공학박사학위논문

Attack-Resilient Feedback Control Systems: Secure State Estimation under Sensor Attacks

외부 공격으로부터 자율 복원 가능한 제어 시스템: 센서 공격에 안전한 상태 추정 기법

2018 년 2 월

서울대학교 대학원

전기컴퓨터공학부

이 찬 화

Attack-Resilient Feedback Control Systems: Secure State Estimation under Sensor Attacks

외부 공격으로부터 자율 복원 가능한 제어 시스템: 센서 공격에 안전한 상태 추정 기법

지도교수 심 형 보

이 논문을 공학박사 학위논문으로 제출함.

2017 년 11 월

서울대학교 대학원

전기컴퓨터공학부

이 찬 화

이 찬 화의 공학박사 학위논문을 인준함.

2017 년 12 월

위 원 장 서 진 헌 (인)

부위원장 심 형 보 (인)

위 원 최 진 영 (인)

위 원 오 성 회 (인)

위 원 은 용 순 (인)

Abstract

Attack-Resilient Feedback Control Systems: Secure State Estimation under Sensor Attacks

Chanhwa Lee
Department of Electrical Engineering and Computer Science
College of Engineering
The Graduate School
Seoul National University

Recent advances in computer and communication technologies make control systems more connected thanks to the developments in networked actuation and sensing devices. As this connectivity increases, the resulting large scale networked control systems, or the cyber-physical systems (CPS), are exposed and can be vulnerable to malicious attacks. In response to the crisis by the malicious adversaries, this dissertation presents sophisticated control algorithms which are more reliable even when some components of the feedback control systems are corrupted. Focusing especially on sensor attacks, security related problems on CPS are carefully analyzed and an attack-resilient state estimation scheme is proposed. First, the notion of *redundant observability* is introduced that explains in a unified manner existing security notions such as dynamic security index, attack detectability, and observability under attacks. The redundant observability is a key concept in this dissertation, and a system is said to be q-redundant observable if it is observable even after eliminating any q measurements. It has been shown that any q-sparse sensor attack is detectable if and only if the given linear time invariant (LTI) system is q-redundant observable. It is also equivalent to the condition that the system is observable under $\lfloor \mathsf{q}/2 \rfloor$-sparse sensor attacks. Moreover, the dynamic security index, which is defined by the minimum number of attacks to be undetectable, can be computed as $\mathsf{q}+1$. In addition, the *redundant detectability* (or, *asymptotic redundant observability*), which is a weaker notion than the redundant observability, is also introduced. While the redundant observability does not care about the magnitudes of sensor attacks and does not mind whether the attacks are disruptive or not, the redundant detectability only deals with attacks that do not converge to zero as time goes on, so that it is more practical in the sense that it can only detect and correct the attacks that are actually harmful to the system. Next, a resilient state estimation scheme is proposed under two assumptions: $\lfloor \mathsf{q}/2 \rfloor$-sparsity of attack vector and q-redundant detectability of the system. The proposed estimator consists of a bank of partial observers operating based on Kalman detectability decomposition and a decoder exploiting error correction techniques. The partial observers are either constructed by Luenberger observers or Kalman filters. The Luenberger observer guarantees the robustness with bounded disturbances/noises, while the Kalman filter shows the suboptimality in the sense of minimum variance with Garussian disturbances/noises. In terms of time complexity, an ℓ_0 minimization problem in the decoder alleviates the computational efforts by reducing the search space to a finite set and by combining a detection algorithm to the optimization process. On the other hand, in terms of space complexity, the required memory is linear with the number of sensors by means of the decomposition used for constructing a bank of partial observers. This resilient state estimation scheme proposed for LTI systems, is further extended for a class of uniformly observable nonlinear systems. Based on the uniform observability decomposition, a high gain observer is constructed for each single measurement to estimate the observable sub-state and it constitutes the partial observer. Finally, the decoder solves a nonlinear error correcting problem by collecting all the information from the high gain observers and by exploiting redundancy.

Keywords: cyber-physical systems, attack resilience, analytical redundancy, security index, attack detection, secure state estimation

Student Number: 2012–30225

사랑하는 가족에게 이 논문을 바칩니다.

Contents

List of Figures

List of Algorithms

Notation and Symbols

$\mathbb{N}$	the set of natural numbers
$\mathbb{R}$	the field of real numbers
$\mathbb{R}^{\mathsf{p}}$	the real Euclidean space of dimension p
$\mathbb{R}^{\mathsf{p}\times\mathsf{n}}$	the space of $\mathsf{p}\times\mathsf{n}$ matrices with real entries
$\mathbb{C}$	the field of complex numbers
$\mathbb{C}^{\mathsf{p}}$	the space of complex vectors of dimension p
$\mathbb{C}^{\mathsf{p}\times\mathsf{n}}$	the space of $\mathsf{p}\times\mathsf{n}$ matrices with complex entries
$I_{\mathsf{n}\times\mathsf{n}}$	the $\mathsf{n}\times\mathsf{n}$ identity matrix (subscript $\mathsf{n}\times\mathsf{n}$ is omitted when there is no confusion.)
$1_{\mathsf{p}\times\mathsf{n}}$	the $\mathsf{p}\times\mathsf{n}$ matrix with all entries equal to one (subscript $\mathsf{p}\times\mathsf{n}$ is omitted when there is no confusion.)
$0_{\mathsf{p}\times\mathsf{n}}$ (or, $O_{\mathsf{p}\times\mathsf{n}}$)	the $\mathsf{p}\times\mathsf{n}$ matrix with all entries equal to zero (subscript $\mathsf{p}\times\mathsf{n}$ is omitted when there is no confusion.)
A^{-1}	the inverse of the square matrix $A\in\mathbb{R}^{\mathsf{n}\times\mathsf{n}}$
$C^{\dagger}$	the pseudoinverse of the matrix $C\in\mathbb{R}^{\mathsf{p}\times\mathsf{n}}$
$C^{\top}$	the transpose of the matrix $C\in\mathbb{R}^{\mathsf{p}\times\mathsf{n}}$
C^{H}	the complex conjugate transpose of the matrix $C\in\mathbb{C}^{\mathsf{p}\times\mathsf{n}}$
$\mathsf{tr}(A)$	the trace of the square matrix $A\in\mathbb{R}^{\mathsf{n}\times\mathsf{n}}$
$\mathsf{det}(A)$	the determinant of the square matrix $A\in\mathbb{R}^{\mathsf{n}\times\mathsf{n}}$
$\mathsf{rank}(C)$	the rank of the matrix $C\in\mathbb{R}^{\mathsf{p}\times\mathsf{n}}$

$\lambda_{\max}(A)$	the maximum eigenvalue of the square matrix $A \in \mathbb{R}^{\mathsf{n}\times\mathsf{n}}$
$\lambda_{\min}(A)$	the minimum eigenvalue of the square matrix $A \in \mathbb{R}^{\mathsf{n}\times\mathsf{n}}$
$\sigma_{\max}(C)$	the maximum singular value of the matrix $C \in \mathbb{R}^{\mathsf{p}\times\mathsf{n}}$
$\sigma_{\min}(C)$	the minimum singular value of the matrix $C \in \mathbb{R}^{\mathsf{p}\times\mathsf{n}}$
$\lvert\alpha\rvert$	the absolute value of the complex number $\alpha \in \mathbb{C}$
$\lvert S\rvert$	the cardinality of the set S
$\mathsf{span}(S)$	the span of the subset S in a vector space, i.e., all linear combinations of elements in S
$[\mathsf{p}]$	the set of natural numbers less than or equal to p, i.e., $\{1, 2, \cdots, \mathsf{p}\}$
$[\mathsf{p}] \setminus \Lambda (= \Lambda^c)$	the complement of the set Λ with respect to the set $[\mathsf{p}]$, i.e., $\{\mathsf{i} \in [\mathsf{p}] : \mathsf{i} \notin \Lambda\}$
y_{i}	the i-th element of the vector $y \in \mathbb{C}^{\mathsf{p}}$
y_{Λ}	the vector in $\mathbb{C}^{\mathsf{p}}$ obtained from $y \in \mathbb{C}^{\mathsf{p}}$ by setting all y_{i}'s such that $\mathsf{i} \in \Lambda^c$ to zero
y_{Λ}^{π}	the vector in $\mathbb{C}^{\lvert\Lambda\rvert}$ obtained from $y \in \mathbb{C}^{\mathsf{p}}$ by eliminating all y_{i}'s such that $\mathsf{i} \in \Lambda^c$
$\mathsf{supp}(y)$	the support of the vector $y \in \mathbb{C}^{\mathsf{p}}$, i.e., $\{\mathsf{i} \in [\mathsf{p}] : y_{\mathsf{i}} \neq 0\}$
$\lVert y\rVert_0$	the ℓ_0 norm of the vector $y \in \mathbb{C}^{\mathsf{p}}$, i.e., $\lVert y\rVert_0 := \lvert\mathsf{supp}(y)\rvert$
Σ_{q}	the set of all q-sparse vectors, i.e., $\{y \in \mathbb{C}^{\mathsf{p}} : \lVert y\rVert_0 \leq \mathsf{q}\}$
$\mathsf{d}_0(y, y')$	the ℓ_0 metric between two vectors y and y', i.e., $\lVert y - y'\rVert_0$
$\lVert y\rVert_1$	the 1-norm of the vector $y \in \mathbb{C}^{\mathsf{p}}$, i.e., $\displaystyle\sum_{\mathsf{i}=1}^{\mathsf{p}} \lvert y_{\mathsf{i}}\rvert$
$\langle y, y'\rangle$	the inner product of two vectors $y, y' \in \mathbb{C}^{\mathsf{p}}$, i.e., $y^{\mathsf{H}}y'$
$\lVert y\rVert_2$	the 2-norm of the vector $y \in \mathbb{C}^{\mathsf{p}}$, i.e., $\sqrt{\langle y, y\rangle} = \sqrt{y^{\mathsf{H}}y}$
$\lVert y\rVert_\infty$	the infinity norm of the vector $y \in \mathbb{C}^{\mathsf{p}}$, i.e., $\max\limits_{\mathsf{i}\in[\mathsf{p}]} \lvert y_{\mathsf{i}}\rvert$
$c_{\mathsf{i},\mathsf{j}}$	the (i, j)-th element of the matrix $C \in \mathbb{R}^{\mathsf{p}\times\mathsf{n}}$
c_{i}	the i-th row of the matrix $C \in \mathbb{R}^{\mathsf{p}\times\mathsf{n}}$

$c_{*,\mathsf{i}}$	the i-th column of the matrix $C \in \mathbb{R}^{\mathsf{p}\times\mathsf{n}}$
C_{Λ}	the matrix in $\mathbb{C}^{\mathsf{p}\times\mathsf{n}}$ obtained from $C \in \mathbb{C}^{\mathsf{p}\times\mathsf{n}}$ by setting all i-th rows such that $\mathsf{i} \in \Lambda^{c}$ to zero row vector
$C_{*,\Lambda}$	the matrix in $\mathbb{C}^{\mathsf{p}\times\mathsf{n}}$ obtained from $C \in \mathbb{C}^{\mathsf{p}\times\mathsf{n}}$ by setting all i-th columns such that $\mathsf{i} \in \Lambda^{c}$ to zero column vector
C_{Λ_1,Λ_2}	the matrix in $\mathbb{C}^{\mathsf{p}\times\mathsf{n}}$ obtained from $C \in \mathbb{C}^{\mathsf{p}\times\mathsf{n}}$ by setting all i-th rows such that $\mathsf{i} \in \Lambda_1^{c}$ to zero row vector and all j-th columns such that $\mathsf{j} \in \Lambda_2^{c}$ to zero column vector
C_{Λ}^{π}	the matrix in $\mathbb{C}^{\lvert\Lambda\rvert\times\mathsf{n}}$ obtained from $C \in \mathbb{C}^{\mathsf{p}\times\mathsf{n}}$ by eliminating all i-th rows such that $\mathsf{i} \in \Lambda^{c}$
$C_{*,\Lambda}^{\pi}$	the matrix in $\mathbb{C}^{\mathsf{p}\times\lvert\Lambda\rvert}$ obtained from $C \in \mathbb{C}^{\mathsf{p}\times\mathsf{n}}$ by eliminating all i-th columns such that $\mathsf{i} \in \Lambda^{c}$
$C_{\Lambda_1,\Lambda_2}^{\pi}$	the matrix in $\mathbb{C}^{\lvert\Lambda_1\rvert\times\lvert\Lambda_2\rvert}$ obtained from $C \in \mathbb{C}^{\mathsf{p}\times\mathsf{n}}$ by eliminating all i-th rows such that $\mathsf{i} \in \Lambda_1^{c}$ and all j-th columns such that $\mathsf{j} \in \Lambda_2^{c}$
$\lVert C\rVert_2$	the induced matrix 2-norm of the matrix $C \in \mathbb{R}^{\mathsf{p}\times\mathsf{n}}$, i.e., $\sqrt{\lambda_{\max}\left(C^{\top}C\right)} = \sigma_{\max}(C)$
$\mathsf{cospark}(C)$	the cospark of the matrix $C \in \mathbb{R}^{\mathsf{p}\times\mathsf{n}}$, i.e., $\min\limits_{x\in\mathbb{R}^{\mathsf{n}},\, x\neq 0_{\mathsf{n}\times 1}} \lVert Cx\rVert_0$
$\mathsf{spark}(F)$	the spark of the matrix $F \in \mathbb{R}^{\mathsf{m}\times\mathsf{p}}$, i.e., $\min\limits_{x\in\mathbb{R}^{\mathsf{p}},\, x\neq 0_{\mathsf{p}\times 1}} \lVert x\rVert_0$ subject to $Fx = 0_{\mathsf{m}\times 1}$
$\Gamma_{\mathsf{i}}^{\mathsf{n}}$	the index set $\{\mathsf{n}(\mathsf{i}-1)+1, \mathsf{n}(\mathsf{i}-1)+2, \cdots, \mathsf{n}\mathsf{i}\}$
Λ^{n}	the index set $\bigcup\limits_{\mathsf{i}\in\Lambda} \Gamma_{\mathsf{i}}^{\mathsf{n}}$
$\mathsf{supp}^{\mathsf{n}}(z)$	the n-stacked support of the n-stacked vector $z = \left[z_1^{\mathsf{n}\top}\ z_2^{\mathsf{n}\top}\ \cdots\ z_{\mathsf{p}}^{\mathsf{n}\top}\right]^{\top} \in \mathbb{C}^{\mathsf{n}\mathsf{p}}$, i.e., $\{\mathsf{i} \in [\mathsf{p}] : z_{\mathsf{i}}^{\mathsf{n}} \neq 0_{\mathsf{n}\times 1}\}$
$\lVert z\rVert_{0^{\mathsf{n}}}$	the n-stacked ℓ_0 norm of $z \in \mathbb{C}^{\mathsf{n}\mathsf{p}}$, i.e., $\lvert\mathsf{supp}^{\mathsf{n}}(z)\rvert$
$\Sigma_{\mathsf{q}}^{\mathsf{n}}$	the set of all n-stacked q-sparse vectors, i.e., $\{z \in \mathbb{C}^{\mathsf{n}\mathsf{p}} : \lVert z\rVert_{0^{\mathsf{n}}} \leq \mathsf{q}\}$
$\mathsf{d}_{0^{\mathsf{n}}}(z, z')$	the n-stacked ℓ_0 metric between two n-stacked vectors z and z', i.e., $\lVert z - z'\rVert_{0^{\mathsf{n}}}$

$\mathsf{cospark}^{\mathsf{n}}(\Phi)$	the n-stacked cospark of the matrix $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$, i.e., $\min_{x\in\mathbb{R}^{\mathsf{n}},\, x\neq 0_{\mathsf{n}\times 1}} \|\Phi x\|_{0^{\mathsf{n}}}$
$\mathcal{Y}^{\perp}$	the orthogonal complement of the subspace $\mathcal{Y} \subset \mathbb{R}^{\mathsf{p}}$, i.e., $\{x \in \mathbb{R}^{\mathsf{p}} : \langle x, y\rangle = 0 \text{ for all } y \in \mathcal{Y}\}$
$\mathsf{dim}(\mathcal{Y})$	the dimension of the vector space $\mathcal{Y}$
$\mathcal{R}(C)$	the range space of the matrix $C \in \mathbb{C}^{\mathsf{p}\times\mathsf{n}}$, i.e., $\{y \in \mathbb{C}^{\mathsf{p}} : y = Cx \text{ for some } x \in \mathbb{C}^{\mathsf{n}}\}$
$\mathcal{N}(C)$	the null space of the matrix $C \in \mathbb{C}^{\mathsf{p}\times\mathsf{n}}$, i.e., $\{x \in \mathbb{C}^{\mathsf{n}} : Cx = 0_{\mathsf{p}\times 1}\}$
$\mathcal{X}_s(A)$	the stable subspace of the matrix $A \in \mathbb{R}^{\mathsf{n}\times\mathsf{n}}$, i.e., $\mathsf{span}\left\{v \in \mathbb{C}^{\mathsf{n}} : (A - \lambda I_{\mathsf{n}\times\mathsf{n}})^k v = 0 \text{ for some } \lvert\lambda\rvert < 1 \text{ and } k \in \mathbb{N}\right\}$
$\mathcal{X}_u(A)$	the unstable subspace of the matrix $A \in \mathbb{R}^{\mathsf{n}\times\mathsf{n}}$, i.e., $\mathsf{span}\left\{v \in \mathbb{C}^{\mathsf{n}} : (A - \lambda I_{\mathsf{n}\times\mathsf{n}})^k v = 0 \text{ for some } \lvert\lambda\rvert \geq 1 \text{ and } k \in \mathbb{N}\right\}$
$\mathcal{V}(A)$	the set of normalized eigenvectors of the matrix $A \in \mathbb{R}^{\mathsf{n}\times\mathsf{n}}$, i.e., $\{v \in \mathbb{C}^{\mathsf{n}} : Av = \lambda v \text{ for some } \lambda \in \mathbb{C},\ \|v\|_2 = 1\}$
$\mathcal{V}_u(A)$	the set of normalized eigenvectors of the matrix $A \in \mathbb{R}^{\mathsf{n}\times\mathsf{n}}$ corresponding to unstable eigenvalues, i.e., $\{v \in \mathbb{C}^{\mathsf{n}} : Av = \lambda v \text{ for some } \lvert\lambda\rvert \geq 1,\ \|v\|_2 = 1\}$
$\overline{\mathcal{O}}(A, C)$	the unobservable subspace of the pair (A, C)
$\overline{\mathcal{O}}_{\mathsf{q}}(A, C)$	the q-redundant unobservable subspace of the pair (A, C)
$\overline{\mathcal{D}}(A, C)$	the undetectable subspace of the pair (A, C)
$\overline{\mathcal{D}}_{\mathsf{q}}(A, C)$	the q-redundant undetectable subspace of the pair (A, C)
$\mathcal{W}(\mathcal{P})$	the weakly unobservable subspace of the system $\mathcal{P}$
$\mathcal{C}(\mathcal{P})$	the controllable weakly unobservable subspace of the system $\mathcal{P}$
$\binom{\mathsf{p}}{\mathsf{q}}$	the binomial coefficient of p and q, i.e., $\frac{\mathsf{p}!}{\mathsf{q}!(\mathsf{p}-\mathsf{q})!}$
$\mathbf{Pr}(E)$	the probability of the event E
$\mathbf{E}[X]$	the expectation of the random variable X
$p_X(x)$	the probability density function of the random variable X

$N(m, P)$	the normal (or, Gaussian) distribution with mean m and covariance P
χ^2_{k}	the chi-squared distribution with k degrees of freedom
δ_{ij}	the Kronecker delta function, i.e., $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$
f^{π}_{Λ}	the canonical projection of the function $f = (f_1, f_2, \cdots, f_{\mathsf{p}}) : \mathbb{R}^{\mathsf{n}} \to \mathbb{R}^{\mathsf{p}}$ by eliminating all i-th entries f_{i}'s such that $\mathsf{i} \in \Lambda^c$
Df	the Jacobian matrix of the function f, i.e., $\dfrac{\partial f}{\partial x}$
id_X	the identity function on the set X, i.e., $\mathsf{id}_X(x) = x$ for $x \in X$
$\mathsf{sat}(x, M)$	the component-wise saturation function of x with the saturation level M and $-M$
$\mathsf{p} \mod \mathsf{q}$	the remainder of the Euclidean division of p by q, i.e., p modulo q
$L_f h(x)$	the Lie derivative of h along the vector field f, i.e., $\dfrac{\partial h}{\partial x} f(x)$
$:=$	defined as
$\equiv$	identically equal
$\Rightarrow$	implies
$\Leftrightarrow$	equivalent to
$\to$	converge to
$\Diamond$	end of theorem, lemma, assumption, remark, and so on
$\Box$	end of proof
$\forall$	for all
i.e.	that is
e.g.	for example
s.t.	such that (or, subject to)
i.i.d.	independent and identically distributed

Notation

- A vector $z \in \mathbb{C}^{\mathsf{np}}$ of length np can be split into p column vectors of length n, i.e., $z = \begin{bmatrix} z_1^{\mathsf{n}\top} & z_2^{\mathsf{n}\top} & \cdots & z_{\mathsf{p}}^{\mathsf{n}\top} \end{bmatrix}^\top \in \mathbb{C}^{\mathsf{np}}$, where $z_{\mathsf{i}}^{\mathsf{n}} \in \mathbb{C}^{\mathsf{n}}$ represents the i-th split column vector of length n in z. Then we call z an n-*stacked vector*. With the index set $\Gamma_{\mathsf{i}}^{\mathsf{n}}$ defined above, it follows that $z_{\mathsf{i}}^{\mathsf{n}} = z^{\pi}_{\Gamma_{\mathsf{i}}^{\mathsf{n}}} \in \mathbb{C}^{\mathsf{n}}$.

- A usual vector $y \in \mathbb{R}^{\mathsf{p}}$ is said to be q-*sparse* if $\|y\|_0 \leq \mathsf{q}$. An n-stacked vector $z \in \mathbb{R}^{\mathsf{np}}$ is said to be n-*stacked* q-*sparse* if $\|z\|_{0^{\mathsf{n}}} \leq \mathsf{q}$.

- With $X \subset \mathbb{R}^{\mathsf{n}}$, a function $f : X \to \mathbb{R}^{\mathsf{p}}$ is said to be *Lipschitz* on X if there exists a constant $\overline{L}$ such that

$$\|f(x) - f(x')\|_\infty \leq \overline{L}\|x - x'\|_\infty, \quad {}^\forall x, x' \in X.$$

  The infimum of such $\overline{L}$ is indicated as $\overline{\mathsf{Lip}}(f)$.

- With $X \subset \mathbb{R}^{\mathsf{n}}$, a differentiable function $f : X \to \mathbb{R}^{\mathsf{p}}$ is called an *immersion* if its Jacobian matrix $Df(x)$ has full column rank for every $x \in X$.

Acronyms

CPS	cyber-physical system
NCS	networked control system
NP	nondeterministic polynomial time
LTI	linear time invariant
CS	compressed sensing
RIP	restricted isometry property
OMP	orthogonal matching pursuit
MVUE	minimum variance unbiased estimator
BLUE	best linear unbiased estimator
WLSE	weighted least squares estimator

PDF	probability density function
ML	maximum likelihood
PBH	Popov-Belevitch-Hautus
DOS	dedicated observer scheme
GOS	generalized observer scheme

Chapter 1

Introduction

1.1 Background

A Cyber-Physical System (CPS) is a system of collaborating computational elements (cyber parts of CPS) controlling physical entities (physical parts of CPS) [3, 43, 68]. Embedded computers and networks monitor and control the physical processes, usually with feedback loops where physical processes affect computations and vice versa. Compared with traditional embedded systems, CPS emphasizes more on an intensive link between the computational and physical elements, while embedded systems focus on the computational device only. Thus, CPS is usually designed as a network of elements that interact with physical inputs and outputs rather than standalone devices. Recent advances in computer and communication technologies have enabled CPS to be prevailing in many engineering areas, such as aerospace, automotive, chemical processes, civil infrastructure, energy, healthcare, manufacturing, transportation, military, robotics, entertainment, and consumer appliances.

In control system community, CPSs are regarded as unprecedented large-scale complex networked control systems (NCS) [26, 98] which are operated over open public networks thanks to increasing connectivity of Internet and recent advances in networked actuation and sensing devices. (See Fig. 1.1.) Reliability of systems under various circumstances is one of the main concerns for control engineers. Robust and fault tolerant control methods are developed to cope with uncertainties in the system model, external disturbances, and failures (or malfunctions)

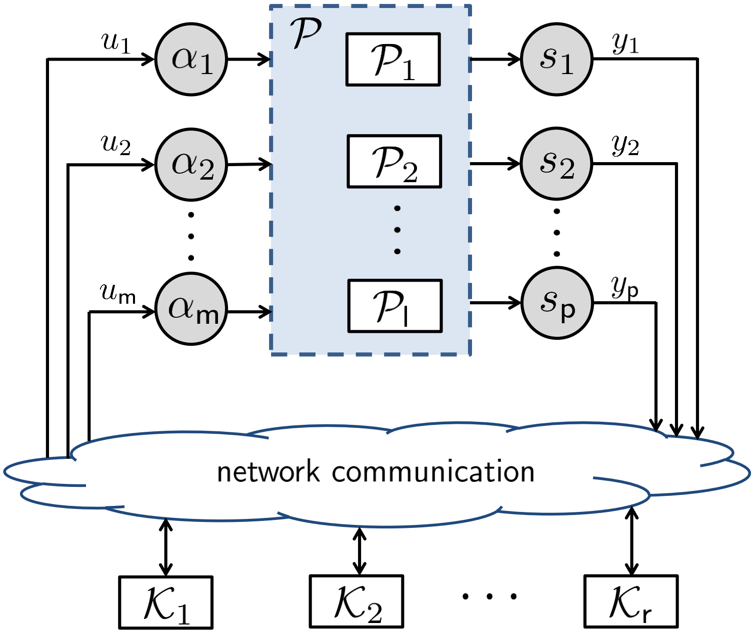


Figure 1.1: Configuration of the networked control system.

in system components. However, new threats or vulnerabilities are reported recently, as advances in computers and communications increase the connectivity between systems whose components are often located remotely through open networks, which is prevalent in NCSs. Indeed, attacks on systems that involve feedback controllers took place in reality [18, 40, 79, 91, 94, 96, 97] and may lead to catastrophic disruptions in critical infrastructure or cause loss of life [91, 94]. For example, the StuxNet worm virus on supervisory control and data acquisition (SCADA) system in Iran nuclear facilities [40], breach at Maroochy Water Services in Austrailia [79], power outage in Ukraine [96], and car hacking [97] are reported. Therefore, attack-resilience of control systems under malicious attacks has become one of the critical system design considerations and is actively studied [54, 70].

From a control systematic perspective, many researchers focus on the system theoretic properties of physical plants and try to enhance security of the system by adopting and modifying advanced control techniques. Various engineering methods are applied to increase security in physical layers of CPS, such as fault detection and isolation [66], robust optimal control [1], estimation theory [12, 42], network graph theory [85], game-theoretic approach [101], and cryptography [36]. Especially, fault detection and isolation techniques [30] are very useful since attacks are similar to faults and can be viewed as unexpected and unwanted data injection to the system components like actuators and sensors. However, the attack signals can be generated in a coordinated way so that they remain stealthy [2,35,63,89,90] since adversaries may infiltrate into the system and inject intentional false data, while the fault signals arise from non-colluding failures and display abnormal behaviors. Consequently, we can regard malicious attacks as a kind of intelligent faults. That is, by exploiting the system model knowledge and data information, attack signals can be manipulated to deceive the target system and remain undetectable. In fault detection and isolation area, model-based residual generation techniques are developed using state estimation methods [20] and "analytical redundancy" is a core element of those techniques [21, 25]. Unlike "physical redundancy" approaches such as [46] and [15], which actually employ additional components and exclude outliers simply by a majority voting logic, the analytical redundancy exploits the inherent redundancy contained in the mathematical model of dynamical systems. In this dissertation, a sort of analytical redundancy in measurements called *redundant observability* is introduced as a key concept that explains in a unified manner existing security notions of control systems under sensor attacks.

Fundamental limitations on security issues such as attack detectability (and identifiability) conditions in consideration of actuator attacks as well as sensor attacks, have been investigated in [66]. In order to assess vulnerability of CPS focusing on sensor attacks, those limitations are quantified by the security index [11, 28, 71] which is the minimum number of attacks to remain undetectable by any type of anomaly detector. That is, the security index is closely related to the notion of attack detectability. Undetectable sensor attacks for a static output map are characterized in [28,39,45,71,81,82], while the security index concept is generalized to a dynamical system under sensor attacks in [11].

Motivated by the considerable works in the field of compressed sensing, error correction, and sparse approximation [8–10, 14, 16, 17, 27, 58, 93], studies on resilient state estimation under sparse sensor attacks, have been carried out recently [12, 19, 31, 51, 62, 75–77]. In [19], basic concepts regarding this problem are introduced and the state reconstruction problem is formulated by an ℓ_0 minimization problem which is NP-hard. By recasting the ℓ_0 minimization problem into a convex optimization problem, it solves the problem under additional restrictive assumptions, however, it can not guarantee real time estimation because the algorithm provides initial state estimates only so that the information is delayed. Bounded noises, disturbances, and modeling errors are considered in [62] and the state estimation error is analyzed, but an explicit error bound is not given. Zero mean Gaussian white noises and disturbances rather than bounded ones, are considered in [44, 51, 55] and Kalman filters are used to guarantee the state estimation performance in a probabilistic manner. Reference [77] proposes an event-triggered projected gradient descent algorithm which is a kind of iterative greedy algorithm [93] with additional restrictive conditions. A satisfiability modulo theory approach, which is a logic to find a healthy combination of sensors by sequentially checking if it satisfies certain binary conditions, is adopted in [76] and [75], but it relies on a heuristic idea for deciding search order from all possible combinations. Authors in [12] prepare all possible candidates of observer combinations to sort out healthy sensors, but the number of observers is fairly large. On the other hand, a computationally efficient estimation scheme using median operation is proposed for the system being observable with every sensor [31].

Although most control systems have nonlinearity in practice, all the above studies are restricted to linear dynamical systems. An attempt to tackle the resilient state estimation problem for nonlinear systems is firstly made in [78], which is a direct extension of the results [76] on linear systems to a class of nonlinear systems, called differentially flat systems. However, by assuming the measurement output to be the "flat" output, the class of systems becomes limited; for example, the given system should not have non-trivial zero dynamics [69]. On the other hand, a secure state estimator is constructed in [29] for a special form of nonlinear systems whose stacked outputs can be represented by a linear function of the initial state and the attack vector.

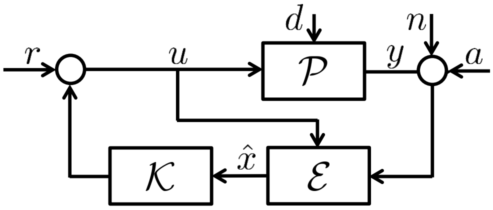


Figure 1.2: Controller-estimatior configuration of the control system under sensor attacks.

1.2 Research Objective and Contributions

Sensors are one of the vulnerable points for security of control systems, and thus, many researchers have studied security problems of the system where measurements are compromised by adversaries [11,12,19,28,31,37,39,45,51,62,64,75–77, 81]. In this dissertation, we consider sensor attacks on feedback control systems whose control connection is the controller-estimator configuration depicted in Fig. 1.2. Note that the design of state feedback ($\mathcal{K}$ in Fig. 1.2) and the design of observer ($\mathcal{E}$ in Fig. 1.2) can be carried out independently for linear time invariant (LTI) systems because the separation principle holds for LTI systems. The main objective of this dissertation is threefold: (i) to analyze the characteristics of linear systems under sensor attacks, focusing on the property of state reconstruction, (ii) to design a secure and attack-resilient estimator which actually recovers the state variable in this situation, and (iii) to extend the results on analysis and design of resilient estimation for linear systems to a class of nonlinear systems.

For discrete-time LTI systems, we first conduct a theoretical analysis and then present a scheme for resilient state estimation. The first part (Chapter 3) of this dissertation deals with the theoretical developments of the relationship between the notion of *redundant observability* and the fundamental limitations on security of control systems under sensor attacks. As mentioned above, attacks have something in common with faults, and a fault detection algorithm primarily utilizes the analytical redundancy. Thus, one can infer that the attack detectability, which directly gives the security index, is linked to the redundancy in some way. It is shown in this dissertation that the redundant observability is the key element and it determines the fundamental limitations of attack detectability. The redundant observability is a sort of analytical redundancy in sensing measurements to reconstruct the system states defined as follows. An LTI system is said to be q-redundant observable if it remains observable even though any q measurement outputs are eliminated. More specifically, it has been shown that any q-sparse sensor attack is detectable if and only if the given LTI system is q-redundant observable, which is again equivalent to the condition that the system is observable under any $\lfloor \mathsf{q}/2 \rfloor$-sparse sensor attacks. In this case, the dynamic security index can be directly obtained as $\mathsf{q}+1$. However, the problem of calculating the security index or the measurement redundancy involves combinatorial logic in nature, and thus, it is NP-hard [28]. To mitigate this computational burden, we also suggest a simple method to compute the dynamic security index by examining only eigenvectors. Furthermore, the *redundant detectability* (or, *asymptotic redundant observability*), which is a weaker notion than the redundant observability, is also introduced. While the redundant observability does not care about the magnitudes of sensor attacks and does not mind whether the attacks are disruptive or not, the redundant detectability only deals with attacks that do not converge to zero as time goes on, so that it is more practical in the sense that it can only detect and correct the attacks that are actually harmful to the system.

In the second part (Chapter 4) of this dissertation, we propose a resilient and robust (or, suboptimal) state estimation scheme under three assumptions: $\lfloor \mathsf{q}/2 \rfloor$-sparsity of attack vector, boundedness (or, Gaussian distribution) of disturbances/ noises, and q-redundant detectability of the system. The proposed estimator consists of a bank of partial observers operating based on the Kalman detectability decomposition and a decoder that exploits error correction techniques. Compared to the existing resilient estimation algorithms in [12, 19, 31, 44, 51, 55, 62, 75–77], advantages of our scheme are as follows. Basically, the proposed scheme assumes the q-redundant detectability that is a weaker notion than the q-redundant observability condition on which other existing resilient estimation algorithms are based. First, it does not require any additional restrictive conditions other than q-redundant detectability (compared with [19,31,75,77]). Second, an observer-based algorithm makes it possible to estimate the current state, not the initial state or delay information (compared with [19,62,76]). Third, the scheme for bounded disturbances/noises, is robust in the sense that a bound on estimation error is explicitly derived from system parameters (compared with [12, 19, 62, 75–77]), and the scheme for Gaussian distributed disturbances/noises, is suboptimal in the sense that it achieves the minimum variance or weighted least square error in more general cases (compared with [44, 51, 55]). Fourth, in terms of time complexity, an ℓ_0 minimization problem in the decoder alleviates the computational efforts by reducing the search space to a finite set and by combining a detection algorithm to the optimization process (compared with other observe-based combinatorial approach such as [12, 51]). Last, the required memory is linear with the number of sensors by means of the decomposition used for constructing a bank of partial observers, and hence, the space complexity is much smaller than that of [12, 51]. Overall, [75] seems to have similar advantages to ours, but it implicitly assumes a certain exhaustive controllability condition for each combination of observers to be transformed into the controllable canonical form.

In the third part (Chapter 5) of this dissertation, a dynamic observer-based resilient state estimation scheme is extended to a class of nonlinear systems. There are a few researches [29, 78] dealing with nonlinear systems for resilient state estimation. For example, authors in [78] first investigated nonlinear differentially flat systems and developed a state estimation scheme under attacks. However, by assuming the measurement output to be the "flat" output, the class of systems becomes very limited because the system should admit a simple static form of observers. In addition, [29] deals with a nonlinear system which has a special structure so that its output can be trivially transformed into a linear function of the initial state and the attack signal. We have presented a resilient estimation algorithm for a class of nonlinear systems called uniformly observable nonlinear systems, which are under much less restrictive class compared with previous results and are often studied in the field of control engineering for nonlinear systems. Assuming that there are enough number of sensors, we show how to counteract the effect of limited number of sensor attacks. In particular, the notion of redundant observability for linear systems is slightly modified for nonlinear systems. Like linear systems, the idea of the resilient estimator implementation is to design partial observers for each output, which is for estimating the observable sub-state only, and then to process the partial information collected from each sensor. For this, the “uniform observability decomposition” from [74], which is an analogous concept of Kalman observability decomposition for linear systems, and a high gain observer [24] are utilized to construct the final detector/estimator. As a by-product of the high gain observer construction, we obtain an assignable convergence rate of the estimation error that converges to zero.

1.3 Outline of the Dissertation

The following outlines this dissertation and briefly presents the contributions of each chapter.

Chapter 2. Error Correction over Reals and its Extensions
The theoretical background of static error correcting problems for a stacked vector case, is presented in this chapter. First, we review some brief backgrounds on error correction techniques for a usual vector case and discuss connection between error correction over reals and compressed sensing. Then, the techniques are extended to the case of stacked vectors, which is an essential mathematical tool in later chapters. This chapter establishes the basic concepts and provides a solid foundation both for theoretical analysis and practical estimator design in attack-resilient estimation. For theoretical studies, the notions of *error detectability* and *error correctability* of the given *coding matrix* are introduced and characterized by suggesting some equivalent conditions. In addition, for practical use, the error detection and state reconstruction schemes are proposed both for bounded noises and Gaussian distributed noises.

Chapter 3. On Redundant Observability
The notion of *redundant observability* is introduced that explains in a unified manner existing security notions such as *dynamic security index*, *attack detectability*, and *observability under attacks*. By following similar procedures used to derive observability in linear system theory, e.g., the observability matrix rank test or Popov-Belevitch-Hautus (PBH) test, we can obtain some equivalent conditions for redundant observability. Moreover, it turns out that an observability matrix behaves like a coding matrix examined in the static error correcting problem, and

hence, its properties (e.g., redundant left invertibility, cospark, error detectability, and error correctability) determine the resilience of control systems under sensor attacks (e.g., redundant observability, dynamic security index, attack detectability, and observability under attacks of the system). Furthermore, the *redundant detectability* (or, *asymptotic redundant observability*), which is a weaker notion than the redundant observability, is also introduced. As the name suggests, the redundant detectability is a concept of redundant observability in an asymptotic sense. That is, if one can recover the state variable asymptotically (in the limit) even after removing any q sensors, the system is called q-redundant detectable. Since the redundant detectability is a kind of asymptotic properties, it treats a signal which converges to zero as a zero signal. Therefore, it only cares about attacks that do not converge to zero as time goes on, while the redundant observability does not care about the magnitudes of sensor attacks. Thus, it is more practical in the sense that it can only detect and correct the attacks that are actually harmful to the system.

Chapter 4. Attack-Resilient State Estimation for Linear Systems

In this chapter, we have developed algorithms which estimate the state variable of the control systems even under sparse sensor attacks. The proposed estimator consists of a bank of partial observers operating based on Kalman detectability (or, observability) decomposition and a decoder exploiting error correction techniques. For bounded disturbances and noises, the partial observers are designed by Luenberger observers and the decoder utilizes the error correction algorithm with bounded noises. The proposed scheme is robust since it guarantees that the estimation error bound is proportional to the noise level. For Gaussian distributed disturbances and noises, the partial observers are constructed by Kalman filters and the decoder operates based on the error correction algorithm with Gaussian noises. The solution of the proposed algorithm is most likely to detect the attacks and is suboptimal since it achieves the minimum variance or weighted least square error for the selected set of sensors. In terms of time complexity, an ℓ_0 minimization problem in the decoder alleviates the computational efforts by reducing the search space to a finite set and by combining a detection algorithm to the optimization process. On the other hand, in terms of space complexity, the required memory is linear with the number of sensors by means of the decomposition used for constructing a bank of partial observers.

Chapter 5. Attack-Resilient State Estimation for Nonlinear Systems

The resilient state estimation algorithm developed for linear systems in the previous chapters, are generalized to uniformly observable nonlinear systems. A high gain observer, which becomes the partial observer whose error decreases to zero exponentially, is constructed for each single output, and it provides the estimate of the observable sub-state. An attack detection scheme, which decides the presence of attack by comparing the residual signal with a time-varying threshold, is proposed. The required condition for this attack detection is less strong than the condition for resilient state estimation, which is natural because the attack can also be revealed/identified once the state is correctly estimated. It will be shown that a detection alarm rings whenever "influential" sensor attacks are injected. If the alarm does not ring then either there is no attack, or the attack is so small that one cannot tell between the sensor attack and the measurement noise. Moreover, by employing the time-varying threshold, the proposed detection scheme also takes into account the transient of the estimation error caused by dynamic observers. Finally, the proposed attack detection algorithm enables resilient state estimation by signaling that the current combination of sensor information is corrupted or not.

Chapter 5. Conclusion

The major contributions of this dissertation is summarized and future research directions are discussed.

Chapter 2

Error Correction over Reals and its Extensions

This chapter provides some brief background on error correction techniques and discuss connection between error correction over reals and compressed sensing. Based on the results of error correcting problems with usual vectors, we extend them to the stacked vector case which forms the foundation of this dissertation.

2.1 Error Correction over Reals and Compressed Sensing

We consider an error correcting problem with real valued input and output which is stated as follows: For the corrupted measurement

$$y = Cx + e \in \mathbb{R}^{\mathsf{p}} \tag{2.1.1}$$

where $C \in \mathbb{R}^{\mathsf{p} \times \mathsf{n}}$ ($\mathsf{p} > \mathsf{n}$) has full column rank and $e \in \mathbb{R}^{\mathsf{p}}$ is q-sparse (i.e., $\|e\|_0 \leq \mathsf{q}$), is it possible to recover x exactly from the data y?

When a *coding matrix* $C \in \mathbb{R}^{\mathsf{p} \times \mathsf{n}}$ has the property that $\|Cx\|_0 > 2\mathsf{q}$ for any $x \in \mathbb{R}^{\mathsf{n}}$ such that $x \neq 0_{\mathsf{n} \times 1}$, it is well known [27, Section 3] that the input x is uniquely recovered by the well-known ℓ_0 minimization problem of

$$\min_{\chi \in \mathbb{R}^{\mathsf{n}}} \|y - C\chi\|_0.$$

In other words, if $\mathsf{cospark}(C) > 2\mathsf{q}$, the input variable x can be recovered by the ℓ_0 minimization decoder $\mathcal{D}_0 : y \mapsto \arg\min_{\chi \in \mathbb{R}^{\mathsf{n}}} \|y - C\chi\|_0$, i.e.,

$$x = \mathcal{D}_0(y) = \arg\min_{\chi \in \mathbb{R}^{\mathsf{n}}} \|y - C\chi\|_0.$$

In recent years, compressed sensing (CS) has attracted remarkable attraction in areas of information theory, coding theory, computer science, statistics, and electrical engineering [9, 14, 16]. CS is an efficient signal sensing protocol which samples data at a lower rate than that of Shannon's theorem suggests, and later recovers the original signal from an incomplete set of measurements. The reconstruction of a signal in CS, i.e., sparse signal recovery, is closely related to the error correcting problem (2.1.1) as follows [9]. We can multiply a full row rank matrix $F \in \mathbb{R}^{(\mathsf{p}-\mathsf{n})\times \mathsf{p}}$, which annihilates the matrix C on the left, i.e., $FC = O_{(\mathsf{p}-\mathsf{n})\times\mathsf{n}}$, to (2.1.1) and obtain

$$\bar{y} = F(Cx + e) = Fe \in \mathbb{R}^{\mathsf{p}-\mathsf{n}}. \tag{2.1.2}$$

Here, $F \in \mathbb{R}^{\mathsf{m}\times\mathsf{p}}$ ($\mathsf{m} = \mathsf{p} - \mathsf{n} < \mathsf{p}$) is called a *sensing matrix* and the error correcting problem of (2.1.1) is transformed to the problem of reconstructing the sparse vector $e \in \mathbb{R}^{\mathsf{p}}$ from the observations $\bar{y} = Fe \in \mathbb{R}^{\mathsf{m}}$. In this situation, if any subsets of $2\mathsf{q}$ columns of F are linearly independent, the q-sparse error vector e can be recovered by searching the sparsest vector $\varepsilon \in \mathbb{R}^{\mathsf{p}}$ that explains $\bar{y} = F\varepsilon$. That is, if $\mathsf{spark}(F) > 2\mathsf{q}$ (note that the spark of the matrix F is the minimal number of columns from F that are linearly dependent), the error variable e can be recovered by

$$e = \arg\min_{\varepsilon \in \mathbb{R}^{\mathsf{p}}} \|\varepsilon\|_0 \quad \text{subject to } \bar{y} = F\varepsilon. \tag{2.1.3}$$

Assuming that e is q sparse, the solvability conditions for (2.1.1) and (2.1.2) are $\mathsf{cospark}(C) > 2\mathsf{q}$ and $\mathsf{spark}(F) > 2\mathsf{q}$, respectively. Actually, when the coding matrix C and the sensing matrix F is related by $FC = O_{\mathsf{m}\times\mathsf{n}}$, we have that

$$\mathsf{cospark}(C) = \mathsf{spark}(F)$$

which easily follows from

$$\min_{x\in\mathbb{R}^{\mathsf{n}},\, x\neq 0_{\mathsf{n}\times 1}} \|Cx\|_0 \quad \Leftrightarrow \quad \min_{y\in\mathbb{R}^{\mathsf{p}},\, y\neq 0_{\mathsf{p}\times 1}} \|y\|_0 \quad \text{subject to } Fy = 0_{\mathsf{m}\times 1}$$

because $Fy = 0_{\mathsf{m}\times 1}$ means that y is of the form Cx (i.e., $y \in \mathcal{R}(C)$) where x is nonzero since $y \neq 0_{\mathsf{p}\times 1}$.

The ℓ_0 minimization of (2.1.3) is a hard combinatorial problem. Hence, in the field of CS, many researchers try to solve (2.1.2) with a computationally feasible algorithm under some additional conditions. The notion of *restricted isometry property* (RIP) introduced by [9], is a key concept in this matter and many sufficient conditions guaranteeing the exact sparse signal recovery is expressed in terms of RIP.

Definition 2.1.1. A sensing matrix $F \in \mathbb{R}^{\mathsf{m}\times\mathsf{p}}$ is said to satisfy the *restricted isometry property* of order k with the *restricted isometry constant* δ_{k} if δ_{k} is the smallest constant such that

$$(1-\delta_{\mathsf{k}})\|e\|_2^2 \leq \|Fe\|_2^2 \leq (1+\delta_{\mathsf{k}})\|e\|_2^2$$

holds for all k-sparse e (i.e., $\|e\|_0 \leq \mathsf{k}$). ♢

One way to reduce the computational effort of solving (2.1.3) is to recast this ℓ_0 minimization problem into a convex optimization which can be easily implemented by linear programming techniques. By replacing $\|\cdot\|_0$ in (2.1.3) with its convex approximation $\|\cdot\|_1$, we obtain a more tractable ℓ_1 minimization problem of

$$\hat{e} = \arg\min_{\varepsilon\in\mathbb{R}^{\mathsf{p}}} \|\varepsilon\|_1 \quad \text{subject to } \bar{y} = F\varepsilon. \tag{2.1.4}$$

The following theorem proved in [10] asserts that the solution of the ℓ_1 minimization problem (2.1.4) is that of the ℓ_0 minimization problem (2.1.3) under some RIP condition with a small enough restricted isometry constant. More precisely, the convex relaxation is exact when $\delta_{2\mathsf{q}} < \sqrt{2}-1$ and $e \in \Sigma_{\mathsf{q}}$.

Theorem 2.1.1. For the measurement $\bar{y} = Fe \in \mathbb{R}^{\mathsf{m}}$ where $F \in \mathbb{R}^{\mathsf{m}\times\mathsf{p}}$ satisfies the RIP of order $2\mathsf{q}$ with the restricted isometry constant

$$\delta_{2\mathsf{q}} < \sqrt{2} - 1$$

and $\|e\|_0 \leq \mathsf{q}$, it follows that

$$e = \arg\min_{\varepsilon \in \mathbb{R}^{\mathsf{p}}} \|\varepsilon\|_1 \quad \text{subject to } \bar{y} = F\varepsilon. \qquad \Diamond$$

While the ℓ_1 relaxation (2.1.4) is called *basis pursuit*, another approach called *matching pursuit* [53, 95] or *greedy algorithm* [93] is also developed. Greedy methods are a kind of iterative approximation and have attracted interest of many researchers for practical benefits. Greedy search algorithms sequentially investigate the support of the sparse signal until a convergence criterion is met, or obtain an improved estimate of the sparse signal at each iteration that attempts to account for the mismatch to the measurements. Orthogonal matching pursuit (OMP) [93] is one of the simple and effective greedy approaches and its algorithm is presented in Algorithm 2.1 where $f_{*,\mathsf{j}}$ represents j-th column of F and $F^{\pi}_{*,\Lambda_{\mathsf{i}}}$ denotes the matrix obtained from F by eliminating all j-th columns such that $\mathsf{j} \in \Lambda_{\mathsf{i}}^c$. In each iteration of OMP, correlations between each column of F and the modified measurements so called residual (r in Algorithm 2.1) are compared each other and the most correlated column is chosen as an element of the support of the sparse signal. The performance guarantees of OMP are similar to those of ℓ_1 minimization (2.1.4) and [53, 95] have shown that OMP exactly recover e from $\bar{y} = Fe$ in a finite iteration with a small restricted isometry constant, stated as follows.

Theorem 2.1.2. For the measurement $\bar{y} = Fe \in \mathbb{R}^{\mathsf{m}}$ where $F \in \mathbb{R}^{\mathsf{m}\times\mathsf{p}}$ satisfies the RIP of order $\mathsf{q}+1$ with the restricted isometry constant

$$\delta_{\mathsf{q}+1} < \frac{1}{\sqrt{\mathsf{q}} + 1}$$

and $\|e\|_0 \leq \mathsf{q}$, OMP described in Algorithm 2.1 perfectly recovers e from $\bar{y} = Fe$ in q iteration. $\Diamond$

Algorithm 2.1 Orthogonal matching pursuit

```
Input: measurement $\bar{y}$, sensing matrix $F$, sparsity $\mathsf{q}$
Output: state estimate $\hat{e}$
Initialization: count $\mathsf{i} = 0$, residual $r_0 = \bar{y}$, support set $\Lambda_0 = \emptyset$
while $\mathsf{i} < \mathsf{q}$ do
    $\mathsf{i} = \mathsf{i} + 1$
    $\Lambda_{\mathsf{i}} = \Lambda_{\mathsf{i}-1} \bigcup \arg\max_{\mathsf{j}} \frac{|\langle f_{*,\mathsf{j}}, r_{\mathsf{i}-1} \rangle|}{\|f_{*,\mathsf{j}}\|_2}$
    $\hat{e}^{\pi}_{\Lambda_{\mathsf{i}}} = \arg\min_{\varepsilon} \|\bar{y} - F^{\pi}_{*,\Lambda_{\mathsf{i}}} \varepsilon\|_2 = \left(F^{\pi}_{*,\Lambda_{\mathsf{i}}}\right)^{\dagger} \bar{y}$
    $r_{\mathsf{i}} = \bar{y} - F^{\pi}_{*,\Lambda_{\mathsf{i}}} \hat{e}^{\pi}_{\Lambda_{\mathsf{i}}}$
end while
$\hat{e} = \arg\min_{\varepsilon \in \{e \colon \mathsf{supp}(e) = \Lambda_{\mathsf{q}}\}} \|\bar{y} - F\varepsilon\|_2$
```

There are another algorithmic approach to sparse signal recovery called combinatorial algorithm. Since combinatorial approach identifies a subset of anomalous elements by investigating all possible combinations [59], it is computationally heavy. However, it does not require any additional assumption other than $\mathsf{spark}(F) > 2\mathsf{q}$, while ℓ_0 minimization and OMP suppose that the sensing matrix satisfy the RIP with a small enough restricted isometry constant. By the way, if we have the freedom to construct and choose the sensing matrix F, which is a common situation in the field of signal processing and information theory, one can generate the sensing matrix by a random matrix in order to satisfy the RIP condition. Authors in [4, 49] have shown that random matrices satisfy the RIP condition with high probability if the elements are selected according to a Gaussian, Bernoulli, or any sub-Gaussian distribution. However, it is not the case in control engineering. Since the coding matrix C comes from the dynamical system, the sensing matrix F which should satisfy $FC = O_{\mathsf{m} \times \mathsf{n}}$ can not be generated by any random matrix. Hence, the RIP condition may not be fulfilled for many control engineering applications. Thus, we have tried to make the best use of the combinatorial algorithm and ℓ_0 minimization in this dissertation because it does not require any additional restrictive condition like the RIP.

2.2 Extension to Stacked Vector Case

In this section, the error correction techniques with a usual vector examined in the previous section, is extended and tailored to the n-stacked vector case which will be encountered frequently in the security problems on control systems. As an extension of (2.1.1), we presume the measurement vector $\hat{z}$ is composed of p measurement data $\hat{z}_{\mathsf{i}}^{\mathsf{n}} \in \mathbb{R}^{\mathsf{n}}$ for $\mathsf{i} \in [\mathsf{p}]$ such that

$$\hat{z} = \begin{bmatrix} \hat{z}_1^{\mathsf{n}} \\ \hat{z}_2^{\mathsf{n}} \\ \vdots \\ \hat{z}_{\mathsf{p}}^{\mathsf{n}} \end{bmatrix}$$

and thus $\hat{z}$ is an n-stacked vector. Given a coding matrix $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$, we want to recover a vector $x \in \mathbb{R}^{\mathsf{n}}$ from the n-stacked measurement vector $\hat{z}$ given by

$$\hat{z} = \Phi x + e \in \mathbb{R}^{\mathsf{np}} \tag{2.2.1}$$

in which the measurement $\hat{z}$ is corrupted by an unknown vector $e \in \mathbb{R}^{\mathsf{np}}$. Here, e is the error vector that is sparse but the magnitude of non-zero elements can be arbitrarily large. Later in this section, we will also take noises into account, and then, the equation (2.2.1) finally has an additional noise vector $v \in \mathbb{R}^{\mathsf{np}}$ which leads the measurement $\hat{z}$ in the following form of

$$\hat{z} = \Phi x + v + e \in \mathbb{R}^{\mathsf{np}} \tag{2.2.2}$$

where the n-stacked measurement $\hat{z}$ is corrupted by the noise vector v as well as the error vector e. The vector v represents noise and is assumed to have bounded magnitude or follows a Gaussian distribution.

2.2.1 Error Detectability and Error Correctability

This section introduces the notions of *error detectability* and *error correctability* when the measurement $\hat{z}$ is noise-free as in (2.2.1). Because one should be able to detect the presence of error in order to reconstruct the original state vector x, two concepts are closely related. We start this section by introducing the error detectability, which is an essential prerequisite for error correctability.

Definition 2.2.1. For a coding matrix $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$, a non-zero error vector $e \in \mathbb{R}^{\mathsf{np}}$ is said to be *undetectable* with respect to Φ if there are two different x and x' in $\mathbb{R}^{\mathsf{n}}$ such that $\Phi x + e = \Phi x'$. ♢

In other words, $e \neq 0_{\mathsf{np}\times 1}$ is undetectable with respect to Φ if and only if $e = \Phi x_e$ for some $x_e \neq 0_{\mathsf{n}\times 1}$. Furthermore, the next lemma indicates that undetectable errors are such signals that are invisible through a residual signal.

Lemma 2.2.1. [45, Theorem 3.2] For the measurement $\hat{z} = \Phi x + e$ where $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$ has full column rank and $x \in \mathbb{R}^{\mathsf{n}}$, let

$$r := \hat{z} - \Phi\Phi^{\dagger}\hat{z} := \left(I_{\mathsf{np}\times\mathsf{np}} - \Phi(\Phi^{\top}\Phi)^{-1}\Phi^{\top}\right)\hat{z}. \tag{2.2.3}$$

Then $e = \Phi x_e \in \mathbb{R}^{\mathsf{np}}$ for some $x_e \in \mathbb{R}^{\mathsf{n}}$ if and only if $r = 0_{\mathsf{np}\times 1}$. In other words, e is undetectable with respect to Φ if and only if $r = 0_{\mathsf{np}\times 1}$. ♢

Note that $\Phi\Phi^{\dagger}$ in Lemma 2.2.1 is an orthogonal projection matrix. Therefore, it projects $\hat{z}$ onto the range space of Φ, $\mathcal{R}(\Phi)$, and we have $\hat{z} \notin \mathcal{R}(\Phi)$ if and only if $\hat{z} \neq \Phi\Phi^{\dagger}\hat{z}$. By the way, Definition 2.2.1 identifies error detectability in respect of the error vector e. As for the coding matrix Φ, this notion can also be defined analogously with a q-sparsity constraint on e as follows.

Definition 2.2.2. A coding matrix $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$ is said to be *(*n*-stacked)* q*-error detectable* if, for all $x, x' \in \mathbb{R}^{\mathsf{n}}$ and $e \in \Sigma_{\mathsf{q}}^{\mathsf{n}}$ such that $\Phi x + e = \Phi x'$, it holds that $x = x'$. ♢

Therefore, the matrix $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$ is not (n-stacked) q-error detectable if and only if there are two different x and x' in $\mathbb{R}^{\mathsf{n}}$, and e in $\Sigma_{\mathsf{q}}^{\mathsf{n}}$ such that $\Phi x + e = \Phi x'$, which corresponds with Definition 2.2.1. Or, when Φ is (n-stacked) q-error detectable and e is (n-stacked) q-sparse, the measurement $\hat{z} = \Phi x + e$ lies in the range space $\mathcal{R}(\Phi)$ if and only if $e = 0_{\mathsf{np}\times 1}$. Now, more equivalent conditions which characterize the error detectability of a coding matrix Φ, are given.

Proposition 2.2.2. The followings are equivalent:

(i) The matrix $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$ is (n-stacked) q-error detectable;

(ii) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p}-\mathsf{q}$, $\Phi_{\Lambda^\mathsf{n}}$ (or, equivalently $\Phi^{\pi}_{\Lambda^\mathsf{n}}$) has full column rank;

(iii) For any $x \in \mathbb{R}^\mathsf{n}$ where $x \neq 0_{\mathsf{n}\times 1}$, $\|\Phi x\|_{0^\mathsf{n}} > \mathsf{q}$;

(iv) For any $x, x' \in \mathbb{R}^\mathsf{n}$ where $x \neq x'$, $\mathsf{d}_{0^\mathsf{n}}(\Phi x, \Phi x') > \mathsf{q}$. ♢

Proof. (i) ⇒ (ii): Suppose that (ii) does not hold, i.e., there exists an index set $\Lambda \subset [\mathsf{p}]$ with $|\Lambda| \geq \mathsf{p}-\mathsf{q}$ and $x \neq 0_{\mathsf{n}\times 1}$ such that $\Phi_{\Lambda^\mathsf{n}}x = 0_{\mathsf{np}\times 1}$. Then it follows that $\|e\|_{0^\mathsf{n}} \leq \mathsf{q}$ where $e := -\Phi x$. Thus, $\Phi x + e = \Phi 0_{\mathsf{n}\times 1}$, and Φ is not q-error detectable.

(ii) ⇒ (iii): Suppose, for the sake of contradiction, that there exists $x \neq 0_{\mathsf{n}\times 1}$ such that $\|\Phi x\|_{0^\mathsf{n}} \leq \mathsf{q}$. Let Λ be the complement of $\mathsf{supp}^\mathsf{n}(\Phi x)$, i.e., $\Lambda = (\mathsf{supp}^\mathsf{n}(\Phi x))^c$. Then it is obvious that $|\Lambda| \geq \mathsf{p}-\mathsf{q}$ and $\Phi_{\Lambda^\mathsf{n}}x = 0_{\mathsf{np}\times 1}$. This contradicts the full column rank condition of $\Phi_{\Lambda^\mathsf{n}}$ in (ii).

(iii) ⇒ (iv): Because $x - x' \neq 0_{\mathsf{n}\times 1}$, we have $\mathsf{d}_{0^\mathsf{n}}(\Phi x, \Phi x') = \|\Phi(x - x')\|_{0^\mathsf{n}} > \mathsf{q}$.

(iv) ⇒ (i): We again prove it by contradiction. Suppose that Φ is not q-error detectable. That is, there exist $x, x' \in \mathbb{R}^\mathsf{n}$ satisfying $x \neq x'$, and $e \in \Sigma^\mathsf{n}_\mathsf{q}$ such that $\Phi x + e = \Phi x'$. It follows from $x - x' \neq 0_{\mathsf{n}\times 1}$ and $e \in \Sigma^\mathsf{n}_\mathsf{q}$ that $\mathsf{d}_{0^\mathsf{n}}(\Phi x', \Phi x) = \|\Phi(x' - x)\|_{0^\mathsf{n}} = \|e\|_{0^\mathsf{n}} \leq \mathsf{q}$. Thus, (iv) fails. □

Remark 2.2.1. In Proposition 2.2.2, the condition (ii) relates q-error detectability to left invertibility of $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$. That is, Φ remains left invertible even if any (n-stacked) q row blocks are eliminated. We may call this property q-*redundant left invertibility.* In other words, one should remove at least $(\mathsf{q}+1)$ row blocks from Φ to break left invertibility, which has something to do with the *sparsest critical* $(\mathsf{q}+1)$-*tuple* found in [81]. On the other hand, the condition (iii) establishes link between the error detectability and the cospark of a coding matrix. More specifically, Φ is q-error detectable if and only if its cospark is larger than q, i.e., $\mathsf{cospark}^\mathsf{n}(\Phi) > \mathsf{q}$. For a usual matrix $C \in \mathbb{R}^{\mathsf{p}\times\mathsf{n}}$, a similar claim of the equivalence between (i) and (ii) is proved in [39, Theorem 1]. A similar claim of the equivalence between (ii) and (iii) is also shown in [81, Theorem 3] for $C \in \mathbb{R}^{\mathsf{p}\times\mathsf{n}}$. Proposition 2.2.2 generalizes those theorems to an (n-stacked) coding matrix $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$. $\Diamond$

On top of the error detectability, the following notion of *error correctability* is now introduced and characterized.

Definition 2.2.3. A coding matrix $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$ is said to be *(*n*-stacked)* q*-error correctable* if, for all $x, x' \in \mathbb{R}^{\mathsf{n}}$ and $e, e' \in \Sigma^{\mathsf{n}}_{\mathsf{q}}$ such that $\Phi x + e = \Phi x' + e'$, it holds that $x = x'$. $\Diamond$

Therefore, when Φ is (n-stacked) q-error correctable, one should be able to recover x (and thus, e as well) from the corrupted measurement $\hat{z} = \Phi x + e$ with $e \in \Sigma^{\mathsf{n}}_{\mathsf{q}}$, because, in principle, one can exhaustively search for all $x' \in \mathbb{R}^{\mathsf{n}}$ and $e' \in \Sigma^{\mathsf{n}}_{\mathsf{q}}$ such that $\hat{z} = \Phi x' + e'$. In other words, if Φ is q-error correctable, then the input $x \in \mathbb{R}^{\mathsf{n}}$ can be uniquely determined from the measurement $\Phi x + e \in \mathbb{R}^{\mathsf{np}}$ whenever the error e is q-sparse. Thus, when the input x and the measurement $\hat{z}$ are related by $\hat{z} = \Phi x + e$ with q-sparse error vector e, there is a decoding map $\mathcal{D} : \mathbb{R}^{\mathsf{np}} \to \mathbb{R}^{\mathsf{n}}$ which can recover the input vector $x \in \mathbb{R}^{\mathsf{n}}$ exactly if and only if the matrix $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$ is q-error correctable. This is not possible if Φ is not q-error correctable because, with Φ not being q-error correctable, there exist $x \neq x' \in \mathbb{R}^{\mathsf{n}}$ and $e, e' \in \Sigma^{\mathsf{n}}_{\mathsf{q}}$ such that $\Phi x + e = \Phi x' + e'$.

Now, by the same arguments as in Proposition 2.2.2, one can easily obtain the following equivalence between error correctability and error detectability.

Proposition 2.2.3. The followings are equivalent:
(i) The matrix $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$ is (n-stacked) q-error correctable;
(ii) The matrix $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$ is (n-stacked) $2\mathsf{q}$-error detectable;
(iii) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - 2\mathsf{q}$, $\Phi_{\Lambda^{\mathsf{n}}}$ (or, equivalently $\Phi^{\pi}_{\Lambda^{\mathsf{n}}}$) has full column rank;
(iv) For any $x \in \mathbb{R}^{\mathsf{n}}$ where $x \neq 0_{\mathsf{n}\times 1}$, $\|\Phi x\|_{0^{\mathsf{n}}} > 2\mathsf{q}$;
(v) For any $x, x' \in \mathbb{R}^{\mathsf{n}}$ where $x \neq x'$, $\mathsf{d}_{0^{\mathsf{n}}}(\Phi x, \Phi x') > 2\mathsf{q}$. $\Diamond$

Proof. (i) $\Rightarrow$ (ii): Assume that $x, x' \in \mathbb{R}^{\mathsf{n}}$ and $e \in \Sigma^{\mathsf{n}}_{2\mathsf{q}}$ satisfying $\Phi x + e = \Phi x'$, are given. Let e_1 and e_2 be such that $e = e_1 - e_2$ where $e_1, e_2 \in \Sigma^{\mathsf{n}}_{\mathsf{q}}$. Thus, we have $\Phi x + e_1 = \Phi x' + e_2$. Since $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$ is q-error correctable, it follows that $x = x'$.

(ii) ⇒ (i): Assume that $x, x' \in \mathbb{R}^{\mathsf{n}}$ and $e, e' \in \Sigma_{\mathsf{q}}^{\mathsf{n}}$ satisfying $\Phi x + e = \Phi x' + e'$, are given. Then, we have $\Phi x + e'' = \Phi x'$ where $e'' = e - e' \in \Sigma_{2\mathsf{q}}^{\mathsf{n}}$. Since $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$ is $2\mathsf{q}$-error detectable, it follows that $x = x'$.
(ii) ⇔ (iii) ⇔ (iv) ⇔ (v): It directly follows from Proposition 2.2.2. □

Proposition 2.2.3 implies that one can detect twice the number of errors that can be corrected and reconstructed. In other words, the error correctability requires twice redundancy in measurements than the error detectability in respect of the redundant left invertibility of a coding matrix.

2.2.2 Error Detection and Correction Scheme for Noiseless Case

Based on the concept of q-error detectability and correctability analyzed in the previous section, we can implement a realizable scheme that actually detects the presence of error and reconstructs the input variable. Those methods for the noiseless measurement (2.2.1) are presented in this section. First, a detection scheme utilizing the residual signal r generated by the discrepancy between the actual measurements $\hat{z}$ and the estimated measurements $\Phi\hat{x}$ where $\hat{x} = \Phi^{\dagger}\hat{z}$ in (2.2.3), is developed. The equivalence between Proposition 2.2.2.(i) and (iii) implies that one can detect non-zero error $e \in \mathbb{R}^{\mathsf{np}}$ as long as $\|e\|_{0^{\mathsf{n}}} < \mathsf{cospark}^{\mathsf{n}}(\Phi)$. Indeed, combining this result with Lemma 2.2.1, one can derive a detection criterion of q-sparse error, which determines whether a q-sparse error compromises the measurements or not. In other words, when ℓ_0 norm of the error e is less than the cospark of Φ, there exists a scheme that can check if $e = 0_{\mathsf{np}\times 1}$.

Lemma 2.2.4. For the measurement $\hat{z} = \Phi x + e$ where $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$ is (n-stacked) q-error detectable, $x \in \mathbb{R}^{\mathsf{n}}$, and $e \in \Sigma_{\mathsf{q}}^{\mathsf{n}}$, let $r = \hat{z} - \Phi\Phi^{\dagger}\hat{z}$. Then $e = 0_{\mathsf{np}\times 1}$ if and only if $r = 0_{\mathsf{np}\times 1}$. Moreover, when $e = 0_{\mathsf{np}\times 1}$, the vector x is recovered by $\hat{x} := \Phi^{\dagger}\hat{z}$. ◊

Proof. Note that no q-sparse error $e \neq 0$ can be represented by Φx_e for some $x_e \in \mathbb{R}^{\mathsf{n}}$ by Proposition 2.2.2.(iii). Therefore, the result directly follows from Lemma 2.2.1. □

In a similar way done in the field of CS, we can multiply an orthogonal full row rank matrix $\Psi \in \mathbb{R}^{(\mathsf{n}-1)\mathsf{p}\times\mathsf{np}}$, which annihilates the matrix Φ on the left of (2.2.1) and obtain another error detection strategy with a slightly modified residual signal.

Corollary 2.2.5. For the measurement $\hat{z} = \Phi x + e$ where $\Phi \in \mathbb{R}^{\mathsf{np} \times \mathsf{n}}$ is (n-stacked) q-error detectable, $x \in \mathbb{R}^{\mathsf{n}}$, and $e \in \Sigma_{\mathsf{q}}^{\mathsf{n}}$, let $r' = \Psi \hat{z}$ where $\Psi \in \mathbb{R}^{(\mathsf{n}-1)\mathsf{p} \times \mathsf{np}}$ is a matrix such that $\Psi\Phi = O_{(\mathsf{n}-1)\mathsf{p} \times \mathsf{n}}$ and $\Psi\Psi^\top = I_{(\mathsf{n}-1)\mathsf{p} \times (\mathsf{n}-1)\mathsf{p}}$. Then $e = 0_{\mathsf{np} \times 1}$ if and only if $r' = 0_{(\mathsf{n}-1)\mathsf{p} \times 1}$. $\Diamond$

Proof. From $\left[\Psi^\top \ (\Phi^\dagger)^\top\right]^\top \left[\Psi^\top \ \Phi\right] = I_{\mathsf{np} \times \mathsf{np}}$, it directly follows that

$$\left[\Psi^\top \ \Phi\right] \left[\Psi^\top \ (\Phi^\dagger)^\top\right]^\top = \Psi^\top \Psi + \Phi\Phi^\dagger = I_{\mathsf{np} \times \mathsf{np}}.$$

Therefore, we have $\Psi^\top \Psi = I_{\mathsf{np} \times \mathsf{np}} - \Phi\Phi^\dagger$ which implies that $\Psi^\top r' = \hat{z} - \Phi\Phi^\dagger \hat{z} = r$ and the proof is completed by Lemma 2.2.4. $\square$

Second, assuming that Φ is (n-stacked) q-error correctable, we will discuss the problem of constructing a decoder that can actually correct (n-stacked) q-sparse error e and recover the input vector x in the noiseless case of (2.2.1). That is, we will find a map $\mathcal{D} : \mathbb{R}^{\mathsf{np}} \to \mathbb{R}^{\mathsf{n}}$ such that $\mathcal{D}(\hat{z}) = x$ where $\hat{z} = \Phi x + e \in \mathbb{R}^{\mathsf{np}}$ and $e \in \Sigma_{\mathsf{q}}^{\mathsf{n}}$. The next lemma says that ℓ_0 minimization introduced for a usual vector error correction problem in Sction 2.1, provides a solution in this situation as well, that is, it indeed works for our n-stacked vector case.

Lemma 2.2.6. For the measurement $\hat{z} = \Phi x + e \in \mathbb{R}^{\mathsf{np}}$ with (n-stacked) q-error correctable $\Phi \in \mathbb{R}^{\mathsf{np} \times \mathsf{n}}$, $x \in \mathbb{R}^{\mathsf{n}}$, and $e \in \Sigma_{\mathsf{q}}^{\mathsf{n}}$,

$$x = \underset{\chi \in \mathbb{R}^{\mathsf{n}}}{\arg\min} \|\hat{z} - \Phi\chi\|_{0^{\mathsf{n}}} \tag{2.2.4}$$

i.e., the decoder $\mathcal{D}_{0^{\mathsf{n}}} : \hat{z} \mapsto \underset{\chi \in \mathbb{R}^{\mathsf{n}}}{\arg\min} \|\hat{z} - \Phi\chi\|_{0^{\mathsf{n}}}$ corrects q errors. $\Diamond$

Proof. If there exists $x' \neq x$ that minimizes $\|\hat{z} - \Phi\chi\|_{0^{\mathsf{n}}}$, then, with $e' := \hat{z} - \Phi x'$, we have that $\hat{z} = \Phi x' + e' = \Phi x + e$ and $\|e'\|_{0^{\mathsf{n}}} \leq \|e\|_{0^{\mathsf{n}}} \leq \mathsf{q}$ because e' is a minimal solution. This implies that Φ is not q-error correctable and thus completes the proof by contradiction. $\square$

Lemma 2.2.6 asserts that, in order to reconstruct x, one should investigate the whole space $\mathbb{R}^{\mathsf{n}}$ to solve (2.2.4). However, the search space in fact can be reduced to a finite set defined by

$$\mathcal{F}_{\mathsf{p},\mathsf{r}}(\hat{z}) := \left\{x \in \mathbb{R}^{\mathsf{n}} : x = (\Phi_{\Lambda^{\mathsf{n}}})^{\dagger}\, \hat{z}_{\Lambda^{\mathsf{n}}} \text{ where } \Lambda \subset [\mathsf{p}] \text{ and } |\Lambda| = \mathsf{p} - \mathsf{r}\right\} \tag{2.2.5}$$

in which, r is any integer satisfying $\mathsf{q} \leq \mathsf{r} \leq 2\mathsf{q}$. Note that $|\mathcal{F}_{\mathsf{p},\mathsf{r}}(\hat{z})| \leq \binom{\mathsf{p}}{\mathsf{p}-\mathsf{r}} = \binom{\mathsf{p}}{\mathsf{r}}$. When it comes to solving (2.2.4), the following theorem claims that it is enough to search over the finite set $\mathcal{F}_{\mathsf{p},\mathsf{r}}(\hat{z})$, not $\mathbb{R}^{\mathsf{n}}$. For this, one can choose r between q and $2\mathsf{q}$ to minimize $\binom{\mathsf{p}}{\mathsf{r}}$.

Theorem 2.2.7. For the measurement $\hat{z} = \Phi x + e \in \mathbb{R}^{\mathsf{np}}$ with (n-stacked) q-error correctable $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$, $x \in \mathbb{R}^{\mathsf{n}}$, and $e \in \Sigma_{\mathsf{q}}^{\mathsf{n}}$,

$$x = \operatorname*{arg\,min}_{\chi \in \mathcal{F}_{\mathsf{p},\mathsf{r}}(\hat{z})} \|\hat{z} - \Phi\chi\|_{0^{\mathsf{n}}} \tag{2.2.6}$$

where r is any integer satisfying $\mathsf{q} \leq \mathsf{r} \leq 2\mathsf{q}$. ◊

Proof. It is enough to show that the vector x belongs to $\mathcal{F}_{\mathsf{p},\mathsf{r}}(\hat{z})$. Pick any subset $\Lambda \subset (\mathsf{supp}^{\mathsf{n}}(e))^c$ satisfying $|\Lambda| = \mathsf{p} - \mathsf{r}$. Because $\Phi_{\Lambda^{\mathsf{n}}}$ has full column rank by Proposition 2.2.3.(iii), it follows that $\chi = (\Phi_{\Lambda^{\mathsf{n}}})^{\dagger}\, \hat{z}_{\Lambda^{\mathsf{n}}} = (\Phi_{\Lambda^{\mathsf{n}}})^{\dagger}\, \Phi_{\Lambda^{\mathsf{n}}} x = x$. Hence, $x \in \mathcal{F}_{\mathsf{p},\mathsf{r}}(\hat{z})$. □

Remark 2.2.2. The ℓ_0 minimization problem (2.2.4) over $\mathbb{R}^n$ is shown to be NP-hard [58] in terms of time computational complexity. This means that the algorithm may not be practically feasible since, in control systems, the computation should be done repeatedly in real-time. Thus, much research effort has been devoted for a relaxation of (2.2.4) by imposing some additional conditions (e.g., applications of basis pursuit and greedy algorithm in control systems as in [19, 77]). It is emphasized that the algorithm proposed in Theorem 2.2.7 actually relieves the computational complexity, not by imposing additional conditions, but by reducing the search space to a finite set. It is a kind of combinatorial approach which tests only $\binom{\mathsf{p}}{\mathsf{r}} \leq \mathsf{p}^{\mathsf{r}}$ (or $\binom{\mathsf{p}}{\mathsf{p}-\mathsf{r}} \leq \mathsf{p}^{\mathsf{p}-\mathsf{r}}$) candidates with the freedom of selecting r between q and $2\mathsf{q}$, while a naive brute force search algorithm without any information on error correctability has no choice but to test all $\binom{\mathsf{p}}{1}+\binom{\mathsf{p}}{2}+\cdots+\binom{\mathsf{p}}{\mathsf{p}}\approx 2^{\mathsf{p}}$ combinations. In our case, the computational efforts decrease drastically by selecting $\mathsf{r}=\mathsf{q}$ when $\mathsf{q}\ll\mathsf{p}$ (or selecting $\mathsf{r}=2\mathsf{q}$ when $\mathsf{q}\approx\mathsf{p}/2$) for example. Compared to other combinatorial algorithms in [42,66,76], Theorem 2.2.7 is more relaxed by introducing r that can varies between q and $2\mathsf{q}$. $\Diamond$

Finally, the following lemma presents a simple criterion to verify whether a given vector $\hat{x}\in\mathbb{R}^{\mathsf{n}}$ coincides with the original input x.

Lemma 2.2.8. For the measurement $\hat{z}=\Phi x+e\in\mathbb{R}^{\mathsf{np}}$ with (n-stacked) q-error correctable $\Phi\in\mathbb{R}^{\mathsf{np}\times\mathsf{n}}$, $x\in\mathbb{R}^{\mathsf{n}}$, and $e\in\Sigma_{\mathsf{q}}^{\mathsf{n}}$,

$$\|\hat{z}-\Phi\hat{x}\|_{0^{\mathsf{n}}}\leq\mathsf{q}\quad\text{if and only if}\quad x=\hat{x}. \qquad \Diamond$$

Proof. (if): This is trivial because $\|\hat{z}-\Phi\hat{x}\|_{0^{\mathsf{n}}}=\|e\|_{0^{\mathsf{n}}}\leq\mathsf{q}$.
(only if): Define $\hat{e}:=\hat{z}-\Phi\hat{x}$, then $\hat{z}=\Phi\hat{x}+\hat{e}=\Phi x+e$ where $e,\hat{e}\in\Sigma_{\mathsf{q}}^{\mathsf{n}}$. Since Φ is q-error correctable, it follows from Definition 2.2.3 that $x=\hat{x}$. □

2.2.3 Error Detection and Correction Scheme for Noisy Case

The previous section showed that the input vector x can be recovered precisely from the noiseless measurements by an optimization problem when the measurement is corrupted by sparse adversaries. However, the measurements are prone to be contaminated by noises in most practical situations. For example, since almost all sensors are not perfect to measure the actual outputs, they may suffer from the quantization errors in digital control systems. In addition, external disturbances can be a source of noises and modeling errors may contribute to deviate the measurements. In other words, consideration of the noise is inevitable because the sensor measurement is not perfect in practice. Moreover, consideration of the noise is also required by the intrinsic property of the dynamic estimator. It will be seen that the dynamic estimator will compute the state estimate that asymptotically converges to the true state as time tends to infinity, but at any finite time, there is small estimation error. This small estimation error will also be treated as the noise in Chapter 4. If the signal recovery method is sensitive to the noise, the solution is not reliable in practical systems. Therefore, stable error detection and signal recovery algorithm need to be devised in the presence of noise.

2.2.3.1 Bounded Noise

Suppose that the measurement $\hat{z} = \Phi x + v + e$ is given where Φ is (n-stacked) q-error correctable, e is (n-stacked) q-sparse, and $v \in \mathbb{R}^{\mathsf{np}}$ satisfies

$$\|v_{\mathsf{i}}^{\mathsf{n}}\|_2 \le v_{\max} \quad \text{for all } \mathsf{i} \in [\mathsf{p}]. \tag{2.2.7}$$

First, let us now consider a detection scheme for the case when the bounded noise $v \in \mathbb{R}^{\mathsf{np}}$ corrupts the measurements as in (2.2.2). For this, let

$$\begin{aligned}
\rho_{\mathsf{p,q}}(\Phi) &:= \min\left\{\sigma_{\min}\left(\Phi_{\Lambda^{\mathsf{n}}}\right) : \Lambda \subset [\mathsf{p}],\ |\Lambda| = \mathsf{p}-\mathsf{q}\right\} \\
&= 1\Big/ \max\left\{\left\| \left(\Phi_{\Lambda^{\mathsf{n}}}\right)^{\dagger} \right\|_2 : \Lambda \subset [\mathsf{p}],\ |\Lambda| = \mathsf{p}-\mathsf{q}\right\}, \\
\eta_{\mathsf{p,q}}(\Phi) &:= \max\left\{\left\|\Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}\left(\Phi_{\Lambda^{\mathsf{n}}}\right)^{\dagger}\right\|_2 : \mathsf{i} \in [\mathsf{p}] \setminus \Lambda,\ \Lambda \subset [\mathsf{p}],\ |\Lambda| = \mathsf{p}-\mathsf{q}\right\}, \\
\kappa_{\mathsf{p,q}}^{d}(\Phi) &:= (\sqrt{\mathsf{p}}+1)\sqrt{\mathsf{p}-\mathsf{q}}/\rho_{\mathsf{p,q}}(\Phi), \\
\kappa_{\mathsf{p,q}}^{e}(\Phi) &:= \left(\eta_{\mathsf{p,q}}(\Phi)\sqrt{\mathsf{p}-\mathsf{q}}+1\right)(\sqrt{\mathsf{p}}+1).
\end{aligned}$$

The proposed method is inspired by the error detection scheme in Lemma 2.2.4, and the following theorem says that one can "practically" detect the q-sparse error in the noisy situation with the residual r given in (2.2.3).

Theorem 2.2.9. For the measurement $\hat{z} = \Phi x + v + e$ where $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$ is (n-stacked) q-error detectable, $x \in \mathbb{R}^{\mathsf{n}}$, $e \in \Sigma_{\mathsf{q}}^{\mathsf{n}}$, and $v \in \mathbb{R}^{\mathsf{np}}$ satisfying $\|v_{\mathsf{i}}^{\mathsf{n}}\|_2 \le v_{\max}$, $^{\forall}\mathsf{i} \in [\mathsf{p}]$, let $\hat{x} = \Phi^{\dagger}\hat{z}$ and $r = \hat{z} - \Phi\hat{x}$. Then,

(i) $e \neq 0_{\mathsf{np}\times 1}$ if

$$\|r_{\mathsf{i}}^{\mathsf{n}}\|_2 = \|\hat{z}_{\mathsf{i}}^{\mathsf{n}} - \Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}\hat{x}\|_2 > \sqrt{\mathsf{p}}\ v_{\max} \quad \text{for some } \mathsf{i} \in [\mathsf{p}],$$

(ii) $\|e_{\mathsf{i}}^{\mathsf{n}}\|_2 \le \kappa_{\mathsf{p},\mathsf{q}}^{e}(\Phi)v_{\max}$, $^{\forall}\mathsf{i} \in [\mathsf{p}]$, if

$$\|r_{\mathsf{i}}^{\mathsf{n}}\|_2 = \|\hat{z}_{\mathsf{i}}^{\mathsf{n}} - \Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}\hat{x}\|_2 \le \sqrt{\mathsf{p}}\; v_{\max} \quad \text{for all } \mathsf{i} \in [\mathsf{p}].$$

In the case of (ii), $\|\hat{x} - x\|_2 \le \kappa_{\mathsf{p},\mathsf{q}}^{d}(\Phi)v_{\max}$. ◊

Proof. (i): This can be shown by contraposition. If $e = 0_{\mathsf{np}\times 1}$, then we have, for all $\mathsf{i} \in [\mathsf{p}]$,

$$\begin{aligned}\|\hat{z}_{\mathsf{i}}^{\mathsf{n}} - \Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}\hat{x}\|_2 &= \|\Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}x + v_{\mathsf{i}}^{\mathsf{n}} - \Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}(\Phi^{\dagger}(\Phi x + v))\|_2 = \|v_{\mathsf{i}}^{\mathsf{n}} - \Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}(\Phi^{\dagger}v)\|_2 \\ &\le \|(I_{\mathsf{np}\times\mathsf{np}} - \Phi\Phi^{\dagger})v\|_2 \le \sqrt{\mathsf{p}}\; v_{\max},\end{aligned}$$

which follows from the fact that $\|I_{\mathsf{np}\times\mathsf{np}} - \Phi\Phi^{\dagger}\|_2 \le 1$.

(ii): Let Λ be a subset of $(\mathsf{supp}^{\mathsf{n}}(e))^{c}$ satisfying $|\Lambda| = \mathsf{p} - \mathsf{q}$. Since $\sqrt{\mathsf{p}}\; v_{\max} \ge \|\hat{z}_{\mathsf{i}}^{\mathsf{n}} - \Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}\hat{x}\|_2$ for all $\mathsf{i} \in \Lambda$ from the assumption, we have

$$\begin{aligned}\sqrt{\mathsf{p}(\mathsf{p}-\mathsf{q})}v_{\max} &\ge \|\hat{z}_{\Lambda^{\mathsf{n}}} - \Phi_{\Lambda^{\mathsf{n}}}\hat{x}\|_2 = \|\Phi_{\Lambda^{\mathsf{n}}}x + v_{\Lambda^{\mathsf{n}}} - \Phi_{\Lambda^{\mathsf{n}}}\hat{x}\|_2 = \|\Phi_{\Lambda^{\mathsf{n}}}(x - \hat{x}) + v_{\Lambda^{\mathsf{n}}}\|_2 \\ &\ge \|\Phi_{\Lambda^{\mathsf{n}}}(x - \hat{x})\|_2 - \|v_{\Lambda^{\mathsf{n}}}\|_2,\end{aligned}$$

which leads to the result that

$$\|\Phi_{\Lambda^{\mathsf{n}}}(x - \hat{x})\|_2 \le (\sqrt{\mathsf{p}} + 1)\sqrt{\mathsf{p}-\mathsf{q}}v_{\max}.$$

Therefore, it is obtained that

$$\|\hat{x} - x\|_2 \le (\sqrt{\mathsf{p}} + 1)\sqrt{\mathsf{p}-\mathsf{q}}v_{\max}/\rho_{\mathsf{p},\mathsf{q}}(\Phi) = \kappa_{\mathsf{p},\mathsf{q}}^{d}(\Phi)v_{\max}.$$

Now, for any $\mathsf{i} \in \Lambda^{c}$, it follows again from the assumption that

$$\begin{aligned}\sqrt{\mathsf{p}}\; v_{\max} &\ge \|\hat{z}_{\mathsf{i}}^{\mathsf{n}} - \Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}\hat{x}\|_2 = \|\Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}x + v_{\mathsf{i}}^{\mathsf{n}} + e_{\mathsf{i}}^{\mathsf{n}} - \Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}\hat{x}\|_2 = \|\Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}(x - \hat{x}) + v_{\mathsf{i}}^{\mathsf{n}} + e_{\mathsf{i}}^{\mathsf{n}}\|_2 \\ &\ge -\|\Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}(x - \hat{x})\|_2 - v_{\max} + \|e_{\mathsf{i}}^{\mathsf{n}}\|_2.\end{aligned}$$

Therefore,

$$\begin{aligned}\|e_{\mathsf{i}}^{\mathsf{n}}\|_2 &\le \|\Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}(x-\hat{x})\|_2 + \sqrt{\mathsf{p}}v_{\max} + v_{\max} = \|\Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}\Phi_{\Lambda^{\mathsf{n}}}^{\dagger}\Phi_{\Lambda^{\mathsf{n}}}(x-\hat{x})\|_2 + (\sqrt{\mathsf{p}}+1)v_{\max}\\ &\le \eta_{\mathsf{p,q}}(\Phi)(\sqrt{\mathsf{p}}+1)\sqrt{\mathsf{p}-\mathsf{q}}v_{\max} + (\sqrt{\mathsf{p}}+1)v_{\max} = \kappa_{\mathsf{p,q}}^{e}(\Phi)v_{\max}, \quad {}^{\forall}\mathsf{i}\in\Lambda^c.\end{aligned}$$

Since $\|e_{\mathsf{i}}^{\mathsf{n}}\|_2 = 0$ for all $\mathsf{i}\in\Lambda$, this completes the proof. □

In fact, when the size of e is small, one cannot tell between the noise v and the error e. The item (ii) in Theorem 2.2.9 reflects this fact and guarantees that the estimation error is small and $\hat{x}$ approximately estimates x.

Now, an algorithm for reconstructing x, which is robust to the noise, is presented.[1] By robustness, we mean that the algorithm can recover the input x from $\hat{z}$ under sparse attack e and bounded noise v, with a guaranteed error bound that is proportional to the noise level. We will first prove that any solution (χ^*, ε^*) to the following relaxed ℓ_0 minimization problem yields an approximation of x as $\hat{x} = \chi^*$:

$$\begin{aligned}&\min_{\chi\in\mathcal{F}_{\mathsf{p,r}}(\hat{z}),\ \varepsilon\in\mathbb{R}^{\mathsf{np}}} \|\varepsilon\|_{0^{\mathsf{n}}}\\ &\text{subject to } \|\hat{z}_{\mathsf{i}}^{\mathsf{n}} - \Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}\chi - \varepsilon_{\mathsf{i}}^{\mathsf{n}}\|_2 \le v'_{\max}, \quad {}^{\forall}\mathsf{i}\in[\mathsf{p}]\end{aligned} \tag{2.2.8}$$

where r is any integer such that $\mathsf{q}\le\mathsf{r}\le 2\mathsf{q}$ and

$$\begin{aligned} v'_{\max} &:= \vartheta_{\mathsf{p,q,r}}(\Phi)v_{\max}\\ &:= \max\left\{\eta'_{\mathsf{p,q,r}}(\Phi)\sqrt{\mathsf{p}-\mathsf{r}}+1, \sqrt{\mathsf{p}-\mathsf{r}}\right\}v_{\max},\\ \eta'_{\mathsf{p,q,r}}(\Phi) &:= \max_{\substack{\Lambda\subset[\mathsf{p}]\\|\Lambda|=\mathsf{p}-\mathsf{q}}}\ \min_{\substack{\bar{\Lambda}\subset\Lambda\\|\bar{\Lambda}|=\mathsf{p}-\mathsf{r}}}\ \max_{\mathsf{i}\in\Lambda\setminus\bar{\Lambda}} \left\|\Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}\left(\Phi_{\bar{\Lambda}^{\mathsf{n}}}\right)^{\dagger}\right\|_2. \end{aligned} \tag{2.2.9}$$

It will be clarified in the proof of Theorem 2.2.11 why $v'_{\max}$, not $v_{\max}$, is used in (2.2.8). The above optimization problem is not easily implementable because there are two optimization variables χ and ε where the variable ε is searched over $\mathbb{R}^{\mathsf{np}}$ under constraints. Hence, we present another optimization problem, which is

[1] Robust signal recovery has also been studied in the literature, but most of them require additional conditions such as uniform uncertainty (or restricted isometry property) [8] or mutual incoherence [17]. Roughly speaking, they are criteria to measure how close the given matrix is to an orthogonal matrix. See [8] and [17].

in more accessible form of:

$$\hat{x} = \operatorname*{arg\,min}_{\chi \in \mathcal{F}_{\mathsf{p},\mathsf{r}}(\hat{z})} \left| \left\{ \mathsf{i} \in [\mathsf{p}] : \|\hat{z}_{\mathsf{i}}^{\mathsf{n}} - \Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi} \chi\|_2 > v'_{\max} \right\} \right|, \tag{2.2.8$'$}$$

or

$$\hat{x} = \operatorname*{arg\,max}_{\chi \in \mathcal{F}_{\mathsf{p},\mathsf{r}}(\hat{z})} \left| \left\{ \mathsf{i} \in [\mathsf{p}] : \|\hat{z}_{\mathsf{i}}^{\mathsf{n}} - \Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi} \chi\|_2 \le v'_{\max} \right\} \right|. \tag{2.2.8$''$}$$

When robustness of the given signal reconstruction scheme is analyzed, the problem (2.2.8) is more useful than (2.2.8$'$) and (2.2.8$''$) because the error vector $\hat{e}$ and the noise vector $\hat{v}$ is directly determined from the solution $\hat{x}$. However, note that (2.2.8) has two optimization variables χ and ε, while (2.2.8$'$) and (2.2.8$''$) have only one optimization variable χ. Hence, in order to solve (2.2.8), one should first minimize $\|\varepsilon\|_{0^{\mathsf{n}}}$ for any given χ under the constraints. By denoting that minimal solution for given χ as $\hat{e}_\chi$, one then calculate the optimal solution $\hat{x}$ minimizing $\|\hat{e}_\chi\|_{0^{\mathsf{n}}}$ among all $\chi \in \mathcal{F}_{\mathsf{p},\mathsf{r}}(\hat{z})$. Consequently, when we implement the algorithm, the unconstrained problem (2.2.8$'$) or (2.2.8$''$) is preferable to (2.2.8) since one can directly calculate the value of the objective function for any given χ without additional optimization process. Actually, (2.2.8$'$) can be interpreted as a relaxation of the problem (2.2.4). The following proposition shows the equivalence of (2.2.8), (2.2.8$'$), and (2.2.8$''$).

Proposition 2.2.10. For the measurement $\hat{z} = \Phi x + e + v \in \mathbb{R}^{\mathsf{np}}$ with (n-stacked) q-error correctable $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$, $x \in \mathbb{R}^{\mathsf{n}}$, $e \in \Sigma_{\mathsf{q}}^{\mathsf{n}}$, and $v \in \mathbb{R}^{\mathsf{np}}$ such that $\|v_{\mathsf{i}}^{\mathsf{n}}\|_2 \le v_{\max}$, $^\forall \mathsf{i} \in [\mathsf{p}]$, three optimization problems (2.2.8), (2.2.8$'$), and (2.2.8$''$) are equivalent (that is, a solution $\hat{x}$ to one optimization problem is also a solution to another optimization problem and vice versa). ◊

Proof. It is trivial that (2.2.8$'$) and (2.2.8$''$) are equivalent. Therefore, we will show that (2.2.8) and (2.2.8$'$) are equivalent. Let $\hat{x} = \chi^*$ and $\hat{e} = \varepsilon^*$ be any solution to (2.2.8), and let $\hat{v} := \hat{z} - \Phi\hat{x} - \hat{e}$. Then, for any $\mathsf{i} \in [\mathsf{p}]$, it automatically holds that $\|\hat{v}_{\mathsf{i}}^{\mathsf{n}}\|_2 \le v'_{\max}$ by the constraint in (2.2.8). Similarly, let $\hat{x}'$ be the solution to (2.2.8$'$). Define $\hat{e}'^{\mathsf{n}}_{\mathsf{j}} := \hat{z}_{\mathsf{j}}^{\mathsf{n}} - \Phi_{\Gamma_{\mathsf{j}}^{\mathsf{n}}}^{\pi}\hat{x}'$ and $\hat{v}'^{\mathsf{n}}_{\mathsf{j}} := 0_{\mathsf{n}\times 1}$ for $\mathsf{j} \in \left\{\mathsf{i} \in [\mathsf{p}] : \|\hat{z}_{\mathsf{i}}^{\mathsf{n}} - \Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}\hat{x}'\|_2 > v'_{\max}\right\}$, and define $\hat{e}'^{\mathsf{n}}_{\mathsf{j}} := 0_{\mathsf{n}\times 1}$ and $\hat{v}'^{\mathsf{n}}_{\mathsf{j}} := \hat{z}_{\mathsf{j}}^{\mathsf{n}} - \Phi_{\Gamma_{\mathsf{j}}^{\mathsf{n}}}^{\pi}\hat{x}'$ for $\mathsf{j} \in \left\{\mathsf{i} \in [\mathsf{p}] : \|\hat{z}_{\mathsf{i}}^{\mathsf{n}} - \Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}\hat{x}'\|_2 \le v'_{\max}\right\}$. Then, $(\chi, \varepsilon) = (\hat{x}', \hat{e}')$ satisfies the constraint in (2.2.8).

We claim that $\hat{x}$ with $\hat{e}$, the solution of (2.2.8), is also a solution of (2.2.8$'$) and vice versa. Indeed, directly from the above definition of $\hat{e}'$, it is obtained that

$$\left|\left\{\mathsf{i}\in[\mathsf{p}] : \|\hat{z}_\mathsf{i}^\mathsf{n} - \Phi^\pi_{\Gamma_\mathsf{i}^\mathsf{n}}\hat{x}'\|_2 > v'_{\max}\right\}\right| = \|\hat{e}'\|_{0^\mathsf{n}}. \tag{2.2.10}$$

On the other hand, because $\|\hat{z}_\mathsf{i}^\mathsf{n} - \Phi^\pi_{\Gamma_\mathsf{i}^\mathsf{n}}\hat{x} - \hat{e}_\mathsf{i}^\mathsf{n}\|_2 \leq v'_{\max}$ for all $\mathsf{i}\in[\mathsf{p}]$, it follows that $\|\hat{z}_\mathsf{j}^\mathsf{n} - \Phi^\pi_{\Gamma_\mathsf{j}^\mathsf{n}}\hat{x}\|_2 \leq v'_{\max}$ for any $\mathsf{j}\in\big(\mathsf{supp}^\mathsf{n}(\hat{e})\big)^c$. Thus, we have

$$\left|\left\{\mathsf{i}\in[\mathsf{p}] : \|\hat{z}_\mathsf{i}^\mathsf{n} - \Phi^\pi_{\Gamma_\mathsf{i}^\mathsf{n}}\hat{x}\|_2 > v'_{\max}\right\}\right| \leq \|\hat{e}\|_{0^\mathsf{n}}. \tag{2.2.11}$$

Since $\hat{e}$ is the minimal solution of (2.2.8), it holds that

$$\|\hat{e}\|_{0^\mathsf{n}} \leq \|\hat{e}'\|_{0^\mathsf{n}}. \tag{2.2.12}$$

Finally, because $\hat{x}'$ is the solution of (2.2.8$'$), it follows that

$$\left|\left\{\mathsf{i}\in[\mathsf{p}] : \|\hat{z}_\mathsf{i}^\mathsf{n} - \Phi^\pi_{\Gamma_\mathsf{i}^\mathsf{n}}\hat{x}'\|_2 > v'_{\max}\right\}\right| \leq \left|\left\{\mathsf{i}\in[\mathsf{p}] : \|\hat{z}_\mathsf{i}^\mathsf{n} - \Phi^\pi_{\Gamma_\mathsf{i}^\mathsf{n}}\hat{x}\|_2 > v'_{\max}\right\}\right|. \tag{2.2.13}$$

Combining (2.2.10), (2.2.11), (2.2.12), and (2.2.13) together results in

$$\begin{aligned}\|\hat{e}\|_{0^\mathsf{n}} &= \left|\left\{\mathsf{i}\in[\mathsf{p}] : \|\hat{z}_\mathsf{i}^\mathsf{n} - \Phi^\pi_{\Gamma_\mathsf{i}^\mathsf{n}}\hat{x}\|_2 > v'_{\max}\right\}\right| \\ =\|\hat{e}'\|_{0^\mathsf{n}} &= \left|\left\{\mathsf{i}\in[\mathsf{p}] : \|\hat{z}_\mathsf{i}^\mathsf{n} - \Phi^\pi_{\Gamma_\mathsf{i}^\mathsf{n}}\hat{x}'\|_2 > v'_{\max}\right\}\right|.\end{aligned}$$

Consequently, $\hat{x}$ is a solution of (2.2.8$'$) and $\hat{x}'$ is a solution of (2.2.8). This concludes the claim. □

While the problem (2.2.8), (2.2.8$'$), or (2.2.8$''$) need not have unique solution, the following theorem establishes a robust estimation scheme which utilizes an optimization problem over a finite set and presents an upper bound of $\|\hat{x} - x\|_2$ for any solution $\hat{x}$ of (2.2.8), (2.2.8$'$), or (2.2.8$''$) with a proportional constant

$$\kappa^c_{\mathsf{p},\mathsf{q},\mathsf{r}}(\Phi) := (\vartheta_{\mathsf{p},\mathsf{q},\mathsf{r}}(\Phi) + 1)\sqrt{\mathsf{p} - 2\mathsf{q}}/\rho_{\mathsf{p},2\mathsf{q}}(\Phi).$$

Theorem 2.2.11. For the measurement $\hat{z} = \Phi x + e + v \in \mathbb{R}^{\mathsf{np}}$ with (n-stacked) q-error correctable $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$, $x \in \mathbb{R}^{\mathsf{n}}$, $e \in \Sigma_{\mathsf{q}}^{\mathsf{n}}$, and $v \in \mathbb{R}^{\mathsf{np}}$ such that $\|v_{\mathsf{i}}^{\mathsf{n}}\|_2 \leq v_{\max}$, $^{\forall}\mathsf{i} \in [\mathsf{p}]$, it holds that

$$\|\hat{x} - x\|_2 \leq \kappa_{\mathsf{p,q,r}}^{c}(\Phi)\ v_{\max},$$

for any solution $\hat{x}$ of the optimization problem (2.2.8), (2.2.8′), or (2.2.8″). ◊

Proof. Let $(\hat{x}, \hat{e})$ be a solution (χ^*, ε^*) of (2.2.8). Then, we first show that $\|\hat{e}\|_{0^{\mathsf{n}}} \leq \mathsf{q}$. Let Λ be a subset of $(\mathsf{supp}^{\mathsf{n}}(e))^c$ satisfying $|\Lambda| = \mathsf{p} - \mathsf{q}$. Then, there always exists a subset $\bar{\Lambda} \subset \Lambda$ such that $|\bar{\Lambda}| = \mathsf{p} - \mathsf{r}$ and

$$\max_{\mathsf{i}\in\Lambda\setminus\bar{\Lambda}} \left\|\Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}} \left(\Phi_{\bar{\Lambda}^{\mathsf{n}}}\right)^{\dagger}\right\|_2 \leq \eta'_{\mathsf{p,q,r}}(\Phi).$$

Let $\bar{x} := \left(\Phi_{\bar{\Lambda}^{\mathsf{n}}}\right)^{\dagger} \hat{z}_{\bar{\Lambda}^{\mathsf{n}}}$, which belongs to $\mathcal{F}_{\mathsf{p,r}}(\hat{z})$. Then it follows that $\bar{x} = x + \left(\Phi_{\bar{\Lambda}^{\mathsf{n}}}\right)^{\dagger} v_{\bar{\Lambda}^{\mathsf{n}}}$ because $\Phi_{\bar{\Lambda}^{\mathsf{n}}}$ has full column rank and thus $\left(\Phi_{\bar{\Lambda}^{\mathsf{n}}}\right)^{\dagger} \Phi_{\bar{\Lambda}^{\mathsf{n}}} = I_{\mathsf{n}\times\mathsf{n}}$. With $\bar{x}$ at hand, let us define a noise vector $\bar{v} := \hat{z}_{\Lambda^{\mathsf{n}}} - \Phi_{\Lambda^{\mathsf{n}}}\bar{x} \in \mathbb{R}^{\mathsf{np}}$ and an error vector $\bar{e} := \hat{z} - \Phi\bar{x} - \bar{v}$. Here, the vector $\bar{v}$ can be decomposed as

$$\begin{aligned}
\bar{v} &= \Phi_{\Lambda^{\mathsf{n}}} x + v_{\Lambda^{\mathsf{n}}} - \Phi_{\Lambda^{\mathsf{n}}}(x + \left(\Phi_{\bar{\Lambda}^{\mathsf{n}}}\right)^{\dagger} v_{\bar{\Lambda}^{\mathsf{n}}}) \\
&= v_{(\Lambda\setminus\bar{\Lambda})^{\mathsf{n}}} + v_{\bar{\Lambda}^{\mathsf{n}}} - \left(\Phi_{(\Lambda\setminus\bar{\Lambda})^{\mathsf{n}}} + \Phi_{\bar{\Lambda}^{\mathsf{n}}}\right) \left(\Phi_{\bar{\Lambda}^{\mathsf{n}}}\right)^{\dagger} v_{\bar{\Lambda}^{\mathsf{n}}} \\
&= v_{(\Lambda\setminus\bar{\Lambda})^{\mathsf{n}}} - \Phi_{(\Lambda\setminus\bar{\Lambda})^{\mathsf{n}}} \left(\Phi_{\bar{\Lambda}^{\mathsf{n}}}\right)^{\dagger} v_{\bar{\Lambda}^{\mathsf{n}}} + (I_{\mathsf{np}\times\mathsf{np}} - \Phi_{\bar{\Lambda}^{\mathsf{n}}}(\Phi_{\bar{\Lambda}^{\mathsf{n}}})^{\dagger}) v_{\bar{\Lambda}^{\mathsf{n}}},
\end{aligned}$$

and thus, it follows that

$$\begin{aligned}
\|\hat{z}_{\mathsf{i}}^{\mathsf{n}} - \Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}\bar{x} - \bar{e}_{\mathsf{i}}^{\mathsf{n}}\|_2 &= \|\bar{v}_{\mathsf{i}}^{\mathsf{n}}\|_2 \\
&\leq \max\left\{\eta'_{\mathsf{p,q,r}}(\Phi)\sqrt{\mathsf{p}-\mathsf{r}} + 1, \sqrt{\mathsf{p}-\mathsf{r}}\right\} v_{\max} = v'_{\max},\ ^{\forall}\mathsf{i} \in [\mathsf{p}],
\end{aligned}$$

in which, we use the fact that $\|I_{\mathsf{np}\times\mathsf{np}} - \Phi_{\bar{\Lambda}^{\mathsf{n}}}(\Phi_{\bar{\Lambda}^{\mathsf{n}}})^{\dagger}\|_2 \leq 1$ and $\|v_{\bar{\Lambda}^{\mathsf{n}}}\|_2 \leq \sqrt{\mathsf{p}-\mathsf{r}}\, v_{\max}$. From the construction, it is clear that $\bar{x}$ and $\bar{e}$ satisfy the constraint in (2.2.8), i.e., $\|\hat{z}_{\mathsf{i}}^{\mathsf{n}} - \Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}\bar{x} - \bar{e}_{\mathsf{i}}^{\mathsf{n}}\|_2 \leq v'_{\max}$ for all $\mathsf{i} \in [\mathsf{p}]$. Moreover, again from the construction, $\|\bar{e}\|_{0^{\mathsf{n}}} \leq \mathsf{q}$. Finally, noting that $\hat{e}$ is the minimal solution of (2.2.8), we have that $\|\hat{e}\|_{0^{\mathsf{n}}} \leq \|\bar{e}\|_{0^{\mathsf{n}}} \leq \mathsf{q}$.

Now, the solution $(\hat{x}, \hat{e})$ of (2.2.8) yields the corresponding noise vector $\hat{v}$ as $\hat{v} := \hat{z} - \Phi\hat{x} - \hat{e}$, which satisfies $\|\hat{v}_{\mathsf{i}}^{\mathsf{n}}\|_2 \leq v'_{\max}$ for all $\mathsf{i} \in [\mathsf{p}]$ by the constraint of (2.2.8). Therefore, we have two expressions for the measurement $\hat{z}$, that is, $\hat{z} = \Phi x + v + e = \Phi\hat{x} + \hat{v} + \hat{e}$, and we are interested in the difference $\tilde{x} := \hat{x} - x$. Let $\tilde{e} := \hat{e} - e$ and $\tilde{v} := \hat{v} - v$. Then, $\|\tilde{e}\|_{0^{\mathsf{n}}} \leq 2\mathsf{q}$ and $\|\tilde{v}_{\mathsf{i}}^{\mathsf{n}}\|_2 \leq v'_{\max} + v_{\max} = (\vartheta_{\mathsf{p,q,r}}(\Phi) + 1)\ v_{\max}$ for all $\mathsf{i} \in [\mathsf{p}]$. Let $\tilde{\Lambda}$ be any subset of $(\mathsf{supp}^{\mathsf{n}}(\tilde{e}))^c$ such that $|\tilde{\Lambda}| = \mathsf{p} - 2\mathsf{q}$. Then, it follows from $\Phi\tilde{x} + \tilde{e} = -\tilde{v}$ that $\Phi_{\tilde{\Lambda}^{\mathsf{n}}}\tilde{x} = -\tilde{v}_{\tilde{\Lambda}^{\mathsf{n}}}$. Since $\Phi_{\tilde{\Lambda}^{\mathsf{n}}}$ has full column rank by Proposition 2.2.3.(iii), it follows that $\tilde{x} = -\left(\Phi_{\tilde{\Lambda}^{\mathsf{n}}}\right)^{\dagger}\tilde{v}_{\tilde{\Lambda}^{\mathsf{n}}}$. Therefore, one can compute the bound of $\|\tilde{x}\|_2$ as

$$\|\tilde{x}\|_2 \leq \left\| \left(\Phi_{\tilde{\Lambda}^{\mathsf{n}}}\right)^{\dagger} \right\|_2 \|\tilde{v}_{\tilde{\Lambda}^{\mathsf{n}}}\|_2 \leq (\vartheta_{\mathsf{p,q,r}}(\Phi)+1)\sqrt{\mathsf{p}-2\mathsf{q}}\,v_{\max}/\rho_{\mathsf{p,2q}}(\Phi) = \kappa^c_{\mathsf{p,q,r}}(\Phi) v_{\max}.$$

□

As in Lemma 2.2.8, a simple criterion to check if a given vector $\hat{x} \in \mathbb{R}^{\mathsf{n}}$ is close to the original x with noisy measurements, is also derived in the following theorem.

Theorem 2.2.12. For the measurement $\hat{z} = \Phi x + e + v \in \mathbb{R}^{\mathsf{np}}$ with (n-stacked) q-error correctable $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$, $x \in \mathbb{R}^{\mathsf{n}}$, $e \in \Sigma^{\mathsf{n}}_{\mathsf{q}}$, and $v \in \mathbb{R}^{\mathsf{np}}$ such that $\|v_{\mathsf{i}}^{\mathsf{n}}\|_2 \leq v_{\max}$, $^{\forall}\mathsf{i} \in [\mathsf{p}]$, it holds that

(i) $\|\hat{x} - x\|_2 \leq \kappa^c_{\mathsf{p,q,r}}(\Phi)\ v_{\max}$ if $\hat{x}$ satisfies

$$\left|\left\{\mathsf{i} \in [\mathsf{p}] : \|\hat{z}_{\mathsf{i}}^{\mathsf{n}} - \Phi^{\pi}_{\Gamma^{\mathsf{n}}_{\mathsf{i}}}\hat{x}\|_2 > v'_{\max}\right\}\right| \leq \mathsf{q},$$

(ii) $\|\hat{x} - x\|_2 > \kappa^{c'}_{\mathsf{p,q,r}}(\Phi)\ v_{\max}$ if $\hat{x}$ satisfies

$$\left|\left\{\mathsf{i} \in [\mathsf{p}] : \|\hat{z}_{\mathsf{i}}^{\mathsf{n}} - \Phi^{\pi}_{\Gamma^{\mathsf{n}}_{\mathsf{i}}}\hat{x}\|_2 > v'_{\max}\right\}\right| > \mathsf{q},$$

where $\kappa^{c'}_{\mathsf{p,q,r}}(\Phi) := (\vartheta_{\mathsf{p,q,r}}(\Phi) - 1)/\max_{\mathsf{i}\in[\mathsf{p}]}\|\Phi_{\Gamma^{\mathsf{n}}_{\mathsf{i}}}\|_2$. ◊

Proof. (i): With $\hat{x}$ given, construct the error vector $\hat{e}$ and the noise vector $\hat{v}$ as follows. For $\mathsf{j} \in \left\{\mathsf{i} \in [\mathsf{p}] : \|\hat{z}_{\mathsf{i}}^{\mathsf{n}} - \Phi^{\pi}_{\Gamma^{\mathsf{n}}_{\mathsf{i}}}\hat{x}\|_2 > v'_{\max}\right\}$, define $\hat{e}_{\mathsf{j}}^{\mathsf{n}} := \hat{z}_{\mathsf{j}}^{\mathsf{n}} - \Phi^{\pi}_{\Gamma^{\mathsf{n}}_{\mathsf{j}}}\hat{x}$ and $\hat{v}_{\mathsf{j}}^{\mathsf{n}} := 0_{\mathsf{n}\times 1}$. For $\mathsf{j} \in \left\{\mathsf{i} \in [\mathsf{p}] : \|\hat{z}_{\mathsf{i}}^{\mathsf{n}} - \Phi^{\pi}_{\Gamma^{\mathsf{n}}_{\mathsf{i}}}\hat{x}\|_2 \leq v'_{\max}\right\}$, define $\hat{e}_{\mathsf{j}}^{\mathsf{n}} := 0_{\mathsf{n}\times 1}$ and $\hat{v}_{\mathsf{j}}^{\mathsf{n}} := \hat{z}_{\mathsf{j}}^{\mathsf{n}} - \Phi^{\pi}_{\Gamma^{\mathsf{n}}_{\mathsf{j}}}\hat{x}$. Then, we have two expressions for the measurement $\hat{z} = \Phi x + v + e = \Phi\hat{x} + \hat{v} + \hat{e}$, in which $\|\hat{v}_{\mathsf{i}}^{\mathsf{n}}\|_2 \leq v'_{\max}$ for all $\mathsf{i} \in [\mathsf{p}]$ and $\|\hat{e}\|_{0^{\mathsf{n}}} \leq \mathsf{q}$. Therefore, the same argument in the proof of Theorem 2.2.11 applies and concludes the claim.

(ii): This is shown by contradiction. Let $\|\hat{x} - x\|_2 \leq \kappa^{c'}_{\mathsf{p,q,r}}(\Phi)\ v_{\max}$. Then, for $\mathsf{i} \in (\mathsf{supp}^{\mathsf{n}}(e))^c$,

$$\begin{aligned}\|\hat{z}^{\mathsf{n}}_{\mathsf{i}} - \Phi^{\pi}_{\Gamma^{\mathsf{n}}_{\mathsf{i}}}\hat{x}\|_2 = \|\Phi^{\pi}_{\Gamma^{\mathsf{n}}_{\mathsf{i}}}(x-\hat{x}) + v^{\mathsf{n}}_{\mathsf{i}}\|_2 &\leq \Big(\max_{\mathsf{i}\in[\mathsf{p}]}\|\Phi_{\Gamma^{\mathsf{n}}_{\mathsf{i}}}\|_2\Big)\kappa^{c'}_{\mathsf{p,q,r}}(\Phi)v_{\max} + v_{\max} \\ &= \vartheta_{\mathsf{p,q,r}}(\Phi)v_{\max} = v'_{\max}.\end{aligned}$$

Therefore, we have $\left|\left\{\mathsf{i} \in [\mathsf{p}] : \|\hat{z}^{\mathsf{n}}_{\mathsf{i}} - \Phi^{\pi}_{\Gamma^{\mathsf{n}}_{\mathsf{i}}}\hat{x}\|_2 > v'_{\max}\right\}\right| \leq \mathsf{q}$ because $\|e\|_{0^{\mathsf{n}}} \leq \mathsf{q}$. □

2.2.3.2 Gaussian Noise

Suppose that the measurement $\hat{z} = \Phi x + v + e$ is given where Φ is (n-stacked) q-error correctable, e is (n-stacked) q-sparse, and $v \in \mathbb{R}^{\mathsf{np}}$ is a Gaussian measurement noise which is assumed that

$$v \sim N(0_{\mathsf{np}\times 1}, P) \quad \text{with } P > 0. \tag{2.2.14}$$

With this stochastic noise, we will first extend the results on bounded noise case to the detection and estimation properties in probability. Then, more sophisticated algorithm will further be developed by the statistical decision theory.

Recall that the assumption on the noise v in the previous section was (2.2.7). Thus, from Gaussian distribution (2.2.14), the probability of satisfying the condition (2.2.7) is calculated and defined as follows:

$$\begin{aligned}p_v &:= \mathbf{Pr}\Big(\|v^{\mathsf{n}}_{\mathsf{i}}\|_2 \leq v_{\max}, ^{\forall}\mathsf{i} \in [\mathsf{p}]\Big) \\ &= \int_{\left\{v\in\mathbb{R}^{\mathsf{np}}:\ \|v^{\mathsf{n}}_{\mathsf{i}}\|_2 \leq v_{\max}, ^{\forall}\mathsf{i}\in[\mathsf{p}]\right\}} \frac{\exp\left(-\frac{1}{2}v^{\top}P^{-1}v\right)}{\sqrt{(2\pi)^{\mathsf{np}}\mathsf{det}(P)}}dv.\end{aligned} \tag{2.2.15}$$

Therefore, with this probability p_v, Theorems 2.2.9, 2.2.11, and 2.2.12 lead to the following corollaries of the detection and estimation in probability.

Corollary 2.2.13. For the measurement $\hat{z} = \Phi x + v + e$ where $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$ is (n-stacked) q-error detectable, $x \in \mathbb{R}^{\mathsf{n}}$, $e \in \Sigma^{\mathsf{n}}_{\mathsf{q}}$, and $v \in \mathbb{R}^{\mathsf{np}}$ satisfying $v \sim N(0_{\mathsf{np}\times 1}, P)$ with $P > 0$, let $\hat{x} = \Phi^{\dagger}\hat{z}$ and $r = \hat{z} - \Phi\hat{x}$. Then,

(i) $\mathbf{Pr}\left(e \neq 0_{\mathsf{np}\times 1}\right) \geq p_v$ if

$$\|r_{\mathsf{i}}^{\mathsf{n}}\|_2 = \|\hat{z}_{\mathsf{i}}^{\mathsf{n}} - \Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}\hat{x}\|_2 > \sqrt{\mathsf{p}}\ v_{\max} \quad \text{for some } \mathsf{i} \in [\mathsf{p}],$$

(ii) $\mathbf{Pr}\left(\|e_{\mathsf{i}}^{\mathsf{n}}\|_2 \leq \kappa_{\mathsf{p,q}}^{e}(\Phi)v_{\max},\ ^{\forall}\mathsf{i} \in [\mathsf{p}]\right) \geq p_v$, if

$$\|r_{\mathsf{i}}^{\mathsf{n}}\|_2 = \|\hat{z}_{\mathsf{i}}^{\mathsf{n}} - \Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}\hat{x}\|_2 \leq \sqrt{\mathsf{p}}\ v_{\max} \quad \text{for all } \mathsf{i} \in [\mathsf{p}].$$

In the case of (ii), $\mathbf{Pr}\left(\|\hat{x} - x\|_2 \leq \kappa_{\mathsf{p,q}}^{d}(\Phi)v_{\max}\right) \geq p_v$. ◊

Corollary 2.2.14. For the measurement $\hat{z} = \Phi x + e + v \in \mathbb{R}^{\mathsf{np}}$ with (n-stacked) q-error correctable $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$, $x \in \mathbb{R}^{\mathsf{n}}$, $e \in \Sigma_{\mathsf{q}}^{\mathsf{n}}$, and $v \in \mathbb{R}^{\mathsf{np}}$ satisfying $v \sim N(0_{\mathsf{np}\times 1}, P)$ with $P > 0$, it holds that

$$\mathbf{Pr}\left(\|\hat{x} - x\|_2 \leq \kappa_{\mathsf{p,q,r}}^{c}(\Phi)\ v_{\max}\right) \geq p_v,$$

for any solution $\hat{x}$ of the optimization problem (2.2.8), (2.2.8′), or (2.2.8″). ◊

Corollary 2.2.15. For the measurement $\hat{z} = \Phi x + e + v \in \mathbb{R}^{\mathsf{np}}$ with (n-stacked) q-error correctable $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$, $x \in \mathbb{R}^{\mathsf{n}}$, $e \in \Sigma_{\mathsf{q}}^{\mathsf{n}}$, and $v \in \mathbb{R}^{\mathsf{np}}$ satisfying $v \sim N(0_{\mathsf{np}\times 1}, P)$ with $P > 0$, it holds that
(i) $\mathbf{Pr}\left(\|\hat{x} - x\|_2 \leq \kappa_{\mathsf{p,q,r}}^{c}(\Phi)\ v_{\max}\right) \geq p_v$ if $\hat{x}$ satisfies

$$\left|\left\{\mathsf{i} \in [\mathsf{p}] : \|\hat{z}_{\mathsf{i}}^{\mathsf{n}} - \Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}\hat{x}\|_2 > v'_{\max}\right\}\right| \leq \mathsf{q},$$

(ii) $\mathbf{Pr}\left(\|\hat{x} - x\|_2 > \kappa_{\mathsf{p,q,r}}^{c'}(\Phi)\ v_{\max}\right) \geq p_v$ if $\hat{x}$ satisfies

$$\left|\left\{\mathsf{i} \in [\mathsf{p}] : \|\hat{z}_{\mathsf{i}}^{\mathsf{n}} - \Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}^{\pi}\hat{x}\|_2 > v'_{\max}\right\}\right| > \mathsf{q},$$

where $\kappa_{\mathsf{p,q,r}}^{c'}(\Phi) := (\vartheta_{\mathsf{p,q,r}}(\Phi) - 1)/\max_{\mathsf{i}\in[\mathsf{p}]}\|\Phi_{\Gamma_{\mathsf{i}}^{\mathsf{n}}}\|_2$. ◊

Putting these basic results with probability aside, more elaborate derivations based on the statistical estimation and detection theory [32,33], is now presented. First, the minimum variance unbiased estimator (MVUE) for the measurement (2.2.2) with $e = 0_{\mathsf{np}\times 1}$ and v satisfying $v \sim N(0_{\mathsf{np}\times 1}, P)$ is introduced as follows.

Lemma 2.2.16. [32, Theorem 4.2] For the measurement $\hat{z} = \Phi x + v \in \mathbb{R}^{\mathsf{np}}$ with $x \in \mathbb{R}^{\mathsf{n}}$ and $v \in \mathbb{R}^{\mathsf{np}}$ such that $v \sim N(0_{\mathsf{np}\times 1}, P)$ for some $P > 0$, the minimum variance unbiased estimator (MVUE) of x is

$$\hat{x}_{\text{MVUE}} = \left(\Phi^\top P^{-1}\Phi\right)^{-1} \Phi^\top P^{-1}\hat{z} \tag{2.2.16}$$

and the corresponding covariance matrix of $\hat{x}_{\text{MVUE}}$ is

$$P_{\hat{x}_{\text{MVUE}}} = \left(\Phi^\top P^{-1}\Phi\right)^{-1} \tag{2.2.17}$$

which achieves the minimum covariance in the sense that $P_{\hat{x}_{\text{MVUE}}} \leq P_{\hat{x}}$ for any type of estimator $\hat{x}$. ◊

Note that Gauss-Markov Theorem [32, Theorem 6.1] gives the best linear unbiased estimator (BLUE) for the measurement $\hat{z} = \Phi x + v$ where v is a random variable, whose probability density function (PDF) is not restricted to a Gaussian distribution, with zero mean and covariance P. Since the BLUE is also the MVUE for Gaussian data, the results of Lemma 2.2.16 also follows directly from Gauss-Markov Theorem. A special case of Lemma 2.2.16 is considered in [83, Theorem 1] and [84, Theorem 1] for an information fusion scheme, and it can be easily proved by the Lagrangian method [6].

Furthermore, it is not difficult to show that the MVUE of x in (2.2.16), is also the weighted least squares estimator (WLSE) with a performance index $J = (\hat{z} - \Phi x)^\top P^{-1} (\hat{z} - \Phi x)$, when the measurement is given by $\hat{z} = \Phi x + v \in \mathbb{R}^{\mathsf{np}}$. That is, we have

$$\begin{aligned} \hat{x}_{\text{WLSE}} &:= \arg\min_{\chi \in \mathbb{R}^{\mathsf{n}}} (\hat{z} - \Phi\chi)^\top P^{-1} (\hat{z} - \Phi\chi) \\ &= \left(\Phi^\top P^{-1}\Phi\right)^{-1} \Phi^\top P^{-1}\hat{z} = \hat{x}_{\text{MVUE}}, \end{aligned}$$

which is summarized in the following lemma.

Lemma 2.2.17. [32, Section 8.4] For the solution $(\hat{x}_{\text{WLSE}}, \hat{v})$ of the minimization problem

$$\min_{\chi \in \mathbb{R}^{\mathsf{n}},\ v \in \mathbb{R}^{\mathsf{np}}} v^\top P^{-1} v \quad \text{subject to } \hat{z} = \Phi\chi + v,$$

it holds that

$$\hat{x}_{\text{WLSE}} = \left(\Phi^\top P^{-1}\Phi\right)^{-1} \Phi^\top P^{-1}\hat{z}. \tag{2.2.18}$$

The optimal solution $\hat{x}_{\text{WLSE}}$ is called the weighted least squares estimator (WLSE) of x with weight P^{-1}. $\Diamond$

For the measurement $\hat{z} = \Phi x + v + e$ with a coding matrix $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$, the input signal x and the error signal e can be seen as unknown deterministic variables while the noise signal v can be considered as a random variable whose distribution is $N(0_{\mathsf{np}\times 1}, P)$. With the estimate $\hat{x}$ of x obtained by MVUE or WLSE, we can calculate the estimated output $\Phi\hat{x}$ and generate a residual signal r which is a difference between the real measurement and the estimated output, i.e., $r := \hat{z} - \Phi\hat{x}$. Then the residual r becomes another random variable whose distribution is also Gaussian. Finally, the mean and covariance of the Gaussian distributed random variable r is given in the following theorem.

Theorem 2.2.18. For the measurement $\hat{z} = \Phi x + v + e$ where $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$ has full column rank and v satisfies $v \sim N(0_{\mathsf{np}\times 1}, P)$ with $P > 0$, let $\hat{x} = \left(\Phi^\top P^{-1}\Phi\right)^{-1} \Phi^\top P^{-1}\hat{z} =: \Psi\hat{z}$ and

$$\begin{aligned} r := \hat{z} - \Phi\hat{x} &= (I_{\mathsf{np}\times\mathsf{np}} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})\hat{z} \\ &= (I_{\mathsf{np}\times\mathsf{np}} - \Phi\Psi)\hat{z}. \end{aligned} \tag{2.2.19}$$

Then, the residual r is Gaussian distributed with mean $(I_{\mathsf{np}\times\mathsf{np}} - \Phi\Psi)e$ and covariance $(I_{\mathsf{np}\times\mathsf{np}} - \Phi\Psi)P$, i.e.,

$$r \sim N\left((I_{\mathsf{np}\times\mathsf{np}} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})e, P - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top\right). \tag{2.2.20}$$

Furthermore, $e = \Phi x_e \in \mathbb{R}^{\mathsf{np}}$ for some $x_e \in \mathbb{R}^{\mathsf{n}}$ if and only if the mean of r, $\mathbf{E}[r](= (I_{\mathsf{np}\times\mathsf{np}} - \Phi\Psi)e)$, satisfies $\mathbf{E}[r] = 0_{\mathsf{np}\times 1}$. In other words, e is undetectable with respect to Φ if and only if $\mathbf{E}[r] = 0_{\mathsf{np}\times 1}$. $\Diamond$

Proof. First, the mean of r is computed as follows.

$$\begin{aligned}\mathbf{E}[r] &= \mathbf{E}[(I_{\mathsf{np}\times\mathsf{np}} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})\hat{z}] \\ &= (I_{\mathsf{np}\times\mathsf{np}} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})\mathbf{E}[\Phi x + v + e] \\ &= (I_{\mathsf{np}\times\mathsf{np}} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})(\Phi x + e) \\ &= (I_{\mathsf{np}\times\mathsf{np}} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})e = (I_{\mathsf{np}\times\mathsf{np}} - \Phi\Psi)e\end{aligned}$$

Second, because it easily follows that

$$\begin{aligned}r - \mathbf{E}[r] &= (I_{\mathsf{np}\times\mathsf{np}} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})(\hat{z} - e) \\ &= (I_{\mathsf{np}\times\mathsf{np}} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})(\Phi x + v) \\ &= (I_{\mathsf{np}\times\mathsf{np}} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})v = (I_{\mathsf{np}\times\mathsf{np}} - \Phi\Psi)v,\end{aligned}$$

the covariance matrix is calculated as

$$\begin{aligned}&\mathbf{E}[(r - \mathbf{E}[r])(r - \mathbf{E}[r])^\top] = \mathbf{E}[(I_{\mathsf{np}\times\mathsf{np}} - \Phi\Psi)vv^\top(I_{\mathsf{np}\times\mathsf{np}} - \Phi\Psi)^\top] \\ &= (I_{\mathsf{np}\times\mathsf{np}} - \Phi\Psi)\mathbf{E}[vv^\top](I_{\mathsf{np}\times\mathsf{np}} - \Phi\Psi)^\top = (I_{\mathsf{np}\times\mathsf{np}} - \Phi\Psi)P(I_{\mathsf{np}\times\mathsf{np}} - \Phi\Psi)^\top \\ &= (I_{\mathsf{np}\times\mathsf{np}} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})P(I_{\mathsf{np}\times\mathsf{np}} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})^\top \\ &= P - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top = (I_{\mathsf{np}\times\mathsf{np}} - \Phi\Psi)P.\end{aligned}$$

Moreover, note that

$$\mathbf{E}[\hat{z}] = \mathbf{E}[\Phi x + v + e] = \Phi x + e$$

and

$$\mathbf{E}[r] = (I_{\mathsf{np}\times\mathsf{np}} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})\mathbf{E}[\hat{z}].$$

Since $\Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1}$ is a projection matrix and it projects $\mathbf{E}[\hat{z}]$ onto the range space of Φ, $\mathcal{R}(\Phi)$, we have $\mathbf{E}[\hat{z}] = \Phi x + e \notin \mathcal{R}(\Phi)$ if and only if $\mathbf{E}[\hat{z}] \neq \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1}\mathbf{E}[\hat{z}]$. This implies that $e \notin \mathcal{R}(\Phi)$ if and only if $(I_{\mathsf{np}\times\mathsf{np}} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})\mathbf{E}[\hat{z}] \neq 0_{\mathsf{np}\times 1}$. Note that $\Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1}$ is not generally an orthogonal projection matrix since it is not symmetric, while $\Phi\Phi^\dagger = \Phi(\Phi^\top\Phi)^{-1}\Phi^\top$ in Lemma 2.2.1 is an orthogonal projection matrix. This completes the proof. □

Theorem 2.2.18 is nothing but a counterpart of Lemma 2.2.1 when Gaussian noise is taken into account. It clarifies the mean and covariance of the Gaussian random variable r, and further, characterization of undetectable attacks with statistical analysis is given. As we have done in Lemma 2.2.4, one can derive a detection criterion of q-sparse errors, assuming that $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$ is (n-stacked) q-error detectable and e is actually (n-stacked) q-sparse. This detection strategy is summarized in the following theorem.

Theorem 2.2.19. For the measurement $\hat{z} = \Phi x + v + e$ where $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$ is (n-stacked) q-error detectable, $x \in \mathbb{R}^{\mathsf{n}}$, $e \in \Sigma_{\mathsf{q}}^{\mathsf{n}}$, and $v \in \mathbb{R}^{\mathsf{np}}$ satisfying $v \sim N(0_{\mathsf{np}\times 1}, P)$ with $P > 0$, let $r = (I_{\mathsf{np}\times\mathsf{np}} - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1})\hat{z}$. Then $e = 0_{\mathsf{np}\times 1}$ if and only if $\mathbf{E}[r] = 0_{\mathsf{np}\times 1}$. Moreover, when $e = 0_{\mathsf{np}\times 1}$, the vector x is recovered by the expectation value of $\hat{x} = (\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1}\hat{z}$, i.e., $x = \mathbf{E}[\hat{x}]$. $\Diamond$

Proof. Note that no q-sparse error $e \neq 0$ can be represented by Φx_e for some $x_e \in \mathbb{R}^{\mathsf{n}}$ by Proposition 2.2.2.(iii). Therefore, the result directly follows from Theorem 2.2.18. □

From the observation of Theorems 2.2.18 and 2.2.19, the problem of detecting a non-zero (n-stacked) q-sparse error signal e with an (n-stacked) q-error detectable coding matrix $\Phi \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$, can be rephrased as: Given the residual signal r which comes from the Gaussian distribution $N(\mathbf{E}[r], P - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top)$, determine if $\mathbf{E}[r] = 0_{\mathsf{np}\times 1}$ or $\mathbf{E}[r] \neq 0_{\mathsf{np}\times 1}$. Therefore, the statistical decision theory [33] is helpful in this situation. More precisely, the χ^2 test for fault detection [7, 48] or Wald test for two-sided vector parameter [33, Chapter 6] can be applied.

Among them, the χ^2 test is widely used to enhance the security of control systems such as [47, 50, 56, 57]. One can simply apply the χ^2 test to detect the presence of error signals in the n-stacked measurement $\hat{z}$ of (2.2.2) and its operating scheme is summarized in Algorithm 2.2. Initially, the attack detection alarm indicator f is set to 0, and then, the residual r is computed according to the equation (2.2.19). Without any error signal (i.e., $e = 0_{\mathsf{np}\times 1}$), the residual r follows a Gaussian distribution $N(0, P - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top)$ which is shown in (2.2.20). Now, define the standardized residual $\zeta := \big(P - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top)\big)^{-\frac{1}{2}} r$ whose distri-

Algorithm 2.2 Detection scheme based on χ^2 test for a static error equation

```
Input: $\hat{z}$
Output: $f$
Initialization: $f = 0$
$\hat{x}_{\mathrm{MVUE}} = (\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1}\hat{z}$
$r = \hat{z} - \Phi\hat{x}_{\mathrm{MVUE}}$
$\zeta = \left(P - \Phi(\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top)\right)^{-\frac{1}{2}} r$
$g = \zeta^\top\zeta$
if $g \leq \Delta_{TH}$ then
    $f = 0$
else if $g > \Delta_{TH}$ then
    $f = 1$
end if
```

bution becomes $N(0_{\mathsf{np}\times 1}, I_{\mathsf{np}\times\mathsf{np}})$. Thus, the 2-norm of ζ denoted by $g := \zeta^\top\zeta$ is an observation from a random variable $\mathbf{g}$ which satisfies a χ^2 distribution with np degrees of freedom, i.e.,

$$\mathbf{g} \sim \chi^2_{\mathsf{np}}.$$

This means that g can not be far away from 0. Finally, when g is greater than the threshold Δ_{TH}, the attack detection alarm is triggered by setting $f = 1$. Here, Δ_{TH} is the predetermined threshold value and it decides the probability of false alarm and the probability of error detection. For example, when the threshold Δ_{TH} is chosen such that

$$\int_0^{\Delta_{TH}} p_{\mathbf{g}}(x)dx = 1 - \delta,$$

where $p_{\mathbf{g}}(x)$ denotes the PDF of the χ^2_{np} distribution, the probability of false alarm becomes δ. As the probability of false alarm δ gets smaller, the probability of error detection also decreases, which implies that there is a trade-off between the small false alarm and the high error detection ratio. Thus, one needs to choose Δ_{TH} as a good compromise between these two conflicting requirements.

Now, a suboptimal state reconstruction algorithm is developed when the measurement $\hat{z}$ of (2.2.2) is corrupted by a Gaussian noise $v \sim N(0_{\mathsf{np}\times 1}, P)$ and an

(n-stacked) q-sparse error e where the coding matrix Φ is (n-stacked) q-error correctable. The algorithm is described in Algorithm 2.3 and it operates on the basis of the multiple hypothesis testing [33, Section 3.8] in the field of statistical decision theory. Once the set of attack-free sensors, $\Lambda^* \subset [\mathsf{p}]$, is identified, the input state x can be recovered from the measurement $\hat{z}_{\Lambda^*}$. To this end, the set of sensors that is most likely to be attack-free, is identified. Specifically, we will first search all subsets Λ's of sensors whose cardinal number is $\mathsf{p} - \mathsf{q}$. Then, we compute the MVUE (or WLSE) $\hat{x}^\Lambda$ of each subset only with the measurement data from the subset, assuming that the subset is attack-free. Recall that this is similar to the procedure in (2.2.5). In Algorithm 2.3, $P^\pi_{\Lambda^\mathsf{n},\Lambda^\mathsf{n}}$ denotes the matrix obtained from P by eliminating all i-th rows and all j-th columns such that $\mathsf{i} \in (\Lambda^\mathsf{n})^c$ and $\mathsf{j} \in (\Lambda^\mathsf{n})^c$. Then, for each subset, the detection scheme in Algorithm 2.2 is applied. That is, the residual r^Λ, the standardized residual ζ^Λ, and its 2-norm g^Λ are calculated for each subset Λ. Finally, the optimal subset Λ^* is decided by the maximum likelihood (ML) decision rule with the values of g^Λ's.

For a detailed explanation, let

$$\left\{\Lambda_1, \Lambda_2, \cdots, \Lambda_{\binom{\mathsf{p}}{\mathsf{q}}}\right\}$$

be the set $\{\Lambda \subset [\mathsf{p}] : |\Lambda| = \mathsf{p} - \mathsf{q}\}$. Now, we wish to distinguish between $\binom{\mathsf{p}}{\mathsf{q}}$ hypotheses, $\mathcal{H}_1, \mathcal{H}_2, \cdots, \mathcal{H}_{\binom{\mathsf{p}}{\mathsf{q}}}$, which are given as follows:

$$\mathcal{H}_\mathsf{i} : \text{ the set } \Lambda_\mathsf{i} \text{ is attack-free, i.e., } e^\mathsf{n}_\mathsf{j} = 0_{\mathsf{n}\times 1} \text{ for all } \mathsf{j} \in \Lambda_\mathsf{i} \ .$$

Let us denote $\mathbf{g}$ as a random variable such that

$$\mathbf{g} \sim \chi^2_{\mathsf{n}(\mathsf{p}-\mathsf{q})},$$

and $p_\mathbf{g}$ as the PDF of the $\chi^2_{\mathsf{n}(\mathsf{p}-\mathsf{q})}$ distribution. Moreover, $\mathbf{g}_\mathsf{i}$ represents a random variable such that g^{Λ_i} is a single observation from $\mathbf{g}_\mathsf{i}$. Note that, if the sensors indexed by Λ_i is attack-free, i.e., $e^\pi_{\Lambda^\mathsf{n}_\mathsf{i}} = 0_{\mathsf{n}(\mathsf{p}-\mathsf{q})\times 1}$, then the random variable $\mathbf{g}_\mathsf{i}$ follows $\chi^2_{\mathsf{n}(\mathsf{p}-\mathsf{q})}$ which can be derived from (2.2.20). The ML decision rule chooses the hypothesis $\mathcal{H}_{\mathsf{i}^*}$ that maximizes the likelihood $p_{\mathbf{g}_\mathsf{i}}\left(g^{\Lambda_\mathsf{i}}; \mathcal{H}_\mathsf{i}\right)$, which is the PDF

Algorithm 2.3 Correction scheme based on multiple hypothesis testing

```
Input: $\hat{z}$
Output: $\hat{x}$
Initialization: $\hat{x} = (\Phi^\top P^{-1}\Phi)^{-1}\Phi^\top P^{-1}\hat{z}$
for $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| = \mathsf{p} - \mathsf{q}$ do
    $\hat{x}^\Lambda = \left(\Phi^{\pi\top}_{\Lambda^\mathsf{n}}(P^\pi_{\Lambda^\mathsf{n},\Lambda^\mathsf{n}})^{-1}\Phi^\pi_{\Lambda^\mathsf{n}}\right)^{-1}\Phi^{\pi\top}_{\Lambda^\mathsf{n}}(P^\pi_{\Lambda^\mathsf{n},\Lambda^\mathsf{n}})^{-1}\hat{z}^\pi_{\Lambda^\mathsf{n}}$
    $r^\Lambda = \hat{z}^\pi_{\Lambda^\mathsf{n}} - \Phi^\pi_{\Lambda^\mathsf{n}}\hat{x}^\Lambda$
    $\zeta^\Lambda = \left(P^\pi_{\Lambda^\mathsf{n},\Lambda^\mathsf{n}} - \Phi^\pi_{\Lambda^\mathsf{n}}(\Phi^{\pi\top}_{\Lambda^\mathsf{n}}(P^\pi_{\Lambda^\mathsf{n},\Lambda^\mathsf{n}})^{-1}\Phi^\pi_{\Lambda^\mathsf{n}})^{-1}\Phi^{\pi\top}_{\Lambda^\mathsf{n}})\right)^{-\frac{1}{2}} r^\Lambda$
    $g^\Lambda = {\zeta^\Lambda}^\top \zeta^\Lambda$
end for
$\Lambda^* = \arg\max_{\substack{\Lambda\subset[\mathsf{p}]\\ |\Lambda|=\mathsf{p}-\mathsf{q}}} p_{\mathbf{g}}\left(g^\Lambda\right)$
$\hat{x} = \hat{x}^{\Lambda^*}$
```

of $\mathbf{g}_\mathsf{i}$ being equal to the observation g^{Λ_i} under the condition that there is no error signal in the measurements indexed by Λ_i. Therefore, we have

$$\mathsf{i}^* = \arg\max_{\mathsf{i}\in\left[\binom{\mathsf{p}}{\mathsf{q}}\right]} p_{\mathbf{g}_\mathsf{i}}\left(g^{\Lambda_\mathsf{i}}; \mathcal{H}_\mathsf{i}\right) = \arg\max_{\mathsf{i}\in\left[\binom{\mathsf{p}}{\mathsf{q}}\right]} p_{\mathbf{g}}\left(g^{\Lambda_\mathsf{i}}\right) \tag{2.2.21}$$

where the second equality comes from the fact that $\mathbf{g}_\mathsf{i} \sim \chi^2_{\mathsf{n}(\mathsf{p}-\mathsf{q})}$ under the hypothesis $\mathcal{H}_\mathsf{i}$. Therefore, from the index set Λ_{i^*} corresponding to the ML hypothesis $\mathcal{H}_{\mathsf{i}^*}$, a suboptimal estimate of x becomes the MVUE (or WLSE) $\hat{x}^{\Lambda_{\mathsf{i}^*}}$ of the set Λ_{i^*}.

Remark 2.2.3. Since the PDF of the $\chi^2_{\mathsf{n}(\mathsf{p}-\mathsf{q})}$ distribution, $p_{\mathbf{g}}$, is not monotonically decreasing for $\mathsf{n}(\mathsf{p}-\mathsf{q}) > 2$, the ML rule of (2.2.21) generally does not pick the index set which has the smallest g^Λ. This is different from the case of bounded noises studied in the previous section, where as small residual as possible is desirable. The underlying philosophy of the ML rule (2.2.21) is that it does not select an index set with abnormally small residuals. Although the set with small residual may give a small state estimation error, it is likely to be injected by an unwanted error signal e. Hence, adversaries who inject the error signal e can suddenly change their signals to be harmful to the system at any time. $\Diamond$

Chapter 3

On Redundant Observability

Motivated by [21], which says that analytical redundancy is an essential technique in fault detection and isolation, we introduce redundancy in measurements (i.e., *redundant observability*) and relate that concept to *dynamic security index*, *attack detectability*, and *observability under sensor attacks*. It will soon be revealed that an observability matrix behaves like a coding matrix examined in the previous chapter, and hence, its properties (e.g., redundant left invertibility, cospark, error detectability, and error correctability) determine the resilience of control systems under sensor attacks (e.g., redundant observability, dynamic security index, attack detectability, and observability under sensor attacks of the system), as summarized in Proposition 3.3.1 which appears at the end of this chapter. Furthermore, the *redundant detectability* (or, *asymptotic redundant observability*), which is a weaker notion than the redundant observability, is also introduced. While the redundant observability does not care about the magnitudes of sensor attacks and does not mind whether the attacks are disruptive or not, the redundant detectability only deals with attacks that do not converge to zero as time goes on, so that it is more practical in the sense that it can only detect and correct the attacks that are actually harmful to the system.

3.1 Redundant Observability

3.1.1 Definition and Characterization

Consider a discrete-time LTI system

$$\overline{\mathcal{P}} : \begin{cases} x(k+1) = Ax(k) & (3.1.1a) \\ \bar{y}(k) = y(k) + a(k) = Cx(k) + a(k) & (3.1.1b) \end{cases}$$

where $x \in \mathbb{R}^{\mathsf{n}}$ is the state variables, $y \in \mathbb{R}^{\mathsf{p}}$ is the original sensor outputs, and $\bar{y} \in \mathbb{R}^{\mathsf{p}}$ is the measurement data under additional input signal $a \in \mathbb{R}^{\mathsf{p}}$. The input signal a does not have to be an attack signal, but we implicitly suppose that a is generated by an adversary and injected into the system for the purpose of disrupting the system. There are total p sensors which measure the system outputs and the i-th measurement data at time k is denoted by $\bar{y}_{\mathsf{i}}(k) = c_{\mathsf{i}}x(k) + a_{\mathsf{i}}(k)$ where c_{i} is the i-th row of C. From a control theoretical viewpoint, the notion of *redundant observability* is defined formally as follows.

Definition 3.1.1. The dynamical system (3.1.1) or the pair (A, C) is said to be q-*redundant observable*[1] if the pair (A, C_{Λ}^{π}) is observable for any $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p}-\mathsf{q}$. That is, the system (3.1.1) is q-redundant observable if it is observable after removing any q sensor outputs. $\Diamond$

One of the most popular and well-known method to determine the observability of a given LTI system is to check the rank condition of the observability matrix, i.e., the system is observable if and only if its observability matrix has full column rank. Thus, we will generalize it to the redundant observability concept. In order to derive necessary and sufficient conditions for the redundant observability, we first rearrange the observability matrix $G'^{(k)} \in \mathbb{R}^{k\mathsf{p}\times\mathsf{n}}$ of the pair (A, C) which is

[1] The same concept was introduced in [77] with q-sparse observability notion, but we have named this q-redundant observability because q-sparse observability was formerly defined in [87] which concerns q-sparse initial values.

defined by

$$G'^{(k)} := \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{k-1} \end{bmatrix}. \tag{3.1.2}$$

By simply exchanging its rows, we have a new observability matrix $G^{(k)} \in \mathbb{R}^{k\mathsf{p}\times\mathsf{n}}$ as follows:

$$G^{(k)} := \begin{bmatrix} G_1^{(k)} \\ G_2^{(k)} \\ \vdots \\ G_{\mathsf{p}}^{(k)} \end{bmatrix} \tag{3.1.3}$$

where $G_{\mathsf{i}}^{(k)}$ is an observability matrix of the pair (A, c_{i}) given by

$$G_{\mathsf{i}}^{(k)} := \begin{bmatrix} c_{\mathsf{i}} \\ c_{\mathsf{i}}A \\ \vdots \\ c_{\mathsf{i}}A^{k-1} \end{bmatrix}. \tag{3.1.4}$$

When $k = \mathsf{n}$ (the dimension of state variable x) in the equations above, we conventionally drop the superscript (k) from the observability matrix, that is,

$$G' = G'^{(\mathsf{n})}, \quad G = G^{(\mathsf{n})}, \quad \text{and} \quad G_{\mathsf{i}} = G_{\mathsf{i}}^{(\mathsf{n})}.$$

With this new observability matrix G at hand, we can easily verify the following equivalence between redundant observability of the pair (A, C) and error detectability of its observability matrix G.

Proposition 3.1.1. The followings are equivalent:

(i) The pair (A, C) is q-redundant observable;

(ii) The observability matrix G is (n-stacked) q-error detectable;

(iii) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$, $G_{\Lambda^{\mathsf{n}}}$ (or, equivalently $G^{\pi}_{\Lambda^{\mathsf{n}}}$) has full column rank;

(iv) For any $x \in \mathbb{R}^{\mathsf{n}}$ where $x \neq 0_{\mathsf{n}\times 1}$, $\|Gx\|_{0^{\mathsf{n}}} > \mathsf{q}$;

(v) For any $x, x' \in \mathbb{R}^{\mathsf{n}}$ where $x \neq x'$, $\mathsf{d}_{0^{\mathsf{n}}}(Gx, Gx') > \mathsf{q}$. ◊

Proof. (i) ⇔ (iii): This can be verified from the fact that $G^{\pi}_{\Lambda^{\mathsf{n}}}$ is the observability matrix (after some row exchange operations) of the pair (A, C^{π}_{Λ}). Let any index set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$ be given. The observability matrix G'_{Λ} of the pair (A, C^{π}_{Λ}), which is defined by

$$G'_{\Lambda} := \begin{bmatrix} C^{\pi}_{\Lambda} \\ C^{\pi}_{\Lambda} A \\ \vdots \\ C^{\pi}_{\Lambda} A^{\mathsf{n}-1} \end{bmatrix},$$

is computed after several row exchange operations on $G^{\pi}_{\Lambda^{\mathsf{n}}}$. Since the row exchange can not alter the rank of a matrix, it follows that $\mathsf{rank}(G'_{\Lambda}) = \mathsf{n}$ if and only if $\mathsf{rank}(G^{\pi}_{\Lambda^{\mathsf{n}}}) = \mathsf{n}$, and the proof is completed.
(ii) ⇔ (iii) ⇔ (iv) ⇔ (v): This is proved in Proposition 2.2.2. □

Hence, it follows from Proposition 3.1.1 that the pair (A, C) is q-redundant observable if and only if its observability matrix G is q-redundant left invertible (or, equivalently $\mathsf{cospark}^{\mathsf{n}}(G) > \mathsf{q}$). While Proposition 3.1.1 establishes powerful criteria on the redundant observability which tests the observability matrix G, we can also give another equivalent condition for the redundant obsevability based on the Popov-Belevitch-Hautus (PBH) observability test. Recall that the pair (A, C) is observable if and only if

$$\mathsf{rank}\left(\begin{bmatrix} \lambda I_{\mathsf{n}\times\mathsf{n}} - A \\ C \end{bmatrix}\right) = \mathsf{n}$$

for any eigenvalue λ of A. In other words, the pair (A, C) is not observable if and only if there exists a non-zero eigenvector v such that

$$\begin{bmatrix} \lambda I_{\mathsf{n}\times\mathsf{n}} - A \\ C \end{bmatrix} v = 0_{(\mathsf{n}+\mathsf{p})\times 1}.$$

The following proposition suggests another necessary and sufficient condition for the redundant observability using the PBH test.

Proposition 3.1.2. The followings are equivalent:
(i) The pair (A, C) is q-redundant observable;
(ii) For any $v \in \mathcal{V}(A)$, $\|Cv\|_0 > \mathsf{q}$. ♢

Proof. This can be proved by the contraposition of the statement. The pair (A, C) is not q-redundant observable if and only if there exists an index set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$ such that the pair (A, C_Λ^π) is not observable. By the PBH test, it is equivalent to the condition that there exists a non-zero eigenvector $v \in \mathcal{V}(A)$ such that $Av = \lambda v$ and $C_\Lambda^\pi v = 0$ for an index set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$. Since $v \in \mathcal{V}(A)$ is an eigenvector of A, we can delete the first condition $Av = \lambda v$. Finally, it follows that there exists an index set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$ such that $C_\Lambda^\pi v = 0$ for some $v \in \mathcal{V}(A)$, which means $\|Cv\|_0 \leq \mathsf{q}$ for some $v \in \mathcal{V}(A)$. □

3.1.2 Relationship with Strong Observability

In this section, we regard the additional input signal a of the system $\overline{\mathcal{P}}$ in (3.1.1) as an unknown input, and give a relationship between the q-redundant observability and the *strong observability.* For this end, let us first consider an LTI system of

$$\overline{\mathcal{P}}_{(A,O,C,D)} : \begin{cases} x(k+1) = Ax(k) & (3.1.5a) \\ \bar{y}(k) = y(k) + Da(k) = Cx(k) + Da(k), & (3.1.5b) \end{cases}$$

which is the same as (3.1.1) except the direct feedthrough matrix D. Note that the system $\overline{\mathcal{P}}_{(A,O,C,D)}$ is said to be *strongly observable* if, for all initial condition $x(0) \in \mathbb{R}^{\mathsf{n}}$ and for every input function $a(\cdot)$, $y(k) \equiv 0$ implies $x(0) = 0$ [5], [92, Chapter 7]. Accordingly, the *weakly unobservable subspace* of the system $\overline{\mathcal{P}}_{(A,O,C,D)}$, denoted as $\mathcal{W}\left(\overline{\mathcal{P}}_{(A,O,C,D)}\right)$, is defined as the set of all initial condition $x(0)$ such that there exists an input function $a(\cdot)$ which makes $y(k) \equiv 0$. Hence, it is trivial by the definitions that the system $\overline{\mathcal{P}}_{(A,O,C,D)}$ is strongly observable if and only if the weakly unobservable subspace of the system $\overline{\mathcal{P}}_{(A,O,C,D)}$ is trivial, i.e., $\mathcal{W}\left(\overline{\mathcal{P}}_{(A,O,C,D)}\right) = \{0\}$.

Now, it is supposed for the system (3.1.1) that the additional input signal can not be injected into all the sensors, but a part of them. We suppose that up to q out of p measurement outputs can be compromised by the signal a. Therefore, a formal condition on the sparsity of the input vector a can be given as follows.

Assumption 3.1.1. There exist at least $\mathsf{p}-\mathsf{q}$ sensors which are not attacked for all $k \geq 0$, i.e.,

$$\left|\left\{\mathsf{i} \in [\mathsf{p}] : a_{\mathsf{i}}(k) = 0,\ {}^{\forall}k \geq 0\right\}\right| \geq \mathsf{p}-\mathsf{q}. \qquad \diamondsuit$$

Characterization of the observability of the system (3.1.1) with an unknown signal a under Assumption 3.1.1, is the main subject of this section. In fact, instead of imposing q-sparsity assumption on a (i.e., Assumption 3.1.1) in the system (3.1.1), we can replace D with I_{Λ^c} in (3.1.5b) without any assumption on $a(\cdot)$ where $I \in \mathbb{R}^{\mathsf{p}\times\mathsf{p}}$ is an identity matrix and $\Lambda \subset [\mathsf{p}]$ is any index set satisfying $|\Lambda| \geq \mathsf{p}-\mathsf{q}$ (or, equivalently $|\Lambda^c| \leq \mathsf{q}$). In short, we consider the following LTI system

$$\overline{\mathcal{P}}_\Lambda : \begin{cases} x(k+1) = Ax(k) & (3.1.6a) \\ \bar{y}(k) = y(k) + I_{\Lambda^c}a(k) = Cx(k) + I_{\Lambda^c}a(k) & (3.1.6b) \end{cases}$$

where I_{Λ^c} is any q-sparse identity matrix with unknown Λ. With this system in mind, the relationship between the redundant observability and the strong observability is derived as follows.

Proposition 3.1.3. The followings are equivalent:
(i) The pair (A, C) is q-redundant observable;
(ii) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p}-\mathsf{q}$, the dynamical system $\overline{\mathcal{P}}_\Lambda$ is strongly observable;
(iii) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p}-\mathsf{q}$, the weakly unobservable subspace of the system $\overline{\mathcal{P}}_\Lambda$ is trivial, i.e., $\mathcal{W}\left(\overline{\mathcal{P}}_\Lambda\right) = \{0\}$;
(iv) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p}-\mathsf{q}$ and for all $F \in \mathbb{R}^{\mathsf{p}\times\mathsf{n}}$, the pair $(A, C + I_{\Lambda^c}F)$ is observable. $\diamondsuit$

Proof. (i) ⇒ (iv): First, pick any $F \in \mathbb{R}^{\mathsf{p}\times\mathsf{n}}$ and any $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p}-\mathsf{q}$. We claim that

$$\mathsf{rank}\left(\begin{bmatrix} \lambda I_{\mathsf{n}\times\mathsf{n}} - A \\ C + I_{\Lambda^c}F \end{bmatrix}\right) = \mathsf{n}$$

for any eigenvalue λ of A. It is enough to show that $(C + I_{\Lambda^c}F)v \neq 0$ for any eigenvector v of A by the PBH test. Since the pair (A, C_Λ) is observable, $C_\Lambda v \neq 0$ for any eigenvector v of A. By simple calculations on matrix, we easily have that

$$(C + I_{\Lambda^c}F)v = (C_\Lambda + C_{\Lambda^c} + F_{\Lambda^c})v = C_\Lambda v + (C + F)_{\Lambda^c} v \neq 0.$$

Finally, it follows from the PBH test that the pair $(A, C + I_{\Lambda^c}F)$ is observable.
(iv) ⇒ (i): Pick any $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$, and we claim that (A, C_Λ) is observable. Let $F = -C_{\Lambda^c}$, we simply have

$$C + I_{\Lambda^c}F = C_\Lambda + C_{\Lambda^c} + F_{\Lambda^c} = C_\Lambda + C_{\Lambda^c} - C_{\Lambda^c} = C_\Lambda.$$

Thus, the claim is satisfied and the proof is completed from Definition 3.1.1.
(ii) ⇔ (iii) ⇔ (iv): This is proved in [92, Theorem 7.16]. □

Proposition 3.1.3 presents a clear relationship between the q-redundant observability and the strong observability. By characterizing the q-redundant observability in terms of the strong observability, we could relate the redundant observability concept to the dynamical system (3.1.1) under the unknown (attack) signal a satisfying Assumption 3.1.1. In other words, Definition 3.1.1 itself has nothing to do with the input signal a, but it turns out to be equivalent to the observability with unknown input a by Proposition 3.1.3. Later on, this forms the foundation of Section 3.3.

3.1.3 Redundant Unobservable Subspace

For an LTI system given by (3.1.1), the *unobservable subspace* is defined by the set of initial conditions that produce identically zero output $y(k) \equiv 0$. (Note that it is the original output $y(k)$, not $\bar{y}(k)$.) The unobservable subspace of the pair (A, C), denoted as $\overline{\mathcal{O}}(A, C)$, is equivalent to the null space of the observability matrix G' in (3.1.2), i.e.,

$$\overline{\mathcal{O}}(A, C) = \mathcal{N}(G').$$

This concept of unobservable subspace can be applied to the q-redundant observability and *redundant unobservable subspace* is defined as follows.

Definition 3.1.2. The subspace spanned by the set of elements that belong to any unobservable subspace of the pair (A, C_Λ^π) for $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$, is called the q-*redundant unobservable subspace* of the dynamical system (3.1.1) or the pair (A, C). $\Diamond$

We denote the q-redundant unobservable subspace of the pair (A, C) as $\overline{\mathcal{O}}_\mathsf{q}(A, C)$, and it can be directly computed from Definition 3.1.2 as the sum of some unobservable subspaces.

Proposition 3.1.4. The q-redundant unobservable subspace of the pair (A, C), $\overline{\mathcal{O}}_\mathsf{q}(A, C)$, can be computed by

$$\overline{\mathcal{O}}_\mathsf{q}(A, C) = \sum_{\substack{\Lambda \subset [\mathsf{p}] \\ |\Lambda| = \mathsf{p} - \mathsf{q}}} \overline{\mathcal{O}}(A, C_\Lambda^\pi) = \sum_{\substack{\Lambda \subset [\mathsf{p}] \\ |\Lambda| = \mathsf{p} - \mathsf{q}}} \mathcal{N}(G_{\Lambda^\mathsf{n}}^\pi). \qquad \Diamond$$

Proof. It should be noted that the sum is conducted over $|\Lambda| = \mathsf{p} - \mathsf{q}$ instead of $|\Lambda| \geq \mathsf{p} - \mathsf{q}$. This is possible because, for $\Lambda_1 \subset \Lambda_2$, it follows that $\mathcal{N}(G_{\Lambda_1^\mathsf{n}}^\pi) \supset \mathcal{N}(G_{\Lambda_2^\mathsf{n}}^\pi)$. □

Since each unobservable subspace $\mathcal{N}(G_{\Lambda^\mathsf{n}}^\pi)$ is A-invariant, i.e., $x \in \mathcal{N}(G_{\Lambda^\mathsf{n}}^\pi)$ implies $Ax \in \mathcal{N}(G_{\Lambda^\mathsf{n}}^\pi)$, the sum of those subspaces is also A-invariant. Hence, the q-redundant unobservable subspace $\overline{\mathcal{O}}_\mathsf{q}(A, C)$ is invariant under A.

Based on Proposition 3.1.4, the q-redundant observability can be characterized in terms of the q-redundant unobservable subspace as follows.

Corollary 3.1.5. The followings are equivalent:
(i) The pair (A, C) is q-redundant observable;
(ii) The q-redundant unobservable subspace of the pair (A, C) is trivial, that is, $\overline{\mathcal{O}}_\mathsf{q}(A, C) = \{0\}$. $\Diamond$

Proof. From the following equivalence,

$$\begin{aligned} \overline{\mathcal{O}}_\mathsf{q}(A, C) = \{0\} \quad &\Leftrightarrow \quad \sum_{\substack{\Lambda \subset [\mathsf{p}] \\ |\Lambda| = \mathsf{p} - \mathsf{q}}} \mathcal{N}(G_{\Lambda^\mathsf{n}}^\pi) = \{0\} \\ &\Leftrightarrow \quad \mathcal{N}(G_{\Lambda^\mathsf{n}}^\pi) = \{0\}, \quad {}^\forall \Lambda \subset [\mathsf{p}] \ \text{ s.t. } \ |\Lambda| = \mathsf{p} - \mathsf{q}, \end{aligned}$$

the proof is completed. □

Furthermore, the quotient space of the unobservable subspace, $\mathbb{R}^{\mathsf{n}}/\overline{\mathcal{O}}(A,C)$, is sometimes called, with abuse of terminology, the observable subspace. Since the quotient space $\mathbb{R}^{\mathsf{n}}/\overline{\mathcal{O}}(A,C)$ is isomorphic to the orthogonal complement $\overline{\mathcal{O}}(A,C)^{\perp}$ of $\overline{\mathcal{O}}(A,C)$, the observable subspace of the pair (A,C) denoted as $\mathcal{O}(A,C)$, is equivalent to the range space of the matrix $G'^{\top}$ where G' is an observability matrix given in (3.1.2), i.e., we have

$$\mathcal{O}(A,C) = \mathcal{N}(G')^{\perp} = \mathcal{R}(G'^{\top}).$$

By defining q-*redundant observable subspace* as the quotient space of the q–redundant unobservable subspace, which is isomorphic to the orthogonal complement of the q-redundant unobservable subspace, the following result on how to calculate the redundant observable subspace is directly obtained from Proposition 3.1.4.

Proposition 3.1.6. The q-redundant observable subspace of the pair (A,C), $\mathcal{O}_{\mathsf{q}}(A,C)$, can be computed by

$$\mathcal{O}_{\mathsf{q}}(A,C) = \left(\sum_{\substack{\Lambda\subset[\mathsf{p}] \\ |\Lambda|=\mathsf{p}-\mathsf{q}}} \mathcal{N}(G^{\pi}_{\Lambda^{\mathsf{n}}}) \right)^{\perp} = \bigcap_{\substack{\Lambda\subset[\mathsf{p}] \\ |\Lambda|=\mathsf{p}-\mathsf{q}}} \mathcal{N}(G^{\pi}_{\Lambda^{\mathsf{n}}})^{\perp} = \bigcap_{\substack{\Lambda\subset[\mathsf{p}] \\ |\Lambda|=\mathsf{p}-\mathsf{q}}} \mathcal{R}({G^{\pi}_{\Lambda^{\mathsf{n}}}}^{\top}). \quad \diamond$$

3.1.4 Asymptotic Redundant Observability (Redundant Detectability)

In classical control theory, *detectability*[2] (or, *asymptotic observability*) is a slightly weaker notion than observability. The LTI system or the pair (A,C) is said to be *detectable* if its unobservable subspace $\overline{\mathcal{O}}(A,C)$ is contained in its stable subspace $\mathcal{X}_s(A)$[3]. Therefore, we can naturally generalize the concept of redundant

[2] Please do not be confused by the notation of detectability with the error detectability or the attack detectability. The detectability is sometimes called the *asymptotic observability* [80].

[3] For an LTI system, the stable subspace $\mathcal{X}_s(A)$ is defined as the subspace spanned by the eigenvectors and generalized eigenvectors corresponding to the stable eigenvalues of A (e.g., with negative real parts for continuous-time systems and located in the open unit disk for discrete-

observability to the redundant detectability as follows.

Definition 3.1.3. The dynamical system (3.1.1) or the pair (A, C) is said to be q-*redundant detectable* (or, *asymptotically* q-*redundant observable*) if the pair (A, C_Λ^π) is detectable for any $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$. That is, the system (3.1.1) is q-redundant detectable if it is detectable after removing any q sensor outputs. $\Diamond$

Proposition 3.1.7. The dynamical system (3.1.1) or the pair (A, C) is q-redundant detectable if and only if its q-redundant unobservable subspace $\overline{\mathcal{O}}_\mathsf{q}(A, C)$ is contained in its stable subspace $\mathcal{X}_s(A)$. $\Diamond$

Proof. From the following equivalence,

$$\begin{aligned} \overline{\mathcal{O}}_\mathsf{q}(A, C) \subset \mathcal{X}_s(A) \quad &\Leftrightarrow \quad \sum_{\substack{\Lambda \subset [\mathsf{p}] \\ |\Lambda| = \mathsf{p} - \mathsf{q}}} \mathcal{N}(G_{\Lambda^\mathsf{n}}^\pi) \subset \mathcal{X}_s(A) \\ &\Leftrightarrow \quad \mathcal{N}(G_{\Lambda^\mathsf{n}}^\pi) \subset \mathcal{X}_s(A), \quad {}^\forall \Lambda \subset [\mathsf{p}] \;\text{ s.t. }\; |\Lambda| = \mathsf{p} - \mathsf{q}, \end{aligned}$$

the proof is completed. □

With the observability matrix $G^{(k)}$ in (3.1.3), we can derive some equivalent conditions for the redundant detectability in the following proposition, which is a counterpart of Proposition 3.1.1.

Proposition 3.1.8. The followings are equivalent:
(i) The pair (A, C) is q-redundant detectable (or, asymptotically q-redundant observable);
(ii) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$, $\mathcal{N}\left(G_{\Lambda^\mathsf{n}}\right) \subset \mathcal{X}_s(A)$ (or, equivalently $\mathcal{N}\left(G_{\Lambda^\mathsf{n}}^\pi\right) \subset \mathcal{X}_s(A)$);
(iii) For any $x \notin \mathcal{X}_s(A)$, $\|Gx\|_{0^\mathsf{n}} > \mathsf{q}$. $\Diamond$

Proof. (i) $\Leftrightarrow$ (ii): This easily follows from the facts that G_{Λ^n} is the observability matrix (after some row exchange operations) of the pair (A, C_Λ) and the unobservable subspace of the pair (A, C_Λ), $\overline{\mathcal{O}}(A, C_\Lambda)$, is the same as the null space of

time systems). Similarly, the unstable subspace $\mathcal{X}_u(A)$ is defined as the subspace spanned by the eigenvectors and generalized eigenvectors corresponding to the unstable eigenvalues of A (e.g., with non-negative real parts for continuous-time systems and located outside of the open unit disk for discrete-time systems).

the observability matrix $G_{\Lambda^{\mathsf{n}}}$, $\mathcal{N}(G_{\Lambda^{\mathsf{n}}})$.
(ii) ⇒ (iii): Suppose, for the sake of contradiction, that there exists $x \notin \mathcal{X}_s(A)$ satisfying $\|Gx\|_{0^{\mathsf{n}}} \leq \mathsf{q}$. Let Λ be the complement of $\mathsf{supp}^{\mathsf{n}}(Gx)$, i.e., $\Lambda = (\mathsf{supp}^{\mathsf{n}}(Gx))^c$. Then it is obvious that $|\Lambda| \geq \mathsf{p}-\mathsf{q}$ and $G_{\Lambda^{\mathsf{n}}}x = 0_{\mathsf{np}\times 1}$. Thus, there exists $x \in \mathbb{R}^{\mathsf{n}}$ such that $x \in \mathcal{N}(G_{\Lambda^{\mathsf{n}}})$ and $x \notin \mathcal{X}_s(A)$ so that $\mathcal{N}(G_{\Lambda^{\mathsf{n}}}) \not\subset \mathcal{X}_s(A)$. This contradicts the condition (ii).
(iii) ⇒ (ii): We again prove it by contradiction. Suppose that (ii) does not hold, i.e., there exists an index set $\Lambda \subset [\mathsf{p}]$ with $|\Lambda| \geq \mathsf{p}-\mathsf{q}$ and $x \in \mathbb{R}^{\mathsf{n}}$ satisfying $x \in \mathcal{N}(G_{\Lambda^{\mathsf{n}}})$ and $x \notin \mathcal{X}_s(A)$. Then it follows from $G_{\Lambda^{\mathsf{n}}}x = 0_{\mathsf{np}\times 1}$ and $|\Lambda| \geq \mathsf{p}-\mathsf{q}$ that $\|Gx\|_{0^{\mathsf{n}}} \leq \mathsf{q}$. Thus, (iii) fails because there exists $x \notin \mathcal{X}_s(A)$ such that $\|Gx\|_{0^{\mathsf{n}}} \leq \mathsf{q}$. □

Similar to the procedure in Proposition 3.1.2, the PBH detectability test produces the following result of characterizing the redundant detectability. Note that, by the PBH detectability test, the pair (A, C) is detectable if and only if

$$\mathsf{rank}\left(\begin{bmatrix} \lambda I_{\mathsf{n}\times\mathsf{n}} - A \\ C \end{bmatrix}\right) = \mathsf{n}$$

for any unstable eigenvalue λ of A. In other words, the pair (A, C) is not detectable if and only if there exists a non-zero eigenvector v of A corresponding to the unstable eigenvalue λ such that

$$\begin{bmatrix} \lambda I_{\mathsf{n}\times\mathsf{n}} - A \\ C \end{bmatrix} v = 0_{(\mathsf{n}+\mathsf{p})\times 1}.$$

Proposition 3.1.9. The followings are equivalent:
(i) The pair (A, C) is q-redundant detectable (or, asymptotically q-redundant observable);
(ii) For any $v \in \mathcal{V}_u(A)$, $\|Cv\|_0 > \mathsf{q}$. ♢

Proof. This can be proved by the contraposition of the statement. The pair (A, C) is not q-redundant detectable if and only if there exists an index set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p}-\mathsf{q}$ such that the pair (A, C_Λ^π) is not detectable. By the PBH test, it is equivalent to the condition that there exists a non-zero eigenvector

$v \in \mathcal{V}_u(A)$ corresponding to the unstable eigenvalue λ such that $Av = \lambda v$ and $C_\Lambda^\pi v = 0$ for an index set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$. Since $v \in \mathcal{V}_u(A)$ is an eigenvector of A, we can delete the first condition $Av = \lambda v$. Finally, it follows that there exists an index set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$ such that $C_\Lambda^\pi v = 0$ for some $v \in \mathcal{V}_u(A)$, which means $\|Cv\|_0 \leq \mathsf{q}$ for some $v \in \mathcal{V}_u(A)$. □

As done in Section 3.1.2, we can also give some equivalent conditions of the redundant detectability in terms of the *strong detectability*. Note that the system $\overline{\mathcal{P}}_{(A,O,C,D)}$ in (3.1.5) is said to be *strongly detectable* if, for all initial condition $x(0) \in \mathbb{R}^{\mathsf{n}}$ and for every input function $a(\cdot)$, $y(k) \equiv 0$ implies $\lim_{k\to\infty} x(k) = 0$ [92, Chapter 7]. Accordingly, the *controllable weakly unobservable subspace* of the system $\overline{\mathcal{P}}_{(A,O,C,D)}$, denoted as $\mathcal{C}\left(\overline{\mathcal{P}}_{(A,O,C,D)}\right)$, is defined as the set of all initial condition $x(0)$ such that there exists an input function $a(\cdot)$ which makes $y(k) \equiv 0$ and $x(\mathsf{k}) = 0$ for some finite time k. It is shown in [92, Exercise 7.9] that the system $\overline{\mathcal{P}}_{(A,O,C,D)}$ is strongly detectable if and only if the controllable weakly unobservable subspace of the system $\overline{\mathcal{P}}_{(A,O,C,D)}$ is trivial, i.e., $\mathcal{C}\left(\overline{\mathcal{P}}_{(A,O,C,D)}\right) = \{0\}$, and the system $\overline{\mathcal{P}}_{(A,O,C,D)}$ is of minimum phase[4]. Now, the relationship between the redundant detectability and the strong detectability is derived as follows.

[4] The minimum phaseness of multi-input-multi-output LTI system can be defined as follows [92, Chapter 7]: Consider the system $\mathcal{P}_{(A,B,C,D)}$ given by

$$\mathcal{P}_{(A,B,C,D)} : \begin{cases} x(k+1) = Ax(k) + Bu(k) \\ y(k) = Cx(k) + Du(k). \end{cases}$$

The *system matrix* $S_{\mathcal{P}_{(A,B,C,D)}}(s)$ of the system $\mathcal{P}_{(A,B,C,D)}$ is defined by

$$S_{\mathcal{P}_{(A,B,C,D)}}(s) := \begin{bmatrix} sI - A & -B \\ C & D \end{bmatrix}.$$

The *invariant factors* of the system matrix $S_{\mathcal{P}_{(A,B,C,D)}}(s)$, the diagonal entries of the *Simith normal form* of $S_{\mathcal{P}_{(A,B,C,D)}}(s)$, are called the *transmission polynomials* of $\mathcal{P}_{(A,B,C,D)}$. A transmission polynomial is called non-trivial if it is not zero. The product of the non-trivial transmission polynomials of $\mathcal{P}_{(A,B,C,D)}$ is called the *zero polynomial* of the system. Any complex root of the zero polynomial is called a *zero* of the system $\mathcal{P}_{(A,B,C,D)}$. The system $\mathcal{P}_{(A,B,C,D)}$ is called a *minimum phase system* if all zeros of the system are contained in the stable region (e.g., the open unit disk for discrete-time systems).

Proposition 3.1.10. The followings are equivalent:
(i) The pair (A, C) is q-redundant detectable (or, asymptotically q-redundant observable);
(ii) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$, the dynamical system $\overline{\mathcal{P}}_\Lambda$ is strongly detectable;
(iii) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$, the controllable weakly unobservable subspace of the system $\overline{\mathcal{P}}_\Lambda$ is trivial, i.e., $\mathcal{C}\left(\overline{\mathcal{P}}_\Lambda\right) = \{0\}$, and the system $\overline{\mathcal{P}}_\Lambda$ is of minimum phase;
(iv) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$ and for all $F \in \mathbb{R}^{\mathsf{p}\times\mathsf{n}}$, the pair $(A, C + I_{\Lambda^c}F)$ is detectable. ♢

Proof. (i) ⇒ (iv): First, pick any $F \in \mathbb{R}^{\mathsf{p}\times\mathsf{n}}$ and any $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p}-\mathsf{q}$. Now, we claim that

$$\mathsf{rank}\left(\begin{bmatrix}\lambda I_{\mathsf{n}\times\mathsf{n}} - A \\ C + I_{\Lambda^c}F\end{bmatrix}\right) = \mathsf{n}$$

for any unstable eigenvalue λ of A. It is enough to show that $(C + I_{\Lambda^c}F)v \neq 0$ for any eigenvector v of A corresponding to the unstable eigenvalue by the PBH test. Since the pair (A, C_Λ) is detectable, $C_\Lambda v \neq 0$ for any eigenvector v of A corresponding to the unstable eigenvalue. By simple calculations on matrix, we easily have that

$$(C + I_{\Lambda^c}F)v = (C_\Lambda + C_{\Lambda^c} + F_{\Lambda^c})v = C_\Lambda v + (C + F)_{\Lambda^c}v \neq 0.$$

Finally, it follows from the PBH test that the pair $(A, C + I_{\Lambda^c}F)$ is detectable.
(iv) ⇒ (i): Pick any $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$, and we claim that (A, C_Λ) is detectable. Let $F = -C_{\Lambda^c}$, it is obvious that

$$C + I_{\Lambda^c}F = C_\Lambda + C_{\Lambda^c} + F_{\Lambda^c} = C_\Lambda + C_{\Lambda^c} - C_{\Lambda^c} = C_\Lambda.$$

Thus, the claim is satisfied and the proof is completed from Definition 3.1.3.
(ii) ⇔ (iii) ⇔ (iv): This is shown in [92, Exercise 7.9]. □

Now, the undetectable subspace of an LTI system is discussed and the concept is extended to the *redundant undetectable subspace*. The discussion below follows the contents of Section 3.1.3 as well. The undetectable subspace of the pair (A, C), denoted by $\overline{\mathcal{D}}(A, C)$, is obtained by the intersection of the unobservable subspace $\overline{\mathcal{O}}(A, C)$ and the unstable subspace $\mathcal{X}_u(A)$, i.e.,

$$\overline{\mathcal{D}}(A, C) := \overline{\mathcal{O}}(A, C) \cap \mathcal{X}_u(A) = \mathcal{N}(G') \cap \mathcal{X}_u(A)$$

where G' is an observability matrix given in (3.1.2). This notion of undetectable subspace can be applied to the q-redundant detectability and *redundant undetectable subspace* is defined as follows.

Definition 3.1.4. The subspace spanned by the set of elements that belong to any undetectable subspace of the pair (A, C^{π}_{Λ}) for $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$, is called the q-*redundant undetectable subspace* of the dynamical system (3.1.1) or the pair (A, C). ♢

The q-redundant undetectable subspace of the pair (A, C) is denoted by $\overline{\mathcal{D}}_{\mathsf{q}}(A, C)$, and it easily follows from Definition 3.1.4 that we can compute it as the sum of some undetectable subspaces.

Proposition 3.1.11. The q-redundant undetectable subspace of the pair (A, C), $\overline{\mathcal{D}}_{\mathsf{q}}(A, C)$, can be computed by

$$\overline{\mathcal{D}}_{\mathsf{q}}(A, C) = \overline{\mathcal{O}}_{\mathsf{q}}(A, C) \cap \mathcal{X}_u(A) = \left(\sum_{\substack{\Lambda \subset [\mathsf{p}] \\ |\Lambda| = \mathsf{p} - \mathsf{q}}} \mathcal{N}(G^{\pi}_{\Lambda^{\mathsf{n}}}) \right) \cap \mathcal{X}_u(A). \qquad ♢$$

Proof. The result easily follows from

$$\begin{aligned} \overline{\mathcal{D}}_{\mathsf{q}}(A, C) &= \sum_{\substack{\Lambda \subset [\mathsf{p}] \\ |\Lambda| = \mathsf{p} - \mathsf{q}}} \overline{\mathcal{D}}(A, C^{\pi}_{\Lambda}) = \sum_{\substack{\Lambda \subset [\mathsf{p}] \\ |\Lambda| = \mathsf{p} - \mathsf{q}}} \Big(\mathcal{N}(G^{\pi}_{\Lambda^{\mathsf{n}}}) \cap \mathcal{X}_u(A) \Big) \\ &= \left(\sum_{\substack{\Lambda \subset [\mathsf{p}] \\ |\Lambda| = \mathsf{p} - \mathsf{q}}} \mathcal{N}(G^{\pi}_{\Lambda^{\mathsf{n}}}) \right) \cap \mathcal{X}_u(A), \end{aligned}$$

where the sum is conducted over $|\Lambda| = \mathsf{p} - \mathsf{q}$ instead of $|\Lambda| \geq \mathsf{p} - \mathsf{q}$, as done in the proof of Proposition 3.1.4. □

Since each undetectable subspace $\overline{\mathcal{D}}(A, C_{\Lambda}^{\pi})$ is A-invariant, the sum of those subspaces is also A-invariant. Therefore, the q-redundant undetectable subspace $\overline{\mathcal{D}}_{\mathsf{q}}(A, C)$ is invariant under A.

Based on Proposition 3.1.11, the q-redundant detectability can be characterized in terms of the q-redundant undetectable subspace as follows.

Corollary 3.1.12. The followings are equivalent:
(i) The pair (A, C) is q-redundant detectable (or, asymptotically q-redundant observable);
(ii) The q-redundant undetectable subspace of the pair (A, C) is trivial, that is, $\overline{\mathcal{D}}_{\mathsf{q}}(A, C) = \{0\}$. ◊

Proof. From the following equivalence,

$$\begin{aligned}\overline{\mathcal{D}}_{\mathsf{q}}(A, C) = \{0\} \quad &\Leftrightarrow \quad \sum_{\substack{\Lambda \subset [\mathsf{p}] \\ |\Lambda| = \mathsf{p}-\mathsf{q}}} \overline{\mathcal{D}}(A, C_{\Lambda}^{\pi}) = \{0\} \\ &\Leftrightarrow \quad \overline{\mathcal{D}}(A, C_{\Lambda}^{\pi}) = \{0\}\,, \quad {}^{\forall}\Lambda \subset [\mathsf{p}] \ \text{ s.t. } \ |\Lambda| = \mathsf{p} - \mathsf{q},\end{aligned}$$

the proof is completed. □

Moreover, the quotient space of the undetectable subspace, $\mathbb{R}^{\mathsf{n}}/\overline{\mathcal{D}}(A, C)$, is sometimes called, with abuse of terminology, the detectable subspace. Since the quotient space $\mathbb{R}^{\mathsf{n}}/\overline{\mathcal{D}}(A, C)$ is isomorphic to the orthogonal complement $\overline{\mathcal{D}}(A, C)^{\perp}$ of $\overline{\mathcal{D}}(A, C)$, the detectable subspace of the pair (A, C) denoted as $\mathcal{D}(A, C)$, becomes the sum of two subspaces $\mathcal{R}(G'^{\top})$ and $\mathcal{X}_u(A)^{\perp}$ as given in the following equation of

$$\mathcal{D}(A, C) = \overline{\mathcal{D}}(A, C)^{\perp} = \big(\mathcal{N}(G') \cap \mathcal{X}_u(A)\big)^{\perp} = \mathcal{R}(G'^{\top}) + \mathcal{X}_u(A)^{\perp}.$$

Define q-*redundant detectable subspace* as the quotient space of the q-redundant undetectable subspace, which is isomorphic to the orthogonal complement of the q-redundant undetectable subspace. Then, the following proposition shows how to compute the redundant detectable subspace, which is a direct consequence from Propositions 3.1.6 and 3.1.11.

Proposition 3.1.13. The q-redundant detectable subspace of the pair (A, C), $\mathcal{D}_{\mathsf{q}}(A, C)$, can be computed by

$$\mathcal{D}_{\mathsf{q}}(A, C) = \mathcal{O}_{\mathsf{q}}(A, C) + \mathcal{X}_u(A)^{\perp} = \bigcap_{\substack{\Lambda \subset [\mathsf{p}] \\ |\Lambda| = \mathsf{p}-\mathsf{q}}} \mathcal{R}({G^{\pi}_{\Lambda^{\mathsf{n}}}}^{\top}) + \mathcal{X}_u(A)^{\perp}. \qquad \Diamond$$

3.2 Attack Detectability and Dynamic Security Index

Regarding *attack detectability*, undetectable attacks are introduced in [45] and [39] for a static output map (3.1.1b) (without the dynamics part (3.1.1a)) with applications to power systems. In short, for the static measurement $\bar{y} = Cx + a \in \mathbb{R}^{\mathsf{p}}$, the attack vector a is undetectable if and only if $a = Cx_a$ for some $x_a \in \mathbb{R}^{\mathsf{n}}$. This is because the residual signal $r := \bar{y} - CC^{\dagger}\bar{y}$ becomes $0_{\mathsf{p}\times 1}$ if and only if $a \in \mathcal{R}(C)$ by [45, Theorem 3.2]. We can generalize this concept to a dynamical system (3.1.1) (both (3.1.1a) and (3.1.1b)). To this end, we first extend the results on the static output map directly without considering the properties of dynamics. This direct extension, which later is shown to be related to the redundant observability, detects the presence of all types of attacks regardless of their impact on dynamical systems. In other words, attacks which do not have any disruptive influence on the system (e.g., attacks which vanish as time goes on), are even detected. Just as we weakened the concept of observability to that of the detectability, the result of direct extension will be slightly modified in consideration of the dynamic properties. Hence, the modified extension, which is closely associated with the redundant detectability, only concerns the disruptive attacks which may be unstable and do not converge to zero as time goes on.

Now, the output measurements of the system (3.1.1) for a finite time period k are collected and the stacked output sequence is computed as

$$\bar{y}^{[0:k-1]} := \begin{bmatrix} \bar{y}_1^{[0:k-1]} \\ \bar{y}_2^{[0:k-1]} \\ \vdots \\ \bar{y}_{\mathsf{p}}^{[0:k-1]} \end{bmatrix} = \begin{bmatrix} G_1^{(k)} \\ G_2^{(k)} \\ \vdots \\ G_{\mathsf{p}}^{(k)} \end{bmatrix} x(0) + \begin{bmatrix} a_1^{[0:k-1]} \\ a_2^{[0:k-1]} \\ \vdots \\ a_{\mathsf{p}}^{[0:k-1]} \end{bmatrix} = G^{(k)} x(0) + a^{[0:k-1]} \qquad (3.2.1)$$

where

$$\bar{y}_{\mathsf{i}}^{[0:k-1]} := \begin{bmatrix} \bar{y}_{\mathsf{i}}(0) \\ \bar{y}_{\mathsf{i}}(1) \\ \vdots \\ \bar{y}_{\mathsf{i}}(k-1) \end{bmatrix} \quad \text{and} \quad \bar{a}_{\mathsf{i}}^{[0:k-1]} := \begin{bmatrix} \bar{a}_{\mathsf{i}}(0) \\ \bar{a}_{\mathsf{i}}(1) \\ \vdots \\ \bar{a}_{\mathsf{i}}(k-1) \end{bmatrix}.$$

Note that, with $k = \mathsf{n}$ in (3.2.1), the situation is exactly the same as the noiseless case in Section 2.2.1 and $a^{[0:\mathsf{n}-1]}$ is (n-stacked) q-sparse by Assumption 3.1.1. Thus, by comparing this finding with the results in Section 2.2.1, we can introduce the notion of *attack detectability* of the system $\overline{\mathcal{P}}$ as follows.

Definition 3.2.1. For a dynamical system (3.1.1), a non-trivial attack signal $a^{[0:\mathsf{n}-1]} \in \mathbb{R}^{\mathsf{np}}$ is said to be *undetectable* with respect to the pair (A, C) if there are two different $x(0)$ and $x'(0)$ in $\mathbb{R}^{\mathsf{n}}$ such that $Gx(0) + a^{[0:\mathsf{n}-1]} = Gx'(0)$. ♢

In other words, the non-trivial attack signal $a^{[0:\mathsf{n}-1]} \neq 0_{\mathsf{np}\times 1}$ is undetectable with respect to (A, C) if and only if $a^{[0:\mathsf{n}-1]} = Gx_a$ for some $x_a \neq 0_{\mathsf{n}\times 1}$. By the way, Definition 3.2.1 identifies attack detectability in respect of the attack signal a. As for the dynamical system (3.1.1), this notion can also be defined analogously with the q-sparsity assumption on a (i.e., Assumption 3.1.1) as follows.

Definition 3.2.2. The dynamical system (3.1.1) or the pair (A, C) is said to be q-*attack detectable* if, for all $x(0), x'(0) \in \mathbb{R}^{\mathsf{n}}$ and $a^{[0:\mathsf{n}-1]} \in \Sigma_{\mathsf{q}}^{\mathsf{n}}$ such that $Gx(0) + a^{[0:\mathsf{n}-1]} = Gx'(0)$, it holds that $x(0) = x'(0)$. ♢

Furthermore, the direct comparison between Definitions 2.2.2 and 3.2.2 simply leads to the following proposition.

Proposition 3.2.1. The followings are equivalent:
(i) The pair (A, C) is q-attack detectable;
(ii) The observability matrix G is (n-stacked) q-error detectable;
(iii) The pair (A, C) is q-redundant observable;
(iv) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$, $G_{\Lambda^{\mathsf{n}}}$ (or, equivalently $G^{\pi}_{\Lambda^{\mathsf{n}}}$) has full column rank;
(v) For any $x \in \mathbb{R}^{\mathsf{n}}$ where $x \neq 0_{\mathsf{n}\times 1}$, $\|Gx\|_{0^{\mathsf{n}}} > \mathsf{q}$;
(vi) For any $x, x' \in \mathbb{R}^{\mathsf{n}}$ where $x \neq x'$, $\mathsf{d}_{0^{\mathsf{n}}}(Gx, Gx') > \mathsf{q}$;
(vii) For any $v \in \mathcal{V}(A)$, $\|Cv\|_0 > \mathsf{q}$. ♢

As a tool for vulnerability analysis of a system, the *security index* quantifies fundamental limitations on the attack detectability. For example, the *static security index*, $\alpha_s(C)$, of the output map (3.1.1b) is defined as the minimum number of sensor attacks for adversaries to remain undetectable and is computed in [28] by

$$\alpha_s(C) = \min_{x \in \mathbb{R}^{\mathsf{n}},\, x \neq 0_{\mathsf{n}\times 1}} \|Cx\|_0 = \mathsf{cospark}(C). \tag{3.2.2}$$

Just as it is for the attack detectability, the concept of static security index can also be extended to the dynamical system (3.1.1). That is, the *dynamic security index*, $\alpha_d(A, C)$, of the system (3.1.1) is defined by the minimum number of sensor attacks for adversaries to remain undetectable in consideration of (3.1.1a) as well as (3.1.1b). Since a non-zero $a^{[0:\mathsf{n}-1]}$ in (3.2.1) is undetectable if and only if $a^{[0:\mathsf{n}-1]} = Gx_a$ for some $x_a \neq 0_{\mathsf{n}\times 1}$ by Definition 3.2.1, the dynamic security index can be computed by

$$\alpha_d(A, C) := \min_{a = Gx,\, x \neq 0_{\mathsf{n}\times 1}} \|a\|_{0^{\mathsf{n}}} = \min_{x \in \mathbb{R}^{\mathsf{n}},\, x \neq 0_{\mathsf{n}\times 1}} \|Gx\|_{0^{\mathsf{n}}} = \mathsf{cospark}^{\mathsf{n}}(G). \tag{3.2.3}$$

However, (3.2.3) is computationally intensive due to the combinatorial nature of the ℓ_0 optimization problem. Thus, another method to obtain the dynamic security index, which requires less computational burden, is presented in the following proposition.

Proposition 3.2.2. For $\alpha_d(A, C)$ given in (3.2.3), it holds that

$$\alpha_d(A, C) = \min_{v \in \mathcal{V}(A)} \|Cv\|_0. \tag{3.2.4}$$

$\Diamond$

Proof. The equality can simply be inferred from Proposition 3.1.1.(iv) and Proposition 3.1.2.(ii) (or, from Proposition 3.2.1.(v) and (vii)). However, a direct proof is given for the readers' convenience as follows. When $Av = \lambda v$, one can trivially check that

$$\min_{v \in \mathcal{V}(A)} \|Cv\|_0 = \min_{v \in \mathcal{V}(A)} \|Gv\|_{0^{\mathsf{n}}}$$

since $G_{\mathsf{i}}v = \begin{bmatrix} c_{\mathsf{i}}v & \lambda c_{\mathsf{i}}v & \cdots & \lambda^{\mathsf{n}-1}c_{\mathsf{i}}v \end{bmatrix}^{\top}$. Noting that

$$\min_{x \in \mathbb{C}^{\mathsf{n}},\, x \neq 0_{\mathsf{n}\times 1}} \|Gx\|_{0^{\mathsf{n}}} = \min_{x \in \mathbb{R}^{\mathsf{n}},\, x \neq 0_{\mathsf{n}\times 1}} \|Gx\|_{0^{\mathsf{n}}}$$

because $G \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$ is a real matrix, it suffices to show that

$$\min_{v \in \mathcal{V}(A)} \|Gv\|_{0^{\mathsf{n}}} = \min_{x \in \mathbb{C}^{\mathsf{n}},\, x \neq 0_{\mathsf{n}\times 1}} \|Gx\|_{0^{\mathsf{n}}}.$$

Now, we claim that there exists $v^* \in \mathcal{V}(A)$ such that

$$\|Gv^*\|_{0^{\mathsf{n}}} = \min_{x \in \mathbb{C}^{\mathsf{n}},\, x \neq 0_{\mathsf{n}\times 1}} \|Gx\|_{0^{\mathsf{n}}}.$$

Let us denote the optimal value of the problem (3.2.3) by

$$\alpha^* := \min_{x \in \mathbb{R}^{\mathsf{n}},\, x \neq 0_{\mathsf{n}\times 1}} \|Gx\|_{0^{\mathsf{n}}}.$$

By the equivalence between Proposition 3.2.1.(iv) and (v), there exists an index set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| = \mathsf{p} - \alpha^*$ such that the observability matrix $G^{\pi}_{\Lambda^{\mathsf{n}}}$ does not have full column rank but the observability matrix $G^{\pi}_{(\Lambda\cup\{\mathsf{i}\})^{\mathsf{n}}}$ has full column rank for every $\mathsf{i} \in \Lambda^c$. That is, the pair (A, C^{π}_{Λ}) is not observable but the pair $(A, C^{\pi}_{\Lambda\cup\{\mathsf{i}\}})$ is observable for every $\mathsf{i} \in \Lambda^c$. Applying the PBH observability test, we conclude that there exist $\lambda^* \in \mathbb{C}$ and $v^* \in \mathcal{V}(A)$ such that

$$\begin{bmatrix} \lambda^* I_{\mathsf{n}\times\mathsf{n}} - A \\ C^{\pi}_{\Lambda} \end{bmatrix} v^* = \begin{bmatrix} 0_{\mathsf{n}\times 1} \\ 0_{(\mathsf{p}-\alpha^*)\times 1} \end{bmatrix} \quad \text{and} \quad c_{\mathsf{i}}v^* \neq 0, \ {}^{\forall}\mathsf{i} \in \Lambda^c.$$

The claim easily follows by verifying that $\|Gv^*\|_{0^{\mathsf{n}}} = \alpha^*$. □

Remark 3.2.1. In [11], the dynamic security index is computed the same as (3.2.4) by examining the system's strong observability and the weakly unobservable subspace. However, Proposition 3.2.2 has more meaning than the results in [11] since it effectively relates the dynamic security index with the redundant observability through the cospark of the observability matrix. Note that $\alpha_d(A, C)$ obtained by (3.2.4) has a computational advantage compared with (3.2.3). That is, (3.2.4) only investigates the minimum ℓ_0 norm of Cv among $\mathcal{V}(A) \ni v$, while (3.2.3) needs to examine the whole space $\mathbb{R}^{\mathsf{n}}$. Therefore, if the geometric multiplicity of each eigenvalue of A is one (e.g., A has n distinct eigenvalues), we need to test at most n eigenvectors in order to compute (3.2.4) while the problem of calculating (3.2.3) directly is computationally infeasible for large p, just like the computation of $\alpha_s(C)$ in (3.2.2) is NP-hard [28, Section III]. $\Diamond$

Now, we slightly modify Definition 3.2.2 in accordance with dynamic properties. Instead of restricting the time period as a finite time n, an asymptotic property that holds in the limit as the time tends to infinity, is investigated, and the notion of *asymptotic attack detectability* is given as follows.

Definition 3.2.3. For a dynamical system (3.1.1), a non-trivial attack signal $a(\cdot)$ is said to be *asymptotically undetectable* with respect to the pair (A, C) if there exist $x(0)$ and $x'(0)$ in $\mathbb{R}^{\mathsf{n}}$ with

$$\lim_{k\to\infty} x(k) \neq \lim_{k\to\infty} x'(k)$$

where $x(k) = A^{k-1}x(0)$ and $x'(k) = A^{k-1}x'(0)$, satisfying $G^{(k)}x(0) + a^{[0:k-1]} = G^{(k)}x'(0)$ for all $k \geq 0$. $\Diamond$

In other words, the signal $a^{[0:k-1]} \neq 0_{k\mathsf{p}\times 1}$ is asymptotically undetectable with respect to (A, C) if and only if there exists x_a satisfying $\lim\limits_{k\to\infty} A^{k-1}x_a \neq 0_{\mathsf{n}\times 1}$ and $a^{[0:k-1]} = G^{(k)}x_a$. By the way, Definition 3.2.3 identifies asymptotic attack detectability in respect of the attack signal a. As for the dynamical system (3.1.1), this notion can also be defined analogously with the q-sparsity assumption on a (i.e., Assumption 3.1.1) as follows.

Definition 3.2.4. The dynamical system (3.1.1) or the pair (A, C) is said to be *asymptotically* q*-attack detectable* if, for any $x(0), x'(0) \in \mathbb{R}^{\mathsf{n}}$ and $a^{[0:k-1]} \in \Sigma_{\mathsf{q}}^{k}$ such that $G^{(k)}x(0) + a^{[0:k-1]} = G^{(k)}x'(0)$ for all $k \geq 0$, it holds that

$$\lim_{k\to\infty} x(k) = \lim_{k\to\infty} x'(k)$$

where $x(k) = A^{k-1}x(0)$ and $x'(k) = A^{k-1}x'(0)$. $\Diamond$

Just as the attack detectability is equivalent to the redundant observability in Proposition 3.2.1, the asymptotic attack detectability can be characterized by the redundant detectability as follows.

Proposition 3.2.3. The followings are equivalent:
(i) The pair (A, C) is asymptotically q-attack detectable;
(ii) The pair (A, C) is q-redundant detectable (or, asymptotically q-redundant observable);
(iii) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$, $\mathcal{N}(G_{\Lambda^{\mathsf{n}}}) \subset \mathcal{X}_s(A)$ (or, equivalently $\mathcal{N}(G^{\pi}_{\Lambda^{\mathsf{n}}}) \subset \mathcal{X}_s(A)$);
(iv) For any $x \notin \mathcal{X}_s(A)$, $\|Gx\|_{0^{\mathsf{n}}} > \mathsf{q}$;
(v) For any $v \in \mathcal{V}_u(A)$, $\|Cv\|_0 > \mathsf{q}$. ◊

Proof. (i) ⇒ (ii): Pick any $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$, and we claim that

$$\mathsf{rank}\left(\begin{bmatrix} \lambda I_{\mathsf{n}\times\mathsf{n}} - A \\ C_\Lambda \end{bmatrix}\right) = \mathsf{n}$$

for any unstable eigenvalue λ of A. It is enough to show that $C_\Lambda v \neq 0$ for any eigenvector v of A corresponding to the unstable eigenvalue by the PBH test. Assume to the contrary that there exists an eigenvector v^* of A corresponding to the unstable eigenvalue λ^* such that $C_\Lambda v^* = 0$. Thus, we have $G^{(k)}_{\Lambda^k} v^* = 0_{k\mathsf{p}\times 1}$ because

$$G^{(k)}_{\mathsf{i}} v^* = \begin{bmatrix} c_{\mathsf{i}} \\ c_{\mathsf{i}} A \\ \vdots \\ c_{\mathsf{i}} A^{k-1} \end{bmatrix} v^* = \begin{bmatrix} c_{\mathsf{i}} v^* \\ c_{\mathsf{i}} A v^* \\ \vdots \\ c_{\mathsf{i}} A^{k-1} v^* \end{bmatrix} = \begin{bmatrix} c_{\mathsf{i}} v^* \\ \lambda^* c_{\mathsf{i}} v^* \\ \vdots \\ \lambda^{*k-1} c_{\mathsf{i}} v^* \end{bmatrix} = \begin{bmatrix} 1 \\ \lambda^* \\ \vdots \\ \lambda^{*k-1} \end{bmatrix} c_{\mathsf{i}} v^* = 0_{k\times 1}$$

for all $\mathsf{i} \in \Lambda$. Finally, it is obtained that $G^{(k)} v^*$ is (k-stacked) q-sparse, i.e., $G^{(k)} v^* \in \Sigma^k_{\mathsf{q}}$. Let $x'(0) = x(0) + v^*$ and $a^{[0:k-1]} = G^{(k)} v^* \in \Sigma^k_{\mathsf{q}}$. It follows that

$$G^{(k)} x(0) + a^{[0:k-1]} = G^{(k)}(x(0) + v^*) = G^{(k)} x'(0),$$

which implies

$$\lim_{k\to\infty} A^{k-1}v^* = 0_{\mathsf{n}\times 1}$$

due to the asymptotic q-attack detectability of the pair (A, C) and Definition 3.2.4. This contradicts the fact that v^* is an eigenvector of A corresponding to the unstable eigenvalue λ^*, which results in

$$\lim_{k\to\infty} A^{k-1}v^* = \lim_{k\to\infty} \lambda^{*k-1}v^* \neq 0_{\mathsf{n}\times 1}.$$

(ii) ⇒ (i): Suppose that (i) does not hold, i.e., the pair (A, C) is not asymptotically q-attack detectable. Then, there exist $x_a \neq 0_{\mathsf{n}\times 1}$ and $a^{[0:k-1]} \in \Sigma_{\mathsf{q}}^k$ such that $G^{(k)}x_a = a^{[0:k-1]}$ for all $k \geq 0$ with

$$\lim_{k\to\infty} A^{k-1}x_a \neq 0_{\mathsf{n}\times 1}.$$

Let $\Lambda := \left(\mathsf{supp}^k\left(a^{[0:k-1]}\right)\right)^c$, and it is obvious that $G^{(k)}_{\Lambda^k}x_a = 0_{k\mathsf{p}\times 1}$ and $|\Lambda| \geq \mathsf{p}-\mathsf{q}$. In short, there exist $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p}-\mathsf{q}$ and $x_a \neq 0_{\mathsf{n}\times 1}$ such that $G^{(k)}_{\Lambda^k}x_a = 0_{k\mathsf{p}\times 1}$ for all $k \geq 0$ and $\lim_{k\to\infty} A^{k-1}x_a \neq 0_{\mathsf{n}\times 1}$. With this index set Λ and state vector x_a at hand, think about the linear system

$$\begin{cases} x(k+1) = Ax(k) \\ \quad y_\Lambda(k) = C_\Lambda x(k) \end{cases}$$

for the pair (A, C_Λ). Here, we can carry out the change of variable $\begin{bmatrix} z \\ w \end{bmatrix} = Tx$ to obtain the Kalman observability decomposition[5]

$$\begin{cases} \begin{bmatrix} z(k+1) \\ w(k+1) \end{bmatrix} = TAT^{-1}Tx(k) = \begin{bmatrix} A_o & O \\ A_{21} & A_{\bar{o}} \end{bmatrix} \begin{bmatrix} z(k) \\ w(k) \end{bmatrix} \\ \qquad y_\Lambda(k) = C_\Lambda T^{-1}Tx(k) = \begin{bmatrix} C_o & O \end{bmatrix} \begin{bmatrix} z(k) \\ w(k) \end{bmatrix}. \end{cases}$$

[5] The detailed procedure for the Kalman decomposition will be presented in Section 4.2.1.

Let us denote $\begin{bmatrix} z_a \\ w_a \end{bmatrix} := Tx_a$ and it follows from $G^{(k)}_{\Lambda^k} x_a = 0_{k\mathsf{p}\times 1}$ that

$$\begin{bmatrix} C_\Lambda \\ C_\Lambda A \\ \vdots \\ C_\Lambda A^{k-1} \end{bmatrix} x_a = \begin{bmatrix} C_\Lambda T^{-1} \\ C_\Lambda T^{-1}(TAT^{-1}) \\ \vdots \\ C_\Lambda T^{-1}(TAT^{-1})^{k-1} \end{bmatrix} Tx_a = \begin{bmatrix} C_o & O \\ C_o A_o & O \\ \vdots & \\ C_o A_o^{k-1} & O \end{bmatrix} \begin{bmatrix} z_a \\ w_a \end{bmatrix} = 0_{k\mathsf{p}\times 1}.$$

Since the pair (A_o, C_o) is observable by the intrinsic property of the Kalman observability decomposition, we must have $z_a = 0$. Furthermore, because T is a nonsingular matrix and $\lim_{k\to\infty} A^{k-1} x_a \neq 0_{\mathsf{n}\times 1}$, it is easily obtained that

$$\lim_{k\to\infty} TA^{k-1}x_a = \lim_{k\to\infty} \begin{bmatrix} A_o & O \\ A_{21} & A_{\bar{o}} \end{bmatrix}^{k-1} \begin{bmatrix} 0 \\ w_a \end{bmatrix} = \begin{bmatrix} 0 \\ \lim_{k\to\infty} A_{\bar{o}}^{k-1} w_a \end{bmatrix} \neq 0_{\mathsf{n}\times 1}.$$

Finally, $A_{\bar{o}}^{k-1} w_a$ doe not converge to zero for some w_a, which means that $A_{\bar{o}}$ is not stable. Therefore, the pair (A, C_Λ) is not detectable for some $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$, and hence, the pair (A, C) is not q-redundant detectable.
(ii) ⇔ (iii) ⇔ (iv) ⇔ (v): This is proved in Propositions 3.1.8 and 3.1.9. □

Note that the property of Proposition 3.2.1.(vii) determines the dynamic security index as computed in Proposition 3.2.2 by exploring $\|Cv\|_0$ for "all" eigenvector v's of A. While the dynamic security index is defined as the minimum number of "any" type of sensor attacks for adversaries to remain undetectable, another useful index which quantifies practical risks of the attack by considering its "disruptive" characteristics as well as the undetectable property, can be proposed. Recall from Definition 3.2.3 that a non-trivial signal $a(\cdot)$ is asymptotically undetectable if and only if there exists x_a satisfying

$$\lim_{k\to\infty} A^{k-1} x_a \neq 0_{\mathsf{n}\times 1}$$

and $a^{[0:k-1]} = G^{(k)} x_a$ for all $k \geq 0$. Since $G^{(k)}_{\mathsf{i}} x_a = 0_{k\times 1}$ for all $k \geq 0$ is equivalent to the condition $G_{\mathsf{i}} x_a = 0_{\mathsf{n}\times 1}$, the *asymptotic dynamic security index* can be computed by

$$\begin{aligned}\alpha_a(A,C) &:= \min_{a=Gx,\ A^{k-1}x\nrightarrow 0_{\mathsf{n}\times 1}} \|a\|_{0^{\mathsf{n}}} = \min_{x\in\mathbb{R}^{\mathsf{n}},\ A^{k-1}x\nrightarrow 0_{\mathsf{n}\times 1}} \|Gx\|_{0^{\mathsf{n}}} \\ &= \min_{x\notin\mathcal{X}_s(A)} \|Gx\|_{0^{\mathsf{n}}}.\end{aligned} \tag{3.2.5}$$

That is, the asymptotic dynamic security index, $\alpha_a(A,C)$, of the system (3.1.1) is defined by the minimum number of sensor attacks which remain undetectable and does not converge to zero as time goes on. Please note that $x \in \mathbb{R}^{\mathsf{n}}$ satisfying $A^{k-1}x \nrightarrow 0_{\mathsf{n}\times 1}$ means that $x \notin \mathcal{X}_s(A)$. Similar to Proposition 3.2.2, it can be computed from the eigenvectors of A corresponding to the "unstable" eigenvalues.

Proposition 3.2.4. For $\alpha_a(A,C)$ given in (3.2.5), it holds that

$$\alpha_a(A,C) = \min_{v\in\mathcal{V}_u(A)} \|Cv\|_0. \tag{3.2.6}$$

◊

Proof. The equality can simply be inferred from Proposition 3.1.8.(iii) and Proposition 3.1.9.(ii) (or, from Proposition 3.2.3.(iv) and (v)) because $x \in \mathbb{R}^{\mathsf{n}}$ satisfying $A^{k-1}x \nrightarrow 0_{\mathsf{n}\times 1}$ means that $x \notin \mathcal{X}_s(A)$. However, a direct proof is given for the readers' convenience as follows. When $Av = \lambda v$, one can trivially check that

$$\min_{v\in\mathcal{V}_u(A)} \|Cv\|_0 = \min_{v\in\mathcal{V}_u(A)} \|Gv\|_{0^{\mathsf{n}}}$$

since $G_{\mathsf{i}}v = \begin{bmatrix} c_{\mathsf{i}}v & \lambda c_{\mathsf{i}}v & \cdots & \lambda^{\mathsf{n}-1}c_{\mathsf{i}}v \end{bmatrix}^\top$. Noting that

$$\min_{x\in\mathbb{C}^{\mathsf{n}},\ A^{k-1}x\nrightarrow 0_{\mathsf{n}\times 1}} \|Gx\|_{0^{\mathsf{n}}} = \min_{x\in\mathbb{R}^{\mathsf{n}},\ A^{k-1}x\nrightarrow 0_{\mathsf{n}\times 1}} \|Gx\|_{0^{\mathsf{n}}}$$

because $G \in \mathbb{R}^{\mathsf{np}\times\mathsf{n}}$ is a real matrix, it suffices to show that

$$\min_{v\in\mathcal{V}_u(A)} \|Gv\|_{0^{\mathsf{n}}} = \min_{x\in\mathbb{C}^{\mathsf{n}},\ A^{k-1}x\nrightarrow 0_{\mathsf{n}\times 1}} \|Gx\|_{0^{\mathsf{n}}}.$$

Since $A^{k-1}v \nrightarrow 0_{\mathsf{n}\times 1}$ for any $v \in \mathcal{V}_u(A)$, we have $\mathcal{V}_u(A) \subset \left\{x \in \mathbb{C}^{\mathsf{n}} : A^{k-1}x \nrightarrow 0_{\mathsf{n}\times 1}\right\}$.

Now, we claim that there exists $v^* \in \mathcal{V}_u(A)$ such that

$$\|Gv^*\|_{0^{\mathsf{n}}} = \min_{x \in \mathbb{C}^{\mathsf{n}},\, A^{k-1}x \nrightarrow 0_{\mathsf{n}\times 1}} \|Gx\|_{0^{\mathsf{n}}}.$$

Let us denote the optimal value of the problem (3.2.5) by

$$\alpha^* := \min_{x \in \mathbb{R}^{\mathsf{n}},\, A^{k-1}x \nrightarrow 0_{\mathsf{n}\times 1}} \|Gx\|_{0^{\mathsf{n}}}.$$

By the equivalence between Proposition 3.2.3.(iii) and (iv), there exists an index set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| = \mathsf{p} - \alpha^*$ such that $\mathcal{N}(G_{\Lambda^{\mathsf{n}}})$, the null space of $G_{\Lambda^{\mathsf{n}}}$, is not contained in $\mathcal{X}_s(A)$ but $\mathcal{N}(G_{(\Lambda\cup\{\mathsf{i}\})^{\mathsf{n}}})$, the null space of $G_{(\Lambda\cup\{\mathsf{i}\})^{\mathsf{n}}}$, is contained in $\mathcal{X}_s(A)$ for every $\mathsf{i} \in \Lambda^c$. That is, the pair (A, C_Λ^π) is not detectable but the pair $(A, C_{\Lambda\cup\{\mathsf{i}\}}^\pi)$ is detectable for every $\mathsf{i} \in \Lambda^c$. Applying the PBH detectability test, we conclude that there exist unstable $\lambda^* \in \mathbb{C}$ (i.e., $|\lambda^*| \geq 1$) and $v^* \in \mathcal{V}_u(A)$ such that

$$\begin{bmatrix} \lambda^* I_{\mathsf{n}\times\mathsf{n}} - A \\ C_\Lambda^\pi \end{bmatrix} v^* = \begin{bmatrix} 0_{\mathsf{n}\times 1} \\ 0_{(\mathsf{p}-\alpha^*)\times 1} \end{bmatrix} \quad \text{and} \quad c_{\mathsf{i}} v^* \neq 0, \ {}^\forall \mathsf{i} \in \Lambda^c.$$

The claim easily follows by verifying that $\|Gv^*\|_{0^{\mathsf{n}}} = \alpha^*$. □

Note that the asymptotic dynamic security index investigates $\|Cv\|_0$ only for "unstable" eigenvector $v \in \mathcal{V}_u(A)$ while the dynamic security index compares $\|Cv\|_0$ for "all" eigenvector $v \in \mathcal{V}(A)$.

3.3 Observability under Sparse Sensor Attacks

In order to analyze attack-resilience of state estimation, this section introduces the observability notion of a control system under sensor attacks and gives some equivalent conditions. Now, a notion of *observability under* q*-sparse sensor attacks* is given as follows.

Definition 3.3.1. The dynamical system (3.1.1) or the pair (A, C) is said to be *observable under* q*-sparse sensor attacks* if the initial state $x(0)$ can be determined from the output y over a finite number of sampling steps with any input signal a satisfying Assumption 3.1.1. ◊

The following proposition presents necessary and sufficient conditions for the observability under q-sparse sensor attacks, and further, it finally summarizes the relationship among newly introduced notions in this chapter.

Proposition 3.3.1. The followings are equivalent:
(i) The pair (A, C) is observable under q-sparse sensor attacks;
(ii) The observability matrix G is (n-stacked) q-error correctable;
(iii) The observability matrix G is (n-stacked) $2\mathsf{q}$-error detectable;
(iv) The pair (A, C) is $2\mathsf{q}$-redundant observable;
(v) The pair (A, C) is $2\mathsf{q}$-attack detectable;
(vi) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - 2\mathsf{q}$, $G_{\Lambda^{\mathsf{n}}}$ (or, equivalently $G^{\pi}_{\Lambda^{\mathsf{n}}}$) has full column rank;
(vii) For any $x \in \mathbb{R}^{\mathsf{n}}$ where $x \neq 0_{\mathsf{n}\times 1}$, $\|Gx\|_{0^{\mathsf{n}}} > 2\mathsf{q}$;
(viii) For any $x, x' \in \mathbb{R}^{\mathsf{n}}$ where $x \neq x'$, $\mathsf{d}_{0^{\mathsf{n}}}(Gx, Gx') > 2\mathsf{q}$;
(ix) For any $v \in \mathcal{V}(A)$, $\|Cv\|_0 > 2\mathsf{q}$. ◊

Proof. (i) ⇔ (ii): Note that the output sequence $\bar{y}^{[0:\mathsf{n}-1]}$ is given by $\bar{y}^{[0:\mathsf{n}-1]} = Gx(0) + a^{[0:\mathsf{n}-1]} \in \mathbb{R}^{\mathsf{np}}$ in (3.2.1) and $a^{[0:\mathsf{n}-1]} \in \Sigma^{\mathsf{n}}_{\mathsf{q}}$ by Assumption 3.1.1, the result directly follows from Definition 2.2.3 which says that G is (n-stacked) q-error correctable if and only if $x(0)$ can be reconstructed from the output measurements $\bar{y}^{[0:\mathsf{n}-1]}$.
(ii) ⇔ (iii) : This is proved in Proposition 2.2.3.
(iii) ⇔ (iv) ⇔ (v) ⇔ (vi) ⇔ (vii) ⇔ (viii) ⇔ (ix): This is proved in Proposition 3.2.1. □

This proposition shows that newly introduced notions in this chapter are closely related to each other. One can test the observability under attacks by examining the well-known standard observability rank condition after eliminating any $2\mathsf{q}$ sensor outputs, which is eventually equivalent to q-error correctability of the observability matrix G.

As we have slightly weakened the notion of observability to that of detectability (or, asymptotic observability), the observability under q-sparse sensor attacks can be modified in accordance with dynamic properties. Instead of restricting the time period as a finite time n, an asymptotic property that holds in the limit as the time tends to infinity, is discussed. Finally, the notion of *asymptotic observability under* q*-sparse sensor attacks* is given as follows.

Definition 3.3.2. The dynamical system (3.1.1) or the pair (A, C) is said to be *detectable under* q*-sparse sensor attacks* (or, *asymptotically observable under* q*-sparse sensor attacks*) if the state variable $x(k)$ can be recovered asymptotically from the measurement output signal y with any sensor attack a satisfying Assumption 3.1.1. $\diamondsuit$

Finally, equivalent conditions for the detectability under q-sparse sensor attacks is explained in terms of the asymptotic attack detectability and the redundant detectability, as follows.

Proposition 3.3.2. The followings are equivalent:
(i) The pair (A, C) is detectable under q-sparse sensor attacks (or, asymptotically observable under q-sparse sensor attacks);
(ii) The pair (A, C) is $2\mathsf{q}$-redundant detectable (or, asymptotically $2\mathsf{q}$-redundant observable);
(iii) The pair (A, C) is asymptotically $2\mathsf{q}$-attack detectable;
(iv) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - 2\mathsf{q}$, $\mathcal{N}(G_{\Lambda^{\mathsf{n}}}) \subset \mathcal{X}_s(A)$ (or, equivalently $\mathcal{N}(G^{\pi}_{\Lambda^{\mathsf{n}}}) \subset \mathcal{X}_s(A)$);
(v) For any $x \notin \mathcal{X}_s(A)$, $\|Gx\|_{0^{\mathsf{n}}} > 2\mathsf{q}$;
(vi) For any $v \in \mathcal{V}_u(A)$, $\|Cv\|_0 > 2\mathsf{q}$. $\diamondsuit$

Proof. (i) $\Rightarrow$ (iii): Assume that $x(0), x'(0) \in \mathbb{R}^{\mathsf{n}}$ and $a''^{[0:k-1]} \in \Sigma^k_{2\mathsf{q}}$ such that $G^{(k)}x(0) + a''^{[0:k-1]} = G^{(k)}x'(0)$ for all $k \geq 0$, are given. Let $a^{[0:k-1]}$ and $a'^{[0:k-1]}$ be such that $a''^{[0:k-1]} = a^{[0:k-1]} - a'^{[0:k-1]}$ where $a^{[0:k-1]}, a^{[0:k-1]} \in \Sigma^k_{\mathsf{q}}$. Thus, we have $G^{(k)}x(0) + a^{[0:k-1]} = G^{(k)}x'(0) + a'^{[0:k-1]}$ for all $k \geq 0$. Note that $\bar{y}^{[0:k-1]} := G^{(k)}x(0) + a^{[0:k-1]}$ is the measurement output signal of the system (3.1.1) with the initial state $x(0)$ and the attack signal $a^{[0:k-1]}$, and $\bar{y}'^{[0:k-1]} := G^{(k)}x'(0) + a'^{[0:k-1]}$ is the measurement output signal of the system (3.1.1) with the initial state $x'(0)$ and the attack signal $a'^{[0:k-1]}$. Since those two output signals $\bar{y}^{[0:k-1]}$ and $\bar{y}'^{[0:k-1]}$ are equal, they should recover the same state variable asymptotically by the definition of the detectability under q-sparse sensor attacks. Therefore, we have

$$\lim_{k\to\infty} x(k) = \lim_{k\to\infty} x'(k)$$

where $x(k) = A^{k-1}x(0)$ and $x'(k) = A^{k-1}x'(0)$. This is because if they are different from each other, then their estimates can never be the same.

(iii) ⇒ (i): We first claim that the limit of the state variable $x(k)$ of the system (3.1.1), $\lim_{k\to\infty} x(k)$, can be uniquely determined from the measurement $\bar{y}^{[0:k-1]} = G^{(k)}x(0) + a^{[0:k-1]}$ whenever the attack signal $a^{[0:k-1]}$ satisfies Assumption 3.1.1. Suppose that two output signals of the system (3.1.1), which are possibly generated by different initial states $x(0), x'(0) \in \mathbb{R}^{\mathsf{n}}$ and attack signals $a^{[0:k-1]}, a'^{[0:k-1]} \in \Sigma_{\mathsf{q}}^k$, are given. Let those output signals be $\bar{y}^{[0:k-1]} := G^{(k)}x(0) + a^{[0:k-1]}$ and $\bar{y}'^{[0:k-1]} := G^{(k)}x'(0) + a'^{[0:k-1]}$. Assume that $\bar{y}^{[0:k-1]} = \bar{y}'^{[0:k-1]}$ for all $k \geq 0$. Then, we have $G^{(k)}x(0) + a''^{[0:k-1]} = G^{(k)}x'(0)$ where $a''^{[0:k-1]} = a^{[0:k-1]} - a''^{[0:k-1]} \in \Sigma_{2\mathsf{q}}^k$ for all $k \geq 0$. Since the pair (A, C) is asymptotically $2\mathsf{q}$-attack detectable, it follows that

$$\lim_{k\to\infty} x(k) = \lim_{k\to\infty} x'(k)$$

where $x(k) = A^{k-1}x(0)$ and $x'(k) = A^{k-1}x'(0)$. Thus, the claim is proved, that is, $\lim_{k\to\infty} x(k)$ can be uniquely determined from the measurement $\bar{y}^{[0:k-1]} = G^{(k)}x(0) + a^{[0:k-1]}$. Therefore, when the pair (A, C) is asymptotically $2\mathsf{q}$-attack detectable, one should be able to recover $\lim_{k\to\infty} x(k)$ from the measurement output signal $\bar{y}^{[0:k-1]} = G^{(k)}x(0) + a^{[0:k-1]}$ with $a^{[0:k-1]}$ satisfying Assumption 3.1.1, because, in principle, one can exhaustively search for all $x'(0) \in \mathbb{R}^{\mathsf{n}}$ and $a'^{[0:k-1]} \in \Sigma_{\mathsf{q}}^k$ such that $\bar{y}^{[0:k-1]} = G^{(k)}x'(0) + a'^{[0:k-1]}$. This completes the proof. Although this proof does not give a concrete scheme to recover $\lim_{k\to\infty} x(k)$, the design procedure for state estimation is the main subject of the next chapter and the detailed algorithm will be proposed there.

(ii) ⇔ (iii)⇔ (iv) ⇔ (v) ⇔ (vi) : This is proved in Proposition 3.2.3. □

Chapter 4

Attack-Resilient State Estimation under Sensor Attacks for Linear Systems

Sensors are one of the vulnerable points for security of networked control systems, and thus, security problems of control systems whose measurements are compromised by adversaries are actively studied these days. In this chapter, it is supposed that all sensor information of control systems is collected at one place and we have developed an algorithm which estimates the state variable of the control systems even under sparse sensor attacks. This attack-resilient state estimator is required even when all distributed sensors send their measurement data to a sensor fusion center (or, information fusion center) through communication networks. Due to the insecure communication links, the measurement data in the networked control systems may be corrupted by adversaries. In addition, the measurement data can also be compromised by an attacker who physically tampers with the sensor itself. The proposed estimator consists of a bank of partial observers operating based on Kalman detectability (or, observability) decomposition and a decoder exploiting error correction techniques. In terms of time complexity, an ℓ_0 minimization problem in the decoder alleviates the computational efforts by reducing the search space to a finite set and by combining a detection algorithm to the optimization process. On the other hand, in terms of space complexity, the required memory is linear with the number of sensors by means of the decomposition used for constructing a bank of partial observers.

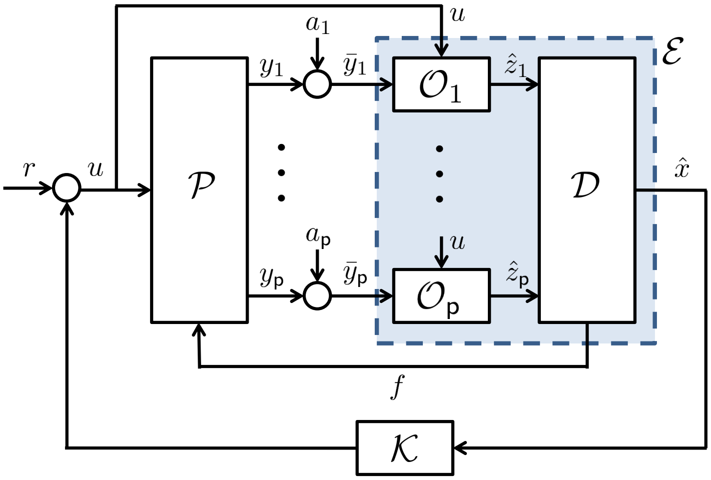


Figure 4.1: Total configuration of the feedback control system.

4.1 Problem Formulation

Overall configuration of the proposed control scheme is given in Fig. 4.1. The plant and the attack model under consideration are presented and the problem formulation is given in this section. We consider a discrete-time LTI plant given by

$$\mathcal{P}: \begin{cases} x(k+1) = Ax(k) + Bu(k) + d(k) \\ y(k) = Cx(k) + n(k) \end{cases} \tag{4.1.1}$$

where $x \in \mathbb{R}^\mathsf{n}$ is the state variables, $u \in \mathbb{R}^\mathsf{m}$ is the control inputs, and $y \in \mathbb{R}^\mathsf{p}$ is the sensor outputs. The dynamics are disrupted by the process disturbance $d \in \mathbb{R}^\mathsf{n}$ and sensors are corrupted by the measurement noise $n \in \mathbb{R}^\mathsf{p}$. The block diagram of the plant (4.1.1) is shown in Fig. 4.2. There are total p sensors which measure the system outputs and the i-th sensor's measurement at time k is denoted by

$$y_\mathsf{i}(k) = c_\mathsf{i}x(k) + n_\mathsf{i}(k)$$

where c_i is the i-th row of C. It is assumed that the pair (A, C) is detectable (or, observable), but the pair (A, c_i) is not necessarily detectable (or, observable).

Among various attack scenarios [89], we consider false data injection attacks on sensors. That is, adversarial attackers can inject arbitrary inputs to some (not

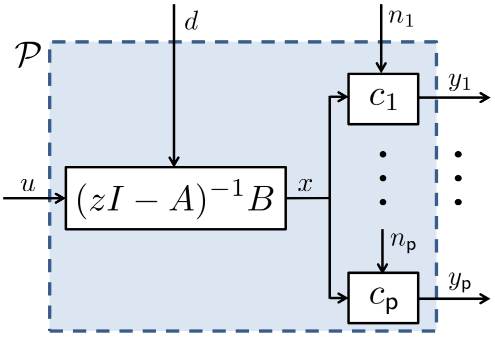
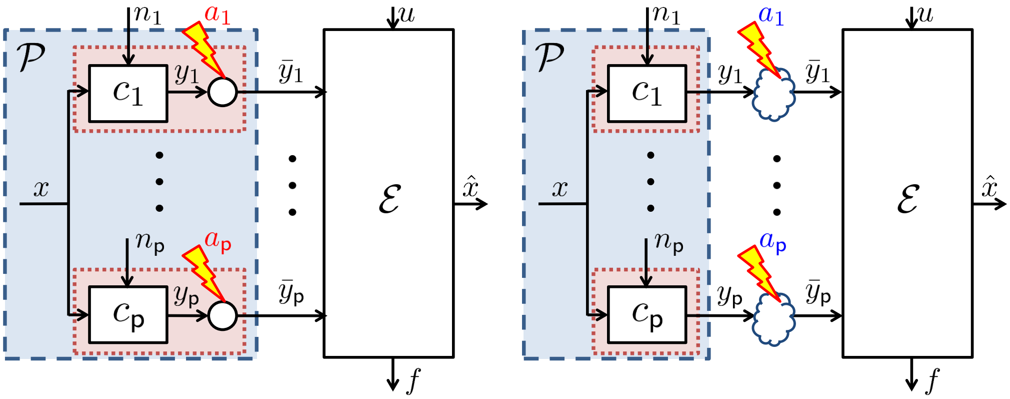


Figure 4.2: Configuration of the plant $\mathcal{P}$.



(a) Sensor attack

(b) Communication attack

Figure 4.3: Two scenarios of the measurement data attack.

all) sensors so that a part of measurements is compromised. Two scenarios are possible as illustrated in Fig. 4.3: first, these additive inputs may be induced by cyber or physical tampering with the sensors (Fig. 4.3a); second, adversaries may penetrate into the communication network on the output side of the plant because those communication links are not secure (Fig. 4.3b). In both cases, the attack is characterized by the attack vector a as in

$$\bar{y}(k) = y(k) + a(k) \tag{4.1.2}$$

where $\bar{y} \in \mathbb{R}^\mathsf{p}$ denotes the sensor data on the controller's side. Therefore, $\bar{y}(k)$, not $y(k)$, is used for state estimation.

Here, it is assumed that the adversaries have a complete knowledge about the plant, but can compromise only a part of the sensors, not all of them. Assuming that the attacker's resources are limited, we suppose that up to q out of p measurement outputs can be compromised. Therefore, a formal condition on the sparsity of the attack vector a can be given as follows.

Assumption 4.1.1. There exist at least $\mathsf{p}-\mathsf{q}$ sensors which are not attacked for all $k \geq 0$, i.e.,

$$\left|\left\{\mathsf{i} \in [\mathsf{p}] : a_{\mathsf{i}}(k) = 0, \ ^{\forall}k \geq 0\right\}\right| \geq \mathsf{p}-\mathsf{q}. \qquad \Diamond$$

This assumption tells more than $\|a(k)\|_0 \leq \mathsf{q}$ for all $k \geq 0$, in the sense that the compromised sensor channels are not altered for all time. In practice, this may not be the case. However, from the point of view of malicious attackers, it takes quite a long time and much effort to infiltrate into a new sensor. Thus, without loss of generality, it can be assumed that the attack channels remain the same in the long term although it is not revealed to the controller which channels are attacked. Hence, if the change of the compromised channels is not so frequent (that they do not change during the transient period of the algorithms to be presented), then the proposed attack detection and state estimation algorithms are still applicable.

The final goal of this chapter is to control (e.g., stabilizing or reference tracking) the plant $\mathcal{P}$ with an elaborate state estimator $\mathcal{E}$ and a state feedback control law[1] $\mathcal{K}$. Since the separation principle holds for LTI systems, the state feedback controller $\mathcal{K}$ can be selected independently assuming that state estimation is successfully conducted. Therefore, the primary objective of this paper is to design an estimator $\mathcal{E}$ which detects the attacked sensors and estimates the state $x(k)$ of the given system $\mathcal{P}$ under Assumption 4.1.1. To this end, we construct an attack-resilient estimator $\mathcal{E}$ which is composed of p partial observers[2] $\mathcal{O}_{\mathsf{i}}$'s and a decoder $\mathcal{D}$ as shown in Fig. 4.1 (i.e., the shaded block in Fig. 4.1). Additionally, the decoder $\mathcal{D}$ also provides a fault count signal f which may corresponds to the

[1] The configuration in Fig. 4.1 is one example of various control schemes. The controller $\mathcal{K}$ may also be composed of both a feedforward controller and a feedback controller, which utilize state information.

[2] In this chapter, the terms "observer" and "estimator" are used to indicate the block of $\mathcal{O}_{\mathsf{i}}$'s and $\mathcal{E}$ in Fig. 4.1, respectively. That is, two terminologies should be distinguished.

estimated number of attacked sensors so that the system can detect the attacks and counteract it in a timely fashion (e.g., giving an attack notification alarm when $f > 0$ and initiating a new attack identification process when $f > \mathsf{q}$).

4.2 Components of Attack-Resilient State Estimator and Their Functions

In this section, we introduce two main components of the attack-resilient state estimator $\mathcal{E}$: the partial observers $\mathcal{O}_{\mathsf{i}}$'s and the decoder $\mathcal{D}$. (See Fig. 4.1.) Then, the basic mechanism on how these components are operating is briefly explained. More precisely, first, the partial observers $\mathcal{O}_{\mathsf{i}}$'s are designed by applying the Kalman detectability (or, observability) decomposition to each sensor output. Second, the previously developed error correction technique in Section 2.2 tailored into this specific problem is then implemented in order to recover the original state variable x and it constitutes the decoder $\mathcal{D}$. By combining the partial observers $\mathcal{O}_{\mathsf{i}}$'s and the decoder $\mathcal{D}$, the final estimator $\mathcal{E}$ is obtained. One key part of the proposed estimator is to construct partial observers by means of the Kalman decomposition. This idea originates from the field of observer design for switched systems [67,86], and it turns out later on that it substantially reduces the number of observers in the final estimator.

4.2.1 Partial Observer: Kalman Detectability Decomposition with Single Sensor

Motivation of the idea to design partial observers $\mathcal{O}_{\mathsf{i}}$'s, begins with the fact that, for conventional observers for the system (4.1.1) and the output (4.1.2) which has the form of

$$\begin{aligned}\hat{x}(k+1) &= A\hat{x}(k) + Bu(k) + L(\bar{y}(k) - C\hat{x}(k)) \\ &= A\hat{x}(k) + Bu(k) + L(Cx(k) + n(k) + a(k) - C\hat{x}(k)),\end{aligned} \tag{4.2.1}$$

the effect of any single non-zero component of $a(k) \in \mathbb{R}^{\mathsf{p}}$ (which is the attack to one sensor channel) may affect all component of $\hat{x}$ (because of L). However, if we employ individual observer to each sensor, then the attack to particular sensor affects only the observer corresponding to the attacked sensor, and we can still maintain a few healthy information processed by the other observers. Of course, each observer may not estimate the full state x in general since the full state x may not be detectable (or, observable) from the output information of one single sensor. Therefore, we introduce the partial observer which estimates only the detectable (or, observable) portion of the full state x.

With only one measurement $y_{\mathsf{i}}(k)$ of the plant (4.1.1), a single-output system is obtained as follows:

$$\mathcal{P}_{\mathsf{i}}: \begin{cases} x(k+1) = Ax(k) + Bu(k) + d(k) \\ y_{\mathsf{i}}(k) = c_{\mathsf{i}}x(k) + n_{\mathsf{i}}(k). \end{cases} \tag{4.2.2}$$

The observability matrix of (4.2.2) which is given by G_{i} in (3.1.4), is used to divide n-dimensional state space $\mathbb{R}^{\mathsf{n}}$ into two subspaces: unobservable subspace $\overline{\mathcal{O}}(A, c_{\mathsf{i}})$ and its orthogonal complement $\overline{\mathcal{O}}(A, c_{\mathsf{i}})^{\perp}$. As seen in Section 3.1.3, the null space of G_{i}, $\mathcal{N}(G_{\mathsf{i}})$, which is A-invariant, becomes the unobservable subspace $\overline{\mathcal{O}}(A, c_{\mathsf{i}})$, and further, the orthogonal complement $\overline{\mathcal{O}}(A, c_{\mathsf{i}})^{\perp}$ of $\overline{\mathcal{O}}(A, c_{\mathsf{i}})$, which is isomorphic to the quotient space $\mathbb{R}^{\mathsf{n}}/\overline{\mathcal{O}}(A, c_{\mathsf{i}})$, is denoted by $\mathcal{O}(A, c_{\mathsf{i}})$. It easily follows from linear algebra that $\mathcal{O}(A, c_{\mathsf{i}}) = \overline{\mathcal{O}}(A, c_{\mathsf{i}})^{\perp} = \mathcal{N}(G_{\mathsf{i}})^{\perp} = \mathcal{R}(G_{\mathsf{i}}^{\top})$. Up to now, we have examined the observability decomposition of the state space $\mathbb{R}^{\mathsf{n}}$. In order to obtain the detectability decomposition, the unobservable subspace is further divided into two subspace: undetectable subspace

$$\overline{\mathcal{D}}(A, c_{\mathsf{i}}) := \overline{\mathcal{O}}(A, c_{\mathsf{i}}) \cap \mathcal{X}_u(A) = \mathcal{N}(G_{\mathsf{i}}) \cap \mathcal{X}_u(A)$$

and its orthogonal complement $\overline{\mathcal{O}}(A, c_{\mathsf{i}}) \cap \overline{\mathcal{D}}(A, c_{\mathsf{i}})^{\perp}$. Here, $\mathcal{X}_u(A)$ denotes the unstable subspace of A which is defined by the subspace spanned by the eigenvectors and generalized eigenvectors corresponding to the unstable eigenvalues of A. Note that $\overline{\mathcal{D}}(A, c_{\mathsf{i}})$ is also A-invariant because both $\mathcal{N}(G_{\mathsf{i}})$ and $\mathcal{X}_u(A)$ are invariant under A. Finally, the n-dimensional state space $\mathbb{R}^{\mathsf{n}}$ is divided into three subspaces: $\mathcal{O}(A, c_{\mathsf{i}})$, $\overline{\mathcal{O}}(A, c_{\mathsf{i}}) \cap \overline{\mathcal{D}}(A, c_{\mathsf{i}})^{\perp}$, and $\overline{\mathcal{D}}(A, c_{\mathsf{i}})$.

To derive a transformation matrix, first, let ν_{i} be the dimension of observable subspace $\mathcal{O}(A, c_\mathsf{i})$, i.e., $\nu_\mathsf{i} := \mathsf{dim}(\mathcal{O}(A, c_\mathsf{i})) = \mathsf{rank}(G_\mathsf{i})$, and $\mu_\mathsf{i} - \nu_\mathsf{i}$ be the dimension of the subspace $\overline{\mathcal{O}}(A, c_\mathsf{i}) \cap \overline{\mathcal{D}}(A, c_\mathsf{i})^\perp$, i.e., $\mu_\mathsf{i} := \nu_\mathsf{i} + \mathsf{dim}(\overline{\mathcal{O}}(A, c_\mathsf{i}) \cap \overline{\mathcal{D}}(A, c_\mathsf{i})^\perp)$. Then the dimension of the undetectable subspace $\overline{\mathcal{D}}(A, c_\mathsf{i})$ is $\mathsf{n} - \mu_\mathsf{i}$. The matrices $Z_\mathsf{i}^o \in \mathbb{R}^{\mathsf{n} \times \nu_\mathsf{i}}$, $Z_\mathsf{i}^d \in \mathbb{R}^{\mathsf{n} \times (\mu_\mathsf{i} - \nu_\mathsf{i})}$, and $W_\mathsf{i} \in \mathbb{R}^{\mathsf{n} \times (\mathsf{n} - \mu_\mathsf{i})}$ are selected such that their columns are orthonormal bases of $\mathcal{O}(A, c_\mathsf{i})$, $\overline{\mathcal{O}}(A, c_\mathsf{i}) \cap \overline{\mathcal{D}}(A, c_\mathsf{i})^\perp$, and $\overline{\mathcal{D}}(A, c_\mathsf{i})$, respectively. Note that the matrix $\begin{bmatrix} Z_\mathsf{i}^o & Z_\mathsf{i}^d & W_\mathsf{i} \end{bmatrix}$ is orthogonal, i.e.,

$$\begin{bmatrix} Z_\mathsf{i}^o & Z_\mathsf{i}^d & W_\mathsf{i} \end{bmatrix}^\top \begin{bmatrix} Z_\mathsf{i}^o & Z_\mathsf{i}^d & W_\mathsf{i} \end{bmatrix} = I_{\mathsf{n} \times \mathsf{n}},$$

and we have

$$\begin{aligned} &Z_\mathsf{i}^{o\top} A Z_\mathsf{i}^d = O_{\nu_\mathsf{i} \times (\mu_\mathsf{i} - \nu_\mathsf{i})}, \quad Z_\mathsf{i}^{o\top} A W_\mathsf{i} = O_{\nu_\mathsf{i} \times (\mathsf{n} - \mu_\mathsf{i})}, \quad Z_\mathsf{i}^{d\top} A W_\mathsf{i} = O_{(\mu_\mathsf{i} - \nu_\mathsf{i}) \times (\mathsf{n} - \mu_\mathsf{i})} \\ &\quad c_\mathsf{i} Z_\mathsf{i}^d = 0_{1 \times (\mu_\mathsf{i} - \nu_\mathsf{i})}, \quad \text{and} \quad c_\mathsf{i} W_\mathsf{i} = 0_{1 \times (\mathsf{n} - \mu_\mathsf{i})}, \end{aligned}$$

from the construction of Z_i^o, Z_i^d, and W_i.

Now, we make the change of state variables as defined by the transformation

$$\begin{bmatrix} z_\mathsf{i}^o \\ z_\mathsf{i}^d \\ w_\mathsf{i} \end{bmatrix} = \begin{bmatrix} Z_\mathsf{i}^{o\top} \\ Z_\mathsf{i}^{d\top} \\ W_\mathsf{i}^\top \end{bmatrix} x. \tag{4.2.3}$$

In terms of this new state, the original single-output system (4.2.2) can be written in the decomposed form of

$$\mathcal{P}_\mathsf{i}' : \begin{cases} \begin{bmatrix} z_\mathsf{i}^o(k+1) \\ z_\mathsf{i}^d(k+1) \\ w_\mathsf{i}(k+1) \end{bmatrix} = \begin{bmatrix} Z_\mathsf{i}^{o\top} A Z_\mathsf{i}^o & O_{\nu_\mathsf{i} \times (\mu_\mathsf{i} - \nu_\mathsf{i})} & O_{\nu_\mathsf{i} \times (\mathsf{n} - \mu_\mathsf{i})} \\ Z_\mathsf{i}^{d\top} A Z_\mathsf{i}^o & Z_\mathsf{i}^{d\top} A Z_\mathsf{i}^d & O_{(\mu_\mathsf{i} - \nu_\mathsf{i}) \times (\mathsf{n} - \mu_\mathsf{i})} \\ W_\mathsf{i}^\top A Z_\mathsf{i}^o & W_\mathsf{i}^\top A Z_\mathsf{i}^d & W_\mathsf{i}^\top A W_\mathsf{i} \end{bmatrix} \begin{bmatrix} z_\mathsf{i}^o(k) \\ z_\mathsf{i}^d(k) \\ w_\mathsf{i}(k) \end{bmatrix} \\ \qquad + \begin{bmatrix} Z_\mathsf{i}^{o\top} B \\ Z_\mathsf{i}^{d\top} B \\ W_\mathsf{i}^\top B \end{bmatrix} u(k) + \begin{bmatrix} Z_\mathsf{i}^{o\top} \\ Z_\mathsf{i}^{d\top} \\ W_\mathsf{i}^\top \end{bmatrix} d(k) \\ y_\mathsf{i}(k) = \begin{bmatrix} c_\mathsf{i} Z_\mathsf{i}^o & 0_{1 \times (\mu_\mathsf{i} - \nu_\mathsf{i})} & 0_{1 \times (\mathsf{n} - \mu_\mathsf{i})} \end{bmatrix} \begin{bmatrix} z_\mathsf{i}^o(k) \\ z_\mathsf{i}^d(k) \\ w_\mathsf{i}(k) \end{bmatrix} + n_\mathsf{i}(k). \end{cases} \tag{4.2.4}$$

Finally, by the Kalman detectability decomposition, the state x is decomposed into the detectable sub-state $\begin{bmatrix} z_{\mathsf{i}}^o \\ z_{\mathsf{i}}^d \end{bmatrix} \in \mathbb{R}^{\mu_{\mathsf{i}}}$ and the undetectable sub-state $w_{\mathsf{i}} \in \mathbb{R}^{\mathsf{n}-\mu_{\mathsf{i}}}$ with the similarity transformation (4.2.3). Or, by the Kalman observability decomposition, the state x is decomposed into the observable sub-state $z_{\mathsf{i}}^o \in \mathbb{R}^{\nu_{\mathsf{i}}}$ and the unobservable sub-state $\begin{bmatrix} z_{\mathsf{i}}^d \\ w_{\mathsf{i}} \end{bmatrix} \in \mathbb{R}^{\mathsf{n}-\nu_{\mathsf{i}}}$ with the same similarity transformation (4.2.3).

Based-on the decomposed form of (4.2.4), one can construct an observer which can asymptotically estimate the detectable sub-state $\begin{bmatrix} z_{\mathsf{i}}^o \\ z_{\mathsf{i}}^d \end{bmatrix} \in \mathbb{R}^{\mu_{\mathsf{i}}}$ since the pair

$$\left(\begin{bmatrix} Z_{\mathsf{i}}^{o\top} A Z_{\mathsf{i}}^o & O_{\nu_{\mathsf{i}} \times (\mu_{\mathsf{i}} - \nu_{\mathsf{i}})} \\ Z_{\mathsf{i}}^{d\top} A Z_{\mathsf{i}}^o & Z_{\mathsf{i}}^{d\top} A Z_{\mathsf{i}}^d \end{bmatrix}, \begin{bmatrix} c_{\mathsf{i}} Z_{\mathsf{i}}^o & 0_{1 \times (\mu_{\mathsf{i}} - \nu_{\mathsf{i}})} \end{bmatrix} \right)$$

is detectable. Of course, one can design an observer which can recover the observable sub-state $z_{\mathsf{i}}^o \in \mathbb{R}^{\nu_{\mathsf{i}}}$ since the pair $\left(Z_{\mathsf{i}}^{o\top} A Z_{\mathsf{i}}^o,\ c_{\mathsf{i}} Z_{\mathsf{i}}^o\right)$ is observable. Let Z_{i} and z_{i} be $\begin{bmatrix} Z_{\mathsf{i}}^o & Z_{\mathsf{i}}^d \end{bmatrix}$ and $\begin{bmatrix} z_{\mathsf{i}}^o \\ z_{\mathsf{i}}^d \end{bmatrix}$, respectively, when the detectability decomposition is in one's mind. Otherwise, let Z_{i} and z_{i} be Z_{i}^o and z_{i}^o, respectively, when the observability decomposition is taken into consideration. By dropping the undetectable sub-state w_{i} (or, unobservable sub-state $\begin{bmatrix} z_{\mathsf{i}}^d \\ w_{\mathsf{i}} \end{bmatrix}$) from (4.2.4), the detectable (or, observable) quotient subsystem of (4.2.4) is obtained as

$$\mathcal{P}_{\mathsf{i}}^d : \begin{cases} z_{\mathsf{i}}(k+1) = S_{\mathsf{i}} z_{\mathsf{i}}(k) + Z_{\mathsf{i}}^\top B u(k) + Z_{\mathsf{i}}^\top d(k) \\ y_{\mathsf{i}}(k) = t_{\mathsf{i}} z_{\mathsf{i}}(k) + n_{\mathsf{i}}(k) \end{cases} \tag{4.2.5}$$

where $S_{\mathsf{i}} := Z_{\mathsf{i}}^\top A Z_{\mathsf{i}}$ and $t_{\mathsf{i}} := c_{\mathsf{i}} Z_{\mathsf{i}}$.

Since the pair $(S_{\mathsf{i}}, t_{\mathsf{i}})$ is detectable (or, observable), one can design an observer which can successfully estimate the detectable (or, observable) portion z_{i} of the full state $x \in \mathbb{R}^{\mathsf{n}}$. Consequently, the main function of each partial observer $\mathcal{O}_{\mathsf{i}}$ is to provide the estimates $\hat{z}_{\mathsf{i}}$ for the detectable sub-state $z_{\mathsf{i}} = \begin{bmatrix} z_{\mathsf{i}}^o \\ z_{\mathsf{i}}^d \end{bmatrix} \in \mathbb{R}^{\mu_{\mathsf{i}}}$ (or, observable sub-state $z_{\mathsf{i}} = z_{\mathsf{i}}^o \in \mathbb{R}^{\nu_{\mathsf{i}}}$). Although we do not specify the type of observers in this section, the details on how to construct such an observer will be discussed in Section 4.3.

4.2.2 Decoder: Error Correction for Stacked Vector

This section discusses the basic functions of the decoder and presents how the decoder operates. Recall that each partial observer $\mathcal{O}_{\mathsf{i}}$ provides the estimates $\hat{z}_{\mathsf{i}}$ for the detectable sub-state $\begin{bmatrix} z_{\mathsf{i}}^o \\ z_{\mathsf{i}}^d \end{bmatrix} \in \mathbb{R}^{\mu_{\mathsf{i}}}$ (or, observable sub-state $z_{\mathsf{i}}^o \in \mathbb{R}^{\nu_{\mathsf{i}}}$). The decoder collects all the data $\hat{z}_{\mathsf{i}}$'s from the partial observers $\mathcal{O}_{\mathsf{i}}$'s and performs the error correction techniques developed in Section 2.2. In order to apply those techniques into the state estimation problem of (4.1.1), first, this problem should be formulated in the form of (2.2.2). To this end, the following equivalence

$$Z_{\mathsf{i}}^\top x = z_{\mathsf{i}}, \qquad {}^\forall \mathsf{i} \in [\mathsf{p}], \tag{4.2.6}$$

which is a direct consequence of (4.2.3), is used. With the state estimation error defined by

$$\tilde{z}_{\mathsf{i}} := \hat{z}_{\mathsf{i}} - z_{\mathsf{i}},$$

one can divide $\tilde{z}_{\mathsf{i}}$ into two parts, v_{i} and e_{i}, so that

$$\tilde{z}_{\mathsf{i}} = v_{\mathsf{i}} + e_{\mathsf{i}},$$

where e_{i} is affected only by the attack a_{i} while v_{i} is induced by all other sources such as the initial estimation error, disturbance d, and noise n_{i}. Details on computing v_{i} and e_{i} will be given in Section 4.3. Appending $\mathsf{n} - \mu_{\mathsf{i}}$ (or, $\mathsf{n} - \nu_{\mathsf{i}}$) zero row vectors, $0_{1\times\mathsf{n}}$, to each $Z_{\mathsf{i}}^\top$ in (4.2.6) and stacking them all, we have the following equation of

$$\begin{bmatrix} Z_1^{\mathsf{n}\top} \\ \vdots \\ Z_{\mathsf{p}}^{\mathsf{n}\top} \end{bmatrix} x(k) = \begin{bmatrix} z_1^{\mathsf{n}}(k) \\ \vdots \\ z_{\mathsf{p}}^{\mathsf{n}}(k) \end{bmatrix} = \begin{bmatrix} \hat{z}_1^{\mathsf{n}}(k) \\ \vdots \\ \hat{z}_{\mathsf{p}}^{\mathsf{n}}(k) \end{bmatrix} - \begin{bmatrix} \tilde{z}_1^{\mathsf{n}}(k) \\ \vdots \\ \tilde{z}_{\mathsf{p}}^{\mathsf{n}}(k) \end{bmatrix} \tag{4.2.7}$$

where

$$Z_{\mathsf{i}}^{\mathsf{n}\top} := \begin{bmatrix} Z_{\mathsf{i}}^{\top} \\ O_{(\mathsf{n}-\mu_{\mathsf{i}})\times\mathsf{n}} \end{bmatrix}, \qquad z_{\mathsf{i}}^{\mathsf{n}}(k) := \begin{bmatrix} z_{\mathsf{i}}(k) \\ 0_{(\mathsf{n}-\mu_{\mathsf{i}})\times 1} \end{bmatrix},$$
$$\hat{z}_{\mathsf{i}}^{\mathsf{n}}(k) := \begin{bmatrix} \hat{z}_{\mathsf{i}}(k) \\ 0_{(\mathsf{n}-\mu_{\mathsf{i}})\times 1} \end{bmatrix}, \qquad \tilde{z}_{\mathsf{i}}^{\mathsf{n}}(k) := \begin{bmatrix} \tilde{z}_{\mathsf{i}}(k) \\ 0_{(\mathsf{n}-\mu_{\mathsf{i}})\times 1} \end{bmatrix}. \tag{4.2.8}$$

This augmentation of zeros is to match the size of each matrix so that it agrees with the n-stacked vector considered in Section 2.2. Finally, (4.2.7) is written in a compact form as

$$\hat{z}(k) = \Phi x(k) + \tilde{z}(k) = \Phi x(k) + v(k) + e(k) \in \mathbb{R}^{\mathsf{np}} \tag{4.2.9}$$

where the coding matrix

$$\Phi := \begin{bmatrix} Z_1^{\mathsf{n}\top} \\ \vdots \\ Z_{\mathsf{p}}^{\mathsf{n}\top} \end{bmatrix} \tag{4.2.10}$$

is composed of the similarity transformation matrices Z_{i}'s and zero elements. It is also supposed that additional zero elements are also appended to $v_{\mathsf{i}}(k)$'s and $e_{\mathsf{i}}(k)$'s as in (4.2.8). Due to the q-sparsity assumption on a (i.e., Assumption 4.1.1) and the fact that e_{i} only depends on a_{i}, it is obvious that $e \in \mathbb{R}^{\mathsf{np}}$ is n-stacked q-sparse. Therefore, the equation (4.2.9) directly matches the error correcting problem of (2.2.2). In conclusion, we can apply those techniques developed in Section 2.2 according to the characteristics of the noise signal $v(k)$, and the results are presented in the next section.

4.3 Design of Attack-Resilient State Estimator

In this section, we have detailed the design of the attack-resilient state estimator $\mathcal{E}$. Control systems are classified into two groups according to the property of the disturbance d and the noise n: bounded disturbance/noise case and Gaussian disturbance/noise case. For each case, a suitable form of observer is adopted for the partial observer $\mathcal{O}_{\mathsf{i}}$, and an appropriate error correction method is applied for a particular context in the decoder $\mathcal{D}$. First, when we have bounded disturbance and noise, the partial observer can be designed by a Luenberger observer and the decoder exploits the error correcting method for the bounded noise case developed in Section 2.2.3.1. Second, when systems suffer from the Gaussian disturbance and noise, one can construct a Kalman filter which serves as the partial observer and the decoder operates based on the error correcting method for the Gaussian noise developed in Section 2.2.3.2.

However, such an attack-resilient estimator design is not possible all the time. The system (4.1.1) should satisfy a certain class of observability and it is shown in Section 3.3 that the redundant detectability is an equivalent condition for the system to be asymptotically observable under sensor attacks. Thus, the following assumption of redundant detectability is made.

Assumption 4.3.1. The pair (A, C) is $2\mathsf{q}$-redundant detectable (or, asymptotically $2\mathsf{q}$-redundant observable). Equivalently, the dynamical system (4.1.1) is detectable under q-sparse sensor attacks (or, asymptotically observable under q-sparse sensor attacks). $\Diamond$

4.3.1 Deterministic Estimator with Bounded Disturbance and Noise

In this section, both the disturbance d and the noise n_i of the system (4.2.5) are supposed to be uniformly bounded as follows.

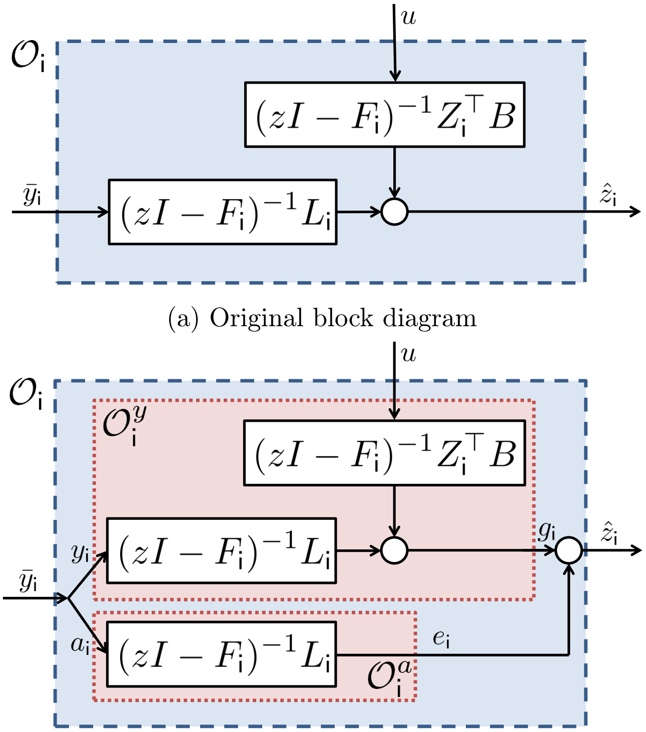
Assumption 4.3.2. The process disturbance d and each measurement noise n_i are uniformly bounded, i.e.,

$$\|d(k)\|_2 \leq d_{\max}, \quad \|n_\mathsf{i}(k)\|_2 \leq n_{\max}, \quad {}^\forall k \geq 0, \quad {}^\forall \mathsf{i} \in [\mathsf{p}]. \qquad \Diamond$$

First, the partial observer $\mathcal{O}_\mathsf{i}$ is designed by a Luenberger observer for the detectable subsystem (4.2.5) which is in the following form of

$$\begin{aligned} \mathcal{O}_\mathsf{i}: \ \hat{z}_\mathsf{i}(k+1) &= S_\mathsf{i}\hat{z}_\mathsf{i}(k) + Z_\mathsf{i}^\top Bu(k) + L_\mathsf{i}\left(\bar{y}_\mathsf{i}(k) - t_\mathsf{i}\hat{z}_\mathsf{i}(k)\right) \\ &=: F_\mathsf{i}\hat{z}_\mathsf{i}(k) + Z_\mathsf{i}^\top Bu(k) + L_\mathsf{i}(y_\mathsf{i}(k) + a_\mathsf{i}(k)) \end{aligned} \tag{4.3.1}$$

where the injection gain L_i is chosen so that $F_\mathsf{i} := S_\mathsf{i} - L_\mathsf{i}t_\mathsf{i}$ is Schur stable. Here, note that $\bar{y}_\mathsf{i}(k)$ is injected instead of $y_\mathsf{i}(k)$ by the attack model (4.1.2). Once we



(a) Original block diagram

(b) Conceptually divided block diagram

Figure 4.4: Configuration of the partial observer $\mathcal{O}_i$.

consider the following systems of

$$\mathcal{O}_i^y: \; g_i(k+1) = F_i g_i(k) + Z_i^\top B u(k) + L_i y_i(k) \tag{4.3.2a}$$

$$\mathcal{O}_i^a: \; e_i(k+1) = F_i e_i(k) + L_i a_i(k) \tag{4.3.2b}$$

where $g_i(0) = \hat{z}_i(0)$ and $e_i(0) = 0_{\mu_i \times 1}$, it is easy to check that $\hat{z}_i = g_i + e_i$. Fig. 4.4 depicts the relationship between the observers (4.3.1) and (4.3.2). Now, define the attack-free estimation error $v_i := g_i - z_i$, and it trivially follows that

$$v_i(k+1) = F_i v_i(k) + L_i n_i(k) - Z_i^\top d(k). \tag{4.3.3}$$

The final state estimation error defined by $\tilde{z}_i := \hat{z}_i - z_i$, satisfies

$$\tilde{z}_i(k) = v_i(k) + e_i(k), \tag{4.3.4}$$

and its dynamic equation is governed by

$$\mathcal{F}_{\mathsf{i}} : \tilde{z}_{\mathsf{i}}(k+1) = F_{\mathsf{i}}\tilde{z}_{\mathsf{i}}(k) + L_{\mathsf{i}}n_{\mathsf{i}}(k) - Z_{\mathsf{i}}^{\top}d(k) + L_{\mathsf{i}}a_{\mathsf{i}}(k). \tag{4.3.5}$$

The dynamics (4.3.3) and (4.3.2b) with initial conditions $v_{\mathsf{i}}(0) = \tilde{z}_{\mathsf{i}}(0)$ and $e_{\mathsf{i}}(0) = 0_{\mu_{\mathsf{i}}\times 1}$, ensure that

$$v_{\mathsf{i}}(k) := F_{\mathsf{i}}^{k}\tilde{z}_{\mathsf{i}}(0) + \sum_{j=0}^{k-1} F_{\mathsf{i}}^{k-1-j}\left(L_{\mathsf{i}}n_{\mathsf{i}}(j) - Z_{\mathsf{i}}^{\top}d(j)\right), \tag{4.3.6a}$$

$$e_{\mathsf{i}}(k) := \sum_{j=0}^{k-1} F_{\mathsf{i}}^{k-1-j}L_{\mathsf{i}}a_{\mathsf{i}}(j). \tag{4.3.6b}$$

Here, the attack induced estimation error vector $e_{\mathsf{i}}(k)$ may have arbitrary values.

For all $k \geq 0$ and $\mathsf{i} \in [\mathsf{p}]$, there exist $\eta_F \geq 1$ and $0 < \beta < 1$ such that $\|F_{\mathsf{i}}^{k}\|_2 \leq \eta_F\beta^k$ since F_{i} is Schur stable. In addition, for some η_L and η_Z, it holds that $\|F_{\mathsf{i}}^{k}L_{\mathsf{i}}\|_2 \leq \eta_L\beta^k$ and $\|F_{\mathsf{i}}^{k}Z_{\mathsf{i}}^{\top}\|_2 \leq \eta_Z\beta^k$. Then, one can easily show that

$$\|v_{\mathsf{i}}(k)\|_2 \leq \eta_F\|\tilde{z}_{\mathsf{i}}(0)\|_2\beta^k + w_{\max} \leq v_{\max}(k) \tag{4.3.7}$$

where

$$w_{\max} := \frac{\eta_L n_{\max} + \eta_Z d_{\max}}{1-\beta},$$
$$v_{\max}(k) := \max_{\mathsf{i}\in[\mathsf{p}]}\left\{\eta_F\|\tilde{z}_{\mathsf{i}}(0)\|_2\beta^k + w_{\max}\right\}.$$

As k increases, $v_{\max}(k)$ converges to $w_{\max}$.

Then, the decoder collects all the data $\hat{z}_{\mathsf{i}}$'s from the partial observers $\mathcal{O}_{\mathsf{i}}$'s and the problem of estimating $x(k)$ is formulated in the form of (4.2.9) as we have seen in the previous section. Since (4.2.9) exactly matches with (2.2.2) where the static error correcting problem is considered, one can directly apply the error correction technique developed in Section 2.2.3.1 into (4.2.9). Theorems 2.2.11 and 2.2.12 are mainly employed so as to recover $x(k)$. However, before applying them, one should check that three conditions on those theorems are satisfied for the given system (4.1.1): boundedness of $v(k)$, q-sparsity of $e(k)$, and q-error correctability of Φ. The first two conditions are easily satisfied by Assumptions

4.3.2 and 4.1.1. That is, the noise vector $v_{\mathsf{i}}(k)$ is bounded by $v_{\max}(k)$ for all $\mathsf{i} \in [\mathsf{p}]$ by (4.3.7), which is induced from Assumption 4.3.2. Since the error vector $e_{\mathsf{i}}(k)$ depends only on the attack element $a_{\mathsf{i}}(j)$ for $0 \leq j \leq k-1$ by (4.3.6b) and $a(j)$ is q-sparse according to Assumption 4.1.1, the vector $e(k)$ is (n-stacked) q-sparse. The last condition, the q-error correctability of Φ, is actually fulfilled by the $2\mathsf{q}$-redundant detectability of the system (4.1.1) (i.e., Assumption 4.3.1), as asserted in the following proposition. Therefore, all three conditions on Theorems 2.2.11 and 2.2.12 hold.

Proposition 4.3.1. The followings are equivalent:
(i) The pair (A, C) is $2\mathsf{q}$-redundant detectable (or, asymptotically $2\mathsf{q}$-redundant observable);
(iii) The matrix $\Phi \in \mathbb{R}^{\mathsf{np} \times \mathsf{n}}$ given in (4.2.10) is (n-stacked) q-error correctable. $\Diamond$

Proof. Pick any $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - 2\mathsf{q}$, and we claim $\mathcal{N}(G_{\Lambda^{\mathsf{n}}}) \subset \mathcal{X}_s(A)$ if and only if $\mathcal{N}(\Phi_{\Lambda^{\mathsf{n}}}) = \{0\}$. This can be shown by the following equivalences

$$\begin{aligned}
\mathcal{N}(G_{\Lambda^{\mathsf{n}}}) \subset \mathcal{X}_s(A) \quad &\Leftrightarrow \quad \mathcal{N}(G_{\Lambda^{\mathsf{n}}}) \cap \mathcal{X}_u(A) = \{0\} \\
&\Leftrightarrow \quad \left(\bigcap_{\mathsf{i} \in \Lambda} \mathcal{N}(G_{\mathsf{i}}) \right) \cap \mathcal{X}_u(A) = \{0\} \\
&\Leftrightarrow \quad \bigcap_{\mathsf{i} \in \Lambda} \left(\mathcal{N}(G_{\mathsf{i}}) \cap \mathcal{X}_u(A) \right) = \{0\} \\
&\Leftrightarrow \quad \bigcap_{\mathsf{i} \in \Lambda} \left(\overline{\mathcal{O}}(A, c_{\mathsf{i}}) \cap \mathcal{X}_u(A) \right) = \{0\} \\
&\Leftrightarrow \quad \bigcap_{\mathsf{i} \in \Lambda} \overline{\mathcal{D}}(A, c_{\mathsf{i}}) = \{0\} \\
&\Leftrightarrow \quad \bigcap_{\mathsf{i} \in \Lambda} \mathcal{R}(W_{\mathsf{i}}) = \{0\} \\
&\Leftrightarrow \quad \bigcap_{\mathsf{i} \in \Lambda} \mathcal{N}\left(\begin{bmatrix} {Z_{\mathsf{i}}^{o}}^{\top} \\ {Z_{\mathsf{i}}^{d}}^{\top} \end{bmatrix} \right) = \{0\} \\
&\Leftrightarrow \quad \bigcap_{\mathsf{i} \in \Lambda} \mathcal{N}({Z_{\mathsf{i}}}^{\top}) = \{0\} \\
&\Leftrightarrow \quad \mathcal{N}(\Phi_{\Lambda^{\mathsf{n}}}) = \{0\} ,
\end{aligned}$$

where the structure of Φ and its elements ${Z_{\mathsf{i}}}^{\top}$'s from the Kalman detectability

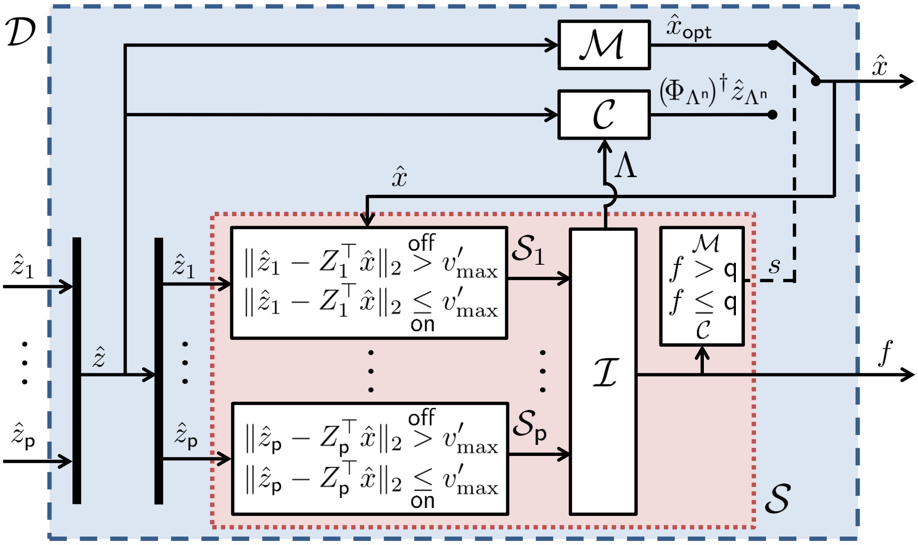

$\mathcal{D}$
$\mathcal{M}$
$\hat{x}_{\mathsf{opt}}$
$\hat{x}$
$\mathcal{C}$
$(\Phi_{\Lambda^{\mathsf{n}}})^{\dagger}\hat{z}_{\Lambda^{\mathsf{n}}}$
Λ
$\hat{x}$
$\hat{z}_1$
$\hat{z}$
$\hat{z}_{\mathsf{p}}$
$\hat{z}_1$
$\hat{z}_{\mathsf{p}}$
$\|\hat{z}_1 - Z_1^{\top}\hat{x}\|_2 \overset{\mathsf{off}}{>} v'_{\max}$
$\|\hat{z}_1 - Z_1^{\top}\hat{x}\|_2 \underset{\mathsf{on}}{\leq} v'_{\max}$
$\mathcal{S}_1$
$\|\hat{z}_{\mathsf{p}} - Z_{\mathsf{p}}^{\top}\hat{x}\|_2 \overset{\mathsf{off}}{>} v'_{\max}$
$\|\hat{z}_{\mathsf{p}} - Z_{\mathsf{p}}^{\top}\hat{x}\|_2 \underset{\mathsf{on}}{\leq} v'_{\max}$
$\mathcal{S}_{\mathsf{p}}$
$\mathcal{I}$
$\mathcal{M}$
$f > \mathsf{q}$
$f \leq \mathsf{q}$
$\mathcal{C}$
s
f
$\mathcal{S}$

Figure 4.5: Configuration of the decoder $\mathcal{D}$ with bounded disturbance/noise.

decomposition, are used. □

The decoder's configuration is sketched in Fig. 4.5 and its operation is described in Algorithm 4.1. Initially, an attack-free index set Λ, a state estimate $\hat{x}$, and a fault count signal f, are set to $[\mathsf{p}]$, $\Phi^{\dagger}\hat{z}$, and 0, respectively. With the incoming data $\hat{z}_{\mathsf{i}}$ and information of $\hat{x}$, each selector module $\mathcal{S}_{\mathsf{i}}$ compares $\|\hat{z}_{\mathsf{i}} - Z_{\mathsf{i}}^{\top}\hat{x}\|_2$ with $v'_{\max}$ given in (2.2.9), and provides the on-off signal based on the value of $\|\hat{z}_{\mathsf{i}} - Z_{\mathsf{i}}^{\top}\hat{x}\|_2$. The signal is "on" if $\|\hat{z}_{\mathsf{i}} - Z_{\mathsf{i}}^{\top}\hat{x}\|_2 \leq v'_{\max}$, and "off" otherwise. The new index set Λ^+ corresponding to "on" signals, more precisely, $\Lambda^+ := \left\{\mathsf{i} \in [\mathsf{p}] : \|\hat{z}_{\mathsf{i}} - Z_{\mathsf{i}}^{\top}\hat{x}\|_2 \leq v'_{\max}\right\}$, is obtained and the index generator $\mathcal{I}$ counts the number of "off" signals which becomes the new fault count signal f^+, that is, $f^+ := \mathsf{p} - |\Lambda^+|$. Now, Λ and f are updated to new Λ^+ and f^+, respectively. There are two cases according to the fault count signal f: first, when $f \leq \mathsf{q}$, the switch s is placed on the calculator $\mathcal{C}$'s side and $\mathcal{C}$ computes the state estimates $\hat{x}$ by $(\Phi_{\Lambda^{\mathsf{n}}})^{\dagger}\,\hat{z}_{\Lambda^{\mathsf{n}}}$; second, if $f > \mathsf{q}$, the switch s is placed on the minimizer $\mathcal{M}$'s side and $\mathcal{M}$ solves the optimization problem (2.2.8′) to generate $\hat{x}_{\mathsf{opt}}$ which becomes the state estimates $\hat{x}$.

During the operation of the decoder, the monitoring scheme that the selector $\mathcal{S}$ and the switch s perform, is running on the basis of Theorem 2.2.12, while the

Algorithm 4.1 Operation of the decoder with bounded disturbance/noise

```
Input: $\hat{z}_1, \hat{z}_2, \cdots, \hat{z}_\mathsf{p}$
Output: $\hat{x}$, $f$
Initialization: $\Lambda = [\mathsf{p}]$, $\hat{x} = \Phi^\dagger \hat{z}$, $f = 0$
while system (4.1.1) is running do
    each selector $\mathcal{S}_\mathsf{i}$ compares $\|\hat{z}_\mathsf{i} - Z_\mathsf{i}^\top \hat{x}\|_2$ with $v'_{\max}$
    index generator $\mathcal{I}$ collects i's s.t. $\|\hat{z}_\mathsf{i} - Z_\mathsf{i}^\top \hat{x}\|_2 \le v'_{\max}$ and updates $\Lambda$ and $f$
        by $\Lambda = \{\mathsf{i} \in [\mathsf{p}] : \|\hat{z}_\mathsf{i} - Z_\mathsf{i}^\top \hat{x}\|_2 \le v'_{\max}\}$ and $f = \mathsf{p} - |\Lambda|$
    if $f \le \mathsf{q}$ then
        switch $s$ selects the line from $\mathcal{C}$
        Calculator $\mathcal{C}$ computes $\hat{x} = (\Phi_{\Lambda^\mathsf{n}})^\dagger \, \hat{z}_{\Lambda^\mathsf{n}}$
    else if $f > \mathsf{q}$ then
        switch $s$ selects the line from $\mathcal{M}$
        minimizer $\mathcal{M}$ solves (2.2.8′) and produces $\hat{x} = \hat{x}_\mathsf{opt}$
    end if
end while
```

calculator $\mathcal{C}$ and the minimizer $\mathcal{M}$ has its roots on Theorems 2.2.9 and 2.2.11, respectively. If $f \le \mathsf{q}$, the successful state estimation is ensured by Theorem 2.2.12.(i). More specifically, we have $\|\hat{x} - x\|_2 \le \kappa^c_{\mathsf{p},\mathsf{q},\mathsf{r}}(\Phi) v_{\max}$. In this case, the index set Λ is assumed to be attack-free, and hence, the calculator $\mathcal{C}$ can recover the original state x approximately by $\hat{x} = (\Phi_{\Lambda^\mathsf{n}})^\dagger \, \hat{z}_{\Lambda^\mathsf{n}}$, which is attributed to Theorem 2.2.9. On the other hand, if $f > \mathsf{q}$, the state estimates $\hat{x}$ is not close enough to the original states x by Theorem 2.2.12.(ii). Hence, the algorithm goes to the minimizer step (i.e., the switch s chooses the minimizer $\mathcal{M}$'s side) to figure out new healthy sensors and the state estimates $\hat{x}$ by $\hat{x}_\mathsf{opt}$. Furthermore, Theorem 2.2.11 guarantees that $\|\hat{x} - x\|_2 \le \kappa^c_{\mathsf{p},\mathsf{q},\mathsf{r}}(\Phi) v_{\max}$. These results are summarized in the following theorem.

Theorem 4.3.2. Under Assumptions 4.3.2, 4.1.1, and 4.3.1, the estimator $\mathcal{E}$ equipped with the observers $\mathcal{O}_\mathsf{i}$'s given by (4.3.1) and the decoder $\mathcal{D}$ employing Algorithm 1, guarantees that

$$\|\hat{x}(k) - x(k)\|_2 \le \kappa^c_{\mathsf{p},\mathsf{q},\mathsf{r}}(\Phi) \; v_{\max}(k), \quad {}^\forall k \ge 0.$$

Furthermore, for any $\delta > 0$, there exists $T(\delta) > 0$ such that

$$\|\hat{x}(k) - x(k)\|_2 \leq \kappa^c_{\mathsf{p},\mathsf{q},\mathsf{r}}(\Phi)\ w_{\max} + \delta, \quad ^\forall k \geq T(\delta). \qquad \Diamond$$

4.3.2 Suboptimal Estimator with Gaussian Disturbance and Noise

Gaussian process disturbance $d(k)$ and measurement noise $n(k)$ are considered in this section, and a suboptimal estimator is developed. We first design a decentralized Kalman filter with each single sensor output. This decentralized Kalman filter constitutes the partial observer $\mathcal{O}_{\mathsf{i}}$. Then, an information fusion scheme collects all the information on state estimates and error covariance matrices from the decentralized Kalman filter, as the decoder $\mathcal{D}$ does in the previous section. Now, the information fusion scheme selects a subset of sensors which is most likely to be attack-free by the ML decision rule. Finally, it computes the optimal (i.e., in the sense of MVUE or WLSE) estimates only with those sensors which are identified as the most likely to be attack-free. To this end, stochastic assumptions on the disturbance $d(k)$, the noise $n(k)$, and the initial state $x(0)$ of the system (4.1.1) are formally stated as follows.

Assumption 4.3.3. The disturbance $d(k)$ and measurement noise $n(k)$ are independent and identically distributed (i.i.d.) white Gaussian process with zero mean and covariance matrices Q and R, respectively, i.e.,

$$\begin{aligned}
d(k) &\sim N(0, Q), \\
n(k) &\sim N(0, R), \\
\mathbf{E}[d(k)] = 0, &\quad \mathbf{E}[d(k)d^\top(t)] = Q\delta_{kt}, \\
\mathbf{E}[n(k)] = 0, &\quad \mathbf{E}[n(k)n^\top(t)] = R\delta_{kt}, \\
&\quad \mathbf{E}[n(k)d^\top(t)] = 0,
\end{aligned}$$

where δ_{kt} is the Kronecker delta function. Furthermore, the initial state $x(0)$ is a Gaussian distributed random variable with mean $\bar{x}_0$ and covariance matrix P_0

independent of $d(k)$ and $n(k)$, i.e.,

$$\begin{aligned} x(0) &\sim N(\bar{x}_0, P_0), \\ \mathbf{E}[x(0)] = \bar{x}_0, &\quad \mathbf{E}[(x(0)-\bar{x}_0)(x(0)-\bar{x}_0)^\top] = P_0. \end{aligned} \qquad \lozenge$$

With the covariance matrix R of the measurement noise $n(k)$ partitioned as

$$R = \begin{bmatrix} R_1 & R_{12} & \cdots & R_{1\mathsf{p}} \\ R_{21} & R_2 & \cdots & R_{2\mathsf{p}} \\ \vdots & \vdots & \ddots & \vdots \\ R_{\mathsf{p}1} & R_{\mathsf{p}2} & \cdots & R_{\mathsf{p}} \end{bmatrix},$$

the assumption above can also be written for each measurement noise $n_{\mathsf{i}}(k)$ for $\mathsf{i} \in [\mathsf{p}]$, as follows:

$$\begin{aligned} n_{\mathsf{i}}(k) &\sim N(0, R_{\mathsf{i}}), \\ \mathbf{E}[n_{\mathsf{i}}(k)] = 0, &\quad \mathbf{E}[n_{\mathsf{i}}(k)n_{\mathsf{i}}^\top(t)] = R_{\mathsf{i}}\delta_{kt}, \\ &\quad \mathbf{E}[n_{\mathsf{i}}(k)n_{\mathsf{j}}^\top(t)] = R_{\mathsf{ij}}\delta_{kt}, \quad \text{if } \mathsf{i} \neq \mathsf{j}, \\ &\quad \mathbf{E}[n_{\mathsf{i}}(k)d^\top(t)] = 0. \end{aligned}$$

First, the partial observer $\mathcal{O}_{\mathsf{i}}$ is designed by a Kalman filter for the detectable subsystem (4.2.5) with the attack model (4.1.2). To this end, let $\hat{z}_{\mathsf{i}}(k|k-1)$ be the estimate of $z_{\mathsf{i}}(k)$ based on observations from $\bar{y}(0)$ to $\bar{y}(k-1)$. Similarly, $\hat{z}_{\mathsf{i}}(k|k)$ is the estimate of $z_{\mathsf{i}}(k)$ after we process the measurement $\bar{y}(k)$ at time k. Then, the Kalman filter has the following form of

$$\begin{aligned} \mathcal{O}_{\mathsf{i}} :\ & \hat{z}_{\mathsf{i}}(k+1|k+1) \qquad\qquad (4.3.8) \\ =\ & S_{\mathsf{i}}\hat{z}_{\mathsf{i}}(k|k) + Z_{\mathsf{i}}^\top Bu(k) + K_{\mathsf{i}}(k+1)\left(\bar{y}_{\mathsf{i}}(k+1) - t_{\mathsf{i}}\left(S_{\mathsf{i}}\hat{z}_{\mathsf{i}}(k|k) + Z_{\mathsf{i}}^\top Bu(k)\right)\right) \\ =\ & (I - K_{\mathsf{i}}(k+1)t_{\mathsf{i}})\left(S_{\mathsf{i}}\hat{z}_{\mathsf{i}}(k|k) + Z_{\mathsf{i}}^\top Bu(k)\right) + K_{\mathsf{i}}(k+1)\bar{y}_{\mathsf{i}}(k+1) \\ =\ & (I - K_{\mathsf{i}}(k+1)t_{\mathsf{i}})\left(S_{\mathsf{i}}\hat{z}_{\mathsf{i}}(k|k) + Z_{\mathsf{i}}^\top Bu(k)\right) + K_{\mathsf{i}}(k+1)(y_{\mathsf{i}}(k+1) + a_{\mathsf{i}}(k+1)) \end{aligned}$$

where

$$\hat{z}_{\mathsf{i}}(k+1|k+1) = \hat{z}_{\mathsf{i}}(k+1|k) + K(k+1)\left(\bar{y}_{\mathsf{i}}(k+1) - t_{\mathsf{i}}\hat{z}_{\mathsf{i}}(k+1|k)\right) \tag{4.3.9a}$$

$$\hat{z}_{\mathsf{i}}(k+1|k) = S_{\mathsf{i}}\hat{z}_{\mathsf{i}}(k|k) + Z_{\mathsf{i}}^{\top}Bu(k) \tag{4.3.9b}$$

$$K_{\mathsf{i}}(k+1) = P_{\mathsf{i}}(k+1|k)t_{\mathsf{i}}^{\top}\left(t_{\mathsf{i}}P_{\mathsf{i}}(k+1|k)t_{\mathsf{i}}^{\top} + R_{\mathsf{i}}\right)^{-1} \tag{4.3.9c}$$

$$P_{\mathsf{i}}(k+1|k) = S_{\mathsf{i}}P_{\mathsf{i}}(k|k)S_{\mathsf{i}}^{\top} + Z_{\mathsf{i}}^{\top}QZ_{\mathsf{i}} \tag{4.3.9d}$$

$$P_{\mathsf{i}}(k+1|k+1) = (I - K_{\mathsf{i}}(k+1)t_{\mathsf{i}})P_{\mathsf{i}}(k+1|k) \tag{4.3.9e}$$

with

$$\hat{z}_{\mathsf{i}}(0|-1) = Z_{\mathsf{i}}^{\top}\bar{x}_0, \quad P_{\mathsf{i}}(0|-1) = Z_{\mathsf{i}}^{\top}P_0Z_{\mathsf{i}}.$$

Here, note that $\bar{y}_{\mathsf{i}}(k)$ is injected instead of $y_{\mathsf{i}}(k)$ by the attack model (4.1.2). As we have done in the previous section, $\mathcal{O}_{\mathsf{i}}$ can be divided into two parts with the following definitions of

$$\begin{aligned}
g_{\mathsf{i}}(k+1|k) &:= S_{\mathsf{i}}g_{\mathsf{i}}(k|k) + Z_{\mathsf{i}}^{\top}Bu(k),\\
e_{\mathsf{i}}(k+1|k) &:= S_{\mathsf{i}}e_{\mathsf{i}}(k|k),\\
g_{\mathsf{i}}(k+1|k+1) &:= g_{\mathsf{i}}(k+1|k) + K(k+1)\left(y_{\mathsf{i}}(k+1) - t_{\mathsf{i}}g_{\mathsf{i}}(k+1|k)\right),\\
e_{\mathsf{i}}(k+1|k+1) &:= e_{\mathsf{i}}(k+1|k) + K(k+1)\left(a_{\mathsf{i}}(k+1) - t_{\mathsf{i}}e_{\mathsf{i}}(k+1|k)\right).
\end{aligned}$$

By setting the initial conditions as $g_{\mathsf{i}}(0|-1) = \hat{z}_{\mathsf{i}}(0|-1) = Z_{\mathsf{i}}^{\top}\bar{x}_0$ and $e_{\mathsf{i}}(0|-1) = 0_{\mu_{\mathsf{i}}\times 1}$, it easily follows from (4.3.9a) and (4.3.9b) that

$$\begin{aligned}
\hat{z}_{\mathsf{i}}(k+1|k) &= g_{\mathsf{i}}(k+1|k) + e_{\mathsf{i}}(k+1|k),\\
\hat{z}_{\mathsf{i}}(k+1|k+1) &= g_{\mathsf{i}}(k+1|k+1) + e_{\mathsf{i}}(k+1|k+1).
\end{aligned}$$

Finally, $\mathcal{O}_{\mathsf{i}}$ in (4.3.8) is divided into $\mathcal{O}_{\mathsf{i}}^{y}$ and $\mathcal{O}_{\mathsf{i}}^{a}$, as follows:

$$\begin{aligned}
\mathcal{O}_{\mathsf{i}}^{y}:\ g_{\mathsf{i}}(k+1|k+1) &= (I - K_{\mathsf{i}}(k+1)t_{\mathsf{i}})\left(S_{\mathsf{i}}g_{\mathsf{i}}(k|k) + Z_{\mathsf{i}}^{\top}Bu(k)\right)\\
&\quad + K_{\mathsf{i}}(k+1)y_{\mathsf{i}}(k+1),
\end{aligned} \tag{4.3.10a}$$

$$\mathcal{O}_{\mathsf{i}}^{a}:\ e_{\mathsf{i}}(k+1|k+1) = (I - K_{\mathsf{i}}(k+1)t_{\mathsf{i}})S_{\mathsf{i}}e_{\mathsf{i}}(k|k) + K_{\mathsf{i}}(k+1)a_{\mathsf{i}}(k+1). \tag{4.3.10b}$$

Now, define the attack-free estimation error

$$\begin{aligned} v_{\mathsf{i}}(k+1|k) &:= g_{\mathsf{i}}(k+1|k) - z_{\mathsf{i}}(k+1) \\ v_{\mathsf{i}}(k+1|k+1) &:= g_{\mathsf{i}}(k+1|k+1) - z_{\mathsf{i}}(k+1) \end{aligned}$$

and we have that

$$\begin{aligned} v_{\mathsf{i}}(k+1|k) &= \Big(S_{\mathsf{i}} g_{\mathsf{i}}(k|k) + Z_{\mathsf{i}}^{\top} Bu(k)\Big) - \Big(S_{\mathsf{i}} z_{\mathsf{i}}(k) + Z_{\mathsf{i}}^{\top} Bu(k) + Z_{\mathsf{i}}^{\top} d(k)\Big) \\ &= S_{\mathsf{i}} v_{\mathsf{i}}(k|k) - Z_{\mathsf{i}}^{\top} d(k) \qquad (4.3.11a) \\ v_{\mathsf{i}}(k+1|k+1) &= (I - K_{\mathsf{i}}(k+1)t_{\mathsf{i}}) v_{\mathsf{i}}(k+1|k) + K_{\mathsf{i}}(k+1) n_{\mathsf{i}}(k+1) \qquad (4.3.11b) \\ &= (I - K_{\mathsf{i}}(k+1)t_{\mathsf{i}}) S_{\mathsf{i}} v_{\mathsf{i}}(k|k) + K_{\mathsf{i}}(k+1) n_{\mathsf{i}}(k+1) \\ &\quad - (I - K_{\mathsf{i}}(k+1)t_{\mathsf{i}}) Z_{\mathsf{i}}^{\top} d(k). \qquad (4.3.11c) \end{aligned}$$

The final state estimation error defined by

$$\tilde{z}_{\mathsf{i}}(k|k) := \hat{z}_{\mathsf{i}}(k|k) - z_{\mathsf{i}}(k),$$

satisfies

$$\tilde{z}_{\mathsf{i}}(k|k) = v_{\mathsf{i}}(k|k) + e_{\mathsf{i}}(k|k), \qquad (4.3.12)$$

and, from (4.3.10b) and (4.3.11c), its dynamic equation is governed by

$$\begin{aligned} \mathcal{F}_{\mathsf{i}}: \ \tilde{z}_{\mathsf{i}}(k+1|k+1) &= (I - K_{\mathsf{i}}(k+1)t_{\mathsf{i}}) S_{\mathsf{i}} \tilde{z}_{\mathsf{i}}(k|k) + K_{\mathsf{i}}(k+1) n_{\mathsf{i}}(k+1) \\ &\quad - (I - K_{\mathsf{i}}(k+1)t_{\mathsf{i}}) Z_{\mathsf{i}}^{\top} d(k) + K_{\mathsf{i}}(k+1) a_{\mathsf{i}}(k+1). \end{aligned} \qquad (4.3.13)$$

Recall that, in conventional Kalman filter theory, the term $P_{\mathsf{i}}(k|k)$ is used to denote the covariance of the estimation error of $\hat{z}_{\mathsf{i}}(k|k)$ when there is no attack. Since $\hat{z}_{\mathsf{i}}(k|k)$ with $a_{\mathsf{i}}(k) \equiv 0$ is the same as $g_{\mathsf{i}}(k|k)$ by its construction, $P_{\mathsf{i}}(k|k)$ can be thought of as

$$P_{\mathsf{i}}(k|k) = \mathbf{E}[(g_{\mathsf{i}}(k|k) - z_{\mathsf{i}}(k))(g_{\mathsf{i}}(k|k) - z_{\mathsf{i}}(k))^{\top}] = \mathbf{E}[v_{\mathsf{i}}(k|k) v_{\mathsf{i}}^{\top}(k|k)].$$

and equations (4.3.9d) and (4.3.9e) ensures that the covariance matrix $P_{\mathsf{i}}(k|k)$ is in the following recursive form:

$$\mathcal{L}_{\mathsf{i}}: \; P_{\mathsf{i}}(k+1|k+1) = (I - K_{\mathsf{i}}(k+1)t_{\mathsf{i}})(S_{\mathsf{i}}P_{\mathsf{i}}(k|k)S_{\mathsf{i}}^{\top} + Z_{\mathsf{i}}^{\top}QZ_{\mathsf{i}}) \tag{4.3.14}$$

where the initial value $P_{\mathsf{i}}(0|0)$ can be calculated by substituting $P_{\mathsf{i}}(0|-1) = Z_{\mathsf{i}}^{\top}P_0Z_{\mathsf{i}}$ for (4.3.9e) and (4.3.9c). The errors $v_{\mathsf{i}}(k|k)$ and $v_{\mathsf{j}}(k|k)$ for $\mathsf{i} \neq \mathsf{j}$ may be correlated, and thus, by using (4.3.11c), the error cross-covariance between $v_{\mathsf{i}}(k|k)$ and $v_{\mathsf{j}}(k|k)$ can be computed recursively. From the recursive form of (4.3.11c), note that $v_{\mathsf{i}}(k|k)$ is a linear combination of elements in

$$\{v_{\mathsf{i}}(0|0), d(0), \cdots, d(k-1), n_{\mathsf{i}}(0), \cdots, n_{\mathsf{i}}(k)\}.$$

Note that, from Assumption 4.3.3, (i) $n_{\mathsf{i}}(k+1)$ and $d(k)$ are orthogonal, (ii) $v_{\mathsf{i}}(k|k)$ and $d(k)$ are orthogonal, and (iii) $v_{\mathsf{i}}(k|k)$ and $n_{\mathsf{j}}(k+1)$ are orthogonal. Using these facts, one can derive the recursive form of the error cross covariance between $v_{\mathsf{i}}(k|k)$ and $v_{\mathsf{j}}(k|k)$ as follows:

$$\begin{aligned}
\mathcal{L}_{\mathsf{ij}}: \; & P_{\mathsf{ij}}(k+1|k+1) = \mathbf{E}[v_{\mathsf{i}}(k+1|k+1)v_{\mathsf{j}}^{\top}(k+1|k+1)] \\
&= (I - K_{\mathsf{i}}(k+1)t_{\mathsf{i}})\left(S_{\mathsf{i}}\mathbf{E}[v_{\mathsf{i}}(k|k)v_{\mathsf{j}}^{\top}(k|k)]S_{\mathsf{j}}^{\top} + Z_{\mathsf{i}}^{\top}QZ_{\mathsf{j}}\right)(I - K_{\mathsf{j}}(k+1)t_{\mathsf{j}})^{\top} \\
&\quad + K_{\mathsf{i}}(k+1)\mathbf{E}[n_{\mathsf{i}}(k+1)n_{\mathsf{j}}^{\top}(k+1)]K_{\mathsf{j}}^{\top}(k+1) \qquad (4.3.15) \\
&= (I - K_{\mathsf{i}}(k+1)t_{\mathsf{i}})\left(S_{\mathsf{i}}P_{\mathsf{ij}}(k|k)S_{\mathsf{j}}^{\top} + Z_{\mathsf{i}}^{\top}QZ_{\mathsf{j}}\right)(I - K_{\mathsf{j}}(k+1)t_{\mathsf{j}})^{\top} \\
&\quad + K_{\mathsf{i}}(k+1)R_{\mathsf{ij}}K_{\mathsf{j}}^{\top}(k+1).
\end{aligned}$$

The initial value $P_{\mathsf{ij}}(0|0)$ can be calculated by $P_{\mathsf{ij}}(0|-1) = Z_{\mathsf{i}}^{\top}P_0Z_{\mathsf{j}}$ and the recursive form of

$$\begin{aligned}
& P_{\mathsf{ij}}(k+1|k+1) = \mathbf{E}[v_{\mathsf{i}}(k+1|k+1)v_{\mathsf{j}}^{\top}(k+1|k+1)] \\
&= (I - K_{\mathsf{i}}(k+1)t_{\mathsf{i}})\mathbf{E}[v_{\mathsf{i}}(k+1|k)v_{\mathsf{j}}^{\top}(k+1|k)](I - K_{\mathsf{j}}(k+1)t_{\mathsf{j}})^{\top} \\
&\quad + K_{\mathsf{i}}(k+1)\mathbf{E}[n_{\mathsf{i}}(k+1)n_{\mathsf{j}}^{\top}(k+1)]K_{\mathsf{j}}^{\top}(k+1) \\
&= (I - K_{\mathsf{i}}(k+1)t_{\mathsf{i}})P_{\mathsf{ij}}(k+1|k)(I - K_{\mathsf{j}}(k+1)t_{\mathsf{j}})^{\top} + K_{\mathsf{i}}(k+1)R_{\mathsf{ij}}K_{\mathsf{j}}^{\top}(k+1),
\end{aligned}$$

where the second equality is obtained by (4.3.11b) and the orthogonal properties mentioned above.

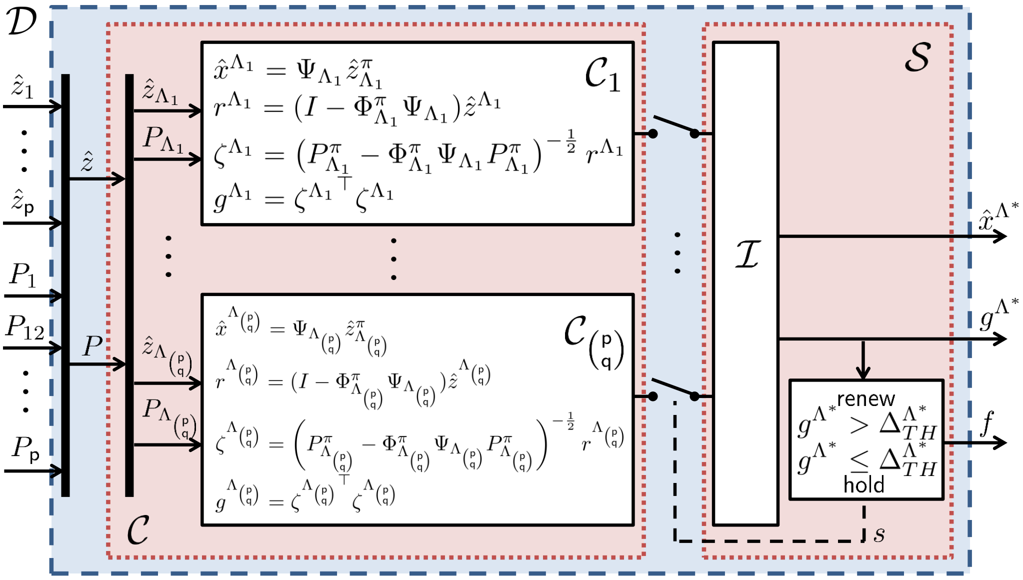

$\mathcal{D}$
$\hat{z}_1$
$\hat{z}_p$
P_1
P_{12}
P_p
$\hat{z}$
P
$\hat{z}_{\Lambda_1}$
P_{Λ_1}
$\hat{x}^{\Lambda_1} = \Psi_{\Lambda_1}\hat{z}^{\pi}_{\Lambda_1}$
$r^{\Lambda_1} = (I - \Phi^{\pi}_{\Lambda_1}\Psi_{\Lambda_1})\hat{z}^{\Lambda_1}$
$\zeta^{\Lambda_1} = \left(P^{\pi}_{\Lambda_1} - \Phi^{\pi}_{\Lambda_1}\Psi_{\Lambda_1}P^{\pi}_{\Lambda_1}\right)^{-\frac{1}{2}} r^{\Lambda_1}$
$g^{\Lambda_1} = \zeta^{\Lambda_1\top}\zeta^{\Lambda_1}$
$\mathcal{C}_1$
$\hat{z}_{\Lambda_{\binom{p}{q}}}$
$P_{\Lambda_{\binom{p}{q}}}$
$\hat{x}^{\Lambda_{\binom{p}{q}}} = \Psi_{\Lambda_{\binom{p}{q}}}\hat{z}^{\pi}_{\Lambda_{\binom{p}{q}}}$
$r^{\Lambda_{\binom{p}{q}}} = (I - \Phi^{\pi}_{\Lambda_{\binom{p}{q}}}\Psi_{\Lambda_{\binom{p}{q}}})\hat{z}^{\Lambda_{\binom{p}{q}}}$
$\zeta^{\Lambda_{\binom{p}{q}}} = \left(P^{\pi}_{\Lambda_{\binom{p}{q}}} - \Phi^{\pi}_{\Lambda_{\binom{p}{q}}}\Psi_{\Lambda_{\binom{p}{q}}}P^{\pi}_{\Lambda_{\binom{p}{q}}}\right)^{-\frac{1}{2}} r^{\Lambda_{\binom{p}{q}}}$
$g^{\Lambda_{\binom{p}{q}}} = \zeta^{\Lambda_{\binom{p}{q}}\top}\zeta^{\Lambda_{\binom{p}{q}}}$
$\mathcal{C}_{\binom{p}{q}}$
$\mathcal{C}$
$\mathcal{I}$
$\mathcal{S}$
$\hat{x}^{\Lambda^*}$
g^{Λ^*}
renew
$g^{\Lambda^*} > \Delta^{\Lambda^*}_{TH}$
$g^{\Lambda^*} \leq \Delta^{\Lambda^*}_{TH}$
hold
f
s

Figure 4.6: Configuration of the decoder $\mathcal{D}$ with Gaussian disturbance/noise.

In summary, by (4.3.10b), the attack induced estimation error vector $e_i(k|k)$ may have arbitrary values if $a_i(k) \not\equiv 0$, while $e_i(k|k) = 0$ when $a_i(k) \equiv 0$. On the other hand, by (4.3.10a), the estimate without any attack, $g_i(k|k)$, is an unbiased estimate and its error, $v_i(k|k)$, is Gaussian distributed with zero mean and covariance matrix $P_i(k|k)$. Finally, with

$$v(k|k) = \begin{bmatrix} v_1(k|k) \\ v_2(k|k) \\ \vdots \\ v_p(k|k) \end{bmatrix} \text{ and } P(k|k) = \begin{bmatrix} P_1(k|k) & P_{12}(k|k) & \cdots & P_{1p}(k|k) \\ P_{21}(k|k) & P_2(k|k) & \cdots & P_{2p}(k|k) \\ \vdots & \vdots & \ddots & \vdots \\ P_{p1}(k|k) & P_{p2}(k|k) & \cdots & P_p(k|k) \end{bmatrix}, \quad (4.3.16)$$

which can be recursively computed by (4.3.14) and (4.3.15), we have

$$v(k|k) \sim N\left(0_{\mu\times 1}, P(k|k)\right),$$

where $\mu := \sum_{i=1}^{p} \mu_i$.

For notational simplicity, $\hat{z}_i(k|k)$, $v_i(k|k)$, $e_i(k|k)$, and $P(k|k)$ are denoted by $\hat{z}_i(k)$, $v_i(k)$, $e_i(k)$, and $P(k)$, respectively. Then, all the data $\hat{z}_i$'s from the partial observers $\mathcal{O}_i$'s in (4.3.8) are collected by the decoder $\mathcal{D}$ and the problem of estimat-

Algorithm 4.2 Operation of the decoder with Gaussian disturbance/noise

```
Input: $\hat{z}_1, \hat{z}_2, \cdots, \hat{z}_{\mathsf{p}}, P_1, P_{12}, \cdots, P_{\mathsf{p}(\mathsf{p}-1)}, P_{\mathsf{p}}$
Output: $\Lambda^*, \hat{x}, g, f$
Initialization: $\Lambda^* = [\mathsf{p}], \hat{x} = \Psi\hat{z}, g = 0, f = 0$
while system (4.1.1) is running do
    $\hat{x} = \Psi_{\Lambda^*}\hat{z}^{\pi}_{\Lambda^*}$
    $r = \hat{z}^{\pi}_{\Lambda^*} - \Phi^{\pi}_{\Lambda^*}\hat{x}$
    $\zeta = (P^{\pi}_{\Lambda^*} - \Phi^{\pi}_{\Lambda^*}\Psi_{\Lambda^*}P^{\pi}_{\Lambda^*})^{-\frac{1}{2}}\, r$
    $g = \zeta^{\top}\zeta$
    if $g \leq \Delta^{\Lambda^*}_{TH}$ then
        $f = 0$
    else if $g > \Delta^{\Lambda^*}_{TH}$ then
        $f = 1$
        for $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| = \mathsf{p} - \mathsf{q}$ do
            $\hat{x}^{\Lambda} = \Psi_{\Lambda}\hat{z}^{\pi}_{\Lambda}$
            $r^{\Lambda} = \hat{z}^{\pi}_{\Lambda} - \Phi^{\pi}_{\Lambda}\hat{x}^{\Lambda}$
            $\zeta^{\Lambda} = (P^{\pi}_{\Lambda} - \Phi^{\pi}_{\Lambda}\Psi_{\Lambda}P^{\pi}_{\Lambda})^{-\frac{1}{2}}\, r^{\Lambda}$
            $g^{\Lambda} = {\zeta^{\Lambda}}^{\top}\zeta^{\Lambda}$
        end for
        $\Lambda^* = \underset{\substack{\Lambda \subset [\mathsf{p}] \\ |\Lambda| = \mathsf{p}-\mathsf{q}}}{\arg\max}\; p_{\mathbf{g}_{\Lambda}}\left(g^{\Lambda}\right)$
    end if
end while
```

ing $x(k)$ is formulated in the form of (4.2.9) as we have seen in Section 4.2.2. Some differences from the previous section, are that the noise v is a Gaussian distributed random variable, not bounded by a certain value, and we do not append any additional zeros. Furthermore, the Kalman filter which constitutes the partial observer $\mathcal{O}_{\mathsf{i}}$, also updates and provides the covariance matrix P of v. More specifically, the error covariance matrix P_{i} of v_{i} is computed by $\mathcal{L}_{\mathsf{i}}$ in (4.3.14), and the error cross covariance P_{ij} of v_{i} and v_{j} is obtained by $\mathcal{L}_{\mathsf{ij}}$ in (4.3.15). Since (4.2.9) exactly matches with (2.2.2) where the static error correcting problem with a Gaussian noise is considered, one can directly apply the error correction technique developed in Section 2.2.3.2 into (4.2.9). The decoder's configuration is sketched in Fig. 4.6

and its operation is described in Algorithm 4.2. Before explaining the operation of the decoder, let Ψ denote $(\Phi'^{\top} P^{-1} \Phi')^{-1} \Phi'^{\top} P^{-1}$ where P is given in (4.3.16) and $\Phi' := [Z_1 \; Z_2 \; \cdots \; Z_{\mathsf{p}}]^{\top} \in \mathbb{R}^{\mu \times \mathsf{n}}$ is Φ in (4.2.10) without any additional zeros. Furthermore, the notation for sub-matrix is slightly abused for simplicity. For example, P^{π}_{Λ}, Φ^{π}_{Λ}, and Ψ_{Λ} denote $P^{\pi}_{\bar{\Lambda},\bar{\Lambda}}$, $\Phi'^{\pi}_{\bar{\Lambda}}$, and $\left(\Phi'^{\pi\top}_{\bar{\Lambda}} (P^{\pi}_{\bar{\Lambda},\bar{\Lambda}})^{-1} \Phi'^{\pi}_{\bar{\Lambda}}\right)^{-1} \Phi'^{\pi\top}_{\bar{\Lambda}} (P^{\pi}_{\bar{\Lambda},\bar{\Lambda}})^{-1}$, respectively, where $\bar{\Lambda} := \bigcup_{\mathsf{j} \in \Lambda} \left\{ \sum_{\mathsf{k}=1}^{\mathsf{j}-1} \mu_{\mathsf{k}} + 1, \sum_{\mathsf{k}=1}^{\mathsf{j}-1} \mu_{\mathsf{k}} + 2, \cdots, \sum_{\mathsf{k}=1}^{\mathsf{j}} \mu_{\mathsf{k}} \right\}$. Recall that $P^{\pi}_{\bar{\Lambda},\bar{\Lambda}}$ denotes the matrix obtained from P by eliminating all i-th rows and all j-th columns such that $\mathsf{i} \in \bar{\Lambda}^c$ and $\mathsf{j} \in \bar{\Lambda}^c$.

Actually, we have the measurement in the form of $\hat{z} = \Phi' x + v + e \in \mathbb{R}^{\mu}$ where $\Phi' \in \mathbb{R}^{\mu \times \mathsf{n}}$ is (μ_{i}-stacked) q-error correctable, $e \in \mathbb{R}^{\mu}$ is (μ_{i}-stacked) q-sparse, and $v \in \mathbb{R}^{\mu}$ satisfies $v \sim N(0_{\mu \times 1}, P)$. Algorithm 4.2 can be seen as a combination of the attack detection scheme (i.e., Algorithm 2.2) in the selected index set of sensors and the state reconstruction scheme (i.e., Algorithm 2.3) when any attack is detected in the selected index set of sensors. Initially, an attack-free index set Λ^*, a state estimate $\hat{x}$, a standardized residual's norm g, and a fault alarm signal f, are set to $[\mathsf{p}]$, $\Psi \hat{z}$, 0, and 0, respectively. The algorithm continually checks if there is any attack in the index set Λ^*. That is, for the given index set Λ^*, the algorithm basically calculates the MVUE (or WLSE) $\hat{x} = \Psi_{\Lambda^*} \hat{z}^{\pi}_{\Lambda^*}$, the residual $r = \hat{z}^{\pi}_{\Lambda^*} - \Phi^{\pi}_{\Lambda^*} \hat{x}$, the standardized residual $\zeta = (P^{\pi}_{\Lambda^*} - \Phi^{\pi}_{\Lambda^*} \Psi_{\Lambda^*} P^{\pi}_{\Lambda^*})^{-\frac{1}{2}} r$, and its 2-norm $g = \zeta^{\top} \zeta$ only with the measurement and covariance data from the subset Λ^*. Recall from Theorem 2.2.18 that if $e_{\mathsf{i}} = 0_{\mu_{\mathsf{i}} \times 1}$ for all $\mathsf{i} \in \Lambda^*$, we have $r \sim N(0_{\mu_{\Lambda^*} \times 1}, P^{\pi}_{\Lambda^*} - \Phi^{\pi}_{\Lambda^*} \Psi_{\Lambda^*} P^{\pi}_{\Lambda^*})$ where $\mu_{\Lambda^*} := \sum_{\mathsf{i} \in \Lambda^*} \mu_{\mathsf{i}}$, and thus, $g \sim \chi^2_{\mu_{\Lambda^*}}$. Therefore, g is used to detect the presence of attack in Λ^* by the χ^2 test. We compare g with the threshold $\Delta^{\Lambda^*}_{TH}$ which is designed before running the algorithm and determines the probability of false alarm and the probability of detection. If $g \leq \Delta^{\Lambda^*}_{TH}$, the index set Λ^* is declared to be attack-free by setting $f = 0$ and the algorithm just maintains the selected optimal index set Λ^*. Otherwise, when g is greater than the threshold $\Delta^{\Lambda^*}_{TH}$, the attack detection alarm is triggered by setting $f = 1$ and the algorithm starts the process of searching new attack-free index set.

In order to find a new attack-free index set and consequently to recover the state x from the new index set, we search all subsets Λ's in $[\mathsf{p}]$ whose cardinal number is $\mathsf{p} - \mathsf{q}$. To this end, let

$$\left\{\Lambda_1, \Lambda_2, \cdots, \Lambda_{\binom{\mathsf{p}}{\mathsf{q}}}\right\}$$

be the set $\{\Lambda \subset [\mathsf{p}] : |\Lambda| = \mathsf{p} - \mathsf{q}\}$. For each subset Λ_i where $\mathsf{i} \in \left[\binom{\mathsf{p}}{\mathsf{q}}\right]$, the computing module $\mathcal{C}_\mathsf{i}$ calculates the MVUE (or WLSE) $\hat{x}^{\Lambda_\mathsf{i}} = \Psi_{\Lambda_\mathsf{i}} \hat{z}^\pi_{\Lambda_\mathsf{i}}$, the residual $r^{\Lambda_\mathsf{i}} = \hat{z}^\pi_{\Lambda_\mathsf{i}} - \Phi^\pi_{\Lambda_\mathsf{i}} \hat{x}^{\Lambda_\mathsf{i}}$, the standardized residual $\zeta^{\Lambda_\mathsf{i}} = \left(P^\pi_{\Lambda_\mathsf{i}} - \Phi^\pi_{\Lambda_\mathsf{i}} \Psi_{\Lambda_\mathsf{i}} P^\pi_{\Lambda_\mathsf{i}}\right)^{-\frac{1}{2}} r^{\Lambda_\mathsf{i}}$, and its 2-norm $g^{\Lambda_\mathsf{i}} = {\zeta^{\Lambda_\mathsf{i}}}^\top \zeta^{\Lambda_\mathsf{i}}$ only with the measurement and covariance data from the subset Λ_i. Then, the selector $\mathcal{S}$ chooses the optimal index set Λ^* by the ML decision rule studied in Section 2.2.3.2. Let us denote $\mathbf{g}_\mathsf{i}$ as a random variable such that g^{Λ_i} is a single observation from $\mathbf{g}_\mathsf{i}$ and $\mathbf{g}_{\Lambda_\mathsf{i}}$ as a random variable such that

$$\mathbf{g}_{\Lambda_\mathsf{i}} \sim \chi^2_{\mu_{\Lambda_\mathsf{i}}}$$

where $\mu_{\Lambda_\mathsf{i}} := \sum_{\mathsf{j} \in \Lambda_\mathsf{i}} \mu_\mathsf{j}$. Note that, if the sensors indexed by Λ_i is attack-free, then the random variable $\mathbf{g}_\mathsf{i}$ as well as $\mathbf{g}_{\Lambda_\mathsf{i}}$ follows the χ^2 distribution with μ_{Λ_i} degrees of freedom. The ML decision rule choose the optimal index set Λ^* that maximize the likelihood $p_{\mathbf{g}_{\Lambda_\mathsf{i}}}\left(g^{\Lambda_\mathsf{i}}\right)$, which is the probability density function of $\mathbf{g}_\mathsf{i}$ being equal to the observation g^{Λ_i} under the condition that there is no attack signal in the measurements indexed by Λ_i. Therefore, we have

$$\Lambda^* = \underset{\substack{\Lambda \subset [\mathsf{p}] \\ |\Lambda| = \mathsf{p} - \mathsf{q}}}{\arg\max}\ p_{\mathbf{g}_\Lambda}\left(g^\Lambda\right),$$

and the MVUE (or WLSE) of the newly selected optimal index set Λ^*, $\hat{x}^{\Lambda^*}$, becomes the final suboptimal estimate of x.

4.4 Remarks on Proposed Attack-Resilient Estimator

4.4.1 Comparison with Fault Detection and Isolation

From a system theoretical point of view, faults and attacks are basically the same except the fact that the attacks may be undetectable because they are devised in a coordinated way by malicious adversaries while the faults can not collude with each other so that they are mostly assumed to show abnormal behavior. The estimator $\mathcal{E}$ can be interpreted as an advanced type of observer-based fault detection and isolation scheme under sparse sensor attacks. Unlike faults, attacks may be undetectable through any type of detector because they are designed craftily by adversaries. However, the sparsity assumption excludes the existence of undetectable attacks in our situation because undetectable attacks must compromise at least a certain number of sensors, which was quantified as the (asymptotic) dynamic security index.

In the case of all (A, c_{i})'s are observable (i.e., the system (4.1.1) is $(\mathsf{p}-1)$-redundant observable), the partial observer $\mathcal{O}_{\mathsf{i}}$ in (4.3.1) becomes the full order state observer as follows:

$$\mathcal{O}_{\mathsf{i}}' : \hat{x}_{\mathsf{i}}(k+1) = A\hat{x}_{\mathsf{i}}(k) + Bu(k) + L_{\mathsf{i}}(\bar{y}_{\mathsf{i}}(k) - c_{\mathsf{i}}\hat{x}_{\mathsf{i}}(k)). \tag{4.3.1$'$}$$

A bank of observers $\mathcal{O}_{\mathsf{i}}'$'s is nothing but the dedicated observer scheme (DOS) and the ℓ_0 minimizer $\mathcal{M}$ with $\mathsf{r} = \mathsf{p} - 1$ decides the final state estimates $\hat{x}$ by a majority voting logic among all $\hat{x}_{\mathsf{i}}$'s for $\mathsf{i} \in [\mathsf{p}]$ [13]. However, if a_{i} is treated as a mere fault, we do not even need the majority voting logic. More specifically, DOS normally detects or isolates the faults based on the output error signal $\tilde{y}_{\mathsf{i}}$, which is used as a residual, of the following system

$$\mathcal{F}_{\mathsf{i}}' : \begin{cases} \tilde{x}_{\mathsf{i}}(k+1) = (A - L_{\mathsf{i}}c_{\mathsf{i}})\tilde{x}_{\mathsf{i}}(k) + L_{\mathsf{i}}n_{\mathsf{i}}(k) - d(k) + L_{\mathsf{i}}a_{\mathsf{i}}(k), \\ \tilde{y}_{\mathsf{i}}(k) = c_{\mathsf{i}}\hat{x}_{\mathsf{i}}(k) - \bar{y}(k) = c_{\mathsf{i}}\tilde{x}_{\mathsf{i}}(k) - n_{\mathsf{i}}(k) - a_{\mathsf{i}}(k) \end{cases} \tag{4.3.5$'$}$$

where $\tilde{x}_{\mathsf{i}} := \hat{x}_{\mathsf{i}} - x$. Note that a_{i} is not generated in a coordinated way because it is considered as a simple fault. Without loss of generality, we can assume that the fault signal a_{i} may not be a zero dynamics signal of $\mathcal{F}_{\mathsf{i}}'$, and thus, it is detectable through the residual $\tilde{y}_{\mathsf{i}}$. Therefore, by excluding the sensor information which has a large residual $\tilde{y}_{\mathsf{i}}$, one can detect and identify the faults.

The generalized observer scheme (GOS) [21] is a variation of DOS and it also utilizes a bank of observers. Contrary to DOS, the i-th observer of GOS is driven by all outputs except the i-th sensor, and thus, it allows to detect and isolate only a single fault. Therefore, GOS is suitable for the system which is 1-redundant observable. Actually, the ℓ_0 minimizer $\mathcal{M}$ with $\mathsf{r} = 1$ works similarly to GOS. However, if the fault acts like an attack, 1-redundant observability guarantees the fault to be detectable, but it is not enough to ensure a single sensor attack to be correctable (or identified) by Proposition 3.3.1. Consequently, GOS works well for a single "fault," but not for a single "attack."

4.4.2 Analysis of Time and Space Complexity

As we have seen in Remark 2.2.2, the NP-hardness of the ℓ_0 minimization problem in the decoder $\mathcal{D}$ could be mitigated by reducing the search space to a finite set without imposing additional conditions other than $2\mathsf{q}$-redundant observability. On top of reducing search space to a finite set, the proposed algorithm further alleviates the computational efforts by combining a detection algorithm to the optimization process. This advantage in terms of time complexity is summarized in the following remark.

Remark 4.4.1. By virtue of the alarm signal f, we can reduce the computational effort significantly. For the attack-resilient state estimation problem, most computational burden originates from the process of solving optimization problem for the decoder in Luenberger observer (or, ML decision rule with combinatorial number of candidates for the decoder in Kalman filter). However, the proposed decoder relieves the computational effort by combining the attack detection mechanism to the optimization process (or, the ML decision rule in Kalman filter). Hence, it only requires to solve the minimization problem (or, conduct the ML decision in Kalman filter) for a very short time interval when the attacker first attempt to inject false data so that the decoder has $f > \mathsf{q}$ (or, $f = 1$ for Kalman filter) at that instant. During normal operation when $f \leq \mathsf{q}$ (or, $f = 0$ for Kalman filter) is guaranteed, the estimator works as if there is no attack. Furthermore, the computational burden to solve the ℓ_0 minimization problem could also be reduced as explained in Remark 2.2.2. That is, the proposed algorithm to solve the ℓ_0 optimization actually relieves the computational complexity by reducing the search space to a finite set. $\Diamond$

Conventional observers which do not take attacks at all, e.g., a full order

Luenberger observer such as (4.2.1), requires n dimension. The cost to pay for attack-resilience is more dimension for total observer dynamics. The proposed estimator requires $\sum_{\mathsf{i}=1}^{\mathsf{p}} \mu_{\mathsf{i}} \leq \mathsf{np}$ dimension, which is in general larger than n, where μ_{i} is the detectability index of the pair (A, c_{i}). Nevertheless, we would like to point out that the required dimension is still much smaller than that of the other observer-based resilient estimators. That is, the required memory for the proposed estimator is linear with the number of sensors by constructing the estimator with a bank of partial observers. However, the other observer-based resilient estimators such as [66] and [12], requires $\mathsf{n}\binom{\mathsf{p}}{\mathsf{q}}$ dimension because they run $\binom{\mathsf{p}}{\mathsf{q}}$ observers and each observer requires n dimension. This advantage in terms of space complexity is summarized in the following remark.

Remark 4.4.2. The idea of constructing the estimator $\mathcal{E}$ with the partial observers $\mathcal{O}_{\mathsf{i}}$'s and the decoder $\mathcal{D}$, which is originally proposed in [42], dramatically reduces the number of the state estimator. Other observer-based resilient state estimators such as [66] and [12], usually consist of all possible combinations of estimator candidates. Thus, they need to run $\binom{\mathsf{p}}{\mathsf{q}}$ estimators so that the required memory size is $\mathsf{n}\binom{\mathsf{p}}{\mathsf{q}}$ for each time step. On the other hand, with the help of Kalman detectability decomposition, the total memory size of the proposed partial observers, $\sum_{\mathsf{i}=1}^{\mathsf{p}} \mu_{\mathsf{i}}$, is not greater than np because the size of each partial observer $\mathcal{O}_{\mathsf{i}}$ is only $\mu_{\mathsf{i}} \leq \mathsf{n}$ for all $\mathsf{i} \in [\mathsf{p}]$. Hence, the proposed estimator is scalable in terms of memory space complexity, that is, it requires a linear space complexity with the number of sensors p. ♢

4.5 Simulation Results: Three-Inertia System

In order to verify the effectiveness of the proposed scheme, simulations with a three-inertia system are conducted in this section. The configuration of the three-inertia system is described in Fig. 4.7 and its dynamics can be represented by a continuous-time state-space equation

$$\mathcal{P}_c : \begin{cases} \dot{x}(t) = A_c x(t) + B_c u(t) + d(t) \\ y(t) = C_c x(t) + n(t) \end{cases} \tag{4.5.1}$$

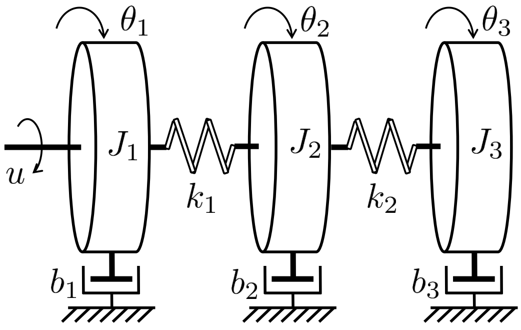


Figure 4.7: Three-inertia system.

with the matrices

$$A_c = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -\frac{k_1}{J_1} & -\frac{b_1}{J_1} & \frac{k_1}{J_1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \frac{k_1}{J_2} & 0 & -\frac{k_1+k_2}{J_2} & -\frac{b_2}{J_2} & \frac{k_2}{J_2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{k_2}{J_3} & 0 & -\frac{k_2}{J_3} & -\frac{b_3}{J_3} \end{bmatrix},$$

$$B_c = \begin{bmatrix} 0 \\ \frac{1}{J_1} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad C_c = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \end{bmatrix},$$

where $J_1 = J_2 = J_3 = 0.01\,\mathrm{kg{\cdot}m^2}$, $b_1 = b_2 = b_3 = 0.007\,\mathrm{N/(rad/s)}$, and $k_1 = k_2 = 1.37\,\mathrm{N/rad}$. Here, the state variables are $[\theta_1 \quad \dot{\theta}_1 \quad \theta_2 \quad \dot{\theta}_2 \quad \theta_3 \quad \dot{\theta}_3]^\top$ and the output measurements are $y := [\theta_1 \quad \theta_2 \quad \theta_3 \quad \theta_1 - \theta_2 \quad \theta_2 - \theta_3]^\top$. On top of the absolute angular position of each inertia, θ_i's, two more measurements which represent the relative angle of adjacent inertias are added to exploit the redundancy of sensors. In addition, the plant is corrupted by the uniformly bounded process disturbance d and measurement noise n with $d_{\max} = n_{\max} = 0.001$. To conduct a discrete-time simulation, the zero-order hold equivalent model of (4.5.1) is considered, that is,

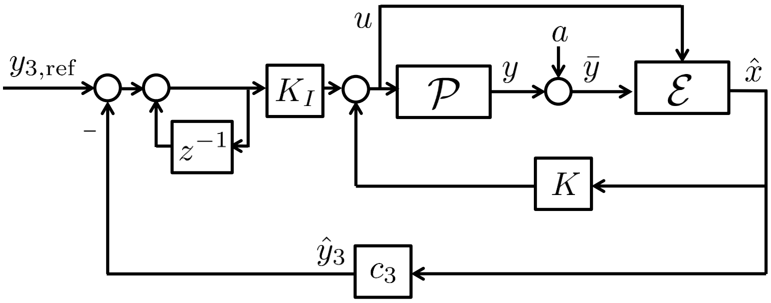


Figure 4.8: Block diagram of the observer-based state feedback integral control scheme.

the system matrices of the discrete-time system (4.1.1) are given by

$$A := e^{A_c T_s}, \quad B := \Big(\int_0^{T_s} e^{A_c \tau} d\tau \Big) B_c, \quad C := C_c \tag{4.5.2}$$

where $T_s :=$ 1ms denotes the sampling time. Note that the pair (A, C) in (4.5.2) is 2-redundant detectable, which implies that one can correct 1-sparse attack signal and its dynamic security index becomes 3. The control objective is to make the output θ_3 follow the step reference $\theta_{3,\mathrm{ref}}$. To this end, an observer-based feedback integral control scheme, as illustrated in [60, Section 6-7] and also in Fig. 4.8, is adopted. First, the state feedback gains K and K_I are chosen as

$$K := -[2.32 \;\; 0.25 \;\; -2.47 \;\; 0.04 \;\; 1.70 \;\; 0.12], \quad K_I := 0.002$$

as if the state x is available. Then, instead of using the conventional Luenberger observer, the proposed estimator $\mathcal{E}$ provides the estimate $\hat{x}$ of x. The injection gain L_{i} of partial observer (4.3.1) in $\mathcal{E}$ is arbitrarily chosen such that $F_{\mathsf{i}} = S_{\mathsf{i}} - L_{\mathsf{i}} t_{\mathsf{i}}$ is Schur stable.

Attack signals are illustrated in Fig. 4.9, which describes that adversaries launch a measurement data injection attack at $t =$ 2sec so that the first sensor is compromised. Figures 4.10 and 4.11 show state trajectories $\theta_1(t)$, $\dot{\theta}_2(t)$, and their estimates. It demonstrates the attack-resilient property of our estimation algorithm. Finally, Fig. 4.12 shows a good reference tracking performance of the proposed control scheme.

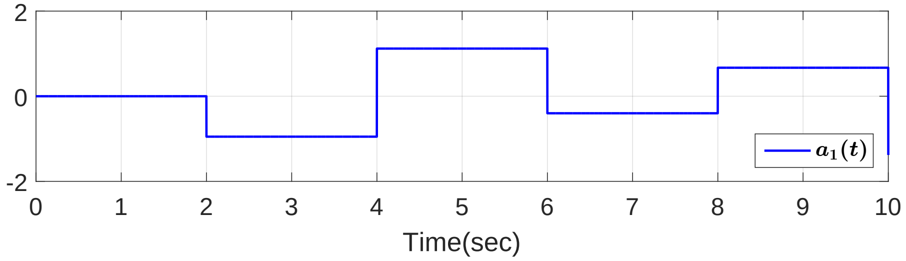


Figure 4.9: Plot of attack $a_1(t)$.

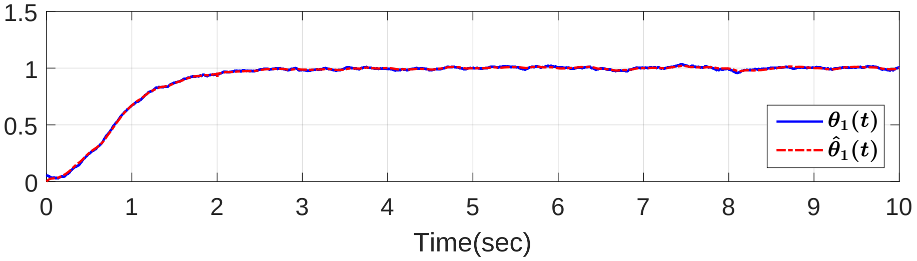


Figure 4.10: Plot of state $\theta_1(t)$ and its estimate $\hat{\theta}_1(t)$.

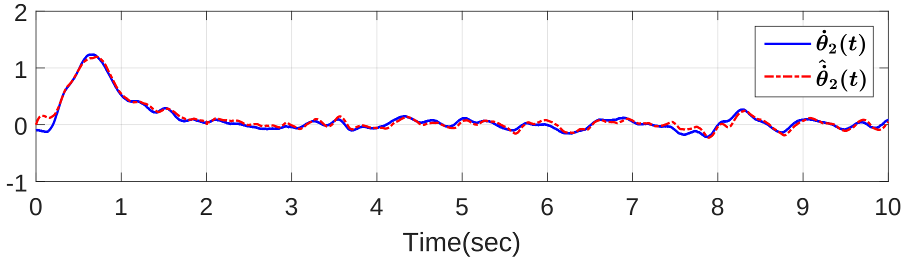


Figure 4.11: Plot of state $\dot{\theta}_2(t)$ and its estimate $\hat{\dot{\theta}}_2(t)$.

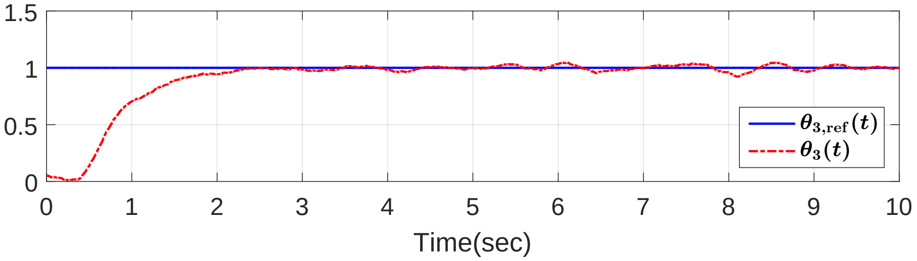


Figure 4.12: Plot of reference signal $\theta_{3,\mathrm{ref}}(t)$ and output $\theta_3(t)$.

Chapter 5

Attack-Resilient State Estimation under Sensor Attacks for Uniformly Observable Nonlinear Systems

Although most control systems have nonlinearity in practice, most of the previous studies on attack-resilient state estimation are restricted to linear dynamical systems. In this chapter, we have extended the results on resilient state estimation for linear systems developed in the previous chapter to a class of nonlinear systems called *uniformly observable nonlinear systems.* Similar to the case of linear systems, it is assumed that the system has sensor redundancy while adversaries can corrupt a subset of sensors with possibly unbounded signals. We design the partial observer by a high gain observer for each measurement output so that only observable portion of system state is obtained. Then, a nonlinear error correcting problem is solved by collecting all the information from those partial observers and by exploiting redundancy. A computationally efficient on-line monitoring scheme is presented for attack detection, and an algorithm for resilient state estimation is provided based on the attack detection scheme.

5.1 Problem Formulation and Preliminaries

5.1.1 Problem Formulation

We consider a smooth continuous-time nonlinear system given by

$$\mathcal{P} : \begin{cases} \dot{x}(t) = f(x(t)) + g(x(t))u(t) \\ y(t) = h(x(t)) \end{cases} \tag{5.1.1}$$

where $x \in \mathbb{R}^{\mathsf{n}}$ is the state variables, $u \in \mathbb{R}$ is the control inputs, and $y \in \mathbb{R}^{\mathsf{p}}$ is the sensor outputs. It is assumed that the state and the input of system (5.1.1) are bounded. More specifically, $u(t) \in U$ for all $t \geq 0$ where U is a compact set, and $x(t) \in X := \{x \in \mathbb{R}^{\mathsf{n}} : \|x\|_\infty \leq M_x\}$ for $t \geq 0$, with a constant $M_x > 0$. There are total p sensors to measure (a smooth function of) the state and the i-th sensor's measurement at time t is denoted by

$$y_{\mathsf{i}}(t) = h_{\mathsf{i}}(x(t)).$$

Sensors themselves or the communication links in the measurement networks are vulnerable to malicious attacks and the measurement data injection attack can be represented by

$$\bar{y}(t) = y(t) + a(t) = h(x(t)) + a(t), \tag{5.1.2}$$

where $\bar{y} \in \mathbb{R}^{\mathsf{p}}$ denotes the sensor data on the controller's side. Thus, $\bar{y}(t)$, not $y(t)$, is used for state estimation. For each sensor output, the attack model is written as

$$\bar{y}_{\mathsf{i}}(t) = y_{\mathsf{i}}(t) + a_{\mathsf{i}}(t) = h_{\mathsf{i}}(x(t)) + a_{\mathsf{i}}(t), \quad \mathsf{i} \in [\mathsf{p}].$$

The attack signal $a_{\mathsf{i}}(t)$ is not assumed to be a bounded signal, and a craftily designed $a_{\mathsf{i}}(t)$ can corrupt $y_{\mathsf{i}}(t)$ so that $\bar{y}_{\mathsf{i}}(t)$ may have arbitrary value. This fact makes it difficult to detect whether there is an attack or not, and even more difficult to estimate the state x from the measured outputs.

Instead of imposing any restrictions on the attack signal $a(t)$ itself, we assume q-sparsity on the set of attack signals $a \in \mathbb{R}^{\mathsf{p}}$. This is motivated by the rationale that the attack resource is limited so that only a portion of the sensors are compromised. Therefore, we suppose that up to q out of p measurement outputs can be compromised and a formal condition on the sparsity of the attack vector a can be given as follows.

Assumption 5.1.1. There exist at least $\mathsf{p}-\mathsf{q}$ sensors which are not attacked for all $t \geq 0$, i.e.,

$$\left|\left\{\mathsf{i} \in [\mathsf{p}] : a_{\mathsf{i}}(t) = 0,\ {}^{\forall}t \geq 0\right\}\right| \geq \mathsf{p} - \mathsf{q}. \qquad \Diamond$$

Based on this sparsity assumption, two problems are of interest in this chapter. The first one is real-time detection of sensor attacks only from the information of the system model (5.1.1), the input u, and the output $\bar{y}$ up to time t. The second problem is to generate a state estimate $\hat{x}(t)$ that converges to the true state $x(t)$ in spite of the attack satisfying Assumption 5.1.1. Similar to the linear systems, it will be seen that the detection of q-sparse sensor attack is solved if system (5.1.1) satisfies q-*redundant observability*, which basically implies observability of (5.1.1) even when any q sensors out of p sensors are removed. Moreover, to solve the resilient state estimation problem, we will ask stronger condition $2\mathsf{q}$-*redundant observability* for system (5.1.1). One of the difficulties in studying observability for nonlinear systems is that, unlike linear systems, a nonlinear system can be both observable and unobservable depending on the input signal $u(t)$ in general. Hence, we will introduce a notion of *uniform observability for any inputs* in Section 5.2, which enables us to design nonlinear observers in most cases. Accordingly, the notion of *redundant observability* for nonlinear systems is slightly modified from that for linear systems.

5.1.2 Bi-Lipschitz Function and Lipschitz Left Inverse

The notion of bi-Lipschitz function and its left inverse will be actively used in this chapter. With $X \subset \mathbb{R}^{\mathsf{n}}$, a fucntion $\phi : X \to \mathbb{R}^{\mathsf{p}}$ is *Lipschitz* on X if there exists a constant $\overline{L}$ such that

$$\|\phi(x) - \phi(x')\|_\infty \leq \overline{L}\|x - x'\|_\infty, \quad {}^{\forall}x, x' \in X.$$

The infimum of such $\overline{L}$ is indicated as $\overline{\mathsf{Lip}}(\phi)$. It is *bi-Lipschitz* on X if, in addition, there exists a positive constant $\underline{L}$ such that

$$\underline{L}\|x - x'\|_\infty \leq \|\phi(x) - \phi(x')\|_\infty, \quad {}^\forall x, x' \in X.$$

The supremum of such $\underline{L}$ is indicated as $\underline{\mathsf{Lip}}(\phi)$. For a given bi-Lipschitz function $\phi : X \to \mathbb{R}^\mathsf{p}$, a function $\psi : \mathbb{R}^\mathsf{p} \to X$ is called a *Lipschitz-extended left inverse* of ϕ if it is Lipschitz on $\mathbb{R}^\mathsf{p}$ and satisfies $\psi(\phi(x)) = x$ for all $x \in X$. It is obvious that a bi-Lipschitz map is injective, and so, its inverse exists on its image $Y := \phi(X)$ and the inverse is also Lipschitz on Y. However, it should be noted that the Lipschitz-extended left inverse ψ is defined on the whole codomain $\mathbb{R}^\mathsf{p}$ and its image $\psi(\mathbb{R}^\mathsf{p})$ is $X \subset \mathbb{R}^\mathsf{n}$.

A differentiable function $\phi : X \to \mathbb{R}^\mathsf{p}$ is called an *immersion* if its Jacobian matrix has full column rank for every $x \in X$. It is well-known that a continuously differentiable function is Lipschitz on any compact subset. Likewise, the following lemma claims that an injective function becomes bi-Lipschitz on every compact set if it is an immersion.

Lemma 5.1.1. Let X be an open subset of $\mathbb{R}^\mathsf{n}$ and $\phi : X \to \mathbb{R}^\mathsf{p}$ be a continuously differentiable function. If ϕ is an injective immersion, then the restriction $\phi|_K : K \to \mathbb{R}^\mathsf{p}$ is bi-Lipschitz for every compact subset $K \subset X$. $\Diamond$

Proof. Note that ϕ is Lipschitz on K because it is continuously differentiable. Thus, it is enough to show

$$\inf_{\substack{x \neq x' \\ x, x' \in K}} \frac{\|\phi(x) - \phi(x')\|_\infty}{\|x - x'\|_\infty} > 0.$$

Suppose, for the sake of contradiction, there exist sequences $\{x_i\}_{i=1}^\infty$ and $\{x_i'\}_{i=1}^\infty$ in K such that $x_i \neq x_i'$ and

$$\lim_{i \to \infty} \frac{\|\phi(x_i) - \phi(x_i')\|_\infty}{\|x_i - x_i'\|_\infty} = 0. \tag{5.1.3}$$

By Bolzano-Weierstrass theorem, without loss of generality (by taking any convergent subsequence if necessary), we may assume that $\{x_i\}_{i=1}^\infty$ and $\{x_i'\}_{i=1}^\infty$ converge to some points x_∞ and x'_∞ in K, respectively. If $x_\infty \neq x'_\infty$, it turns out $\phi(x_\infty) = \phi(x'_\infty)$ and it contradicts injectivity of ϕ. If $x_\infty = x'_\infty$, by continuous differentiability of ϕ, it is derived that

$$\lim_{i\to\infty} \frac{\|\phi(x_i) - \phi(x'_i) - D\phi(x_\infty)\cdot(x_i - x'_i)\|_\infty}{\|x_i - x'_i\|_\infty} = 0, \tag{5.1.4}$$

where $D\phi(x_\infty)$ denotes Jacobian matrix of ϕ at x_∞. Hence, it follows from (5.1.4) together with (5.1.3) that

$$\lim_{i\to\infty} \frac{\|D\phi(x_\infty)\cdot(x_i - x'_i)\|_\infty}{\|x_i - x'_i\|_\infty} = 0,$$

which contradicts the fact that $D\phi(x_\infty)$ has full column rank. □

For example, if $X = [-1,1]\times[-1,1] \subset \mathbb{R}^2$ and $\phi(x) = Tx$ with a matrix $T \in \mathbb{R}^{3\times 2}$ of full column rank, then ϕ is an injective immersion and thus it is a bi-Lipschitz function on X. One of its Lipschitz-extended left inverses is given by $\psi(y) = \mathsf{sat}(T^\dagger y, 1)$ where $T^\dagger \in \mathbb{R}^{2\times 3}$ is the pseudoinverse of T and $\mathsf{sat}(\cdot, 1)$ denotes the component-wise saturation function with the saturation level 1.

5.1.3 Nonlinear Error Detectability and Error Correctability

In this section, we extend the notions of error detectability and error correctability studied in Section 2.2.1 to a nonlinear coding function. Suppose that a nonlinear *coding function* $\Phi : X \to \mathbb{R}^{\mathsf{np}}$ is defined by $\Phi = (\Phi_1, \Phi_2, \cdots, \Phi_{\mathsf{p}})$ where $\Phi_{\mathsf{i}} : X \to \mathbb{R}^{\mathsf{n}}$ for $\mathsf{i} \in [\mathsf{p}]$. We consider a problem of reconstructing the state vector $x \in X$ from the n-stacked measurements $\hat{z}$ given by

$$\hat{z} = \Phi(x) + e \ \in \mathbb{R}^{\mathsf{np}} \tag{5.1.5}$$

where $\hat{z}$ is corrupted by an unknown n-stacked q-sparse vector $e \in \Sigma^{\mathsf{n}}_{\mathsf{q}}$. First, the notion of error detectability of the coding function Φ, is investigated. Similar to Definition 2.2.2, we can define the error detectability as follows. Please recall that the function $\Phi^{\pi}_{\Lambda^{\mathsf{n}}}$ denotes the canonical projection of the function Φ by eliminating all Φ_{i}'s such that $\mathsf{i} \in \Lambda^c$.

Definition 5.1.1. A coding function $\Phi : X \to \mathbb{R}^{\mathsf{np}}$ is said to be *(n-stacked) q-error detectable* if, for all $x, x' \in X$ and $e \in \Sigma_{\mathsf{q}}^{\mathsf{n}}$ such that $\Phi(x) + e = \Phi(x')$, it holds that $x = x'$. Furthermore, the function Φ is said to be *infinitesimally (n-stacked) q-error detectable* if its Jacobian matrix $D\Phi(x)$ is (n-stacked) q-error detectable for all $x \in X$. Finally, the function Φ is said to be *strongly (n-stacked) q-error detectable* if it is both (n-stacked) q-error detectable and infinitesimally (n-stacked) q-error detectable. $\Diamond$

In Definition 5.1.1, the infinitesimal property of the error detectability when there is an infinitely small change in the variable x, is also considered. Similar to the result of Proposition 2.2.2, we can also characterize those error detectability concepts introduced in Definition 5.1.1, as follows.

Proposition 5.1.2. The followings are equivalent:
(i) The function $\Phi : X \to \mathbb{R}^{\mathsf{np}}$ is (n-stacked) q-error detectable;
(ii) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$, $\Phi_{\Lambda^{\mathsf{n}}}^{\pi}$ is injective (or, one-to-one);
(iii) For any $x, x' \in X$ where $x \neq x'$, $\mathsf{d}_{0^{\mathsf{n}}}(\Phi(x), \Phi(x')) > \mathsf{q}$. $\Diamond$

Proof. (i) $\Rightarrow$ (ii): Suppose that (ii) does not hold, i.e., there exists an index set $\Lambda \subset [\mathsf{p}]$ with $|\Lambda| \geq \mathsf{p} - \mathsf{q}$ and $x \neq x'$ such that $\Phi_{\Lambda^{\mathsf{n}}}^{\pi}(x) = \Phi_{\Lambda^{\mathsf{n}}}^{\pi}(x')$. Then it follows that $\|e\|_{0^{\mathsf{n}}} \leq \mathsf{q}$ where $e := \Phi(x') - \Phi(x)$. Thus, $\Phi(x) + e = \Phi(x')$, and Φ is not q-error detectable.
(ii) $\Rightarrow$ (iii): Suppose, for the sake of contradiction, that there exists $x \neq x'$ such that $\mathsf{d}_{0^{\mathsf{n}}}(\Phi(x), \Phi(x')) \leq \mathsf{q}$. Let Λ be the complement of $\mathsf{supp}^{\mathsf{n}}(\Phi(x') - \Phi(x))$, i.e., $\Lambda = (\mathsf{supp}^{\mathsf{n}}(\Phi(x') - \Phi(x)))^c$. Then it is obvious that $|\Lambda| \geq \mathsf{p} - \mathsf{q}$ and $\Phi_{\Lambda^{\mathsf{n}}}^{\pi}(x) = \Phi_{\Lambda^{\mathsf{n}}}^{\pi}(x')$. This contradicts the injectivity condition of $\Phi_{\Lambda^{\mathsf{n}}}^{\pi}$ in (ii).
(iii) $\Rightarrow$ (i): We again prove it by contradiction. Suppose that Φ is not q-error detectable. That is, there exist $x, x' \in X$ satisfying $x \neq x'$, and $e \in \Sigma_{\mathsf{q}}^{\mathsf{n}}$ such that $\Phi(x) + e = \Phi(x')$. It follows from $x \neq x'$ and $e \in \Sigma_{\mathsf{q}}^{\mathsf{n}}$ that $\mathsf{d}_{0^{\mathsf{n}}}(\Phi(x'), \Phi(x)) = \|\Phi(x') - \Phi(x)\|_{0^{\mathsf{n}}} = \|e\|_{0^{\mathsf{n}}} \leq \mathsf{q}$. Thus, (iii) fails. $\Box$

Proposition 5.1.3. The followings are equivalent:
(i) The function $\Phi : X \to \mathbb{R}^{\mathsf{np}}$ is infinitesimally (n-stacked) q-error detectable;
(ii) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$, the Jacobian matrix $D\Phi_{\Lambda^{\mathsf{n}}}^{\pi}(x)$ has

full column rank for all $x \in X$;
(iii) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$, $\Phi^{\pi}_{\Lambda^{\mathsf{n}}}$ is an immersion. ◊

Proof. It directly follows from the definition of the immersion and Proposition 2.2.2. □

Proposition 5.1.4. The followings are equivalent:
(i) The function $\Phi : X \to \mathbb{R}^{\mathsf{np}}$ is strongly (n-stacked) q-error detectable;
(ii) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$, $\Phi^{\pi}_{\Lambda^{\mathsf{n}}}$ is an injective immersion. ◊

Similarly, the notion of error correctability is also introduced and characterized in the subsequent paragraphs. Note that they are slight variations from Definition 2.2.3 and Proposition 2.2.3.

Definition 5.1.2. A coding function $\Phi : X \to \mathbb{R}^{\mathsf{np}}$ is said to be *(n-stacked) q-error correctable* if, for all $x, x' \in X$ and $e, e' \in \Sigma^{\mathsf{n}}_{\mathsf{q}}$ such that $\Phi(x) + e = \Phi(x') + e'$, it holds that $x = x'$. Furthermore, the function Φ is said to be *infinitesimally (n-stacked) q-error correctable* if its Jacobian matrix $D\Phi(x)$ is (n-stacked) q-error correctable for all $x \in X$. Finally, the function Φ is said to be *strongly (n-stacked) q-error correctable* if it is both (n-stacked) q-error correctable and infinitesimally (n-stacked) q-error correctable. ◊

Proposition 5.1.5. The followings are equivalent:
(i) The function $\Phi : X \to \mathbb{R}^{\mathsf{np}}$ is (n-stacked) q-error correctable;
(ii) The function $\Phi : X \to \mathbb{R}^{\mathsf{np}}$ is (n-stacked) $2\mathsf{q}$-error detectable;
(iii) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - 2\mathsf{q}$, $\Phi^{\pi}_{\Lambda^{\mathsf{n}}}$ is injective (or, one-to-one);
(iv) For any $x, x' \in X$ where $x \neq x'$, $\mathsf{d}_{0^{\mathsf{n}}}(\Phi(x), \Phi(x')) > 2\mathsf{q}$. ◊

Proof. (i) ⇒ (ii): Assume that $x, x' \in X$ and $e \in \Sigma^{\mathsf{n}}_{2\mathsf{q}}$ satisfying $\Phi(x) + e = \Phi(x')$, are given. Let e_1 and e_2 be such that $e = e_1 - e_2$ where $e_1, e_2 \in \Sigma^{\mathsf{n}}_{\mathsf{q}}$. Thus, we have $\Phi(x) + e_1 = \Phi(x') + e_2$. Since $\Phi : X \to \mathbb{R}^{\mathsf{np}}$ is q-error correctable, it follows that $x = x'$.
(ii) ⇒ (i): Assume that $x, x' \in X$ and $e, e' \in \Sigma^{\mathsf{n}}_{\mathsf{q}}$ satisfying $\Phi(x) + e = \Phi(x') + e'$, are given. Then, we have $\Phi(x) + e'' = \Phi(x')$ where $e'' = e - e' \in \Sigma^{\mathsf{n}}_{2\mathsf{q}}$. Since $\Phi : X \to \mathbb{R}^{\mathsf{np}}$ is $2\mathsf{q}$-error detectable, it follows that $x = x'$.
(ii) ⇔ (iii) ⇔ (iv): It directly follows from Proposition 5.1.2. □

Proposition 5.1.6. The followings are equivalent:

(i) The function $\Phi : X \to \mathbb{R}^{\mathsf{np}}$ is infinitesimally (n-stacked) q-error correctable;

(ii) The function $\Phi : X \to \mathbb{R}^{\mathsf{np}}$ is infinitesimally (n-stacked) $2\mathsf{q}$-error detectable;

(iii) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - 2\mathsf{q}$, the Jacobian matrix $D\Phi^{\pi}_{\Lambda^{\mathsf{n}}}(x)$ has full column rank for all $x \in X$;

(iv) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - 2\mathsf{q}$, $\Phi^{\pi}_{\Lambda^{\mathsf{n}}}$ is an immersion. ◊

Proof. It directly follows from the definition of the immersion and Proposition 2.2.3. □

Proposition 5.1.7. The followings are equivalent:

(i) The function $\Phi : X \to \mathbb{R}^{\mathsf{np}}$ is strongly (n-stacked) q-error correctable;

(ii) The function $\Phi : X \to \mathbb{R}^{\mathsf{np}}$ is strongly (n-stacked) $2\mathsf{q}$-error detectable;

(iii) For every set $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - 2\mathsf{q}$, $\Phi^{\pi}_{\Lambda^{\mathsf{n}}}$ is an injective immersion. ◊

5.2 Uniformly Observable Nonlinear Systems for Any Input

5.2.1 Uniform Observability Decomposition

The basic idea of constructing the partial observer is the same as that of linear systems presented in Section 4.2.1. That is, we design p nonlinear observers to each individual system with the single measurement $\bar{y}_{\mathsf{i}}$ given by

$$\overline{\mathcal{P}}_{\mathsf{i}} : \begin{cases} \dot{x}(t) = f(x(t)) + g(x(t))u(t) \\ \bar{y}_{\mathsf{i}}(t) = h_{\mathsf{i}}(x(t)) + a_{\mathsf{i}}(t), \end{cases} \tag{5.2.1}$$

for all $\mathsf{i} \in [\mathsf{p}]$. Because there is no guarantee that the state x is observable from the single output $\bar{y}_{\mathsf{i}}$, each observer cannot recover the full state x in general. Instead, each observer can recover observable portion of the state only. By observable portion, we mean the observable sub-state in a special coordinate. For linear systems, this sub-state corresponds to the observable subsystem in the well-known Kalman observabiliy decomposition, i.e., the state $z_{\mathsf{i}} = z^{o}_{\mathsf{i}} \in \mathbb{R}^{\nu_{\mathsf{i}}}$ in (4.2.5) is the observable sub-state. For nonlinear systems, we assume that the observable subsystem is *uniformly observable for any input* [22, 24, 88]. While general nonlinear systems can be both observable and unobservable depending on the input signal $u(t)$, a uniformly observable nonlinear system is observable for every inputs. In other words, it is observable uniformly in inputs and one can design a nonlinear observer for a class of uniformly observable nonlinear systems. The following assumption asks uniform observability of the observable portion of the individual system (5.2.1) with a single output $\bar{y}_{\mathsf{i}}$.

Assumption 5.2.1. For each $\mathsf{i} \in [\mathsf{p}]$, there exist a natural number $\nu_{\mathsf{i}} \leq \mathsf{n}$ and a diffeomorphism $\mathcal{T}_{\mathsf{i}} : \mathbb{R}^{\mathsf{n}} \to \mathbb{R}^{\nu_{\mathsf{i}}} \times \mathbb{R}^{\mathsf{n}-\nu_{\mathsf{i}}}$ such that, by $\begin{bmatrix} z_{\mathsf{i}} \\ w_{\mathsf{i}} \end{bmatrix} := \mathcal{T}_{\mathsf{i}}(x)$ with $z_{\mathsf{i}} \in \mathbb{R}^{\nu_{\mathsf{i}}}$ and $w_{\mathsf{i}} \in \mathbb{R}^{\mathsf{n}-\nu_{\mathsf{i}}}$, the system (5.2.1) is transformed into the form

$$\dot{z}_{\mathsf{i}} = \mathcal{F}_{\mathsf{i}}(z_{\mathsf{i}}) + \mathcal{G}_{\mathsf{i}}(z_{\mathsf{i}})u \tag{5.2.2a}$$

$$\dot{w}_{\mathsf{i}} = \mathcal{F}_{\mathsf{i}}'(z_{\mathsf{i}}, w_{\mathsf{i}}) + \mathcal{G}_{\mathsf{i}}'(z_{\mathsf{i}}, w_{\mathsf{i}})u \tag{5.2.2b}$$

$$\bar{y}_{\mathsf{i}} = y_{\mathsf{i}} + a_{\mathsf{i}} = \mathcal{H}_{\mathsf{i}}(z_{\mathsf{i}}) + a_{\mathsf{i}} \tag{5.2.2c}$$

where the z_{i}-subsystem (5.2.2a) with the attack-free output $y_{\mathsf{i}} := \mathcal{H}_{\mathsf{i}}(z_{\mathsf{i}})$ in (5.2.2c) is uniformly observable on $\mathbb{R}^{\nu_{\mathsf{i}}}$, i.e., the ν_{i}-dimensional subsystem (5.2.2a) and (5.2.2c) takes the form

$$\dot{z}_{\mathsf{i}} = \begin{bmatrix} \dot{z}_{\mathsf{i},1} \\ \dot{z}_{\mathsf{i},2} \\ \vdots \\ \dot{z}_{\mathsf{i},\nu_{\mathsf{i}}} \end{bmatrix} = \begin{bmatrix} z_{\mathsf{i},2} \\ \vdots \\ z_{\mathsf{i},\nu_{\mathsf{i}}} \\ \alpha_{\mathsf{i}}(z_{\mathsf{i}}) \end{bmatrix} + \begin{bmatrix} \beta_{\mathsf{i},1}(z_{\mathsf{i},1}) \\ \beta_{\mathsf{i},2}(z_{\mathsf{i},1}, z_{\mathsf{i},2}) \\ \vdots \\ \beta_{\mathsf{i},\nu_{\mathsf{i}}}(z_{\mathsf{i},1}, \cdots, z_{\mathsf{i},\nu_{\mathsf{i}}}) \end{bmatrix} u \tag{5.2.3a}$$

$$\bar{y}_{\mathsf{i}} = y_{\mathsf{i}} + a_{\mathsf{i}} = z_{\mathsf{i},1} + a_{\mathsf{i}}. \tag{5.2.3b}$$

Moreover, the functions $\alpha_{\mathsf{i}} : \mathbb{R}^{\nu_{\mathsf{i}}} \to \mathbb{R}$ and $\beta_{\mathsf{i},\mathsf{j}} : \mathbb{R}^{\mathsf{j}} \to \mathbb{R}$ for $\mathsf{j} \in [\nu_{\mathsf{i}}]$, are globally Lipschitz. $\Diamond$

Remark 5.2.1. The first ν_{i} component of the diffeomorphism $\mathcal{T}_{\mathsf{i}}$ is given by

$$z_{\mathsf{i}} = \begin{bmatrix} z_{\mathsf{i},1} \\ z_{\mathsf{i},2} \\ z_{\mathsf{i},3} \\ \vdots \\ z_{\mathsf{i},\nu_{\mathsf{i}}} \end{bmatrix} = \begin{bmatrix} h_{\mathsf{i}}(x) \\ L_f h_{\mathsf{i}}(x) \\ L_f^2 h_{\mathsf{i}}(x) \\ \vdots \\ L_f^{\nu_{\mathsf{i}}-1} h_{\mathsf{i}}(x) \end{bmatrix} =: \Phi_{\mathsf{i}}(x) \tag{5.2.4}$$

which is easily verified by comparing (5.2.1) and (5.2.3) with $u \equiv 0$. Here, $L_f h_{\mathsf{i}}(x) = \dfrac{\partial h_{\mathsf{i}}}{\partial x} f(x)$ is the Lie derivative of h_{i} along the vector field f and the notation $L_f^{\mathsf{k}} h_{\mathsf{i}}(x)$ represents the repetition of the calculation as

$$L_f^{\mathsf{k}} h_{\mathsf{i}}(x) = L_f L_f^{\mathsf{k}-1} h_{\mathsf{i}}(x) = \frac{\partial (L_f^{\mathsf{k}-1} h_{\mathsf{i}})}{\partial x} f(x). \qquad \Diamond$$

Remark 5.2.2. The sub-state z_{i} corresponds to the observable sub-state from the output y_{i} while w_{i} corresponds to the unobservable sub-state. This is obvious from the structure of (5.2.3) and (5.2.2b). Therefore, the system (5.2.2), which is decomposed into the observable subsystem (5.2.2a) and the unobservable subsystem (5.2.2b) by the diffeomorphism $\mathcal{T}_{\mathsf{i}}$, is called *uniform observability decomposition* [74]. For linear systems, Assumption 5.2.1 always holds as we have seen in Section 4.2.1 that the Kalman observability decomposition is always possible. On the other hand, the triangular structure of $\beta_{\mathsf{i}} = [\beta_{\mathsf{i},1}, \cdots, \beta_{\mathsf{i},\nu_{\mathsf{i}}}]^\top$ is a necessary and sufficient condition for uniform observability of z_{i}-subsystem (5.2.3) (see [24] for the proof of this statement). $\Diamond$

Remark 5.2.3. Asking global Lipschitz properties for α_{i} and $\beta_{\mathsf{i},\mathsf{j}}$ for $\mathsf{j} \in [\nu_{\mathsf{i}}]$, is not a restriction thanks to boundedness of $x(t)$. Indeed, noting that $x(t) \in X$, find a constant $M_{z,\mathsf{i}}$ such that $\|z_{\mathsf{i}}\|_\infty = \|\Phi_{\mathsf{i}}(x)\|_\infty \le M_{z,\mathsf{i}}$ for all $x \in X$. Then, one can modify α_{i} and $\beta_{\mathsf{i},\mathsf{j}}$ outside the set $\mathcal{Z}^{\mathsf{i}} := \{z_{\mathsf{i}} : \|z_{\mathsf{i}}\|_\infty \le M_{z,\mathsf{i}}\}$ so that α_{i} and $\beta_{\mathsf{i},\mathsf{j}}$ are globally Lipschitz while they remain the same in $\mathcal{Z}^{i}$. In theory, this modification is always possible by Kirszbraun's Lipschitz extension theorem [72, p. 21]. That is, for a function $f : X \to \mathbb{R}$ which is Lipschitz on X, a Lipschitz extension is given by

$$\overline{f}(x) := \inf_{y \in X} (f(y) + \overline{\mathsf{Lip}}(f)\|x - y\|_\infty)$$

where $\overline{\mathsf{Lip}}(f)$ is a Lipschitz constant of f on X. For a vector-valued function f, this extension is applied to each component. However, in practice, one can employ a simpler way. For example, $\alpha_{\mathsf{i}}(z_{\mathsf{i}})$ is replaced by

$$\overline{\alpha}_{\mathsf{i}}(z_{\mathsf{i}}) = \alpha_{\mathsf{i}}(\mathsf{sat}(z_{\mathsf{i}}, M_{z,\mathsf{i}})) \tag{5.2.5}$$

where sat is the component-wise saturation function, i.e., for $z_{\mathsf{i}} \in \mathbb{R}^{\nu_{\mathsf{i}}}$,

$$\mathsf{sat}(z_{\mathsf{i}}, M) := \begin{bmatrix} \mathsf{sat}(z_{\mathsf{i},1}, M) \\ \vdots \\ \mathsf{sat}(z_{\mathsf{i},\nu_{\mathsf{i}}}, M) \end{bmatrix} \in \mathbb{R}^{\nu_{\mathsf{i}}},$$

where, for $z_{\mathsf{i},\mathsf{j}} \in \mathbb{R}$,

$$\mathsf{sat}(z_{\mathsf{i},\mathsf{j}}, M) := \begin{cases} M, & z_{\mathsf{i},\mathsf{j}} > M, \\ z_{\mathsf{i},\mathsf{j}}, & |z_{\mathsf{i},\mathsf{j}}| \le M, \\ -M, & z_{\mathsf{i},\mathsf{j}} < -M. \end{cases}$$

See [73, Section 3.3] for more details. ◊

5.2.2 Design of High Gain Observer

For each $\mathsf{i} \in [\mathsf{p}]$, a high gain observer only for the observable subsystem (5.2.3) is constructed by

$$\dot{\hat{z}}_{\mathsf{i}} = \begin{bmatrix} \dot{\hat{z}}_{\mathsf{i},1} \\ \dot{\hat{z}}_{\mathsf{i},2} \\ \vdots \\ \dot{\hat{z}}_{\mathsf{i},\nu_{\mathsf{i}}} \end{bmatrix} = \begin{bmatrix} \hat{z}_{\mathsf{i},2} \\ \vdots \\ \hat{z}_{\mathsf{i},\nu_{\mathsf{i}}} \\ \alpha_{\mathsf{i}}(\hat{z}_{\mathsf{i}}) \end{bmatrix} + \begin{bmatrix} \beta_{\mathsf{i},1}(\hat{z}_{\mathsf{i},1}) \\ \beta_{\mathsf{i},2}(\hat{z}_{\mathsf{i},1}, \hat{z}_{\mathsf{i},2}) \\ \vdots \\ \beta_{\mathsf{i},\nu_{\mathsf{i}}}(\hat{z}_{\mathsf{i},1}, \cdots, \hat{z}_{\mathsf{i},\nu_{\mathsf{i}}}) \end{bmatrix} u + P_{\mathsf{i}}^{-1} C_{\mathsf{i}}^\top (\bar{y}_{\mathsf{i}} - C_{\mathsf{i}} \hat{z}_{\mathsf{i}}) \tag{5.2.6}$$

where $\hat{z}_{\mathsf{i}}$ is the state estimate of z_{i}, the matrix C_{i} has the form of

$$C_{\mathsf{i}} := \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{1 \times \nu_{\mathsf{i}}},$$

and $P_{\mathsf{i}} = P_{\mathsf{i}}(\theta_{\mathsf{i}}) \in \mathbb{R}^{\nu_{\mathsf{i}} \times \nu_{\mathsf{i}}}$ is the unique positive definite solution of

$$0 = -\theta_{\mathsf{i}} P_{\mathsf{i}} - A_{\mathsf{i}}^T P_{\mathsf{i}} - P_{\mathsf{i}} A_{\mathsf{i}} + C_{\mathsf{i}}^T C_{\mathsf{i}}. \tag{5.2.7}$$

In (5.2.7), θ_{i} is a constant to be determined, and the matrix A_{i} is given by

$$A_{\mathsf{i}} := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{\nu_{\mathsf{i}} \times \nu_{\mathsf{i}}}.$$

We suppose the initial condition $\hat{z}_{\mathsf{i}}(0)$ of the high gain observer (5.2.6) is set such that $\|\hat{z}_{\mathsf{i}}(0)\|_\infty \leq M_{z,\mathsf{i}}$. Finally, the parameter θ_{i} is determined by the following lemma. In practice, θ_{i} is often taken by a sufficiently large number from simulations.

Lemma 5.2.1. [24, Theorem 3], [73, Lemma 3.2.2] Consider the system (5.2.3) where the functions α_{i} and $\beta_{\mathsf{i},\mathsf{j}}$ are globally Lipschitz and the attack signal a_{i} is identically zero, i.e., $a_{\mathsf{i}} \equiv 0$. Let the observer be given by (5.2.6). Then, there exists a positive constant $\theta_{\mathsf{i}}^* \geq 1$ such that, for any $\theta_{\mathsf{i}} \geq \theta_{\mathsf{i}}^*$, the observer (5.2.6) guarantees

$$\|\hat{z}_{\mathsf{i}}(t) - z_{\mathsf{i}}(t)\|_\infty \leq \eta_{\mathsf{i}}(\theta_{\mathsf{i}}) e^{-\frac{\theta_{\mathsf{i}}}{4}t} \|\hat{z}_{\mathsf{i}}(0) - z_{\mathsf{i}}(0)\|_\infty. \tag{5.2.8}$$

for $t \geq 0$ with some function $\eta_{\mathsf{i}}(\theta_{\mathsf{i}})$. Moreover, for a fixed time $\tau > 0$,

$$\eta_{\mathsf{i}}(\theta_{\mathsf{i}}) e^{-\frac{\theta_{\mathsf{i}}}{4}\tau} \to 0 \qquad \text{as} \qquad \theta_{\mathsf{i}} \to \infty. \qquad \Diamond$$

5.3 Redundant Observability for Uniformly Observable Nonlinear Systems

Recall that we have appended additional $\mathsf{n} - \nu_{\mathsf{i}}$ zeros to the observable sub-state z_{i} in (4.2.8) so that the size of those sub-states from different sensors matches each other. We follow the same procedure for nonlinear systems, too. That is, additional zeros are augmented to (5.2.4) so that it becomes n dimensional, which is written by

$$z_{\mathsf{i}}^{\mathsf{n}} := \begin{bmatrix} z_{\mathsf{i}} \\ 0_{(\mathsf{n}-\nu_{\mathsf{i}})\times 1} \end{bmatrix} = \begin{bmatrix} \Phi_{\mathsf{i}}(x) \\ 0_{(\mathsf{n}-\nu_{\mathsf{i}})\times 1} \end{bmatrix} =: \Phi_{\mathsf{i}}'(x) \in \mathbb{R}^{\mathsf{n}}. \tag{5.3.1}$$

It is also supposed that additional zero elements are also appended to $\hat{z}_{\mathsf{i}}$'s of the high gain observer (5.2.6) just like (5.3.1). Now, let us collect all z_{i}'s for $\mathsf{i} \in [\mathsf{p}]$ and stack them at once by

$$z := \begin{bmatrix} z_1^{\mathsf{n}} \\ \vdots \\ z_{\mathsf{p}}^{\mathsf{n}} \end{bmatrix} = \begin{bmatrix} \Phi_1'(x) \\ \vdots \\ \Phi_p'(x) \end{bmatrix} =: \Phi(x) \in \mathbb{R}^{\mathsf{np}}, \tag{5.3.2}$$

which is defined on X. The estimates $\hat{z}_{\mathsf{i}}$'s obtained from (5.2.6) for all $\mathsf{i} \in [\mathsf{p}]$, are also collected and form a stacked vector of

$$\hat{z} := \begin{bmatrix} \hat{z}_1^{\mathsf{n}} \\ \vdots \\ \hat{z}_{\mathsf{p}}^{\mathsf{n}} \end{bmatrix} \in \mathbb{R}^{\mathsf{np}} \tag{5.3.3}$$

where $\hat{z}_{\mathsf{i}}^{\mathsf{n}}$ is augmented with additional zeros to $\hat{z}_{\mathsf{i}}$. Note that the state estimate $\hat{z}$ converges to the real state z by Lemma 5.2.1 when there is no attack. In order to recover the state x from the collection of estimates $\hat{z}$, the function $\Phi : X \to \Phi(X) \subset \mathbb{R}^{\mathsf{np}}$ should have injectivity so that it has a left inverse Φ^{-1}, defined at least on its image $\Phi(X)$. Furthermore, let the estimate of x be the left inverse of $\hat{z}$ if $\hat{z} = z$. On top of injectivity, we require the mapping Φ be an immersion in order to ensure bi-Lipschitzness (which will be used later) on the domain X. Asking Φ to be an injective immersion is in fact an extension of linear case since Jacobian of Φ corresponds to the observability matrix, which has full column rank. Moreover, since up to q estimates among all $\hat{z}_{\mathsf{i}}$'s might be compromised, we require some redundancy in the map Φ that the map remains as an injective immersion even if any q components Φ_{i}' are eliminated from Φ. The following definition precisely states this requirement and it coincides with the strong error detectability of Φ

in Proposition 5.1.4.

Definition 5.3.1. The dynamical system (5.1.1) is said to be q-*redundant observable* if, for the mapping $\Phi : X \to \mathbb{R}^{\mathsf{np}}$ in (5.3.2), the function $\Phi^{\pi}_{\Lambda^{\mathsf{n}}} : X \to \mathbb{R}^{\mathsf{n}|\Lambda|}$ is an injective immersion for any $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$. ◊

In Definition 5.3.1, the function $\Phi^{\pi}_{\Lambda^{\mathsf{n}}}$ denotes the canonical projection of the function Φ by eliminating all Φ_{i}''s such that $\mathsf{i} \in \Lambda^c$. In term of the definition above, 0-redundant observability can be regarded as conventional observability of the system (5.1.1). More specifically, the system (5.1.1) is said to be *strongly differentially observable* if the mapping Φ is an injective immersion [23, Definition I.2.4.2].

Now, it is noted that, although q-redundant observability of (5.1.1) guarantees existence of a left inverse $(\Phi^{\pi}_{\Lambda^{\mathsf{n}}})^{-1}$ of $\Phi^{\pi}_{\Lambda^{\mathsf{n}}}$ where $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| \geq \mathsf{p} - \mathsf{q}$, the inverse $(\Phi^{\pi}_{\Lambda^{\mathsf{n}}})^{-1}$ is defined only on the image $\Phi^{\pi}_{\Lambda^{\mathsf{n}}}(X)$. While it is true that $z^{\pi}_{\Lambda^{\mathsf{n}}}(t) \in \Phi^{\pi}_{\Lambda^{\mathsf{n}}}(X) \subset \mathbb{R}^{\mathsf{n}|\Lambda|}$, there is no guarantee that the estimate $\hat{z}^{\pi}_{\Lambda^{\mathsf{n}}}(t)$, that converges to $z^{\pi}_{\Lambda^{\mathsf{n}}}(t)$, belongs to $\Phi^{\pi}_{\Lambda^{\mathsf{n}}}(X)$. In order to use the left inverse of $\Phi^{\pi}_{\Lambda^{\mathsf{n}}}$ on the whole space $\mathbb{R}^{\mathsf{n}|\Lambda|}$, let us define the *Lipschitz-extended left inverse* of $\Phi^{\pi}_{\Lambda^{\mathsf{n}}}$ as

$$\begin{aligned} \Psi^{\Lambda} : \mathbb{R}^{\mathsf{n}|\Lambda|} &\to X \\ z^{\Lambda} &\mapsto \mathsf{sat}\left(\overline{(\Phi^{\pi}_{\Lambda^{\mathsf{n}}})^{-1}}(z^{\Lambda}),\ M_x\right) \end{aligned} \tag{5.3.4}$$

in which, $\overline{(\Phi^{\pi}_{\Lambda^{\mathsf{n}}})^{-1}}$ is a Lipschitz extension of $(\Phi^{\pi}_{\Lambda^{\mathsf{n}}})^{-1}$ from $\Phi^{\pi}_{\Lambda^{\mathsf{n}}}(X)$ to $\mathbb{R}^{\mathsf{n}|\Lambda|}$ (please refer to Remark 5.2.3), and the saturation function is employed in order to map the image of $\overline{(\Phi^{\pi}_{\Lambda^{\mathsf{n}}})^{-1}}$ into the set X. Indeed, this function Ψ^{Λ} is globally Lipschitz on $\mathbb{R}^{\mathsf{n}|\Lambda|}$ because $\Phi^{\pi}_{\Lambda^{\mathsf{n}}}$ is bi-Lipschitz on X by Lemma 5.1.1, and so, a left inverse of $\Phi^{\pi}_{\Lambda^{\mathsf{n}}}$ exists on $\Phi^{\pi}_{\Lambda^{\mathsf{n}}}(X)$ which is Lipschitz on $\Phi^{\pi}_{\Lambda^{\mathsf{n}}}(X)$. It is then extended to be globally Lipschitz on $\mathbb{R}^{\mathsf{n}|\Lambda|}$, and the saturation function in the end preserves Lipschitz property. With the global Lipschitz inverse function Ψ^{Λ} at hand, let the estimate of $x(t)$ be

$$\hat{x}^{\Lambda}(t) := \Psi^{\Lambda}(\hat{z}^{\pi}_{\Lambda^{\mathsf{n}}}(t)) \ \in X.$$

Remark 5.3.1. For simple construction of Lipschitz extension $\overline{(\Phi^{\pi}_{\Lambda^{\mathsf{n}}})^{-1}}$ in practice, one may want to employ a method using saturation functions as in (5.2.5).

Let $M_z := \max_{\mathsf{i}\in[\mathsf{p}]} M_{z,\mathsf{i}}$, and $\mathcal{Z}^\Lambda := \{z^\Lambda \in \mathbb{R}^{\mathsf{n}|\Lambda|} : \|z^\Lambda\|_\infty \le M_z\}$ which contains $\Phi(X)$ by construction. If there is a smooth function $\Phi'^{-1}_{\Lambda^{\mathsf{n}}}$ defined on $\mathcal{Z}^\Lambda$ such that $\Phi'^{-1}_{\Lambda^{\mathsf{n}}}(z^\Lambda) = (\Phi^\pi_{\Lambda^{\mathsf{n}}})^{-1}(z^\Lambda)$ for all $z^\Lambda \in \Phi(X)$, then a Lipschitz extension $\overline{(\Phi^\pi_{\Lambda^{\mathsf{n}}})^{-1}}$ is easily obtained by

$$\overline{(\Phi^\pi_{\Lambda^{\mathsf{n}}})^{-1}}(z^\Lambda) = \Phi'^{-1}_{\Lambda^{\mathsf{n}}}(\mathsf{sat}(z^\Lambda,\, M_z))). \tag{5.3.5}$$

$\Diamond$

Suppose that system (5.1.1) is q-redundant observable. Since up to q sensors are compromised, there is at least one index set $\Lambda^* \subset [\mathsf{p}]$ with $|\Lambda^*| = \mathsf{p} - \mathsf{q}$ such that $\Lambda^* \subset (\mathsf{supp}(a(t)))^c$ for all $t \ge 0$. In this case, we have

$$\begin{aligned}\|\hat{x}^{\Lambda^*}(t) - x(t)\|_\infty &= \|\Psi^{\Lambda^*}(\hat{z}^\pi_{\Lambda^{*\mathsf{n}}}(t)) - \Psi^{\Lambda^*}(z^\pi_{\Lambda^{*\mathsf{n}}}(t))\|_\infty \\ &\le \overline{\mathsf{Lip}}(\Psi^{\Lambda^*}) \max_{\mathsf{i}\in\Lambda^*}\left\{2M_{z,\mathsf{i}}\eta_{\mathsf{i}}(\theta_{\mathsf{i}})e^{-\frac{\theta_{\mathsf{i}}}{4}t}\right\}\end{aligned}$$

which follows from Lemma 5.2.1, and thus, $x(t)$ is asymptotically recovered by $\hat{x}^{\Lambda^*}(t)$. However, since the set Λ^* is not known, let us discuss how to find Λ^* such that $\Lambda^* \subset (\mathsf{supp}(a(t)))^c$ for all $t \ge 0$, in the subsequent sections.

5.4 Attack Detection and Resilient State Estimation for Uniformly Observable Nonlinear Systems

5.4.1 Detection of Sensor Attacks

The state estimation error defined by $\tilde{z}(t) := \hat{z}(t) - z(t)$ satisfies

$$\tilde{z}(t) = \hat{z}(t) - \Phi(x(t)) = v(t) + e(t) \ \in \mathbb{R}^{\mathsf{np}} \tag{5.4.1}$$

where the vector v is the transient estimation error caused by the high gain observers and the vector e represents the error caused by injected sensor attack. The situation of (5.4.1) is similar to that of (4.3.4) and (4.3.12) for linear systems. If there is no attack, we have $e(t) \equiv 0$ and $v_{\mathsf{i}}(t) = \hat{z}_{\mathsf{i}}(t) - \Phi_{\mathsf{i}}(x(t)) \in \mathbb{R}^{\nu_{\mathsf{i}}}$ converges to zero as in (5.2.8) of Lemma 5.2.1 for all $\mathsf{i} \in [\mathsf{p}]$. Under the q-sparsity assumption on a (i.e., Assumption 5.1.1), with the unknown attack-free index set

$$\underline{\Lambda} := \left\{ \mathsf{i} \in [\mathsf{p}] : a_{\mathsf{i}}(t) = 0, \ {}^{\forall} t \geq 0 \right\},$$

it follows that $e_{\underline{\Lambda}^{\mathsf{n}}}(t) \equiv 0$, and thus, the vector $\tilde{z}_{\underline{\Lambda}^{\mathsf{n}}}(t) = \hat{z}_{\underline{\Lambda}^{\mathsf{n}}}(t) - \Phi_{\underline{\Lambda}^{\mathsf{n}}}(x(t)) = v_{\underline{\Lambda}^{\mathsf{n}}}(t)$ goes to zero. On the other hand, the vector $e_{([\mathsf{p}]\setminus\underline{\Lambda})^{\mathsf{n}}}(t)$ may not be zero and the estimation error $\tilde{z}_{([\mathsf{p}]\setminus\underline{\Lambda})^{\mathsf{n}}}(t) = \hat{z}_{([\mathsf{p}]\setminus\underline{\Lambda})^{\mathsf{n}}}(t) - z_{([\mathsf{p}]\setminus\underline{\Lambda})^{\mathsf{n}}}(t) = v_{([\mathsf{p}]\setminus\underline{\Lambda})^{\mathsf{n}}}(t) + e_{([\mathsf{p}]\setminus\underline{\Lambda})^{\mathsf{n}}}(t)$ may not converge to zero. Finally, we can conclude that

$$\begin{aligned} &\|v(t)\|_\infty \leq v_{\max}(t) := \max_{\mathsf{i}\in[\mathsf{p}]} \left\{ 2M_{z,\mathsf{i}}\eta_{\mathsf{i}}(\theta_{\mathsf{i}}) e^{-\frac{\theta_{\mathsf{i}}}{4}t} \right\}, \\ &\|e(t)\|_{0^{\mathsf{n}}} \leq \mathsf{q}, \end{aligned} \tag{5.4.2}$$

by Lemma 5.2.1 and Assumption 5.1.1. As t increases, $v_{\max}(t)$ converges to zero.

Now, we present a detection mechanism for "influential" attacks. For this end, note that id_X denotes the identity function on the set X and let

$$\begin{aligned} \kappa^{n,d}_{|\Lambda|,\mathsf{q}}(\Phi^{\pi}_{\Lambda^{\mathsf{n}}}) &:= \frac{\overline{\mathsf{Lip}}(\mathsf{id}_{\mathbb{R}^{\mathsf{n}|\Lambda|}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}} \circ \Psi^{\Lambda}) + 1}{\min\left\{ \underline{\mathsf{Lip}}(\Phi^{\pi}_{\bar{\Lambda}^{\mathsf{n}}}) : \bar{\Lambda} \subset \Lambda, \ |\bar{\Lambda}| = |\Lambda| - \mathsf{q} \right\}}, \\ \kappa^{n,e}_{|\Lambda|,\mathsf{q}}(\Phi^{\pi}_{\Lambda^{\mathsf{n}}}) &:= \left(\overline{\mathsf{Lip}}(\mathsf{id}_{\mathbb{R}^{\mathsf{n}|\Lambda|}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}} \circ \Psi^{\Lambda}) + 1\right)\left(1 + \frac{\overline{\mathsf{Lip}}(\Phi^{\pi}_{\Lambda^{\mathsf{n}}})}{\min\left\{ \underline{\mathsf{Lip}}(\Phi^{\pi}_{\bar{\Lambda}^{\mathsf{n}}}) : \bar{\Lambda} \subset \Lambda, |\bar{\Lambda}| = |\Lambda| - \mathsf{q} \right\}}\right). \end{aligned}$$

The following theorem can be seen as a nonlinear counterpart of Theorem 2.2.9.

Theorem 5.4.1. Under Assumptions 5.1.1 and 5.2.1, suppose that the system (5.1.1) is $2\mathsf{q}$-redundant observable. For a given $\Lambda \subset [\mathsf{p}]$ with $|\Lambda| = \mathsf{p} - \mathsf{q}$, let $\hat{x}^{\Lambda}(t) := \Psi^{\Lambda}(\hat{z}^{\pi}_{\Lambda^{\mathsf{n}}}(t))$ and $r^{\Lambda}(t) := \hat{z}^{\pi}_{\Lambda^{\mathsf{n}}}(t) - \Phi^{\pi}_{\Lambda^{\mathsf{n}}}(\hat{x}^{\Lambda}(t))$. Then, it holds that

(i) $e^{\pi}_{\Lambda^{\mathsf{n}}}(t) \neq 0$, if

$$\left\|r^{\Lambda}(t)\right\|_\infty = \left\|\hat{z}^{\pi}_{\Lambda^{\mathsf{n}}}(t) - \Phi^{\pi}_{\Lambda^{\mathsf{n}}}(\hat{x}^{\Lambda}(t))\right\|_\infty > \overline{\mathsf{Lip}}\left(\mathsf{id}_{\mathbb{R}^{\mathsf{n}(\mathsf{p}-\mathsf{q})}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}} \circ \Psi^{\Lambda}\right) v_{\max}(t), \tag{5.4.3}$$

(ii) $\|e^{\pi}_{\Lambda^{\mathsf{n}}}(t)\|_\infty \leq \kappa^{n,e}_{|\Lambda|,\mathsf{q}}(\Phi^{\pi}_{\Lambda^{\mathsf{n}}}) v_{\max}(t)$, if

$$\left\|r^{\Lambda}(t)\right\|_\infty = \left\|\hat{z}^{\pi}_{\Lambda^{\mathsf{n}}}(t) - \Phi^{\pi}_{\Lambda^{\mathsf{n}}}(\hat{x}^{\Lambda}(t))\right\|_\infty \leq \overline{\mathsf{Lip}}\left(\mathsf{id}_{\mathbb{R}^{\mathsf{n}(\mathsf{p}-\mathsf{q})}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}} \circ \Psi^{\Lambda}\right) v_{\max}(t).$$

In the case of (ii), $\|\hat{x}^{\Lambda}(t) - x(t)\|_{\infty} \leq \kappa^{n,d}_{|\Lambda|,\mathsf{q}}(\Phi^{\pi}_{\Lambda^{\mathsf{n}}})v_{\max}(t)$. ◊

Proof. (i): It follows that

$$\begin{aligned}
&\|\hat{z}^{\pi}_{\Lambda^{\mathsf{n}}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}}(\Psi^{\Lambda}(\hat{z}^{\pi}_{\Lambda^{\mathsf{n}}}))\|_{\infty} \\
= \;& \|\hat{z}^{\pi}_{\Lambda^{\mathsf{n}}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}}(\Psi^{\Lambda}(\hat{z}^{\pi}_{\Lambda^{\mathsf{n}}})) - (\Phi^{\pi}_{\Lambda^{\mathsf{n}}}(x) - \Phi^{\pi}_{\Lambda^{\mathsf{n}}}(\Psi^{\Lambda}(\Phi^{\pi}_{\Lambda^{\mathsf{n}}}(x))))\|_{\infty} \\
= \;& \|(\mathsf{id}_{\mathbb{R}^{\mathsf{n(p-q)}}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}} \circ \Psi^{\Lambda})(\hat{z}^{\pi}_{\Lambda^{\mathsf{n}}}) - (\mathsf{id}_{\mathbb{R}^{\mathsf{n(p-q)}}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}} \circ \Psi^{\Lambda})(\Phi^{\pi}_{\Lambda^{\mathsf{n}}}(x))\|_{\infty} \\
\leq \;& \overline{\mathsf{Lip}}(\mathsf{id}_{\mathbb{R}^{\mathsf{n(p-q)}}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}} \circ \Psi^{\Lambda})\|\hat{z}^{\pi}_{\Lambda^{\mathsf{n}}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}}(x)\|_{\infty} \\
= \;& \overline{\mathsf{Lip}}(\mathsf{id}_{\mathbb{R}^{\mathsf{n(p-q)}}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}} \circ \Psi^{\Lambda})\|v^{\pi}_{\Lambda^{\mathsf{n}}} + e^{\pi}_{\Lambda^{\mathsf{n}}}\|_{\infty}.
\end{aligned}$$

Hence, if $e^{\pi}_{\Lambda^{\mathsf{n}}}(t) = 0$, then

$$\begin{aligned}
\|\hat{z}^{\pi}_{\Lambda^{\mathsf{n}}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}}(\Psi^{\Lambda}(\hat{z}^{\pi}_{\Lambda^{\mathsf{n}}}))\|_{\infty} &\leq \overline{\mathsf{Lip}}(\mathsf{id}_{\mathbb{R}^{\mathsf{n(p-q)}}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}} \circ \Psi^{\Lambda})\|v^{\pi}_{\Lambda^{\mathsf{n}}}(t)\|_{\infty} \\
&\leq \overline{\mathsf{Lip}}(\mathsf{id}_{\mathbb{R}^{\mathsf{n(p-q)}}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}} \circ \Psi^{\Lambda})v_{\max}(t).
\end{aligned}$$

This proves the claim.

(ii): Since $e^{\pi}_{\Lambda^{\mathsf{n}}}$ is (n-stacked) q-sparse, i.e., $\|e^{\pi}_{\Lambda^{\mathsf{n}}}\|_{0^{\mathsf{n}}} \leq \mathsf{q}$ where $|\Lambda| = \mathsf{p} - \mathsf{q}$, there is an index set $\bar{\Lambda} \subset \Lambda$ such that $|\bar{\Lambda}| = \mathsf{p} - 2\mathsf{q}$ and $e^{\pi}_{\bar{\Lambda}^{\mathsf{n}}} = 0$. Then, it follows from (5.4.1) that $\hat{z}^{\pi}_{\bar{\Lambda}^{\mathsf{n}}} = \Phi^{\pi}_{\bar{\Lambda}^{\mathsf{n}}}(x) + v^{\pi}_{\bar{\Lambda}^{\mathsf{n}}}$ and we have

$$\begin{aligned}
\|\hat{z}^{\pi}_{\Lambda^{\mathsf{n}}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}}(\Psi^{\Lambda}(\hat{z}^{\pi}_{\Lambda^{\mathsf{n}}}))\|_{\infty} &\geq \|\hat{z}^{\pi}_{\bar{\Lambda}^{\mathsf{n}}} - \Phi^{\pi}_{\bar{\Lambda}^{\mathsf{n}}}(\Psi^{\Lambda}(\hat{z}^{\pi}_{\Lambda^{\mathsf{n}}}))\|_{\infty} \\
&= \|\Phi^{\pi}_{\bar{\Lambda}^{\mathsf{n}}}(x) + v^{\pi}_{\bar{\Lambda}^{\mathsf{n}}} - \Phi^{\pi}_{\bar{\Lambda}^{\mathsf{n}}}(\hat{x}^{\Lambda})\|_{\infty} \\
&\geq \underline{\mathsf{Lip}}(\Phi^{\pi}_{\bar{\Lambda}^{\mathsf{n}}})\|x - \hat{x}^{\Lambda}\|_{\infty} - \|v^{\pi}_{\bar{\Lambda}^{\mathsf{n}}}\|_{\infty} \\
&\geq \underline{\mathsf{Lip}}(\Phi^{\pi}_{\bar{\Lambda}^{\mathsf{n}}})\|x - \hat{x}^{\Lambda}\|_{\infty} - v_{\max}.
\end{aligned}$$

Therefore, from the assumption, we obtain

$$\overline{\mathsf{Lip}}(\mathsf{id}_{\mathbb{R}^{\mathsf{n(p-q)}}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}} \circ \Psi^{\Lambda})v_{\max} \geq \underline{\mathsf{Lip}}(\Phi^{\pi}_{\bar{\Lambda}^{\mathsf{n}}})\|\hat{x}^{\Lambda} - x\|_{\infty} - v_{\max}.$$

From $2\mathsf{q}$-redundant observability, it follows that $\underline{\mathsf{Lip}}(\Phi^{\pi}_{\bar{\Lambda}^{\mathsf{n}}}) > 0$ for any $\bar{\Lambda}$ such that $|\bar{\Lambda}| = \mathsf{p} - 2\mathsf{q}$ since $\Phi^{\pi}_{\bar{\Lambda}^{\mathsf{n}}}$ is bi-Lipschitz on X by Lemma 5.1.1. Therefore, we have

$$\|\hat{x}^{\Lambda} - x\|_{\infty} \leq \frac{\overline{\mathsf{Lip}}(\mathsf{id}_{\mathbb{R}^{\mathsf{n(p-q)}}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}} \circ \Psi^{\Lambda}) + 1}{\underline{\mathsf{Lip}}(\Phi^{\pi}_{\bar{\Lambda}^{\mathsf{n}}})} v_{\max}.$$

On the other hand, it is also easily obtained that

$$\begin{aligned}
\overline{\mathsf{Lip}}(\mathsf{id}_{\mathbb{R}^{\mathsf{n(p-q)}}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}} \circ \Psi^{\Lambda}) v_{\max} &\geq \|\hat{z}^{\pi}_{\Lambda^{\mathsf{n}}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}}(\Psi^{\Lambda}(\hat{z}^{\pi}_{\Lambda^{\mathsf{n}}}))\|_{\infty} \\
&= \|\Phi^{\pi}_{\Lambda^{\mathsf{n}}}(x) + v^{\pi}_{\Lambda^{\mathsf{n}}} + e^{\pi}_{\Lambda^{\mathsf{n}}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}}(\Psi^{\Lambda}(\hat{z}^{\pi}_{\Lambda^{\mathsf{n}}}))\|_{\infty} \\
&\geq -\|\Phi^{\pi}_{\Lambda^{\mathsf{n}}}(x) - \Phi^{\pi}_{\Lambda^{\mathsf{n}}}(\hat{x}^{\Lambda})\|_{\infty} - v_{\max} + \|e^{\pi}_{\Lambda^{\mathsf{n}}}\|_{\infty} \\
&\geq -\overline{\mathsf{Lip}}(\Phi^{\pi}_{\Lambda^{\mathsf{n}}})\|x - \hat{x}^{\Lambda}\|_{\infty} - v_{\max} + \|e^{\pi}_{\Lambda^{\mathsf{n}}}\|_{\infty}.
\end{aligned}$$

Hence,

$$\|e^{\pi}_{\Lambda^{\mathsf{n}}}\|_{\infty} \leq \left(\overline{\mathsf{Lip}}(\mathsf{id}_{\mathbb{R}^{\mathsf{n(p-q)}}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}} \circ \Psi^{\Lambda}) + 1\right) \left(1 + \frac{\overline{\mathsf{Lip}}(\Phi^{\pi}_{\Lambda^{\mathsf{n}}})}{\underline{\mathsf{Lip}}(\Phi^{\pi}_{\Lambda^{\mathsf{n}}})}\right) v_{\max}.$$

This completes the proof. □

If Λ in Theorem 5.4.1 is replaced by $[\mathsf{p}]$ and the condition $|\bar{\Lambda}| = \mathsf{p} - 2\mathsf{q}$ in the proof is replaced by the condition $|\bar{\Lambda}| = \mathsf{p} - \mathsf{q}$, we can easily derive the following corollary.

Corollary 5.4.2. Under Assumptions 5.1.1 and 5.2.1, suppose that the system (5.1.1) is q-redundant observable. Let $\hat{x}(t) := \Psi(\hat{z}(t))$ and $r(t) := \hat{z}(t) - \Phi(\hat{x}(t))$. Then, it holds that
(i) $e(t) \neq 0$, if

$$\|r(t)\|_{\infty} = \|\hat{z}(t) - \Phi(\hat{x}(t))\|_{\infty} > \overline{\mathsf{Lip}}(\mathsf{id}_{\mathbb{R}^{\mathsf{np}}} - \Phi \circ \Psi) v_{\max}(t), \tag{5.4.4}$$

(ii) $\|e(t)\|_{\infty} \leq \kappa^{n;e}_{\mathsf{p,q}}(\Phi) v_{\max}(t)$, if

$$\|r(t)\|_{\infty} = \|\hat{z}(t) - \Phi(\hat{x}(t))\|_{\infty} \leq \overline{\mathsf{Lip}}(\mathsf{id}_{\mathbb{R}^{\mathsf{np}}} - \Phi \circ \Psi) v_{\max}(t).$$

In the case of (ii), $\|\hat{x}(t) - x(t)\|_{\infty} \leq \kappa^{n;d}_{\mathsf{p,q}}(\Phi) v_{\max}(t)$. ◊

Inequality (5.4.4) is the key to the detection of sensor attack. It is noted that both sides of (5.4.4) can be readily evaluated since all the quantities are available at all time $t \geq 0$. By checking (5.4.4), one can detect sensor attack. Of course, violation of (5.4.4) does not necessarily imply no sensor attack. However, even when there is an attack, its effect on the state estimation is limited as seen in the theorem since $v_{\max}(t)$ converges to zero as t increases by (5.4.2). This case happens when the size of error e is so small that the distinction between the transient error v and the error e caused by attack is not possible.

Remark 5.4.1. Since $v_{\max}(t)$ tends to zero as time goes to infinity by (5.4.2), one may want to replace $v_{\max}(t)$ in Theorem 5.4.1 and Corollary 5.4.2 with $\bar{v}_{\max}(t) := \max\{v_{\max}(t), \delta\}$ where δ is a small positive constant. This is because there is measurement noise in practice and one does not want the detection by (5.4.3) or (5.4.4) to be corrupted by noise when t is large so that $v_{\max}(t)$ is very small. The value of δ is chosen such that the effect of the measurement noise on estimation error is less than δ. $\diamondsuit$

5.4.2 Attack-Resilient State Estimation

Note that Theorem 5.4.1 explains the detection of sensor attack for a given subset $\Lambda \subset [\mathsf{p}]$, while Corollary 5.4.2 detects for the whole set $[\mathsf{p}]$. Thus, the same discussion on (5.4.4) also applies to (5.4.3). When (5.4.3) is violated, we suppose that there is no influential attack on $\bar{y}_{\mathsf{i}}$ for $\mathsf{i} \in \Lambda$, and the state estimates $\hat{z}_{\mathsf{i}}$ for $\mathsf{i} \in \Lambda$ are trustful. By repeating (5.4.3) with all subsets $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| = \mathsf{p} - \mathsf{q}$, one can always find trustful set of sensors since at most q sensors are compromised. This is the main idea of the resilient state estimation scheme presented in this section.

In practice where the proposed estimator is implemented in a digital computer, the inequality (5.4.3) for attack detection is checked at every sampling instant. Hence, one idea to estimate the state $x(t)$ under q-sparse sensor attack is to prepare all different $\binom{\mathsf{p}}{\mathsf{p}-\mathsf{q}}$ index sets $\Lambda \subset [\mathsf{p}]$ such that $|\Lambda| = \mathsf{p}-\mathsf{q}$, and test (5.4.3) for all of them during each sampling period. Then, one can always find at least one index set Λ^* that violates (5.4.3), which implies that there is no influential attack on $\bar{y}_{\mathsf{i}}$ for $\mathsf{i} \in \Lambda^*$. Therefore, the true state $x(t)$ is estimated by $\hat{x}(t) = \Psi^{\Lambda^*}(\hat{z}^{\pi}_{\Lambda^{*\mathsf{n}}}(t))$ with the estimation error discussed in Theorem 5.4.1. However, in order to monitor any changes of influential attacks, one should keep testing (5.4.3) at every sampling instant with all index sets $\Lambda \subset [\mathsf{p}]$ satisfying $|\Lambda| = \mathsf{p} - \mathsf{q}$, which is computationally heavy. This burden may be relieved by introducing a simple switching algorithm as in the following theorem with a proportional constant

$$\kappa^{n,c}_{\mathsf{p},\mathsf{q}}(\Phi) := \frac{\max\left\{\overline{\mathsf{Lip}}\left(\mathsf{id}_{\mathbb{R}^{\mathsf{n}(\mathsf{p}-\mathsf{q})}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}} \circ \Psi^{\Lambda}\right) : |\Lambda| = \mathsf{p}-\mathsf{q}\right\} + 1}{\min\left\{\underline{\mathsf{Lip}}(\Phi^{\pi}_{\bar{\Lambda}^{\mathsf{n}}}) : |\bar{\Lambda}| = \mathsf{p}-2\mathsf{q}\right\}}.$$

Theorem 5.4.3. Under Assumptions 5.1.1 and 5.2.1, suppose that system (5.1.1) is $2\mathsf{q}$-redundant observable. Define the index sets $\Lambda_{\mathsf{i}} \subset [\mathsf{p}]$ for $\mathsf{i} = 1, 2, \cdots, \binom{\mathsf{p}}{\mathsf{q}}$ such that

$$\left\{\Lambda_1, \Lambda_2, \cdots, \Lambda_{\binom{\mathsf{p}}{\mathsf{q}}}\right\}$$

is the same as the set $\{\Lambda \subset [\mathsf{p}] : \ |\Lambda| = \mathsf{p}-\mathsf{q}\}$. Let $\hat{x}^{\Lambda_{\mathsf{i}}}(t) := \Psi^{\Lambda_{\mathsf{i}}}(\hat{z}^{\pi}_{\Lambda^{\mathsf{n}}_{\mathsf{i}}}(t))$ and $r^{\Lambda_{\mathsf{i}}}(t) := \hat{z}^{\pi}_{\Lambda^{\mathsf{n}}_{\mathsf{i}}}(t) - \Phi^{\pi}_{\Lambda^{\mathsf{n}}_{\mathsf{i}}}(\hat{x}^{\Lambda_{\mathsf{i}}}(t))$. Consider a switching signal $\sigma(t)$ generated from $\sigma(0) = 1$ by the update rule

$$\sigma(t^+) \leftarrow \left(\sigma(t) \ \mathsf{mod} \ \binom{\mathsf{p}}{\mathsf{p}-\mathsf{q}}\right) + 1$$

whenever

$$\|r^{\Lambda_{\sigma(t)}}(t)\|_\infty > \overline{\mathsf{Lip}}\left(\mathsf{id}_{\mathbb{R}^{\mathsf{n}(\mathsf{p}-\mathsf{q})}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}_{\sigma(t)}} \circ \Psi^{\Lambda_{\sigma(t)}}\right) v_{\max}(t). \tag{5.4.5}$$

Then, the state estimate for $x(t)$ is given by

$$\hat{x}(t) = \hat{x}^{\Lambda_{\sigma(t)}}(t)$$

which has the property

$$\|\hat{x}(t) - x(t)\|_\infty \leq \kappa^{n,c}_{\mathsf{p},\mathsf{q}}(\Phi) v_{\max}(t)$$

for all $t \geq 0$ except at the switching times of $\sigma(t)$. ♢

Proof. According to Assumption 5.1.1, the signal $e(t)$ is q-sparse for all t and there exists a natural number $\mathsf{m} \leq \binom{\mathsf{p}}{\mathsf{p}-\mathsf{q}}$ such that $e_{\Lambda_{\mathsf{m}}}(t)$ is identically zero. It implies that the inequality

$$\|r^{\Lambda_{\mathsf{m}}}(t)\|_\infty \leq \overline{\mathsf{Lip}}(\mathsf{id}_{\mathbb{R}^{\mathsf{n}(\mathsf{p}-\mathsf{q})}} - \Phi^{\pi}_{\Lambda^{\mathsf{n}}_{\mathsf{m}}} \circ \Psi^{\Lambda_{\mathsf{m}}}) v_{\max}(t)$$

holds for every $t \geq 0$. Therefore, according to the update rule of σ, there will be at most $\binom{\mathsf{p}}{\mathsf{p}-\mathsf{q}} - 1$ times of consecutive switching of $\sigma(t)$ at the same time t until (5.4.5) is violated. (Note that σ does not necessarily become the same as m.) Then, the proof is completed from the upper bound of the estimation error in Theorem 5.4.1, by considering that the set $\Lambda \subset [\mathsf{p}]$ with $|\Lambda| = \mathsf{p} - \mathsf{q}$ is arbitrary. □

Update of the switching signal σ in Theorem 5.4.3 is understood as follows. Whenever the value of $\sigma(t)$ is updated at time t, the condition (5.4.5) is checked again at the same time t with the updated $\sigma(t^+)$ until the inequality is violated (i.e., consecutive update can occur). This repeated update does not occur infinitely as shown in the proof. In practice, since the estimator is implemented by digital computer, a few sampling delay will occur by the consecutive updates and, during this delay, the state estimation is corrupted, which can be seen in the simulation results in the next section.

5.5 Simulation Results: Numerical Example

We consider a numerical example of the system (5.1.1) given as

$$\dot{x} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} -2x_1 - x_2^3 \\ -x_2 \\ -x_2 \cos x_2 + \sin x_2 - x_3 \end{bmatrix} + \begin{bmatrix} 1 + 3x_2^2 \\ 1 \\ \cos x_2 \end{bmatrix} u =: f(x) + g(x)u$$

$$\bar{y} = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \bar{y}_3 \\ \bar{y}_4 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 - x_2^3 - \sin x_2 + x_3 \\ x_1 + \sin x_2 - x_2^3 - x_3 \\ -x_1 + x_2^3 + x_2 \\ -x_2 - \sin x_2 + x_3 \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} =: h(x) + a$$

where $u(t) = 0.25 \sin(0.2\pi t) - 0.1$, for which it is verified that the state x remains in $X = \{x \in \mathbb{R}^3 : \|x\|_\infty \leq 0.5\}$ with sufficiently small initial conditions. To make the situation more realistic, on top of the attack signal a_{i}, a Gaussian distributed noise n_{i} of zero mean with the power spectral density of 10^{-8} is additionally introduced to corrupt the sensor measurement $\bar{y}_{\mathsf{i}}$. For this system, the function

Table 5.1: Left inverse functions of Φ_{Λ_i}, $i = 1, 2, 3, 4$

$\Lambda_1 = \{2,3,4\}$ $\Phi^{-1}_{\{2,3,4\}} : \mathbb{R}^5 \to \mathbb{R}^3$	$\begin{bmatrix} z_{3,1} + z_{3,2} + (2z_{3,1} + z_{3,2})^3 \\ 2z_{3,1} + z_{3,2} \\ -2z_{2,1} - z_{2,2} + \sin(2z_{3,1} + z_{3,2}) \end{bmatrix}$
$\Lambda_2 = \{1,3,4\}$ $\Phi^{-1}_{\{1,3,4\}} : \mathbb{R}^5 \to \mathbb{R}^3$	$\begin{bmatrix} -z_{1,1} - z_{1,2} + (2z_{3,1} + z_{3,2})^3 \\ 2z_{3,1} + z_{3,2} \\ \frac{1}{2}(2z_{1,1} + z_{1,2} + z_{4,1}) + \sin(2z_{3,1} + z_{3,2}) \end{bmatrix}$
$\Lambda_3 = \{1,2,4\}$ $\Phi^{-1}_{\{1,2,4\}} : \mathbb{R}^5 \to \mathbb{R}^3$	$\begin{bmatrix} -z_{1,1} - z_{1,2} + (-2z_{2,1} - z_{2,2} - z_{4,1})^3 \\ -2z_{2,1} - z_{2,2} - z_{4,1} \\ -2z_{2,1} - z_{2,2} + \sin(-2z_{2,1} - z_{2,2} - z_{4,1}) \end{bmatrix}$
$\Lambda_4 = \{1,2,3\}$ $\Phi^{-1}_{\{1,2,3\}} : \mathbb{R}^6 \to \mathbb{R}^3$	$\begin{bmatrix} -z_{1,1} - z_{1.2} + (2z_{3,1} + z_{3,2})^3 \\ 2z_{3,1} + z_{3,2} \\ -2z_{2,1} - z_{2,2} + \sin(2z_{3,1} + z_{3,2}) \end{bmatrix}$

$\Phi : X \to \mathbb{R}^7$ in (5.2.4) and (5.3.2) is computed by

$$\Phi(x) = \begin{bmatrix} \Phi_1(x) \\ \Phi_2(x) \\ \Phi_3(x) \\ \Phi_4(x) \end{bmatrix} = \begin{bmatrix} h_1(x) \\ L_f h_1(x) \\ h_2(x) \\ L_f h_2(x) \\ h_3(x) \\ L_f h_3(x) \\ h_4(x) \end{bmatrix} = \begin{bmatrix} x_1 + x_2 - x_2^3 - \sin x_2 + x_3 \\ -2x_1 + \sin x_2 - x_2 + 2x_2^3 - x_3 \\ x_1 + \sin x_2 - x_2^3 - x_3 \\ -2x_1 - \sin x_2 + 2x_2^3 + x_3 \\ -x_1 + x_2^3 + x_2 \\ 2x_1 - x_2 - 2x_2^3 \\ -x_2 - \sin x_2 + x_3 \end{bmatrix} =: \begin{bmatrix} z_{1,1} \\ z_{1,2} \\ z_{2,1} \\ z_{2,2} \\ z_{3,1} \\ z_{3,2} \\ z_{4,1} \end{bmatrix} =: z$$

where additional zeros in (5.3.1) are excluded for simplicity. Accordingly, the notation in this simulation section is slightly abused by eliminating the additional zeros. It can be seen that each Φ_i transforms the system into uniformly observable subsystem with respect to $\bar{y}_i$, and the stack of all observable parts z remains in the set $\mathcal{Z} := \{z \in \mathbb{R}^7 : \|z\|_\infty \le 2\}$. One can also ensure that the above system is 2-redundant observable by verifying that $\Phi_{\bar{\Lambda}^n}$ is an injective immersion for every $|\bar{\Lambda}| = 2$.

Since the system is 2-redundant observable, resilient state estimation is possible under up to 1-sparse attack. Therefore, let us suppose an attack scenario depicted in Fig. 5.1. A square wave a_2 is injected to the second sensor at $t = 4$sec. Partial observers for individual uniformly observable subsystems are designed with $\theta_{\mathsf{i}} = 16$ for $\mathsf{i} \in [4]$, which yields $v_{\max}(t) = 168e^{-4t}$. For the recovery of state x, we choose four left inverse functions $\Phi_{\Lambda_{\mathsf{i}}}^{-1}$ for each $\mathsf{i} = 1, \cdots, 4$, where $\Lambda_{\mathsf{i}} = [4] - \{\mathsf{i}\}$, as in Table 5.1. With these functions, as in (5.3.4) and (5.3.5), Lipschitz-extended left inverse of $\Phi_{\Lambda_{\mathsf{i}}}$ is obtained by

$$\begin{aligned} \Psi^{\Lambda_{\mathsf{i}}} &: \mathbb{R}^{N_{\mathsf{i}}} \to X \\ &z^{\Lambda_{\mathsf{i}}} \mapsto \mathsf{sat}(\Phi_{\Lambda_{\mathsf{i}}}^{-1}(\mathsf{sat}(z^{\Lambda_{\mathsf{i}}}, 2)), 0.5) \end{aligned}$$

for each i, where $z^{\Lambda_{\mathsf{i}}}$ is a stacked vector of $z_{\mathsf{j},\mathsf{k}}$'s for all $\mathsf{j} \in \Lambda_{\mathsf{i}}$ and N_{i} is the dimension of $z^{\Lambda_{\mathsf{i}}}$. It is noted that the Lipschitz constant of $\Psi^{\Lambda_{\mathsf{i}}}$ on $\mathbb{R}^{N_{\mathsf{i}}}$ is less than or equal to the Lipschitz constant of $\Phi_{\Lambda_{\mathsf{i}}}^{-1}$ on $\mathcal{Z}^{\Lambda_{\mathsf{i}}} = \{z^{\Lambda_{\mathsf{i}}} \in \mathbb{R}^{N_{\mathsf{i}}} : \|z^{\Lambda_{\mathsf{i}}}\|_\infty \leq 2\}$ due to the two saturation functions, and the Lipschitz constant of Φ is greater than or equal to the Lipschitz constant of $\Phi_{\Lambda_{\mathsf{i}}}$. Hence, for simplicity, we take a conservative bound for the right hand side of the condition (5.4.5) as

$$\overline{\mathsf{Lip}}\left(\mathsf{id}_{\mathbb{R}^{N_\sigma}} - \Phi_{\Lambda_\sigma} \circ \Psi^{\Lambda_\sigma}\right) \leq 1 + \overline{\mathsf{Lip}}(\Phi) \times \max_{\mathsf{i}\in[4]} \left\{\overline{\mathsf{Lip}}(\Phi_{\Lambda_{\mathsf{i}}}^{-1}|_{\mathcal{Z}^{\Lambda_{\mathsf{i}}}})\right\} \leq 1 + 7 \times 770.$$

By this simplification, the upper bound of the estimation error in Theorem 5.4.3 is increased, but it will be seen in the simulations that this does not sacrifice much after a sufficiently long time because the exponential term in $v_{\max}$ dominantly converges to zero. For the simulation, $\bar{v}_{\max}(t)$ is used instead of $v_{\max}(t)$ with $\delta = 0.05$ due to the presence of measurement noise (see Remark 5.4.1).

An attack signal is illustrated in Fig. 5.1, which depicts that adversaries inject the attack at $t = 4$sec so that the second sensor is compromised. Therefore, the switching signal $\sigma(t)$ jumps from 1 to 2 at $t = 4$sec, that is, it changes the selected index set from $\Lambda_1 = \{1, 2, 3\}$ to $\Lambda_2 = \{1, 3, 4\}$ immediately after the attack is detected. As a result, Figures 5.2, 5.3, and 5.4 show state trajectories $x_1(t)$, $x_2(t)$, $x_3(t)$, and their estimates. They demonstrate the attack-resilient property of our estimation algorithm. For a short period of time after the adversaries start to attack, the state estimates have sharp peak by the attack vector, but it is restored soon by the proposed observation scheme.

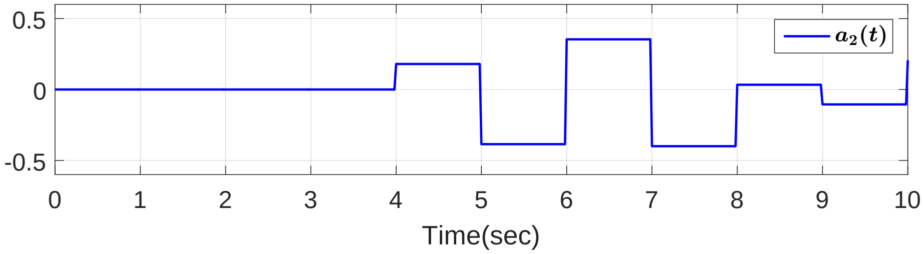


Figure 5.1: Plot of attack $a_2(t)$.

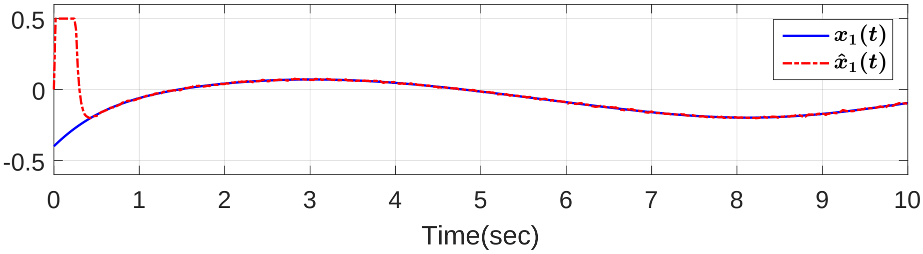


Figure 5.2: Plot of state $x_1(t)$ and its estimate $\hat{x}_1(t)$.

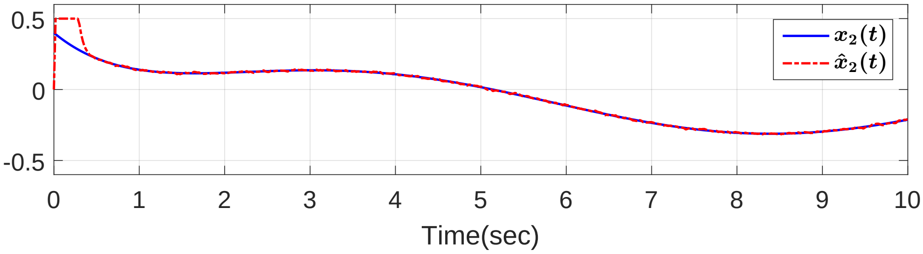


Figure 5.3: Plot of state $x_2(t)$ and its estimate $\hat{x}_2(t)$.

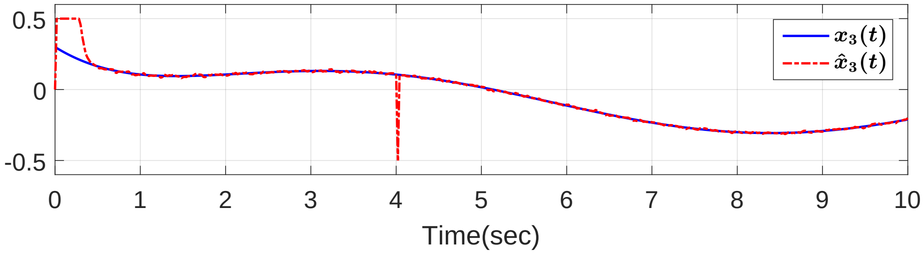


Figure 5.4: Plot of state $x_3(t)$ and its estimate $\hat{x}_3(t)$.

Chapter 6

Conclusion

This chapter summarizes the main results of this dissertation that have been addressed so far, and provides the future issues.

6.1 Summary

This dissertation is concerned with security of control systems under sensor attacks. Specifically, for linear systems, the notion of redundant detectability (or, asymptotic redundant observability) is introduced that explains in a unified manner existing security notions such as dynamic security index, attack detectability, and observability under attacks when only disruptive sensor attacks are taken into account. Indeed, equivalent conditions between the redundant detectability and the existing security related notions are derived and presented. Then, by utilizing a bank of partial observers based on Kalman detectability decomposition and a decoder exploiting the redundant detectability, a practical and efficient estimator design algorithm is proposed to enhance the resilience of control systems in the presence of sensor attacks as well as process disturbances and measurement noises. The main assumption on the attack signal is q-sparsity while both of bounded and Gaussian distributed disturbances/noises are considered: A Luenberger observer is used for the bounded case; A Kalman filter is designed for the Gaussian distributed case. The proposed state estimation scheme substantially improves computational efficiency with much less required memory compared to those of the existing results. In addition, the linear resilient state estimation algorithm is also generalized for a class of nonlinear systems called uniformly observable nonlinear systems. Partial observers are designed by a high gain observer and a nonlinear error correcting problem is solved.

A theoretical analysis to examine the security problems on CPSs under sensor attacks, is conducted in Chapter 3. It has been shown that the measurement redundancy for the left invertibility of the observability matrix determines the redundant observability of a given LTI system. Furthermore, the redundant detectability, which is a weaker notion than the redundant observability, is introduced and it is closely related to the security problems of control systems under disruptive sensor attacks, i.e., when the attack signal which does not converge to zero, is considered, the redundant detectability plays a key role in security related problems. To summarize, $2\mathsf{q}$-redundant detectability implies that the numbers of detectable and correctable sensor attacks are $2\mathsf{q}$ and q, respectively. In addition, the dynamic security index, the minimum number of sensor attacks to remain undetectable, is $2\mathsf{q}+1$ and a simple method to compute the index utilizing unstable eigenvectors only, is also suggested.

In Chapter 4, assuming that the measurement data injection attack is q-sparse and the disturbances/noises are bounded (or, Gaussian distributed), an attack-resilient and robust (or, suboptimal) state estimation scheme based on a bank of partial observers has been proposed under $2\mathsf{q}$-redundant detectability. By reducing the search space to a finite set in the optimization process and combining the attack monitoring mechanism to the error correction algorithm, we can mitigate the NP-hardness of ℓ_0 minimization problem in terms of time computational complexity. Furthermore, with the help of the Kalman detectability decomposition used to construct the partial observers, the proposed estimator is scalable in terms of memory space complexity. For bounded disturbances/noises, a Luenburger observer is designed for the partial observer, and the estimation error bound is explicitly given by the bounds on disturbances/noises, which guarantees the robustness of the proposed estimator. Lastly, the estimator equipped with the Kalman filter based information fusion scheme, identifies the attack-free sensors based-on the maximum likelihood decision rule and computes the minimum variance unbiased estimator so that the final estimate turns out to be suboptimal.

In Chapter 5, we have proposed a solution to the resilient state estimation problem for uniformly observable nonlinear systems with redundant sensors. A switching algorithm that makes use of the detection algorithm of sensor attacks, is designed to search for a combination of uncompromised sensors successfully and to generate correct estimates which are insensitive to sparse malicious attacks. The uniform observability decomposition which is an analogous concept of Kalman observability decomposition for linear systems, is utilized to design a high gain observer for each single output and it estimates the observable portion of system state. Then, a nonlinear error correcting problem is solved by collecting all the information from those partial observers and by exploiting redundancy. Finally, a computationally efficient on-line monitoring scheme is presented for attack detection, and an algorithm for resilient state estimation is provided based on the attack detection scheme.

6.2 Future Works

One of the key future research related to the study of this dissertation is to develop a distributed attack-resilient estimator. Due to the universal use of sensor networks, there is an increasing need to estimate the state of dynamical systems through geographically dispersed sensors. Distributed estimation is to estimate the state of dynamical systems via information exchange with its neighbors [34,38,52,61,65]. Each sensor node assumes no information about the global structure and measurement data, but can access its own measurements and local information through its neighbors. Assuming that there is no centralized device to monitor the measurement data of all sensors, a distributed state estimation scheme needs to be developed when some of sensors are corrupted by adversaries. Related results include [85] and [66], but they assume that all nodes require some global information, such as complete knowledge of the network topology. One of the main assumptions of the centralized resilient estimation algorithm is redundancy of measurements called $2\mathsf{q}$-redundant observability. The redundancy concept extends to the notion of *network robustness* for arbitrary directed graphs [41,99,100]. That is, network robustness is a fundamental property

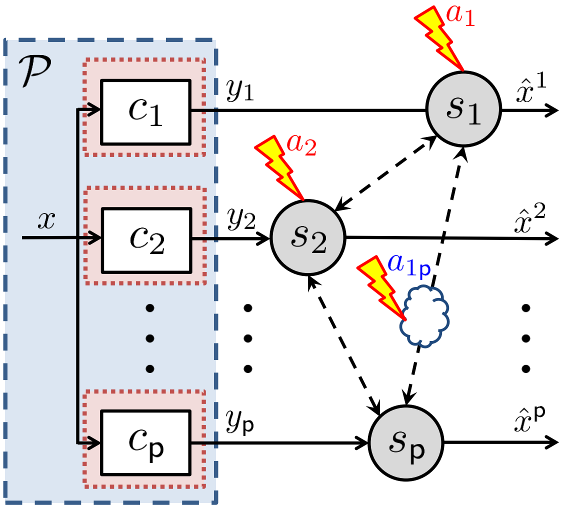


Figure 6.1: Configuration of the distributed sensor network and attack scenario.

for formulating the redundancy concept of direct information exchange between nodes in a network and for analyzing the behavior of distributed algorithms that use only local information.

The configuration of the distributed state estimation problem is depicted in Fig. 6.1. Note that sensor nodes labeled s_i's, which measure the sensing outputs y_i's, are geographically dispersed and connected to nearby nodes through network communication characterized by the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V} = [\mathsf{p}]$. In addition to sensing, these sensor nodes are equipped with computing devices that can implement data fusion protocols. In this situation, an iterative algorithm based on robust graphs which updates the state estimate and the unobservable subspace at each time step relying solely on information obtained from neighbors, needs to be developed and it is an interesting research direction.

Bibliography

[1] S. Amin, A. A. Cárdenas, and S. S. Sastry, "Safe and secure networked control systems under denial-of-service attacks," in *Proceedings of 12th International Workshop on Hybrid Systems: Computation and Control*, 2009, pp. 31–45.

[2] J. Back, J. Kim, C. Lee, G. Park, and H. Shim, "Enhancement of security against zero dynamics attack via generalized hold," in *Proceedings of 56th IEEE Conference on Decision and Control*, 2017, pp. 1350–1355.

[3] R. Baheti and H. Gill, "Cyber-physical systems," *The Impact of Control Technology*, vol. 12, pp. 161–166, 2011.

[4] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[5] G. Basile and G. Marro, "On the observability of linear, time-invariant systems with unknown inputs," *Journal of Optimization Theory and Applications*, vol. 3, no. 6, pp. 410–415, 1969.

[6] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, 2004.

[7] B. Brumback and M. Srinath, "A chi-square test for fault-detection in Kalman filters," *IEEE Transactions on Automatic Control*, vol. 32, no. 6, pp. 552–554, 1987.

[8] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[9] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[10] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9–10, pp. 589–592, 2008.

[11] Y. Chen, S. Kar, and J. M. F. Moura, "Cyber-physical systems: Dynamic sensor attacks and strong observability," in *Proceedings of 40th IEEE International Conference on Acoustics, Speech and Signal Processing*, 2015, pp. 1752–1756.

[12] M. S. Chong, M. Wakaiki, and J. P. Hespanha, "Observability of linear systems under adversarial attacks," in *Proceedings of 2015 American Control Conference*, 2015, pp. 2439–2444.

[13] R. N. Clark, "Instrument fault detection," *IEEE Transactions on Aerospace Electronic Systems*, vol. 14, pp. 456–465, 1978.

[14] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, "Introduction to compressed sensing," in *Compressed Sensing: Theory and Applications.* Cambridge University Press, 2012.

[15] D. Dolev, N. A. Lynch, S. S. Pinter, E. W. Stark, and W. E. Weihl, "Reaching approximate agreement in the presence of faults," *Journal of the ACM*, vol. 33, no. 3, pp. 499–516, 1986.

[16] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[17] D. L. Donoho, M. Elad, and V. N. Temlyakov, “Stable recovery of sparse overcomplete representations in the presence of noise,” *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 6–18, 2006.

[18] A. Dutta and C. Langbort, “Confiscating flight control system by stealthy output injection attack,” *Journal of Aerospace Information Systems*, vol. 14, no. 4, pp. 203–213, 2017.

[19] 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.

[20] P. M. Frank, “Fault diagnosis in dynamic systems via state estimation–A survey,” *System Fault Diagnostics, Reliability and Related Knowledge-based Approaches*, vol. 1, pp. 35–98, 1987.

[21] ——, “Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: A survey and some new results,” *Automatica*, vol. 26, no. 3, pp. 459–474, 1990.

[22] J. P. Gauthier and G. Bornard, “Observability for any $u(t)$ of a class of nonlinear systems,” *IEEE Transactions on Automatic Control*, vol. 26, no. 4, pp. 922–926, 1981.

[23] J. P. Gauthier and I. Kupka, *Deterministic Observation Theory and Applications.* Cambridge University Press, 2001.

[24] J. P. Gauthier, H. Hammouri, and S. Othman, “A simple observer for nonlinear systems applications to bioreactors,” *IEEE Transactions on Automatic Control*, vol. 37, no. 6, pp. 875–880, 1992.

[25] J. Gertle, “Analytical redundancy methods in fault detection and isolation,” in *Proceedings of IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes*, 1991, pp. 9–21.

[26] R. A. Gupta and M.-Y. Chow, “Networked control system: Overview and research trends,” *IEEE Transactions on Industrial Electronics*, vol. 57, no. 7, pp. 2527–2535, 2010.

[27] V. Guruswami, J. R. Lee, and A. Wigderson, “Euclidean sections of ℓ_1^n with sublinear randomness and error-correction over the reals,” in *Proceedings of 11th International Workshop on APPROX and 12th International Workshop on RANDOM*, vol. 5171 of *Lecture Notes in Computer Science*, Springer-Verlag, 2008, pp. 444–454.

[28] J. M. Hendrickx, K. H. Johansson, R. M. Jungers, H. Sandberg, and K. C. Sou, “Efficient computations of a security index for false data attacks in power networks,” *IEEE Transactions on Automatic Control*, vol. 59, no. 12, pp. 3194–3208, 2014.

[29] Q. Hu, D. Fooladivanda, Y. H. Chang, and C. J. Tomlin, “Secure state estimation for nonlinear power systems under cyber attacks,” in *Proceedings of 2017 American Control Conference*, 2017, pp. 2779—-2784.

[30] I. Hwang, S. Kim, Y. Kim, and C. E. Seah, “A survey of fault detection, isolation, and reconfiguration methods,” *IEEE Transactions on Control Systems Technology*, vol. 18, no. 3, pp. 636–653, 2010.

[31] H. Jeon, S. Aum, H. Shim, and Y. Eun, “Resilient state estimation for control systems using multiple observers and median operation,” *Mathematical Problems in Engineering*, vol. 2016, Article ID 3750264, 2016.

[32] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory.* Prentice Hall PTR, 1993.

[33] ——, *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory.* Prentice Hall PTR, 1993.

[34] J. Kim, H. Shim, and J. Wu, "On distributed optimal Kalman-Bucy filtering by averaging dynamics of heterogeneous agents," in *Proceedings on 55th IEEE Conference on Decision and Control*, 2016, pp. 6309–6314.

[35] J. Kim, G. Park, H. Shim, and Y. Eun, "Zero-stealthy attack for sampled-data control systems: The case of faster actuation than sensing," in *Proceedings of 55th IEEE Conference on Decision and Control*, 2016, pp. 5956–5961.

[36] J. Kim, C. Lee, H. Shim, J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Encrypting controller using fully homomorphic encryption for security of cyber-physical pystems," in *Proceedings of 6th IFAC Workshop on Distributed Estimation and Control in Networked Systems*, vol. 49, no. 22, 2016, pp. 175–180.

[37] J. Kim, C. Lee, H. Shim, Y. Eun, and J. H. Seo, "Detection of sensor attack and resilient state estimation for uniformly observable nonlinear systems," in *Proceedings of 55th IEEE Conference on Decision and Control*, 2016, pp. 1297–1302.

[38] T. Kim, H. Shim, and D. D. Cho, "Distributed luenberger observer design," in *Proceedings on 55th IEEE Conference on Decision and Control*, 2016, pp. 6928–6933.

[39] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 645–658, 2011.

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

[41] H. J. LeBlanc, H. Zhang, X. Koutsoukos, and S. Sundaram, "Resilient asymptotic consensus in robust networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 4, pp. 766–781, 2013.

[42] C. Lee, H. Shim, and Y. Eun, "Secure and robust state estimation under sensor attacks, measurement noise and process disturbances: Observer-based combinatorial approach," in *Proceedings of 14th European Control Conference*, 2015, pp. 1866–1871.

[43] E. A. Lee, "Cyber physical systems: Design challenges," in *Proceedings of 11th IEEE International Symposium on Object Oriented Real-Time Distributed Computing*, 2008, pp. 363–369.

[44] X. Liu, Y. Mo, and E. Garone, "Secure dynamic state estimation by decomposing Kalman filter," in *Proceedings of 20th IFAC World Congress*, 2017, pp. 7612–7617.

[45] 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. 13:1–13:33, 2011.

[46] R. E. Lyons and W. Vanderkulk, "The use of triple-modular redundancy to improve computer reliability," *IBM Journal of Research and Development*, vol. 6, no. 2, pp. 200–209, 1962.

[47] K. Manandhar, X. Cao, F. Hu, and Y. Liu, "Detection of faults and attacks including false data injection attack in smart grid using Kalman filter," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 4, pp. 370–379, 2014.

[48] R. K. Mehra and J. Peschon, "An innovations approach to fault detection and diagnosis in dynamic systems," *Automatica*, vol. 7, no. 5, pp. 637–640, 1971.

[49] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Uniform uncertainty principle for Bernoulli and subgaussian ensembles," *Constructive Approximation*, vol. 28, no. 3, pp. 277–289, 2008.

[50] F. Miao, Q. Zhu, M. Pajic, and G. J. Pappas, "Coding schemes for securing cyber-physical systems against stealthy data injection attacks,"

IEEE Transactions on Control of Network Systems, vol. 4, no. 1, pp. 106–117, 2017.

[51] S. Mishra, Y. Shoukry, N. Karamchandani, S. Diggavi, and P. Tabuada, "Secure state estimation: Optimal guarantees against sensor attacks in the presence of noise," in *Proceedings of 2015 IEEE International Symposium on Information Theory*, 2015, pp. 2929–2933.

[52] A. Mitra and S. Sundaram, "An approach for distributed state estimation of LTI systems," in *Proceedings on 54th Annual Allerton Conference on Communication, Control, and Computing*, 2016, pp. 1088–1093.

[53] Q. Mo and Y. Shen, "A remark on the restricted isometry property in orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3654–3656, 2012.

[54] Y. Mo, T. H.-J. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli, "Cyber-physical security of a smart grid infrastructure," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 195–209, 2012.

[55] Y. Mo and E. Garone, "Secure dynamic state estimation via local estimators," in *Proceedings of 55th IEEE Conference on Decision and Control*, 2016, pp. 5073–5078.

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

[57] ——, "False data injection attacks in control systems," in *Proceedings of First Workshop on Secure Control Systems, 13th International Conference on Hybrid Systems: Computation and Control*, 2010.

[58] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, 1995.

[59] H. Q. Ngo and D.-Z. Du, “A survey on combinatorial group testing algorithms with applications to dna library screening,” *Discrete Mathematical Problems with Medical Applications*, vol. 55, pp. 171–182, 2000.

[60] K. Ogata, *Discrete-Time Control Systems*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 1995.

[61] R. Olfati-Saber, “Distributed Kalman filtering for sensor networks,” in *Proceedings on 46th IEEE Conference on Decision and Control*, 2007, pp. 5492–5498.

[62] M. Pajic, J. Weimer, N. James, P. Tabuada, O. Sokolsky, I. Lee, and G. Pappas, “Robustness of attack-resilient state estimators,” in *Proceedings of IEEE/ACM 5th International Conference on Cyber-Physical Systems*, 2014, pp. 163–174.

[63] G. Park, H. Shim, C. Lee, Y. Eun, and K. H. Johansson, “When adversary encounters uncertain cyber-physical systems: Robust zero-dynamics attack with disclosure resources,” in *Proceedings of 55th IEEE Conference on Decision and Control*, 2016, pp. 5085–5090.

[64] J. Park, R. Ivanov, J. Weimer, M. Pajic, S. H. Son, and I. Lee, “Security of cyber-physical systems in the presence of transient sensor faults,” *ACM Transactions on Cyber-Physical Systems*, vol. 1, no. 3, pp. 15:1–15:23, 2017.

[65] S. Park and N. C. Martins, “Design of distributed LTI observers for state omniscience,” *IEEE Transactions on Automatic Control*, vol. 62, no. 2, pp. 561–576, 2017.

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

[67] Z. Ping, C. Lee, and H. Shim, "Robust estimation algorithm for both switching signal and state of switched linear systems," *International Journal of Control, Automation and Systems*, vol. 15, no. 1, pp. 95–103, 2017.

[68] R. R. Rajkumar, I. Lee, L. Sha, and J. Stankovic, "Cyber-physical systems: The next computing revolution," in *Proceedings of the 47th Design Automation Conference*, 2010, pp. 731–736.

[69] G. G. Rigatos, "Differential flatness theory and flatness-based control," *Nonlinear Control and Filtering Using Differential Flatness Approaches*, pp. 47–101, 2015.

[70] H. Sandberg, S. Amin, and K. H. Johansson, "Cyberphysical security in networked control systems," *IEEE Control Systems*, vol. 35, no. 1, pp. 20–23, 2015.

[71] H. Sandberg, A. Teixeira, and K. H. Johansson, "On security indices for state estimators in power networks," in *Proceedings of First Workshop on Secure Control Systems, 13th International Conference on Hybrid Systems: Computation and Control*, 2010.

[72] J. T. Schwartz, *Nonlinear Functional Analysis.* Gordon and Breach Science Publishers, 1969.

[73] H. Shim, "A passivity-based nonlinear observer and a semi-global separation principle," Ph.D. dissertation, Seoul National University, 2000.

[74] H. Shim and A. Tanwani, "Hybrid-type observer design based on a sufficient condition for observability in switched nonlinear systems," *International Journal of Robust and Nonlinear Control*, vol. 24, no. 6, pp. 1064–1089, 2014.

[75] Y. Shoukry, M. Chong, M. Wakaiki, P. Nuzzo, A. L. Sangiovanni-Vincentelli, S. A. Seshia, J. P. Hespanha, and P. Tabuada, "SMT-based observer design for cyber-physical systems under sensor at-

tacks," in *Proceedings of IEEE/ACM 7th International Conference on Cyber-Physical Systems*, 2016, pp. 1–10.

[76] Y. Shoukry, P. Nuzzo, A. Puggelli, A. L. Sangiovanni-Vincentelli, S. A. Seshiz, and P. Tabuada, "Secure state estimation for cyber physical systems under sensor attacks: A satisfiability modulo theory approach," *IEEE Transactions on Automatic Control*, vol. 62, no. 10, pp. 4917–4932, 2017.

[77] Y. Shoukry and P. Tabuada, "Event-triggered state observers for sparse sensor noise/attacks," *IEEE Transactions on Automatic Control*, vol. 61, no. 8, pp. 2079–2091, 2016.

[78] Y. Shoukry, P. Nuzzo, N. Bezzo, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Secure state reconstruction in differentially flat systems under sensor attacks using satisfiability modulo theory solving," in *Proceedings of 54th IEEE Conference on Decision and Control*, 2015, pp. 3804–3809.

[79] J. Slay and M. Miller, "Lessons learned from the Maroochy water breach," *Critical Infrastructure Protection*, pp. 73–82, 2007.

[80] E. D. Sontag, *Mathematical Control theory: Deterministic Finite Dimensional Systems*, 2nd ed. Springer Science & Business Media, 1998.

[81] K. C. Sou, H. Sandberg, and K. H. Johansson, "Computing critical k-tuples in power networks," *IEEE Transactions on Smart Grid*, vol. 27, no. 3, pp. 1511–1520, 2012.

[82] ——, "On the exact solution to a smart grid cyber-security analysis problem," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 856–865, 2013.

[83] S.-L. Sun, "Multi-sensor optimal information fusion Kalman filters with applications," *Aerospace Science and Technology*, vol. 8, no. 1, pp. 57–62, 2004.

[84] S.-L. Sun and Z.-L. Deng, “Multi-sensor optimal information fusion Kalman filter,” *Automatica*, vol. 40, no. 6, pp. 1017–1023, 2004.

[85] S. Sundaram and C. N. Hadjicostis, “Distributed function calculation via linear iterative strategies in the presence of malicious agents,” *IEEE Transactions on Automatic Control*, vol. 56, no. 7, pp. 1495–1508, 2011.

[86] A. Tanwani, H. Shim, and D. Liberzon, “Observability for switched linear systems: Characterization and observer design,” *IEEE Transactions on Automatic Control*, vol. 58, no. 4, pp. 891–904, 2013.

[87] N. Tarfulea, “Observability for initial value problems with sparse initial data,” *Applied and Computational Harmonic Analysis*, vol. 30, no. 3, pp. 423–427, 2011.

[88] A. Teel and L. Praly, “Global stabilizability and observability imply semi-global stabilizability by output feedback,” *Systems & Control Letters*, vol. 22, no. 5, pp. 313–325, 1994.

[89] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, “A secure control framework for resource-limited adversaries,” *Automatica*, vol. 15, pp. 135–148, 2015.

[90] ——, “Revealing stealthy attacks in control systems,” in *Proceedings of 50th Annual Allerton Conference on Communication, Control, and Computing*, 2012, pp. 1806–1813.

[91] C.-W. Ten, C.-C. Liu, and G. Manimaran, “Vulnerability assessment of cybersecurity for SCADA systems,” *IEEE Transactions on Power Systems*, vol. 23, no. 4, pp. 1836–1846, 2008.

[92] H. Trentelman, A. A. Stoorvogel, and M. Hautus, *Control Theory For Linear Systems.* Springer Science & Business Media, 2012.

[93] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[94] R. J. Turk, "Cyber incidents involving control systems," Technical Report, Idao National Laboratory, pp. 1836–1846, 2005.

[95] J. Wang and B. Shim, "On the recovery limit of sparse signals using orthogonal matching pursuit," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4973–4976, 2012.

[96] D. E. Whitehead, K. Owens, D. Gammel, and J. Smith, "Ukraine cyber-induced power outage: Analysis and practical mitigation strategies," in *Proceedings of 70th Annual Conference for Protective Relay Engineers*, 2017, pp. 1–8.

[97] A. Wright, "Hacking cars," *Communications of the ACM*, vol. 54, no. 11, pp. 18–19, 2011.

[98] T. C. Yang, "Networked control system: A brief survey," *IEE Proceedings-Control Theory and Applications*, vol. 153, no. 4, pp. 403–412, 2006.

[99] H. Zhang, E. Fata, and S. Sundaram, "A notion of robustness in complex networks," *IEEE Transactions on Control of Network Systems*, vol. 2, no. 3, pp. 310–320, 2015.

[100] H. Zhang and S. Sundaram, "Robustness of information diffusion algorithms to locally bounded adversaries," in *Proceedings of 2012 American Control Conference*, 2012, pp. 5855–5861.

[101] Q. Zhu and T. Basar, "Game-theoretic methods for robustness, security, and resilience of cyberphysical control systems: Games-in-games principle for optimal cross-layer resilient control systems," *IEEE Control Systems*, vol. 35, no. 1, pp. 46–65, 2015.

국문초록

외부 공격으로부터 자율 복원 가능한 제어 시스템: 센서 공격에 안전한 상태 추정 기법

Attack-Resilient Feedback Control Systems: Secure State Estimation under Sensor Attacks

최근 컴퓨터와 네트워크 통신을 통해서 실제 시스템을 관리 및 운용하고 제어하게 되면서, 외부로부터의 악의적인 공격이 보다 쉽게 제어 시스템에 접근할 수 있게 되었다. 이런 외부와의 연결에서 기인한 제어 시스템의 공격은 국가 기반 시설을 파괴하거나 심지어 사람의 생명을 앗아갈 수도 있으며, 이는 기존의 제어 공학에서는 고려하지 않았던 새로운 형태의 문제이다. 이에 본 논문에서는 외부로부터 공격이 가해지더라도, 시스템이 자율적으로 복원하여 정상 동작할 수 있도록 하는 제어 기법을 고안하고, 이를 제어 이론적 관점에서 고찰한다. 특히, 제어 시스템의 일부 센서에 외부 공격으로부터 악의적인 입력이 가해지는 상황에서 시스템의 상태 변수를 잘 추정하는 관측기를 설계하는 방법을 제안한다. 먼저 이론적 분석을 진행하며, 여기서의 핵심은 중복 관측 가능성 (redundant observability) 이란 개념이다. 시불변 선형 시스템에서 임의의 q 개 센서를 제외하더라도 여전히 상태 변수를 추정할 수 있으면, 그 시스템이 q 중복 관측 가능하다고 한다. 이러한 중복 관측 가능성의 개념은 센서 공격에 노출된 제어 시스템의 안전과 관계된 많은 문제를 한번에 설명할 수 있다. 즉, q 중복 관측 가능한 시스템은, 어떤 q 개의 센서가 공격받더라도, 이를 검출할 수 있으며, $\lfloor q/2 \rfloor$ 개의 센서가 공격받는다면, 어떤 센서가 공격받았는지 정확하게 확인할 수 있다. 시스템에 들키지 않고 (undetectable) 주입할 수 있는 최소한의 센서 공격의 개수를 동적 안전 지표 (dynamic security index) 라 정의할 수 있으며, 이 경우 동적 안전 지표는 $q + 1$ 이 된다. 또한 중복 관측 가능성은 점근 중복 관측 가능성 (asymptotic redundant observability) 의 개념으로 확장될 수 있는데, 이는 시간이 지남에 따라 0 에 수렴하는 신호는 0 과 같이 취급하게 되어, 실질적으로 0 에 수렴하지 않는 공격 신호에만 관심을 갖고 다루게 된다. 즉, 시스템을 파괴할 수 있는 영향력을 지닌 공격 신호만 검출하고, 확인하기 때문에 조금 더 실용적인 개념이라 할 수 있다. 그 다음으로, 센서 공격에도 자율

복원 가능한 관측기 설계 기법을 제안한다. 제안된 관측기는, 칼만 점근 관측 가능성 분해(Kalman detectability decomposition)를 활용하여 각각 개별 센서에서 점근 관측 가능 부분 상태 변수(detectable sub-state)를 추정하는 부분 관측기와, 모든 부분 관측기로부터 정보를 취합하여 중복성(redundancy)에 기반한 오류 정정 문제를 풀어서 전체 상태 변수를 복원하는 해독기로 구성된다. 희박성(sparsity)을 지닌 센서 공격 신호 외에, 유한한(bounded) 외란과 잡음이 들어오는 경우, 부분 관측기는 루엔버거 관측기(Luenberger observer)로 구성되며, 이는 외란에 대한 강인성(robustness)을 보장하고, 정규 분포를 갖는 외란과 잡음이 들어오는 경우에는, 부분 관측기를 칼만 필터(Kalman filter)로 설계하여, 오차의 분산을 최소화할 수 있다. 본 논문에서 제시된 상태 추정 알고리즘은 최적화 문제를 유한 집합에서 계산하고 최적화 문제에 검출 기법을 함께 적용하여 기존의 결과에 비해 시간 복잡도(time complexity)를 낮춰서 계산 시간을 줄일 수 있었으며, 칼만 점근 관측 가능성 분해를 활용한 부분 관측기 설계를 통해 공간 복잡도(space complexity)를 센서 개수에 대한 선형으로 획기적으로 줄여서 메모리 공간을 절약하였다. 또한, 선형 시스템에 대해 개발된 중복 관측 가능성의 개념과 자율 복원 상태 추정 기법을 제어 입력 값에 의존하지 않는 균등 관측 가능 (uniformly observable) 비선형 시스템으로 확장한다. 선형 시스템의 칼만 관측 가능성 분해에 대응하는 균등 관측 가능성 분해(uniform observability decomposition)를 활용하여, 비선형 시스템의 개별 센서에 대해 관측 가능 부분 상태 변수(observable sub-state)를 추정하는 부분 관측기를 고이득 관측기(high gain observer)로 구성하고, 해독기는 이것들로부터 모은 정보를 활용하여 중복성에 기반한 비선형 오류 정정 문제를 풀어서 상태 변수를 추정한다. 이를 위하여, 공격 검출 기법을 활용해서 센서들의 부분 집합 중에서 공격이 검출되지 않는 부분 집합을 찾고, 이렇게 찾은 부분 집합의 센서들이 제공하는 상태 변수 추정값으로부터 최종적으로 상태 변수를 계산한다.

주요어: 사이버 물리 시스템, 공격에 대한 자율 복원성, 해석적 중복성, 안전 지표, 공격 검출, 안전한 상태 추정
학　번: 2012–30225