Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

RESEARCH ARTICLE

Distributionally robust trade-off design of parity relation based fault detection systems

Yiming Wan*[1,2] | Yujia Ma[1] | Maiying Zhong[3]

[1]School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China
[2]Key Laboratory of Image Processing and Intelligent Control, Huazhong University of Science and Technology, Ministry of Education, China
[3]Shandong Key Laboratory of Big-data Driven Safety Control Technology for Complex Systems, College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao, China

Correspondence
*Yiming Wan, Email: ywan@hust.edu.cn

Abstract

The fault detection (FD) system design aims at optimizing the trade-off between false alarm rate (FAR) and fault detection rate (FDR) under stochastic disturbances or uncertainties. A challenging difficulty in practice is the inexact information of stochastic disturbance distribution, i.e., the actual distribution deviates from the one used in the design. To address this challenge, a distributionally robust optimization (DRO) approach that accounts for the inexact distribution information is proposed for the parity relation based FD of stochastic discrete-time linear systems. It introduces moment-based and entropy-based ambiguity sets to describe the inexact disturbance distribution. Over such ambiguity sets, the FD system design for a scalar residual maximizes the worst-case FDR with respect to a reference fault mode, while ensuring a predefined worst-case FAR. To address the limitation of a scalar residual, the FD test of a vector residual is constructed with respect to a parameterized set of multiple fault modes. The resulting FD tests can be expressed in the same structure as the celebrated generalized likelihood ratio test (GLRT), while only the detection threshold is adjusted to compensate for distribution ambiguity. Moreover, the worst-case FDR in the presence of any given fault is evaluated by solving another DRO problem. Using a continuous stirred tank reactor example with inexact distribution information, it is demonstrated that the proposed designs achieve desirable performance trade-off and provide effective worst-case FDR evaluations, whilst the GLRT fails to ensure the predefined FAR.

KEYWORDS:
Fault detection; false alarm rate; fault detection rate; distribution ambiguity; distributionally robust optimization

1 | INTRODUCTION

The importance of safety-critical systems has attracted tremendous research interests in model-based fault detection (FD), see [1,2,3] and references therein. A typical FD system is composed of a residual generator and a residual evaluator[2]. The generated residual captures the deviation between the observed system behavior and the nominal system model. Then the size of the residual is evaluated and compared against a chosen threshold to detect latent faults. The FD system design involves determining both the residual generator and the detection threshold. A typical residual generator is based on observers, filters or parity relations[2]. The parity relation approach derives an input-output model over a sliding window by decoupling the unknown initial state with a so-called parity vector or matrix[2,4].

The desirable FD performance is a high fault detection rate (FDR) and a low false alarm rate (FAR). Generally, FDR is increased at the cost of compromising FAR. Therefore, an optimal trade-off between FDR and FAR is preferred in the FD system design[2,5,6].

The fundamental research problem in model-based FD is to optimize the FD performance in the presence of model uncertainties including external disturbances or unknown parameters. Existing robust FD techniques can be classified into two categories, depending on the description of model uncertainties. The first category assumes unknown-but-bound uncertainties, and adopts a deterministic worst-case approach. Inspired by robust control concepts, one worst-case approach employs system norms to describe robustness to disturbances and sensitivity to faults. It first constructs a residual generator that minimizes a ratio between such disturbance robustness and fault sensitivity, and then determines a detection threshold using the worst-case upper bound of the residual norm[7,8,9,10,11]. Alternatively, the set-membership method exploits a set-valued observer[12,13,14,15,16], parity relation[17], or feasibility based model invalidation[18] to perform a set-based consistency test. In these aforementioned methods, zero FAR is ensured by accounting for the worst-case uncertainty, whilst the resulting FDR could be overly compromised. This is due to adopting a rather conservative detection threshold or set to account for the worst-case uncertainty that occurs with vanishingly low probability.

In the second category of robust FD techniques, model uncertainties are assumed to be stochastic. With uncertainty distribution that distinguishes between low and high probability of occurrence, a much less conservative detection threshold can be set by allowing an acceptable FAR, which results in an improved FDR. For linear systems with additive Gaussian disturbances, a generalized likelihood ratio test (GLRT) is proposed in the scheme of parity relation or Kalman filter[19,20,21]. In the observer-based FD scheme, a probabilistic relaxation of the norm-based threshold computation[22,23] or FAR-constrained optimal design[5,6] is proposed to exploit the probability distributions of parametric uncertainties and disturbances. Recently, the set-membership approach has been merged with the probabilistic paradigm in[24,25,26] by admitting a risk level in the set-based consistency test.

Despite the significant progress briefly reviewed above, robust FD still poses a remarkable challenge in the integration of a residual generator and a residual evaluator to optimize the trade-off between FAR and FDR. The main difficulty lies in constructing a tractable optimization problem that quantifies how FAR and FDR depend on design parameters. Most existing literature separately considers the residual generator design and threshold computation, which renders the overall FD performance sub-optimal. Towards the integrated trade-off design, several results have been reported in [27,28] for linear systems with additive unknown-but-bounded disturbances. The proposed approach minimizes the set of undetectable faults under the constraint of zero FAR. In the probabilistic paradigm, the GLRT is widely used due to its asymptotic optimality in terms of FAR and miss detection rate [19,29].

Another challenging difficulty comes from the inexactness of distribution information about stochastic uncertainties in practice [30,31]. All the aforementioned probabilistic FD designs commit to an assumed or estimated distribution of stochastic uncertainties, thus their performance could be sensitive to any deviation from the assumed or estimated distribution. With ambiguous distribution information, FAR and FDR cannot be directly quantified, thus the integrated trade-off design becomes even more challenging. As a recent progress towards this challenge, a probabilistic robust parity relation approach has been proposed in [32,33] by utilizing the minimum error minimax probability machine. It minimizes a weighted sum of FAR and miss detection rate in a worst-case setting, i.e., under all stochastic disturbances with a given mean and covariance matrix. Similarly, a distributionally robust approach is proposed in [34] to ensure the predefined worst-case FAR for disturbance distributions within a Wasserstein ball while maximizing fault sensitivity.

The above two challenges are addressed in this paper. The integrated trade-off design of the parity relation based FD is investigated for stochastic discrete-time linear systems. The inaccurate probability measure of disturbance distribution is described by the moment-based and entropy-based ambiguity sets which can be constructed with empirical data or domain-specific knowledge [30,31,35]. Over such ambiguity sets, a distributionally robust optimization (DRO) approach is proposed for the integrated trade-off design and the worst-case performance evaluation. Our main contributions include:

(i) The FD system design for a scalar residual is formulated as maximizing the worst-case FDR with respect to a reference fault mode, while being constrained by a predefined worst-case FAR, over the moment-based or entropy-based ambiguity set. With the obtained solution, the FD test of a vector residual is constructed with respect to a parameterized set of multiple fault modes. The derived FD test has the same structure as in the GLRT, but adjusts its threshold to compensate for distribution ambiguity.

(ii) With the considered two types of distribution ambiguity sets, DRO-based systematic methods are proposed for threshold computation that ensures the predefined worst-case FAR. Similarly, the worst-case FDR in the presence of any given fault

signal is evaluated by solving associated DRO problems. For the above threshold computation and FDR evaluation, closed-form expressions are derived under the moment-based distribution ambiguity, while efficient bisection search algorithms are developed to address the entropy-based distribution ambiguity.

The rest of this paper is organized as follows. In Section 2, the research motivation and problem description are stated. Then, the DRO approach is proposed in Sections 3 and 4 to address the integrated trade-off design, parametric tuning, and worst-case FDR evaluation subject to the moment-based and entropy-based ambiguity sets. Extensive simulation study and some concluding remarks are given in Sections 5 and 6.

Notations: We denote by $\mathbf{0}$ and $\mathbf{I}$ a zero matrix and an identity matrix of appropriate dimensions, respectively. The diagonal concatenation of matrices X_1 and X_2 is denoted by $\mathrm{diag}(X_1, X_2)$. $X^\dagger$ represents the pseudoinverse of a matrix X. For a symmetric matrix X, its positive definiteness and positive semidefiniteness are denoted by $X > 0$ and $X \geq 0$, respectively. For a vector $x \in \mathbb{R}^n$, its infinity norm and 2-norm are defined as $\|x\|_\infty = \max\limits_{1\leq i\leq n} |x_i|$ and $\|x\|_2 = \sqrt{x^\top x}$, respectively. A Gaussian distribution with mean μ and covariance $\Sigma > 0$ is represented by $\mathcal{N}(\mu, \Sigma)$. The standard normal cumulative distribution function (CDF) is denoted by $\Phi(\cdot)$, while the standard normal inverse CDF is denoted by $\Phi^{-1}(\cdot)$. χ^2_n represents a central Chi-square distribution with n degrees of freedom; and its CDF and inverse CDF are denoted by $\Gamma_n(\cdot)$ and $\Gamma_n^{-1}(\cdot)$, respectively. Similarly, $\chi^2_{n,\lambda}$ represents a non-central Chi-square distribution with n degrees of freedom and the noncentrality parameter λ; and its corresponding CDF and inverse CDF are $\Gamma_{n,\lambda}(\cdot)$ and $\Gamma^{-1}_{n,\lambda}(\cdot)$, respectively.

2 | PROBLEM STATEMENT

2.1 | System description

Consider the following linear time-invariant system

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + B_d d(k) + B_f f(k) \\ y(k) &= Cx(k) + Du(k) + D_d d(k) + D_f f(k) \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^{n_x}$, $u \in \mathbb{R}^{n_u}$, $y \in \mathbb{R}^{n_y}$, $d \in \mathbb{R}^{n_d}$, and $f \in \mathbb{R}^{n_f}$ are the state, the control input, the measured output, the stochastic disturbance, and the latent fault, respectively. The system matrices A, B, B_d, B_f, C, D, D_d, and D_f are known and time-invariant, with appropriate dimensions.

Assumption 1. Without loss of generality, assume that D_d is of full row rank, and the stochastic disturbance $d(k)$ have a zero mean.

From the system model (1), the stacked output equation over a time window $[k-h+1, k]$ is

$$\mathbf{y}_k = \mathbf{H}_o x(k-h+1) + \mathbf{H}_u \mathbf{u}_k + \mathbf{H}_d \mathbf{d}_k + \mathbf{H}_f \mathbf{f}_k, \tag{2}$$

where

$$\mathbf{y}_k = \begin{bmatrix} y(k-h+1) \\ y(k-h+2) \\ \vdots \\ y(k) \end{bmatrix}, \ \mathbf{H}_o = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{h-1} \end{bmatrix}, \ \mathbf{H}_u = \begin{bmatrix} D & \mathbf{0} & \cdots & \mathbf{0} \\ CB & D & \ddots & \vdots \\ \vdots & \vdots & \ddots & \mathbf{0} \\ CA^{h-2}B & CA^{h-3}B & \cdots & D \end{bmatrix}, \tag{3}$$

$\mathbf{u}_k$, $\mathbf{d}_k$, and $\mathbf{f}_k$ are constructed similarly to $\mathbf{y}_k$. With the same structure as in $\mathbf{H}_u$, $\mathbf{H}_d$ and $\mathbf{H}_f$ are defined by replacing (B, D) with (B_d, D_d) and (B_f, D_f), respectively.

2.2 | Brief review of GLRT for parity relation based FD

In this subsection, the celebrated GLRT approach is briefly reviewed, which will be compared against our proposed approaches.

Let $\mathbf{N}_o \in \mathbb{R}^{(n_y h - n_x) \times n_y h}$ denote the basis matrix of the left null space of the observability matrix $\mathbf{H}_o$, i.e., $\mathbf{N}_o \mathbf{H}_o = \mathbf{0}$ and $\text{rank}(\mathbf{N}_o) = n_y h - n_x$. Hence, a parity matrix $V_r \in \mathbb{R}^{(n_y h - n_x) \times n_y h}$ that ensures $V_r \mathbf{H}_o = \mathbf{0}$ can be expressed as

$$V_r = W \mathbf{N}_o \tag{4}$$

with W being a nonsingular matrix to be designed. According to (2) and (4), the parity relation approach generates a residual

$$\bar{r}_k = V_r(\mathbf{y}_k - \mathbf{H}_u \mathbf{u}_k)$$

which is governed by

$$\bar{r}_k = W z_k, \ z_k = \mathbf{N}_o(\mathbf{y}_k - \mathbf{H}_u \mathbf{u}_k) = \bar{\mathbf{H}}_d \mathbf{d}_k + \bar{\mathbf{H}}_f \mathbf{f}_k \tag{5}$$

with $\bar{\mathbf{H}}_d = \mathbf{N}_o \mathbf{H}_d$ and $\bar{\mathbf{H}}_f = \mathbf{N}_o \mathbf{H}_f$. In the above equation, z_k is the primary residual, which is transformed into $\bar{r}_k$ by the design matrix W. We define $z_{0,k}$ and $z_{f,k}$ as below to indicate z_k being generated in the fault-free and faulty cases, respectively:

$$\begin{aligned} z_{0,k} &= z_k|_{\mathbf{f}_k = 0} = \bar{\mathbf{H}}_d \mathbf{d}_k, \\ z_{f,k} &= z_k|_{\mathbf{f}_k \neq 0} = \bar{\mathbf{H}}_d \mathbf{d}_k + \bar{\mathbf{H}}_f \mathbf{f}_k. \end{aligned} \tag{6}$$

In the GLRT approach, the disturbance $d(k)$ is assumed to be a white Gaussian noise, with zero mean and covariance Q_d. Then, the covariance matrix of the stacked disturbance vector $\mathbf{d}_k$ is

$$\bar{\mathbf{Q}}_d = \text{diag}(Q_d, Q_d, \cdots, Q_d).$$

Since the stacked fault $\mathbf{f}_k$ is deterministic, the covariance of $z_{0,k}$ in the fault-free case is

$$\Sigma_0 = \bar{\mathbf{H}}_d \bar{\mathbf{Q}}_d \bar{\mathbf{H}}_d^\top, \tag{7}$$

which is the same as $z_{f,k}$ in the faulty case. Note that $\bar{\mathbf{H}}_d = \mathbf{N}_o\mathbf{H}_d$ has full row rank, and Σ_0 is nonsingular, since both $\mathbf{N}_o$ and $\mathbf{H}_d$ are of full row rank.

With the above notations, FD is seen as the hypothesis test

$$\mathscr{H}_0 : \ \bar{r}_k \sim \mathcal{N}(\mathbf{0}, W\Sigma_0 W^\top),$$
$$\mathscr{H}_1 : \ \bar{r}_k \sim \mathcal{N}(W\bar{\mathbf{H}}_f\mathbf{f}_k, W\Sigma_0 W^\top).$$

In the GLRT, the log-likelihood ratio is

$$\mathcal{L}(\bar{r}_k) = 2\ln \frac{\exp\left\{-\frac{1}{2}\left(\bar{r}_k - W\bar{\mathbf{H}}_f\mathbf{f}_k\right)^\top (W\Sigma_0 W^\top)^{-1}\left(\bar{r}_k - W\bar{\mathbf{H}}_f\mathbf{f}_k\right)\right\}}{\exp\left\{-\frac{1}{2}\bar{r}_k^\top (W\Sigma_0 W^\top)^{-1}\bar{r}_k\right\}}.$$

Due to (5) and the nonsingularity of W and Σ_0, the above ratio can be simplified as

$$\mathcal{L}(\bar{r}_k) = -\mathbf{f}_k^\top \bar{\mathbf{H}}_f^\top \Sigma_0^{-1}\bar{\mathbf{H}}_f\mathbf{f}_k + 2\mathbf{f}_k^\top \bar{\mathbf{H}}_f^\top \Sigma_0^{-1} z_k,$$

and it is maximized when $\mathbf{f}_k = (\bar{\mathbf{H}}_f^\top \Sigma_0^{-1}\bar{\mathbf{H}}_f)^\dagger \bar{\mathbf{H}}_f^\top \Sigma_0^{-1} z_k$. The corresponding maximum log-likelihood is

$$\mathcal{L}^*(\bar{r}_k) = z_k^\top \Sigma_0^{-1}\bar{\mathbf{H}}_f(\bar{\mathbf{H}}_f^\top \Sigma_0^{-1}\bar{\mathbf{H}}_f)^\dagger \bar{\mathbf{H}}_f^\top \Sigma_0^{-1} z_k. \tag{8}$$

Note that $\Sigma_0^{-\frac{1}{2}} z_k$ in the fault-free case follows the standard normal distribution, and $\mathcal{L}^*(\bar{r}_k)$ can be regarded as performing an orthogonal projection of $\Sigma_0^{-\frac{1}{2}} z_k$ onto the column space of $\Sigma_0^{-\frac{1}{2}}\bar{\mathbf{H}}_f$. Thus $\mathcal{L}^*(\bar{r}_k)$ follows the $\chi^2_{m_f}$ distribution whose degrees of freedom is

$$m_f = \operatorname{rank}(\Sigma_0^{-\frac{1}{2}}\bar{\mathbf{H}}_f). \tag{9}$$

Finally, the detection decision follows

$$\begin{cases} \mathcal{L}^*(\bar{r}_k) > \Gamma_{m_f}^{-1}(1-\gamma) \ \Rightarrow \ \text{fault alarm}, \\ \mathcal{L}^*(\bar{r}_k) \le \Gamma_{m_f}^{-1}(1-\gamma) \ \Rightarrow \ \text{no fault alarm}, \end{cases} \tag{10}$$

where $\Gamma_{m_f}^{-1}(\cdot)$ represents the $\chi^2_{m_f}$ inverse CDF, and γ denotes the predefined FAR. It is worth noting that the obtained GLRT does not depend on the selection of the nonsingular matrix W. The same GLRT for parity relation based FD was reported in [19], with a different derivation.

2.3 | Residual distribution ambiguity

In practice, any statistical FD method including GLRT has to address the issue that the disturbance distribution and the resulting residual distribution are typically unknown a priori. A basic idea is to estimate the primary residual covariance Σ_0 from data, such that the GLRT in (8) and (10) can be computed without knowing the disturbance covariance $\bar{\mathbf{Q}}_d$. Specifically, we first compute samples of $z_{0,k}$ using available fault-free input and output data according to (5), and then use the empirical covariance

of $z_{0,k}$ as an estimate of Σ_0. We assume the availability of a sufficient amount of fault-free data, such that the above estimate of Σ_0 is accurate. Hence Σ_0 is regarded as known in the rest of this paper. The residual mean is zero according to Assumption 1.

Although the residual covariance is accurately estimated from data, the underlying residual distribution is still unknown, which could be different from the Gaussian distribution assumed in the conventional GLRT. Moreover, a moderate data size for an accurate covariance estimate is often insufficient for a reliable estimation of residual distribution[36]. Therefore, the inexactness or ambiguity of residual distribution should be fully taken into account in a statistical FD test, which otherwise results in a poor FD performance. For instance, the discrepancy between the assumed Gaussian distribution and the actual one can lead to an unacceptable FAR in the conventional GLRT.

The ambiguity of probability distribution can be described by a set of possible distributions, called an ambiguity set $\mathbb{P}$. In this paper, two types of ambiguity sets for the inexact distribution of the primary residual z_k in (5) and (6) are constructed from data:

- The moment-based ambiguity set $\mathbb{P}_m(\mu, \Sigma)$ consists of all probability distributions with mean μ and covariance Σ. In this paper, with Σ_0 be estimated as the empirical covariance of the fault-free primary residual samples $z_{0,k}$, the distributions of $z_{0,k}$ and $z_{f,k}$ in (6) are assumed to belong to $\mathbb{P}_m(\mathbf{0}, \Sigma_0)$ and $\mathbb{P}_m(\bar{\mathbf{H}}_f \mathbf{f}_k, \Sigma_0)$ respectively, which are denoted by $z_{0,k} \sim \mathbb{P}_m(\mathbf{0}, \Sigma_0)$ and $z_{f,k} \sim \mathbb{P}_m(\bar{\mathbf{H}}_f \mathbf{f}_k, \Sigma_0)$.

- The entropy-based ambiguity set $\mathbb{P}_e(P_0, \eta)$ includes any distribution P whose KL divergence (i.e., negative relative entropy) with respect to a nominal distribution P_0 is less than or equal to a predefined nonnegative value η, i.e.,

$$\mathbb{P}_e(P_0, \eta) = \left\{ P \in \mathbb{D} | \mathrm{KL}(P, P_0) \leq \eta \right\}, \tag{11}$$

  where $\mathbb{D}$ denotes the set of all probability distributions, and

$$\mathrm{KL}(P, P_0) = \int \log \frac{dP}{dP_0} dP$$

  represents the KL divergence from P to P_0. In this paper, assume $z_{0,k} \sim \mathbb{P}_e(P_{z_0}, \eta)$ in the fault-free case, i.e., the true distribution Q_{z_0} of the fault-free primary residual $z_{0,k}$ satisfies $\mathrm{KL}(Q_{z_0}, P_{z_0}) \leq \eta$, with the nominal distribution P_{z_0} being $\mathcal{N}(\mathbf{0}, \Sigma_0)$. In the faulty case, let Q_{z_f} and $P_{z_f} = \mathcal{N}(\bar{\mathbf{H}}_f \mathbf{f}_k, \Sigma_0)$ represent the true distribution and nominal distribution of the faulty primary residual, respectively. According to (6), $Q_{z_f}(z) = Q_{z_0}(z - \bar{\mathbf{H}}_f \mathbf{f}_k)$ and $P_{z_f}(z) = P_{z_0}(z - \bar{\mathbf{H}}_f \mathbf{f}_k)$ hold, thus we have $\mathrm{KL}(Q_{z_f}, P_{z_f}) = \mathrm{KL}(Q_{z_0}, P_{z_0}) \leq \eta$. This implies that the divergence bound η is valid for both fault-free and faulty cases. Since the true distribution Q_{z_0} is not exactly known, we determine η using the sample distribution $\hat{Q}_{z_0}$ of the fault-free primary residual. Firstly, as an empirical estimate of $\mathrm{KL}(Q_{z_0}, P_{z_0})$, $\hat{\eta} = \mathrm{KL}(\hat{Q}_{z_0}, P_{z_0})$ is computed via Monte-Carlo simulation or k-nearest-neighbor methods[37]. Such an estimate is asymptotically normal distributed, with its bias c_N and variance δ_N^2 converging with the sample size N[37,38]. Therefore, we set $\eta = \hat{\eta} + c_N + 2\sigma_N$ such that $\mathrm{KL}(Q_{z_0}, P_{z_0}) \leq \eta$ holds for the unknown true distribution Q_{z_0} with a 95% confidence level.

A distribution of any type, whether it be Gaussian, Gamma, uniform, etc., belongs to the above moment-based or entropy-based ambiguity sets if it satisfies the associated requirements in the above definitions. It should be noted that the moment-based ambiguity set $\mathbb{P}_m(\bar{\mathbf{H}}_f\mathbf{f}_k, \Sigma_0)$ and the nominal distribution $P_{z_f} = \mathcal{N}(\bar{\mathbf{H}}_f\mathbf{f}_k, \Sigma_0)$ in the entropy-based ambiguity set $\mathbb{P}_e(P_{z_f}, \eta)$ show the fault effect on the residual distribution, but the FD test design does not require to know the true fault signal $\mathbf{f}_k$.

Remark 1. If non-negligible covariance estimation errors are present in Σ_0, the underlying true residual distributions of $z_{0,k}$ and $z_{f,k}$ are not contained in the moment-based ambiguity sets $\mathbb{P}_m(\mathbf{0}, \Sigma_0)$ and $\mathbb{P}_m(\bar{\mathbf{H}}_f\mathbf{f}_k, \Sigma_0)$, respectively. How to address such covariance uncertainty in the moment-based FD test design is left to future research. However, with such non-negligible covariance uncertainty, the entropy-base ambiguity sets are still applicable, as long as $\mathbb{P}_e(P_{z_0}, \eta)$ and $\mathbb{P}_e(P_{z_f}, \eta)$ contain the true residual distributions of $z_{0,k}$ and $z_{f,k}$ by adopting a suitable value of η.

In the parity relation based FD scheme, this paper addresses the trade-off design considering the above two types of distribution ambiguity sets. Specifically, it includes the design of a parity matrix, the threshold computation that ensures a predefined worst-case FAR, and the worst-case FDR evaluation in the presence of any given fault signal. Such a distributionally robust trade-off design is non-trivial, since it has to cope with a set of possible distributions. In contrast, most statistical FD test designs in literature commit to a specific distribution, thus would give a poor performance if the actual distribution becomes different from the one adopted in the design.

3 | DISTRIBUTIONALLY ROBUST DESIGN OVER MOMENT-BASED AMBIGUITY SET

In this section, a DRO approach is proposed for parity relation based FD over the moment-based ambiguity set of disturbance distributions. It starts with a scalar residual, and then proceeds to a vector residual as a complete solution.

3.1 | Distributionally robust design of a scalar test

With w being a column vector of appropriate dimension, a parity vector

$$v = w^\top \mathbf{N}_o \tag{12}$$

is used to generate a scalar residual

$$r_k = v(\mathbf{y}_k - \mathbf{H}_u\mathbf{u}_k) = w^\top z_k \tag{13}$$

according to (5) and (12). In the derivations throughout this paper, the time index k in the subscript is sometimes omitted for the sake of brevity.

Although the occurrence of faults is unpredictable, partial knowledge about faults is often available to describe fault modes that are of interest in a specific application. For this reason, we predefine a reference fault mode as $\delta_k \mathbf{f}_{\text{ref}} \in \mathbb{R}^{n_f h}$, where $\mathbf{f}_{\text{ref}}$ is a unit vector that ensures

$$\bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}} \neq \mathbf{0}, \tag{14}$$

and the scalar $\delta_k > 0$ denotes the time-varying magnitude. Such a reference fault mode specifies how the fault signal varies over a sliding time window $[k-h+1, k]$. Note that no single reference fault mode can represent all possible fault scenarios. The following derivations using one reference fault mode lay the foundation for choosing and addressing a parameterized set of multiple reference fault modes in Section 3.2.

For a selected reference fault mode, two cases are considered:

(i) The faulty residual $r_k = w_1^\top z_k$ has a positive mean, i.e., $w_1^\top \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}} > 0$. Accordingly, a one-sided detection test is constructed as

$$\begin{cases} w_1^\top z_k > b_1 \Rightarrow \text{ fault alarm}, \\ w_1^\top z_k \leq b_1 \Rightarrow \text{ no fault alarm}, \end{cases} \tag{15}$$

where $b_1 > 0$ is the detection threshold.

(ii) The faulty residual $r_k = w_2^\top z_k$ has a negative mean, i.e., $w_2^\top \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}} < 0$. Accordingly, a one-sided detection test is constructed as

$$\begin{cases} w_2^\top z_k < b_2 \Rightarrow \text{ fault alarm}, \\ w_2^\top z_k \geq b_2 \Rightarrow \text{ no fault alarm}, \end{cases} \tag{16}$$

where $b_2 < 0$ is the detection threshold.

In the following, the above two one-sided tests are first designed separately, and then combined into a two-sided test.

For the one-sided test (15) under the reference fault mode $\delta_k \mathbf{f}_{\text{ref}}$, the integrated design of w_1 and b_1 is formulated as a DRO problem

$$\max_{w_1 \neq 0, \beta_1, b_1} \quad \beta_1 \tag{17a}$$

$$\text{s.t.} \quad \inf_{z_{0,k} \sim \mathbb{P}_{z_0}} \Pr\left\{ w_1^\top z_{0,k} \leq b_1 \right\} \geq \alpha, \tag{17b}$$

$$\inf_{z_{f,k} \sim \mathbb{P}_{z_f}} \Pr\left\{ w_1^\top z_{f,k} \geq b_1 \right\} \geq \beta_1, \tag{17c}$$

$$w_1^\top \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}} > 0, b_1 > 0, 0 \leq \beta_1 < 1. \tag{17d}$$

For the one-sided test (15), (17b) ensures the worst-case FAR less than $1-\alpha$. Similarly, (17c) implies that the resulting FDR is lower bounded by β_1. Therefore, for the reference fault mode $\delta_k \mathbf{f}_{\text{ref}}$ that satisfies $w_1^\top \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}} > 0$, the DRO problem (17) of the one-sided test (15) maximizes the worst-case FDR β_1 while ensuring a predefined FAR $1-\alpha$.

In this section, the inexact distributions of $z_{0,k}$ and $z_{f,k}$ in (17) are described by the moment-based ambiguity sets $\mathbb{P}_{z_0} = \mathbb{P}_m(\mathbf{0}, \Sigma_0)$ and $\mathbb{P}_{z_f} = \mathbb{P}_m(\bar{\mathbf{H}}_f \delta_k \mathbf{f}_{\text{ref}}, \Sigma_0)$. How to cope with the distributionally robust chance constraints (17b) and (17c) is then discussed as follows.

Lemma 1. [39,40] Assume that the distribution of a random vector ξ belongs to a moment-based ambiguity set $\mathbb{P}_m(\bar{\xi}, \Xi)$, with $\Xi > 0$. For $q \neq \mathbf{0}$, $\epsilon \in [0, 1)$, and a given scalar c satisfying $q^\top \bar{\xi} \leq c$, the distributionally robust chance constraint

$$\inf_{\xi \sim \mathbb{P}_m(\bar{\xi}, \Xi)} \Pr\{q^\top \xi \leq c\} \geq \epsilon$$

is equivalent to

$$c - q^\top \bar{\xi} \geq \kappa_\epsilon \sqrt{q^\top \Xi q}, \ \kappa_\epsilon = \sqrt{\frac{\epsilon}{1-\epsilon}}. \tag{18}$$

For the case $q^\top \bar{\xi} > c$, $\inf_{\xi \sim \mathbb{P}_m(\bar{\xi}, \Xi)} \Pr\{q^\top \xi \leq c\} = 0$ holds.

By applying Lemma 1, the DRO (17) can be equivalently transformed into a deterministic constrained optimization problem

$$\max_{w_1 \neq 0, \beta_1, b_1} \quad \beta_1 \tag{19a}$$

$$\text{s.t.} \quad b_1 \geq \kappa_\alpha \sqrt{w_1^\top \Sigma_0 w_1}, \tag{19b}$$

$$-b_1 + w_1^\top \bar{\mathbf{H}}_f \delta_k \mathbf{f}_{\text{ref}} \geq \kappa_{\beta_1} \sqrt{w_1^\top \Sigma_0 w_1}, \tag{19c}$$

$$w_1^\top \bar{\mathbf{H}}_f \delta_k \mathbf{f}_{\text{ref}} \geq b_1, b_1 > 0, 0 \leq \beta_1 < 1, \tag{19d}$$

with Σ_0 defined in (7), $\kappa_\alpha = \sqrt{\frac{\alpha}{1-\alpha}}$ and $\kappa_{\beta_1} = \sqrt{\frac{\beta_1}{1-\beta_1}}$. The constraint $w_1^\top \bar{\mathbf{H}}_f \delta_k \mathbf{f}_{\text{ref}} \geq b_1$ in (19d) is imposed, since the right-hand side of (19c) is nonnegative according to Lemma 1. Then, it can be derived from (19b) and (19c) that

$$\kappa_\alpha \sqrt{w_1^\top \Sigma_0 w_1} \leq b_1 \leq w_1^\top \bar{\mathbf{H}}_f \delta_k \mathbf{f}_{\text{ref}} - \kappa_{\beta_1} \sqrt{w_1^\top \Sigma_0 w_1},$$

which is further simplified as

$$\left(\kappa_\alpha + \kappa_{\beta_1}\right) \sqrt{w_1^\top \Sigma_0 w_1} \leq w_1^\top \bar{\mathbf{H}}_f \delta_k \mathbf{f}_{\text{ref}} \tag{20}$$

by eliminating b_1. Since $\kappa_{\beta_1} = \sqrt{\frac{\beta_1}{1-\beta_1}}$ increases monotonically with β_1, maximizing β_1 is equivalent to maximizing κ_{β_1}. Therefore, according to (20), the optimal solution is achieved when

$$\kappa_{\beta_1} = \frac{w_1^\top \bar{\mathbf{H}}_f \delta_k \mathbf{f}_{\text{ref}}}{\sqrt{w_1^\top \Sigma_0 w_1}} - \kappa_\alpha \tag{21}$$

holds. Since α is fixed a priori, κ_α is then determined, and maximizing β_1 is equivalent to solving

$$\max_{w_1 \neq 0} \quad \delta_k^2 \frac{w_1^\top \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}} \mathbf{f}_{\text{ref}}^\top \bar{\mathbf{H}}_f^\top w_1}{w_1^\top \Sigma_0 w_1}. \tag{22}$$

It can be further transformed into

$$\max_{\tilde{w}_1 \neq 0} \quad \frac{\tilde{w}_1^\top \Sigma_0^{-\frac{1}{2}} \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}} \mathbf{f}_{\text{ref}}^\top \bar{\mathbf{H}}_f^\top \Sigma_0^{-\frac{1}{2}} \tilde{w}_1}{\tilde{w}_1^\top \tilde{w}_1}$$

with $\tilde{w}_1 = \Sigma_0^{\frac{1}{2}} w_1$ and the symmetric matrix $\Sigma_0^{\frac{1}{2}}$ being the square root of Σ_0. Then, it is straightforward to derive the optimal solution as

$$\tilde{w}_1^* = \frac{\Sigma_0^{-\frac{1}{2}} \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}}}{\sqrt{\mathbf{f}_{\text{ref}}^\top \bar{\mathbf{H}}_f^\top \Sigma_0^{-1} \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}}}}, \; w_1^* = \frac{\Sigma_0^{-1} \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}}}{\sqrt{\mathbf{f}_{\text{ref}}^\top \bar{\mathbf{H}}_f^\top \Sigma_0^{-1} \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}}}}, \tag{23a}$$

$$b_1^* = \kappa_\alpha = \sqrt{\frac{\alpha}{1-\alpha}}, \tag{23b}$$

$$\kappa_{\beta_1^*} = \delta_k \sqrt{\mathbf{f}_{\text{ref}}^\top \bar{\mathbf{H}}_f^\top \Sigma_0^{-1} \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}}} - \kappa_\alpha, \quad \beta_1^* = \frac{\kappa_{\beta_1^*}^2}{1 + \kappa_{\beta_1^*}^2}. \tag{23c}$$

Required by (19c) and (23c), the condition to ensure a non-zero worst-case FDR for the fault signal $\mathbf{f}_k = \delta_k \mathbf{f}_{\text{ref}}$ is

$$\delta_k \sqrt{\mathbf{f}_{\text{ref}}^\top \bar{\mathbf{H}}_f^\top \Sigma_0^{-1} \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}}} \geq \kappa_\alpha. \tag{24}$$

This condition becomes different for the general fault signal $\mathbf{f}_k \neq \mathbf{f}_{\text{ref}}$, which will be further specified in (45). From (23) and (24), it can be seen that the value of δ_k does not influence the parity vector w_1^* and the threshold b_1^*, but only affects the worst-case FDR.

Similarly to the above derivations, the integrated design of w_2 and b_2 for the second one-sided test (16) is formulated as

$$\begin{aligned} \max_{w_2 \neq 0, \beta_2, b_2} \quad & \beta_2 \\ \text{s.t.} \quad & \inf_{z_{0,k} \sim \mathbb{P}_m(\mathbf{0}, \Sigma_0)} \Pr\left\{ w_2^\top z_{0,k} \geq b_2 \right\} \geq \alpha, \\ & \inf_{z_{f,k} \sim \mathbb{P}_m(\bar{\mathbf{H}}_f \delta_k \mathbf{f}_{\text{ref}}, \Sigma_0)} \Pr\left\{ w_2^\top z_{f,k} \leq b_2 \right\} \geq \beta_2, \\ & w_2^\top \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}} < 0, b_2 < 0, 0 \leq \beta_2 < 1, \end{aligned}$$

and the optimal solution is

$$w_2^* = -w_1^*, \; b_2^* = -\kappa_\alpha, \; \beta_2^* = \beta_1^* \tag{25}$$

with w_1^* and β_1^* given in (23).

The two one-sided tests given in (23) and (25) are symmetric, and can be combined into the following two-sided test

$$\begin{cases} |r_k| > \kappa_\alpha \;\Rightarrow\; \text{fault alarm}, \\ |r_k| \leq \kappa_\alpha \;\Rightarrow\; \text{no fault alarm}, \end{cases} \tag{26}$$

for the residual $r_k = (w_1^*)^\top z_k$.

Remark 2. From (23), it can be seen that the two-sided test (26) does not depend on the magnitude δ_k of the reference fault mode since it is completely described by w_1^* and b_1^*. However, its worst-case FDR for $\delta_k \mathbf{f}_{\text{ref}}$ is indeed related to the magnitude δ_k, as indicated by β_1^* in (23c).

Remark 3. It can be also seen that the solution w_1^* in (23a) depends on the direction $\mathbf{f}_{\text{ref}}$ of the reference fault mode. If the actual fault $\mathbf{f}_k$ does not match with the specific fault mode, i.e., $\mathbf{f}_k = \delta_k \mathbf{f}_{\text{ref}}$, the resulting fault contribution $(w_1^*)^\top \bar{\mathbf{H}}_f \mathbf{f}_k$ could be rather small in the residual, even though the magnitude of $\mathbf{f}_k$ is large. Therefore, the two-sided test (26) might have a poor FDR for a fault $\mathbf{f}_k \neq \delta_k \mathbf{f}_{\text{ref}}$. This limitation will be addressed in Section 3.2 by proposing a vector test.

Remark 4. Similarly to our scalar test in this section, the probabilistic robust parity relation approach proposed in [32,33] also introduces a reference fault mode, and copes with the same moment-based ambiguity set of distributions. However, the method in [32,33] formulates a different minimax problem that minimizes a weighted sum of FAR and miss detection rate in a worst-case setting. The solution obtained in [32,33] relies on numerical iterations, while our scalar test has the closed-form solution in (23). Due to its dependence on the reference fault mode, the FD test derived in [32,33] has the same limitation pointed out in Remark 3 for our scalar test (26).

3.2 | Distributionally robust design of a vector test

As in Remark 3, the FD test of the scalar residual in Section 3.1 relies on only one reference fault mode, hence might give a poor FDR due to the discrepancy between the actual fault and the selected reference fault mode. To address this limitation, the FD test of a vector residual is constructed in this subsection, with respect to a parameterized family of multiple reference fault modes.

Consider multiple fault modes represented by $\{\delta_{k,i}\mathbf{f}_{\text{ref},i}, i = 1, 2, \cdots, m_{\text{ref}}\}$, where $\mathbf{f}_{\text{ref,i}}$ is a unity vector and $\delta_{k,i}$ is the associated magnitude. By solving a DRO problem in the form of (17), each fault mode $\delta_{k,i}\mathbf{f}_{\text{ref},i}$ is used to construct a scalar residual

$$\begin{aligned} r_{k,i} &= \phi_i^\top \tilde{z}_k,\ \tilde{z}_k = \Sigma_0^{-\frac{1}{2}} z_k = \Sigma_0^{-\frac{1}{2}} (\bar{\mathbf{H}}_d \mathbf{d}_k + \bar{\mathbf{H}}_f \mathbf{f}_k), \\ \phi_i &= \frac{\Sigma_0^{-\frac{1}{2}} \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref},i}}{\sqrt{\mathbf{f}_{\text{ref},i}^\top \bar{\mathbf{H}}_f^\top \Sigma_0^{-1} \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref},i}}}, \end{aligned} \tag{27}$$

according to (5) and (23a), with $\tilde{w}_1^*$ replaced by ϕ_i. These scalar residuals form a vector residual

$$\begin{aligned} \mathbf{r}_k &= \begin{bmatrix} r_{k,1} & r_{k,2} & \cdots & r_{k,m_{\text{ref}}} \end{bmatrix}^\top = \Phi^\top \tilde{z}_k, \\ \Phi &= \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{m_{\text{ref}}} \end{bmatrix}. \end{aligned} \tag{28}$$

Before evaluating $\mathbf{r}_k$ for FD, we first need to choose the reference fault modes $\{\delta_{k,i}\mathbf{f}_{\text{ref},i}, i = 1, 2, \cdots, m_{\text{ref}}\}$. Let the singular value decomposition (SVD) of $\Sigma_0^{-\frac{1}{2}}\bar{\mathbf{H}}_f$ be denoted by

$$\Sigma_0^{-\frac{1}{2}}\bar{\mathbf{H}}_f = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} V_1^\top \\ V_2^\top \end{bmatrix}, \tag{29}$$

where $S \in \mathbb{R}^{m_f \times m_f}$ is a diagonal matrix with positive diagonal elements, and $m_f = \text{rank}(\Sigma_0^{-\frac{1}{2}}\bar{\mathbf{H}}_f)$ is defined in (9). In the following, col(X) and null(X) represent the column space and left nullspace of a matrix X, respectively. Since (27) and (28) reveal the link between the reference fault modes $\{\mathbf{f}_{\mathbf{ref},i}\}_{i=1}^{m_{\text{ref}}}$ and the columns $\{\phi_i\}_{i=1}^{m_{\text{ref}}}$ of Φ, the following requirements for choosing these fault modes come from the necessary conditions imposed on the matrix Φ:

i) All columns of Φ in (28) are nonzero. This requires $\bar{\mathbf{H}}_f\mathbf{f}_{\text{ref},i} \neq 0$ according to (27). Otherwise, certain elements of the vector residual $\mathbf{r}_k$ are constantly zero, which is undesirable.

ii) The matrix Φ is of full column rank, i.e., the scalar elements $\{r_{k,i}\}_{i=1}^{m_{\text{ref}}}$ of the vector residual $\mathbf{r}_k$ are linearly independent. This requires all reference fault modes to be linearly independent.

iii) The transformation from the normalized primary residual $\tilde{z}_k$ to the vector residual $\mathbf{r}_k$ in the first equation of (28) should not enlarge the set of completely undetectable faults. According to (27) and (28), the sets of completely undetectable faults for $\tilde{z}_k$ and $\mathbf{r}_k$ are null $\left(\Sigma_0^{-\frac{1}{2}}\bar{\mathbf{H}}_f\right)$ and null $\left(\Phi^\top\Sigma_0^{-\frac{1}{2}}\bar{\mathbf{H}}_f\right)$, respectively. Hence, the selection of reference fault modes should ensure null $\left(\Phi^\top\Sigma_0^{-\frac{1}{2}}\bar{\mathbf{H}}_f\right)$ = null $\left(\Sigma_0^{-\frac{1}{2}}\bar{\mathbf{H}}_f\right)$ = col(V_2), with V_2 defined in (29).

With straightforward linear algebras omitted here, it can be seen from the above requirements that the number of reference fault modes should be $m_{\text{ref}} = m_f = \text{rank}\left(\Sigma_0^{-\frac{1}{2}}\bar{\mathbf{H}}_f\right)$. With these considerations, we introduce a family of fault modes parametrized by an orthogonal matrix $\Xi_k \in \mathbb{R}^{m_f \times m_f}$ and a scalar $\sigma_k > 0$, i.e.,

$$\begin{bmatrix} \delta_{k,1}\mathbf{f}_{\text{ref},1} & \delta_{k,2}\mathbf{f}_{\text{ref},2} & \cdots & \delta_{k,m_f}\mathbf{f}_{\text{ref},m_f} \end{bmatrix} = \sigma_k V_1 S^{-1}\Xi_k,\ \Xi_k^\top\Xi_k = \mathbf{I}_{m_f}. \tag{30}$$

Since Ξ_k and σ_k can be time varying, what the expression (30) defines is not a fixed set of reference fault modes, but a parameterized set of fault modes that may change with time.

Next, we discuss how to determine Ξ_k and σ_k for the reference fault modes defined in (30). Firstly, for any time-varying fault signal $\mathbf{f}_k$, we can select Ξ_k and σ_k as in Appendix A, such that one reference fault mode $\delta_{k,i}\mathbf{f}_{\text{ref},i}$ in (30) exactly captures the effect of $\mathbf{f}_k$ on the primary residual $\tilde{z}_k$ in (27), i.e.,

$$\Sigma_0^{-\frac{1}{2}}\bar{\mathbf{H}}_f\delta_{k,i}\mathbf{f}_{\text{ref},i} = \Sigma_0^{-\frac{1}{2}}\bar{\mathbf{H}}_f\mathbf{f}_k. \tag{31}$$

Thus this fault signal $\mathbf{f}_k$ can be exactly addressed by using the above reference fault mode $\delta_{k,i}\mathbf{f}_{\text{ref},i}$ to construct the scalar residual $r_{k,i}$ in (27). However, the actual fault $\mathbf{f}_k$ is unknown, hence the values of Ξ_k and σ_k that ensure (31) cannot be explicitly computed. Fortunately, the vector test constructed below in (38) remains invariant to any values of Ξ_k and σ_k that are specified to ensure (31) for the time-varying fault $\mathbf{f}_k$, which is a desirable feature.

Thanks to the chosen fault modes in (30), $\{\phi_i, i = 1, \cdots, m_f\}$ form a set of orthogonal bases, since we derive

$$\Phi = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{m_f} \end{bmatrix} = U_1 \Xi_k \tag{32}$$

by substituting (29) and (30) into the expression of ϕ_i in (27). Following the two-sided scalar test in (26), a FD test of the vector residual $\mathbf{r}_k$ in (28) can be constructed by monitoring each residual component $r_{k,i}$, i.e.,

$$\begin{cases} \exists i, |r_{k,i}| > \kappa_\alpha \Rightarrow \text{ fault alarm}, \\ \forall i, |r_{k,i}| \le \kappa_\alpha \Rightarrow \text{ no fault alarm}, \end{cases} \tag{33}$$

which is compactly written as

$$\begin{cases} \|\mathbf{r}_k\|_\infty > \kappa_\alpha \Rightarrow \text{ fault alarm}, \\ \|\mathbf{r}_k\|_\infty \le \kappa_\alpha \Rightarrow \text{ no fault alarm}. \end{cases} \tag{34}$$

This test (34) defines the m_f-dimensional hypercube

$$\mathcal{C} = \left\{ \mathbf{r}_k \middle| \, \|\mathbf{r}_k\|_\infty \le \kappa_\alpha \right\} \tag{35}$$

as the fault-free region.

The above derivations have a clear geometric interpretation. In (27), ϕ_i is a unit vector, and $\tilde{z}_k$ is a normalized primary residual with its covariance being an identity matrix. Then, residual components $\{r_{k,i}\}$ can be regarded as the projection of $\tilde{z}_k$ onto multiple orthogonal directions $\{\phi_i\}$, and each projected variance is 1. With (28), (27), and (32), the vector residual $\mathbf{r}_k$ is expressed as

$$\mathbf{r}_k = \Phi^\top \tilde{z}_k = \Xi_k^\top U_1^\top \tilde{z}_k = \Xi_k^\top U_1^\top \Sigma_0^{-\frac{1}{2}} z_k. \tag{36}$$

As the orthogonal matrix Ξ_k changes, the orthogonal projection directions ϕ_i rotate according to (32), thus the hypercube described in (35) rotates accordingly. These rotating hypercubes has the inscribed ball

$$\mathcal{B} = \left\{ \mathbf{r}_k \middle| \, \|\mathbf{r}_k\|_2^2 \le \kappa_\alpha^2 \right\}, \tag{37}$$

as a new fault-free region demonstrated by a 2-dimensional example in Figure 1.

The above geometric interpretation inspires another FD test

$$\begin{cases} \|\mathbf{r}_k\|_2^2 > \kappa_\alpha^2 \;\Rightarrow\; \text{fault alarm}, \\ \|\mathbf{r}_k\|_2^2 \le \kappa_\alpha^2 \;\Rightarrow\; \text{no fault alarm}, \end{cases} \tag{38}$$

where $\|\mathbf{r}_k\|_2^2$ is further expressed as

$$\begin{aligned} \|\mathbf{r}_k\|_2^2 &= \tilde{z}_k^\top U_1 \Gamma \Gamma^\top U_1^\top \tilde{z}_k = \tilde{z}_k^\top U_1 U_1^\top \tilde{z}_k \\ &= z_k^\top \Sigma_0^{-1} \bar{\mathbf{H}}_f \left(\bar{\mathbf{H}}_f^\top \Sigma_0^{-1} \bar{\mathbf{H}}_f \right)^\dagger \bar{\mathbf{H}}_f^\top \Sigma_0^{-1} z_k, \end{aligned} \tag{39}$$

using (29) and (36). It can be seen from (39) that the obtain test (38) remains invariant to the values of Ξ_k, although the vector residual signal in (36) indeed depends on Ξ_k.

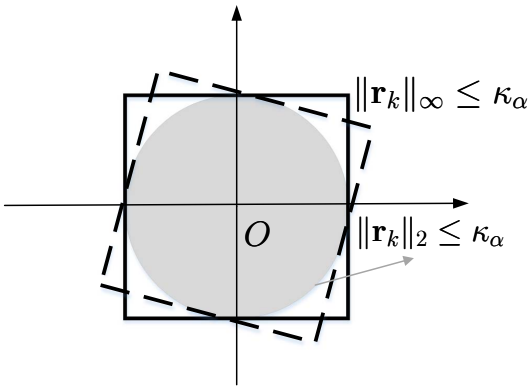


FIGURE 1 A 2-dimensional illustration of rotating hypercubes and their inscribed ball.

It is worth noting the difference between the vector tests (34) and (38). With all possible values of Ξ_k, (30) is not a single set of fault modes, but represents a parameterized family of fault mode sets. The test (34) gives a different residual evaluation as Ξ_k changes, thus its FD performance dependents on Ξ_k. But there is no universally optimal choice of Ξ_k since any set of reference fault modes cannot represent all possible fault scenarios. In contrast, the vector test (38) remains invariant with respect to the entire family of fault mode sets parameterized by Ξ_k in (30). Furthermore, the vector test (38) has a larger set of detectable faults than the test (34), since the ball $\mathcal{B}$ in (37) is a subset of the hypercube $\mathcal{C}$ in (35). Because of these benefits in the above comparisons, only the vector test (38) will be discussed in the rest of this paper.

Compared to the scalar test (26), the vector test (38) has the following two advantages. Firstly, it copes with multiple reference fault modes instead of considering only one reference fault mode. Secondly, it is not restricted to one set of reference fault modes, but remains invariant with respect to a parameterized family of fault mode sets described in (30). A further comparison between the scalar test (26) and the vector test (38) in terms of worst-case FDRs will be given at the end of Section 3.3.

It is also interesting to compare the proposed vector test (38) with the celebrated GLRT (10). The residual signals in the GLRT (10) and the vector test (38) have different dimensions. But their residual evaluation functions turn out to be exactly the same, as

shown in (8) and (39). The only difference between these two FD tests lies in how the detection thresholds are determined. For the GLRT (10), the detection threshold is computed according the assumed Gaussian distribution, and the resulting FAR might be much higher than the predefined level γ in the presence of distribution ambiguity. One may argue that even without knowing any distribution information, the detection threshold can be tuned in a trial-and-error process such that the GLRT gives a low FAR with a satisfying FDR. However, such a trial-and-error method requires training data to tune the threshold, thus still fails to ensure a low FAR if the distribution of real-time data deviates from the data distribution in the trial-and-error process. In contrast, our proposed approaches provide systematic procedures for threshold computation as in (44) and Algorithm 2, so that the worst-case FAR is equal to the predefined level γ as long as the disturbance distribution belongs to the considered ambiguity set.

3.3 | Worst-case performance analysis

For the vector test (38), the residual is generated by (36) with the parity matrix being $W = \Xi_k^\top U_1^\top \Sigma_0^{-\frac{1}{2}}$. With such a parity matrix, we discuss in this subsection about how to choose the detection threshold κ_α^2 in the vector test (38) such that its worst-case FDR is maximized while ensuring a predefined worst-case FAR. Note that in the following we directly determine κ_α^2 without the need to know α.

In the fault-free case, the residual $\mathbf{r}_k$ defined in (36) has a distribution belonging to $\mathbb{P}_m(\mathbf{0}, \mathbf{I}_{m_f})$. Then, the resulting worst-case FAR is

$$\sup_{\mathbf{r}_k \sim \mathbb{P}_m(\mathbf{0}, \mathbf{I}_{m_f})} \Pr\left\{ \|\mathbf{r}_k\|_2^2 \geq \kappa_\alpha^2 \right\} = \frac{m_f}{\kappa_\alpha^2}, \tag{40}$$

which follows the multivariate Chebyshev inequality[41].

According to (27), the vector residual $\mathbf{r}_k$ in (36) in the faulty case has a distribution belonging to $\mathbb{P}_m(\mu_k, \mathbf{I}_{m_f})$, with

$$\mu_k = \Xi_k^\top U_1^\top \Sigma_0^{-\frac{1}{2}} \bar{\mathbf{H}}_f \mathbf{f}_k. \tag{41}$$

The achieved worst-case FDR is then given as follows.

Theorem 1. For the vector test (38), its worst-case FDR for a fault $\mathbf{f}_k$ is

$$\inf_{\mathbf{r}_k \sim \mathbb{P}_m(\mu_k, \mathbf{I}_{m_f})} \Pr\left\{ \|\mathbf{r}_k\|_2^2 > \kappa_\alpha^2 \right\} = \begin{cases} \frac{\left(\|\mu_k\|_2 - \kappa_\alpha\right)^2}{1 + \left(\|\mu_k\|_2 - \kappa_\alpha\right)^2} & \text{if } \|\mu_k\|_2 > \kappa_\alpha, \\ 0 & \text{if } \|\mu_k\|_2 \leq \kappa_\alpha, \end{cases} \tag{42}$$

with

$$\|\mu_k\|_2 = \left(\mathbf{f}_k^\top \bar{\mathbf{H}}_f^\top \Sigma_0^{-1} \bar{\mathbf{H}}_f \mathbf{f}_k \right)^{\frac{1}{2}}. \tag{43}$$

The proof is given in Appendix B.

Given the vector test (38), threshold computation aims at fining the minimal κ_α^2 that ensures the worst-case FAR while maximizing the worst-case FDR. For this purpose, we set the detection threshold to be

$$\kappa_\alpha^2 = \frac{m_f}{\gamma} \tag{44}$$

such that the worst-case FAR in (40) is equal to γ, where γ is the predefined level of FAR allowed in a specific application.

Next, we further compare the vector test (38) with the scalar test (26) in terms of worst-case FDRs under the same fault $\mathbf{f}_k$. Following the proof of Theorem 1, it can be derived that the worst-case FDR of the scalar test (26) is

$$\inf_{r_k \sim \mathbb{P}_m(\mu_{1,k},1)} \Pr\{|r_k|^2 > \kappa_\alpha^2\} = \begin{cases} \frac{(|\mu_{1,k}|-\kappa_\alpha)^2}{1+(|\mu_{1,k}|-\kappa_\alpha)^2} & \text{if } |\mu_{1,k}| > \kappa_\alpha, \\ 0 & \text{if } |\mu_{1,k}| \le \kappa_\alpha, \end{cases} \tag{45}$$

with

$$|\mu_{1,k}| = \left| (\tilde{w}_1^*)^\top \Sigma_0^{-\frac{1}{2}} \bar{\mathbf{H}}_f \mathbf{f}_k \right|. \tag{46}$$

Consider a fault signal described as $\mathbf{f}_k = V_1 S^{-1} \zeta_k$, with V_1 and S defined in (29). In this case, $\|\mu_k\|_2$ in (43) and $|\mu_{1,k}|$ in (46) are obtained as $\|\mu_k\|_2 = \|\zeta_k\|_2$ and $|\mu_{1,k}| = \left|(\tilde{w}_1^*)^\top U_1 \zeta_k\right|$, respectively. According to (42), the vector test (38) produces a zero worst-case FDR if ζ_k lies in the bounded m_f-dimensional ball $\|\zeta_k\|_2 \le \kappa_\alpha$. In contrast, according to (45), a larger set of ζ_k, i.e., an unbounded strip $\left|(\tilde{w}_1^*)^\top U_1 \zeta_k\right| \le \kappa_\alpha$, gives a zero worst-case FDR by the scalar test (45), which implies a poorer FD performance compared to the vector test (38).

Remark 5. With the worst-case FAR fixed as γ, it is interesting to investigate how the worst-case FDR varies if the time window length h increases. According to (9) and (44), m_f grows with h, thus the detection threshold κ_α^2 needs to be increased accordingly to give the fixed worst-case FAR γ. From (43), it can be seen that $\|\mu_k\|_2$ also increases with h. Therefore, when fixing the worst-case FAR and increasing h, both κ_α^2 and $\|\mu_k\|_2$ grow, and how the worst-case FDR in (42) varies depends on $(\|\mu_k\|_2 - \kappa_\alpha)^2$ whose value is not guaranteed to grow with h.

Remark 6. The moment-based ambiguity set can be conservative due to using only mean and covariance information. Moreover, the worst-case probability distribution in the moment-based DRO problem (17) is generally discrete [42,43], which is unrealistic in many applications. Due to this reason, the moment-based test design might result in a poor FDR, as will be seen in the simulation results of Section 5. Such conservatism is also a limitation of the recent FD literature leveraging the minimax probability machine [32,33], because the same mean and covariance information was used therein.

4 | DISTRIBUTIONALLY ROBUST DESIGN OVER ENTROPY-BASED AMBIGUITY SET

To address the conservatism of the moment-based design (see Remark 6), residual distributions in this section are described by entropy-based ambiguity sets. Then, entropy-based DRO problems are solved for the FD test design and performance evaluation, although the basic idea is the same as in Section 3.

4.1 | Distributionally robust design

Similarly to Lemma 1, the following proposition is derived for the distributionally robust design over the entropy-based distribution ambiguity.

Proposition 1. Assume that the distribution of a random vector ξ belongs to an entropy-based ambiguity set $\mathbb{P}_e(P_0, \eta)$. For a given scalar c, the distributionally robust chance constraint

$$\inf_{\xi \sim \mathbb{P}_e(P_0,\eta)} \Pr\{q^\top \xi \le c\} \ge \epsilon \tag{47}$$

is equivalent to

$$\Pr_{\xi \sim P_0}\{q^\top \xi \le c\} \ge 1 - \bar{g}_\eta(\epsilon) \tag{48}$$

with

$$\bar{g}_\eta(\epsilon) = \sup_{t>0} g_\eta(t, \epsilon),\ g_\eta(t, \epsilon) = \frac{e^{-\eta}(t+1)^{1-\epsilon} - 1}{t}, \tag{49}$$

Suppose the nominal distribution P_0 is $\mathcal{N}(\bar{\xi}, \Xi)$ with $\Xi > 0$, then the chance constraint (48) can be further simplified as

$$c - q^\top \bar{\xi} \ge \varpi_{\eta,\epsilon} \sqrt{q^\top \Xi q}, \tag{50}$$

where

$$\varpi_{\eta,\epsilon} = \Phi^{-1}(1 - \bar{g}_\eta(\epsilon)), \tag{51}$$

and $\Phi^{-1}(\cdot)$ represents the inverse cumulative distribution function of the standard normal distribution.

The proof is given in Appendix C. Note that the equivalent deterministic constraint (50) is in the same form as (18) in Lemma 1, except that $\varpi_{\eta,\epsilon}$ is used in (50) instead of κ_ϵ in (18).

Corollary 1. With any given $\eta > 0$ and the definitions in (49) and (51), $\varpi_{\eta,\epsilon}$ increases monotonically with ϵ.

The proof is given in Appendix D.

Again, the optimal design of the one-sided test (15) is considered by formulating the DRO problem (17) with respect to the reference fault mode $\delta_k \mathbf{f}_{\text{ref}}$. Instead of using the moment-based ambiguity sets, residual distributions are now described by the entropy-based ambiguity sets $\mathbb{P}_{z_0} = \mathbb{P}_e(P_{z_0}, \eta)$ and $\mathbb{P}_{z_f} = \mathbb{P}_e(P_{z_f}, \eta)$, where $P_{z_0} = \mathcal{N}(\mathbf{0}, \Sigma_0)$ and $P_{z_f} = \mathcal{N}(\bar{\mathbf{H}}_f \delta_k \mathbf{f}_{\text{ref}}, \Sigma_0)$

represent nominal Gaussian distributions. In this case, by applying Proposition 1, the DRO problem (17) with $\mathbb{P}_{z_0} = \mathbb{P}_e(\boldsymbol{P}_{z_0}, \eta)$ and $\mathbb{P}_{z_f} = \mathbb{P}_e(\boldsymbol{P}_{z_f}, \eta)$ is equivalently expressed as

$$\begin{aligned} \max_{w_1 \neq 0, \beta_1, b_1} \quad & \beta_1 \\ \text{s.t.} \quad & b_1 \geq \varpi_{\eta,\alpha} \sqrt{w_1^\top \Sigma_0 w_1}, \\ & -b_1 + w_1^\top \bar{\mathbf{H}}_f \delta_k \mathbf{f}_{\text{ref}} \geq \varpi_{\eta,\beta_1} \sqrt{w_1^\top \Sigma_0 w_1}, \\ & w_1^\top \bar{\mathbf{H}}_f \delta_k \mathbf{f}_{\text{ref}} > b_1, b_1 > 0, 0 \leq \beta_1 < 1, \end{aligned} \tag{52}$$

with $\varpi_{\eta,\alpha}$ and ϖ_{η,β_1} defined in the same form as in (51). It is not surprising to see that the optimization problems (52) and (19) are in the same structure, except that κ_α and κ_{β_1} in (19) are respectively replaced by $\varpi_{\eta,\alpha}$ and ϖ_{η,β_1} in (52). Since both ϖ_{η,β_1} in (52) (see Theorem 1) and κ_{β_1} in (19) monotonically increases with β_1, the derivations (20)–(24) can be directly followed to solve (52). As such, the optimal solution to (52) is achieved when

$$\varpi_{\eta,\beta_1} = \frac{w_1^\top \bar{\mathbf{H}}_f \delta_k \mathbf{f}_{\text{ref}}}{\sqrt{w_1^\top \Sigma_0 w_1}} - \varpi_{\eta,\alpha}$$

holds, and maximizing β_1 is again equivalent to solving (22). Consequently, the obtained optimal solution $(w_1^*, b_1^*, \beta_1^*)$ is expressed as

$$w_1^* = \frac{\Sigma_0^{-1} \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}}}{\sqrt{\mathbf{f}_{\text{ref}}^\top \bar{\mathbf{H}}_f^\top \Sigma_0^{-1} \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}}}}, \; b_1^* = \varpi_{\eta,\alpha},$$

$$\varpi_{\eta,\beta_1^*} = \delta_k \sqrt{\mathbf{f}_{\text{ref}}^\top \bar{\mathbf{H}}_f^\top \Sigma_0^{-1} \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}}} - \varpi_{\eta,\alpha}.$$

By following the derivations in Section 3.2, it is straightforward to obtain the vector test below from the above one-sided scalar test:

$$\begin{cases} \|\mathbf{r}_k\|_2^2 > \varpi_{\eta,\alpha}^2 \;\Rightarrow\; \text{fault alarm} \\ \|\mathbf{r}_k\|_2^2 \leq \varpi_{\eta,\alpha}^2 \;\Rightarrow\; \text{no fault alarm} \end{cases} \tag{53}$$

with $\|\mathbf{r}_k\|_2^2$ defined in (39). As in the previous vector test (38), the derived test (53) again has the same residual evaluation function as in the GLRT (10), but its threshold is different to account for the entropy-based distribution ambiguity.

4.2 | Worst-case performance analysis

For the entropy-based vector test (53), the residual generation and evaluation are given in (36) and (39), which are the same as in the moment-based vector test (38). However, the detection threshold $\varpi_{\eta,\alpha}^2$ of the entropy-based vector test (53) has a different functional dependence on α. In this subsection, we discuss how to evaluate the worst-case FAR and FDR with a given threshold $\varpi_{\eta,\alpha}^2$ under the entropy-based distribution ambiguity.

Following Part (i) of Lemma 2 in Appendix C, the worst-case FAR of the vector test (53) is

$$\begin{aligned}&\sup_{z_k\sim\mathbb{P}_e(P_{z_0},\eta)} \Pr\left\{\|\mathbf{r}_k\|_2^2 > \varpi_{\eta,\alpha}^2\right\}\\ &= \min_{\rho\geq 0}\ \rho\ln\left[\gamma_0(e^{1/\rho}-1)+1\right]+\rho\eta,\end{aligned} \tag{54}$$

where $P_{z_0}=\mathcal{N}(\mathbf{0},\Sigma_0)$ is the nominal Gaussian distribution in the fault-free case, $\Gamma_{m_f}(\cdot)$ denotes the $\chi^2_{m_f}$ CDF, ρ is the decision variable to be solved, and

$$\gamma_0 = \Pr_{z_k\sim P_{z_0}}\left\{\|\mathbf{r}_k\|_2^2 > \varpi_{\eta,\alpha}^2\right\} = 1-\Gamma_{m_f}(\varpi_{\eta,\alpha}^2). \tag{55}$$

Note that η is the divergence bound defined in (11) to describe the entropy-based distribution ambiguity, and how to determine η is discussed in the paragraph below (11). The second equation in (55) holds, because $\|\mathbf{r}_k\|_2^2$ defined in (39) follows the $\chi^2_{m_f}$ distribution in the fault-free case. With a given threshold $\varpi^2_{\eta,\alpha}$, γ_0 can be computed according to (55), then the worst-case FAR is determined by the optimization problem (54) whose solution will be discussed later.

Similarly, the worst-case FDR of the vector test (53) under an unknown fault $\mathbf{f}_k$ is

$$\begin{aligned}&\inf_{z_k\sim\mathbb{P}_e(P_{z_f},\eta)} \Pr\left\{\|\mathbf{r}_k\|_2^2 > \varpi_{\eta,\alpha}^2\right\}\\ &= 1-\sup_{z_k\sim\mathbb{P}_e(P_{z_f},\eta)}\left\{\|\mathbf{r}_k\|_2^2 \leq \varpi_{\eta,\alpha}^2\right\}\\ &= 1-\min_{\rho\geq 0}\ \left\{\rho\ln\left[\gamma_f(e^{1/\rho}-1)+1\right]+\rho\eta\right\}\end{aligned} \tag{56}$$

where $P_{z_f}=\mathcal{N}(\bar{\mathbf{H}}_f\mathbf{f}_k,\Sigma_0)$ is the nominal Gaussian distribution in the faulty case, $\Gamma_{m_f,\lambda}(\cdot)$ denotes the non-central $\chi^2_{m_f,\lambda}$ CDF with $\lambda=\mathbf{f}_k^\top\bar{\mathbf{H}}_f^\top\Sigma_0^{-1}\bar{\mathbf{H}}_f\mathbf{f}_k$, and

$$\gamma_f = \Pr_{z_k\sim P_{z_f}}\left\{\|\mathbf{r}_k\|_2^2 \leq \varpi_{\eta,\alpha}^2\right\} = \Gamma_{m_f,\lambda}(\varpi_{\eta,\alpha}^2). \tag{57}$$

For a given fault $\mathbf{f}_k$ and a given threshold $\varpi^2_{\eta,\alpha}$, γ_f can be computed according to (57), and the worst-case FDR is determined by the optimization problem (56).

It can be seen that computing the worst-case FAR in (54) and the worst-case FDR in (56) requires solving the same minimization problem except that γ_0 and γ_f take different values. Thus only the minimization problem in (54) is discussed next. Firstly, $\rho=0$ is excluded, because the objective function tends to 1 as ρ approaches zero, and this makes no sense for computing FAR or FDR. Then, inspired by[44,31], a bisection search algorithm in the case of $\rho>0$ is proposed according to the analysis in Appendix E, and its details are presented in Algorithm 1. The procedure for computing the worst-case FDR with the given fault $\mathbf{f}_k$ and threshold $\varpi^2_{\eta,\alpha}$ is similar to Algorithm 1, thus is omitted.

Algorithm 1 Worst-case FAR evaluation given the threshold $\varpi_{\eta,\alpha}^2$ for the test (53)

Compute $\gamma_0 = 1 - \Gamma_{m_f}(\varpi_{\eta,\alpha}^2)$ according to (55).
Initialization for solving (54): $\nu_l = \gamma_0 e^{\eta}$, $\nu_u = 1$.
while $\nu_u - \nu_l > \varsigma$ (ς is a predefined small positive value such as 10^{-6}) **do**
  Compute $\tilde{\nu} = \frac{\nu_l + \nu_u}{2}$, $t^* = \max\left\{0, \left(\frac{\tilde{\nu}}{\gamma_0 e^{\eta}}\right)^{\frac{1}{1-\tilde{\nu}}} - 1\right\}$, and $\bar{\psi} = e^{-\eta}(t^*+1)^{\tilde{\nu}} - \gamma_0 t^* - 1$.
  if $\bar{\psi} > 0$ **then** update $\nu_u = \tilde{\nu}$
  else if $\bar{\psi} < 0$ **then** update $\nu_l = \tilde{\nu}$
  else update $\nu_l = \nu_u = \tilde{\nu}$, and return.
  end if
end while
Output $\tilde{\nu}$ as the worst-case FAR.

4.3 | Parameter tuning

For the moment-based vector test (38), the threshold κ_α^2 is determined by setting $\alpha = \frac{m_f}{m_f+\gamma}$, as detailed in Section 3.3. However, for the entropy-based vector test (53), we do not have a closed-form expression for determining the threshold $\varpi_{\eta,\alpha}^2$, but resort to a bisection search algorithm (Algorithm 2) for the threshold computation.

For a fixed η, the FAR-related optimization problem (54) is leveraged to determine the minimal threshold $\varpi_{\eta,\alpha}^2$ that ensures the predefined worst-case FAR γ. The basic idea follows two steps. Firstly, find γ_0^* with the procedure described below, such that the worst-case FAR in (54) with $\gamma_0 = \gamma_0^*$ is γ. Secondly, according to (55), the threshold $\varpi_{\eta,\alpha}^2$ is set to $\Gamma_{m_f}^{-1}(1-\gamma_0^*)$. Note that α is not explicitly needed here for computing the threshold $\varpi_{\eta,\alpha}^2$.

In the first step, we find γ_0^* such that

$$\min_{\rho \geq 0} \ \rho \ln\left[\gamma_0^*(e^{1/\rho} - 1) + 1\right] + \rho\eta = \gamma,$$

where the left-hand side is the worst-case FAR in (54) with $\gamma_0 = \gamma_0^*$. For this purpose, we need to further analyze the associated optimization problem in (54). On one hand, the worst-case FAR γ (i.e., the minimal objective function of (54)) monotonically increases with γ_0, which can be proved by following the proof of Theorem 1. On the other hand, the inequality

$$\sup_{z_k \sim \mathbb{P}_e(P_{z_0}, \eta)} \Pr\left\{\left\|\mathbf{r}_k\right\|_2^2 > \varpi_{\eta,\alpha}^2\right\} = \min_{\rho \geq 0} \ \rho \ln\left[\gamma_0(e^{1/\rho} - 1) + 1\right] + \rho\eta \leq \gamma$$

implies that γ_0 is upper-bounded as

$$\gamma_0 = \Pr_{z_k \sim P_{z_0}}\left\{\left\|\mathbf{r}_k\right\|_2^2 > \varpi_{\eta,\alpha}^2\right\} \leq \sup_{t>0} \frac{e^{-\eta}(t+1)^{\gamma} - 1}{t} \tag{58}$$

according to Part (ii) of Lemma 2 in Appendix C. Therefore, the worst-case FAR achieves the predefined level γ at the upper bound of γ_0 in (58), which implies

$$\gamma_0^* = \sup_{t>0} \frac{e^{-\eta}(t+1)^{\gamma} - 1}{t}. \tag{59}$$

Similarly to solving (54), a bisection search algorithm for solving (59) is presented in Algorithm 2, according to the analysis in Appendix F.

Algorithm 2 Threshold computation that ensures the worst-case FAR γ for the test (53)

Initialization for solving (59): $\gamma_l = 0$, $\gamma_u = \begin{cases} \gamma e^{-\eta} & \text{if } \gamma e^{-\eta} < 1 \\ 1 & \text{if } \gamma e^{-\eta} \geq 1 \end{cases}$.
while $\gamma_u - \gamma_l > \varsigma$ (ς is a predefined small positive value such as 10^{-6}) **do**
  Compute $\tilde{\gamma} = \frac{\gamma_l + \gamma_u}{2}$, $t^* = \left(\frac{\gamma}{\tilde{\gamma} e^{\eta}}\right)^{\frac{1}{1-\gamma}} - 1$ and $\bar{\varphi} = e^{-\eta}(t^*+1)^{\gamma} - 1 - \tilde{\gamma} t^*$.
  if $\bar{\varphi} > 0$ **then** update $\gamma_l = \tilde{\gamma}$
  else if $\bar{\varphi} < 0$ **then** update $\gamma_u = \tilde{\gamma}$
  else update $\gamma_l = \gamma_u = \tilde{\gamma}$, and return
  end if
end while
Set $\gamma_0^* = \tilde{\gamma}$, and compute the threshold $\varpi_{\eta,\alpha}^2 = \Gamma_{m_f}^{-1}(1 - \gamma_0^*)$.

5 | SIMULATION STUDY

In this section, a well-mixed continuous stirred tank reactor (CSTR) is considered. It includes an exothermic reaction A→B, where A is the reactant species, and B is the desired product. The reactant A is fed to the reactor with a flow rate F, concentration C_{Af}, and temperature T_{Af}. In the equipped cooling jacket, a cooling stream with temperature T_{cf} flows at a rate F_c. The process dynamics is described by[45]

$$\begin{aligned}
\dot{C}_{\mathrm{A}}(t) =& \frac{F}{V}\left(C_{\mathrm{Af}} - C_{\mathrm{A}}(t)\right) - k\exp\left(\frac{-E}{RT(t)}\right)C_{\mathrm{A}}(t) + w_1(t),\\
\dot{C}_{\mathrm{B}}(t) =& -\frac{F}{V}C_{\mathrm{B}}(t) + k\exp\left(\frac{-E}{RT(t)}\right)C_{\mathrm{A}}(t) + w_2(t),\\
\dot{T}(t) =& \frac{F}{V}\left(T_{\mathrm{Af}} - T(t)\right) + \frac{-\Delta H}{\rho c_p}k\exp\left(\frac{-E}{RT(t)}\right)C_{\mathrm{A}}(t)\\
&- \frac{U\mathrm{A}}{\rho c_p V}\left(T(t) - T_{\mathrm{c}}(t)\right) + w_3(t),\\
\dot{T}_{\mathrm{c}}(t) =& \frac{F_c}{V_c}\left(T_{\mathrm{cf}}(t) - T_{\mathrm{c}}(t)\right) + \frac{\mathrm{UA}}{\rho_c c_{pc} V_c}\left(T(t) - T_c(t)\right) + w_4(t)
\end{aligned} \tag{60}$$

where the species A concentration C_{A}, species B concentration C_{B}, reactor temperature T, and cooling jacket temperature T_c constitute the system states, and the feed concentration C_{Af}, feed temperature T_{Af}, and cooling stream temperature T_{Cf} are the system inputs. Process disturbances $w_1(t)$, $w_2(t)$, $w_3(t)$, and $w_4(t)$ are zero-mean Gaussian white noises with variances $3 \times 10^{-5}(\mathrm{mol/L})^2$, $3 \times 10^{-5}(\mathrm{mol/L})^2$, $6 \times 10^{-4}\mathrm{K}^2$, and $6 \times 10^{-4}\mathrm{K}^2$, respectively. System parameters in (60) are described in Table 1. In this section, "K" represents the unit of Kelvin temperature.

Let the nonlinear system (60) be operated around a stead-state point described as $C_{\mathrm{Af}}^0 = 4$ mol/L, $T_{\mathrm{Af}}^0 = T_{\mathrm{Cf}}^0 = 360$ K, $C_{\mathrm{A}}^0 = 3.8506$ mol/L, $C_{\mathrm{B}}^0 = 0.1494$ mol/L, $T^0 = 363.2828$ K, $T_c^0 = 362.6622$ K. Around this steady-state point, linearization

TABLE 1 System parameters

Parameter	Description	Value	Unit
F	Inlet flow rate	83.3	L/min
F_c	Flow rate of cooling stream	50	L/min
V	Tank volume	1000	L
V_c	Volume of cooling jacket	100	L
ΔH	Enthalpy of A→B	-5×10^4	J/mol
UA	Heat transfer coefficient	9.01×10^5	J/(min·K)
k	Pre-exponential constant of A→B	5×10^4	min^{-1}
E	Activation energy A→ B	5×10^4	J/mol
R	Gas constant	8.314	J/(min·K)
ρ	Fluid density in reactor	1000	g/L
c_p	Fluid heat capacity in reactor	0.231	J/(g·K)
ρ_c	Fluid density in cooling jacket	1000	g/L
c_{pc}	Fluid heat capacity in cooling jacket	4.2	J/(g·K)

and discretization (60) with a sampling period $t_s = 0.1$s yields the linearized discrete-time linear system (1), with

$$A = \begin{bmatrix} 0.9914 & 0 & 0 & 0 \\ 0.0003 & 0.9917 & 0 & 0 \\ 0.0585 & 0 & 0.7096 & 0.2858 \\ 0.0061 & 0 & 0.1572 & 0.7986 \end{bmatrix}, \; B = \begin{bmatrix} 0.0083 & 0 & 0 \\ 0 & 0 & 0 \\ 0.0003 & 0.0070 & 0.0079 \\ 0 & 0.0007 & 0.0445 \end{bmatrix}.$$

In this simulation example, C_B, T, and T_c are measurable, hence the matrices C and D in (1) are $C = \begin{bmatrix} \mathbf{0}_{3\times1} & \mathbf{I}_3 \end{bmatrix}$ and $D = \mathbf{0}_{3\times3}$, respectively. The measurement noises in the sensors of C_B, T, and T_c are zero-mean white Gaussian with variances $3 \times 10^{-4}(\text{mol/L})^2$, $6 \times 10^{-3}\text{K}^2$, $6 \times 10^{-3}\text{K}^2$, respectively. Let $\begin{bmatrix} w^\top & v^\top \end{bmatrix}^\top$ be the disturbance d in the linearized system (1), with $w \in \mathbb{R}^4$ and $v \in \mathbb{R}^3$ respectively representing the discrete-time process noises and the measurement noises. Then, the associated matrices B_d and D_d in (1) are $B_d = \begin{bmatrix} \mathbf{I}_4 & \mathbf{0}_{4\times3} \end{bmatrix}$ and $D_d = \begin{bmatrix} \mathbf{0}_{3\times4} & \mathbf{I}_3 \end{bmatrix}$.

The following two fault scenarios are included:

- A drift sensor fault on the reactor temperature T:
  $B_f = \mathbf{0}_{4\times1}$ and $D_f = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^\top$ in (1),

$$f(k) = \begin{cases} 0 & 0 \le k \le 300 \\ 0.005(k - 300) & k > 300; \end{cases}$$

- An oscillatory actuator fault on the cooling stream temperature T_{Cf}:
  $B_f = \begin{bmatrix} 0 & 0 & 0.0079 & 0.0445 \end{bmatrix}^\top$ and $D_f = \mathbf{0}_{3\times 1}$ in (1),

$$f(k) = \begin{cases} 0 & 0 \le k \le 300 \\ 0.002k \sin \frac{\pi k}{50} - 3 & k > 300. \end{cases}$$

For performance comparisons, the following four parity relation based FD methods are implemented using the same parity order $h = 7$, the predefined FAR $\gamma = 10\%$, and the primary residual z_k in (5):

- the GLRT (8)–(10): its residual evaluation $\mathcal{L}^*(\bar{r}_k)$ is computed as in (8) using the primary residual z_k, and its threshold is set to $\Gamma^{-1}_{m_f}(1-\gamma)$ defined in (10) with $m_f = \mathrm{rank}(\Sigma_0^{-\frac{1}{2}}\bar{\mathbf{H}}_f) = 7$ in (9).

- the scalar test (26) using the moment-based DRO approach: with the reference fault mode $\delta_k \mathbf{f}_{\mathrm{ref}}$ selected to be in the direction

$$\mathbf{f}_{\mathrm{ref}} = \begin{bmatrix} 0.3657 & 0.3697 & 0.3738 & 0.3779 & 0.3819 & 0.3860 & 0.3900 \end{bmatrix}^\top, \tag{61}$$

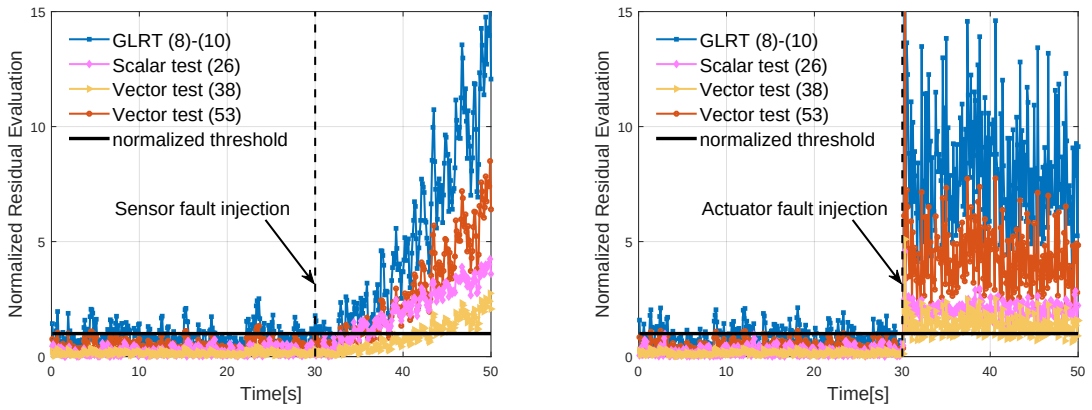
  the parity vector w_1^* is computed according to (23), the residual signal is generated as $r_k = (w_1^*)^\top z_k$, and the associated threshold is determined as $\kappa_\alpha^2 = 1/\gamma = 10$ according to (44);

- the vector test (38) using the moment-based DRO approach: for the ambiguity sets $\mathbb{P}_m(0, \Sigma_0)$ and $\mathbb{P}_m(\bar{\mathbf{H}}_f \mathbf{f}_k, \Sigma_0)$, Σ_0 is the 17×17 empirical covariance matrix of the fault-free primary residual, whose value can be determined according to procedures described in Section 2.3 but not given in detail here due to space limit; the residual evaluation $\|\mathbf{r}_k\|_2^2$ is computed as in (39) from the primary residual z_k, and its threshold is $\kappa_\alpha^2 = m_f/\gamma = 70.00$ according to (44) and $m_f = \mathrm{rank}(\Sigma_0^{-\frac{1}{2}}\bar{\mathbf{H}}_f) = 7$ in (9); note that the vector test (38) turns out to be independent of the parametrized set of reference fault modes in (30);

- the vector test (53) using the entropy-based DRO approach: the residual evaluation $\|\mathbf{r}_k\|_2^2$ still follows (39); the bound $\eta = 0.30$ for the ambiguity sets $\mathbb{P}_e(P_{z_0}, \eta)$ and $\mathbb{P}_e(P_{z_f}, \eta)$ is determined by the KL divergence between the empirical distribution of the fault-free primary residual z_k and the nominal distribution $P_{z_0} = \mathcal{N}(\mathbf{0}, \Sigma_0)$, with detailed procedures in Section 2.3; then, its threshold $\varpi^2_{\eta=0.30,\alpha} = 22.65$ is computed using Algorithm 2.

With the considered fault scenarios in one simulation run, the normalized residual evaluations are depicted in Figure 2, where all detection thresholds of the above four methods are normalized to unity. Since the actual disturbance distribution deviates from the assumed Gaussian distribution due to linearization and discretization errors, the FARs of the GLRT (8)–(10) are significantly higher than the predefined level $\gamma = 10\%$, i.e., 32.08% in the sensor fault scenario and 21.50% in the actuator fault scenario. In contrast, the vector test (53) using the entropy-based DRO approach has only 2.73% FAR in the sensor fault scenario and

zero FAR in the actuator fault. As expected, the vector test (38) using the moment-based DRO approach results in the most conservative threshold, and gets the lowest FDR among the implemented FD tests.



(a) Sensor fault scenario

(b) Actuator fault scenario

FIGURE 2 Normalized residual evaluation of one simulation run in the considered fault scenarios.

Next, we illustrate how the selection of reference fault modes affects the FD performance by comparing the scalar test (26) with the vector test (38). For this purpose, we consider the temperature sensor fault

$$\mathbf{f}_k = \begin{bmatrix} 0.450 & 0.455 & 0.460 & 0.465 & 0.470 & 0.475 & 0.480 \end{bmatrix}^{\mathsf{T}} \tag{62}$$

over the time window $[390, 396]$. The scalar test (26) is implemented using two different reference fault modes: 1) the direction of the reference fault mode is $\mathbf{f}_{\text{ref}}$ in (61), which is in the same direction as the true sensor fault signal in (62); 2) the direction of the reference fault mode is

$$\mathbf{f}_{\text{ref}} = \begin{bmatrix} -0.3657 & 0.3697 & -0.3738 & 0.3779 & -0.3819 & 0.3860 & -0.3900 \end{bmatrix}^{\mathsf{T}}, \tag{63}$$

which is almost orthogonal to the true sensor fault signal in (62). The vector test (38) using the moment-based DRO approach has the same parameter setting as given before. In Figure 3, the actual FARs and FDRs of the implemented FD methods are plotted against the predefined FARs γ whose 34 different values are equally spaced within the interval $(0, 1)$. Each dot on the curves of Figure 3 is determined by 4000 fault-free or faulty Monte Carlo simulations. For all implemented methods, the actual FARs remain below the predefined FAR level γ. The scalar test (26) achieves the highest FDRs only when its reference fault mode in (61) happens to perfectly match with the direction of the true fault signal in (62). However, it also gives the poorest FDRs when using the improper reference fault mode in (63). Note that there is no guarantee that one single reference fault mode remains close to the direction of a time-varying fault signal, which is the major limitation of the scalar test (26). In contrast, the

vector test (38) gives 99.03% FDR at the cost of 3.91% FAR in Figure 3(b), and it is independent of the parameterized set of reference fault modes in (30).

Similarly to Figure 3, Figure 4 compares the GLRT and the two vector tests using the moment-based and entropy-based DRO approaches. The same temperature sensor fault signal in (62) over the time window [390, 396] is considered again. For the vector test (53) using the entropy-based DRO approach, $\eta = 0.3$ and $\eta = 0.7$ are both implemented to illustrate the effect of η on the FD performance.

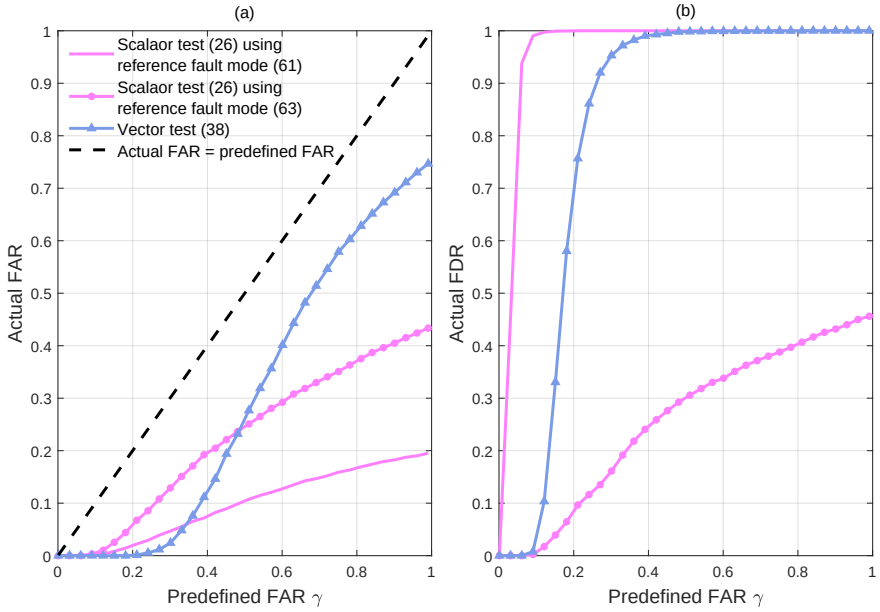


FIGURE 3 FDRs and FARs of scalar and vector tests using moment-based DRO approaches.

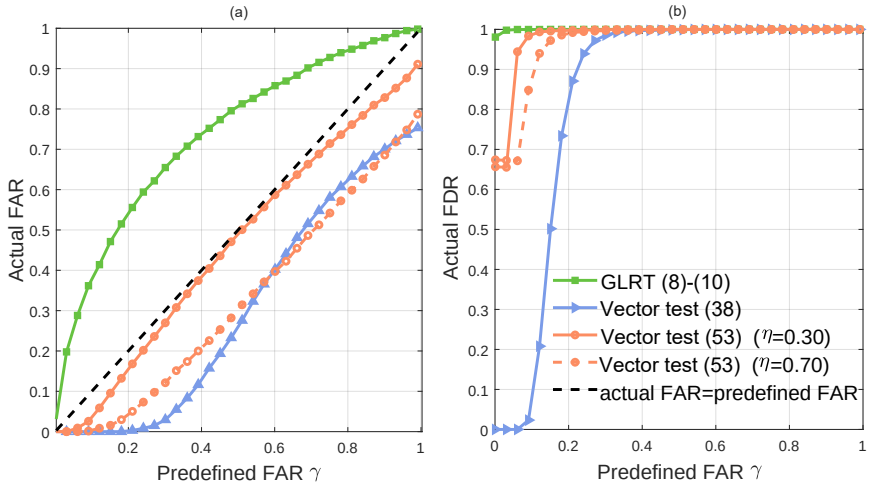


FIGURE 4 FDRs and FARs of different vector tests under different predefined FARs.

As seen from Figure 4, the GRLT (8)–(10) has its actual FARs all higher than the predefined FARs, although it achieves the highest FDRs. This is again due to ambiguous disturbance distribution caused by linearization and discretization errors. These errors are compensated by the ambiguity set in both DRO approaches, hence FARs of these two approaches are all lower than the predefined FARs. Due to the conservatism of the moment-based ambiguity set explained in Remark 6, the moment-based DRO approach needs to use a conservative threshold to ensure the predefined FAR for the worst-case distribution. As a result, its actual FAR is much lower than the predefined worst-case FAR, but its achieved FAR is the poorest compared to other methods. The performance of the entropy-based DRO approach depends on the value of η. It achieves a better performance tradeoff by setting $\eta = 0.3$, because the actual FARs are closer to the predefined worst-case FARs, as illustrated by the dotted orange curve and the dashed black line in Figure 4(a).

With different values of the reactor temperature sensor bias, the theoretical worst-case FDRs of the vector tests using two DRO approaches are compared against their actual FDRs in 2000 Monte-Carlo simulations. The predefined worst-case FAR γ is still set to 10%. Then, the detection threshold for the moment-based vector test (38) is 70.00, while the entropy-based vector test (53) has its detection threshold set to 22.65 or 32.31 for the divergence parameter $\eta = 0.3$ or $\eta = 0.7$, respectively. The theoretical worst-case FDRs are computed by solving (56) with similar procedures in Algorithm 1. As depicted in Figure 5, for a fault amplitude less than 0.8, the moment-based DRO approach achieves much lower FDRs than the entropy-based DRO approach. As expected, there exists a gap between the theoretical worst-case FDRs and the actual FDRs. This gap in the moment-based DRO approach is larger than in the entropy-based DRO approach. Moreover, this gap in the entropy-based DRO approach becomes narrower as η reduces from 0.7 to 0.3.

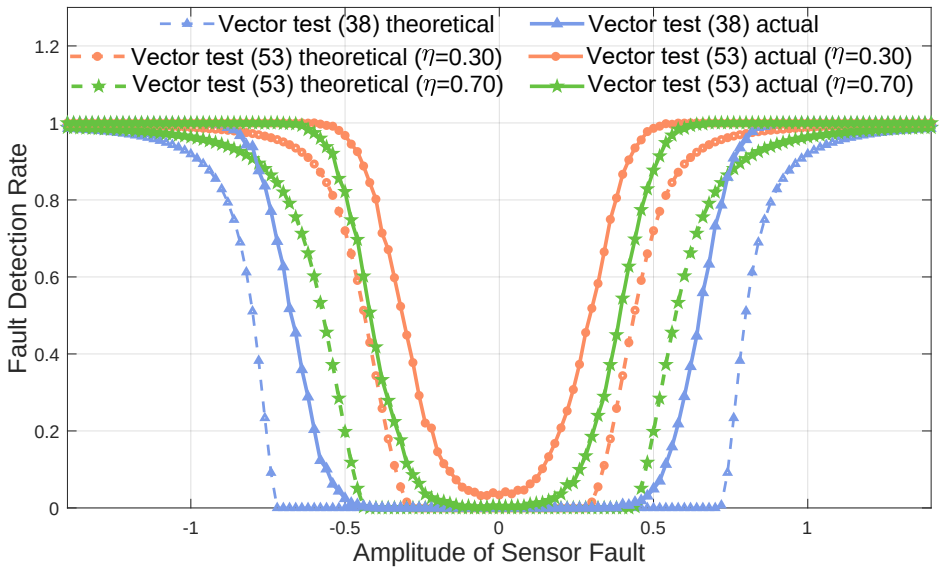


FIGURE 5 FDRs for the amplitude of fault varying from -1.4 to 1.4.

6 | CONCLUSIONS

This paper considers the parity relation based FD problem for linear systems in the presence of stochastic disturbances whose distribution information is ambiguous. With the moment-based and entropy-based distribution ambiguity sets, a DRO approach is proposed for the trade-off design. The obtained FD tests have the same structure as the conventional GLRT, and compensate for the distribution ambiguity by their detection thresholds. By solving associated DRO problems, we develop systematic methods for the threshold computation that ensures a predefined worst-case FAR, and for the worst-case FDR evaluation in the presence of any given fault signal. With simulation results on a CSTR example, it is shown that the GLRT fails to ensure a low FAR due to not addressing ambiguous distribution information. In contrast, our proposed entropy-based DRO design achieves desirable trade-off between FAR and FDR, and provides effective worst-case performance evaluation. However, the moment-based DRO design gives a poor performance trade-off due to the conservatism of moment-based distribution ambiguity description.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grants 61803163 and 61733009, the Research Fund for the Taishan Scholar Project of Shandong Province of China, and Shandong Key Laboratory of Big-data Driven Safety Control Technology for Complex Systems under Grant SKDN202002. The authors would also like to thank Prof. Tamas Keviczky at Delft University of Technology for his helpful suggestions.

Conflict of interest

The authors declare no potential conflict of interests.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Gao Z, Cecati C, Ding SX. A survey of fault diagnosis and fault-tolerant techniques – Part I: fault diagnosis with model-based and signal-based approaches. *IEEE Transactions on Industrial Electronics* 2015; 62(6): 3757–3767.

2. Ding SX. *Model-Based Fault Diagnosis Techniques: Design Schemes, Algorithms and Tools*. Springer-Verlag London. 2nd ed. 2013.

3. Chen H, Jiang B, Ding SX, Huang B. Data-driven fault diagnosis for traction systems in high-speed trains: A survey, challenges, and perspectives. *IEEE Transactions on Intelligent Transportation Systems* 2020. doi: 10.1109/TITS.2020.3029946

4. Patton RJ, Chen J. A review of parity space approach to fault diagnosis. *IFAC Proceedings* 1991; 24: 65–81.

5. Ding SX, Li L, Kruger M. Application of randomized algorithms to assessment and design of observer-based fault detection systems. *Automatica* 2019; 107: 175–182.

6. Esfahani PM, Lygeros J. A tractable fault detection and isolation approach for nonlinear systems with probabilistic performance. *IEEE Transactions on Automatic Control* 2016; 61(3): 633–647.

7. Han W, Wang Z, Shen Y. H_-/L_∞ fault detection observer for linear parameter-varying systems with parametric uncertainty. *International Journal of Robust and Nonlinear Control* 2019; 29: 2912–2926.

8. Li L, Luo H, Ding SX, Yang Y, Peng K. Performance-based fault detection and fault-tolerant control for automatic control systems. *Automatica* 2019; 99: 308–316.

9. Zhai X, Xu H, Wang G. Robust H_-/H_∞ fault detection observer design for polytopic spatially interconnected systems over finite frequency domain. *International Journal of Robust and Nonlinear Control* 2021; 31(2): 404–426.

10. Zhang Y, Fang H, Zheng Y, Li X. Torus-event-based fault diagnosis for stochastic multirate time-varying systems with constrained fault. *IEEE Transactions on Cybernetics* 2019; 50(6): 2803–2813.

11. Zhang J, Christofides P, He X, Wu Z, Zhang Z, Zhou D. Event-triggered filtering and intermittent fault detection for time-varying systems with stochastic parameter uncertainty and sensor saturation. *International Journal of Robust and Nonlinear Control* 2018; 28: 4666–4680.

12. Raimondo DM, Marseglia GR, Braatz RD, Scott JK. Closed-Loop Input Design for Guaranteed Fault Diagnosis using Set-Valued Observers. *Automatica* 2016; 74: 107–117.

13. Scott JK, Raimondo DM, Marseglia GR, Braatz RD. Constrained Zonotopes: A New Tool for Set-Based Estimation and Fault Detection. *Automatica* 2016; 69: 126–136.

14. Silvestre D, Rosa P, Hespanha JP, Silvestre C. Fault detection for LPV systems using set-valued observers: A coprime factorization approach. *Systems & Control Letters* 2017; 106: 32–39.

15. Xu F, Tan J, Y. W, X. W, Liang B, Yuan B. Robust Fault Detection and Set-Theoretic UIO for Discrete-Time LPV Systems With State and Output Equations Scheduled by Inexact Scheduling Variables. *IEEE Transactions on Automatic Control* 2019; 64(2): 4982–4997.

16. Wang Y, Puig V, Xu F, Cembrano G. Robust fault detection and isolation based on zonotopic unknown input observers for discrete-time descriptor systems. *Journal of the Franklin Institute* 2019; 356(10): 5293–5314.

17. Blesa J, Puig V, Saludes J, Fernández-Cantí RM. Set-membership parity space approach for fault detection in linear uncertain dynamic systems. *International Journal of Adaptive Control and Signal Processing* 2016; 30: 186-205.

18. Harirchi F, Ozay N. Guaranteed model-based fault detection in cyber–physical systems: A model invalidation approach. *Automatica* 2018; 93: 476–488.

19. Gustafsson F. Statistical signal processing approaches to fault detection. *Annual Reviews in Control* 2007; 31: 41-54.

20. Sheriff MZ, Mansouri M, Karim MN, Nounou H, Nounou M. Fault detection using multiscale PCA-based moving window GLRT. *Journal of Process Control* 2017; 54: 47–64.

21. Zhang Q, Basseville M. Statistical detection and isolation of additive faults in linear time-varying systems. *Automatica* 2014; 50: 2527–2538.

22. Zhou J, Yang Y, Zhao Z, Ding SX. A fault detection scheme for ship propulsion systems using randomized algorithm techniques. *Control Engineering Practice* 2018; 81: 65–72.

23. Zhong M, Liu C, Zhou D, Xue T. Probability analysis of fault diagnosis performance for satellite attitude control systems. *IEEE Transactions on Industrial Informatics* 2019; 15(11): 5867–5876.

24. Combastel C. An Extended Zonotopic and Gaussian Kalman Filter (EZGKF) merging set-membership and stochastic paradigms: Toward non-linear filtering and fault detection. *Annual Reviews in Control* 2016; 42: 232–243.

25. Rostampour V, Ferrari R, Keviczky T. A set based probabilistic approach to threshold design for optimal fault detection. In: Proceedings of 2017 American Control Conference. ; 2017; Seattle, USA: 5422–5429.

26. Wan Y, Puig V, Ocampo-Martinez C, Wang Y, Harinath E, Braatz RD. Fault detection for uncertain LPV systems using probabilistic set-membership parity relation. *Journal of Process Control* 2020; 87: 27–36.

27. Zhang P, Ding SX. An integrated trade-off design of observer based fault detection systems. *Automatica* 2008; 44(7): 1886–1894.

28. Wan Y, Dong W, Wu H, Ye H. Integrated fault detection system design for linear discrete time-varying systems with bounded power disturbances. *International Journal of Robust and Nonlinear Control* 2013; 23: 1781–1802.

29. Kay SM. *Fundamentals of Statistical Signal Processing: Detection Theory*. New Jersey, US: Prentice-Hall . 1998.

30. Rahimian H, Mehrotra S. Distributionally robust optimization: a review. *arXiv preprint arXiv:1908.05659* 2019.

31. Jiang R, Guan Y. Data-driven chance constrained stochastic program. *Mathematical Programming* 2016; 158(1-2): 291-327.

32. Zhong M, Xue T, Song Y, Ding SX, Ding EL. Parity Space Vector Machine Approach to Robust Fault Detection for Linear Discrete-Time Systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 2021; 51(7): 4251–4261.

33. Song Y, Zhong M, Xue T, Ding SX, Li W. Parity space-based fault isolation using minimum error minimax probability machine. *Control Engineering Practice* 2020; 95: 104242.

34. Shang C, Ding SX, Ye H. Distributionally robust fault detection design and assessment for dynamical systems. *Automatica* 2021; 125: 109434.

35. Yu C, Ljung L, Wills A, Verhaegen M. Constrained Subspace Method for the Identification of Structured State-Space Models. *IEEE Transactions on Automatic Control* 2019; 65(10): 4201–4214.

36. Scott DW. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons. 2nd ed. 2015.

37. Song KS. Goodness-of-fit tests based on Kullback-Leibler discrimination information. *IEEE Transactions on Information Theory* 2002; 48(5): 1103–1117.

38. Zhao P, Lai L. Minimax optimal estimation of KL divergence for continuous distributions. *IEEE Transactions on Information Theory* 2020; 66(12): 7787–7811.

39. Calafiore GC, Ghaoui LE. On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications* 2006; 130(1): 1–22.

40. Lanckriet GR, Ghaoui LE, Bhattacharyya C, Jordan MI. A robust minimax approach to classification. *Journal of Machine Learning Research* 2002; 3(Dec): 555–582.

41. Navarro J. A very simple proof of the multivariate Chebyshev's inequality. *Communications in Statistics – Theory and Methods* 2016; 45(12): 3458–3463.

42. Bertsimas D, Popescu I. Optimal inequalities in probability theory: a convex optimization approach. *SIAM Journal of Optimization* 2005; 15(3): 780-804.

43. Delage E, Ye Y. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research* 2010; 58(3): 595–612.

44. Hu Z, Hong JL. Kullback–Leibler divergence constrained distributionally robust optimization. 2012. Optimization Online.

45. Mhaskar P, Liu J, Christofides PD. *Fault-Tolerant Process Control: Methods and Applications*. Springer-Verlag . 2013.

□

APPENDIX

A SELECTION OF Ξ_K AND σ_K SUCH THAT EQUATION (31) HOLDS

By extracting the ith column from both sides of (30), we have $\delta_{k,i}\mathbf{f}_{\text{ref},i} = \sigma_k V_1 S^{-1}\Xi_{k,i}$, where $\Xi_{k,i}$ is the ith column of Ξ_k. Combining this equation with (29), (31) is rewritten as

$$\Sigma_0^{-\frac{1}{2}}\bar{\mathbf{H}}_f\delta_{k,i}\mathbf{f}_{\text{ref},i} = \sigma_k U_1\Xi_{k,i} = U_1 S V_1^\top \mathbf{f}_k. \tag{A1}$$

Since $\Xi_{k,i}$ is a column vector within the orthogonal matrix Ξ_k, $\Xi_{k,i}$ is also a unity vector. Then, it can be derived from (A1) that $\sigma_k = \left\| SV_1^\top \mathbf{f}_k \right\|_2$ and $\Xi_{k,i} = SV_1^\top \mathbf{f}_k/\sigma_k$. Given the above $\Xi_{k,i}$, the other columns of Ξ_k can be arbitrary chosen as long as Ξ_k is an orthogonal matrix.

B PROOF OF THEOREM 1

Using Theorem 6.1 in[42], the worst-case FDR of the test (38) is expressed as

$$1 - \sup_{\mathbf{r}_k \sim \mathbb{P}_m(\mu_k, \mathbf{I}_{n_r})} \Pr\left\{ \left\|\mathbf{r}_k\right\|_2^2 \le \kappa_\alpha^2 \right\} = 1 - \frac{1}{1+\tau} \tag{B2}$$

with

$$\tau = \inf_{\|\mathbf{r}\|_2^2 \le \kappa_\alpha^2} \left(\mathbf{r} - \mu_k\right)^\top \left(\mathbf{r} - \mu_k\right).$$

Geometrically, τ is the distance from μ_k to the ball $\mathcal{B} = \left\{ \mathbf{r}_k \middle| \left\|\mathbf{r}_k\right\|_2^2 \le \kappa_\alpha^2 \right\}$ defined in (37). For $\mu_k \in \mathcal{B}$, such a distance is zero since the optimal solution is exactly $\mathbf{r}^* = \mu_k$. For $\mu_k \notin \mathcal{B}$, the optimal solution lies on the surface of $\mathcal{B}$, hence it is obtained as

$$\mathbf{r}^* = \frac{\kappa_\alpha \mu_k}{\left\|\mu_k\right\|_2} \text{ and } \tau^* = (\left\|\mu_k\right\|_2 - \kappa_\alpha)^2$$

by using the method of Lagrangian multipliers. After substituting the above solution into (B2), the worst-case FDR in (42) is obtained. Using (29), (30), and (41), (43) can be derived.

C PROOF OF PROPOSITION 1

For the sake of self-containedness, we include the following lemma on distributionally robust chance constraint over the entropy-based distribution ambiguity, whose detailed proof is referred to[44] and[31]. It is directly applied to prove Proposition 1.

Lemma 2. [44,31] Assume that the distribution of a random vector ξ belongs to an entropy-based ambiguity set $\mathbb{P}_e(P_0, \eta)$. Let $H(\theta, \xi)$ denote a function depending on the random vector ξ and a fixed parameter θ.

(i) The DRO problem $\sup_{\xi\sim\mathbb{P}_e(P_0,\eta)} \{H(\theta,\xi) > 0\}$ is equivalent to the deterministic optimization problem

$$\min_{\rho\geq 0} \ \rho \ln\left[h_\theta(e^{1/\rho} - 1) + 1\right] + \rho\eta \tag{C3}$$

with

$$h_\theta = \Pr_{\xi\sim P_0}\{H(\theta,\xi) > 0\}.$$

(ii) The distributionally robust chance constraint

$$\sup_{\xi\sim\mathbb{P}_e(P_0,\eta)} \Pr\{H(\theta,\xi) > 0\} \leq \epsilon \tag{C4}$$

is equivalent to the chance constraint

$$\Pr_{\xi\sim P_0}\{H(\theta,\xi) > 0\} \leq \bar{\epsilon} \tag{C5}$$

with $t = e^{1/\rho} - 1 > 0$ and

$$\bar{\epsilon} = \sup_{t>0} \frac{e^{-\eta}(t+1)^\epsilon - 1}{t} < \epsilon. \tag{C6}$$

In Part (ii) of Lemma 2, the distributionally robust chance constraint (C4) is equivalently transformed into the standard chance constraint (C5) with respect to the nominal distribution P_0. Note that compared to the risk level ϵ in (C4), the risk level $\bar{\epsilon}$ in (C5) is more restrictive to compensate for distribution ambiguity, according to (C6).

In Proposition 1, (48) and (49) are the result of directly applying Part (ii) of Lemma 2, with the general constraint $H(\theta, \xi) > 0$ replaced by $q^\top\xi \leq c$. And (50) is obtained since

$$\Pr_{\xi\sim P_0}\left\{q^\top\xi \leq c\right\} = \Pr_{\xi\sim P_0}\left\{\frac{q^\top\xi - q^\top\bar{\xi}}{\sqrt{q^\top\Xi q}} \leq \frac{c - q^\top\bar{\xi}}{\sqrt{q^\top\Xi q}}\right\} = \Phi\left(\frac{c - q^\top\bar{\xi}}{\sqrt{q^\top\Xi q}}\right).$$

D PROOF OF COROLLARY 1

It is obvious that the inverse cumulative distribution function $\Phi^{-1}(1-\bar{g}_\eta(\epsilon))$ decreases monotonically with $\bar{g}_\eta(\epsilon)$. Next, we just need to prove that $\bar{g}_\eta(\epsilon)$ has a monotone-decreasing dependence on ϵ by showing that

$$\bar{g}_\eta(\epsilon_1) \geq \bar{g}_\eta(\epsilon_2) \text{ holds for } \epsilon_1 \leq \epsilon_2. \tag{D7}$$

Assume that the supremums $\sup_{t>0} g_\eta(t,\epsilon_1)$ and $\sup_{t>0} g_\eta(t,\epsilon_2)$ are attained at t_1^* and t_2^*, respectively. Then, it is straightforward that

$$\bar{g}_\eta(\epsilon_1) = \sup_{t>0} g_\eta(t,\epsilon_1) \geq g_\eta(t_2^*,\epsilon_1) \tag{D8}$$

holds. Also, we have

$$g_\eta(t_2^*,\epsilon_1) \geq g_\eta(t_2^*,\epsilon_2) = \bar{g}_\eta(\epsilon_2), \tag{D9}$$

since $g_\eta(t,\epsilon)$ decreases monotonically with ϵ according to its definition (51). Combining (D8) and (D9) results in (D7).

E BISECTION SEARCH ALGORITHM FOR SOLVING EQUATION (54)

With η and γ_0 given, solving (54) relies on checking whether the set

$$\mathcal{P}_{\tilde{v}} = \left\{ \rho \,\middle|\, \rho \ln\left[\gamma_0(e^{1/\rho}-1)+1\right] + \rho\eta \leq \tilde{v},\ \rho > 0 \right\} \tag{E10}$$

is empty for a given $\tilde{v} \in (0,1)$. If $\mathcal{P}_{\tilde{v}}$ is empty, then the optimal solution v^* is larger than $\tilde{v}$ and the algorithm needs to search above $\tilde{v}$. Otherwise, the algorithm needs to search below $\tilde{v}$.

Define $t = e^{1/\rho} - 1 > 0$ and

$$\psi(t,\tilde{v}) = e^{-\eta}(t+1)^{\tilde{v}} - \gamma_0 t - 1. \tag{E11}$$

For any fixed $\tilde{v} \in (0,1)$, the function $\psi(t,\tilde{v})$ over $t>0$ is concave, and its supremum

$$\bar{\psi}(\tilde{v}) = \sup_{t>0} \psi(t,\tilde{v}) \tag{E12}$$

is attained at

$$t^*(\tilde{v}) = \max\left\{0, \left(\frac{\tilde{v}}{\gamma_0 e^{\eta}}\right)^{\frac{1}{1-\tilde{v}}} - 1\right\}. \tag{E13}$$

With these notations above, checking non-emptiness of $\mathcal{P}_{\tilde{v}}$ is equivalent to checking the non-emptiness of

$$\mathcal{T}_{\tilde{v}} = \{t \,|\psi(t,\tilde{v}) \geq 0,\ t>0\}, \tag{E14}$$

which can be further transformed into checking whether

$$\bar{\psi}(\tilde{v}) \geq 0 \text{ and } t^*(\tilde{v}) > 0 \tag{E15}$$

hold, with $\bar{\psi}(\tilde{v})$ and $t^*(\tilde{v})$ defined in (E12) and (E13). Note that $t^*(\tilde{v}) = 0$ is infeasible, because in this case $\bar{\psi}(\tilde{v}) = e^{-\eta} - 1 < 0$ holds for $\eta > 0$.

For a selected value $\tilde{v}$, if $\bar{\psi}(\tilde{v}) > 0$ holds, then the set $\mathcal{T}_{\tilde{v}}$ in (E14) is non-empty, and the minimal value v^* is below $\tilde{v}$. If $\bar{\psi}(\tilde{v}) = 0$, then v^* is equal to $\tilde{v}$. Otherwise, the set $\mathcal{T}_{\tilde{v}}$ in (E14) is empty, and v^* is above $\tilde{v}$.

Actually, the bisection search of $\tilde{v}$ does not need to consider the entire interval $(0, 1)$. It will be proved in the following that the feasible interval for $\tilde{v}$ is $(\gamma_0 e^{\eta}, 1)$. Then it is sufficient to search with this smaller interval to further speed up computation.

Proof to show the feasible interval of $\tilde{v}$: The analysis for (E15) reveals the infeasibility of $t^*(\tilde{v}) = 0$. Hence $\left(\frac{\tilde{v}}{\gamma_0 e^{\eta}}\right)^{\frac{1}{1-\tilde{v}}} - 1 \leq 0$ is not allowed. For this reason, $\gamma_0 e^{\eta} < 1$ has to be satisfied, and a feasible $\tilde{v}$ must lie within the interval $(\gamma_0 e^{\eta}, 1)$ so that $\left(\frac{\tilde{v}}{\gamma_0 e^{\eta}}\right)^{\frac{1}{1-\tilde{v}}} - 1 > 0$ holds.

F BISECTION ALGORITHM FOR SOLVING EQUATION (59)

The basic idea is to check whether the set

$$\begin{aligned} \mathcal{Q}_{\tilde{\gamma}} &= \left\{ t \,\middle|\, \frac{e^{-\eta}(t+1)^{\gamma} - 1}{t} > \tilde{\gamma} \right\} \\ &= \{ t \,|\, e^{-\eta}(t+1)^{\gamma} - 1 - \tilde{\gamma} t > 0 \} \end{aligned}$$

is empty for a given $\tilde{\gamma}$. This is equivalent to checking the non-emptiness of $\mathcal{T}_{\tilde{v}}$, according to the definitions (E11) and (E14). Similarly to the analysis in the last two paragraph of Appendix E, the feasible interval of $\tilde{\gamma}$ is $(0, \gamma e^{-\eta})$ if $\gamma e^{-\eta} < 1$; and otherwise, this feasible interval is $(0, 1)$.