Asymptotic Optimality of D-CuSum for Quickest Change Detection under Transient Dynamics

Shaofeng Zou* Georgios Fellouris*† Venugopal V. Veeravalli*♮†
*Coordinated Science Lab, ♮ ECE Department, †Department of Statistics
University of Illinois at Urbana-Champaign
Email: {szou3, fellouri, vvv}@illinois.edu

***Abstract*—The problem of quickest change detection (QCD) under transient dynamics is studied, in which the change from the initial distribution to the final persistent distribution does not happen instantaneously, but after a series of cascading transient phases. It is assumed that the durations of the transient phases are deterministic but unknown. The goal is to detect the change as quickly as possible subject to a constraint on the average run length to false alarm. The dynamic CuSum (D-CuSum) algorithm is investigated, which is based on reformulating the QCD problem into a dynamic composite hypothesis testing problem, and has a recursion that facilitates implementation. We show that this algorithm is adaptive to the unknown change point, as well as the unknown transient duration. And under mild conditions of the pre-change and post-change distributions, its asymptotic optimality is demonstrated for all possible asymptotic regimes as the transient duration and the average run length to false alarm go to infinity.**

I. Introduction

The quickest change detection (QCD) problem has a wide range of applications, including but not limited to biomedical signal and image processing, quality control, surveillance system and target detection. In this problem, it is assumed that at an unknown time an abrupt change occurs and the statistical behavior of the observations changes from their typical pattern. The goal is to design algorithms that can detect such a change as quickly as possible subject to false alarm constraints.

In classical QCD problems [1]–[3], the statistical behavior of a system is characterized by one pre-change distribution and one post-change distribution that generate the data before and after the change, respectively. After an event happens, the statistical behavior of the observations changes from the pre-change distribution to the post-change distribution instantaneously. However, there are many practical applications with more involved statistical behaviors after the change. For example, in power systems, if a line outage occurs, the system goes through a transient phase before entering a persistent phase [4]. Therefore, the statistical behavior of the system after the change is characterized by multiple distributions, corresponding to the transient phase and the persistent phase, and these two phases are cascaded temporally. The goal is to detect the change from the initial phase to the transient phase as quickly as possible.

These kinds of applications motivated us to study the problem of QCD under transient dynamics. In this problem, the initial distribution does not change to the persistent distribution instantaneously, but after a series of cascading transient phases, where the observations within each phase are generated by a different distribution. The goal is to detect the change from the initial distribution to the first transient phase as quickly as possible.

This problem is fundamentally different from the problem of detecting the transient changes in [5] and [6], in which the system goes back to the pre-change distribution after a single transient phase, and the goal is to detect the change with high probability within the transient phase. In our problem, the system changes to a distinct persistent distribution after a series of cascading transient phases. Therefore, the samples after the transient phases still provide useful information about whether the change has occurred. Furthermore, we aim to detect the change as quickly as possible but not necessarily within the transient phases. Thus, a decision to stop after the transients is still valid for our problem.

The proposed QCD problem with transient post-change dynamics was studied in [7] in the special case that it is known in advance that there is only one transient phase, with duration one. For this problem, a generalization of Page's CuSum algorithm [8] was proposed and shown to be exactly optimal under Lorden's minimax criterion [9]. A generalization of the Shiryaev-Roberts [10] test was proposed in [11], under the assumption that the number of transient phases is known (not necessarily equal to one), and the durations of the transient phases are geometrically distributed, a convenient assumption that leads to a recursive form for the detection statistic. However, this scheme was not supported by a theoretical analysis.

In the present paper, the number of transient phases is still considered to be known in advance, but we now assume that the durations of the transient phases are deterministic, *can take any possible integer value, and are completely unknown*. From this point of view, the QCD problem with transient dynamics can be considered as a special case of the QCD problem with composite post-change distributions, in which the post-change distribution is indexed by the durations of the transient phases. Previous studies for the QCD problem under a composite setting [12]–[14] usually assume that the distributions can be embedded to an exponential family, or that the parameter space of the post-change distribution is compact. Hence, these algorithms and their analysis cannot be directly

978-1-5090-4096-4/17/$31.00 ©2017 IEEE

applied to our problem.

In this paper, we study a generalization of the classical CuSum algorithm that was proposed in [4], and to which we will refer as the *dynamic CuSum (D-CuSum) algorithm*. This algorithm is derived by formulating the QCD problem as a dynamic composite hypothesis testing problem. Most importantly, its detection statistic admits a recursion, thus facilitating its implementation. Furthermore, numerical results in [4] demonstrate that this algorithm is adaptive to different durations of the transients. However, no theoretical analysis of the D-CuSum algorithm was provided in [4]. Our goal in this paper is to provide a theoretical justification for the D-CuSum algorithm. We focus on the special case with one transient phase and demonstrate the asymptotic optimality of D-CuSuM algorithm under mild conditions of the pre-change and post-change distributions as the average run length to false alarm and the transient duration go to infinity at any relative rate.

II. Problem Model

Consider a process $\{X_i\}_{i=1}^{\infty}$, observed sequentially by a decision maker. At an unknown point v_1, an event occurs and $\{X_i\}_{i=v_1}^{\infty}$ undergo a change in distribution from the initial distribution f_0. It is assumed that this change goes through $L-1$ transient phases with unknown durations before entering a persistent phase. Each phase k, for $1 \leq k \leq L$, is associated with an unknown starting point v_k, and the observations within this phase are generated by a known distribution f_k. The duration of phase k is denoted by $D_k = v_{k+1} - v_k$, which is unknown. Conditioned on the change points $v_1, \ldots, v_L$, the observations $\{X_i\}_{i=1}^{\infty}$ are assumed to be independent. More specifically, the observations are distributed as follows:

$$X_i \sim f_k, \text{ if } v_k \leq i < v_{k+1}, \tag{1}$$

for $0 \leq k \leq L$, where $v_0 = 1$, and $v_{L+1} = \infty$.

We use $\mathbb{P}_{v_1}$ to denote the probability measure with the first change at v_1, and $\mathbb{E}_{v_1}$ to denote the corresponding expectation. Then $\mathbb{P}_\infty$, $\mathbb{E}_\infty$ stand for the probability measure and expectation when $v_1 = \infty$, i.e., the change does not occur. For any stopping time τ, we define the Average Run Length (ARL) to false alarm and the worst-case Average Detection Delay (ADD) under Pollak's criterion [10] as follows:

$$\mathrm{ARL}(\tau) = \mathbb{E}_\infty[\tau], \tag{2}$$

$$J_{\mathrm{P}}(\tau) = \sup_{v_1 \geq 1} \mathbb{E}_{v_1}[(\tau - v_1)^+ | \tau \geq v_1], \tag{3}$$

where $(\tau - v_1)^+ = \max\{\tau - v_1, 0\}$.

The goal is to minimize $J_{\mathrm{P}}(\tau)$ subject to a constraint on ARL:

$$\underset{\tau:\mathrm{ARL}(\tau)\geq\gamma}{\text{minimize}}\ J_{\mathrm{P}}(\tau). \tag{4}$$

We note that our results also hold under Lorden's criterion [9] where the worst-case ADD is defined as

$$J_{\mathrm{L}}(\tau) = \sup_{v_1 \geq 1} \operatorname{ess\,sup} \mathbb{E}_{v_1}[(\tau - v_1)^+ | X_1, \ldots, X_{v_1-1}]. \tag{5}$$

We use $I_j = \int f_j \log \frac{df_j}{df_0}$ to denote the Kullback-Leibler (KL) divergence between f_j and f_0, which we assume to be positive and finite, $j = 1, \ldots, L$. Moreover, we use $Z_j(X_i) = \log \frac{f_j(X_i)}{f_0(X_i)}$ to denote the log likelihood ratio between f_j and f_0 for the sample X_i, $j = 1, \ldots, L$, $i = 1, 2, \ldots$.

III. The D-CuSum Algorithm

The D-CuSum algorithm is derived by formulating the QCD problem as a dynamic composite hypothesis testing problem at each time instant [4], which is to distinguish between the following two hypotheses at each time k:

$$\mathcal{H}_0^k : k < v_1, \tag{6}$$

$$\mathcal{H}_1^k : k \geq v_1. \tag{7}$$

The hypothesis $\mathcal{H}_0^k$ corresponds to the case that at time k the change from f_0 to f_1 has not occurred yet. Thus, under $\mathcal{H}_0^k$ the samples $X_1, \ldots, X_k$ are distributed according to f_0. The alternative hypothesis $\mathcal{H}_1^k$ corresponds to the case that at time k, the change from f_0 to f_1 has occurred. Most importantly, $\mathcal{H}_1^k$ is a composite hypothesis, since it depends on the unknown values of $v_1, \ldots, v_L$.

The D-CuSum detection statistic at time k is the generalized likelihood ratio for the above testing problem

$$W[k] = \max_{v_1 \leq v_2 \leq \cdots \leq v_L} \left\{ \sum_{j=v_1}^{\min\{v_2-1,k\}} \log \frac{f_1(X_j)}{f_0(X_j)} + \cdots + \sum_{j=\min\{v_L,k+1\}}^{k} \log \frac{f_L(X_j)}{f_0(X_j)} \right\}. \tag{8}$$

It was shown in [4] that $W[k]$ has a recursive structure:

$$W[k] = \max\left\{\Omega^1[k], \Omega^2[k], \ldots, \Omega^L[k]\right\}, \tag{9}$$

where for each $1 \leq i \leq L$ we define

$$\Omega^i[k] = \max\left\{\Omega^1[k-1], \ldots, \Omega^i[k-1], 0\right\} + Z_i(X_k). \tag{10}$$

The corresponding stopping time is given by comparing $W[k]$ against a pre-determined positive threshold:

$$\tau(b) = \inf\{k \geq 1 : W[k] > b\}. \tag{11}$$

An important observation for the analysis of this stopping rule is that the L-dimensional random vector $\{\Omega[k] = \Omega^1[k], \ldots, \Omega^L[k]\}$ depends on $X_1, \ldots, X_{k-1}$ only through $\{\Omega^1[k-1], \ldots, \Omega^L[k-1]\}$. Thus, $\{\Omega[k]\}_{k\geq 1}$ is a Markov process.

IV. Main Results

In this section, we demonstrate the asymptotic optimality of the D-CuSum in the special case of a single transition ($L = 2$). For our asymptotic analysis to be non-trivial, we let not only the prescribed lower bound on the ARL, γ, but also the duration of the transient, D_1, go to infinity. Specifically, we assume that there is a constant $c \in [0, \infty]$ such that

$$D_1 \sim c \frac{\log \gamma}{I_1}. \tag{12}$$

It turns out that the optimal asymptotic performance will depend on whether $c \geq 1$ or $c < 1$. This dichotomy can be seen in the following universal asymptotic lower bound on the worst-case ADD.

Theorem 1 (Lower Bound)**.** *Suppose that* (12) *holds.*
(i) If $c \geq 1$*, then as* $\gamma \to \infty$

$$\inf_{T:\mathbb{E}_\infty[T]\geq\gamma} J_L(T) \geq \inf_{T:\mathbb{E}_\infty[T]\geq\gamma} J_P(T) \geq \frac{\log\gamma}{I_1}(1-o(1)); \tag{13}$$

(ii) if $c < 1$*, then as* $\gamma \to \infty$

$$\inf_{T:\mathbb{E}_\infty[T]\geq\gamma} J_L(T) \geq \inf_{T:\mathbb{E}_\infty[T]\geq\gamma} J_P(T) \geq \log\gamma\left(\frac{1-c}{I_2}+\frac{c}{I_1}\right)(1-o(1)). \tag{14}$$

Outline of Proof. Since $\mathbb{E}_\infty[T] \geq \gamma$, then for each $m < \gamma$, there exists some $v_1 \geq 1$, such that

$$\mathbb{P}_\infty(T > v_1) > 0 \text{ and } \mathbb{P}_\infty(T < v_1 + m | T \geq v_1) \leq \frac{m}{\gamma}, \tag{15}$$

which can be shown by contradiction as in [12].

By Markov's inequality, for any $\epsilon > 0$,

$$\begin{aligned}&\mathbb{E}_{v_1}[T - v_1 | T \geq v_1]\\ &\geq \mathbb{P}_{v_1}(T - v_1 \geq (1-\epsilon)K_\gamma | T \geq v_1)(1-\epsilon)K_\gamma.\end{aligned} \tag{16}$$

If $c \geq 1$, we choose $K_\gamma = \frac{\log\gamma}{I_1}$. And if $c < 1$, we choose $K_\gamma = (\frac{1-c}{I_2} + \frac{c}{I_1})\log\gamma$. The goal is to show

$$\mathbb{P}_{v_1}(T - v_1 \geq (1-\epsilon)K_\gamma | T \geq v_1) \to 1, \text{ as } \gamma \to \infty, \tag{17}$$

for the two cases.

Case 1: If $c \geq 1$, $K_\gamma = \frac{\log\gamma}{I_1}$. Changing the measure $\mathbb{P}_\infty$ to $\mathbb{P}_{v_1}$ [15, Proof of Theorem 7.1.3], we obtain

$$\begin{aligned}&\mathbb{P}_{v_1}(T - v_1 < (1-\epsilon)K_\gamma | T \geq v_1)\\ &\quad\leq e^a \mathbb{P}_\infty(T < v_1 + (1-\epsilon)K_\gamma | T \geq v_1)\\ &\quad + \mathbb{P}_{v_1}\left(\max_{1\leq i\leq(1-\epsilon)K_\gamma}\sum_{j=v_1}^{v_1+i} Z_1(X_j) \geq a\right).\end{aligned} \tag{18}$$

Choose m in (15) to be $\frac{(1-\epsilon)\log\gamma}{I_1}$ and $a = (1-\epsilon^2)\log\gamma$. By (15), there exists v_1, such that

$$\begin{aligned}&e^a\mathbb{P}_\infty(T < v_1 + (1-\epsilon)K_\gamma | T \geq v_1)\\ &\quad\leq \frac{(1-\epsilon)\log\gamma}{\gamma^{\epsilon^2}} \to 0, \text{ as } \gamma \to \infty.\end{aligned} \tag{19}$$

Moreover, by the strong law of large number, we have

$$\mathbb{P}_{v_1}\left(\max_{1\leq i\leq(1-\epsilon)K_\gamma}\sum_{j=v_1}^{v_1+i} Z_1(X_j) \geq a\right) \to 0, \text{ as } \gamma \to \infty. \tag{20}$$

Hence,

$$\mathbb{P}_{v_1}(T - v_1 < (1-\epsilon)K_\gamma | T \geq v_1) \to 0, \text{ as } \gamma \to \infty. \tag{21}$$

Case 2: If $c < 1$, $K_\gamma = (\frac{1-c}{I_2} + \frac{c}{I_1})\log\gamma$. By changing measures similarly to case 1, we obtain

$$\begin{aligned}&\mathbb{P}_{v_1}(T < v_1 + (1-\epsilon)K_\gamma | T \geq v_1)\\ &\quad\leq e^{a'}\mathbb{P}_\infty(T < v_1 + (1-\epsilon)K_\gamma | T \geq v_1)\\ &\quad + \mathbb{P}_{v_1}\Bigg(\max_{v_1\leq i\leq v_1+(1-\epsilon)K_\gamma}\sum_{j=v_1}^{\min\{v_1+D_1-1,i\}} Z_1(X_j)\\ &\qquad + \sum_{j=v_1+D_1}^{i} Z_2(X_j) > a'\Bigg).\end{aligned} \tag{22}$$

We choose $m = (\frac{1-c}{I_2} + \frac{c}{I_1})(1-\epsilon)\log\gamma$ in (15) and $a' = (1-\epsilon^2)\log\gamma$. By (15), there exists v_1, such that as $\gamma \to \infty$

$$e^{a'}\mathbb{P}_\infty(T < v_1 + (1-\epsilon)K_\gamma | T \geq v_1) \leq \frac{(1-\epsilon)K_\gamma}{\gamma^{\epsilon^2}} \to 0. \tag{23}$$

Similarly, using the strong law of large number, we can show that

$$\begin{aligned}&\mathbb{P}_{v_1}\Bigg(\max_{v_1\leq i\leq v_1+(1-\epsilon)K_\gamma}\sum_{j=v_1}^{\min\{v_1+D_1-1,i\}} Z_1(X_j)\\ &\qquad + \sum_{j=v_1+D_1}^{i} Z_2(X_j) > a'\Bigg) \to 0, \text{ as } \gamma \to \infty.\end{aligned} \tag{24}$$

Hence,

$$\mathbb{P}_{v_1}(T - v_1 < (1-\epsilon)K_\gamma | T \geq v_1) \to 0, \text{ as } \gamma \to \infty. \tag{25}$$

□

As we can see from Theorem 1, the lower bound on the worst-case ADD can take different forms depending on the scaling behavior between ARL and the duration of the transient. This implies that in order to achieve the lower bound, a stopping rule should be able to adapt to different values of the duration of the transient D_1.

We next derive the upper bound on the worst-case ADD for the D-CuSum algorithm. We will show that although the D-CuSum algorithm does not employ the knowledge of D_1, it adapts to different values of D_1.

Here, we assume that there is a constant $c' \in [0, \infty]$ such that

$$D_1 \sim c'\,\frac{b}{I_1}. \tag{26}$$

Theorem 2 (Upper Bound)**.** *Suppose that* (26) *holds. For the D-CuSum algorithm in* (11)*, as* $b \to \infty$*, if* $c' \geq 1$*,*

$$J_L(\tau(b)) = J_P(\tau(b)) \leq \frac{b}{I_1}(1+o(1)); \tag{27}$$

if $c' < 1$*,*

$$J_L(\tau(b)) = J_P(\tau(b)) \leq b\left(\frac{c'}{I_1}+\frac{1-c'}{I_2}\right)(1+o(1)). \tag{28}$$

The main idea of the proof is to construct a stopping rule that employs the knowledge of D_1. Such a stopping rule always stops later than the D-CuSum algorithm. Therefore, the ADD of this stopping rule naturally serves as an upper bound for the ADD of the D-CuSum algorithm.

Outline of Proof. Due to the regenerative property of the test statistics, it is clear that

$$J_{\mathrm{L}}(\tau(b)) = J_{\mathrm{P}}(\tau(b)) \leq \mathbb{E}_1[\tau(b)]. \quad (29)$$

Define a test statistic that employs the knowledge of D_1 as

$$W'[k] = \max_{1\leq v_1 \leq k} \left(\sum_{j=v_1}^{\min\{v_1+D_1-1,k\}} Z_1(X_j) + \sum_{j=\min\{v_1+D_1,k+1\}}^{k} Z_2(X_j) \right). \quad (30)$$

Define the stopping rule based on $W'[k]$ as

$$\tau'(b) = \inf\{k \geq 1 : W'[k] > b\}. \quad (31)$$

It can be verified that $W[k] \geq W'[k]$. Hence,

$$\tau(b) \leq \tau'(b). \quad (32)$$

Next, we develop the upper bound for $\mathbb{E}_1[\tau'(b)]$, which is also an upper bound for $\mathbb{E}_1[\tau(b)]$.

Case 1: If $c' \geq 1$, our goal is to show

$$\mathbb{E}_1[\tau'(b)] \leq \frac{b}{I_1}(1 + o(1)), \quad (33)$$

for asymptotically large b.

We denote the largest integer that is smaller than c' as $\lfloor c' \rfloor$. For $i \leq \lfloor c' \rfloor$, we can show

$$\mathbb{P}_1\left(\frac{\tau'(b)}{\frac{b}{I_1}(1+\epsilon)} > i\right) \leq \delta^i, \quad (34)$$

where $\delta \to 0$ as $b \to \infty$. Similarly, for $i \geq \lfloor c' \rfloor + 1$, we can also show

$$\mathbb{P}_1\left(\frac{\tau'(b)}{\frac{b}{I_1}(1+\epsilon)} > i\right) \leq (\delta')^i, \quad (35)$$

where $\delta' \to 0$ as $b \to \infty$. Therefore,

$$\mathbb{E}_1\left[\frac{\tau'(b)}{\frac{b}{I_1}(1+\epsilon)}\right] \leq \sum_{i=0}^{\infty} \mathbb{P}_1\left(\frac{\tau'(b)}{\frac{b}{I_1}(1+\epsilon)} > i\right) \leq \frac{1}{1-\max(\delta,\delta')}, \quad (36)$$

which implies that

$$\mathbb{E}_1[\tau'(b)] \leq \frac{b(1+\epsilon)}{I_1(1-\max(\delta,\delta'))}. \quad (37)$$

Case 2: If $c' < 1$, our goal is to show

$$\mathbb{E}_1[\tau'(b)] \leq b\left(\frac{c'}{I_1} + \frac{1-c'}{I_2}\right)(1+o(1)). \quad (38)$$

Using similar ideas as in the first case, we can show that

$$\mathbb{P}_1\left(\frac{\tau'(b)}{\left(\frac{c'}{I_1} + \frac{1-c'}{I_2}\right)(1+\epsilon)} > i\right) \leq (\delta'')^i, \quad (39)$$

where $\delta'' \to 0$, as $b \to \infty$. Similarly, by (39), we obtain

$$\mathbb{E}_1[\tau'(b)] \leq b\left(\frac{c'}{I_1} + \frac{1-c'}{I_2}\right)\frac{1+\epsilon}{1-\delta''}. \quad (40)$$

□

The two cases in Theorem 2 correspond to the following two scenarios: (1) when D_1 is large, the D-CuSum algorithm stops within the transient phase with high probability; (2) when D_1 is small, the D-CuSum algorithm stops after the transient phase with high probability. This implies that although the D-CuSum algorithm does not employ the knowledge of D_1, it can still adapt to different value of D_1.

Finally, we demonstrate the asymptotic optimality of the D-CuSum algorithm. For technical convenience, we assume that the pre-change and post-change distributions satisfy the following condition:

$$\mathbb{P}_\infty(Y > m) \leq e^{-\alpha m}, \forall m > 0, \quad (41)$$

where $Y = \inf\{k \geq 1 : W[k] \leq 0\}$, and $\alpha > 0$ is any constant. We believe that our results will also hold without the condition in (41). And it can be easily verified that the condition in (41) is satisfied if $\mathbb{E}_{f_0} \log\left(\frac{\max(f_1(X), f_2(X))}{f_0(X)}\right) < 0$. We derive a lower bound on the ARL of D-CuSum, which provides a suitable threshold that attains the ARL lower bound asymptotically. For this choice of threshold, the asymptotic upper bound in Theorem 2 matches the asymptotic lower bound in Theorem 1.

Theorem 3 (Asymptotic optimality)**.** *Assume the distributions f_0, f_1, f_2 satisfy the condition in* (41)*. Let b_γ be such that $\mathbb{E}_\infty[\tau(b_\gamma)] \geq \gamma$. Then,*

$$J_L(\tau(b_\gamma)) \sim \inf_{T:\mathbb{E}_\infty[T]\geq\gamma} J_L(T), \quad (42)$$

and

$$J_P(\tau(b_\gamma)) \sim \inf_{T:\mathbb{E}_\infty[T]\geq\gamma} J_P(T). \quad (43)$$

Thus, the D-CuSum algorithm in (11) *is asymptotically optimal with respect to both Lorden's and Pollak's criteria.*

Outline of Proof. Let b_γ be such that $\mathbb{E}_\infty[\tau(b_\gamma)] \geq \gamma$. From Theorems 1 and 2 it suffices to show that $b_\gamma \leq \log\gamma(1+o(1))$. By the condition (41), the process $W[k]$ is regenerative, which implies that

$$\mathbb{E}_\infty[\tau(b_\gamma)] = \frac{\mathbb{E}_\infty[Y]}{\mathbb{P}_\infty(\tau(b_\gamma) < Y)} \geq \frac{1}{\mathbb{P}_\infty(\tau(b_\gamma) < Y)}. \quad (44)$$

It can be shown that for any $m > 0$,

$$\begin{aligned}
&\mathbb{P}_\infty(\tau(b_\gamma) < Y) \\
&= \mathbb{P}_\infty(\tau(b_\gamma) < Y, Y < m) + \mathbb{P}_\infty(\tau(b_\gamma) < Y, Y > m) \\
&\leq \mathbb{P}_\infty(\tau(b_\gamma) < m) + \mathbb{P}(Y > m) \\
&\overset{(a)}{\leq} m^3 e^{-b\gamma} + e^{-\alpha m} \\
&= (\frac{b_\gamma^3}{\alpha^3} + 1)e^{-b_\gamma},
\end{aligned} \quad (45)$$

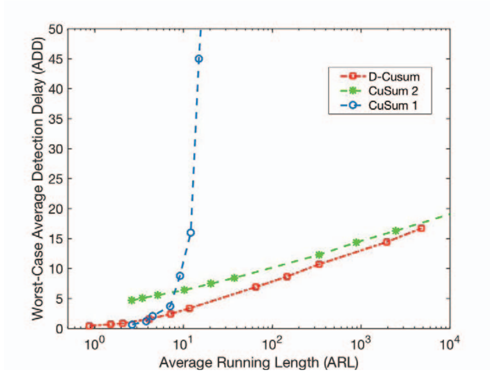

Fig. 1. Comparison of Performance between D-CuSum and CuSum without Considering Transient Dynamics: $D_1 = 5$

where the first term in (a) is due to the Boole's inequality and the Chernoff bound, and the last step is by choosing $m = \frac{b_\gamma}{\alpha}$. Hence,

$$\mathbb{E}_\infty[\tau(b_\gamma)] \geq e^{b_\gamma} \frac{1}{\frac{b_\gamma^3}{\alpha^3} + 1}. \tag{46}$$

Therefore, it follows that $b_\gamma = \log \gamma(1 + o(1))$. $\square$

V. Numerical Results

In this section, we compare the performance of the D-CuSum algorithm with that of the (traditional) CuSum algorithm [8], which is known to be optimal for the classical QCD problem without transient dynamics [17]. More specifically, we compare the performance of the D-CuSum algorithm with the performance of the CuSum algorithms for detecting a change from f_0 to f_1 (CuSum 1) and detecting a change from f_0 to f_2 (CuSum 2).

We set $L = 2$, i.e., one transient phase. We choose $f_0 = \mathcal{N}(0, 1)$, $f_1 = \mathcal{N}(1, 1)$ and $f_2 = \mathcal{N}(-1, 1)$. We set the duration of the transient phase $D_1 = 5$. In this particular example, $I_1 = I_2 = 1/2 = I$, and the optimal asymptotic performance is the same in all regimes for the D-CuSum algorithm. We plot the worst-case ADD as a function of ARL in Fig. 1. We note that the worst-case ADD is achieved when the change happens at time 0 under both criteria and for all three algorithms (shown in the proof of Theorem 2), hence we are plotting the ADD when the change happens at time 1 on the y-axis.

From Fig. 1, we can see that the D-CuSum algorithm has better performance than the two CuSum algorithms. More specifically, when the thresholds are chosen such that the ADDs of the three algorithms are smaller than D_1, CuSum 1 has the best performance. This is because when the ADD is smaller than D_1, the algorithms stop within the transient phase with high probability.

When the ADD is larger than D_1, CuSum 1 does not work anymore. Although in this case, all the algorithms stop after the transient phase, D-CuSum still has a better performance than CuSum 2. This is because the D-CuSum algorithm incorporates the distribution in the transient phase, while CuSum 2 does not. However, the performance gap between these two algorithms gets smaller as the ADD increases. The reason is that only finite information can be extracted from the transient phase ($D_1 = 5$) by the D-CuSum and as more information is collected from the persistent phase, the effect of the transient phase becomes negligible.

VI. Extensions

The results given in this paper are generalizable to the case where there is more than one transient phase. This generalization, while being somewhat tedious, is straightforward. It is also of interest to consider the extension to the case where the observations within each transient phase are not i.i.d. as in the observation model considered by Lai [12] for the standard QCD problem.

Acknowledgment

The work of Shaofeng Zou and Venugopal V. Veeravalli was supported by the National Science Foundation (NSF) under grants CIF 1514245 and ECCS 146231. The work of Georgios Fellouris was supported by the NSF grant CIF 1514245.

References

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

[2] H. V. Poor and O. Hadjiliadis, *Quickest Detection*, Cambridge University Press, 2009.

[3] M. Basseville and I. V. Nikiforov, *Detection of abrupt changes: theory and application*, vol. 104, Prentice Hall Englewood Cliffs, 1993.

[4] G. Rovatsos, J. Jiang, A. D. Domínguez-García, and V. V. Veeravalli, "Statistical power system line outage detection under transient dynamics," *submitted to IEEE Transactions on Signal Processing*, 2016.

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

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

[7] G. V. Moustakides and V. V. Veeravalli, "Sequentially detecting transitory changes," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*. IEEE, 2016, pp. 26–30.

[8] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–115, 1954.

[9] G. Lorden, "Procedures for reacting to a change in distribution," *The Annals of Mathematical Statistics*, pp. 1897–1908, 1971.

[10] M. Pollak, "Optimal detection of a change in distribution," *The Annals of Statistics*, pp. 206–227, 1985.

[11] G. Rovatsos, S. Zou, and V. V. Veeravalli, "Quickest change detection under transient dynamics," *to appear in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017.

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

[13] T. L. Lai, "Sequential changepoint detection in quality control and dynamical systems," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 613–658, 1995.

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

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

[16] S. Asmussen, *Applied probability and queues*, vol. 51, Springer Science & Business Media, 2008.

[17] George V. Moustakides, "Optimal stopping times for detecting changes in distributions," vol. 14, no. 4, pp. 1379–1387, Dec. 1986.