Preprints of the 23rd IFAC World Congress
Busan, Republic of Korea, August 23-28, 2026

A General Framework for Design and Analysis of Optimal Fault Detection

Xinrui Gao [1], **Yuri A.W. Shardt** [1], **Bhushan Gopaluni** [2]

[1] *Department of Automation Engineering, Technical University of Ilmenau, Ilmenau, Thuringia, Germany, 98694*

[2] *Department of Chemical and Biological Engineering, University of British Columbia, Vancouver, BC V6T 1Z3, Canada.*

(e-mail: {xinrui.gao,yuri.shardt}@tu-ilmenau.de, bhushan.gopaluni@ubc.ca)

Abstract: Fault detection and isolation (FDI) have been extensively studied in control engineering and process monitoring, yet a unified theoretical framework connecting different approaches remains elusive. This paper presents a general framework for design and analysis of optimal fault detection (FD), which bridges paradigms that are traditionally separate. Starting from a measure-theoretic perspective, FD is formulated as a unified optimisation problem defined on general signal spaces that encompasses both stochastic and deterministic systems. The duality between two complementary formulations of the optimisation problem is analysed using Lagrangian relaxation to show the intrinsic connections and differences. Several cases of implementations of optimal FD design derived from the framework are also presented.

Keywords: **fault detection and isolation**; **duality theory**; **optimisation**.

1. INTRODUCTION

Fault detection and isolation (FDI) is vital for ensuring safety, reliability, and efficiency in modern industrial systems. Over the past decades, research has developed two methodological streams: model-based methods that use analytical redundancy through mathematical modelling (Ding, 2013), and data-driven approaches that extract features from historical process data (Yin *et al.*, 2014). Model-based methods, first developed using observer-based residual generation (Beard, 1971), have been extended with numerous residual generators and optimal design formulations (Ding, 2013). Recent work shows a unified approach across various system types for model-based fault-detection (FD) design based on least-squares estimation (Ding, 2018). It has also been shown that incorporating system dynamics in residual generators fundamentally improves detection by shrinking the deemed-normal region (Gao et al., 2025). Data-driven methods have been widely adopted in industrial practice (Yin et al., 2014). Recent advances in slow-feature analysis have shown promise for monitoring dynamic processes (Gao & Shardt, 2021).

Despite significant progress, theoretical gaps remain in unifying different types of approaches across different system assumptions. Model-based and data-driven methods take distinct paradigms in formulating and solving the problem with a limited exploration of their common ground. Moreover, many methods are designed based on different heuristic formulations according to specific system assumptions, e.g., deterministic or probabilistic, linear or nonlinear, dynamic or static, time-invariant or time-varying, and the fundamental connections between these formulations remain unclear. In addition, many approaches, especially data-driven ones, treat modelling and decision-making as separate steps, which could lead to suboptimal overall performance. Therefore, there is a need to address these gaps by showing a unified framework for optimal FD that encompasses both probabilistic and deterministic systems. The resulting formulations constitute well-posed optimisation problems with clear performance metrics, whose structural connections and differences are investigated using Lagrangian analysis. The framework provides a theoretical basis for systematically deriving, comparing, and unifying existing FD methods.

Thus, this paper seeks to address the gaps by developing a unified measure-theoretic framework for optimal fault detection that encompasses both probabilistic and deterministic systems. The resulting formulations, which are well-defined optimisation problems, possess clear performance metrics. The connections and differences between the two complementary problem formulations is analysed using Lagrangian relaxation. The detailed properties of the framework in such systems are presented.

2. SYSTEM DESCRIPTION AND PROBLEM FORMULATION

This section introduces a general system description and defines the unified framework for optimal fault detection. The relationship between the two basic formulations is analysed.

2.1 General Descriptions for Systems and Faults

Consider a dynamic system of the general form

$$\boldsymbol{y} = \mathcal{S}\left(\boldsymbol{u}, \boldsymbol{d}, \boldsymbol{f}\right) \tag{1}$$

where $\boldsymbol{y} \in \mathbb{D}_y$, $\boldsymbol{u} \in \mathbb{D}_u$, $\boldsymbol{d} \in \mathbb{D}_d$, and $\boldsymbol{f} \in \mathbb{D}_f$ are, respectively, the output, input, disturbance, and fault signals in corresponding domains, and $\mathcal{S}$: $\mathbb{D}_u \times \mathbb{D}_d \times \mathbb{D}_f \to \mathbb{D}_y$ is the

©2026 the authors. Accepted by IFAC for publication under a Creative Commons Licence CC-BY-NC-ND

system map that can be regarded as a mathematical operator between signal spaces. The exact form of the map and the type of signals vary depending on the system, e.g., in linear static systems, the signals can be deterministic or random vectors of finite dimensions, and the system map would be a constant matrix of proper dimensions. For stable linear time-invariant (LTI) systems, the system map can be transfer functions in a $\mathcal{H}_\infty$ space (Zhou & Doyle, 1998) or impulse responses in a Lebesgue space $\mathcal{L}_1$ (Desoer & Vidyasagar, 1975), while the signals are deterministic functions of time/frequency or stochastic processes.

The definition of two types of faults is given below.

Definition 1. The fault $\boldsymbol{f}$ is called an additive fault for the system $\mathcal{S}$ if the output can be decomposed into a sum of the disturbance and the fault components, that is,

$$\boldsymbol{y} = \mathcal{S}(\boldsymbol{u}, \boldsymbol{d}, \boldsymbol{f}) = \mathcal{S}_d(\boldsymbol{d}) + \mathcal{S}_f(\boldsymbol{f}) \tag{2}$$

where $\mathcal{S}_d$ and $\mathcal{S}_f$ are the maps from, respectively, the disturbance and fault domains to the system output domain.

Definition 2. The fault $\boldsymbol{f}$ is called a multiplicative fault for the system $\mathcal{S}$ if the output can be decomposed into a product of the disturbance and the fault components, that is,

$$\boldsymbol{y} = \mathcal{S}(\boldsymbol{u}, \boldsymbol{d}, \boldsymbol{f}) = \mathcal{S}_d(\boldsymbol{d})\mathcal{S}_f(\boldsymbol{f}) \tag{3}$$

The product in (3) is interpreted in the sense of the operator $\mathcal{S}$, *e.g.*, elementwise or convolution, which depends on the system class. The effect of $\boldsymbol{u}$ on $\boldsymbol{y}$ is not explicitly considered in (2) and (3) because it can always be absorbed into $\mathcal{S}_d$ and $\mathcal{S}_f$.

It is worth noting that there is no universally accepted definition for additive and multiplicative faults. Different definitions can be given, especially when considering different systems, for example, in Ding (2013), multiplicative faults in a linear system refer to the malfunctions that cause undesired changes in parameters of the plant, sensors, or actuators, while additive faults have a direct additive influence on system states or outputs. This definition makes physical sense. When using a state-space description, it is clear that multiplicative faults in this definition cause *multiplicative* effects to the plant, actuators, and sensors (see Equations (3.38) and (3.39) in Ding (2013)). As well, it is nice that additive faults of this definition introduce a clearly decoupled *additive* term to the nominal input-output description of the system (see Equation (3.31) in Ding (2013)). However, this definition focuses more on the physical components of the system rather than on the nature of faults, which can lead to a lack of clarity in the analysis, for example, with specific systems, the two types of faults have the same form of mathematical description after some re-arrangements, which limits the analysis of the theoretical properties of the FD methods.

2.2 Problem Formulation and Performance Assessment

This subsection introduces two formulations of the optimal FD framework to analyse the connections and differences.

2.2.1 A General View of Fault Detection

Consider a potential fault $\boldsymbol{f}$ that needs to be detected, so that the system in Equation (1) satisfies

$$\text{fault-free} \Leftrightarrow \boldsymbol{f} = 0; \text{faulty} \Leftrightarrow \boldsymbol{f} \neq 0 \tag{4}$$

Then, from the perspective of online implementation, FD maps a measurement $(\boldsymbol{u}, \boldsymbol{y})$ of the system to a decision space such that the decision is either that the system is faulty or that it is fault-free. Clearly, this includes two steps: mapping and decision making. The mapping step builds a decision space using some function $\mathcal{J}$

$$\mathbb{D}_J = \left\{ J \middle| J = \mathcal{J}(\boldsymbol{u}, \boldsymbol{y}), (\boldsymbol{u}, \boldsymbol{y}) \in \mathbb{D}_m \right\} \tag{5}$$

where the measurement space

$$\mathbb{D}_m = \left\{ (\boldsymbol{u}, \boldsymbol{y}) \middle| \boldsymbol{y} = \mathcal{S}(\boldsymbol{u}, \boldsymbol{d}, \boldsymbol{f}) \right\} \tag{6}$$

reduces to different subspaces in the fault-free and faulty cases, that is,

$$\mathbb{D}_{m,n} = \left\{ (\boldsymbol{u}, \boldsymbol{y}) \middle| \boldsymbol{y} = \mathcal{S}(\boldsymbol{u}, \boldsymbol{d}, \boldsymbol{f} = 0) \right\} \tag{7}$$

$$\mathbb{D}_{m,f} = \left\{ (\boldsymbol{u}, \boldsymbol{y}) \middle| \boldsymbol{y} = \mathcal{S}(\boldsymbol{u}, \boldsymbol{d}, \boldsymbol{f} \neq 0) \right\} \tag{8}$$

Then, the decision is made in space $\mathbb{D}_J$ based on the logic

$$J \in \mathbb{D}_{J,0} \Rightarrow \text{no alarm}; J \in \mathbb{D}_{J,1} \Rightarrow \text{alarm} \tag{9}$$

where the nonalarm region $\mathbb{D}_{J,0}$ and the alarm region $\mathbb{D}_{J,1}$ are two complementary subsets of $\mathbb{D}_J$, that is,

$$\mathbb{D}_{J,0} = \mathbb{D}_{J,1}^c, \mathbb{D}_{J,0} \bigcup \mathbb{D}_{J,1} = \mathbb{D}_J \tag{10}$$

In multivariate statistical process control, the mapping $\mathcal{J}$ is called the test statistic or evaluation function (Ding, 2021), and is often required to be positive definite (Gao et al., 2025) or in class $\mathcal{K}_\infty$ (*i.e.*, strictly increasing and $\mathcal{J}(0) = 0$) if $\mathcal{J}$ is defined on some norm of a vector (Esfahani & Lygeros, 2015). This will simplify the problem and lead to a decision space with the desired properties for which it is easier to make a decision. In FD methods based on machine learning, $\mathcal{J}$ can be flexible and the resulting J is called the feature (Gao et al., 2020).

Due to the properties of signals $\boldsymbol{d}$ and $\boldsymbol{f}$, *e.g.*, overlapping of their domains, and the mapping effect of $\mathcal{S}$, $\mathbb{D}_{m,n}$ and $\mathbb{D}_{m,f}$ are not necessarily disjoint. They are very likely to overlap. This is one of source of the complexity of FD problems, for example, if $\mathbb{D}_{m,n}$ and $\mathbb{D}_{m,f}$ are not separable for any $\mathcal{J}$, FD has to be a trade-off between missed alarms and false alarms.

2.2.2 Performance Assessment and Optimal Designs

To implement an FD system as shown in (5) and (9), one needs to find the mapping $\mathcal{J}$ and determine the nonalarm region $\mathbb{D}_{J,0}$ in the first place. This is the task of system design for fault detection. To obtain the optimal FD solution, the two steps of finding $\mathcal{J}$ and $\mathbb{D}_{J,0}$ need to be considered simultaneously rather than separately. This section introduces a general framework for performance assessment, and then, based on it,

defines unified formulations for the problem of optimal FD design.

Corresponding to the measurement space of the fault-free and faulty cases shown in (7) and (8), the decision space $\mathbb{D}_J$ reduces to, respectively,

$$\mathbb{D}_J \to \mathbb{D}_{J,n} = \left\{ J_0 \middle| J_0 = \mathcal{J}\left(\boldsymbol{u},\boldsymbol{y}\right), \left(\boldsymbol{u},\boldsymbol{y}\right) \in \mathbb{D}_{m,n} \right\} \quad (11)$$

$$\mathbb{D}_J \to \mathbb{D}_{J,f} = \left\{ J_1 \middle| J_1 = \mathcal{J}\left(\boldsymbol{u},\boldsymbol{y}\right), \left(\boldsymbol{u},\boldsymbol{y}\right) \in \mathbb{D}_{m,f} \right\} \quad (12)$$

Then, the sets of decisions for fault-free and faulty cases are defined to characterise the FD performance.

Definition 3. Set of false alarms (false positive)

$$\mathbb{D}_{fp} = \left\{ J \middle| J = \mathcal{J}\left(\boldsymbol{u},\boldsymbol{y}\right) \in \mathbb{D}_{J,1}, \left(\boldsymbol{u},\boldsymbol{y}\right) \in \mathbb{D}_{m,n} \right\} \quad (13)$$

Definition 4. Set of missed alarms (false negative)

$$\mathbb{D}_{fn} = \left\{ J \middle| J = \mathcal{J}\left(\boldsymbol{u},\boldsymbol{y}\right) \in \mathbb{D}_{J,0}, \left(\boldsymbol{u},\boldsymbol{y}\right) \in \mathbb{D}_{m,f} \right\} \quad (14)$$

Definition 5. Set of true alarms (true positive)

$$\mathbb{D}_{tp} = \left\{ J \middle| J = \mathcal{J}\left(\boldsymbol{u},\boldsymbol{y}\right) \in \mathbb{D}_{J,1}, \left(\boldsymbol{u},\boldsymbol{y}\right) \in \mathbb{D}_{m,f} \right\} \quad (15)$$

It follows from (11) to (15) that

$$\mathbb{D}_{fp} = \mathbb{D}_{J,1} \bigcap \mathbb{D}_{J,n} \quad (16)$$

$$\mathbb{D}_{fn} = \mathbb{D}_{J,0} \bigcap \mathbb{D}_{J,f} \quad (17)$$

$$\mathbb{D}_{tp} = \mathbb{D}_{J,1} \bigcap \mathbb{D}_{J,f} \quad (18)$$

Since $\mathbb{D}_{tp}$ and $\mathbb{D}_{fn}$ are two complementary subsets of $\mathbb{D}_{J,f}$, either of them is sufficient to assess FD performance for the faulty case. For the same reason, FD performance for the fault-free case can be assessed using $\mathbb{D}_{fp}$. Define measures on the decision space of fault-free and faulty cases

$$\mathcal{M}_n : \mathbb{D}_{J,n} \mapsto \mathbb{A}, \mathbb{A} \subseteq \left[0, +\infty\right) \quad (19)$$

$$\mathcal{M}_f : \mathbb{D}_{J,f} \mapsto \mathbb{B}, \mathbb{B} \subseteq \left[0, +\infty\right) \quad (20)$$

Then, the "size" of the three sets defined by (13) to (15) is reflected by the value of their measure, which is used to assess the corresponding aspects of FD performance.

From the perspective of triggering alarms, the problem of optimal FD can be formulated as finding the mapping $\mathcal{J}$ and the alarm region $\mathbb{D}_{J,1}$ to maximise true alarms while constraining false alarms to be no larger than a tolerance limit α, that is,

$$\left\{ \mathcal{J}, \mathbb{D}_{J,1} \right\} = \arg\max_{J, \mathbb{D}_{J,1}} \mathcal{M}_f\left(\mathbb{D}_{tp}\right), \text{ s.t. } \mathcal{M}_n\left(\mathbb{D}_{fp}\right) \le \alpha \quad (21)$$

From the perspective of reducing alarms, the optimal FD can also be formulated as finding the mapping $\mathcal{J}$ and nonalarm region $\mathbb{D}_{J,0}$ that minimises the false alarms while constraining missed alarms to be no larger than a tolerance limit β, that is,

$$\left\{ \mathcal{J}, \mathbb{D}_{J,0} \right\} = \arg\min_{J, \mathbb{D}_{J,0}} \mathcal{M}_n\left(\mathbb{D}_{fp}\right), \text{ s.t. } \mathcal{M}_f\left(\mathbb{D}_{fn}\right) \le \beta \quad (22)$$

Remark 1. Problems (21) and (22) are two parameterisations of the same FDI problem that is actually a multi-objective optimisation (MOO) problem. The tolerance parameters α and β in (21) and (22) are independently specified design constraints. However, once the optimal solution to (21) is obtained under a given α, it induces the tightest β^*. Conversely, solving (22) under a given β induces the tightest α^*. This correspondence could be useful when specifying the limits in practice, and the induced pair (α, β) of both problems theoretically traces the same Pareto front of the MOO problem.

2.2.3 Theoretical Analysis of the Two General Formulations

Problems (21) and (22) give two general formulations of the problem of optimal FD design. They are symmetrical in both physical interpretations and mathematical structures, which becomes apparent if we rewrite (21) in its equivalent form

$$\left\{ \mathcal{J}, \mathbb{D}_{J,1} \right\} = \arg\min_{J, \mathbb{D}_{J,1}} \mathcal{M}_f\left(\mathbb{D}_{fn}\right), \text{ s.t. } \mathcal{M}_n\left(\mathbb{D}_{fp}\right) \le \alpha \quad (23)$$

Hence, it is natural to be curious about their connections in, for example, the structure of the problem and the relationship of the resulting optimal solutions.

Define the Lagrangian functions associated with Problems (22) and (23), as, respectively

$$L_0\left(\mathcal{J}, \mathbb{D}_{J,0}, \lambda_0\right) = \mathcal{M}_n\left(\mathbb{D}_{fp}\right) + \lambda_0\left(\mathcal{M}_f\left(\mathbb{D}_{fn}\right) - \beta\right) \quad (24)$$

$$L_1\left(\mathcal{J}, \mathbb{D}_{J,1}, \lambda_1\right) = \mathcal{M}_f\left(\mathbb{D}_{fn}\right) + \lambda_1\left(\mathcal{M}_n\left(\mathbb{D}_{fp}\right) - \alpha\right) \quad (25)$$

where λ_0 and λ_1 are Lagrangian multipliers that are non-negative. The infimum of L_0 and L_1 gives the Lagrangian dual function, that is,

$$\theta_0\left(\lambda_0\right) = \inf_{\left(\mathcal{J}, \mathbb{D}_{J,0}\right)} L_0\left(\mathcal{J}, \mathbb{D}_{J,0}, \lambda_0\right) \quad (26)$$

$$\theta_1\left(\lambda_1\right) = \inf_{\left(\mathcal{J}, \mathbb{D}_{J,1}\right)} L_1\left(\mathcal{J}, \mathbb{D}_{J,1}, \lambda_1\right) \quad (27)$$

By definition, the dual function is a lower bound of the Lagrangian function, and thus, also that of the loss function in the primal problem ((22) and (23)). Hence, the optimal solution of Problems (22) and (23) can be approximated by the largest lower bound over the domain of Lagrangian multipliers, that is,

$$\max_{\lambda_0 \ge 0} \theta_0\left(\lambda_0\right) \quad (28)$$

$$\max_{\lambda_1 \ge 0} \theta_1\left(\lambda_1\right) \quad (29)$$

Problem (28) (respectively, (29)) is called the dual problem of the primal problem (22) ((23)) (Boyd & Vandenberghe, 2004). It has been proven that the optimal value of the dual problem is always no larger than that of the primal problem, that is,

$$d_0^* \le p_0^*, d_1^* \le p_1^* \quad (30)$$

where p_0^*, p_1^*, d_0^*, and d_1^* are, respectively, the optimal values of Problems (22), (23), (28), and (29). Inequality (30) is called weak duality, while strong duality is the situation where equality always holds (Boyd & Vandenberghe, 2004).

a. The Two Formulations are Distinct

Even though the two Lagrangian Functions (24) and (25) seem to be symmetric in a certain way, they are far from equivalent. Since they are defined on different domains $\mathbb{D}_{J,f}$ and $\mathbb{D}_{J,n}$, $\mathcal{M}_n$ and $\mathcal{M}_f$ are different measures. Even if we let $\lambda_0 = \lambda_1$ and $\alpha = \beta$, the two functions are still distinct. Hence, the formulations given by (21) and (22) define two essentially different optimisation problems.

Under some strict conditions, *e.g.*, $\lambda_0 = \lambda_1 = 1$ and $\alpha = \beta$, it follows from (24) and (25) that the Lagrangian functions are the same, and the dual functions have the same value. Namely, the two loss functions in Problems (21) and (22) share the same lower bound, but not necessarily the same infimum.

b. The Two Formulations have Connections

It is shown that a general relationship between Lagrangian functions in (24) and (25) exists, that is,

$$\begin{aligned} L_1\left(\mathcal{J},\mathbb{D}_{J,1},\lambda_1\right) &= \lambda_1\left(\frac{1}{\lambda_1}\mathcal{M}_f\left(\mathbb{D}_{fn}\right)+\mathcal{M}_n\left(\mathbb{D}_{fp}\right)-\alpha\right) \\ &= \lambda_1 L_0\left(\mathcal{J},\mathbb{D}_{J,0},\frac{1}{\lambda_1}\right)+\beta-\lambda_1\alpha \end{aligned} \tag{31}$$

Equation (31) further indicates a similar relationship between the dual functions, namely,

$$\theta_1\left(\lambda_1\right) = \lambda_1\theta_0\left(\frac{1}{\lambda_1}\right)+\beta-\lambda_1\alpha \tag{32}$$

This implies that the optimal FD design based on (21) by solving its Lagrangian relaxation problem (27) or its dual problem (29) can be translated into solving a problem related to those (Problem (26) or (28)) of (22), and *vice versa*.

However, this general relationship at the Lagrangian (dual) functions level does not mean that the optimal values (and the resulting optimal solutions) have the same correspondence as in (31). Let λ_0^* and λ_1^* be the optimal multipliers for the dual Problems (28) and (29). It follows that

$$\begin{aligned} d_1^* &= \theta_1\left(\lambda_1^*\right) = \lambda_1^*\theta_0\left(\frac{1}{\lambda_1^*}\right)+\beta-\lambda_1^*\alpha \\ &\le \lambda_1^*\theta_0\left(\lambda_0^*\right)+\beta-\lambda_1^*\alpha = \lambda_1^* d_0^*+\beta-\lambda_1^*\alpha \end{aligned} \tag{33}$$

Equality holds if $\lambda_0^* = 1/\lambda_1^*$. From (24) and (25), the value of Lagrangian functions can be interpreted as an overall extent of the detection errors of an FD system. Hence, Inequality (33) can be used to compare which formulation will result in a better overall FD solution. If strong duality holds, it also follows that for the primal problems

$$p_1^* \le \lambda_1^* p_0^* + \beta - \lambda_1^*\alpha \tag{34}$$

3. A GENERAL FRAMEWORK FOR PROBABILISTIC AND DETERMINISTIC PROCESSES

This section examines the optimal FD design for probabilistic and deterministic systems based on the general formulations given in the last section.

3.1 Probabilistic Systems

In the probabilistic setting, signals $\boldsymbol{u}$, $\boldsymbol{d}$, $\boldsymbol{f}$, and $\boldsymbol{y}$ are stochastic processes, which leads to a probability decision space $\mathbb{D}_J$. As a result, fault detection based on the decision logic (9) is essentially a hypothesis test with "fault-free" being the null hypothesis and "faulty" the alternative one (Ding, 2021). The Probability Measures (19) and (20) on the three decision subsets $\mathbb{D}_{fp}$, $\mathbb{D}_{fn}$, and $\mathbb{D}_{tp}$ correspond to, respectively, the false alarm rate (FAR), the missed detection rate (MDR), and the fault detection rate (FDR).

Definition 6. False alarm rate

$$\mathrm{FAR} = \mathcal{M}_n\left(\mathbb{D}_{fp}\right) = \Pr\left\{\mathcal{J}\left(\boldsymbol{u},\boldsymbol{y}\right)\in\mathbb{D}_{J,1}, \forall\left(\boldsymbol{u},\boldsymbol{y}\right)\in\mathbb{D}_{m,n}\right\} \tag{35}$$

Definition 7. Missed detection rate

$$\mathrm{MDR} = \mathcal{M}_f\left(\mathbb{D}_{fn}\right) = \Pr\left\{\mathcal{J}\left(\boldsymbol{u},\boldsymbol{y}\right)\in\mathbb{D}_{J,0}, \forall\left(\boldsymbol{u},\boldsymbol{y}\right)\in\mathbb{D}_{m,f}\right\} \tag{36}$$

Definition 8. Fault detection rate

$$\mathrm{FDR} = \mathcal{M}_f\left(\mathbb{D}_{tp}\right) = \Pr\left\{\mathcal{J}\left(\boldsymbol{u},\boldsymbol{y}\right)\in\mathbb{D}_{J,1}, \forall\left(\boldsymbol{u},\boldsymbol{y}\right)\in\mathbb{D}_{m,f}\right\} \tag{37}$$

Therefore, the problem of optimal FD design in (21) and (22) becomes an optimisation problem with a chance constraint on FAR and MDR, respectively

$$\left\{\mathcal{J},\mathbb{D}_{J,1}\right\} = \arg\max_{\mathcal{J},\mathbb{D}_{J,1}} \mathrm{FDR},\ \text{s.t. FAR} \le \alpha \tag{38}$$

$$\left\{\mathcal{J},\mathbb{D}_{J,0}\right\} = \arg\min_{J,\mathbb{D}_{J,0}} \mathrm{FAR},\ \text{s.t. MDR} \le \beta \tag{39}$$

3.2 Deterministic Systems

In deterministic systems, assume that the disturbance is a deterministic signal with a finite bound δ_d. Then, the measurement space is a subset of the general one defined in Equation (6), that is,

$$\mathbb{D}_m = \left\{\left(\boldsymbol{u},\boldsymbol{y}\right)\middle|\boldsymbol{y}=\mathcal{S}\left(\boldsymbol{u},\boldsymbol{d},\boldsymbol{f}\right), \|\boldsymbol{d}\|\le\delta_d\right\} \tag{40}$$

where $\|\boldsymbol{d}\|$ is a norm of $\boldsymbol{d}$. As well, according to the presence of faults, the two subspaces of $\mathbb{D}_m$ are

$$\mathbb{D}_{m,n} = \left\{\left(\boldsymbol{u},\boldsymbol{y}\right)\middle|\boldsymbol{y}=\mathcal{S}\left(\boldsymbol{u},\boldsymbol{d},\boldsymbol{f}=0\right), \|\boldsymbol{d}\|\le\delta_d\right\} \tag{41}$$

$$\mathbb{D}_{m,f} = \left\{\left(\boldsymbol{u},\boldsymbol{y}\right)\middle|\boldsymbol{y}=\mathcal{S}\left(\boldsymbol{u},\boldsymbol{d},\boldsymbol{f}\ne 0\right), \|\boldsymbol{d}\|\le\delta_d\right\} \tag{42}$$

Lemma 1. *For a stable, deterministic system with a bounded input, there exists an FD solution* $(\mathcal{J}, \mathbb{D}_{J,0})$ *that raises no false alarms if*

$$\mathbb{D}_{J,n} \subseteq \mathbb{D}_{J,0} \tag{43}$$

Proof. First, we will show there exists a detection solution ($\mathcal{J}$, $\mathbb{D}_{J,0}$) that satisfies Condition (43). The bounded input (including the disturbance and the input signals) indicates a bounded measurement space $\mathbb{D}_{m,n}$ for the fault-free case if the system is stable. Hence, given some map $\mathcal{J}$, *e.g.*, a locally bounded function, the decision space $\mathbb{D}_{J,n}$, which is defined in (11), is also bounded. As a result, one can always find a nonalarm space $\mathbb{D}_{J,0}$ that contains the full decision space $\mathbb{D}_{J,n}$ of the fault-free case. Therefore, the FD solution ($\mathcal{J}$, $\mathbb{D}_{J,0}$) satisfying Condition (43) exists. Second, based on (16) and the fact that $\mathbb{D}_{J,0}$ and $\mathbb{D}_{J,1}$ are complementary, Condition (43) implies that $\mathbb{D}_{fp}$ is empty, *i.e.*, there are no false alarms. Thus, Lemma 1 is proven. *Q.E.D.*

Definition 9. Set of undetectable faults (Ding, 2021)

$$\mathbb{D}_{f,\text{undetc}} = \left\{ \boldsymbol{f} \neq 0 \middle| \left(\boldsymbol{u}, \boldsymbol{y}\left(\boldsymbol{u}, \boldsymbol{d}, \boldsymbol{f}\right)\right) \in \mathbb{D}_{m,n}, \forall \left\| \boldsymbol{d} \right\| \leq \delta_d \right\} \quad (44)$$

To be more specific, Equation (44) means that

$$\begin{aligned} &\forall \boldsymbol{f} \in \mathbb{D}_{f,\text{undetc}}, \exists \left(\boldsymbol{u}, \boldsymbol{y}\left(\boldsymbol{u}, \boldsymbol{d}\right)\right) \in \mathbb{D}_{m,n} \\ &\text{s.t. } \boldsymbol{y}\left(\boldsymbol{u}, \boldsymbol{d}\right) = \boldsymbol{y}\left(\boldsymbol{u}, \boldsymbol{d}, \boldsymbol{f}\right) \end{aligned} \quad (45)$$

This implies that the faults in $\mathbb{D}_{f,\text{undetc}}$ cause no changes distinct from disturbance effects to the system, and thus, cannot be reliably detected regardless of what design is used for FD. In other words, these faults can only trigger alarms if false alarms are allowed, *i.e.*, if Condition (43) does not hold.

Theorem 1. *Let the image domain of the undetectable faults be*

$$\mathbb{D}_{m,\text{undetc}} = \left\{ \left(\boldsymbol{u}, \boldsymbol{y}\right) \middle| \boldsymbol{y} = \mathcal{S}\left(\boldsymbol{u}, \boldsymbol{d}, \boldsymbol{f}\right), \boldsymbol{f} \in \mathbb{D}_{f,\text{undetc}}, \left\| \boldsymbol{d} \right\| \leq \delta_d \right\} \quad (46)$$

and the corresponding decision set

$$\mathbb{D}_{J,\text{undetc}} = \left\{ J \middle| J = \mathcal{J}\left(\boldsymbol{u}, \boldsymbol{y}\right), \left(\boldsymbol{u}, \boldsymbol{y}\right) \in \mathbb{D}_{m,\text{undetc}} \right\} \quad (47)$$

Then, the optimal FD problem given by (21) *can be reformulated as*

$$\begin{aligned} &\left\{ \mathcal{J}, \mathbb{D}_{J,0} \right\} = \arg \min_{J, \mathbb{D}_{J,0}} \left(\mathcal{M}_f\left(\mathbb{D}_{fn}\right) - \mathcal{M}_f\left(\mathbb{D}_{J,\text{undetc}}\right)\right) \\ &\text{s.t. } \mathbb{D}_{J,n} \subseteq \mathbb{D}_{J,0} \end{aligned} \quad (48)$$

whose optimal solution ($\mathcal{J}$, $\mathbb{D}_{J,0}$) satisfies (i) $\mathcal{J}$ is injective on $\mathbb{D}_{m,f}$ and $\mathbb{D}_{m,n}$, and (ii) $\mathbb{D}_{J,0} = \mathbb{D}_{J,n}$.

Proof. From (42), it is clear that $\mathbb{D}_{m,\text{undetc}}$ is a subset of $\mathbb{D}_{m,f}$. On the other hand, by the definition in (44), it is also a subset of $\mathbb{D}_{m,n}$. Furthermore, according to (44) and (46), $\mathbb{D}_{m,\text{undetc}}$ contains all the measurements that are from the system having certain faults and could also be generated by the same system in the fault-free case. Namely, $\mathbb{D}_{m,\text{undetc}}$ contains all the common elements of $\mathbb{D}_{m,f}$ and $\mathbb{D}_{m,n}$, and thus, equals their intersection, that is,

$$\mathbb{D}_{m,\text{undetc}} = \mathbb{D}_{m,f} \bigcap \mathbb{D}_{m,n} \quad (49)$$

Therefore, for an arbitrary decision map $\mathcal{J}$, it follows that $\mathbb{D}_{J,\text{undetc}}$ is a subset of $\mathbb{D}_{J,n}$ and $\mathbb{D}_{J,f}$, and thus, a subset of their intersection. Supposing that Condition (43) holds, $\mathbb{D}_{J,\text{undetc}}$ is then a subset of $\mathbb{D}_{fn}$ according to Equation (17), that is,

$$\mathbb{D}_{J,\text{undetc}} \subseteq \left(\mathbb{D}_{J,n} \bigcap \mathbb{D}_{J,f} \right) \subseteq \left(\mathbb{D}_{J,0} \bigcap \mathbb{D}_{J,f} \right) = \mathbb{D}_{fn} \quad (50)$$

Hence, $\mathbb{D}_{J,\text{undetc}}$ is the smallest set of missed alarms that a FD method can possibly achieve under the constraint that no false alarms are raised.

Equation (50) indicates that the Optimal FD Problem (21) can be reformulated as minimising the difference between $\mathbb{D}_{fn}$ and $\mathbb{D}_{J,\text{undetc}}$ in the sense of the measure $\mathcal{M}_f$, while being subjected to Condition (43), that is,

$$\begin{aligned} &\left\{ \mathcal{J}, \mathbb{D}_{J,0} \right\} = \arg \min_{J, \mathbb{D}_{J,0}} \left(\mathcal{M}_f\left(\mathbb{D}_{fn}\right) - \mathcal{M}_f\left(\mathbb{D}_{J,\text{undetc}}\right)\right) \\ &\text{s.t. } \mathbb{D}_{J,n} \subseteq \mathbb{D}_{J,0} \end{aligned} \quad (51)$$

This formulation leads to an FD solution that has the smallest number of missed alarms (equivalently, the largest number of true alarms), while raising no false alarms. The optimal solution ($\mathcal{J}$, $\mathbb{D}_{J,0}$) is obtained only when the equalities in (50) hold, that is,

$$\mathbb{D}_{J,\text{undetc}} = \mathbb{D}_{J,n} \bigcap \mathbb{D}_{J,f} \quad (52)$$

$$\mathbb{D}_{J,0} = \mathbb{D}_{J,n} \quad (53)$$

Equation (52) further implies that

$$\mathcal{J}\left(\mathbb{D}_{m,f} \bigcap \mathbb{D}_{m,n} \right) = \mathcal{J}\left(\mathbb{D}_{m,f} \right) \bigcap \mathcal{J}\left(\mathbb{D}_{m,n} \right) \quad (54)$$

which will hold if and only if $\mathcal{J}$ is injective on $\mathbb{D}_{m,f}$ and $\mathbb{D}_{m,n}$ (Halmos, 1974, Chapter 10). *Q.E.D.*

4. IMPLEMENTATIONS OF THE GENERAL FRAMEWORK

The section presents several examples of optimal FD design for specific systems. In all cases, the threshold J_{th} is determined by the FAR constraint α as in Problem (21).

4.1 Static Systems with Additive Faults

Given a static system

$$\boldsymbol{y} = h_d\left(\boldsymbol{d}\right) + h_f\left(\boldsymbol{f}\right) \quad (55)$$

where $\boldsymbol{d}$ is a random disturbance, $\boldsymbol{f}$ is a deterministic fault vector, and h_d and h_f are general functions. According to the General Formulation (38) for probabilistic systems, fault detection for this system can be designed as (Gao et al., 2023)

$$\left\{ \mathcal{J}, \mathbb{D}_{J,1} \right\} = \arg \max_{\mathcal{J}, \mathbb{D}_{J,1}} \log \frac{\text{FDR}}{\text{FAR}} = \arg \max_{\mathcal{J}, \mathbb{D}_{J,1}} \frac{L\left(\boldsymbol{y} \middle| \boldsymbol{f} \neq 0\right)}{L\left(\boldsymbol{y} \middle| \boldsymbol{f} = 0\right)} \quad (56)$$

where L is the likelihood function of the system being identified as faulty. If the random component is normally

distributed, *i.e.*, $h_d(\boldsymbol{d}) \sim \mathcal{N}(0, \boldsymbol{\Sigma})$, then the resulting optimal FD solution for System (55) is

$$\begin{aligned} &\mathcal{J}^*(\boldsymbol{y}) = \boldsymbol{y}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{y}, \\ &\mathbb{D}^*_{J,1} = \left\{ J \middle| J = \mathcal{J}^*(\boldsymbol{y}) > J_{th}, J_{th} = \chi^2(k_m, \alpha) \right\} \end{aligned} \tag{57}$$

where k_m is the dimension of $\boldsymbol{y}$, which is also the degrees of freedom of the χ^2-distribution.

4.2 Gaussian Dynamic Systems

Consider a type of Gaussian dynamic systems

$$\boldsymbol{y} = \mathcal{S}(\boldsymbol{u}, \boldsymbol{d}, \boldsymbol{f}) \tag{58}$$

where $\boldsymbol{d}$ is Gaussian noise, and $\boldsymbol{u}$ and $\boldsymbol{f}$ are deterministic vectors for system inputs and potential faults. The commonly used method for this type of systems is first to design a residual generator so that the faults and the nominal uncertainties can be decoupled into an additive form, that is,

$$\boldsymbol{r} = \mathcal{R}(\boldsymbol{u}, \boldsymbol{y}) = \mathcal{R}_d(\boldsymbol{d}) + \mathcal{R}_f(\boldsymbol{f}) \tag{59}$$

where $\mathcal{R}_d$ and $\mathcal{R}_f$ are components of $\mathcal{R}$ for $\boldsymbol{d}$ and $\boldsymbol{f}$. If the generated residual is a zero-mean white Gaussian process for fault-free cases with a marginal distribution of $\mathcal{R}_d(t) \sim \mathcal{N}(0, \boldsymbol{\Sigma}_r(t))$ (*e.g.*, residual generators based on Kalman filter for Gaussian LTI systems (Gao et al., 2025)), then the same strategy as in (56) can be used. The resulting optimal FD solution resembles the one in Section 4.1, that is,

$$\begin{aligned} &\mathcal{J}^*(t) = \mathcal{R}(\boldsymbol{u}, \boldsymbol{y})^\top \boldsymbol{\Sigma}_r^{-1}(t) \mathcal{R}(\boldsymbol{u}, \boldsymbol{y}), \\ &\mathbb{D}^*_{J,1} = \left\{ J \middle| J = \mathcal{J}^*(t) > J_{th}, J_{th} = \chi^2(k_r, \alpha) \right\} \end{aligned} \tag{60}$$

where $\boldsymbol{\Sigma}_r$ is the covariance of $\boldsymbol{r}$, and k_r is the dimension of $\boldsymbol{r}$.

For this FD strategy, FD performance also depends on the residual generator that can be built based on system models or measurements. It is worth noting that a better residual generator, which normally refers to better performance of system modelling, does not necessarily mean better FD. A detailed analysis of how a good residual generator improves FD in Gaussian LTI systems is given in Gao et al. (2025).

4.3 LTI systems with Uncertain Distributions

For systems with uncertainties in disturbance distribution, the Optimal FD Problem (38) can be reformulated as a distributionally robust optimisation problem

$$\left\{ \mathcal{J}, \mathbb{D}_{J,1} \right\} = \arg\max_{\mathcal{J}, \mathbb{D}_{J,1}} \text{iFDR}, \text{ s.t. sFAR} \le \alpha \tag{61}$$

using the infimum of FDR and the supremum of FAR that are defined, respectively, as

$$\text{iFDR} = \inf_{P_\xi \in \mathbb{D}_f} P_\xi \left\{ \mathcal{J}(\boldsymbol{u}, \boldsymbol{y}) \in \mathbb{D}_{J,1}, (\boldsymbol{u}, \boldsymbol{y}) \in \mathbb{D}_{m,f} \right\} \tag{62}$$

$$\text{sFAR} = \sup_{P_\xi \in \mathbb{D}_n} P_\xi \left\{ \mathcal{J}(\boldsymbol{u}, \boldsymbol{y}) \in \mathbb{D}_{J,1}, (\boldsymbol{u}, \boldsymbol{y}) \in \mathbb{D}_{m,n} \right\} \tag{63}$$

where $\mathbb{D}_f$ and $\mathbb{D}_n$ are sets of distributions for the faulty and fault-free cases, respectively. Problem (39) can be formulated similarly. Using various ambiguity sets and residual generators, different specific formulations for optimal FD design can be obtained (Shang et al., 2021; Xue et al., 2020).

5. CONCLUSIONS

This paper has presented a unified framework for design and analysis of optimal FD. The theoretical connections and differences between the two basic formulations were investigated. The general framework of optimal FD for deterministic and probabilistic systems was derived and instantiated through three implementation cases covering static, Gaussian dynamic, and distribution-uncertain systems. The existence conditions of the optimal detection solution for deterministic systems were proven. The proposed framework provides theoretical guidance for designing and analysing optimal FD systems. Extension to more complicated systems, including numerical study comparing framework-derived solutions on benchmark processes, will be considered.

REFERENCES

Beard, R. V. (1971). *Failure accomodation in linear systems through self-reorganization*. Massachusetts Institute of Technology.

Boyd, S. P., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

Desoer, C. A., & Vidyasagar, M. (1975). *Feedback Systems: Input-Output Properties*. Academic Press.

Ding, S. X. (2013). *Model-based fault diagnosis techniques: design schemes, algorithms, and tools* (Second edi). Springer London.

Ding, S. X. (2018). Optimal fault detection and estimation: a unified scheme and least squares solutions. *IFAC-PapersOnLine*, *51*(24), 465–472.

Ding, S. X. (2021). *Advanced methods for fault diagnosis and fault-tolerant control*. Springer.

Esfahani, P. M., & Lygeros, J. (2015). A tractable fault detection and isolation approach for nonlinear systems with probabilistic performance. *IEEE Transactions on Automatic Control*, *61*(3), 633–647.

Gao, X., Hu, A., & Shardt, Y. A. W. (2025). How Dynamics Improve Fault Detection: A Gaussian LTI Case. *IFAC-PapersOnLine*, *59*(6), 121–126.

Gao, X., & Shardt, Y. A. W. (2021). Dynamic system modelling and process monitoring based on long-term dependency slow feature analysis. *Journal of Process Control*, *105*, 27–47.

Gao, X., Xie, J., & Shardt, Y. A. W. (2023). Concurrent Monitoring and Isolation of Static Deviations and Dynamic Anomalies with a Sparsity Constraint. *IFAC-PapersOnLine*, *56*(2), 11790–11795.

Gao, X., Yang, F., & Feng, E. (2020). A process fault diagnosis method using multi-time scale dynamic feature extraction based on convolutional neural network. *The Canadian Journal of Chemical Engineering*, *98*(6), 1280–1292.

Halmos, P. R. (1974). *Naive Set Theory*. Springer. https://doi.org/10.1007/978-1-4757-1645-0

Shang, C., Ding, S. X., & Ye, H. (2021). Distributionally robust fault detection design and assessment for dynamical systems. *Automatica*, *125*, 109434.

Wise, B. M., Ricker, N. L., Veltkamp, D. F., & Kowalski, B. R. (1990). A theoretical basis for the use of principal component models for monitoring multivariate processes. *Process Control and Quality*, *1*(1), 41–51.

Xue, T., Zhong, M., Luo, L., Li, L., & Ding, S. X. (2020). Distributionally robust fault detection by using Kernel density estimation. *IFAC-PapersOnLine*, *53*(2), 652–657.

Yin, S., Ding, S. X., Xie, X., & Luo, H. (2014). A review on basic data-driven approaches for industrial process monitoring. *IEEE Transactions on Industrial Electronics*, *61*(11), 6418–6428.

Zhou, K., & Doyle, J. C. (1998). *Essentials of robust control* (Vol. 104). Prentice hall Upper Saddle River, NJ.