Abstract—A generalized multisensor sequential change detection problem is considered, in which a number of (possibly correlated) sensors monitor an environment in real time, the joint distribution of their observations is determined by a global parameter vector, and at some unknown time there is a change in an unknown subset of components of this parameter vector. The goal is to detect the change as soon as possible, while controlling the rate of false alarms. We establish the second-order asymptotic optimality (with respect to Lorden’s criterion) of various generalizations of the CUSUM rule; that is, we show that their additional expected worst case detection delay (relative to the one that could be achieved if the affected subset was known) remains bounded as the rate of false alarm goes to 0, for any possible subset of affected components. This general framework incorporates the traditional multisensor setup in which only an unknown subset of sensors is affected by the change. The latter problem has a special structure which we exploit in order to obtain feasible representations of the proposed schemes. We present the results of a simulation study where we compare the proposed schemes with scalable detection rules that are only first-order asymptotically optimal. Finally, in the special case that the change affects exactly one sensor, we consider the scheme that runs in parallel the local CUSUM rules and study the problem of specifying the local thresholds.

Index Terms—Asymptotic optimality, CUSUM, sequential, change detection, multichannel, multisensor.

I. INTRODUCTION

Suppose that a decision maker sequentially collects data from multiple streams, coming for example from sensors monitoring an environment. At some unknown point in time there is an abrupt change in the system that is perceived by at least a subset of the deployed sensors. The problem then is to combine the sequentially acquired observations from all streams in order to detect the change as quickly as possible, while controlling the rate of false alarms below an acceptable level. In what follows, we will refer to the various streams as “sensors”, although this general framework can be applied to various setups beyond environmental monitoring, such as intruder detection in computer security [1]–[3], or epidemic detection in bio-surveillance [4], [5].

If the regime before and after the change is completely specified, one may apply a classical sequential change detection algorithm, such as the Cumulative Sum (CUSUM) rule, which was proposed by Page [6] and was shown by Moustakides [7] to be the best possible rule in Lorden’s [8] minimax setup in the case of iid observations before and after the change. However, in many applications that motivate this multisensor sequential change detection problem it is reasonable to assume that there is only partial information regarding the post-change regime. For example, the change may affect the distribution of only an unknown subset of sensors. In this context, we need to design detection rules that account for this uncertainty, and also quantify how much is lost relative to the ideal case that the actual affected subset is known.

The simplest multisensor detection rule is the multichart CUSUM, according to which each sensor runs locally the CUSUM algorithm and the fusion center stops the first time an alarm is raised by any sensor. This is a one-shot scheme, suitable for decentralized implementation, as each possibly remote sensor needs to communicate with the fusion center at most once, an important advantage in sensor networks that are characterized by limited bandwidth and energy. Tartakovsky et al. [1], [2] showed that this rule has a strong asymptotic optimality property when the change affects exactly one sensor (see also [9, Sec. 9.2]). However, this rule is very inefficient when a large, or even a moderate, number of sensors are actually affected. This calls for detection rules that are robust, in the sense that they should perform well for a large class of scenarios regarding the subset of affected sensors.

In order to address this problem, Mei [10] proposed raising an alarm when the sum of the local CUSUM statistics exceeds a threshold, a rule to which we will refer as SUM-CUSUM. Under the assumptions of iid observations before/after the change, completely specified but arbitrary pre/post-change distributions (up to an integrability condition) and independent sensors it was shown that SUM-CUSUM is asymptotically optimal for every possible subset of affected sensors. Moreover, based on simulation experiments it was argued that SUM-CUSUM performs better than the multichart CUSUM unless a very small number of sensors is actually affected. This scheme was further refined by Mei in [11], where it was suggested to use the sum of only the top \(L \) local
CUSUM statistics when at most $L$ sensors can be affected. Liu et al. [12] and Banerjee and Veeravalli [13] considered further modifications that allow for only a subset of sensors to be employed at any given time.

A different, CUSUM-based, multisensor detection scheme was proposed by Xie and Siegmund [14]. Specifically, the proposed rule in this work was a modification of the CUSUM rule that corresponds to the case that the change occurs with probability $\pi$ in each sensor, where $\pi$ is the actual proportion of affected sensors. Asymptotic approximations for the operating characteristics of this detection rule were obtained in the case of independent sensors, each of which observes Gaussian iid observations whose mean may change from 0 to some unknown value. Moreover, it was argued based on simulation experiments that this detection rule is more efficient than SUM-CUSUM and this phenomenon was explained on the basis that SUM-CUSUM does not incorporate the assumption that all affected sensors have the same change-point.

The main motivation for our work was to propose and study multisensor detection rules that enjoy a stronger asymptotic optimality property than SUM-CUSUM and are more efficient in practice. However, we consider a more general setup that incorporates applications that are not captured by the traditional multisensor framework, such as the power outage detection problem treated by Chen et al. [15]. Specifically, we assume that there is a global parameter vector that determines the joint distribution of the (possibly correlated) data streams and that the change modifies a subset of the components of this parameter vector. The affected subset of components is unknown, but it is assumed to belong to a given class of subsets, $\mathcal{P}$, that reflects our prior information regarding the location of the change. For example, in the extreme case that there is no information regarding the affected subset, $\mathcal{P}$ consists of all non-empty subsets of components.

In this generalized multisensor setup, we consider first of all the CUSUM-type rule that employs the Generalized Likelihood Ratio (GLR) statistic with respect to the unknown subset, to which we refer as GLR-CUSUM. Moreover, we consider two alternative, mixture-based CUSUM-type schemes. We show that all these detection rules enjoy a second-order asymptotic optimality property, i.e., that their additional worst-case detection delay (relative to the one that could be achieved if the affected subset was known) remains bounded as the rate of false alarms goes to 0 for any possible subset of affected components. This should be contrasted with the usual, weaker asymptotic optimality property, according to which the asymptotic relative efficiency goes to 1, and to which we refer as first-order asymptotic optimality.

Our general framework includes as a special case the traditional setup in which the change affects only an unknown subset of sensors, leaving the remaining ones completely unaffected. In this special case, assuming further independent sensors and an upper bound on the size of the affected subset, we obtain feasible representations of the GLR-CUSUM and one of the proposed mixture-based rules. Moreover, we show that a detection rule in [14] is also second-order asymptotically optimal for any choice of the parameter $\pi$, which however should not be interpreted as the proportion of affected sensors. Thus, we provide a theoretical explanation for the empirically observed superiority of this rule over SUM-CUSUM, which is only first-order asymptotically optimal, as well as its robustness with respect to $\pi$.

In the special case that it is known in advance that the change affects exactly one sensor, GLR-CUSUM reduces to the multichart CUSUM, according to which an alarm is raised when a local CUSUM statistic exceeds a local threshold. The selection of identical thresholds seems to be a default choice in the literature, although non-identical thresholds have also been considered (see, e.g. [9, p. 467]). We address the problem of threshold selection in this setup and we show that seemingly reasonable threshold selections may lead to non-robust behavior, as the multichart CUSUM may fail to be even first-order asymptotically optimal. We then argue in favor of a specific selection of thresholds that equals the asymptotic relative performance loss.

Let us discuss at this point the connection of our work with the literature of sequential change detection. Lorden [8] assumed (in a “single-sensor” setup) that the distribution of the observations belongs to an exponential family and is specified after the change up to an unknown parameter. In this context, he established the first-order asymptotic optimality of a CUSUM-type rule that is based on (a truncated away from 0 version of) the GLR statistic. In the special case of detecting a shift in the mean of Gaussian observations, Siegmund and Venkatraman [16] obtained approximations for the operating characteristics of the corresponding GLR-CUSUM. Our work differs from these two papers in that in our setup there is a finite number of scenarios in the post-change regime. Nevertheless, we also use the term GLR-CUSUM to describe the CUSUM-type rule whose detection statistic is based on maximizing with respect to the affected subset, which is the unknown parameter in our setup.

We should also mention an alternative formulation of the multisensor sequential change detection problem, in which the change-points in the various sensors are in general different and the goal is to detect the first of them. This problem was considered from a Bayesian point of view by Ragavan and Veeravalli [17] and Ludkovski [18], where a model for change propagation was assumed. A minimax approach was considered by Hadjiliadis et al. [19] and Zhang et al. [20], where it was shown that the multichart CUSUM has a strong asymptotic optimality property with respect to an extended Lorden criterion in which the worst case is taken with respect to all individual change-points and the history of observations up to the minimum of the change-points.

Another closely related problem to the one considered in this paper is the joint change detection and isolation problem, in which one is interested in isolating the affected subset upon detecting the change. While the GLR-CUSUM provides a natural estimator for this subset, a proper treatment of this problem requires a different formulation, in which the misclassification rate is also controlled together with the rate of false alarms. More details on the joint sequential change detection and isolation problem we refer to Nikiforov [21], [22], Oskiper and Poor [23], Tartakovsky [24].
The rest of the paper is organized as follows: in Section II we formulate a general multisensor sequential change detection problem and establish the second-order asymptotic optimality of various CUSUM-based detection rules. In Section III we restrict ourselves to the traditional setup of independent sensors where only a subset of them is affected by the change. In Section IV we focus on the case that the change affects exactly one sensor and the design of the multichart CUSUM. In Section V we present the findings of a simulation study that illustrates our asymptotic results. We conclude in Section VI, and present the proofs of most results in the Appendix.

Finally, in what follows we set \( x^+ = \max(x, 0) \), use \(| \cdot |\) to denote set cardinality, and write \( x \sim y \) when \( x/y \to 1 \), \( x \asymp y \) when \( x - y \to 0 \).

II. A GENERAL MULTISENSOR SEQUENTIAL QUICKSET DETECTION PROBLEM

A. Problem Formulation

Let \( K \) be the number of sensors (streams) and \( X_k^t \) the observation in sensor \( k \) at time \( t \), where \( 1 \leq k \leq K \) and \( t \in \mathbb{N} := \{1, 2, \ldots \} \). We assume that the random vectors \( \{X_t := (X_1^t, \ldots, X_K^t)\}_{t \in \mathbb{N}} \) are independent over time, but we allow observations from different sensors at the same time instant to be dependent. Moreover, we assume that the distribution of the sequence \( \{X_t\}_{t \in \mathbb{N}} \) may change at some unknown, deterministic time \( \nu \) that takes values in \( \{0, 1, \ldots\} \). If the change never occurs, we write \( \nu = \infty \). To be more specific, let \( f(\cdot; \theta) \) be a family of densities with respect to some dominating \( \sigma\)-algebra generated by all observations up to time \( t \), i.e., \( \mathcal{F}_t := \sigma(X_t; 1 \leq s \leq t) \). A sequential change detection rule is an \( \{\mathcal{F}_t\} \)-stopping time, \( T, \) at which the fusion center raises an alarm declaring that the change has occurred. Following Lorden’s approach [8] we measure the performance of a detection rule \( T \) when the change occurs in subset \( A \) with the worst (with respect to the change point) conditional expected delay given the worst possible history of observations up to the change point:

\[
J_A[T] := \sup_{\nu \geq 0} \mathbb{E}_v^A[(T - \nu)^+ | \mathcal{F}_v].
\]

Moreover, we want to control the expected time for a false alarm above a user-specified level \( \gamma \geq 1 \), thus, we restrict ourselves to sequential detection rules in the class \( \mathcal{C}_\gamma = \{T : \mathbb{E}_\infty[T] \geq \gamma\} \). We will be interested in sequential detection rules that attain the optimal performance \( \inf_{T \in \mathcal{C}_\gamma} J_A[T] \) for every \( A \in \mathcal{P} \) asymptotically as \( \gamma \to \infty \).

Definition 1: Suppose that a detection rule \( T^* \) can be designed so that \( T^* \in \mathcal{C}_\gamma \) for any given \( \gamma \geq 1 \). We will say that \( T^* \) is asymptotically optimal of

(i) first-order, if for every \( A \in \mathcal{P} \) we have as \( \gamma \to \infty \)

\[
J_A[T^*] \sim \inf_{T \in \mathcal{C}_\gamma} J_A[T],
\]

(ii) second-order, if for every \( A \in \mathcal{P} \) we have

\[
J_A[T^*] - \inf_{T \in \mathcal{C}_\gamma} J_A[T] = O(1),
\]

where \( O(1) \) is a bounded term as \( \gamma \to \infty \).

Remark: If \( T^* \) is first-order asymptotically optimal, then it is possible that for some \( A \in \mathcal{P} \) we have \( J_A[T^*] - \inf_{T \in \mathcal{C}_\gamma} J_A[T] \to \infty \), since both terms go to infinity as \( \gamma \to \infty \). In such a case, the additional (worst-case) expected number of observations that \( T^* \) requires (relative to the corresponding optimal detection delay that would be achievable if the actual affected subset was known) is unbounded as the rate of false alarms goes to 0. This motivates the stronger notion of second-order asymptotic optimality.

It is the main goal of this work to propose detection rules that are second-order asymptotically optimal. In order to do so, we need to characterize the optimal performance that is attained by the oracle detection rule.

B. The Oracle Rule

For any given subset \( A \), the solution to the constrained optimization problem \( \inf_{T \in \mathcal{C}_\gamma} J_A[T] \) is well known. In order to describe it, let us set

\[
\Lambda_t^A \equiv \frac{d \mathbb{P}_0^A}{d \mathbb{P}_\infty}(\mathcal{F}_t), \quad Z_t^A \equiv \log \Lambda_t^A,
\]

and note that due to the assumption of independence over time we have

\[
\Lambda_t^A = \Lambda_{t-1}^A \exp(f_t^A), \quad Z_t^A = Z_{t-1}^A + f_t^A, \quad Z_0^A := 0,
\]
where $\ell^A_t$ is the log-likelihood ratio

$$\ell^A_t := \log \frac{dP^A_t}{dP^\infty}(X_t). \quad (4)$$

The CUSUM rule for detecting a change in subset $A$ is

$$S^A_b = \inf \{ t \in \mathbb{N} : Y^A_t \geq b \}, \quad (5)$$

where $b \in \mathbb{R}$ is a deterministic threshold and $Y^A_t$ is the CUSUM statistic

$$Y^A_t := Z^A_t - \min_{0 \leq s < t} Z^A_s, \quad t \in \mathbb{N} \quad (6)$$

which may be equivalently defined through the following recursion

$$Y^A_t = \max\{Y^A_{t-1}, 0\} + \ell^A_t, \quad Y^A_0 := 0. \quad (7)$$

Moustakides [7] showed that, for any given $\gamma \geq 1$, $S^A_b$ optimizes Lorden’s criterion, $\mathcal{J}_A$, in the class $\mathcal{C}_\gamma$, when threshold $b$ is chosen so that $E^\infty[S^A_b] = \gamma$. Earlier, Lorden [8] had established the first-order asymptotic optimality of the CUSUM rule by showing that

$$\inf_{T \in \mathcal{C}_\gamma} \mathcal{J}_A[T] \sim \log \frac{\gamma}{I_A} + O(1), \quad I_A := E^A_0[\ell^A_1]. \quad (8)$$

under the assumption that $\ell^A_1$ is $P^A_0$, $P^\infty$- integrable. The exact optimality of the CUSUM rule can be used to obtain a characterization of the optimal performance up to a second-order asymptotic approximation. Indeed, if additionally we have $E^A_0[(\ell^A_1)^2] < \infty$, it is well known (see, e.g., [9], [25]) that

$$E^A_0[S^A_b] = \frac{b}{I_A} + O(1),$$

where $O(1)$ is a term that is bounded from above and below as $b \to \infty$. Moreover, from [26] it follows that $E^\infty[S^A_b] \sim e^{b/v_A}$ as $b \to \infty$, where $v_A$ is a model-dependent, renewal-theoretic constant. Therefore, we conclude that

$$\inf_{T \in \mathcal{C}_\gamma} \mathcal{J}_A[T] \sim \log \frac{\gamma}{I_A} + O(1).$$

We summarize these characterizations of first and second order asymptotic optimality in the following Lemma.

**Lemma 1:**

(i) Suppose that $\ell^A_1$ is integrable with respect to $P^\infty$ and $P^A_0$ for every $A \in \mathcal{P}$ and that, for any given $\gamma \geq 1$, a detection rule $T^*$ can be designed so that $T^* \in \mathcal{C}_\gamma$. If

$$\mathcal{J}_A[T^*] \leq \log \frac{\gamma}{I_A} (1 + o(1)) \quad \forall \ A \in \mathcal{P},$$

then $T^*$ is first-order asymptotically optimal with respect to class $\mathcal{P}$.

(ii) Suppose additionally that $\ell^A_1$ is square-integrable with respect to $P^A_0$ for every $A \in \mathcal{P}$. If

$$\mathcal{J}_A[T^*] \leq \log \frac{\gamma}{I_A} + O(1) \quad \forall \ A \in \mathcal{P},$$

then $T^*$ is second-order asymptotically optimal with respect to class $\mathcal{P}$.

**Remark:** It is useful to stress that the CUSUM statistic is often defined in the literature as

$$Y^A_t := Z^A_t - \min_{0 \leq s \leq t} Z^A_s, \quad (9)$$

or equivalently through the following recursion

$$Y^A_t = \max\{Y^A_{t-1} + \ell^A_t, 0\}, \quad Y^A_0 := 0. \quad (10)$$

Thus, $Y^A_t$ differs from $\bar{Y}^A_t$ only when the latter is negative, and the two statistics lead to the same stopping time when the threshold $b$ is positive. However, with the non-negative version of the CUSUM statistic it is not in general possible to satisfy the false alarm constraint for any $\gamma > 1$. Since in this work we focus on asymptotic optimality results, we will use both versions of the CUSUM statistic, depending on which one is more technically convenient each time.

**C. Generalized and Mixture-Based CUSUM Rules**

We now construct various multichannel detection rules that will be shown to be second-order asymptotically optimal with respect to any given class $\mathcal{P}$. In order to do so, suppose for the moment that the affected subset is $A \in \mathcal{P}$. At some time $t > v$, the likelihood ratio of $P^A_v$ versus $P^\infty$ is given by

$$\frac{\Lambda^A_v}{\Lambda_v} = \frac{dP^A_v}{dP^\infty}(\mathcal{F}_t) = \prod_{s=v+1}^{t} \exp(\ell^A_s),$$

and the corresponding log-likelihood ratio takes the form

$$Z^A_t - Z^A_v = \log \frac{\Lambda^A_v}{\Lambda_v} = \sum_{s=v+1}^{t} \ell^A_s,$$

where $\Lambda^A_v$, $Z^A_v$, and $\ell^A$ are defined in (3) and (4), respectively. Maximizing with respect to both the unknown change-point, $v$, and the unknown subset, $A$, suggests raising an alarm when the statistic

$$\max_{0 \leq s \leq t} Z^A_t - Z^A_s \quad \text{becomes larger than some threshold } b. \quad \text{Interchanging the two maximizations reveals that this detection statistic can be equivalently expressed as}$$

$$\max_{A \in \mathcal{P}} \max_{0 \leq s < t} Z^A_t - Z^A_s = \max_{A \in \mathcal{P}} \bar{Y}^A_t,$$

where $\bar{Y}^A$ is defined in (6). Therefore, this rule stops the first time $t$ there is a subset $A$ whose corresponding CUSUM statistic $\bar{Y}^A_t$ exceeds threshold $b$. If we only consider positive thresholds, we may equivalently consider the detection statistic $\max_{A \in \mathcal{P}} Y^A_t$, where $Y^A_t$ is defined in (9) (see the remark at the end of the previous subsection).

We can generalize this scheme by allowing the thresholds that correspond to the various CUSUM statistics to differ. Thus, if $\{b_A, A \in \mathcal{P}\}$ is a family of positive, constant thresholds, we obtain the following detection rule, to which we refer as GLR-CUSUM:

$$\min_{A \in \mathcal{P}} S^A_b = \inf \left\{ t \in \mathbb{N} : \max_{A \in \mathcal{P}} (Y^A_t - b_A) \geq 0 \right\}. \quad (10)$$
We will show in Proposition 3 that some intuitively reasonable threshold specifications, such as selecting each threshold \(b_A\) proportionally to the Kullback-Leibler divergence \(I_A\), may fail to guarantee even the first-order asymptotic optimality of GLR-CUSUM. For this reason, in what follows we restrict our attention to thresholds of the form

\[
b_A = b - \log p_A,
\]

where \(b\) is a positive threshold that is determined by the false alarm constraint and depends on \(\gamma\), whereas \(\{p_A, A \in \mathcal{P}\}\) are constants (weights) that do not depend on \(\gamma\) and satisfy

\[
p_A > 0 \quad \forall A \in \mathcal{P} \quad \text{and} \quad \sum_{A \in \mathcal{P}} p_A = 1.
\]

With this threshold specification the GLR-CUSUM in (10) takes the form

\[
S_b := \inf \left\{ t : \max_{A \in \mathcal{P}} (Y_t^A + \log p_A) \geq b \right\}
\]

and may be equivalently expressed as follows:

\[
S_b = \inf \left\{ t : \max_{0 \leq s \leq t} \left( Z_t^A - Z_s^A + \log p_A \right) \geq b \right\}
\]

When the \(p_A\)'s are identical, we will refer to \(S_b\) as the unweighted GLR-CUSUM. However, the introduction of weights allows us to present the GLR-CUSUM together with competitive mixture-based CUSUM tests. Specifically, if we do not maximize but average with respect to the affected subset in (12), we obtain the following stopping time:

\[
\tilde{S}_b := \inf \left\{ t : \max_{0 \leq s \leq t} \sum_{A \in \mathcal{P}} p_A \exp (Z_t^A - Z_s^A) \geq e^b \right\}.
\]

If we further interchange maximization and summation in the detection statistic in (13), we obtain the alternative detection statistic \(\sum_{A \in \mathcal{P}} p_A \exp (Y_t^A)\). For technical convenience, we replace \(Y^A\) with \(\tilde{Y}^A\) and set

\[
\tilde{S}_b := \inf \left\{ t : \max_{0 \leq s \leq t} \sum_{A \in \mathcal{P}} p_A \exp (\tilde{Y}_t^A) \geq e^b \right\}.
\]

To sum up, we have introduced three distinct multichannel detection rules: the GLR-CUSUM, \(S_b\), which can be equivalently defined by either (11) or (12), and two mixture-based CUSUM rules, \(S_b\) and \(\tilde{S}_b\), defined in (13) and (14), respectively. In the following theorem, whose proof can be found in the Appendix, we establish their second-order asymptotic optimality property with respect to an arbitrary class \(\mathcal{P}\).

**Theorem 1:** For any given class \(\mathcal{P}\) and weights \(\{p_A, A \in \mathcal{P}\}\),

(i) if \(b = \log \gamma\), then \(S_b, \tilde{S}_b, \bar{S}_b \in C_\gamma\);

(ii) if \(\ell^A\) is square-integrable with respect to \(\mathcal{P}_A^0\) for every \(A \in \mathcal{P}\), then \(S, \tilde{S}, \bar{S}\) are second-order asymptotically optimal with respect to \(\mathcal{P}\).

Although it is not needed in the proof of the previous theorem, it is important (and useful for computational purposes) to note that the worst-case scenario for all the above detection rules is when \(\nu = 0\). This is the content of the following proposition, whose proof can be found in the Appendix.

**Proposition 1:** For any given class \(\mathcal{P}\), threshold \(b\) and weights \(\{p_A, A \in \mathcal{P}\}\), we have (i) \(J_A[S_b] = E_0^A[S_b]\); (ii) \(J_A[\tilde{S}_b] = E_0^A[\tilde{S}_b]\); (iii) \(J_A[S_b] = E_0^A[S_b]\) for every \(A \in \mathcal{P}\).

**D. Implementation**

The two representations of the GLR-CUSUM, (11) and (12), suggest two possible approaches regarding its implementation. According to the first one, at each time \(t\) we compute and maximize over the CUSUM statistics, \(Y_t^A, A \in \mathcal{P}\). This approach requires running \(|\mathcal{P}|\) recursions and is convenient when the cardinality of \(\mathcal{P}\) is small. This is for example the case when only one component of the parameter vector \(\theta\) is affected by the change (\(\mathcal{P} = \{1\}\)), in which case \(|\mathcal{P}| = N\). However, such an implementation may not be feasible for large values of \(N\) when we have complete ignorance regarding the affected subset (\(\mathcal{P} = \mathcal{P}_N\)). This approach also applies to the mixture-based CUSUM rule defined in (14).

According to the second approach, at each time \(t\) we need to compute and maximize over \(Z_t^A - Z_s^A + \log p_A, 0 \leq s < t, A \in \mathcal{P}\). In the next section we will see that this task is simplified in the special case that we have independent sensors and only a subset of them is affected by the change. Moreover, in the next lemma we show that, without any loss, we can always restrict the maximization to an adaptive time-window. Since a similar approach applies to the mixture-based CUSUM defined in (13), we state the following lemma in some generality.

**Lemma 2:** Let \(\mathcal{P}\) be an arbitrary class and consider the following sequence of stopping times:

\[
r_{n+1} = \inf \left\{ t > r_n : Z_t^A = \min_{0 \leq s \leq t} Z_s^A \forall A \in \mathcal{P} \right\};
\]

\[
r_0 = 0.
\]

Then for any function \(G: \mathbb{R}^{\mathcal{P}} \rightarrow \mathbb{R}\) that is non-decreasing in each of its arguments,

\[
\inf \left\{ t : \max_{0 \leq s \leq t} G \left( \left( Z_t^A - Z_s^A \right)_{A \in \mathcal{P}} \right) \geq b \right\}
\]

\[
= \inf \left\{ t : \max_{r(t) \leq s \leq t} G \left( \left( Z_t^A - Z_s^A \right)_{A \in \mathcal{P}} \right) \geq b \right\},
\]

where \((\cdot)_{A \in \mathcal{P}}\) denotes a \(\mathcal{P}\)-tuple and \(r(t) = \max \{r_n : r_n \leq t\}\).

**Proof:** Fix \(t > 0\). Since \(G\) is non-decreasing in each of its arguments, it suffices to show that for every \(0 \leq s \leq r(t)\) and \(A \in \mathcal{P}\) we have \(Z_t^A - Z_s^A \leq Z_t^A - Z_r(t)\), or equivalently \(Z_r(t) \leq Z_s^A\), which follows from the definition of \(r(t)\). □

III. A SPECIAL MULTISENSOR SEQUENTIAL CHANGE-DETECTION PROBLEM

A. A Special Case of the General Framework

In this section, we focus on the special case that the change affects the distribution of only a subset of sensors. In order to be more specific, let us recall that at each time \(t\) we observe...
a random vector $X_t = (X_t^1, \ldots, X_t^K)$ with joint density $X_t \sim f(\cdot|\theta_t)$, where $\theta_t = (\theta_t^1, \ldots, \theta_t^K)$ is an unobserved parameter vector. Let $f_k$ be the marginal density of the $k$th stream, i.e., $X_t^k \sim f_k(\cdot|\theta_t)$. In what follows, we assume that the dimensions of $X_t$ and $\theta_t$ coincide, i.e., $N = K$, and that each marginal density $f_k(\cdot|\theta_t)$ is determined only by the $k$th component of $\theta_t$, $\theta_t^k$. Specifically, we assume that

$$f_k(\cdot|\theta_t^k = 0) = h_k, \quad f_k(\cdot|\theta_t^k = 1) = g_k,$$

where $h_k$ and $g_k$ are completely specified densities with respect to some $\sigma$-finite measure $\lambda_k$, such that their Kullback-Leibler information number,

$$I_k := \int \log\left(\frac{g_k}{h_k}\right) g_k d\lambda_k,$$

is positive and finite. Note that $h_k$ and $g_k$ do not need to belong to the same parametric family. In this setup we have a change in the marginal distributions of a subset of sensors, while the remaining ones are completely unaffected, and the change detection problem (1) takes the form

$$X_t^k \sim h_k \quad 1 \leq k \leq K, \quad t \leq v,$$

$$X_t^k \sim \begin{cases} g_k, & k \in \mathcal{A}, \\ h_k, & k \notin \mathcal{A}, \end{cases} \quad t > v,$$

where $\mathcal{A} \subset \{1, \ldots, K\}$ is now a subset of sensors that belongs to class $\mathcal{P}$. When in particular $\mathcal{P} = \mathcal{P}_L$ (resp. $\mathcal{P} = \mathcal{P}_L$), the change affects exactly (resp. at most) $L$ sensors, where the classes $\mathcal{P}_L$ and $\mathcal{P}_L$ are defined in (2).

In the literature of the multisensor quickest detection problem it is typically assumed that observations from different streams are independent. However, this is not necessary for the results of the previous section to hold. This is illustrated by the following example.

**Example: Correlated normal streams.**

Let $\Sigma$ be an invertible covariance matrix of dimension $K$ with diagonal $(\sigma_1^2, \ldots, \sigma_K^2)$ and let $\mu_1, \ldots, \mu_K$ be non-zero constants. For every non-empty subset $A \subset \{1, \ldots, K\}$ we define the $K$-dimensional vector $\mu_A := (\mu_A^1, \ldots, \mu_A^K)$ such that

$$\mu_A^k := \begin{cases} \mu_k, & k \in A, \\ 0, & k \notin A, \end{cases}$$

and suppose that under $\mathcal{P}_v^A$ we have

$$X_t \sim \begin{cases} \mathcal{N}(0, \Sigma), & t \leq v, \\ \mathcal{N}(\mu_A, \Sigma), & t > v. \end{cases}$$

This is a special case of (17) with $h_k = \mathcal{N}(0, \sigma_k^2)$ and $g_k = \mathcal{N}(\mu_k, \sigma_k^2)$, for which the log-likelihood ratio $\ell_t^A$ in (4) takes the form:

$$\ell_t^A = \theta_A \cdot X_t - \frac{1}{2} \theta_A \cdot \mu_A, \quad \text{where} \quad \theta_A := \Sigma^{-1} \mu_A.$$

### B. The Case of Independent Sensors

We now restrict ourselves to the case of independent sensors. Specifically, we assume that the local filtrations $\{\mathcal{F}_t^1\}, \ldots, \{\mathcal{F}_t^K\}$ are independent, where $\mathcal{F}_t^k := \sigma(X_t^k, 1 \leq s \leq t), 1 \leq k \leq K$. Under this assumption, the log-likelihood ratio statistic $\ell_t^A$ in (4) admits the following decomposition

$$\ell_t^A = \sum_{k \in \mathcal{A}} \ell_t^k, \quad \ell_t^k := \log \frac{g_k}{h_k}(X_t^k),$$

which implies that the likelihood and log-likelihood ratio statistics, $\Lambda_t^A$ and $Z_t^A$, defined in (3), take the form

$$\Lambda_t^A = \prod_{k \in \mathcal{A}} \Lambda_t^k, \quad \Lambda_t^k := \Lambda_t^k \exp(\ell_t^k), \quad \Lambda_0^k = 1,$$

$$Z_t^A = \sum_{k \in \mathcal{A}} Z_t^k, \quad Z_t^k := Z_{t-1}^k + \ell_t^k, \quad Z_0^k = 0.$$

Moreover, we have $I_A = \sum_{k \in \mathcal{A}} I_k$, where $I_A$ is the Kullback-Leibler information number defined in (8) and $I_k$ is the local Kullback-Leibler information number defined in (16). If we further select weights of the form

$$p_A := \frac{p^{[A]}}{\sum_{B \subseteq \mathcal{P}} p^{[B]}},$$

where $p$ is some arbitrary positive parameter, then from (12) and (13) it follows that the GLR-CUSUM and the mixture-based CUSUM rule, $\tilde{S}_p$, can be expressed as follows:

$$S_p := \inf \left\{ t : \max_{0 \leq s \leq t} \sum_{k \in \mathcal{A}} \left( Z_t^k - Z_s^k + \log p \right) \right\},$$

$$\tilde{S}_p := \inf \left\{ t : \max_{0 \leq s \leq t} \sum_{A \subseteq \mathcal{P}} p^{[A]} \prod_{k \in A} \exp(Z_t^k - Z_s^k) \right\},$$

where $C(\mathcal{P}) := \sum_{B \subseteq \mathcal{P}} p^{[B]}$. In the following proposition, whose proof is presented in the Appendix, we show that the latter expressions can be further simplified for classes of the form $\mathcal{P}_L$ and $\mathcal{P}_L$. In order to do so, for any $0 \leq s \leq t$ we set

$$Z^{(s,t)} := Z_t^k - Z_s^k, \quad 1 \leq k \leq K,$$

and we introduce the corresponding order statistics $Z^{(s,t)}_1 \geq \ldots \geq Z^{(s,t)}_K$.

**Proposition 2:** Consider the change detection problem (17) and suppose that the independence assumption (18) holds.

(i) The GLR-CUSUM in (20) takes the form

$$\inf \left\{ t : \max_{0 \leq s \leq t} \sum_{k=1}^L Z^{(s,t)}_k \geq b + \log |\mathcal{P}_L| \right\} \text{ when } \mathcal{P} = \mathcal{P}_L,$$

(ii) The mixture-based CUSUM in (21) takes the form

$$\inf \left\{ t : \max_{0 \leq s \leq t} \left( \sum_{k=1}^L Z^{(s,t)}_k + \log p \right)^+ \geq b + \log \left( \sum_{k=1}^L |\mathcal{P}_k| p^k \right) \right\} \text{ when } \mathcal{P} = \mathcal{P}_L.$$
(ii) When $\mathcal{P} = \mathcal{P}_K$, the mixture rule $\hat{S}_b$ in (21) takes the form

$$
\hat{S}_b(\pi) := \inf \left\{ t : \max_{0 \leq s \leq t} \prod_{k=1}^{K} \left[ 1 - \pi + \pi \exp(Z^k_t - Z^k_s) \right] \right\}
$$

$$
\geq \exp(\hat{b})
$$

where $\pi = p/(1 + p)$ and

$$
\exp(\hat{b}) := \left( 1 - \pi \right)^{-K} \left( \exp(b) \sum_{B \in \mathcal{P}} p^{|B|} + 1 \right).
$$

Remark: From Lemma 2 we know that we can replace $\max_{0 \leq s \leq t}$ in the above detection rules by $\max_{0 \leq s \leq t}$, where $r(t) = \max[r_n : r_n \leq t]$ and the sequence of stopping times $(r_n)$ is defined in (15). In the case of independent sensors that we consider in this section we have $\gamma^k_A \leq \sum_{k \in A} t^k$, for every $t$ and $A$, which implies that for a class of the form $\mathcal{P}_L$ the times $(r_n)$ can be defined as follows:

$$
r_n = \inf \left\{ t > r_{n-1} : Z^k_t = \min_{0 \leq s \leq t} Z^k_s \quad \forall 1 \leq k \leq K \right\},
$$

where $r_0 := 0$. From [10, Proposition 1] it follows that $(r_n - r_{n-1})_{n \in \mathbb{N}}$ are iid random variables with finite expectation under $\mathbb{P}_\infty$, which however grows exponentially in $K$. This implies that the above window may not be very useful for computational purposes when $K$ is large. A more appropriate approach for practical implementation in this case is to use the window $\sigma(t) := \max[\sigma_n : \sigma_n \leq t]$, where $\sigma_0 := 0$ and

$$
\sigma_n := \inf \left\{ t > \sigma_{n-1} : Z^k_t < Z^k_{\sigma_{n-1}} \quad \forall 1 \leq k \leq K \right\}.
$$

Again, $(\sigma_n - \sigma_{n-1})_{n \in \mathbb{N}}$ is a sequence of iid random variables with finite expectation under $\mathbb{P}_\infty$, which however seems to grow logarithmically in $K$ (based on empirical observations). However, we should emphasize that the resulting detection rules are not equivalent to the original ones. For similar adaptive window-limited modifications of CUSUM-type rules we refer to Yashchin [27]–[29].

Remark: When the weights are selected according to (19), all subsets of the same size have the same weight. It is straightforward to generalize the previous results when $p_A$ is proportional to $\prod_{k \in A} p_k$, where $p_k$, $1 \leq k \leq K$ are arbitrary positive parameters. However, setting $p_k = p$ offers an intuitive way of selecting the parameter $\pi$, or equivalently $p$, in the mixture-based CUSUM, $\hat{S}(\pi)$, in (24). Indeed, $\hat{S}(\pi)$ can be obtained by repeated application of the one-sided sequential test

$$
\hat{T}_b(\pi) := \inf \left\{ t \in \mathbb{N} : \hat{\Lambda}_t \geq e^b \right\},
$$

where

$$
\hat{\Lambda}_t := \prod_{k=1}^{K} \left( 1 - \pi + \pi \Lambda^k_t \right).
$$

This is the one-sided SPRT for testing $\mathbb{P}_\infty$ against the auxiliary probability measure

$$
\hat{P} := \bigotimes_{k=1}^{K} \left( \left( 1 - \pi \right) \mathbb{P}^k_\infty + \pi \mathbb{P}^k_0 \right),
$$

according to which the density in sensor $k$ is $g_k$ with probability $\pi$ and $h_k$ with probability $1 - \pi$. This implies that the parameter $\pi$ in $\hat{S}(\pi)$ can be interpreted as the proportion of affected sensors and suggests setting $\pi = L/K$ when we know in advance that exactly $L$ sensors are affected ($\mathcal{P} = \mathcal{P}_L$). On the other hand, it is known that at most $L$ sensors are affected ($\mathcal{P} = \mathcal{P}_L$), a reasonable default choice seems to be $\pi = L/(2K)$.

Remark: It is interesting to compare the detection rule $\hat{S}_b$, defined in (24), with the detection rule

$$
\hat{S}_b(\pi) := \inf \left\{ t \in \mathbb{N} : \max_{0 \leq s \leq t} \sum_{k=1}^{K} \log \left( 1 - \pi + \pi \exp ((Z^k_t - Z^k_s)^+) \right) \geq b \right\},
$$

which was proposed in [14]. In the following theorem, whose proof is presented in the Appendix, we show that the latter is also second-order asymptotically optimal with respect to $\mathbb{P}_K$ for any choice of $\pi$ in $(0, 1)$, a result that explains the (empirically observed in [14]) robustness of this rule with respect to $\pi$.

Theorem 2: Consider the change detection problem (17) and suppose that the independence assumption (18) holds.

Consider the detection rules $\hat{S}_b(\pi)$ with some $\pi \in (0, 1)$ and $\hat{S}_b(\pi)$ with some $\pi \in (0, 1)$, defined in (24) and (26), respectively. Then,

(i) $\hat{S}_b(\pi) \in \mathcal{C}_T$ if $b = \log \gamma$ and $\hat{S}_b(\pi) \in \mathcal{C}_T$ if $b = \log \gamma + \log |\mathcal{P}_K|$;

(ii) if also $\ell^A_1$ is square-integrable under $\mathbb{P}^0_A$ for every $A \in \mathcal{P}_K$, both $\hat{S}_b(\pi)$ and $\hat{S}_b(\pi)$ are second-order asymptotically optimal with respect to $\mathcal{P}_K$.

Remark: We stress that the parameter $\pi$ in the detection rule proposed in [14], $\hat{S}_b(\pi)$, cannot be interpreted as the proportion of affected sensors. Indeed, if we set $\pi = 1$ in (26) we recover the unweighted GLR-CUSUM in the case of complete uncertainty, that is the detection rule obtained by setting $L = K$ and $p = 1$ in (23). On the other hand, if we set $\pi = 1$ in $\hat{S}_b(\pi)$, defined in (24), we recover the optimal CUSUM rule in the case that all sensors are affected, which is exactly what one would expect if $\pi$ is to be interpreted as the proportion of affected sensors.

C. Scalable Schemes

We close this section by highlighting the connection between the GLR-CUSUM and the SUM-CUSUM proposed by Mei in [10]. In order to do so, we first observe that the detection statistic of the unweighted GLR-CUSUM in the case of complete ignorance, which can be obtained by setting $L = K$ and $p = 1$ in (23), is

$$
\max_{0 \leq s \leq t} \sum_{k=1}^{K} \left( Z^k_t - Z^k_s \right)^+.
$$

If we interchange max and sum in this statistic, we obtain the sum of the local CUSUM statistics, $\sum_{k=1}^{K} Y^k_t$. This is the detection statistic of SUM-CUSUM, which was shown in [10]
to be first-order asymptotically optimal with respect to $\mathcal{P}^*_k$, i.e., in the case of complete ignorance. The main advantage of this rule is that it is much easier to implement than the corresponding GLR-CUSUM and mixture-based CUSUM rules, as it requires running only $K$ recursions. However, it is reasonable to expect that SUM-CUSUM should be less efficient, since it is only first-order asymptotically optimal. This intuition will be corroborated by the results of a simulation study in Section V.

Remark: If we interchange, in a similar way, max and product in the mixture-based CUSUM rule (24), we obtain

$$M_b(\pi) := \inf \left\{ t : \sum_{k=1}^{K} \log \left( 1 - \pi + \pi \exp(Y_{1k}^k) \right) \geq b \right\}.$$  

This is another scalable rule comparable with SUM-CUSUM, however the parameter $\pi$ can no longer be interpreted as proportion of affected sensors.

Remark: From Mei [11] it follows that the detection rule

$$M_b(L) = \inf \left\{ t : \sum_{k=1}^{L} Y_{1k}^{(k)} \geq b \right\},$$

where $Y_{1k}^{(1)} \geq Y_{1k}^{(2)} \geq \cdots \geq Y_{1k}^{(K)}$, is first-order asymptotically optimal with respect to $\mathcal{P}^*_L$ (at most $L$ sensors affected). This rule should be compared with the corresponding GLR-CUSUM and mixture-based CUSUM that correspond to class $\mathcal{P}^*_L$.

### IV. Multichart CUSUM

In this section we still consider the change-detection problem (17), where the change affects only a subset of sensors, and assume that the sensors are independent, i.e., (18) holds. Our focus will now be on the GLR-CUSUM, defined in (10), in the special case that the change can affect exactly one sensor ($\mathcal{P} = \mathcal{P}_1$), where it takes the form

$$S = \min_{1 \leq k \leq K} c_k^* b_k,$$

(28)

each $b_k$ being a positive threshold. The detection rule (28) is also known as multichart CUSUM in the literature of statistical quality control and has been well studied (see, e.g., [1], [2]), especially in the case that all thresholds are equal. However, unless the sensors are homogeneous in the sense that $b_k = h$ and $g_k = g$, it is intuitively clear that $\mathcal{P}$ should be preferable to have unequal thresholds. The results of Section II imply that there is a large family of non-identical thresholds for which the multichart CUSUM is second-order asymptotically optimal. Specifically, if we set

$$b_k = b - \log p_k, \quad 1 \leq k \leq K,$$

(29)

where $b$ is determined by the false alarm constraint and the $p_k$’s are arbitrary positive constants that add up to 1 and do not depend on $\gamma$, then from Theorem 1 it follows that the resulting detection rule is second-order asymptotically optimal for any selection of the $p_k$’s. Our first goal in this section is to show that there are reasonable, alternative threshold specifications with which the multichart CUSUM may fail to be even first-order asymptotically optimal. The second goal is to compare various choices for the $p_k$’s when the thresholds are selected according to (29).

In order to establish these results, we will rely on some well known facts about the performance characteristics of the multichart CUSUM. Thus, let $T_k^b = \inf \{ t \geq 0 : Z_{1k}^k \geq b \}$ and let $\eta_k^b$ be the corresponding overshoot, i.e., $\eta_k^b := Z_{1k}^k - b$.

Since $Z^k$ is a random walk with a finite second moment, the limiting expected overshoot is well defined and we set

$$\rho_k := \lim_{b \to \infty} E_0[\eta_k^b].$$

Moreover, we have the following representation (see [25, p. 32] or [9, p. 44]) for the expected infimum of $Z^k$ under $\mathbb{P}^k_0$,

$$\beta_k := E_0^k \left[ \inf_{t \geq 0} Z_t^k \right] = \rho_k - E_0^k \left[ (Z_1^k)^2 \right] / 2.$$  

Finally, we introduce the Laplace transform of the limiting distribution of the overshoot under $\mathbb{P}^k_0$:

$$\delta_k := \lim_{b \to \infty} E_0^k \left[ \exp(-\eta_k^b) \right].$$

Then, it is well known (see, e.g., [9, p. 471]) that as $b_k \to \infty$

$$\mathcal{J}_k[S] = E_0^k[S] \simeq \frac{b_k + \rho_k + \beta_k}{l_k},$$

(30)

where $x \simeq y$ means $x - y = o(1)$. On the other hand, under $\mathbb{P}_\infty$ we have (see, e.g., [26], [9, p. 467])

$$E_\infty[S] \sim \frac{1}{\sum_{k=1}^{K} e^{-b_k} l_k \delta_k^2}.$$  

(31)

From (30) it is clear that if the $b_k$’s are chosen proportionally to the Kullback-Leibler information numbers, i.e., $b_k \propto l_k$, then the first-order asymptotic performance of the multichart CUSUM under the various scenarios is equalized, i.e., $\mathcal{J}_k[S] \sim \mathcal{J}_m[S]$ for every $1 \leq k \neq m \leq K$. However, as we show in the following proposition, with this threshold specification the multichart CUSUM loses even its first-order asymptotic optimality property with respect to $\mathcal{P}_1$, unless the $l_k$’s are identical.

**Proposition 3:** Suppose that $b_k = c_k l_k$ for every $1 \leq k \leq K$, where $c_k$ is a constant that does not depend on $k$ and is chosen so that the false alarm constraint is satisfied with equality, i.e., $E_\infty[S] = \gamma$. Then, for every $1 \leq k \leq K$ we have:

$$\lim_{\gamma \to \infty} \inf_{T \in \mathcal{C}_k} \mathcal{J}_k[T] = \min_{1 \leq j \leq K} I_j.$$  

**Proof:** Setting $b_k = c_k l_k$ in (30) and (31) we obtain

$$\mathcal{J}_k[S] = c_k + O(1),$$

where $O(1)$ is an asymptotically bounded term as $\gamma \to \infty$, and

$$\gamma \sim E_\infty[S] \sim \frac{1}{\sum_{k=1}^{K} e^{-c_k l_k} l_k \delta_k^2}.$$  

(32)

Let $I^* := \min_{1 \leq j \leq K} I_j$. Then (32) becomes

$$\gamma \sim \frac{c_k^*}{\sum_{k:h_k = l_k} l_k \delta_k^2 + \sum_{k:h_k > l_k} e^{-c_k(l_k - l_k)} l_k \delta_k^2},$$

which implies that $c_k = (\log \gamma) / l_k + O(1)$ and completes the proof. □
Fig. 1. The horizontal axis in both graphs represents $\theta_2$, the post-change mean in the second sensor when it is the one affected by the change. The post-change mean in the first sensor, $\theta_1$, when this is the one affected by the change, is fixed to 1. The vertical axis in (a) (resp. (b)) represents the asymptotic relative performance loss of the multichart CUSUM when the first (resp. second) sensor is affected.

### Table I

| $|\mathcal{A}|$ | $b$ | $E_{\infty}(T)$ | $E_{\infty}^A(T)$ | SE |
| :---: | :---: | :---: | :---: | :---: |
| $S_k^A$ | 2 | 9.88 | 100090 | 450 | 10.64 | 0.02 |
| $M_k(5)$ | 2 | 17.1 | 100010 | 455 | 15.30 | 0.03 |
| $M_k(2)$ | 2 | 14.2 | 100065 | 450 | 14.21 | 0.03 |
| $S_k(P_5)$ | 2 | 9.85 | 100065 | 450 | 13.47 | 0.03 |
| $S_k(P_3)$ | 2 | 9.35 | 100105 | 450 | 13.57 | 0.03 |
| $S_k(P_2)$ | 2 | 9.91 | 100105 | 450 | 13.45 | 0.03 |
| $S_k(P_2)$ | 2 | 9.86 | 100115 | 465 | 13.12 | 0.03 |
| $S_k(P_2)$ | 2 | 9.58 | 100005 | 445 | 13.38 | 0.03 |
| $S_k(P_2)$ | 2 | 9.78 | 100005 | 460 | 13.15 | 0.03 |
| $S_k^A$ | 3 | 9.94 | 100065 | 450 | 7.369 | 0.02 |
| $M_k(5)$ | 3 | 17.1 | 100010 | 455 | 10.59 | 0.02 |
| $M_k(3)$ | 3 | 15.9 | 100025 | 450 | 10.44 | 0.02 |
| $S_k(P_5)$ | 3 | 9.85 | 100065 | 450 | 9.04 | 0.02 |
| $S_k(P_3)$ | 3 | 9.63 | 100075 | 460 | 9.458 | 0.02 |
| $S_k(P_3)$ | 3 | 9.91 | 100105 | 450 | 9.054 | 0.02 |
| $S_k(P_3)$ | 3 | 9.90 | 100115 | 450 | 9.098 | 0.02 |
| $S_k(P_3)$ | 3 | 9.58 | 100005 | 445 | 9.136 | 0.02 |
| $S_k(P_3)$ | 3 | 9.67 | 100060 | 450 | 9.150 | 0.02 |
| $S_k^A$ | 4 | 9.93 | 100010 | 450 | 5.716 | 0.02 |
| $M_k(5)$ | 4 | 17.1 | 100010 | 455 | 8.197 | 0.02 |
| $M_k(4)$ | 4 | 16.8 | 100050 | 450 | 8.192 | 0.02 |
| $S_k(P_5)$ | 4 | 9.85 | 100065 | 450 | 6.821 | 0.02 |
| $S_k(P_4)$ | 4 | 9.75 | 100050 | 450 | 7.068 | 0.02 |
| $S_k(P_3)$ | 4 | 9.91 | 100105 | 450 | 6.826 | 0.02 |
| $S_k(P_2)$ | 4 | 9.91 | 100070 | 450 | 6.870 | 0.02 |
| $S_k(P_2)$ | 4 | 9.58 | 100005 | 445 | 6.977 | 0.02 |
| $S_k(P_2)$ | 4 | 9.60 | 100010 | 450 | 7.006 | 0.02 |

Let us now focus on specification (29) for the thresholds. While we have seen in Theorem 1 that setting $b = \log \gamma$ guarantees $S \in C_\gamma$, it is clear that this choice can be very conservative, as it does not take into account the particular pre/post-change distributions. The asymptotic expression (31) can be used to provide a much more accurate approximation for $E_{\infty}[S]$, even though it may not guarantee that $E_{\infty}[S] \geq \gamma$. Specifically, from (30) and (31) it follows that if we select threshold $b$ in (29) as

$$b = \log \gamma + \log \left( \sum_{k=1}^{K} p_k I_k \delta^2_k \right),$$

then $E_{\infty}[S] \sim \gamma$ and

$$\frac{\log \gamma + \log \left( \sum_{j=1}^{K} p_j I_j \delta^2_j \right) - \log p_k + \rho_k + \beta_k}{I_k}.\,$$

We can then use this high-order approximation in order to obtain an intuitive selection of the $p_k$’s. Indeed, such a selection can be obtained if we equalize, up to a first order, the asymptotic relative performance loss under each scenario. To be more specific, let us set

$$\tilde{J}_k[S] := \frac{\tilde{J}_k[S] - \tilde{J}_k[S^k]}{\tilde{J}_k[S^k]}, \quad 1 \leq k \leq K,$$

where $S^k$ is the optimal CUSUM test for detecting a change in sensor $k$. Assuming that the thresholds for the multichart CUSUM are selected according to (29) and (33) and that $S^k$ is also designed such that $E_\infty[S^k] \sim \gamma$, from (34) we obtain

$$\tilde{J}_k[S] \sim \frac{C_k(p)}{\log \gamma + \rho_k + \beta_k + \log(I_k \delta^2_k)},$$

and consequently

$$\tilde{J}_k[S] \sim \frac{C_k(p)}{\log \gamma},$$

where $C_k(p)$ is defined as follows:

$$C_k(p) := \log \left( \frac{\sum_{j=1}^{K} p_j I_j \delta^2_j}{p_k I_k \delta^2_k} \right).$$

This reveals that $\tilde{J}_k[S] \sim \tilde{J}_m[S]$ for every $1 \leq k \neq m \leq K$ when

$$p_k \propto \left( I_k \delta^2_k \right)^{-1}.$$
In fact, the latter specification was proposed in [24, p. 223] on the basis of the observation that it guarantees 
\[ E_\infty [S^k_{b_i}] \sim E_\infty [S^m_{b_i}], \] for every \( 1 \leq k \neq m \leq K \).

An alternative specification for the \( p_k \)'s has been proposed in [9, p. 467], according to which thresholds are selected so that 
\[ I_k \tilde{J}_k[S] \sim I_m \tilde{J}_m[S] \] for every \( 1 \leq k \neq m \leq K \).

From (34) it is clear that this is the case when 
\[ p_k \propto e^{\beta_k}. \]

Finally, if we want to have larger thresholds for sensors with larger Kullback-Leibler information numbers, which is the underlying logic in the threshold specification of Proposition 3, we can set 
\[ p_k \propto 1/I_k. \]

In the remainder of this section we compare the effect of the above specifications on the asymptotic relative performance loss of a multichart CUSUM whose thresholds are selected according to (29) and (33). Specifically, we assume that 
\( h_k = N(0, 1) \) and 
\( g_k = N(\theta_k, 1), \) where \( \theta_k \neq 0. \) We let \( K = 2, \gamma = 10^3 \) and we compute \( \tilde{J}_1[S] \) and \( \tilde{J}_2[S], \) based on the asymptotic approximation (35), fixing the post-change mean in the first sensor at \( \theta_1 = 1, \) and letting the post-change mean in the second sensor, \( \theta_2, \) vary. In this way, we examine the relative performance loss as the relative magnitude of the anticipated change in the two sensors varies. In the Gaussian case we have the following, easily computable expressions (see, e.g., [25, p. 32]) for the renewal-theoretic quantities that are included in this asymptotic approximation:

\[ \beta_k = -\sum_{n=1}^{\infty} \left( \frac{\theta_k}{\sqrt{n}} \phi(c_{n,k}) - \frac{\theta_k^2}{2} \Phi(-c_{n,k}) \right), \]
\[ \delta_k = \frac{2}{\theta_k} \exp \left( \sum_{n=1}^{\infty} \frac{-2}{n} \Phi(-c_{n,k}) \right), \]

where \( c_{n,k} := \theta_k \sqrt{n}/2. \) The results of this computation are summarized in Fig. 1, where we see that selecting the \( p_k \)'s according to (36) leads to a more robust behavior in comparison to the other specifications when \( \theta_2 \) differs significantly from \( \theta_1. \) Of course, all specifications approach the behavior of identical thresholds when \( \theta_2 \) is close to \( \theta_1. \)
V. A Simulation Study

A. Description

In this section we present the results of a simulation study whose main goal is to compare the GLR-CUSUM, $S$, given by (11), the mixture-based CUSUM rules, $\tilde{S}$ and $\hat{S}$, given by (14) and (24), respectively, and the SUM-CUSUM, $M$, defined in (27). We set $K = 5$ and $h_k = \mathcal{N}(0, 1)$, $g_k = \mathcal{N}(1, 1)$. That is, all sensors initially observe iid Gaussian observations with variance 1, and at the time of the change the mean changes from 0 to 1 in an unknown subset of these sensors. For our comparisons to be fair, we need to guarantee that all detection rules have access to the same amount of prior information.

We consider the following regimes: (i) no prior information ($\mathcal{P} = \mathcal{P}_{\{K\}}$); (ii) knowing that at most $L$ sensors can be affected ($\mathcal{P} = \mathcal{P}_{\{L\}}$), where $1 \leq L \leq K$.

For the implementation of $S$ and $\tilde{S}$ we follow the first approach described in Subsection II-D. For the implementation of $\hat{S}$ we use the second approach, together with the adaptive-window based on regeneration times (15).

For all detection rules we have considered in this work, the worst case scenario is when the change occurs at time 0. Thus, when the actual affected subset is $A$, in order to compute the worst-case detection delay of each rule we simply need to simulate it under $\mathcal{P}_0^A$.

In order to illustrate the second-order asymptotic optimality property, we plot the additional number of observations required by each rule in order to detect the change compared to the optimal CUSUM test for which the affected subset is known in advance. Thus, if $T$ represents a generic detection rule and $A$ the true subset of affected sensors, we plot $E_{\mathcal{P}_0}[T] - E_{\mathcal{P}_0}[S]^A$ against $\text{ARL}[T] := E_{\mathcal{P}_0}[T]$ for various threshold choices. In order to illustrate the first-order asymptotic optimality property, we plot the ratio $E_{\mathcal{P}_0}[T]/E_{\mathcal{P}_0}[S]^A$ against $\text{ARL}[T]$ again for various threshold choices. In Table I, we present numerical results for all detection rules when we choose their thresholds so that the expected time to false alarm is approximately $\gamma = 10^5$. Standard errors are given based on 50,000 Monte Carlo simulation runs.

B. Results

1) No Prior Information: We present the results that correspond to the case of no prior information in Fig. 2–Fig. 3,
illustrating the notion of first-order and second-order asymptotic optimality, respectively, with respect to class \( \mathcal{P}_k \). We see that when 2 sensors are affected, the mixture-based CUSUM procedures are slightly worse than the GLR-CUSUM, whereas when 4 sensors are affected, the mixture CUSUMs take the lead. In both cases, the proposed schemes perform much better than SUM-CUSUM, \( M(5) \), whose inflicted performance loss (relative to the optimal) increases much faster as \( \gamma \) increases. Nevertheless, its ratio over the optimal performance decreases, which supports the result that SUM-CUSUM is first-order asymptotically optimal.

2) An Upper Bound on the Number of Affected Sensors Is Known: In Fig. 4–5 we assume that we know in advance that at most \( L \) sensors can be affected and we compare the GLR-CUSUM procedure, \( S \), the mixture-based CUSUM procedures, \( \hat{S} \) and \( \hat{S}(L/(2K)) \), and the scalable scheme \( M(L) \). Again, we see that \( M(L) \) performs uniformly worse than the proposed schemes in all cases.

VI. Conclusions

We considered a generalized multisensor sequential change detection problem, in which a number of possibly correlated sensors monitor a system in real time, a global parameter vector determines their joint distribution, and there is a change in an unknown subset of components of this parameter vector. In this context, we established a strong asymptotic optimality property for various CUSUM-based detection rules. In the special case that the sensors are independent and only a subset of them is affected by the change, we proposed feasible versions of the above procedures. We showed using simulation experiments that the proposed detection rules always outperform scalable detection rules, such as the one proposed in [10], and we argued that this phenomenon can be explained theoretically by the fact that the latter scheme enjoys a weaker form of asymptotic optimality. Moreover, we proposed a modification of a detection rule in [14] that is able to incorporate prior information. Finally, we considered the design of the multichart CUSUM in the special case that the change is known to affect exactly one sensor.

APPENDIX

Proof of Theorem 1: From the definition of the three detection rules it is clear that \( \bar{S}_b \leq S_b \) and \( S_b \leq \bar{S}_b \), therefore for any \( b \) and \( A \) we have:

\[ J_A[\bar{S}_b] \leq J_A[S_b] \quad \text{and} \quad E_\infty[S_b] \geq E_\infty[\bar{S}_b] \]

and

\[ J_A[\bar{S}_b] \leq J_A[S_b] \quad \text{and} \quad E_\infty[S_b] \geq E_\infty[\bar{S}_b]. \]

Consequently, in order to prove (i), it suffices to show that for any \( b > 0 \)

\[ E_\infty[\bar{S}_b] \geq e^b \quad \text{and} \quad E_\infty[\bar{S}_b] \geq e^b, \quad (37) \]

and in order to establish (ii), it suffices to show that as \( b \to \infty \)

\[ J_A[S_b] \leq \frac{b}{I_A} + O(1). \quad (38) \]

(i) In order to prove the first inequality in (37), we introduce the mixture probability measure

\[ \hat{P} := \sum_{A \in \mathcal{P}} p_A P_A^0, \]

and denote by \( \bar{E} \) the corresponding expectation. The mixture-based CUSUM rule \( \bar{S}_b \) results from repeated application, in the spirit of Lorden’s [8] construction, of the one-sided sequential test:

\[ \bar{T}_b := \inf \{ t \geq 0 : \bar{\Lambda}_t \geq \bar{e}^b \}, \]

where \( \bar{\Lambda} \) is the likelihood ratio process of \( \hat{P} \) versus \( P_\infty \), i.e.,

\[ \bar{\Lambda}_t := \int \frac{d\hat{P}}{dP_\infty}(\mathcal{F}_t) = \sum_{A \in \mathcal{P}} p_A \Lambda_t^A, \quad t \in \mathbb{N}. \]

Consequently, from [8, Th. 2] it follows that for every \( b > 0 \) we have

\[ E_\infty[\bar{S}_b] \geq \frac{1}{P_\infty(\bar{T}_b < \infty)}, \]

and from Wald’s likelihood ratio identity we obtain

\[ P_\infty(\bar{T}_b < \infty) = \mathbb{E} \left[ \Lambda_0^{-1} I_{\{\bar{\Lambda}_T < \infty\}} \right] \leq e^{-b}, \]

which completes the proof of the first inequality in (37). In order to establish the second inequality, we introduce for each subset \( A \in \mathcal{P} \) the corresponding Shiryaev-Roberts statistic

\[ R_t^A := \sum_{s=0}^{t-1} \exp(Z_t^A - Z_s^A), \quad R_0^A := 0 \]

and observe that

\[ R_t^A \geq \max_{0 \leq s < t} \exp(Z_t^A - Z_s^A) = \exp(\bar{Y}_t^A). \]

We also introduce the mixture Shiryaev-Roberts statistic \( R_t := \sum_{A \in \mathcal{P}} p_A R_t^A \) and the corresponding stopping time \( \bar{R}_b := \inf \{ t \geq 0 : R_t \geq e^b \} \). Then, for any \( b > 0 \) we obtain

\[ \bar{S}_b = \inf \left\{ t \in \mathbb{N} : \sum_{A \in \mathcal{P}} p_A \exp(\bar{Y}_t^A) \geq e^b \right\} \geq \inf \left\{ t \in \mathbb{N} : \sum_{A \in \mathcal{P}} p_A R_t^A \geq e^b \right\} = \bar{R}_b, \]

thus, it suffices to show that \( E_\infty[\bar{R}_b] \geq e^b \). In order to see this, note that for any \( b \) and \( t \) we have \( 0 \leq R_t < e^b \) on the event \( \{ \bar{R}_b > t \} \), therefore

\[ \int_{\{ \bar{R}_b > t \}} |R_t - t| dP_\infty \]

\[ \leq \int_{\{ \bar{R}_b > t \}} R_t dP_\infty + \int_{\{ \bar{R}_b > t \}} t dP_\infty \]

\[ \leq e^b P_\infty(\bar{R}_b > t) + \int_{\{ \bar{R}_b > t \}} \bar{R}_b dP_\infty. \]

For any given \( b \), the upper bound goes to 0 as \( t \to \infty \) as long as \( E_\infty[\bar{R}_b] < \infty \). But this can be assumed without any loss of generality, since otherwise the desired inequality
holds trivially. Therefore, we can apply the optional sampling theorem to the $P_\infty$-martingale $\{ R_i - t \}_{i \in \mathbb{N}}$ and the stopping time $\mathcal{R}_B$ and obtain $E_{\infty}[\mathcal{R}_B] = E_{\infty}[R_B] \geq e^b$, where the inequality holds from the definition of the stopping rule $\mathcal{R}_B$. This completes the proof of the second inequality in (37).

(ii) For any $b > 0$ and $A \in \mathcal{P}$, from the definition of $S_B$ in (11) it follows that $S_B \leq S_B^{\log p_A}$, where $S_B^{\log p_A}$ is the CUSUM rule defined in (5) with threshold $b - \log p_A$. Consequently, as $b \to \infty$ we have

$$J_A[S_B] \leq \frac{b - \log p_A}{I_A} + O(1) = \frac{b}{I_A} + O(1),$$

which completes the proof of (38).

Proof of Proposition 1: Fix $b > 0$ and $A \in \mathcal{P}$. (i) We will first prove the result for the GLR-CUSUM, $S_B$. On the event $\{ S_B > v \}$ we have

$$S_B = \inf \left\{ t > v : \max_{b \in \mathcal{P}} \left( Y^B_t - Y^B_v + Y^B_v + \log p_B \right) \geq b \right\}.$$

For each $B \in \mathcal{P}$, $(Y^B_t - Y^B_v)_{t \geq v}$ depends only on the post-change observations. Therefore, the detection statistic depends on the pre-change observations only through the non-negative random variables $Y^B_v$, $B \in \mathcal{P}$. Moreover, the detection statistic is increasing in these quantities, which means that, for any possible change point $v$, the worst-possible scenario for the history of observations up to $v$ is that $Y^B_v = 0$ for every $B \in \mathcal{P}$:

$$\text{ess sup} \ E^A_{\infty}[(S_B - v)^+ | \mathcal{F}_v] = E^A_{\infty}[(S_B - v)^+ | Y^B_v = 0 \ \forall \ B \in \mathcal{P}].$$

However, when $Y^B_v = 0$ for every $B \in \mathcal{P}$, each sequence $(Y^B_t - Y^B_v)_{t \geq v}$ under $P^A_{\infty}$ has the same distribution as $(Y^B_t)_{t \geq 0}$ under $P_0^A$. Thus, for every change-point $v$ we have

$$E^A_{\infty}[(S_B - v)^+ | Y^B_v = 0 \ \forall \ B \in \mathcal{P}] = E^A_0[S_B],$$

which proves that $J_A[S_B] = E^A_0[S_B]$. (ii) In order to prove the corresponding result for $\tilde{S}_B$, we note that on the event $\{ S_B > v \}$ we have

$$\tilde{S}_B = \inf \left\{ t > v : \max_{b \in \mathcal{P}} \sum_{B \in \mathcal{P}} p_B \exp(Z^B_t - Z^B_v) \geq e^b \right\} \leq \inf \left\{ t > v : \max_{b \in \mathcal{P}} \sum_{B \in \mathcal{P}} p_B \exp(Z^B_t - Z^B_v) \geq e^b \right\} := \tilde{S}_B^{(v)}.$$

For every change point $v$, $\tilde{S}_B^{(v)}$ under $P^A_{\infty}$ has the same distribution as $\tilde{S}_B$ under $P^A_{\infty}$ and is independent of $\mathcal{F}_v$. Therefore,

$$E^A_{\infty}[(\tilde{S}_B - v)^+ | \mathcal{F}_v] \leq E^A_{\infty}[(\tilde{S}_B^{(v)})] = E^A_0[\tilde{S}_B],$$

which proves that $J_A[\tilde{S}_B] \leq E^A_0[\tilde{S}_B]$, and consequently $J_A[\tilde{S}_B] = E^A_0[\tilde{S}_B]$. (iii) Finally, on the event $\{ S_B > v \}$ we have

$$\tilde{S}_B = \inf \left\{ t > v : \sum_{b \in \mathcal{P}} p_B \exp(\tilde{Y}^B_t - \tilde{Y}^B_v + \tilde{Y}^B_v) \geq b \right\}.$$

Fix $B \in \mathcal{P}$. When $\tilde{Y}^B_v < 0$, then from recursion (7) it follows that $(\tilde{Y}^B_t)_{t \geq v}$ is independent of $\mathcal{F}_v$ and that its distribution under $P^A_{\infty}$ is the same as that of $\{ \tilde{Y}^B_t \}_{t \geq 0}$ under $P^A_0$. When $\tilde{Y}^B_v \geq 0$, a similar argument as in (i) shows that the worst case is $\tilde{Y}^B_v = 0$. Thus, we conclude that $J_A[\tilde{S}_B] = E^A_0[\tilde{S}_B]$.

Proof of Proposition 2: Recall the notation $C(\mathcal{P}) := \sum_{B \in \mathcal{P}} P[B]$. (i) We observe that $\log C(\mathcal{P}_L) = L \log p + \log |\mathcal{P}_L|$ and

$$\max_{A \in \mathcal{P}_L} \sum_{k \in A} (Z^k_t - Z^k_v + \log p) = \sum_{k=1}^{L} (Z^k_{s_{k+1}} + \log p),$$

which proves (22). Moreover, we observe that $\log C(\mathcal{P}_L) = \log \left( \sum_{k=1}^{L} |\mathcal{P}_k| p^k \right)$ and that

$$\max_{A \in \mathcal{P}_k} \sum_{k \in A} (Z^k_t - Z^k_v + \log p) = \sum_{k=1}^{L} (Z^k_{s_{k+1}} + \log p)$$

whenever there is at least one $k$ such that $Z^k_{s_{k+1}} + \log p > 0$. On the other hand, if $Z^k_t - Z^k_v + \log p \leq 0$ for every $1 \leq k \leq K$, the detection statistic cannot reach a positive threshold, which completes the proof of (23).

(ii) For any $\pi \in (0, 1)$ and $p = \pi / (1 - \pi)$ we have

$$K \prod_{k=1}^{K} \left( 1 - \pi + \pi \frac{A_k}{A_L} \right) = (1 - \pi)^K + \sum_{A \in \mathcal{P}_k} (1 - \pi)^{K - |A| / |A_L|} \prod_{k \in A} \frac{A_k}{A_L},$$

which proves (24).

Proof of Theorem 2: (i) Fix $b$. We need to show that

$$E_{\infty}[\hat{S}_B(\pi)] \geq e^b \quad \text{and} \quad E_{\infty}[\tilde{S}_B(\pi)] \geq e^b / |\mathcal{P}_k|$$

where $\pi \in (0, 1)$ in the first and $\pi \in (0, 1)$ in the second inequality. In order to prove the first inequality, we recall that $\hat{S}_B(\pi)$ can be obtained with the repeated application of the one-sided sequential test $\hat{Y}_B(\pi)$, defined in (25). Therefore, from [8, Th. 2] it follows that

$$E_{\infty}[\hat{S}_B(\pi)] \geq \frac{1}{P_{\infty}(\hat{Y}_B(\pi) < \infty)}$$

and from Wald’s likelihood ratio identity we obtain

$$P_{\infty}(\hat{Y}_B(\pi) < \infty) = E_\pi \left[ \exp \left( -\hat{Z}_B(\pi) \mathbbm{1}_{[\hat{Y}_B(\pi) < \infty]} \right) \right] \leq e^{-b},$$

where $\hat{Z}$ is the logarithm of the likelihood ratio $\hat{A}_\pi$ defined in (25), i.e.,

$$\hat{Z}_t := \log \hat{A}_\pi = \sum_{k=1}^{K} \log \left( 1 - \pi + \pi \frac{A_k}{A_L} \right).$$

This completes the proof of the first inequality in (39). In order to prove the second inequality, it suffices to observe that for any $b > 0$ and $\pi \in (0, 1)$ we have

$$\tilde{S}_B(\pi) \geq \tilde{S}_B(\pi) \geq \inf \left\{ t : \max_{0 \leq s \leq t} \sum_{k=1}^{L} (Z^k_t - Z^k_v) \geq b \right\}.$$
and that the lower bound coincides (up to a constant) with the unweighted GLR-CUSUM in the case of complete uncertainty, that is the detection rule obtained by setting $L = K$ and $\pi = 1$ in (23).

(ii) When $\pi = 1$, $\hat{S}_b(\pi)$ coincides with the GLR-CUSUM in the case of complete uncertainty, whose asymptotic optimality has already been established. Therefore, without loss of generality, we assume that $\pi \in (0,1)$. Due to the pathwise inequality $\hat{S}_b(\pi) \geq \hat{S}_b(\pi)$, it suffices to show that for every subset $A$ we have

$$J_A[\hat{S}_b(\pi)] \leq \frac{b}{I_A} + \mathcal{O}(1),$$

where $\mathcal{O}(1)$ is a bounded term as $b \to \infty$. From [8, Th. 2] it follows that $J_A[\hat{S}_b(\pi)] \leq E_0[A[T_b(\pi)]]$ for every $b > 0$, whereas for every $t$ we observe that

$$\hat{Z}_t = \sum_{k=1}^{K} \log \left( 1 - \pi + \pi \exp(Z^{\text{opt}}_t) \right)$$

$$\geq \sum_{k \in A} (Z^{\text{opt}}_k + \log \pi) + \sum_{k \notin A} \log(1 - \pi)$$

$$= Z^{\text{opt}}_t + c_A,$$

where $c_A := \mathbb{1}_A \log \pi + (K - |A|) \log(1 - \pi)$. Therefore, $\hat{T}_b(\pi) \leq \inf \{ t : Z^{\text{opt}}_t \geq b + c_A \}$ for every $b > 0$, and consequently

$$E_0[A[T_b(\pi)]] \leq \frac{b + c_A}{I_A} + \mathcal{O}(1),$$

which completes the proof.

Acknowledgments

The authors would like to thank Drs. Alexander Tartakovsky, George Moustakides, Venu Veeravalli and Emmanuel Yashchin for stimulating discussions and helpful suggestions, as well as the two referees whose comments helped us improve significantly an earlier version of this paper.

References

[1] A. G. Tartakovsky, B. L. Rozovskii, R. B. Blázek, and H. Kim, “A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods,” IEEE Trans. Signal Process., vol. 54, no. 9, pp. 3372–3382, Sep. 2006.

[2] A. G. Tartakovsky, B. L. Rozovskii, R. B. Blázek, and H. Kim, “Detection of intrusions in information systems by sequential change-point methods,” Statist. Methodol., vol. 3, no. 3, pp. 252–293, 2006.

[3] C. Lévy-Leduc and F. Roueff, “Detection and localization of change-points in high-dimensional network traffic data,” Ann. Appl. Statist., vol. 3, no. 2, pp. 637–662, 2009.

[4] C. Sonesson and D. Bock, “A review and discussion of prospective statistical surveillance in public health,” J. Roy. Statist. Soc. A (Statist. Soc.), vol. 166, no. 1, pp. 5–21, 2003.

[5] G. Shmueli and H. Burkom, “Statistical challenges facing early outbreak detection in biosurveillance,” Technometrics, vol. 52, no. 1, pp. 39–51, Feb. 2010.

[6] E. S. Page, “Continuous inspection schemes,” Biometrica, vol. 41, no. 1–2, pp. 100–115, 1954.

[7] G. V. Moustakides, “Optimal stopping times for detecting changes in distributions,” Ann. Statist., vol. 14, no. 4, pp. 1379–1387, 1986.

[8] G. Lorden, “Procedures for reacting to a change in distribution,” Ann. Math. Statist., vol. 42, no. 6, pp. 1897–1908, 1971.

[9] A. Tartakovsky, I. Nikiforov, and M. Basseville, Sequential Analysis: Hypothesis Testing and Changepoint Detection, Boca Raton, FL, USA: CRC Press, 2014.

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

Y. Mei, “Quickest detection in censoring sensor networks,” in Proc. IEEE Int. Symp. Inf. Theory, Jul./Aug. 2011, pp. 2148–2152.

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.

T. Banerjee and V. V. Veeravalli, “Data-efficient quickest change detection in sensor networks,” IEEE Trans. Signal Process., vol. 63, no. 14, pp. 3727–3735, Jul. 2015.

Y. Xie and D. Siegmund, “Sequential multi-sensor change-point detection,” Ann. Statist., vol. 41, no. 2, pp. 670–692, 2013.

Y. C. Chen, T. Banerjee, A. D. Dominguez-Garcia, and V. V. Veeravalli, “Quickest line outage detection and identification,” IEEE Trans. Power Syst., vol. 31, no. 1, pp. 749–758, Jan. 2016.

D. Siegmund and E. S. Venkatraman, “Using the generalized likelihood ratio statistic for sequential detection of a change-point,” Ann. Statist., vol. 23, no. 1, pp. 255–271, 1995.

V. Raghavan and V. V. Veeravalli, “Quickest change detection of a Markov process across a sensor array,” IEEE Trans. Inf. Theory, vol. 46, no. 4, pp. 1961–1981, Apr. 2010.

M. Lajkovich, “Bayesian quickest detection in sensor arrays,” Sequential Anal., Design Methods Appl., vol. 31, no. 4, pp. 481–504, 2012.

O. Hadjiliadis, H. Zhang, and H. V. Poor, “One shot schemes for decentralized quickest change detection,” IEEE Trans. Inf. Theory, vol. 55, no. 7, pp. 3346–3359, Jul. 2009.

H. Zhang, N. Rodosthenous, and O. Hadjiliadis, “Robustness of the N-CUSUM stopping rule in a Wiener disorder problem,” Ann. Appl. Probab., vol. 25, no. 6, pp. 3405–3433, Dec. 2015.

I. V. Nikiforov, “A generalized change detection problem,” IEEE Trans. Inf. Theory, vol. 41, no. 1, pp. 171–187, Jan. 1995.

I. V. Nikiforov, “Two strategies in the problem of change detection and isolation,” IEEE Trans. Inf. Theory, vol. 43, no. 2, pp. 770–776, Mar. 1997.

T. Oskiper and H. V. Poor, “Online activity detection in a multiuser environment using the matrix CUSUM algorithm,” IEEE Trans. Inf. Theory, vol. 48, no. 2, pp. 477–493, Feb. 2002.

A. G. Tartakovsky, “Multidecision quickest change-point detection: Previous achievements and open problems,” Sequential Anal., vol. 27, no. 2, pp. 201–231, 2008.

M. Woodroofe, Nonlinear Renewal Theory in Sequential Analysis. Philadelphia, PA, USA: SIAM, 1982.

R. A. Khan, “Detecting changes in probabilities of a multi-component process,” Sequential Anal., vol. 14, no. 4, pp. 375–388, 1995.

E. Yashchin, “Regenerative likelihood ratio control schemes,” in Frontiers in Statistical Quality Control (Frontiers in Statistical Quality Control), vol. 11, S. Knoth and W. Schmid, Eds. Cham, Switzerland: Springer International Publishing, 2015.

E. Yashchin, “Change-point models in industrial applications,” in Proc. 2nd World Congr. Nonlinear Anal., Dec. 1997, vol. 30, no. 7, pp. 3997–4006.

E. Yashchin, “Likelihood ratio methods for monitoring parameters of a nested random effect model,” J. Amer. Statist. Assoc., vol. 90, no. 430, pp. 729–738, Jun. 1995.