IDENTIFYING CORRELATED COMPONENTS IN HIGH-DIMENSIONAL MULTIVARIATE GAUSSIAN MODELS

Jun Geng[1] *and Weiyu Xu*[2] *and Lifeng Lai*[3]

[1] School of Elec. and Info. Engr., Harbin Inst. of Tech., Harbin, China, jgeng@hit.edu.cn
[2] Dept. of Elec. and Comp. Engr., University of Iowa, Iowa City, IA, USA, weiyu-xu@uiowa.edu
[3] Dept. of Elec. and Comp. Engr., University of California, Davis, CA, USA, lflai@ucdavis.edu

ABSTRACT

In this paper, the problem of identifying correlated components in a high-dimensional Gaussian vector is considered. In the setup considered, instead of having to take a full-vector observation at each time index, the observer is allowed to observe any subset or full set of components in the vector, and he has the freedom to design his sampling strategies over time. The observer aims to find an optimal sampling strategy and a decision rule to maximize the error exponent (per sample). We focus on sequential strategies, in which the sampling actions depend on the observations taken so far. We first derive performance bounds of any sequential sampling strategy. We then design a low complexity procedure called sequential diagonal procedure. We show that this low complexity sequential procedure substantially outperforms the optimal non-adaptive strategy when the strength of the signal is strong.

Index Terms— correlation structure; error exponent; sequential method; spiked Gaussian model

1. INTRODUCTION

As a crucial modern data analysis technique, covariance matrix structure inference arises in a range of applications in bioinformatics, image processing, wireless communication, climate studies, etc. In this paper, the problem of correlated components identification in Gaussian vectors is considered. This paper extends our previous results in [1] by considering adaptive sensing procedures.

Specifically, the underlying Gaussian vector is generated from a spiked signal model [2–5]; hence, it possesses a special correlation structure. It is assumed that s out of p components are correlated in the underlying vector, and the goal is to identify these s correlated components via a sequence of observations. In contrast to most of existing works, in which the observer makes his inference based on a sequence of full-vector observations, this paper considers the scenario that the observer has the freedom to design his sampling strategy and is capable of observing any subset of components in the underlying signal. The observer aims to select the most efficient sampling strategy to maximize the error exponent of the false identification (per sample). In this paper, we focus on adaptive sampling strategy, which is a strategy that casually depends on observations. Since the observer is allowed to adjust his sampling strategy whenever he obtains a new observation, adaptive strategies are expected to outperform non-adaptive ones. In this paper, we first present performance bounds of the optimal adaptive sampling strategy, and then we propose a low complexity adaptive algorithm named as sequential diagonal method. We show that the error exponent of the proposed adaptive method is in inverse proportion to $(2 + s)$, which significantly outperforms the optimal non-adaptive method, when the signal is strong.

There are extensive works related to the research in this paper. We only mention a few of them here. [6] adopts covariance thresholding for sparse principal component analysis. [7–9] focus on detecting a sparse correlation structure in covariance matrix. [10] considers the sequential detection problem in Gauss-Markov random fields via adaptive sampling. Different from these works, in our work the observer focuses on estimating the support of correlated components rather than detecting the existence of the correlation structure. The adaptive sampling strategies studied in this paper is also related to the active sensing problem such as [11], the sequential testing problem such as [12] and the sparse signal detection and estimation [13].

The remainder of this paper is organized as follows. The mathematical model is given in Section 2. Section 3 reviews the relevant result of non-adaptive strategy. Section 4 discusses adaptive sampling strategies. Section 5 illustrates numerical simulation results. Finally, Section 6 offers concluding remarks.

Notations: (random) vectors will be presented by bold face lower-case letters, and their component by $[\cdot]_n$. Matrices will be presented by bold face upper-case letters, and their component by $[\cdot]_{m,n}$. A set is denoted by a calligraphic character, and its cardinality by $|\cdot|$; $||\cdot||_1$ denotes the ℓ_1 norm of a vector.

2. MODEL

In this paper, we aim to identify all correlated components in a spiked signal model. Particularly, the underlying Gaussian vector $\mathbf{x}_k \in \mathbb{R}^p$ is modeled as

$$\mathbf{x}_k = \sigma g_k \mathbf{v} + \mathbf{z}_k, \quad k = 1, 2, \ldots, \tag{1}$$

The work of J. Geng was supported by the Fundamental Research Funds for the Central Universities under grant AUGA5710013915. The work of L. Lai was supported by the National Science Foundation under grant CNS-1660128.

where σ is the strength of the signal, $g_k \sim \mathcal{N}(0,1)$ models the uncertainty of signal amplitude, and $\mathbf{z}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is the observation noise with $\mathbf{I}$ being the identity matrix. g_k and $\mathbf{z}_k$ are independent to each other, and both of them are independent and identically distributed (i.i.d.) over k. $\mathbf{v} \in \mathbb{R}^p$ describes the direction of the underlying signal $\mathbf{x}_k$. In this paper, we assume that the components in $\mathbf{v}$ take values in $\{0,1\}$, and $\mathbf{v}$ has s non-zero components. Denote $\mathcal{S}$ as the set of indices of the non-zero components in $\mathbf{v}$. It is easy to verify that the mean of $\mathbf{x}_k$ is $\mathbf{0}$ and the covariance matrix $\mathbf{C}_{\mathcal{S}}$ is

$$[\mathbf{C}_{\mathcal{S}}]_{m,n} = \begin{cases} 1, & m=n, \quad m,n \notin \mathcal{S} \\ 1+\sigma^2, & m=n, \quad m,n \in \mathcal{S} \\ \sigma^2, & m \neq n, \quad m,n \in \mathcal{S} \\ 0, & \text{otherwise} \end{cases}. \tag{2}$$

Hence, $\mathbf{x}_k$ contains s correlated components.

The observer does not know which components are in $\mathcal{S}$. Hence, there are $\binom{p}{s}$ possible choices for $\mathcal{S}$. The problem then is modeled as a multi-hypothesis testing problem:

$$\mathcal{H}_i : \mathcal{S} = \mathcal{S}_i, \quad i = 1, \ldots, \binom{p}{s}.$$

In addition, denote P_i as the probability measure of $\mathbf{x}_k$ under $\mathcal{H}_i$, and denote $\pi_{i,0}$ as the prior probability of $\mathcal{H}_i$. We assume that $\pi_{i,0} > 0$ for all i. To ease the notation, we use $\mathbf{C}_i$ instead of $\mathbf{C}_{\mathcal{S}_i}$ to denote the covariance matrix under $\mathcal{H}_i$.

Instead of having to take a full-vector observation of $\mathbf{x}_k$ in each time slot, we assume that the observer is allowed to observe any subset of the components in $\mathbf{x}_k$. Let

$$\mathcal{U} := \{\mathbf{u} : \mathbf{u} \in \{0,1\}^p, \mathbf{u} \neq \mathbf{0}\}$$

be the set of all possible sampling strategies for the observer. For a strategy $\mathbf{u} \in \mathcal{U}$, $[\mathbf{u}]_n = 1$ indicates that the n^{th} component of the underlying signal is observed, while $[\mathbf{u}]_n = 0$ indicates the opposite. Note that $|\mathcal{U}| = 2^p - 1$. We denote the members in $\mathcal{U}$ as $\mathbf{u}_r, r = 1, \ldots, 2^p - 1$.

The observer takes observations in a sequential manner. Let $\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_\tau\}$ be a sampling action taken by the observer, in which $\mathbf{a}_k \in \mathcal{U}$ is the sampling strategy adopted in time slot k, and τ is the time when the observer stops taking observations. Throughout the sampling action, we assume that the observer can totally observe at most N components:

$$\sum_{k=1}^{\tau} ||\mathbf{a}_k||_1 = N. \tag{3}$$

Let $\{\mathbf{y}_k, k = 1, \ldots, \tau\}$ be the observation sequence, in which

$$[\mathbf{y}_k]_n = \begin{cases} [\mathbf{x}_k]_n & \text{if } [\mathbf{a}_k]_n = 1 \\ \phi & \text{otherwise} \end{cases}. \tag{4}$$

$\mathcal{A}$ is called adaptive if $\mathbf{a}_k$ casually depends on $\{\mathbf{y}_1, \ldots, \mathbf{y}_{k-1}\}$; non-adaptive if $\mathbf{a}_k$ is pre-determined. Denote $\mathbb{A}_N$ as the set of all possible sampling actions satisfying (3).

Our goal is to find the most efficient sampling actions to maximize the error exponent for this multi-hypothesis testing problem, that is, to solve

$$\lim_{N\to\infty} -\frac{1}{N} \log \min_{\mathcal{A}\in\mathbb{A}_N, \delta_\tau} P_e(\delta_\tau, \mathcal{A}), \tag{5}$$

in which δ_τ is the terminal decision rule, and $P_e(\delta_\tau, \mathcal{A}) := \sum_{i=1}^{\binom{p}{s}} \pi_{i,0} P_i(\delta_\tau \neq i)$ is the error probability. We note that the error exponent defined in (5) modifies the classic definition of the error exponent, which normalizes the logarithm of error probability by the number of full observations, by replacing the normalization factor with the number of total measurements N.

3. BACKGROUND FROM OPTIMAL NON-ADAPTIVE SAMPLING STRATEGY

Non-adaptive sampling strategies for the problem formulated in Section 2 has been discussed in our previous paper [1]. In this section, we review the results that are relevant to this paper.

Let $\pi_{i,k} = P(\mathcal{H}_i \text{ is true} | \mathbf{y}_1, \ldots, \mathbf{y}_k)$ be the posterior probability of $\mathcal{H}_i$ being true. Since the proposed problem is under Bayesian framework, it can be shown that the optimal decision rule (for both adaptive and non-adaptive strategies) is the maximum posterior probability rule, i.e.,

$$\delta_\tau^* = \operatorname*{argmax}_i \pi_{i,\tau}. \tag{6}$$

Let $f_i^{\mathbf{u}_r}(\cdot)$ be the conditional pdf for $\mathbf{y}_k$ given $\mathbf{u}_r$ under $\mathcal{H}_i$. Under the optimal decision rule, it can be further shown that

$$\lim_{N\to\infty} -\frac{1}{N} \log \min_{\mathcal{A}\in\mathbb{A}_N} P_e(\delta_\tau^*, \mathcal{A}) = \lim_{N\to\infty} \frac{1}{N} \max_{\mathcal{A}\in\mathbb{A}_N} \min_{i<j} \max_{\lambda\in[0,1]} \sum_{k=1}^{\tau} -\log \int f_i^{\mathbf{a}_k}(\mathbf{y}_k)^\lambda f_j^{\mathbf{a}_k}(\mathbf{y}_k)^{1-\lambda} d\mathbf{y}_k. \tag{7}$$

We note that

$$C(f_i^{\mathbf{u}_r}, f_j^{\mathbf{u}_r}; \lambda) := -\log \int f_i^{\mathbf{u}_r}(\mathbf{y})^\lambda f_j^{\mathbf{u}_r}(\mathbf{y})^{1-\lambda} d\mathbf{y} \tag{8}$$

is the Chernoff bound between $\mathcal{H}_i$ and $\mathcal{H}_j$ (given $\mathbf{u}_r$). Therefore, an intuitive explanation of (7) is that *to maximize the error exponent in* (5) *is equivalent to find the best strategy* $\mathcal{A}$ *that maximizes the worst pair-wise error exponent*.

Pairwise hypotheses are classified into different categories by the number of different elements in their support sets $\mathcal{S}_i$ and $\mathcal{S}_j$. In particular, let

$$\mathcal{D}_d := \{(\mathcal{H}_i, \mathcal{H}_j) : |\mathcal{S}_i \triangle \mathcal{S}_j| = 2d\}, \tag{9}$$

where $\triangle$ denotes the symmetric difference of two sets, be the set of pair-wise hypotheses with d pair different components. [1] has discussed the optimal solution for the following problem

$$\max_{\mathcal{A}\in\mathbb{A}_N} \min_{(i,j)\in\mathcal{D}_1} \max_{\lambda\in[0,1]} \sum_{k=1}^{\tau} -\log \int f_i^{\mathbf{a}_k}(\mathbf{y}_k)^\lambda f_j^{\mathbf{a}_k}(\mathbf{y}_k)^{1-\lambda} d\mathbf{y}_k. \tag{10}$$

Specifically, the optimal value of (10) is given in the following theorem:

Theorem 3.1. *1) The optimal value of* (10) *is* NT_{l^*}*, in which* $l^* = \operatorname{argmax} T_l$*.* $T_l, l = 1, \ldots, p$*, is calculated as*

$$T_l = \frac{1}{l}\frac{1}{\binom{p}{l}}\left[\sum_{m=0}^{s-1}\binom{2}{2}\binom{s-1}{m}\binom{p-s-1}{l-m-2}K_{2m+2} + \sum_{m=0}^{s-1}\binom{2}{1}\binom{s-1}{m}\binom{p-s-1}{l-m-1}K_{2m+1}\right], \quad (11)$$

in which

$$K_{2m-1} = \frac{1}{2}\log\frac{4+2(2m-1)\sigma^2+(m-1)\sigma^4}{4(1+(m-1)\sigma^2)^{1/2}(1+m\sigma^2)^{1/2}},$$
$$K_{2m-2} = \frac{1}{2}\log\frac{(2+\sigma^2)[2+(2m-3)\sigma^2]}{4[1+(m-1)\sigma^2]}.$$

2)(Asymptotic Performance) When $\sigma^2 \to 0$*, we have* $l^* = p$*, and*

$$T_p = \frac{1}{p}\frac{2s-1}{8}\sigma^4 + o(\sigma^4).$$

When $\sigma^2 \to \infty$*, we have*

$$T_{l^*} \approx \left(\frac{3}{2} - \frac{1}{2^s}\right)\frac{1}{2p}\log\sigma^2(1+o(1)). \quad (12)$$

Note that (7) considers the worst case error exponent over all pairwise hypotheses but (10) only focuses on the pairwise hypotheses within $\mathcal{D}_1$; hence the error exponent for the optimal non-adaptive strategy is upper-bounded by the optimal solution presented in Theorem 3.1. Specifically, we have

$$\lim_{N\to\infty} -\frac{1}{N}\log\min_{\mathcal{A}\in\mathbb{A}_N,\delta_\tau} P_e(\delta_\tau, \mathcal{A}) \le T_{l^*}. \quad (13)$$

Hence, we can conclude that *the error exponent for the optimal non-adaptive strategy is in inverse proportion to* p *when* σ^2 *is large*.

4. ADAPTIVE SAMPLING STRATEGIES

We note that for any fixed N, the optimal adaptive sampling action and the optimal value can be solved by the dynamic programming method (DP). However, DP usually leads to a complex recursive structure, which provides little insight on the implementation of adaptive sampling strategies. In this section, we first present performance bounds of the optimal adaptive sampling action; and then we propose a low complexity adaptive sampling action for the case of large σ^2.

4.1. Performance Bounds

Lemma 4.1. *If* $\mathcal{A}^*$ *is the optimal adaptive sampling action, then*

$$\lim_{N\to\infty} -\frac{1}{N}\log P_e(\delta_\tau^*, \mathcal{A}^*) \le \lim_{N\to\infty}\frac{1}{N}\min_{i,j}\max_{\mathcal{A}\in\mathbb{A}_N}\max_{\lambda\in[0,1]}\sum_{r=1}^{2^p-1}\rho_r C(f_i^{\mathbf{u}_r}, f_j^{\mathbf{u}_r}; \lambda). \quad (14)$$

The optimal non-adaptive sampling action certainly provides a lower bound for the optimal adaptive sampling action. Comparing with (7), the difference of the upper bound and the lower bound is to interchange the order $\max_{\mathcal{A}}$ and $\min_{i,j}$. Let

$$\mathcal{Q}_N = \frac{1}{N}\min_{(i,j)\in\mathcal{D}_1}\max_{\mathcal{A}\in\mathbb{A}_N}\max_{\lambda\in[0,1]}\sum_{r=1}^{2^p-1}\rho_r C(f_i^{\mathbf{u}_r}, f_j^{\mathbf{u}_r}; \lambda). \quad (15)$$

The limit of $\mathcal{Q}_N$ provides an upper bound for the optimal adaptive sampling action. The optimal solution of (15) and corresponding value for $\mathcal{Q}_N$ is presented in the following theorem.

Theorem 4.2. *1)(Optimal solution) For an arbitrarily given pair* $(\mathcal{H}_i, \mathcal{H}_j) \in \mathcal{D}_1$*, let*

$$\lambda_r^{(i,j)*} := \operatorname*{argmax}_{\lambda} C(f_i^{\mathbf{u}_r}, f_j^{\mathbf{u}_r}; \lambda) \quad (16)$$

for $r = 1, \ldots, 2^p - 1$*, and let*

$$r^{(i,j)*} := \operatorname*{argmax}_{r}\left\{C\left(f_i^{\mathbf{u}_r}, f_j^{\mathbf{u}_r}; \lambda_r^{(i,j)*}\right)/||\mathbf{u}_r||_1\right\}. \quad (17)$$

Then $(\mathcal{H}_i, \mathcal{H}_j)$*,* $\lambda^{(i,j)*}$ *and* $\mathcal{A}^{(i,j)*} = \left\{\rho_1^{(i,j)*}, \ldots, \rho_{2^p-1}^{(i,j)*}\right\}$ *is an optimal solution for* $\mathcal{Q}_N$*, in which*

$$\lambda^{(i,j)*} := \lambda_{r^{(i,j)*}}^{(i,j)*}, \quad (18)$$

$$\rho_r^{(i,j)*} := \begin{cases} N/||\mathbf{u}_r||_1 & \text{if } r = r^{(i,j)*} \\ 0 & \text{if } r \neq r^{(i,j)*} \end{cases}. \quad (19)$$

2)(Asymptotic performance bound) When $\sigma^2 \to 0$*,*

$$\lim_{N\to\infty}\mathcal{Q}_N = \frac{1}{s+1}\frac{2s-1}{8}\sigma^4 + o(\sigma^4). \quad (20)$$

When $\sigma^2 \to \infty$*,*

$$\lim_{N\to\infty}\mathcal{Q}_N = \frac{1}{2}\log\sigma^2(1+o(1)). \quad (21)$$

Proof Outline: For an arbitrarily given pair $(\mathcal{H}_i, \mathcal{H}_j) \in \mathcal{D}_1$, we have

$$\begin{aligned}&\max_{\mathcal{A}\in\mathbb{A}_N}\max_{\lambda\in[0,1]}\sum_{r=1}^{2^p-1}\rho_r C(f_i^{\mathbf{u}_r}, f_j^{\mathbf{u}_r}; \lambda)\\ &\le \max_{\mathcal{A}\in\mathbb{A}_N}\sum_{r=1}^{2^p-1}\rho_r||\mathbf{u}_r||_1\frac{\max_{\lambda\in[0,1]} C(f_i^{\mathbf{u}_r}, f_j^{\mathbf{u}_r}; \lambda)}{||\mathbf{u}_r||_1}\\ &\le N\max_{\mathbf{u}_r\in\mathcal{U}}\max_{\lambda\in[0,1]}\frac{C(f_i^{\mathbf{u}_r}, f_j^{\mathbf{u}_r}; \lambda)}{||\mathbf{u}_r||_1}. \end{aligned} \quad (22)$$

Since both two inequalities in (22) become equalities for the solutions presented in this theorem, (18) and (19) are optimal for $\mathcal{Q}_N$. Since the Chernoff bound of two Gaussian distributions has close form expressions [14], the asymptotic upper bound can be then derived following a direct calculation.

We comment that in general this upper bound is not tight as the optimal strategy is designed for a specific $(\mathcal{H}_i, \mathcal{H}_j) \in \mathcal{D}_1$. However, this result indicates that adaptive sampling actions has the potential to significantly improve the error exponent.

4.2. A Low Complexity Adaptive Algorithm

In this subsection, we propose a low complexity adaptive sampling strategy. The motivating observation is that each component of the underlying signal $\mathbf{x}_k$ only has two possible pdfs: if $n \in \mathcal{S}$, the distribution of $[\mathbf{x}_k]_n$, denoted as h_1, is $\mathcal{N}(0, 1+\sigma^2)$; otherwise, the distribution, denoted as h_0, is $\mathcal{N}(0,1)$. In the proposed algorithm, the observer adopts the sequential probability ratio test (SPRT) to decide the distribution of each component. In particular, let τ_{n-1} be the time that the distribution of the $(n-1)^{th}$ component is decided, τ_n is defined as follows:

$$\tau_n := \inf\left\{\kappa > \tau_{n-1} : \prod_{k=\tau_{n-1}+1}^{\kappa} \frac{h_1(y_k)}{h_0(y_k)} \notin (A, B)\right\}$$

for $n = 1, \ldots, p$, in which A and B are properly chosen thresholds, the initial time $\tau_0 = 0$ and the observation

$$y_k = [\mathbf{x}_k]_n \quad \text{for } \tau_{n-1} < k \leq \tau_n. \tag{23}$$

The distribution of the n^{th} component in $\mathbf{x}$ is decided by

$$\delta_{\tau_n} := \begin{cases} \text{claim } h_1 & \text{if } \prod_{k=\tau_{n-1}+1}^{\tau_n} \frac{h_1(y_k)}{h_0(y_k)} \geq B \\ \text{claim } h_0 & \text{if } \prod_{k=\tau_{n-1}+1}^{\tau_n} \frac{h_1(y_k)}{h_0(y_k)} \leq A \end{cases}. \tag{24}$$

That is, the observer keeps observing the n^{th} component in the underlying signal as long as the statistic of SPRT within (A, B). The observer claims that the n^{th} component belongs to $\mathcal{S}$ if the SPRT statistic climbs over the upper bound and does not belong to if it falls below the lower bound. The observer switches to observe the next component whenever he makes a decision on the current observing component. *The observer stops the detection procedure when either all the components in the underlying signal are decided or all the sampling rights are exhausted. The observer makes errors when he 1) makes wrong decisions on the distribution of any components or 2) exhausts all the sampling rights before all the components are declared.*

Johnstone proposed a diagonal thresholding method in [15], in which the components with largest s sample variance are selected to estimate the eigenvectors, for the spiked signal model. Above proposed method can be viewed as a sequential version of Johnstone's method since the SPRT statistic in our scenario can be explicitly written in terms of sample variance. The performance of the proposed method is given in the following theorem:

Theorem 4.3. *When $\sigma^2 \to \infty$, by setting thresholds*

$$A = B^{-1} = \sigma \exp\{-N/(s+2)\}, \tag{25}$$

the error exponent of the sequential diagonal method is

$$\lim_{N\to\infty} -\frac{1}{N} \log P_e \geq \frac{1}{2(2+s)} \log \sigma^2 (1 + o(1)). \tag{26}$$

Proof Outline: Let α and γ be the Type I and Type II error of SPRT, respectively. By union bound inequality, the error probability is bounded by

$$P_e \leq P(\tau_p > N) + s\gamma + (p-s)\alpha. \tag{27}$$

We can show that the detection delay of SPRT decays exponentially given thresholds A and B. In particular,

$$P_1(\tau_n - \tau_{n-1} > N) \leq \exp\{-\lambda_1 \log \alpha\} \exp\{-Nt_1\},$$
$$P_0(\tau_n - \tau_{n-1} > N) \leq \exp\{-\lambda_0 \log \gamma\} \exp\{-Nt_0\},$$

where λ_0 and λ_1 are two constants within $(0,1)$. t_0 and t_1 are constants determined by λ_0, λ_1 and σ. As a result, $P(\tau_p > N) = P\left(\sum_{n=1}^{p}(\tau_n - \tau_{n-1}) > N\right)$ also decays exponentially. Its bound can be characterized by α, γ, λ_0, λ_1 and σ. By setting $\lambda_0 = 0.5$, $\lambda_1 = 1/\sigma$ and $\alpha = \gamma$, this theorem can be achieved by solving (27).

Note that the error exponent of the optimal non-adaptive strategy is in inverse proportion to p when σ^2 is large; however that of the sequential diagonal method is in inverse proportion to $(2+s)$. When $s << p$, the proposed adaptive strategy outperforms the optimal non-adaptive strategy significantly.

5. SIMULATION

In this section, we conduct a numerical simulation to compare the performance of the optimal non-adaptive strategy and the proposed sequential diagonal thresholding strategy. The simulation result is shown in Figure 1. In this simulation, we set $p = 7$ and $s = 3$. The red dot-dash line is the performance of the optimal value presented in Theorem 3.1, and the blue dash line is that of the diagonal thresholding method. When σ is small, non-adaptive strategy outperforms the sequential diagonal method since non-adaptive sampling strategy explores the correlation information within the observation vector. However, the sequential diagonal method outperforms the optimal non-adaptive strategy when σ is large. This is because the information within the margin distribution of each element is sufficient to classify the its distribution, and to sample one element in each time is a more efficient sampling method.

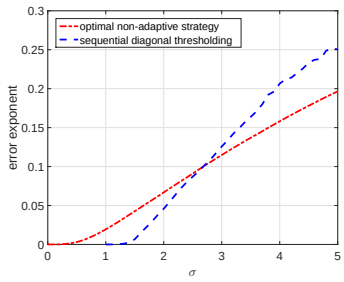

Fig. 1. Performance of the sequential diagonal thresholding method

6. CONCLUSION

In this paper, we have considered the problem of identifying correlated components in a Gaussian vector. We have derived performance bounds of any sequential sampling strategy. We have proposed a low complexity procedure named sequential diagonal procedure, which substantially outperforms the optimal non-adaptive strategy when the signal is strong.

7. REFERENCES

[1] J. Geng, W. Xu, and L. Lai, "Detection of correlated components in multivariate Gaussian models," in *Proc. IEEE Inform. Theory Workshop*, (Jeju Island, Korea), pp. 224–228, Oct. 2015.

[2] D. Paul, "Asymptotics of sample eigenstructure for a large dimensional spiked covariance model," *Statistica Sinica*, vol. 17, pp. 1617–1642, Apr. 2007.

[3] A. A. Amini and M. J. Wainwright, "High-dimensional analysis of semidefinite relaxations for sparse principal compenents," *Annals of Statistics*, vol. 37, pp. 2877–2921, Mar. 2009.

[4] T. T. Cai, Z. Ma, and Y. Wu, "Optimal estimation and rank detection for sparse spiked covariance matrices," *Probability Theory and Related Fields*, vol. 161, pp. 781–815, Apr. 2014.

[5] I. M. Johnstone, "On the distribution of the largest engenvalue in principal component analysis," *Annals of Statistics*, vol. 29, pp. 295–327, Apr. 2001.

[6] Y. Deshpande and A. Montanari, "Sparse PCA via covariance thresholding," in *Advances in Neural Information Processing Systems*, (Montreal, CA), pp. 334–342, Dec. 2014.

[7] E. A. Castro, S. Bubeck, and G. Lugosi, "Detection of correlations," *Annals of Statistics*, vol. 40, pp. 412–435, June 2011.

[8] E. A. Castro, S. Bubeck, and G. Lugosi, "Detecting positive correlations in a multivariate sample," *Bernoulli*, vol. 21, pp. 209–241, Feb. 2015.

[9] R. M. Castro, G. Lugosi, and P. A. Savalle, "Detection of correlations with adaptive sensing," *IEEE Trans. Inform. Theory*, vol. 60, pp. 7913–7927, Nov. 2014.

[10] J. Heydari, A. Tajer, and H. V. Poor, "Quickest detection of Gauss-Markov random fields," (Allerton, IL, USA), pp. 808–814, Oct. 2015.

[11] S. Nitinawarat, G. Atia, and V. V. Veeravalli, "Controlled sensing for multihypothesis testing," submitted, available at arXiv:1205.0858.

[12] M. Malloy and R. Nowak, "Sequential testing for sparse recovery," *IEEE Trans. Inform. Theory*, vol. 60, pp. 7862–7873, Oct. 2014.

[13] R. Castro, "Adaptive sensing performance lower bounds for sparse estimation and testing," *Bernoulli*, vol. 20, pp. 2217–2246, 2014.

[14] F. Nielsen, "Chernoff information of exponential families," 2011. submitted, available at arXiv:1102.2684.

[15] I. M. Johnstone and A. Y. Lu, "Sparse principal components analysis," *Unpublised*, 2004.