MULTISTREAM QUICKEST CHANGE DETECTION: ASYMPTOTIC OPTIMALITY UNDER A SPARSE SIGNAL

Georgios Fellouris[‡] *George V. Moustakides*[⋆] *Venu V. Veeravalli*[†]

[‡] Univ. of Illinois, Urbana-Champaign
Statistics Dept. and Coordinated Science Lab
Champaign, IL, 61820
[⋆] Rutgers Univ. / Univ. of Patras
Comp. Science Dept./ ECE Dept.
New Brunswick, NJ / Rion, Greece.
[†] Univ. of Illinois, Urbana-Champaign
ECE Dept. and Coordinated Science Lab
Urbana, IL, 61801

ABSTRACT

In multichannel sequential change detection, multiple sensors monitor a system in which an abrupt change occurs at some unknown time and is perceived by an unknown subset of sensors. The goal is to detect this change quickly, while controlling the rate of false alarms. In the traditional asymptotic analysis of this problem, the false alarm rate goes to 0 while all other parameters remain fixed. We argue that this framework is not very informative, as the corresponding asymptotic optimality property cannot differentiate between universal and parsimonious rules. We propose an asymptotic framework in which the number of sensors also goes to infinity, and we show that in this context universal rules may fail to be asymptotically optimal when the number of streams is not very small. On the other hand, parsimonious rules are shown to be asymptotically optimal under reasonable sparsity conditions.

Index Terms— Sequential Change Detection, CUSUM, Multichannel, Multisensor, Sparse.

1. INTRODUCTION

The problem of efficiently detecting changes in stochastic processes, often referred to as *sequential (or quickest) change detection*, arises in various branches of science and engineering [1, 2]. In this paper, we consider an important case of this problem in which multiple streams of data are collected in various sensors (channels), and an unknown subset of the streams undergo a change in distribution at some unknown point in time. The goal is to combine the information from all sensors in order to quickly detect the change, while controlling the rate of false alarms. This problem has been studied extensively when the change is perceived by exactly *one* sensor whose identity is unknown [3, 4, 5, 6, 7]. More recently, the assumption of a unique affected sensor has been removed and various works have allowed for the change to affect an unknown subset of sensors [8, 9, 10, 11, 12].

A multichannel sequential change-detection rule is typically understood to be efficient when it achieves the performance of the oracle procedure that assumes knowledge of the true affected subset, to a first-order term as the false alarm rate goes to 0, for any possible affected subset. It is well known that there are procedures that are completely ignorant regarding the affected subset, yet still enjoy asymptotic optimality [8, 11]. Therefore, the classical notion of asymptotic optimality fails to distinguish between *universal* procedures that do not make any assumptions regarding the size or location of the signal, and *parsimonious* procedures that utilize prior information regarding the affected subset of sensors, and perform better in practice, especially when the number of streams is large [11].

Motivated by this observation, we propose a notion of asymptotic optimality in which the number of sensors goes to infinity as the false alarm rate goes to zero. We focus in particular on the case of *sparse signal*, where the maximum proportion of affected sensors goes to 0. Our main contribution in this work is that we obtain sufficient conditions for the asymptotic optimality in this novel sense of existing multichannel detection rules [10, 11]. These conditions allow us to distinguish parsimonious and universal procedures, especially when a large number of streams is monitored.

The rest of this paper is organized as follows: In Section 2 we formulate the problem mathematically and we review various procedures that have been proposed in the literature. In Section 3 we state our main result and discuss its ramifications. In Section 4 we conclude and discuss extensions of this work.

2. PROBLEM FORMULATION

Suppose we collect data sequentially from K sensors. For each $k \in [K]$, let $X^k \equiv \{X_t^k,\, t = 1, 2 \dots\}$ be the sequence of observations in the k^{th} sensor, where $[K] := \{1, \dots, K\}$. We assume that the sequences $X^1, \dots, X^K$ are independent, each X^k is a sequence of independent random variables, and that there is an unknown, deterministic point in time $\nu \in \{0, 1, \dots\}$ at which the distribution changes in an unknown subset of sensors, $\mathcal{B} \subset [K]$, i.e.,

$$X_t^k \sim f_k, k \notin \mathcal{B}, \quad X_t^k \sim \begin{cases} f_k, & t \leq \nu \\ g_k, & t > \nu, \end{cases} \quad k \in \mathcal{B}, \qquad (1)$$

The first and third author were supported by the US National Science Foundation under grant CIF 1514245. The second author was supported by the US National Science Foundation under grant CIF 1513373.

where $t = 1, 2 \ldots$. Here, f_k and g_k are known densities with respect to a σ-finite measure λ_k such that the Kullback-Leibler information number,

$$\mathcal{I}_k := \int \log\left(\frac{g_k(x)}{f_k(x)}\right) g_k(x)\, \lambda_k(dx),$$

is positive and finite. We denote by $\mathsf{P}_\nu^{\mathcal{B}}$ the underlying probability measure when the change occurs at time ν in subset $\mathcal{B}$, and by P_∞ the corresponding measure when there is no change in any sensor. The goal is to stop sampling as quickly as possible after the change has occurred, based on the data from all sensors up to this time. Therefore, a *multichannel sequential change detection rule* T is an $\{\mathscr{F}_t\}$-stopping time, where

$$\mathscr{F}_t = \sigma(X_s^k : 1 \leq s \leq t,\ k \in [K]).$$

That is, the decision to stop and raise an alarm at time t is determined by the observations in all sensors up to time t. Following Lorden's approach [13], we quantify the detection delay of an arbitrary detection rule T when the change occurs in subset $\mathcal{B} \subset [K]$ with the following criterion:

$$\mathcal{J}_{\mathcal{B}}[T] := \sup_{\nu} \text{ essup } \mathsf{E}_\nu^{\mathcal{B}}\left[(T-\nu)^+ | \mathscr{F}_\nu\right],$$

where $\mathsf{E}_\nu^{\mathcal{B}}$ is expectation under $\mathsf{P}_\nu^{\mathcal{B}}$. In other words, we consider the worst-case scenario with respect to change-point ν and the observations from all sensors until the time of the change. Moreover, we denote by $\mathcal{C}_\gamma$ the class of sequential change-detection rules for which the expected time to false alarm is bounded below by γ, i.e., $\mathcal{C}_\gamma = \{T : \mathsf{E}_\infty[T] \geq \gamma\}$, where $\gamma > 1$ is a user-specified level and E_∞ refers to expectation under P_∞. The problem then is to minimize $\mathcal{J}_{\mathcal{B}}$ among detection rules in class $\mathcal{C}_\gamma$ *for every possible affected subset* $\mathcal{B}$.

2.1. Oracle rule

For each time t and sensor k we denote by Z_t^k the cumulative log-likelihood ratio of the observations in sensor k up to time t, i.e.,

$$Z_t^k = Z_{t-1}^k + \ell_t^k, \qquad \ell_t^k := \log\left(\frac{g_k(X_t^k)}{f_k(X_t^k)}\right). \tag{2}$$

Let $W^{\mathcal{B}}$ denote the Cumulative Sums (CUSUM) statistic [14] for detecting a change in subset $\mathcal{B} \subset [K]$, which is defined by the following recursion

$$W_t^{\mathcal{B}} := \left(W_{t-1}^{\mathcal{B}} + \sum_{k\in\mathcal{B}} \ell_t^k\right)^+, \quad W_0^{\mathcal{B}} := 0,$$

where $W_0^{\mathcal{B}} := 0$. Let $S_{\mathcal{B}}(h)$ be the corresponding CUSUM stopping time, i.e., the first time $W^{\mathcal{B}}$ exceeds a fixed threshold h, i.e.,

$$S_{\mathcal{B}}(h) := \inf\left\{t \geq 1 : W_t^{\mathcal{B}} \geq h\right\}. \tag{3}$$

It is well known that $S_{\mathcal{B}}$ optimizes $\mathcal{J}_{\mathcal{B}}$ within the class of detection rules in $\mathcal{C}_\gamma$ whenever its threshold h satisfies the false alarm constraint with equality [15]. Moreover, we have the following first-order approximation to the optimal performance [13] as $\gamma \to \infty$:

$$\inf_{T\in\mathcal{C}_\gamma} \mathcal{J}_{\mathcal{B}}[T] \sim \frac{\log\gamma}{\sum_{k\in\mathcal{B}} \mathcal{I}_k}, \tag{4}$$

where by $x_\gamma \sim y_\gamma$ we mean that $x_\gamma / y_\gamma \to 1$ as $\gamma \to \infty$. Finally, if the following second moment condition

$$\int \left(\log\left(\frac{g_k(x)}{f_k(x)}\right)\right)^2 g_k(x)\, \lambda_k(dx) < \infty \tag{5}$$

is satisfied, from the exact optimality of the CUSUM test and renewal theory (see, e.g., [1]) it can be deduced that

$$\inf_{T\in\mathcal{C}_\gamma} \mathcal{J}_{\mathcal{B}}[T] = \frac{\log\gamma}{\sum_{k\in\mathcal{B}} \mathcal{I}_k} + \Theta(1), \tag{6}$$

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

2.2. Universal detection rules

A multichannel sequential change-detection rule is typically said to be *asymptotically optimal* if it achieves the first-order asymptotic approximation to the optimal performance (4) *for any possible affected subset* $\mathcal{B}$ as $\gamma \to \infty$, while all other parameters of the problem remain fixed. It is well known that it is possible to design asymptotically optimal rules even when there is complete ignorance regarding the affected subset. Indeed, this is the case for *Sum-CUSUM*, the procedure that stops when *the sum of all local CUSUM statistics* is above a positive threshold $h > 0$, i.e.,

$$\check{S}(h) := \inf\left\{t \geq 1 : \sum_{k=1}^{K} W_t^k \geq h\right\}, \tag{7}$$

which was suggested in [8]. Here, W_t^k corresponds to the statistic $W_t^{\mathcal{B}}$ with $\mathcal{B} = \{k\}$. In fact, it has been shown that the optimal performance can be achieved up to the second-order term (6) by the following universal procedure

$$\hat{S}(h) := \inf\left\{t \geq 1 : \max_{0\leq s\leq t} \sum_{k=1}^{K} (Z_t^k - Z_s^k)^+ \geq h\right\}, \tag{8}$$

where $x^+ = \max\{x, 0\}$. We will refer to this procedure as *GLR-CUSUM*, as it essentially performs a maximization at any given time over the completely unknown affected subset of sensors [11]. The same second-order asymptotic optimality property can be achieved by the procedure

$$\tilde{S}_p(h) := \inf\left\{t \geq 1 : \max_{0\leq s\leq t} \sum_{k=1}^{K} g_p(Z_t^k - Z_s^k) \geq h\right\}, \tag{9}$$

where $g_p(x) = \log(1 - p + p\exp\{x\})$ [11]. We will refer to this procedure as *Mix-p-CUSUM*, as it is motivated by a mixture model according to which the change occurs with probability p in each sensor [10, 11]. It is important to underline that, in this classical framework, *Mix-p-CUSUM* is second-order asymptotically optimal *for any choice of* $p \in (0, 1)$, not only when p is the true proportion of affected sensors [11].

2.3. Incorporating prior information

It is natural and straightforward to modify the above rules in order to incorporate prior information, such as that signal is present in at most L sensors, where L is some integer between 1 and K. In this context, *Sum-CUSUM* can be replaced by the first time the sum of the L largest CUSUM statistics is above a positive threshold $h > 0$, i.e.,

$$\check{S}_L(h) := \inf \left\{ t \geq 1 : \sum_{k=1}^{L} W_t^{(k)} \geq h \right\}, \tag{10}$$

where $W_t^{(1)} \leq \ldots \leq W_t^{(K)}$, a procedure to which we will refer as *Top-L-Sum-CUSUM* [9]. Similarly, as it was shown in [11], GLR-CUSUM can be modified as follows

$$\hat{S}_L(h) := \inf \left\{ t \geq 1 : \max_{0 \leq s \leq t} \sum_{k=1}^{L} \left(Z_{s:t}^{(k)} \right)^+ \geq h \right\}, \tag{11}$$

where $Z_{s:t}^k = Z_t^k - Z_s^k$ and $Z_{s:t}^{(1)} \leq \ldots \leq Z_{s:t}^{(K)}$. We will refer to the latter scheme as *Top-L-GLR-CUSUM*.

Whenever the size of the affected subset is smaller or equal to L, *Top-L-Sum-CUSUM* and *Top-L-GLR-CUSUM* preserve the asymptotic optimality properties of *Sum-CUSUM* and *GLR-CUSUM* respectively, and perform better in practice [8, 10, 11]. Of course, this is quite expected, since these procedures utilize information that the universal procedures in (7)-(8) do not possess.

3. ASYMPTOTIC OPTIMALITY UNDER A SPARSE SIGNAL

The discussion of the previous section suggests that the classical notion of asymptotic optimality, either first-order or second-order, is not informative enough to distinguish between universal procedures and their parsimonious modifications when prior information about the signal is available. In order to address this issue, we suggest a novel definition of asymptotic optimality, according to which a multichannel sequential change-detection rule is asymptotically optimal when it achieves the optimal performance, to a first-order approximation, not only as the rate of false alarms goes to 0, but also as the number of sensors goes to infinity. Furthermore, we focus on the case of a sparse signal that $L/K \to 0$, where L is the *maximum possible* number of affected sensors.

Definition 1. *We say that a multichannel sequential change-detection rule $T^* \in \mathcal{C}_\gamma$ is asymptotically optimal in the case of a sparse signal when for every $\mathcal{B} \subset [\mathcal{N}]$ such that $|\mathcal{B}| \leq L$ we have*

$$\mathcal{J}_\mathcal{B}[T^*] \sim \inf_{T \in \mathcal{C}_\gamma} \mathcal{J}_\mathcal{B}[T]$$

as $\gamma \to \infty$ and $L/K \to 0$.

In the following theorem we obtain sufficient conditions for *(Top-L)-GLR-CUSUM* and *Mix-p-CUSUM* to be asymptotically optimal in the above sense. We also provide a sketch of the proof, which will be presented in full detail elsewhere.

Theorem 1. *Suppose that condition* (5) *holds. As $\gamma \to \infty$ and $L/K \to 0$*

(i) GLR-CUSUM and Mix-p-CUSUM with fixed $p \in (0, 1)$ are both asymptotically optimal when $K = o(\log \gamma)$,

(ii) Top-L-Sum-CUSUM and Mix-p-CUSUM with $p = L/K$ are both asymptotically optimal when

$$L \log(K/L) = o(\log \gamma).$$

Proof. The first step in this proof relies on the following non-asymptotic lower bound on the optimal performance

$$\inf_{T \in \mathcal{C}_\gamma} \mathcal{J}_\mathcal{B}[T] \geq \frac{\log \gamma}{\sum_{k \in \mathcal{B}} \mathcal{I}_k} + \Theta(1), \tag{12}$$

where $\Theta(1)$ is a bounded term as $\gamma, |\mathcal{B}| \to \infty$. For the *GLR-CUSUM* we can show that if the threshold is selected such that the false alarm constraint be satisfied with equality, then

$$\mathcal{J}_\mathcal{B}[\hat{S}] \leq \frac{\log \gamma + \log |\mathcal{P}|}{\sum_{k \in \mathcal{B}} \mathcal{I}_k} + \mathcal{O}(1), \tag{13}$$

where $\mathcal{O}(1)$ is a bounded term as $\gamma, L, K \to \infty$ and $|\mathcal{P}|$ is the number of all subsets of $[K]$, i.e., 2^K. For *Top-L-GLR-CUSUM* we have the same upper bound with the difference that $|\mathcal{P}|$ is now the number of all subsets of $[K]$ whose size is at most L, thus,

$$|\mathcal{P}| = \sum_{j=1}^{L} \binom{K}{j} \sim L \log(K/L)$$

as $L/K \to 0$. A comparison of (12) and (13) completes the proof for *GLR-CUSUM* and *Top-L-GLR-CUSUM*.

Now, when the threshold of the *Mix-p-CUSUM* is selected such that the false alarm constraint be satisfied with equality, we have

$$\mathcal{J}_\mathcal{B}[\tilde{S}_p] \leq \frac{\log \gamma + K\, H(\pi, p)}{\sum_{k \in \mathcal{B}} \mathcal{I}_k} + \mathcal{O}(1), \tag{14}$$

where $\pi := L/K$ and

$$H(x, y) := -x \log y - (1 - x) \log(1 - y), \quad x, y \in (0, 1).$$

When p is fixed, $H(p, \pi) = \mathcal{O}(1)$ for every L, K. However, when we set $p = L/K \equiv \pi$, then $H(\pi, \pi)$ becomes the entropy of a Bernoulli random variable with parameter π, and $K\, H(\pi, \pi) \sim L \log(K/L)$ as $\pi \equiv L/K \to 0$. Comparing (14) with (12) completes the proof for *Mix-p-CUSUM*. □

3.1. Discussion

Theorem 1 applies to two existing procedures in the literature, *Mix-p-CUSUM* and *(Top-L)-GLR-CUSUM*, and suggests that their asymptotic optimality under a sparse signal requires certain sparsity conditions, contrary to the classical asymptotic framework where (second-order) asymptotic optimality is always guaranteed.

Specifically, the asymptotic optimality of universal detection rules, such as the *GLR-CUSUM* and *Mix-p-CUSUM* with fixed p, is guaranteed when $K = o(\log \gamma)$. With a false alarm rate of the order $\gamma \approx 10^5$, this condition is satisfied with a very small number of sensors, i.e., $K \ll 10$. On the other hand, parsimonious versions of these procedures become asymptotically optimal under a much more reasonable sparsity constraint. These results agree with our intuition, as well as empirical findings, that universal rules become less efficient in practice than parsimonious rules that utilize sparsity information when the number of streams increases.

Another interesting observation is that while the asymptotic optimality of *Mix-p-CUSUM* cannot be achieved under a sparse signal with an arbitrary, fixed p, at the same time it does *not* require p to agree with the true proportion of signals. It suffices to select p equal to the *maximum* proportion of signals, which is a much weaker and easily satisfied requirement. This result suggests that *Mix-p-CUSUM* will be robust with respect to p not only when the number of streams is small [10, 11], but also with a larger number of streams.

Finally, we expect that *Sum-CUSUM* should also fail to be asymptotically optimal in the sparse setup under consideration unless K is very small, and that the corresponding condition for the parsimonious *Top-L-Sum-CUSUM* will be much weaker. The determination of these conditions, and their comparison to the corresponding conditions of Theorem 1, will shed light to the performance loss that is inflicted upon these procedures when the number of sensors is not very small. This is a more complicated task that is left for future work. It is worth emphasizing however that current results, which assume a fixed number of sensors as the rate of false alarms goes to zero, cannot provide an answer to this question.

4. CONCLUSIONS

One of the main goals of an asymptotic optimality property is to distinguish among competing procedures, when the optimal procedure is either unknown or infeasible. In multichannel sequential change detection, a procedure is typically called asymptotically optimal if it achieves *under every possible scenario regarding the affected subset* the same first-order asymptotic performance as the corresponding oracle rule that assumes knowledge of the true affected subset. However, the optimal asymptotic performance in this classical definition is typically considered as the rate of false alarms goes to 0, *while all parameters of the problem remain fixed*. We argued that this is a weak property that fails to distinguish between universal procedures with no information regarding the signal and parsimonious procedures that utilize such information, and therefore should be preferred in practice. To remedy this problem, we proposed a stronger notion of asymptotic optimality, in which the number of streams also goes to infinity, and the maximum proportion of affected sensors goes to 0. In this regime, existing procedures in the literature, such as the *Top-L-GLR-CUSUM* and *Mix-p-CUSUM* with L and p equal to the maximum possible number and proportion of affected sensors respectively, remain asymptotically optimal under reasonable sparsity conditions.

While *Sum-CUSUM* and its parsimonious modifications are attractive in practice due to to their recursive nature, their performance loss may increase dramatically in a sparse setup, suggesting the use of more statistically efficient procedures such as *Top-L-GLR-CUSUM* and *Mix-p-CUSUM*. The exact implementation of the latter requires storing a K-dimensional vector of log-likelihood ratios for a random time-interval, whose expected length is finite, but grows with the number of sensors [11]. Nevertheless, it is possible to design adaptive windows of deterministic length that can make the above schemes computationally feasible in practice, while preserving their powerful detection ability.

5. REFERENCES

[1] Alexander Tartakovsky, Igor Nikiforov, and Michèle Basseville, *Sequential Analysis: Hypothesis Testing and Changepoint Detection*, CRC Press, 2014.

[2] H. Vincent Poor and Olympia Hadjiliadis, *Quickest Detection*, Cambridge University Press, 2008.

[3] Alexander G. Tartakovsky, "Asymptotically minimax multialternative sequential rule for disorder detection," *Proceedings of the Steklov Institute of Mathematics*, vol. 202, no. 4, pp. 229–236, 1994.

[4] Rasul A. Khan, "Detecting changes in probabilities of a multicomponent process," *Sequential Analysis*, vol. 14, no. 4, pp. 375–388, 1995.

[5] Alexander G. Tartakovsky, Boris L. Rozovskii, Rudolf B. Blažek, and Hongjoong Kim, "A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods," *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 3372–3382, 2006.

[6] Alexander G. Tartakovsky, Boris L. Rozovskii, Rudolf B. Blažek, and Hongjoong Kim, "Detection of intrusions in information systems by sequential changepoint methods (with discussion)," *Statistical Methodology*, vol. 3, no. 3, pp. 252–340, 2006.

[7] Alexander G. Tartakovsky and Venugopal V. Veeravalli, "Change-point detection in multichannel and distributed systems," in *Applied Sequential Methodologies: Real-World Examples with Data Analysis*, Nitis Mukhopadhyay, Sujay Datta, and Saibal Chattopadhyay, Eds., vol. 173, pp. 339–370. Marcel Dekker, Inc., New York, USA, 2004.

[8] Yajun Mei, "Efficient scalable schemes for monitoring a large number of data streams," *Biometrika*, vol. 97, no. 2, pp. 419–433, 2010.

[9] Yajun Mei, "Quickest detection in censoring sensor networks," in *IEEE International Symposium on Information Theory*, 2011, pp. 2148–2152.

[10] Yao Xie and David Siegmund, "Sequential multi-sensor change-point detection," *The Annals of Statistics*, vol. 41, no. 2, pp. 670–692, 2013.

[11] G. Fellouris and G. Sokolov, "Second-order asymptotic optimality in multisensor sequential change detection," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3662 – 3675, June 2016.

[12] S. Banerjee and G. Fellouris, "Decentralized sequential change detection with ordered cusums," in *Information Theory (ISIT), 2016 IEEE International Symposium on*, Barcelona, Spain, July 2016.

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

[14] Ewan S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1, pp. 100–115, 1954.

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