

# Sequential Change Detection of a Correlation Structure under a Sampling Constraint

Anamitra Chaudhuri  
Department of Statistics  
University of Illinois at Urbana-Champaign  
Champaign, Illinois 61820  
E-mail: ac34@illinois.edu

Georgios Fellouris  
Department of Statistics  
Coordinated Science Laboratory  
University of Illinois at Urbana-Champaign  
Champaign, Illinois 61820  
E-mail: fellouri@illinois.edu

Ali Tajer  
Department of ECSE  
Rensselaer Polytechnic Institute  
Troy, New York 12180  
E-mail: tajer@ecse.rpi.edu

**Abstract**—The problem of sequentially detecting a change in the correlation structure of multiple Gaussian information sources is considered when it is possible to sample only two of them at each time instance. It is assumed that all sources are initially independent and that at least two of them become positively correlated after the change. The problem is to stop sampling as quickly as possible after the change, while controlling the false alarm rate and without assuming any prior information on the number of sources that become correlated. A joint sampling and change-detection rule is proposed and is shown to achieve the smallest possible worst-case conditional expected detection delay among all processes that satisfy the same constraints, to a first order approximation as the false alarm rate goes to 0, for any possible number of post-change correlated sources.

## I. INTRODUCTION

In statistical decision-making, growth in the complexity, dimension, and scale of the data, in turn, increases the data-acquisition and computational costs, if not rendering them prohibitive altogether. Enforcing proper data-acquisition (sampling) constraints is a natural measure to contain such costs. For example, when the data of interest become available at multiple locations or sources, it may be practical and economically efficient to take observations from only a small fraction of these locations. There are many areas in science and engineering where these type of scenarios arise, e.g., sensor networks, surveillance system, cyber security, power grids etc. Such sampling constraints have been considered in the sequential anomaly detection problem [1], [2], [3], [4], where the processes of interest are assumed to be statistically independent and the goal is to identify the anomalous ones. Similar constraints are imposed in [5], [6], where only one source is sampled at each time instance and observations from different sources are assumed to exhibit temporal dependence. Moreover, such sampling constraints can be embedded into the more general framework of the sequential design of experiments in sequential testing [7], [8].

A related problem to sequential testing is that of sequential change detection, where the goal is to detect as quickly as possible a change in the distribution of the underlying process [9], [10]. Sequential change detection problem with multiple sources of observations under sampling constraints have been considered in [11], [12], [13]. In the first two references, the

sampled processes are statistically independent, whereas in the latter the change can affect the dependence structure of a subset of these processes. Some other relevant works in sequential setup that involve the covariance structure of the processes are [14], [15], [16].

In this paper we assume that there are multiple (Gaussian) sources, which are initially independent, and at some unknown time at least two of them become positively correlated, while their marginal distributions remain unaltered. The problem is to detect this change as quickly as possible, while controlling the false alarm rate, when it is possible to sample *only two* sources at each sampling instance, and without assuming any prior information on the number of sources that become correlated.

We propose a joint sampling and detection rule for this problem and obtain an explicit, non-asymptotic upper bound for its worst-case conditional expected detection delay, as measured by Lorden's criterion [17]. Most importantly, we show that this upper bound agrees, to a first-order asymptotic approximation as the false alarm rate goes to 0, with the best performance that can be achieved by rules that satisfy the same sampling and false alarm constraints, *for any possible number of sources that may become correlated*. Finally, we present the results of a simulation study that reveals that the upper bound is quite sharp, especially for low false alarm rates, for any number of the unknown post-change correlated sources.

The proposed sequential procedure is similar to the one proposed in [12], however there are also some important differences. To be more precise, in [12] it is assumed that the observed processes are independent, that it is possible to sample only one of them at each sampling instance, and that the *marginal* distribution of *only one of them* changes. Moreover, the proposed scheme in this work is shown to achieve *up to a constant term*, as the false alarm rate goes to 0, the worst-case detection delay that is achievable even when all processes are observed at all times.

On the contrary, the observed processes in our work are not independent, as the change affects their joint distribution. Moreover, we assume complete ignorance regarding the number of sources that become correlated. As a result, in order to establish the desired asymptotic optimality property for every

possible set of post-change correlated sources, we develop a lower bound for the performance of any rule that satisfies the same false alarm *and sampling* constraints, i.e., we do not compare the performance of the proposed procedure with the optimal performance in the full-sampling case.

## II. PROBLEM FORMULATION

Consider  $p$  information sources that generate a sequence of independent, Gaussian,  $p$ -dimensional random vectors

$$\mathbf{X}(t) := (X^1(t), \dots, X^p(t)), \quad t \in \mathbb{N} := \{1, 2, \dots\}.$$

We assume that  $X^i(t)$  has mean zero and variance 1 for every  $i \in [p] := \{1, \dots, p\}$  and every  $t \in \mathbb{N}$ . All sources are initially independent and *at least two* of them become positively correlated at some unknown time  $\nu$ . Specifically, if  $\text{Var}[\mathbf{X}(t)]$  represents the true covariance matrix of  $\mathbf{X}(t)$ , we assume that

$$\text{Var}[\mathbf{X}(t)] = \begin{cases} I_p & \text{if } t \leq \nu \\ \Sigma & \text{if } t > \nu \end{cases}. \quad (1)$$

Here,  $I_p$  is the  $p \times p$  identity matrix and  $\Sigma$  is a  $p \times p$  covariance matrix such that

$$|\mathcal{A}(\Sigma)| \geq 1 \quad \text{and} \quad \Sigma_{ij} \in \{0, \rho\} \quad \text{for all } (i, j) \in \mathcal{E}, \quad (2)$$

where  $\rho \in (0, 1)$ ,  $\mathcal{E}$  is the family of *all* ordered pairs of sources, and  $\mathcal{A}(\Sigma)$  is the subfamily of correlated pairs under  $\Sigma$ , i.e.,

$$\begin{aligned} \mathcal{E} &:= \{(i, j) : 1 \leq i < j \leq p\}, \\ \text{and } \mathcal{A}(\Sigma) &:= \{(i, j) \in \mathcal{E} : \Sigma_{ij} \neq 0\}. \end{aligned} \quad (3)$$

We assume that the post-change correlation,  $\rho$ , is specified and we interpret it as the minimum correlation value whose detection is of interest. On other hand, we do not make any assumption about how many or which entries of  $\Sigma$  are non-zero after the change. When the change happens at time  $\nu \geq 0$  and the post-change covariance matrix is  $\Sigma$ , we denote the underlying probability measure by  $\mathbb{P}_\nu^\Sigma$  and the corresponding expectation by  $\mathbb{E}_\nu^\Sigma$ . Thus, under  $\mathbb{P}_0^\Sigma$ ,  $\mathbf{X}(t), t \in \mathbb{N}$  are i.i.d. random vectors with covariance matrix  $\Sigma$ . Moreover, we denote by  $\mathbb{P}_\infty/\mathbb{E}_\infty$  the underlying probability measure/expectation when there is no change, i.e., when  $\mathbf{X}(t)$  has covariance matrix  $I_p$  for every  $t \in \mathbb{N}$ .

Note that the probability assigned by  $\mathbb{P}_0^\Sigma$  to an event that involves *only two* distinct sources is independent of  $\Sigma$ . In order to emphasize this, we will denote the underlying probability measure/expectation by  $\mathbb{P}_0/\mathbb{E}_0$  when the two sources are correlated and by  $\mathbb{P}_\infty/\mathbb{E}_\infty$  when they are independent.

The problem we consider in this work is to detect the change as quickly as possible when the data become available sequentially and we are allowed to sample *only two* sources at each time. Thus, our first task is to specify a *sampling rule*, i.e., a sequence  $S \equiv (S_t, t \in \mathbb{N})$  such that  $S_t \in \mathcal{E}$  represents the two sources that we sample at time  $t$  and it is a function of the collected observations up to the previous time instance.

More formally, let  $\mathbf{X}(t; e)$  denote the data at time  $t$  from  $e \in \mathcal{E}$ , i.e., if  $e = (k, l)$ , then

$$\mathbf{X}(t; e) := (X^k(t), X^l(t)).$$

The observed filtration induced by a sampling rule  $S$  is defined recursively as follows: given  $t \in \mathbb{N}$  and  $S_1, \dots, S_t$ , then

$$\mathcal{F}_t^S = \sigma(S_u, \mathbf{X}(u; S_u); 1 \leq u \leq t),$$

and, by assumption,  $S_{t+1}$  has to be  $\mathcal{F}_t^S$ -measurable. Our second task is to specify a rule for declaring the change with minimal delay after it occurs, while controlling the false alarm rate. Specifically, given a sampling rule  $S$ , we have to specify an  $\{\mathcal{F}_t^S\}$ -stopping time, i.e., a random time  $T$ , such that

$$\{T = t\} \in \mathcal{F}_t^S \quad \text{for every } t \in \mathbb{N},$$

at which we declare that the change has already occurred. We will refer to such a pair  $(S, T)$  as a detection policy and focus on the class of policies  $\Delta(\gamma)$  such that  $\mathbb{E}_\infty[T] \geq \gamma$ , where  $\gamma$  is a user-specified level.

Following Lorden's approach [17], for any policy  $(S, T)$  we denote by  $D^\Sigma(S, T)$  the worst-case (with respect to the change-point) conditional expected detection delay given the worst-possible history of observations up to the change-point when the *post-change covariance matrix* is  $\Sigma$ , i.e.,

$$D^\Sigma(S, T) := \sup_{\nu \geq 0} \text{esssup}_{\mathbb{P}_\nu^\Sigma} [T - \nu | \mathcal{F}_\nu^S, T > \nu]. \quad (4)$$

One of the main results of this work is that there is a policy  $(\tilde{S}, \tilde{T})$  that achieves

$$\inf_{(S, T) \in \Delta(\gamma)} D^\Sigma(S, T), \quad (5)$$

to a first order asymptotic approximation as  $\gamma \rightarrow \infty$ , for every *post-change covariance matrix*  $\Sigma$  that satisfies (2).

## III. NOTATIONS AND STATISTICS

For any pair of sources  $e \in \mathcal{E}$ , we denote by  $\lambda(t; e)$  the log-likelihood ratio of  $\mathbb{P}_0$  versus  $\mathbb{P}_\infty$  based on the observations from the sources in  $e$  at time  $t$ , i.e.,

$$\lambda(t; e) \equiv \log \frac{d\mathbb{P}_0}{d\mathbb{P}_\infty} (\mathbf{X}(t; e)).$$

For each time  $t \in \mathbb{N}$ ,  $\mathbf{X}(t; e)$  is a bivariate Gaussian random vector with zero mean. Its covariance matrix is equal to

$$\Sigma_1 := \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

under  $\mathbb{P}_0$  and equal to the identity matrix  $I_2$  under  $\mathbb{P}_\infty$ . As a result, for each  $t \in \mathbb{N}$ , we have

$$\lambda(t; e) = -\frac{1}{2} \log(1 - \rho^2) - \frac{1}{2} (\mathbf{X}(t; e))^T (\Sigma_1^{-1} - I_2) \mathbf{X}(t; e).$$

These log-likelihood ratios are identically distributed under  $\mathbb{P}_0$ , and as a result their corresponding means and variances do not depend on  $e \in \mathcal{E}$  and  $t \in \mathbb{N}$ :

$$\begin{aligned} \mathcal{I} &:= \mathbb{E}_0[\lambda(t; e)] = -\frac{1}{2} \log(1 - \rho^2), \\ \mathcal{V} &:= \mathbb{E}_0[(\lambda(t; e) - \mathcal{I})^2] = \rho^2. \end{aligned}$$

For each  $e \in \mathcal{E}$  and  $A > 0$  we also define the following stopping times

$$\begin{aligned}\delta_-^e &:= \inf \left\{ t \geq 1 : \sum_{u=1}^t \lambda(u; e) < 0 \right\}, \\ \delta_+^e &:= \inf \left\{ t \geq 1 : \sum_{u=1}^t \lambda(u; e) > 0 \right\}, \\ \delta_A^e &:= \inf \left\{ t \geq 1 : \sum_{u=1}^t \lambda(u; e) > A \right\},\end{aligned}$$

whose distributions under  $P_0$  or  $P_\infty$  do not depend on  $e$ . Thus, the following probabilities will also be independent of  $e$ :

$$p_+ := P_\infty(\delta_+^e = \infty), \quad p_- := P_0(\delta_-^e = \infty).$$

#### IV. UNIVERSAL ASYMPTOTIC LOWER BOUND

In this section we establish a first-order asymptotic lower bound on (5) as  $\gamma \rightarrow \infty$  that is *universal* in  $\Sigma$ . The proof is based on the approach developed in [18].

*Theorem 4.1:* For any  $\Sigma$  that satisfies (2) we have

$$\inf_{(S,T) \in \Delta(\gamma)} D^\Sigma(S, T) \geq \frac{\log \gamma}{\mathcal{I}} (1 + o(1)),$$

where  $o(1)$  is a term that goes to 0 as  $\gamma \rightarrow \infty$ .

*Proof:* Fix any  $\Sigma$  that satisfies (2) and  $(S, T) \in \Delta(\gamma)$ . Let  $Z_t$  denote the log-likelihood ratio of  $\mathbb{P}_0^\Sigma$  versus  $\mathbb{P}_\infty$  based on the observed data up to time  $t$ , i.e.,

$$Z_t := \log \frac{d\mathbb{P}_0^\Sigma}{d\mathbb{P}_\infty}(\mathcal{F}_t^S), \quad t \in \mathbb{N},$$

which takes the following form:

$$Z_t = \sum_{u=1}^t \lambda(u; S_u) \mathbb{1}\{S_u \in \mathcal{A}(\Sigma)\}, \quad t \in \mathbb{N}.$$

Furthermore, define

$$A_t := \mathcal{I} \sum_{u=1}^t \mathbb{1}\{S_u \in \mathcal{A}(\Sigma)\}, \quad t \in \mathbb{N}.$$

Note that, under  $\mathbb{P}_0^\Sigma$ , the process  $\{Z_t - A_t : t \in \mathbb{N}\}$  is a zero mean martingale and  $A_t \leq t\mathcal{I}$  for any  $t \in \mathbb{N}$ . For  $0 < \epsilon < 1$ , let

$$N_{\gamma, \epsilon} := \frac{\log \gamma}{\mathcal{I}} (1 - \epsilon).$$

For any  $\nu \geq 0$ , applying Markov's inequality in (4) we get

$$\begin{aligned}D^\Sigma(S, T) &\geq \mathbb{E}_\nu^\Sigma[T - \nu | T > \nu] \\ &\geq N_{\gamma, \epsilon} \mathbb{P}_\nu^\Sigma(T - \nu > N_{\gamma, \epsilon} | T > \nu).\end{aligned}\quad (6)$$

Let also  $\eta_\gamma > 0$  be defined as follows:

$$\log \eta_\gamma := (1 - \epsilon^2) \log \gamma = (1 + \epsilon) N_{\gamma, \epsilon} \mathcal{I}, \quad (7)$$

and observe that

$$\mathbb{P}_\nu^\Sigma(T - \nu \leq N_{\gamma, \epsilon} | T > \nu) \leq p(\nu, T) + q(\nu, T),$$

where

$$\begin{aligned}p(\nu, T) &:= \mathbb{P}_\nu^\Sigma(T - \nu \leq N_{\gamma, \epsilon}, e^{Z_T - Z_\nu} > \eta_\gamma | T > \nu) \\ q(\nu, T) &:= \mathbb{P}_\nu^\Sigma(T - \nu \leq N_{\gamma, \epsilon}, e^{Z_T - Z_\nu} \leq \eta_\gamma | T > \nu).\end{aligned}$$

Now,

$$\begin{aligned}p(\nu, T) &= \mathbb{P}_\nu^\Sigma(T \leq \nu + N_{\gamma, \epsilon}, e^{Z_T - Z_\nu} > \eta_\gamma | T > \nu) \\ &\leq \mathbb{P}_\nu^\Sigma\left(\max_{\nu \leq k \leq \nu + N_{\gamma, \epsilon}} Z_k - Z_\nu > \log \eta_\gamma | T > \nu\right) \\ &= \mathbb{P}_\nu^\Sigma\left(\max_{\nu \leq k \leq \nu + N_{\gamma, \epsilon}} Z_k - Z_\nu > \log \eta_\gamma\right) \\ &= \mathbb{P}_0^\Sigma\left(\max_{1 \leq k \leq N_{\gamma, \epsilon}} Z_k > \log \eta_\gamma\right) \\ &\stackrel{(7)}{=} \mathbb{P}_0^\Sigma\left(\max_{1 \leq k \leq N_{\gamma, \epsilon}} (Z_k - A_k) + A_k > N_{\gamma, \epsilon}(1 + \epsilon)\mathcal{I}\right) \\ &= \mathbb{P}_0^\Sigma\left(\max_{1 \leq k \leq N_{\gamma, \epsilon}} (Z_k - A_k) + A_k > N_{\gamma, \epsilon}\mathcal{I} + \epsilon N_{\gamma, \epsilon}\mathcal{I}\right) \\ &\leq \mathbb{P}_0^\Sigma\left(\max_{1 \leq k \leq N_{\gamma, \epsilon}} (Z_k - A_k) > \epsilon N_{\gamma, \epsilon}\mathcal{I}\right) \\ &\leq \frac{\mathcal{V}}{\epsilon^2 N_{\gamma, \epsilon} \mathcal{I}^2} \equiv \xi_\epsilon(\gamma).\end{aligned}$$

The second equality follows from the facts that  $\{T > \nu\} \in \mathcal{F}_{\nu-1}$  and the observations are independent over time. The second inequality holds because  $A_k \leq k\mathcal{I} \leq N_{\gamma, \epsilon}\mathcal{I}$  for any  $1 \leq k \leq N_{\gamma, \epsilon}$ , and the last one follows from the Doob's maximal inequality for martingales. Now for any given  $\epsilon > 0$ ,  $N_{\gamma, \epsilon} \rightarrow \infty$  as  $\gamma \rightarrow \infty$ . Therefore,  $\xi_\epsilon(\gamma) \rightarrow 0$  as  $\gamma \rightarrow \infty$ .

Furthermore,

$$\begin{aligned}q(\nu, T) &= \mathbb{P}_\nu^\Sigma(T \leq \nu + N_{\gamma, \epsilon}, e^{Z_T - Z_\nu} \leq \eta_\gamma | T > \nu) \\ &= \frac{\mathbb{P}_\nu^\Sigma(\nu < T \leq \nu + N_{\gamma, \epsilon}, e^{Z_T - Z_\nu} \leq \eta_\gamma)}{\mathbb{P}_\infty(T > \nu)} \\ &= \frac{\mathbb{E}_\infty[e^{Z_T - Z_\nu} \mathbb{1}\{\nu < T \leq \nu + N_{\gamma, \epsilon}, e^{Z_T - Z_\nu} \leq \eta_\gamma\}]}{\mathbb{P}_\infty(T > \nu)} \\ &\leq \frac{\eta_\gamma \mathbb{P}_\infty(\nu < T \leq \nu + N_{\gamma, \epsilon})}{\mathbb{P}_\infty(T > \nu)} \\ &= \eta_\gamma \mathbb{P}_\infty(T \leq \nu + N_{\gamma, \epsilon} | T > \nu) \\ &\leq \eta_\gamma \frac{N_{\gamma, \epsilon}}{\gamma} = \gamma^{-\epsilon^2} (1 - \epsilon) \frac{\log \gamma}{\mathcal{I}} \equiv \xi'_\epsilon(\gamma).\end{aligned}$$

The last inequality follows from Theorem 1 in [18]. Now, it is not difficult to observe that  $\xi'_\epsilon(\gamma) \rightarrow 0$  as  $\gamma \rightarrow \infty$  for any given  $\epsilon > 0$ .

Finally from (6), we have

$$\begin{aligned}D^\Sigma(S, T) &\geq N_{\gamma, \epsilon} \mathbb{P}_\nu^\Sigma(T - \nu > N_{\gamma, \epsilon} | T > \nu) \\ &\geq \frac{\log \gamma}{\mathcal{I}} (1 - \epsilon) (1 - \xi_\epsilon(\gamma) - \xi'_\epsilon(\gamma)).\end{aligned}$$

Now let  $\gamma \rightarrow \infty$  and  $\epsilon \rightarrow 0$ , and the proof is complete.  $\blacksquare$

## V. PROPOSED PROCEDURE

In this section, we introduce a policy,  $(\tilde{S}, \tilde{T})$ , that achieves the universal asymptotic lower bound of the previous section for every covariance matrix  $\Sigma$  that satisfies (2). In order to do so, we fix an arbitrary permutation,  $(e_1, \dots, e_K)$ , of the elements of  $\mathcal{E}$ , defined in (3), where  $K$  is the size of  $\mathcal{E}$ , i.e.,

$$K := |\mathcal{E}| = \binom{p}{2}.$$

Furthermore, we fix a threshold  $A > 0$  that will be selected so that the false alarm constraint be satisfied. The proposed sampling rule follows a similar pattern as in [12]. It starts by sampling the first pair of sources in the above permutation, i.e.,  $\tilde{S}_1 = e_1$ , and then at each time  $t \in \mathbb{N}$ :

- it computes the following statistic:

$$Y_t = \max\{Y_{t-1}, 0\} + \lambda(t; S_t), \quad Y_0 \equiv 0,$$

- if  $Y_t \in (0, A)$ , it keeps sampling the same pair, i.e.,

$$\tilde{S}_{t+1} = \tilde{S}_t,$$

- if  $Y_t \notin (0, A)$ , it moves to the next one, i.e.,

$$\tilde{S}_{t+1} = \begin{cases} e_{d+1} & \text{if } \tilde{S}_t = e_d \text{ and } d < K \\ e_1 & \text{if } \tilde{S}_t = e_K \end{cases}.$$

The alarm is raised the first time  $t$  that  $Y_t$  is equal or larger than  $A$ , i.e., the proposed stopping time is:

$$\tilde{T} := \inf\{t \geq 1 : Y_t \geq A\}.$$

*Theorem 5.1:* If  $A = \log \gamma$ , then  $(\tilde{S}, \tilde{T}) \in \Delta(\gamma)$  for every  $\gamma > 0$ , and for every covariance matrix  $\Sigma$  that satisfies (2) we have

$$D^\Sigma(\tilde{S}, \tilde{T}) \leq \frac{\log \gamma}{\mathcal{I}} + \left(\mathcal{I} + \frac{\mathcal{V}}{\mathcal{I}}\right) + \frac{K - |\mathcal{A}(\Sigma)|}{p_+ (1 - (1 - p_-)^{|\mathcal{A}(\Sigma)|})}.$$

*Proof:* We start by defining the following sequence of random times and indicators

$$\begin{aligned} \delta_{n+1} &:= \inf\{t \in \mathbb{N} : \tilde{S}_{t+\delta_n+1} \neq \tilde{S}_{\delta_n+1}\}, \quad n \geq 0 \\ z_n &= \mathbb{1}\{Y_{\delta_n} \geq A\}, \quad n \geq 1. \end{aligned}$$

where  $\delta_0 \equiv 0$ . Thus,  $\delta_n$  represents the duration of the  $n^{\text{th}}$  “stage” of the proposed sampling procedure and  $z_n$  is equal to 1 if the process is terminated at the end of the  $n^{\text{th}}$  stage and 0 otherwise. Let also  $M$  denote the stage in which the process is terminated, i.e.,

$$M := \inf\{n \in \mathbb{N} : z_n = 1\}.$$

Then, we have the following representation:

$$\tilde{T} = \sum_{n=1}^M \delta_n,$$

It is not difficult to observe that, under  $\mathbb{P}_\infty$ ,  $(\delta_n)$  are i.i.d. with the common expectation  $\mathbb{E}_\infty[\delta]$ , and  $\delta$  being defined as

$$\delta := \min\{\delta_-, \delta_A\},$$

where  $(\delta_-, \delta_A)$  is an arbitrary term from the set  $\{(\delta_-^e, \delta_A^e) : e \in \mathcal{E}\}$ . Moreover, under  $\mathbb{P}_\infty$ ,  $M$  is a Geometric random variable with parameter  $\mathbb{P}_\infty(z = 1)$ , where  $z$  is an arbitrary term from the sequence  $(z_n)$ . Thus, by Wald’s identity,

$$\mathbb{E}_\infty[\tilde{T}] = \frac{\mathbb{E}_\infty[\delta]}{\mathbb{P}_\infty(z = 1)}.$$

The right-hand side is equal to the expected time to false alarm of a CUSUM test (see [19]) for detecting whether the correlation of a two-dimensional Gaussian random vector has changed from 0 to  $\rho$ . In particular,  $\mathbb{E}_\infty[\tilde{T}] \geq e^A$  (see [10]), which proves that setting  $A = \log \gamma$  guarantees the desired error control.

To obtain an upper bound on the worst case detection delay, we start by observing that the worst case scenario occurs when  $Y_\nu \leq 0$ , where  $\nu$  is the time of the change. Thus, it suffices to upper bound the expectation of  $\tilde{T}$  under  $\mathbb{P}_0^\Sigma$ . In order to do so, first of all observe that  $\tilde{T} = T_1 + T_2$ , where

$$T_1 := \sum_{n=1}^M \delta_n \mathbb{1}\{w_n = 1\}, \quad T_2 := \sum_{n=1}^M \delta_n \mathbb{1}\{w_n = 0\},$$

and  $w_n$  is equal to 1 if the sources sampled at stage  $n$  are actually correlated under  $\Sigma$  and 0 otherwise. Note that  $(w_n)$  are deterministic quantities. We will bound the expectation of each of these two terms separately.

We start with  $T_1$ . Let  $(\delta'_n, z'_n)$  denote the subsequence of  $(\delta_n, z_n)$  that corresponds to stages where *correlated* sources are sampled and denote by  $(\delta', z')$  an arbitrary term of this subsequence. Both  $(\delta'_n)$  and  $(z'_n)$  are i.i.d. under any  $\mathbb{P}_0^\Sigma$ , with expectations that do not depend on  $\Sigma$ ,  $\mathbb{E}_0[\delta']$  and  $\mathbb{P}_0(z' = 1)$ , respectively. Let also  $M'$  denote the number of stages *at which correlated sources are sampled* until the process is terminated at a stage *where correlated sources are sampled*, i.e.,

$$M' := \inf\{n \in \mathbb{N} : z'_n = 1\}.$$

Since the actual process may also be terminated at a stage where an uncorrelated pair is being sampled, we have

$$T_1 \leq \sum_{n=1}^{M'} \delta'_n.$$

Then,  $M'$  is a Geometric random variable with parameter  $\mathbb{P}_0(z' = 1)$ , which is a stopping time with respect to the filtration generated by  $(\delta'_n, z'_n)$ . Thus, by Wald’s identity we have

$$\mathbb{E}_0^\Sigma[T_1] \leq \frac{\mathbb{E}_0[\delta']}{\mathbb{P}_0(z' = 1)}.$$

The right-hand side is equal to the expected detection delay, when the change occurs at 0, of a CUSUM test that detects a change in the correlation of a two-dimensional Gaussian random vector from 0 to  $\rho$ . Thus, it is equal to  $A/\mathcal{I} + O(1)$ , where  $O(1)$  is a term that does not depend on  $\gamma$ . This  $O(1)$  can be upper bounded by (see [20])

$$\frac{\mathbb{E}_0[\lambda(1; e)^+]^2}{\mathcal{I}} \leq \frac{\mathbb{E}_0[|\lambda(1; e)|^2]}{\mathcal{I}} \leq \frac{\mathbb{E}_0[|\lambda(1; e)|^2]}{\mathcal{I}} = \frac{\mathcal{V} + \mathcal{I}^2}{\mathcal{I}}.$$

In order to bound the expectation of  $T_2$  under  $\mathbb{P}_0^\Sigma$ , it will be convenient to introduce some additional terminology and notation. We refer to the stages from  $m+1$  to  $m+K$  as the  $m^{\text{th}}$  “cycle” of the sampling process and denote by  $\sigma_m$  the total duration of only those stages in the  $m^{\text{th}}$  cycle in which *uncorrelated* sources are sampled, i.e.,

$$\sigma_m = \sum_{n=1}^K \delta_{K(m-1)+n} \mathbb{1}\{w_n = 0\}, \quad m \in \mathbb{N}.$$

Let  $(\delta_n'', z_n'')$  denote the subsequence of  $(\delta_n, z_n)$  that corresponds to stages where *uncorrelated* sources are sampled and denote by  $(\delta'', z'')$  an arbitrary term of this subsequence. Both  $(\delta_n'')$  and  $(z_n'')$  are i.i.d. under any  $\mathbb{P}_0^\Sigma$ , with expectations that do not depend on  $\Sigma$ ,  $\mathbb{E}_\infty[\delta'']$  and  $\mathbb{P}_\infty(z'' = 1)$ , respectively. Then,  $(\sigma_n)$  are also i.i.d. under  $\mathbb{P}_0^\Sigma$  with expectation  $(K - |\mathcal{A}(\Sigma)|) \mathbb{E}_\infty[\delta'']$ . Furthermore, we have

$$\mathbb{E}_\infty[\delta''] \leq \mathbb{E}_\infty[\delta_-] = \frac{1}{\mathbb{P}_\infty(\delta_+ = \infty)} = \frac{1}{p_+}, \quad (8)$$

where the first equality follows from Corollary 8.39 in [21]. Let also  $N$  denote the cycle at which the sampling process is terminated. Then:

$$T_2 \leq \sum_{m=1}^N \sigma_m.$$

Under  $\mathbb{P}_0^\Sigma$ ,  $N$  is a Geometric random variable with parameter

$$\begin{aligned} & \mathbb{P}_0^\Sigma(z_i = 1 \text{ for some } 1 \leq i \leq n) \\ & \geq \mathbb{P}_0^\Sigma(z_i = 1 \text{ for some } 1 \leq i \leq n \text{ with } w_i = 1) \\ & = 1 - \mathbb{P}_0^\Sigma(z_i = 0 \text{ for every } 1 \leq i \leq n \text{ with } w_i = 1) \\ & = 1 - \prod_{1 \leq i \leq n: w_i = 1} \mathbb{P}_0(z'_i = 0) = 1 - \mathbb{P}_0(z' = 0)^{|\mathcal{A}(\Sigma)|} \\ & = 1 - (1 - \mathbb{P}_0(z' = 1))^{|\mathcal{A}(\Sigma)|} \geq 1 - (1 - p_-)^{|\mathcal{A}(\Sigma)|}, \end{aligned} \quad (9)$$

where the last inequality follows from the fact that

$$\mathbb{P}_0(z' = 1) \geq \mathbb{P}_0(\delta_- > \delta_A) \geq \mathbb{P}_0(\delta_- = \infty) = p_-.$$

Finally, by Wald’s identity, using (8) and (9) we have

$$\mathbb{E}_0^\Sigma[T_2] = \mathbb{E}_0^\Sigma[\sigma_1] \mathbb{E}_0^\Sigma[N] \leq \frac{K - |\mathcal{A}(\Sigma)|}{p_+ (1 - (1 - p_-)^{|\mathcal{A}(\Sigma)|})},$$

and the proof is complete.

## VI. SIMULATION STUDY

In this section we present a simulation study in which we explore how the worst-case conditional expected detection delay of the proposed procedure,  $D^\Sigma(\tilde{S}, \tilde{T})$ , varies with the true number of post-change correlated pairs,  $|\mathcal{A}(\Sigma)|$ , when  $p = 10$ ,  $\rho = 0.7$ , and  $\gamma \in \{10^2, 10^5\}$ . For the implementation of the proposed procedure, we fix the value of  $\gamma$  and select threshold  $A$  according to Theorem 5.1, i.e.,  $A = \log \gamma$ . Moreover, we fix the following permutation of  $\mathcal{E}$ :

$$((1, 2), (1, 3), \dots, (1, p), (2, 3), \dots, (2, p), \dots, (p-1, p)),$$

i.e., we set  $e_1 = (1, 2), e_2 = (1, 3), \dots, e_K = (p-1, p)$ .

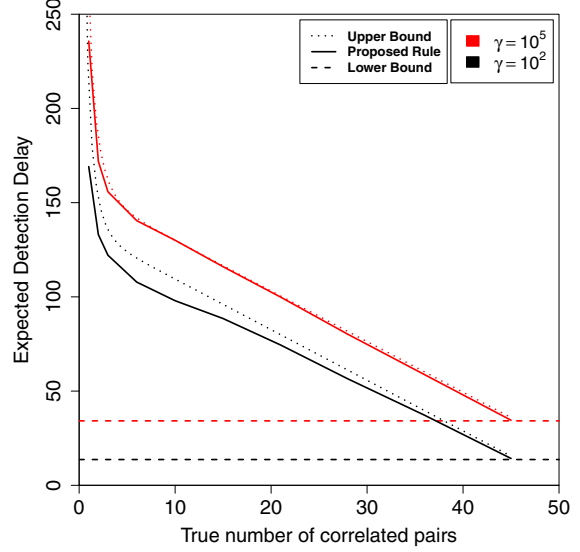


Figure 1. Expected detection delay  $D^\Sigma(\tilde{S}, \tilde{T})$  with change in true number of correlated pairs  $|\mathcal{A}(\Sigma)|$ , under different values of  $\gamma$

For the post-change regime, we assume that the sources in  $\{p-s+1, p-s+2, \dots, p\}$  become correlated to each other at time  $\nu = 0$  with correlation  $\rho$ , where  $s$  is an integer such that  $2 \leq s \leq p$ . We consider different values for  $s$  in the set  $\{2, \dots, p\}$ , and in this way the true number of correlated pairs,  $|\mathcal{A}(\Sigma)|$  varies in  $\{\binom{s}{2} : 2 \leq s \leq p\}$ . In view of the adopted permutation of  $\mathcal{E}$ , we need to perform all stages corresponding to the uncorrelated pairs prior to the correlated ones. Therefore, for any true subset of correlated sources, the above setup corresponds to the worst-case scenario for the proposed rule. Moreover, in addition to the true value of the expected detection delay when  $\nu = 0$ , which is obtained via Monte Carlo simulation, we also plot in Figure 1 the corresponding upper bound from Theorem 5.1, after estimating the probabilities  $p_+$  and  $p_-$  again via Monte Carlo simulation.

From Figure 1, we observe that the upper bound of Theorem 5.1 is very close to  $D^\Sigma(\tilde{S}, \tilde{T})$ , especially for the larger value of  $\gamma$ . Moreover, we can see that when  $|\mathcal{A}(\Sigma)|$  is large enough,  $D^\Sigma(\tilde{S}, \tilde{T})$  decreases linearly with  $|\mathcal{A}(\Sigma)|$ . This can be predicted by the upper bound of Theorem 5.1. Indeed, the denominator in the third term of this upper bound changes very slowly and behaves like a constant as  $|\mathcal{A}(\Sigma)|$  increases to  $K$ , hence this upper bound exhibits an almost linear trend with negative slope. Finally, we observe that the performance of the proposed rule essentially agrees with the asymptotic lower bound in both cases for  $\gamma$  when  $|\mathcal{A}(\Sigma)| = K$ , i.e., when all processes become correlated after the change. ■

## VII. ACKNOWLEDGEMENTS

This work was supported in part by the U.S. National Science Foundation through grant NSF ATD 1737962.

## REFERENCES

- [1] K. Cohen and Q. Zhao, "Active hypothesis testing for anomaly detection," *IEEE Transactions on Information Theory*, vol. 61, no. 3, pp. 1432–1450, 2015.
- [2] B. Huang, K. Cohen, and Q. Zhao, "Active anomaly detection in heterogeneous processes," *IEEE Transactions on Information Theory*, vol. 65, no. 4, pp. 2284–2301, 2019.
- [3] A. Tsopelakos, G. Fellouris, and V. Veeravalli, "Sequential anomaly detection with observation control," in *Proc. IEEE International Symposium on Information Theory*, Paris, France, Jul. 2019, pp. 2389–2393.
- [4] A. Tsopelakos and G. Fellouris, "Sequential anomaly detection with observation control under a generalized error metric," in *Proc. IEEE International Symposium on Information Theory*, Los Angeles, CA, Jun. 2020, pp. 1165–1170.
- [5] J. Heydari, A. Tajer, and H. V. Poor, "Quickest detection of markov networks," in *Proc. IEEE International Symposium on Information Theory*, Barcelona, Spain, Jul. 2016, pp. 1341–1345.
- [6] J. Heydari and A. Tajer, "Quickest search for local structures in random graphs," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 526–538, Sep. 2017.
- [7] H. Chernoff, "Sequential design of experiments," *Ann. Math. Statist.*, vol. 30, no. 3, pp. 755–770, 09 1959.
- [8] S. Nitinawarat, G. K. Atia, and V. V. Veeravalli, "Controlled sensing for multihypothesis testing," *IEEE Transactions on Automatic Control*, vol. 58, no. 10, pp. 2451–2464, 2013.
- [9] H. Poor and O. Hadjiladis, *Quickest detection*. United Kingdom: Cambridge University Press, Jan. 2008, vol. 9780521621045.
- [10] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. CRC Press, 08 2014.
- [11] K. Liu, Y. Mei, and J. Shi, "An adaptive sampling strategy for online high-dimensional process monitoring," *Technometrics*, vol. 57, no. 3, pp. 305–319, 2015.
- [12] Q. Xu, Y. Mei, and G. V. Moustakides, "Second-order asymptotically optimal change-point detection algorithm with sampling control," in *Proc. IEEE International Symposium on Information Theory*, Los Angeles, CA, Jun. 2020, pp. 1136–1140.
- [13] J. Heydari and A. Tajer, "Quickest change detection in structured data with incomplete information," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, New Orleans, LA, Mar. 2017, pp. 6434–6438.
- [14] A. Chaudhuri and G. Fellouris, "Sequential detection and isolation of a correlated pair," in *Proc. IEEE International Symposium on Information Theory*, Los Angeles, CA, Jun. 2020, pp. 1141–1146.
- [15] L. Xie, Y. Xie, and G. V. Moustakides, "Sequential subspace change point detection," *Sequential Analysis*, vol. 39, no. 3, pp. 307–335, 2020.
- [16] L. K. Chan and J. Zhang, "Cumulative sum control charts for the covariance matrix," *Statistica Sinica*, vol. 11, no. 3, pp. 767–790, 2001.
- [17] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, vol. 42, no. 6, pp. 1897–1908, Dec. 1971.
- [18] Tze Leung Lai, "Information bounds and quick detection of parameter changes in stochastic systems," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2917–2929, 1998.
- [19] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.
- [20] G. Lorden, "On excess over the boundary," *Ann. Math. Statist.*, vol. 41, no. 2, pp. 520–527, Apr. 1970.
- [21] D. Siegmund, *Sequential Analysis Tests and Confidence Intervals*, ser. Springer Series in Statistics. New York, NY: Springer New York, 1985.