Active Quickest Detection When Monitoring Multi-streams with Two Affected Streams

Qunzhi Xu and Yajun Mei

***Abstract*—We study the multi-stream quickest detection problem under the active learning setup, It is assumed that there are p local streams in a system and $s \leq p$ unknown local streams are affected by an undesired event at some unknown time, but one is only able to take observations from r of these p local streams at each time instant. The objective is how to adaptively sample from these p local streams and how to use the observed data to raise a correct global alarm as quickly as possible. In this paper, we develop the first asymptotic optimality theory in the active quickest detection literature for the case when $s = r = 2$. To be more concrete, we propose to combine three ideas to develop efficient active quickest detection algorithms: (1) win-stay, lose-switch sampling strategy; (2) local CUSUM statistics for local monitoring; and (3) the SUM-Shrinkage technique to fuse local statistics into a global decision. We show that our proposed algorithms are asymptotically optimal in the sense of minimizing detection delay up to the second order subject to the false alarm constraint. Numerical studies are conducted to validate our theoretical results.**

***Index Terms*—Asymptotic optimality, quickest detection, active learning, win-stay lose-switch, CUSUM.**

I. Introduction

Active learning has received extensive attentions in modern big data age, partly due to the resource constraints either at the data acquisition level or at the communication level. Often it is required that the decision maker actively selects partial samples from the underlying raw data so as to make right decision with as few attempts as possible. Some examples of active learning include compressed sensing (see Donoho [1] and Candes, Romberg and Tao [2]), distributed sensor networks (see Li and Jin [3] and Laszka, Abbas and Koutsoukos [4]), information extraction (see Barrio and Gravano [5] and Boicea, Truică, Rădulescu and Buse [6]), etc.

One important active learning problem when monitoring streaming data is active quickest detection problem, which has a wide range of real-world applications in industrial quality control, biosurveillance, network security, etc. Under a general setting, there are p local streams in a system, and at some unknown time τ, an occurring event impacts s of the available streams by changing the distribution of their samples. However, under the sampling control constraint, one is allowed to sample from only r of the p local streams at each time instant. The objective of active quickest detection is to decide how to adaptively sample from these p local streams and how to use the observed data to raise a correct global alarm as quickly as possible once the change occurs subject to both the false alarm and sampling control constraints.

It is useful to point out that the quickest detection problem has been extensively studied under the classical *passive learning* setup where one has fully access to all the available data to make decisions on when to raise an alarm, see Basseville and Nikiforov [7], Poor and Hadjiliadis [8] and Tartakovsky, Nikiforov and Basseville [9]. For more recent references, see Mei [10], Xie and Siegmund [11] and Chan [12]. However, theoretical research on active quickest detection problem is limited, despite that the simplest version of $s = r = 1$ was formulated as early as 1963 by Shiryaev [13].

Indeed, so far the asymptotic optimality theory has only been developed for active quickest detection in the simplest case of $r = s = 1$, see Xu, Mei and Moustakides [14] and Xu and Mei [15]. Intuitively, such asymptotic optimality theory heavily depend on the structure of the simplest case of $r = s = 1$ where there are not many reasonable options available for sampling or decision policies. This is the reason why one can develop asymptotically optimal algorithm for $r = s = 1$ case by simply adopting the greedy sampling policy and the naive decision policy that raises a global alarm if any local stream raises a local alarm. However, when the true number of local affected streams $s \geq 2$, such algorithm will lose statistical efficiency and thus new ideas are needed to develop asymptotically optimal algorithms.

In this article, we establish the asymptotic optimality theory for the active quickest detection problem with $s = 2$ local affected streams when monitoring a fixed number p of local streams under the sampling control $r = 2$, as the average run length to false alarm constraint goes to ∞. Our asymptotic optimality result is the first of its kind in the literature of active quickest change for $s \geq 2$, and our idea for $s = r = 2$ can also be extended to a more general case of $s = r > 2$. Our key idea is to develop efficient active quickest detection algorithm for $s \geq 2$ by combining three ideas together: (i) win-stay, lose-switch sampling strategy which was first proposed by Robbins [16] for the multi-armed bandit problems (MAB); (2) local CUSUM statistics for local monitoring one-dimensional data; and (3) the SUM-Shrinkage technique in Liu, Zhang, and Mei [17] to fuse local statistics into a global decision. We show that our proposed algorithm holds a second-order asymptotically optimal property in the quickest detection framework. This means that even with the sampling rate of $2/p$ each time instant as compared to the optimal passive quickest detection algorithm with full data, our proposed algorithm has the same detection delay performance up to second-order subject to the

This work was supported in part by the National Science Foundation (NSF) under grants DMS 2015405. Q. Xu is partially supported by the ARC/ACO fellowship. The authors are with Georgia Institute of Technology (e-mails: {xuqunzhi,ymei3}@gatech.edu).

false alarm constraint.

Finally, we need to note that the sampling control has been combined with sequential data under different settings, see Banerjee, Taposh and Veeravalli [18], Heydari, Tajer, and Poor [19]), etc. Moreover, sampling control has been extensively studied in two other related well-known problems: multi-armed bandit problems and sequential hypothesis testing. For a limited list of references, see Lai and Robbins [20], Li, Li, Wang and Liu [21], Tsopelakos, Fellouris, and Veeravalli [22], etc. In those contexts, all observations provide some information for decision making. Here we should emphasize that quickest detection problems under sampling control are much more challenging, because observations from the affected stream will not provide any information to the change-point if it is taken *before* the change occurs, but the change-point itself is unknown and is the focus of quickest detection.

The remainder of the paper is organized as follows. In Section II, we state the mathematical formulation of our problem. In Section III, we present our proposed algorithm and provide its theoretical properties in Section IV. Numerical studies are presented in Section V to illustrate the performance properties of our proposed algorithm. The high-sketch of the proofs are presented in Appendix.

II. Problem Formulation

Suppose there are p statistically independent local streams in a system, and denote with X_t^i the observation from the i-th stream at time t, where $i = 1, \ldots, p$ and $t = 1, 2, \ldots$. Initially, the system is in the in-control state and the data stream $\{X_t^i\}$ from the i-th stream produces i.i.d. samples following the density $f_i(X)$. At some unknown time τ, a triggering event occurs to the system and affects two out of p data streams in the sense that if the i-th stream is affected, the density function of its local samples X_t^i changes from f_i to g_i at time $t \geq \tau$. Specifically, denote the index set of the affected streams as $O_C = \{i_1, i_2\}$. Then if index $i \in O_C$,

$$X_t^i \sim \begin{cases} f_i(X), & \text{if } t \leq \tau \\ g_i(X), & \text{if } t > \tau, \end{cases} \tag{1}$$

whereas $X_t^i \sim f_i(X)$ for $i \notin O_C$ and all $t > 0$.

Under the sampling control constraint, we are only able to observe $r = 2$ out of p local streams at each time instant t. To be rigorous, define the *Sample set* $O_t = \{i_{1,t}, i_{2,t}\}$ which points to the pair of two streams that will be sampled during time instant t, and the sampling constraint can be expressed as

$$O_t \subset \{1, 2, \cdots, p\} \text{ and } |O_t| = 2, \tag{2}$$

where $|A|$ denotes the cardinality of set A.

In the active quickest detection problem under sampling control, an algorithm consists two elements: one is the sampling policies, e.g., choose the sample set O_t for all time t subject to (2), and the other is the decision policy that is defined as the stopping time T with respect to the observed data sequence.

Denote by $\mathsf{P}_\tau^{(i_1,i_2)}$ and $\mathsf{E}_\tau^{(i_1,i_2)}$ the probability measure and expectation when change occurs at time τ with the index set of affected streams $O_C = \{i_1, i_2\}$. Denote by P_∞ and E_∞ the same when there are no changes, or equivalently, when the change occurs at time ∞. Following the performance measures for quickest detection proposed by Pollak [23], we are interested in finding a procedure $(\{O_t\}_{t=1,\cdots,\infty}, \mathsf{T})$ that minimizes the worst average detection delay conditioned on the event that we stop after the change time t,

$$D^{(i_1,i_2)}(\mathsf{T}) = \sup_{t \geq 0} \mathsf{E}_t^{(i_1,i_2)}[\mathsf{T} - t | \mathsf{T} > t]. \tag{3}$$

for any combinations of affected streams $O_C = \{i_1, i_2\} \subset \{1, 2, \cdots, p\}$, subject to the sampling control constraint in (2) and the average run length (ARL) to false-alarm constraint

$$\mathsf{E}_\infty[\mathsf{T}] \geq \gamma > 1. \tag{4}$$

where γ is a pre-specified constant.

III. Our Proposed Algorithm

In this section, we develop our algorithm, denoted by T_{WSLS}, based on the win-stay, lose-switch sampling policy. At a high level, we propose to sample a subset of two streams until we are confident to decide whether a change has occurred or not. If we detect a change, then we stop and raise a global alarm. If we decide there is no change or we have sampled from these two streams for a long time, then we switch to sample from another pair of two streams. We repeat these steps until we raise an alarm.

For better presentation, the current section is divided into three subsections: in Section III-A we define local statistics, which will be the cornerstone of our algorithm. We propose the win-stay, lose-switch sampling policy in Section III-B, and the decision policy in Section III-C.

A. Local Statistics

For the sake of clarity, we define two sets of local monitoring statistics, $\widetilde{W}_t^i$ and W_t^i for the i-th local stream at time t. The former is used to update the observed data, and the latter also takes into account of possible switch of sampling different data streams.

Let us first define the local statistics $\widetilde{W}_t^i$. When the i-th local stream is observed, we can simply update it based on the well-known CUSUM statistics. When the i-th local stream is not observed, then we treat it as missing data and the corresponding log-likelihood-ratio of missing data as 0. Mathematically, at each time instant $t = 1, 2, \cdots$, the local statistics $\widetilde{W}_t^i$ can be defined recursively as

$$\begin{aligned} \widetilde{W}_t^i &= \widetilde{W}_{t-1}^i + \mathbb{1}_{\{i \in O_t\}} \log \frac{g_i(X_t^i)}{f_i(X_t^i)} \\ &= \begin{cases} \max\{\widetilde{W}_{t-1}^i, 0\}, & \text{if } i \notin O_t \\ \max\{\widetilde{W}_{t-1}^i, 0\} + \log \frac{g_i(X_t^i)}{f_i(X_t^i)}, & \text{if } i \in O_t, \end{cases} \end{aligned} \tag{5}$$

with the initial values $\widetilde{W}_0^i = 0$ for all $i = 1, \cdots, p$. Here O_t is the set of local streamed that will be sampled at time t, and will be defined later in subsection III-B.

Next, we define the local statistics W_t^i that is a modification of $\widetilde{W}_t^i$ by taking into account of possible switch of sampling different data streams. Assume we are sampling the $i_{1,t}$-th and $i_{2.t}$-th local streams at time t and consider the values of local statistics $\widetilde{W}_t^{i_{1,t}}$ and $\widetilde{W}_t^{i_{2,t}}$. In general, we reset all local statistics back to 0 if we switch to different pair of two streams (ie: $O_{t+1} \neq O_t$). The only exception is when one of $\widetilde{W}_t^{i_{1,t}}$ and $\widetilde{W}_t^{i_{2,t}}$ is zero and the other is positive. While it is reasonable to continue to sample the local stream with positive $\widetilde{W}_t^i$ value and switch only the local stream with zero value, it is highly non-trivial to mathematically analyze such case. To get around the mathematical challenges, we propose to take in advantage of the renewal properties of the local CUSUM statistics by resetting the positive local statistics to a constant $\delta \geq 0$. That will allow us to use the independent structure of data to prove that the corresponding algorithm is asymptotically optimal. Mathematically, we define

$$W_t^i = \begin{cases} \min(\widetilde{W}_t^i, \delta), & \text{if } O_{t+1} \neq O_t \\ \widetilde{W}_t^i, & \text{if } O_{t+1} = O_t. \end{cases} \tag{6}$$

for $i = 1, 2, \cdots, p$, where $\delta \geq 0$ is a pre-specificed constant. In addition, at each time step, we will further re-set $\widetilde{W}_t^i = W_t^i$ after updating its value from $\widetilde{W}_{t-1}^i$ in (5) and before updating the values in the time $t + 1$.

B. Win-Stay, Lose-Switch Sampling Policy

Here we adopt the win-stay, lose-switch sampling policy with a twist of avoiding of sampling a local stream for too long. On the one hand, if the local statistics $\widetilde{W}_t^i$ of two sampled data streams are positive, then we should continue to sample the same two local streams, and if one of them become zero, then we should switch to sample new local streams. On the other hand, if we sample two local streams for a very long time but the corresponding local statistics are positive but small values, then it might suggest that there are no strong evidences for these two local streams involving changes, and we might want to explore new local streams.

For that purpose, we propose to introduce a new variable $\beta(t)$ that indicates the accumulated time we have stayed in the current pair of two streams $\{i_{1,t}, i_{2,t}\}$ being sampled at time instant t. The variable $\beta(t)$ is updated as:

$$\beta(t) = \begin{cases} \beta(t-1) + 1, & \text{if } O_t = O_{t-1} \\ 1, & \text{otherwise} \end{cases} \tag{7}$$

with the initial values $\beta(0) = 0$.

Next, we introduce a new controlling parameter q, which can be thought of as the maximum consecutive time we can tolerate to stay in the same pair of two streams. In our results below, we set $q = q(A) = Ce^{\epsilon A}$ for some constant $C \in (0, \infty)$ and $\epsilon \in (0, \frac{1}{2})$.

Now we are ready to define our sampling policy. If if $\widetilde{W}_t^{i_{1,t}} > 0, \widetilde{W}_t^{i_{2,t}} > 0$ and $\beta(t) < q(A)$, then we adopt the win-stay, lose-switch sampling policy by continuing sampling the same two local streams, e.g., $O_{t+1} = O_t$. Otherwise, the sample set O_{t+1} will be different from O_t, as we need to avoid sampling those local streams whose local statistic value is zero or stays for too long. Mathematically, we randomly pick two streams $\{j_{1,t}, j_{2,t}\}$ from the unobservable set $\{1, \cdots, p\} \setminus O_t$ and define O_{t+1} as

$$= \begin{cases} O_t & \text{if } \widetilde{W}_t^{i_{1,t}}, \widetilde{W}_t^{i_{2,t}} > 0 \text{ and } \beta(t) < q(A) \\ \{j_{1,t}, j_{2,t}\} & \text{if } \widetilde{W}_t^{i_{1,t}}, \widetilde{W}_t^{i_{2,t}} \leq 0 \text{ or } \beta(t) \geq q(A), \\ \{i_{1,t}, j_{2,t}\} & \text{if } \widetilde{W}_t^{i_{1,t}} > 0, \widetilde{W}_t^{i_{2,t}} \leq 0 \text{ and } \beta(t) < q(A) \\ \{j_{1,t}, i_{2,t}\} & \text{if } \widetilde{W}_t^{i_{1,t}} \leq 0, \widetilde{W}_t^{i_{2,t}} > 0 \text{ and } \beta(t) < q(A) \end{cases} \tag{8}$$

C. Decision Policy

Our proposed decision policy is inspired by the SUM-Shrinkage technique in Liu, Zhang, and Mei [17], and one efficient way to fuse local statistics into a global decision by considering the sum of the s-largest values of local statistics when s out of p local streams are affected by the change.

Under our context when we are only allowed to sample from $r = 2$ out of p data streams, in our algorithm, there are at most two local statistics W_t^i for observed local streams whose value might be positive, as all other W_t^i values are zero for unobserved local streams. Hence, we propose to raise an alarm at the stopping time

$$\mathsf{T}_{\text{WSLS}}(A) = \inf\left\{t > 0 : W_t^{i_{1,t}} + W_t^{i_{2,t}} \geq A\right\}, \tag{9}$$

where $O_t = \{i_{1,t}, i_{2,t}\}$ points to the pair of two streams that is sampled during time t.
In summary, our proposed algorithm defined by the stopping time $\mathsf{T}_{\text{WSLS}}(A)$ in (9) can be summarized as follows:

Algorithm 1 Our Proposed Algorithm T_{WSLS}

1: Initialize $O_1 = \{1, 2\}, \beta(0) = 0$ and $W_0^i = 0$ for $i = 1, \cdots, p$.
2: **for** each time t **do**
3: Sample the streams in the sample set $O_t = \{i_{1,t}, i_{2,t}\}$.
4: Update the local summary statistics $\widetilde{W}_t^i$ for $i = 1, \cdots, p$ as in (5).
5: Update the accumulated time $\beta(t)$ as in (7).
6: Update the sample set O_{t+1} at time $t + 1$ according to the rule in (8).
7: Update the local summary statistics W_t^i as in (6) and reset $\widetilde{W}_t^i = W_t^i$ for $i = 1, \cdots, p$.
8: **if** $W_t^{i_{1,t}} + W_t^{i_{2,t}} \geq A$ **then**
9: Raise an alarm at $\mathsf{T}_{\text{WSLS}}(A) = t$.
10: **end if**
11: **end for**

for $i = 1, \cdots, p$

IV. Asymptotic Optimality

In this section, we will investigate the theoretical properties of our proposed algorithm $\mathsf{T} = \mathsf{T}_{\text{WSLS}}$ in (9).

First, let us make some standard assumptions in the quickest detection literature. We assume that Kullback-Leibler information numbers are positive and finite for all $i = 1, 2, \cdots, p$:

$$I(f_i, g_i) \quad = \quad \int \log \frac{f_i(X)}{g_i(X)} f_i(X) dX \in (0, \infty),$$

$$I(g_i, f_i) = \int \log \frac{g_i(X)}{f_i(X)} g_i(X) dX \in (0, \infty). \quad (10)$$

Moreover, we assume that the second order moments of log likelihood ratios are bounded away from ∞.

$$\int (\log \frac{f_i(X)}{g_i(X)})^2 f_i(X)) dX \in (0, \infty),$$
$$\int (\log \frac{g_i(X)}{f_i(X)})^2 g_i(X)) dX \in (0, \infty). \quad (11)$$

Now we are ready to present the theoretical properties of our proposed algorithm. The following theorem summarizes the non-asymptotic properties of $\mathsf{T}_{\mathrm{WSLS}}$ on the ARL to false alarm and detection delay for any threshold $A > 0$.

Theorem 1 *For our proposed algorithm $\mathsf{T}_{\mathrm{WSLS}}$ in (9), we have*

$$\mathsf{E}_\infty[\mathsf{T}_{\mathrm{WSLS}}] \geq e^A. \quad (12)$$

Moreover, for any combinations of affected streams $O_C = \{i_1, i_2\} \in \{1, 2, \cdots, p\}$, its detection delay satisfies

$$D^{(i_1,i_2)}(\mathsf{T}) \leq \frac{A}{I(g_{i_1}, f_{i_1}) + I(g_{i_2}, f_{i_2})} + C_0 p(p-1) + C_1, \quad (13)$$

where C_0, C_1 are constants depending only on the distributions, not on A.

The high-level sketch of the proof of Theorem 1 will be postponed in the Appendix. By Theorem 1, the following corollary establishes the second-order asymptotic optimality properties of our proposed algorithm $\mathsf{T}_{\mathrm{WSLS}}$ in (9) in the quickest detection framework when the average run length to false alarm constraint γ in (4) go to ∞.

Corollary 1 *Let $A = \log\gamma$, then our proposed algorithm $\mathsf{T}_{\mathrm{WSLS}}$ in (9) satisfies both the false alarm constraint in (4) and the sampling control constraint in (2). Moreover, for any combinations of affected streams $O_C = \{i_1, i_2\} \in \{1, 2, \cdots, p\}$, its detection delay satisfies*

$$0 \leq D^{(i_1,i_2)}(\mathsf{T}_{\mathrm{WSLS}}) - D^{(i_1,i_2)}_{\mathrm{orc}} \leq C_0 p(p-1) + C_2 \quad (14)$$

where the parameters C_0 and C_2 are constants depending only on the distributions, not on γ. Here $D^{(i_1,i_2)}_{\mathrm{orc}}$ is the oracle detection delay achieved by the classical CUSUM procedure for monitoring changes in distribution of the i_1-th stream and i_2-th stream subject to the false alarm constraint in (4).

It is useful to add a couple of remarks to better our results. First, while the second-order asymptotic optimality property is proved under the scenario that the number p of local streams is fixed, our results also hold under the scenario when $p \to \infty$ in such a way that $p^2 = o(\log\gamma)$ as $\gamma \to \infty$, since the relationship (14) hold for any given p. Of course, when $p^2 \gg \log\gamma$, our techniques will break down, and it remains an open problem to develop asymptotic optimality theory. Second, our proposed algorithm can be easily extended to the general case of $s = r > 2$ case, where the second-order asymptotic optimality still holds.

V. Numerical Studies

In this section, we report the results of Monte Carlo simulations to illustrate the usefulness of our proposed algorithm $\mathsf{T}_{\mathrm{WSLS}}$ in (9). For the purpose of comparisons, we consider the following baseline algorithm:

- The random sampling algorithm under the sampling constraint, denoted as $\mathsf{T}_{\mathrm{RND}}$, where at each time we randomly pick up two streams, update the corresponding local statistics as in the standard CUSUM statistics and raise an alarm when the sum of top two local statistics is large. Mathematically, $\mathsf{T}_{\mathrm{RND}}$ raise an alarm at

$$\mathsf{T}_{\mathrm{RND}}(A) = \inf\{t : t > 0, \widetilde{W}_t^{(1)} + \widetilde{W}_t^{(2)} \geq A\} \quad (15)$$

where $\widetilde{W}_t^{(1)}, \widetilde{W}_t^{(2)}$ is the top two local statistics of $\{\widetilde{W}_t^1, \cdots, \widetilde{W}_t^p\}$ with $\widetilde{W}_t^i$ being defined as in (5).

In our simulations, we assume that $f_i \equiv f \sim N(0, 1)$ and $g_i \equiv g \sim N(1, 1)$. We consider the number of streams $p = 3$ and a wide range of false alarm constraint $\gamma \in [10^3, 10^4]$. We set the controlling parameter $q_A = 140$, and the reset value $\delta = 1$ for local statistics. To illustrate our asymptotic results, for the low bounds, we report the oracle detection delay achieved by the standard CUSUM procedure. The corresponding results are summarized in Figure 1.

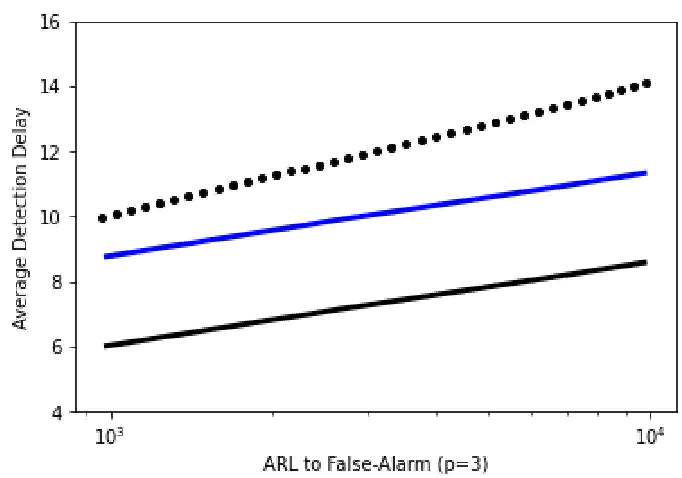

Fig. 1. Average detection delay as a function of ARL to false-alarm for proposed $\mathsf{T}_{\mathrm{WSLS}}$(blue), $\mathsf{T}_{\mathrm{RND}}$(dotted black), and oracle CUSUM procedure(solid black).

From Figure 1 we observe that the gap between the blue line and the solid black line remains as a constant as the ARL to false alarm increases from $\gamma = 10^3$ to 10^4. This is consistent with our theoretical result that the proposed algorithm $\mathsf{T}_{\mathrm{WSLS}}$ is second-order asymptotically optimal when the dimension p is fixed and $\gamma \to \infty$. We can also observe that the gap between the blue line and the dotted black line grows larger as γ increases, and this shows that our proposed algorithm is quite efficient.

APPENDIX: PROOF OF THEOREM 1

In this section we provide a high level sketch of the proof of Theorem 1. The proof of Corollary 1 follows directly from Theorem 1 and the classical result on the oracle detection delay of the classical CUSUM procedures, and thus its proof is omitted due to page limit. To highlight our main ideas, below we focus on the homogeneous case ($f_i \equiv f, g_i \equiv g$ for $i = 1, \cdots, p$) and the constant $\delta = 0$, as the proof for the non-homogeneous case and the positive $\delta > 0$ is similar.

In the proof of Theorem 1, the key idea is to decompose our proposed algorithm T_{WSLS} in (9) for the quickest change problem into a series of sequential tests for the sequential hypothesis testing problem, which is the reason why we introduce the reset concept to take advantage of the renewal properties of the local statistics. Mathematically, T_{WSLS} can be written as the sum,

$$\mathsf{T}_{\text{WSLS}} = \mathsf{T}_1 + \mathsf{T}_2 + \cdots + \mathsf{T}_k, \tag{16}$$

where each T_i is a sequential test and k is a random integer-valued variable to be defined in a moment.

First, all sequential tests T_i's are independent copies of the following sequential test $\mathcal{T}$:

$$\mathcal{T} = \min\{\mathcal{T}_1, \mathcal{T}_2, q_A\}, \tag{17}$$

where $\mathcal{T}_1$ indicates that a change has occurred and $\mathcal{T}_2$ indicates that local statistics at some observed stream is zero. They are defined rigorously by

$$\begin{aligned}
\mathcal{T}_1 &= \inf\{t : t > 0, S_t^1 + S_t^2 \geq A\}, \\
\mathcal{T}_2 &= \min(\mathcal{T}_{2a}, \mathcal{T}_{2b}), \\
\mathcal{T}_{2a} &= \inf\{t : t > 0, S_t^1 \leq 0\}, \\
\mathcal{T}_{2b} &= \inf\{t : t > 0, S_t^2 \leq 0\},
\end{aligned}$$

with $S_t^i = \sum_{\ell=1}^{t} \log(g(X_\ell^i)/f(X_\ell^i))(i = 1, 2)$.

The sequential tests T_i's are essentially applications of $\mathcal{T}$ in (17) to different blocks of data over time when we are sampling the same two streams. The random variable k in (16) is then defined as the first time when the sum of two log likelihoods in the sequential test hits the upper threshold A.

Second, we need to investigate the statistical properties of $\mathcal{T}$ in (17) under various probability measures. When monitoring two local streams, there are three possible scenarios: both are in pre-change, both are in post-change, and exactly one is pre-change. Thus we need to investigate the properties of the sequential test $\mathcal{T}$ under three kinds of distributions of $\{X_\ell^1, X_\ell^2\}_{\ell=1}^\infty$:

- $\{X_\ell^1, X_\ell^2\}_{\ell=1}^\infty \sim f$, denote as $\mathsf{E}_{f,f}$ and $\mathsf{P}_{f,f}$.
- $\{X_\ell^1, X_\ell^2\}_{\ell=1}^\infty \sim g$, denote as $\mathsf{E}_{g,g}$ and $\mathsf{P}_{g,g}$.
- $\{X_\ell^1\}_{\ell=1}^\infty \sim f, \{X_\ell^2\}_{\ell=1}^\infty \sim g$ or $\{X_\ell^1\}_{\ell=1}^\infty \sim g, \{X_\ell^2\}_{\ell=1}^\infty \sim f$, denote as $\mathsf{E}_{f,g}$ and $\mathsf{P}_{f,g}$.

Third, we are ready to prove Theorem 1 by exploring the relation (16) and the properties of $\mathcal{T}$ under the above-mentioned three possible probability measures.

1) False alarm property (12) in Theorem 1: note that when there are no changes, all the sequential test $\mathsf{T}_i(i = 1, 2, \cdots)$ have the same distribution of $\mathcal{T}$ under $\mathsf{P}_{f,f}$. By (16), we have

$$\begin{aligned}
\mathsf{E}_\infty[\mathsf{T}_{\text{WSLS}}] &= \mathsf{E}_{f,f}[\mathcal{T}] + (1-\alpha)\mathsf{E}_{f,f}[\mathcal{T}] \\
&\quad + (1-\alpha)^2\mathsf{E}_{f,f}[\mathcal{T}] + \cdots \\
&= \frac{1}{\alpha}\mathsf{E}_{f,f}[\mathcal{T}]
\end{aligned} \tag{18}$$

where $\alpha = \mathsf{P}_{f,f}(S_{\mathcal{T}}^1 + S_{\mathcal{T}}^2 \geq A)$. Since $\mathsf{E}_{f,f}[\mathcal{T}] \geq 1$, it suffices to show that $\alpha \leq \exp(-A)$. To do so, in parallel to $\mathcal{T}_1$ and $\mathcal{T}_2$ in (17),we now consider another $\mathcal{T}_3$:

$$\mathcal{T}_3 = \inf\{t : t > 0, S_t^1 + S_t^2 \leq 0\}. \tag{19}$$

Clearly $\mathcal{T}_2 \leq \mathcal{T}_3$. Also note that $\min(\mathcal{T}_1, \mathcal{T}_3)$ is the standard SPRT based on the log-likelihood ratio $S_t^1 + S_t^2$ of two-dimensional data (X_t^1, X_t^2). It is clear that $\mathsf{P}_{f,f}(\mathcal{T}_1 \leq \mathcal{T}_3)$ is the Type-I error probability of this SPRT and we have $\mathsf{P}_{f,f}(\mathcal{T}_1 \leq \mathcal{T}_3) \leq \exp(-A)$ by the standard tool of changing measure, see Page 10 of Siegmund [26]. Hence

$$\begin{aligned}
\alpha = \mathsf{P}_{f,f}(\mathcal{T}_1 \leq \mathcal{T}_2, \mathcal{T}_1 \leq q_A) &\leq \mathsf{P}_{f,f}(\mathcal{T}_1 \leq \mathcal{T}_2) \\
&\leq \mathsf{P}_{f,f}(\mathcal{T}_1 \leq \mathcal{T}_3) \\
&\leq \exp(-A),
\end{aligned}$$

which complete the proof of (12).

2) Detection delay property (13) in Theorem 1: assume that the change occurs to the i_1-th and i_2-th streams at time $\tau \geq 0$. In this case, it is important for us to investigate the impact of change time τ to the detection delay. Mathematically, besides the stopping time k in (16) that indicates when our algorithm raises a global alarm, we also need to define another stopping time which SPRT block our algorithm stays when the change occurs at time τ. That is, as in (16), we define another stopping time $\upsilon(\tau)$ via the sequential tests T_i's:

$$\upsilon(\tau) = \inf\{i > 0, \mathsf{T}_1 + \cdots + \mathsf{T}_i \geq \tau\}. \tag{20}$$

Then conditional on $\mathsf{T}_{\text{WSLS}} > \tau$, the detection delay of our proposed algorithm can be written as

$$\begin{aligned}
\mathsf{T}_{\text{WSLS}} - \tau &= \Big(\mathsf{T}_1 + \cdots + \mathsf{T}_{\upsilon(\tau)} - \tau\Big) \\
&\quad + \Big(\mathsf{T}_{\upsilon(\tau)+1} + \cdots + \mathsf{T}_k\Big).
\end{aligned}$$

It remains to analyze the expectations of $M_1 = \mathsf{T}_1 + \cdots + \mathsf{T}_{\upsilon(\tau)} - \tau$ and $M_2 = \mathsf{T}_{\upsilon(\tau)+1} + \cdots + \mathsf{T}_k$ conditional on $\{\mathsf{T}_{\text{WSLS}} > \tau\}$. The analysis for the conditional expectation of M_2 is tedious but straightforward, while it is highly non-trivial to investigate the conditional expectation of M_1, as it is possible that we are observing two unaffected data streams but the local statistics for these two observed unaffected local streams might be large (though smaller than A) at time τ. Here we bypass this difficulty through the controlling parameter $q = q_A$ which makes sure that the local statistics for these two observed unaffected local streams cannot be too large. The detailed arguments will be presented elsewhere due to page limits.

REFERENCES

[1] D. L. Donoho, “Compressed sensing”, *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp.1289–1306, 2006.

[2] E. J. Candes, J. K. Romberg and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements”, *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 59, no. 8, pp. 1207–1223, 2006.

[3] J. Li and J. Jin, “Optimal sensor allocation by integrating causal models and set-covering algorithms”, *IIE Transaction*, vol. 42, no. 8, pp. 564–576, 2010.

[4] A. Laszka, W. Abbas, and X. Koutsoukos, “Scheduling battery-powered sensor networks for minimizing detection delays”*IEEE COMMUNICATIONS LETTERS*, vol. 21, no. 4, pp. 789–792, 2017.

[5] P. Barrio and L. Gravano, “Sampling strategies for information extraction over the deep web”, *Information Processing & Management*, vol. 53, no. 2, pp.!309–331, 2017.

[6] A. Boicea, C. O. Truică, F. Rădulescu and E. C. Buse, “Sampling strategies for extracting information from large data sets”, *Data & Knowledge Engineering*, vol. 115, pp. 1–15, 2018.

[7] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Applications*, Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[8] H. V. Poor and O. Hadjiliadis, “Quickest Detection”, Cambridge University Press, 2008.

[9] A. Tartakovsky, I. Nikiforov, M. Basseville, *Sequential Analysis: Hypothesis Testing and Changepoint Detection*, Chapman and Hall/CRC, 2014.

[10] Y. Mei, “Efficient scalable schemes for monitoring a large number of data streams,” *Biometrika*, vol. 97, no. 2, pp. 419-433, 2010.

[11] Y. Xie and D. Siegmund, “Sequential multi-sensor change-point detection,” *The Annals of Statistics*, vol. 41, no. 2, pp. 670-692, 2013.

[12] H. P. Chan, “Optimal sequential detection in multi-stream data,” *The Annals of Statistics*, vol. 45, no. 6, pp. 2736-2763, 2017.

[13] A. N. Shiryaev, “On the detection of disorder in a manufacturing process. I,” *Theory of Probability and its Applications,* vol. 8, no. 3, pp. 247-265, 1963.

[14] Q. Xu, Y. Mei and G. V. Moustakides, “Optimum multi-stream sequential change-point detection with sampling control,” *IEEE Transactions on Information Theory,* vol. 67, no. 11, pp. 7627–7636, 2021.

[15] Q. Xu and Y. Mei, “Multi-stream quickest detection with unknown post-change parameters under sampling control,,” 2021 IEEE International Symposium on Information Theory (ISIT), pp. 112-117, 2021.

[16] H. Robbins, “Some aspects of the sequential design of experiments,” *Bulletin of the American Mathematical Society*, vol. 58, no. 5, pp. 527–535, 1952.

[17] K. Liu, R. Zhang and Y. Mei, “Scalable SUM-Shrinakge schemes for distributed monitoring large-scale data streams,” *Statistica Sinica,* vol. 29, no. 1, pp. 1–22, 2019.

[18] T. Banerjee and V. V. Veeravalli,“Data-efficient quickest change detection in minimax settings,” *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6917-6931, 2013.

[19] J. Heydari, A. Tajer, and H. V. Poor. “Quickest detection of Gauss-Markov random fields.” 2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp. 808-814, 2015.

[20] T. L. Lai and H. Robbins, “Asymptotically efficient adaptive allocation rules”, *Advances in Applied Mathematics*, vol.6, no. 1, pp. 4-22, 1985.

[21] S. Li, X. Li, X. Wang and J. Liu, “Sequential hypothesis test with online usage-constrained sensor selection”, *IEEE Transactions on Information Theory*, vol. 65, no. 7, pp. 4392–4410, 2019.

[22] A. Tsopelakos, G. Fellouris, and V. V. Veeravalli, “Sequential anomaly detection with observation control,” 2019 IEEE International Symposium on Information Theory (ISIT), pp. 2389-2393, 2019.

[23] M. Pollak, “Optimal detection of a change in distribution”, *The Annals of Statistics*, vol. 13, no. 1, pp. 206-277, 1985.

[24] G. Lorden, “Procedures for reacting to a change in distribution,” *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1897–1908, 1971.

[25] G. V. Moustakides, “Optimal stopping times for detecting changes in distributions,” *The Annals of Statistics*, vol. 14, no. 4, pp. 1379–1387, 1986.

[26] D. Siegmund, *Sequential analysis: Tests and confidence intervals*, Springer, 1985.