2019 IEEE 58th Conference on Decision and Control (CDC)
Palais des Congrès et des Expositions Nice Acropolis
Nice, France, December 11-13, 2019

Attack-resilient Estimation for Linear Discrete-time Stochastic Systems with Input and State Constraints

Wenbin Wan[†], Hunmin Kim[†], Naira Hovakimyan[†], and Petros G. Voulgaris[‡]

Abstract— **In this paper, an attack-resilient estimation algorithm is developed for linear discrete-time stochastic systems with inequality constraints on the actuator attacks and states. The proposed algorithm consists of optimal estimation and information aggregation. The optimal estimation provides minimum-variance unbiased (MVU) estimates, and then they are projected onto the constrained space in the information aggregation step. It is shown that the estimation errors and their covariances from the proposed algorithm are less than those from the unconstrained algorithm. Moreover, we proved that the state estimation errors of the proposed estimation algorithm are practically exponentially stable. A simulation on mobile robots demonstrates the effectiveness of the proposed algorithm compared to an existing algorithm.**

I. Introduction

Cyber-Physical Systems (CPS) have been of paramount importance in power systems, critical infrastructures, transportation networks and industrial control systems for many decades [1]. Recent cases of CPS attacks have clearly illustrated the vulnerability of CPS and raised awareness of the security challenges in these systems. These include attacks on large-scale systems, such as the StuxNet virus attack on an industrial supervisory control and data acquisition (SCADA) system [2], German steel mill cyber attack [3], and attacks on modern vehicles [4], [5].

Literature review. Traditionally, cyber-attack detection has been studied by monitoring the cyber-space misbehavior [6]. With the emergence of CPS, it becomes vitally important to monitor the physical misbehavior as well, because the attacks on CPS always have an impact on physical system. Model-based detection has been intensively studied in recent years. Attack detection has been formulated as an ℓ_0/ℓ_∞ optimization problem, which is non-deterministic polynomial-time hard (NP-hard) in [7], [8], [9]. A convex relaxation has been studied in [7], [9]. On top of this, the worst case estimation error has been analysed in [9]. A residual-based detector has been designed for power systems against false data injection attacks, and the impact of attacks has been analyzed in [10]. Linear algebraic conditions, as well as graph-theoretic conditions for detectability and identifiability have been provided in [11]. A switching mode resilient detection and estimation framework for GPS spoofing attacks has been studied in [12]. A multi-rate controller to detect zero-dynamic attacks has been designed in [13]. While most of the detection techniques were passive, some papers have studied active detection [14], [15], where the control input is watermarked with a pre-designed scheme that sacrifices optimality. The attack detection problem has been formulated as a simultaneous estimation problem of the state and the unknown input in [16]. The approach has been extended to nonlinear systems in [17], constrained systems in [18], and stochastic random set methods in [19]. The aforementioned detection algorithms rely on stochastic thresholds. For accurate detection, a smaller covariance is desired.

To reduce the covariance, the current paper focuses on information aggregation. In particular, we consider inequality state constraints and input constraints. There is a rich literature on Kalman filter with constraints [20], [21], [22]. We refer to [23] for more details for constrained filtering. Unknown input estimation algorithm with input constraints is introduced in [18]. This paper considers both inequality state and input constraints for unknown input estimation.

Contribution. We design an attack-resilient estimation algorithm given inequality constraints on the states and the attacks. The proposed algorithm consists of actuator attack estimation and state estimation. For each step, we design an optimal linear estimator without considering the constraints and then project the estimates onto the constrained space. We prove that the projection reduces the estimation error, as well as the error covariance. The practical exponential stability of the estimation error is proved formally. A numerical simulation on mobile robots shows the performance of the proposed attack-resilient estimation algorithm. The proofs of the lemmas and theorems are omitted due to the page limit. The complete version of the current paper can be found in [24].

II. Preliminaries

This section discusses some preliminary knowledge including notations, motivation, and problem statement.

A. Notations

The following notations are adopted: We use the subscript k of x_k to denote the time index; $\mathbb{R}^n$ denotes the n-dimensional Euclidean space; $\mathbb{R}^{n\times m}$ denotes the set of all $n\times m$ real matrices; $A^\top$ A^{-1}, $A^\dagger$, $diag(A)$, $tr(A)$ and $rk(A)$ denote the transpose, inverse, Moore-Penrose pseudoinverse, diagonal, trace and rank of matrix A, respectively; I denotes the identity matrix with an appropriate dimension; $\|\cdot\|$

This work has been supported by the National Science Foundation (ECCS-1739732 and CMMI-1663460).

[†]Wenbin Wan, Hunmin Kim, and Naira Hovakimyan are with the Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. {wenbinw2, hunmin, nhovakim}@illinois.edu

[‡]Petros G. Voulgaris is with the Department of Aerospace Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. {voulgari}@illinois.edu

978-1-7281-1398-2/19/$31.00 ©2019 IEEE

denotes the standard Euclidean norm for vector or an induced matrix norm; $\mathbb{E}[\cdot]$ denotes the expectation operator. For a symmetric matrix S, $S > 0$ and $S \geq 0$ indicates that S is positive definite and positive semi-definite, respectively. For a vector a, $(a)(i) = a(i)$ denotes the i^{th} element in the vector a. Finally a, $\hat{a}$, $\tilde{a} \triangleq a - \hat{a}$ denote the true value, estimate and estimation error of a.

B. Motivation

1) χ^2 test for detection: In attack detection for stochastic systems, the χ^2 test is widely used [15], [25].

Given a fixed attack input $v \neq 0$ and attack input estimate $\hat{v} \neq 0$, a smaller covariance induces a larger normalized test value $\hat{v}^\top \Sigma_v^{-1} \hat{v}$, which decreases false negative rates. To reduce the covariance, the minimum variance estimation method is being considered intensively [26], [27], [28]. The current paper pursues an optimal filter design technique.

2) Constraints: It has been shown that constraints can be used to further reduce the covariance in optimal filtering; i.e., state constraints in Kalman filter (KF) [22], [23], and input constraints in input and state estimation (ISE) [18]. We consider linear filtering with both input and state constraints to reduce false negative rates in attack detection and to achieve accurate state estimation. The constraints are induced by unmodeled dynamics and operational processes. Some of these examples include vision-aided inertial navigation [29], target tracking [30] and power systems [18], [31].

C. Problem Statement

Consider the linear time-varying discrete-time stochastic system [1]:

$$\begin{aligned} x_{k+1} &= A_k x_k + B_k u_k + G_k d_k + w_k \\ y_k &= C_k x_k + v_k, \end{aligned} \tag{1}$$

where $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^m$, $d_k \in \mathbb{R}^p$ and $y_k \in \mathbb{R}^l$ are the state, the known input, the unknown actuator attack and sensor measurement, respectively. Noises w_k and v_k are assumed to be independent identically distributed (i.i.d.) Gaussian random variables with zero means and covariances $Q_k \triangleq \mathbb{E}[w_k w_k^\top] \geq 0$ and $R_k \triangleq \mathbb{E}[v_k v_k^\top] > 0$ respectively. Moreover, v_k is also uncorrelated with the initial state x_0 and process noise w_k. We assume that $rk(C_k G_{k-1}) = p$ as in [32], [33].

In the cyber-space, digital attack signals could be unconstrained, but their impact on the physical world is restricted by physical and operational constraints. Any physical constraints and ability limitations on states and actuator attacks are presented by known inequality constraints:

$$\mathscr{A}_k d_k \leq b_k, \ \mathscr{B}_k x_k \leq c_k. \tag{2}$$

We assume that the feasible sets of the constraints $\mathscr{A}_k d_k \leq b_k$ and $\mathscr{B}_k x_k \leq c_k$ are non-empty. The vectors b_k and c_k, matrices $\mathscr{A}_k, \mathscr{B}_k, A_k, B_k, C_k$ and G_k are known and bounded.

The estimator design problem, addressed in this paper, can be stated as: *Given a linear discrete-time stochastic system* (1) *with constraints on the input and state* (2)*, design an attack-resilient and stable filtering algorithm that simultaneously estimates the system state and the actuator attack.*

III. Algorithm Design

In this section, we design an attack-resilient estimation algorithm with inequality constraints. The algorithm design is motivated by unknown input estimation [32], [33], [34], and a projection method for inequality constraint [18], [23]. We design an estimation algorithm as in [32], [33], [34] without considering the constraint, then project the estimates using inequality constraints as in [18], [23].

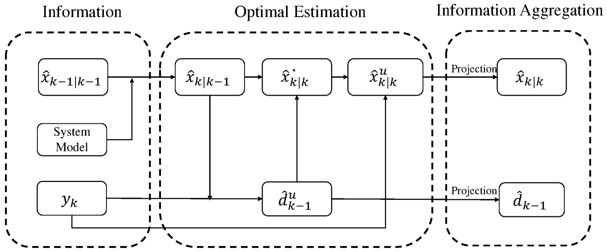

Fig. 1: The algorithm consists of two parts: optimal estimation and information aggregation.

A. Algorithm Statement

The proposed algorithm can be summarized as follows:

1) *Prediction*:

$$\hat{x}_{k|k-1} = A_{k-1}\hat{x}_{k-1|k-1} + B_{k-1}u_{k-1} \tag{3}$$

2) *Actuator attack estimation*:

$$\hat{d}^u_{k-1} = M_k(y_k - C_k \hat{x}_{k|k-1}) \tag{4}$$

$$\begin{aligned} \hat{d}_{k-1} = \underset{d}{\operatorname{argmin}} & (d - \hat{d}^u_{k-1})^\top (P^{d,u}_{k-1})^{-1} (d - \hat{d}^u_{k-1}) \\ & \text{subject to } \mathscr{A}_{k-1} d \leq b_{k-1} \end{aligned} \tag{5}$$

3) *Time update*:

$$\hat{x}^\star_{k|k} = \hat{x}_{k|k-1} + G_{k-1}\hat{d}^u_{k-1} \tag{6}$$

4) *Measurement update*:

$$\hat{x}^u_{k|k} = \hat{x}^\star_{k|k} + L_k(y_k - C_k \hat{x}^\star_{k|k}) \tag{7}$$

$$\begin{aligned} \hat{x}_{k|k} = \underset{x}{\operatorname{argmin}} & (x - \hat{x}^u_{k|k})^\top (P^{x,u}_k)^{-1} (x - \hat{x}^u_{k|k}) \\ & \text{subject to } \mathscr{B}_k x \leq c_k, \end{aligned} \tag{8}$$

Given the previous state estimate $\hat{x}_{k-1|k-1}$, the defender can predict the current state $\hat{x}_{k|k-1}$ under the assumption that the unknown actuator attack is absent (i.e., $d_{k-1} = 0$) in (3). The estimation of the unconstrained actuator attack $\hat{d}^u_{k-1}$ can be obtained by observing the difference between the predicted output $C_k \hat{x}_{k|k-1}$ and the measured output y_k in (4), and M_k is the filter gain that is chosen to minimize the input error covariances $P^{d,u}_k$. Then, we apply the constraints on the unconstrained actuator attack estimate in (5) and obtain the constrained actuator attack estimation $\hat{d}_{k-1}$. The state prediction $\hat{x}_{k|k-1}$ can be updated incorporating the actuator

[1] The current paper considers a general formulation for the attack input matrix. If d_k is injected into the input, then $G_k = B_k$. If d_k is directly injected into the system, then $G_k = I$.

attack estimate $\hat{d}^u_k$ in (6). In (7), the output y_k is used to correct the current state estimate as in KF, where L_k is the filter gain that is chosen to minimize the state error covariance $P^{x,u}_k$. The state constraints are applied in (8) to obtain the constrained state estimation $\hat{x}_{k|k}$. The algorithm is summarized in Fig. 1 and presented in Algorithm 1.

Algorithm 1 Attack-resilient Estimation with State and Input Constraint: $\mathscr{A}_k d_k \leq b_k$ and $\mathscr{B}_k x_k \leq c_k$

```
***Input:*** $\hat{x}_{k-1|k-1}$; $P^x_{k-1}$;
***Output:*** $\hat{d}_{k-1}$; $P^d_{k-1}$; $\hat{x}_{k|k}$; $P^x_k$.
    ▷ Prediction
    $\hat{x}_{k|k-1} = A_{k-1}\hat{x}_{k-1|k-1} + B_{k-1}u_{k-1}$;
    $P^x_{k|k-1} = A_{k-1}P^x_{k-1}A^\top_{k-1} + Q_{k-1}$;
    ▷ Actuator attack estimation
    $\tilde{R}_k = C_k P^x_{k|k-1} C^\top_k + R_k$;
    $M_k = (G^\top_{k-1}C^\top_k \tilde{R}^{-1}_k C_k G_{k-1})^{-1} G^\top_{k-1} C^\top_k \tilde{R}^{-1}_k$;
    $\hat{d}^u_{k-1} = M_k(y_k - C_k\hat{x}_{k|k-1})$;
    $P^{d,u}_{k-1} = (G^\top_{k-1}C^\top_k \tilde{R}^{-1}_k C_k G_{k-1})^{-1}$;
    $P^{xd}_{k-1} = -P^x_{k-1}A^\top_{k-1}C^\top_k M^\top_k$
    $\hat{d}_{k-1} = \underset{d}{\operatorname{argmin}}(d - \hat{d}^u_{k-1})^\top (P^{d,u}_{k-1})^{-1}(d - \hat{d}^u_{k-1})$
            subject to $\mathscr{A}_{k-1}d \leq b_{k-1}$;
    $\bar{\mathscr{A}}_{k-1}$ and $\bar{b}_{k-1}$ corresponding to active set;
    $\gamma^d_{k-1} = P^{d,u}_{k-1}\bar{\mathscr{A}}^\top_{k-1}(\bar{\mathscr{A}}_{k-1}P^{d,u}_{k-1}\bar{\mathscr{A}}^\top_{k-1})^{-1}$;
    $P^d_{k-1} = (I - \gamma^d_{k-1}\bar{\mathscr{A}}_{k-1})P^{d,u}_{k-1}(I - \gamma^d_{k-1}\bar{\mathscr{A}}_{k-1})^\top$;
    ▷ Time update
    $\hat{x}^\star_{k|k} = \hat{x}_{k|k-1} + G_{k-1}\hat{d}^u_{k-1}$;
    $P^{\star x}_k = A_{k-1}P^x_{k-1}A^\top_{k-1} + A_{k-1}P^{xd}_{k-1}G^\top_{k-1}$
            $+ G_{k-1}(P^{xd}_{k-1})^\top A^\top_{k-1} + G_{k-1}P^d_{k-1}G^\top_{k-1}$
            $- G_{k-1}M_kC_kQ_{k-1} - Q_{k-1}C^\top_k M^\top_k G^\top_{k-1} + Q_{k-1}$;
    $\tilde{R}^\star_k = C_k P^{\star x}_k C^\top_k + R_k - C_kG_{k-1}M_kR_k - R_kM^\top_k G^\top_{k-1}C^\top_k$;
    ▷ Measurement update
    $L_k = (P^{\star x}_k C^\top_k - G_{k-1}M_kR_k)\tilde{R}^{\star\dagger}_k$;
    $\hat{x}^u_{k|k} = \hat{x}^\star_{k|k} + L_k(y_k - C_k\hat{x}^\star_{k|k})$;
    $P^{x,u}_k = (I - L_kC_k)G_{k-1}M_kR_kL^\top_k + L_kR_kM^\top_k G^\top_{k-1}(I - L_kC_k)^\top$
            $+ (I - L_kC_k)P^{\star x}_k(I - L_kC_k)^\top + L_kR_kL^\top_k$;
    $\hat{x}_{k|k} = \underset{x}{\operatorname{argmin}}(x - \hat{x}^u_{k|k})^\top (P^{x,u}_k)^{-1}(x - \hat{x}^u_{k|k})$
            subject to $\mathscr{B}_k x \leq c_k$;
    $\bar{\mathscr{B}}_k$ and $\bar{c}_k$ corresponding to active set;
    $\gamma^x_k = P^{x,u}_k \bar{\mathscr{B}}^\top_k(\bar{\mathscr{B}}_k P^{x,u}_k \bar{\mathscr{B}}^\top_k)^{-1}$;
    $P^x_k = (I - \gamma^x_k\bar{\mathscr{B}}_k)P^{x,u}_k(I - \gamma^x_k\bar{\mathscr{B}}_k)^\top$;
```

B. Algorithm Derivation

1) Prediction: Given the previous state estimate $\hat{x}_{k-1|k-1}$, the current state can be predicted by (3). Its error covariance matrix is

$$P^x_{k|k-1} \triangleq \mathbb{E}[\tilde{x}_{k|k-1}\tilde{x}^\top_{k|k-1}] = A_{k-1}P^x_{k-1}A^\top_{k-1} + Q_{k-1},$$

where $P^x_k \triangleq \mathbb{E}[\tilde{x}_{k|k}\tilde{x}^\top_{k|k}]$ is the state estimation error covariance.

2) Actuator attack estimation: The actuator attack estimator in (4) utilizes the difference between the measured output y_k and the predicted output $C_k\hat{x}_{k|k-1}$. Substituting (1) and (3) into (4), we have

$$\hat{d}^u_{k-1} = M_k(C_kA_{k-1}\tilde{x}_{k-1|k-1} + C_kG_{k-1}d_{k-1} + C_kw_{k-1} + v_k),$$

which is a linear function of the actuator attack d_k. Applying the method of least squares from [35], which gives linear minimum-variance unbiased estimates, we can get the optimal gain in actuator attack estimation:

$$M_k = (G^\top_{k-1}C^\top_k\tilde{R}^{-1}_kC_kG_{k-1})^{-1}G^\top_{k-1}C^\top_k\tilde{R}^{-1}_k,$$

where $\tilde{R}_k \triangleq C_kP^x_{k|k-1}C_k + R_k$. It error covariance matrix is

$$P^d_{k-1} = M_k\tilde{R}_kM^\top_k = (G^\top_{k-1}C^\top_k\tilde{R}^{-1}_kC_kG_{k-1})^{-1}.$$

When apply the constraint in (2), the problem is formulated as the constrained convex optimization problem:

$$\begin{aligned}\hat{d}_{k-1} = \underset{d}{\operatorname{argmin}}&(d - \hat{d}^u_{k-1})^\top W^d_{k-1}(d - \hat{d}^u_{k-1})\\ \text{subject to }&\mathscr{A}_{k-1}d \leq b_{k-1},\end{aligned} \tag{9}$$

where W^d_{k-1} can be any positive definite symmetric weighting matrix. In the current paper, we choose $W^d_{k-1} = (P^{d,u}_{k-1})^{-1}$ which results in the smallest error covariance as shown in [20]. From Karush-Kuhn-Tucker (KKT) conditions of optimality, we can find the corresponding active constraints. We denote by $\bar{\mathscr{A}}_k$ and $\bar{b}_k$ the rows of $\mathscr{A}_k$ and the elements of b_k corresponding to the active constraints. Then (9) becomes

$$\begin{aligned}\hat{d}_{k-1} = \underset{d}{\operatorname{argmin}}&(d - \hat{d}^u_{k-1})^\top W^d_{k-1}(d - \hat{d}^u_{k-1})\\ \text{subject to }&\bar{\mathscr{A}}_{k-1}d = \bar{b}_{k-1}.\end{aligned}$$

The solution of the above program can be found by

$$\hat{d}_{k-1} = \hat{d}^u_{k-1} - \gamma^d_{k-1}(\bar{\mathscr{A}}_{k-1}\hat{d}^u_{k-1} - \bar{b}_{k-1}),$$

where $\gamma^d_{k-1} \triangleq (W^d_{k-1})^{-1}\bar{\mathscr{A}}^\top_{k-1}(\bar{\mathscr{A}}_{k-1}(W^d_{k-1})^{-1}\bar{\mathscr{A}}^\top_{k-1})^{-1}$. Its estimation error is

$$\tilde{d}_{k-1} = (I - \gamma^d_{k-1}\bar{\mathscr{A}}_{k-1})\tilde{d}^u_{k-1} + \gamma^d_{k-1}(\bar{\mathscr{A}}_{k-1}d_{k-1} - \bar{b}_{k-1}). \tag{10}$$

The error covariance matrix can be found by

$$P^d_{k-1} \triangleq \mathbb{E}[\tilde{d}_{k-1}\tilde{d}^\top_{k-1}] = (I - \gamma^d_{k-1}\bar{\mathscr{A}}_{k-1})P^{d,u}_{k-1}(I - \gamma^d_{k-1}\bar{\mathscr{A}}_{k-1})^\top,$$

under the assumption that $\gamma^d_{k-1}(\bar{\mathscr{A}}_{k-1}d_{k-1} - \bar{b}_{k-1}) = 0$ in (10).

The cross error covariance matrix of the state estimate and the actuator attack estimate is

$$P^{xd}_{k-1} = -P^x_{k-1}A^\top_{k-1}C^\top_kM^\top_k.$$

3) Time update: Given the actuator attack estimate $\hat{d}^u_{k-1}$, the state prediction $\hat{x}_{k|k-1}$ can be updated as in (6). We can derive the error covariance matrix of $\hat{x}^\star_{k|k}$ as

$$\begin{aligned}P^{\star x}_k \triangleq \mathbb{E}[(\tilde{x}^\star_{k|k})(\tilde{x}^\star_{k|k})^\top] &= A_{k-1}P^x_{k-1}A^\top_{k-1} + A_{k-1}P^{xd}_{k-1}G^\top_{k-1}\\ &+ G_{k-1}P^{dx}_{k-1}A^\top_{k-1} + G_{k-1}P^d_{k-1}\hat{G}^\top_{k-1} + Q_{k-1}\\ &- G_{k-1}M_kC_kQ_{k-1} - Q_{k-1}C^\top_kM^\top_kG^\top_{k-1},\end{aligned}$$

where $P^{dx}_{k-1} = (P^{xd}_{k-1})^\top$.

4) Measurement update: In this step, the measurement y_k is used to update the propagated estimate $\hat{x}^\star_{k|k}$ as shown in (7). The covariance matrix of the state estimation error is

$$\begin{aligned}P^{x,u}_k \triangleq \mathbb{E}[(\tilde{x}^u_{k|k})(\tilde{x}^u_{k|k})^\top] &= (I - L_kC_k)G_{k-1}M_kR_kL^\top_k + L_kR_kL^\top_k\\ &+ L_kR_kM^\top_kG^\top_{k-1}(I - L_kC_k)^\top + (I - L_kC_k)P^{\star x}_k(I - L_kC_k)^\top.\end{aligned}$$

The gain matrix L_k is chosen by minimizing the trace norm of $P^{x,u}_k$: $\min_{L_k} tr(P^{x,u}_k)$. The solution of the program is given by

$$L_k = (P_k^{\star x} C_k^\top - G_{k-1} M_k R_k) \tilde{R}_k^{\star\dagger},$$

where $\tilde{R}_k^\star \triangleq C_k P_k^{\star x} C_k^\top + R_k - C_k G_{k-1} M_k R_k - R_k M_k^\top G_{k-1}^\top C_k^\top$.

Now we apply the constraint in (2) to the state estimate $\hat{x}_{k|k}^u$. We formalize the state estimation with the constraints as the constrained convex optimization problem:

$$\begin{aligned} \hat{x}_{k|k} = \underset{x}{\operatorname{argmin}} (x - \hat{x}_{k|k}^u)^\top W_k^x (x - \hat{x}_{k|k}^u) \\ \text{subject to } \mathscr{B}_k x \leq c_k, \end{aligned} \tag{11}$$

where $W_k^x = (P_k^{x,u})^{-1}$ for the smallest error covariance.

We denote by $\bar{\mathscr{B}}_k$ and $\bar{c}_k$ the rows of $\mathscr{B}_k$ and the elements of c_k corresponding to the active constraints of (11). Using the active constraints, we reformulate the problem (11) as

$$\begin{aligned} \hat{x}_{k|k} = \underset{x}{\operatorname{argmin}} (x - \hat{x}_{k|k}^u)^\top W_k^x (x - \hat{x}_{k|k}^u) \\ \text{subject to } \bar{\mathscr{B}}_k x = \bar{c}_k. \end{aligned}$$

The solution of the above problem is given by

$$\hat{x}_{k|k} = \hat{x}_{k|k}^u - \gamma_k^x (\bar{\mathscr{B}}_k \hat{x}_{k|k}^u - \bar{c}_k),$$

where $\gamma_k^x \triangleq (W_k^x)^{-1} \bar{\mathscr{B}}_k^\top (\bar{\mathscr{B}}_k (W_k^x)^{-1} \bar{\mathscr{B}}_k^\top)^{-1}$. Under the assumption that $\gamma_k^x (\bar{\mathscr{B}}_k x_k - \bar{c}_k) = 0$ holds, the state estimation error covariance matrix can be expressed as $P_k^x = \bar{\Gamma}_k P_k^{x,u} \bar{\Gamma}_k^\top$, where $\bar{\Gamma}_k \triangleq I - \gamma_k^x \bar{\mathscr{B}}_k$.

IV. Analysis

In Section IV-A, we show that the projection induced by inequality constraints improves attack-resilient estimation and detection by decreasing the state estimation error and false negative rates. However, the projection induces a biased estimate as well [18]. In this context, we will seek to prove practical exponential stability, as shown in Section IV-B. All the proofs of the lemmas and theorems can be found in [24].

A. Performance Improvement through Constraints

The projection reduces the estimation errors and the covariance, as formulated in Theorem 4.1.

Theorem 4.1: We have $\|\tilde{x}_{k|k}\| \leq \|\tilde{x}_{k|k}^u\|$ and $\|\tilde{d}_k\| \leq \|\tilde{d}_k^u\|$; $P_k \leq P_k^u$, and $P_k^d \leq P_k^{d,u}$. Strict inequality holds if $rk(\bar{\mathscr{B}}_k) \neq 0$, and $rk(\bar{\mathscr{A}}_k) \neq 0$, respectively.

The properties in Theorem 4.1 are desired for accurate estimation as well as attack detection. In particular, if the size of attack is smaller than the statistical threshold, the χ^2 detector cannot distinguish the attack from the noise. Given $d_k \neq 0$, the covariance reduction implies the threshold reduction:

$$d_k^\top (P_k^{d,u})^{-1} d_k \leq d_k^\top (P_k^d)^{-1} d_k,$$

where the test value $d_k^\top (P_k^d)^{-1} d_k$ may reject the null hypothesis, while $d_k^\top (P_k^{d,u})^{-1} d_k$ cannot. Moreover, the estimation error reduction implies an accurate test value:

$$\begin{aligned} &\|d_k^\top (P_k)^{-1} d_k - (\hat{d}_k)^\top (P_k)^{-1} \hat{d}_k\| \\ &\leq \|d_k^\top (P_k)^{-1} d_k - (\hat{d}_k^u)^\top (P_k)^{-1} \hat{d}_k^u\|, \end{aligned}$$

which further reduces false negative rates.

B. Stability Analysis

Although the projection reduces the estimation errors and the covariance as shown in Theorem 4.1, it trades the unbiased estimation off according to Proposition 6 in [18]. This is because we can guarantee $\bar{\mathscr{B}}_k x_k \leq \bar{c}_k$ instead of $\bar{\mathscr{B}}_k x_k = \bar{c}_k$, but the unconstrained estimate $\hat{x}_{k|k}^u$ is projected onto $\bar{\mathscr{B}}_k x_k = \bar{c}_k$. In the absence of the projection, Algorithm 1 reduces to the algorithm in [33], which is unbiased.

It is essential to construct an update law $\tilde{x}_{k|k}$ from $\tilde{x}_{k-1|k-1}$ to analyze stability of the estimation error. However, the construction is not straight forward comparing to that in filtering with equality constraints [18], [20] or filtering without constraints [33], [36]. Especially, since $\bar{\mathscr{B}}_k x_k \leq \bar{c}_k$, it is difficult to find the exact relation between $\tilde{x}_{k|k}$ and $\tilde{x}_{k|k}^u$:

$$\tilde{x}_{k|k} = \tilde{x}_{k|k}^u - \gamma_k^x (\bar{\mathscr{B}}_k \hat{x}_{k|k}^u - \bar{c}_k) \neq (I - \gamma_k^x \bar{\mathscr{B}}_k) \tilde{x}_{k|k}^u.$$

To address this issue, we first decompose the estimation error $\tilde{x}_{k|k}$ into two orthogonal spaces

$$\tilde{x}_{k|k} = (I - \gamma_k^x \bar{\mathscr{B}}_k) \tilde{x}_{k|k} + \gamma_k^x \bar{\mathscr{B}}_k \tilde{x}_{k|k} \tag{12}$$

and then, we apply the following lemmas to each term.

Lemma 4.1: It holds that $(I - \gamma_k^x \bar{\mathscr{B}}_k) \tilde{x}_{k|k} = (I - \gamma_k^x \bar{\mathscr{B}}_k) \tilde{x}_{k|k}^u$.

Lemma 4.2: It holds that $\gamma_k^x \bar{\mathscr{B}}_k \tilde{x}_{k|k} = \alpha_k \gamma_k^x \bar{\mathscr{B}}_k \tilde{x}_{k|k}^u$, where $\alpha_k = diag(\alpha_k^1, \cdots, \alpha_k^n)$ and $\alpha_k^i \triangleq (\gamma_k^x \bar{\mathscr{B}}_k \tilde{x}_k)(i)((\gamma_k^x \bar{\mathscr{B}}_k \tilde{x}_k^u)(i))^\dagger \in [0, 1)$ for $i = 1, \cdots, n$.

According to Lemmas 4.1 and 4.2, the errors in the space $I - \gamma_k^x \bar{\mathscr{B}}_k$ remain identical after the projection, while the errors in the space $\gamma_k^x \bar{\mathscr{B}}_k$ reduce through the projection. By Lemmas 4.1 and 4.2, (12) becomes

$$\tilde{x}_{k|k} = \Gamma_k \tilde{x}_{k|k}^u,$$

where $\Gamma_k \triangleq (I - \gamma_k^x \bar{\mathscr{B}}_k) + \alpha_k \gamma_k^x \bar{\mathscr{B}}_k$. Note that α_k is an unknown matrix and thus cannot be used for the algorithm. We use it only for analytical purposes.

Now under the following assumptions, we present the stability of Algorithm 1.

Assumption 4.1: It holds that $rk(\mathscr{B}) < n$. There exist $\bar{a}$, $\bar{c}_y$, $\bar{g}$, $\bar{m}$, $\underline{q}$, $\underline{\beta}$, $\bar{\beta}$ > 0, such that the following holds for all $k \geq 0$: $\|A_k\| \leq \bar{a}, \|C_k\| \leq \bar{c}_y, \|G_k\| \leq \bar{g}, \|M_k\| \leq \bar{m}, Q_k \geq \underline{q}I$.

It is assumed that $rk(\mathscr{B}) < n$; i.e., the number of the state constrains are less than the number of state variables. The rest of Assumption 4.1 is widely used in literature on extended KF [37] and nonlinear ISE [17].

Theorem 4.2: Consider Assumption 4.1 and assume that there exist non-negative constants $\underline{p}$ and $\bar{p}$ such that $\underline{p}I \leq P_k^{x,u} \leq \bar{p}I$ holds for all k. Then the estimation errors $\tilde{x}_{k|k}$ and $\tilde{d}_k$ are practically exponentially stable in mean square; i.e., there exist constants $a_x, a_d, b_x, b_d, c_x, c_d$ such that, for all k,

$$\begin{aligned} \mathbb{E}[\|\tilde{x}_k\|^2] &\leq a_x e^{-b_x k} \mathbb{E}[\|\tilde{x}_0\|^2] + c_x \\ \mathbb{E}[\|\tilde{d}_k\|^2] &\leq a_d e^{-b_d k} \mathbb{E}[\|\tilde{d}_0\|^2] + c_d. \end{aligned}$$

Theorem 4.2 holds under the assumption of boundedness of $P_k^{x,u}$. One of the sufficient conditions is the uniform detectability of the transformed system as shown in Theorem 4.3.

Theorem 4.3: If the pair $(C_k, \tilde{A}_{k-1})$ is uniformly detectable, then there exist non-negative constants $\underline{p}$ and $\bar{p}$ such that for all k $\underline{p}I \leq P_k^{x,u} \leq \bar{p}I$, where $\tilde{A}_{k-1} \triangleq (I - G_k M_k (C_k G_{k-1} M_k)^{-1} \tilde{C}_k) \bar{A}_{k-1} \bar{\Gamma}_{k-1}$ and $\bar{A}_{k-1} = (I - G_{k-1} M_k C_k) A_{k-1}$.

V. Numerical Simulation

We simulate a scenario shown in Fig. 2, where a two-agent system that has state and input constraints gets attacked and moves to the attacker's desired place.

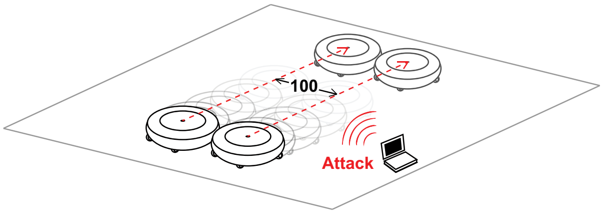

Fig. 2: Illustration of the simulation scenario: (i) red dash line denoted the path after attack; (ii) 100 denotes the minimum distance difference between two agents by physical state constraint.

A. Single Agent Model

We consider a double integrator dynamic model for each agent $i \in \{1, \cdots, n\}$, where n denotes the number of agents in the system. In this simulation, the subscript (i) is used to represent the agent i's vector/matrix; e.g., $x_k^{(i)}$ and $A_k^{(i)}$ denote the state and the system matrix of agent i. Its discrete time state vector $x_k^{(i)}$ that considers planar position and velocity at time step k, is given by

$$x_k^{(i)} = [r_{x,k}^{(i)}, r_{y,k}^{(i)}, v_{x,k}^{(i)}, v_{y,k}^{(i)}]^\top,$$

where $r_{x,k}^{(i)}$, $r_{y,k}^{(i)}$ denote x, y position coordinates and $v_{x,k}^{(i)}$, $v_{y,k}^{(i)}$ denote velocity coordinates. The actuator attack in this simulation is constrained by the acceleration limit, and the state is constrained due to the speed limit and required minimum distance between the two agents:

$$\begin{gathered} |d_k^{(i)}(n)| \leq 20, \quad |v_{x,k}^{(i)}| \leq 80, \quad |v_{y,k}^{(i)}| \leq 80; \\ |r_{x,k}^{(i)} - r_{x,k}^{(j)}| \geq 100 \quad \text{or} \quad |r_{y,k}^{(i)} - r_{y,k}^{(j)}| \geq 100, \end{gathered}$$

where (n) denotes the n^{th} element in the vector.

Each model is discretized into the following matrices with sampling time of 0.1 seconds:

$$A_k^{(i)} = \begin{bmatrix} 1 & 0 & 0.1 & 0 \\ 0 & 1 & 0 & 0.1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad B_k^{(i)} = G_k^{(i)} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0.1 & 0 \\ 0 & 0.1 \end{bmatrix},$$

and the output $y_k^{(i)}$ is the sensor measurement of positions and velocity; i.e., $C_k^{(i)} = I$. The covariance matrices of noises are chosen as $Q_k^{(i)} = 0.1I$, and $R_k^{(i)} = 0.01I$.

B. Multi-agent System Model

The multi-agent system of n agents, where $n \in \mathbb{N}$, can be written in the form of system (1), where A_k and C_k are diagonal matrices as follows: $diag(A_k) = (A^{(1)}, \cdots, A^{(i)})$, $diag(C_k) = (C_k^{(1)}, \cdots, C_k^{(i)})$; $B_k = G_k \triangleq [B^{(1)}, \cdots, B^{(i)}]^\top$. The state vector, input vector, actuator attack and sensor measurement are denoted by $x_k \triangleq [x_k^{(1)}, \cdots, x_k^{(i)}]^\top$, $u_k \triangleq [u_k^{(1)}, \cdots, u_k^{(i)}]^\top$, $d_k \triangleq [d_k^{(1)}, \cdots, d_k^{(i)}]^\top$ and $y_k \triangleq [y_k^{(1)}, \cdots, y_k^{(i)}]^\top$, respectively.

C. Attack Scenario

We consider the scenario that the attacker injects the identical actuator attack to the both agents so that they move horizontally to the right at same time. The unknown actuator attacks are

$$d_k(1) = d_k(3) = \begin{cases} 20 & \text{if } 100n \leq k < 40 + 100n, \\ 0 & \text{if } 40 + 100n \leq k < 60 + 100n, \\ -20 & \text{if } 60 + 100n \leq k < 100 + 100n, \end{cases}$$
$$d_k(2) = d_k(4) = 0,$$

where $n \in \{1, 2, \cdots, 9\}$.

D. Simulation Result

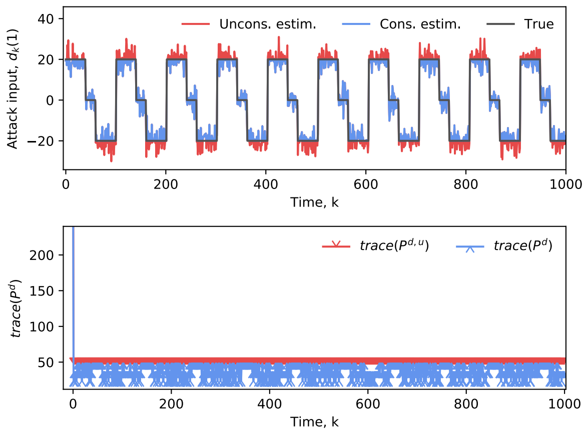

Fig. 3: Unconstrained and constrained estimation of the actuator attack $d_k(1)$. Trace of unconstrained estimate error covariance of the actuator attack $tr(P^{d,u})$ and constrained estimate error covariance of the actuator attack $tr(P^d)$.

Figures 3 and 4 show a comparison of the actuator attack and state estimation with and without the constraints. When the actuator attack estimate and the state estimate are projected to the constrained space, the constrained estimations have smaller estimation error and smaller error covariance as expected.

VI. Conclusion

This paper studies attack-resilient estimation algorithm for time-varying stochastic systems given inequality constraints on the states and actuator attacks. We formally prove that estimation errors and their covariances are less than those from unconstrained algorithms, which is a desired condition

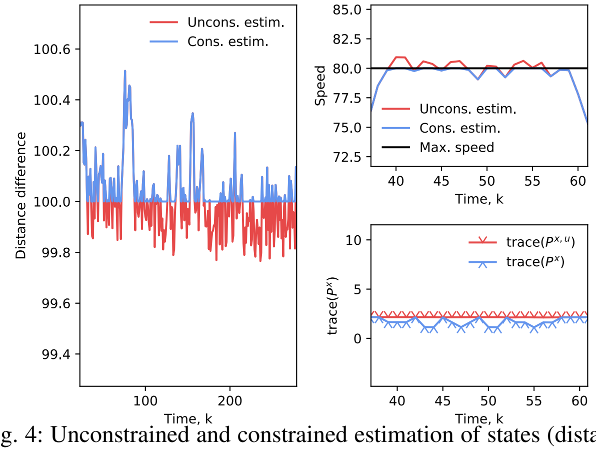

Fig. 4: Unconstrained and constrained estimation of states (distance difference of two agents and speed of one agent). Trace of unconstrained estimate error covariance of state $tr(P^{x,u})$ and constrained estimate error covariance of state $tr(P^x)$.

for attack detection in stochastic systems. We prove that the estimation errors are practically exponentially stable. A simulation is presented to reveal the attack-resilient property and efficiency of the proposed algorithm in attack detection.

References

[1] R. Rajkumar, I. Lee, L. Sha, and J. Stankovic, "Cyber-physical systems: the next computing revolution," in *Design Automation Conference*, pp. 731–736, 2010.

[2] R. Langner, "Stuxnet: Dissecting a cyber warfare weapon," *IEEE Security & Privacy*, vol. 9, no. 3, pp. 49–51, 2011.

[3] R. M. Lee, M. J. Assante, and T. Conway, "German steel mill cyber attack," *Industrial Control Systems*, vol. 30, p. 22, 2014.

[4] K. Koscher, A. Czeskis, F. Roesner, S. Patel, T. Kohno, S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, *et al.*, "Experimental security analysis of a modern automobile," in *2010 IEEE Symposium on Security and Privacy*, pp. 447–462, 2010.

[5] S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, *et al.*, "Comprehensive experimental analyses of automotive attack surfaces," in *USENIX Security Symposium*, vol. 4, pp. 447–462, 2011.

[6] J. Raiyn, "A survey of cyber attack detection strategies," *International Journal of Security and Its Applications*, vol. 8, no. 1, pp. 247–256, 2014.

[7] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, 2014.

[8] M. Pajic, J. Weimer, N. Bezzo, P. Tabuada, O. Sokolsky, I. Lee, and G. J. Pappas, "Robustness of attack-resilient state estimators," in *ACM/IEEE International Conference on Cyber-Physical Systems*, pp. 163–174, 2014.

[9] M. Pajic, I. Lee, and G. J. Pappas, "Attack-resilient state estimation for noisy dynamical systems," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 82–92, 2017.

[10] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security*, vol. 14, no. 1, pp. 21–32, 2011.

[11] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[12] H.-J. Yoon, W. Wan, H. Kim, N. Hovakimyan, L. Sha, and P. G. Voulgaris, "Towards resilient UAV : Escape time in GPS denied environment with sensor drift," *arXiv preprint arXiv:1906.05348*, 2019.

[13] H. Jafarnejadsani, H. Lee, N. Hovakimyan, and P. Voulgaris, "A multirate adaptive control for MIMO systems with application to cyber-physical security," in *2018 IEEE Conference on Decision and Control*, pp. 6620–6625, 2018.

[14] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *2009 47th Annual Allerton Conference on Communication, Control, and Computing*, pp. 911–918, 2009.

[15] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on SCADA systems," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 4, pp. 1396–1407, 2014.

[16] S. Z. Yong, M. Zhu, and E. Frazzoli, "Resilient state estimation against switching attacks on stochastic cyber-physical systems," in *2015 54th IEEE Conference on Decision and Control*, pp. 5162–5169, 2015.

[17] H. Kim, P. Guo, M. Zhu, and P. Liu, "Attack-resilient estimation of switched nonlinear cyber-physical systems," in *2017 American Control Conference*, pp. 4328–4333, 2017.

[18] S. Z. Yong, M. Zhu, and E. Frazzoli, "Simultaneous input and state estimation of linear discrete-time stochastic systems with input aggregate information," *2015 54th IEEE Conference on Decision and Control*, pp. 461–467, 2015.

[19] N. Forti, G. Battistelli, L. Chisci, and B. Sinopoli, "A Bayesian approach to joint attack detection and resilient state estimation," in *2016 IEEE 55th Conference on Decision and Control*, pp. 1192–1198, 2016.

[20] D. Simon and T. L. Chia, "Kalman filtering with state equality constraints," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 1, pp. 128–136, 2002.

[21] S. J. Julier and J. J. LaViola, "On Kalman filtering with nonlinear equality constraints," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2774–2784, 2007.

[22] S. Ko and R. R. Bitmead, "State estimation for linear systems with state equality constraints," *Automatica*, vol. 43, no. 8, pp. 1363–1368, 2007.

[23] D. Simon, "Kalman filtering with state constraints: A survey of linear and nonlinear algorithms," *IET Control Theory & Applications*, vol. 4, no. 8, pp. 1303–1318, 2010.

[24] W. Wan, H. Kim, N. Hovakimyan, and P. G. Voulgaris, "Attack-resilient estimation for linear discrete-time stochastic systems with input and state constraints," *arXiv preprint arXiv:1903.08282*, 2019.

[25] A. Teixeira, S. Amin, H. Sandberg, K. H. Johansson, and S. S. Sastry, "Cyber security analysis of state estimators in electric power systems," in *49th IEEE Conference on Decision and Control*, pp. 5991–5998, 2010.

[26] P. K. Kitanidis, "Unbiased minimum-variance linear state estimation," *Automatica*, vol. 23, no. 6, pp. 775–778, 1987.

[27] S. Gillijns and B. De Moor, "Unbiased minimum-variance input and state estimation for linear discrete-time systems," *Automatica*, vol. 43, no. 1, pp. 111–116, 2007.

[28] S. Gillijns and B. De Moor, "Unbiased minimum-variance input and state estimation for linear discrete-time systems with direct feedthrough," *Automatica*, vol. 43, no. 5, pp. 934–937, 2007.

[29] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, pp. 3565–3572, 2007.

[30] L. Wang, Y. Chiang, and F. Chang, "Filtering method for nonlinear systems with constraints," *IEEE Proceedings-Control Theory and Applications*, vol. 149, no. 6, pp. 525–531, 2002.

[31] A. J. Wood, B. F. Wollenberg, and G. B. Sheblé, *Power generation, operation, and control*. John Wiley & Sons, 2013.

[32] M. Darouach and M. Zasadzinski, "Unbiased minimum variance estimation for systems with unknown exogenous inputs," *Automatica*, vol. 33, no. 4, pp. 717–719, 1997.

[33] S. Z. Yong, M. H. Zhu, and E. Frazzoli, "A unified filter for simultaneous input and state estimation of linear discrete-time stochastic systems," *Automatica*, vol. 63, pp. 321–329, 2016.

[34] S. Gillijns and B. De Moor, "Unbiased minimum-variance input and state estimation for linear discrete-time systems," *Automatica*, vol. 43, no. 1, pp. 111–116, 2007.

[35] A. H. Sayed, *Fundamentals of adaptive filtering*. John Wiley & Sons, 2003.

[36] B. D. O. Anderson and J. B. Moore, "Detectability and stabilizability of time-varying discrete-time linear-systems," *SIAM Journal on Control and Optimization*, vol. 19, no. 1, pp. 20–32, 1981.

[37] S. Kluge, K. Reif, and M. Brokate, "Stochastic stability of the extended Kalman filter with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 55, no. 2, pp. 514–518, 2010.