A Performance Guarantee for Adaptive Estimation of Sparse Signals

Dennis Wei and Alfred O. Hero, III
Department of Electrical Engineering and Computer Science
University of Michigan
Ann Arbor, MI, USA

***Abstract*—This paper studies adaptive sensing for estimating the nonzero amplitudes of a sparse signal. We consider a previously proposed optimal two-stage policy for allocating sensing resources. We derive an upper bound on the mean squared error resulting from the optimal two-stage policy and a corresponding lower bound on the improvement over non-adaptive sensing. It is shown that the adaptation gain is related to the detectability of nonzero signal components as characterized by a Bhattacharyya coefficient, thus quantifying analytically the dependence on the sparsity level of the signal, the signal-to-noise ratio, and the sensing resource budget. The bound is shown to be a good approximation to the optimal two-stage gain through numerical simulations.**

I. Introduction

Adaptive or controlled sensing refers to the control of the signal acquisition process in response to information that has already been learned about the signal. Recently, attention has been focused on adaptive sensing for sparse signals [1]–[7], i.e., signals that occupy a small number of dimensions in a much larger ambient space. The common theme in these works is to gradually learn the support of the signal and allocate sensing resources accordingly to improve the signal-to-noise ratio (SNR) for inference tasks.

In [1], adaptive sensing is considered for the estimation of the nonzero amplitudes of a sparse signal. For the case of two sensing stages, policies for allocating sensing resources are derived that are optimal for a variety of estimation loss functions, generalizing an optimal two-stage policy proposed by [2] under slightly different motivations. For more than two stages, [1] also provides allocation policies based on the approximate dynamic programming method of open-loop feedback control (OLFC) [8] with performance that improves monotonically as the number of stages increases.

Empirical results in [1], [2] show that multistage adaptive sensing can lead to dramatically better estimates of nonzero signal components compared to non-adaptive sensing. However, analytical quantification of the gain due to adaptation in this setting has so far been lacking. In this paper, we make a contribution in this direction by obtaining an upper bound on the mean squared error (MSE) resulting from two-stage adaptive sensing, and by extension a lower bound on the adaptation gain. While our focus is on two-stage policies because of their relative tractability, the result extends to the multistage policies of [1] because of the monotonicity property noted above. Intuitively, the gain depends on the extent to which sensing resources can be concentrated on the signal support, which depends in turn on the ability to detect nonzero signal components. It is of interest therefore to determine the impact on performance of several key quantities: the sparsity level of the signal, which limits the degree of concentration; the SNR, which affects detectability; and the total sensing resource budget. Our analysis shows that many of these dependences can be summarized by a single measure of detectability, specifically the Bhattacharyya coefficient between the distribution of the observations when a signal component is present and the unconditional distribution of the observations. We note that Bhattacharyya coefficients have been used before in adaptive sensor management [9], specifically to bound the probability of classification error and thus assess the utility of further observations.

In related work on adaptive sensing of sparse signals, performance guarantees have been derived for other inference tasks, notably signal support recovery. Under a Gaussian observation model similar to the one in this work, bounds are obtained on the SNR required by a procedure known as distilled sensing [3] to recover the support with vanishing false discovery and non-discovery rates as the signal dimension increases. The method of sequential thresholding [4] generalizes distilled sensing and is shown to recover the support exactly in the high-dimensional limit provided that the number of observations grows as a function of the sparsity level and the SNR as measured by a Kullback-Leibler divergence. Distilled sensing has also been generalized to a gamma observation model appropriate for spectrum sensing and the detection error probability is characterized in terms of the channel occupancy, primary user power, and sensing budget [5]. Additionally, similar support recovery guarantees have been given for sensing that is both adaptive and compressive [6], [7], in contrast to the non-compressive component-wise observations considered in [3]–[5] and herein. For amplitude estimation however, the performance of adaptive methods has been less well studied. The present work also differs from the above references in its focus on an optimal policy (albeit for two stages) with more complicated structure than the policies analyzed previously.

This paper is concerned with the problem of estimating nonzero signal amplitudes as posed in [1]. Sections II and III review relevant material from [1], including the signal and observation models, the MSE cost function, and a two-stage policy for sensing resource allocation that minimizes the MSE. In Section IV, we develop the main results of this paper, namely a bound on the MSE of the two-stage policy in Section III and its improvement compared to non-adaptive sensing. The bound is validated numerically in Section V and the paper concludes in Section VI.

II. Problem formulation

The two-stage sensing problem considered in this paper is a special case of the one in [1] and is also related to [2]. Let $\mathbf{x}$ be a sparse N-dimensional signal whose support is represented by a vector $\mathbf{I}$ of Bernoulli indicator variables I_i, $i = 1, \ldots, N$, with $\mathbb{P}(I_i = 1) = p$ a priori. If $I_i = 0$, $x_i = 0$ whereas if $I_i = 1$, x_i is distributed a priori as a Gaussian random variable with mean μ and variance σ^2, independent of other components of $\mathbf{x}$. While the policies in [1] apply to signal priors that are non-uniform over i, the results in the present work depend on the assumption of uniformity.

The signal $\mathbf{x}$ is observed in two stages in the same basis in which it is sparse. Given sensing effort parameters $\lambda_i(t-1)$, which may represent observation time, number of samples, or other resources depending on the application, the observation $y_i(t)$ of component i

in stage t is given by

$$y_i(t) = x_i + \frac{n_i(t)}{\sqrt{\lambda_i(t-1)}}, \quad i = 1, \ldots, N, \quad t = 1, 2, \tag{1}$$

where $n_i(t)$ denotes i.i.d. zero-mean Gaussian noise with variance ν^2 and we adopt the convention that the observation is not taken if $\lambda_i(t-1) = 0$. In adaptive sensing, the second-stage effort allocation $\boldsymbol{\lambda}(1) = (\lambda_1(1), \ldots, \lambda_N(1))$ can be chosen as a function of the first-stage observations $\mathbf{y}(1) = (y_1(1), \ldots, y_N(1))$ to selectively increase measurement precision (inverse variance). The overall sensing budget is normalized as

$$\sum_{i=1}^{N} (\lambda_i(0) + \lambda_i(1)) = N. \tag{2}$$

For a sensing budget different from N, the effect of the normalization is to scale ν^2 since it is always the ratio $\nu^2/\lambda_i(t-1)$ that determines the effective noise variance in (1).

The observations $\mathbf{y}(1), \mathbf{y}(2)$ are used to produce an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$. As in [1], we assume that the nonzero components of $\mathbf{x}$ are of primary interest and seek to minimize the mean squared error (MSE) over the signal support:

$$\mathbb{E}\left\{\sum_{i=1}^{N} I_i \left|\hat{x}_i - x_i\right|^2\right\}, \tag{3}$$

where the expectation is taken over $\mathbf{I}$, $\mathbf{x}$, $\mathbf{y}(1)$ and $\mathbf{y}(2)$. In addition to accounting for estimation error directly, the cost function (3) also promotes better performance in detecting nonzero signal components, as demonstrated in [2], [10], since achieving low MSE requires identifying the true signal support.

III. Optimal two-stage effort allocation policy

In [1], an effort allocation policy is derived that minimizes the MSE (3) over all two-stage policies subject to the budget constraint (2). This optimal two-stage policy is the subject of study in the present work and is summarized here along with related facts for later reference.

The policy in [1] is obtained using dynamic programming [8] and depends on the observation history $\mathbf{Y}(t) \equiv \{\mathbf{y}(1), \ldots, \mathbf{y}(t)\}$ through a collection of state variables $\mathbf{s}(t) = (\mathbf{p}(t), \boldsymbol{\mu}(t), \boldsymbol{\sigma}^2(t), \Lambda(t))$, where $p_i(t) = \mathbb{P}(I_i = 1 \mid \mathbf{Y}(t))$, $\mu_i(t) = \mathbb{E}[x_i \mid I_i = 1, \mathbf{Y}(t)]$, and $\sigma_i^2(t) = \mathrm{var}(x_i \mid I_i = 1, \mathbf{Y}(t))$ for $i = 1, \ldots, N$, and $\Lambda(t)$ is the remaining sensing budget after t stages. The state variables are initialized uniformly over i to the prior values p, μ, σ^2, and $\Lambda(0)$ at $t = 0$, and evolve according to the following relations:

$$p_i(t+1) = \frac{p_i(t) f_1(y_i(t+1))}{p_i(t) f_1(y_i(t+1)) + (1 - p_i(t)) f_0(y_i(t+1))}, \tag{4a}$$

$$\mu_i(t+1) = \frac{\nu^2 \mu_i(t) + \lambda_i(t)\sigma_i^2(t) y_i(t+1)}{\nu^2 + \lambda_i(t)\sigma_i^2(t)}, \tag{4b}$$

$$\sigma_i^2(t+1) = \frac{\nu^2 \sigma_i^2(t)}{\nu^2 + \lambda_i(t)\sigma_i^2(t)}, \tag{4c}$$

$$\Lambda(t+1) = \Lambda(t) - \sum_{i=1}^{N} \lambda_i(t), \tag{4d}$$

where

$$f_0(y_i(t+1)) = \phi(y_i(t+1); 0, \nu^2/\lambda_i(t)), \tag{5a}$$

$$f_1(y_i(t+1)) = \phi(y_i(t+1); \mu_i(t), \sigma_i^2(t) + \nu^2/\lambda_i(t)), \tag{5b}$$

and $\phi(\cdot; \mu, \sigma^2)$ denotes a Gaussian probability density function (PDF) with mean μ and variance σ^2. With these definitions, it is shown in [1] that the conditional mean estimator $\hat{\mathbf{x}} = \boldsymbol{\mu}(2)$ minimizes (3) and the effort allocation problem can be expressed as a two-stage optimization problem,

$$\begin{aligned} J_1^*(\mathbf{s}(1)) = \nu^2 \min_{\boldsymbol{\lambda}(1)} \quad & \sum_{i=1}^{N} \frac{p_i(1)}{\nu^2/\sigma_i^2(1) + \lambda_i(1)} \\ \text{s.t.} \quad & \sum_{i=1}^{N} \lambda_i(1) = \Lambda(1), \quad \lambda_i(1) \geq 0 \ \ \forall\, i, \end{aligned} \tag{6}$$

$$J_0^*(\mathbf{s}(0)) = \min_{0 \leq \lambda \leq 1} \ \mathbb{E}\left\{J_1^*(\mathbf{s}(1)) \mid \mathbf{s}(0), \lambda \mathbf{1}\right\}. \tag{7}$$

The optimal two-stage MSE is given by $J_0^*(\mathbf{s}(0))$. The second-stage optimization (6) is conditioned on the values of the first-stage observations $\mathbf{y}(1)$ through the state $\mathbf{s}(1)$. The first-stage optimization (7) is then defined recursively in terms of $J_1^*(\mathbf{s}(1))$. Under a uniform prior, the first-stage allocation $\boldsymbol{\lambda}(0)$ is also uniform by symmetry, i.e., $\boldsymbol{\lambda}(0) = \lambda \mathbf{1}$ where $\mathbf{1}$ is a vector of ones, thus making (7) a one-dimensional optimization. The expectation in (7) is taken over $\mathbf{y}(1)$, which has i.i.d. components with distribution

$$f_p(y_i(1)) = p f_1(y_i(1)) + (1-p) f_0(y_i(1)) \tag{8}$$

parameterized by $\mathbf{s}(0) = \{p, \mu, \sigma^2, \Lambda(0)\}$ and λ. This expectation can be evaluated offline through Monte Carlo sampling.

The optimal solution to (6) can be given explicitly as detailed in [1], [2]. Here it suffices to note that the solution takes the form

$$\lambda_i^*(1) = \max\left\{C\sqrt{p_i(1)} - \frac{\nu^2}{\sigma_i^2(1)}, 0\right\}, \tag{9}$$

where C is a normalization constant that ensures $\sum_{i=1}^{N} \lambda_i^*(1) = \Lambda(1)$. It can be seen from (9) that the optimal allocations are thresholded to zero after a certain point in the rank order of the quantities $\sqrt{p_i(1)}\sigma_i^2(1)$. Furthermore, the nonzero allocations increase as the square root of the probabilities $p_i(1)$ and decrease with the precisions $1/\sigma_i^2(1)$.

IV. Performance guarantees

Previous work [1], [2] has shown empirically that the optimal two-stage policy in Section III can achieve substantially lower MSE than the non-adaptive strategy of uniform effort allocation. The contribution of this paper is to provide an analytical guarantee on the gain due to adaptation. Of particular interest is the dependence of the gain on several key parameters: the sparsity of the signal as represented by the fraction p, the SNR, and the sensing budget. To summarize the latter two quantities, we define the ratios $r \equiv \sigma^2/\nu^2$ and $s \equiv \mu^2/\sigma^2$. The ratio r characterizes the strength of the nonzero signal amplitudes relative to the effective noise variance ν^2, which depends on both the intrinsic noise level as well as the sensing budget through normalization. The ratio s represents the prior degree of certainty regarding the nonzero signal amplitudes.

It can be seen from the cost function (3) that the improvement due to adaptation is determined by the ability to concentrate sensing effort on the signal support and thus increase the effective SNR. This in turn depends on the detectability of nonzero signal components, which is characterized by the contrast between the signal-absent measurement distribution f_0 (5a) and the signal-present distribution f_1 (5b). Theorem 1 shows that the relevant measure of contrast is the Bhattacharyya coefficient

$$\mathrm{BC}_p = \int_{-\infty}^{\infty} \sqrt{f_1(y_i(1)) f_p(y_i(1))}\, dy_i(1) \tag{10}$$

between $f_1(y_i(1))$ and the unconditional distribution $f_p(y_i(1))$ (8), which is a mixture of $f_0(y_i(1))$ and $f_1(y_i(1))$. The Bhattacharyya coefficient measures the overlap between two distributions and takes values between 0 and 1, with 0 corresponding to no overlap and 1 to perfect overlap. It is related to the Hellinger distance H between two distributions via $H = \sqrt{1 - \mathrm{BC}}$.

Our main result is an upper bound on the optimal two-stage MSE $J_0^*(\mathbf{s}(0))$. Similar to [3]–[7], we consider the limit as the signal dimension $N \to \infty$, which simplifies the form of the bound.

Theorem 1: Assume that the signal $\mathbf{x}$ and observations $\mathbf{y}(1), \mathbf{y}(2)$ follow the model in Section II. Then in the limit of high signal dimension N, the optimal two-stage MSE $J_0^*(\mathbf{s}(0))$ is bounded from above as

$$\lim_{N\to\infty} \frac{1}{N\sigma^2} J_0^*(\mathbf{s}(0)) \leq \frac{p}{1 + r + r \max\limits_{\lambda\in[0,1]} (\mathrm{BC}_p^{-2} - 1)(1-\lambda)}. \tag{11}$$

Proof: First we bound the cost-to-go $J_1^*(\mathbf{s}(1))$ in (6). Although the optimal solution to (6) can be given explicitly [1], [2], it depends on the rank order of the random variables $\sqrt{p_i(1)}$. Since use of a suboptimal allocation can only increase the cost-to-go, we can bound $J_1^*(\mathbf{s}(1))$ by evaluating (6) with the following suboptimal allocation first proposed in [2]:

$$\lambda_i(1) = \Lambda(1) \frac{\sqrt{p_i(1)}}{\sum_{j=1}^N \sqrt{p_j(1)}}. \tag{12}$$

The allocation (12) preserves the square root dependence on the probabilities $p_i(1)$ as in the optimal allocation (9) but does not threshold any of the $\lambda_i(1)$ to zero. Empirical results in [2] show that the performance gap between the optimal and suboptimal allocations is small.

Substituting (12) into (6), we obtain

$$\begin{aligned}\frac{1}{N\sigma^2} J_1^*(\mathbf{s}(1)) &\leq \frac{\nu^2}{N\sigma^2} \sum_{i=1}^N \frac{p_i(1)}{\nu^2/\sigma_i^2(1) + \Lambda(1)\sqrt{p_i(1)}/\sum_{j=1}^N \sqrt{p_j(1)}} \\ &= \frac{\nu^2}{N\sigma^2} \sum_{i=1}^N \frac{p_i(1)}{\nu^2/\sigma^2 + \lambda + N(1-\lambda)\sqrt{p_i(1)}/\sum_{j=1}^N \sqrt{p_j(1)}} \\ &= \frac{1}{N} \sum_{i=1}^N \frac{p_i(1)}{1 + r\lambda + r(1-\lambda)\sqrt{p_i(1)} \left(\frac{1}{N}\sum_{j=1}^N \sqrt{p_j(1)}\right)^{-1}}.\end{aligned} \tag{13}$$

In the first equality above, we have used (4c) and (4d), $\lambda_i(0) = \lambda$, and the normalization $\Lambda(0) = N$, while in the second equality (13), we have used the definition of r. From (4a) and (8),

$$p_i(1) = \frac{p f_1(y_i(1))}{p f_1(y_i(1)) + (1-p) f_0(y_i(1))} = p \frac{f_1(y_i(1))}{f_p(y_i(1))}, \tag{14}$$

and since $\{y_i(1)\}_{i=1}^N$ are i.i.d., so too are $\left\{\sqrt{p_i(1)}\right\}_{i=1}^N$. Thus as $N \to \infty$, the law of large numbers ensures that the sample averages on the right-hand side of (13) converge almost surely to expectations over $y_i(1)$. Using (14) and (10) we have

$$\mathbb{E}\left[\sqrt{p_i(1)}\right] = \int_{-\infty}^{\infty} \sqrt{p} \sqrt{\frac{f_1(y_i(1))}{f_p(y_i(1))}} f_p(y_i(1)) dy_i(1) = \mathrm{BC}_p \sqrt{p},$$

and hence

$$\lim_{N\to\infty} \frac{1}{N\sigma^2} J_1^*(\mathbf{s}(1)) \leq \mathbb{E}\left\{ \frac{p_i(1)}{1 + r\lambda + r(1-\lambda)\sqrt{p_i(1)} \left(\mathrm{BC}_p \sqrt{p}\right)^{-1}} \right\} \tag{15}$$

Next we further simplify the right-hand side of (15). Using (14), the expectation in (15), which is taken with respect to the unconditional density $f_p(y_i(1))$, can be converted into an expectation with respect to $f_1(y_i(1))$, i.e., conditioned on $I_i = 1$. Thus the right-hand side of (15) becomes

$$\mathbb{E}\left\{ \frac{p}{1 + r\lambda + r(1-\lambda)\sqrt{p_i(1)} \left(\mathrm{BC}_p \sqrt{p}\right)^{-1}} \mid I_i = 1 \right\},$$

which is of the form $p/(a + b/p_i(1)^{-1/2})$, a concave function of $p_i(1)^{-1/2}$ with $a, b \geq 0$. Jensen's inequality then yields

$$\begin{aligned}&\lim_{N\to\infty} \frac{1}{N\sigma^2} J_1^*(\mathbf{s}(1)) \\ &\quad \leq \frac{p}{1 + r\lambda + r(1-\lambda)\left[\left(\mathrm{BC}_p \sqrt{p}\right) \mathbb{E}\left\{p_i(1)^{-1/2} \mid I_i = 1\right\}\right]^{-1}} \\ &\quad = \frac{p}{1 + r\lambda + r(1-\lambda)\mathrm{BC}_p^{-2}},\end{aligned} \tag{16}$$

where we have used (14) to obtain

$$\begin{aligned}\mathbb{E}\left\{p_i(1)^{-1/2} \mid I_i = 1\right\} &= \int_{-\infty}^{\infty} \sqrt{\frac{f_p(y_i(1))}{p f_1(y_i(1))}} f_1(y_i(1)) dy_i(1) \\ &= \frac{\mathrm{BC}_p}{\sqrt{p}}.\end{aligned}$$

We now use (7) to relate the bound on $J_1^*(\mathbf{s}(1))$ in (16) to $J_0^*(\mathbf{s}(0))$. Since the right-hand side of (16) is deterministic, the only addition to (16) is a minimization over $\lambda \in [0, 1]$, or equivalently a maximization over λ in the denominator. A rearrangement of the denominator completes the proof. ■

To determine the gain relative to non-adaptive sensing, we compute the non-adaptive MSE as a special case of two-stage sensing by setting $\lambda = 0$ in (7), i.e., by skipping the first observation stage. Then the state remains unchanged from its initial value, $\mathbf{s}(1) = \mathbf{s}(0)$, and the non-adaptive MSE $J_0^{\mathrm{na}}(\mathbf{s}(0))$ is equal to $J_1^*(\mathbf{s}(0))$. By symmetry, the optimal solution to (6) is to allocate one unit of effort to each component. Hence we obtain

$$\frac{1}{N\sigma^2} J_0^{\mathrm{na}}(\mathbf{s}(0)) = \frac{\nu^2}{\sigma^2} \frac{p}{\nu^2/\sigma^2 + 1} = \frac{p}{1+r} \tag{17}$$

using the definition of r. Comparing (11) and (17) results in the following bound on the optimal two-stage gain $G = J_0^{\mathrm{na}}(\mathbf{s}(0))/J_0^*(\mathbf{s}(0))$:

$$\lim_{N\to\infty} G \geq 1 + \frac{r}{r+1} \max_{\lambda\in[0,1]} (\mathrm{BC}_p^{-2} - 1)(1-\lambda). \tag{18}$$

Further analysis of the two-stage MSE $J_0^*(\mathbf{s}(0))$ and gain G requires computing the Bhattacharyya coefficient BC_p. Unfortunately, a closed-form expression for BC_p is not available when $p \in (0, 1)$. We consider first the limiting behavior at low and high SNR. At low SNR ($\nu^2 \to \infty$) or as $\lambda \to 0$, it can be seen from (5) that the distributions $f_1(y_i(1))$ and $f_0(y_i(1))$ become more diffuse and similar to each other. Therefore the Bhattacharyya coefficient BC_p and the gain G both approach 1 at low SNR as expected. At high SNR ($\nu^2 \to 0$, $r \to \infty$), $f_1(y_i(1))$ and $f_0(y_i(1))$ become increasingly concentrated and their overlap tends to zero. Hence from (10) and (8), BC_p approaches $\sqrt{p}$. If we take $\lambda \to 0$ at a slower rate than ν^2 so that ν^2/λ still approaches zero, then it can be seen from (18) that G tends to $1/p$ at high SNR. This is the same gain as that of the oracle, which has full knowledge of the true signal support and distributes effort only over the support. Thus we have confirmed analytically that two-stage adaptive sensing can approach the performance of the oracle at high SNR, as observed previously [1], [2].

A more detailed analytical result can be obtained by bounding BC_p in terms of BC_0, the Bhattacharyya coefficient between $f_1(y_i(1))$ and $f_0(y_i(1))$. Using the definitions of r and s, the latter coefficient can be computed in closed form as

$$\mathrm{BC}_0 = \sqrt{\frac{2\sqrt{1+r\lambda}}{2+r\lambda}} \exp\left(-\frac{sr\lambda}{4(2+r\lambda)}\right), \tag{19}$$

since it involves two Gaussian distributions as specified in (5). The following proposition provides upper bounds on BC_p in terms of BC_0. Substituting the upper bounds in place of BC_p in Theorem 1 and (18) yields bounds on the two-stage MSE and gain that are weaker than before but more easily computed.

Proposition 1: For any $p \in [0,1]$, the Bhattacharrya coefficient BC_p (10) is bounded in terms of BC_0 (19) as follows:

$$\begin{aligned}\mathrm{BC}_p \leq \min\Bigg\{ & \frac{p}{\sqrt{p}+\sqrt{1-p}} + \frac{\sqrt{p(1-p)}}{\sqrt{p}+\sqrt{1-p}}\,\mathbb{P}_1(\mathcal{Y}_1) \\ & + \mathrm{BC}_0\left(\frac{1-p}{\sqrt{p}+\sqrt{1-p}} + \frac{\sqrt{p(1-p)}}{\sqrt{p}+\sqrt{1-p}}\,\mathbb{P}_{01}(\mathcal{Y}_0)\right), 1\Bigg\} \\ \leq & \sqrt{p} + \mathrm{BC}_0\sqrt{1-p},\end{aligned}$$

where $\mathcal{Y}_0 = \{y_i(1) : f_1(y_i(1)) < f_0(y_i(1))\}$, $\mathcal{Y}_1 = \{y_i(1) : f_1(y_i(1)) > f_0(y_i(1))\}$, $\mathbb{P}_1$ denotes probability under f_1, and $\mathbb{P}_{01}$ denotes probability under

$$f_{01}(y) = \mathcal{N}\left(y; \frac{\mu}{2+r\lambda}, \frac{2(1+r\lambda)}{2+r\lambda}\frac{\nu^2}{\lambda}\right).$$

Expressions for $\mathbb{P}_1(\mathcal{Y}_1)$ and $\mathbb{P}_{01}(\mathcal{Y}_0)$ can be given in terms of the standard Gaussian cumulative distribution function [11] but are omitted here for brevity. On the other hand, the outer bound in Proposition 1 is simpler in that it does not require computing the Gaussian CDF. A proof of Proposition 1 can be found in [11].

V. Numerical validation

We validate the bound (18) on the adaptation gain through numerical simulations. The signal dimension N is set to 10000, the prior mean μ is normalized to 1, and the prior standard deviation σ is set to $1/4$ corresponding to $s = 16$. In Fig. 1, the blue curves with crosses show the MSE reduction relative to non-adaptive sensing (the adaptation gain) of the optimal two-stage policy with second stage given by (9) and first stage by (7). The red curves show the lower bound on the gain given in (18) using the first inequality in Proposition 1 to further bound BC_p. Two sparsity levels are considered and the curves are plotted as functions of SNR $\mu^2/\nu^2 = sr$. It is seen that the analytical bound, which is an asymptotic result, approximates the finite-dimensional optimal gain fairly well. At low and high SNR, both curves approach values of 1 and $1/p$ respectively as predicted. At intermediate SNR, the maximum deviation of the two curves is 1.9 dB for $p = 0.1$ and 3.6 dB for $p = 0.01$.

VI. Conclusion and future work

This paper has provided an analytical guarantee on the performance of two-stage adaptive sensing for estimating the nonzero amplitudes in a sparse signal. The improvement compared to non-adaptive sensing is shown to be related to the detectability of nonzero components as measured by a Bhattacharyya coefficient, thus quantifying the dependence on the sparsity level, the SNR, and the sensing resource budget. Future work will consider the extension of the techniques in this paper to estimation loss functions other than MSE, to analogous policies for adaptive spectrum sensing [10], to other tasks such as signal support recovery, and to direct analysis of policies with more than two stages. In addition, it would be desirable to determine fundamental limits on adaptive sensing, along the lines of [12], [13], for the problem considered in this work.

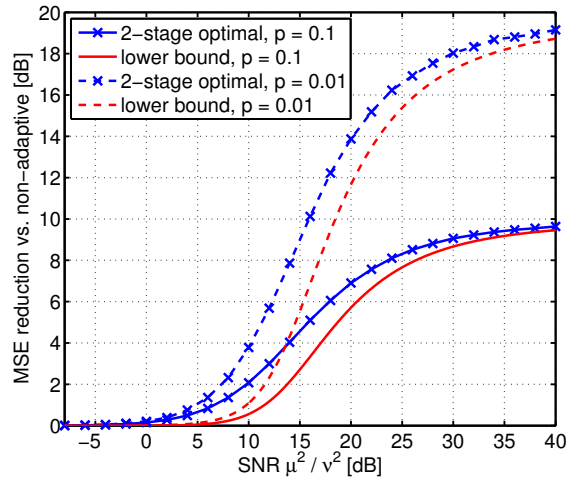

Fig. 1. Optimal two-stage MSE reduction (blue curves with crosses) and corresponding lower bounds ((18), red curves) as functions of SNR $\mu^2/\nu^2 = sr$ for two different sparsity levels.

Acknowledgement

This research was partially supported by ARO grant W911NF-11-1-0391.

References

[1] D. Wei and A. O. Hero, "Multistage adaptive estimation of sparse signals," *IEEE J. Sel. Topics Signal Process.*, 2013, to appear.

[2] E. Bashan, R. Raich, and A. O. Hero, "Optimal two-stage search for sparse targets using convex criteria," *IEEE Trans. Signal Process.*, vol. 56, pp. 5389–5402, Nov. 2008.

[3] J. Haupt, R. M. Castro, and R. Nowak, "Distilled sensing: Adaptive sampling for sparse detection and estimation," *IEEE Trans. Inf. Theory*, vol. 57, pp. 6222–6235, Sep. 2011.

[4] M. Malloy and R. Nowak, "On the limits of sequential testing in high dimensions," in *Conf. Rec. Asilomar Conf. Signals Syst. Comput.*, Nov. 2011, pp. 1245–1249.

[5] A. Tajer, R. M. Castro, and X. Wang, "Adaptive sensing of congested spectrum bands," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 6110–6125, Sep. 2012.

[6] M. A. Iwen and A. H. Tewfik, "Adaptive strategies for target detection and localization in noisy environments," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2344–2353, May 2012.

[7] M. L. Malloy and R. D. Nowak, "Near-optimal adaptive compressed sensing," in *Conf. Rec. Asilomar Conf. Signals Syst. Comput.*, Nov. 2012, pp. 1935–1939.

[8] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed. Nashua, NH: Athena Scientific, 2005, vol. 1.

[9] K. L. Jenkins and D. A. Castañón, "Adaptive sensor management for feature-based classification," in *Proc. IEEE Conf. Decision and Control (CDC)*, Dec. 2010, pp. 522–527.

[10] D. Wei and A. O. Hero, "Adaptive spectrum sensing and estimation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Vancouver, Canada, May 2013.

[11] ——, "Performance guarantees for adaptive estimation of sparse signals," 2013, in preparation.

[12] E. Arias-Castro, E. J. Candes, and M. A. Davenport, "On the fundamental limits of adaptive sensing," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 472–481, Jan. 2013.

[13] R. M. Castro, "Adaptive sensing performance lower bounds for sparse signal estimation and testing," Mar. 2013, arXiv:1206.0648.