A. Robust Quickest Change Detection

In many real world cases, the linear relationship of the system, which is depicted as $\mathbf{H} \in \mathbb{R}^{M\times N}$ a.k.a system matrix in this paper, can only be estimated. Therefore, the system matrix can be inaccurate, uncertain, or time-varying (Yang et al., 2014). In this section, by exploiting the robust optimization technique, we investigate the robust quickest change detection problem in the presence of uncertainties in $\mathbf{H}$. We next show that such a problem can also be cast as an MIQP problem. And therefore, both proposed detectors can be applied.

Before taking into account the uncertainties in $\mathbf{H}$, we first derive an equivalent form of our MIQP problem (16). In (16) the constraint $(\mathbf{I}-\mathbf{P}_{\mathbf{H}}^{\perp})(\boldsymbol{\mu}_+^{(t)}-\boldsymbol{\mu}_-^{(t)}) = \mathbf{0}$ is equivalent to $\mathbf{H}^T(\boldsymbol{\mu}_+^{(t)}-\boldsymbol{\mu}_-^{(t)}) = \mathbf{0}$, since they are essentially ensuring that $(\boldsymbol{\mu}_+^{(t)}-\boldsymbol{\mu}_-^{(t)})$ lie in the complementary space of $\mathcal{R}(\mathbf{H})$. And based on this constraint and the definition of $\tilde{\boldsymbol{x}}^{(t)}$ in (32), the item $\left(\boldsymbol{\mu}_+^{(t)}-\boldsymbol{\mu}_-^{(t)}\right)^T\tilde{\boldsymbol{x}}^{(t)}$ in the objective function is equivalent to $\left(\boldsymbol{\mu}_+^{(t)}-\boldsymbol{\mu}_-^{(t)}\right)^T\boldsymbol{x}^{(t)}$. Thus, an equivalent formulation of (16) can be written as follows:

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

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 & 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. $\boldsymbol{h}_i \in \mathcal{U}_i$. By simply relaxing the constraint $\mathbf{H}^T(\boldsymbol{\mu}_+^{(t)}-\boldsymbol{\mu}_-^{(t)}) = \mathbf{0}$, we get the new constraints as follows:

$$
-\varepsilon_i \le \boldsymbol{h}_i^T(\boldsymbol{\mu}_+^{(t)}-\boldsymbol{\mu}_-^{(t)}) \le \varepsilon_i,\ \forall\, \boldsymbol{h}_i \in \mathcal{U}_i,\ i=1,\cdots,N, \tag{A.2}
$$

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 (Duchi, 2018), but also some very 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. $\mathcal{U}_i = \{\boldsymbol{h}_i \mid \mathbf{D}_i\boldsymbol{h}_i \le \boldsymbol{d}_i\},\ i=1,\cdots,N$. According to (Ben-Tal et al., 2009), by exploiting the robust optimization techniques, the constraints A.2 under the polyhedron uncertainty set can be equivalently reformulated as

$$
\begin{aligned}
&\boldsymbol{p}_i^T\boldsymbol{d}_i \le \varepsilon_i, \quad \mathbf{D}_i^T\boldsymbol{p}_i = \boldsymbol{\mu}_+^{(t)}-\boldsymbol{\mu}_-^{(t)},\ \ \boldsymbol{p}_i \ge 0, \quad i=1,\cdots,N,\\
&\boldsymbol{n}_i^T\boldsymbol{d}_i \le \varepsilon_i, \quad -\mathbf{D}_i^T\boldsymbol{n}_i = \boldsymbol{\mu}_+^{(t)}-\boldsymbol{\mu}_-^{(t)},\ \ \boldsymbol{n}_i \ge 0, \quad i=1,\cdots,N.
\end{aligned}
\tag{A.3}
$$

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

$$
\begin{aligned}
&\min \frac{1}{2\sigma^2}\left\{\left\|(\boldsymbol{\mu}_+^{(t)}-\boldsymbol{\mu}_-^{(t)})\right\|_2^2-2(\boldsymbol{\mu}_+^{(t)}-\boldsymbol{\mu}_-^{(t)})^T\boldsymbol{x}^{(t)}\right\}\\
&s.t.\quad \begin{cases}
\boldsymbol{p}_i^T\boldsymbol{d}_i\le\varepsilon_i,\quad \mathbf{D}_i^T\boldsymbol{p}_i=\boldsymbol{\mu}_+^{(t)}-\boldsymbol{\mu}_-^{(t)},\ \ \boldsymbol{p}_i\ge 0,\quad i=1,\cdots,N\\
\boldsymbol{n}_i^T\boldsymbol{d}_i\le\varepsilon_i,\quad -\mathbf{D}_i^T\boldsymbol{n}_i=\boldsymbol{\mu}_+^{(t)}-\boldsymbol{\mu}_-^{(t)},\ \ \boldsymbol{n}_i\ge 0,\quad i=1,\cdots,N\\
\rho_L\boldsymbol{u}^{(t)}\le\boldsymbol{\mu}_+^{(t)}+\boldsymbol{\mu}_-^{(t)}\le\rho_U\boldsymbol{u}^{(t)}\\
\boldsymbol{\mu}_+^{(t)}\ge\mathbf{0}\\
\boldsymbol{\mu}_-^{(t)}\ge\mathbf{0}\\
\boldsymbol{\mu}_+^{(t)}+\rho_U\boldsymbol{b}^{(t)}\le\rho_U\mathbf{1}\\
\boldsymbol{\mu}_-^{(t)}-\rho_U\boldsymbol{b}^{(t)}\le\mathbf{0}
\end{cases}\\
&\text{var}:\boldsymbol{u}^{(t)}\in\{0,1\}^M,\boldsymbol{b}^{(t)}\in\{0,1\}^M,\boldsymbol{\mu}_+^{(t)},\boldsymbol{\mu}_-^{(t)},\{\boldsymbol{p}_i\},\{\boldsymbol{n}_i\}.
\end{aligned}
\tag{A.4}
$$

And the linear constraints introduced by the uncertainty of $\mathbf{H}$ do not change the nature of the problem, so that both our detectors can still be used. For convenient subsequent comparisons, we rename the SDPCUSUM and BBCUSUM considering the uncertainty of $\mathbf{H}$ as RSDPCUSUM and RBBCUSUM, repectively.

B. Numerical Results

We compare the performance of different detectors on **Data I**, and the results show that RSDPCUSUM and RBBCUSUM achieve better performance when the matrix $\mathbf{H}$ is inaccurate.

We assume the $\mathbf{H}$ with the value given in (24) models the true linear relationship and is denoted as $\mathbf{H}_r$. The inaccurate estimate of $\mathbf{H}$ is denoted as $\mathbf{H}_e$, and its value in our experiment is set to be

$$
\mathbf{H}=\begin{bmatrix} 1.4 & 0\\ 0 & 1.3\\ -1.2 & 1.2\\ 0.8 & 0.6\end{bmatrix}. \tag{1}
$$

And we use the polyhedron uncertainty set to describe the uncertainty of $\mathbf{H}$, i.e,

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

and

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

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

$$
\begin{aligned}
\boldsymbol{d}_1&=[1.5,0.5,-0.5,1.5,-0.5,0.5,1.5,-0.5]^T,\\
\boldsymbol{d}_2&=[0.5,1.5,1.5,1.5,0.5,-0.5,-0.5,-0.5]^T.
\end{aligned}
$$

The experimental results are shown in TABLE 3. It is seen from TABLE 3 that under different (ρ_L, ρ_H), the average detection delay of RSDPCUSUM and RBBCUSUM are smaller, which implies that RSDPCUSUM and RBBCUSUM achieve better performance that SDPCUSUM and RBBCUSUM in the presence inaccurate $\mathbf{H}$.

C. Performance Analysis

We have provided a guaranteed sub-optimality gap for SDPCUSUM. And it turns out that we can derive a similar sub-optimality gap for RSDPCUSUM.

Table 3. Performance Comparison on Simulated Data

	BBCUSUM AVERAGE DETECTION DELAY	RSDPCUSUM AVERAGE DETECTION DELAY	RBBCUSUM AVERAGE DETECTION DELAY
$\rho_L = 0.5, \rho_H = 3.0$ $\sigma^2 = 2.0$ FALSE ALARM PERIOD = 500.0	461.5	419.6	406.7
$\rho_L = 0.1, \rho_H = 1.5$ $\sigma^2 = 2.0$ FALSE ALARM PERIOD = 500.0	478.2	402.7	408.8

Theorem C.1. *Let $\boldsymbol{s}_M$ denote the optimal solution to problem A.1 and $\boldsymbol{s}_P$ denote the suboptimal solution generated by the RSDPCUSUM. Let the function $g : \mathbb{R}^M \to \mathbb{R}$ denote the objective function of problem A.1. The objective function value of the suboptimal solution $\boldsymbol{s}_P$ can be bounded as follows :*

$$g(\boldsymbol{s}_M) \leq g(\boldsymbol{s}_P) \leq g(\boldsymbol{s}_M) + \frac{2\sqrt{M}\rho_H}{\sigma^2}\|\boldsymbol{x}\|_2, \tag{2}$$

where σ is the variance of noise, M is the dimension of the observation vector, and ρ_H is the upper bound of the detectable part of the injected anomaly.

The proof is similar to that of **Theorem 1**.

References

Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. *Robust optimization*, volume 28. Princeton university press, 2009.

Bertsimas, D. and Sim, M. The price of robustness. *Operations research*, 52(1):35–53, 2004.

Duchi, J. Optimization with uncertain data. *Notes*, 2018, 2018.

Yang, K., Huang, J., Wu, Y., Wang, X., and Chiang, M. Distributed robust optimization (dro), part i: Framework and example. *Optimization and Engineering*, 15:35–67, 2014.