City Research Online

City St George's, University of London

Citation: Yang, G., Gallo, A. J., Barboni, A., Ferrari, R. M. G., Serrani, A. & Parisini, T. (2025). On the Output Redundancy of LTI Systems: A Geometric Approach With Application to Privacy. IEEE Transactions on Automatic Control, 70(11), pp. 7509-7522. doi: 10.1109/tac.2025.3579219

This is the published version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/38007/

Link to published version: https://doi.org/10.1109/tac.2025.3579219

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

On the Output Redundancy of LTI Systems: A Geometric Approach With Application to Privacy

Guitao Yang, Alexander J. Gallo, Angelo Barboni, Riccardo M. G. Ferrari, *Senior Member, IEEE*, Andrea Serrani, *Member, IEEE*, and Thomas Parisini, *Fellow, IEEE*

***Abstract*—This article examines the properties of output-redundant systems, that is, systems possessing a larger number of outputs than inputs, through the lense of the geometric approach of Wonham et al. We begin by formulating a simple output allocation synthesis problem, which involves "concealing" input information from a malicious eavesdropper having access to the system output, while still allowing for a legitimate user to reconstruct it. It is shown that the solvability of this problem requires the availability of a redundant set of outputs. This very problem is instrumental to unveiling the fundamental geometric properties of output-redundant systems, which form the basis for our subsequent constructions and results. As a direct application, we demonstrate how output allocation can be employed to effectively protect the input information from certain output eavesdroppers with guaranteed results.**

***Index Terms*—Geometric approach, LTI systems, output redundancy, system invertibility.**

I. Introduction

In this article, we analyze the properties and structure of output-redundant systems within a geometric framework. We demonstrate how these structural properties may be leveraged by analysing the problem of safeguarding input information against eavesdropping attacks via the injection of output disturbances.

Received 13 September 2024; revised 29 March 2025; accepted 24 May 2025. Date of publication 12 June 2025; date of current version 27 October 2025. This work was supported by European Union's Horizon 2020 Research and Innovation Program under Grant 739551 (KIOS CoE). Recommended by Associate Editor S. Galeani. *(Corresponding author: Guitao Yang.)*

Guitao Yang is with the Wolfson School of Mechanical Electrical and Manufacturing Engineering, Loughborough University, LE11 3TU Loughborough, U.K. (e-mail: guitao.yang@gmail.com).

Alexander J. Gallo is with the Department of Electronics, Information, and Bioengineering, Politecnico di Milano, 20133 Milano, Italy (e-mail: alexanderjulian.gallo@polimi.it).

Angelo Barboni is with Zurich Insurance Group, 8002 Zurich, Switzerland (e-mail: angelobarboni.eng@gmail.com).

Riccardo M. G. Ferrari is with the Delft Centre for Systems and Control, Technical University of Delft, 2628 CD Delft, The Netherlands (e-mail: r.ferrari@tudelft.nl).

Andrea Serrani is with the Department of Electrical, Electronic and Information Engineering "G. Marconi," Università di Bologna, 40126 Bologna, Italy (e-mail: andrea.serrani@unibo.it).

Thomas Parisini is with the Department of Electrical and Electronic Engineering, Imperial College London, SW7 2AZ London, U.K., also with the Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark, and also with the Department of Engineering and Architecture, University of Trieste, 34127 Trieste, Italy (e-mail: t.parisini@imperial.ac.uk).

Digital Object Identifier 10.1109/TAC.2025.3579219

The progressive increase in complexity of cyber–physical systems, stemming from the integration of widespread communication and computation technologies, have made them prime targets for all sorts of cyberattacks and adversarial interventions [1]. Among the possible actions taken by malicious agents are eavesdropping attacks [2], i.e., attacks which breach the confidentiality of the information being transmitted within the control system, enabling them to access private user information. This problem is particularly relevant for cases, such as smart grids or heating, ventilation, and air conditioning systems, where measurement outputs may contain sensitive information about user behavior, such as energy consumption, or room occupancy [3], [4], [5]. Among the schemes proposed by the control-theoretic community to achieve privacy is that of achieving obfuscation of the relevant information by affine transformations of the output, whereby a disturbance is injected onto the transmitted signal. Without the pretense of providing a complete overview, examples of this strategy are reported in [6], [7], and [8].

In this article, the very problem of obfuscating exogenous inputs is used to introduce and investigate fundamental geometric concepts associated to the property of output redundancy in linear systems. To set the stage for our discussion, we consider LTI systems of the form

$$\begin{aligned} x^+ &= Ax + Bu \\ y &= Cx \end{aligned} \tag{1}$$

where x, u, and y are the state, input, and output, respectively, seen as elements of appropriate finite-dimensional vector spaces in the field of real numbers. The operator $+$ in x^+ denotes the evolution of the state variable in a general sense, which encompasses continuous-time evolution (dx/dt), discrete-time evolution $x(t+1)$, or even variables evolving at nonequally spaced time instances ($x(t_{k+1})$). In (1), A, B and C represent maps among vector spaces. Assuming access to the measured output y, as well as the system parameters A, B, and C, under a condition of left invertibility of the system, the attacker may recover the exact system input, which constitutes the private and sensitive information being transmitted within the control systems. However, if the output signal is modified as

$$y_d = y + Dd$$

where d represents a disturbance intentionally injected through the map D, the eavesdropper might be misled by the resulting

© 2025 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

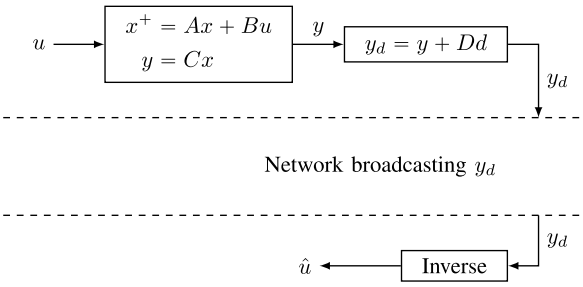


Fig. 1. Schematic diagram of our considered setup: the upper part presents the exogenous input, the system, and the obfuscation strategy; the middle section represents the network over which y_d is broadcast, with the eavesdropper; the lower part of the diagram illustrates the actions taken by a legitimate agent to reconstruct an estimate of u, denoted by $\hat{u}$.

corruption of the available information, and may lose its capability to recover sensitive information. In the considered scenario, we want to evaluate this simple strategy provided that a legitimate user is able to reconstruct the required information from the corrupted data, whereas the eavesdropper can not. The considered setup is illustrated in the schematic diagram in Fig. 1, where an "inverse system" (for lack of better terminology) is introduced to reconstruct the exogenous input. For this to be possible, we show that it is necessary for the system to be *output-redundant*, i.e., that the number of scalar measurements, $p \in \mathbb{N}$, be strictly larger than the dimension $m \in \mathbb{N}$ of the input signal $u(\cdot)$ to be concealed from the eavesdropper. We then investigate the geometric properties of the ensuing output-redundant system (1), evaluating under what conditions it is possible to design the map D, such that two objectives can be achieved simultaneously: first, the input and state can be reconstructed by some legitimate user (a *defender*) with some knowledge of D (e.g., the matrix form of D or the column space of D); second, that a malicious agent, without knowledge of D, obtains a biased estimate of the input. The problem is posed as an *output allocation problem* via the selection of the *output allocation map* D, conducted on the basis of the geometric interpretation of output redundancy. Intuitively, the difference between the number of inputs and outputs of a system provides more "flexibility" to inject a disturbance capable of obfuscating the input from a malicious agent, without affecting reconstruction by a legitimate defender. We investigate the properties associated with output redundancy of a system in order to make this notion of "flexibility" precise. We note that the need for "flexibility" finds correspondence in the method proposed in [8], whereby it is the lifting of the output signal to be made private into a larger space (and therefore the inclusion of additional "flexibility") that enables the design of a scheme that ensures privacy, without any performance loss.

To make the presentation reasonably self-contained, apart from output redundancy, additional properties will be assumed to hold for system (1), namely minimality and left-invertibility.

A. Brief Literature Review

Existing studies have examined output redundancy from various perspectives. A somewhat customary assumption dictates that the sensors measuring the system state are duplicated, with subsets of these sensors potentially susceptible to faults [9], [10], [11]. In the cited references, output redundancy is leveraged to mitigate the impact of faulty sensors, whereas the overall system performance continues to rely on the nonfaulty sensors. The studies [12], [13] consider setups where certain measurements are linearly dependent on others, which presents a broader scenario compared to duplicated sensors. Specifically, Ren et al. [12] demonstrated that the inclusion of redundant sensors improves estimation performance by reducing the variance of the estimated state error. On the other end, Cristofaro and Zaccarian [13] proposed a methodology to dynamically select a weighted combination of available measurements to reject constant biases. In contrast to the notion of output redundancy explored in the aforementioned works, Corona and Cristofaro [14] defined output redundancy as a dual concept of input redundancy, carrying over a taxonomy that distinguishes between the notions of strong output redundancy and weak output redundancy. In that article, a system is termed to be strongly output redundant (SOR) when the output matrix lacks full rank, whereas weakly output redundant (WOR) occurs when the number of outputs exceeds the number of inputs. While this perspective enriches the concept of output redundancy beyond the scope of previous works, which primarily fall under the SOR paradigm, the focus of Corona and Cristofaro [14] is centered on the output regulation problem (hence, in a certain sense on right-invertibility), and does not consider aspects related to left-invertibility, which we believe to be of fundamental importance in the considered setup. To address this gap, this article provides a study of output redundancy and ensuing properties from a coordinate-free geometric perspective; algorithms generated from this approach apply to a variety of formulations, encompassing both continuous and discrete-time linear systems.

As mentioned, output redundancy is naturally associated with the null space of a rank-deficient output map, where it is often assumed that the rank of the output map equals the dimension of the input space. However, as mentioned in [14], this characterization does not encompass all potential scenarios, as it disregards epic (i.e., surjective) output operators with rank exceeding the dimension of the input space. This latter scenario signifies an intrinsic redundancy within the system, where only a proper subset of outputs is sufficient to uniquely determine the input trajectory. In this article, we employ geometric tools to provide a comprehensive analysis of output-redundant LTI systems and delve into their properties through suitable decomposition of both the state space and the output space. Subsequently, we rigorously formulate the problem that we introduced at the beginning of the article and solve it by leveraging the properties of the output-redundant system that we derived. It is worth noticing that the findings presented here may have broader applications beyond the scope of the considered privacy preservation problem, for instance to fault detection and isolation [16].

Unsurprisingly, the problem dealt with in this article can be regarded as a dual of the input allocation problem formulated in [15], where input redundancy (defined as properties of the system transfer matrix) is exploited to optimize additional metrics defined on the system input, without impacting output tracking

TABLE I
DUALITY BETWEEN DYNAMIC INPUT AND OUTPUT ALLOCATION

	Input allocation	Output allocation
System	$\begin{cases} x^+ = Ax + Bu \\ y = Cx \end{cases}$	$\begin{cases} \hat{x}^+ = A\hat{x} - L(y_d - \hat{y}) \\ \hat{y} = C\hat{x} \end{cases}$
Control variable	input u	output error $y_d - \hat{y}$
Control objective	$\|y_{\text{ref}} - y\| \to$ small	$\|u - \hat{u}\| \to$ small
Allocation mechanism	$u = u_c + B^{\perp}d$ [15]	$y_d = y + Dd$ (this article)

In the input allocation section, u_c denotes the input provided by a given tracking controller, and y_{ref} a reference output to be tracked.

performance. State–space input allocation strategies are also developed in [17] by leveraging a geometric analysis, which provides an alternative point of view to input redundancy. In our dual settings, however, we shift focus to an observation problem, treating the estimated state as regulated variables influenced by a disturbance that has been intentionally injected (i.e., *allocated*) into the system output, hence affecting the state of a Luenberger-type observer. By dynamically allocating redundant measurements, we aim to impair the eavesdropper's ability to reconstruct the system's input. We refer to this strategy as the allocating mechanism or simply the allocator. The underlying principles of this duality are visually summarized in Table I. It should be noted that a more general definition of input redundancy and related taxonomy (alternative to those of [15] and [17]) that disposes of the assumption of right-invertibility has been recently proposed in [18]. As system invertibility is central to our discussion, a dualization of the more general definitions proposed therein has not been pursued in this work.

B. Contributions and Outline of the Article

The main contribution of this article is as follows.

1) We demonstrate and substantiate the necessity of output redundancy in a system to solve the problem of concealing the system's input while preserving the ability of a legitimate user to reconstruct it, as stated in the introduction.
2) We provide a more comprehensive approach to output redundancy for LTI systems, where we formally distinguish between a strong and weak characterization in a geometric sense. This is a step beyond the frameworks in [9], [12], and [13], which are limited to a notion of SOR.
3) We explore the properties of both strongly output-redundant (SOR) and weakly output-redundant (WOR) systems under a geometric approach framework, which applies to various types of systems. Moreover, the properties are coordinate-free due to the merit of the geometric approach.
4) We illustrate how the derived properties are leveraged to solve the problem of preventing the output eavesdropper from reconstructing the true system's input, while maintaining reconstructability of this latter by a legitimate user.

The rest of this article is organized as follows. In Section II, we state the notation and provide essential concepts and notions of the geometric approach used in our work. We formalize the problem in Section III, and prove the necessity of output redundancy for its solvability. In Sections IV and V, we present the properties and solutions to the synthesis problem by demonstrating the utilization of the derived properties of SOR and WOR systems, respectively. Section VI showcases simulation results validating the solutions and analysis in Section V. Finally, Section VII concludes this article.

II. NOTATION AND PRELIMINARIES

A. Notation

Throughout this article, we use $\mathtt{A}$ to denote the matrix representation of a linear map $A : \mathscr{X} \to \mathscr{Y}$ among finite-dimensional vector spaces $\mathscr{X}$, $\mathscr{Y}$ over the field $\mathbb{R}$, i.e., $\text{Mat}(A) = \mathtt{A}$. For a finite-dimensional vector space $\mathscr{X}$, $\mathscr{X}'$ denotes its dual space, i.e., the space of linear functionals $x' : \mathscr{X} \to \mathbb{R}$. For a map $A : \mathscr{X} \to \mathscr{Y}$, $A' : \mathscr{Y}' \to \mathscr{X}'$ denotes the dual map satisfying $\text{Mat}(A') = \mathtt{A}^\top$. The external (internal) direct sum of vector spaces $\mathscr{V}$, $\mathscr{W}$ (subspaces $\mathscr{V}, \mathscr{W} \subseteq \mathscr{X}$) is denoted by $\mathscr{V} \oplus \mathscr{W}$, with elements $v \oplus w$, where $v \in \mathscr{V}$ and $w \in \mathscr{W}$ [19, Section 0.3]. The symbols I and $\mathtt{I}$ stand respectively for the identity map and the identity matrix of compatible dimension, whereas $\mathtt{0}$ is an all-zeros matrix with compatible dimension. For a symmetric square matrix $\mathtt{M} \in \mathbb{R}^{n\times n}$, $\mathtt{M} \succ (\prec)\ 0$ denotes positive (negative) definiteness. $\mathtt{M}^\dagger$ represents the pseudoinverse of $\mathtt{M}$. An LTI system is usually represented as $\{C, A, B\}$ with output map C, state map A, and input map B, respectively. For an endomorphism $A : \mathscr{X} \to \mathscr{X}$, $\sigma(A)$ and $\sigma(\mathtt{A})$ stand for the spectrum of the map A and the eigenvalues of the matrix $\mathtt{A}$, respectively. To streamline the presentation, maps that are instrumental for the completion of commutative diagrams will be defined in the caption of the figures presenting the diagrams themselves. These maps may also be referred to in the text.

B. Preliminaries on Geometric Approach

This preliminary section builds on the concepts introduced in [19], [20], and [21]. While our notation aligns more closely with that of [19] and [20], the pioneering work of [21] highlights the duality between controlled and conditioned invariants, a perspective that has inspired subsequent research, including our own.

Let $C : \mathscr{X} \to \mathscr{Y}$ be a map; then, C is an *epimorphism* (or C is *epic*, onto) if $\text{Im}\, C = \mathscr{Y}$, and C is a *monomorphism* (or C is *monic*, one-to-one) if $\text{Ker}\, C = 0$. We say two finite-dimensional linear spaces $\mathscr{X}_1$ and $\mathscr{X}_2$ are *isomorphic*, $\mathscr{X}_1 \simeq \mathscr{X}_2$, if they are dimensionally equal. If $\mathscr{S} \subseteq \mathscr{Y}$, $C^{-1}\mathscr{S}$ denotes the *preimage* of $\mathscr{S}$ under C. Let $C : \mathscr{X} \to \mathscr{Y}$, and let $\mathscr{V} \subset \mathscr{X}$ be a subspace with *insertion map* $V : \mathscr{V} \to \mathscr{X}$, i.e., $\mathscr{V} = \text{Im}\, V$ and V is monic. The *domain restriction* of C to $\mathscr{V}$ is denoted by $C|\mathscr{V} = CV$. Moreover, suppose $\text{Im}\, C \subseteq \mathscr{W} \subseteq \mathscr{Y}$. The *codomain restriction* of C from $\mathscr{W}$, denoted as $\mathscr{W}|C$ satisfies $W(\mathscr{W}|C) = C$, where $W : \mathscr{W} \to \mathscr{Y}$ is the insertion map of $\mathscr{W}$.

A subspace $\mathscr{V} \subseteq \mathscr{X}$ is said to be *invariant* with respect to a map $A : \mathscr{X} \to \mathscr{X}$ if $A\mathscr{V} \subseteq \mathscr{V}$. For an invariant subspace $\mathscr{V}$, we denote by $A|\mathscr{V} : \mathscr{V} \to \mathscr{V}$ the *restriction of A to $\mathscr{V}$*, i.e., the

unique map satisfying $AV = V(A|\mathscr{V})$, where $V : \mathscr{V} \to \mathscr{X}$ is the insertion map of $\mathscr{V}$. Furthermore, we denote by $A|\mathscr{X}/\mathscr{V}$ or simply by $\bar{A} : \mathscr{X}/\mathscr{V} \to \mathscr{X}/\mathscr{V}$ the map induced on $\mathscr{X}/\mathscr{V}$ by A, that is, the unique map satisfying $\bar{A}P = PA$, where $P : \mathscr{X} \to \mathscr{X}/\mathscr{V}$ is the canonical projection on $\mathscr{X}/\mathscr{V}$, the factor space $\mathscr{X}$ modulo $\mathscr{V}$. For a map $A : \mathscr{X} \to \mathscr{X}$ and subspaces $\mathscr{B} \subseteq \mathscr{X}, \mathscr{K} \subseteq \mathscr{X}$, we define the smallest A-invariant subspace that contains $\mathscr{B}$ as $\langle A \mid \mathscr{B} \rangle$ and the largest A-invariant subspace that is contained in $\mathscr{K}$ as $\langle \mathscr{K} \mid A \rangle$ [20, Section 2.1].

Let $A : \mathscr{X} \to \mathscr{X}$, $B : \mathscr{U} \to \mathscr{X}$, and $C : \mathscr{X} \to \mathscr{Y}$. We say a subspace $\mathscr{V} \subseteq \mathscr{X}$ is (A, B)-*invariant* if there exists a map $F : \mathscr{X} \to \mathscr{U}$ such that $(A + BF)\mathscr{V} \subseteq \mathscr{V}$. If this is the case, we say that the state-feedback map F is a friend of $\mathscr{V}$, and denote the class of all friends of $\mathscr{V}$ by $\mathbf{F}(\mathscr{V})$. We say a subspace $\mathscr{W} \subseteq \mathscr{X}$ is (C, A)-*invariant* if there exists a map $L : \mathscr{Y} \to \mathscr{X}$ such that $(A + LC)\mathscr{W} \subseteq \mathscr{W}$. If this is the case, we also say that the output-injection map L is a friend of $\mathscr{W}$, and denote the class of all friends of $\mathscr{W}$ by $\mathbf{L}(\mathscr{W})$. The notation $L \in \mathbf{L}(\mathscr{W})$ reads "L is a friend of $\mathscr{W}$." Moreover, let $\mathscr{B}, \mathscr{K} \subseteq \mathscr{X}$. We write $\underline{\mathscr{V}}(\mathscr{K})$ to denote the class of (A, B)-invariant subspaces *contained in* $\mathscr{K}$, and by $\underline{\mathscr{W}}(\mathscr{B})$ the class of (C, A)-invariant subspaces *containing* $\mathscr{B}$. The supremal and infimal elements of $\underline{\mathscr{V}}(\mathscr{K})$ and $\underline{\mathscr{W}}(\mathscr{B})$ are denoted by $\mathscr{V}^*(A, B; \mathscr{K})$ and $\mathscr{W}^*(C, A; \mathscr{B})$, respectively, or simply by $\mathscr{V}^*$ and $\mathscr{W}^*$ [19, Section 4.2], [20, Section 2.2].

We say a subspace $\mathscr{R} \subseteq \mathscr{X}$ is a *controllability subspace* if $\mathscr{R} = \langle A + BF | \operatorname{Im} BG \rangle$ for some state feedback map $F : \mathscr{X} \to \mathscr{U}$ and input selection map $G : \mathscr{U} \to \mathscr{U}$. We say a subspace $\mathscr{S} \subseteq \mathscr{X}$ is an *unobservability subspace* if $\mathscr{S} = \langle \operatorname{Ker} HC \mid A + LC \rangle$ for some output-injection map $L : \mathscr{Y} \to \mathscr{X}$ and measurement mixing map $H : \mathscr{Y} \to \mathscr{Y}$. For $\mathscr{B}, \mathscr{K} \subseteq \mathscr{X}$, we write $\underline{\mathscr{R}}(\mathscr{K})$ to denote the class of controllability subspaces *contained in* $\mathscr{K}$, and $\underline{\mathscr{S}}(\mathscr{B})$ the class of unobservability subspaces *containing* $\mathscr{B}$. The supremal and infimal elements of $\underline{\mathscr{R}}(\mathscr{K})$ and $\underline{\mathscr{S}}(\mathscr{B})$ are denoted by $\mathscr{R}^*(A, B; \mathscr{K})$ and $\mathscr{S}^*(C, A; \mathscr{B})$, respectively, or simply by $\mathscr{R}^*$ and $\mathscr{S}^*$, respectively [19, Section 5.1], [20, Section 2.3], [21].

III. Problem Formalization and the Necessity of Output Redundancy

We recall that the motivating problem described in the Introduction consists in injecting a disturbance signal on the output measurements of an LTI system to provide obfuscation of the input against eavesdroppers, whilst ensuring that a legitimate user can reconstruct the input from the perturbed output. This latter aspect of the problem is concerned with left inversion of a system affected by an unmeasurable output disturbance. Specifically, consider the system

$$x^+ = Ax + Bu, \quad x(0) = x_0 \tag{2a}$$

$$y_d = Cx + Dd \tag{2b}$$

where $x, x_0 \in \mathscr{X} \simeq \mathbb{R}^n$, $u \in \mathscr{U} \simeq \mathbb{R}^m$, $d \in \mathscr{D} \simeq \mathbb{R}^q$, $y_d \in \mathscr{Y} \simeq \mathbb{R}^p$ are the system states, input, output disturbance, and perturbed output, respectively; $A : \mathscr{X} \to \mathscr{X}, B : \mathscr{U} \to \mathscr{X}$, $D : \mathscr{D} \to \mathscr{Y}$, and $C : \mathscr{X} \to \mathscr{Y}$ are the corresponding maps. For reasons that will become clear in the sequel, we define the aggregate input $u_a := u \oplus d \in \mathscr{U} \oplus \mathscr{D} =: \mathscr{U}_a$, where $\mathscr{U} \oplus \mathscr{D}$ denotes the external direct sum of $\mathscr{U}$ and $\mathscr{D}$.

We make the following standing assumptions on system (2).

Assumption 1: The pair (A, B) is controllable, and the pair (C, A) is observable.

Assumption 2: The triplet $\{C, A, B\}$ is left-invertible (see [22], [21, Section 4.3] for definitions and properties.)

Among these properties, we recall here (see [22, Theorem 5] and [20, Proposition 9]) that system (2) is left-invertible if and only if

$$\begin{aligned} &B \text{ is monic, and} \\ &\mathscr{V}^*(A, B; \operatorname{Ker} C) \cap \operatorname{Im} B = 0. \end{aligned} \tag{3}$$

Remark 1: Assumption 1 is made for the sake of simplicity, and can be removed at the cost of undue (and unnecessary) complications. On the other hand, Assumption 2, ensures reconstructibility (generally, by means of a noncausal system) of the input signal when $d = 0$, which is clearly a necessary condition in the considered setup [23].

In the statements of the forthcoming results, we shall refrain from mentioning explicitly that Assumptions 1 and 2 hold for (2), as these assumptions will henceforth be considered valid with no exception.

Definition 1: Consider system (2) with initial condition $x_0 \in \mathscr{X}$, input signal $u(\cdot) \in \mathbb{U}$, disturbance signal $d(\cdot) \in \mathbb{D}$, and output signal $y_d(\cdot) \in \mathbb{Y}$, where $\mathbb{U}$, $\mathbb{D}$, and $\mathbb{Y}$ are suitably defined spaces of admissible functions defined over a given interval $[0, T]$, $T > 0$. Define

$$\theta_{x_0}(u, d) : \mathbb{U} \times \mathbb{D} \to \mathbb{Y} \tag{4}$$

as the operator[1] that maps signals $u(\cdot), d(\cdot)$ to outputs $y_d(\cdot)$ over the interval $[0, T]$, parameterized by the initial condition x_0. For example, for continuous-time systems

$$\theta_{x_0}(u, d)(t) = Ce^{At}x_0 + \int_0^t Ce^{A(t-\tau)}Bu(\tau)d\tau + Dd(t)$$

for $t \in [0, T]$.

In order to formalize the problem dealt with in this article, we introduce a definition of left-invertibility for system (2) that explicitly considers the presence of the disturbance.

Definition 2: System (2) is said to be left-invertible with respect to u under the output disturbance d if for any $x_0 \in \mathscr{X}$

$$\theta_{x_0}(u_1, d_1) = \theta_{x_0}(u_2, d_2) \tag{5}$$

implies

$$u_1(t) = u_2(t) \quad \forall t \in [0, T]. \tag{6}$$

Note that the above-mentioned definition is not equivalent to left-invertibility with respect to both u and d, as it is not necessary for the condition $\theta_{x_0}(u, d_1) = \theta_{x_0}(u, d_2)$ to imply that $d_1(\cdot) = d_2(\cdot)$. However, left-invertibility with respect to the aggregate input $u_a = u \oplus d$ trivially implies left-invertibility with respect to u under the disturbance d. Equivalent definitions of left-invertibility of systems defined by quadruplets $\{C, A, B, D\}$ are provided in [24, Chapter 8].

Remark 2: It should be clear that, due to linearity, system (2) is left-invertible with respect to u under the output disturbance

[1] For the sake of simplicity, we have used (and will henceforth use) the more economical expression (4) in place of $\theta_{x_0}(u(\cdot), d(\cdot))$.

d if and only if for $x_0 = 0$

$$\theta_0(u, d) = 0 \tag{7}$$

implies $u(t) = 0$ for all $t \in [0, T]$. This equivalent formulation of Definition 2 will often be used in the sequel.

The problem dealt with in this work is then stated as follows.

Problem 1: For system (2), find $\mathscr{D} \neq 0$ and the map $D : \mathscr{D} \to \mathscr{Y}$ such that system (2) is left-invertible with respect to u under the output disturbance d, according to Definition 2.

Remark 3: We stress that solvability of Problem 1 is necessary to solve the motivating problem defined in the introduction, as a legitimate user must always be able to reconstruct $u(\cdot)$ when the output signal is intentionally corrupted. Conversely, the question of whether privacy can be guaranteed against an eavesdropping attacker for a given choice of the space $\mathscr{D}$ and the map D, must be evaluated separately, as it depends on the information available to the attacker.

Up to this point, no specific properties have been postulated for the map D or the space $\mathscr{D}$. However, the defense strategy should avoid the occurrence of disturbance signals producing no effect on the output, that is, the possibility that D has a nontrivial null space. As a result, it makes sense to assume that D be monic, and, consequently, $\dim \mathscr{D} \leq p$. This choice has an important implication on the property stated in Definition 2:

Proposition 1: Assume that D is monic. Then, system (2) is left-invertible with respect to u under the output disturbance d if and only if it is left-invertible with respect to the aggregate input $u_a = u \oplus d$.

Proof: We shall prove necessity, as sufficiency is obvious. Let $u(\cdot) \in \mathbb{U}$, $d(\cdot) \in \mathbb{D}$ be such that (7) hold. By virtue of the assumption of left-invertibility with respect to u under d, it follows that $u(t) = 0$ for all $t \in [0, T]$, and, consequently

$$Dd(t) = 0 \qquad \forall\, t \in [0, T]. \tag{8}$$

As D is monic, (8) implies $d(\cdot) = 0$, from which left-invertibility of the system with respect to u_a follows. □

Once we have restricted D to be monic, henceforth we shall identify $\mathscr{D}$ with a subspace of $\mathscr{Y}$, $\mathscr{D} \subseteq \mathscr{Y}$, and let $D : \mathscr{D} \to \mathscr{Y}$ denote the insertion map of $\mathscr{D}$ in $\mathscr{Y}$ (which is monic by definition). As a result, one disposes of the need to define $\mathscr{D}$ and D separately. In the light of these observations and the result of Proposition 1, Problem 1 is replaced by the following.

Problem 2: For system (2), find a nontrivial subspace $\mathscr{D} \subseteq \mathscr{Y}$ (with associated insertion map $D : \mathscr{D} \to \mathscr{Y}$) such that system (2) is left-invertible with respect to the aggregate input $u_a = u \oplus d$.

In order to tackle Problem 2, we introduce the following definitions.

Definition 3 (Output redundant system): System (2) is said to be output redundant if $p = \dim \mathscr{Y} > \dim \mathscr{U} = m$.

Note that, for an output-redundant system, Assumption 2 implies that $\dim(\operatorname{Im} C) \geq m$, as it is necessary for the number of indipendent outputs to be at least the same as the number of inputs for the map from $\mathscr{U}$ to $\mathscr{Y}$ to be injective. It is precisely the dimension of $\operatorname{Im} C$ that prompts the following taxonomy.

Definition 4 (SOR system): System (2) is said to be SOR if it is output-redundant and $\dim(\operatorname{Im} C) = m$.

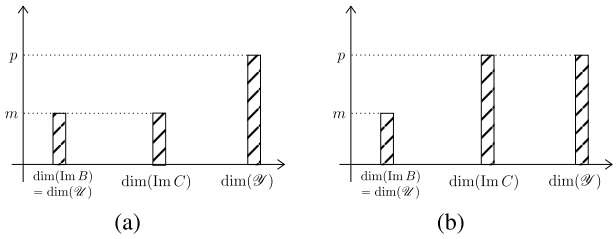


Fig. 2. Relation between the dimension of input space, image of output map, and output space for: (a) SOR systems; (b) WOR systems.

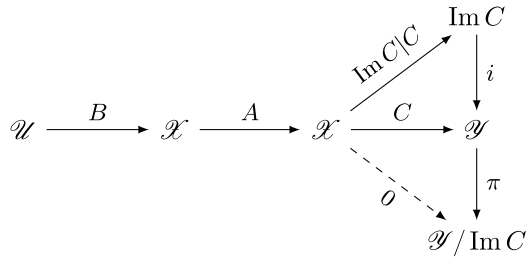


Fig. 3. Commutative diagram for SOR systems, where $\operatorname{Im} C|C : \mathscr{X} \to \operatorname{Im} C$ is the codomain restriction of C to $\operatorname{Im} C$, $i : \operatorname{Im} C \to \mathscr{Y}$ is the insertion map, and $\pi : \mathscr{Y} \to \mathscr{Y} / \operatorname{Im} C$ is the canonical projection mod $\operatorname{Im} C$.

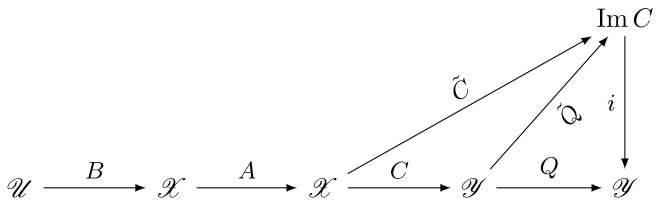


Fig. 4. Example of a commutative diagram for system (2) as discussed in the proof of Theorem 2.

Definition 5 (WOR system): System (2) is termed WOR if it is output-redundant and $\dim(\operatorname{Im} C) = p$.

Remark 4: Our definition of WOR is conceptually equivalent to that in [14]. However, our approach adopts a geometric perspective, which leads us to express redundancy in terms of $\dim \operatorname{Im}$ rather than rank. Furthermore, the distinction between strong output redundancy and strong output redundancy arises from the dualization of the concepts of strong and weak input redundancy, as discussed in works, such as [15] and [17].

A somewhat obvious necessary condition for the solvability of Problem 2 follows.

Theorem 1: To solve Problem 2, it is necessary that system (2) be *output-redundant*. Furthermore, $q = \dim \mathscr{D} \leq p - m$.

Proof: As $\dim \mathscr{U}_a = m + q$, a necessary condition for left-invertibility with respect to u_a is $m + q \leq p$, from which the statements of the theorem follow from the fact that $q \geq 1$ by assumption. □

The relation between the dimension of $\mathscr{U}$, $\mathscr{Y}$, and the image of the output map is illustrated in Fig. 2 for both the SOR and WOR cases. Note that it may be possible for system (2) to satisfy $m < \dim(\operatorname{Im} C) < p$.

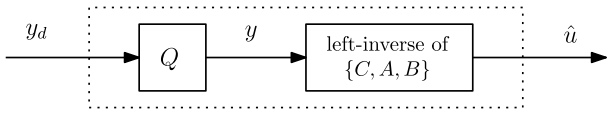


Fig. 5. Strategy to reconstruct u for SOR systems.

IV. Strong Output Redundancy

In this section, we first introduce properties of interest for SOR systems, and then we illustrate the solution to the problem formalized in Section III for this specific class.

A. Properties of SOR Systems

Let system (1) be SOR. According to Definition 4, the map C satisfies

$$p = \dim(\mathscr{Y}) > \dim(\operatorname{Im} C) = m \tag{9}$$

which implies that C is not epic. Consequently, it is possible to factor out $\operatorname{Im} C$ from $\mathscr{Y}$ by canonical projection. The commutative diagram in Fig. 3 depicts the factorization of $\mathscr{Y}$ into $\operatorname{Im} C$ and $\mathscr{Y}/\operatorname{Im} C$. As $\operatorname{Ker} \pi = \operatorname{Im} C$, the combined map $\pi C : \mathscr{X} \to \mathscr{Y}/\operatorname{Im} C$ is the zero map. This in turn implies that no input $u(\cdot) \in \mathbb{U}$ can be retrieved from $\mathscr{Y}/\operatorname{Im} C$; this simple observation will prove instrumental to the design of $\mathscr{D}$ solving Problem 2.

B. Design of $\mathscr{D}$ for SOR Systems

Owing to the aforementioned properties of SOR systems, the following result provides a necessary and sufficient condition on $\mathscr{D}$ for the solution to Problem 2, namely left-invertibility w.r.t. the aggregate input u_a.

Theorem 2: Assume that system (2) is SOR. A nontrivial subspace $\mathscr{D} \subseteq \mathscr{Y}$ solves Problem 2 if and only if

$$\mathscr{D} \cap \operatorname{Im} C = 0. \tag{10}$$

As we restrict D as an insertion map in Problem 2, we treat $\mathscr{D}$ as $\operatorname{Im} D$.

Proof: Owing to (9), $\mathscr{Y}$ can be decomposed as

$$\mathscr{Y} = \operatorname{Im} C \oplus \mathscr{W} \tag{11}$$

where $\mathscr{W} \simeq \mathscr{Y}/\operatorname{Im} C$ is an arbitrary complementary subspace to $\operatorname{Im} C$. Let $Q : \mathscr{Y} \to \mathscr{Y}$ be the projection on $\operatorname{Im} C$ along $\mathscr{W}$, and $R := I - Q$ be the projection on $\mathscr{W}$ along $\operatorname{Im} C$. Also, let $\tilde{Q} := \operatorname{Im} C | Q$ and $\tilde{R} = \mathscr{W} | R$ denote the corresponding natural projections. An example of the ensuing commutative diagrams is shown in Fig. 4. Accordingly, $y_d \in \mathscr{Y}$ is expressed uniquely as

$$y_d = \tilde{Q} y_d \oplus \tilde{R} y_d = (\tilde{C}x + \tilde{D}_1 d) \oplus \tilde{D}_2 d \tag{12}$$

where $\tilde{C} := \tilde{Q}C$, $\tilde{D}_1 := \tilde{Q}D$, and $\tilde{D}_2 := \tilde{R}D$ (note that $\tilde{R}C = 0$). We prove sufficiency by construction. Select a nontrivial subspace $\mathscr{D} \subseteq \mathscr{Y}$ satisfying (10) and let the subspace $\mathscr{W}$ in (11) satisfy $\mathscr{D} \subseteq \mathscr{W}$. Since (10) implies $\tilde{D}_1 = 0$, from (12) one obtains

$$y_d = \tilde{C}x \oplus \tilde{D}_2 d. \tag{13}$$

Given this decomposition, it is sufficient (it is indeed also necessary) for $\{\tilde{C}, A, B\}$ to be left invertible for the system to be left invertible w.r.t. the input u under disturbance d. Indeed, u can be reconstructed from $\tilde{Q}y_d = \tilde{C}x$. If additionally $\tilde{D}_2$ is monic, following Proposition 1, under these conditions the system is left invertible w.r.t. the aggregated input u_a, solving Problem 2. Thus, let us show that the system defined by the triplet $\{\tilde{C}, A, B\}$ is left-invertible, and that the map $\tilde{D}_2$ is monic. Recalling conditions (3), to show that $\{\tilde{C}, A, B\}$ is left-invertible it suffices to prove that $\operatorname{Im} B \cap \mathscr{V}^*(A, B; \operatorname{Ker} \tilde{C}) = 0$. To this end, note that

$$\operatorname{Ker} \tilde{C} = C^{-1}\left(\operatorname{Ker} \tilde{Q} \cap \operatorname{Im} C\right) = C^{-1}\left(\mathscr{W} \cap \operatorname{Im} C\right) = \operatorname{Ker} C$$

hence $\mathscr{V}^*(A, B; \operatorname{Ker} \tilde{C}) = \mathscr{V}^*(A, B; \operatorname{Ker} C)$, from which left-invertibility of $\{\tilde{C}, A, B\}$ follows. The assertion that $\tilde{D}_2$ is monic is proved in a similar fashion by noticing that

$$\operatorname{Ker} \tilde{D}_2 = D^{-1}(\operatorname{Ker} \tilde{R} \cap \mathscr{D}) = D^{-1}(\operatorname{Im} C \cap \mathscr{D}) = \operatorname{Ker} D = 0.$$

Necessity is proven by contradiction. Assume that (10) does not hold and system (2) is left-invertible with respect to u_a. Let $(u(\cdot), d(\cdot)) \in \mathbb{U} \times \mathbb{D}$ be such that $\theta_0(u, d) = 0$. This implies $\tilde{D}_2 d(t) = 0$ for all $t \in [0, T]$, hence $d(t) \in D^{-1} \operatorname{Ker} \tilde{R} \Rightarrow d(t) \in D^{-1} \operatorname{Im} C \Rightarrow Dd(t) \in \mathscr{D} \cap \operatorname{Im} C$ for all $t \in [0, T]$. Consequently, one can restrict the co-domain of D to $\mathscr{D} \cap \operatorname{Im} C$ whenever $\theta_0(u, d) = 0$. Accordingly, let $\tilde{D}_3 := \tilde{D}_1 | (\mathscr{D} \cap \operatorname{Im} C)$ denote the domain restriction of $\tilde{D}_1 = \tilde{Q}D$ to $\mathscr{D} \cap \operatorname{Im} C$. Clearly, $\tilde{D}_3 \neq 0$. By inspection, system (2) is left-invertible with respect to $u_a \in \mathscr{U} \oplus \mathscr{D}$ only if the system

$$\begin{aligned} x^+ &= Ax + Bu, \quad x(0) = x_0 \\ \tilde{y}_d &= \tilde{C}x + \tilde{D}_3 \tilde{d} \end{aligned} \tag{14}$$

with output $\tilde{y}_d \in \operatorname{Im} C$ and disturbance $\tilde{D}_3 \tilde{d} \in \mathscr{D} \cap \operatorname{Im} C$, is left-invertible with respect to the aggregate input $\tilde{u}_a = (u, \tilde{d}) \in \mathscr{U} \oplus (\mathscr{D} \cap \operatorname{Im} C)$. This is not possible, given that

$$\begin{aligned} \dim\left(\mathscr{U} \oplus (\mathscr{D} \cap \operatorname{Im} C)\right) &= m + \dim(\mathscr{D} \cap \operatorname{Im} C) \\ &> m = \dim(\operatorname{Im} C). \end{aligned}$$

□

Remark 5 (Maximum coverage): The *maximum coverage of $\mathscr{Y}$ from $\mathscr{D}$*, i.e., the subspace with largest dimension solving Problem 2, is achieved when $\mathscr{D} \simeq \mathscr{Y}/\operatorname{Im} C$.

It is worth noticing that the necessary and sufficient condition given in Theorem 2 relies solely on the definition of $\mathscr{Y}$ and $\operatorname{Im} C$, and is not related to the mappings from the input space $\mathscr{U}$ to the state space $\mathscr{X}$. As a result, for SOR systems, the dynamics of (2) plays no role in the solution of Problem 2.

Strategy for reconstructing the input for SOR systems: The proof of Theorem 2 suggests a simple strategy to reconstruct $u(\cdot)$ from the perturbed output $y_d(\cdot)$. Once the legitimate user has knowledge of the complementary subspace $\mathscr{W}$, and $\mathscr{D} \subseteq \mathscr{W}$, the effect of the disturbance can be removed from the output altogether by projection on $\operatorname{Im} C$ along $\mathscr{W}$. This allows reconstruction of $u(\cdot)$ via left-inversion of system (1). The ensuing scheme is depicted in Fig. 5.

C. Attacker's Sensitivity to Disturbance

Recalling our motivating example, let us provide some insight into the sensitivity of an eavesdropper attack to the injected disturbance. To this end, one must focus on a specific strategy to be implemented by the attacker. This is accomplished by defining an *attacker model*, that is, the combination of system knowledge, resources, and objectives the attacker has at its disposal [25].

Assumption 3 (Attack model): Assume that the attacker:

1) knows the maps (C, A, B) of system (1) *(system knowledge)*;
2) has access to the corrupted output signal y_d *(attacker's resources)*;
3) attempts to reconstruct the input u from observations of y_d using, as a guess for $\mathscr{D}$, a subspace $\mathscr{D}_a$ providing maximum coverage (i.e., satisfying $\operatorname{Im} C \oplus \mathscr{D}_a = \mathscr{Y}$), with associated projection Q_a on $\operatorname{Im} C$ along $\mathscr{D}_a$ *(attacker's objective)*.

The following result follows immediately from Theorem 2.

Corollary 1: Assume that system (2) is SOR. Suppose there is an eavesdropper attacker as in Assumption 3. The disturbance injection d biases the attacker's reconstructed input only if

$$\mathscr{D} \not\subseteq \mathscr{D}_a. \tag{15}$$

Proof: The proof follows easily by contradiction. Suppose (15) does not hold, i.e., that $\mathscr{D} \subseteq \mathscr{D}_a$. Since $\operatorname{Ker} Q_a = \mathscr{D}_a$, it follows that $Q_a D d = 0$, hence $y_a := Q_a y_d = y$. As system (2) is left-invertible when $d = 0$, the attacker can reconstruct the input u without any bias. □

Remark 6: One possible solution to achieving the attacker's objective, and possibly the most intuitive to use in a discrete time setting, is to exploit a least squares (LS) algorithm to compute an estimate of u, see, e.g., [26], [27], [28]. Indeed, for left-invertible systems (without the presence of an output disturbance), it has been shown that the input can be reconstructed exactly with some delay, depending on the relative degree of the system [28]. For that solution strategy, the definition of the LS algorithm implies

$$\mathscr{D}_a := \operatorname{Ker} C' \tag{16}$$

where $C' : \mathscr{Y}' \to \mathscr{X}'$ is the dual map of C and $\mathscr{Y}'$, $\mathscr{X}'$ are dual spaces to $\mathscr{Y}$, $\mathscr{X}$, respectively.

In Section IV-D, we show an example of an SOR system, showcasing how satisfying (15) implies $Q_a D \neq 0$, although $Q_a C = C$.

D. Illustrating Example for SOR Systems

We present a simple illustrating example concerning SOR systems. Consider system (2) with $n = 4$, $m = 2$, $p = 3$ and map C represented in given bases for $\mathscr{X}$ and $\mathscr{Y}$ as

$$\mathtt{C} := \operatorname{Mat}(C) = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \end{bmatrix}.$$

Since $p = 3 > \operatorname{rank} C = m = 2$, the system is SOR. We demonstrate the existence of an insertion map D, expressed in matrix form, that satisfies condition (10). Furthermore, we provide the matrix representation of the projection Q, ensuring that $QC = C$ and $QD = 0$, thereby eliminating the effect of the injected signal $d(t)$. In addition, from the attacker's perspective, we define D_a and show that $\operatorname{Im} D \not\subseteq \operatorname{Im} D_a$. Finally, we derive the corresponding projection Q_a and establish that $Q_a D \neq 0$, which implies the presence of bias in the attacker's estimate.

The matrix representations of the maps $\operatorname{Im} C|C$ and i in the diagram of Fig. 3 read as

$$\operatorname{Mat}(\operatorname{Im} C|C) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \ \operatorname{Mat}(i) = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 2 & 1 \end{bmatrix}.$$

As $p - \operatorname{rank} C = 1$, the only possible nonzero subspace $\mathscr{D}$ satisfying (10) is a complementary subspace to $\operatorname{Im} C$ in $\mathscr{Y}$. As an arbitrary choice (among infinitely many) for $\mathscr{D}$, select the subspace with insertion map D with coordinate representation

$$\mathtt{D} := \operatorname{Mat}(D) = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^\top. \tag{17}$$

Select a basis adapted to the decomposition $\mathscr{Y} = \operatorname{Im} C \oplus \mathscr{Z}$, where $\mathscr{Z} \simeq \mathscr{Y} / \operatorname{Im} C$ serves as a generic complementary subspace to $\operatorname{Im} C$. This yields the following representation for the canonical projection $\pi : \mathscr{Y} \to \mathscr{Y} / \operatorname{Im} C$ in Fig. 3:

$$\operatorname{Mat}(\pi) = \begin{bmatrix} 1 & 1 & -1 \end{bmatrix}.$$

It can be easily verified that the projection $Q : \mathscr{Y} \to \mathscr{Y}$ on $\operatorname{Im} C$ along $\mathscr{D}$ has representation

$$\mathtt{Q} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

yielding $\mathtt{QC} = \mathtt{C}$ and $\mathtt{QD} = 0$, hence $\mathtt{Q}\pi y = \mathtt{C}x$; this implies that Problem 2 can be solved. Assuming that the attacker employs an LS algorithm, one obtains

$$\mathscr{D}_a = \operatorname{Ker} C' \simeq \operatorname{Ker} \mathtt{C}^\top = \operatorname{span}\left\{ \begin{bmatrix} -1 & -1 & 1 \end{bmatrix}^\top \right\}$$

and, for the projection $Q_a : \mathscr{Y} \to \mathscr{Y}$ on $\operatorname{Im} C$ along $\mathscr{D}_a$ (apart from round-off errors)

$$\mathtt{Q}_a = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix}.$$

Consequently, $\mathtt{Q}_a \mathtt{D} \neq 0$, leading to a bias of the attacker's estimate.

V. Weak Output Redundancy

In this section, we first introduce the properties of WOR systems and develop an appropriate defense strategy against the malicious eavesdropper for this class of systems.

A. Properties of WOR Systems

Assume that system (1) is WOR. Let $\mathscr{S}^*$ denote the infimal unobservability subspace containing $\operatorname{Im} B$, i.e., $\mathscr{S}^* := \mathscr{S}^*(C, A; \operatorname{Im} B)$. The following (fundamental) result has been proved in [16].

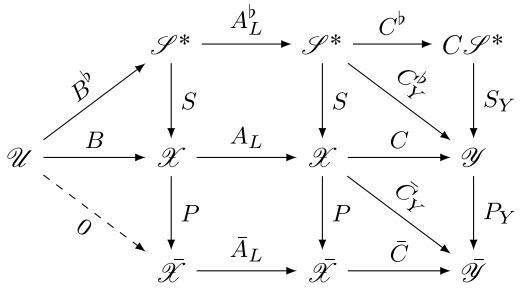


Fig. 6. Commutative diagram for WOR systems. In the diagram, $\bar{\mathscr{X}} := \mathscr{X}/\mathscr{S}^*$ and $\bar{\mathscr{Y}} := \mathscr{Y}/(C\mathscr{S}^*)$. $S : \mathscr{S}^* \to \mathscr{X}$ is the insertion map of $\mathscr{S}^*$ in $\mathscr{X}$, $S_Y : C\mathscr{S}^* \to \mathscr{Y}$ is the insertion map of $C\mathscr{S}^*$ in $\mathscr{Y}$, $P : \mathscr{X} \to \bar{\mathscr{X}}$ is the canonical projection modulo $\mathscr{S}^*$, and $P_Y : \mathscr{Y} \to \bar{\mathscr{Y}}$ is the canonical projection modulo $C\mathscr{S}^*$. The codomain restriction $B^\flat : \mathscr{U} \to \mathscr{S}^*$ is well-defined since $\operatorname{Im} B \subseteq \mathscr{S}^*$, and the map $\bar{C} : \bar{\mathscr{X}} \to \bar{\mathscr{Y}}$ is well-defined due to the fact that $\operatorname{Ker} \bar{C}_Y = \operatorname{Ker} P_Y C = \mathscr{S}^* + \operatorname{Ker} C \supseteq \mathscr{S}^*$. The map $A_L^\flat : \mathscr{S}^* \to \mathscr{S}^*$ is the restriction of A_L to the A_L-invariant subspace $\mathscr{S}^*$. The map $\bar{A}_L : \bar{\mathscr{X}} \to \bar{\mathscr{X}}$ is the map induced on $\bar{\mathscr{X}}$ by A_L.

Lemma 1 ([16]): For a WOR system, $\mathscr{S}^*$ is nontrivial, and satisfies $m \le \dim(\mathscr{S}^*) \le n - p + m$.

Since the class of friends of $\mathscr{S}^*$ is nonempty for WOR systems, select $L \in \mathbf{L}(\mathscr{S}^*)$ and consider a classical Luenberger observer for (1)

$$\hat{x}^+ = (A + LC)\hat{x} - Ly \tag{18}$$

where $L : \mathscr{Y} \to \mathscr{X}$ is the output-injection map. Define the state and output observation errors as

$$\tilde{x} := x - \hat{x}, \qquad \tilde{y} := y - C\hat{x}$$

with associated *error system*

$$\begin{aligned} \tilde{x}^+ &= A_L\tilde{x} + Bu \\ \tilde{y} &= C\tilde{x} \end{aligned} \tag{19}$$

where $A_L := A + LC$. *We regard (19) as a new system altogether, obtained from (1) via an output-injection transformation using a friend of* $\mathscr{S}^*$. The maps defining system (19) are then represented according to the commutative diagram in Fig. 6, from which one can define two distinct systems as follows.

- System $\Sigma^\flat := \{C^\flat, A_L^\flat, B^\flat\}$ has input space $\mathscr{U}$, state space $\mathscr{S}^*$, and output space $C\mathscr{S}^*$.
- System $\bar{\Sigma} := \{\bar{C}, \bar{A}_L, 0\}$ is autonomous with respect to the input space $\mathscr{U}$, and has state space $\bar{\mathscr{X}} = \mathscr{X}/\mathscr{S}^*$ and output space $\bar{\mathscr{Y}} = \mathscr{Y}/(C\mathscr{S}^*)$.

Relevant properties of the systems $\Sigma^\flat$ and $\bar{\Sigma}$ are presented in the following lemmas.

Lemma 2 ([20] Proposition 2): Let system (1) be WOR. Assume that (C, A) is observable. Then, the pairs $(C^\flat, A_L^\flat)$ and $(\bar{C}, \bar{A}_L)$ are observable for any $L \in \mathbf{L}(\mathscr{S}^*)$.

Lemma 3 ([16]): Let system (1) be WOR. Assume that $\{C, A, B\}$ is left-invertible. Then, system $\Sigma^\flat$ is square and invertible. Consequently, $\dim \bar{\mathscr{Y}} = p - m$.

Remark 7: Lemma 2 establishes that if the original system is observable, then both subsystems, $\Sigma^\flat$ and $\bar{\Sigma}$, remain observable. This result plays a crucial role in the proof of Proposition 3. In addition, Lemma 3 demonstrates the invertibility of the system $\Sigma^\flat$, which is fundamental to input reconstruction in WOR systems, as discussed in Section V-B.

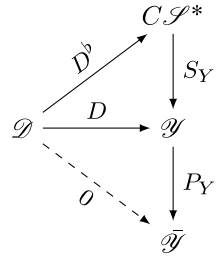


Fig. 7. Decomposition of the map D in Lemma 4. The codomain restriction $D^\flat = C\mathscr{S}^*|D$, which is well-defined since $\operatorname{Im} D \subseteq C\mathscr{S}^*$ by assumption, is also the insertion map of $\mathscr{D}$ in $C\mathscr{S}^*$.

Remark 8: When (C, A) is observable, the spectra of $A_L^\flat$ and $\bar{A}_L$ can both be freely assigned by a friend of $\mathscr{S}^*$ ([20] Proposition 13 and Theorem 16). Consequently, we shall always find $L \in \mathbf{L}(\mathscr{S}^*)$ such that $A_L^\flat$ and $\bar{A}_L$ are both stable. An algorithm for calculating such L is provided in Appendix A.

Under the assumption that the original system (1) is WOR, we have derived, via output-injection, an "error system" (19) where the unobservability subspace $\mathscr{S}^*$ has been rendered invariant. The error system is represented as the interconnection of the *square and invertible* system $\Sigma^\flat$ with the *autonomous* system $\bar{\Sigma}$. The ensuing structure strategically confines all influences from the input space $\mathscr{U}$ through the map B into $\mathscr{S}^*$, while the system $\bar{\Sigma}$ is decoupled from the input. This distinctive feature of WOR systems will be explored in the subsequent section, when the output disturbance d is introduced.

B. Design of $\mathscr{D}$ for WOR Systems

We now consider the presence of the output disturbance d on a WOR system. Let system (2) be WOR, and apply an output-injection transformation with $L \in \mathbf{L}(\mathscr{S}^*)$. This yields, in place of (19), the new error system

$$\begin{aligned} \tilde{x}^+ &= A_L\tilde{x} + Bu + D_L d \\ \tilde{y}_d &= C\tilde{x} + Dd \end{aligned} \tag{20}$$

where $D_L := LD$ denotes the input disturbance map. It is noted that the output disturbance influences both the state and the output of the error system (20).

The next step is to define a preliminary condition on $\mathscr{D}$ with useful implications.

Lemma 4: Let system (2) be WOR. Assume that $\mathscr{D} \subseteq \mathscr{Y}$ satisfies

$$\mathscr{D} \subseteq C\mathscr{S}^*. \tag{21}$$

Then, the map $P_Y D : \mathscr{D} \to \bar{\mathscr{Y}}$ is the zero map.

Proof: As P_Y is the canonical projection modulo $C\mathscr{S}^*$ on $\mathscr{Y}$, $\operatorname{Ker} P_Y = C\mathscr{S}^*$. Thus, $P_Y D = 0$ follows directly from (21).□

Lemma 4 is represented by the commutative diagram in Fig. 7, showing that d does not affect $\bar{\mathscr{Y}}$. Assuming that $\mathscr{D} \neq 0$ satisfies (21), applying the same decomposition as in Fig. 6 to the perturbed system (20) yields the following equations for the

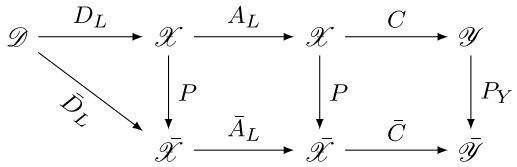


Fig. 8. Decomposition of system $\{C, A_L, D_L\}$ resulting from $L \in \mathbf{L}(\mathscr{S}^*)$, where $\bar{D}_L := PLD$.

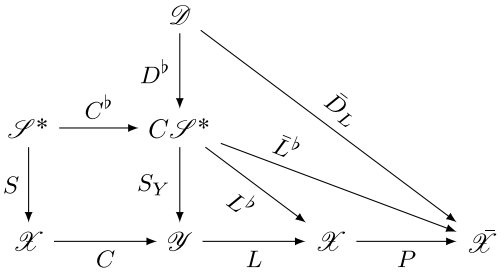


Fig. 9. Diagram showing the action of $D_L : \mathscr{D} \to \mathscr{X}$ on $\bar{\mathscr{X}}$ via its codomain restriction on $C\mathscr{S}^*$, whenever (21) holds. Note that this diagram can be thought of as "zooming in" on the map $\bar{D}_L$ in Fig. 8.

dynamics of the system defined on the factor space $\bar{\mathscr{X}}$:

$$\begin{aligned} \bar{x}^+ &= \bar{A}_L \bar{x} + \bar{D}_L d \\ \bar{y} &= \bar{C} \bar{x} \end{aligned} \tag{22}$$

where $\bar{x} \in \bar{\mathscr{X}}$, $\bar{y} \in \bar{\mathscr{Y}}$, $d \in \mathscr{D}$, and the maps $\bar{A}_L$, $\bar{D}_L$, and $\bar{C}$ are defined in Fig. 8. The main result of this section will be devoted to show that it is always possible to design $\mathscr{D} \neq 0$ satisfying (21) such that system (22) is left-invertible (with respect to the input d). In turn, this result will lead to the solution of Problem 2 for system (20) by way of reconstructability of $d(\cdot)$ from $\tilde{y}(\cdot)$ via canonical projection modulo $C\mathscr{S}^*$. To this end, we recall conditions (3) applied to system (22) for it to be left-invertible as follows.

1) $\bar{D}_L : \mathscr{D} \to \bar{\mathscr{X}}$ is monic.
2) $\mathrm{Im}(\bar{D}_L) \cap \bar{\mathscr{V}}^* = 0$, where

$$\bar{\mathscr{V}}^* := \mathscr{V}^*(\bar{A}_L, \bar{D}_L; \mathrm{Ker}\, \bar{C}). \tag{23}$$

Using the diagram in Fig. 9, it follows that:

$$\bar{D}_L := PLD = \bar{L}^\flat D^\flat$$

where $\bar{L}^\flat : C\mathscr{S}^* \to \bar{\mathscr{X}}$ is defined as $\bar{L}^\flat := PLS_Y$, the domain restriction of $\bar{L} := PL$ on $C\mathscr{S}^*$. Before proceeding further with the design of $\mathscr{D}$, one must prove that $\bar{L}^\flat$ is a nonzero map, which is clearly a necessary condition for condition 1 above to hold.

Proposition 2: Let system (2) be WOR, and let $L : \mathscr{Y} \to \mathscr{X}$ be a friend of $\mathscr{S}^*$. Then, $\bar{L}^\flat \neq 0$.

Proof: Let $L \in \mathbf{L}(\mathscr{S}^*)$. As $\mathscr{S}^*$ is A_L-invariant, it follows that $\mathrm{Im}\, A_L S \subseteq \mathscr{S}^*$, hence

$$0 = PA_L S = P(A + LC)S = PAS + PLCS. \tag{24}$$

Using the diagram in Fig. 9, it is seen that

$$PLCS = \bar{L}^\flat C^\flat. \tag{25}$$

Consequently, $\bar{L}^\flat = 0$ implies $PAS = 0$, which is equivalent to

$$A\mathscr{S}^* \subseteq \mathrm{Ker}\, P = \mathscr{S}^*$$

showing that $\bar{L}^\flat = 0$ implies that $\mathscr{S}^*$ is A-invariant as well. Since $\mathscr{S}^*$ is a proper subset of $\mathscr{X}$ by virtue of Lemma 1, and $\mathscr{S}^* \supseteq \mathrm{Im}\, B$ by definition, it follows that $\bar{L}^\flat = 0$ contradicts controllability of the pair (A, B), thus violating Assumption 1.□

We are now in the position to state and prove the main result of this section as follows.

Proposition 3: Assume that system (2) is WOR. For any $L \in \mathbf{L}(\mathscr{S}^*)$, there exists a nontrivial subspace $\mathscr{D} \subseteq C\mathscr{S}^*$ such that system (22) is left-invertible.

Proof: Fix, arbitrarily, $L \in \mathbf{L}(\mathscr{S}^*)$. We first show that it is always possible to select $\mathscr{D} \subseteq C\mathscr{S}^*$ such that its insertion map $D^\flat : \mathscr{D} \to C\mathscr{S}^*$ yields a monic map $\bar{D}_L : \mathscr{D} \to \bar{\mathscr{X}}$, i.e., that condition 1 is satisfied. To this end, select $\mathscr{D}$ as a complementary subspace of $\mathrm{Ker}\, \bar{L}^\flat$ in $C\mathscr{S}^*$, that is, let $\mathscr{D} \subseteq \mathscr{Y}$ satisfy

$$C\mathscr{S}^* = \mathrm{Ker}\, \bar{L}^\flat \oplus \mathscr{D}. \tag{26}$$

This is possible by virtue of the fact that $\bar{L}^\flat \neq 0$ implies $\dim(\mathrm{Im}\, \bar{L}^\flat) > 0$, hence, by the first isomorphism theorem, $\dim(C\mathscr{S}^* / \mathrm{Ker}\, \bar{L}^\flat) > 0$. Since any complementary subspace of $\mathrm{Ker}\, \bar{L}^\flat$ in $C\mathscr{S}^*$ is isomorphic to $C\mathscr{S}^* / \mathrm{Ker}\, \bar{L}^\flat$, it follows that $\dim \mathscr{D} > 0$ in (26). Since

$$\mathrm{Ker}\, \bar{D}_L = \mathrm{Ker}(\bar{L}^\flat D^\flat) = {D^\flat}^{-1} \left(\mathrm{Ker}\, \bar{L}^\flat \cap \mathrm{Im}\, D^\flat \right)$$

and $\mathrm{Ker}\, \bar{L}^\flat \cap \mathscr{D} = 0$ by (26), one obtains

$$\mathrm{Ker}\, \bar{D}_L = \mathrm{Ker}\, D^\flat = 0$$

since $D^\flat$ is monic by definition. Consequently, $\bar{D}_L$ is monic.

Next, we show that $\mathscr{D}$ can be chosen such that, in addition, condition 2 holds. Clearly, if $\mathscr{D}$ selected at the previous step is such that $\mathrm{Im}(\bar{D}_L) \cap \bar{\mathscr{V}}^* = 0$, the proof is complete. Conversely, assume that this is not the case, that is, $\mathrm{Im}(\bar{D}_L) \cap \bar{\mathscr{V}}^* \neq 0$ for the previous choice of $\mathscr{D}$ satisfying (26). We claim that, necessarily

$$\mathrm{Im}\, \bar{D}_L \not\subseteq \bar{\mathscr{V}}^*. \tag{27}$$

By contradiction, assume that (27) does not hold. Then, from the definition of $\bar{\mathscr{V}}^*$, it follows that:

$$\bar{A}_L \bar{\mathscr{V}}^* \subseteq \bar{\mathscr{V}}^* + \mathrm{Im}\, \bar{D}_L = \bar{\mathscr{V}}^*$$

which implies $\bar{\mathscr{V}}^*$ is also $\bar{A}_L$-invariant. By definition, $\bar{\mathscr{V}}^* \subseteq \mathrm{Ker}\, \bar{C}$, which implies that $\bar{\mathscr{V}}^* \neq 0$ is contained in the unobservable subspace of the pair $(\bar{C}, \bar{A}_L)$. This contradicts the fact that $(\bar{C}, \bar{A}_L)$ is observable, as asserted in Lemma 2. Therefore, we conclude that (27) must hold. From (27), it is possible to define a nonzero subspace $\bar{\mathscr{N}} \subset \bar{\mathscr{X}}$ such that

$$\mathrm{Im}\, \bar{D}_L = \left(\bar{\mathscr{V}}^* \cap \mathrm{Im}\, \bar{D}_L \right) \oplus \bar{\mathscr{N}}. \tag{28}$$

Define a new subspace $\mathscr{D}_1$ as

$$\mathscr{D}_1 := \left(\bar{D}_L \right)^{-1} \bar{\mathscr{N}}$$

and note that $\mathscr{D}_1 \subseteq \mathscr{D} \subseteq C\mathscr{S}^*$. Let $D_1^\flat : \mathscr{D}_1 \to C\mathscr{S}^*$ denote the domain restriction of $D^\flat : \mathscr{D} \to C\mathscr{S}^*$ to $C\mathscr{S}^*$, and $\bar{D}_L^1 : \mathscr{D}_1 \to \bar{\mathscr{X}}$ the domain restriction of $\bar{D}_L$ to $\mathscr{D}_1$ (see Fig. 10). Notice that $\bar{D}_L^1$ is still monic since $\bar{D}_L$ is monic and $\mathscr{D}_1 \subseteq \mathscr{D}$. Moreover, since

$$\bar{D}_L^1 \mathscr{D}_1 = \bar{D}_L^1 \left(\bar{D}_L^{-1} \bar{\mathscr{N}} \right) = \bar{\mathscr{N}}$$

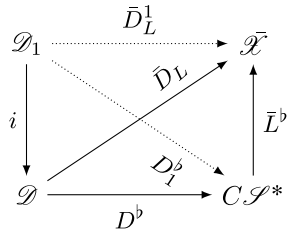


Fig. 10. Commutative diagrams illustrating the newly defined maps (in dotted arrows) for $\mathscr{D}_1 := (\bar{D}_L)^{-1}\bar{\mathscr{N}}$. The nondescript notation i is used for the insertion map $i : \mathscr{D}_1 \to \mathscr{D}$.

one obtains

$$\operatorname{Im} \bar{D}_L^1 = \bar{\mathscr{N}}. \tag{29}$$

Since $\bar{\mathscr{N}} \cap \bar{\mathscr{V}}^* = 0$ by virtue of (28), from (29) it follows that:

$$\bar{\mathscr{V}}^* \cap \operatorname{Im} \bar{D}_L^1 = 0. \tag{30}$$

Note that $\bar{\mathscr{V}}^*$ defined in (23) is computed using $\bar{D}_L$, not $\bar{D}_L^1$; consequently, (30) is not yet sufficient to conclude that system (22) is left-invertible when $\bar{D}_L^1$ replaces $\bar{D}_L$ (alternatively, $\mathscr{D}_1$ replaces $\mathscr{D}$). To show that this is indeed the case, define

$$\bar{\mathscr{V}}_1^* := \mathscr{V}^*(\bar{A}_L, \bar{D}_L^1; \operatorname{Ker} \bar{C})$$

and note that, by the properties of (A, B)-invariant subspaces

$$\bar{A}_L \bar{\mathscr{V}}_1^* \subseteq \bar{\mathscr{V}}_1^* + \operatorname{Im} \bar{D}_L^1. \tag{31}$$

Since, by definition of $\bar{D}_L^1$

$$\operatorname{Im} \bar{D}_L^1 \subseteq \operatorname{Im} \bar{D}_L$$

it follows from (31) that:

$$\bar{A}_L \bar{\mathscr{V}}_1^* \subseteq \bar{\mathscr{V}}_1^* + \operatorname{Im} \bar{D}_L$$

which implies that $\bar{\mathscr{V}}_1^*$ is also a $(\bar{A}_L, \bar{D}_L)$-invariant subspace contained in $\operatorname{Ker} \bar{C}$. Since $\mathscr{V}^*$ is the supremal element of this class, one obtains

$$\bar{\mathscr{V}}_1^* \subseteq \bar{\mathscr{V}}^*$$

which, according to (30), yields

$$\bar{\mathscr{V}}_1^* \cap \operatorname{Im} \bar{D}_L^1 = 0.$$

As a result, left-invertibility of system (22) follows from selecting $\mathscr{D}_1$ as the disturbance subspace. □

The following result, providing the solution of Problem 2 for WOR systems, is essentially a corollary of Proposition 3.

Theorem 3: Assume that system (2) is WOR. Then, there exists a nontrivial subspace $\mathscr{D} \subseteq \mathscr{Y}$ solving Problem 2.

Proof: We begin by recalling that left-invertibility is invariant under output-injection transformations, as $\mathscr{V}^*$ is invariant under output injection [29]. Consequently, system (2) is left-invertible with respect to the aggregate input u_a if and only if so is system (20). Choose, arbitrarily, $L \in \mathbf{L}(\mathscr{S}^*)$, and select $\mathscr{D} \subseteq C\mathscr{S}^*$ as in Proposition 3. Let $(u(\cdot), d(\cdot)) \in \mathbb{U} \times \mathbb{D}$ be such that $\theta_0(u, d) = 0$ for system (20). In particular, this implies that $\bar{y}(t) = P_Y \tilde{y}_d(t) = 0$ for all $t \in [0, T]$. Since system (22) is left-invertible by construction, it follows that $d(\cdot) = 0$. Letting $d = 0$ in (20), one obtains $u(\cdot) = 0$ by way of left-invertibility of the triplet $\{C, A_L, B\}$. Consequently, system (20) is left-invertible with respect to $u_a = (u, d)$, and so is system (2). □

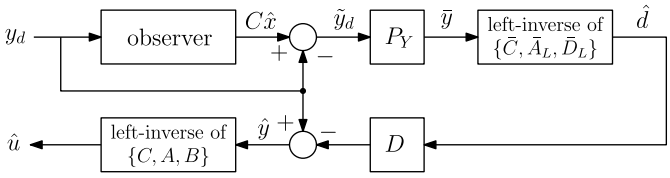


Fig. 11. Strategy to reconstruct u for WOR systems.

Remark 9: A numerical algorithm for the computation of the matrix representation of the insertion map D, based on the method presented in the proof of Proposition 3, is provided in Appendix B. We stress that other solutions may exist, given the fact that Theorem 3 provides a sufficient condition. It is indeed an open problem whether the proposed algorithm to find $\mathscr{D}$ defines the supremal element of all disturbance spaces solving Problem 2.

Strategy for reconstructing the input in WOR systems: The strategy for producing a reconstruction $\hat{u}$ of u, which stems from the results presented in this section, is illustrated by the scheme in Fig. 11. The observer (18) with injection gain $L \in \mathbf{L}(\mathscr{S}^*)$ (selected such that the spectrum of A_L is stable) is used to generate the "output error" $\tilde{y}_d$ of system (20). Projection modulo $C\mathscr{S}^*$ of $\tilde{y}_d$ yields the output $\bar{y}$ of system (22), which is used to reconstruct the disturbance $d(\cdot)$ by way of left-inversion of the triplet $\{\bar{C}, \bar{A}_L, \bar{D}_L\}$. The resulting signal $\hat{d}(\cdot)$ is used to reconstruct the output $y(\cdot)$ of the unperturbed system (1) as $\hat{y} = y_d - D\hat{d}$. The signal $\hat{y}(\cdot)$, in turn, is used to generate $\hat{u}(\cdot)$ via left-inversion of system (1). Alternatively, the lower dimensional model $\Sigma^\flat = \{C^\flat, A_L^\flat, B^\flat\}$ can be used in place of the original system, as this option may be advantageous from a computational standpoint. Specifically, once the disturbance signal has been reconstructed, a reconstruction of the output $\tilde{y}$ of the error system (19) is provided by $\hat{\tilde{y}} = \tilde{y}_d - D\hat{d}$. The signal $\hat{\tilde{y}}(\cdot)$ is then used to reconstruct $u(\cdot)$ by way of left-inversion of $\Sigma^\flat$.

C. Attacker's Sensitivity to Disturbance

In the case of WOR systems, since $\mathscr{Y} = \operatorname{Im} C$, the only subspace $\mathscr{D}_a \subseteq \mathscr{Y}$ that can be chosen by the attacker such that $\mathscr{Y} = \operatorname{Im} C \oplus \mathscr{D}_a$ is the trivial subspace, i.e., $\mathscr{D}_a = 0$. This implies that no projection methods can be used to eliminate the effect of the disturbance on the output. There is, however, one possible strategy for the attacker to correctly reconstruct the input u *if* it knows $\mathscr{D}$, as it can implement the same strategy as the defense. This class of attacker may be considered as "subspace aware," and strategies addressing this eventuality will be considered in future work. Further analysis of the obfuscation properties of our scheme, namely if it is necessary at all for the attacker to have knowledge of the exact $\mathscr{D}$ used by the defender, is also left for future work. In Section VI, we show how our proposed strategy may indeed be a solution to the obfuscation problem, by considering an attacker using an LS algorithm to reconstruct u from y_d.

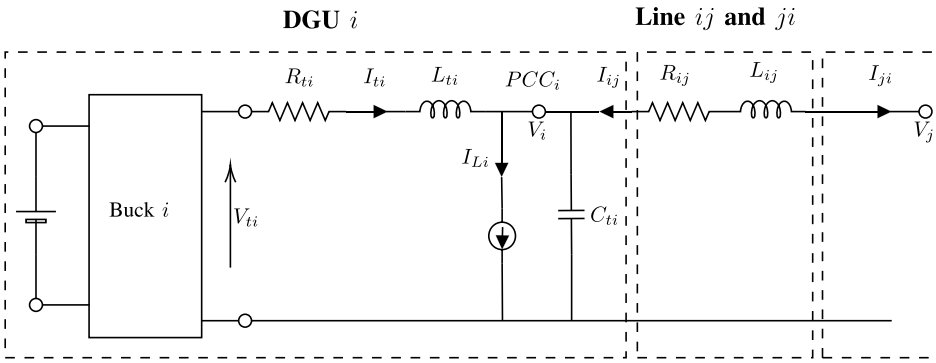


Fig. 12. Schematic diagram of the coupled DGUs.

VI. Simulation Example: WOR Systems

In this section, we provide an example for WOR systems, showing that Problem 1 is indeed solved for our design of $\mathscr{D}$, as well as demonstrating how the disturbance injection can bias the estimate of the input by an eavesdropping attacker. We consider an islanded dc microgrid composed of two coupled distributed generation units (DGUs), where buck converters are coupled via RLC filters (see Fig. 12). Each DGU can be modeled as the following dynamical system [30]:

$$\begin{aligned} C_i\dot{V}_i &= I_{ti} - I_{Li} + \sum_{j\in\mathcal{N}_i} \frac{V_j - V_i}{R_{ij}} \\ L_{ti}\dot{I}_{ti} &= V_{ti} - V_i - R_{ti}I_{ti} \end{aligned} \tag{32}$$

where V_i, I_{ti} are the voltage at the point of common connection and terminal current, V_{ti} is the terminal voltage, and I_{Li} is a load current. The terminal voltage V_{ti} is the control input, which is obtained via the decentralized PI controller defined in [30], ensuring that V_i asymptotically tracks an externally defined reference. The parameters C_i, R_{ti}, L_{ti} are the capacitance, inductance and resistance of the DGU's RLC filter. The system composed of two DGUs can be cast as (1) by taking the closed-loop dynamics from [30], and $u = I_L = \mathrm{col}_{i\in\{1,2\}} I_{Li}$, and $y = \mathrm{col}_{i\in\{1,2\}} \begin{bmatrix} V_i & I_{ti} \end{bmatrix}^\top$. In this context, the estimation of u can be used to achieve improved performance and system balancing by the system operator. However, an eavesdropping attacker may exploit estimated values of the load currents to infer users' personal habits and behaviors [3].

We consider an attacker capable of storing $r \in \mathbb{N}$ successive samples of y_d, and then performing an LS algorithm, such as the one proposed in [28], to construct an estimate $\hat{u}_a$ of the input. Specifically, the attacker solves the following LS problem:

$$\hat{u}_a \doteq \arg\min_{\underline{\hat{u}}(k-1)} \frac{1}{2} \left\| \underline{y_d}(k) - \Phi \underline{\hat{u}}(k-1) \right\|^2 \tag{33}$$

where $\underline{y_d}(k) = \mathrm{col}(y_d(k-r), \ldots, y_d(k))$, $\underline{\hat{u}}(k) = \mathrm{col}(\hat{u}(k-r), \ldots, \hat{u}(k))$, and Φ is composed of the Markov parameters of the system with respect to the input matrix. For r larger than the relative degree of the system, and for $\mathscr{D} = 0$, this algorithm gives exact reconstruction of u with some delay. To illustrate the effects of injecting Dd onto the communicated values of y, we consider two cases: a. the case in which $d = 0$; b. the case in which a disturbance signal is applied.

A. No Disturbance

The results of the first scenario, in which no dithering signal is injected, are presented in Fig. 13. In the figure, a comparison of the unknown input with various estimates is provided, element-wise. We show the estimates $\hat{u}$, defined by the defender, and $\hat{u}_a$, defined by the attacker as in Definition 3. As expected, given the left-invertibility of the system, all estimates are accurate.

B. Injected Disturbance

In this second scenario, we consider the case in which the dithering signal Dd is injected into the output. Here we choose D following the steps stated in Appendix B, which yields:

$$\mathtt{D} = \begin{bmatrix} 0.7957 & 0.5072 \\ -0.0220 & -0.0140 \\ 0 & 0 \\ -0.5445 & -0.4655 \\ 0.0135 & 0.0116 \\ 0 & 0 \end{bmatrix}.$$

Recall that the results presented in this article do not rely on any specific realization of $d(k)$, so long as a valid $\mathscr{D}$ (and therefore $\mathtt{D}$) is designed. To substantiate the results in the remainder of this section, however, let us propose the following definition:

$$\begin{aligned} \breve{x}^+ &= \mathtt{A}\breve{x} + \mathtt{B}\breve{u} \\ d &= \mathtt{D}^\dagger(\mathtt{C}\breve{x} - y) \end{aligned} \tag{34}$$

where $\breve{x}$ is a *simulated* state that has the same dynamics as the system, $\breve{x}(0) = 0$, and $\breve{u}$ is a fictitious input meant to obfuscate the true value of u. The design of this injected signal is inspired by *covert attacks*, a class of data injection attacks that can disrupt certain anomaly detection algorithms [25]. As mentioned, the definition (34) is provided as a practical example, and further investigation of how to *best* design $d(k)$ for obfuscation purposes is left for future work.

As shown in Fig. 14, the eavesdropper using the LS technique is no longer capable of estimating the unknown input without a

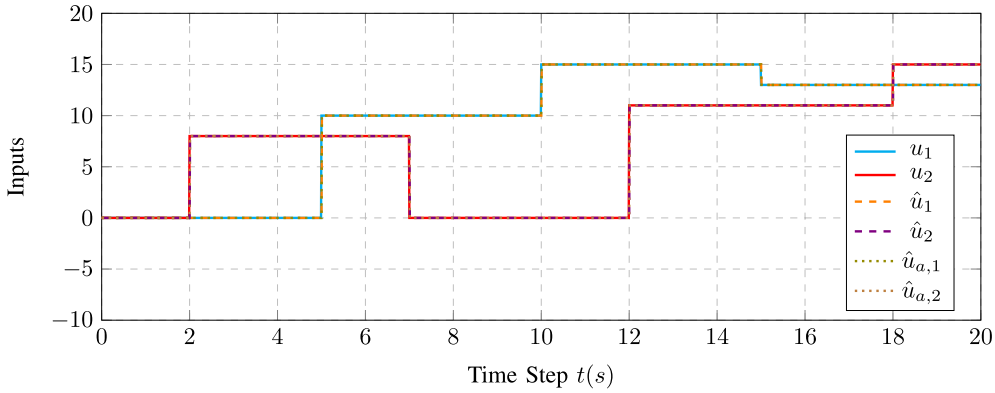


Fig. 13. Actual input ($u_i, i \in \{1,2\}$) and its reconstruction from the defender ($\hat{u}_i, i \in \{1,2\}$) and the LS eavesdropper ($\hat{u}_{a,i}, i \in \{1,2\}$), without injected signal, i.e., $Dd(k) = 0$ $forallk$.

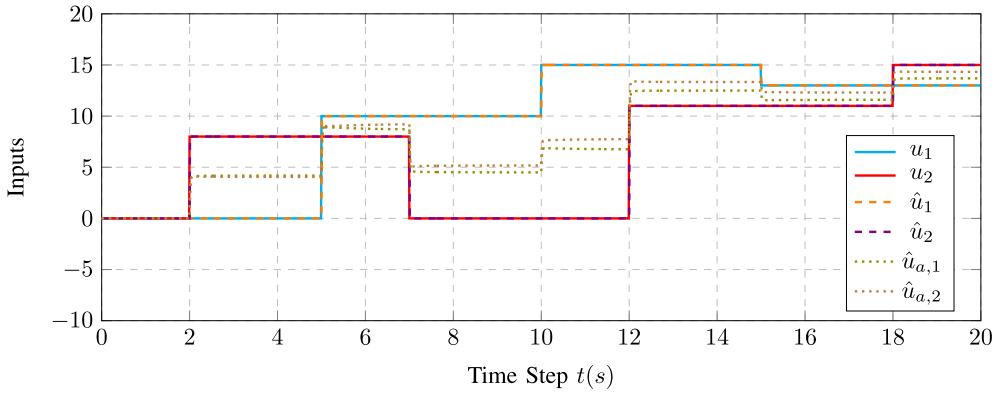


Fig. 14. Actual input ($u_i, i \in \{1,2\}$) and its reconstruction from the defender ($\hat{u}_i, i \in \{1,2\}$) and the LS eavesdropper ($\hat{u}_{a,i}, i \in \{1,2\}$), with injected signal, i.e., $Dd(k) \neq 0$ $forallk$.

bias, but the defender is still capable of reconstructing the input u precisely.

VII. Conclusions and Future Work

In this article, we have introduced the concept of output redundancy in LTI systems and investigated the resulting properties from a geometric approach perspective. We have differentiated between strong output redundancy and weak output redundancy based on the dimensions of the image of the output map and that of the output space. For the SOR systems, we have shown that the output redundancy resides in the quotient space of the output space modulo the image of the output map. Conversely, in the WOR systems, output redundancy is associated with the infimal unobservability subspace that contains the image of the input map. Throughout this article, we have focused on a synthesis problem involving the concealment of input information by means of output allocation. We have shown that output redundancy is indeed necessary, and provided constructive solutions for the considered classes of output-redundant systems. For future work, we aim to analyze the effectiveness of the proposed output allocation strategy against attackers employing more advanced reconstruction strategies. In addition, exploring how output redundancy can be leveraged to enhance output-feedback controller design offers another stimulating direction.

Appendix

A. Numerical Algorithm for Calculating the Matrix $\mathtt{L}$

In this section, we introduce the algorithm for calculating $\mathtt{L}$, the matrix representation of $L \in \mathbf{L}(\mathscr{S}^*)$ regarding system (1) while enabling the spectrum of $\bar{A}_L$ and $A_L^\flat$ ($A_L := A + LC$) to be freely assigned. More specifically, we intend to find an $\mathtt{L} \in \mathbb{R}^{n\times p}$ such that

$$PA_L\mathscr{S}^* = 0, \quad \sigma\left(\bar{A}_L\right) = \bar{\Lambda}, \quad \sigma(A_L^\flat) = \Lambda^\flat \tag{35}$$

where $\bar{\Lambda}$ and $\Lambda^\flat$ denote symmetric sets of arbitrary complex numbers with compatible set numbers, and $P : \mathscr{X} \to \bar{\mathscr{X}}$ is the canonical projection, as shown in Fig. 6.

We first calculate a basis of $\mathscr{S}^*$ ($\dim(\mathscr{S}^*) = s$) using the unobservability subspace algorithm [20, Theorem 18], which is a matrix representation $\mathtt{S} \in \mathbb{R}^{n\times s}$ of the insertion map $S : \mathscr{S}^* \to \mathscr{X}$. Then, it is easy to calculate $\mathtt{P} \in \mathbb{R}^{(n-s)\times n}$ based on $\mathtt{S}$

$$\mathtt{PS} = \mathtt{0}_{(n-s)\times s}, \quad \operatorname{rank}\mathtt{P} = n - s.$$

Subsequently, we find an arbitrary $\mathtt{L}_0 \in \mathbb{R}^{n\times p}$ satisfying

$$\mathtt{P}(\mathtt{A} + \mathtt{L}_0\mathtt{C})\mathtt{S} = \mathtt{0}_{(n-s)\times s}.$$

We thereby compute $\mathtt{A}_0$ as

$$\mathtt{A}_0 = \mathtt{P}(\mathtt{A} + \mathtt{L}_0\mathtt{C})\mathtt{P}^\dagger.$$

Similar to calculating $\mathtt{P}$, we compute $\mathtt{P}_{\mathtt{Y}} \in \mathbb{R}^{(p-z)\times p}$ ($\dim(C\mathscr{S}^*) = z$) by

$$\mathtt{P}_{\mathtt{Y}}\mathtt{S}_{\mathtt{Y}} = \mathtt{0}_{(p-z)\times z},\ \operatorname{rank}\mathtt{P}_{\mathtt{Y}} = p - z$$

where $\mathtt{S}_{\mathtt{Y}} \in \mathbb{R}^{p\times z}$ is a basis of the subspace $C\mathscr{S}^*$. Moreover, calculate $\mathtt{C}_0 := \mathtt{P}_{\mathtt{Y}}\mathtt{C}\mathtt{P}^\dagger$, and choose $\mathtt{L}_{0\mathtt{a}}$ such that $\sigma(\mathtt{A}_0 + \mathtt{L}_{0\mathtt{a}}\mathtt{C}_0) = \bar{\Lambda}$. After that, we let $\mathtt{L}_1 = \mathtt{L}_0 + \mathtt{P}^\dagger\mathtt{L}_{0\mathtt{a}}\mathtt{P}_{\mathtt{Y}}$, where it can be verified that

$$\sigma(\mathtt{P}(\mathtt{A} + \mathtt{L}_1\mathtt{C})\mathtt{P}^\dagger) = \bar{\Lambda}.$$

Now we define

$$\mathtt{A}_1 := \mathtt{S}^\dagger(\mathtt{A} + \mathtt{L}_1\mathtt{C})\mathtt{S}$$
$$\mathtt{C}_1 := \mathtt{C}\mathtt{S}.$$

Clearly, $\mathtt{A}_1$ and $\mathtt{C}_1$ are the matrix representation of the map $(A + L_1C)|\mathscr{S}^*$ and $C|\mathscr{S}^*$, respectively. Since $(\mathtt{C}_1, \mathtt{A}_1)$ is observable (Proposition 2), there exists an $\mathtt{L}_{1\mathtt{a}} \in \mathbb{R}^{s\times p}$ such that

$$\sigma(\mathtt{A}_1 + \mathtt{L}_{1\mathtt{a}}\mathtt{C}_1) = \Lambda^\flat.$$

Now we let $\mathtt{L} = \mathtt{L}_1 + \mathtt{S}\mathtt{L}_{1\mathtt{a}}$, which completes the computation of $\mathtt{L}$ satisfying (35).

B. Numerical Algorithm for Calculating the matrix $\mathtt{D}$

Suppose we have calculated $\mathtt{L}$ as described in Appendix A with the desired spectrum for $A_L^\flat$ and $\bar{A}_L$. Then, we calculate $\bar{\mathtt{L}}^\flat = \mathtt{P}\mathtt{L}\mathtt{S}_{\mathtt{Y}} \in \mathbb{R}^{(n-s)\times z}$, and we obtain the rank of $\bar{\mathtt{L}}^\flat$ as τ. Then we select a matrix $\mathtt{D}_{\mathtt{a}}^\flat \in \mathbb{R}^{z\times\tau}$ such that

$$\operatorname{rank}(\bar{\mathtt{L}}^\flat\mathtt{D}_{\mathtt{a}}^\flat) = \tau.$$

We calculate a basis of $\bar{\mathscr{V}}^*$ [defined in (23)] based on [19, Theorem 4.3], denoted by $\mathtt{V} \in \mathbb{R}^{(n-p)\times v}$, where

$$\bar{\mathtt{A}}_{\mathtt{L}} = \mathtt{P}(\mathtt{A} + \mathtt{L}\mathtt{C})\mathtt{P}^\dagger,\quad \bar{\mathtt{D}}_{\mathtt{L}_{\mathtt{a}}} = \bar{\mathtt{L}}^\flat\mathtt{D}_{\mathtt{a}}^\flat,\quad \bar{\mathtt{C}} = \mathtt{P}_{\mathtt{Y}}\mathtt{C}\mathtt{P}^\dagger.$$

Subsequently, we check the solution of

$$\mathtt{V}\theta_i = \bar{\mathtt{D}}_{\mathtt{L}_{\mathtt{a}}}[i],\ i \in \{1, \ldots, \tau\} \tag{36}$$

where $\bar{\mathtt{D}}_{\mathtt{L}_{\mathtt{a}}}[i]$ represents the ith column of the matrix $\bar{\mathtt{D}}_{\mathtt{L}_{\mathtt{a}}}$. If (36) cannot be solved for any $i \in \{1, \ldots, \tau\}$, then we calculate $\mathtt{D} = \mathtt{S}_{\mathtt{Y}}\mathtt{D}_{\mathtt{a}}^\flat$, and we obtain the desired $\mathtt{D}$.

Otherwise, we record the index i where (36) has a solution for θ_i, and we discard the corresponding columns of $\bar{\mathtt{D}}_{\mathtt{L}_{\mathtt{a}}}$, resulting in a new matrix $\bar{\mathtt{D}}_{\mathtt{L}_{\mathtt{b}}} \in \mathbb{R}^{(n-p)\times\tau_b}$ with fewer number of columns ($\tau_b < \tau$). Then, we compute $\mathtt{D} = \mathtt{S}_{\mathtt{Y}}\mathtt{D}_{\mathtt{a}}^\flat$, and we obtain the desired $\mathtt{D}$.

REFERENCES

[1] K. E. Hemsley and D. R. E. Fisher, "History of industrial control system cyber incidents," Idaho Nat. Lab. (INL), Idaho Falls, ID, USA, Tech. Rep. NL/CON-18-44411-Rev002, 2018.

[2] M. S. Chong, H. Sandberg, and A. M. Teixeira, "A tutorial introduction to security and privacy for cyber-physical systems," in *Proc. 18th Eur. Control Conf.*, 2019, pp. 968–978.

[3] P. Kumar, Y. Lin, G. Bai, A. Paverd, J. S. Dong, and A. Martin, "Smart grid metering networks: A survey on security, privacy and open research issues," *IEEE Commun. Surv. Tut.*, vol. 21, no. 3, pp. 2886–2927, Jul.–Sep. 2019.

[4] Z. Chen, C. Jiang, and L. Xie, "Building occupancy estimation and detection: A review," *Energy Buildings*, vol. 169, pp. 260–270, 2018.

[5] R. Alisic, M. Molinari, P. E. Paré, and H. Sandberg, "Ensuring privacy of occupancy changes in smart buildings," in *Proc. IEEE Conf. Control Technol. Appl.*, 2020, pp. 871–876.

[6] A. Sultangazin and P. Tabuada, "Symmetries and isomorphisms for privacy in control over the cloud," *IEEE Trans. Autom. Control*, vol. 66, no. 2, pp. 538–549, Feb. 2021.

[7] A. M. Naseri, W. Lucia, and A. Youssef, "A privacy preserving solution for cloud-enabled set-theoretic model predictive control," in *Proc. Eur. Control Conf.*, 2022, pp. 894–899.

[8] H. Hayati, C. Murguia, and N. Van De Wouw, "Privacy-preserving federated learning via system immersion and random matrix encryption," in *Proc. 61st IEEE Conf. Decis. Control*, 2022, pp. 6776–6781.

[9] G. Yang, H. Rezaee, and T. Parisini, "Sensor redundancy for robustness in nonlinear state estimation," in *Proc. 58th IEEE Conf. Decis. Control*, 2019, pp. 3865–3870.

[10] N. P. Winkler, P. P. Neumann, E. Schaffernicht, and A. J. Lilienthal, "Using redundancy in a sensor network to compensate sensor failures," in *Proc. IEEE Sensors*, 2021, pp. 1–4.

[11] G. Yang, H. Rezaee, A. Serrani, and T. Parisini, "Sensor fault-tolerant state estimation by networks of distributed observers," *IEEE Trans. Autom. Control*, vol. 67, no. 10, pp. 5348–5360, Oct. 2022.

[12] Y. Ren, Z. Duan, P. Duan, and L. Shi, "On the effects and optimal design of redundant sensors in sensor networks," 2023, *arXiv:2306.09215*.

[13] A. Cristofaro and L. Zaccarian, "Adaptive pseudo-inverse observers for output redundant discrete-time linear systems," *IEEE Trans. Autom. Control*, vol. 69, no. 11, pp. 7403–7417, Nov. 2024.

[14] D. Corona and A. Cristofaro, "Some remarks on optimal output regulation for weakly dual redundant plants," in *Proc. 24th Mediterranean Conf. Control Automat.*, 2016, pp. 1205–1211.

[15] L. Zaccarian, "Dynamic allocation for input redundant control systems," *Automatica*, vol. 45, no. 6, pp. 1431–1438, 2009.

[16] G. Yang, A. Barboni, H. Rezaee, A. Serrani, and T. Parisini, "Joint state, disturbance and fault estimation for weakly output redundant discrete-time linear systems," in *Proc. 61st IEEE Conf. Decis. Control*, 2022, pp. 904–909.

[17] A. Serrani, "Output regulation for over-actuated linear systems via inverse model allocation," in *Proc. 51st IEEE Conf. Decis. Control*, 2012, pp. 4871–4876.

[18] J. Kreiss and J.-F. Trègouët, "Input redundancy: Definitions, taxonomy, characterizations and application to over-actuated systems," *Syst. Control Lett.*, vol. 158, 2021, Art. no. 105060.

[19] W. M. Wonham, *Linear Multivariable Control: A Geometric Approach*, 3rd ed. New York, NY, USA: Springer, 1985.

[20] M.-A. Massoumnia, "A geometric approach to failure detection and identification in linear systems," Ph.D. Diss., Massachusetts Inst. Technol., 1986.

[21] G. Basile and G. Marro, *Controlled and Conditioned Invariants in Linear System Theory*. Englewood Cliffs, NJ, USA: Prentice Hall, 1992.

[22] A. Morse and W. Wonham, "Status of noninteracting control," *IEEE Trans. Autom. Control*, vol. AC-16, no. 6, pp. 568–581, Dec. 1971.

[23] L. Silverman, "Inversion of multivariable linear systems," *IEEE Trans. Autom. Control*, vol. AC-14, no. 3, pp. 270–276, Jun. 1969.

[24] H. Trentelman, A. Stoorvogel, and M. Hautus, *Control Theory for Linear Systems*. London, U.K.: Springer, 2001.

[25] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135–148, 2015.

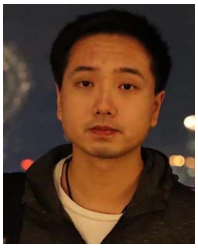
[26] M. Sain and J. Massey, "Invertibility of linear time-invariant dynamical systems," *IEEE Trans. Autom. Control*, vol. AC-14, no. 2, pp. 141–149, Apr. 1969.

[27] M. E. Valcher, "State observers for discrete-time linear systems with unknown inputs," *IEEE Trans. Autom. Control*, vol. 44, no. 2, pp. 397–401, Feb. 1999.

[28] A. Ansari and D. S. Bernstein, "Deadbeat unknown-input state estimation and input reconstruction for linear discrete-time systems," *Automatica*, vol. 103, pp. 11–19, 2019.

[29] A. S. Morse, "Structural invariants of linear multivariable systems," *SIAM J. Control*, vol. 11, no. 3, pp. 446–465, 1973.

[30] M. Tucci, S. Riverso, and G. Ferrari-Trecate, "Line-independent plug-and-play controllers for voltage stabilization in DC microgrids," *IEEE Trans. Control Syst. Technol.*, vol. 26, no. 3, pp. 1115–1123, May 2018.



Guitao Yang received the B.Eng. degree in electrical and electronic engineering from The University of Manchester, Manchester, U.K., in 2016, and the M.Sc. and Ph.D. degrees in control systems from Imperial College London, London, U.K., in 2017 and 2022, respectively.

From 2025, he was a Research Associate with Wolfson School of Mechanical Electrical and Manufacturing Engineering, Loughborough University, Loughborough, U.K. His research interests include distributed state estimation, fault-tolerant observers, and the geometric approach.

Alexander J. Gallo received the M.Eng. degree in electrical and electronic engineering and the Ph.D. degree in control engineering from Imperial College London, London, U.K., in 2016 and 2021, respectively.

September 2024, he was a Postdoctoral Researcher with the Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy. His research interests include distributed cybersecurity and fault tolerant control for large-scale interconnected systems, with a particular focus on energy distribution networks, as well as health-aware and fault tolerant control of wind turbines.

Angelo Barboni received the M.Sc. degree in electrical and control engineering (with honors) from the University of Trieste, Trieste, Italy, in 2015, and the Ph.D. degree in control engineering from Imperial College London, London, U.K., in 2021.

From 2021 to 2023, he was a Research Associate with the Control and Power group in the Department of Electrical and Electronic Engineering, Imperial College London. Since 2023, he has been a Researcher with Group Accumulation Management, Zurich Insurance Group, Zurich, Switzerland, working on models for risk quantification in the cybersecurity domain. His research interests include (distributed) estimation, anomaly detection, fault-tolerant control, and optimization. Dr. Barboni was the recipient of the scholarship with the High-Performance Embedded and Distributed Systems (HiPEDS) Centre for Doctoral Training, Imperial College London.

Riccardo M. G. Ferrari (Senior Member, IEEE) received the Laurea degree (*cum laude* and printing honours) in electronic engineering, and the Ph.D. degree in information engineering from the University of Trieste, Trieste, Italy, in 2004 and 2009, respectively.

He held both academic and industrial R&D positions, in particular as Researcher in the field of process instrumentation and control for the steel-making sector. He is a Marie Curie alumnus and currently an Associate Professor with the Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands. His research interests include fault tolerant control and fault diagnosis and attack detection in large-scale cyber–physical systems, with applications to wind energy generation, electric mobility, and cooperative autonomous vehicles.

Dr. Ferrari was a co-recipient of the O. Hugo Schuck Award 2023.

Andrea Serrani (Member, IEEE) received the B.Eng (*Laurea*) degree in electrical engineering and the Ph.D. degree from the University of Ancona, Ancona, Italy, in 1993 and 1997, respectively. He received the M.S. and D.Sc. degrees in systems science and mathematics from Washington University in St. Louis, St. Louis, MO, USA, in 1996 and 2000, respectively.

From 1994 to 1999, he was a Fulbright Fellow with Washington University in St. Louis. From 2002 to 2024, he was with The Ohio State University, Columbus, OH, USA. Since 2024, he has been with the University of Bologna, Bologna, Italy, where he is currently a Full Professor. He has held multiple summer faculty fellowships at AFRL. He has authored or coauthored more than 150 articles in journals, proceedings of international conferences and book chapters, and has coauthored the book *Robust Autonomous Guidance: An Internal Model Approach* published by Springer-Verlag. His research focuses on the intersection of nonlinear, adaptive, and geometric control theory with applications in aerospace and marine systems, fluidic systems, robotics and automotive engineering. His work has been supported by AFRL, NSF, Ford Motor Co., and NASA, among others.

Dr. Serrani was a Distinguished Lecturer of the IEEE CSS, and was the Editor-in-Chief of the IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY (2017–2024) and the Associate Editor for the same journal (2010–2016), *Automatica* (2008–2014), and the *International Journal of Robust and Nonlinear Control* (2006–2014). He was on the Conference Editorial Boards of IEEE CSS and IFAC, as Program Chair for the 2019 ACC, and as General Co-Chair for the 2022 IEEE CDC. He currently serves as VP-Publications and on the Board of Governors of the IEEE CSS.

Thomas Parisini (Fellow, IEEE) received the Ph.D. degree in electronic engineering and computer science from the University of Genoa, Genoa, Italy, in 1993.

Since 2001, he has been the Danieli Endowed Chair of automation engineering with the University of Trieste, Trieste, Italy, where from 2009 to 2012, he was the Deputy Rector. In 2023, he held a "Scholar-in-Residence" visiting position with Digital Futures-KTH, Stockholm, Sweden. He currently holds the Chair of Industrial Control, and is the Head of the Control and Power Research Group, Imperial College London, London, U.K. He also holds a Distinguished Professorship with Aalborg University, Aalborg, Denmark. He has authored or coauthored a research monograph in the Communication and Control Series, Springer Nature, and more than 400 research papers in archival journals, book chapters, and international conference proceedings.

Dr. Parisini was the recipient of the Knighthood of the Order of Merit of the Italian Republic for scientific achievements abroad awarded by the Italian President of the Republic in 2023. In 2018, he was the recipient of the Honorary Doctorate from the University of Aalborg, Denmark, and in 2024, the IEEE CSS Transition to Practice Award. Moreover, he was the recipient of the 2007 IEEE Distinguished Member Award, and was co-recipient of the IFAC Best Application Paper Prize of the *Journal of Process Control*, *Elsevier*, for the three-year period 2011–2013 and of the 2004 Outstanding Paper Award of IEEE TRANSACTIONS ON NEURAL NETWORKS. In 2016, he was awarded as Principal Investigator with Imperial of the H2020 European Union flagship Teaming Project KIOS Research and Innovation Centre of Excellence led by the University of Cyprus with an overall budget of over 40 million Euros. He was the 2021–2022 President of the IEEE Control Systems Society and he was the Editor-in-Chief for IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY (2009–2016). He was the Chair of the IEEE CSS Conference Editorial Board (2013–2019). He was the Associate Editor for several journals including the IEEE TRANSACTIONS ON AUTOMATIC CONTROL and the IEEE TRANSACTIONS ON NEURAL NETWORKS. He is currently an Editor for *Automatica* and the Editor-in-Chief for the *European Journal of Control*. He was the Program Chair of the 2008 IEEE Conference on Decision and Control and General Co-Chair of the 2013 IEEE Conference on Decision and Control. He is a Fellow of IFAC. He is a Member of IEEE TAB Periodicals Review and Advisory Committee.