ANNUAL REVIEWS

Annual Review of Control, Robotics, and Autonomous Systems

Fault Diagnosis in Dynamical Systems: Geometric Interpretation and Tractable Algorithms

Mohammad Amin Sheikhi,[1] Gabriel de Albuquerque Gleizer,[1] Tamás Keviczky,[1] and Peyman Mohajerin Esfahani[1,2]

[1]Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands; email: m.a.sheikhi@tudelft.nl, g.gleizer@tudelft.nl, t.keviczky@tudelft.nl

[2]Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario, Canada; email: p.mohajerinesfahani@utoronto.ca

Annu. Rev. Control Robot. Auton. Syst. 2026. 9:147–79

First published as a Review in Advance on December 10, 2025

The *Annual Review of Control, Robotics, and Autonomous Systems* is online at control.annualreviews.org

https://doi.org/10.1146/annurev-control-030123-015422

Keywords

fault detection and isolation, fault estimation, behavioral set, robust linear residual generators, nonlinear systems, active estimation, convex optimization

Abstract

This survey reviews recent developments in fault diagnosis for both linear and nonlinear dynamical systems, covering model-based and data-driven approaches as well as passive and active detection and estimation methods. A central focus is placed on the geometric interpretation of diagnosis filters and their connection to the concept of behavioral sets, providing an intuitive view of their performance. We also review optimization-based techniques that enhance the robustness of linear filters when applied to nonlinear or uncertain systems. Furthermore, we point out recent progress in active fault diagnosis, where input design plays a key role in improving detectability and estimation accuracy. To bridge theory and practice, we include a set of real-world industrial applications that demonstrate the implementation and effectiveness of these methods in realistic settings.

1. INTRODUCTION

Fault: an undesirable deviation from the nominal condition, caused by sources such as internal malfunctions, hardware degradation, parameter variations, external disturbances, or stealthy (adversarial) manipulations

Health monitoring and control of engineering systems under potential faults or anomalies is a critical concern in modern industry. This broad area of research, commonly referred to as fault diagnosis, includes a diverse set of problems (1). This survey provides a structured overview of the literature, organized around key problem formulations, system dynamics and information settings, and various solution approaches developed to address them.

1.1. Problems

Fault diagnosis typically comprises four interrelated problems: fault detection, fault isolation, fault estimation, and fault mitigation or fault-tolerant control. Fault detection refers to the ability to determine whether an unexpected event or anomaly has occurred in a system (2, 3). Once a fault is detected, fault isolation aims to identify the specific source or location of the fault within the system (4, 5); it extends fault detection by distinguishing among multiple potential fault sources to determine which has occurred. Fault estimation is the next level of fault isolation and aims to quantify the magnitude, dynamics, or temporal evolution of the fault (6, 7). Finally, fault mitigation involves designing corrective actions or control strategies to minimize the impact of the fault or restore system functionality (8, 9). While the diagnosis problems exhibit increasing conceptual and technical complexity from detection to mitigation, each poses distinct theoretical and practical challenges depending on the system setting and available information. This survey focuses on fault detection, isolation, and estimation (skipping mitigation), with particular emphasis on the geometric interpretation of diagnosis filters and the intuition behind how they partition the measurement space.

1.2. Settings

The analysis and design of fault diagnosis filters depend on several key components, including (*a*) the underlying system dynamics (linear versus nonlinear dynamics), (*b*) the available information (exact versus uncertain models), and (*c*) the degree of freedom (passive versus active diagnosis). With regard to the dynamics, the majority of fault diagnosis techniques have traditionally been developed for linear time-invariant (LTI) systems (10, 11), where the system's behavior can be fully captured by a fixed set of matrices. However, real-world systems often exhibit nonlinear behaviors, possibly influenced by multiplicative noise (12), time-varying parameters (13), or other general nonlinearities (e.g., 14–16). In such cases, fault diagnosis becomes significantly more complex. Another important consideration is the extent of model information available to the diagnosis algorithm. In some cases, the model is only partially known or contains structured uncertainties, requiring robust or adaptive approaches (17, 18). In the extreme case, no model information is available, and data-driven methods are required (19, 20). Furthermore, prior knowledge about the class of possible fault signals, such as their temporal profiles (21), magnitudes, or sparsity (13, 22), can significantly influence the design of detection and estimation schemes. Depending on the assumptions about fault signal behavior and uncertainty, fault diagnosis can be approached from a robust/adversarial perspective or a probabilistic/statistical framework (17), each of which offers different guarantees and trade-offs.

An alternative lens through which one can strike a balance between these knowns and unknowns at the modeling level is game theory, where the interaction between the diagnosis system and the fault signal can be modeled as a game between two players: a defender (the diagnosis filter) and an attacker (the fault signal). In this formulation, the diagnosis algorithm must be robust against worst-case scenarios, where the adversarial signal may actively seek to evade

detection. It is worth noting that this game-theoretic viewpoint is typically less conservative than the classical robust approach, as it explicitly accounts for the strategic interaction between attacker and defender. The game-theoretic perspective is particularly suited for security-critical applications, such as cyber-physical systems (23) and autonomous vehicles (24), where faults may be deliberately introduced by intelligent attackers. However, game-theoretic formulations are beyond the scope of this survey and are not covered.

While most existing work has focused on passive fault diagnosis, where the system is monitored without intervention, a promising yet underexplored direction is active fault diagnosis. This approach involves designing excitation inputs to improve fault detectability or enhance diagnostic performance. Active strategies are particularly valuable in low signal-to-noise ratio settings (25) or when fault effects on measurements closely resemble those caused by normal disturbances (26). Despite improving diagnosis performance, the design of optimal probing inputs poses its own set of additional challenges, as it often results in intractable, nonconvex optimization problems.

Residual: a diagnostic signal that indicates the presence of a fault, remains decoupled from other inputs, and is computed from known signals

1.3. Solution Approaches

A core aspect of fault diagnosis is the design of filters or classifiers that process input–output measurements and generate residuals or decision signals indicative of faults. These filters are typically modeled as either linear or nonlinear systems and are often parameterized to optimize performance criteria such as sensitivity, robustness, or detection delay. While linear residual generators are well-established for LTI systems (27, 28), nonlinear observers (14, 29) and data-driven classifiers—particularly those based on machine learning (30)—are gaining attention in handling nonlinear dynamics. The filter design inherently involves trade-offs between fault sensitivity and robustness to modeling uncertainties and disturbances. A unifying and intuitive way to interpret diagnosis filters is through their geometric effect on the measurement space (31). Specifically, the filter induces a partition of the input–output measurement space, where each region corresponds to a particular fault mode (32). In the case of linear filters, this often results in subspace partitioning, where the residual space is structured to separate nominal and faulty behaviors (33).

The survey discussion is organized based on system dynamics, available information, filter parameterization, and solution methods, with particular attention to their geometric insights. We also review emerging frameworks in game-theoretic and active detection, offering a comprehensive overview of current methodologies and outlining key directions for future research. Furthermore, we illustrate the relevance of these methods through several modern applications, including cyber-security in power systems, anomaly detection in autonomous vehicles, and health monitoring of high-end industrial printers, supported by real-world case studies and, where applicable, experimental implementations. **Table 1** summarizes the reviewed literature.

1.4. Notation

Sets $\mathbb{N}$, $\mathbb{R}$ ($\mathbb{R}_+$), and $\mathbb{C}$ denote the nonnegative integers (including zero), the (positive) real numbers, and the complex numbers, respectively. $\mathbf{I}$ denotes the identity matrix, with $\mathbf{I}_n$ indicating its size n. For any matrix $\mathbf{X}$, the notations $\overset{\diagdown}{\mathbf{X}}$ and $\overset{\diagup}{\mathbf{X}}$ imply its block Toeplitz and Hankel structures, respectively, while $\mathbf{X}^\dagger$ represents the Moore–Penrose inverse. $\|\cdot\|_\infty$ and $\|\cdot\|_2$ correspond to the infinity norm and 2-norm of a vector, respectively, whereas $\|\cdot\|_{\mathcal{L}_2}$ denotes the $\mathcal{L}_2$-norm of a signal. Signals in this article are (generalized) functions of either discrete or continuous time, depending on context, and often omit the argument—e.g., a multivariate signal $\mathbf{x} : \mathcal{T} \to \mathbb{R}^{n_x}$, where the time set $\mathcal{T}$ is either $\mathbb{R}_+$ for continuous-time systems or $\mathbb{N}$ for discrete-time systems, is generally

Table 1 Summary of the literature reviewed in the survey

Setting	Dynamics	Diagnosis problem		
		Fault detection	Fault isolation	Fault estimation
Perfect model	Linear	11, 23, 27, 28, 34–37	4, 5, 8, 10, 22, 31, 32, 38–42; 26, 43–46	6, 7, 12, 47–51; 25
	Nonlinear	14	14, 52–62	13, 63–66
Imperfect model	Linear	18, 67	68	69–71; 72
	Nonlinear	15, 17, 21, 73, 74	15, 16, 29, 73	24
Data-driven	Linear	20, 30, 75–79	33, 80; 81	19, 82–94
	Nonlinear	95, 96	97–100	101, 102

Non-underlined reference numbers are for works on passive diagnosis; underlined numbers are for works on active diagnosis.

represented in system equations simply by $\mathbf{x}$. We use the operator $\mathfrak{q}$ for the differentiation or the forward time shift, i.e., $\mathfrak{q}\mathbf{x}(t) = \mathbf{x}(t+1)$ for a discrete-time signal $\mathbf{x}$ and $\mathfrak{q}\mathbf{x}(t) = \mathrm{d}\mathbf{x}(t)/\mathrm{d}t$ for a continuous-time signal. Since $\mathfrak{q}$ is a linear and commutative operator, we treat polynomials of $\mathfrak{q}$ as typical LTI systems and occasionally evaluate these polynomials at some value of $\mathfrak{q}$ in $\mathbb{C}$ to reveal system properties such as poles, zeros, and matrix rank.

2. PROBLEM DEFINITION

Given a dynamical system of interest, a diagnosis filter is an engineered system that processes available measurements and provides the required information for detection, isolation, and/or estimation of faults present in the system. The standard setup is illustrated in **Figure 1**. Each block is subject to different multivariate signals. The signals $\mathbf{u} : \mathcal{T} \to \mathbb{R}^{n_u}$ and $\mathbf{y} : \mathcal{T} \to \mathbb{R}^{n_y}$ denote the control input and the dynamical system output, respectively, both of which are known and measurable. The unmeasured signals $\mathbf{f} : \mathcal{T} \to \mathbb{R}^{n_f}$, $\mathbf{w} : \mathcal{T} \to \mathbb{R}^{n_w}$, and $\mathbf{d} : \mathcal{T} \to \mathbb{R}^{n_d}$ represent the fault, noise, and natural disturbance, respectively. While the fault is the signal of interest, the remaining unknown signals are expected to be filtered out by the diagnosis filter. The distinction between noise and disturbance is common in the literature but subtle, as it depends more on the problem than on the nature of the signals $\mathbf{d}$ and $\mathbf{w}$. The disturbance signal $\mathbf{d}$ is generally low-dimensional but has a large influence on the output, and as such it is expected to be decoupled at the diagnosis filter. In contrast, the signal $\mathbf{w}$ includes process and measurement noise that can have a larger dimension but a limited impact, typically being characterized by a random process or a bounded signal. The diagnosis filter processes the available measurements $\mathbf{u}$ and $\mathbf{y}$ to generate the residual $\mathbf{r}$, which is the signal that enables the diagnosis task of interest. In general, the residual must be (approximately) zero irrespective of the natural disturbance $\mathbf{d}$, as long as the fault is zero (i.e., $\mathbf{f} = 0$). At this level of generality, we can formally introduce the residual as the function of the exogenous four signals [i.e., $\mathbf{r}(\mathbf{f}, \mathbf{w}, \mathbf{d}, \mathbf{u})$] and translate the fault detection problem using the

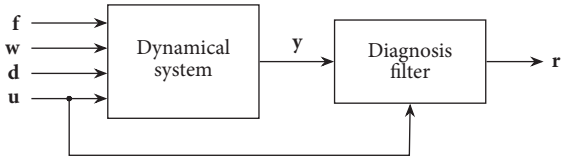

Figure 1

General configuration of systems and diagnosis filters. The filter processes the available control inputs $\mathbf{u}$ and output measurement $\mathbf{y}$ and generates the residual $\mathbf{r}$ that offers information on the fault signal $\mathbf{f}$, regardless of natural disturbance $\mathbf{d}$ and exogenous noise $\mathbf{w}$.

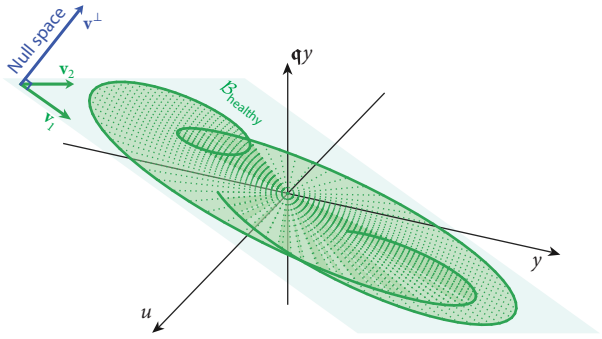

Figure 2

Two characterizations of a hyperplane (*mint green*) containing the healthy input–output trajectories $\mathcal{B}_{\text{healthy}}$, when the diagnosis filter is a linear dynamical system: the span of the hyperplane using the basis $[\mathbf{v}_1, \mathbf{v}_2]$ (*green*) and the span of the null space (i.e., the orthogonal complement) using the basis $\mathbf{v}^\perp$ (*blue*). The trajectory resulting from a sinusoidal input is shown in dark green.

following mappings:

$$\mathbf{d} \mapsto \mathbf{r}(\mathbf{0}, \mathbf{0}, \mathbf{d}, \mathbf{u}) \equiv 0, \quad \forall \mathbf{u}, \mathbf{d} \qquad \text{(disturbance decoupling)}, \tag{1a}$$

$$\mathbf{f} \mapsto \mathbf{r}(\mathbf{f}, \mathbf{0}, \mathbf{d}, \mathbf{u}) \not\equiv 0, \quad \forall \mathbf{u}, \mathbf{d} \qquad \text{(fault sensitivity)}. \tag{1b}$$

Equation 1a ensures perfect decoupling of $\mathbf{d}$ in the noise- and fault-free condition regardless of $\mathbf{u}$, whereas Equation 1b guarantees sensitivity to the fault, even when $\mathbf{d}$ is present.

2.1. Geometric Interpretation

Consider the configuration illustrated in **Figure 1** and the detection problem characterized through the mappings in Equation 1. The disturbance decoupling in Equation 1a essentially characterizes a set containing all the healthy trajectories of the pairs $(\mathbf{u}, \mathbf{y})$ for the underlying dynamical system in the absence of faults $\mathbf{f}$. The minimal among such sets is often referred to as the behavioral set (103) and is denoted hereafter by $\mathcal{B}_{\text{healthy}}$.

The geometry of the set characterized by Equation 1a is determined by the diagnosis filter, i.e., the mapping from $(\mathbf{u}, \mathbf{y})$ to the residual $\mathbf{r}$. When this diagnosis filter is itself a linear dynamical system, the set $\mathcal{B}_{\text{healthy}}$ takes the form of a hyperplane, as depicted in **Figure 2**. There are two ways to characterize $\mathcal{B}_{\text{healthy}}$: (*a*) describing the hyperplane using its basis and (*b*) describing its null space. The former is primarily a subject of the system identification literature. Let us elaborate on this further through a simple example.

Example (characterization of healthy behavior). Consider a first-order stable continuous-time system described with $\mathfrak{q}y + y = u$, where $\mathfrak{q}$ represents the differentiation operator. Equation 1a indicates that $\mathbf{r} \equiv \mathbf{0}$ if $\mathfrak{q}y + y = u$. As such, it defines a hyperplane in the space with coordinates $(u, y, \mathfrak{q}y)$, where one such trajectory results from a sinusoidal input (see **Figure 2**). More specifically, any system trajectory $(u, y, \mathfrak{q}y)$ is healthy if any of the following hold:

$$\begin{bmatrix} u \\ y \\ \mathfrak{q}y \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{bmatrix}}_{[\mathbf{v}_1, \mathbf{v}_2]} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \text{for some signals } \alpha_1, \alpha_2 \iff \underbrace{\begin{bmatrix} -1 & 1 & 1 \end{bmatrix}}_{\mathbf{v}^\perp} \begin{bmatrix} u \\ y \\ \mathfrak{q}y \end{bmatrix} = 0.$$

The pair $[\mathbf{v}_1, \mathbf{v}_2]$ serves as a basis for the healthy hyperplane, while the vector $\mathbf{v}^\perp$ is a basis for the null space. An advantage of the null-space representation is that it readily yields a potential filter $r = \mathfrak{q}y + y - u$.

Differential–algebraic equations (DAEs): a system of equations that includes both differential operations and algebraic terms, representing dynamical systems in a general form

While these characterizations perfectly fit healthy data points in linear systems, this may not hold for nonlinear dynamics, a topic further discussed in Sections 2.3 and 3.1.

2.2. System Representations

To provide a systematic way to characterize the healthy behavior in **Figure 2**, we consider a class of system dynamics as the mapping from $(\mathbf{f}, \mathbf{w}, \mathbf{d}, \mathbf{u})$ to $\mathbf{y}$ (see **Figure 1**) described via differential–algebraic equations (DAEs) as

$$\mathbf{H}(\mathfrak{q})\boldsymbol{\xi} + \mathbf{L}(\mathfrak{q})\mathbf{z} + \mathbf{F}(\mathfrak{q})\mathbf{f} + \mathbf{W}(\mathfrak{q})\mathbf{w} + \boldsymbol{\eta}(\boldsymbol{\xi}, \mathbf{z}) = \mathbf{0}, \tag{2.}$$

where $\boldsymbol{\xi} : \mathcal{T} \to \mathbb{R}^{n_\xi}$ is the unknown latent signal (e.g., internal states or natural exogenous disturbances), $\mathbf{z} : \mathcal{T} \to \mathbb{R}^{n_z}$ is the known measurement signal (e.g., input–output trajectories), $\mathbf{f} : \mathcal{T} \to \mathbb{R}^{n_f}$ is the fault signal (diagnosis target), and $\mathbf{w} : \mathcal{T} \to \mathbb{R}^{n_w}$ is the unknown stochastic or bounded signal (e.g., process and output noise). The matrices $\mathbf{H}(\mathfrak{q})$, $\mathbf{L}(\mathfrak{q})$, $\mathbf{F}(\mathfrak{q})$, and $\mathbf{W}(\mathfrak{q})$ are polynomial matrices in $\mathfrak{q}$. The nonlinearity is characterized by $\boldsymbol{\eta}(\boldsymbol{\xi}, \mathbf{z})$, and without loss of generality, we assume that the nonlinear function $\boldsymbol{\eta}(\cdot)$ has no affine terms around the origin. The DAE is a rich modeling framework in the sense that it can encompass LTI systems and a class of nonlinear systems. For instance, consider the state-space representation

$$\begin{aligned} \mathfrak{q}\mathbf{x} &= \mathbf{A}\mathbf{x} + \mathbf{B}_u\mathbf{u} + \mathbf{B}_f\mathbf{f} + \mathbf{B}_d\mathbf{d} + \mathbf{B}_w\mathbf{w}, \\ \mathbf{y} &= \mathbf{C}\mathbf{x} + \mathbf{D}_u\mathbf{u} + \mathbf{D}_f\mathbf{f} + \mathbf{D}_d\mathbf{d} + \mathbf{D}_w\mathbf{w}, \end{aligned} \tag{3.}$$

in which the system state is represented by $\mathbf{x} : \mathcal{T} \to \mathbb{R}^{n_x}$. The system matrices $(\mathbf{A}, [\mathbf{B}_u, \mathbf{B}_f, \mathbf{B}_d, \mathbf{B}_w], \mathbf{C}, [\mathbf{D}_u, \mathbf{D}_f, \mathbf{D}_d, \mathbf{D}_w])$ form a state-space realization, which can be readily converted to transfer functions as well (104, 105), yielding

$$\mathbf{y} = \mathbf{G}_u(\mathfrak{q})\mathbf{u} + \mathbf{G}_f(\mathfrak{q})\mathbf{f} + \mathbf{G}_d(\mathfrak{q})\mathbf{d} + \mathbf{G}_w(\mathfrak{q})\mathbf{w}, \quad \mathbf{x}(0) = 0, \tag{4.}$$

where $\mathbf{G}_u(\mathfrak{q}) \in \mathbb{R}^{n_y \times n_u}$, $\mathbf{G}_f(\mathfrak{q}) \in \mathbb{R}^{n_y \times n_f}$, $\mathbf{G}_d(\mathfrak{q}) \in \mathbb{R}^{n_y \times n_d}$, and $\mathbf{G}_w(\mathfrak{q}) \in \mathbb{R}^{n_y \times n_w}$ are transfer function (i.e., rational) matrices in $\mathfrak{q}$. The conversion from state space in Equation 3 to DAE in Equation 2 is obtained by defining the matrices

$$\mathbf{H}(\mathfrak{q}) := \begin{bmatrix} \mathbf{A} - \mathfrak{q}\mathbf{I} & \mathbf{B}_d \\ \mathbf{C} & \mathbf{D}_d \end{bmatrix}, \quad \boldsymbol{\xi} := \begin{bmatrix} \mathbf{x} \\ \mathbf{d} \end{bmatrix}, \quad \mathbf{L}(\mathfrak{q}) := \begin{bmatrix} \mathbf{B}_u & \mathbf{0} \\ \mathbf{D}_u & -\mathbf{I} \end{bmatrix}, \quad \mathbf{z} := \begin{bmatrix} \mathbf{u} \\ \mathbf{y} \end{bmatrix},$$

$$\mathbf{F}(\mathfrak{q}) := \begin{bmatrix} \mathbf{B}_f \\ \mathbf{D}_f \end{bmatrix}, \qquad \mathbf{W}(\mathfrak{q}) := \begin{bmatrix} \mathbf{B}_w \\ \mathbf{D}_w \end{bmatrix}.$$

The nonlinear extension is straightforward, for which we direct readers to Reference 106.

2.3. Residual Generation

Residual signals are typically generated using either hardware redundancy or analytical redundancy. While hardware redundancy is commonly recommended in safety-critical applications (1), this review focuses on the use of analytical redundancy to generate residuals. An example of analytical redundancy is a mathematical model describing the system under consideration. Taking the difference between the actual system output and the model output readily yields a residual signal

for diagnosis. However, in many cases, such as in the presence of the external disturbance signal $\mathbf{d}$, satisfying the conditions in Equations 1a and 1b requires a more involved residual generation process.

Consider the noise-free linear system following the DAE model in Equation 2 by setting $\boldsymbol{\eta}(\boldsymbol{\xi}, \mathbf{z}) = \mathbf{0}$ and $\mathbf{w} = \mathbf{0}$. In this formulation, the aim is to exclude the contribution of the signal $\boldsymbol{\xi}$ from the residual. We begin by formally defining the healthy behavioral sets $\mathcal{B}_{\text{healthy}}$ (visualized in **Figure 2**) as the sets of available measurement signal $\mathbf{z}$ consistent with specific properties dictated by the model. The healthy and faulty behaviors of a linear deterministic system [$\boldsymbol{\eta}(\boldsymbol{\xi}, \mathbf{z}) = \mathbf{0}$, $\mathbf{w} = \mathbf{0}$] are described as

$$\mathcal{B}_{\text{healthy}} := \{\mathbf{z} \mid \exists \boldsymbol{\xi} \,:\, \mathbf{H}(\mathfrak{q})\boldsymbol{\xi} + \mathbf{L}(\mathfrak{q})\mathbf{z} = \mathbf{0}\},$$

$$\mathcal{B}_{\text{faulty}} := \{\mathbf{z} \mid \exists \mathbf{f} \not\equiv \mathbf{0},\, \boldsymbol{\xi} \,:\, \mathbf{H}(\mathfrak{q})\boldsymbol{\xi} + \mathbf{L}(\mathfrak{q})\mathbf{z} + \mathbf{F}(\mathfrak{q})\mathbf{f} = \mathbf{0}\},$$

which reflects a similar notion introduced in the behavioral approach theory (107).

We are now ready to parameterize the residual generator (diagnosis filter block in **Figure 1**) for a given model in Equation 2. Let $\mathbf{N}_H(\mathfrak{q})$ be the polynomial matrix whose rows span the left null space of $\mathbf{H}(q)$, forming a minimal polynomial basis (11). The healthy behavior set $\mathcal{B}_{\text{healthy}}$ can then be equivalently characterized by

$$\mathcal{B}_{\text{healthy}} := \{\mathbf{z} \mid \mathbf{N}_H(\mathfrak{q})\mathbf{L}(\mathfrak{q})\mathbf{z} = \mathbf{0}\},$$

allowing us to generate the residual signal using polynomial matrix equations. Left-multiplying Equation 2 by $\mathbf{N}_H(\mathfrak{q})$ yields the residual signal

$$\mathbf{r} = \mathbf{N}_H(\mathfrak{q})\mathbf{L}(\mathfrak{q})\mathbf{z} = -\mathbf{N}_H(\mathfrak{q})\mathbf{F}(\mathfrak{q})\mathbf{f},$$

provided that $\mathbf{N}_H(\mathfrak{q})\mathbf{F}(\mathfrak{q}) \not\equiv \mathbf{0}$. This parameterization effectively enforces the decoupling and sensitivity conditions described in Equations 1a and 1b, respectively. To ensure the filter is realizable, an arbitrary stable polynomial $a(\mathfrak{q})$ of sufficient order is introduced,[1] leading to the residual generator

$$\mathbf{r} = \mathbf{R}(\mathfrak{q})\mathbf{z} = a^{-1}(\mathfrak{q})\mathbf{N}_H(\mathfrak{q})\mathbf{L}(\mathfrak{q})\mathbf{z}, \qquad 5.$$

which serves as a baseline for dealing with different classes of fault diagnosis. This practical introduction of poles creates transient behavior in the residual, even in the absence of faults. Therefore, the decoupling criterion of Equation 1a holds only asymptotically. That is, the residual satisfies the following two properties: $\lim_{t\to\infty} \mathbf{r}(t) = \mathbf{0}$ for all $\mathbf{z} \in \mathcal{B}_{\text{healthy}}$, and $\mathbf{r}(t) \not\equiv \mathbf{0}$ for all $\mathbf{z} \in \mathcal{B}_{\text{faulty}}$. Such a linear filter falls into the category of null-space-based residual generators that decouple the contributions of internal states and possibly natural disturbances (105). For the more restricted class of state-space systems, the filter in Equation 5 is a kernel representation of the system (34, 108). The kernel-representation literature typically parameterizes the filter via state-space or transfer-function approaches. For detailed presentations, we direct readers to References 35, 104, and 105.

Another method to generate residuals is through observers. Unknown input observers (UIOs) were initially proposed for unbiased state estimation in the presence of inputs that are not directly measurable (109). This methodology was extended for residual generation in fault diagnosis

Kernel representation: a formulation that expresses governing equations in a form where both independent (inputs) and dependent (outputs) signals appear as arguments of a zero equation

Unknown input observer (UIO): a system state observer whose estimation error is independent of a subset of inputs, typically unmeasurable ones such as disturbances and input noises

[1]Once $\mathbf{N}_H(\mathfrak{q})$ has been designed and its degree is known, $a(\mathfrak{q})$ can be selected as a stable monic polynomial of equal or higher degree, with poles chosen depending on, e.g., the dominant frequency characteristics of the signals involved.

schemes (32). For brevity, we keep the noise-free assumption. Then, an observer-based residual generator for the system in Equation 3 takes the form

$$\begin{aligned} \mathfrak{q}\hat{\mathbf{x}} &= \mathbf{A}_{\mathrm{UIO}}\hat{\mathbf{x}} + (\mathbf{TB}_u - \mathbf{KD}_u)\mathbf{u} + \mathbf{Ky}, \\ \mathbf{r} &= \mathbf{C}_{\mathrm{UIO}}\hat{\mathbf{x}} + \mathbf{D}_{\mathrm{UIO}}(\mathbf{y} - \mathbf{D}_u\mathbf{u}), \end{aligned} \quad \Longrightarrow \quad \begin{aligned} \mathfrak{q}\mathbf{e} &= \mathbf{A}_{\mathrm{UIO}}\mathbf{e} + (\mathbf{KD}_f - \mathbf{TB}_f)\mathbf{f}, \\ \mathbf{r} &= \mathbf{C}_{\mathrm{UIO}}\mathbf{e} + \mathbf{D}_{\mathrm{UIO}}\mathbf{D}_f\mathbf{f}, \end{aligned} \tag{6.}$$

where $\mathbf{e} := \hat{\mathbf{x}} - \mathbf{Tx}$ denotes the estimation error, and the design parameters $(\mathbf{T}, \mathbf{K}, \mathbf{A}_{\mathrm{UIO}}, \mathbf{C}_{\mathrm{UIO}}, \mathbf{D}_{\mathrm{UIO}})$ are subject to the following matrix equality constraint:

$$\begin{bmatrix} \mathbf{T} & -\mathbf{K} \\ \mathbf{0} & \mathbf{D}_{\mathrm{UIO}} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{B}_d \\ \mathbf{C} & \mathbf{D}_d \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{\mathrm{UIO}}\mathbf{T} & \mathbf{0} \\ -\mathbf{C}_{\mathrm{UIO}}\mathbf{T} & \mathbf{0} \end{bmatrix}, \text{ and } \mathbf{A}_{\mathrm{UIO}} \text{ is Schur/Hurwitz stable.} \tag{7.}$$

The necessary and sufficient conditions for the existence of a solution to Equation 7, assuming $\mathbf{B}_d$ is full column rank without loss of generality, are given by the following (32): rank $\mathbf{D}_{\mathrm{UIO}}\mathbf{C}$ is equal to rank $\mathbf{B}_d$, and $(\bar{\mathbf{A}}, \mathbf{D}_{\mathrm{UIO}}\mathbf{C})$ is a detectable pair, where $\bar{\mathbf{A}} := \mathbf{A} - \mathbf{B}_d(\mathbf{D}_{\mathrm{UIO}}\mathbf{CB}_d)^{\dagger}\mathbf{D}_{\mathrm{UIO}}\mathbf{CA}$. A residual generator satisfying Equation 7 simultaneously decouples both $\mathbf{d}$ and $\mathbf{x}$ in the same way that the null-space-based filter in Equation 5 does. Ding et al. (28) demonstrated the equivalence between the observer and transfer-function parameterizations.

Finally, residual generators can also be built by exploiting the left null space of the extended observability matrix, which can be fully parameterized by the pair $(\mathbf{A}, \mathbf{C})$. This approach is the so-called parity relation (10, 36). A classical result from Patton & Chen (110) shows that the parity relation design is equivalent to the use of a deadbeat observer. The simplicity of the design renders parity-based filters attractive for fault diagnosis applications, as further detailed in the next section.

Schur stable: describes a matrix whose eigenvalues lie in the open unit circle of the complex plane

Hurwitz stable: describes a matrix whose eigenvalues lie in the open left half-plane of the complex plane

Parity relation: a residual generation method that exploits temporal redundancy between inputs and outputs to check the consistency of the system's mathematical equations with the measurements over a time window

3. DETECTION

In this section, we present the problem of fault detection and review existing formulations. We also compare different detection schemes from a design perspective.

Fault detection problems are binary classification problems in which the decision-making system determines whether the operational state is faulty or healthy. According to the residual properties, a nonzero value of $\mathbf{r}$ is sufficient to indicate a fault. Therefore, the diagnostic residual signal does not necessarily need to be multivariate for the purpose of detection. In fact, any linear combination of rows in $\mathbf{N}_H(\mathfrak{q})$, leading to a scalar residual, can serve as a basis for designing the detection filter. To take this into consideration, the filter transfer function $\mathbf{R}(\mathfrak{q})$ in Equation 5 is modified to

$$\mathbf{R}(\mathfrak{q}) = a^{-1}(\mathfrak{q})\boldsymbol{\gamma}(\mathfrak{q})\mathbf{N}_H(\mathfrak{q})\mathbf{L}(\mathfrak{q}) = a^{-1}(\mathfrak{q})\mathbf{N}(\mathfrak{q})\mathbf{L}(\mathfrak{q}), \tag{8.}$$

where $\boldsymbol{\gamma}(\mathfrak{q})$ is a polynomial row vector representing a linear combination of the rows of $\mathbf{N}_H(\mathfrak{q})$. Defining $\mathbf{N}(\mathfrak{q}) := \boldsymbol{\gamma}(\mathfrak{q})\mathbf{N}_H(\mathfrak{q})$ implicitly incorporates the role of $\boldsymbol{\gamma}(\mathfrak{q})$ into the design while simultaneously resulting in a scalar residual signal. Hence, filter design amounts to designing the row polynomial vector $\mathbf{N}(\mathfrak{q})$ and the realization poles $a(\mathfrak{q})$. However, in this review, we keep $a(\mathfrak{q})$ fixed and focus on determining an optimal $\mathbf{N}(\mathfrak{q})$ based on the given filter performance criterion. The filter design, in its simplest form, can be expressed as a polynomial matrix inequality:

$$\begin{aligned} \mathbf{N}(\mathfrak{q})\mathbf{H}(\mathfrak{q}) &= \mathbf{0}, \\ \mathbf{N}(\mathfrak{q})\mathbf{F}(\mathfrak{q}) &\neq \mathbf{0}. \end{aligned} \tag{9.}$$

Whether this problem can be solved or not is assessed via the property of fault detectability for linear DAE systems, formally defined as follows.

Definition (fault detectability). A nonzero fault signal $\mathbf{f}$ is detectable if for any $\mathbf{z}$ and $\boldsymbol{\xi}$ satisfying $\mathbf{H}(\mathfrak{q})\boldsymbol{\xi} + \mathbf{L}(\mathfrak{q})\mathbf{z} + \mathbf{F}(\mathfrak{q})\mathbf{f} = \mathbf{0}$, $\mathbf{z}$ lies outside the healthy behavior set—i.e.,

$\mathbf{z} \notin \mathcal{B}_{\text{healthy}}$. Moreover, a system described by Equation 2 is fault detectable if any fault $\mathbf{f}$ satisfying $\mathbf{F}(\mathfrak{q})\mathbf{f} \not\equiv \mathbf{0}$ is detectable.

A necessary and sufficient condition guaranteeing both the fault detectability and the feasibility problem in Equation 9 is provided by (38)

$$\text{normal rank} \quad \Big[\mathbf{H}(\mathfrak{q}) \ \mathbf{F}(q)\Big] > \text{normal rank} \quad \Big[\mathbf{H}(\mathfrak{q})\Big].$$

If we use the methods presented in the sidebar titled From Polynomial Matrix Algebra to Linear Algebra, then the problem in Equation 9 is equivalent, up to a scalar factor, to the following:

$$\begin{aligned} &\bar{\mathbf{N}}\bar{\bar{\mathbf{H}}}_p = \mathbf{0} && \text{(disturbance decoupling)},\\ &\|\bar{\mathbf{N}}\bar{\bar{\mathbf{F}}}_p\|_\infty \geq 1 && \text{(fault sensitivity)}, \end{aligned} \tag{10.}$$

where p must be chosen sufficiently large, so that $p+1$ is the degree of $\mathbf{N}(\mathfrak{q})$. Mohajerin Esfahani & Lygeros (17) reformulated this nonconvex problem as a series of linear programs. Here, we provide a simpler solution via the linear algebraic approach outlined in the sidebar titled From Polynomial Matrix Algebra to Linear Algebra. First, find $\bar{\mathbf{N}}_H$ as the left null space of $\bar{\bar{\mathbf{H}}}_p$. Next, take $\boldsymbol{\mu}$ as the leading left singular vector of $\bar{\mathbf{N}}_H\bar{\bar{\mathbf{F}}}_p$, which ensures that $\boldsymbol{\mu}^{\mathrm{T}}\bar{\mathbf{N}}_H\bar{\bar{\mathbf{F}}}_p \neq \mathbf{0}$, provided that the system is fault detectable and p is sufficiently large. Finally, rescale $\boldsymbol{\mu}$ so that $\|\boldsymbol{\mu}^{\mathrm{T}}\bar{\mathbf{N}}_H\bar{\bar{\mathbf{F}}}_p\|_\infty \leq 1$. By construction, $\bar{\boldsymbol{\gamma}} := \boldsymbol{\mu}^{\mathrm{T}}\bar{\mathbf{N}}_H$ satisfies the conditions in Equation 9.

Once the residual generator is designed, detection logic is applied in the diagnosis filter:

$$\begin{cases} J(r) \leq J_{\text{th}} \implies \text{no fault}, \\ J(r) > J_{\text{th}} \implies \text{fault occurrence}, \end{cases} \tag{11.}$$

where $J(r)$ denotes an evaluation function applied to the residual signal, and J_{th} is the detection threshold that must be determined during the training phase (18, 62, 112). In the ideal scenario (i.e., a noise-free linear system), J_{th} is set to zero, and any nonzero residual is interpreted as an indication of a fault.

Normal rank: the maximal rank of a polynomial matrix in $\mathfrak{q}$ over all possible values of $\mathfrak{q} \in \mathbb{C}$

Linear program: a convex optimization problem in which both the objective function and the constraints are affine

Leading left singular vector: the first column of the matrix $\mathbf{U}$ from the singular value decomposition $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{T}}$

3.1. Nonlinear and Robustification Approaches

The set of feasible residual generators, as solutions to Equation 10, results in detection filters that are effective under the noise-free linear assumption [i.e., $\mathbf{w} = \mathbf{0}$ and $\boldsymbol{\eta}(\boldsymbol{\xi}, \mathbf{z}) = \mathbf{0}$]. However, this assumption rarely holds in practical scenarios due to inevitable model imperfections. To address this, one may consider enhancing the robustness of the filter. We begin by identifying the sources of such imperfections, followed by proposing a mechanism to mitigate their impacts.

On the one hand, model imperfections may arise from various sources, including unmodeled dynamics, inherent model uncertainties (24), and natural noise components that cannot be entirely decoupled; in addition, nonlinearities can be treated as a form of imperfection within the class of linear diagnosis filters. On the other hand, the mechanism for filter robustification remains consistent, regardless of the specific cause of these imperfections. In the general DAE formulation given by Equation 2, these effects are accounted for by the terms $\mathbf{W}(\mathfrak{q})\mathbf{w} + \boldsymbol{\eta}(\boldsymbol{\xi}, \mathbf{z})$, which we refer to here as model mismatch. If we left-multiply Equation 2 by $a^{-1}(\mathfrak{q})\mathbf{N}(\mathfrak{q})$, the scalar residual r becomes

$$r = \frac{\mathbf{N}(\mathfrak{q})}{a(\mathfrak{q})}\mathbf{L}(\mathfrak{q})\mathbf{z} = \underbrace{-\frac{\mathbf{N}(\mathfrak{q})}{a(\mathfrak{q})}\mathbf{H}(\mathfrak{q})\boldsymbol{\xi}}_{=0} \underbrace{-\frac{\mathbf{N}(\mathfrak{q})}{a(\mathfrak{q})}\mathbf{F}(\mathfrak{q})\mathbf{f}}_{r_f} \underbrace{-\frac{\mathbf{N}(\mathfrak{q})}{a(\mathfrak{q})}\mathbf{W}(\mathfrak{q})\mathbf{w} - \frac{\mathbf{N}(\mathfrak{q})}{a(\mathfrak{q})}\boldsymbol{\eta}(\boldsymbol{\xi}, \mathbf{z})}_{r_{\text{MM}}}. \tag{12.}$$

FROM POLYNOMIAL MATRIX ALGEBRA TO LINEAR ALGEBRA

Representing polynomial matrices as certain types of real-(complex-)valued matrices is particularly useful for fault diagnosis filter design. First, it enables efficient methods for performing operations with them (111), such as products, inverses, and null spaces. Second, it further allows one to place polynomial matrix constraints and objectives in standard convex optimization methods.

Let $\mathbf{X}(\mathfrak{q})$ and $\mathbf{Y}(\mathfrak{q})$ be polynomial matrices in operator $\mathfrak{q}$ of sizes $n_\mathrm{r} \times n_\mathrm{c}$ and $n_\mathrm{rr} \times n_\mathrm{r}$, respectively. Then, we have

$$\mathbf{X}(\mathfrak{q}) := \sum_{i=0}^{p_X} \mathbf{X}_i \mathfrak{q}^i, \mathbf{Y}(\mathfrak{q}) := \sum_{i=0}^{p_Y} \mathbf{Y}_i \mathfrak{q}^i,$$

where $\mathbf{Y}_i \in \mathbb{R}^{n_\mathrm{rr} \times n_\mathrm{r}}$ and $\mathbf{X}_i \in \mathbb{R}^{n_\mathrm{r} \times n_\mathrm{c}}$ are constant matrices. We define the following representations, which are in a block row form and a block Toeplitz form, respectively:

$$\bar{\mathbf{Y}} := \begin{bmatrix} \mathbf{Y}_0 & \mathbf{Y}_1 & \cdots & \mathbf{Y}_{p_Y} \end{bmatrix} \in \mathbb{R}^{n_\mathrm{rr} \times (p_Y+1) n_\mathrm{r}},$$

$$\overline{\overline{\mathbf{X}}}_p := \begin{bmatrix} \mathbf{X}_0 & \mathbf{X}_1 & \cdots & \mathbf{X}_{p_X} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_0 & \mathbf{X}_1 & \cdots & \mathbf{X}_{p_X} & \mathbf{0} & \vdots \\ \vdots & & \ddots & \ddots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{X}_0 & \mathbf{X}_1 & \cdots & \mathbf{X}_{p_X} \end{bmatrix} \in \mathbb{R}^{p n_\mathrm{r} \times (p+p_X) n_\mathrm{c}},$$

where p determines the number of block rows of $\overline{\overline{\mathbf{X}}}_p$. The product of two polynomial matrices satisfies

$$\mathbf{Y}(\mathfrak{q})\mathbf{X}(\mathfrak{q}) = \bar{\mathbf{Y}} \overline{\overline{\mathbf{X}}}_{p_Y+1} \begin{bmatrix} \mathbf{I} & \mathfrak{q}\mathbf{I} & \mathfrak{q}^2\mathbf{I} & \cdots & \mathfrak{q}^{p_Y+p_X}\mathbf{I} \end{bmatrix}^\mathrm{T}.$$

For example, the polynomial matrix $\mathbf{X}(\mathfrak{q}) = \begin{bmatrix} 1-\mathfrak{q} \\ 2 \end{bmatrix}$ can be decomposed as

$$\mathbf{X}(\mathfrak{q}) = \underbrace{\begin{bmatrix} -1 \\ 0 \end{bmatrix}}_{\mathbf{X}_1} \mathfrak{q} + \underbrace{\begin{bmatrix} 1 \\ 2 \end{bmatrix}}_{\mathbf{X}_0} = \sum_{i=0}^{p_X=1} \mathbf{X}_i \mathfrak{q}^i.$$

Letting $p = 2$, the Toeplitz format of $\mathbf{X}(\mathfrak{q})$ is given by

$$\overline{\overline{\mathbf{X}}}_2 = \begin{bmatrix} 1 & -1 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 2 & 0 \end{bmatrix}.$$

Its left null space is spanned by the vector $[\,2\ -1\ 0\ 1\,]$, which gives that $\mathbf{N}(\mathfrak{q}) = [\,2\ -1+\mathfrak{q}\,]$ satisfies $\mathbf{N}(\mathfrak{q})\mathbf{X}(\mathfrak{q}) \equiv \mathbf{0}$. Additionally, computing a left inverse, i.e., solving $\mathbf{Y}(\mathfrak{q})\mathbf{X}(\mathfrak{q}) = 1$, is the same as finding $\bar{\mathbf{Y}}$ such that $\bar{\mathbf{Y}}\overline{\overline{\mathbf{X}}}_2 = [\,1\ 0\ 0\,]$, which gives the obvious solution $\bar{\mathbf{Y}} = [\,0\ 0.5\ 0\ 0\,]$, i.e., $\mathbf{Y}(\mathfrak{q}) = [\,0\ 0.5\,]$.

Ideally, in null-space-based filters, only the fault contribution r_f should be present in the residual signal; however, a simultaneous contribution from model mismatch r_MM degrades the filter performance. We hereafter refer to r_MM as the "mismatch signature." Therefore, the objective of residual generation is twofold: (*a*) to maximize fault sensitivity and (*b*) to minimize the impact

of model mismatch. These objectives are inherently conflicting—i.e., minimizing the influence of model mismatch often leads to reduced fault sensitivity and vice versa. This interplay is commonly referred to as the FAR/MAR (false-alarm rate/missed-alarm rate) trade-off, which can be adjusted by appropriately setting the detection threshold in Equation 11 (17, 67).

Quadratic program: a convex optimization problem in which the objective function is quadratic and subject to affine constraints

To robustify the residual generator against model mismatch, it is necessary to quantify the contribution of model mismatch to the residual signal. To achieve this, Mohajerin Esfahani & Lygeros (17) proposed a semisupervised approach that integrates available model information with fault-free data obtained from the real system (24), a high-fidelity simulator (21), or a digital twin (113). The underlying intuition is straightforward: The data may compensate for the missing knowledge of the system. As a result, the extent to which the acquired data are representative of the model mismatch can directly affect the filter performance.

The terms $\mathbf{W}(\mathfrak{q})$ and $\boldsymbol{\eta}(\boldsymbol{\xi}, \mathbf{z})$ may or may not be explicitly known. In either case, the primary assumption is that r_{MM} can be computed over a finite time window under the fault-free condition. Depending on the model information available in the training phase, there are two ways to calculate the mismatch signature. Given the real high-fidelity system trajectory $\mathbf{z}^{\mathrm{real}}$ and the linear dynamics matrix $\mathbf{H}(\mathfrak{q})$, we have

$$r_{\mathrm{MM}} = \begin{cases} -a^{-1}(\mathfrak{q})\mathbf{N}(\mathfrak{q})\big(\mathbf{W}(\mathfrak{q})\mathbf{w} + \boldsymbol{\eta}(\boldsymbol{\xi}, \mathbf{z}^{\mathrm{real}})\big) & \boldsymbol{\xi}, \mathbf{w}, \mathbf{W}(\mathfrak{q}), \boldsymbol{\eta}(\cdot), \\ a^{-1}(\mathfrak{q})\mathbf{N}(\mathfrak{q})\mathbf{L}(\mathfrak{q})\big(\mathbf{z}^{\mathrm{real}} - \mathbf{z}^{\mathrm{lin}}\big) & \underbrace{\mathbf{z}^{\mathrm{lin}}, \mathbf{L}(\mathfrak{q}),}_{\text{information}} \end{cases} \tag{13.}$$

where $\mathbf{z}^{\mathrm{lin}}$ corresponds to the linear model. In the first case, full knowledge of the real system is assumed to be available during training (17), which is rarely practical in real-world applications. In contrast, the second case relies only on measurable signals, namely $\mathbf{z}^{\mathrm{real}}$ and $\mathbf{z}^{\mathrm{lin}}$, thus enabling broader applicability. For instance, $\mathbf{z}$ typically denotes an augmented vector of input–output signals, which, in a special case, reduces to the output difference—i.e., $\mathbf{z}^{\mathrm{real}} - \mathbf{z}^{\mathrm{lin}} \equiv \mathbf{y}^{\mathrm{real}} - \mathbf{y}^{\mathrm{lin}}$, since both systems are driven by the same input signal $\mathbf{u}$ (21). The goal is to suppress the contribution of r_{MM} to the residual signal, which can be achieved by minimizing the $\mathcal{L}_2$-norm of r_{MM} recorded over a finite time interval $[0, T] \in \mathcal{T}$. Accordingly, the filter robustification is translated into an optimization framework:

$$\begin{aligned} \underset{\bar{\mathbf{N}}}{\text{minimize}} \quad & \|r_{\mathrm{MM}}\|_{\mathcal{L}_2}^2 \\ \text{subject to} \quad & \bar{\mathbf{N}}\bar{\mathbf{H}}_p = \mathbf{0}, \\ & \left\|\bar{\mathbf{N}}\bar{\mathbf{F}}_p\right\|_\infty \geq 1, \end{aligned} \tag{14.}$$

where the optimization problem can be presented and solved as a sequence of quadratic programs with linear constraints (17, 21, 24). The key point in deriving the quadratic program reformulation is to express the cost function as a quadratic form in terms of the filter parameter $\bar{\mathbf{N}}$—i.e., $\|r_{\mathrm{MM}}\|_{\mathcal{L}_2}^2 = \bar{\mathbf{N}}\mathbf{Q}_{\mathrm{MM}}\bar{\mathbf{N}}^{\mathrm{T}}$, in which the positive semidefinite $\mathbf{Q}_{\mathrm{MM}}$ is referred to as the mismatch signature matrix (17). In continuous time, a set of basis functions is proposed to approximate $\mathbf{Q}_{\mathrm{MM}}$, whereas in discrete time, the computation is less involved as the basis functions reduce to simple time-shift bases (21).

The optimization can be extended to accommodate multiple sets of signature signals resulting from separate experiments. Let N_{exp} stand for the number of experiments. Then, the optimization

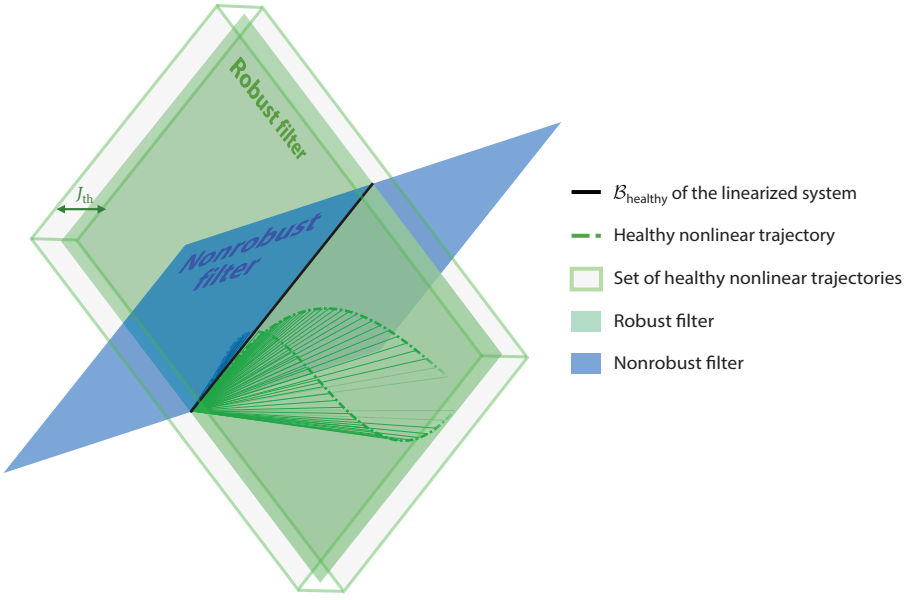

Figure 3

Geometry of robust linear filters. The black solid line is the healthy behavior $\mathcal{B}_{\text{healthy}}$ of the linearized dynamic system, the dashed green lines are two sample trajectories of the healthy nonlinear system, and the green and blue hyperplanes represent two linear filters. Both hyperplanes contain $\mathcal{B}_{\text{healthy}}$; the green hyperplane is the robust one containing the healthy nonlinear system trajectories with a smaller threshold J_{th}.

problem in Equation 14 can be modified to

$$\begin{aligned} &\underset{\bar{\mathbf{N}}}{\text{minimize}} && \ell(\bar{\mathbf{N}}) \\ &\text{subject to} && \bar{\mathbf{N}}\bar{\bar{\mathbf{H}}}_p = \mathbf{0}, \\ &&& \left\|\bar{\mathbf{N}}\bar{\bar{\mathbf{F}}}_p\right\|_\infty \geq 1, \end{aligned} \qquad \text{with} \quad \ell := \begin{cases} \bar{\mathbf{N}}\left(\dfrac{1}{N_{\text{exp}}}\displaystyle\sum_{i=1}^{N_{\text{exp}}} \mathbf{Q}_{\text{MM}}^{(i)}\right)\bar{\mathbf{N}}^{\text{T}} & \text{(average cost)}, \\ \displaystyle\max_{i\leq N_{\text{exp}}} \bar{\mathbf{N}}\mathbf{Q}_{\text{MM}}^{(i)}\bar{\mathbf{N}}^{\text{T}} & \text{(worst-case cost)}. \end{cases}$$

Although the constraint $\left\|\bar{\mathbf{N}}\bar{\bar{\mathbf{F}}}_p\right\|_\infty \geq 1$ is nonconvex, it can be reformulated as a finite set of linear constraints, whose number of constraints scales linearly with the filter degree (17). Building on this, the proposed optimization framework offers three key advantages: (*a*) Convexity ensures computational tractability and the ability to use commercial solvers (e.g., MOSEK, Gurobi, or CPLEX); (*b*) the formulation scales efficiently, making it suitable for high-dimensional complex systems; and (*c*) the underlying convex structure (i.e., having feasible sets as a union of convex sets) allows the use of tools from the scenario approach literature (e.g., 114) in order to provide statistical guarantees for unseen mismatch signature matrices, which represent plausible future experimental settings (17, 115). **Figure 3** illustrates the robustified linear filters obtained through this technique, together with a threshold defined by Equation 11.

Van der Ploeg et al. (24) presented a more specific formulation tailored to structural model uncertainties, demonstrating that the framework in Equation 14 can be adapted to various scenarios given the available model information. It is important to emphasize that the filter design assumes no prior knowledge of the fault signal. However, in certain applications, such as power grids (21) and wind turbines (116, 117), partial information about fault characteristics may be available. In such cases, the filter performance can be enhanced by incorporating this fault information into

the filter synthesis process, specifically within the optimization framework of Equation 14. Dong et al. (71) exploited the frequency content of faults by optimizing a mixed $\mathcal{H}_-/\mathcal{H}_2$ performance index over certain frequency ranges. The results given by Pan et al. (23) assume that the fault signals can be represented as a linear combination of predefined basis functions. Regarding these parameterizations, the fault feasible set $\mathcal{F}$ is summarized as

$$\mathcal{F} := \begin{cases} \{\mathbf{f} \in \mathbb{R}^{n_f} \mid \mathbf{f} = \int_\Theta \mathrm{FT}\{\mathbf{f}\}\, \mathrm{e}^{\mathrm{j}\theta}\, \mathrm{d}\theta,\ \Theta \subset [\pi, -\pi]\} & \text{(frequency content)}, \\ \{\mathbf{f} \in \mathbb{R}^{n_f} \mid \mathbf{f} = \sum_{i=1}^{n_\mathrm{b}} \alpha_i \phi_\mathrm{b}^{(i)},\ \alpha \in \mathbb{R}, \text{basis } \phi_\mathrm{b}^{(i)}\} & \text{(basis functions)}. \end{cases}$$

Hence, addressing the underlying diagnosis problem in an optimization framework provides the flexibility to incorporate additional prior information through new constraints or an alternative parameterization or by modifying the objective function.

For the specific case of linear time-varying systems, the detection threshold is robustified using a projection-based filter (74). More recently, distributionally robust techniques have been proposed to enhance filter performance, particularly when the noise is assumed to be drawn from an uncertain but structured family of distributions (18, 67). At the same time, set-based methods have received more attention, owing to recent advances in reachability analysis tools such as zonotopic methods and the suitability of such methods for bounded cyberattack estimation (37). Mu et al. (118) recently performed a comparison of several methods. Similar to the case of fault isolation, set-based methods have the advantage of being applicable to a broad class of nonlinear systems but suffer from higher online computational cost and difficulty with scalability.

3.2. Model-Free Approaches

Model-free approaches utilize historical data instead of models to solve diagnosis problems. These methods can generally be categorized into unsupervised and supervised approaches. In the unsupervised category, training data are used primarily either to learn statistical models (30, 95) or to extract latent statistical structures for performing classical variational analysis and hypothesis testing. Techniques such as principal component analysis (76), partial least squares (77), canonical variate analysis (78), and their variants are widely employed for anomaly detection (for a survey, see 119). In the supervised category, data-driven solutions are predominantly based on system identification methods, which assume that the observed data belong to a known system class. Conventional two-stage, or indirect, designs typically involve first identifying the system model parameters and then applying model-based diagnostic methods. Subspace identification methods are well-established techniques for linear system modeling, with notable variants including N4SID (numerical algorithms for subspace state-space system identification) (120), MOESP (multivariable output-error state space) (121), and PBSID (predictor-based subspace identification) (122). Compared with the standard prediction-error method (123) and maximum likelihood estimation (124), subspace identification methods offer advantages such as avoiding nonconvex optimization and providing better numerical stability. Direct designs, on the other hand, leverage subspace identification methods to construct observers or diagnosis filters directly from input–output data, bypassing explicit system modeling. These approaches have been successfully applied in several fault detection studies (80, 125).

When dealing only with sampled data, it is more natural to frame the problem in discrete time. For brevity, we adopt the following innovation form as the data-generating system, representing a minimal state-space realization of a finite-dimensional LTI system (123):

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{A}\mathbf{x}_t + \mathbf{B}_u\mathbf{u}_t + \mathbf{B}_f\mathbf{f}_t + \mathbf{K}\mathbf{e}_t, \\ \mathbf{y}_t &= \mathbf{C}\mathbf{x}_t + \mathbf{D}_u\mathbf{u}_t + \mathbf{D}_f\mathbf{f}_t + \mathbf{e}_t. \end{aligned} \qquad 15.$$

Persistency of excitation: a condition on the input signal to ensure consistent system identification; in discrete time, an input $\bar{\mathbf{u}}_{[0,N-1]}$ has persistency of excitation of order L if $\hat{\mathbf{U}}_{0,L}$ has full row rank

Process and measurement noises are modeled through the zero-mean innovation signal $\mathbf{e}_t \in \mathbb{R}^{n_y}$ with a steady-state Kalman gain $\mathbf{K} \in \mathbb{R}^{n_x \times n_y}$ (126). In data-driven settings, the Toeplitz and Hankel data structures are instrumental in problem formulation and are introduced below. Let N represent the number of collected samples in the training dataset. Define L as the depth of the Hankel matrix constructed from N data points of an arbitrary signal $\mathbf{s}_t$, as follows:

$$\hat{\mathbf{S}}_{t,L} := \begin{bmatrix} \mathbf{s}_t & \mathbf{s}_{t+1} & \cdots & \mathbf{s}_{t+N-L} \\ \mathbf{s}_{t+1} & \mathbf{s}_{t+2} & \cdots & \mathbf{s}_{t+N-L+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{s}_{t+L-1} & \mathbf{s}_{t+L} & \cdots & \mathbf{s}_{t+N-1} \end{bmatrix}.$$

The data collected over the time interval $[t_1, t_2]$ are stacked into a column vector $\bar{\mathbf{s}}_{[t_1,t_2]} := [\mathbf{s}_{t_1}^\top \ \mathbf{s}_{t_1+1}^\top \cdots \mathbf{s}_{t_2}^\top]^\top$. Similarly, the state trajectory over this interval is gathered in the matrix $\mathbf{X}_{[t_1,t_2]} := [\mathbf{x}_{t_1} \ \mathbf{x}_{t_1+1} \cdots \mathbf{x}_{t_2}]$. The data equation corresponding to the system in Equation 15 for a dataset of size N over a sliding time window of length L follows:

$$\begin{bmatrix} \hat{\mathbf{U}}_{t,L} \\ \hat{\mathbf{Y}}_{t,L} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{T}_L^u & \mathbf{O}_L \end{bmatrix}}_{\mathbf{G}_L} \begin{bmatrix} \hat{\mathbf{U}}_{t,L} \\ \mathbf{X}_{[t,t+N-L]} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{T}_L^f \hat{\mathbf{F}}_{k,L} + \mathbf{T}_L^e \hat{\mathbf{E}}_{k,L} \end{bmatrix}, \tag{16.}$$

in which $\mathbf{O}_L$ represents the extended observability and $\mathbf{T}_L^\star$ is the lower triangular block-Toeplitz matrix of Markov parameters (impulse responses) structured as

$$\mathbf{O}_L = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{L-1} \end{bmatrix}, \quad \mathbf{T}_L^\star = \begin{bmatrix} \mathbf{M}_0^\star & 0 & \cdots & 0 \\ \mathbf{M}_1^\star & \mathbf{M}_0^\star & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ \mathbf{M}_{L-1}^\star & \mathbf{M}_{L-2}^\star & \cdots & \mathbf{M}_0^\star \end{bmatrix},$$

with $\star$ representing u, f, or e. The corresponding Markov parameters are given by

$$\mathbf{M}_i^u = \begin{cases} \mathbf{D}_u & i = 0 \\ \mathbf{CA}^{i-1}\mathbf{B}_u & i > 0 \end{cases}, \ \mathbf{M}_i^f = \begin{cases} \mathbf{D}_f & i = 0 \\ \mathbf{CA}^{i-1}\mathbf{B}_f & i > 0 \end{cases}, \ \mathbf{M}_i^e = \begin{cases} \mathbf{I}_{n_y} & i = 0 \\ \mathbf{CA}^{i-1}\mathbf{K} & i > 0 \end{cases}.$$

Define $\mathcal{K}_L$ such that its rows span the left null space of $\mathbf{G}_L$, implying $\mathcal{K}_L \mathbf{G}_L = \mathbf{0}$. For a sufficiently large value of L, the existence of a solution for $\mathcal{K}_L$ is guaranteed by the Cayley–Hamilton theorem, exploiting the inherent temporal redundancy of LTI systems (10). One should note that $\mathbf{G}_L$ and $\mathcal{K}_L$ respectively parameterize the hyperplane (green) and the normal subspace (blue) illustrated in **Figure 3** given a fault- and noise-free condition. In this context, data-driven null-space-based filters aim to estimate $\mathcal{K}_L$ directly from the recorded data (20, 75), whereas conventional subspace identification methods focus on identifying $\mathbf{G}_L$ (127, 128).

Regardless of the objective, two fundamental assumptions are typically required in data-driven designs: (*a*) The training dataset $\mathcal{D}_{\text{healthy}} := \{\mathbf{u}_i, \mathbf{y}_i\}_{i=0}^{N-1}$ is fault-free and (*b*) the data are recorded such that $\operatorname{rank} \begin{bmatrix} \mathbf{X}_{[0,N-L]} \\ \hat{\mathbf{U}}_{0,L} \end{bmatrix} = n_x + Ln_u$ (a condition that can be ensured by designing an input signal with a persistency of excitation of sufficient order; see 103). According to Willems's fundamental lemma (103), any healthy noise-free trajectory of the system with length L satisfies

$$\forall \bar{\mathbf{u}}_{[t,t+L-1]}, \mathbf{x}_t, \quad \mathcal{K}_L \begin{bmatrix} \bar{\mathbf{u}}_{[t,t+L-1]} \\ \bar{\mathbf{y}}_{[t,t+L-1]} \end{bmatrix} = 0, \text{ and } \mathcal{K}_L \begin{bmatrix} \hat{\mathbf{U}}_{t,L} \\ \hat{\mathbf{Y}}_{t,L} \end{bmatrix} = 0, \tag{17.}$$

which can be seen as the data-based kernel representation (108). The corresponding residual generator is designed as

$$\mathbf{r} = \underbrace{\begin{bmatrix}\mathcal{K}_L^u & \mathcal{K}_L^y\end{bmatrix}}_{\mathcal{K}_L}\begin{bmatrix}\bar{\mathbf{u}}_{[t,t+L-1]}\\ \bar{\mathbf{y}}_{[t,t+L-1]}\end{bmatrix} = \mathcal{K}_L^y(\mathbf{T}_L^f\bar{\mathbf{f}}_{[t,t+L-1]} + \mathbf{T}_L^e\bar{\mathbf{e}}_{[t,t+L-1]}), \tag{18.}$$

where $\mathcal{K}_L^y\mathbf{O}_L = \mathbf{0}$ and $\mathcal{K}_L^u = -\mathcal{K}_L^y\mathbf{T}_L^u$.

Parity space: the left null space of the system's extended observability matrix of a given order, spanned by parity vectors

The proposed residual generator is a finite impulse response filter with horizon L as a design parameter. In the diagnosis literature, $\mathcal{K}_L^y$ characterizes the so-called parity space (36, 112). As proved by Chen & Patton (109), Patton & Chen (110), and Ding (104), for any parity-relation-based residual generator, one can derive an observer-based diagnosis filter with identical dynamics. Thus, we suggest using the parity space at the design stage for its simplicity and then realizing the solution in the observer form for its recursive implementation, which is online friendly. From Equation 18, it follows that this approach provides decoupling only regarding the system states and not the innovation (noise) signal. As a result, ensuring satisfactory fault detection performance requires the detection threshold to be robust against the noise level (17, 18, 71).

Chen et al. (96) extended the kernel-representation-based design to a class of Lipschitz nonlinear systems, where $\mathcal{K}_L$ is a nonlinear operator. Krishnan & Pasqualetti (79) characterized the fault detectability in a data-driven sense based on the observed data informativity. Another class of data-driven diagnosis schemes extends the UIO design (Equation 6) to data-driven settings by additionally assuming access to the state trajectory $\mathbf{X}_{[0,N-L]}$ (86, 129), unlike the data-based kernel-representation solutions (Equation 17). This assumption is justified in applications where the state information can be obtained or estimated reliably (130).

4. ISOLATION

This section discusses the isolation task as a natural extension to a multi-classification problem. Assuming $n_f > 1$, the objective is to identify the specific source responsible for triggering the fault alarm.

Whereas a scalar residual signal is sufficient for the detection, isolating the fault generally necessitates multiple residuals or a residual vector. Designing a set of dedicated filters, each tailored to be sensitive to a subset of faults while remaining insensitive to others, leads to what is known as the bank-of-filters approach (2, 80). This can be implemented in the DAE framework by augmenting $\boldsymbol{\xi}$ with fault signals intended to be decoupled from the residual signal (38). In other words, to isolate the ith additive fault f_i from the remaining faults $\tilde{\mathbf{f}} := [\, f_1 \;\; f_2 \;\cdots\; f_{i-1} \;\; f_{i+1} \;\cdots\,]$, the system can be represented as

$$\begin{bmatrix}\mathbf{H}(\mathfrak{q}) & \widetilde{\mathbf{F}}(\mathfrak{q})\end{bmatrix}\begin{bmatrix}\boldsymbol{\xi}\\ \tilde{\mathbf{f}}\end{bmatrix} + \mathbf{L}(\mathfrak{q})\mathbf{z} + \mathbf{F}_i(\mathfrak{q})f_i + \mathbf{W}(\mathfrak{q})\mathbf{w} + \boldsymbol{\eta}(\boldsymbol{\xi},\mathbf{z}) = \mathbf{0},$$

where $\mathbf{F}_i(\mathfrak{q})$ is the ith column of $\mathbf{F}(\mathfrak{q})$, and the remaining columns are kept in $\widetilde{\mathbf{F}}(\mathfrak{q})$ with respect to $\tilde{\mathbf{f}}$. The rest of the design follows similar steps outlined in Section 3. The necessary and sufficient condition for the fault isolability of f_i is given by (38)

$$\text{normal rank}\begin{bmatrix}\mathbf{H}(\mathfrak{q}) & \widetilde{\mathbf{F}}(q) & \mathbf{F}_i(\mathfrak{q})\end{bmatrix} > \text{normal rank}\begin{bmatrix}\mathbf{H}(\mathfrak{q}) & \widetilde{\mathbf{F}}(q)\end{bmatrix}.$$

In this scheme, each filter generates a scalar residual signal that indicates the presence of the corresponding fault. A key idea in deriving the isolation scheme from existing approaches is to treat the fault signals that the filter should ignore as additional disturbances to be completely rejected (not to be confused with noise to be mitigated). In the null-space-based filters, the transfer

function $\mathbf{G}_d(\mathfrak{q})$ is built on the augmented polynomial matrix $\left[\mathbf{H}(\mathfrak{q})\ \widetilde{\mathbf{F}}(\mathfrak{q})\right]$ (35). Accordingly, the solvability constraint in Equation 7 must be adapted so that $(\mathbf{B}_d, \mathbf{D}_d)$ includes the corresponding columns of $(\mathbf{B}_f, \mathbf{D}_f)$ in observer-based approaches (32).

In a similar fashion, parity-relation-based designs leverage the multidimensionality of the generated residual to meet additional requirements posed by the isolation problem (80)—for instance, in the case of f_i,

$$\alpha_i \mathcal{K}_L^y \mathbf{T}_L^{\tilde{f}} = 0,\ \alpha_i \mathcal{K}_L^y \mathbf{T}_L^{f_i} \neq 0.$$

Here, α_i is a newly introduced design parameter taking a linear combination of rows in $\mathcal{K}_L^y$. The block Toeplitz matrices $\mathbf{T}_L^{\tilde{f}}$ and $\mathbf{T}_L^{f_i}$ are constructed based on the associated columns in $(\mathbf{B}_f, \mathbf{D}_f)$. The existence of a solution to the bank of filters is linked to the left invertibility of the fault subsystem, i.e., the system from $\mathbf{f}$ to $\mathbf{y}$ (6, 50). In the DAE framework, this is expressed as

$$\text{normal rank}\left[\mathbf{H}(\mathfrak{q})\ \mathbf{F}(\mathfrak{q})\right] = n_f + \text{normal rank}\ \mathbf{H}(\mathfrak{q}),$$

which carries equivalent implications in both the state-space and the transfer-function model descriptions (131, 132). An immediate consequence of the left invertibility is that $n_y \geq n_f$.

In cases where designing dedicated filters for individual faults is not feasible, a possible relaxation is to group a subset of fault signals and reconsider the isolation problem. This leads to the concept of a structured residual set (4). A different perspective to achieve fault isolability involves assigning a unique directional residual induced by a particular fault in the residual space, referred to as the signature direction (8, 32). The isolation task, therefore, amounts to determining which signature direction the generated residual signal most closely aligns with. The relative signature directions can be optimized to maximize their separability. Although handling simultaneous faults is straightforward with a bank of filters, it is generally more challenging in the directional residual method. On the other hand, a bank of filters imposes stricter requirements on the system structure. Further, Beard (8) and Massoumnia (31) investigated the underlying problem from a geometrical standpoint and derived the necessary conditions for fault isolability in terms of output separability.

4.1. Nonlinear Systems

Fault isolation for nonlinear systems was first addressed using linearization at local operating points, followed by decoupling the disturbances together with the higher-order terms from the residuals (see, e.g., 73; for a survey, see 133). Zhang et al. (29) used a nonlinear DAE approach, relying on differential-polynomial representations. With that, Ritt–Wu's algorithm is used to decouple the effect of particular faults. A seminal work by De Persis & Isidori (14) used a geometric approach to formulate the theory of diagnosability in nonlinear systems, considering continuous-time input- and fault-affine systems of the form

$$\mathfrak{q}\mathbf{x} = \mathbf{g}(\mathbf{x}) + \mathbf{g}_u(\mathbf{x})\mathbf{u} + \mathbf{g}_f(\mathbf{x})\mathbf{f},$$

$$\mathbf{y} = \mathbf{h}(\mathbf{x}).$$

In this setup, the authors found geometric conditions for the problem of designing an observer-based residual generator that is sensitive to only one of the faults $\mathbf{f}$, proceeding to show how to design such observers using a high-gain approach.

Both the algebraic method of Zhang et al. (29) and the geometric approach of De Persis & Isidori (14) suffer from sheer design complexity, sometimes requiring pencil-and-paper calculations or computations with nontrivial complexity, such as Ritt–Wu's method. Therefore, linear approximations may be necessary for larger-scale problems. LTI approximations can be (implicitly) used via the robustification approaches presented in Section 3 (see 17, 21). Alternatively,

linear time-varying approximations or the particular case of linear parameter-varying approximations can be used. Bokor & Balas (15) and van der Ploeg et al. (52) developed the robustification approach for linear parameter-varying systems. In the specific case of actuator faults, Zhang (53) provided an adaptive Kalman filter for linear time-varying systems with stability and minimum variance guarantees. Venkateswaran et al. (54) showed that, depending on some differential conditions on $\mathbf{g}(\cdot)$ and $\mathbf{h}(\cdot)$, linear residual generators exist that provide perfect decoupling of nonlinear systems; in this case, a systematic approach for designing diagnosis filters is available.

Several works have used nonlinear state estimators to approximate the nonlinear terms (16, 55, 56). Boem et al. (57) addressed large-scale nonlinear systems subject to stochastic noise, which give probabilistic false-alarm guarantees.

4.2. Multimodel Hypothesis: Passive and Active Methods

Instead of trying to understand which channel from a specified fault matrix is active, certain classes of fault detection and isolation problems are better modeled by a multimodel hypothesis. That is, in the DAE framework, a model of the type

$$\mathbf{H}(\mathfrak{q})\mathbf{x} + \mathbf{L}(\mathfrak{q})\mathbf{z} + \mathbf{W}(\mathfrak{q})\mathbf{w} + \boldsymbol{\eta}(\mathbf{x}, \mathbf{z}) = \mathbf{0}$$

is considered healthy, while m models of the type

$$\mathbf{H}_i(\mathfrak{q})\mathbf{x} + \mathbf{L}_i(\mathfrak{q})\mathbf{z} + \mathbf{W}_i(\mathfrak{q})\mathbf{w} + \boldsymbol{\eta}_i(\mathbf{x}, \mathbf{z}) = \mathbf{0} \quad \text{with} \quad i \in \{1, \ldots, m\}$$

are considered to be the potential m fault modes. The objective of the fault isolation filter in this multimodel classification scenario is to identify which of the $m + 1$ modes is active. Halimi et al. (58) and Küsters & Trenn (59) established the necessary rank conditions to guarantee that any two subsystems can be distinguished from each other.

4.2.1. Passive isolation. To address the isolation problem, the bank-of-filters approach has been developed using methods such as parity space (39), UIOs (40), and sliding mode observers (60, 61). However, a drawback of this approach is that the computational complexity of the filters, in both design and implementation, increases significantly as the system dimension and the number of modes increase. Dong et al. (62) designed the filters via DAE methods, enabling reduced-order designs that are tractable via convex optimization formulations. For linear systems under sub-Gaussian stochastic noise, they provided false-alarm rate bounds that admit a logarithmic dependency with respect to the desired reliability level, improving on the polynomial rate in the work by Boem et al. (57).

Another approach for mode detection uses set-membership methods, which rely on reachability analysis. These methods are better suited under bounded noise assumptions instead of probabilistic ones. In set-based methods, the current output is compared with the reachable sets of each mode, which enables the exclusion of modes whenever the output is out of the corresponding mode's reachable set. Harirchi & Ozay (68) reduced the set-membership check to the feasibility of a mixed-integer linear programming problem.

4.2.2. Active methods. Multimodel classification performance is highly dependent on the input. Therefore, designing optimal inputs can help better distinguish between the possible modes. Heirung & Mesbah (134) published a survey on input design for the multimodel classification problem that extensively reviews methods for both stochastic and set-based settings. For a few cases—e.g., single input–single output LTI systems with a two-model hypothesis and simple input constraints—analytical solutions have been derived. The focus in the past two decades has been on more complex cases that require computational methods and dealing with the tractability of the nonconvex optimization problems they form.

Scott et al. (43) and Marseglia & Raimondo (26) used an active set-based diagnosis approach, where an optimally separating input sequence is designed to maximize the chances of pinpointing a single mode. Set-based methods are known to be computationally demanding, and input design problems are particularly expensive. Blanchini et al. (44) partially addressed this issue by placing the uncertainty propagation burden offline. More recent research has focused on the tractability and generality of the set computations involved, and for that, sets such as constrained zonotopes (41) for discrete-time ordinary differential equations and line zonotopes (42) for discrete-time DAEs have been developed.

Despite having better computation tractability, the stochastic case also involves nonconvex optimization problems. Noom et al. (81) have recently made advances in this area; they used disciplined concave minimization by operating on a subdomain of the input space and showed that a concave function serves as a good envelope approximation of the Bhattacharyya separation criterion. A recent alternative is to use chance-constrained optimization. Guo et al. (45) proposed a new separation criterion based on the Cauchy–Schwarz divergence, requiring a three-stage optimization to obtain the optimal input. Qiu et al. (46) combined bounded and Gaussian uncertainties using set-based confidence levels; the resulting input design problem is a mathematical program with complementarity constraints.

4.3. Model-Free Approaches

Providing data-driven solutions to the fault isolation problem depends on the type of fault information available prior to design. In general, it is assumed that only fault-free system trajectories are collected and that $\mathbf{F}(\mathfrak{q})$ is not known explicitly. Based on healthy data, it is reasonable to assume that any present fault enters the system through actuators and/or sensors. The literature on data-driven fault isolation design was developed predominantly for a class of additive actuator/sensor faults (e.g., 33, 80, 88). Accordingly, in the state-space representation, the pair $(\mathbf{B}_f, \mathbf{D}_f)$ takes values from the columns of $(\mathbf{B}_u, \mathbf{D}_u)$ for actuator faults and from $(\mathbf{0}, \mathbf{I}_{n_y})$ for sensor faults. For the DAE formulation, we have

$$\mathbf{F}(\mathfrak{q}) = \begin{bmatrix} \mathbf{L}_{11} & \mathbf{0} \\ \mathbf{L}_{21} & \mathbf{I} \end{bmatrix}.$$

Once fault matrices are considered this way, one approach is to identify $(\mathbf{B}_u, \mathbf{D}_u)$ using subspace identification techniques in order to apply model-based methods for isolation. Instead of the mentioned indirect approach, a direct estimation of $\mathcal{K}_L^y \mathbf{T}_L^u$ from the data is sufficient to implement a parity-relation-based bank of filters, as proposed by Ding et al. (80).

Regarding directional residual signals, Sheikhi et al. (33) showed that the signature matrix $\mathcal{K}_L^y \mathbf{T}_L^{f_i}$ corresponding to an actuator or sensor fault can be estimated directly from the given data, up to a similarity transformation. In fact, signature matrices characterize the subspaces spanned by the columns of $\mathcal{K}_L^y \mathbf{T}_L^{f_i}$ and can be directly retrieved using subspace identification techniques. Thanks to the multidimensionality of the residual signal for a sufficiently large filter horizon L, the directional residual approach partitions the residual space according to each fault-induced direction, as illustrated in **Figure 4**. In this space, each hyperplane corresponds to a specific fault-induced signature matrix. The intersection of these hyperplanes is related to the underlying system zeros and is fundamental for characterizing mutual fault isolability (5), which can be verified using only data (33). Although Gleizer (94) recently showed that the span of fault matrices can be identified, data-driven fault isolation for general fault matrices remains a relatively underresearched area. For nonlinear systems, data-driven fault isolation solutions have been proposed for the class of nonlinear systems described by Takagi–Sugeno fuzzy models (97), basis functions (98), deep neural networks (99), and Koopman operators with guaranteed input-to-state stability (100).

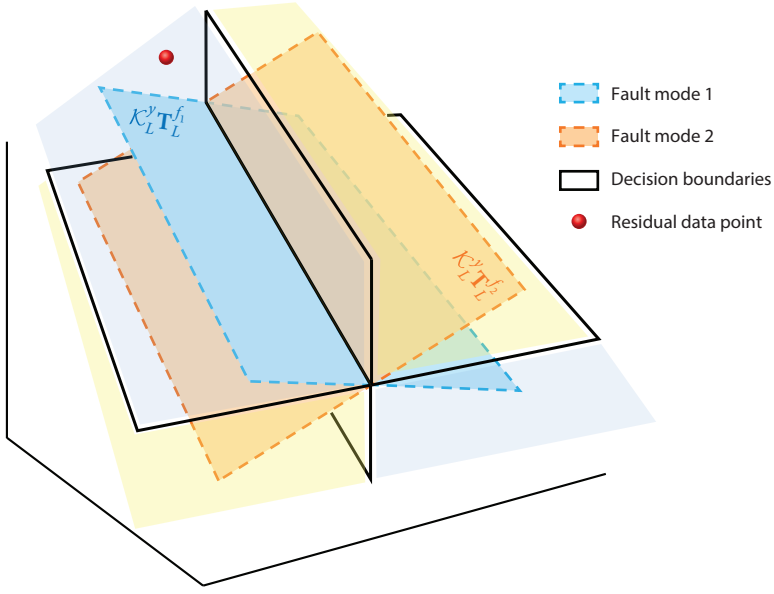

Figure 4

Conic partitioning of the residual space, a lower-dimensional projection of the input–output measurement space, for the isolation problem using a linear filter. The dashed blue and orange hyperplanes depict the column spaces of the signature matrices corresponding to fault modes 1 and 2, respectively ($\mathcal{K}_L^y \mathbf{T}_L^{f_i}, i \in \{1, 2\}$). The solid lines denote the decision boundaries set by the isolation logic. Based on these boundaries, the blue and orange cones indicate the partitions associated with fault modes 1 and 2, respectively. The red circle is a sample residual signal, identified as fault mode 1 since it lies within the corresponding conic region, i.e., closer to the blue hyperplane.

5. ESTIMATION

In estimation problems, the objective of the diagnosis filter (**Figure 1**) is that the residual $\mathbf{r}$ tracks or reconstructs the fault signal $\mathbf{f}$. Referring to Equation 2, we can see that in the absence of noise (i.e., $\mathbf{w} = \mathbf{0}$), this leads to the following objectives:

- Tracking: For any $\mathbf{x}$, $\lim_{t\to\infty} |\mathbf{f}(t) - \mathbf{r}(t)| = \mathbf{0}$.
- Reconstruction: For any $\mathbf{x}$, $\mathbf{r} = \mathfrak{q}^{-\tau}\mathbf{f}$ for some prescribed delay τ.

Note that delayed reconstruction problems are only well-defined in discrete time and can be understood as deadbeat implementations of tracking filters.

The filter design can make use of additional information about the fault signal. For example, the fault may be assumed to piecewise constant or polynomial (12, 25, 63), band limited (71), slow (84), periodic (87, 101), generated by a known autonomous system (135, section 8.5), or sparse (13, 22, 88). Fault estimation problems can be further characterized by the nature of the fault's effect on the system. The standard DAE description in Equation 2 presents what is called an additive fault signal. In some cases, the faulty behavior is better described by multiplicative faults, i.e., when the fault signal $\mathbf{f}$ multiplies other system variables, such as $\mathbf{x}$ and $\mathbf{z}$. Estimation of multiplicative faults has its particular challenges and is therefore addressed separately in Section 5.3 for passive methods and in Section 5.4 for active methods. The remaining subsections are devoted to additive fault estimation: Sections 5.1 and 5.2 are devoted to model-based estimation methods for linear and nonlinear systems, respectively, and Section 5.5 is devoted to model-free methods.

5.1. Linear Systems

Using Equation 2 with $\boldsymbol{\eta}(\cdot,\cdot) \equiv \mathbf{0}$, the fault estimation filter $-a(\mathfrak{q})^{-1}\mathbf{N}(\mathfrak{q})\mathbf{L}(\mathfrak{q})$ must satisfy one of the criteria established above for fault tracking or reconstruction. First, note that for any $\mathbf{N}(\mathfrak{q})$ subject to $\mathbf{N}(\mathfrak{q})\mathbf{H}(\mathfrak{q}) = \mathbf{0}$, we have that $a(\mathfrak{q})\mathbf{r} = \mathbf{N}(\mathfrak{q})\mathbf{F}(\mathfrak{q})\mathbf{f}$. If no more information on the fault signal is known, this gives the following estimation condition:

$$\mathbf{N}(\mathfrak{q})\mathbf{F}(\mathfrak{q}) \equiv a(\mathfrak{q})\mathbf{I}. \tag{19.}$$

Equation 19 may be difficult to solve in practice. The first observation is that one must find a pair $\mathbf{N}(\mathfrak{q}), a(\mathfrak{q})$ such that the degree of $\mathbf{N}(\mathfrak{q})\mathbf{L}(\mathfrak{q})$ is no greater than that of $a(\mathfrak{q})$. For state-space systems with $\mathbf{B}_f \neq \mathbf{0}$ and $\mathbf{D}_f = \mathbf{0}$, this is not possible, and the estimator needs differentiators in the continuous-time case (7, 136). In discrete time, differentiation is replaced by a time shift, and as such, it is possible to relax the criterion in Equation 19 to a τ-delay fault estimation (50) criterion,

$$\mathbf{N}(\mathfrak{q})\mathbf{F}(\mathfrak{q}) \equiv \mathbf{I}, \tag{20.}$$

which results in the residual satisfying $\mathbf{r} = a(\mathfrak{q})^{-1}\mathbf{f}$. Thus, by picking $a(\mathfrak{q}) = \mathfrak{q}^{\tau}$, where τ is at least the degree of $\mathbf{N}(\mathfrak{q})\mathbf{L}(\mathfrak{q})$, we achieve a maximal-bandwidth delayed input reconstruction, also referred to as deadbeat estimation (51). Gillijns & De Moor (49), Kirtikar et al. (50), and Yong et al. (69) showed for minimal state-space realizations that simultaneous τ-delay initial state and input reconstruction requires that the fault-output subsystem be (τ-delay) left invertible and have no invariant zeros, a condition that translates, in the DAE framework, to the existence of a matrix $\mathbf{M}(\mathfrak{q})$ such that $\mathbf{M}(\mathfrak{q})[\mathbf{H}(\mathfrak{q})\ \mathbf{F}(\mathfrak{q})] = \mathbf{I}$ for all $\mathfrak{q} \in \mathbb{C}$ (25). We can see that the combined conditions $\mathbf{N}(\mathfrak{q})\mathbf{H}(\mathfrak{q}) = \mathbf{0}$ and $\mathbf{N}(\mathfrak{q})\mathbf{F}(\mathfrak{q}) = \mathbf{I}$ are a strict subset of the latter via the following trivial example:

$$\begin{aligned} \mathfrak{q}\mathbf{x} + \mathbf{x} + \mathbf{u} + \mathbf{d} + \mathbf{f} &= 0, \\ \mathbf{y} + \mathbf{f} &= 0, \end{aligned}$$

where obviously the filter $\mathbf{r} \equiv -\mathbf{y}$ provides delay-free fault estimation even though the state is not observable. On the other hand, for systems with at least one invariant zero, asymptotic estimation can be considered (47–49, 69, 70).

Given how difficult it is to find residual generators satisfying Equation 19 or 20, several recent works have focused on estimation when some information about the fault is known. For example, van der Ploeg et al. (12) considered the case of piecewise-constant faults, which reduces the estimation criterion to the DC-gain version $\mathbf{N}(1)\mathbf{F}(1) = a(1)\mathbf{I}$ in discrete time. The fault can also be assumed to have a more generic frequency content. This was considered by Dong et al. (71) in the discrete-time case, where it is assumed that the fault spectrum lies in a subset $\Omega \subset \mathbb{R}_+$. Then, the corresponding stringent condition

$$\mathbf{N}(e^{j\omega})\mathbf{F}(e^{j\omega}) = a(e^{j\omega})\mathbf{I}, \quad \forall \omega \in \Omega,$$

which may be as difficult to solve as Equation 19, is replaced by

$$||\mathbf{N}(e^{j\omega})\mathbf{F}(e^{j\omega}) - d(e^{j\omega})\mathbf{I}||_2^2 \leq \eta_1, \quad \text{for a finite collection of } \omega \in \Omega, \tag{21.}$$

making it amenable to use in optimization problems. In particular, Dong et al. (71) considered a combination of the approximate tracking condition in Equation 21 with the $\mathcal{H}_2$-norm of the transfer function from noise $\mathbf{w}$ to residual, which can be formulated as a simple quadratic constraint. The result is a quadratic program problem, which makes it suitable for large-scale problems.

When the fault is assumed to be a function of time with unknown parameters, fault estimation can often be recast as a state observer problem using a generator subsystem approach. For example, a fault that is a polynomial function of time can be modeled as a chain of integrators whose

initial states must be reconstructed in what is called an ultralocal approach (63). This approach has been well-developed for nonlinear systems (see Section 5.2). Many other functions of time can be modeled; e.g., a sinusoid $f(t) = f_0 \sin(t) + f_1 \cos(t)$ can be modeled as the outcome of the autonomous system $\ddot{f} + f = 0$.

Most fault reconstruction works rely on a state–fault parameterization of the behavioral set, by first removing the contribution of the input $\mathbf{u}$. Referring to Equation 3 in the absence of noise gives the condition

$$\mathbf{r} := \mathbf{y} - \mathbf{T}_L^u \mathbf{u} = \mathbf{O}_L \mathbf{x}_0 + \mathbf{T}_L^f \mathbf{f}. \tag{22}$$

Therefore, simply using a left inverse of $[\,\mathbf{O}_L \;\; \mathbf{T}_L^f\,]$ provides a perfect L-delay fault reconstruction, provided such a left inverse exists. This coincides with the τ-delay initial state and input reconstructability conditions of Kirtikar et al. (50). Dong & Verhaegen (19, 82) noted that the inversion of $[\,\mathbf{O}_L \;\; \mathbf{T}_L^f\,]$ can be avoided if one refrains from estimating or annihilating the initial state; instead, by using a stable Kalman predictor representation, the filter is built by inverting only $\mathbf{T}_L^f$, ignoring the presence of $\mathbf{x}_0$ in Equation 22. The critical observation is that the estimation bias depends on the transmission zeros from fault to output: When there are minimum-phase zeros, the bias decreases exponentially with L, and when there are non-minimum-phase zeros (93), a mixed causal–anticausal representation (as proposed in 85) enables minimal bias somewhere in the middle of the reconstruction window; thus, tuning the delay becomes a fundamental step, and faults in non-minimum-phase systems can only be estimated accurately with very long delays. For zeros located on the unit circle, the bias cannot be controlled with the delay; e.g., for a zero at 1, the fault reconstruction has a constant bias. Finally, the Markov-parameter approach facilitates the consideration of time-domain properties of the fault signal. For example, Zhang (13) and Noom et al. (88) considered the scenario where the fault is assumed to be the combination of a few signals from a dictionary of possible fault signals. The sparsity in terms of the number of active dictionary entries is achieved by adding a 1-norm regularizer on the fault signal. In large-scale systems, not all components of the fault signal are necessarily nonzero at the same time; rather, only a subset may be active during the filter horizon. This induces sparsity at the signal level, which Anguluri et al. (22) exploited to recover the nonzero components.

Minimum-phase zero: a system zero located inside the unit circle (discrete time) or in the left half-plane (continuous time)

Non-minimum-phase zero: a system zero located outside the unit circle (discrete time) or in the right half-plane (continuous time)

5.2. Nonlinear Systems

When the nominal system has nonlinear dynamics, there are two dominant approaches in the literature. One is to design a nonlinear fault estimator that effectively captures the effect of the nonlinearities in the behavior; the other is to design a linear fault estimation filter that is robust against the nonlinearities. The latter may use techniques similar to those discussed in Section 3. For certain classes of nonlinearities, linear filters suffice. For instance, if the nonlinear term in Equation 2 is just a function of $\mathbf{z}$, then a new known signal $\mathbf{z}' = E(\mathbf{z})$ can be generated online, effectively turning the nonlinear DAE into a linear DAE for estimation purposes. In the cases where $E(\mathbf{z})$ multiplies the fault signal, this becomes a multiplicative fault problem, which is reviewed in Section 5.3.

Ghanipoor et al. (63) extended the ultralocal approach in Section 5.1 to nonlinear systems, approximating the fault signal as the outcome of a chain of integrators:

$$\begin{cases} \dot{\bar{\zeta}}_j = \bar{\zeta}_{j+1}, & 0 < j < r, \\ \dot{\bar{\zeta}}_r = \mathbf{0}, \\ \bar{\mathbf{f}} = \bar{\zeta}_1, \end{cases} \tag{23}$$

where $\bar{\zeta}_j \in \mathbb{R}^{n_f}$. This technique requires that the fault be r times differentiable and can provide the rth-order Taylor expansion of the fault signal at a given point. The fault estimation filter then becomes a nonlinear observer, in which the original system state is augmented with the states of the actual fault internal state. The solution to the design problem is obtained by solving linear matrix inequalities, providing an input-to-state stable nonlinear filter. These inequalities can include objective functions, yielding semidefinite programs that minimize either the $\mathcal{L}_2$ gain or the $\mathcal{L}_2$-$\mathcal{L}_\infty$ induced norm from uncertainties to residual. Increasing r leads to high-dimensional observers and scalability issues yet also provides more degrees of freedom for optimal synthesis.

Witczak et al. (64) also used linear matrix inequalities for observer design; the noise $\mathbf{w}$ and the time difference of the fault $\mathbf{f}_{k+1} - \mathbf{f}_k$ are assumed to be in ℓ_2, and the nonlinearities satisfy certain Lipschitz assumptions. In addition, the so-called observer matching condition $\operatorname{rank}(\mathbf{B}_f) = \operatorname{rank}(\mathbf{C}\mathbf{B}_f) = n_f$ is imposed. Zhu et al. (65) circumvented this condition; their approach also relies on linear matrix inequalities to build an intermediate estimator, instead of imposing that there are no invariant zeros between the fault and output of the linear part of the system. More classical results rely on adaptive high-gain observers (66), where the additive fault is assumed to be the sum of known function templates whose weights are to be estimated.

5.3. Multiplicative Faults

Faults are often better represented by changes in system parameters or unexpected dynamics that enter the system. It is more appropriate to model such faults as signals that multiply the system's signals, such as inputs, outputs, and internal states. Van der Ploeg et al. (52) and Gleizer et al. (25) addressed the estimation of multiplicative faults under the DAE framework. The combined approaches enable the estimation of faults represented as

$$\left(\mathbf{H}(\mathfrak{q}) + \sum_{i=1}^{m} f_i^{\mathrm{m}} \mathbf{H}_i'(\mathfrak{q})\right)\xi + \left(\mathbf{L}(\mathfrak{q}) + \sum_{i=1}^{m} f_i^{\mathrm{m}} \mathbf{L}_i'(\mathfrak{q})\right)\mathbf{z} + \mathbf{F}(\mathfrak{q})(\mathbf{f}^{\mathrm{a}} + E(\mathbf{z})\mathbf{f}^{\mathrm{m}}) + \mathbf{W}(\mathfrak{q})\mathbf{w} = 0, \qquad 24.$$

where $\mathbf{f}^{\mathrm{m}}$ and $\mathbf{f}^{\mathrm{a}}$ represent multiplicative and additive faults, respectively. In much of the literature, multiplicative faults are recast as intermediate additive faults. For example, instead of trying to separately estimate $\mathbf{f}^{\mathrm{a}}$ and $\mathbf{f}^{\mathrm{m}}$, one may choose to estimate the combined signal $\mathbf{f}' := \mathbf{f}^{\mathrm{a}} + E(\mathbf{z})\mathbf{f}^{\mathrm{m}}$ and at a later stage distinguish the individual components. This approach, however, can make the estimation problem infeasible. Suppose, for example, that only one scalar multiplicative fault is modeled, with $\mathbf{H}'(\mathfrak{q}) = \mathbf{0}$ and $f^{\mathrm{m}}\mathbf{L}'(\mathfrak{q})\mathbf{z} = \begin{bmatrix} f^{\mathrm{m}}u & f^{\mathrm{m}}y_1 & f^{\mathrm{m}}y_2 \end{bmatrix}^{\mathrm{T}}$. By simply recasting $f_1^{\mathrm{a}} := f^{\mathrm{m}}u$, $f_2^{\mathrm{a}} := f^{\mathrm{m}}y_1$, and $f_3^{\mathrm{a}} := f^{\mathrm{m}}y_2$, one must estimate three fault signals instead of one, which, in fact, is structurally infeasible if the system has only two outputs. In addition to these technical challenges, estimating the multiplicative component is generally more informative than estimating the combined fault signal.

Multiplicative fault estimation poses distinct challenges. Even with knowledge of $E(\mathbf{z})$, it is not possible to separate the individual contributions from the additive and multiplicative components of $\mathbf{f}^{\mathrm{a}} + E(\mathbf{z})\mathbf{f}^{\mathrm{m}}$ without extra assumptions about the fault signals, such as frequency content. Furthermore, the multiplicative nature of the fault signal makes the map from $\mathbf{f}^{\mathrm{m}}$ to the residual a nonlinear map, thus requiring nonlinear components in the estimation design. Van der Ploeg et al. (12) assumed that the faults are piecewise constant, and Gleizer et al. (25) extended this work to the case where the faults are a combination of known time signals—i.e., $f_i^{\mathrm{m,a}}(t) = \sum_j p_{ij}\phi_{ij}(t)$, where p_{ij} are parameters to be identified. This includes polynomial functions, which represent an ultralocal approach.

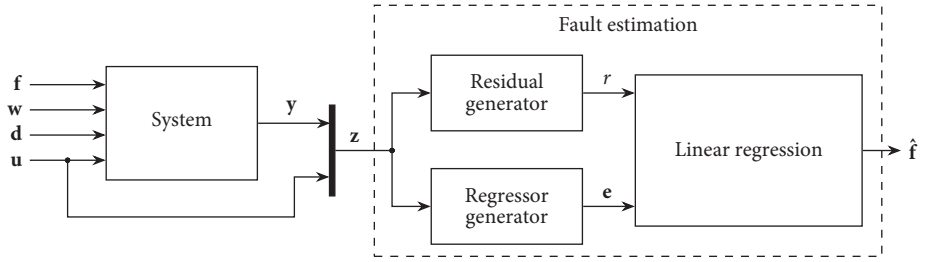

Figure 5

Fault estimation diagram for multiplicative faults. The regressor generator produces signals e_i representing the contribution of a unit fault. A moving-horizon linear regression computes the fault estimates $\hat{\mathbf{f}}$.

The fault estimation method proposed by van der Ploeg et al. (12) and Gleizer et al. (25) involves three main components, as depicted in **Figure 5**: a residual generator, a so-called regressor generator, and a regression block. To explain the concept, consider the case of (piecewise) constant faults and one additive fault, where all faults only multiply known signals [i.e., $\mathbf{H}'_i(\mathfrak{q}) = \mathbf{0}$ for all i]. Let $\mathbf{z}' := E(\mathbf{z})$ be the result of the nonlinearity applied to the known signals $\mathbf{z}$. Then, referring back to Equation 24, we can see that if $\mathbf{N}(\mathfrak{q})\mathbf{H}(\mathfrak{q}) = \mathbf{0}$, then the residual described by $a(\mathfrak{q})r = -\mathbf{N}(\mathfrak{q})\mathbf{L}(\mathfrak{q})\mathbf{z}$ satisfies

$$r = a(\mathfrak{q})^{-1}\mathbf{N}(\mathfrak{q})\Big(\sum_i (\mathbf{G}'_i(\mathfrak{q})\mathbf{z} + \mathbf{F}(\mathfrak{q})\mathbf{z}')f_i^{\mathrm{m}} + \mathbf{F}(\mathfrak{q})f^{\mathrm{a}}\Big) =: \sum_{i=1}^{m} f_i^{\mathrm{m}}\mathbf{M}_i^{\mathrm{m}}(\mathfrak{q})\begin{bmatrix}\mathbf{z}\\ \mathbf{z}'\end{bmatrix} + f^{\mathrm{a}}\mathbf{M}^{\mathrm{a}}(\mathfrak{q})\mathbf{1}, \quad 25.$$

where $\mathbf{G}_i(\mathfrak{q}) = \mathbf{L}'_i(\mathfrak{q})$ for all $i = 1, \ldots, m$, and $\mathbf{1}(t) \equiv 1$ is the constant signal equal to 1. That is, the residual is a linear combination, weighted by the faults, of the output of multiple linear filters of the signals $\mathbf{z}$, $\mathbf{z}' = E(\mathbf{z})$, and $\mathbf{1}$. The outputs are collected in the signal $\mathbf{e}$ in **Figure 5**. With this, a linear regression can be performed from $\mathbf{e}$ to r over a user-specified time horizon to retrieve the fault estimates $\hat{\mathbf{f}}$. When $\mathbf{H}'_i(\mathfrak{q}) \not\equiv \mathbf{0}$ (i.e., the faults also multiply functions of the latent variables $\mathbf{x}$), Equation 25 holds approximately provided that the matrix $\mathbf{H}(\mathfrak{q})$ admits a polynomial left inverse and includes extra terms in $\mathbf{G}'_i(\mathfrak{q})$ (see 25).

5.4. Active Methods and Input Design

For linear systems with perfect model information, the performance of additive fault estimation does not depend on the input. This is not the case in the presence of model uncertainties or multiplicative faults. Similar to the multimodel isolation case (Section 4.2), active methods can dramatically improve diagnosis performance.

Tan et al. (72) used a set-based approach to address the input design of additive faults in linear parameter-varying systems with inexact scheduling variables. In such a case, the optimal input tries to concentrate the system behavior where the scheduling variable uncertainty is the least. The one-step-ahead problem is nonconvex, but the global optimal input can be computed by solving 2^{n_u} quadratic programs. The authors also investigated how to combine fault estimation and mitigation within the same framework.

Considering the estimation framework in Section 5.3, Gleizer et al. (25) proposed an input design method for multiplicative fault estimation and found that it requires that the regressors be persistently exciting of sufficiently high order. Moreover, when the noise $\mathbf{w}$ is nonzero, the bias and variance of the estimation decrease as the richness of the regressor signal increases. The authors proposed designing $\mathbf{u}$ to maximally excite the regressor $\mathbf{e}$, which is a nonconvex problem; therefore, local optimization methods are exploited, together with a convex relaxation to provide suboptimality bounds.

5.5. Model-Free Approaches

Most of the literature on model-free (or data-driven) fault estimation considers LTI systems and assumes that a fault-free dataset is initially available. This essentially allows a model to be identified either directly or indirectly. Referring to the state-space description in Equation 3, one can adopt a similar fault modeling approach to that in Section 4.3—i.e., if the class of faults is restricted to sensors ($\mathbf{B}_f = \mathbf{0}$, $\mathbf{D}_f = \mathbf{I}$), actuators ($\mathbf{B}_f = \mathbf{B}_u$, $\mathbf{D}_f = \mathbf{D}_u$) (90, 91), or load disturbances ($\mathbf{B}_f = \mathbf{I}$, $\mathbf{D}_f = \mathbf{0}$) (87), then the full faulty system can be identified from healthy data. Existing literature has addressed this identification problem through system matrices and/or Markov parameters (19, 85, 93), basis functions (91), or behavioral subspaces (92); Ding (125) provided an overview of the more established methods.

Once a model is (implicitly) identified, model-based methods can be used. A challenge that appears in the data-driven settings, though, is that identification errors can render the left inverse unstable, even when the real system admits stable left inverses. Wan et al. (89) explicitly addressed this challenge and showed how to design stable approximate left inverses for fault estimation by using high-order vector autoregressive with exogenous inputs (VARX) approximations of the Kalman predictor representation of the system. Their method requires the fault-output system to have no non-minimum-phase zeros, an assumption later dropped in a work by Yu & Verhaegen (85). Challenges related to forward and inverse stability are overcome by design in bilateral finite impulse response approaches (e.g., 91).

The problem of simultaneously identifying a system and its external input (or fault) has received less attention. Several works have addressed the cases where several assumptions can be imposed on the fault signal, such as being constant, slow-varying, or periodic (84, 87, 101). The more fundamental problem where no assumptions are made about the fault signal first appeared in a work by Palanthandalam-Madapusi & Bernstein (83). They used subspace methods to fully estimate ($\mathbf{A}$, $[\mathbf{B}_u, \mathbf{B}_f]$, $\mathbf{C}$, $[\mathbf{D}_u, \mathbf{D}_f]$) directly from faulty data, assuming only that the unknown input space dimension is known and the fault-output system is input and state observable. Gleizer (94) recently extended this work and showed that the input–output system can be identified irrespective of the presence of faults and without requiring unknown input observability, provided that the faults are open-loop and the input is random and zero-mean. This work then provides methods that enable finding the smallest unknown input space dimension and the corresponding span of fault matrices ($\mathbf{B}_f, \mathbf{D}_f$) that explain the data. These methods are exact in the absence of noise and rely on simple linear algebra but can suffer from severe noise sensitivity. When considering the fault signals to be a sparse linear combination of known signals, Noom et al. (88) also provided a model-free version, where the identification of system parameters is done simultaneously with the fault identification. This involves a combination of convex relaxations: nuclear norms as a relaxation of the rank for the identification part and the 1-norm as a relaxation of the 0-norm (sparsity) for the fault part. Liu et al. (102) suggested using a fuzzy model to design an adaptive filter for a class of Lipschitz nonlinear systems.

6. APPLICATIONS

This section illustrates how the theoretical developments of fault diagnosis reviewed in the preceding sections can translate into practice through several real-world applications. To this end, we highlight distinct challenges encountered in different modern industrial applications and the tailored methodologies developed in response. For clarity, each subsection includes a subtitle that reflects the key difficulty or defining feature of the application (such as high noise levels, model uncertainty, or sensing limitations), which serves as the primary motivation for the design of specialized fault diagnosis techniques. Each subsection presents one or two specific use cases, along

Table 2 Example applications and their settings

Use case	Setting	Dynamics	Problem	Fault model
Power systems (Section 6.1)	Imperfect model	Nonlinear	Fault detection	Additive
Autonomous vehicles (Section 6.2)	Perfect model	Nonlinear	Fault estimation	Additive + multiplicative
Mechatronics (Section 6.3)	Perfect model	Linear	Fault estimation	Multiplicative
Industrial printing (Section 6.4)	Data-driven	Linear	Fault isolation	Multimodel

with references that report detailed experimental validations and quantitative results; Section 6.2 additionally includes a link to a short video demonstration of an experimental setup and its results. These applications are mapped to our proposed categories in **Table 2**.

6.1. Power Systems Case Study: High-Dimensional Nonlinear Dynamics

Fault detection in large-scale power systems is a critical task, particularly in the face of cyber-physical threats and significant natural disturbances. These systems are characterized by nonlinear dynamics and high levels of ambient noise due to demand fluctuations, renewable integration, and load uncertainties. A representative scenario is the detection of cyberattacks on automatic generation control loops in multiarea power networks (106). In such settings, faults or malicious disturbances are designed to mimic nominal behavior under natural fluctuations, making detection challenging.

Mohajerin Esfahani & Lygeros (17) utilized the robust fault detection methodology in Section 3.1 to address these conditions, constructing residual-based detectors that are robust against natural disturbances while remaining sensitive to adversarial faults. Svetozarevic et al. (116) applied a similar detection framework to the nonlinear dynamics of wind turbine systems that considers nonlinear aerodynamic dynamics and multiplicative noise. These results demonstrate the potential of robust linear detectors to handle fault detection in high-dimensional, nonlinear, and uncertain environments typical of modern power and energy systems.

6.2. Autonomous Vehicle Case Study: Inseparable Faults

Autonomous driving platforms require accurate real-time fault estimation to ensure safe operation under dynamic and uncertain conditions. This problem is typically addressed in systems modeled as linear but subject to structured uncertainties, actuator delays, and multiplicative noise whose dynamic effects are inseparable from some additive faults. Of particular interest are faults affecting steering performance, such as actuator degradation (multiplicative faults) or command biases (additive faults), which can degrade tracking performance or lead to unsafe maneuvers.

Van der Ploeg et al. (24) developed fault estimation techniques and validated them experimentally on a Renault Grand Scenic autonomous driving platform at the Netherlands Organization for Applied Scientific Research (TNO); **Figure 6*a*** shows the experimental setup.[2] The filter builds on the architecture proposed in Reference 12, in combination with a robustification layer to account for structured uncertainty and delay. The filter succeeds in estimating joint additive and multiplicative faults within 0.15 s, which is significantly faster than typical human reaction

[2]A short video of the experimental test is available at **https://mohajerinesfahani.github.io/Publications/journal/2024/FDI_experiment.mp4**.

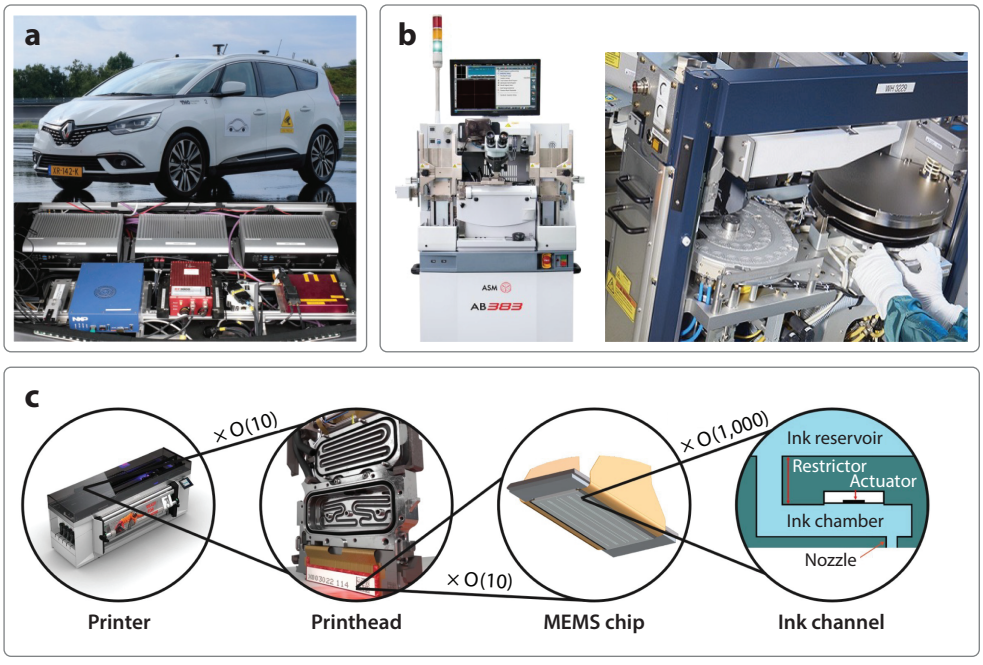

Figure 6

Fault diagnosis filters in real-world applications: (*a*) TNO Renault Grand Scenic autonomous driving test platform. Panel adapted with permission from Reference 24; copyright 2025 IEEE. (*b*) ASMPT AB383 wire bonder (*left*) (137) and ASML/VDL-ETG wafer-handling robot (*right*) (138, 139). Left image adapted with permission from Reference 137; right image adapted with permission from Reference 138. (*c*) Canon Production Printing varioPRESS iV7 industrial printer, with a schematic representation of the ink channel use case (140, 141). Panel adapted with permission from Reference 140. Abbreviations: MEMS, microelectromechanical system; TNO, Netherlands Organization for Applied Scientific Research.

times. The results demonstrated in this work show reliable estimation of both additive and multiplicative steering faults under realistic driving conditions, highlighting the applicability of robust estimation in automotive settings.

6.3. Mechatronics Case Study: Low Signal-to-Noise Ratio

Health monitoring of high-tech manufacturing systems, such as ASMPT's AB383 wire bonder, faces unique challenges in fault diagnosis due to these systems' ultrahigh precision requirements. Faults in such systems are often extremely small (sometimes <0.1% deviation) and evolve slowly over time (e.g., the case of wear and tear). This calls for fault estimation techniques capable of handling low signal-to-noise ratio conditions, often using active detection strategies to enhance sensitivity.

De Reij (137) implemented a robust estimation architecture on a 20-state linearized model of the wire bonder, estimating five multiplicative and three additive faults using a combination of the methods by Gleizer et al. (25) and frequency-domain techniques. Multiplicative faults correspond to parametric degradation (e.g., stiffness or damping changes), and additive faults represent factory floor vibrations; **Figure 6*b*** illustrates the experimental device of this use case. Van Esch et al. (139) studied a hybrid fault model for selective compliance assembly robot arm (SCARA) manipulators used in chip handling, where nonlinear tilt faults of small angular

magnitude proved particularly difficult to detect. They found that the detection threshold was around 2°, with lower-magnitude faults (e.g., milliradian-level tilts) evading reliable detection. These studies reinforce the importance of combining active detection with geometric and robust estimation frameworks for successful deployment in high-precision manufacturing systems [see the recent work by Gleizer et al. (25)].

6.4. High-End Industrial Printers Case Study: Limited Sensors

High-end industrial inkjet printers operate under extreme precision requirements, yet a wide range of faults can lead to degraded performance. Fault isolation in this context is challenging for several reasons. First, sensing is very limited at the nozzle level, with only a single sensor available for both actuation and measurement. Second, most of the faults are multiplicative or even increase system dimensionality. And third, reliable physics models—especially for the several fault modes—are lacking. For this last reason, the fault diagnosis problem is thus approached in a data-driven manner.

Van Peijpe et al. (142) developed a fault isolation framework based on piezoelectric self-sensing and synthetic residual generation for classification. The nozzle system is modeled as a four-state LTI system, with parameters obtained via system identification. The diagnosis approach employs a supervised learning method, where synthetic fault signals are injected to train the classifiers. A total of 11 fault classes are considered, with fault activity constrained using simplex priors; **Figure 6*c*** shows the relevant components of the experimental setup alongside a schematic representation of the ink channels. Amini (141) further enhanced the classification performance using data analytics techniques and employed system-theoretic approaches to estimate severity levels within fault categories (e.g., partial versus complete dried nozzle). The short timescale (50 μs per jetting) and noisy data conditions make this a benchmark application for real-time, data-driven fault isolation.

SUMMARY POINTS

1. Problems and scalable solutions: We have reviewed classical problems in fault diagnosis, namely detection, isolation, and estimation, emphasizing a common framework based on differential–algebraic equation system descriptions and revealing its connection to alternative system representations. This framework enables scalable filter design based on convex optimization even when the system is nonlinear.
2. Geometric interpretation: We have provided a geometric interpretation of the underlying diagnosis problems, which establishes a connection with general classification problems and provides insight into novel solution approaches.
3. Active and model-free design: We have discussed recent advances in active and model-free fault diagnosis, while noting that these directions are still underexplored.
4. Modern use cases: We have presented four contemporary applications where each one has its own unique set of challenges, closing the gap between theory and practice.

FUTURE ISSUES

1. Computational scalability in active design: A formidable challenge in active design is the inherent nonconvexity of the resulting optimization problems. Recent developments in

nonconvex optimization algorithms (e.g., distributed optimization) can be helpful in this direction.

2. New robust design paradigms: Robust fault diagnosis often leads to conservative designs. Promising directions to alleviate this conservatism include game-theoretic and distributionally robust optimization approaches.
3. Direct model-free design: Most data-driven methods in fault diagnosis still rely on a two-step approach, although direct data-driven design has only recently begun to emerge in the diagnosis. However, performance characterization of direct approaches under noisy conditions is still missing. Furthermore, access to fault-free data is often a prerequisite in these approaches, which may not be available in certain applications.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was supported by the Digital Twin project (through project number P18-03 of the research program TTW-Perspective, which is partly financed by the Dutch Research Council) and European Research Council grant TRUST-949796.

LITERATURE CITED

1. Venkatasubramanian V, Rengaswamy R, Yin K, Kavuri SN. 2003. A review of process fault detection and diagnosis: part I: quantitative model-based methods. *Comput. Chem. Eng.* 27(3):293–311
2. Frank PM. 1990. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: a survey and some new results. *Automatica* 26(3):459–74
3. Zhang XJ. 1989. *Auxiliary Signal Design in Fault Detection and Diagnosis*. Springer
4. Gertler J, Singer D. 1990. A new structural framework for parity equation-based failure detection and isolation. *Automatica* 26(2):381–88
5. Massoumnia MA, Verghese GC, Willsky AS. 2002. Failure detection and identification. *IEEE Trans. Autom. Control* 34(3):316–21
6. Hou M, Patton RJ. 1998. Input observability and input reconstruction. *Automatica* 34(6):789–94
7. Park Y, Stein JL. 1988. Closed-loop, state and input observer for systems with unknown inputs. *Int. J. Control* 48(3):1121–36
8. Beard RV. 1971. *Failure accommodation in linear systems through self-reorganization*. PhD Thesis, Massachusetts Institute of Technology
9. Zhang Y, Jiang J. 2008. Bibliographical review on reconfigurable fault-tolerant control systems. *Annu. Rev. Control* 32(2):229–52
10. Chow E, Willsky A. 2003. Analytical redundancy and the design of robust failure detection systems. *IEEE Trans. Autom. Control* 29(7):603–14
11. Frisk E, Nyberg M. 2001. A minimal polynomial basis solution to residual generation for fault diagnosis in linear systems. *Automatica* 37(9):1417–24
12. van der Ploeg C, Alirezaei M, van de Wouw N, Mohajerin Esfahani P. 2022. Multiple faults estimation in dynamical systems: tractable design and performance bounds. *IEEE Trans. Autom. Control* 67(9):4916–23
13. Zhang Q. 2021. Dynamic system fault diagnosis under sparseness assumption. *IEEE Trans. Signal Process.* 69:2499–508

14. De Persis C, Isidori A. 2002. A geometric approach to nonlinear fault detection and isolation. *IEEE Trans. Autom. Control* 46(6):853–65
15. Bokor J, Balas G. 2004. Detection filter design for LPV systems—a geometric approach. *Automatica* 40(3):511–18
16. Zhang X, Polycarpou MM, Parisini T. 2008. Design and analysis of a fault isolation scheme for a class of uncertain nonlinear systems. *Annu. Rev. Control* 32(1):107–21
17. Mohajerin Esfahani P, Lygeros J. 2015. A tractable fault detection and isolation approach for nonlinear systems with probabilistic performance. *IEEE Trans. Autom. Control* 61(3):633–47
18. Shang C, Ding SX, Ye H. 2021. Distributionally robust fault detection design and assessment for dynamical systems. *Automatica* 125:109434
19. Dong J, Verhaegen M. 2011. Identification of fault estimation filter from I/O data for systems with stable inversion. *IEEE Trans. Autom. Control* 57(6):1347–61
20. Ding SX, Yang Y, Zhang Y, Li L. 2014. Data-driven realizations of kernel and image representations and their application to fault detection and control system design. *Automatica* 50(10):2615–23
21. Pan K, Palensky P, Mohajerin Esfahani P. 2021. Dynamic anomaly detection with high-fidelity simulators: a convex optimization approach. *IEEE Trans. Smart Grid* 13(2):1500–15
22. Anguluri R, Kosut O, Sankar L. 2023. Localization and estimation of unknown forced inputs: a group LASSO approach. *IEEE Trans. Control Netw. Syst.* 10(4):1997–2009
23. Pan K, Palensky P, Mohajerin Esfahani P. 2019. From static to dynamic anomaly detection with application to power system cyber security. *IEEE Trans. Power Syst.* 35(2):1584–96
24. van der Ploeg C, Vieira Oliveira P, Silvas E, Mohajerin Esfahani P, van de Wouw N. 2025. Robust fault estimation with structured uncertainty: scalable algorithms and experimental validation in automated vehicles. *IEEE Trans. Control Syst. Technol.* 33(5):1651–66
25. Gleizer GA, Mohajerin Esfahani P, Keviczky T. 2025. Active estimation of multiplicative faults in dynamical systems. Preprint, arXiv:2506.23769v1 [eess.SY]
26. Marseglia GR, Raimondo DM. 2017. Active fault diagnosis: a multi-parametric approach. *Automatica* 79:223–30
27. Nyberg M, Frisk E. 2006. Residual generation for fault diagnosis of systems described by linear differential-algebraic equations. *IEEE Trans. Autom. Control* 51(12):1995–2000
28. Ding X, Guo L, Frank PM. 2002. Parameterization of linear observers and its application to observer design. *IEEE Trans. Autom. Control* 39(8):1648–52
29. Zhang Q, Basseville M, Benveniste A. 1998. Fault detection and isolation in nonlinear dynamic systems: a combined input–output and local approach. *Automatica* 34(11):1359–73
30. Chen H, Chai Z, Dogru O, Jiang B, Huang B. 2021. Data-driven designs of fault detection systems via neural network-aided learning. *IEEE Trans. Neural Netw. Learn. Syst.* 33(10):5694–705
31. Massoumnia MA. 2003. A geometric approach to the synthesis of failure detection filters. *IEEE Trans. Autom. Control* 31(9):839–46
32. Chen J, Patton RJ, Zhang HY. 1996. Design of unknown input observers and robust fault detection filters. *Int. J. Control* 63(1):85–105
33. Sheikhi MA, Gleizer GA, Mohajerin Esfahani P, Keviczky T. 2025. Data-driven fault isolation in linear time-invariant systems: a subspace classification approach. *IEEE Control Syst. Lett.* 9:1598–603
34. Ding X, Frank PM. 1990. Fault detection via factorization approach. *Syst. Control Lett.* 14(5):431–36
35. Varga A. 2013. New computational paradigms in solving fault detection and isolation problems. *Annu. Rev. Control* 37(1):25–42
36. Gertler J. 1997. Fault detection and isolation using parity relations. *Control Eng. Pract.* 5(5):653–61
37. Li J, Wang Z, Shen Y, Xie L. 2023. Attack detection for cyber-physical systems: a zonotopic approach. *IEEE Trans. Autom. Control* 68(11):6828–35
38. Frisk E, Krysander M, Åslund J. 2009. Sensor placement for fault isolation in linear differential-algebraic systems. *Automatica* 45(2):364–71
39. Cocquempot V, El Mezyani T, Staroswiecki M. 2004. Fault detection and isolation for hybrid systems using structured parity residuals. In *2004 5th Asian Control Conference*. IEEE
40. Wang D, Lum KY. 2007. Adaptive unknown input observer approach for aircraft actuator fault detection and isolation. *Int. J. Adapt. Control Signal Process.* 21(1):31–48

41. Scott JK, Raimondo DM, Marseglia GR, Braatz RD. 2016. Constrained zonotopes: a new tool for set-based estimation and fault detection. *Automatica* 69:126–36
42. Rego BS, Raimondo DM, Raffo GV. 2025. Line zonotopes: a tool for state estimation and fault diagnosis of unbounded and descriptor systems. *Automatica* 179:112380
43. Scott JK, Findeisen R, Braatz RD, Raimondo DM. 2014. Input design for guaranteed fault diagnosis using zonotopes. *Automatica* 50(6):1580–89
44. Blanchini F, Casagrande D, Giordano G, Miani S, Olaru S, Reppa V. 2017. Active fault isolation: a duality-based approach via convex programming. *SIAM J. Control Optim.* 55(3):1619–40
45. Guo Y, Liu Q, Wang Z, Zhang Z, He X. 2023. Active fault diagnosis for linear stochastic systems subject to chance constraints. *Automatica* 156:111194
46. Qiu H, Xu F, Liang B, Wang X. 2023. Active fault diagnosis under hybrid bounded and Gaussian uncertainties. *Automatica* 147:110703
47. Floquet T, Barbot JP. 2006. State and unknown input estimation for linear discrete-time systems. *Automatica* 42(11):1883–89
48. Marro G, Zattoni E. 2010. Unknown-state, unknown-input reconstruction in discrete-time nonminimum-phase systems: geometric methods. *Automatica* 46(5):815–22
49. Gillijns S, De Moor B. 2007. Unbiased minimum-variance input and state estimation for linear discrete-time systems. *Automatica* 43(1):111–16
50. Kirtikar S, Palanthandalam-Madapusi H, Zattoni E, Bernstein DS. 2011. *l*-delay input and initial-state reconstruction for discrete-time linear systems. *Circuits Syst. Signal Process.* 30:233–62
51. Ansari A, Bernstein DS. 2019. Deadbeat unknown-input state estimation and input reconstruction for linear discrete-time systems. *Automatica* 103:11–19
52. van der Ploeg C, Silvas E, van de Wouw N, Mohajerin Esfahani P. 2021. Real-time fault estimation for a class of discrete-time linear parameter-varying systems. *IEEE Control Syst. Lett.* 6:1988–93
53. Zhang Q. 2018. Adaptive Kalman filter for actuator fault diagnosis. *Automatica* 93:333–42
54. Venkateswaran S, Liu Q, Wilhite BA, Kravaris C. 2022. Design of linear residual generators for fault detection and isolation in nonlinear systems. *Int. J. Control* 95(3):804–20
55. Boem F, Ferrari RM, Parisini T. 2011. Distributed fault detection and isolation of continuous-time non-linear systems. *Eur. J. Control* 17(5–6):603–20
56. Ferrari RM, Parisini T, Polycarpou MM. 2011. Distributed fault detection and isolation of large-scale discrete-time nonlinear systems: an adaptive approximation approach. *IEEE Trans. Autom. Control* 57(2):275–90
57. Boem F, Riverso S, Ferrari-Trecate G, Parisini T. 2018. Plug-and-play fault detection and isolation for large-scale nonlinear systems with stochastic uncertainties. *IEEE Trans. Autom. Control* 64(1):4–19
58. Halimi M, Millérioux G, Daafouz J. 2014. Model-based modes detection and discernibility for switched affine discrete-time systems. *IEEE Trans. Autom. Control* 60(6):1501–14
59. Küsters F, Trenn S. 2018. Switch observability for switched linear systems. *Automatica* 87:121–27
60. Mincarelli D, Pisano A, Floquet T, Usai E. 2016. Uniformly convergent sliding mode-based observation for switched linear systems. *Int. J. Robust Nonlinear Control* 26(7):1549–64
61. Zhang Z, Li S, Yan H, Fan Q. 2019. Sliding mode switching observer-based actuator fault detection and isolation for a class of uncertain systems. *Nonlinear Anal. Hybrid Syst.* 33:322–35
62. Dong J, Kolarijani AS, Mohajerin Esfahani P. 2023. Multimode diagnosis for switched affine systems with noisy measurement. *Automatica* 151:110898
63. Ghanipoor F, Murguia C, Mohajerin Esfahani P, van de Wouw N. 2025. Robust fault estimators for nonlinear systems: an ultra-local model design. *Automatica* 171:111920
64. Witczak M, Buciakowski M, Puig V, Rotondo D, Nejjari F. 2016. An LMI approach to robust fault estimation for a class of nonlinear systems. *Int. J. Robust Nonlinear Control* 26(7):1530–48
65. Zhu JW, Yang GH, Wang H, Wang F. 2016. Fault estimation for a class of nonlinear systems based on intermediate estimator. *IEEE Trans. Autom. Control* 61(9):2518–24
66. Xu A, Zhang Q. 2004. Nonlinear system fault diagnosis based on adaptive estimation. *Automatica* 40(7):1181–93
67. Shang C, Ye H, Huang D, Ding SX. 2022. From generalized Gauss bounds to distributionally robust fault detection with unimodality information. *IEEE Trans. Autom. Control* 68(9):5333–48

68. Harirchi F, Ozay N. 2018. Guaranteed model-based fault detection in cyber–physical systems: a model invalidation approach. *Automatica* 93:476–88
69. Yong SZ, Zhu M, Frazzoli E. 2016. A unified filter for simultaneous input and state estimation of linear discrete-time stochastic systems. *Automatica* 63:321–29
70. Hsieh CS. 2017. Unbiased minimum-variance input and state estimation for systems with unknown inputs: a system reformation approach. *Automatica* 84:236–40
71. Dong J, Pan K, Pequito S, Mohajerin Esfahani P. 2025. Robust multivariate detection and estimation with fault frequency content information. *Automatica* 173:112049
72. Tan J, Zheng H, Meng D, Wang X, Liang B. 2023. Active input design for simultaneous fault estimation and fault-tolerant control of LPV systems. *Automatica* 151:110903
73. Seliger R, Frank PM. 1991. Fault-diagnosis by disturbance decoupled nonlinear observers. In *Proceedings of the 30th IEEE Conference on Decision and Control*. IEEE
74. Li L, Ding SX, Zhong M, Peng K. 2025. Orthogonal projection-based fault detection for linear discrete-time varying systems. *IEEE Trans. Autom. Control* 70(5):3478–85
75. Xue T, Ding SX, Zhong M, Zhou D. 2022. An integrated design scheme for SKR-based data-driven dynamic fault detection systems. *IEEE Trans. Ind. Inform.* 18(10):6828–39
76. Jiang Q, Yan X, Huang B. 2015. Performance-driven distributed PCA process monitoring based on fault-relevant variable selection and Bayesian inference. *IEEE Trans. Ind. Electron.* 63(1):377–86
77. Yin S, Zhu X, Kaynak O. 2014. Improved PLS focused on key-performance-indicator-related fault diagnosis. *IEEE Trans. Ind. Electron.* 62(3):1651–58
78. Juricek BC, Seborg DE, Larimore WE. 2004. Fault detection using canonical variate analysis. *Ind. Eng. Chem. Res.* 43(2):458–74
79. Krishnan V, Pasqualetti F. 2020. Data-driven attack detection for linear systems. *IEEE Control Syst. Lett.* 5(2):671–76
80. Ding SX, Zhang P, Naik A, Ding EL, Huang B. 2009. Subspace method aided data-driven design of fault detection and isolation systems. *J. Process Control* 19(9):1496–510
81. Noom J, Soloviev O, Smith C, Verhaegen M. 2024. Online input design for discrimination of linear models using concave minimization. Preprint, arXiv:2401.05782v1 [eess.SY]
82. Dong J, Verhaegen M. 2009. Subspace based fault detection and identification for LTI systems. *IFAC Proc. Vol.* 42(8):330–35
83. Palanthandalam-Madapusi HJ, Bernstein DS. 2009. A subspace algorithm for simultaneous identification and input reconstruction. *Int. J. Adapt. Control Signal Process.* 23(12):1053–69
84. Hou J, Liu T, Wang QG. 2018. Recursive subspace identification subject to relatively slow time-varying load disturbance. *Int. J. Control* 91(3):622–38
85. Yu C, Verhaegen M. 2018. Data-driven fault estimation of non-minimum phase LTI systems. *Automatica* 92:181–87
86. Fattore G, Valcher ME. 2024. A data-driven approach to UIO-based fault diagnosis. In *2024 IEEE 63rd Conference on Decision and Control (CDC)*. IEEE
87. Liu T, Hou J, Qin SJ, Wang W. 2020. Subspace model identification under load disturbance with unknown transient and periodic dynamics. *J. Process Control* 85:100–11
88. Noom J, Soloviev O, Verhaegen M. 2024. Proximal-based recursive implementation for model-free data-driven fault diagnosis. *Automatica* 165:111656
89. Wan Y, Keviczky T, Verhaegen M. 2017. Fault estimation filter design with guaranteed stability using Markov parameters. *IEEE Trans. Autom. Control* 63(4):1132–39
90. Naderi E, Khorasani K. 2017. A data-driven approach to actuator and sensor fault detection, isolation and estimation in discrete-time linear systems. *Automatica* 85:165–78
91. Sheikhi MA, Mohajerin Esfahani P, Keviczky T. 2024. A kernel-based approach to data-driven actuator fault estimation. *IFAC-PapersOnLine* 58(4):318–23
92. Yan J, Markovsky I, Lygeros J. 2025. Secure data reconstruction: a direct data-driven approach. Preprint, arXiv:2502.00436v2 [eess.SY]
93. Wan Y, Keviczky T, Verhaegen M, Gustafsson F. 2016. Data-driven robust receding horizon fault estimation. *Automatica* 71:210–21

94. Gleizer GA. 2025. Output behavior equivalence and simultaneous subspace identification of systems and faults. *IEEE Control Syst. Lett.* 9:1285–90
95. Chen Z, Liang K, Ding SX, Yang C, Peng T, Yuan X. 2021. A comparative study of deep neural network-aided canonical correlation analysis-based process monitoring and fault detection methods. *IEEE Trans. Neural Netw. Learn. Syst.* 33(11):6158–72
96. Chen H, Liu Z, Huang B. 2023. Data-driven fault detection for Lipschitz nonlinear systems: from open to closed-loop systems. *Automatica* 155:111161
97. El-Koujok M, Benammar M, Meskin N, Al-Naemi M, Langari R. 2014. Multiple sensor fault diagnosis by evolving data-driven approach. *Inf. Sci.* 259:346–58
98. Kallas M, Mourot G, Maquin D, Ragot J. 2018. Data-driven approach for fault detection and isolation in nonlinear system. *Int. J. Adapt. Control Signal Process.* 32(11):1569–90
99. Bakhtiaridoust M, Irani FN, Yadegar M, Meskin N. 2023. Data-driven sensor fault detection and isolation of nonlinear systems: deep neural-network Koopman operator. *IET Control Theory Appl.* 17(2):123–32
100. Irani FN, Yadegar M, Meskin N. 2024. Koopman-based deep iISS bilinear parity approach for data-driven fault diagnosis: experimental demonstration using three-tank system. *Control Eng. Pract.* 142:105744
101. Hou J, Liu T, Wang QG. 2021. Subspace identification of Hammerstein-type nonlinear systems subject to unknown periodic disturbance. *Int. J. Control* 94(4):849–59
102. Liu L, Wang Z, Zhang H. 2017. Data-based adaptive fault estimation and fault-tolerant control for MIMO model-free systems using generalized fuzzy hyperbolic model. *IEEE Trans. Fuzzy Syst.* 26(6):3191–205
103. Willems JC, Rapisarda P, Markovsky I, De Moor BL. 2005. A note on persistency of excitation. *Syst. Control Lett.* 54(4):325–29
104. Ding SX. 2008. *Model-Based Fault Diagnosis Techniques: Design Schemes, Algorithms, and Tools*. Springer
105. Varga A. 2023. *Solving Fault Diagnosis Problems*. Springer
106. Mohajerin Esfahani P, Vrakopoulou M, Andersson G, Lygeros J. 2012. A tractable nonlinear fault detection and isolation technique with application to the cyber-physical security of power systems. In *2012 IEEE 51st Conference on Decision and Control*. IEEE
107. Polderman JW, Willems JC. 1998. *Introduction to Mathematical Systems Theory: A Behavioral Approach*. Springer
108. Willems JC. 1986. From time series to linear system—part I. Finite dimensional linear time invariant systems. *Automatica* 22(5):561–80
109. Chen J, Patton RJ. 2012. *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Springer
110. Patton RJ, Chen J. 1991. A review of parity space approaches to fault diagnosis. *IFAC Proc. Vol.* 24(6):65–81
111. Henrion D, Sebek M. 2000. An algorithm for polynomial matrix factor extraction. *Int. J. Control* 73(8):686–95
112. Isermann R. 2005. *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*. Springer
113. Wang J, Ye L, Gao RX, Li C, Zhang L. 2019. Digital twin for rotating machinery fault diagnosis in smart manufacturing. *Int. J. Prod. Res.* 57(12):3920–34
114. Campi MC, Garatti S. 2008. The exact feasibility of randomized solutions of uncertain convex programs. *SIAM J. Optim.* 19(3):1211–30
115. Mohajerin Esfahani P, Sutter T, Lygeros J. 2014. Performance bounds for the scenario approach and an extension to a class of non-convex programs. *IEEE Trans. Autom. Control* 60(1):46–58
116. Svetozarevic B, Mohajerin Esfahani P, Kamgarpour M, Lygeros J. 2013. A robust fault detection and isolation filter for a horizontal axis variable speed wind turbine. In *2013 American Control Conference*. IEEE
117. Habibi H, Howard I, Simani S. 2019. Reliability improvement of wind turbine power generation using model-based fault detection and fault tolerant control: a review. *Renew. Energy* 135:877–96
118. Mu B, Yang X, Scott JK. 2022. Comparison of advanced set-based fault detection methods with classical data-driven and observer-based methods for uncertain nonlinear processes. *Comput. Chem. Eng.* 166:107975

119. Qin SJ. 2012. Survey on data-driven industrial process monitoring and diagnosis. *Annu. Rev. Control* 36(2):220–34
120. Van Overschee P, De Moor B. 1994. N4SID: subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica* 30(1):75–93
121. Verhaegen M. 1994. Identification of the deterministic part of MIMO state space models given in innovations form from input-output data. *Automatica* 30(1):61–74
122. Chiuso A. 2007. The role of vector autoregressive modeling in predictor-based subspace identification. *Automatica* 43(6):1034–48
123. Ljung L. 1998. *System Identification*. Springer
124. Yin M, Iannelli A, Smith RS. 2021. Maximum likelihood estimation in data-driven modeling and control. *IEEE Trans. Autom. Control* 68(1):317–28
125. Ding SX. 2014. Data-driven design of monitoring and diagnosis systems for dynamic processes: a review of subspace technique based schemes and some recent results. *J. Process Control* 24(2):431–49
126. Kailath T. 1980. *Linear Systems*. Prentice Hall
127. Verhaegen M, Verdult V. 2007. *Filtering and System Identification: A Least Squares Approach*. Cambridge University Press
128. Van Overschee P, De Moor B. 2012. *Subspace Identification for Linear Systems: Theory—Implementation—Applications*. Springer
129. Disarò G, Valcher ME. 2024. On the equivalence of model-based and data-driven approaches to the design of unknown-input observers. *IEEE Trans. Autom. Control* 70(3):2074–81
130. Turan MS, Ferrari-Trecate G. 2021. Data-driven unknown-input observers and state estimation. *IEEE Control Syst. Lett.* 6:1424–29
131. Willsky A. 2003. On the invertibility of linear systems. *IEEE Trans. Autom. Control* 19(3):272–74
132. Kadam SD, Palanthandalam-Madapusi HJ. 2022. A note on invertibility and relative degree of MIMO LTI systems. *IFAC J. Syst. Control* 20:100193
133. Bokor J, Szabó Z. 2009. Fault detection and isolation in nonlinear systems. *Annu. Rev. Control* 33(2):113–23
134. Heirung TAN, Mesbah A. 2019. Input design for active fault diagnosis. *Annu. Rev. Control* 47:35–50
135. Markovsky I. 2018. *Low Rank Approximation*. Springer
136. Basile G, Marro G. 1992. *Controlled and Conditioned Invariants in Linear System Theory*. Prentice Hall
137. de Reij S. 2023. *Multivariate fault estimation for a high-fidelity model of the AB383 wire bonder: a model-based approach*. MS Thesis, Delft University of Technology
138. van de Wouw N. 2025. Digital Twin Research Project 4: technology health management. *Digital Twin*. **https://www.digital-twin-research.nl/research/research-project-4**
139. van Esch T, Ghanipoor F, Murguia C, van de Wouw N. 2024. Hybrid model-data fault diagnosis for wafer handler robots: tilt and broken belt cases. Preprint, arXiv:2412.09114v1 [cs.RO]
140. van Peijpe CD 2024. *Fault detection and isolation for high-end industrial printers: a hybrid model- and data-based approach*. MS Thesis, Delft University of Technology
141. Amini A. 2025. *Fault diagnosis and estimation in inkjet printers using self-sensing piezo actuators*. MS Thesis, Delft University of Technology
142. van Peijpe C, Ghanipoor F, de Loore Y, Hacking P, van de Wouw N, Mohajerin Esfahani P. 2024. Fault isolation for the ink deposition process in high-end industrial printers. Preprint, arXiv:2412.07545v1 [eess.SY]