

CHANGE DETECTION WITH UNKNOWN POST-CHANGE PARAMETER USING KIEFER-WOLFOWITZ METHOD

Vijay Singamasetty, Navneeth Nair, Srikrishna Bhashyam and Arun Pachai Kannu

Department of Electrical Engineering

Indian Institute of Technology Madras, Chennai - 600036, India.

ee13s075@ee.iitm.ac.in, navneethnair@gmail.com, skrishna@ee.iitm.ac.in, arunpachai@ee.iitm.ac.in

ABSTRACT

We consider a change detection problem with an unknown post-change parameter. The optimal algorithm in minimizing worst case detection delay subject to a constraint on average run length, referred as parallel CUSUM, is computationally expensive. We propose a low complexity algorithm based on parameter estimation using Kiefer-Wolfowitz (KW) method with CUSUM based change detection. We also consider a variant of KW method where the tuning sequences of KW method are reset periodically. We study the performance under the Gaussian mean change model. Our results show that reset KW-CUSUM performs close to the parallel CUSUM in terms of worst case delay versus average run length. Non-reset KW-CUSUM algorithm has smaller probability of false alarm compared to the existing algorithms, when run over a finite duration.

Index Terms— Kiefer-Wolfowitz algorithm, CUSUM, false alarm probability, average run length, Detection delay

1. INTRODUCTION AND SYSTEM MODEL

The problem of change detection using statistical tests has been studied over several decades [1–5]. In this paper, we consider the problem of change detection when the post-change distribution has an unknown parameter. Let $\{f_\theta, \theta \in \mathbb{R}\}$ denote a family of probability density functions parameterized by θ . The observation at time n is denoted as Y_n . Let $\nu \in \mathbb{N}^+$ denote the change point such that the observations before and after change follow different statistics. We assume ν to be a unknown non-random value. Specifically, the observations $\{Y_n\}$, $n \geq 1$ are independent and follow the statistics

$$Y_n \sim \begin{cases} f_{\theta_0} & n < \nu \\ f_{\bar{\theta}} & n \geq \nu \end{cases} . \quad (1)$$

Here, the value of θ_0 is known but $\bar{\theta}$ is unknown apriori. However, it is known that $\bar{\theta} \in \Theta_1$ where $\Theta_1 \subset \mathbb{R}$ is a known set. For ease of presentation, we first consider the case of Θ_1 being a *discrete* set. Later in Section 3.3, we address the case of Θ_1 being a continuous set. We are interested in developing algorithms to detect the change quickly and reliably. Specifically, we are interested in the following quantities, probability of false alarm P_f , average run length τ_r and average detection delay τ_d and worst-case detection delay τ_w . We use P_ν to denote the probability measure (and E_ν for expectation) when the change point is ν . We use $\nu = \infty$ for the no change case. With T being the time at which the algorithm declares the change (which is random), we have

$$P_f = P_\infty\{T < \infty\}, \quad (2)$$

$$\tau_r = E_\infty\{T\}, \quad (3)$$

$$\tau_d = E_\nu\{(T - \nu)|T \geq \nu\}. \quad (4)$$

P_f is a well studied criterion when the change point is random with a prior distribution [6]. In the non-random setting for change point, several algorithms including the conventional CUSUM algorithm will always raise false alarm ($P_f = 1$) when run over infinite time duration [3]. In such cases, τ_r - the average time the algorithm takes to raise a false alarm (2), is a quantity of interest [1]. Alternative minimax formulations on false alarm criteria have been proposed in [7]. The detection delay τ_d defined in (4) may depend on the change point. Hence, another quantity of significance is the worst-case detection delay τ_w defined as

$$\tau_w = \sup_{\nu \geq 1} E_\nu\{(T - \nu)|T \geq \nu\} = \sup_{\nu \geq 1} \tau_d. \quad (5)$$

The above definition (5) of worst case detection delay was originally studied in [2].

2. PRELIMINARIES AND RELATED WORK

2.1. Regression Function

For convenience, we define $\Theta = \theta_0 \cup \Theta_1$, which is a discrete set containing all the possible parameters for the observation statistics. Also, let $I \subset \mathbb{R}$ be the *smallest interval* such that $\Theta \subset I$ and I may be finite or not. For $\theta \in I$, we define the log-likelihood ratio (LLR) of the observations parameterized by θ as

$$L^\theta(Y_n) = \log \frac{f_\theta(Y_n)}{f_{\theta_0}(Y_n)}. \quad (6)$$

We define the expected value of LLR as

$$G(\theta) = E_\nu\{L^\theta(Y_n)\}, \forall \theta \in I, \quad (7)$$

where the expectation is taken over the *true* distribution of Y_n . Since the true distribution differs before and after change, we sometimes use the explicit notation $G_b(\theta)$ and $G_a(\theta)$ to denote the regression function $G(\theta)$ before and after change, respectively. We denote the Kullback-Leibler (KL) distance between pdf f_1 and f_2 as $D(f_1||f_2) = \int_{\mathbb{R}} f_1(x) \log \frac{f_1(x)}{f_2(x)} dx$. We know that $D(f_1||f_2) \geq 0$ with equality if and only if $f_1 = f_2$. The properties of the function $G(\theta)$ are summarized in the following lemma and they can be easily verified.

Lemma 1. *Regression functions before and after change are*

$$G_b(\theta) = -D(f_{\theta_0}||f_\theta) \leq 0. \quad (8)$$

$$G_a(\theta) = D(f_{\bar{\theta}}||f_{\theta_0}) - D(f_{\bar{\theta}}||f_\theta) = G_a(\bar{\theta}) - D(f_{\bar{\theta}}||f_\theta). \quad (9)$$

Hence, it follows that, $G_b(\theta)$ reaches maximum at θ_0 and $G_a(\theta)$ reaches maximum at $\bar{\theta}$. The change detection algorithms use this property to simultaneously identify the unknown parameter and detect the change. For subsequent use, let us denote $\theta_{\max} = \arg \max_{\theta} G(\theta)$.

2.2. Existing Algorithms

We discuss two existing change detection algorithms, parallel [8] and adaptive [9] CUSUM, for the case of unknown post-change parameter. Parallel CUSUM algorithm proceeds as follows: For all $\theta \in \Theta_1$, $n \geq 1$, with the initialization $W_0^\theta = 0$, we compute $W_n^\theta = \max(W_{n-1}^\theta + L^\theta(Y_n), 0)$ and declare change when $\max_{\theta \in \Theta_1} W_n^\theta > A$, else we continue. Here A is a fixed threshold. In [1], asymptotic optimal properties of parallel CUSUM in minimizing the worst case detection delay (with a definition slightly different from (5)) subject to a constraint on the average run length has been established. With T_{pc} being the time at which parallel CUSUM declares change, it easily follows that parallel CUSUM will always raise false alarm, if run indefinitely, i.e., $P_\infty(T_{pc} < \infty) = 1$.

Parallel CUSUM computes the CUSUM metric for all $\theta \in \Theta_1$. In order to reduce this complexity, an adaptive CUSUM algorithm is proposed in [9], in which the unknown parameter is tracked adaptively. If $G_a(\cdot)$ is a strictly concave function, for any ϵ , we can always find p such that $G_a(p) = G_a(p + \epsilon)$ and $\bar{\theta}$ lies in the interval $[p, p + \epsilon]$. Choosing a small ϵ , and small step size μ , the algorithm proceeds adaptively as follows: Initialize $p_0 \in I$. Iterate as $p_n = p_{n-1} + \mu[L^{p_{n-1}+\epsilon}(Y_n) - L^{p_{n-1}}(Y_n)]$ and truncate/restrict p_n within I . By setting $\hat{\theta}_n = p_n + \frac{\epsilon}{2}$, we estimate the unknown parameter as $\bar{\theta}_n = \text{close}(\hat{\theta}_n, \Theta)$. Here, $\text{close}(x, \mathcal{A})$ chooses the element from the set \mathcal{A} which has the smallest Euclidean distance to the input x . With initialization $W_0 = 0$, CUSUM metric at time $n \geq 1$ is computed as

$$W_n = \max(W_{n-1} + L^{\bar{\theta}_n}(Y_n), 0) \quad (10)$$

and decision rule \mathcal{R} with threshold $A \in (0, \infty)$ is given by

$$\mathcal{R} = \begin{cases} \text{Declare change at time } n & \text{if } W_n > A \\ \text{Continue} & \text{otherwise.} \end{cases} \quad (11)$$

Adaptive CUSUM computes the LLR for *only one* parameter value ($\bar{\theta}_n$) at each time instant. In [9], it has been argued that, if $G(\theta)$ is strictly concave and symmetric about its maximum, then $E\{\hat{\theta}_n\} \rightarrow \theta_{\max}$ as $n \rightarrow \infty$, with suitably chosen step size μ .

3. KIEFER-WOLFOWITZ BASED CUSUM

3.1. Kiefer Wolfowitz Method

In this method, we employ stochastic approximation [10] based Kiefer-Wolfowitz (KW) method [11] to estimate the unknown post-change parameter. Then we perform CUSUM using the estimated parameter, in a manner similar to adaptive CUSUM. Towards that, we define the *tuning* sequences $\{a_n\}$ and $\{c_n\}$ which satisfy the following requirements:

$$(C1): c_n \rightarrow 0, \text{ as } n \rightarrow \infty \quad (C2): \sum_{n=1}^{\infty} a_n = \infty \quad (12)$$

$$(C3): \sum_{n=1}^{\infty} \frac{a_n^2}{c_n} < \infty \quad (C4): \sum_{n=1}^{\infty} a_n c_n < \infty \quad (13)$$

The polynomial-like sequences of the form $a_n = n^{-a}$ and $c_n = n^{-c}$ with suitably chosen $a, c > 0$ can satisfy the requirements (12)-(13). Standard KW CUSUM method proceeds as follows:

1. Initialize the parameter estimate as $\hat{\theta}_0$ such that $\hat{\theta}_0 \pm c_1$ belongs to I .
2. Initialize the CUSUM metric as $W_0 = 0$.
3. Update the parameter at each time instant as

$$\hat{\theta}_n = \hat{\theta}_{n-1} + a_n \left(\frac{L^{\hat{\theta}_{n-1}+c_n}(Y_n) - L^{\hat{\theta}_{n-1}-c_n}(Y_n)}{c_n} \right). \quad (14)$$

4. Truncate/Restrict $\hat{\theta}_n$ such that $\hat{\theta}_n \pm c_{n+1}$ belongs to I .
5. Round off the estimate as $\bar{\theta}_n = \text{close}(\hat{\theta}_n, \Theta)$
6. At each time instant n , compute the CUSUM metric as (10) and declare change with threshold A in the same manner as (11).

The update equation (14) is a stochastic gradient method with $2a_n$ being step-size parameter and the gradient of regression function evaluated at $\hat{\theta}_{n-1}$, which is $\frac{G(\hat{\theta}_{n-1}+c_n) - G(\hat{\theta}_{n-1}-c_n)}{2c_n}$, being replaced with its instantaneous approximation. In order to proceed further, we make the following assumptions. We assume that variance of $L^\theta(Y_n)$ is bounded for all θ , specifically,

$$\sup_{\theta \in I} \text{Var}\{L^\theta(Y_n)\} < \infty. \quad (15)$$

Further, we assume the regularity conditions: For all $\theta \in I$,

$$\exists K_0 > 0, K_1 > 0 : K_0|\theta - \theta_{\max}| \leq |G'(\theta)| \leq K_1|\theta - \theta_{\max}|. \quad (16)$$

and

$$G'(\theta)(\theta - \theta_{\max}) < 0, \forall \theta \neq \theta_{\max}. \quad (17)$$

Theorem 1 (Broadie et al. [12]). *If the tuning sequences satisfy the conditions (12)-(13) and the regularity conditions (15)-(17) are satisfied, for the KW estimates $\hat{\theta}_n$ in (14), we have $E\{(\hat{\theta}_n - \theta_{\max})^2\} \rightarrow 0$ as $n \rightarrow \infty$. We also have $P(|\hat{\theta}_n - \theta_{\max}| > \epsilon) \rightarrow 0$ as $n \rightarrow \infty$, for any $\epsilon > 0$.*

In our model, $\hat{\theta}_n$ converges to θ_0 in the absence of change. If the change happens at finite ν , then $\hat{\theta}_n$ converges to $\bar{\theta}$. KW recursion (14) in the original paper [11] used two i.i.d. samples to compute $L^{\hat{\theta}_{n-1}+c_n}(\cdot)$ and $L^{\hat{\theta}_{n-1}-c_n}(\cdot)$, for instance Y_n and Y_{n-1} . However, convergence has been established in [12] even if the same sample Y_n is repeatedly used, as done in (14). The mean square convergence of KW method is stronger than the mean convergence guarantee [9] for adaptive CUSUM. Also, unlike adaptive CUSUM, symmetry of regression function $G(\cdot)$ is *not* required for convergence of KW method.

3.2. Reset KW CUSUM

As the step tuning sequences $\{a_n\}$ and $\{c_n\}$ converge to 0, when the change happens at a large ν , it takes longer time for the KW parameter estimates to converge to the post-change value and hence the detection delay of standard KW CUSUM method will be correspondingly large (as illustrated in Fig. 1). To overcome this limitation, we consider resetting the tuning sequences a_n (and c_n) every P time instants as $a_n = a_k$ for $n > P$ with $k = \text{mod}(n-1, P) + 1$ and a_n for $1 \leq n \leq P$ can be chosen arbitrarily (c_n is reset in similar manner).

3.3. Other Models

Reset and standard (non-reset) KW CUSUM can be applied (without any modification) to the following variations of our original model (1), for the case of $\Theta_1 = \{\bar{\theta}\}$ (which corresponds to the standard change detection problem with known pdfs) and for the case of Θ_1 being a *continuous* set. Performance of these KW-CUSUM methods are analyzed in the following section.

4. PERFORMANCE ANALYSIS

4.1. Gaussian Mean Change Model

For the performance analysis, we consider the following specific model,

$$Y_n \sim \begin{cases} \mathcal{N}(Y_n : 0, \sigma^2) & n < \nu \\ \mathcal{N}(Y_n : M, \sigma^2) & n \geq \nu \end{cases} \quad (18)$$

where $\mathcal{N}(Y_n : \theta, \sigma^2)$ denotes a Gaussian pdf with variable Y_n with mean θ and variance σ^2 . Here M is an unknown integer with $M \in \{1, \dots, K\}$. This model appears in many applications, for instance in the physical layer fusion model in sensor networks [13] with an unknown number of affected sensors sending symbol 1 after the change [14]. For this model, $G(\theta)$ will be a quadratic function and regularity conditions (15)-(17) will be satisfied. Hence standard KW method convergence guarantee in Theorem 1 is valid. Further, for this specific model, we present additional convergence guarantees in the following section.

4.2. Additional Convergence Guarantees

In the Gaussian mean change model, we have $\Theta = \{0, 1, \dots, K\}$. Also, the Kiefer-Wolfowitz (KW) method parameter update (14) becomes

$$\hat{\theta}_n = \hat{\theta}_{n-1} + \frac{2a_n}{\sigma^2} (Y_n - \hat{\theta}_{n-1}). \quad (19)$$

Interestingly, the KW parameter update equation does not depend on the sequence c_n . By setting $\hat{\theta}_0 = 0$ and choosing $a_n = \frac{\sigma^2}{2n}$, which meets the KW convergence requirements, the parameter estimation (19) can be re-written as

$$\hat{\theta}_n = \frac{\sum_{k=1}^n Y_k}{n}, \quad (20)$$

which is the empirical mean of the observations. Usually, we *truncate* the KW parameter estimate to remain within the interval $[0, K]$. We ignore this truncation in the following asymptotic analysis. For subsequent use, let $\mu_o(k)$ (and $\mu_1(k)$) denote empirical mean of k pre-change (post-change) i.i.d. observations.

Lemma 2. *For the observation model (18), when there is no change, KW estimate in (20) converges to 0 almost surely. On the other hand, when change happens at a finite point ν , we have $\hat{\theta}_n \rightarrow M$ almost surely, as $n \rightarrow \infty$.*

Proof. Under no change, the convergence result follows from strong law of large numbers [15]. Now, consider that the change happens at finite ν . Now, the KW parameter estimate using $N + 1$ post-change observations (at time $\nu + N$), is given by $\hat{\theta}_{\nu+N} = \frac{\sum_{k=1}^{\nu+N} Y_k}{\nu+N}$ and it

can be re-written as

$$\begin{aligned} \hat{\theta}_{\nu+N} &= \left(\frac{\sum_{k=1}^{\nu-1} Y_k}{\nu-1} \right) \frac{\nu-1}{N+\nu} + \left(\frac{\sum_{k=\nu}^{\nu+N} Y_k}{N+1} \right) \frac{N+1}{N+\nu}, \\ &= \mu_0(\nu-1) \frac{\nu-1}{N+\nu} + \mu_1(N+1) \frac{N+1}{N+\nu}. \end{aligned} \quad (21)$$

For fixed ν , as $N \rightarrow \infty$, we have $\mu_0(\nu-1) \frac{\nu-1}{N+\nu} \rightarrow 0$ almost surely and $\mu_1(N+1) \frac{N+1}{N+\nu} \rightarrow M$ almost surely, from strong law of large numbers. \square

From above lemma, in the absence of change, for every realization of the observation sequence $\{Y_n\}$, KW parameter estimate satisfies (with probability one) $\text{close}(\hat{\theta}_n, \Theta) = 0, \forall n \geq N_0$. Here N_0 may depend on the particular realization (sample path). For that particular sample path, if non-reset KW-CUSUM does not cross threshold by time N_0 , then it will not raise false alarm subsequently. We note in our simulation results (Fig. 2) that non-reset KW-CUSUM algorithm has smaller probability of false alarm compared to the existing algorithms, when run over finite duration. Now, we study the behavior of KW estimate (20) with respect to the number of post-change observations, when the change point $\nu \rightarrow \infty$.

Lemma 3. *As $\nu \rightarrow \infty$, KW estimate $\hat{\theta}_{\nu+N} \rightarrow 0$ almost surely, if N grows sub-linearly with ν , that is $N = o(\nu)$. On the other hand, if N grows linearly with ν as $N = \beta\nu$ for a constant β , we have $\hat{\theta}_{\nu+N} \rightarrow M \frac{\beta}{\beta+1}$ almost surely, as $\nu \rightarrow \infty$.*

Proof follows easily by applying limits to the expression in (21). Hence, when ν is large and N is very small compared to ν , we have $\text{close}(\hat{\theta}_{\nu+N}, \Theta) = 0$ and KW CUSUM metric does not increase at time $\nu + N$. On the other hand, if $N = \beta\nu$ for large enough β , we have $\text{close}(\hat{\theta}_{\nu+N}, \Theta) = M$, expected value of $L^{\hat{\theta}_{\nu+N}}(Y_{\nu+N})$ is positive and KW CUSUM metric tends to increase. This suggests that the number of post change observations need to grow linearly with ν for non-reset KW CUSUM to detect the change. The linear increase in detection delay with respect to ν is observed in our simulations (Fig. 1).

4.3. Reset KW CUSUM Convergence

For the reset KW method where we reset a_n periodically (every P time instants), conditions (12),(13) are not met and hence the parameter estimate *will not* converge in the mean square sense to the true parameter. On the other hand, we have the following guarantee on mean convergence.

Lemma 4. *If $\Pi_{n=1}^P (1 - \frac{2a_n}{\sigma^2}) < 1$, for reset KW method, we have $E\{\hat{\theta}_n\} \rightarrow \theta_{\max}$ as $n \rightarrow \infty$ where $\theta_{\max} = 0$ in the absence of change and $\theta_{\max} = M$ when change point is finite.*

Proof. We prove for the case of finite ν . Taking the expectation on (19), for $n \geq 0$, we have $E\{\hat{\theta}_{\nu+n+1}\} = E\{\hat{\theta}_{\nu+n}\} + \frac{2a_{\nu+n+1}}{\sigma^2} (E\{Y_{\nu+n+1}\} - E\{\hat{\theta}_{\nu+n}\})$. Noting that $E\{Y_{\nu+n+1}\} = M$, we have $M - E\{\hat{\theta}_{\nu+n+1}\} = M - E\{\hat{\theta}_{\nu+n}\} - \frac{2a_{\nu+n+1}}{\sigma^2} \times (M - E\{\hat{\theta}_{\nu+n}\})$. Defining the error $e(\nu + n) = M - E\{\hat{\theta}_{\nu+n}\}$, we have $e(\nu + n + 1) = \left(1 - \frac{2a_{\nu+n+1}}{\sigma^2}\right) e(\nu + n)$. It follows that $e(\nu + n + 1) = \left(\prod_{k=1}^{n+1} [1 - \frac{2a_{\nu+k}}{\sigma^2}]\right) e(\nu)$. With

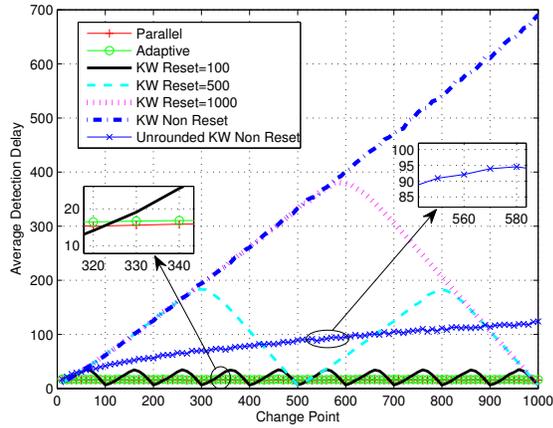


Fig. 1. Plot of τ_d versus ν with a fixed threshold.

$\prod_{n=1}^P \left(1 - \frac{2a_n}{\sigma^2}\right) \triangleq \xi$, we have $\prod_{k=1}^{n+1} \left(1 - \frac{2a_{\nu+k}}{\sigma^2}\right) = C\xi^{\lfloor \frac{n+1}{P} \rfloor}$ where $C = \prod_{l=1}^{\text{mod}(n+1, P)} \left(1 - \frac{2a_{\nu+l}}{\sigma^2}\right)$. As $\xi < 1$ and C is upper bounded by a finite constant, we have $\prod_{k=1}^{n+1} \left(1 - \frac{2a_{\nu+k}}{\sigma^2}\right) \rightarrow 0$ as $n \rightarrow \infty$ and hence $e(\nu + n + 1) \rightarrow 0$ and $E\{\hat{\theta}_{\nu+n+1}\} \rightarrow M$. \square

5. SIMULATION RESULTS

We consider Gaussian mean change model with $K = 100$, $\sigma^2 = 4$ and $M = 2$. For KW method, we use $a_n = 1/n$ and reset it every P time instants for reset KW method. Using Monte Carlo averaging over multiple trials, by varying the threshold A , we find the values of P_f , τ_r and τ_d for various algorithms. The time duration of each trial is restricted to 10^5 samples. In Fig. 1, we show the behavior of average detection delay τ_d with respect to the change point ν , for a fixed threshold. Parallel and adaptive CUSUM do not have any significant variation of τ_d with ν . In accordance with our discussions after Lemma 3, τ_d of non-reset KW CUSUM linearly increases with ν . Interestingly, τ_d of reset KW CUSUM (legend “Reset” specifies the value of P) exhibits oscillatory behavior, with the maximum delay occurring when ν is near the middle of the resetting window.

In Fig. 2, with $\nu = 10$, we plot τ_d versus P_f by varying the threshold, where false alarm probability P_f is obtained with algorithms run over a finite duration. While P_f for adaptive and parallel CUSUM remains close to one, we find that P_f of standard KW CUSUM decreases with increase in threshold. Intuitively, for large P , resetting KW CUSUM performs close to standard KW CUSUM and for small P , it performs close to parallel and adaptive CUSUM. We also study the unrounded standard KW CUSUM where metric (10) is computed with KW estimate $\hat{\theta}_n$ instead of the rounded value $\bar{\theta}_n = \text{close}(\hat{\theta}_n, \Theta)$. With a large change point ν , $\hat{\theta}_{\nu-1}$ is typically close to 0 and $\hat{\theta}_{\nu+k}$ starts drifting towards M with k but rounding $\bar{\theta}_{\nu+k} = \text{close}(\hat{\theta}_{\nu+k}, \Theta)$ may bring it back to 0 for several values of k . Hence the unrounded standard KW CUSUM has smaller τ_d compared to the (rounded) standard KW CUSUM in Fig. 1.

We plot the worst case detection delay τ_w defined in (5) versus average run length τ_r in Fig. 3. It can be argued that the worst case delay occurs at $\nu = 1$ for parallel and adaptive CUSUM (similar to Lemma 2 in [5]). For the resetting KW CUSUM, the worst-case

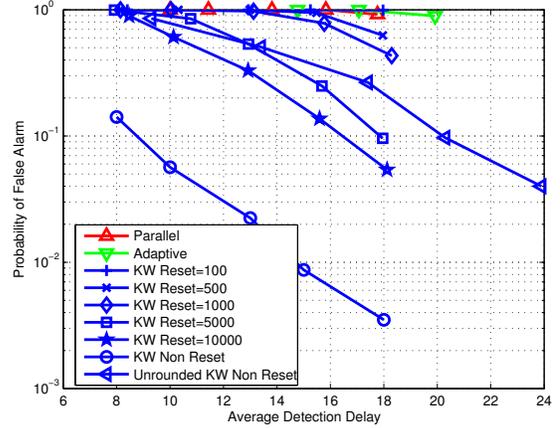


Fig. 2. Plot of P_f versus τ_d with $\nu = 10$.

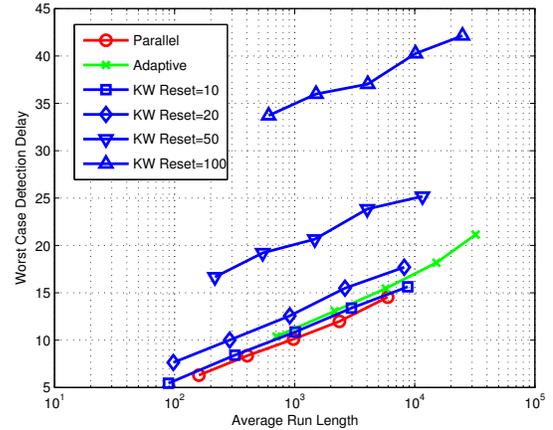


Fig. 3. Worst case delay (τ_w) versus average run length (τ_r).

delay can be computed as $\max_{1 \leq \nu \leq P} \tau_d$. As the resetting window size gets smaller, resetting KW CUSUM performs comparably to the parallel and adaptive CUSUM.

6. CONCLUSIONS

We considered the change detection problem with unknown post change parameter. We proposed low complexity (reset/non-reset) KW CUSUM algorithms and studied their performance for the Gaussian mean change model. Reset KW CUSUM performs close to the optimal parallel CUSUM in terms of worst-case detection delay versus average run length. Our results also show that non-reset KW-CUSUM algorithm has smaller probability of false alarm compared to the existing algorithms, when run over a finite duration.

7. REFERENCES

[1] G. Lorden, “Procedures for reacting to a change in distribution,” *The Annals of Mathematical Statistics*, vol. 42,

- no. 6, pp. 1897–1908, Dec. 1971. [Online]. Available: <http://www.jstor.org/stable/2240115>
- [2] M. Pollak, “Optimal detection of a change in distribution,” *The Annals of Statistics*, vol. 13, no. 1, pp. 206–227, Mar. 1985. [Online]. Available: <http://www.jstor.org/stable/2241154>
- [3] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.
- [4] Y. Mei, “Quickest detection in censoring sensor networks,” in *Intl. Symp. on Info. theory (ISIT)*, Aug. 2011, pp. 2148–2152.
- [5] V. V. Veeravalli and T. Banerjee, “Chapter 6 - quickest change detection,” in *Academic Press Library in Signal Processing: Volume 3 Array and Statistical Signal Processing*. Elsevier, 2014, pp. 209–255.
- [6] A. N. Shiryaev, “On optimum methods in quickest detection problems,” *Theory of Probability & Its Applications*, vol. 8, no. 1, pp. 22–46, 1963.
- [7] Y. Mei, “Is average run length to false alarm always an informative criterion?” *Sequential Analysis*, vol. 27, no. 4, pp. 354–376, 2008.
- [8] I. V. Nikiforov, “A suboptimal quadratic change detection scheme,” *IEEE Transactions on Information Theory*, vol. 46, no. 6, pp. 2095–2107, Sep. 2000.
- [9] C. Li, H. Dai, and H. Li, “Adaptive quickest change detection with unknown parameter,” in *Intl. Conf. Acoustics, Speech and Signal Proc. (ICASSP)*, Apr. 2009, pp. 3241–3244.
- [10] H. Robbins and S. Monro, “A stochastic approximation method,” *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400–407, Sep. 1951. [Online]. Available: <http://www.jstor.org/stable/2236626>
- [11] J. Kiefer and J. Wolfowitz, “Stochastic estimation of the maximum of a regression function,” *The Annals of Mathematical Statistics*, vol. 23, no. 3, pp. 462–466, Sep. 1952. [Online]. Available: <http://www.jstor.org/stable/2236690>
- [12] M. Broadie, D. Cicek, and A. Zeevi, “General bounds and finite-time improvement for the Kiefer-Wolfowitz stochastic approximation algorithm,” *Operations Research*, vol. 59, no. 5, pp. 1211–1224, Oct. 2011.
- [13] T. Banerjee, V. Sharma, V. Kavitha, and A. K. JayaPrakasam, “Generalized analysis of a distributed energy efficient algorithm for change detection,” *IEEE Transactions on Wireless Communications*, vol. 10, no. 1, pp. 91–101, Jan. 2011.
- [14] P. S. Kumar, B. S. Kiran, A. P. Kannu, and S. Bhashyam, “Algorithms for change detection with unknown number of affected sensors,” in *Natl. Conf. on Communications (NCC)*, Feb 2013, pp. 1–5.
- [15] P. Billingsley, *Probability and measure*. John Wiley & Sons, 2008.