

Monitoring High-dimensional Streaming Data via Fusing Nonparametric Shiryaev-Roberts Statistics

Xinyuan Zhang* and Yajun Mei†

*School of Industrial and Systems Engineering, Georgia Institute of Technology, {xzhang967, ymei3}@gatech.edu

†Department of Biostatistics, School of Global Public Health, New York University

Abstract—Monitoring high-dimensional streaming data has a wide range of applications in science, engineering, and industry. In this work, we propose an efficient and robust sequential change-point detection algorithm for monitoring high-dimensional streaming data. It has two components. At the local level, we adopt a window-limited nonparametric Shiryaev-Roberts (WL-NPSR) statistic for detecting potential distribution changes at each dimension of the streaming data. At the global level, we fuse local WL-NPSR statistics together to construct a global monitoring statistic via quantile filtering and sum-shrinkage functions. Theoretical analysis and extensive numerical experiments demonstrate the efficiency and robustness of our proposed algorithm.

I. INTRODUCTION

The explosive growth of high-dimensional data has become a hallmark of the modern information era. Real-time monitoring of high-dimensional data streams extends its reach across a multitude of domains like traffic networks, multistage manufacturing, epidemic surveillance, etc. A key objective in monitoring streaming data is to detect undesirable events as soon as possible, which can be uncovered through the distribution of data. From the statistical point of view, this can be modeled as the sequential change-point detection problem. Specifically, in a system, suppose one observes K -dimensional random vectors over time, whose distributions might change at some unknown time. The objective is to utilize the observed data to raise an alarm as quickly as possible if a change occurs while controlling the false alarm rate if there is no change.

There are two approaches to developing efficient global monitoring algorithms. The first one is to model the joint distributions of the K -dimensional random vectors under the pre- and post-change hypotheses, and then use the corresponding likelihood ratios to raise a global alarm, see [1]–[3]. While this approach often yields theoretically efficient procedures under parametric models, it is computationally infeasible when K is large, not to mention the difficulty of specifying the joint distributions of high-dimensional data. The second approach is to monitor each local dimension individually and then combine local detection statistics to form a global monitoring statistic. By doing so, the difficult problem of monitoring K -dimensional data is reduced to the easier problem of monitoring K one-dimensional local data simultaneously, which turns out to be able to balance the tradeoff between computational and statistical efficiency when

the local data are independent or weakly dependent, see [4]–[7]. Here we will focus on the second approach.

In this paper, we propose a window-limited nonparametric Shiryaev-Roberts (WL-NPSR) algorithm for monitoring high-dimensional data. At the global level, we propose a novel global monitoring statistic that combines the sum-shrinkage technique in [8] with the clever idea of quantiles filtering in [6]. Utilizing sum-shrinkage and quantiles enables more flexibility and robustness when facing different model assumptions or post-change scenarios. At the local level, we extend and modify a nonparametric Shiryaev-Roberts (SR) statistic for univariate data. Recall that many classical procedures have been developed for monitoring univariate data in the literature, including Shewhart Chart, CUSUM procedure, Shiryaev-Roberts procedure, EWMA chart, Scan Statistics, Generalized Likelihood Ratio (GLR) under the parametric models, see [9]–[14]; for nonparametric methods, rich research has also been done, often under the assumption that the post-change distribution is stochastically larger or smaller than the pre-change distribution, see [15]–[19]. We take advantage of such extensive literature to develop robust univariate sequential change-point detection statistics that are based on the likelihood ratios of sequential ranks, which can detect both stochastically larger and smaller post-change distributions simultaneously.

The rest of the paper is organized as follows. In Section II, we present the problem formulation and briefly conduct the literature review. In Section III, we propose our monitoring algorithm and provide its theoretical properties. In IV, numerical experiments are conducted to illustrate the performance of our proposed algorithm. Technical proofs and more simulation studies are included in the Appendix.

II. PROBLEM FORMULATION AND LITERATURE REVIEW

We present problem formulations and a brief literature review in this section. This section is split into three subsections. Subsection II-A includes the mathematical formulation of the sequential change-point detection problem of monitoring high-dimensional data, and Subsection II-B reviews the parametric and nonparametric Shiryaev-Roberts statistics for univariate processes. In subsection II-C we review the sum-shrinkage technique in [8] and quantiles filtering in [6] when constructing global monitoring statistics.

A. Mathematical Formulation

Suppose that we are monitoring a system with K data streams. Denote by $X_{k,t}$ the observation from the k th local data stream at time t for $k = 1, \dots, K$ and $t = 1, 2, \dots$. Observations are assumed to be independent both among and within data streams. Initially, local data $\{X_{k,t}\}$ are i.i.d with the probability density function (pdf) $g_{k,0}$, which is assumed to be symmetrical about 0. At some unknown time point $\nu \in \{1, 2, \dots\}$, an undesirable event occurs and changes the pdf of the k th data stream to be $g_{k,1}$. We do not know how many and which data streams have changes in their distributions. Denote the pre- and post-change cumulative distribution function (cdf) as $G_{k,0}$ and $G_{k,1}$, respectively. The actual parametric forms of $G_{k,0}$ and $G_{k,1}$ are unknown, and the post-change distribution can be stochastically larger or smaller than the pre-change distribution. Denote by $\mathbb{P}_j$ and $\mathbb{E}_j$ the probability measure and expectation when the change happens at $\nu = j$, and denote the same by $\mathbb{P}_\infty$ and $\mathbb{E}_\infty$ when there is no change.

In sequential change-point detection problems, a statistical procedure or algorithm is defined as an integer-valued stopping time T with respect to the observation sequence $\{X_{k,t} : k = 1, \dots, K, t = 1, 2, \dots\}$. On the one hand, when there is no change, one would prefer to monitor the data as long as possible without interruptions. Mathematically, this is equivalent to having the stopping time T to be stochastically large under $\mathbb{P}_\infty$, and it is standard to formally state this as the false alarm constraint of the form

$$\mathbb{E}_\infty(T) \geq A, \quad (1)$$

where A is a pre-specified constant. On the other hand, once a change occurs, one wants to raise an alarm as quickly as possible or wants the stopping time T to have a small detection delay, defined by $(T - \nu + 1)^+$, under $\mathbb{P}_\nu$. There are several standard ways to define expected detection delay under the minimax formulations. One is under the conditional expectation, see Lorden [20] or Pollak [21], e.g., $D_{\text{POL}}(T) = \sup_{1 \leq \nu < \infty} \mathbb{E}_\nu(T - \nu + 1 | T \geq \nu)$. The other is through the marginal/unconditional distribution, defined in Lai [14],

$$D_{\text{LAI}}(T) = \mathbb{E}_\nu(T - \nu + 1)^+. \quad (2)$$

Here we adopt Lai's unconditional detection delay $D_{\text{LAI}}(T)$ in (2) for theoretical and numerical analyses, as it is still open problems to investigate the theoretical properties under Lorden or Pollak's detection delay criteria that needs to address the conditional performance under the rare events.

B. Shiryaev-Roberts Statistic

In the simplest sequential change-point detection problem of monitoring i.i.d univariate data $\{X_t\}$ whose distribution might change from g_0 to g_1 , the well-known Shiryaev-Roberts (SR) procedure is defined as the stopping time $T_{\text{SR}} = \inf\{t : R_t \geq A\}$, where the SR statistic R_t is given by

$$R_t = \sum_{j=1}^t \Lambda_j^t \quad \text{where} \quad \Lambda_j^t = \prod_{i=j}^t \frac{g_1(X_i)}{g_0(X_i)}. \quad (3)$$

This procedure has a natural Bayesian interpretation that the candidate change time ν occurs uniformly over $j = 1, 2, \dots, t$. It was shown in [21], [22] that the SR procedure is asymptotically optimal under Pollak's detection delay $D_{\text{POL}}(T)$ subject to the false alarm constraint in (1), as the lower bound A goes to infinity.

A nonparametric Shiryaev-Roberts (NPSR) detection procedure is proposed by [15]. The key technique in NPSR is to compute the likelihood ratios of signs and signed-ranks induced by observations under pre-specified distributions. To be more specific, at each time step t , two-dimensional summary statistics $Z_i^t := (\sum_{j=1}^t \mathbb{1}_{\{|X_j| \leq |X_i|\}}, \mathbb{1}_{\{X_i > 0\}})$ for each $i = 1, \dots, t$ are recorded. This follows the well-known Wilcoxon's signed-rank tests when testing whether the difference of two paired samples is symmetric about 0 or not. Next, two hypothesized working distributions are pre-specified, one for pre-change and the other for post-change:

$$\begin{aligned} f_0(x) &= \frac{1}{2} \exp\{-|x|\} \\ f_1(x) &= pa \exp(-ax) \mathbb{1}_{\{x \geq 0\}} + qb \exp(bx) \mathbb{1}_{\{x < 0\}}. \end{aligned} \quad (4)$$

For any continuous-valued data X that is symmetrical about 0, we consider the monotonic transformation

$$Q(x) := (1 - 2\mathbb{1}_{\{x > 0\}}) \log(2 - 2G_0(|x|)) \quad (5)$$

on it. It is straightforward to show that the transformed data $Q(X)$ will have the hypothesized working distribution f_0 in (4) while preserving the same summary statistics Z_i^t which are based on signs and orderings. The positive parameters $a, b, p, q = 1 - p$ in (4) are tuning parameters that can be chosen to resemble the true post-change distribution of the transformed data $Q(X)$. Finally, the NPSR computes the likelihood ratio of two-dimensional vectors $Z_i^t, i = 1, \dots, t$ under hypothesized working distributions when the change occurs at the time $\nu = j$, i.e., Λ_j^t in the SR statistics (3) is replaced by

$$\Lambda_j^t(Z_1^t, \dots, Z_t^t) := \frac{\mathbb{P}_{\nu=j}(Z_1^t, \dots, Z_t^t)}{\mathbb{P}_\infty(Z_1^t, \dots, Z_t^t)}. \quad (6)$$

This nonparametric likelihood ratio has a closed form for any finite sequence of samples. It was shown in [15] that the NPSR is robust and efficient, though it is computationally expensive.

C. Sum-Shrinkage Method and Quantiles filtering

Motivated by communications in sensor networks, Liu et al. [8] introduce the sum-shrinkage technique to construct global monitoring statistic

$$G_t = \sum_{k=1}^K h_k(W_{k,t}), \quad (7)$$

where $W_{k,t}$ is the local detection statistic. There are many choices for the shrinkage function h_k such as hard-thresholding $h_k(x) = x \mathbb{1}_{\{x > b_k\}}$ or soft-thresholding $h_k(x) = \max\{x - b_k, 0\}$ for some constant b_k . The shrinkage transformations play the role of increasing signal-to-noise by filtering out those in-control local data streams and only gathering

information from likely out-of-control data streams. Unfortunately, it is highly non-trivial to choose a suitable constant filtering parameter b_k that works well in a general context. To develop a robust global monitoring statistic and get around the challenges of choosing the parameter b_k 's, Li [6] proposes an adaptive global monitoring statistic that can automatically adjust to different post-change scenarios. The main idea is to utilize the quantiles of local detection statistics under the pre-change hypothesis, and then compare them with the order statistics of observed local detection statistics. In particular, the corresponding global monitoring statistic is defined as

$$G_t = \sum_{k=1}^K (W_{(k),t} - q_{(k),t})^2 \mathbb{1}_{\{W_{(k),t} > q_{(k),t}\}}, \quad (8)$$

where $q_{(k),t}$ is the $(k-0.75)/(K-0.5)$ quantile of the local detection statistics $W_{k,t}$ under the pre-change hypothesis, which can be either theoretically derived or numerically simulated.

III. OUR PROPOSED ALGORITHM

We are now ready to present our proposed robust algorithm for monitoring high-dimensional streaming data. This section includes three components. In Subsection III-A we develop a window-limited version of the NPSR that is robust and computationally efficient. In subsection III-B we combine the sum-shrinkage technique with quantile-filtering to construct a novel global monitoring statistic that is efficient under our context. Due to page limits, brief theoretical properties are summarized in Subsection III-C without proofs.

A. Local Detection Statistics

We propose to extend the NPSR by choosing two sets of tuning parameters in (4), (a^+, b^+, p^+) and (a^-, b^-, p^-) for detecting stochastically larger and smaller local changes, respectively. As shown in [15], the general requirements would be $0 \leq a^+ \leq 1 \leq b^+ < \infty$, $a^+ < b^+$ and $p^+ a^+ \geq q^+ b^+$, whereas $0 \leq b^- \leq 1 \leq a^- < \infty$, $a^- > b^-$ and $p^- a^- \leq q^- b^-$. Under symmetric cases, one can choose $p^+ = 1 - p^-$, $a^+ = b^-$ and $b^+ = a^-$.

In addition, to reduce the computational burden in (6), we apply a window-limited approach where only observations from a sliding window of fixed length will be considered. Denote the fixed window length with ℓ . That is, we only utilize the window-limited data $X_{(t-\ell)+1}, \dots, X_t$ to compute the two-dimensional summary statistics

$$Z_i^{\ell,t} := \left(\sum_{j=1}^{t \wedge \ell} \mathbb{1}_{\{|X_{j+(t-\ell)+1}| \leq |X_{i+(t-\ell)+1}|\}}, \mathbb{1}_{\{X_{i+(t-\ell)+1} > 0\}} \right) \quad (9)$$

for $i = 1, \dots, t \wedge \ell$. Then we compute the likelihood ratio $\Lambda_j^t(Z_1^{\ell,t}, \dots, Z_{t \wedge \ell}^{\ell,t})$ using formula (6).

Define by $\Lambda_j^{+,t}$ and $\Lambda_j^{-,t}$ the nonparametric likelihood ratios computed by (6) with (a^+, b^+, p^+) and (a^-, b^-, p^-) , respectively. Also define by R_t^+ and R_t^- the corresponding NPSR statistics in (3). Then our proposed local detection statistic is defined as.

$$R_t^\ell := \max\{R_t^+, R_t^-\}, \quad (10)$$

where superscript ℓ is used to emphasize the window length.

It is natural to apply R_t^ℓ in (10) to all K local data streams, thereby yielding K local detection statistics $R_{1,t}^\ell, \dots, R_{K,t}^\ell$ at each time t . When the dimension K is large, unfortunately, there are often too many ties in these local detection statistics, and thus we propose to follow the multi-armed bandit problems to introduce randomness to break ties. To be more specific, let $U_{k,t} \stackrel{\text{i.i.d.}}{\sim} \text{Uniform}[0, 1]$ and denote that

$$\mathcal{R}_{k,t} := R_{k,t}^\ell + U_{k,t} \quad (11)$$

for $k = 1, \dots, K$. It turns out that adding jitters will stabilize the performance of our proposed algorithm in the initial time steps while having little effects for the large time steps.

B. Global Monitoring Statistics

At time t , with the observed K local statistics in (11) after adding jitters, denote by $\mathcal{R}_{(1),t} \leq \dots \leq \mathcal{R}_{(K),t}$ the corresponding order statistics. We combine the sum-shrinkage technique in [8] with quantile filtering in [6] to define our global monitoring statistic as

$$G_t = \sum_{k=1}^K |\mathcal{R}_{(k),t} - q_{(k),t}| \mathbb{1}_{\{\mathcal{R}_{(k),t} > q_{(k),t}\}}, \quad (12)$$

where $q_{(k),t}$ is the k/K quantile of $\mathcal{R}_{(k),t}$ to be discussed a little bit later. In our context, our experiences show that the L_1 norm in (12) will yield better performance than the L_2 norm in (8). Part of the reason is probably that we adopt the SR statistic, which is on the likelihood ratio scale, whereas Li [6] considers the CUSUM statistics in (12) which is on the log-likelihood ratio scale.

Our proposed Window-Limited Nonparametric Shiryaev-Roberts algorithm raises a global alarm at the time

$$T_A^G = \inf\{t \geq 1 : G_t \geq A\}. \quad (13)$$

Let us now discuss how to derive the pre-change quantiles $q_{(k),t}$, $k = 1, \dots, K$. Theoretically, this is very challenging; but fortunately, it is straightforward to compute them numerically via Monte Carlo simulations. To be more concrete, we generate $t \wedge \ell$ independent local raw data $X_1^*, \dots, X_{t \wedge \ell}^*$ from any given distributions symmetric at 0, e.g., standard normal, also sample $U_1^*, \dots, U_{t \wedge \ell}^*$ from Uniform[0, 1], and then use them to compute the local detection statistics $\mathcal{R}_{k,t}$ in (11). This process is repeated B times so that we get B realizations. Then we take the k/K sample quantile $\hat{q}_{(k),t}$ from these B realizations to approximate the quantiles $q_{(k),t}$, for $k = 1, \dots, K$. It is useful to highlight that the quantile calculation step can be conducted offline, and only needs to be done once. This means that the corresponding computation burden or memory requirement is marginal for online monitoring.

C. Theoretical Properties

In this subsection, we present some theoretical properties of our proposed algorithm.

The following lemma demonstrates that different parts R_t^+, R_t^- of the local detection statistics R_t^ℓ in (10) will play

different roles when facing stochastically larger or smaller post-change distributions, which is due to the monotonicity of the window-limited likelihood ratio Λ_k^t in (6).

Lemma 1: Denote $R_t(x_1, \dots, x_t)$ as the nonparametric SRs computed on samples $x_1, \dots, x_t$. Let ψ^+, ψ^- be functions such that $\psi^+(x) \leq x \leq \psi^-(x)$. For any $i \leq t$, if $0 \leq a \leq 1 \leq b$ and $pa \geq qb$,

$$\begin{aligned} & R_t(x_1, \dots, x_{i-1}, \psi^+(x_i), \dots, \psi^+(x_t)) \\ & \leq R_t(x_1, \dots, x_t) \\ & \leq R_t(x_1, \dots, x_{i-1}, \psi^-(x_i), \dots, \psi^-(x_t)). \end{aligned} \quad (14)$$

If $0 \leq b \leq 1 \leq a < \infty$ and $qb \leq pa$, the above inequalities will all be reversed.

By Lemma 1, when $G_1(x) \leq G_0(x)$, i.e., when the post-change distribution is stochastically larger than the pre-change distribution, if we set $\psi^+(x) = G_0^{-1}(G_1(x))$, then the value of R_t^+ will likely increase and R_t^- will likely decrease as compared to their null pre-change values; the conclusions will reverse when the post-change distribution is stochastically smaller than the pre-change distribution.

Next, we present the performance of our proposed algorithm when the dimension $K = 1$ with full data, i.e., $\ell = \infty$, and the corresponding stopping time is denoted as T_A . Recall that f_1, f_0 are Laplace densities given in Subsection II-B, $g_1(x)$ is the (unknown) true post-change density function, and Q is the transformation in (5).

Lemma 2: Suppose under $\mathbb{P}_\infty$, the i.i.d observations $X_1, X_2, \dots$ follow some continuous distribution symmetrical about 0. Then

$$\mathbb{E}_\infty\{T_A\} \geq \frac{A-1}{2} \quad (15)$$

and there exists a constant C such that

$$\lim_{A \rightarrow \infty} \frac{\mathbb{E}_\infty\{T_A\}}{A} \leq C. \quad (16)$$

Moreover, when the Kullback–Leibler divergence

$$D_Q = \int_{\mathbb{R}} g_1(x) \log \frac{f_1(Q(x))}{f_0(Q(x))} dx, \quad (17)$$

is finite and positive, we have

$$\lim_{A \rightarrow \infty} \limsup_{k \rightarrow \infty} \frac{\mathbb{E}_k\{(T_A - k + 1)^+\}}{\log(A)} \leq D_Q^{-1}. \quad (18)$$

The proof of Lemma 2 is based on a non-trivial modification of those proofs in [15] by taking into account of two-sided tests in (10) and added jettors in (10). Note that the constant C in (16) can be explicitly characterized as $C = \min\{1/a^+, 1/b^-\} > 1$ if $p^+a^+ \vee q^+b^+ \leq 1/2$ and $p^-a^- \vee q^-b^- \leq 1/2$.

Finally, we are ready to present the performance of our proposed algorithm under the general dimension K and finite window length ℓ . The following theorem provides a rough lower bound for the average run length to the false alarm.

Theorem 1: Assume $A > K$. Then

$$\liminf_{\ell \rightarrow \infty} \mathbb{E}_\infty\{T_A^G\} \geq \frac{A-K}{2K}. \quad (19)$$

Combining Lemma 2 and Theorem 1 along with the detection delay technique in [23], when $\log(A) \gg \log(K)$, our proposed algorithm T_A^G has efficient detection delays under Lai's unconditional definition in (18). With this said, unfortunately, it remains an open problem to investigate the detection delays under Lorden's or Pollak's definitions, which involve the conditional detection delays under rare events.

IV. NUMERICAL EXPERIMENTS

In this section, we present the results of our numerical studies to illustrate the performance of our proposed algorithm. Our simulation settings are as follows. There are $K = 100$ data streams in the system and $n_c \in \{100, 80, 50, 20, 10, 8, 5, 3, 1\}$ of them might have location shifts δ at the time $\nu = 51$. When the system is in-control, data stream $\{X_{k,t}\}$ follows $N(0,1)$ if its index k is odd, otherwise it follows standardized $T(2.5)$. We consider two sub-cases:

- 1) **Equal magnitudes:** $\{X_{k,t}\}, k \leq \lceil n_c/2 \rceil$ have location shifts $\delta = 1$; $\{X_{k,t}\}, \lceil n_c/2 \rceil + 1 \leq k \leq n_c$ have location shifts $\delta = -1$.
- 2) **Mixed magnitudes:** $\{X_{k,t}\}, k \leq \lceil n_c/2 \rceil$ have location shifts of magnitude $|\delta_1| = 0.5$; $\{X_{k,t}\}, \lceil n_c/2 \rceil + 1 \leq k \leq n_c$ have location shifts of magnitude $|\delta_2| = 1$. For the same magnitude of $|\delta_i|, i = 1, 2$, half of them are positive and the rest are negative.

When $n_c = 1$, only $\{X_{1,t}\}$ will have a location shift of 1 in the first case or 0.5 in the second case. For the purpose of comparison, besides the proposed algorithm T_A^G in (13), we also consider three alternatives: T_L in (8) with sample quantiles, $T_{sum} = \inf\{t : \sum_k \mathcal{R}_{k,t} \geq A\}$ and $T_{max} = \inf\{t : \max_k \mathcal{R}_{k,t} \geq A\}$. The average run length to false alarm (ARL_0) of all algorithms are set to be 1000, and threshold A values for these four algorithms are obtained by Monte Carlo simulations to satisfy this ARL_0 constraint. In particular, the threshold A values for these four algorithms, $T_A^G, T_L, T_{sum}, T_{max}$, are chosen to be 10515.49, 75773623.25, 116444.61, 115250.06, respectively. All results are estimated from 2500 simulation runs and standard errors are reported in parenthesis. The average detection delay is estimated by the sample mean of $(T - \nu + 1)^+$, where T is the stopping time and the change-point ν is set to be $\nu = 51$.

The remainder of this section contains two subsections. In Subsection IV-A we discuss the choice of parameters. In Subsection IV-B we report the simulation results.

A. Choice of Parameters in Our Proposed Algorithm

As in [15], by Lemma 2, we choose post-change parameters $(a^\pm, b^\pm, p^\pm)$ to maximize D_Q in (17). Specifically, by taking partial derivatives of (17) with respect to p, a, b we will have

$$\begin{aligned} p &= 1 - G_1(0), \\ a &= \frac{p}{-\int_0^\infty g_1(x) \log(2 - 2G_0(x)) dx}, \\ b &= \frac{p-1}{\int_{-\infty}^0 g_1(x) \log(2 - 2G_0(-x)) dx}. \end{aligned} \quad (20)$$

TABLE I
AVERAGE DETECTION DELAY FOR $K = 100$ DATA STREAMS.

Shift Case	n_c	T_A^G	T_L	T_{sum}	T_{max}
1)	1	34.80(0.46)	35.26(0.47)	30.41(0.37)	30.63(0.38)
	3	13.85(0.10)	14.09(0.10)	15.25(0.11)	16.00(0.12)
	5	10.57(0.06)	11.11(0.07)	12.70(0.07)	13.50(0.08)
	8	8.36(0.05)	9.11(0.05)	10.86(0.06)	11.87(0.06)
	10	7.69(0.04)	8.54(0.04)	10.38(0.05)	11.52(0.06)
	20	6.01(0.03)	7.03(0.04)	8.73(0.04)	10.24(0.05)
	50	4.41(0.02)	5.72(0.03)	7.07(0.03)	9.00(0.04)
	80	3.71(0.02)	5.08(0.02)	6.27(0.03)	8.30(0.04)
	100	3.45(0.02)	4.82(0.02)	5.93(0.03)	8.08(0.04)
2)	1	238.4(4.24)	241.3(4.23)	226.3(4.00)	232.0(4.08)
	3	23.39(0.24)	24.02(0.25)	24.63(0.24)	25.48(0.25)
	5	14.47(0.11)	14.84(0.11)	15.89(0.13)	16.45(0.13)
	8	10.36(0.06)	10.98(0.07)	12.80(0.07)	13.43(0.08)
	10	9.29(0.05)	10.08(0.05)	11.67(0.06)	12.63(0.07)
	20	7.20(0.04)	8.31(0.04)	10.08(0.05)	11.32(0.06)
	50	5.19(0.02)	6.59(0.03)	8.07(0.04)	9.76(0.05)
	80	4.38(0.02)	5.83(0.03)	7.14(0.03)	9.16(0.05)
	100	4.02(0.02)	5.53(0.03)	6.82(0.03)	8.83(0.04)

For simplicity, we take G_0 to be standard normal, use G_1^+ as $N(1, 1)$ to compute (a^+, b^+, p^+) , G_1^- as $N(-1, 1)$ to compute (a^-, b^-, p^-) . Then by numerical integration, $p^+ = 1 - p^- = 0.8413$, $a^+ = b^- = 0.531$, $b^+ = a^- = 1.703$. Moreover, for computationally simplicity, we choose the window length $\ell = 30$ for our proposed algorithm T_A^G here, although our extensive numerical experiences suggest that the window length $\ell = 50$ or $\ell = 100$ also yield similar results. Thus we use these parameters for all the numerical experiments below.

B. Detection Delay

Table I summarizes the expected detection delays along with the standard errors for our proposed algorithm T_A^G with the window length $\ell = 30$ as well as three baseline alternatives. It is clear from the table that our proposed algorithm, T_A^G , yields the smallest detection delays in almost all situations, except when only one local stream is affected. This is consistent with our intuition. Note that our local statistics will not accumulate all temporal information since they only utilize the most recent ℓ observations. When ℓ is fixed and $t > \ell$, we are testing whether the most recent ℓ samples come from a fixed distribution. If there is only one data stream having a small change, the type-II error can be large and it will be difficult to test out the deviation from the pre-change distribution. This explains why all algorithms using the proposed local statistics (11) would yield large detection delays when there is only one data stream having a 0.5 location shift.

It is also interesting to note that the analog global statistic in (8), using squared difference, always yields slightly larger detection delay than T_A^G . Thus using L_1 -norm to fuse all local detection statistics seems to be more efficient as compared to the L_2 -norm.

A very surprising phenomenon is that for the SR-type local detection statistics, the "SUM" rule appears to yield smaller detection delay than the "MAX" rule in all of our numerical setting. This seems to different from extensive existing research that the "SUM" rule of local CUSUM statistics would

yield larger detection delays than the "MAX" rule when the number of affected number is small, see [4], [6], [8].

At the high-level, such disparity might be because while SR and CUSUM local statistics yield similar stopping times when monitoring one-dimensional data, their properties after fusing at the global level are completely different! In particular, the SR statistic $R_{k,t}$ is in the likelihood scale, whereas the CUSUM statistics $W_{k,t}$ is in the log-likelihood scale. Below we sketch a high-level explanation to clarify such inconsistency. Indeed, in the sequential change-point detection literature, we would treat $W_{k,t} \approx \log(R_{k,t})$, or equivalently, $R_{k,t} \approx \exp(W_{k,t})$. In particular, if the k -th local stream is affected under the post-change hypothesis, $W_{k,t}$ grows at the rate of $I_k d$ whereas $R_{k,t}$ grows at the rate of $\exp(I_k d)$ as the detection delay $d \rightarrow \infty$. Assume there are $n_c \geq 2$ affected local streams, and without loss of generality, the first n_c local streams are affected with signal-to-noise ratios in the decreasing order: $I_1 > I_2 > \dots > I_{n_c}$. When the local CUSUM statistics are the building blocks, the asymptotic ratios of the detection delays of the "MAX" and "SUM" rules would be

$$\frac{\sum_{k=1}^{n_c} I_k d}{\max_{1 \leq k \leq n_c} I_k d} = 1 + \frac{\sum_{k=2}^{n_c} I_k}{I_1} > 1, \quad (21)$$

which is consistent with the literature that the "SUM" and "MAX" algorithms have very different asymptotic properties when the local CUSUM statistics are used.

Meanwhile, when the local SR statistics are used as building blocks, the corresponding ratio would become

$$\frac{\sum_{k=1}^{n_c} \exp(I_k d)}{\max_{1 \leq k \leq n_c} \exp(I_k d)} = 1 + \sum_{k=2}^{n_c} e^{(I_k - I_1)d} \rightarrow 1 \quad (22)$$

as $d \rightarrow \infty$. In other words, we might expect that $\sum_{k=1}^K R_{k,t}$ has similar asymptotic properties as $\max_{1 \leq k \leq K} R_{k,t} = \exp(\max_{1 \leq k \leq K} W_{k,t})$. While we provide some asymptotic arguments, it is still an open problem to investigate the finite-sample properties of the "SUM" or "MAX" rules on the local SR-type statistics, also see [24] for similar observations of fusing local SR statistics in different contexts.

V. CONCLUDING REMARKS

We proposed an algorithm for monitoring high-dimensional streaming data by fusing local two-sided nonparametric SR statistics into efficient a global monitoring statistic via sum-shrinkage and quantile-filtering. The resulting algorithm is robust to model assumptions and yields satisfactory performance in detection delays. A key tuning parameter is the window length ℓ , which deserved further finite-sample analysis, so as to balance the tradeoff between the computational efficiency and statistical efficiency.

ACKNOWLEDGEMENT

This research was supported in part by an NSF-DMS grant 2015405 and by an NIH grant 1R21AI157618-01A1. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

- [1] C. Zou and P. Qiu, "Multivariate statistical process control using lasso," *Journal of the American Statistical Association*, vol. 104, no. 488, pp. 1586–1596, 2009.
- [2] Y. Xie and D. Siegmund, "Sequential multi-sensor change-point detection1," *The Annals of Statistics*, vol. 41, no. 2, pp. 670–692, 2013.
- [3] Y. Wang and Y. Mei, "Large-scale multi-stream quickest change detection via shrinkage post-change estimation," *IEEE Transactions on Information Theory*, vol. 61, no. 12, pp. 6926–6938, 2015.
- [4] Y. Mei, "Efficient scalable schemes for monitoring a large number of data streams," *Biometrika*, vol. 97, no. 2, pp. 419–433, 2010.
- [5] J. Li, "A two-stage online monitoring procedure for high-dimensional data streams," *Journal of Quality Technology*, vol. 51, no. 4, pp. 392–406, 2019.
- [6] —, "Efficient global monitoring statistics for high-dimensional data," *Quality and Reliability Engineering International*, vol. 36, no. 1, pp. 18–32, 2020.
- [7] M. Pollak, "A rule of thumb: Run lengths to false alarm of many types of control charts run in parallel on dependent streams are asymptotically independent," *The Annals of Statistics*, 2021.
- [8] K. Liu, R. Zhang, and Y. Mei, "Scalable sum-shrinkage schemes for distributed monitoring large-scale data streams," *Statistica Sinica*, vol. 29, no. 1, pp. 1–22, 2019.
- [9] W. A. Shewhart, "The application of statistics as an aid in maintaining quality of a manufactured product," *Journal of the American Statistical Association*, vol. 20, no. 152, pp. 546–548, 1925.
- [10] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.
- [11] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory of Probability & Its Applications*, vol. 8, no. 1, pp. 22–46, 1963.
- [12] S. Roberts, "A comparison of some control chart procedures," *Technometrics*, vol. 8, no. 3, pp. 411–430, 1966.
- [13] J. S. Hunter, "The exponentially weighted moving average," *Journal of quality technology*, vol. 18, no. 4, pp. 203–210, 1986.
- [14] T. L. 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.
- [15] L. Gordon and M. Pollak, "An efficient sequential nonparametric scheme for detecting a change of distribution," *The Annals of Statistics*, pp. 763–804, 1994.
- [16] P. Qiu and D. Hawkins, "A nonparametric multivariate cumulative sum procedure for detecting shifts in all directions," *Journal of the Royal Statistical Society Series D: The Statistician*, vol. 52, no. 2, pp. 151–164, 2003.
- [17] N. Das, "A new multivariate non-parametric control chart based on sign test," *Quality Technology & Quantitative Management*, vol. 6, no. 2, pp. 155–169, 2009.
- [18] C. Zou and F. Tsung, "A multivariate sign ewma control chart," *Technometrics*, vol. 53, no. 1, pp. 84–97, 2011.
- [19] J. Li, "Nonparametric adaptive cusum chart for detecting arbitrary distributional changes," *Journal of Quality Technology*, vol. 53, no. 2, pp. 154–172, 2021.
- [20] G. Lorden, "Procedures for reacting to a change in distribution," *Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1897–1908, 1971.
- [21] M. Pollak, "Optimal detection of a change in distribution," *The Annals of Statistics*, pp. 206–227, 1985.
- [22] —, "Average run lengths of an optimal method of detecting a change in distribution," *The Annals of Statistics*, pp. 749–779, 1987.
- [23] R. Zhang and Y. Mei, "Asymptotic statistical properties of communication-efficient quickest detection schemes in sensor networks," *Sequential Analysis*, vol. 37, no. 3, pp. 375–396, 2018.
- [24] Y. Wu, "A combined sr-cusum procedure for detecting common changes in panel data," *Communications in Statistics-Theory and Methods*, vol. 48, no. 17, pp. 4302–4319, 2019.