Chapter 16
Control Systems Under Attack: The Securable and Unsecurable Subspaces of a Linear Stochastic System

Bharadwaj Satchidanandan and P. R. Kumar

Abstract The ideas of controllable and unobservable subspaces of a linear dynamical system introduced by Kalman play a central role in the theory of control systems. They determine, among other aspects, the existence of solutions to many control problems of interest. Analogous to the notions of controllable and unobservable subspaces are the notions of *"securable"* and *"unsecurable"* subspace of a linear dynamical system, which have operational significance in the context of secure control. We examine what guarantees can be provided with respect to securable subspace, especially in the case when there is process noise in the system.

16.1 Introduction

The ideas of controllable and unobservable subspaces of a linear dynamical system, introduced by Kalman in [1], play a central role in the theory of control systems. They provide, for example, necessary and sufficient conditions for the existence of a stabilizing control law for any linear dynamical system of interest. Analogous to the notions of controllable and unobservable subspaces, we examine, in this paper, the notions of *"securable"* and *"unsecurable"* subspaces of a linear dynamical system, which we show have operational significance in the context of secure control.

Consider a multiple-input, multiple-output, discrete-time linear dynamical system, an arbitrary subset of whose sensors and actuators may be "malicious." The

This material is based upon work partially supported by NSF under Contract Nos. ECCS-1646449, ECCS-1547075, CCF-1619085 and Science & Technology Center Grant CCF-0939370, the U.S. Army Research Office under Contract No. W911NF-15-1-0279, the Power Systems Engineering Research Center (PSERC), and NPRP grant NPRP 8-1531-2-651 from the Qatar National Research Fund, a member of Qatar Foundation.

B. Satchidanandan (✉) · P. R. Kumar
Texas A&M University, College Station, Texas, USA
e-mail: bharadwaj.s1990@tamu.edu

P. R. Kumar
e-mail: prk.tamu@gmail.com

R. Tempo et al. (eds.), *Emerging Applications of Control and Systems Theory*, Lecture Notes in Control and Information Sciences - Proceedings,
https://doi.org/10.1007/978-3-319-67068-3_16

malicious sensors may not truthfully report the measurements that they observe, and the malicious actuators may not apply their control inputs as per the specified control law. In such a setting, even if the system is controllable and observable, the desired control objective may not be achievable. The honest nodes in the system may believe the state trajectory to be a certain sequence $\{\mathbf{x}[0], \mathbf{x}[1], \ldots\}$ whereas the actual state trajectory of the system may be very different. It is against this backdrop that we define the notions of *securable* and *unsecurable* subspaces of a linear dynamical system. The unsecurable subspace is defined, roughly, as the set of states that the system could actually be in, or ever reach, without the honest sensors ever being able to detect, based on their measurements, that the system had visited that state, or that there was any malicious activity in the system. Theorems 16.1 and 16.2 in the paper characterize the securable and unsecurable subspaces of a linear system. These results are analogous to those reported in [5], and in [2–4] for continuous-time linear dynamical systems, which examine what sorts of attacks are possible on control systems while remaining undetected. We formalize the results as characterizations of securable and unsecurable subspaces. They may be regarded as the analogs of the controllable and unobservable subspaces reexamined in an era where there is intense interest in cybersecurity of control systems. We then turn to the case of systems with noise, i.e., linear stochastic dynamical systems. We show that the securable and unsecurable subspaces defined in the context of deterministic systems also have operational meaning in the context of stochastic systems.

One way to view these results is as negative or impossibility results which state that given a linear control system with certain malicious sensors and actuators, it is impossible for the honest sensors to distinguish certain state trajectories from others. Consequently, it may be impossible to guarantee that the system does not reach certain states that are considered "unsafe." An alternate viewpoint is to look at these results from a system designer's perspective. These results could be regarded as providing guidelines for designing secure control systems. For example, for a specified amount of resilience required of the control system, typically quantified by the number of Byzantine nodes that the system should tolerate, or for a specification that the system should not visit certain "unsafe" states, the results can be translated into conditions that the securable and unsecurable subspaces should satisfy in order to meet the security specifications. This can potentially constitute a principled approach to design systems that are secure by construction, as opposed to designing systems to maximize a performance metric, and only subsequently installing ad-hoc security measures as an afterthought.

As mentioned before, many of the results in this paper pertaining to deterministic linear dynamical systems are mathematically isomorphic to some of the results contained in [2–5]. In addition, we report preliminary results on the extension of the above results to the context of stochastic linear dynamical systems where only noisy measurements of states are available.

16.2 Problem Formulation

Consider a *pth* order discrete-time linear dynamical system with m inputs and n outputs described by

$$\begin{aligned} \bar{\mathbf{x}}[t+1] &= A\bar{\mathbf{x}}[t] + B\bar{\mathbf{u}}[t], \\ \bar{\mathbf{y}}[t+1] &= C\bar{\mathbf{x}}[t+1], \\ \bar{\mathbf{x}}[0] &= \mathbf{x}_0. \end{aligned} \tag{16.1}$$

where $\bar{\mathbf{x}}[t] \in \mathbb{R}^p$ denotes the state of the system at time t, $\bar{\mathbf{u}}[t] \in \mathbb{R}^m$ denotes the input applied to the system at time t, $\bar{\mathbf{y}}[t] \in \mathbb{R}^n$ denotes the output of the system at time t, and A, B, and C are real matrices of appropriate dimensions.

We denote by $\mathbf{z}[t]$ the values reported by the sensors at time t. If sensor i, $i \in \{1, 2, ..., n\}$, is honest, then $z_i[t] = \bar{y}_i[t]$ for all t. We assume that an arbitrary, known, possibly history-dependent control policy $g = \{g_1, g_2, ...\}$ is in place, and denote by $\bar{\mathbf{u}}^g[t]$ the control policy-specified input at time t, so that $\bar{\mathbf{u}}^g[t] = g_t(\mathbf{z}^t)$, where $\mathbf{z}^t := [\mathbf{z}^T[0] \;\; \mathbf{z}^T[1] \;\; \ldots \;\; \mathbf{z}^T[t]]^T$. If actuator i is honest, then $\bar{u}_i[t] = \bar{u}_i^g[t]$ for all t.

We assume the adversarial nodes in the system to be near-omniscient, in the sense that at time $t = 0$, they have perfect knowledge of the initial state $\mathbf{x}_0$ of the system. On the other hand, the honest nodes in the system, at any time t, have access only to the measurements $\mathbf{z}^t$ that are reported until that time. Clearly, this assumption represents a worst-case scenario from the point of view of the honest nodes in the system. Consequently, the results presented in this paper serve as fundamental bounds that apply regardless of the capabilities of the attacker, and in particular, even for systems where the adversary's knowledge may be more limited.

Note that if all the nodes in the system are honest, and if the pair (A, C) is observable, then the nodes can correctly estimate the initial state $\mathbf{x}_0$ of the system by time $p - 1$. Consequently, they can correctly estimate the state $\bar{\mathbf{x}}[t]$ of the system at any time t. However, when there are malicious sensors and/or actuators present in the system, this need not be the case. Specifically, the honest nodes in the system could be under the impression that the state of the system at some time t is $\widehat{\mathbf{x}}[t]$, while in reality, the system could be in state $\bar{\mathbf{x}}[t] \neq \widehat{\mathbf{x}}[t]$. This brings us to the central question that is addressed in this paper: *Suppose that there are malicious nodes present in the system and that they act in a fashion that keeps them undetected. Suppose also that the honest nodes believe the system's state evolution to be* $\{\widehat{\mathbf{x}}[0], \widehat{\mathbf{x}}[1], \widehat{\mathbf{x}}[2], \ldots\}$. *Under these conditions, what are the set of states that the system can actually be in, or ever reach? This set essentially contains the set of states that the malicious nodes can steer the system to.* For this reason, we term this set as the *"unsecurable"* subspace of the system (A, B, C) for state $\widehat{\mathbf{x}}[0]$. The orthogonal complement of this is called the *"securable"* subspace. The projection of the uncertain state on this subspace is actually what the honest sensors and actuators believe it is, whether the system is not under attack or is under a stealthy attack. It is the largest such subspace. A formal definition of securable and unsecurable subspaces is presented in the next section.

16.3 Securable and Unsecurable Subspaces of Linear Control Systems

In order to determine if malicious nodes are present in the system or not, each honest sensor i subjects the reported measurement sequence $\{\mathbf{z}\}$ to the following test. If and only if the test fails (at any time t) does the sensor declare that malicious nodes are present in the system.

The rest of the paper follows the notation specified in the appendix of the paper.
Test: At each time t, check if the reported sequence of measurements up to that time $\mathbf{z}^t$ satisfies the following condition: $\exists \widehat{\mathbf{x}}_0 \in \mathbb{R}^p$ such that,

$$\mathbf{z}^t - F[t-1]\bar{\mathbf{u}}^{g^{t-1}} = \Gamma[t]\widehat{\mathbf{x}}_0. \tag{16.2}$$

Proposition 16.1 *If all the nodes in the system are honest, the reported measurements* $\{\mathbf{z}\}$ *pass (16.2) at each time* t*. Conversely, if the reported measurements* $\{\mathbf{z}\}$ *pass (16.2) at each time* t*, then there exists an initial state* $\mathbf{x}[0]$ *such that* $\mathbf{z}[t]$ *is the output of the system at time* t *under control* $\{\bar{\mathbf{u}}^g\}$*, and so, there is no definitive reason for the honest sensor to declare that malicious nodes are present in the system.*

Proof Omitted. ■

In what follows, we assume that the measurements reported by the malicious sensors pass the above test, and examine the limits of what the malicious nodes can do under this constraint.

Since the reported measurements $\{\mathbf{z}\}$ pass (16.2), it follows in particular that $\exists \widehat{\mathbf{x}}_0 \in \mathbb{R}^p$ such that $\forall t$,

$$\mathbf{z}^{t-1} - F[t-2]\bar{\mathbf{u}}^{g^{t-2}} = \Gamma[t-1]\widehat{\mathbf{x}}_0, \tag{16.3}$$

$$\bar{\mathbf{y}}_H[t] - \sum_{i=0}^{t-1} C_H A^i B \bar{\mathbf{u}}^g[t-1-i] = C_H A^t \widehat{\mathbf{x}}_0. \tag{16.4}$$

The following proposition is a (partial) converse of the above statement.

Proposition 16.2 *Suppose that there exist* $\widehat{\mathbf{x}}_0$*,* $\mathbf{z}_M^{\tau-1}$*, and* $\bar{\mathbf{d}}^{\tau-1}$ *such that (16.3) and (16.4) hold for* $t = \tau$*. Then, there is a vector* $\mathbf{z}_M[\tau]$ *that satisfies Test (16.2) at time* τ*.*

Proof Consider $\mathbf{z}_M[\tau] = C_M A^\tau \widehat{\mathbf{x}}_0 + \sum_{i=0}^{\tau-1} C_M A^i B \bar{\mathbf{u}}^g[\tau-1-i]$. It is straightforward to verify that it satisfies (16.2). ■

The above proposition states that it is sufficient for the malicious nodes to consider strategies that only ensure "consistency" at the outputs of the honest sensors. The outputs to be reported by the malicious sensors can be fabricated accordingly.

The next proposition, along with Theorem 16.2, shows that one can consider a simpler system consisting of only malicious actuators, honest sensors, and a control

policy that is identically zero, and translate the conclusion obtained from the analysis of such a system to the more general system (16.1). In other words, one can dispense with the honest actuators and malicious sensors. There is no loss of generality in assuming that the control policy is identically equal to zero, and that the system has only honest sensors and malicious actuators.

Given the system described by (16.1), consisting of honest and malicious nodes as described before, consider the following reduction of the system where all sensors are honest, all actuators are malicious, and the control policy is identically equal to zero:

$$\begin{aligned} \mathbf{x}[t+1] &= A\mathbf{x}[t] + B_M\mathbf{d}[t], \\ \mathbf{y}_H[t+1] &= C_H\mathbf{x}[t+1], \\ \mathbf{x}[0] &= \mathbf{x}_0. \end{aligned} \tag{16.5}$$

where $\mathbf{y}_H[t]$ are the measurements observed by the (honest) sensors at time t, $\mathbf{d}[t]$ are the inputs applied by the (malicious) actuators at time t. We will refer to system (16.5) as the "reduced system" of system (16.1), or simply the "reduced system" when there is no ambiguity. Note that the reduced system has the same state space as its parent system (16.1), and is also initialized with the same state as its parent. It is only the inputs and the outputs of the systems that are different. As before, the malicious actuators are assumed to be near-omniscient so that they have perfect knowledge of the initial state $\mathbf{x}_0$. For the reduced system, Test (16.2) reduces to the following, and is performed by the (honest) sensors.

Test for the reduced system: Check if $\exists \widetilde{\mathbf{x}}_0 \in \mathbb{R}^p$ such that for all t,

$$\mathbf{y}_H^t = \Gamma_H[t]\widetilde{\mathbf{x}}_0. \tag{16.6}$$

Proposition 16.3 *Suppose that there exists a sequence* $\{\mathbf{d}\}$ *for the reduced system satisfying test (16.6). Then, if the malicious actuators in the parent system (16.1) inject* $\{\bar{\mathbf{d}}\} \equiv \{\mathbf{d}\}$*, there exist fabricated measurements* $\{\mathbf{z}_M\}$ *that can be reported by the malicious sensors in the parent system that pass Test (16.2) with* $\widehat{\mathbf{x}}_0 = \widetilde{\mathbf{x}}_0$.

Proof For the reduced system, we have

$$\mathbf{y}_H[t] = C_H A^t \mathbf{x}_0 + \sum_{i=0}^{t-1} C_H A^i B_M \mathbf{d}[t-1-i]. \tag{16.7}$$

Now, suppose for induction that there exist measurements $\mathbf{z}_M[0], \mathbf{z}_M[1], \ldots, \mathbf{z}_M[t-1]$ that the malicious sensors can report for system (16.1) when the malicious actuators inject $\bar{\mathbf{d}}[i] = \mathbf{d}[i]$, $i = 0, 2, ..., t-2$, such that the reported measurements pass test (16.2) up to time $t-1$ with $\widehat{\mathbf{x}}_0 = \widetilde{\mathbf{x}}_0$. The base case of $t = 1$ holds since the malicious sensors in the parent system can report $\mathbf{z}_M[0] = C_M\widetilde{\mathbf{x}}_0$. This amounts to assuming that (16.3) holds with $\widehat{\mathbf{x}}_\mathbf{0} = \widetilde{\mathbf{x}}_0$. Now, if the malicious actuators in the parent system inject, at time $t-1$, $\bar{\mathbf{d}}[t-1] = \mathbf{d}[t-1]$, then,

$$\bar{\mathbf{y}}_H[t] = C_H A^t \mathbf{x}_0 + \sum_{i=0}^{t-1} C_H A^i B \bar{\mathbf{u}}^g[t-1-i] + \sum_{i=0}^{t-1} C_H A^i B_M \mathbf{d}[t-1-i].$$

Substituting (16.7) in the above gives

$$\bar{\mathbf{y}}_H[t] = \mathbf{y}_H[t] + \sum_{i=0}^{t-1} C_H A^i \bar{\mathbf{u}}^g[t-1-i].$$

Since the output of the reduced system satisfies (16.6), we have $\mathbf{y}_H[t] = C_H A^t \widetilde{\mathbf{x}}_0$. Substituting this into the above equation gives $\bar{\mathbf{y}}_H[t] - \sum_{i=0}^{t-1} C_H A^i \bar{\mathbf{u}}^g[t-1-i] = C_H A^t \widetilde{\mathbf{x}}_0$, which satisfies (16.4) for $\widehat{\mathbf{x}}_\mathbf{0} = \widetilde{\mathbf{x}}_0$. The desired result follows from Proposition 16.2. ■

The following definition is of central importance.

Definition 16.1 Consider a system (A, B, C) of the form (16.1) with initial state $\mathbf{x}_0$. The *unsecurable subspace for state* $\mathbf{s}_0$ of the system is the maximal set of states $V(\mathbf{s}_0)$ such that for each $\mathbf{v} \in V(\mathbf{s}_0)$, there exist t, $\{\bar{\mathbf{d}}\}$, $\{\mathbf{z}_M\}$ such that $\bar{\mathbf{x}}[t] = \mathbf{v}$ and (16.2) holds for $\widehat{\mathbf{x}}_0 = \mathbf{s}_0$.

In particular, for the reduced system (A, B_M, C_H), the unsecurable subspace for state $\mathbf{s}_0$ is the maximal set of states $V_R(\mathbf{s}_0)$ such that for each $\mathbf{v} \in V_R(\mathbf{s}_0)$, there exist t, $\{\mathbf{d}\}$ such that $\mathbf{x}[t] = \mathbf{v}$ and (16.6) holds for $\widetilde{\mathbf{x}}_0 = \mathbf{s}_0$.

In other words, the unsecurable space for $\mathbf{s}_0$ is the set of states that the system can be in if the honest nodes are deceived into inferring the initial state as $\mathbf{s}_0$. If the unsecurable subspace is of dimension greater than zero, it (i) states that the malicious nodes cannot distort certain linear combinations of the state without being detected, and (ii) specifies those linear combinations that are "intact."

The following theorem characterizes the unsecurable subspace and suggests an algorithm to compute it.

Theorem 16.1 *Consider a reduced system* (A, B_M, C_H) *of the form (16.5). For such a system,*

(i) The unsecurable subspace $V_R(0)$ *for state* 0 *is the maximal set* $W \subseteq \mathbb{R}^p$ *such that* $\forall \mathbf{w} \in W$,

  a. $C_H \mathbf{w} = 0$, *and*
  b. $\exists \mathbf{d}$ *such that* $A\mathbf{w} + B_M \mathbf{d} \in W$.

(ii) The unsecurable subspace for state $\mathbf{s}_0$, $V_R(\mathbf{s}_0)$, *is*

$$V_R(\mathbf{s}_0) = \{\mathbf{s}_0 + \mathbf{w} : \mathbf{w} \in V_R(0)\}. \tag{16.8}$$

Proof **Lemma 16.1** *The set W is a subspace.*

Proof Omitted. ■

We now show that W is equal to $V_R(0)$. The crux of the argument is that W is an invariant subspace in the following sense.

Lemma 16.2 *If the system's state visits W at any time t, then the malicious actuators can synthesize control actions that keep the state in W at all subsequent times.*

Proof We show this via induction. Let $\mathbf{w} \in W$, and let $\mathbf{x}[t] = \mathbf{w}$, which also serves as the base case for induction. Assume for induction that $\mathbf{x}[\tau] \in W$, where $\tau \geq t$ is a fixed time. Then, $\mathbf{x}[\tau + 1] = A\mathbf{x}[\tau] + B_M\mathbf{d}[\tau]$. Since $\mathbf{x}[\tau] \in W$, it follows from the definition of W that there exists a control choice $\mathbf{d}$ for $\mathbf{d}[\tau]$ such that $A\mathbf{x}[\tau] + B_M\mathbf{d}[\tau] \in W$, implying that $\mathbf{x}[\tau + 1] \in W$. ■

Remark: Owing to the above Lemma, W is called the controlled invariant subspace in linear system theory [6].

Now, suppose that $\mathbf{x}[0] = \mathbf{w}$ and $\mathbf{w} \in W$. We then have from Lemma 16.2 that there exists a sequence $\{\mathbf{d}\}$ that the malicious actuators can apply as inputs such that $\mathbf{x}[t] \in W$ for all t. Since $W \subseteq \mathrm{N}(C_H)$ by definition, we have $\mathbf{y}_H[t] = C_H\mathbf{x}[t] = 0$ for all t. Consequently, (16.6) holds for $\widetilde{\mathbf{x}}_0 = 0$, and it follows from Definition 16.1 that $\mathbf{w} \in V_R(0)$. Hence, $W \subseteq V_R(0)$.

Now suppose that $\mathbf{v} \in V_R(0)$. We then have from Definition 16.1 that $\exists\{\mathbf{d}\}$ that the malicious actuators can apply as inputs to the system such that $\mathbf{x}[t] = \mathbf{v}$ for some t and (16.6) holds for $\widetilde{\mathbf{x}}_0 = 0$. This implies that $\mathbf{y}_H[t] = 0$ for all t. Since $0 = \mathbf{y}_H[t] = C_H\mathbf{x}[t] = C_H\mathbf{v}$, we have that $\mathbf{v} \in \mathrm{N}(C_H)$, satisfying the first condition to be an element of W. Since $\{\mathbf{d}[t], \mathbf{d}[t + 1], ...\}$ is a sequence that the malicious actuators can apply such that $\mathbf{x}[t'] \in \mathrm{N}(C_H)$ for all $t' \geq t$, $\mathbf{v}$ satisfies the second condition to be an element of W. Therefore, $\mathbf{v} \in W$, and $V_R(0) \subseteq W$. Combining the two results, we have $W = V_R(0)$.

(ii) Let $\mathbf{v} \in \{\mathbf{s}_0 + \mathbf{w} : \mathbf{w} \in V_R(0)\}$, and let $\mathbf{x}[0] = \mathbf{v}$. Then, $\mathbf{y}_H^t = \Gamma_H[t]\mathbf{v} + H_M[t - 1]\mathbf{d}^{t-1} = \Gamma_H[t]\mathbf{s}_0 + \Gamma_H[t]\mathbf{w} + H_M[t - 1]\mathbf{d}^{t-1}$ for all t. Since $\mathbf{w} \in V_R(0)$, it follows from the definition of $V_R(0)$ that there exists sequence $\{\mathbf{d}\}$ so that $\Gamma_H[t]\mathbf{w} + H_M[t - 1]\mathbf{d}^{t-1} = 0$ for all t. Therefore, if the actuators inject such a sequence $\{\mathbf{d}\}$, then, $\mathbf{y}_H^t$ reduces to $\mathbf{y}_H^t = \Gamma_H[t]\mathbf{s}_0$. Therefore, (16.6) holds with $\widetilde{\mathbf{x}}_0 = \mathbf{s}_0$, and so, $\{\mathbf{s}_0 + \mathbf{w} : \mathbf{w} \in V(0)\} \subseteq V_R(s_0)$.

Next, let $\mathbf{v} \in V_R(s_0)$. Then, we have from the definition of $V_R(\mathbf{s}_0)$ that $\exists\{\mathbf{d}'\}, \tau$ such that $\mathbf{x}[\tau] = \mathbf{v}$ and $\Gamma_H[t]\mathbf{s}_0 = \mathbf{y}_H^t$ for all t. This in turn implies that $\exists\{\mathbf{d}\}$ such that $\mathbf{x}_0 = \mathbf{v}$ and $\Gamma_H[t]\mathbf{s}_0 = \mathbf{y}_H^t$ for all t. Also, when $\mathbf{x}_0 = \mathbf{v}$, we have for all t, $\mathbf{y}_H^t = \Gamma_H[t]\mathbf{v} + H_M[t - 1]\mathbf{d}^{t-1}$. Combining the two, we have that there exists $\{\mathbf{d}\}$ such that $\Gamma_H[t]\mathbf{s}_0 = \Gamma_H[t]\mathbf{v} + H_M[t - 1]\mathbf{d}^{t-1}$ for all t. This means that $\mathbf{v}$ solves, for all t,

$$[\Gamma_H[t] \;\; H_M[t - 1]] \begin{bmatrix} \mathbf{v} \\ \mathbf{d}^{t-1} \end{bmatrix} = [\Gamma_H[t] \;\; H_M[t - 1]] \begin{bmatrix} \mathbf{s}_0 \\ 0 \end{bmatrix},$$

so that for all t,

$$\begin{bmatrix} \mathbf{v} \\ \mathbf{d}^{t-1} \end{bmatrix} = \begin{bmatrix} \mathbf{s}_0 \\ 0 \end{bmatrix} + \widetilde{\mathbf{w}},$$

where $\widetilde{\mathbf{w}} \in \mathrm{N}([\Gamma_H[t] \;\; H_M[t-1]])$. Denote by $\mathbf{w}$ the first p entries of $\widetilde{\mathbf{w}}$, and it follows from the definition of $V_R(0)$ that $\widetilde{\mathbf{w}} \in V_R(0)$. Hence, $\mathbf{v}$ must be of the form $\mathbf{s}_0 + \mathbf{w}$, $\mathbf{w} \in V_R(0)$, and hence, $V_R(s_0) \subseteq \{\mathbf{s}_0 + \mathbf{w} : \mathbf{w} \in V(0)\}$.

Combining the two results, we have $V_R(\mathbf{s}_0) = \{\mathbf{s}_0 + \mathbf{w} : \mathbf{w} \in V(0)\}$. ■

The following theorem translates the above conclusions obtained from the reduced system (16.5) to the original system (16.1) that is of interest.

Theorem 16.2 *The unsecurable subspace $V(\mathbf{s}_0)$ for the system (A, B, C) is the same as the unsecurable subspace $V_R(\mathbf{s}_0)$ for its reduction (A, B_M, C_H).*

Proof Let $\mathbf{v} \in V_R(\mathbf{s}_0)$. Then, it follows form the definition of $V_R(\mathbf{s}_0)$ that for the reduced system (16.5), there exists $\{\mathbf{d}\}$ that can be applied by the actuators so that (16.6) is satisfied for $\widetilde{\mathbf{x}}_0 = \mathbf{s}_0$ when $\mathbf{x}_0 = \mathbf{v}$. Therefore, by Proposition 16.3, $\mathbf{v} \in V(\mathbf{s}_0)$, and so, $V_R(\mathbf{s}_0) = V(\mathbf{s}_0)$.

Next, let $\mathbf{v} \in V(\mathbf{s}_0)$. Then, from the definition of $V(\mathbf{s}_0)$, we have for system (16.1) that $\exists\{\mathbf{d}\}, \{\mathbf{z}_M\}$ such that for all t, $\mathbf{z}^t = \Gamma[t]\mathbf{s}_0 + F[t-1]\bar{\mathbf{u}}^{g^{t-1}}$ when $\mathbf{x}_0 = \mathbf{v}$. This implies that $\bar{\mathbf{y}}_H^t = \Gamma_H[t]\mathbf{s}_0 + H[t-1]\bar{\mathbf{u}}^{g^{t-1}}$. Since we also have $\bar{\mathbf{y}}_H^t = \Gamma_H[t]\mathbf{v} + H[t-1]\bar{\mathbf{u}}^{g^{t-1}} + H_M[t-1]\mathbf{d}^{t-1}$, substituting this in the previous equation gives

$$\Gamma_H[t]\mathbf{v} + H_M[t-1]\mathbf{d}^{t-1} = \Gamma_H[t]\mathbf{s}_0. \tag{16.9}$$

Now, if the actuators apply the above sequence $\{\mathbf{d}\}$ to the reduced system (with initial state $\mathbf{v}$), we have for each t, $\mathbf{y}_H^t = \Gamma_H[t]\mathbf{v} + H_M[t-1]\mathbf{d}^{t-1} = \Gamma_H[t]\mathbf{s}_0$, where the last equality follows from the (16.9). Hence, (16.6) is satisfied with $\widetilde{\mathbf{x}}_0 = \mathbf{s}_0$, and hence, $V(\mathbf{s}_0) \subseteq V_R(\mathbf{s}_0)$.

Combining the two results, we have $V(\mathbf{s}_0) = V_R(\mathbf{s}_0)$. ■

The characterization of $V_R(0)$ given in Theorem 16.1 allows one to use standard algorithms that compute $(A, \mathscr{R}(B_M))-$controlled invariant subspaces of a linear dynamical system for computing its unsecurable subspace.

Definition 16.2 The *securable subspace* S of a discrete-time linear dynamical system of the form (16.1) is the orthogonal complement of $V(0)$, the unsecurable subspace of the zero state.

The securable subspace has the interpretation of the maximal set of states that the malicious nodes cannot steer the system to without leaving a nonzero trace at the output of the honest sensors. The following section examines the performance of a stochastic linear dynamical system in the securable subspace, which provides further operational meaning to it.

16.4 Performance in the Securable Subspace in the Context of Stochastic Systems

The previous section contained results analogous to some of those developed for continuous-time, deterministic, linear dynamical systems in [2, 3], and also those reported in [5]. In this section, we report preliminary results of an ongoing work which show that the notion of a securable subspace, as defined in the previous section, could also have operational significance in the context of stochastic systems. While we show this in the context of a simple class of stochastic systems in which the process noise and the initial state are the only sources of uncertainty, similar ideas and proof technique could be applied for the more general model with partial and noisy state observations.

Consider a multiple-input, multiple-output stochastic linear dynamical system described by

$$\mathbf{x}[t+1] = A\mathbf{x}[t] + B\mathbf{u}[t] + \mathbf{w}[t+1], \tag{16.10}$$

$$\mathbf{y}[t+1] = \mathbf{x}[t+1], \tag{16.11}$$

where $\mathbf{x}[t] \in \mathbb{R}^p$, $\mathbf{u}[t] \in \mathbb{R}^m$, $\mathbf{w}[t+1]$ has a known covariance Σ_w, and is independent and identically distributed across time,[1] and A and B are known real matrices of appropriate dimensions. As before, let $\mathbf{u}^g[t] = g_t(z^t)$ denote the control policy-specified input at time t, where $\mathbf{z}[t]$ is the measurement vector reported at time t. Let $\mathbf{d}[t] := \mathbf{u}_M[t] - \mathbf{u}_M^g[t]$, where the subscript 'M,' as usual, denotes the indices of the malicious actuators. Note that without loss of generality, we can assume the honest sensors to be indexed from 1 to h_s, and the honest actuators from 1 to h_a (since the rows and columns of $\mathbf{x}$, A, and B can be reordered accordingly). The system evolves in closed loop as

$$\mathbf{x}[t+1] = A\mathbf{x}[t] + B\mathbf{u}^g[t] + B_M\mathbf{d}[t] + \mathbf{w}[t+1], \tag{16.12}$$

$$\mathbf{y}[t+1] = \mathbf{x}[t+1]. \tag{16.13}$$

The honest nodes in the system perform the following test to determine the presence of malicious nodes in the system.

Test: A honest node checks if

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} (\mathbf{z}[k+1] - A\mathbf{z}[k] - B\mathbf{u}^g[k])\,(\mathbf{z}[k+1] - A\mathbf{z}[k] - B\mathbf{u}^g[k])^T = \Sigma_w. \tag{16.14}$$

[1] More generally, this could be generalized to a martingale difference sequence with a one-step ahead conditional covariance that is uniformly bounded below by a positive definite matrix.

Note that (i) the above test is based only on the information available to the honest nodes in the system, and (ii) if all the nodes in the system are honest, the reported measurements would pass the above test almost surely. The following theorem gives an operational meaning to the securable subspace in the context of stochastic systems.

Theorem 16.3 *Let* $\mathbf{m}[t] := \mathbf{z}[t] - \mathbf{x}[t]$ *be the distortion in the state estimate of the honest nodes. Assume that* $Dim(S \cap \mathrm{N}(C_H)) = 1$. *If the reported sequence of measurements* $\{\mathbf{z}\}$ *passes test (16.14), then,*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} ||\mathbf{m}_S[k]||^2 = 0. \tag{16.15}$$

In other words, the state estimation error caused by malicious sensors and actuators can only be of zero power in the securable subspace.

Proof We define $\gamma[t+1] := \mathbf{m}[t+1] - A\mathbf{m}[t] + B_M\mathbf{d}[t]$, so that test (16.14) reduces to $\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} (\mathbf{w}[k+1] + \gamma[k+1])\ (\mathbf{w}[k+1] + \gamma[k+1])^T = \Sigma_w$. In particular, we have $\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} (\mathbf{w}_H[k+1] + \gamma_H[k+1])\ (\mathbf{w}_H[k+1] + \gamma_H[k+1])^T = \Sigma_{w,H}$, where $\Sigma_{w,H}$ denotes the top-left $h_s \times h_s$ matrix of Σ_w. Since $\mathbf{m}_H[t+1] = 0$ for all t, it follows from the definition of $\gamma[t+1]$ that $\gamma_H[t+1] \in \sigma(\mathbf{m}^t, \mathbf{d}^t)$, where $\gamma_H[t]$ and $\mathbf{m}_H[t]$ are defined in the usual manner. Since $\mathbf{w}_H[k+1]$ is independent of $\sigma(\mathbf{m}^t, \mathbf{d}^t)$, the above equality yields

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} \gamma_H[k+1]\ \gamma_H^T[k+1] = 0. \tag{16.16}$$

Now, from the definition of $\gamma[t+1]$, we have

$$\begin{aligned}\mathbf{m}_V[t+1] + \mathbf{m}_S[t+1] = A\mathbf{m}_V[t] + (A\mathbf{m}_S[t])_V + (A\mathbf{m}_S[t])_S \\ + B_M\mathbf{d}_C[t] + B_M\mathbf{d}_U[t] + \gamma_V[t+1] + \gamma_S[t+1],\end{aligned} \tag{16.17}$$

where $\mathbf{m}_V[t]$ denotes the projection of $\mathbf{m}[t]$ on the unsecurable subspace $V := V(0)$ as in Definition 16.1, $\mathbf{m}_S[t], \gamma_V[t], \gamma_S[t], (A\mathbf{m}_S[t])_V$, and $(A\mathbf{m}_S[t])_S$ are defined likewise, $\mathbf{d}_U[t] := \mathbf{d}[t] - \mathbf{d}_C[t]$, and $\mathbf{d}_C[t]$ is a vector such that $A\mathbf{m}_V[t] + B_M\mathbf{d}_C[t] \in V$, which is guaranteed to exist from the characterization of V given in Theorem 16.1(i).

Now, define $\mathscr{H} := \mathrm{Span}(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{h_s})$, where $\mathbf{e}_i \in \mathbb{R}^p$ is a vector all of whose components are zeros except for the *ith* component, which is unity. Then, we have $\gamma_{\mathscr{H}}[t+1] = \mathbf{m}_{\mathscr{H}}[t+1] - ((A\mathbf{m}_S[t])_S + B_M\mathbf{d}_U[t])_{\mathscr{H}}$. Since $\mathbf{m}_H[t] \equiv 0$, we have $\mathbf{m}_{\mathscr{H}}[t] \equiv 0$. It follows from the above that

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} ||((A\mathbf{m}_S[k])_S + B_M\mathbf{d}_U[k])_{\mathscr{H}}||^2 = \lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} ||\gamma_H[k+1]||^2 = 0, \tag{16.18}$$

where the last equality follows by equating the trace of (16.16).

Now, suppose for contradiction that $\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} ||\mathbf{m}_S[k]||^2 = \epsilon$ for some $\epsilon > 0$. Since $Dim(S \cap \mathrm{N}(C_H)) = 1$, it follows that if $\mathbf{m}_S[k] \neq 0$, then for no choice of $\mathbf{d}_U[k]$ will $((A\mathbf{m}_S[k])_S + B_M\mathbf{d}_U[k]) \in \mathscr{H}^C$. Therefore, if $\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} ||\mathbf{m}_S[k]||^2 = \epsilon$ for any $\epsilon > 0$, then $\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} ||((A\mathbf{m}_S[k])_S + B_M\mathbf{d}_U[k])_{\mathscr{H}}||^2 > \epsilon\delta$ for some $\delta > 0$, contradicting (16.18). ■

16.5 Conclusion

We consider the problem of securing control systems from malicious sensors and actuators. Towards this, we formalize the notion of the securable and unsecurable subspace of a linear dynamical system. The unsecurable subspace has the interpretation as a set of states that the system could actually be in, or ever reach, as a consequence of malicious actions of the adversarial nodes, without the honest sensors in the system ever detecting definitively any malicious activity. This is an invariant subspace in the sense that once the state of the system enters this space, the malicious sensor and actuator nodes in the system can collude to keep the system in this space forever without the honest sensors ever being able to confirm any malicious activity based on their own observations or the ones being reported to them. The orthogonal complement of this subspace, the securable subspace, has the interpretation in the context of deterministic systems as the set of states that the malicious nodes cannot steer the system to without leaving a nonzero trace at the output of the honest sensors. These subspaces also have relevance to the case where the system is noisy. We have presented some preliminary results to show that the notion of a securable subspace has operational significance in the broader context of linear stochastic systems. Specifically, in the context of stochastic systems, the securable subspace has the interpretation as the subspace along which the state estimation error of the honest nodes in the system is what it would have been had there been no malicious nodes in the system, in spite of arbitrary attack strategies of malicious sensors and actuators that are actually present in the system. A characterization of these subspaces, and an algorithm to compute them for any linear system and any combination of malicious sensors and actuators is obtained by a standard recourse to geometric control methods.

Notation

The following notation is used throughout the paper:

1. Let $s_1 < s_2 < ... < s_{h_s}$ denote the indices of the honest sensors, $\psi_1, \psi_2, ..., \psi_{m_s}$ denote those of the malicious sensors, and $a_1 < a_2 < ... < a_{m_a}$ denote those of the malicious actuators. Then,
   - C_H is the $h_s \times p$ matrix whose *ith* row is the s_i^{th} row of C, $i = 1, 2, ..., h_s$,
   - B_M is the $p \times m_a$ matrix whose *ith* column is the a_i^{th} column of B, $i = 1, 2, ..., m_a$,

- $\bar{\mathbf{y}}_H[t]$ is the $h_s \times 1$ vector whose *ith* component is the s_i^{th} entry of $\bar{\mathbf{y}}[t]$, $i = 1, 2, ..., h_s$,
- $\mathbf{z}_M[t]$ is the $m_s \times 1$ vector whose *ith* entry is the ψ_i^{th} entry of $\mathbf{z}[t]$, $i = 1, 2, ..., m_s$,
- $\bar{\mathbf{d}}[t]$ is the $m_a \times 1$ vector whose *ith* component is $\bar{d}_i[t] := \bar{u}_{a_i}[t] - \bar{u}^g_{a_i}[t]$, $i = 1, 2, ..., m_a$.

2. $\mathbf{x}^t$ denotes $[\mathbf{x}^T[0] \;\; \mathbf{x}^T[1] \;\; \ldots \;\; \mathbf{x}^T[t]]^T$.
3. $\Gamma[t] := [C^T \;\; (CA)^T \;\; (CA^2)^T \;\; \ldots \;\; (CA^t)^T]^T$.
4. $\Gamma_H[t] := [(C_H)^T \;\; (C_HA)^T \;\; \ldots \;\; (C_HA^t)^T]^T$.
5. $$F[t] := \begin{bmatrix} 0 & 0 & \ldots & 0 \\ CB & 0 & \ldots & 0 \\ CAB & CB & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^tB & CA^{t-1}B & \ldots & CB \end{bmatrix},$$
6. $$H[t] := \begin{bmatrix} 0 & 0 & \ldots & 0 \\ C_HB & 0 & \ldots & 0 \\ C_HAB & C_HB & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C_HA^tB & C_HA^{t-1}B & \ldots & C_HB \end{bmatrix}, \quad H_M[t] := \begin{bmatrix} 0 & 0 & \ldots & 0 \\ C_HB_M & 0 & \ldots & 0 \\ C_HAB_M & C_HB_M & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C_HA^tB_M & C_HA^{t-1}B_M & \ldots & C_HB_M \end{bmatrix}.$$
7. $\mathrm{N}(\cdot)$ denotes the null space of a matrix, and $\mathscr{R}(\cdot)$ denotes the range space of a matrix.

References

1. Kalman, Rudolf: On the general theory of control systems. IRE Trans. Automatic Control **4**(3), 110–110 (1959)
2. Pasqualetti, Fabio, Dorfler, Florian, Bullo, Francesco: Control-theoretic methods for cyberphysical security: geometric principles for optimal cross-layer resilient control systems. IEEE Control Syst. **35**(1), 110–127 (2015)
3. Pasqualetti, Fabio, Drfler, Florian, Bullo, Francesco: Attack detection and identification in cyber-physical systems. IEEE Trans. Autom. Control **58**(11), 2715–2729 (2013)
4. Pasqualetti, F., Dorfler, F., Francesco, B.: Cyber-physical security via geometric control: distributed monitoring and malicious attacks. In: 2012 IEEE 51st Annual Conference on Decision and Control (CDC), pp. 3418–3425. IEEE (2012)
5. Teixeira, A., Shames, I., Henrik, S., Johansson, K.H.: Revealing stealthy attacks in control systems. In: 2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp. 1806–1813. IEEE (2012)
6. Basile, G., Marro, G.: Controlled and conditioned invariant subspaces in linear system theory. J Optimization Theor. Appl. 306–315 (1969)