Uncertainty-Aware Quickest Change Detection

Yancheng Huang[1], **Kai Yang**[1 *], **Chengbo Qiu**[1], **Jiangfan Zhang**[2] and **Xiaodong Wang**[3]
[1]Tongji University
[2]Missouri University of Science and Technology
[3]Columbia University

Abstract

Detecting change points in time series data as quickly as possible, i.e., quickest change detection (QCD), plays an essential role in numerous real-world applications, ranging from network security to financial fraud detection. The overarching goal of this paper is to tackle the QCD problem in linear regression models through the lens of nonlinear optimization methods. Conventional approaches presuppose precise knowledge of the linear regression model, which could be impractical given the presence of estimation and implementation errors. To address this limitation, we first employ the generalized CUSUM method to tackle this problem, which, however, is practically infeasible since its computational complexity increases exponentially with respect to the dimension of the observation vector. As a remedy, we reformulate the test statistic of generalized CUSUM as a Mixed Integer Quadratic Program (MIQP) problem. And then two efficient algorithms are proposed based on Semidefinite Relaxation and branch and bound methods, respectively. Finally, extensive experiments are carried out to underscore the significant benefits of employing SDP relaxation and branch and bound for our quickest change detection problem.

1 Introduction

Based on sequential observations, the quickest change detection problem is to detect the distribution changes as quickly as possible, subject to false alarm constraints. And the QCD problem in linear regression models arises in a variety of fields [Liang *et al.*, 2016; Yang *et al.*, 2016; Dou *et al.*, 2019] including detecting false data injection attacks in smart grids, and identifying anomalies in communication networks as well as IT systems. Compared to the general QCD problems, the QCD problem in linear regression models is particularly challenging because the dynamics of the linear systems itself can cause significant changes in the distribution of observed data, even without external changes. Existing studies on QCD problem in linear regression models [Zhang and Wang, 2021; Huang *et al.*, 2011; Li *et al.*, 2014] presuppose the precise knowledge of the linear regression model. However, in practical applications, it is recognized that the parameters of the acquired linear regression model may exhibit inaccuracies, uncertainties, or temporal fluctuations. Hence, we develop a robust QCD approach to address the presence of uncertain parameters. This holds significant importance, given the inherent challenges associated with acquiring an exact linear regression model in numerous real-world scenarios. The detailed contributions of this paper are summarized as follows:

- The QCD problem in uncertain linear regression models is formulated as an MIQP. And then two efficient algorithms, named SDPCUSUM and BBCUSUM, have been proposed. To the best of our knowledge, this is the first systematic approach that formulates the QCD problem in uncertain linear regression models as an MIQP and subsequently tackles it via SDP and branch and bound methods.
- We conduct extensive experiments to elucidate the effectiveness of the proposed algorithms, which achieve close-to-optimal performance in both deterministic and uncertain linear regression models.

2 Related Work

In the quickest change detection, the objective of a sequential change detector is to minimize the expected average detection delay while satisfying the false alarm constraints. The commonly used performance metric is the worst-case average detection delay, which was proposed by Lorden [Lorden, 1971]. Specifically, let the random variable Γ represent the stopping time at which a change is declared. The Lorden's worst-case average detection delay (ADD) is defined as:

$$J(\Gamma) \triangleq \sup_{\tau} \operatorname*{ess\,sup}_{\mathcal{F}_\tau} \mathbb{E}_\tau \left[(\Gamma - \tau)^+ \mid \mathcal{F}_\tau \right], \qquad (1)$$

where $\mathcal{F}_\tau$ is a sigma-algebra that is generated by all the observations made up to time instant τ. $\mathbb{E}_\tau$ represents the expectation of the detection delay given that a change happens at time instant τ and is evaluated conditioned on the $\mathcal{F}_\tau$. The essential supremum $\operatorname{ess\,sup}$ is obtained over $\mathcal{F}_\tau$, which represents the least favorable history of observations leading up to the change point for the detection delay. The supremum

*Corresponding author

is obtained over τ, indicating that the change occurs at the point where the detection delay is maximized. In summary, $J(\Gamma)$ is called the worst-case delay since it characterizes the least favorable change point and the least favorable history of observations up to the change point.

In addition to detecting changes as quickly as possible, the detector also needs to avoid frequent false alarms when no change occurs. To this end, the running length of the sequential change detector in the absence of a change needs to be large enough. And the quickest change detection is formulated as follows:

$$\inf_{\Gamma} J(\Gamma) \text{ subject to } \mathbb{E}_\infty[\Gamma] \geq \beta. \tag{2}$$

The false alarm period (FAP), denoted as $\mathbb{E}_\infty[\Gamma]$, represents the average stopping time when no change occurs at all ($\tau = \infty$), while β is the required lower bound on the FAP. Let $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ denote the parameter in the pre-distribution f_u and post-distribution f_a, respectively. If both $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ are known, then the well-known CUSUM algorithm [Moustakides, 1986] is the optimal solution to the above minimax problem. The log likelihood ratio(LLR) at time t can be written as

$$\ell_t = \ln \frac{f_a\left(\mathbf{x}^{(t)} \mid \boldsymbol{\theta}_1\right)}{f_u\left(\mathbf{x}^{(t)} \mid \boldsymbol{\theta}_2\right)}. \tag{3}$$

In the CUSUM algorithm, the LLR is considered as the test statistic and is accumulated over time. The change is declared if the accumulated test statistic exceeds a predefined threshold h:

$$T_{\mathrm{CUSUM}} = \min\left\{K : \max_{1 \leq k \leq K} \sum_{t=k}^{K} \ell_t \geq h\right\}. \tag{4}$$

The value of k which maximizes the test statistic in (4) can be considered as an estimate of change point. Since (4) can be written in a recursive way, in practice CUSUM can be implemented with low complexity.

Basesd on the CUSUM algorithm, many algorithms [Siegmund and Venkatraman, 1995; Brodsky and Darkhovsky, 2005; Mei, 2006] have also been proposed in cases where pre- or post-change distributions have unknown parameters. However, none of these can be used to tackle our problem since the pre- and post-change distributions of these work remain constant over time. One typical problem that can be formulated as the quickest QCD problem in linear regression model is the detection of false data injection attacks(FDIA) in smart grid system. Based on QCD, there have been a large number of algorithms have been developed to detect various forms of cyber-attacks in smart grids [Huang *et al.*, 2011; Li *et al.*, 2014; Huang *et al.*, 2014; Kurt *et al.*, 2018b]. However, all these methods presuppose the precise knowledge of the linear regression model and can not be used in the presence of uncertainty in the linear regression model.

3 Problem Statement

In this section, we first detail our uncertain linear regression model and several basic assumptions, and then we formulate the pre- and post-change model in our QCD problem. The uncertain linear regression model is as follows:

$$\mathbf{x}^{(t)} = \mathbf{H}\boldsymbol{\theta}^{(t)} + \mathbf{n}^{(t)}, \tag{5}$$

where $\mathbf{x}^{(t)} \in \mathbb{R}^M$ is the known observation vector and $\boldsymbol{\theta}^{(t)} \in \mathbb{R}^N$ is the vector of unknown parameters. $\mathbf{H} \in \mathbb{R}^{M\times N}$ is the system matrix modeling the linear relationship between $\mathbf{x}^{(t)}$ and $\boldsymbol{\theta}^{(t)}$. And it is assumed to belong to an uncertain set $\mathcal{S}$, i,e., $\mathbf{H} \in \mathcal{S}$. The dimension of the observation vector is assumed to be larger than that of the unknown parameters in order to provide necessary redundancy against the noise effect [Li *et al.*, 2014], i.e., $M > N$. The noise term $\mathbf{n}^{(t)}$ is assumed to be i.i.d. and obey gaussian distribution, i.e., $\mathbf{n}^{(t)} \overset{i.i.d.}{\sim} \mathcal{N}\left(\mathbf{0}, \sigma^2\mathbf{I}_M\right)$.

Assuming that the change occurs at time instant t_a, we model the change as a sequence of unknown vectors $\{\mathbf{a}^{t_a}, \mathbf{a}^{t_a+1}, \mathbf{a}^{t_a+2}, ...\}$ injected on $\mathbf{x}_t$. And then the pre- and post-change observation can written as

$$\begin{cases} \mathbf{x}^{(t)} = \mathbf{H}\boldsymbol{\theta}^{(t)} + \mathbf{n}^{(t)} & \text{if } t < t_a, \\ \mathbf{x}^{(t)} = \mathbf{H}\boldsymbol{\theta}^{(t)} + \mathbf{a}^{(t)} + \mathbf{n}^{(t)} & \text{if } t \geq t_a. \end{cases} \tag{6}$$

We further decompose $\mathbf{a}^{(t)}$ into two parts:

$$\mathbf{a}^{(t)} = \mathbf{H}\mathbf{c}^{(t)} + \boldsymbol{\mu}^{(t)}, \tag{7}$$

where $\mathbf{H}\mathbf{c}^{(t)}$ denote the component of $\mathbf{a}^{(t)}$ lies in the column space of $\mathbf{H}$, i.e., $\mathbf{H}\mathbf{c}^{(t)} \in \mathcal{R}(\mathbf{H})$. And $\boldsymbol{\mu}^{(t)}$ denote the component of $\mathbf{a}^{(t)}$ lies in the complementary space of $\mathcal{R}(\mathbf{H})$, i.e.,

$$\boldsymbol{\mu}^{(t)} = \mathbf{P}_{\mathbf{H}}^{\perp}\mathbf{a}^{(t)} \in \mathcal{R}^{\perp}(\mathbf{H}), \tag{8}$$

where

$$\mathbf{P}_{\mathbf{H}}^{\perp} \triangleq \mathbf{I} - \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T. \tag{9}$$

As illustrated in [Liu *et al.*, 2011], if the change occurs with $\boldsymbol{\mu}^{(t)} = \mathbf{0}$, then there is no way to distinguish $\boldsymbol{\theta}^{(t)}$ from $\boldsymbol{\theta}^{(t)} + \mathbf{c}^{(t)}$ since

$$\mathbf{x}^{(t)} = \mathbf{H}\boldsymbol{\theta}^{(t)} + \mathbf{a}^{(t)} + \mathbf{n}^{(t)} = \mathbf{H}(\boldsymbol{\theta}^{(t)} + \mathbf{c}^{(t)}) + \mathbf{n}^{(t)}. \tag{10}$$

Thus, we assume that $\boldsymbol{\mu}^{(t)} \neq \mathbf{0}$ when the change occurs. Usually not every components of $\mathbf{x}^{(t)}$ is affected by the change, so we let $\mathcal{U}^{(t)}$ denote the set of the index of nonzero elements of $\boldsymbol{\mu}^{(t)}$ at time instant t . For each $m \in \mathcal{U}^{(t)}$, we set upper bound ρ_H and lower bounds ρ_L on the value of $|\mu_m^{(t)}|$, where $\mu_m^{(t)}$ is the m-th element of $\boldsymbol{\mu}^{(t)}$.

Based on the above assumptions, the pre- and post-change model can be rewritten as

$$\begin{aligned} t < t_a : &\quad \mu_m^{(t)} = 0, m = 1, 2, \ldots, M, \\ t \geq t_a : &\quad \begin{cases} \rho_L \leq \left|\mu_m^{(t)}\right| \leq \rho_U, m \in \mathcal{U}^{(t)}, \\ \mu_m^{(t)} = 0, m \notin \mathcal{U}^{(t)}. \end{cases} \end{aligned} \tag{11}$$

And our goal is to detect the injected vector $\boldsymbol{\mu}^{(t)}$ as soon as possible after it actually occurs at time instant t_a. It is worth mentioning that all the assumptions we make in this section are widely used by previous studies [Zhang and Wang, 2021; Li *et al.*, 2014; Huang *et al.*, 2011].

4 Generalized CUSUM Algorithm

Considering the model in (6), the classic generalized CUSUM algorithm [Kurt *et al.*, 2018a] can be used to estimate the unknown parameters $\boldsymbol{\theta}^{(t)}$ and $\mathbf{a}^{(t)}$. Following [Zhang and Wang, 2021], the generalized CUSUM algorithm for our QCD problem in model (11) can be written in a recursive way as follows,

$$\begin{aligned} V_K &= \max_{1\le k\le K} \sum_{t=k}^{K} v_t \\ &= \max\{V_{K-1}, 0\} + v_K, \text{ with } V_0 = 0, \end{aligned} \tag{12}$$

where

$$\begin{aligned} v_t \triangleq \sup_{\mathcal{U}^{(t)}} & \sup_{\boldsymbol{\mu}^{(t)}:\left\{\rho_L \le \left|\mu_m^{(t)}\right| \le \rho_U\right\}_{m\in\mathcal{U}^{(t)}}, \boldsymbol{\mu}^{(t)}\in\mathcal{R}^{\perp}(\mathbf{H})} \frac{1}{2\sigma^2} \\ & \times \left\{2\left(\boldsymbol{\mu}^{(t)}\right)^T \boldsymbol{x}^{(t)} - \left\|\boldsymbol{\mu}^{(t)}\right\|_2^2\right\}. \end{aligned} \tag{13}$$

Once V_K is greater than the specified threshold h, the change is declared by the algorithm. The detailed derivation can be found in [Zhang and Wang, 2021].

It can be seen from (13) that the only difficulty of this problem is to calculate the test statistic v_t. As long as v_t is available for all t, through recursion we can calculate the value of all V_K. However, the calculation of v_t is pretty difficult. Specifically, solving v_t needs to optimize $\boldsymbol{\mu}^{(t)}$ and set $\mathcal{U}^{(t)}$. The constraints $\{\rho_L \le |\mu_m^{(t)}| \le \rho_U\}_{m\in\mathcal{U}^{(t)}}$ on $\boldsymbol{\mu}^{(t)}$ makes this problem non-convex, and the number of possible $\mathcal{U}^{(t)}$ is on the order of 2^M. Moreover, the system matrix $\mathbf{H}$ is not deterministic but belongs to an uncertain set. To address this issue, RGCUSUM proposed in [Zhang and Wang, 2021] relaxes the constraint $\boldsymbol{\mu}^{(t)} \in \mathcal{R}^{\perp}(\mathbf{H})$ and thereby the calculation of the v_t becomes computationally efficient. However, this relaxation may bring much performance loss and the optimal solution is not even feasible. And also, RGCUSUM is completely unable to handle the case where $\mathbf{H}$ is uncertain due to this relaxation.

5 Problem Formulation

In this section, we formulate the calculation of statistic v_t as an MIQP problem. We first derive the MIQP problem when the system matrix $\mathbf{H}$ can be completely determined, which is given in 5.1. And then it is further extended to the case where $\mathbf{H}$ is uncertain, which is provided in 5.2.

5.1 MIQP under Exact H

We first conduct the vectorization of set $\mathcal{U}^{(t)}$. Let the binary variable $\boldsymbol{u}^{(t)} \in \{0,1\}^M$ represent the set $\mathcal{U}^{(t)}$, so that $u_m^{(t)} = 1$ for $m \in \mathcal{U}^{(t)}$ and $u_m^{(t)} = 0$ for $m \notin \mathcal{U}^{(t)}$, where $u_m^{(t)}$ is the m-th element of $\boldsymbol{u}^{(t)}$. Second, we use the method used in [Schmidt *et al.*, 2009] to eliminate the absolute value in non-convex constraint $\{\rho_L \le |\mu_m^{(t)}| \le \rho_U\}_{m\in\mathcal{U}^{(t)}}$. Specifically, $\boldsymbol{\mu}^{(t)}$ is decomposed into two components $\boldsymbol{\mu}_+^{(t)}$ and $\boldsymbol{\mu}_-^{(t)}$, i.e, $\boldsymbol{\mu}^{(t)} = \boldsymbol{\mu}_+^{(t)} - \boldsymbol{\mu}_+^{(t)}$. And we enforce $\boldsymbol{\mu}_+^{(t)} \in \mathbb{R}^M \ge \mathbf{0}$, $\boldsymbol{\mu}_-^{(t)} \in \mathbb{R}^M \ge \mathbf{0}$, and their inner product $\boldsymbol{\mu}_+^{(t)}(\boldsymbol{\mu}_-^{(t)})^T = 0$.

Note that the complementary condition constraint $\boldsymbol{\mu}_+^{(t)}(\boldsymbol{\mu}_-^{(t)})^T = 0$ is still difficult. As a remedy, we employ an active ensemble strategy [Grossmann and Floudas, 1987] to reformulate this constraint as

$$\begin{aligned} &\boldsymbol{\mu}_+^{(t)} + \rho_U \boldsymbol{b}^{(t)} \le \rho_U \mathbf{1}, \\ &\boldsymbol{\mu}_-^{(t)} - \rho_U \boldsymbol{b}^{(t)} \le \mathbf{0}, \end{aligned} \tag{14}$$

where $\boldsymbol{b}^{(t)} \in \{0,1\}^M$ is an auxiliary binary vector. Combining all the above, we can reformulate (13) as the following MIQP problem when the system matrix can be exactly known.

$$\begin{aligned} &\min \frac{1}{2\sigma^2}\left\{\left\|\boldsymbol{\mu}_+^{(t)} - \boldsymbol{\mu}_-^{(t)}\right\|_2^2 - 2(\boldsymbol{\mu}_+^{(t)} - \boldsymbol{\mu}_-^{(t)})^T \boldsymbol{x}^{(t)}\right\} \\ &s.t. \quad \begin{cases} \mathbf{H}^T(\boldsymbol{\mu}_+^{(t)} - \boldsymbol{\mu}_-^{(t)}) = \mathbf{0} \\ \rho_L \boldsymbol{u}^{(t)} \le \boldsymbol{\mu}_+^{(t)} + \boldsymbol{\mu}_-^{(t)} \le \rho_U \boldsymbol{u}^{(t)} \\ \boldsymbol{\mu}_+^{(t)} + \rho_U \boldsymbol{b}^{(t)} \le \rho_U \mathbf{1} \\ \boldsymbol{\mu}_-^{(t)} - \rho_U \boldsymbol{b}^{(t)} \le \mathbf{0} \\ \boldsymbol{\mu}_+^{(t)} \ge \mathbf{0},\ \boldsymbol{\mu}_-^{(t)} \ge \mathbf{0} \end{cases} \\ &\text{var}: \boldsymbol{u}^{(t)} \in \{0,1\}^M, \boldsymbol{b}^{(t)} \in \{0,1\}^M, \boldsymbol{\mu}_+^{(t)}, \boldsymbol{\mu}_-^{(t)}. \end{aligned} \tag{15}$$

It is worth mentioning that the integer variables now include, in addition to $\boldsymbol{u}^{(t)}$, the newly introduced auxiliary variable $\boldsymbol{b}^{(t)}$.

5.2 MIQP under Uncertain H

In this subsection, by exploiting the robust optimization technique, we investigate the robust QCD problem in the presence of uncertainty in $\mathbf{H}$. We demonstrate that such a problem can also be formulated as an MIQP problem.

We assume that the uncertainty set of $\mathbf{H}$ is constraint-wise. Specifically, the uncertainties between different rows of $\mathbf{H}^T$ are decoupled. This assumption is commonly used and can be applied to many practical problems [Yang *et al.*, 2014; Bertsimas and Sim, 2004]. Let $\boldsymbol{h}_i$ denote the i-th column of $\mathbf{H}$. And each $\boldsymbol{h}_i$ belongs to an uncertainty set, i.e. $\boldsymbol{h}_i \in \mathcal{S}_i$. By simply relaxing the constraint $\mathbf{H}^T(\boldsymbol{\mu}_+^{(t)} - \boldsymbol{\mu}_-^{(t)}) = \mathbf{0}$, we get the new constraints as follows:

$$-\varepsilon_i \le \boldsymbol{h}_i^T(\boldsymbol{\mu}_+^{(t)} - \boldsymbol{\mu}_-^{(t)}) \le \varepsilon_i,\ \forall \boldsymbol{h}_i \in \mathcal{S}_i,\ 1 \le i \le N \tag{16}$$

where ε_i represents a small positive constant. And the above constraints can be equivalently transformed into several linear constraints, not only when $\mathcal{U}_i$ represents some simple uncertain sets, including interval-based uncertainty set, but also other general uncertain sets [Yang *et al.*, 2014], including polyhedron uncertainty set and D-norm uncertainty set. We take the general polyhedron uncertainty set as an example, i.e. $\mathcal{S}_i = \{\boldsymbol{h}_i \mid \mathbf{D}_i \boldsymbol{h}_i \le \boldsymbol{d}_i\},\ i = 1, \cdots, N$. According to [Yang *et al.*, 2014], by exploiting the robust optimization technique, the constraints (16) under the polyhedron uncertainty set can be equivalently reformulated as

$$\begin{aligned} &\boldsymbol{p}_i^T \boldsymbol{d}_i \le \varepsilon_i, \quad \mathbf{D}_i^T \boldsymbol{p}_i = \boldsymbol{\mu}_+^{(t)} - \boldsymbol{\mu}_-^{(t)}, \boldsymbol{p}_i \ge 0, 1 \le i \le N, \\ &\boldsymbol{n}_i^T \boldsymbol{d}_i \le \varepsilon_i, -\mathbf{D}_i^T \boldsymbol{n}_i = \boldsymbol{\mu}_+^{(t)} - \boldsymbol{\mu}_-^{(t)}, \boldsymbol{n}_i \ge 0, 1 \le i \le N. \end{aligned} \tag{17}$$

Thus, our MIQP problem when every column of $\mathbf{H}$ belongs to the polyhedral uncertainty set is as follows:

$$
\min \frac{1}{2\sigma^2}\left\{\left\|\boldsymbol{\mu}_+^{(t)}-\boldsymbol{\mu}_-^{(t)}\right\|_2^2-2(\boldsymbol{\mu}_+^{(t)}-\boldsymbol{\mu}_-^{(t)})^T\boldsymbol{x}^{(t)}\right\}
$$

$$
s.t. \quad \begin{cases}
\mathbf{D}_i^T\boldsymbol{p}_i=\boldsymbol{\mu}_+^{(t)}-\boldsymbol{\mu}_-^{(t)}, \quad \forall i\\
-\mathbf{D}_i^T\boldsymbol{n}_i=\boldsymbol{\mu}_+^{(t)}-\boldsymbol{\mu}_-^{(t)}, \quad \forall i\\
\boldsymbol{p}_i^T\boldsymbol{d}_i\le\varepsilon_i,\ \ \boldsymbol{n}_i^T\boldsymbol{d}_i\le\varepsilon_i, \quad \forall i\\
\rho_L\boldsymbol{u}^{(t)}\le\boldsymbol{\mu}_+^{(t)}+\boldsymbol{\mu}_-^{(t)}\le\rho_U\boldsymbol{u}^{(t)}\\
\boldsymbol{\mu}_+^{(t)}+\rho_U\boldsymbol{b}^{(t)}\le\rho_U\mathbf{1}\\
\boldsymbol{\mu}_-^{(t)}-\rho_U\boldsymbol{b}^{(t)}\le\mathbf{0}\\
\boldsymbol{\mu}_+^{(t)}\ge\mathbf{0},\ \ \boldsymbol{\mu}_-^{(t)}\ge\mathbf{0}\\
\boldsymbol{p}_i\ge\mathbf{0},\ \ \boldsymbol{n}_i\ge\mathbf{0}, \quad \forall i
\end{cases} \tag{18}
$$

$$
\text{var}: \boldsymbol{u}^{(t)},\boldsymbol{b}^{(t)},\boldsymbol{\mu}_+^{(t)},\boldsymbol{\mu}_-^{(t)},\{\boldsymbol{p}_i\},\{\boldsymbol{n}_i\}.
$$

The linear constraints introduced by the uncertainty of $\mathbf{H}$ do not change the nature of the problem, and the calculation of $\boldsymbol{v}_t$ in the presence of inaccurate $\mathbf{H}$ is still an MIQP problem.

6 Robust QCD Detectors

In this section, we propose two algorithms that can tackle the QCD problem in the uncertain linear regression models. One is called SDPCUSUM based on the Semidefinite Relaxation of the MIQP problem, and the other is called BBCUSUM based on the branch and bound method.

6.1 SDPCUSUM

In this subsection, we introduce our SDPCUSUM algorithm. The SDP relaxation has been proven to provide a relatively tighter bound in many problems [Goemans and Williamson, 1995; Ma *et al.*, 2002] and can be used to generate suboptimal solutions [Axehill *et al.*, 2010]. Therefore, the calculation of v_t at each time instant in SDPCUSUM consists of two steps. First, SDPCUSUM solves the SDP relaxation of the origin MIQP problem (18). Second, SDPCUSUM generates the suboptimal objective function value of problem (18) based on its SDP relaxation. Denote $[\boldsymbol{u}^{(t)},\boldsymbol{b}^{(t)}]\in\mathbb{R}^{2M}$ as the column vector created by stacking $\boldsymbol{u}^{(t)}$ and $\boldsymbol{b}^{(t)}$ together. The SDP relaxation of problem (18) is as follows :

$$
\min \ \frac{1}{2\sigma^2}\left\{\left\|\boldsymbol{\mu}_+^{(t)}-\boldsymbol{\mu}_-^{(t)}\right\|_2^2-2\left(\boldsymbol{\mu}_+^{(t)}-\boldsymbol{\mu}_-^{(t)}\right)^T\boldsymbol{x}^{(t)}\right\}
$$

$$
s.t. \quad \begin{cases}
\mathbf{D}_i^T\boldsymbol{p}_i=\boldsymbol{\mu}_+^{(t)}-\boldsymbol{\mu}_-^{(t)}, \quad \forall i\\
-\mathbf{D}_i^T\boldsymbol{n}_i=\boldsymbol{\mu}_+^{(t)}-\boldsymbol{\mu}_-^{(t)}, \quad \forall i\\
\boldsymbol{p}_i^T\boldsymbol{d}_i\le\varepsilon_i,\ \ \boldsymbol{n}_i^T\boldsymbol{d}_i\le\varepsilon_i, \quad \forall i\\
\rho_L\boldsymbol{u}^{(t)}\le\boldsymbol{\mu}_+^{(t)}+\boldsymbol{\mu}_-^{(t)}\le\rho_U\boldsymbol{u}^{(t)}\\
\boldsymbol{\mu}_+^{(t)}+\rho_U\boldsymbol{b}^{(t)}\le\rho_U\mathbf{1}\\
\boldsymbol{\mu}_-^{(t)}-\rho_U\boldsymbol{b}^{(t)}\le\mathbf{0}\\
\boldsymbol{A}_{ii}=u_i^{(t)},\ i=1,\ldots,M\\
\boldsymbol{A}_{ii}=b_{i-M}^{(t)},\ i=M+1,\ldots,2M\\
\boldsymbol{\mu}_+^{(t)}\ge\mathbf{0},\ \ \boldsymbol{\mu}_-^{(t)}\ge\mathbf{0}\\
\boldsymbol{p}_i\ge\mathbf{0},\ \ \boldsymbol{n}_i\ge\mathbf{0}, \quad \forall i\\
\begin{bmatrix}\boldsymbol{A} & [\boldsymbol{u}^{(t)},\boldsymbol{b}^{(t)}]\\ [\boldsymbol{u}^{(t)},\boldsymbol{b}^{(t)}]^T & 1\end{bmatrix}\succeq 0
\end{cases} \tag{19}
$$

$$
\text{var}: \ \boldsymbol{u}^{(t)},\ \boldsymbol{b}^{(t)},\ \boldsymbol{A},\ \boldsymbol{\mu}_+^{(t)},\ \boldsymbol{\mu}_-^{(t)},\{\boldsymbol{p}_i\},\{\boldsymbol{n}_i\},
$$

where $\boldsymbol{u}^{(t)}\in\mathbb{R}^M$, $\boldsymbol{b}^{(t)}\in\mathbb{R}^M$, $\boldsymbol{A}\in\mathbb{R}^{2M\times 2M}$. Let $(\boldsymbol{u}^*,\boldsymbol{b}^*,\boldsymbol{A}^*,\boldsymbol{\mu}_+^*,\boldsymbol{\mu}_-^*)$ denote the optimal solution of (19). Based on this optimal solution of the SDP relaxation, we apply heuristic method to generate the suboptimal solution of the integer variables in (18). Specifically, denote w^* as the optimal solution of the integer variables in (19), i.e, $w^*=[\boldsymbol{u}^*,\boldsymbol{b}^*]$. We generate random vectors from a Gaussian distribution with mean $\boldsymbol{w}^*$ and variance $\boldsymbol{A}^*-\boldsymbol{w}^*\boldsymbol{w}^{*T}$. And then we round each element of the generated vector to 0 or 1. The procedure is described by the equations in the following,

$$
\bar{\boldsymbol{w}}\in\mathcal{N}\left(\boldsymbol{w}^*,\boldsymbol{A}^*-\boldsymbol{w}^*\boldsymbol{w}^{*T}\right), \tag{20}
$$

$$
\hat{\boldsymbol{w}}=\text{round}(\bar{\boldsymbol{w}}), \tag{21}
$$

where $\bar{\boldsymbol{w}}$ is the generated random vector and $\hat{\boldsymbol{w}}$ is the rounded vector of integer variables. The round represents a function that rounds each element of a vector to 0 or 1. With this known $\hat{\boldsymbol{w}}$, the MIQP problem (18) degenerates into a QP problem. After solving this QP problem, we can obtain the suboptimal solution of the continuous variables in problem (18). Preferably, a few $\hat{\boldsymbol{w}}$ are generated by (20)-(21) in practice. And for each sample, we solve the corresponding QP problem and obtain its optimal objective function value. The lowest optimal objective function value is kept as the suboptimal objective function value of problem (18). It is worth mentioning that the general SDP can be solved theoretically in polynomial time using the ellipsoid method. Moveover, since the SDP relaxation is a widely used optimization technique, there have been numerous studies focused on solving SDP problems faster, including the parallelizable numerical algorithm proposed in [Madani *et al.*, 2017], heuristics methods and approximate methods proposed in [Lemon *et al.*, 2016]. The choice of SDP solver depends on the practical application, which is beyond the scope of this paper.

6.2 BBCUSUM

Besides the SDP relaxation, branch and bound [Tziligakis, 1999; Yang *et al.*, 2008a] and cutting plane methods [Yang *et al.*, 2008b; Bürger *et al.*, 2012] are common approaches to handle nonlinear optimization problem with integer constraints. Therefore, BBCUSUM further employs the branch and bound method [Tziligakis, 1999; Yang *et al.*, 2008a] to tackle the MIQP problem under investigation. We use the SDP relaxation (19) to provide the lower bound of our problem. And we decide which variable to branch upon next based on the solution of the SDP relaxation in the course of the algorithm. Specifically, we choose the binary variable that is closest to 0.5, i.e., the most uncertain variable to branch upon next. The detailed algorithm is shown in $Algorithm$ 1. It is worth mentioning that the computational complexity of branch and bound is adaptive to accuracy requirements in practice. The selection of algorithm parameters such as search depth and search time all depends on practical requirements.

7 Numerical Results

In this section, we conduct extensive experiments on simulated dataset. The simulated dataset is generated according to

Table 1: Performance Comparison with exact $\mathbf{H}$

$(\rho_L, \rho_H, \sigma^2)$	False Alarm Period	Average Detection Delay			
		RGCUSUM	Adaptive CUSUM	SDPCUSUM	BBCUSUM
(0.8, 1.2, 0.25)	2000.0	185.5	1968.5	135.7	127.3
(0.5, 1.5, 0.1)	5000.0	190.8	4879.8	139.1	136.3

Table 2: Performance Comparison with Uncertain $\mathbf{H}$

$(\rho_L, \rho_H, \sigma^2)$	False Alarm Period	Average Detection Delay			
		RGCUSUM	Adaptive CUSUM	SDPCUSUM	BBCUSUM
(0.2, 1.5, 0.5)	500.0	396.7	488.3	357.2	343.7
(0.1, 3.0, 2.0)	500.0	449.1	495.6	409.2	404.2

the aforementioned model (6). Specifically, we set the $\mathbf{H}$ to be a 4-by-2 matrix

$$\mathbf{H} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix}. \tag{22}$$

The time-varying parameter $\boldsymbol{\theta}^{(t)}$ is set to be

$$\boldsymbol{\theta}^{(t)} = [1 + 0.5\sin(0.02\pi t)] \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \tag{23}$$

and the time-varying parameter $\boldsymbol{\mu}^{(t)}$ is set to be

$$\boldsymbol{\mu}^{(t)} = \begin{cases} [-1, 1, -1, 0]^T, & \text{if } t \text{ is odd }, \\ [1, 1, 0, -1]^T, & \text{if } t \text{ is even }. \end{cases} \tag{24}$$

It can be easily found that $\boldsymbol{\mu}^{(t)} \in \mathcal{R}^{\perp}(\mathbf{H})$ for all t. All experiments in this paper are conducted on a computer equipped with i7-9750H CPU, 16GB RAM.

7.1 Performance Comparison Under Exact H

In this subsection, we compare the proposed SDPCUSUM and BBCUSUM with other existing algorithms on our data. In this experiment, all algorithms use the accurate $\mathbf{H}$ for quickest change detection. The experimental results are shown in Table 1. The number of Monte Carlo runs is 500. It is seen from Table 1 that for any given average FAP under any (ρ_L, ρ_H), the average detection delay of our both algorithms is smaller than that of other algorithms, which implies our algorithms achieve better performance. For a fair comparison, we set (ρ_L,ρ_H) in RGCUSUM [Zhang and Wang, 2021] to the same values as in our detectors. The adaptive CUCUM [Huang *et al.*, 2011] shows the worst results due to the model mismatch. Specifically, the adaptive CUSUM algorithm assumes the state vector $\boldsymbol{\theta}^{(t)}$ obeys a Gaussian distribution with known mean and covariance. However, the distribution of $\boldsymbol{\theta}^{(t)}$ is unknown in our experiment. RGCUCUM also performs worse than both of our algorithms. This is expected since RGCUSUM used worse relaxation and the optimal solution of the relaxed problem is not projected back to the feasible region of the origin problem, which brings some performance loss. BBCUSUM achieves the best performance among all algorithms.

7.2 Performance Comparison Under Uncertain H

In this subsection, we compare the performance of SDPCUSUM and BBCUSUM with other algorithms under the case where $\mathbf{H}$ is estimated inaccurately. For SDPCUSUM and BBCUSUM, the accurate $\mathbf{H}$ is assumed to belong to a polyhedron uncertainty set. The setting of the inaccurate $\mathbf{H}$ and polyhedron uncertainty set can be found in Appendix B. The experimental results are shown in Table 2. It is seen from Table 2 that under different (ρ_L, ρ_H) and different σ^2 of noise, the average detection delay of SDPCUSUM and BBCUSUM are relative smaller, which implies that SDPCUSUM and BBCUSUM achieve better performance in the presence of inaccurate $\mathbf{H}$. The experimental results show the effectiveness of our methods in uncertain linear regression models.

8 Conclusion

This paper aim to tackle the quickest change detection problem in a linear regression model with uncertain parameters. We first formulate the statistic of generalized CUSUM as an MIQP problem. And then two efficient algorithms are proposed based on Semidefinite Relaxation and branch and bound methods, respectively. Finally, extensive experiments are carried out to underscore the significant benefits of employing SDP relaxation and branch and bound for the quickest change detection problem. To the best of our knowledge, this is the first systematic approach that considers the quickest change detection problem in an uncertain linear system.

References

[Axehill *et al.*, 2010] Daniel Axehill, Lieven Vandenberghe, and Anders Hansson. Convex relaxations for mixed integer predictive control. *Automatica*, 46(9):1540–1545, 2010.

[Bertsimas and Sim, 2004] Dimitris Bertsimas and Melvyn Sim. The price of robustness. *Operations research*, 52(1):35–53, 2004.

[Brodsky and Darkhovsky, 2005] Boris Brodsky and Boris Darkhovsky. Asymptotically optimal methods of change-point detection for composite hypotheses. *Journal of Statistical Planning and Inference*, 133(1):123–138, 2005.

[Bürger *et al.*, 2012] Mathias Bürger, Giuseppe Notarstefano, and Frank Allgöwer. Distributed robust optimization via cutting-plane consensus. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pages 7457–7463. IEEE, 2012.

[Dou *et al.*, 2019] Shaoyu Dou, Kai Yang, and H Vincent Poor. Pc2a: Predicting collective contextual anomalies via lstm with deep generative model. *IEEE Internet of Things Journal*, 6(6):9645–9655, 2019.

[Goemans and Williamson, 1995] Michel X Goemans and David P Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995.

[Grossmann and Floudas, 1987] Ignacio E Grossmann and Christodoulos A Floudas. Active constraint strategy for flexibility analysis in chemical processes. *Computers & Chemical Engineering*, 11(6):675–693, 1987.

[Huang *et al.*, 2011] Yi Huang, Husheng Li, Kristy A Campbell, and Zhu Han. Defending false data injection attack on smart grid network using adaptive cusum test. In *2011 45th Annual Conference on Information Sciences and Systems*, pages 1–6. IEEE, 2011.

[Huang *et al.*, 2014] Yi Huang, Jin Tang, Yu Cheng, Husheng Li, Kristy A Campbell, and Zhu Han. Real-time detection of false data injection in smart grid networks: An adaptive cusum method and analysis. *IEEE Systems Journal*, 10(2):532–543, 2014.

[Kurt *et al.*, 2018a] Mehmet Necip Kurt, Yasin Yılmaz, and Xiaodong Wang. Distributed quickest detection of cyber-attacks in smart grid. *IEEE Transactions on Information Forensics and Security*, 13(8):2015–2030, 2018.

[Kurt *et al.*, 2018b] Mehmet Necip Kurt, Yasin Yılmaz, and Xiaodong Wang. Real-time detection of hybrid and stealthy cyber-attacks in smart grid. *IEEE Transactions on Information Forensics and Security*, 14(2):498–513, 2018.

[Lemon *et al.*, 2016] Alex Lemon, Anthony Man-Cho So, Yinyu Ye, et al. Low-rank semidefinite programming: Theory and applications. *Foundations and Trends® in Optimization*, 2(1-2):1–156, 2016.

[Li *et al.*, 2014] Shang Li, Yasin Yılmaz, and Xiaodong Wang. Quickest detection of false data injection attack in wide-area smart grids. *IEEE Transactions on Smart Grid*, 6(6):2725–2735, 2014.

[Liang *et al.*, 2016] Gaoqi Liang, Junhua Zhao, Fengji Luo, Steven R Weller, and Zhao Yang Dong. A review of false data injection attacks against modern power systems. *IEEE Transactions on Smart Grid*, 8(4):1630–1638, 2016.

[Liu *et al.*, 2011] Yao Liu, Peng Ning, and Michael K Reiter. False data injection attacks against state estimation in electric power grids. *ACM Transactions on Information and System Security (TISSEC)*, 14(1):1–33, 2011.

[Lorden, 1971] Gary Lorden. Procedures for reacting to a change in distribution. *The annals of mathematical statistics*, pages 1897–1908, 1971.

[Ma *et al.*, 2002] Wing-Kin Ma, Timothy N Davidson, Kon Max Wong, Zhi-Quan Luo, and Pak-Chung Ching. Quasi-maximum-likelihood multiuser detection using semi-definite relaxation with application to synchronous cdma. *IEEE transactions on signal processing*, 50(4):912–922, 2002.

[Madani *et al.*, 2017] Ramtin Madani, Abdulrahman Kalbat, and Javad Lavaei. A low-complexity parallelizable numerical algorithm for sparse semidefinite programming. *IEEE Transactions on Control of Network Systems*, 5(4):1898–1909, 2017.

[Mei, 2006] Yajun Mei. Sequential change-point detection when unknown parameters are present in the pre-change distribution. 2006.

[Moustakides, 1986] George V Moustakides. Optimal stopping times for detecting changes in distributions. *the Annals of Statistics*, 14(4):1379–1387, 1986.

[Schmidt *et al.*, 2009] Mark Schmidt, Glenn Fung, and Romer Rosales. Optimization methods for l1-regularization. *University of British Columbia, Technical Report TR-2009-19*, 2009.

[Siegmund and Venkatraman, 1995] David Siegmund and ES Venkatraman. Using the generalized likelihood ratio statistic for sequential detection of a change-point. *The Annals of Statistics*, pages 255–271, 1995.

[Tziligakis, 1999] Constantine Nikolaos Tziligakis. *Relaxation and exact algorithms for solving mixed integer-quadratic optimization problems*. PhD thesis, Massachusetts Institute of Technology, 1999.

[Yang *et al.*, 2008a] Kai Yang, Xiaodong Wang, and Jon Feldman. A new linear programming approach to decoding linear block codes. *IEEE Transactions on Information Theory*, 54(3):1061–1072, 2008.

[Yang *et al.*, 2008b] Kai Yang, Yihong Wu, Jianwei Huang, Xiaodong Wang, and Sergio Verdú. Distributed robust optimization for communication networks. In *IEEE INFOCOM 2008-The 27th Conference on Computer Communications*, pages 1157–1165. IEEE, 2008.

[Yang *et al.*, 2014] Kai Yang, Jianwei Huang, Yihong Wu, Xiaodong Wang, and Mung Chiang. Distributed robust optimization (DRO), part I: Framework and example. *Optimization and Engineering*, 15(1):35–67, 2014.

[Yang *et al.*, 2016] Kai Yang, Ruilin Liu, Yanjia Sun, Jin Yang, and Xin Chen. Deep network analyzer (dna): A big data analytics platform for cellular networks. *IEEE Internet of Things Journal*, 4(6):2019–2027, 2016.

[Zhang and Wang, 2021] Jiangfan Zhang and Xiaodong Wang. Low-complexity quickest change detection in linear systems with unknown time-varying pre-and post-change distributions. *IEEE Transactions on Information Theory*, 67(3):1804–1824, 2021.

A BBCUSUM

In this section, we summarize the details of the proposed BB-CUSUM as follows,

Algorithm 1 BBCUSUM for MIQP given in (18)

Input: MIQP problem P_0 with integer variable set X and continuous variable set U
Initialize $L = \{P_0\}$ with root node, set $UB = +\infty$.
Select current node $P_c \leftarrow P_0$.
repeat
  Solve the SDP relaxation of P_c
  if *SDP is feasible and the optimal value of SDP relaxation* $J(P_c) \leq UB$ **then**
    if *SDP solution is not integer-feasible* **then**
      Select branching variable $v \in U$ based on branching strategy.
      Create subproblem P_u by adding constraint $v = 0$ and P_l by adding constraint $v = 1$.
      Append $\{P_l, P_u\}$ to L.
    else
      $UB \leftarrow J(P_c)$.
      store the current solution $(X^*, U^*) \leftarrow (\bar{X}, \bar{U})$.
    end
  end
  Remove current node P_c from the to-do list in L.
  Select the next node based on depth-first(last node in list L).
until *termination*;
Output: MIQP solution vector (X^*, U^*)

B Experimental Settings

In section 7.2, we conduct experiments to compare the performance of SDPCUSUM and BBCUSUM with other algorithms on our dataset. And all algorithms use the inaccurate $\mathbf{H}$ for quickest detection. As previously mentioned, the $\mathbf{H}$ with the value given in (22) models the true linear relationship on our data. In this experiment, the inaccurate estimate of $\mathbf{H}$ is set to be

$$\mathbf{H} = \begin{bmatrix} 1.1 & -0.1 \\ 0.1 & 1.3 \\ -1.2 & 0.9 \\ 0.6 & 0.5 \end{bmatrix}.$$

In SDPCUSUM and BBCUSUM, we use the polyhedron uncertainty set to describe the uncertainty of $\mathbf{H}$. The polyhedron uncertainty is set to be

$$\mathbf{D}_i \boldsymbol{h}_i \leq \boldsymbol{d}_i, \ i = 1, 2,$$

where $\boldsymbol{h}_i$ represents the i-th column of $\mathbf{H}$ as previously defined. $\mathbf{D}_1$ and $\mathbf{D}_2$ are set to be

$$\mathbf{D}_1 = \mathbf{D}_2 = \begin{bmatrix} \mathbf{I} \\ -\mathbf{I} \end{bmatrix},$$

where $\mathbf{I}$ represents the 4×4 identity matrix. $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$ are set to be

$$\boldsymbol{d}_1 = [1.5, 0.5, -0.5, 1.5, -0.5, 0.5, 1.5, -0.5]^T,$$

$$\boldsymbol{d}_2 = [0.5, 1.5, 1.5, 1.5, 0.5, -0.5, -0.5, -0.5]^T.$$