Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Probabilistic Robust Parity Relation based Fault Detection Using Biased Minimax Probability Machine ⋆

Yujia Ma * **Yiming Wan** * **Maiying Zhong** **

* *Key Laboratory of Image Processing and Intelligent Control, Ministry of Education, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China.*
** *College of Electrical Engineering and Automation, Shangdong University of Science and Technology, Qingdao 266590, China.*

e-mail: ywan@hust.edu.cn

Abstract: This paper proposes a probabilistic robust parity relation based approach to fault detection of stochastic linear systems. Instead of assuming exact knowledge of disturbance distribution, the uncertainty of distribution information is taken into account by considering an ambiguity set of disturbance distributions. The biased minimax probability machine scheme is exploited to formulate an integrated design of the parity vector/matrix and the detection threshold. It maximizes the worst-case fault detection rate (FDR) with respect to selected reference faults, while ensuring a predefined worst-case false alarm rate. Firstly, a scalar residual design is derived in an analytical form. The analysis of its FDR in the presence of an arbitrary fault shows its limitation due to using a single reference fault. This issue is further addressed by proposing a vector residual design with a systematic method to select multiple reference faults. The efficacy of the proposed approach is illustrated by a simulation example.

Keywords: Fault detection, parity relation, biased minimax probability machine, false alarm rate, fault detection rate.

1. INTRODUCTION

The importance of safety-critical applications has motivated enormous research interests in model-based fault detection (FD). A typical FD system consists of a residual generator and a residual evaluator (Ding, 2013). The generated residual captures the deviation from the observed system behavior to 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 aims at maximizing the fault detection rate (FDR), while ensuring an acceptable false alarm rate (FAR) in the presence of disturbances or model uncertainties.

As one common approach to residual generation, the parity relation method derives an input-output model over a sliding window by decoupling the unknown initial state with a so-called parity vector or matrix (Patton and Chen, 1991; Ding, 2013). According to the assumption about disturbances, the parity relation based FD algorithms can be classified into two categories. The first category assumes disturbances to be unknown but bounded. The design of a parity vector or matrix is to minimize a ratio between robustness to disturbances and sensitivity to faults. Then, the norm-based residual evaluation is adopted to determine a threshold that ensures zero FAR in the worst-case (Ding, 2013). However, the resulting FDR could be low, because a conservative threshold is used to account for the worst-case disturbance that rarely occurs.

The second category of the parity relation based FD algorithms copes with stochastic systems. For linear systems with additive Gaussian disturbances, a generalized likelihood ratio test is proposed in the parity relation scheme (Basseville and Nikiforov, 1993; Gustafsson, 2007). In the observer-based FD scheme, randomized algorithms are proposed to exploit the probability distributions of parametric uncertainties and disturbances in the threshold computation (Rostampour et al., 2017; Zhou et al., 2018), or FAR-constrained optimal design (Ding et al., 2019; Esfahani and Lygeros, 2016). Recently, the probabilistic paradigm is combined with the set-membership approach in Combastel (2016); Wan et al. (2020) by admitting a risk level in the set-based consistency test. All these aforementioned methods requires accurate distribution information, thus cannot address inaccuracy/ambiguity of these distributions. To this end, Zhong et al. (2019) utilizes the minimum error minimax probability machine (MEMPM) in the parity relation scheme to minimize a weighted sum of FAR and miss detection rate in the worst case over a set of possible disturbance distributions.

This paper focuses on the probabilistic robust design of the parity relation based FD subject to an ambiguity set

⋆ This work is supported by the National Natural Science Foundation of China, Grant No. 61803163. (Corresponding author: Yiming Wan)

of disturbance distributions. Different from the MEMPM scheme in Zhong et al. (2019), the proposed approach exploits a biased minimax probability machine (BMPM) scheme which maximizes the worst-case FDR with respect to selected reference faults while ensuring a predefined worst-case FAR. Firstly, this BMPM-based approach derives a scalar residual design in a simple analytical form. In contrast, the MEMPM scheme in Zhong et al. (2019) leads to a nonlinear optimization problem whose solution requires a numerical iterative algorithm. Then, the analytically derived FDR in the presence of an arbitrary fault reveals its poor FDR when the actual fault does not match with the reference fault. This issue is further addressed by proposing a vector residual design with a systematic method to select multiple reference faults. Although the idea of using reference faults is inspired by Zhong et al. (2019), no systematic discussion was included therein about how to choose reference faults.

This paper is organized as follows. Section 2 reviews the conventional parity relation method and states the motivation of the probabilistic robust design. The proposed BMPM approach of a scalar residual design and a vector residual design are presented in Sections 3 and 4, respectively. The simulation results and some conclusions are presented in Sections 5 and 6.

2. PROBLEM STATEMENT

2.1 System description

Consider the following linear time-invariant system

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

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.

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

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

where

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

$\mathbf{u}_k$, $\mathbf{d}_k$, 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 conventional parity relation based FD

To decouple the unknow initial state, a parity relation based residual generator is constructed by

$$\begin{aligned} r_k &= V(\mathbf{y}_k - \mathbf{H}_u \mathbf{u}_k) \\ &= V(\mathbf{H}_o x(k-h+1) + \mathbf{H}_d \mathbf{d}_k + \mathbf{H}_f \mathbf{f}_k) \\ &= V(\mathbf{H}_d \mathbf{d}_k + \mathbf{H}_f \mathbf{f}_k), \end{aligned} \tag{4}$$

where the parity matrix V is designed to satisfy

$$V\mathbf{H}_o = 0. \tag{5}$$

Let $\mathbf{N}_o \in \mathbf{R}^{n_o \times n_y h}$ denote the basis matrix of the left null space of $\mathbf{H}_o$. Without loss of generality, the parity matrix V can be expressed as

$$V = W\mathbf{N}_o, \tag{6}$$

where W is a nonsingular matrix to be designed. With

$$\bar{\mathbf{H}}_d = \mathbf{N}_o \mathbf{H}_d, \ \bar{\mathbf{H}}_f = \mathbf{N}_o \mathbf{H}_f, \tag{7}$$

the generated residual r_k can be rewritten as

$$r_k = W(\bar{\mathbf{H}}_d \mathbf{d}_k + \bar{\mathbf{H}}_f \mathbf{f}_k). \tag{8}$$

The conventional parity relation approach assumes the disturbance $\mathbf{d}_k$ to be unknown-but-bounded. Then, the design matrix W is chosen such that the ratio between sensitivity to faults and robustness to disturbances is maximized, i.e., (Ding, 2013)

$$\max_W \frac{\|W\bar{\mathbf{H}}_f\|}{\|W\bar{\mathbf{H}}_d\|}, \tag{9}$$

where $\|\cdot\|$ denotes some induced norm of a matrix, such as 2-norm or ∞-norm. See details in Section 7.4 of (Ding, 2013). After generating the residual with (8) and the optimal solution W^*, the next step is residual evaluation which usually adopts the root mean square value of r_k over a sliding window $[k-N+1, k]$:

$$J_k = \left(\frac{1}{N} \sum_{i=k-N+1}^{k} r_i^\top r_i \right)^{\frac{1}{2}}.$$

Finally, the detection decision generally follows

$$\begin{cases} J_k > J_{\text{th}} \Rightarrow \text{ fault alarm}, \\ J_k \le J_{\text{th}} \Rightarrow \text{ no fault alarm}, \end{cases}$$

where J_{th} is determined by $J_{\text{th}} = \sup_{f(k)=0} J_k$ to ensure zero FAR, see details in Chapters 9 in Ding (2013).

2.3 Motivation of this study

The above conventional parity relation approach assumes unknown-but-bounded disturbances. The solution to its formulated sensitivity/robustness optimization problem (9) does not ensure optimal tradeoff between FDR and FAR. Moreover, the parity matrix V, or equivalently W, and the detection threshold J_{th} are separately designed, hence the overall FD performance is not fully optimized.

In this paper, the integrated design of the parity matrix and the detection threshold is investigated in the statistical setting to directly optimize the FD performance in terms of FDR and FAR. Instead of assuming Gaussian disturbances, the uncertainty of the probability distribution of disturbances is taken into account. Specifically, the distribution of $\mathbf{d}_k$ belongs to an ambiguity set $\mathbb{P}(0, \bar{\mathbf{Q}})$ that consists of all probability distributions with zero mean and covariance $\bar{\mathbf{Q}}$. Considering all allowable disturbance distributions in the ambiguity set $\mathbb{P}(0, \bar{\mathbf{Q}})$, the objective is to maximize the worst-case FDR while ensuring a predefined worst-case FAR.

3. FD OF A SCALAR RESIDUAL USING BMPM

In this section, a probabilistic robust parity relation based FD of a scalar residual is proposed by exploiting the BMPM scheme.

3.1 Probabilistic robust design using BMPM

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

$$v = w^\top \mathbf{N}_o \tag{10}$$

is used to generate a scalar residual

$$r_k = v(\mathbf{y}_k - \mathbf{H}_u \mathbf{u}_k) = w^\top z_k, \tag{11}$$

where z_k represents the primary residual

$$z_k = \mathbf{N}_o(\mathbf{y}_k - \mathbf{H}_u \mathbf{u}_k) = \bar{\mathbf{H}}_d \mathbf{d}_k + \bar{\mathbf{H}}_f \mathbf{f}_k, \tag{12}$$

according to (7) and (10). From z_k in (12), denote the primary residuals in the fault-free and faulty cases by

$$\begin{aligned} z_{0,k} &= z_k|_{f=0} = \bar{\mathbf{H}}_d \mathbf{d}_k, \\ z_{f,k} &= z_k|_{f\neq 0} = \bar{\mathbf{H}}_d \mathbf{d}_k + \bar{\mathbf{H}}_f \mathbf{f}_k, \end{aligned} \tag{13}$$

respectively. Since the fault $\mathbf{f}_k$ is assumed to be unknown but deterministic, and the distribution of the stochastic disturbance $\mathbf{d}_k$ belongs to the ambiguity set $\mathbb{P}(0, \bar{\mathbf{Q}})$, which is denoted by $\mathbf{d}_k \sim \mathbb{P}(0, \bar{\mathbf{Q}})$, the distributions of $z_{0,k}$ and $z_{f,k}$ can be respectively expressed as $z_{0,k} \sim \mathbb{P}(0, \Sigma_0)$ and $z_{f,k} \sim \mathbb{P}(\bar{\mathbf{H}}_f \mathbf{f}_k, \Sigma_0)$, where the covariance matrix Σ_0 is

$$\Sigma_0 = \bar{\mathbf{H}}_d \bar{\mathbf{Q}} \bar{\mathbf{H}}_d^\top. \tag{14}$$

A predefined reference fault $\delta \mathbf{f}_{\text{ref}} \in \mathbb{R}^{n_f h}$ is introduced such that the mean of z_f is fixed at $\bar{\mathbf{H}}_f \delta \mathbf{f}_{\text{ref}}$, where $\mathbf{f}_{\text{ref}}$ is a unit vector, and the scalar $\delta > 0$ is the fault magnitude. For such a reference fault, two cases are considered:

(i) The mean of the faulty residual $r_k = w_1^\top z_k$ is positive, 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 mean of the faulty residual $r_k = w_2^\top z_k$ is negative, 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.

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

The design of w_1 and b_1 in the one-sided test (15) is formulated in the BMPM scheme:

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

$$\text{s.t.} \quad \inf_{z_0 \sim \mathbb{P}(0, \Sigma_0)} \Pr\left\{w_1^\top z_0 \leq b_1\right\} \geq \alpha, \tag{17b}$$

$$\inf_{z_f \sim \mathbb{P}(\bar{\mathbf{H}}_f \delta \mathbf{f}_{\text{ref}}, \Sigma_0)} \Pr\left\{w_1^\top z_f \geq b_1\right\} \geq \beta_1, \tag{17c}$$

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

For the one-sided test (15) in the case $w_1^\top \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}} > 0$, the chance constraints (17b) and (17c) ensure that the worst-case FAR is $1 - \alpha$, and the worst-case FDR is β_1 for the reference fault $\delta \mathbf{f}_{\text{ref}}$.

Lemma 1. (Calafiore and El Ghaoui, 2006) Assume that the distribution of a random vector ξ belongs to an ambiguity set $\mathbb{P}(\bar{\xi}, \Xi)$ that has mean $\bar{\xi}$ and covariance $\Xi > 0$. For $q \neq 0$, $\epsilon \in [0, 1)$, and a given scalar c satisfying $q^\top \bar{\xi} \leq c$, the chance constraint

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

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

By applying Lemma 1, the BMPM (17) can be equivalently transformed into the following optimization problem with deterministic constraints:

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

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

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

$$w_1^\top \bar{\mathbf{H}}_f \delta \mathbf{f}_{\text{ref}} > b_1, b_1 > 0, \tag{19d}$$

with $\kappa_\alpha = \sqrt{\frac{\alpha}{1-\alpha}}$ and $\kappa_{\beta_1} = \sqrt{\frac{\beta_1}{1-\beta_1}}$. The constraint $w_1^\top \bar{\mathbf{H}}_f \delta \mathbf{f}_{\text{ref}} > b_1$ in (19d) is imposed to ensure a positive β_1, according to Lemma 1. Then, it can be derived from (19b) and (19c) that

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

which is further simplified as

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

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

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

holds. Since κ_α is fixed, maximizing β_1 is equivalent to solving

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

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

$$\tilde{w}_1^* = \frac{\Sigma_0^{-\frac{1}{2}} \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}}}{\sqrt{\mathbf{f}_{\text{ref}}^\top \bar{\mathbf{H}}_f^\top \Sigma_0^{-1} \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}}}}, \ w_1^* = \frac{\Sigma_0^{-1} \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}}}{\sqrt{\mathbf{f}_{\text{ref}}^\top \bar{\mathbf{H}}_f^\top \Sigma_0^{-1} \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}}}}, \tag{21a}$$

$$b_1^* = \kappa_\alpha = \sqrt{\frac{\alpha}{1-\alpha}}, \tag{21b}$$

$$\kappa_{\beta_1^*} = \delta \sqrt{\mathbf{f}_{\text{ref}}^\top \bar{\mathbf{H}}_f^\top \Sigma_0^{-1} \bar{\mathbf{H}}_f \mathbf{f}_{\text{ref}}} - \kappa_\alpha, \ \beta_1^* = \frac{\kappa_{\beta_1^*}^2}{1 + \kappa_{\beta_1^*}^2}. \tag{21c}$$

Required by (19c) and (21c), the above solution has to satisfy $(w_1^*)^\top \bar{\mathbf{H}}_f \delta \mathbf{f}_{\rm ref} \geq b_1$ and $\kappa_{\beta_1^*} \geq 0$. This results in the condition

$$\delta^2 \mathbf{f}_{\rm ref}^\top \bar{\mathbf{H}}_f^\top \Sigma_0^{-1} \bar{\mathbf{H}}_f \mathbf{f}_{\rm ref} \geq \kappa_\alpha^2, \tag{22}$$

which can be ensured by increasing the fault magnitude δ. Note that the optimal solution w_1^* is related to $\mathbf{f}_{\rm ref}$, without depending on the fault magnitude δ, while b_1^* is determined solely by the predefined α.

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

$$\begin{aligned} \max_{w_2 \neq 0, \beta_2, b_2} \quad & \beta_2 \\ \text{s.t.} \quad & \inf_{z_0 \sim \mathbb{P}(0,\Sigma_0)} \Pr\left\{w_2^\top z_0 \geq b_2\right\} \geq \alpha, \\ & \inf_{z_f \sim \mathbb{P}(\bar{\mathbf{H}}_f \delta \mathbf{f}_{\rm ref}, \Sigma_0)} \Pr\left\{w_2^\top z_f \leq b_2\right\} \geq \beta_2, \\ & w_2^\top \bar{\mathbf{H}}_f \delta \mathbf{f}_{\rm ref} < 0, b_2 < 0, \end{aligned}$$

and its optimal solution is

$$w_2^* = -w_1^*,\ b_2^* = -\kappa_\alpha,\ \beta_2^* = \beta_1^* \tag{23}$$

with w_1^* and β_1^* given in (21). The two one-sided tests (21) and (23) are symmetric, thus can be combined into the following two-sided test for the residual $r_k = (w_1^*)^\top z_k$:

$$\begin{cases} |r_k| > \kappa_\alpha \ \Rightarrow \ \text{fault alarm}, \\ |r_k| \leq \kappa_\alpha \ \Rightarrow \ \text{no fault alarm}. \end{cases} \tag{24}$$

3.2 Statistical FD performance analysis

For the two-sided test (24), the residual $r_k = (w_1^*)^\top z_k$ in the fault-free case has a zero mean and variance $(w_1^*)^\top \Sigma_0 w_1^* = 1$, i.e., $r_k \sim \mathbb{P}(0,1)$. Then, the obtained FAR is upper bounded by

$$\sup_{r_k \sim \mathbb{P}(0,1)} \Pr\{|r_k| > b^*\} = \frac{1-\alpha}{\alpha} \tag{25}$$

which follows the univariate Chebyshev inequality (Navarro, 2016).

The two-sided test (24) is derived with respect to a specified reference fault $\delta \mathbf{f}_{\rm ref}$ in Section 3.1. How its worst-case FDR varies with an arbitrary fault $\mathbf{f}_k \neq \delta \mathbf{f}_{\rm ref}$ is analyzed in the following theorem.

Theorem 2. For the two-sided test (24), its worst-case FDR for any fault $\mathbf{f}_k$ different from $\delta \mathbf{f}_{\rm ref}$ is

$$\begin{cases} \dfrac{(|\mu_k| - \kappa_\alpha)^2}{1 + (|\mu_k| - \kappa_\alpha)^2} & \text{if } |\mu_k| \geq \kappa_\alpha, \\ 0 & \text{if } |\mu_k| < \kappa_\alpha, \end{cases} \tag{26}$$

with

$$\mu_k = (w_1^*)^\top \bar{\mathbf{H}}_f \mathbf{f}_k. \tag{27}$$

Proof. In the faulty case, the residual $r_k = (w_1^*)^\top z_k$ has a distribution that belongs to the ambiguity set $\mathbb{P}(\mu_k, 1)$. Using Theorem 6.1 in Bertsimas and Popescu (2005), the worst-case FDR is expressed as

$$\begin{aligned} &\inf_{r_k \sim \mathbb{P}(\mu_k,1)} \Pr\{|r_k| > \kappa_\alpha\} \\ &= 1 - \sup_{r_k \sim \mathbb{P}(\mu_k,1)} \Pr\{|r_k| \leq \kappa_\alpha\} = \frac{\eta}{1+\eta} \end{aligned} \tag{28}$$

with

$$\eta = \inf_{|r| \leq \kappa_\alpha} (r - \mu_k)^2 = \begin{cases} (|\mu_k| - \kappa_\alpha)^2 & \text{if } |\mu_k| > \kappa_\alpha, \\ 0 & \text{if } |\mu_k| \leq \kappa_\alpha. \end{cases}$$ □

As α increases, κ_α monotonically increases. Consequently, the worst-case FAR of the two-sided test (24) decreases according to (25), while the worst-case FDR for any given fault is non-increasing according to (26).

4. EXTENSION TO FD OF A VECTOR RESIDUAL

According to Theorem 2, the worst-case FDR of the two-sided test (24) could be zero when the actual fault $\mathbf{f}_k \neq \delta \mathbf{f}_{\rm ref}$ satisfies $|(w_1^*)^\top \bar{\mathbf{H}}_f \mathbf{f}_k| < \kappa_\alpha$, even though the magnitude of $\mathbf{f}_k$ is large. Motivated by this observation, the test of a scalar residual in Section 3 is extended to a vector residual to enhance FD performance.

Let

$$\mathbf{r}_k = [r_{k,1}\ r_{k,2}\ \cdots\ r_{k,n_s}]^\top \tag{29}$$

represent a vector residual whose dimension n_s is to be determined later. Each residual component $r_{k,i}$ corresponds to solving a BMPM in the form of (17) by assigning a reference fault $\delta_i \mathbf{f}_{{\rm ref},i}$, where $\mathbf{f}_{{\rm ref},i}$ is a unit vector and the scalar $\delta_i > 0$ is the fault magnitude. As a result, each residual component $r_{k,i}$ is expressed as

$$\begin{aligned} r_{k,i} &= \phi_i^\top \tilde{z}_k,\ \tilde{z}_k = \Sigma_0^{-\frac{1}{2}} z_k, \\ \phi_i &= \frac{\Sigma_0^{-\frac{1}{2}} \bar{\mathbf{H}}_f \mathbf{f}_{{\rm ref},i}}{\sqrt{\mathbf{f}_{{\rm ref},i}^\top \bar{\mathbf{H}}_f^\top \Sigma_0^{-1} \bar{\mathbf{H}}_f \mathbf{f}_{{\rm ref},i}}}, \end{aligned} \tag{30}$$

according to (11) and (21a), by replacing $\tilde{w}_1^*$ with ϕ_i. Then, the associated two-sided test of each residual component $r_{k,i}$ is in the same form of (24), i.e.,

$$\begin{cases} |r_{k,i}| > \kappa_\alpha \ \Rightarrow \ \text{fault alarm}, \\ |r_{k,i}| \leq \kappa_\alpha \ \Rightarrow \ \text{no fault alarm}. \end{cases} \tag{31}$$

Next, how to select the reference faults $\{\mathbf{f}_{{\rm ref},i}\}$ is discussed.

Proposition 3. Let the singular value decomposition (SVD) of $\Sigma_0^{-\frac{1}{2}} \bar{\mathbf{H}}_f$ be denoted by

$$\Sigma_0^{-\frac{1}{2}} \bar{\mathbf{H}}_f = [U_1\ U_2] \begin{bmatrix} S & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^\top \\ V_2^\top \end{bmatrix}. \tag{32}$$

If the reference faults $\{\mathbf{f}_{{\rm ref},i}\}$ are selected as

$$[\mathbf{f}_{{\rm ref},1}\ \mathbf{f}_{{\rm ref},2}\ \cdots\ \mathbf{f}_{{\rm ref},n_s}] = V_1 \tag{33}$$

with $n_s = \operatorname{rank}(\Sigma_0^{-\frac{1}{2}} \bar{\mathbf{H}}_f)$, then

$$\phi_i^\top \phi_j = \begin{cases} 0 & i \neq j, \\ 1 & i = j, \end{cases} \tag{34}$$

is guaranteed.

Proof. By substituting (33) into (30) and exploiting (32),

$$[\phi_1\ \phi_2\ \cdots\ \phi_{n_s}] = U_1 \tag{35}$$

can be derived to prove (34).

Proposition 3 has a clear geometric interpretation. In (30), 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 projecting $\tilde{z}_k$ onto multiple orthogonal directions $\{\phi_i\}$, and each projected variance is 1. Each residual component $r_{k,i}$ is independently monitored by the two-sided test (31), and all the n_s tests in (31) can be combined into

$$\begin{cases} \|\mathbf{r}_k\|_\infty > \kappa_\alpha \ \Rightarrow \ \text{fault alarm}, \\ \|\mathbf{r}_k\|_\infty \leq \kappa_\alpha \ \Rightarrow \ \text{no fault alarm}, \end{cases} \tag{36}$$

for the vector residual

$$\mathbf{r}_k = U_1^\top \Sigma_0^{-\frac{1}{2}} z_k, \tag{37}$$

where $\|\cdot\|_\infty$ represents the infinity norm of a vector, i.e., $\|\mathbf{r}_k\|_\infty = \max_{1\le i\le n_s} |r_{k,i}|$.

5. SIMULATION STUDY

In this section, a satellite attitude control system is considered as a simulation example. Define the system state $x = [\varphi\ \theta\ \psi\ \omega_x\ \omega_y\ \omega_z]^\top$, the system input $u = [T_x\ T_y\ T_z]^\top$, the measured output $y = [\varphi\ \theta\ \psi\ \omega_x\ \omega_y\ \omega_z]^\top$, where φ, θ, ψ denote the roll, pitch and yaw angles, $\omega_x, \omega_y, \omega_z$ represent the roll, pitch and yaw angle rates, and T_x, T_y, T_z are the roll, pitch and yaw control moments, respectively. The system model is then described by (Zhong et al., 2019)

$$\begin{cases} x(k+1) = Ax(k) + Bu(k) + d_1(k) + B_f f(k) \\ y(k) = x(k) + d_2(k) \end{cases} \tag{38}$$

$$A = \begin{bmatrix} A_1 & 0.001 I_3 \\ \mathbf{0}_{3\times 3} & A_2 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 1 & 0 & 1\times 10^{-5} \\ 0 & 1 & 0 \\ -1\times 10^{-5} & 0 & 1 \end{bmatrix},$$

$$A_2 = \begin{bmatrix} 1 & -7.568\times 10^{-8} & -3.344\times 10^{-8} \\ 1.067\times 10^{-7} & 1 & -9.183\times 10^{-8} \\ -1.434\times 10^{-8} & -2.793\times 10^{-8} & 1 \end{bmatrix},$$

$$B = \begin{bmatrix} 8.70\times 10^{-5} I_3 \\ B_1 \end{bmatrix}, \quad B_1 = 10^{-4} \times \begin{bmatrix} 8 & 0 & 0 \\ 0 & 7.299 & 0 \\ 0 & 0 & 6.289 \end{bmatrix},$$

$$B_f = 10^{-4} \times [0.87\ 0\ 0\ 8\ 0\ 0]^\top,$$

where $\mathbf{0}_{m\times n}$ represents a zero matrix of dimension $m \times n$, and I_n is an identity matrix of dimension n. In (38), $d_1 \in \mathbb{R}^6$ denotes a vector of process noises, with each element uniformly distributed over the interval $[-2.598 \times 10^{-5}, 2.598\times 10^{-5}]$; and $d_2 \in \mathbb{R}^6$ is a vector of measurement noises, which follows a normal distribution with zero mean and covariance $\mathrm{diag}\{4\times 10^{-7} I_3, 7\times 10^{-8} I_3\}$.

For performance comparisons, the following three parity relation based FD methods are implemented with the same parity order $h = 6$:

- the conventional approach: its parity matrix is determined by solving (9) as in Section 7.4 of Ding (2013), and its detection threshold is set to ensures a predefined FAR in 1000 fault-free Monte Carlo simulations.
- the BMPM approach of a scalar residual (referred to as "BMPM-scalar"): the parity vector v^* is computed according to (10) and (21a), with the chosen reference fault direction $\mathbf{f}_{\text{ref}} = [1\ 1\ 1\ 1\ 1\ 1]^\top/\sqrt{6}$. Its detection threshold $J_{th} = b^*$ only depends on the predefined α according to (21b) and it increases monotonically with α.
- the BMPM approach of a vector residual (referred to as "BMPM-vector"): with the reference faults in (33), the parity matrix is computed according to (32) and (37). Its detection threshold has the same dependence on α as in the BMPM-scalar approach.

The following two scenarios of roll momentum wheel fault f are considered:

- a constant fault: the constant amplitude is set to a value bewteen -1.6 and 1.6;
- a time-varying fault:

$$f(k) = \begin{cases} \sin(k) + \sin(2k), & 1000 \le k \le 2000 \\ \sin(\frac{2\pi k}{7}), & 3000 \le k \le 4000 \\ \sin(\frac{2\pi k}{7}) + 0.1, & 5000 \le k \le 6000 \\ 0, & \text{otherwise.} \end{cases} \tag{39}$$

In the constant fault scenario, the FDRs of the conventional approach and the BMPM-scalar approach are compared. For each simulated constant fault f, 3000 Monte-Carlo simulations are performed. For the conventional approach, the threshold is set to 2.8×10^{-3} to ensure a predefined FAR 5%. For the BMPM-scalar approach, α is predefined as 0.8, and the resulting detection threshold is $J_{th} = \kappa_\alpha = 2$. How the achieved FDRs vary with the fault amplitude is depicted in Fig. 1. When the fault amplitude equals zero, both two approaches achieve almost the same FAR. For the fault amplitude less than 1.3, the conventional approach results in much lower FDRs than the BMPM-scalar approach. Moreover, for each fault amplitude, the theoretical worst-case FDR provides a lower bound of the actual FDR achieved by the BMPM-scalar approach.

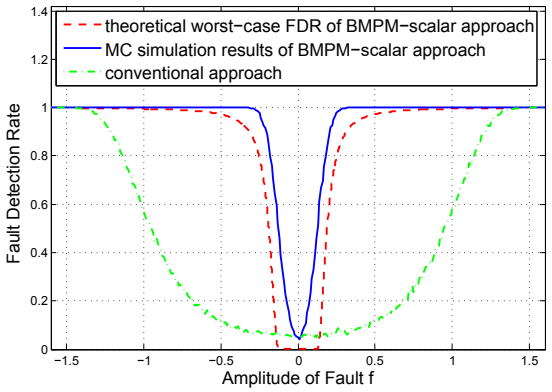

Fig. 1. FDR for the amplitude of constant fault varying from -1.6 to 1.6.

In the time-varying fault scenario, all three parity relation approaches mentioned above are implemented. The detection threshold in conventional approach is the same as in the previous constant fault scenario. For the BMPM-scalar and BMPM-vector approaches, α is set to 0.88, and the corresponding detection thresholds are both $J_{th} = \kappa_\alpha = 2.703$. The detection results are shown in Fig. 2 and Table 1. The poor FDR of the conventional approach is due to its limitations mentioned in Section 2.3. The BMPM-scalar approach also results in a low FDR, because of the mismatch between its selected reference fault and the actual fault. In contrast, the BMPM-vector approach achieves the highest FDR with an acceptable FAR.

Table 1. detection results of three approaches

	Conventional approach	BMPM-scalar	BMPM-vector
FAR(%)	4.85	1.15	4.20
FDR(%)	14.65	55.04	99.50

To further evaluate the tradeoff between FDR and FAR, the ROC (Receiver Operating Characteristic) curves of

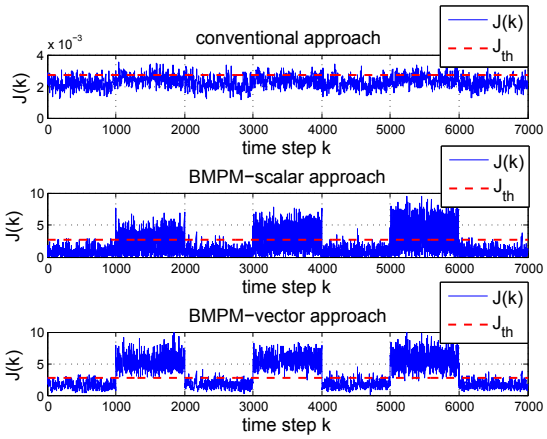

Fig. 2. Residual evaluations and thresholds of three approaches to detect the time-varying fault (39)

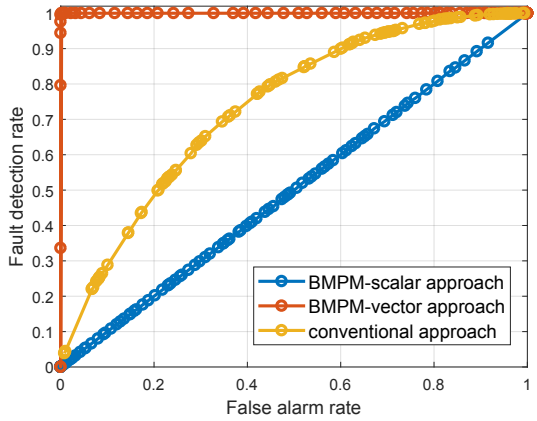

Fig. 3. ROC curves for detecting the fault $\mathbf{f} = [0.78\ 0.97\ 0.43\ -0.43\ -0.97\ -0.78\ 0]$ over a fixed time window.

these implemented FD methods are calculated over a fixed time window. For each FD method, 50 dots on its ROC curve are plotted, each of which corresponds to the FDR and FAR achieved in 2000 fault-free and faulty MC simulations. These 50 dots are determined by 50 values of the predefined FAR or α equally spaced within the interval $(0, 1)$. As depicted in Fig. 3, for the fault signal $\mathbf{f} = [0.78\ 0.97\ 0.43\ -0.43\ -0.97\ -0.78\ 0]$ within the considered time window, the proposed scalar design performs poorly, since the actual fault is orthogonal to its selected reference fault. In contrast, the proposed BMPM-vector approach performs best thanks to using multiple orthogonal reference faults.

6. CONCLUSIONS

This paper presents a BMPM approach to the probabilistic robust design of the parity relation based FD for stochastic linear systems. Over an ambiguity set of disturbance distributions with given mean and covariance, the worst-case FDR is maximized with respect to reference faults while ensuring a predefined worst-case FAR. The derived scalar residual design could result in a poor FDR if the actual fault is different from the single reference fault. Further performance enhancement is achieved by providing a systematic way to select multiple reference faults and construct a vector residual design. Detailed simulation study illustrates the efficacy of the proposed approach over the conventional parity relation approach. Future work will be devoted to analyzing the FDR of the vector residual design for arbitrary faults different from the selected reference faults.

REFERENCES

Basseville, M. and Nikiforov, I.V. (1993). *Detection of Abrupt Changes: Theory and Application*. Prentice-Hall.

Bertsimas, D. and Popescu, I. (2005). Optimal inequalities in probability theory: a convex optimization approach. *SIAM Journal on Optimization*, 15(3), 780–804.

Calafiore, G.C. and El Ghaoui, L. (2006). On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications*, 130(1), 1–22.

Combastel, C. (2016). 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*, 42, 232–243.

Ding, S.X., Li, L., and Kruger, M. (2019). Application of randomized algorithms to assessment and design of observer-based fault detection systems. *Automatica*, 107, 175–182.

Ding, S.X. (2013). *Model-based Fault Diagnosis Techniques*. Springer-Verlag London, 2nd edition.

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

Gustafsson, F. (2007). Statistical signal processing approaches to fault detection. *Annual Reviews in Control*, 31, 41–54.

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

Patton, R.J. and Chen, J. (1991). A review of parity space approach to fault diagnosis. *IFAC Proceedings Volumes*, 24, 65–81.

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

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

Zhong, M., Xue, T., Song, Y., Ding, S.X., and Ding, E.L. (2019). Parity space vector machine approach to robust fault detection for linear discrete-time systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. DOI: 10.1109/TSMC.2019.2930805.

Zhong, M., Yang, S., and Ding, S.X. (2015). Parity space-based fault detection for linear discrete time-varying systems with unknown input. *Automatica*, 59(1), 120–126.

Zhou, J., Yang, Y., Zhao, Z., and Ding, S.X. (2018). A fault detection scheme for ship propulsion systems using randomized algorithm techniques. *Control Engineering Practice*, 81, 65–72.