r(\log_2 \log n)^{1+\delta}}$, for δ arbitrarily small, but fixed. For the same specification on r, non-adaptive sensing can only recover signals with amplitudes larger than $\sqrt{2r \log n}$. Thus, the proposed adaptive procedure is capable of detecting sparse signals that are essentially *exponentially weaker* than those recoverable using non-adaptive sensing.

VI. Discussion and Conclusions

While we only considered the case where the amplitudes of the nonzero signal entries were equal and positive, the proposed adaptive procedure could also be applied to more general classes of signals, such as those for which the signal has both positive and negative values. In this case, one approach is to split the budget of sensing energy in half, and execute the procedure once assuming the nonzero entries are positive (keeping all entries that exceed zero at each step, as described above), and again assuming the entries are negative (retaining indices at each for which the corresponding observation is negative). The final sets of observations could then be independently subjected to Higher Criticism testing.

In general, we have shown that a simple adaptive procedure provably outperforms all non-adaptive sampling techniques when detecting sparse signals in additive noise. We note that analogous improvements can also be realized using the same procedure in estimation problems, and this setting will be treated in a future work.

VII. Appendix

We first state two lemmata that will be useful in the proofs. The first describes a tail bound for the Binomial distribution, which follows from [3].

Lemma 7.1: Let B be a Binomial(n,p) random variable, and assume that $b < \mathbb{E}[B] = np$. Then,

$$\Pr\left(B \leq b\right) \leq \left(\frac{n-np}{n-b}\right)^{n-b} \left(\frac{np}{b}\right)^{b}.$$

We will also need to bound tail probabilities for certain Gaussian random variables. For that, we will utilize the following standard result (see, for example, [4]).

Lemma 7.2: Let $Z \sim \mathcal{N}(z,1)$ for $z > 0$. Then,

$$\Pr\left(Z < 0\right) \leq \frac{1}{z\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right).$$

The proofs of both main theorems are similar in nature. To save space we formally present the proof of Theorem 5.1, then briefly describe the generalizations that are needed to establish Theorem 5.2.

A. Proof of Theorem 5.1

We first start with a lemma that quantifies the effect of each refinement step. Recall that $I^{(j)}$ is the index set of iteration j (refer to Algorithm 1 for details).

Lemma 7.3: Let ℓ_j *and* m_j *denote, respectively, the number of indices corresponding to zero and non-zero elements of* x *in* $I^{(j)}$. *Conditioning on* m_j *and* ℓ_j *we have that, for* n *sufficiently large,*

$$m_{j+1} \;\geq\; \left(1 - \frac{1}{\log n}\right) m_j, \tag{5}$$

and

$$\left(\frac{1}{2} - \frac{1}{\log n}\right)\ell_j \;\leq\; \ell_{j+1} \;\leq\; \left(\frac{1}{2} + \frac{1}{\log n}\right)\ell_j, \tag{6}$$

hold simultaneously with probability at least $1-\xi$, *where* $\xi = \exp\left(-n^{-\alpha} m_j\right) + 2\exp\left(-2\ell_j/\left(\log n\right)^2\right)$, *for any* $\alpha > 0$.

Proof: Note that in any step of the $(k+1)$-step procedure, the observed amplitude of non-zero signal components is no less than $\sqrt{2r \log n/(k+1)}$. It suffices to consider this worst case to establish a general bound.

Let $T_i = \mathbf{1}\{Y_i^{(j)} > 0\}$, $i \in I^{(j)}$, be the indicators of the indices retained in $I^{(j+1)}$. Let $\mathcal{I}_s \subseteq I^{(j)}$ be the collection of indices corresponding to non-zero components (therefore $|\mathcal{I}_s| = m_j$). From Lemma 7.2, we have for $i \in \mathcal{I}_s$

$$\Pr\left(T_i = 0\right) \leq \sqrt{\frac{k+1}{4\pi r \log n}}\, n^{-r/(k+1)}. \tag{7}$$

The number of signal components retained after the refinement step is simply given by $m_{j+1} = \sum_{i \in \mathcal{I}_s} T_i$. Let γ be a small quantity that satisfies $\gamma > \Pr\left(T_i = 0\right)$, and apply Lemma 7.1 to obtain

$$\begin{aligned}
&\Pr\left(m_{j+1} \leq (1-\gamma)m_j \,|\, m_j\right) \\
&\quad\leq \left(\frac{\Pr\left(T_i = 0\right)}{\gamma}\right)^{\gamma m_j} \left(\frac{\Pr\left(T_i = 1\right)}{1-\gamma}\right)^{(1-\gamma)m_j} \\
&\quad\leq \left(\frac{\Pr\left(T_i = 0\right)}{\gamma}\right)^{\gamma m_j} \left(\frac{1}{1-\gamma}\right)^{(1-\gamma)m_j}.
\end{aligned}$$

Now, for $\alpha > 0$, notice that if

$$\gamma \log\left(\frac{\Pr\left(T_i = 0\right)}{\gamma}\right) + (1-\gamma)\log\left(\frac{1}{1-\gamma}\right) \leq -n^{-\alpha} \tag{8}$$

then the probability that (5) does not hold is given by

$$\Pr\left(m_{j+1} \leq (1-\gamma)m_j \,|\, m_j\right) \leq \exp\left(-n^{-\alpha} m_j\right).$$

Let $\gamma = \left(\log n\right)^{-1}$. We will now check that (8) holds, using (7) and the fact that

$$\left(\frac{1-\gamma}{\gamma}\right)\log\left(\frac{1}{1-\gamma}\right) \leq 1$$

for any $\gamma \in (0,1)$. After trivial manipulation we see that (8) is satisfied if

$$\begin{aligned}
&\frac{-r\log n}{k+1} + \frac{\log\log n}{2} + \frac{\log n}{n^{\alpha}} \\
&\qquad + \left(1 + \frac{1}{2}\log\left(\frac{k+1}{4\pi r}\right)\right) \leq 0,
\end{aligned}$$

which is clearly the case for n sufficiently large, since the first term dominates the second, the third decays to zero, and the remainder does not depend on n.

To verify the inequalities in (6), define $\mathcal{I}_0 \subseteq I^{(j)}$ to be the collection of indices corresponding to zero entries (therefore $|\mathcal{I}_0| = \ell_j$). Note that for $i \in \mathcal{I}_0$, T_i is a Bernoulli random variable with parameter $1/2$ (since each corresponding observation $Y_i^{(j)}$ is a zero-mean Gaussian random variable). Applying Hoeffding's inequality to $\ell_{j+1} = \sum_{i \in \mathcal{I}_0} T_i$ we obtain

$$\Pr\left(\left|\ell_{j+1} - \frac{\ell_j}{2}\right| > \gamma \ell_j \middle| \ell_j\right) \leq 2\exp\left(-2\ell_j\gamma^2\right).$$

The probability that both claimed conditions hold simultaneously is obtained by applying the union bound, concluding the proof. ∎

The proof of Theorem 5.1 proceeds by iterating the result of Lemma 7.3. We begin by considering the set $I^{(2)}$. Let $\alpha = (1-\beta)/2$, and note that with probability at least $1 - \xi(2)$, where

$$\xi(2) = \exp\left(-n^{(1-\beta)/2}\right) + 2\exp\left(-\frac{2n(1-n^{-\beta})}{(\log n)^2}\right),$$

the event

$$E_2 = \left\{ \begin{array}{l} \left(1 - \frac{1}{\log n}\right) n^{1-\beta} \leq m_2 \leq n^{1-\beta} \\ \\ \left(\frac{1}{2} - \frac{1}{\log n}\right) n(1-n^{-\beta}) \leq \ell_2 \\ \\ \ell_2 \leq \left(\frac{1}{2} + \frac{1}{\log n}\right) n(1-n^{-\beta}) \end{array} \right\}$$

holds when n is sufficiently large.

Now, conditioned on event E_2, we can apply Lemma 7.3 to $I^{(3)}$. Proceeding similarly for sets $I^{(4)}, \ldots, I^{(k+1)}$, and noting that $\xi(i) < \xi(j)$ when $i < j$, we have that for n sufficiently large the event

$$E_{k+1} = \left\{ \begin{array}{l} \left(1 - \frac{1}{\log n}\right)^k n^{1-\beta} \leq m_{k+1} \leq n^{1-\beta} \\ \\ \left(\frac{1}{2} - \frac{1}{\log n}\right)^k n(1-n^{-\beta}) \leq \ell_{k+1} \\ \\ \ell_{k+1} \leq \left(\frac{1}{2} + \frac{1}{\log n}\right)^k n(1-n^{-\beta}) \end{array} \right\}$$

holds with probability at least $1 - k\xi(k+1)$ where

$$\xi(k+1) = \exp\left(-\left(1 - \frac{1}{\log n}\right)^{k-1} n^{(1-\beta)/2}\right) + 2\exp\left(-\frac{2n(1-n^{-\beta})\left(\frac{1}{2} - \frac{1}{\log n}\right)^{k-1}}{(\log n)^2}\right).$$

Note that, as $n \to \infty$ both of the terms in $\xi(k+1)$ tend to zero, implying that E_{k+1} holds with probability tending to 1. The last step of the proof consists of showing what happens when Higher Criticism testing is applied after the last refinement. Conditionally on event E_{k+1} we are left with an 'effective' signal vector with length $n_e = |I^{(k+1)}|$, where

$$\left(1 - \frac{1}{\log n}\right)^k n^{1-\beta} + \left(\frac{1}{2} - \frac{1}{\log n}\right)^k n(1-n^{-\beta}) \leq n_e \leq n^{1-\beta} + \left(\frac{1}{2} + \frac{1}{\log n}\right)^k n(1-n^{-\beta}).$$

The final observation, $Y_a^{(k+1)}$, allocates $n/(k+1)$ energy among the n_e indices, resulting in an effective signal amplitude of

$$\sqrt{2r\frac{n\log n}{(k+1)n_e}}, \tag{9}$$

for each non-zero signal component. To specify the effective signal parameters r_e and β_e for this reduced dimension problem, we equate (9) with $\sqrt{2r_e \log n_e}$, which implies

$$r_e = \frac{r}{k+1}\frac{n\log n}{n_e \log n_e}.$$

Noting that when $n \to \infty$, $n/n_e \to 2^k$ and $\log n / \log n_e \to 1$, we conclude that

$$r_e \to r\frac{2^k}{k+1}.$$

We proceed similarly for the sparsity parameter β_e, to determine $\beta_e \to \beta$. Thus, the final observation $Y^{(k+1)}$, is equivalent to a single observation of a signal with length n_e, having $n_e^{1-\beta_e}$ nonzero entries of amplitude $\sqrt{2r_e \log n_e}$. Applying Higher Criticism testing to these observations, and noting that $n_e \to \infty$ as $n \to \infty$ results in the threshold behavior described in Theorem 3.1 (but with the new parameters β_e and r_e).

B. Sketch of Proof of Theorem 5.2

The proof mimics that of Theorem 5.1 with differences as noted here. We first generalize the signal model by letting $\mu(n) = \sqrt{2rg(n)}$, and we let $k = k(n)$ denote the total number of refinements, which is now a function of n. An analogous claim to that of Lemma 7.3 can be obtained for a general tolerance $\gamma(n)$. Iterating this result under the following conditions:

- $g(n) = (\log_2 \log n)^{1+\delta}$, for any fixed $\delta > 0$,
- $k(n) = \log_2 \log n$,
- $\gamma(n) = (\log_2 \log n)^{-2}$,

it can be shown that the output of the adaptive procedure is equivalent to observations of a sparse signal of length $n_e \to \infty$ as $n \to \infty$, having $n_e^{1-\beta_e}$ nonzero entries of amplitude at least $\sqrt{2r_e \log n_e}$, where $r_e = r$ and $\beta_e = \beta$, establishing the claim.

REFERENCES

[1] D. Donoho and J. Jin, "Higher criticsm for detecting sparse heterogenous mixtures," *Ann. Statist.*, vol. 32, no. 3, pp. 962–994, 2004.

[2] J. Jin, "Detecting a target in very noisy data from multiple looks," *Inst. Math. Statist. (IMS) Lecture Notes*, vol. 45, pp. 255–286, 2004.

[3] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Statist.*, vol. 23, pp. 493–507, 1952.

[4] S. Verdu, *Multiuser Detection*, Cambridge University Press, New York, 1998.