ADAPTIVE SAMPLING FOR SPARSE RECOVERY

Ali Tajer and H. Vincent Poor

Department of Electrical Engineering, Princeton University, Princeton, NJ 08544
Email: {tajer, poor}@princeton.edu

ABSTRACT

Consider n data sequences, each consisting of independent and identically distributed elements drawn from one of the two possible zero-mean Gaussian distributions with variances A_0 and A_1. The problem of quickly identifying all of the sequences with variance A_1 is considered and an adaptive two-stage experimental design and testing procedure is proposed. The agility and reliability gains in comparison with the existing related methods for quick search over multiple sequences are quantified..

1. PRELIMINARIES

1.1. Model

Consider n observation sequences $\mathcal{X}_1, \ldots, \mathcal{X}_n$, where for each $i \in \{1, \ldots, n\}$, $\mathcal{X}_i \triangleq \{X_i(j); j = 1, 2, \ldots\}$. The observations of each sequence $i \in \{1, \ldots, n\}$ are independent and identically distributed (i.i.d) and obey one of the two hypotheses:

$$\begin{array}{ll} \mathsf{H}_0: & X_i(j) \sim Q_0 \triangleq \mathcal{N}_{\mathbb{C}}(0, A_0), \quad j = 1, 2, \ldots \\ \mathsf{H}_1: & X_i(j) \sim Q_1 \triangleq \mathcal{N}_{\mathbb{C}}(0, A_1), \quad j = 1, 2, \ldots \end{array} \tag{1}$$

where $\mathcal{N}_{\mathbb{C}}(a, b)$ denotes the complex Gaussian distribution with mean a and variance b, and A_0 and A_1 are specified positive real numbers. Each sequence $\mathcal{X}_i$ is distributed according to Q_0 or Q_1 independently of the rest. Also, we assume that the hypothesis H_0 occurs *sparsely* and assume that for each sequence it occurs with prior probability ϵ_n. To model the sparsity we also assume that $\epsilon_n = o(1)$. Also, for $m \in \{0, 1\}$ let us define the sets

$$\mathcal{H}_m \triangleq \left\{ i \in \{1, \ldots, n\} : \mathcal{X}_i \sim Q_m \right\}. \tag{2}$$

1.2. Search Objective

The main objective is to use observations $\mathcal{X}_1, \ldots, \mathcal{X}_n$ in order to identify $T \in \mathbb{N}$ elements of $\mathcal{H}_0$. The conventional non-adaptive sampling procedures use some pre-specified number of observations and locate T sequences of interest. This strategy is non-adaptive, in the sense that the measurement process is fixed *a priori* and does not change during the experiment. In contrast, we devise an adaptive procedure in which the measurement strategy is adjusted sequentially such that future measurements use information gathered from previous ones. We demonstrate that such measurement adaptation substantially improves the *reliability* and *agility* in identifying the T sequences of interest.

Clearly, for detecting multiple sequences drawn from Q_0 there exists a tradeoff between agility and reliability in the sense that achieving a higher level of detection reliability requires using more sampling resources, which in turn results in delay in reliable detection. We characterize this tradeoff in both non-adaptive and the proposed adaptive sampling procedures. Comparing these tradeoffs demonstrates the agility and reliability gains afforded by the adaptive procedure. The analysis provided is asymptotic with respect to a large number of sequences, n.

2. NON-ADAPTIVE SAMPLING

The development of the non-adaptive sampling scheme serves a two-fold purpose. On one hand this detection scheme is also deployed in the detection phase of the adaptive procedure proposed in Section 3, and on the other hand, it offers a baseline for assessing the gain yielded by the adaptive procedure.

2.1. Non-Adaptive Sensing Procedure

Constructing a non-adaptive sampling procedure involves two issues. One pertains to the experimental design, which is the design of the information-gathering process. In our setup the experimental design elucidates the distribution of the sampling budget among the sequences $\mathcal{X}_1, \ldots, \mathcal{X}_n$. The second issue is to design a detector based on some optimality criterion. Prior to the sampling procedure all the sequences are equally likely to be drawn from Q_0. Due to the inherent symmetry and the sparse occurrence of H_0 we assume that the experimental design measures all sequences equally.

Given this experimental design, it is straightforward to construct a detector that is optimal in the sense that it maximizes the *a posterior* probability of successfully detecting T elements of H_0. Suppose that the measurement budget is $M \in \mathbb{N}$ per sequence. The observation vector $\boldsymbol{X}_i \triangleq [X_i(j)]_j$ for sequence i is a sample from a mixture Gaussian distribution. Given the observation set $\mathcal{D}_n \triangleq \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$, the maximum *a posteriori* probability (MAP) rule for identifying T member of $\mathcal{H}_0$ is formalized in the following remark.

This research was supported in part by the U. S. Air Force Office of Scientific Research under MURI Grant FA9550-09-1-0643, and in part by the Qatar National Research Fund under Grant NPRP 08-522-2-211.

Remark 1 *The MAP rule for detecting T elements of H_0 is given by*

$$\widehat{\mathcal{U}}_{\textit{unif}} \triangleq \arg \max_{\mathcal{U}:\ |\mathcal{U}|=T} P(\mathcal{U} \subseteq \mathcal{H}_0 \mid \mathcal{D}_n) = \arg \min_{\mathcal{U}:\ |\mathcal{U}|=T} \sum_{i\in\mathcal{U}} \|\boldsymbol{X}_i\|^2 \,. \quad (3)$$

Hence, in order to locate T sequences of interest the MAP detector requires only the sufficient statistic $Y_i \triangleq \|\boldsymbol{X}_i\|^2$ for $i = 1, \ldots, n$. Corresponding to the sequence of random variables $\{Y_1, \ldots, Y_n\}$ we define $\{Y_{(1)}, \ldots, Y_{(n)}\}$ as the sequence of order statistics in an increasing order, e.g., $Y_{(m)}$ represents the m^{th} smallest element of $\{Y_1, \ldots, Y_n\}$.

2.2. Asymptotic Performance

Recalling the distribution of $X_i(j)$ given in (2), the sufficient statistics Y_i under hypothesis H_m for $m \in \{0, 1\}$ is distributed as

$$Y_i \mid \mathsf{H}_m \sim \text{Gamma}\ (M, A_m) \quad \text{for} \quad i = 1, \ldots, n \,,$$

where Gamma (a, b) denotes a Gamma distribution with parameters a and b. Clearly the detector makes a detection error if $\widehat{\mathcal{U}}_{\text{unif}} \bigcap \mathcal{H}_1 \neq \emptyset$. Let us define u_t and v_t as the indices of the t^{th} smallest elements of the sets $\{Y_i : i \in \{1, \ldots, n\}\}$ and $\{Y_i : i \in \mathcal{H}_0\}$, respectively. From Remark 1 the detection error probability is given by

$$\begin{aligned} P_{\text{unif}}(n) &\triangleq P\left(\widehat{\mathcal{U}}_{\text{unif}} \cap \mathcal{H}_1 \neq \emptyset\right) \\ &= 1 - P\left(\widehat{\mathcal{U}}_{\text{unif}} \cap \mathcal{H}_1 = \emptyset\right) \\ &= 1 - P(\{u_1, \ldots, u_T\} \subseteq \mathcal{H}_0) \,. \end{aligned}$$

Note that the event of having all the T smallest measured Y_i belonging to H_0 is equivalent to

$$\begin{aligned} \{\{u_1, \ldots, u_T\} \subseteq \mathcal{H}_0\} &\equiv \left\{ \max_{t\in\{1,\ldots,T\}} Y_{v_t} \leq \min_{i\in\mathcal{H}_1} Y_i \right\} \\ &\equiv \left\{ Y_{v_T} \leq \min_{i\in\mathcal{H}_1} Y_i \right\} \,. \end{aligned}$$

Therefore,

$$P_{\text{unif}}(n) = 1 - P(Y_{v_T} \leq \min_{i\in\mathcal{H}_1} Y_i) \,. \quad (4)$$

Assessing $P_{\text{NA}}(n)$ as defined above relies on the properties of the order statistics of two sets of random variables. The following lemma is instrumental for characterizing the distributions of these sets of order statistics and evaluating $P_{\text{NA}}(n)$. This is a generalization of a well-studied problem in the context of extreme value theory that considers *the first* order statistic [1]. In this lemma, we give the corresponding results for higher order statistics and at the same time we also allow for distribution evolution, i.e., the *number* and *distribution* of the involved random variables changes *simultaneously*, for which the existing results are not applicable.

Lemma 1 *Let $\{Y_i\}_{i=1}^m$ be a sequence of i.i.d. random variables distributed as Gamma (M, α_m) and denote its corresponding sequence of order statistics by $\{Y_{(i)}\}_{i=1}^m$. Let $b_m \triangleq \alpha_m \left[\frac{\Gamma(M+1)}{m}\right]^{\frac{1}{M}}$ and for some $T \in \mathbb{N}$ define the sequence of random variables $W_{(i)}^m \triangleq \frac{Y_{(i)}}{b_m}$ for $i = 1, \ldots, T$. Then as $m \to \infty$, $W_{(i)}^m$ converges in distribution to a random variable $W_{(i)}$ with cumulative density function (CDF)*

$$Q_{(i)}(w; m) \triangleq P(W_{(i)} < w) \doteq 1 - \exp(-w^M) \sum_{k=0}^{i-1} \frac{w^{kM}}{k!} \,.$$

For the setting of Section 1.1, the following theorem characterizes the asymptotic performance of the MAP detector. It also establishes the tradeoffs among the prior probability ϵ_n, per sequence sampling budget M, and the ratio of the variances of distributions Q_0 and Q_1.

Theorem 1 (Non-Adaptive Tradeoff) *When $\epsilon_n = o(1)$ and $n\epsilon_n = \omega(1)$, the error probability of the MAP for identifying T elements of $\mathcal{H}_0$ is given by*

$$\begin{aligned} P_{\textit{unif}}(n) &= P\left(\widehat{\mathcal{U}}_{\textit{unif}} \cap \mathcal{H}_1 \neq \emptyset\right) \\ &\doteq 1 - \left(1 + [(A_1/A_0)^M \cdot \epsilon_n]^{-1}\right)^{-T} \,. \end{aligned} \quad (5)$$

As expected, there exists a tension between reliability and agility. On one hand, increasing the sampling budget per sequence M favors reliability, as according to (5) it improves the probability of successfully detection, and on the other hand, imposes more delay in detecting T sequences distributed as Q_0. By using the result of Theorem 1, we offer a necessary and sufficient condition on the scaling of the ratios of the variances A_1/A_0 to guarantee asymptotically error-free detection in the non-adaptive sampling setting.

Corollary 1 (Non-Adaptive Variance Scaling) *When $\epsilon_n = o(1)$ and $n\epsilon_n = \omega(1)$, a necessary and sufficient condition for $P_{\textit{unif}}(n) \to 0$ as $n \to \infty$ is that $\frac{A_1}{A_0}$ scales with increasing n as*

$$\frac{A_1}{A_0} = \omega\left(\sqrt[M]{\frac{1}{\epsilon_n}}\right) \,. \quad (6)$$

3. ADAPTIVE SAMPLING

3.1. Adaptive Sampling

Our proposed adaptive sampling procedure has two phases, namely the *exploration* phase and the *detection* phase. The exploration phase, being an iterative procedure, is intended to purify the set of the sequences to be observed carefully for detecting the sequences drawn from Q_0. This phase is accomplished by successively identifying and eliminating a group of sequences deemed to be drawn from Q_1. The detection phase is performed after the exploration phase in order to identify T sequences of interest among the subset of candidate sequences retained after exploration. The detection scheme deployed is identical to the MAP detection scheme of Section 2.

The exploration phase proceeds in an iterative way. In each iteration it further monitors the sequences retained by the previous iteration and eliminates those deemed to be drawn from Q_0 least-likely. The core idea is that it is relatively easy to identify sequences drawn from Q_1 with low-quality measurements (recall that ϵ_n is small). Each iteration carries on by thresholding the observed energy on each sequence retained by the previous iteration. The threshold level depends only on $\frac{A_1}{A_0}$, and is designed such that at each iteration roughly half of the existing sequences distributed as Q_1 are eliminated, while almost all of those distributed as Q_0 are preserved. The output of each exploration phase will have a more condensed proportion of the desired sequences to the non-desired ones. Subsequently, the detector developed for the non-adaptive procedure is applied on this refined set of sequences in order to identify T sequences distributed as Q_0. This entire procedure bears similarities with Distilled Sensing [2], however, the analysis is substantially different. This is due to the different sensing objective (identifying any arbitrary number of sequences as opposed to [2] that aims to identify almost all) as well as the underlying statistical model.

We show that the gains yielded by this adaptive procedure can be interpreted in two ways. First we demonstrate that when targeting at achieving the same level of detection reliability, the adaptive procedure requires substantially less sampling budget, or equivalently it is substantially more agile. Secondly, we show that under the same sampling budget, and targeting identical detection reliability, the adaptive procedure imposes less-stringent conditions on how fast the power of the active users γ_n must scale with increasing n. This essentially indicates that for some choices of γ_n the adaptive procedure can guarantee successful detection while the most non-adaptive procedure fails to do so.

Let us define K as the number of exploration cycles (iterations) in the exploration phase. Also denote the sampling budget per sequence in the k^{th} exploration cycle by M_k. The exploration phase is initialized by including *all* sequences for sampling and resumes as follows. In the first iteration all sequences are allocated the identical sampling budget of M_1. The energy levels of all sequences are compared against $\lambda_1(A_1/A_0)$, where λ_1 is the median of the distribution Gamma $(M_1, 1)$. The sequences for which the measured energy exceed this threshold are discarded and the rest are carried over to the second iteration for further sampling. The same procedure is repeated throughout all K cycles. More specifically, in the k^{th} cycle all the sequences retained by the $(k-1)^{th}$ iteration are allocated the identical sampling budget of M_k. The energy levels of these sequences are compared with $\lambda_k(A_1/A_0)$, where λ_k is the median of the distribution Gamma $(M_k, 1)$ and the exploration is performed via thresholding as in the first iteration. Finally, after the exploration phase, each of the remaining sequences is allocated the sampling budget of M_{K+1} and the MAP detection scheme provided in Remark 1 is applied in order to detect T sequences distributed as Q_0.

Exploration phase

Input $K \in \mathbb{N}$ and $\{M_1, \ldots, M_{K+1}\}$.
Initialize the index set $\mathcal{G}_0 \leftarrow \{1, \ldots, n\}$.
for $k = 1, \ldots, K$ **do**
  Set $Y_i^k = \begin{cases} \|\boldsymbol{X}_i^k\|^2 & \text{for} \quad i \in \mathcal{G}_{k-1} \\ +\infty & \text{for} \quad i \notin \mathcal{G}_{k-1} \end{cases}$.
  Obtain $\mathcal{G}_k \leftarrow \{i \in \mathcal{G}_{k-1} \mid Y_i^k < \lambda_k(A_1/A_0)\}$.
end for

Detection phase

Set $Y_i^{K+1} = \begin{cases} \|\boldsymbol{X}_i^{K+1}\|^2 & \text{for} \quad i \in \mathcal{G}_K \\ +\infty & \text{for} \quad i \notin \mathcal{G}_K \end{cases}$.
$\widehat{\mathcal{U}}_{\text{adap}} \triangleq \{i \in \mathcal{G}_K : Y_i^{K+1} \leq Y_{(T)}^{K+1}\}$.
Output $\widehat{\mathcal{U}}_{\text{adap}}$.

We set $\mathcal{G}_0 \triangleq \{1, \ldots, n\}$ and for $k = 1, \ldots, K$, we define $\mathcal{G}_k$ as the set of the indices of the sequences that are retained by the k^{th} exploration cycle. Clearly we have $\mathcal{G}_K \subseteq \cdots \subseteq \mathcal{G}_1 \subseteq \mathcal{G}_0$ and $\mathcal{G}_K$ contains the set of the indices of the candidate sequences among which T sequences distributed as Q_0 will be detected. The set of measurements defined for the non-adaptive scheme is extended for the proposed adaptive procedure as follows. We define the set of measurements in the k^{th} cycle as

$$\mathcal{D}_n^k \triangleq \left\{ \boldsymbol{X}_i^k : \quad \text{for} \quad i \in \mathcal{G}_{k-1} \right\} \quad \text{for} \quad k = 1, \ldots, K+1 \, .$$

The measurement sets $\mathcal{D}_n^1, \ldots, \mathcal{D}_n^K$ are processed in the exploration phase and the measurement set $\mathcal{D}_n^{K+1}$ is used in the detection phase. Under hypothesis H_m, the observation sample $X_i^k(j)$ for $k = 1, \ldots, K+1$ is distributed as

$$X_i^k(j) \mid \mathsf{H}_m \overset{\text{i.i.d.}}{\sim} \mathcal{N}_{\mathbb{C}}(0, A_m), \quad \text{for} \quad i \in \mathcal{G}_{k-1} \, . \tag{7}$$

We also define

$$Y_i^k \triangleq \|\boldsymbol{X}_i^k\|^2 \quad \text{for} \quad i \in \mathcal{G}_{k-1} \quad \text{and} \quad k = 1, \ldots, K+1 \, . \tag{8}$$

Equations (7) and (8) provide that for $k = 1, \ldots, K+1$

$$Y_i^k \mid \mathsf{H}_m \sim \text{Gamma}\,(M_k, A_m) \quad \text{for} \quad i \in \mathcal{G}_{k-1} \, . \tag{9}$$

For each k, corresponding to the sequence $\{Y_i^k\}_{i \in \mathcal{G}_{k-1}}$ we define the sequence of order statistics $\{Y_{(i)}^k\}_{i \in \mathcal{G}_{k-1}}$ in an increasing order such that $Y_{(i)}^k$ represents the i^{th} smallest element of this sequence. The adaptive sampling procedure is formally described in the table above.

3.2. Asymptotic Performance

We start by assessing the performance for any given value of the exploration cycles K. The analysis of the adaptive sampling procedure follows the approach of [2], albeit with the non-trivial modifications to deal with the different objective and the different observation model. The following lemmas shed light on how the adaptive procedure accomplishes the exploration cycles. Lemma 2 characterizes the proportion of the sequences distributed as Q_0 that are retained in each exploration cycle.

Lemma 2 *Let $m_0 = |\mathcal{H}_0|$ and for $k = 1, \ldots, K$ define m_k as the number of sequences distributed as Q_0 that are retained by the k^{th} exploration cycle. Conditionally on m_{k-1} and for sufficiently large n the event*

$$\left(\frac{A_1/A_0}{1+A_1/A_0}\right) m_{k-1} \leq m_k \leq m_{k-1}, \text{ for } k = 1, \ldots, K, \tag{10}$$

holds with probability at least $1 - \exp\left(-\frac{m_{k-1}}{n^\alpha}\right)$ for any $\alpha > 0$.

The next lemma shows that during each exploration cycle almost half of the sequences distributed as Q_1 are eliminated.

Lemma 3 *Let $\ell_0 = |\mathcal{H}_1|$ and for $k = 1, \ldots, K$ define ℓ_k as the number of the sequences distributed as Q_1 that are retained by the k^{th} exploration cycle. Conditionally on ℓ_{k-1} and for sufficiently large n, for all $k = 1, \ldots, K$, the event*

$$\left(\frac{1}{2} - \frac{1}{\log n}\right) \ell_{k-1} \leq \ell_k \leq \left(\frac{1}{2} + \frac{1}{\log n}\right) \ell_{k-1}\ , \tag{11}$$

holds with probability at least $1 - 2\exp\left(-\frac{2\ell_{k-1}}{(\log n)^2}\right)$.

A careful use of the above lemmas establishes the performance of the adaptive sampling in the following theorem.

Theorem 2 (Adaptive Tradeoff) *When $\epsilon_n = o(1)$ and $n\epsilon_n = \omega(1)$, the error probability of the adaptive sampling procedure for identifying T sequences distributed as Q_0 is given by*

$$\begin{aligned} P_{\text{adap}}(n) &\triangleq P\left(\widehat{\mathcal{U}}_{\text{adap}} \cap \mathcal{H}_1 \neq \emptyset\right) \\ &\doteq 1 - \left(1 + \left[(A_1/A_0)^{M_{K+1}} \cdot 2^K \epsilon_n\right]^{-1}\right)^{-T}. \end{aligned} \tag{12}$$

In order to quantify the gains yielded by the adaptive procedure, we compare the results for the non-adaptive and adaptive schemes provided in Theorems 1 and 2. In particular we first characterize the *agility factor*, which we define as the ratio of the sampling budgets required by the adaptive procedure to that required by the non-adaptive scheme with the aim of attaining identical asymptotic reliability levels in the asymptotic and the non-asymptotic procedure, i.e., $P_{\text{adap}}(n) \doteq P_{\text{unif}}(n)$,

Theorem 3 (Agility) *When $\epsilon_n = o(1)$ and $n\epsilon_n = \omega(1)$, the agility factor of the adaptive sampling approach with Mn sampling budget is asymptotically upper bounded by $\left(\frac{1}{2^K} + \frac{2}{M}\right)$, where K is the number of exploration cycles.*

It is noteworthy that while the number of exploration cycles K can be made arbitrarily large (but fixed as a function of n), increasing it beyond some point will affect the agility very insignificantly. More specifically, for large K, the agility factor will be dominated by the term $\frac{2}{M}$. This underlines the fundamental limit of the agility gain yielded by the adaptive procedure.

An analog of Corollary 1 can be derived for the adaptive procedure, providing a necessary and sufficient condition on the scaling of $\frac{A_1}{A_0}$ for guaranteeing asymptotic error-free detection of T sequences distributed as Q_0. For comparison purposes we assume that both adaptive and non-adaptive procedures are granted the same sampling budget.

Corollary 2 (Adaptive Power Scaling) *When $\epsilon_n = o(1)$ and $n\epsilon_n = \omega(1)$, given that the sampling budget is Mn, a necessary and sufficient condition for $P_{\text{adap}}(n) \to 0$ as $n \to \infty$ is that*

$$\frac{A_1}{A_0} = \omega\left(\sqrt[M']{\frac{1}{2^K \epsilon_n}}\right)\ , \tag{13}$$

where $M' \geq 2^K(M-2)+2$.

Comparing the result above with that of Corollary 1 shows that an adaptive scheme can cope with signals with much smaller variances. More specifically, by noting that M' is substantially larger than M, the variance scaling requirement in the adaptive scenario, which is smaller than $\omega\left(\sqrt[M']{\frac{1}{\epsilon_n}}\right)$ becomes substantially smaller than its counterpart in the non-adaptive scenario $\omega\left(\sqrt[M]{\frac{1}{\epsilon_n}}\right)$. As a result, there are scenarios where non-adaptive schemes fail to successfully identify T sequences of interest, while the adaptive scheme succeeds.

4. CONCLUSION

In this paper we have presented an adaptive sampling methodology for quickly searching over finitely many sequences with the objective of identifying multiple sequences that are distributed according to a given distribution of interest. The core idea of the sampling procedure is to successively and gradually adjust the measurement process using information gleaned from the previous measurements. Compared to the non-adaptive procedures, dramatic gains in terms of reliability and agility are achieved.

5. REFERENCES

[1] H. A. David and H. N. Nagaraja, *Order Statistics*, 3rd ed. New York: Wiley, 2003.

[2] J. Haupt, R. Castro, and R. Nowak, "Distilled sensing: Adaptive sampling for sparse detection and estimation," *arXiv:1001.5311*, 2010, submitted for publication.