ONLINE CHANGE DETECTION OF LINEAR REGRESSION MODELS

Jun Geng[1] *and Bingwen Zhang*[2] *and Lauren M. Huie*[3] *and Lifeng Lai*[2]

[1] School of Elec. and Info. Engr., Harbin Inst. of Tech., Harbin, China, jgeng@hit.edu.cn
[2] Dept. of Elec. and Comp. Engr., Worcester Poly. Inst., Worcester, MA, {bzhang,llai}@wpi.edu
[3] Air Force Research Laboratory, Rome, NY, lauren.huie@us.af.mil

ABSTRACT

In this paper, we consider the problem of quickly detecting an abrupt change of linear coefficients in linear regression models. In particular, the observer sequentially observes a sequence of observations $\{(\mathbf{x}_n, y_n)\}_{n=1}^{\infty}$, which is assumed to obey a linear regression model at each time slot n. Some of the coefficients in the linear model change at a fixed but unknown time t. The post-change linear coefficients are unknown to the observer. The observer aims to design an online algorithm to detect the model change based on his sequential observations. Two performance metrics, namely the worst case detection delay (WADD) and the average run length to false alarm (ARL2FA), are adopted to evaluate the performance of detection algorithms. We design a low complexity algorithm, termed as parallel sum algorithm, for the detection purpose. An asymptotic upper bound on WADD is provided under any given ARL2FA constraint.

***Index Terms*—** Linear model; online change detection; parallel-sum procedure; sequential detection; unknown post-change parameter

1. INTRODUCTION

Linear regression is a basic but important model in statistical machine learning. It has wide range applications in data fitting, signal processing, economic data analysis [1], biomedical science [2], etc. A fundamental question of linear regression is to estimate the coefficients in the linear model based on a group of observed data. A common assumption used in the existing work is that all data come from a single linear model. However, in many applications, such as the ones in dynamic linear systems, the system model changes over time. In such applications, it is of interest to detect the change of linear model.

In our recent work [3], we study the change detection problem for a linear model under an offline setup, in which all observations are collected before the analysis is carried out. In this paper, we focus on an *on-line setup*, in which one collects observations in a sequential manner and carries out the analysis after each observation is collected. In particular, an observer keeps monitoring the system input $\mathbf{x}_n$ and the system output y_n. y_n and $\mathbf{x}_n$ are assumed to obey a linear model at each time slot n. At the very beginning, the relationship between y_n and $\mathbf{x}_n$ is known exactly by the observer. However, some of the linear coefficients change at an unknown time t, and the observer does not know the post-change linear coefficients. Based on his *sequential* observations, the goal of the observer is to design an on-line detection algorithm to detect such a change in the linear model.

We formulate this problem in the framework of non-Bayesian quickest change-point detection, in which the change time t is assumed to be fixed but unknown. In particular, under an average run length to false alarm (ARL2FA) constraint, which implies that the expected duration between two false alarms is larger than γ, we aim to minimize the worst case average detection delay (WADD) $\sup_{t\geq 1} \operatorname{esssup} \mathbb{E}_t[(\tau - t + 1)^+ | \mathbf{x}_1, \ldots, \mathbf{x}_{t-1}, y_1, \ldots, y_{t-1}]$, in which τ is the time instant when an alarm is raised. When the post-change distribution contains unknown parameters, it has been shown by T. L. Lai [4] that the generalized likelihood ratio (GLR) cumulative sum (CUSUM) detection rule is asymptotically optimal as ALR2FA goes to infinity. However, GLR-CUSUM is computationally infeasible for the problem considered in this paper. Hence, in this paper, we propose a low complexity algorithm, in which the observer calculates the correlations between y_n and each individual component in $\mathbf{x}_n$ and then compares the sum of these calculated statistics with a predesigned threshold. If the threshold is exceeded, which indicates that y_n strongly depends on some components in $\mathbf{x}_n$, the observer raises an alarm. This algorithm is termed as the parallel-sum algorithm. This algorithm is computationally feasible. We further analyze the performance of this low-complexity algorithm and provide an upper bound of WADD.

Extensive existing works are related to the problem considered in this paper. Due to limited space, we only mention a few most relevant papers here. [5] proposes a parallel recursive χ^2 test to detect the abrupt change in multivariate Gaussian random signals with unknown mean after change. [6] proposes a SUM algorithm, which is based on the sum of local CUSUMs, to quickly detect the abrupt change in multiple independent data streams. Different from these two works, our paper aims to detect the change in the linear models rather than the change in the distributions of random vectors. [7]

The work of J. Geng was supported by the Fundamental Research Funds for the Central Universities under grant AUGA5710013915. The work of B. Zhang and L. Lai was supported by the National Science Foundation under grants DMS-12-65663 and CNS-14-57076, and an extension grant from Air Force Research Lab, Information Directorate.

considers the change detection problem for the linear model, and proposes an algorithm based on the decomposition of the post-change parameter space. [8] adopts the window-limited GLR CUSUM for the change detection in stochastic dynamic systems. However, the algorithms in these two papers have a very high computation complexity. In this paper, we propose a low complexity algorithm.

The remainder of this paper is organized as follows. The mathematical model is given in Section 2. Section 3 presents a lower bound of WADD and reviews the GLR-CUSUM algorithm. In Section 4, we propose and study the performance of the low complex parallel sum algorithm. Numerical examples are given in Section 5 to illustrate the results obtained in this work. Finally, Section 6 offers concluding remarks.

2. MODEL

We consider the problem of detecting the change-point in a linear regression model. Let $\{(\mathbf{x}_n, y_n)\}_{n=1}^{\infty}$ be a sequence of conditionally (condition on the change-point) independent and identically distributed (i.i.d.) random vectors whose underlying model changes at a fixed but unknown time t, which is referred to as change-point. Before the change happens, i.e. $n < t$, y_n linearly depends on $\mathbf{x}_n$ via a known model $y_n = \mathbf{a}_0^T \mathbf{x}_n + \epsilon_n$, in which $\mathbf{a}_0$ is perfectly known by the observer, and ϵ_n is i.i.d. Gaussian noise with zeros mean and variance one. After the change-point, the dependency between y_n and $\mathbf{x}_n$ changes to a new model $y_n = \mathbf{a}_1^T \mathbf{x}_n + \epsilon_n$, in which $\mathbf{a}_1$ is not exactly known by the observer.

In this paper, we assume that $\mathbf{x}_n = [x_{1,n}, x_{2,n}, \ldots, x_{p,n}]^T \in \mathbb{R}^p$. $\mathbf{a}_0$ and $\mathbf{a}_1$ are also p dimensional real vectors. Moreover, we assume that $\mathbf{x}_n$ has an underlying probability distribution, whose probability density function (pdf) is denoted as $f(\mathbf{x})$. However, $f(\mathbf{x})$ is unknown to the observer. The observer only knows that the mean value of $\mathbf{x}$ is a zero vector, and each entry in the covariance matrix of $\mathbf{x}$ is finite.

To simplify the notation and derivations, we transform the model into a simpler but equivalent form. Since the pre-change coefficient $\mathbf{a}_0$ is known, we can set $y'_n = y_n - \mathbf{a}_0^T \mathbf{x}_n$. Hence, before change-point t, y'_n is normal distributed; after change-point t, $y'_n = (\mathbf{a}_1 - \mathbf{a}_0)^T \mathbf{x}_n + \epsilon_n$. Let $\mathbf{a} := \mathbf{a}_1 - \mathbf{a}_0$, the model mentioned above is equivalent to the following model

$$y_n = \begin{cases} \mathbf{0}^T \mathbf{x}_n + \epsilon_n & n < t \\ \mathbf{a}^T \mathbf{x}_n + \epsilon_n & n \geq t \end{cases}. \tag{1}$$

In practice, it is common that the change only modifies parts of the coefficients in the linear model. Hence, it is reasonable to assume that the post-change linear coefficients $\mathbf{a}$ only contain s non-zero elements, where s could be any number in $\{1, 2, \ldots, p\}$. For the detection problem, if the change happens to more components in $\mathbf{a}$ (i.e., if s is large), it will be easier to detect the change. Therefore, the case with a small s is of more practical interest. We assume that the value of s is known to the observer; however, our proposed algorithm can be easily extended to the case with an unknown s.

Let $\mathbf{a} = [a_1, a_2, \ldots, a_p]^T$ and let $\mathcal{A}$ be the domain of $\mathbf{a}$. To guarantee the change is detectable, the origin should be excluded from $\mathcal{A}$. Specifically, we assume that if the change happens to the i^{th} linear coefficient, then a_i locates in a compact set

$$\mathcal{A}_i = \{\underline{a}_i \leq a_i \leq \bar{a}_i\},$$

where $\underline{a}_i > 0$ for all $1 \leq i \leq p$. Otherwise, $a_i \in \bar{\mathcal{A}}_i := \{a_i = 0\}$. Hence, $\mathcal{A}$ can be written as

$$\mathcal{A} = \cup_{(i_1,\ldots,i_p)\in\mathcal{P}} (\mathcal{A}_{i_1} \times \cdots \times \mathcal{A}_{i_s} \times \bar{\mathcal{A}}_{i_{s+1}} \times \cdots \times \bar{\mathcal{A}}_{i_p}),$$

where $\mathcal{P}$ is the set of all permutations of $\{1, 2, \ldots, p\}$ with cardinality s.

The observer aims to design an online algorithm to detect the change-point t via his sequential observations $(y_n, \mathbf{x}_n), n = 1, 2, \ldots$. Let τ be the stopping time at which the change is declared by the observer. We formulate the above problem in the framework of non-Bayesian quickest detection. In particular, we consider Lorden's setup:

$$\begin{aligned} \text{minimize}_{\tau} \quad & \text{WADD}(\tau; \mathbf{a}) := \\ & \qquad \sup_{t\geq 1} \operatorname{esssup} \mathbb{E}_t^{\mathbf{a}}[(\tau - t + 1)^+ | \mathcal{F}_{t-1}], \\ \text{subject to} \quad & \text{ARL2FA}(\tau) := \mathbb{E}_{\infty}[\tau] \geq \gamma, \end{aligned} \tag{2}$$

where $\mathbb{E}_t^{\mathbf{a}}$ is the expectation under the probability measure that change occurs at t with the post change linear coefficient being $\mathbf{a}$, $\mathbb{E}_{\infty}$ is the expectation under the probability measure that change never happens (i.e., $t = \infty$), and $\mathcal{F}_{t-1}$ is the sigma field generated by $\{(\mathbf{x}_n, y_n)\}_{n=1}^{t-1}$.

We use k, m and n as time index for observations, and use i and j as component index for vectors. For example, the i^{th} component in observation $\mathbf{x}_n$ is denoted as $x_{i,n}$.

3. LOWER BOUND OF THE DETECTION DELAY

Let $f_0(\mathbf{x}_n, y_n)$ be the joint pdf of $(\mathbf{x}_n, y_n)$ before change-point t, and let $f_1(\mathbf{x}_n, y_n; \mathbf{a})$ be the joint pdf after change-point t with the linear coefficient being $\mathbf{a}$. For any given $\mathbf{a}$, the likelihood ratio can be calculated as

$$\begin{aligned} L_n(\mathbf{a}) &:= \frac{f_1(\mathbf{x}_n, y_n; \mathbf{a})}{f_0(\mathbf{x}_n, y_n)} = \frac{f_1(y_n | \mathbf{x}_n; \mathbf{a}) f(\mathbf{x}_n)}{f_0(y_n) f(\mathbf{x}_n)} \\ &= \frac{\exp\{-\frac{1}{2}(y_n - \mathbf{a}^T \mathbf{x}_n)^2\}}{\exp\{-\frac{1}{2} y_n^2\}} \\ &= \exp\left\{\mathbf{a}^T \mathbf{x}_n y_n - \frac{1}{2} \mathbf{a}^T \mathbf{x}_n \mathbf{x}_n^T \mathbf{a}\right\}. \end{aligned} \tag{3}$$

Furthermore, the Kullback-Leibler (KL) divergence can be calculated as

$$\begin{aligned} D(f_1, f_0; \mathbf{a}) &:= \mathbb{E}^{\mathbf{a}}\left[\log L_n(\mathbf{a})\right] \\ &= \mathbf{a}^T \mathbb{E}^{\mathbf{a}}[\mathbf{x}_n y_n] - \frac{1}{2} \mathbf{a}^T \mathbb{E}^{\mathbf{a}}\left[\mathbf{x}_n \mathbf{x}_n^T\right] \mathbf{a} \\ &= \frac{1}{2} \mathbf{a}^T \mathbf{R} \mathbf{a} \\ &= \frac{1}{2} \sum_{i=1}^{p} a_i^2 r_{i,i} + \frac{1}{2} \sum_{i\neq j} a_i a_j r_{i,j}, \end{aligned} \tag{4}$$

where $r_{i,j}$ is the element of covariance matrix $\mathbf{R}$. A lower bound of the worst case detection delay is given in the following theorem.

Theorem 3.1. (Theorem 1 in [4]) *For any* $\mathbf{a} \in \mathcal{A}$*, as* $\gamma \to \infty$*, we have*

$$\inf\{\mathit{WADD}(\tau, \mathbf{a}) : \mathit{ARL2FA}(\tau) \geq \gamma\} \geq \frac{|\log \gamma|}{D(f_1, f_0; \mathbf{a})}(1 + o(1)). \quad (5)$$

It is also shown in [4] that the GLR CUSUM algorithm is asymptotically (in the sense $\gamma \to \infty$) optimal for all $\mathbf{a} \in \mathcal{A}$. The GLR CUSUM statistic is given as

$$\begin{aligned} C_n &:= \max_{1 \leq m \leq n} \frac{\sup_{\mathbf{a} \in \mathcal{A}} \prod_{k=m}^{n} f_1(\mathbf{x}_k, y_k; \mathbf{a})}{\prod_{k=m}^{n} f_0(\mathbf{x}_k, y_k)} \\ &= \max_{1 \leq m \leq n} \sup_{\mathbf{a} \in \mathcal{A}} \prod_{k=m}^{n} L_k(\mathbf{a}), \end{aligned} \quad (6)$$

and the corresponding stopping rule is

$$\tau_{GLR} := \min\{n \geq 0 : C_n \geq B\} \quad (7)$$

for a properly chosen threshold B.

Unfortunately, GLR CUSUM is computationally infeasible for our problem. Specifically, for each $m \in \{1, \ldots, n\}$, GLR CUSUM requires to estimate $\mathbf{a}$ by solving

$$\sup_{\mathbf{a} \in \mathcal{A}} \prod_{k=m}^{n} L_k(\mathbf{a}) = \sup_{\mathbf{a} \in \mathcal{A}} \prod_{k=m}^{n} \frac{\exp\{-\frac{1}{2}(y_k - \mathbf{a}^T \mathbf{x}_k)^2\}}{\exp\{-\frac{1}{2} y_k^2\}}, \quad (8)$$

which is equivalent to solve

$$\inf_{\mathbf{a} \in \mathcal{A}} \sum_{k=m}^{n} (y_k - \mathbf{a}^T \mathbf{x}_k)^2 \quad \text{for } m = 1, \ldots, n. \quad (9)$$

We note that $\mathbf{a}$ only contains s non-zero elements. It is known that to find an $s-$sparse solution of an underdetermined system is NP hard [1].

4. A LOW COMPLEXITY ALGORITHM

In this section, we first propose a low complex algorithm for the change detection, and then present an upper bound for its detect delay. The proposed algorithm is described as follows:

$$W_i(m, n; a_i) := 2a_i \sum_{k=m}^{n} x_{i,k} y_k - a_i^2 \sum_{k=m}^{n} x_{i,k}^2, \quad \text{for } 1 \leq i \leq p, \quad (10)$$

$$U(m, n) := \sup_{\mathbf{a} \in \mathcal{A}} \sum_{i=1}^{s} W_i(m, n; a_i), \quad (11)$$

$$C_n := \sup_{1 \leq m \leq n} U(m, n), \quad (12)$$

$$\tau_{ps} := \inf\{n \geq 0 : C_n \geq \log B\}. \quad (13)$$

[1]Due to the additional constraint that $a_i \in \mathcal{A}_i$ if the change happens to the i^{th} component, it is difficult to say whether the l_1 relaxation technique works for above problem or not. However, even if the l_1 relaxation works, the computational complexity of GLR CUSUM is still very high as the observer needs to solve n LASSOs at each time slot n.

We first provide some intuitive explanations of the proposed algorithm. In the linear regression model, y_k depends on $x_{i,k}$ if $a_i \neq 0$; hence it is reasonable to use the correlation between y_k and $x_{i,k}$ for the detection purpose. When the components of $\mathbf{x}_k$ are independent, we note that

$$\mathbb{E}^{\mathbf{a}}[W_i(m, n; a_i)] = -\mathbb{E}_{\infty}[W_i(m, n; a_i)] = (n - m + 1) a_i^2 r_{i,i}.$$

That is, W_i has a negative trend before the change and has a positive trend after the change. This property of W_i can be used to construct the detection statistic.

In order to take all changing components in $\mathbf{a}$ into consideration, the observer may want to sum up all positive W_i's together. This idea is realized by the supremum operator in (11). Specifically, before the change happens, all W_i's tend to be negative; hence U is negative on average and the false alarm is controlled. On the other hand, if the change occurs, s out of p W_i's tend to turn positive; hence the observer sums up s most likely components to speed up the detection procedure. Since the change-point t is unknown, (12) searches over all possible time instants up to the current time instant. This is a classic technique to construct a quickest detection statistic from a one-sided SPRT statistic [9].

In terms of implementation, the observer runs p parallel procedures to calculate the correlation between y_k and each individual component in $\mathbf{x}_k$, and then sums up the results by selecting the best $\mathbf{a}$ within the feasible set $\mathcal{A}$. Therefore, we term the proposed algorithm as parallel-sum algorithm.

The proposed parallel-sum algorithm can be easily computed. Let

$$a_i^* = \frac{\sum_{k=m}^{n} x_{i,k} y_k}{\sum_{k=1}^{n} x_{i,k}^2}. \quad (14)$$

It is easy to see $W_i(m, n; a_i)$ achieves the maximum at a_i^* if no constraint on a_i is considered. Let

$$\hat{a}_i = \underline{a}_i \mathbf{1}_{\{a_i^* < \underline{a}_i\}} + a_i^* \mathbf{1}_{\{\underline{a}_i \leq a_i^* \leq \bar{a}_i\}} + \bar{a}_i \mathbf{1}_{\{a_i^* > \bar{a}_i\}},$$

and let $\hat{\mathbf{a}}^* = [\hat{a}_1^*, \hat{a}_2^*, \ldots, \hat{a}_p^*]^T$ be the optimal solution in (11). Denote the order statistics of $\{W_i(m, n; \hat{a}_i)\}_{i=1}^{p}$ as

$$W_{(1)}(m, n; \hat{a}_{(1)}) \geq W_{(2)}(m, n; \hat{a}_{(2)}) \geq \cdots \geq W_{(p)}(m, n; \hat{a}_{(p)}).$$

It is easy to see the optimal solution is given as

$$\hat{a}_i^* = \begin{cases} \hat{a}_i & \text{if } W_i(m, n; \hat{a}_i) \geq W_{(s)}(m, n; \hat{a}_{(s)}) \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

Hence, the whole calculation procedure is very simple for each given m.

One potential concern of the parallel-sum algorithm is that the computation complexity increases as n increases. To deal with this difficulty, we may replace $\sup_{1 \leq m \leq n}$ in (12) by $\sup_{n - w_a \leq m \leq n}$, that is, we adopt a time window with length w_a to truncate the proposed algorithm. The window limited idea was firstly purposed in [10] and then studied for the GLR CUSUM algorithm in [4, 8]. In this paper, we do not discuss the window limited algorithm in detail.

The performance of parallel-sum algorithm is presented in the following theorem:

Theorem 4.1. *By setting*

$$\log B = 4s \log \left(p + p\sqrt{2s/\pi} + \mathbb{E}[N_{\max}] \right) + 4s \log \gamma,$$

in which N_{max} is a finite random variable whose distribution relies on $f(\mathbf{x})$, one can guarantee that

$$ARL2FA[\tau_{ps}] \geq \gamma \tag{16}$$

and the detection delay is bounded by

$$WADD(\tau_{ps}; \mathbf{a}) \leq \frac{4s \log \gamma}{\sum_{i=1}^{p} a_i^2 r_{i,i} + 2\sum_{i \neq j} a_i a_j r_{i,j}} (1 + o(1)). \tag{17}$$

as $\gamma \to \infty$.

Proof Outline: This theorem can be proved by exploring the relationship between non-Bayesian quickest detection and one-sided SPRT. In particular, it suffices to consider the detection delay $\mathbb{E}_1^{\mathbf{a}}[\tau_1]$ and the Type I error probability $P_\infty(\tau_1 < \infty)$ for $\tau_1 = \inf\{n \geq 0 : U(1,n) \geq \log B\}$ in the one-sided SPRT. By Lemma 1 in [9], we have $WADD(\tau_{ps}; \mathbf{a}) \leq \mathbb{E}^{\mathbf{a}}[\tau_1]$ and $ARL2FA[\tau_{ps}] \geq 1/P_\infty(\tau_1 < \infty)$.

To establish an upper bound of $\mathbb{E}_1^{\mathbf{a}}[\tau_1]$, we can construct a random walk $\tilde{U}(1,n) := \sum_{k=1}^{n} Y_k$ with $Y_k := 2y_k \sum_{i=1}^{p} a_i x_{i,k} - \sum_{i=1}^{p} (a_i x_{i,k})^2$ for any given $\mathbf{a}$. Let $\tilde{\tau}_1$ be the first stopping time that $\tilde{U}_n$ hits $\log B$. Wald's identity indicates that $\mathbb{E}_1^{\mathbf{a}}[\tilde{\tau}_1] = |\log B| / \mathbb{E}[Y_1](1 + o(1))$. Moreover, we note $\mathbb{E}_1^{\mathbf{a}}[\tau_1] \leq \mathbb{E}_1^{\mathbf{a}}[\tilde{\tau}_1]$ since $\tilde{U}(1,n)$ is dominated by $U(1,n)$. As a result, we have

$$\mathbb{E}_1^{\mathbf{a}}[\tau_1] \leq \frac{|\log B|}{\sum_{i=1}^{p} a_i^2 r_{i,i} + 2\sum_{i \neq j} a_i a_j r_{i,j}} (1 + o(1)).$$

The most challenging part is to establish an upper bound for the error probability. By mathematical manipulations, we can show that the error probability has an exponential tail

$$P_\infty(\tau_1 < \infty) \leq \left(p + p\sqrt{2s/\pi} + \mathbb{E}[N_{max}] \right) B^{-\frac{1}{4s}}.$$

Then, the desired conclusion can be obtained by setting a proper threshold B such that the error probability is bounded by $1/\gamma$. Due to space limitation, the rigorous proof is omitted in this paper.

5. SIMULATION

In this section, we provide a numerical example to illustrate the results obtained in our paper. In this numerical example, we assume that $p = 15$ and $s = 3$, the true post-change linear coefficient $\mathbf{a}$ is given as $a_1 = 0.8$, $a_2 = 0.65$, $a_3 = 0.5$, and $a_i = 0$ for the rest of components in $\mathbf{a}$. The observer has no knowledge about $\mathbf{a}$ except that $\mathcal{A}_i = [0.4, 2.5]$ for all $1 \leq i \leq p$. In the simulation, we set that $\mathbf{x}$ has diagonal covariance matrix $\mathbf{R}$, which is randomly selected as $\mathbf{R} = \text{diag}[1.32, 1.18, 1.04, 0.93, 0.86, 0.84, 0.71, 0.64, 0.52, 0.42, 0.39, 0.28, 0.17, 0.14, 0.03]$.

In this simulation, we illustrate the relationship between WADD and ARL2FA for the proposed parallel-sum algorithm. The simulation result is shown in Figure 1. In this simulation, we numerically calculate the performance of the parallel-sum algorithm for two different distributions of $\mathbf{x}$. In particular, the blue solid line with squares is the performance of the parallel-sum algorithm when $\mathbf{x}$ is normally distributed with zero mean. The green solid line with diamonds is the performance of the proposed algorithm when $\mathbf{x}$ is set be a Possion distributed random vector (the mean of $\mathbf{x}$ is shifted to zero). In Figure 1, the black dot-dash line is the lower bound of WADD for all detection algorithms, which is presented in (5). The black dash-line is the upper bound of the parallel-sum algorithm, which is presented in (17). From the simulation, we can see that the green line and the blue line are close to each other, hence the performance of the parallel-sum algorithm is robust over the distribution of $\mathbf{x}$. This result is consistent with our theoretical analysis presented in this paper as the proposed algorithm does not rely on the distribution of $\mathbf{x}$. In addition, the parallel-sum algorithm is not asymptotically optimal since it diverges from the lower bound as γ increases. However, we note that the detection delay of the parallel-sum algorithm still increases almost linearly with $\log \gamma$, and the computation complexity of this algorithm is extremely low.

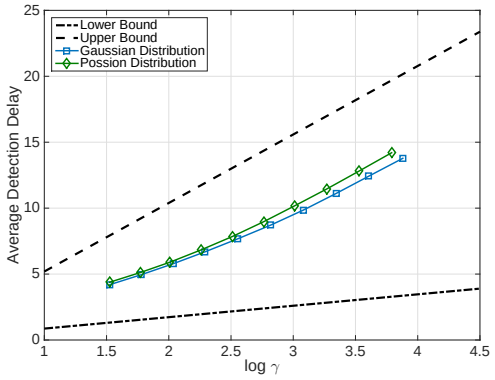

Fig. 1. WADD versus ARL2FA when $p = 15$, $s = 3$

6. CONCLUSION

In this paper, we have considered the problem of quickly detecting an abrupt change in the linear model. We have proposed a low complexity online algorithm, namely the parallel-sum algorithm. We have presented an asymptotic upper bound on the worst case average detection delay under a given average run length to false alarm constraint for the proposed algorithm.

7. REFERENCES

[1] P. Perron, *Dealing With Structutral Breaks Palgrave Handbook of Econometrics*. New York, NY, USA: Palgrave Macmillan, 2006.

[2] N. Zhang and D. Siegmund, "A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data," *Biometrics*, vol. 63, pp. 22–32, Mar. 2007.

[3] B. Zhang, J. Geng, and L. Lai, "Multiple change-points estimation in linear regression models via sparse group lasso," *IEEE Trans. Signal Processing*, vol. 63, pp. 2209–2224, May 2015.

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

[5] I. V. Nikiforov, "A suboptimal quadratic change detection scheme," *IEEE Trans. Inform. Theory*, vol. 46, no. 6, pp. 2095–2017, 2000.

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

[7] I. V. Nikiforov, "A simple change detection scheme," *Signal Processing*, vol. 81, pp. 149–172, 2001.

[8] T. L. Lai, "Sequential changepoint detection in quality control and dynamical systems," *Journal of the Royal Statistical Society*, vol. 57, pp. 613–658, 1995.

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

[10] A. S. Willsky and H. L. Jones, "A generalized likelihood ratio approach to detection and estimation of jumps in linear systems," *IEEE Trans. Automatic Control*, vol. AC-21, pp. 108–112, Feb. 1976.