ASYMPTOTICALLY OPTIMAL QUICKEST CHANGE DETECTION UNDER A NUISANCE CHANGE

Tze Siong Lau and Wee Peng Tay *

Nanyang Technological University
School of Electrical and Electronic Engineering
50 Nanyang Avenue, Singapore 639798

ABSTRACT

This paper considers the problem of quickest detection of a change in distribution where the signal may undergo both nuisance and critical changes. Our goal is to detect the critical change without raising a false alarm over the nuisance change. We formulate the quickest change detection (QCD) problem in the presence of a nuisance change following Lorden's formulation. We propose a window-limited sequential change detection procedure based on the generalized likelihood ratio test statistic for the problem of QCD in which both nuisance and critical changes may occur. We derive a recursive update scheme for our proposed test statistic and show that our test is asymptotically optimal under mild technical conditions. We compare our proposed stopping rules with a naive 2-stage stopping time, which attempts to detect the changes using separate CuSum stopping procedures for the nuisance and critical changes. Simulations suggest that our proposed stopping time outperforms the naive 2-stage procedures.

***Index Terms*—** Quickest change detection, Nuisance change, Optimal Stopping Time, Recursive update, GLRT statistic

1. INTRODUCTION

Quickest change detection (QCD) is a fundamental problem in statistics. Given a sequence of independent and identically distributed (i.i.d.) observations $\{x_t : t \in \mathbb{N}\}$ with distribution f up to an unknown change point ν and are i.i.d. with distribution $g \neq f$ after. Subject to false alarm constraints, the goal is to detect this change as quickly as possible. Traditionally, applications of QCD can be found in manufacturing, in areas such as quality control [1, 2] where any change in the quality of products must be quickly detected. As the cost and size of modern-day sensors decreases, QCD methods have found applications in other areas such as fraud detection [3], cognitive radio [4], network surveillance [5–8], structural health monitoring [9], spam detection [10], bioinformatics [11], power system line outage detection [12], remote sensing [13],spectrum reuse [14], video segmentation [15] etc.

For the QCD problem where f and g are fully specified and the change point ν is unknown but deterministic, Page [16] developed the Cumulative Sum Control Chart (CuSum). Its optimality as the false alarm rate goes to zero was established by Lorden [17]. Later, Moustakides [18] showed that the CuSum test is exactly optimal under Lorden's optimality criterion. The CuSum test is also asymptoticaly optimal [19] under the Pollak's criterion [20], as the false alarm rate goes to zero. When g is not fully specified, Lorden [17] showed that the Generalized Likelihood Ratio (GLR) CuSum test is asymptotically optimal for the case of finite multiple post-change distributions. Other methods for the case when g is unknown to a certain degree were also proposed [19, 21–24]. We refer the reader to [25–27] and the references therein for an overview of the QCD problem. The papers [28–34] consider QCD of transient changes, where the change is either not persistent or multiple changes occur throughout the monitoring process. Unlike our QCD problem which allows some changes to be considered nuisance, the aforementioned papers consider any change to be critical.

In many practical applications, the signal of interest may undergo different types of changes. However, only a subset of these changes may be of interest to the user. One example is the problem of bearing failure detection using accelerometer readings [35]. During normal operations, the bearings are driven at two different activity levels, idle or active. In a typical bearing failure detection scenario, the bearing is initially be driven at the idle state. A change to drive it at the active state results in a change in the statistical properties of the accelerometer readings. However, this change is not of interest to us. We are only interested in the change arising from the failing of the bearing. Furthermore, the statistical properties of the observations obtained when the bearing is faulty depend on the activity level which it is driven at. We distinguish the changes which the signal undergoes using the concept of a nuisance change and a critical change. However, the traditional QCD framework does not allow us to distinguish between critical and nuisance changes.

In this paper, we address the QCD problem under the possibility of a nuisance change where both the nuisance and critical change-points are unknown but deterministic. We propose a window-limited stopping time which is able to identify the critical change quickly while ignoring the nuisance change.

A preliminary version of this work was presented in [36] where we developed an optimal sequential change detection procedure for the Bayesian formulation of the problem and proposed a stopping time which can be updated recursively for the non-Bayesian formulation of the problem. The asymptotic behavior of the GLRT stopping time [36] proposed for the non-Bayesian problem is difficult to derive as the test-statistic used is not a likelihood ratio and standard techniques cannot be applied to analyze its asymptotic behavior. To the best of the authors' knowledge, there are no other works that consider the QCD problem for a signal which may undergo a change that is not of interest.

The rest of this paper is organized as follows. In Section 2, we present our signal model and problem formulation. We propose the W-SGLR stopping time in Section 3 and derive its properties in Section 4. We present numerical simulations to illustrate the performance of our proposed stopping time in Section 5. We conclude in

*This research is supported by the Singapore Ministry of Education Academic Research Fund Tier 1 grant 2017-T1-001-059 (RG20/17).

Section 6.

2. PROBLEM FORMULATION

In this paper, we assume that the signals observed may be subjected to two types of change: a critical change at ν_c and a nuisance change at ν_n. Both the critical and nuisance change points are unknown *a priori*. We are interested in detecting the critical change while the nuisance change is not of interest. We consider the following signal model: let f, f_n, g, g_n be distinct distributions. At each time t, we let $h_{\nu_c,\nu_n,t}$ to be the distribution that generates the observation X_t when the nuisance change point is at ν_n and the critical change point is at ν_c:

$$h_{\nu_c,\nu_n,t} = \begin{cases} f & \text{if } t < \min\{\nu_c, \nu_n\}, \\ f_n & \text{if } \nu_n \leq t < \nu_c, \\ g & \text{if } \nu_c \leq t < \nu_n, \\ g_n & \text{if } \max\{\nu_c, \nu_n\} \leq t. \end{cases} \tag{1}$$

Let $X_1, X_2, \ldots$ be a sequence of independent random variables satisfying $X_t \sim h_{\nu_c,\nu_n,t}$ where $\nu_n, \nu_c \geq 0$ are the unknown but deterministic nuisance and critical change-points respectively. The quickest change detection problem is to detect the critical change ν_c through observing $X_1 = x_1, X_2 = x_2, \ldots$, as quickly as possible while keeping the false alarm rate low. In our signal model, the nuisance change point also affects the distribution which generates the observations after the critical change point. This creates a dependence between the nuisance change point and the distribution after the critical change point. While our QCD problem is to detect a change in distribution from either f or f_n to either g or g_n as quickly as possible, our formulation is different from assuming composite pre-change and post-change distribution families [37] since the nuisance change causes the distribution of X_t to be non-stationary before or after the critical change, depending on whether the nuisance change occurs before or after the critical change, respectively.

In a typical sequential change detection procedure, at each time t, a test statistic $S(t)$ is computed based on the currently available observations $X_1 = x_1, \ldots, X_t = x_t$, and the observer decides that a change has occurred at a stopping time τ where τ is the first t such that $S(t)$ exceeds a pre-determined threshold b:

$$\tau(b) = \inf\{t : S(t) > b\}.$$

Our QCD problem can be formulated as a minimax problem similarly as Lorden's formulation [17] where we seek a stopping time that minimizes the WADD subject to an average run length (ARL) constraint:

$$\begin{aligned} \underset{\tau}{\text{minimize}} \quad & \text{WADD}(\tau) \\ \text{subject to} \quad & \text{ARL}(\tau) \geq \gamma. \end{aligned} \tag{2}$$

where τ is a stopping time with respect to the filtration $\{\sigma(X_1^t)\}_{t\geq 0}$,

$$\text{WADD}(\tau) = \sup_{\nu_c,\nu_n \geq 1} \operatorname{ess\,sup} \mathbb{E}_{\nu_c,\nu_n} \left[(\tau - \nu_c + 1)^+ \,\middle|\, X_1^{\nu_c - 1} \right],$$

$\text{ARL}(\tau) = \inf_{\nu_n \geq 1} \mathbb{E}_{\infty,\nu_n}[\tau]$, ess sup is the essential supremum operator and $\mathbb{E}_{\nu_c,\nu_n}$ is the expectation operator assuming the critical change-point is at ν_c and nuisance change-point is at ν_n. In the next section, we propose a stopping time for (2).

3. TEST STATISTICS FOR QUICKEST CHANGE DETECTION

Suppose that we observe the sequence $X_1 = x_1, X_2 = x_2, \ldots$ sequentially and it is known that the nuisance change does not take place (i.e., $\nu_n = \infty$), then Page's CuSum test statistic [16] is given as

$$S_{\text{CS}}(t) = \max_{1\leq k\leq t+1} \sum_{i=k}^{t} \log \frac{g(x_i)}{f(x_i)} = \log \max_{1\leq k\leq t+1} \prod_{i=k}^{t} \frac{g(x_i)}{f(x_i)}$$

and we declare that a critical change has taken place at τ_{CS} when the CuSum test statistic first exceeds a pre-determined threshold b where,

$$\begin{aligned} \tau_{\text{CS}} &= \inf\{t : S_{\text{CS}}(t) > b\} && (3) \\ &= \inf\left\{ t : \max_{1\leq k\leq t+1} \prod_{i=k}^{t} \frac{g(x_i)}{f(x_i)} > e^b \right\}. && (4) \end{aligned}$$

Based on the stopping time (4), Page's CuSum stopping time can be seen as a repeated application of a one-sided sequential probability ratio test. As there are t sequential probability ratio tests to run at time t, a naive implementation of the stopping time τ_{CS} will require $O(t^2)$ operations to compute $S_{\text{CS}}(t)$ for each t. Fortunately, the CuSum test statistic has a convenient recursion $S_{\text{CS}}(t+1) = \max\left\{ S_{\text{CS}}(t) + \log \frac{g(x_t)}{f(x_t)}, 0 \right\}$ that allows the CuSum stopping time to be implemented efficiently.

If the nuisance change takes places at a time $\nu_n < \infty$ and ν_n is known, a modification of Page's test statistic gives the following:

$$S_{\text{CS}}(t) = \log \max_{1\leq k\leq t+1} \prod_{i=k}^{t} \frac{h_{0,\nu_n,i}(x_i)}{h_{\infty,\nu_n,i}(x_i)}, \tag{5}$$

where $h_{0,\nu_n,i}(x)$ and $h_{\infty,\nu_n,i}(x)$ are as defined in (1). Similar to the case where $\nu_n = \infty$, the CuSum test statistics admits a convenient recursion for efficient implementation. Furthermore, for both the cases mentioned above, τ_{CS} is known to be asymptotically optimal [17].

In our problem formulation, the nuisance change-point ν_n is unknown. Replacing ν_n with its maximum likelihood estimator in both the numerator and denominator, we obtain the following Generalized Likelihood Ratio (GLR) test statistic

$$\begin{aligned} \Lambda_{\text{GLR}}(k,t) &= \frac{\max_{k\leq j\leq t+1} \prod_{i=k}^{t} h_{0,j,i}(x_i)}{\max_{k\leq j\leq t+1} \prod_{i=k}^{t} h_{\infty,j,i}(x_i)}, && (6) \\ S_{\text{GLR}}(t) &= \log \max_{1\leq k\leq t+1} \Lambda_{\text{GLR}}(k,t). && (7) \end{aligned}$$

While the GLR stopping time is commonly used in practical applications, its average run length is challenging to study theoretically since the GLR test statistic $S_{\text{GLR}}(k,t)$ is not a likelihood ratio and standard techniques in Theorem 6.16 of [27] cannot be used to analyze its average run length. In order to develop a stopping time with ARL that can be studied theoretically, we simplify the maximum likelihood estimation at the numerator to only consider two cases $j = k$ and $j = t + 1$. This gives us the Simplified GLR (SGLR) test statistic

$$\begin{aligned} \Lambda_{\text{SGLR}}(k,t) &= \frac{\max\left\{ \prod_{i=k}^{t} g(x_i), \prod_{i=k}^{t} g_n(x_i) \right\}}{\max_{k\leq j\leq t+1} \prod_{i=k}^{t} h_{\infty,j,i}(x_i)} && (8) \\ S_{\text{SGLR}}(t) &= \log \max_{1\leq k\leq t+1} \Lambda_{\text{SGLR}}(k,t). && (9) \end{aligned}$$

As critical changes are typically expected to occur rarely, a large number of samples are expected to be observed before the critical change. Unlike the CuSum test statistic, the SGLR test statistic does not have a convenient recursion. Any implementation of the SGLR stopping time would require at least $O(t)$ operations to compute $S_{\text{SGLR}}(t)$ for each t. This requirement on computational resources is a significant limitation for many practical applications. In order to control the computational resources required, we propose the Window-Limited SGLR (W-SGLR) test statistic and stopping time following the ideas presented in [19],

$$S_{\text{W-SGLR}}(t) = \log \max_{t-m_b \leq k \leq t+1} \Lambda_{\text{SGLR}}(k,t), \tag{10}$$

$$\tau_{\text{W-SGLR}} = \inf \{t \;:\; S_{\text{W-SGLR}}(t) > b\} \tag{11}$$

where the window size m_b satisfies $\liminf_{b\to\infty} \dfrac{m_b}{b} > I^{-1}$, $\log m_b = o(b)$ such that $I = \min\left\{\mathbb{E}_g\left[\log \frac{g(X)}{f(X)}\right], \mathbb{E}_g\left[\log \frac{g(X)}{f_n(X)}\right], \mathbb{E}_{g_n}\left[\log \frac{g_n(X)}{f(X)}\right], \mathbb{E}_{g_n}\left[\log \frac{g_n(X)}{f_n(X)}\right]\right\}$. We note that for each t, it requires $O(m_b^3)$ operations to naively compute $S_{\text{W-SGLR}}(t)$. Window-limited test statistics were first introduced by Willsky and Jones [19]. Lai further discusses their properties and the choice of window size and thresholds in [19]. For the rest of this paper, we require the following assumption:

Assumption 1. *The first four moments with respect to both g and g_n of $\log \frac{f_n(X)}{f(X)}$ exists.*

4. PROPERTIES OF THE W-SGLR STOPPING TIME

In this section, we derive a recursive update scheme for the $S_{\text{W-SGLR}}$ and present the asymptotic properties of $S_{\text{W-SGLR}}$ and $\tau_{\text{W-SGLR}}$. First, we define the following generalized likelihood ratios

$$\Lambda(k,t) = \frac{\prod_{i=k}^{t} g(x_i)}{\max_{k\leq j\leq t+1} \prod_{i=k}^{j-1} f(x_i) \prod_{i=j}^{t} f_n(x_i)},$$
$$\Lambda_n(k,t) = \frac{\prod_{i=k}^{t} g_n(x_i)}{\max_{k\leq j\leq t+1} \prod_{i=k}^{j-1} f(x_i) \prod_{i=j}^{t} f_n(x_i)}$$

and we have $\Lambda_{\text{SGLR}}(k,t) = \max\{\Lambda(k,t), \Lambda_n(k,t)\}$. Next, we present a lemma describing the recursive properties of Λ and Λ_n:

Lemma 1. *For any $k, t \in \mathbb{N}$ such that $0 \leq k \leq t$, we have the following recursion*

$$\Lambda(k,t+1) = \min\left\{\Lambda(k,t)\tfrac{g(x_{t+1})}{f_n(x_{t+1})}, \left(\prod_{i=k}^{t} \tfrac{g(x_i)}{f(x_i)}\right)\tfrac{g(x_{t+1})}{f(x_{t+1})}\right\},$$
$$\Lambda_n(k,t+1) = \min\left\{\Lambda_n(k,t)\tfrac{g_n(x_{t+1})}{f_n(x_{t+1})}, \left(\prod_{i=k}^{t} \tfrac{g_n(x_i)}{f(x_i)}\right)\tfrac{g_n(x_{t+1})}{f(x_{t+1})}\right\}.$$

Proof. Let $m = \arg\max_{k\leq j\leq t+2} \prod_{i=k}^{j-1} f(x_i) \prod_{i=j}^{t+1} f_n(x_i)$, we consider the possible values of m. If $m = t+2$, then $\Lambda(k,t+1) = \prod_{i=k}^{t+1} \frac{g(x_i)}{f(x_i)}$. If $m < t+2$, we obtain

$$\begin{aligned}\Lambda(k,t+1) &= \frac{\prod_{i=k}^{t+1} g(x_i)}{\max_{k\leq j\leq t+2} \prod_{i=k}^{j-1} f(x_i) \prod_{i=j}^{t+1} f_n(x_i)} \\ &= \frac{\prod_{i=k}^{t} g(x_i)}{\max_{k\leq j\leq t+2} \prod_{i=k}^{j-1} f(x_i) \prod_{i=j}^{t} f_n(x_i)} \frac{g(x_{t+1})}{f_n(x_{t+1})} \\ &= \Lambda(k,t)\frac{g(x_{t+1})}{f_n(x_{t+1})}.\end{aligned}$$

Putting everything together, we obtain

$$\Lambda(k,t+1) = \min\left\{\Lambda(k,t)\tfrac{g(x_{t+1})}{f_n(x_{t+1})}, \left(\prod_{i=k}^{t} \tfrac{g(x_i)}{f(x_i)}\right)\tfrac{g(x_{t+1})}{f(x_{t+1})}\right\}.$$

The proof for $\Lambda_n(k,t+1)$ is similar. The proof is now complete. □

In order to derive a recursive implementation of $S_{\text{W-SGLR}}(t)$, we introduce the following:

$$\alpha(t,j) = \Lambda\left(t - m_b - j + 1, t\right), \quad \alpha_n(t,j) = \Lambda_n\left(t - m_b - j + 1, t\right),$$
$$\beta(t,j) = \prod_{i=t-(m_b-j+1)}^{t} \frac{g(x_i)}{f(x_i)}, \quad \beta_n(t,j) = \prod_{i=t-(m_b-j+1)}^{t} \frac{g_n(x_i)}{f(x_i)},$$

for $j \in 1,\ldots,m_b+1$ and $\alpha(t,m_b+2) = \alpha_n(t,m_b+2) = \beta(t,m_b+2) = \beta_n(t,m_b+2) = 1$. By expressing the W-SGLR test statistic in terms of $\alpha(t,k), \alpha_n(t,k)$ and using Lemma 1, we obtain the following proposition.

Proposition 1. *The W-SGLR test statistic can be computed using*

$$S_{\text{W-SGLR}}(t) = \log \max_{1\leq k\leq m_b+2} \max\{\alpha(t,k), \alpha_n(t,k)\}$$

where $\alpha(t,k), \alpha_n(t,k)$ can be computed using the following recursion for $k \in 1,\ldots,m_b+1$:

$$\beta(t+1,j) = \beta(t,j)\frac{g(x_{t+1})}{f(x_{t+1})},$$
$$\alpha(t+1,j) = \min\left\{\alpha(t,j-1)\frac{g(x_{t+1})}{f(x_{t+1})}, \beta(t+1,j)\right\},$$
$$\beta_n(t+1,j) = \beta_n(t,j)\frac{g_n(x_{t+1})}{f(x_{t+1})},$$
$$\alpha_n(t+1,j) = \min\left\{\alpha_n(t,j-1)\frac{g_n(x_{t+1})}{f(x_{t+1})}, \beta(t+1,j)\right\}.$$

As compared to a naive implementation of the W-SGLR stopping time, an implementation using Proposition 1 requires only $O(m_b)$ operations to compute $S_{\text{W-SGLR}}(t)$ for each t.

Since the stopping time $\tau_{\text{W-SGLR}}$ is defined by the first time the test statistic $S_{\text{W-SGLR}}$ crosses the threshold b, the rate of growth of the statistic $\log\Lambda$ and $\log\Lambda_n$ would allow us to understand the increase in detection delay as b increases. Next, we will show that rate of growth $\frac{1}{t-k+1}\log\Lambda(k,t)$ and $\frac{1}{t-k+1}\log\Lambda_n(k,t)$ converge, to different limits, in probability as t tends to infinity. In particular, the limit which the rate of growth converges to depends on the sign of $\rho_g = \mathbb{E}_g[\log\frac{f_n(X)}{f(X)}]$ and $\rho_{g_n} = \mathbb{E}_{g_n}[\log\frac{f_n(X)}{f(X)}]$ where $\mathbb{E}_h$ denotes the expectation operator assuming X has distribution h.

Theorem 1. *For any $\nu_c \leq k < \infty$ and $\nu_n < \infty$, we have the following convergence in probability as summarized in the table below*

Sign of ρ_g, ρ_{g_n}		*under $\mathbb{P}_{\infty,\nu_c}$*	*under $\mathbb{P}_{\nu_n,\nu_c}$*
ρ_g	ρ_{g_n}	$\frac{\log\Lambda(k,t)}{t-k+1}$ *converges to*	$\frac{\log\Lambda_n(k,t)}{t-k+1}$ *converges to*
>0	*>0*	$\mathbb{E}_g\left[\log\frac{g(X)}{f_n(X)}\right]$	$\mathbb{E}_{g_n}\left[\log\frac{g_n(X)}{f_n(X)}\right]$
>0	*<0*	$\mathbb{E}_g\left[\log\frac{g(X)}{f_n(X)}\right]$	$\mathbb{E}_{g_n}\left[\log\frac{g_n(X)}{f(X)}\right]$
<0	*>0*	$\mathbb{E}_g\left[\log\frac{g(X)}{f(X)}\right]$	$\mathbb{E}_{g_n}\left[\log\frac{g_n(X)}{f_n(X)}\right]$
<0	*<0*	$\mathbb{E}_g\left[\log\frac{g(X)}{f(X)}\right]$	$\mathbb{E}_{g_n}\left[\log\frac{g_n(X)}{f(X)}\right]$

where $\mathbb{P}_{\nu_n,\nu_c}$ is probability distribution on $\{X_i\}_{i\in\mathbb{N}}$ assuming the critical change-point is at ν_c and nuisance change-point is at ν_n.

Heuristically, this implies that $\log \Lambda$ and $\log \Lambda_n$ grows linearly with respect to t and that the WADD grows linearly with respect to the threshold b. The next theorem states these results rigorously.

Theorem 2. *For any $\nu_n \in \mathbb{N} \cup \{\infty\}$, we have*

$$\begin{aligned} ARL_{\nu_n}[\tau_{\text{W-SGLR}}] &= \mathbb{E}_{\nu_n,\infty}[\tau_{\text{W-SGLR}}] \geq \frac{1}{2}e^b, \\ WADD_{\nu_n}[\tau_{\text{W-SGLR}}] &= \sup_{\nu_c \geq 1} \operatorname{ess\,sup} \mathbb{E}_{\nu_n,\nu_c}\left[(\tau_{\text{W-SGLR}} - \nu_c + 1)^+ \,\middle|\, X_1^{\nu_c - 1}\right] \\ &\leq (I^{-1} + o(1))b. \end{aligned}$$

Furthermore,

$$\begin{aligned} ARL[\tau_{\text{W-SGLR}}] &\geq \frac{1}{2}e^b, \\ WADD[\tau_{\text{W-SGLR}}] &\leq (I^{-1} + o(1))b, \end{aligned}$$

and if $\rho_g < 0$, the stopping time $\tau_{\text{W-SGLR}}$ is asymptotically optimal for (2)*.*

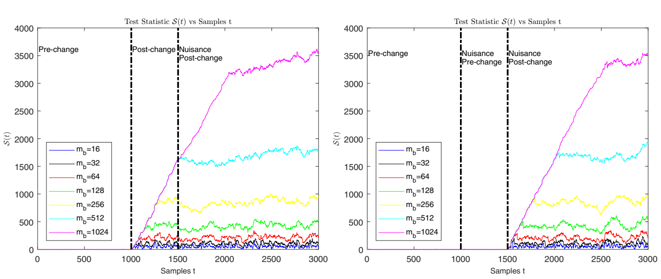
The interested reader may refer to [38] for the proofs for Theorem 1 and 2 which we omit due to space constraints.

In the next section, we present numerical results from simulations to illustrate the performance of the proposed stopping time.

5. NUMERICAL RESULTS

We denote $\mathcal{N}(\mu, \sigma^2)$ as the normal distribution with mean μ and variance σ^2. In our first set of simulations, we let $f = \mathcal{N}(0, 1)$, $g = \mathcal{N}(0, 10)$, $f_n = \mathcal{N}(2, 1)$ and $g_n = \mathcal{N}(2, 10)$ where the critical change is a change in variance and the nuisance change is a change in mean. We ran the simulations with two change-point configurations to illustrate the behaviour of the W-SGLR test statistic for different window-sizes. In Fig. 1(a), we set $\nu_c = 1000$, $\nu_n = 1500$ and in Fig. 1(b), we set $\nu_c = 1500$, $\nu_n = 1000$. In Fig. 1(a) and (b), the test statistic remains low before the critical change-point and grows linearly with the gradient $I = 3.33487$ after the critical change-point. This trend continues until the statistics approximately achieves the value of $I \times m_b = 3.33487 m_b$. In Fig. 1(a) when $m_b = 1024$, we note that the test statistic continues to grow linearly with the gradient $I = 3.33487$ even after the nuisance change point. In Fig. 1(b), we note that the test statistic continues to remain low during the period between the nuisance change point and critical.

Next, we compare the stopping time with a naive 2-stage stopping time. The naive stopping time $\tau_{\text{2-stage}}$ is constructed from stopping times based on the optimal CuSum stopping times described



(a) (b)

Fig. 1. Examples of W-SGLR test statistics $\mathcal{S}(t)$ as a function of t when (a) $\nu_c = 1000$, $\nu_n = 1500$ and (b) $\nu_c = 1500$, $\nu_n = 1000$ with $f = \mathcal{N}(0, 1)$, $f_n = \mathcal{N}(2, 1)$, $g = \mathcal{N}(0, 10)$, $g_n = \mathcal{N}(2, 10)$.

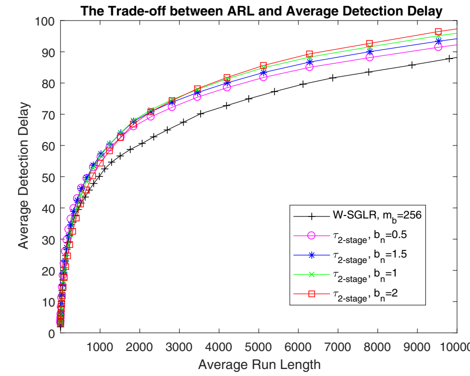


Fig. 2. Comparison of trade-off between the ARL and the average detection delay for the proposed stopping time $\tau_{\text{W-SGLR}}$ and naive stopping time $\tau_{\text{2-stage}}$. The critical and nuisance change points were chosen randomly on the signal of length 2^{16}

in (4) with $\tau_{\{p \to q, b\}} = \inf\left\{t : \max_{1 \leq k \leq t+1} \prod_{i=k}^{t} \frac{p(x_i)}{q(x_i)} > e^b\right\}$ for any pair of distribution p and $q \neq p$ and threshold b. There are four stopping times required $\tau_{f \to f_n, b_n}$, $\tau_{f \to g, b_c}$, $\tau_{f \to g_n, b_c}$, and $\tau_{f_n \to g_n, b_c}$ where the threshold for declaring critical change is b_c and threshold for declaring a nuisance change is b_n. The 2-stage CuSum $\tau_{\text{2-stage}}$ can be described as follows. In the first stage, we apply the stopping times $\tau_{\{f \to g, b_c\}}$, $\tau_{\{f \to g_n, b_c\}}$ and $\tau_{\{f \to f_n, b_n\}}$ to the observations. If $\tau_{\{f \to g, b_c\}}$ or $\tau_{\{f \to g_n, b_c\}}$ stops the process before $\tau_{\{f \to f_n, b_n\}}$, we declare that a critical change has occurred and $\tau_{\text{2-stage}} = \min\{\tau_{\{f \to g, b_c\}}, \tau_{\{f \to g_n, b_c\}}\}$. Otherwise, we apply $\tau_{\{f_n \to g_n, b_c\}}$ to the rest of the observations after the stopping time $\tau_{\{f \to f_n, b_n\}}$ and set $\tau_{\text{2-stage}} = \tau_{\{f_n \to g_n, b_c\}}$. We compare the trade-off between the average run length and the average detection delay of the proposed W-SGLR stopping time against the 2-stage stopping time for varying nuisance change thresholds in Fig. 2. We observe that our proposed W-SGLR stopping achieves a lower average detection delay as compared to each of the 2-stage stopping time for large average run lengths.

6. CONCLUSION

We have studied the non-Bayesian QCD problem where the signal may be subjected to a nuisance change. We proposed a window-limited stopping time $\tau_{\text{W-SGLR}}$ that quickly detects the critical change while ignoring the nuisance change. We derived the stopping time's asymptotic behavior and the condition when the proposed stopping time is asymptotically optimal. Numerical simulations illustrate the behavior of the W-SGLR test statistic for different window sizes and suggest that the W-SGLR stopping time outperforms the naive 2-stage stopping time for large average run lengths.

7. REFERENCES

[1] W. H. Woodall, D. J. Spitzner, D. C. Montgomery, and S. Gupta, "Using control charts to monitor process and product quality profiles," *J. of Quality Technology*, vol. 36, no. 3, p. 309, 2004.

[2] T. L. Lai, "Sequential changepoint detection in quality control and dynamical systems," *J. of the Roy. Statistical Soc.*, pp. 613–658, 1995.

[3] R. J. Bolton and D. J. Hand, "Statistical fraud detection: A review," *Statistical Sci.*, pp. 235–249, 2002.

[4] L. Lai, Y. Fan, and H. V. Poor, "Quickest detection in cognitive radio: A sequential change detection framework," in *IEEE Conf. Global Telecommun.* IEEE, 2008, pp. 1–5.

[5] L. Akoglu and C. Faloutsos, "Event detection in time series of mobile communication graphs," in *Proc. Army Sci. Conf.*, 2010, pp. 77–79.

[6] K. Sequeira and M. Zaki, "ADMIT: anomaly-based data mining for intrusions," in *Proc. Conf. Knowl. Discovery and Data Mining*. ACM, 2002, pp. 386–395.

[7] A. G. Tartakovsky, B. L. Rozovskii, R. B. Blazek, 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, 2006.

[8] A. G. Tartakovsky, "Rapid detection of attacks in computer networks by quickest change-point detection methods," *Data anal. for network cyber-security*, pp. 33–70, 2014.

[9] H. Sohn, J. A. Czarnecki, and C. R. Farrar, "Structural health monitoring using statistical process control," *J. Structural Eng.*, vol. 126, no. 11, pp. 1356–1363, 2000.

[10] S. Xie, G. Wang, S. Lin, and P. S. Yu, "Review spam detection via temporal pattern discovery," in *Proc. Conf. Knowl. Discovery and Data Mining*. ACM, 2012, pp. 823–831.

[11] V. M. R. Muggeo and G. Adelfio, "Efficient change point detection for genomic sequences of continuous measurements," *Bioinformatics*, vol. 27, no. 2, pp. 161–166, 2011.

[12] T. Banerjee, Y. C. Chen, A. D. Dominguez-Garcia, and V. V. Veeravalli, "Power system line outage detection and identification-A quickest change detection approach," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process.* IEEE, 2014, pp. 3450–3454.

[13] P. R. Coppin and M. E. Bauer, "Digital change detection in forest ecosystems with remote sensing imagery," *Remote sensing reviews*, vol. 13, no. 3-4, pp. 207–234, 1996.

[14] N. Kundargi and A. Tewfik, "A nonparametric sequential Kolmogorov-Smirnov test for transmit opportunity detection at the MAC layer," in *Proc. IEEE Workshop on Signal Proc. Advances in Wireless Commun.*, June 2009, pp. 101–105.

[15] Z. Li, H. Lu, and Y.-P. Tan, "Video segmentation based on sequential change detection," in *Proc. IEEE Int. Conf. on Multimedia and Expo*, vol. 3, June 2004, pp. 1955–1958 Vol.3.

[16] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.

[17] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Stat.*, pp. 1897–1908, 1971.

[18] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Stat.*, pp. 1379–1387, 1986.

[19] T. L. Lai, "Information bounds and quick detection of parameter changes in stochastic systems," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2917–2929, 1998.

[20] M. Pollak, "Optimal detection of a change in distribution," *Ann. Stat.*, pp. 206–227, 1985.

[21] D. Siegmund and E. S. Venkatraman, "Using the generalized likelihood ratio statistic for sequential detection of a change-point," *Ann. Stat.*, pp. 255–271, 1995.

[22] T. Banerjee and V. V. Veeravalli, "Data-efficient minimax quickest change detection with composite post-change distribution," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 5172–5184, 2015.

[23] T. S. Lau, W. P. Tay, and V. V. Veeravalli, "Quickest change detection with unknown post-change distribution," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process.* IEEE, 2017, pp. 3924–3928.

[24] T. S. Lau and W. P. Tay, "Optimal sampling policy for quickest change detection," in *Proc. IEEE Global Conf. Signal and Inform. Process.* IEEE, 2017, pp. 78–82.

[25] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential analysis: Hypothesis testing and changepoint detection*. CRC Press, 2014.

[26] V. V. Veeravalli and T. Banerjee, "Quickest change detection," *Academic press library in signal process.: Array and statistical signal process.*, vol. 3, pp. 209–256, 2013.

[27] H. V. Poor and O. Hadjiliadis, *Quickest detection*. Cambridge University Press Cambridge, 2009, vol. 40.

[28] J. Heydari and A. Tajer, "Quickest search for a change point," *arXiv preprint arXiv:1804.09657*, 2018.

[29] E. Ebrahimzadeh and A. Tchamkerten, "Sequential detection of transient changes in stochastic systems under a sampling constraint," in *Proc. IEEE Int. Symp. Inform. Theory*. IEEE, 2015, pp. 156–160.

[30] B. K. Guépié, L. Fillatre, and I. Nikiforov, "Sequential detection of transient changes," *Sequential Analysis*, vol. 31, no. 4, pp. 528–547, 2012.

[31] G. V. Moustakides and V. V. Veeravalli, "Sequentially detecting transitory changes," in *Proc. IEEE Int. Symp. Inform. Theory*, July 2016, pp. 26–30.

[32] S. Zou, G. Fellouris, and V. V. Veeravalli, "Quickest change detection under transient dynamics: Theory and asymptotic analysis," *IEEE Trans. Inf. Theory*, 2018.

[33] V. Krishnamurthy, "Bayesian sequential detection with phase-distributed change time and nonlinear penalty-a POMDP lattice programming approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 7096–7124, Oct 2011.

[34] ——, "Quickest detection pomdps with social learning: Interaction of local and global decision makers," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5563–5587, Aug 2012.

[35] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mechanical Syst. and Signal Process.*, vol. 64, pp. 100–131, 2015.

[36] T. S. Lau and W. P. Tay, "Quickest change detection under a nuisance change," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process.* IEEE, 2018, pp. 6643–6647.

[37] Y. Mei, "Asymptotically optimal methods for sequential change-point detection," Ph.D. dissertation, California Institute of Technology, 2003.

[38] T. S. Lau and W. P. Tay, "Asymptotically optimal quickest change detection under a nuisance change," 2018. [Online]. Available: https://drive.google.com/file/d/1GOdIJvxE70Sgyl7qKSQFTLCgrvi7oNXI/view?usp=sharing