

Article

## Function Based Fault Detection for Uncertain Multivariate Nonlinear Non-Gaussian Stochastic Systems Using Entropy Optimization Principle

Liping Yin <sup>1,2,\*</sup> and Li Zhou <sup>1</sup>

<sup>1</sup> School of Information and Control, Nanjing University of Information Science & Technology, Ninglu Road No. 219, Pukou District, Nanjing, 210044, China

<sup>2</sup> School of Automation Science and Electrical Engineering, Beijing University of Aeronautics & Astronautics, Xueyuan Road No. 37, Haidian District, Beijing, 100191, China;  
E-Mail: lk\_zhouli@163.com

\* Author to whom correspondence should be addressed; E-Mail: lpyin@nuist.edu.cn.

Received: 26 June 2012; in revised form: 16 December 2012 / Accepted: 18 December 2012 /  
Published: 21 December 2012

---

**Abstract:** In this paper, the fault detection in uncertain multivariate nonlinear non-Gaussian stochastic systems is further investigated. Entropy is introduced to characterize the stochastic behavior of the detection errors, and the entropy optimization principle is established for the fault detection problem. The principle is to maximize the entropies of the stochastic detection errors in the presence of faults and to minimize the entropies of the detection errors in the presence of disturbances. In order to calculate the entropies, the formulations of the joint probability density functions (JPDFs) of the stochastic errors are presented in terms of the known JPDFs of both the disturbances and the faults. By using the novel performance indexes and the formulations for the entropies of the detection errors, new fault detection design methods are provided for the considered multivariate nonlinear non-Gaussian plants. Finally, a simulation example is given to illustrate the efficiency of the proposed fault detection algorithm.

**Keywords:** fault detection; multivariate stochastic systems; uncertain; entropy optimization; non-Gaussian system

---

## 1. Introduction

Fault detection (FD) has long been regarded as an important and integrated part in modern control systems for improving reliability [1–10]. The so-far obtained approaches include the system identification technique [2] and the statistic approaches based on the Bayesian theorem, likelihood methods and hypothesis test techniques [11]. Besides, filters or observers have been widely used to generate the residual signal for fault detection and estimation [2,6,8,9,12–20].

Among the above results, for dynamic stochastic systems, the filter-based approach has been shown as an effective way where generally the noises or disturbances are supposed to be Gaussian [21,22] and the filter design objective is to realize a minimum variance residual. However, in many chemical and manufacturing processes, the inputs involved are non-Gaussian [19,20,23–26]. For example, in paper making, the length of fibre and the size of filling molecular weight are important input variables, but none of them would obey a Gaussian distribution. This is simply because most random variables and random processes are bounded in paper web formation system. Actually, even for stochastic systems with Gaussian inputs, nonlinearities in the system may lead to non-Gaussian outputs. For the non-Gaussian variables (vectors), it is well-known that the expectation and variance are insufficient to characterize their statistics. As a result, new measures should be adopted. *Entropy* is a known measure that describes the average information contained in the given PDF (JPDF), which has been widely used in information, thermodynamics and control fields [6,8,27–30]. By minimizing the entropy, higher order moments can be minimized with respect to the random variables [13,26,27,31–33].

Nevertheless, since entropy is an integral operation of PDF while a PDF is a non-linear function with an integral constraint and a positive constraint, it is not easy to find the dynamical relationships between the PDFs of the input and the concerned output and further to formulate the entropies of the distribution of outputs. Besides, for random vectors, the formulation for the JPDF of the output becomes much more complicated than the case of random variables even if the transform matrix is linear. For example, in [6],  $\pi_{k+1} \in [a, b]^n$  is supposed to be a non-Gaussian continuous random vector,  $D_k \in R^{m \times n}$ , after the multivariate mapping  $\theta_{k+1} = D_k \pi_{k+1}$ , the JPDF of  $\theta_{k+1}$  should be discussed for three different cases based on  $D_k$  and  $\theta_{k+1}$  is not continuous any more except that  $D_k$  is invertible. Thus, the existing concept of entropy has to be extended. In fact, the approach presented in [6] holds only for linear systems with special form of system matrices when applied to FD.

In this paper, the filter-based FD approach for uncertain multivariate nonlinear non-Gaussian stochastic system is further investigated using the entropy optimization principle. Firstly, a novel filtering and filter-based FD framework is established to construct a residual such that the fault can be detected from the changes of the residual. Secondly, the formulations between the JPDFs of the disturbances and faults and those of the detection errors are established. The entropy optimization principles are then presented to calculate the gain matrix of the optimal FD filter.

The remainder of this paper is organized as follows. Section 2 presents the preliminaries. In Section 2.1 and Section 2.2, the nonlinear difference model and the corresponding filter model are described. In Section 2.3, the definition of entropy is introduced for the detection errors and the basic relationships are formulated between the entropy of the input and the output. The main results are given in Section 3, where the entropy optimization performance index is proposed in Section 3.1 and

its corresponding formulations for entropies are established in Section 3.2. In order to calculate the JPFDs of the detection errors, the simplified algorithms are provided in Section 3.3 and the FD filtering algorithms are finally given in Section 3.4 to compute the optimal FD filter gain using the proposed entropy optimization principle. A simple simulation example is provided in Section 4 to demonstrate the effectiveness of the main results. Conclusion is given in Section 5.

In the following, unless stated otherwise, matrices are assumed to have compatible dimensions. The identity and zero matrices are denoted by  $I$  and  $0$  respectively with appropriate dimensions. For a square matrix  $M$ , its inverse and determinant are denoted by  $M^{-1}$  and  $\det M$  respectively. For two real vectors  $v_1$  and  $v_2$ , the notation  $v_1 \succeq v_2$  means that every element of  $v_1$  is no less than the corresponding one of  $v_2$  and  $\|v_1\|_2$  is used to denote the Euclidean norm. For a multivariate nonlinear smooth function  $y = g(x)$ ,  $\frac{\partial g(x)}{\partial x}$  denotes its Jacobian matrix. For a random vector  $z$ , the formula  $P\{z \preceq \tau\}$  represents the joint probability of event  $z \preceq \tau$ ,  $F_z(\cdot)$  denotes its joint probability distribution function,  $\gamma_z(\cdot)$  denotes the corresponding joint probability density function (JPDF), and  $H(z)$ ,  $\varepsilon(z)$  are used to denote its entropy and expectation, respectively.

## 2. Preliminary

### 2.1. Plant Models

Consider a multivariate nonlinear stochastic discrete-time system described by

$$\begin{cases} x_{k+1} = F_0(x_k, \delta_k, w_k) + \Delta F \\ y_k = G(x_k, v_k) \end{cases} \quad (1)$$

where  $x_k \in R^n$  is the state,  $y_k \in R^m$  is the output,  $w_k \in R^p$  is the stochastic disturbance input,  $v_k \in R^m$  is the disturbance influencing on the output and  $\delta_k \in R^q$  is the fault to be detected.  $F_0(\cdot, \cdot, \cdot)$  is a known multivariate Borel measurable and smooth nonlinear functions of their arguments,  $\Delta F$  represents the uncertainty satisfying  $\|\Delta F\|_2 \leq \delta_0$  and  $|\det \frac{\partial F_0(\cdot, \cdot, \cdot)}{\partial w_k}| \neq 0$  holds for any  $w_k \in [a, b]^p$ .

It should be pointed out that  $\delta_k$ ,  $w_k$  and  $v_k$  are supposed to be arbitrary bounded independent random vectors rather than Gaussian ones, which is different from the existing FD approaches based on the Kalman filtering theory. It is noted that  $\delta_k$  can also represent the abrupt change of the model parameters. This model is actually a generalization of those where only additive faults or unexpected changes of model parameters are concerned.

To simplify the FD design procedures, the following assumptions are required, which can be satisfied by many practical processes.

**Assumption 1.**  $x_0$ ,  $\delta_k$ ,  $v_k$  and  $w_k$  ( $k = 0, 1, 2, \dots$ ) are bounded, mutually independent random vectors with known JPFDs denoted as  $\gamma_{x_0}(\tau)$ ,  $\gamma_\delta(\tau)$ ,  $\gamma_v(\tau)$  and  $\gamma_w(\tau)$ , respectively.  $\gamma_w(\tau) = \gamma_{w_0}(\tau) + \Delta\gamma_w(\tau)$  where  $\gamma_{w_0}(\tau)$  is a known function defined on  $[a, b]^p$  and  $\Delta\gamma_w(\tau)$  satisfies  $\|\Delta\gamma_w(\tau)\|_2 \leq \delta_1$ .

Generally speaking, two groups of approaches can be used to determine  $\gamma_{x_0}(\tau)$ ,  $\gamma_\delta(\tau)$ ,  $\gamma_v(\tau)$ , and  $\gamma_w(\tau)$ . One is the direct measurement using some advanced instruments such as the digital camera. For example, with the developments in image processing, several digital cameras have been used to

measure the distribution of the flame in the flame combustion system, which can be further transformed into the temperature distribution. The other is the kernel estimation technique based on the open loop test [34]. In some practical processes (e.g., paper web formation control and particle size distribution control), enormous data can be stored and used to analyze the model of both the disturbances and the faults, with which some probabilistic properties can also be obtained, where the involved random vectors (including both the disturbances and the faults) may also obey the non-Gaussian statistic behaviour [14,15,18,32,35].

**Assumption 2.**  $F_0(\cdot, \cdot, \cdot)$  is a known multivariate *Borel* measurable and smooth nonlinear functions of their arguments, where  $F_0(0, 0, 0) = 0$  holds.

**Assumption 3.**  $G(\cdot, \cdot) = C_k x_k + D_k v_k$ , where  $C_k \in R^{m \times n}$  and  $m \leq n$  hold. Further,  $C_k$  is with a full row rank at every sample time, and its first  $m$  columns also have full rank (otherwise, the column can be re-arranged to guarantee this assumption).  $D_k \in R^{m \times m}$  is an invertible matrix.

2.2. Filter and Error Dynamics

For the nonlinear dynamic system given by (1), the filter can be described by

$$\begin{cases} \hat{x}_{k+1} = F_0(\hat{x}_k, 0, 0) + U_k(y_k - \hat{y}_k) \\ \hat{y}_k = C_k \hat{x}_k \end{cases} \tag{2}$$

where  $U_k \in R^{n \times m}$  is the filter matrix gain to be determined. Combining (1) with (2), the resulting estimation error  $e_k = x_k - \hat{x}_k$  satisfies

$$e_{k+1} = \Phi(x_k, \delta_k, w_k) \tag{3}$$

$$= F_0(x_k, \delta_k, w_k) + \Delta F - U_k y_k - F_0(\hat{x}_k, 0, 0) + U_k \hat{y}_k \tag{4}$$

The residual signal for the fault detection is therefore defined by

$$\hat{e}_k = y_k - \hat{y}_k = C_k e_k + D_k v_k \tag{5}$$

Related to (4), the vector denoted by

$$\theta_k = -U_k y_k - F_0(\hat{x}_k, 0, 0) + U_k \hat{y}_k \tag{6}$$

can be regarded as a deterministic term with an unknown matrix gain  $U_k$ . The main difficulty is that the term  $F_0(x_k, \delta_k, w_k)$  in (4) is both nonlinear and multivariate at each sample time. As the detection error, it is supposed that  $\hat{e}_k$  is defined on  $[\alpha, \beta]^m$ , where  $\alpha$  and  $\beta$  can also be respectively chosen as  $-\infty$  and  $+\infty$ . In the following, we will establish the relationships recursively between the JPFDs of  $x_0, \delta_k, v_k$  and  $w_k$  with  $\hat{e}_k$ .

**Remark 1.** In [8],  $e_k$  is considered instead of  $\hat{e}_k$  for simplicity based on a rigorous assumption, which may lead to considerable conservations. As such, in this note,  $\hat{e}_k$  will be concerned directly for which  $\hat{e}_k$  should be affected primarily by the fault  $\delta_k$  and minimally by the disturbances  $v_k$  and  $w_k$ .

2.3. Entropy and Its Formulation

$H(z)$  can be regarded as a measure of randomness of  $z$ , which is defined as follows.

**Definition 1.** The entropy for a continuous random vector  $z$  is defined by

$$H(z) = \begin{cases} - \int_{[\alpha, \beta]^l} \gamma_z(\tau) \ln(\gamma_z(\tau)) d\tau & \gamma_z(\tau) > 0 \\ 0 & \gamma_z(\tau) = 0 \end{cases} \tag{7}$$

where  $\gamma_z(\tau)$  is denoted as the JPDF of  $z$ ,  $z \in [\alpha, \beta]^l$ .

Based on Remark 1, the main task for fault detection is to find  $U_k$  such that  $H(\hat{e}_k)$  can be influenced by  $\delta_k$  maximally and by  $w_k$  or  $v_k$  minimally.

It can be shown that  $H(\hat{e}_k)$ , as the conditional entropies of  $x_{k-1}, \delta_k, v_k, w_k, \Delta F$  and  $U_k$ , is actually a functional of  $\gamma_\delta(\tau), \gamma_v(\tau), \gamma_w(\tau)$  and  $\gamma_{x_{k-1}}(\tau)$  as well as the underdetermined gain  $U_k$ , which can be further represented by  $H(\hat{e}_k | x_{k-1}, \delta_k, w_k, v_k, \Delta F, U_k)$ . In order to calculate the entropies, the JPDFs of the errors have to be formulated in advance. The following result reveals the relationship between the JPDFs of the input and output, subject to a multivariate nonlinear mapping.

**Lemma 1.** For a multivariate Borel smooth function  $\varphi = \Gamma(\sigma)$ , if  $\sigma$  is a random vector with known JPDF  $\gamma_\sigma(\tau)$ , the JPDFs of  $\varphi$  can be given by

$$\gamma_\varphi(\tau) = \frac{d \left[ \int_{\Omega_k(\tau)} \gamma_{\sigma_k}(\rho) d\rho \right]}{d\rho} \tag{8}$$

where for a given  $\tau$ ,  $\Omega_k(\tau)$  is the definition domain of the random vector  $\sigma$  with  $\Omega_k(\tau) = \{\rho | \Gamma(\rho) \preceq \tau\}$ .

**Proof:** For a given  $\tau$ ,  $\Omega_k(\tau)$  is the set consisting of all events  $\rho$  such that  $\Gamma(\rho) \preceq \tau$ . For the concerned function,  $\varphi$  is also a random vector. Based on Probability theory [27], the joint distribution function of  $\varphi$  can be given by

$$F_\varphi(\tau) = P\{\varphi \preceq \tau\} = \int_{\Omega_k(\tau)} \gamma_{\sigma_k}(\rho) d\rho \tag{9}$$

Hence, Equation (8) can be obtained by taking derivatives on both sides of (9). Q.E.D

The next step is to consider the special relationship between  $\hat{e}_k$  and the terms  $e_k, v_k$  in (5), which is related to the linear mapping  $\varphi = \Theta\sigma$ , where  $\Theta \in R^{m \times n}$  is a compatible real matrix.

In the following, Lemma 2 and Lemma 3 will be given respectively according to whether the concerned matrix  $\Theta$  is invertible.

**Lemma 2.** If  $m = n$  and  $\Theta$  is invertible, the following relationship holds

$$\gamma_\varphi(\tau) = \gamma_\sigma(\Theta^{-1}\tau) \left| \det \Theta^{-1} \right| \tag{10}$$

**Proof:** If  $\Theta$  is invertible, it can be verified that for given  $\tau$ , there exists  $\tau_0$  such that  $\tau_0 = \Theta^{-1}\tau$ . Thus, by using Lemma 1, we have

$$\gamma_\varphi(\tau)d\tau = P\{\tau \prec \varphi \preceq \tau + d\tau\} = P\{\tau_0 \prec \sigma \preceq \tau_0 + d\tau_0\} = \gamma_\sigma(\tau_0)d\tau_0 = \gamma_\sigma(\Theta^{-1}\tau) \left| \det \Theta^{-1} \right| d\tau$$

This process also can be regarded as a generalization of the classical relationship for random variables (see, e.g., Chapter 6, [27]). Q.E.D

In the other case,  $\Theta$  is a singular matrix. Corresponding to the structure of  $C_k$ , supposing  $\Theta$  is with a full row rank. In this case, there exist a low-triangle invertible matrix  $T_1$  and an upper-triangle invertible matrix  $T_2$  such that

$$T_1\Theta T_2 = \begin{bmatrix} I_m & 0 \end{bmatrix} \tag{11}$$

where  $I_m$  represents an  $m$ -dimensional identical matrix and  $T_2$  is invertible with positive diagonal elements.

To simplify the presentation, we denote

$$\tilde{\varphi} := \begin{bmatrix} \tilde{\varphi}^{(1)} \\ \tilde{\varphi}^{(2)} \end{bmatrix} = T_1\varphi, \tilde{\sigma} := T_2^{-1}\sigma = \begin{bmatrix} \tilde{\sigma}^{(1)} \\ \tilde{\sigma}^{(2)} \end{bmatrix}, \tilde{\tau} := T_1\tau = \begin{bmatrix} \tilde{\tau}^{(1)} \\ \tilde{\tau}^{(2)} \end{bmatrix} \tag{12}$$

with compatible dimensions. This implies

$$\tilde{\varphi} = \begin{bmatrix} I_m & 0 \end{bmatrix} \tilde{\sigma} = \sigma^{(1)} \tag{13}$$

**Lemma 3.** If  $\Theta \in R^{m \times n}$  is with full row rank and  $m < n$ , then for the linear mapping  $\varphi = \Theta\sigma$ , the following relationship holds

$$\gamma_\varphi(\tau) = \int_{[a,b]^{n-m}} \gamma_\sigma(\eta) \left| \det T_1 \right| \left| \det T_2 \right| d\tilde{\tau}^{(2)} \tag{14}$$

with the auxiliary integral argument  $\eta$  being denoted by  $\eta := \left[ (T_2 T_1 \tau)^T \quad (T_2 \tilde{\tau}^{(2)})^T \right]^T$ , where  $T_1$  and  $T_2$  are denoted by (11) and  $\tilde{\tau}^{(2)}$  is denoted by (12).

**Proof:** To complete the required proof, there are three tasks to be achieved: (1) find  $\gamma_{\tilde{\sigma}}$  in terms of  $\gamma_\sigma$ ; (2) formulate  $\gamma_{\tilde{\varphi}}$  in terms of  $\gamma_{\tilde{\sigma}}$ ; (3) represent  $\gamma_\varphi$  in terms of  $\gamma_{\tilde{\varphi}}$ .

Based on Lemma 1, task 1 and 3 can be solved as follows. From (12), it can be shown that the JPDFs of  $\tilde{\varphi}$  and  $\tilde{\sigma}$  should satisfy

$$\gamma_\varphi(\tau) = \gamma_{\tilde{\varphi}}(T_1\tau) \left| \det T_1 \right|, \gamma_{\tilde{\sigma}}(\epsilon) = \gamma_\sigma(T_2\epsilon) \left| \det T_2 \right| \tag{15}$$

On the other hand, (13) implies that

$$P\{\tilde{\varphi} \preceq \tilde{\tau}^{(1)}\} = P\{\tilde{\sigma}^{(1)} \preceq \tilde{\tau}^{(1)}, \tilde{\sigma}^{(2)} \in [a, b]^{n-m}\} \tag{16}$$

It is noted that both  $\tilde{\tau}^{(1)}$  and  $\tilde{\sigma}^{(1)}$  are  $m$ -dimensional sub-vectors as defined by (12). Thus, we have

$$F_{\tilde{\varphi}}(\tilde{\tau}^{(1)}) = F_{\tilde{\sigma}^{(1)}}(\tilde{\tau}^{(1)}) = \int_{[a,b]^{n-m}} F_{\tilde{\sigma}}(\tilde{\tau}) d\tilde{\tau}^{(2)}$$

which means that

$$\begin{aligned} \gamma_{\tilde{\varphi}}(\tilde{\tau}^{(1)}) &= \frac{\partial F_{\tilde{\varphi}}(\tilde{\tau}^{(1)})}{\partial \tilde{\tau}^{(1)}} = \frac{\partial F_{\tilde{\sigma}^{(1)}}(\tilde{\tau}^{(1)})}{\partial \tilde{\tau}^{(1)}} \\ &= \gamma_{\tilde{\sigma}^{(1)}}(\tilde{\tau}^{(1)}) = \int_{[a,b]^{n-m}} \gamma_{\tilde{\sigma}}(\tilde{\tau}) d\tilde{\tau}^{(2)} \end{aligned} \tag{17}$$

As such, task 2 is completed. Combining (15) with (17) yields (14). Q.E.D

**Remark 2.** In Lemma 3, if  $m = n$ , according to matrix theory,  $T_2 = T_1'$  holds. In this case,  $\Theta = (T_2 T_1)^{-1}$  and

$$\begin{aligned} \gamma_{\varphi}(\tau) &= \gamma_{\sigma}(\eta) |\det T_1| |\det T_2| \\ &= \gamma_{\sigma}(T_2 T_1 \tau) |\det T_2 T_1| \\ &= \gamma_{\sigma}(\Theta^{-1} \tau) |\det \Theta^{-1}| \end{aligned}$$

which implies that Lemma 3 is consistent with Lemma 2 when both the row rank and the column rank of  $\Theta$  is  $m$ .

### 3. Main Results

#### 3.1. Performance Indexes

Following the above procedures, the fault detection objective can be performed by judging the changes of  $\hat{e}_k$  at every sample time, where the entropies of  $\hat{e}_k$  only in the presence of  $v_k$  and  $w_k$  are minimized, but the entropies resulting from  $\delta_k$  are maximized. After the corresponding performance indexes are established, the involved filter gain can be designed based on the entropy optimization principle.

Generally speaking, there are two main tasks to finish in the following procedures. The first one is to provide appropriate optimization principles to achieve the FD objectives, which are represented by the entropies of  $\hat{e}_k$ . The second one is to formulate the entropies and PDFs of  $\hat{e}_k$  in terms of the PDFs of  $x_0$ ,  $v_k$ ,  $w_k$  and  $x_k$ ,  $\delta_k$ , as well as  $\Delta F$ .

In this subsection, we will focus on the entropy optimization principle. The JPDPF of  $\hat{e}_k(e_k)$  in the presence of  $v_k$ ,  $w_k$  and  $\Delta F$  and in the absence of  $\delta_k$  can be represented by  $\gamma_{\hat{e}_k}^0(\gamma_{e_k}^0)$ , while that in the presence of  $\delta_k$ ,  $\Delta F$  and in the absence of  $v_k$ ,  $w_k$  can be represented by  $\gamma_{\hat{e}_k}^1(\gamma_{e_k}^1)$ . Similarly, we denote  $H^0(\hat{e}_k \mid x_{k-1}, v_k, w_k, U_k, \Delta F)$  ( $H^1(\hat{e}_k \mid x_{k-1}, \delta_k, U_k, \Delta F)$ ) as the entropy of  $\hat{e}_k$  in the presence of  $v_k$ ,  $w_k$ ,  $\Delta F$  and in the absence of  $\delta_k$  (in the presence of  $\delta_k$ ,  $\Delta F$  and in the absence of  $v_k$ ,  $w_k$ ). Concretely it can be further denoted that for  $i = 0, 1$ ,

$$H^i(\hat{e}_k) = \begin{cases} - \int_{[\alpha,\beta]^m} \gamma_{\hat{e}_k}^i(\tau) \ln(\gamma_{\hat{e}_k}^i(\tau)) d\tau & \gamma_{\hat{e}_k}^i(\tau) > 0 \\ 0 & \gamma_{\hat{e}_k}^i(\tau) = 0 \end{cases} \tag{18}$$

where  $H^i(\hat{e}_k)$  ( $i = 0, 1$ ) represents the conditional entropies  $H^0(\hat{e}_k \mid x_{k-1}, v_k, w_k, U_k, \Delta F)$  and  $H^1(\hat{e}_k \mid x_{k-1}, \delta_k, U_k, \Delta F)$  respectively for brevity.

Based on (2), (4), (5) and Assumption 3, it can be claimed that only the first  $m$  rows of  $U_k$  are related to  $H^i(\hat{e}_{k+1})$  while the last  $n - m$  rows are irrelevant to the entropy. This means that in this case,

$U_{k,m+1}, \dots, U_{k,n}$  in (19) are redundant vectors for the optimization, where  $U_{k,m+1} = \dots = U_{k,n} = 0$  can be selected under this circumstance.

Since  $U_k$  is an  $n \times m$  matrix, in order to use conventional optimization techniques, we denote

$$U_k = [U_{k1}^T \cdots U_{kn}^T]^T, u_k = [U_{k1} \cdots U_{kn}]^T \quad (19)$$

where  $U_{ki}$  is the row vector of  $U_k$  and thus  $u_k \in R^{mn \times 1}$  is a stretched column vector.

The entropy optimization performance index is proposed as follows

$$J_N(\hat{e}_k) := \sum_{k=1}^N J(\hat{e}_k, x_{k-1}, \delta_k, w_k, v_k, u_k) = \sum_{k=1}^N \left[ -H^0(\hat{e}_k) + \frac{1}{2} u_k^T R_1 u_k + R_2 H^1(\hat{e}_k) \right] \quad (20)$$

where  $R_1$  and  $R_2$  are pre-specified matrix weight and constant weight respectively.

If there exists a filter such that  $J_N(\hat{e}_k)$  is minimized for each sample time  $N$ , then it is called as an *entropy optimization FD filter*.

**Remark 3.** It is noted that entropy optimization is corresponding to variance optimization for Gaussian signals [32]. For example, the entropy optimization principle for FD problems is in parallel to the main results in [2], where linear Gaussian systems were studied and the minimax technology was applied to the variance of errors.

**Remark 4.** To enhance the FD filtering performance, the expectation vector  $\varepsilon(\hat{e}_k)$  can also be included in the cost function described in  $J_N(\hat{e}_k)$ . We can consider the more general performance indexes than  $J_N(\hat{e}_k)$  where the entropies of the error resulting from  $x_0$ ,  $v_k$ ,  $w_k$  and  $\delta_k$  are considered simultaneously. However, noting that there is no added difference between the two performance indexes in optimization context, we will focus on  $J_N(\hat{e}_k)$  for the FD problem in the following subsections.

**Remark 5.** In the entropy optimization performance index,  $R_1$  and  $R_2$  are two pre-specified weights.  $R_1$  corresponds to the energy. If there is no restriction on the energy,  $R_1$  can be selected as 0. Generally, the smaller  $R_1$  is, the easier the fault can be detected.  $R_3$  can also be added as the weight of  $H^0(\hat{e}_k)$ . In that case  $R_2$  and  $R_3$  are relative, so we set  $R_3 = 1$  and regulate  $R_2$ . It is noted that the cumulative performance index is adopted in this paper. If there is no sum operation, the index corresponds to the “greedy” one [36]. However, it is shown that the use of long-range prediction can overcome the problem of stabilizing a non-minimum phase plant with unknown or variable dead-time. As such, in this paper, (20) is employed. Moreover, Entropy optimization is equivalent to the variance optimization for Gaussian signals. It can be verified that the entropy optimization is consistent with the results for linear Gaussian systems.

### 3.2. Formulations for the Error JPDPs

In this subsection, the formulations for the JPDPs related to the FD performance index will be discussed. The JPDPs  $\gamma_{e_k}^i$  have to be calculated in order to formulate  $\gamma_{\hat{e}_k}^i$  based on Lemmas 1–3, and then to provide the representation of  $H^i(\hat{e}_k)$  ( $i = 0, 1$ ).

For a given  $\tau \in [\alpha, \beta]^n$ , the domains of definition for several involved functions are defined as follows

$$\Pi_k^0 := \{(\tau_1, \tau_2) \mid \Phi(\tau_1, 0, \tau_2) \preceq \tau\}, \Pi_k^1 := \{(\tau_1, \tau_2) \mid \Phi(\tau_1, \tau_2, 0) \preceq \tau\} \tag{21}$$

$$\Xi_k^0 := \{(\tau_1, \tau_2) \mid F(\tau_1, 0, \tau_2) \preceq \tau\}, \Xi_k^1 := \{(\tau_1, \tau_2) \mid F(\tau_1, \tau_2, 0) \preceq \tau\} \tag{22}$$

where the notation  $\preceq$  have been defined in the ending of the Introduction section.

To summarize, the following result can be obtained.

**Theorem 1.** Under Assumptions 1-3,  $\gamma_{e_k}^0(\tau)$  and  $\gamma_{e_k}^1(\tau)$  ( $k = 1, 2, \dots$ ) can be calculated from the differentials on the multiple integrations as follows:

$$\gamma_{e_k}^0(\tau) = \frac{d \left[ \int \int_{\Pi_k^0(\tau)} \gamma_{x_{k-1}, w}(\tau_1, \tau_2) d\tau_1 d\tau_2 \right]}{d\tau}, \gamma_{e_k}^1(\tau) = \frac{d \left[ \int \int_{\Pi_k^1(\tau)} \gamma_{x_{k-1}, \delta}(\tau_1, \tau_2) d\tau_1 d\tau_2 \right]}{d\tau}$$

where

$$\gamma_{x_k}^0(\tau) = \frac{d \left[ \int \int_{\Xi_k^0(\tau)} \gamma_{x_{k-1}, w}(\tau_1, \tau_2) d\tau_1 d\tau_2 \right]}{d\tau}, \gamma_{x_k}^1(\tau) = \frac{d \left[ \int \int_{\Xi_k^1(\tau)} \gamma_{x_{k-1}, \delta}(\tau_1, \tau_2) d\tau_1 d\tau_2 \right]}{d\tau},$$

and  $\Pi_k^i(\tau), \Xi_k^i(\tau)$  ( $i = 0, 1$ ) are denoted respectively by (21) and (22).

**Proof:** This result can be obtained similarly to Lemma 1 using the independence property in probability theory. Q.E.D

For the FD problem,  $\hat{e}_k$  is used as the residual to detect the fault. In order to calculate the entropies of  $\gamma_{\hat{e}_k}^i(\tau)$  ( $i = 0, 1$ ), we denote two auxiliary vectors as follows

$$s_{1k} = C_k e_k, s_{2k} = D_k v_k$$

and correspondingly to (11), we suppose that there exist  $T_{1k}$  and  $T_{2k}$  such that

$$T_{1k} C_k T_{2k} = \begin{bmatrix} I_m & 0 \end{bmatrix}$$

**Theorem 2.** Under Assumptions 1-3,  $\gamma_{\hat{e}_k}^i(\tau)$  ( $i = 0, 1; k = 0, 1, 2, \dots$ ) can be calculated from the following

$$\gamma_{\hat{e}_k}^i(\tau) = \int_{[\alpha, \beta]^m} \gamma_{s_{1k}}^i(\tau - \tau_1) \gamma_{s_{2k}}(\tau_1) d\tau_1$$

where

$$\gamma_{s_{1k}}^i(\tau) = \left[ \int_{[\alpha, \beta]^{n-m}} \gamma_{e_k}^i(\eta) |\det T_{1k}| |\det T_{2k}| d\tilde{\tau}^{(2)} \right]$$

and

$$\gamma_{s_{2k}}(\tau) = \gamma_v \left( D_k^{-1} \tau \right) \left| \det D_k^{-1} \right|$$

**Proof:** The formulations of  $\gamma_{s_{1k}}^i(\tau)$  can be obtained using Lemma 2 and Lemma 3. Noting  $\hat{e}_k = s_{1k} + s_{2k}$ , the result can be provided. Q.E.D

### 3.3. Simplified Calculation of the Error JPDFs

In the above results, the formulations of the entropies have been reduced to the differentials of some multiple integrals that depend on their integral domains. To simplify the design methods, the following assumption is introduced.

**Assumption 4.**  $\det\left[\frac{\partial F(x_k, \delta_k, w_k)}{\partial x_k}\right] \neq 0$  holds and  $\det\left[\frac{\partial \Phi(x_k, \delta_k, w_k)}{\partial x_k}\right] \neq 0$  holds for any  $x_k$  in their operation domain.

Different from *Theorem 1*, under *Assumption 4*, simplified algorithms can be further obtained to calculate  $\gamma_{e_k}^i(\tau)$  and then to formulate  $\gamma_{e_k}^i(\tau)$  ( $i = 0, 1$ ) by constructing the following auxiliary multivariate functions.

Under *Assumption 4*, for functions  $\Phi(\cdot, \cdot, \cdot)$  and  $F(\cdot, \cdot, \cdot)$  and a given vector  $\tau_1$ , there exist the inverse functions denoted by

$$\delta = \Phi^{-1}(\tau_1, \tau, 0), w = \Phi^{-1}(\tau_1, 0, \tau) \tag{23}$$

respectively, which satisfy

$$\tau = \Phi(\tau_1, \delta, 0), \tau = \Phi(\tau_1, 0, w)$$

Construct the auxiliary vectors for  $\gamma_{e_{k+1}}$  as

$$\eta_{k+1}^0 = \begin{bmatrix} w_{k+1} \\ e_{k+1} \end{bmatrix}, \Psi^0(w_{k+1}, x_k) = \begin{bmatrix} w_{k+1} \\ \Phi(x_k, 0, w_{k+1}) \end{bmatrix} \tag{24}$$

$$\eta_{k+1}^1 = \begin{bmatrix} \delta_{k+1} \\ e_{k+1} \end{bmatrix}, \Psi^1(\delta_{k+1}, x_k) = \begin{bmatrix} \delta_{k+1} \\ \Phi(x_k, \delta_{k+1}, 0) \end{bmatrix} \tag{25}$$

And for  $\gamma_{x_{k+1}}$ , we denote

$$\xi_{k+1}^0 = \begin{bmatrix} w_{k+1} \\ x_{k+1} \end{bmatrix}, \Upsilon^0(w_{k+1}, x_k) = \begin{bmatrix} w_{k+1} \\ F(x_k, 0, w_{k+1}) \end{bmatrix} \tag{26}$$

$$\xi_{k+1}^1 = \begin{bmatrix} \delta_{k+1} \\ x_{k+1} \end{bmatrix}, \Upsilon^1(\delta_{k+1}, x_k) = \begin{bmatrix} \delta_{k+1} \\ F(x_k, \delta_{k+1}, 0) \end{bmatrix} \tag{27}$$

where  $\Phi(\cdot, \cdot, \cdot)$  is denoted by (3). With (25) and (24), the dynamics of estimation errors in the presence of  $\delta_{k+1}$  ( $w_{k+1}$ ) and in the absence of  $w_{k+1}$  ( $\delta_{k+1}$ ) can also be represented respectively by

$$\eta_{k+1}^0 = \Psi^0(w_{k+1}, x_k), \eta_{k+1}^1 = \Psi^1(\delta_{k+1}, x_k), \xi_{k+1}^0 = \Upsilon^0(w_{k+1}, x_k), \xi_{k+1}^1 = \Upsilon^1(\delta_{k+1}, x_k) \tag{28}$$

Using the auxiliary vectors and functions given by (25), (24), (27), (26) and (28), the following results can be obtained.

**Theorem 3.** Under *Assumptions 1–4*,  $\gamma_{e_k}^0(\tau)$  and  $\gamma_{e_k}^1(\tau)$  can be calculated by ( $k = 0, 1, 2, \dots$ )

$$\gamma_{e_{k+1}}^0(\tau) = \int_{[\alpha, \beta]^p} \gamma_w(\tau_1) \gamma_{x_k}^0(\Phi^{-1}(\tau_1, 0, \tau)) \psi_0(\tau_1, \tau) d\tau_1 \tag{29}$$

and

$$\gamma_{e_{k+1}}^1(\tau) = \int_{[\alpha, \beta]^q} \gamma_\delta(\tau_1) \gamma_{x_k}^1(\Phi^{-1}(\tau_1, \tau, 0)) \psi_1(\tau_1, \tau) d\tau_1 \tag{30}$$

where

$$\gamma_{x_{k+1}}^0(\tau) = \int_{[\alpha, \beta]^p} \gamma_w(\tau_1) \gamma_{x_k}^0(F^{-1}(\tau_1, 0, \tau)) \psi_0(\tau_1, \tau) d\tau_1 \tag{31}$$

$$\gamma_{x_{k+1}}^1(\tau) = \int_{[\alpha, \beta]^q} \gamma_\delta(\tau_1) \gamma_{x_k}^1(F^{-1}(\tau_1, \tau, 0)) \psi_1(\tau_1, \tau) d\tau_1 \tag{32}$$

and

$$\psi_0(\tau_1, \tau) := \left| \det \frac{\partial F(\tau_1, 0, \tau)}{\partial \tau_1} \right|^{-1}, \psi_1(\tau_1, \tau) := \left| \det \frac{\partial F(\tau_1, \tau, 0)}{\partial \tau_1} \right|^{-1} \tag{33}$$

**Proof:** Firstly, from (25) and (28), it can be claimed that  $\gamma_{e_{k+1}}^1(\tau) = \int_{[\alpha, \beta]^q} \gamma_{\eta_{k+1}}^1(\tau_1, \tau) d\tau_1$ . It is noted that  $\Psi^1(\delta_{k+1}, x_k)$  is an one-to-one mapping from  $(\delta_{k+1}, x_k)$  to  $(\delta_{k+1}, e_{k+1})$  under Assumption A.4. Thus, the PDF of  $(\delta_{k+1}, e_{k+1})$  can be represented by that of  $(\delta_{k+1}, x_k)$  (see e.g., [27]). On the other hand, random vectors  $x_k$  and  $\delta_{k+1}$  can be regarded as mutually independent ones under Assumption A.1. As such, based on the special structure described by (25) and the notations in (23), it can be verified that

$$\begin{aligned} \gamma_{\eta_{k+1}}^1(\tau_1, \tau) &= \gamma_{\delta_{k+1}x_k}(\tau_1, \delta) \left| \det \frac{\partial F(\tau_1, \tau, 0)}{\partial \tau_1} \right|^{-1} = \gamma_{\delta_{k+1}}(\tau_1) \gamma_{x_k}(\Phi^{-1}(\tau_1, \tau, 0)) \left| \det \frac{\partial F(\tau_1, \tau, 0)}{\partial \tau_1} \right|^{-1} \\ &= \gamma_{\delta_{k+1}}(\tau_1) \gamma_{x_k}(\Phi^{-1}(\tau_1, \tau, 0)) \psi_1(\tau_1, \tau) \end{aligned} \tag{34}$$

which means that (30) holds. Similarly, by considering the auxiliary vector and function described by  $\eta_{k+1}^0 = \Psi^0(w_{k+1}, x_k)$ , we can obtain (32). This procedure can also be used to prove (29) and (31). Q.E.D

With  $\gamma_{e_k}^i(\tau)$  ( $i = 0, 1$ ),  $\gamma_{\hat{e}_k}^i(\tau)$  can be obtained by using Theorem 2 so that  $H^i(\hat{e}_k)$  can also be formulated for  $J_N(\hat{e}_k)$ .

### 3.4. Optimal FD Filter Design Strategy

Based on the above procedures, it can be claimed that the optimization is required for the performance index  $J_N(\hat{e}_k)$ .

It is noted that (20) leads to

$$J_k(\hat{e}_k) = J_{k-1}(\hat{e}_k) + \left[ \Psi(u_k) + \frac{1}{2} u_k^T R_1 u_k \right], k = 0, 1, 2, \dots, +\infty \tag{35}$$

where

$$\Psi(u_k) := \Psi(\hat{e}_k, u_k) = -H^0(\hat{e}_k) + R_1 H^1(\hat{e}_k) \tag{36}$$

The optimal filtering strategy can be obtained through

$$\frac{\partial \left[ \Psi(\hat{e}_k, u_k) + \frac{1}{2} u_k^T R_1 u_k \right]}{\partial u_k} = 0 \tag{37}$$

where an explicit function from other arguments to  $u_k$  can be further determined. The principle of optimality can thus result in the optimal FD filtering law for the whole process. To simplify the filter structure, a recursive design procedure is formulated in the following.

Denote

$$u_k = u_{k-1} + \Delta u_k, k = 0, 1, 2, \dots, N, \dots, +\infty \tag{38}$$

As a function of  $e_k, w_k$  and  $\tau$ , it can be approximated to give

$$\Psi(\hat{e}_k, u_k) = \Psi_{k0} + \Psi_{k1}\Delta u_k + \frac{1}{2}\Delta u_k^T \Psi_{k2}\Delta u_k + o(\Delta u_k^T u_k) \tag{39}$$

where

$$\Psi_{k0} := \Psi_k(u_k)|_{u_k=u_{k-1}}, \Psi_{k1} := \frac{\partial \Psi_k(u_k)}{\partial u_k} \Big|_{u_k=u_{k-1}}, \Psi_{k2} := \frac{\partial^2 \Psi_k(u_k)}{\partial u_k^2} \Big|_{u_k=u_{k-1}}$$

The recursive algorithm can be provided to determine the gain of the entropy optimization FD filter.

**Theorem 4.** An entropy optimization FD filtering strategy for  $J_N$  subject to nonlinear error model (4) is given by

$$\Delta u_k^* = -\frac{\Psi_{k1} + R_1 u_{k-1}}{\Psi_{k2} + R_1} \tag{40}$$

for a weight  $R_1 > 0$  satisfying

$$\Psi_{k2} + R_1 > 0 \tag{41}$$

**Proof:** Firstly, it can be seen that

$$R_1 u_k^2 = u_{k-1}^T R_1 u_{k-1} + 2u_{k-1}^T R_1 \Delta u_k + \Delta u_k^T R_1 \Delta u_k \tag{42}$$

Substituting (39) and (42) into (37) yields recursive suboptimal control law (40) for all  $k = 0, 1, 2, \dots, +\infty$ , under condition (41). It should be pointed out that the above algorithm given by (40) results from a necessary condition for optimization. To guarantee the sufficiency, the following second-order derivative should also be satisfied

$$\frac{\partial^2 [\Psi_k(u_k) + \frac{1}{2}u_k^T R_1 u_k]}{\partial \Delta u_k^2} = \Psi_{k2} + R_1$$

which can be guaranteed if  $R_1$  is selected sufficiently large.

Q.E.D

**Remark 6.** The real-time suboptimal FD filter design algorithm can be summarized as follows:

- Initialize  $x_0, \hat{x}_0 = \varepsilon(x_0)$  and  $u_0$ ;
- At the sample time  $k$ , compute  $\gamma_{e_{k+1}}^i(\tau), \tau \in [\alpha, \beta]^n$  based on Theorem 3;
- At the sample time  $k$ , compute  $\gamma_{\hat{e}_{k+1}}^i(\tau), \tau \in [\alpha, \beta]^m$  and then obtain  $H^i(\hat{e}_{k+1}) (i = 0, 1)$  via (18);
- Calculate  $\Delta u_k$  and  $u_k$  using equation (40) and (38);
- Increase  $k$  by 1 to the next step.

The schematic diagram and flow chart are given in Figures 1 and 2 to illustrate how to apply our function based fault detection method to multivariate uncertain non-Gaussian system step by step.

Figure 1. Function based fault detection.

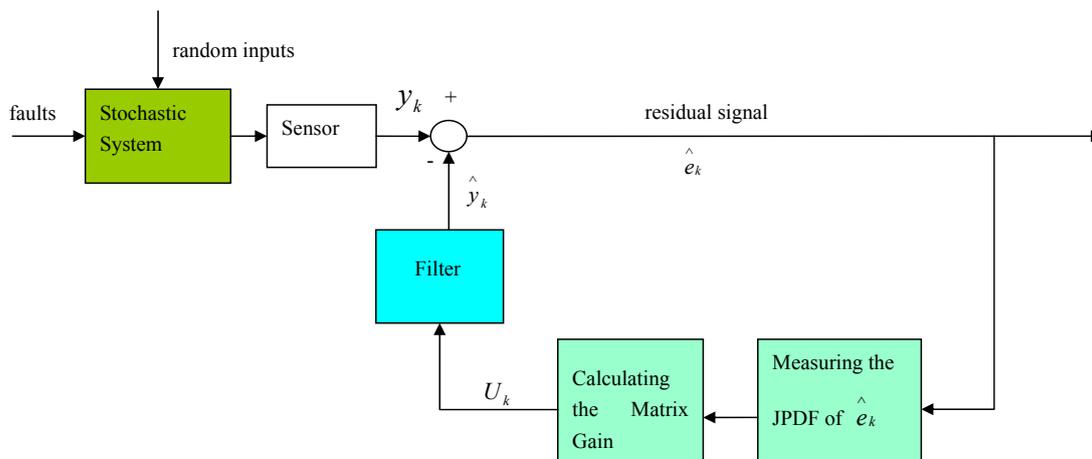
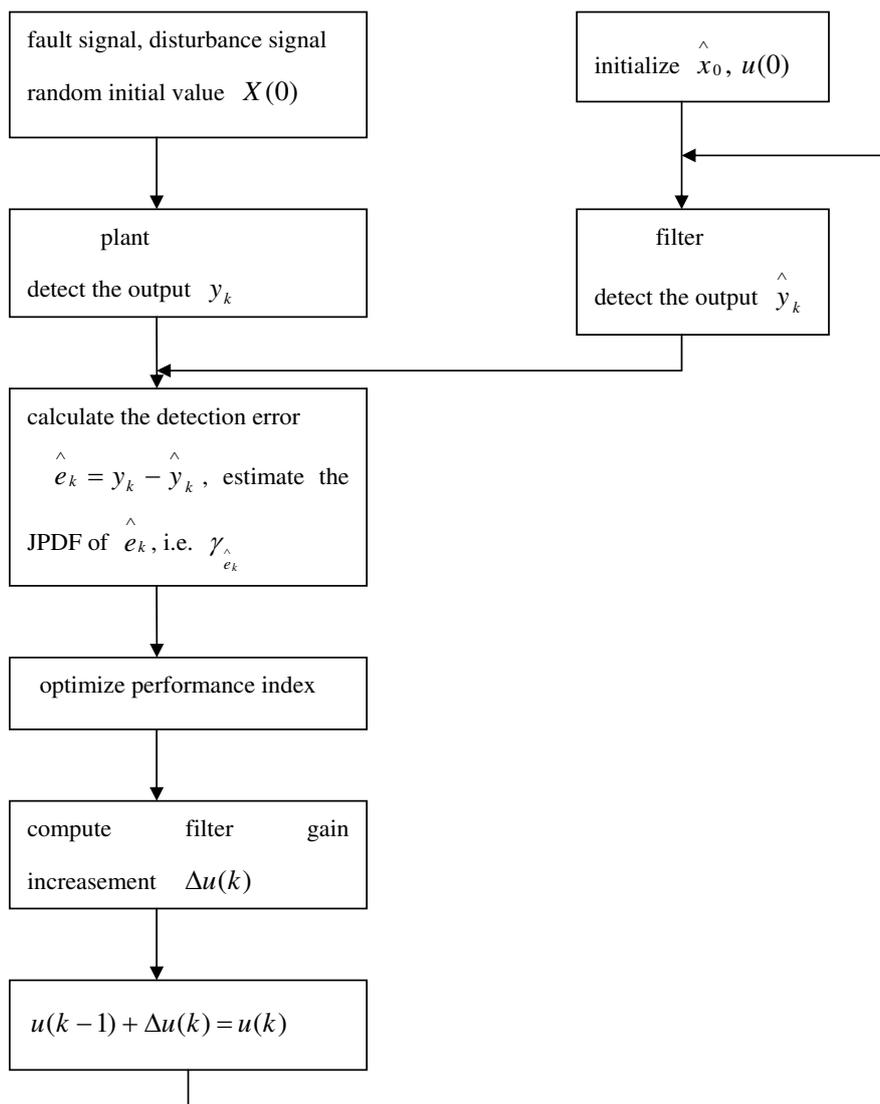


Figure 2. Flow chart of function based fault detection.



### 4. Simulation Results

To demonstrate the FD algorithm, we consider a nonlinear non-Gaussian model described as

$$\begin{cases} \begin{bmatrix} x_{1,k+1} \\ x_{2,k+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & a_0(k) \end{bmatrix} \begin{bmatrix} x_{1,k} \\ x_{2,k} \end{bmatrix} + \begin{bmatrix} \sin x_{1,k} \\ \cos^2 x_{2,k} \end{bmatrix} + \begin{bmatrix} 1 \\ b_0(k) \end{bmatrix} \delta_{k+1} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \omega_{k+1} + \Delta F \\ y_k = x_{1,k} + x_{2,k} + v_k \end{cases}$$

where  $a_0(k) = -\arctan(1+k)^{-1} + 0.5$ ,  $b_0(k) = 0.02(k+1)^{-1/2}$ , the fault  $\delta_k$  and the disturbances  $\omega_k, v_k (k = 1, 2, \dots)$  are assumed to be mutually independent. The asymmetric PDF of  $\delta_k$  is defined by

$$\gamma_\delta(x) = \begin{cases} -48(x^2 - x + \frac{3}{16}) & x \in [0.25, 0.75] \\ 0 & x \in (-\infty, 0.25) \cup (0.75, +\infty) \end{cases}$$

The PDF of the disturbance  $w_k$  is supposed to have a perturbation, i.e.,  $\gamma_w(x) = \gamma_{w_0}(x) + \Delta\gamma_w$  where

$$\gamma_{w_0}(x) = \begin{cases} -\frac{3000(x^2-0.01)}{4} & x \in [-0.1, 0.1] \\ 0 & x \in (-\infty, -0.1) \cup (0.1, +\infty) \end{cases} \text{ and } |\Delta\gamma_w| \leq 0.01.$$

The disturbance  $v_k$  is a Gaussian variable with the distribution  $N(0, 0.01)$ .

According to (2), the filter can be formed as follows:

$$\begin{cases} \begin{bmatrix} \hat{x}_{1,k+1} \\ \hat{x}_{2,k+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & a_0(k) \end{bmatrix} \begin{bmatrix} \hat{x}_{1,k} \\ \hat{x}_{2,k} \end{bmatrix} + \begin{bmatrix} \sin \hat{x}_{1,k} \\ \cos^2 \hat{x}_{2,k} \end{bmatrix} + \begin{bmatrix} u_k \\ 0 \end{bmatrix} (y_k - \hat{y}_k) \\ \hat{y}_k = \hat{x}_{1,k} + \hat{x}_{2,k} \end{cases}$$

Thus, the estimation error system is driven by

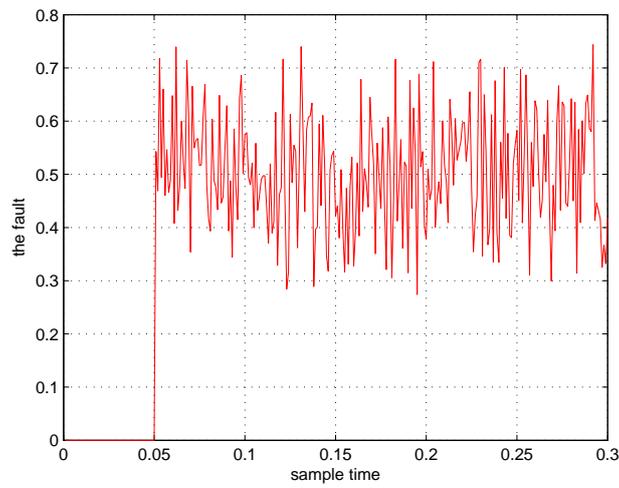
$$\begin{aligned} e_{k+1} &= \begin{bmatrix} x_{1,k} + \sin x_{1,k} \\ a_0(k)x_{2,k} + \cos^2 x_{2,k} \end{bmatrix} + \begin{bmatrix} 1 \\ b_0(k) \end{bmatrix} \delta_{k+1} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \omega_{k+1} - \begin{bmatrix} \hat{x}_{1,k} + \sin \hat{x}_{1,k} \\ a_0(k)\hat{x}_{2,k} + \cos^2 \hat{x}_{2,k} \end{bmatrix} \\ &\quad - \begin{bmatrix} u_k \\ 0 \end{bmatrix} (y_k - \hat{y}_k) + \Delta F. \end{aligned}$$

and thus

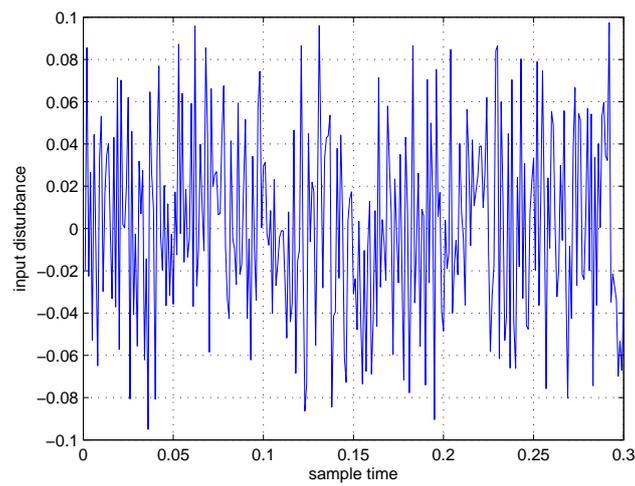
$$\begin{aligned} \hat{e}_{k+1} &= y_{k+1} - \hat{y}_{k+1} \\ &= x_{1,k+1} + x_{2,k+1} + v_{k+1} - \hat{x}_{1,k+1} - \hat{x}_{2,k+1} \\ &= e_{1,k+1} + e_{2,k+1} + v_{k+1} \end{aligned}$$

In the simulations,  $x_0$  is set to be under uniform distribution on the interval  $[0, 1]^2$  and thus  $\hat{x}_0 = \varepsilon(x_0) = [0.5, 0.5]^T$ ,  $u_0$  is set to be  $[0, 0]^T$ , the weights are set to be  $R_1 = 1$  and  $R_2 = 0.1$ . Figures 3–5 show the signals for the fault and disturbances, while Figures 6–8 show the corresponding pdfs of the signal. Figures 9 and 10 are provided to demonstrate the residual signals  $\hat{e}_k$  in the absence and in the presence of fault emerging respectively. When the random fault occurs, it is shown that the detection error increases significantly. Figure 11 is the 3-D mesh of the system output  $y$ , which shows a remarkable change of the pdf  $\gamma_y$  when the fault occurs at sample time 50. Figure 12 is the optimization performance index along time  $t$ . The result shows that a satisfactory FD performance can be obtained through the error dynamics by optimizing the entropies of the detection errors.

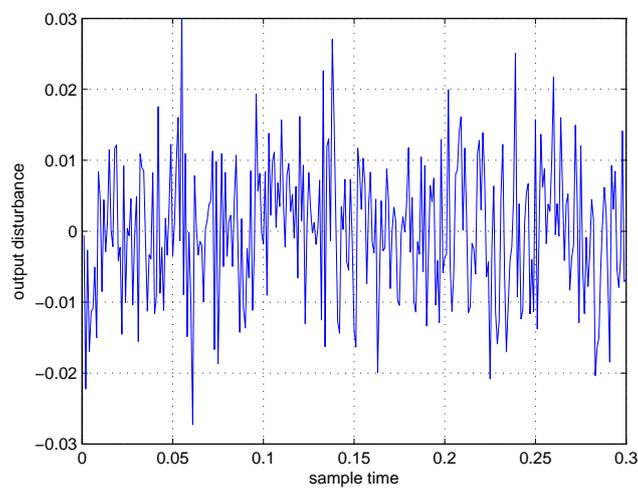
**Figure 3.** The response of the fault  $\delta$ .



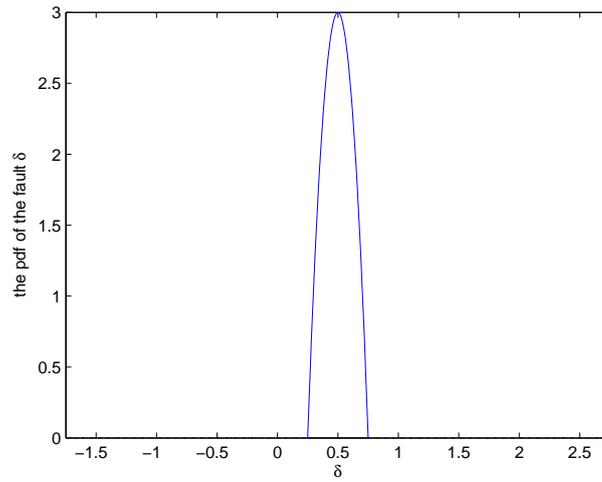
**Figure 4.** The response of the input disturbance  $\omega_k$ .



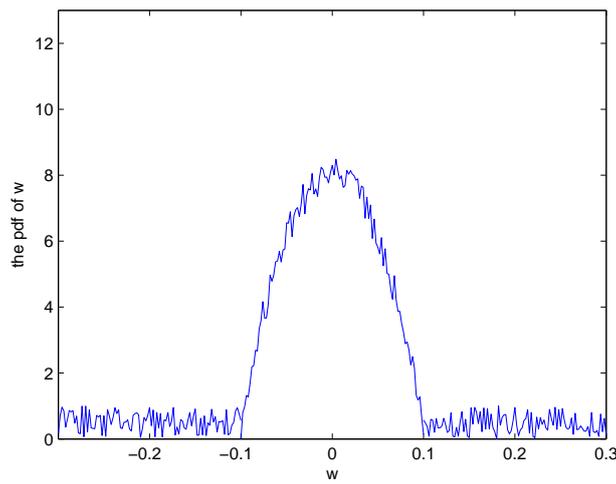
**Figure 5.** The response of the output disturbance  $v_k$ .



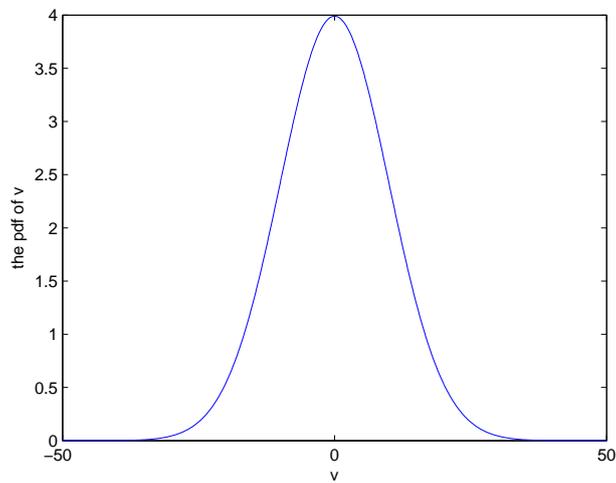
**Figure 6.** The PDF of the fault signal  $\delta$ .



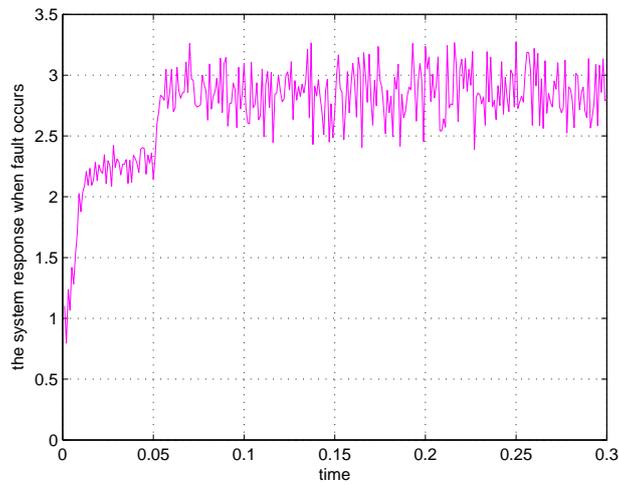
**Figure 7.** The PDF of the input disturbance  $\omega_k$ .



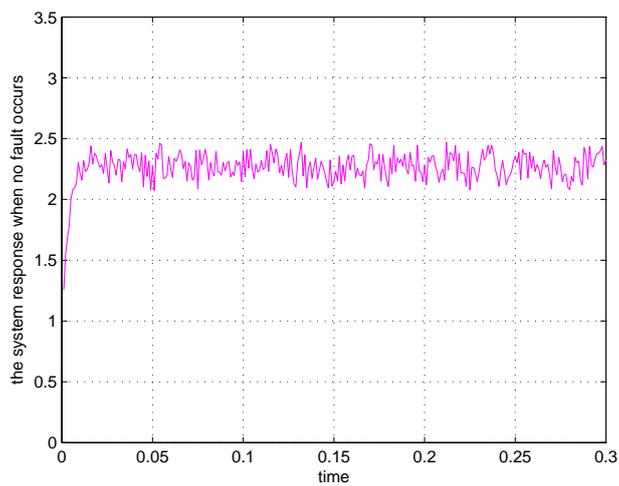
**Figure 8.** The PDF of the output disturbance  $v_k$ .



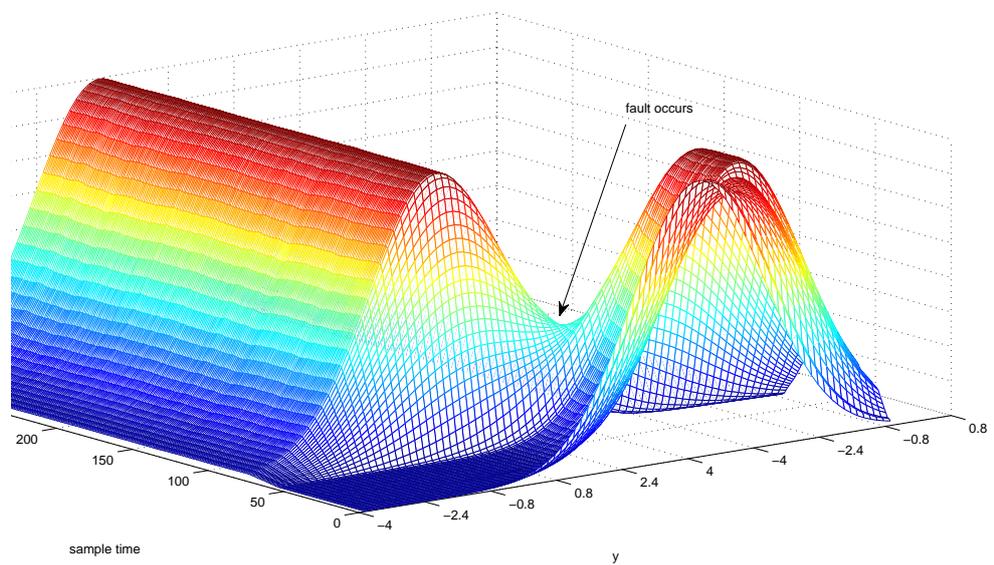
**Figure 9.** The residual value when fault occurs.

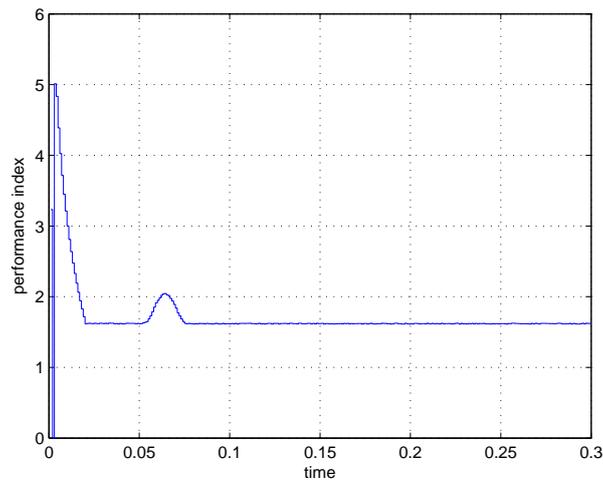


**Figure 10.** The residual value when no fault occurs.



**Figure 11.** 3-D mesh of the system output  $y$ .



**Figure 12.** The optimization performance index  $J$ .

## 5. Conclusions

Since the classical FD approach using the Kalman-filter theory is insufficient to apply to the stochastic systems with non-Gaussian variables, a new FD framework is proposed in this paper for dynamic multivariate nonlinear stochastic systems. The entropy optimization principle is established for the concerned nonlinear detection error system, as represented by a non-Gaussian stochastic system. The main design principle is to maximize the entropies of the residual errors when the faults occur and to minimize the entropies of the residual errors resulting from other stochastic noises. For this purpose, new relationships are provided to calculate the JPDFs of the detection error in terms of the JPDFs of both the disturbances and the faults. As such, recursive approaches can be constructed to formulate the entropies of the detection errors. Combining the formulations with the novel performance indexes based on the entropy optimization principles, the recursive algorithms are provided to calculate the gain of the optimal FD filters. The advantages derived from the proposed fault detection approach are summarized as following

- The detected fault and system noises do not have to be Gaussian.
- The entropy optimization principle for FD problems is in parallel to the main results to [2], where linear Gaussian systems were studied and the minimax technology was applied to the variance of errors. It has therefore generalized variance optimization for Gaussian signal.
- The fault detection algorithm only uses the JPDF of the residual signal to calculate the filter gain matrix. There is no need to measure the system output pdf. This constitutes a major advantage compared with [37–39], which require the output pdf to be measurable.
- This fault detection approach is applicable to multivariate and uncertain systems. It is a generalization of the method in [8] where only single-input-signal-output system is concerned.

However, several issues still need to be studied in the future, *i.e.*, how to determine the threshold and how to perform fault diagnosis. Moreover, as a branch of stochastic distribution control (SDC), fault

detection using entropy optimization principle still remains in the theoretic research stage [8,37,40]. Its real world application requires more research efforts.

### Acknowledgment

This work is jointly supported by the National Science Foundation of China under grant 60925012,61104123,61104073,61105115 and China Postdoctoral Science Foundation funded project (2012M520141). These are gratefully acknowledged.

### References

1. Basseville, M. On-board component fault detection and isolation using the statistic local approach. *Automatic* **1998**, *34*, 1391–1415.
2. Chen, R.H.; Mingori, D.L.; Speyer, J.L. Optimal stochastic fault detection filter. *Automatica* **2003**, *39*, 377–390.
3. Chen, R.H.; Speyer, J.L. A generalized least-squares fault detection filter. *Int. J. Adapt. Contr. Signal Process.* **2000**, *14*, 747–757.
4. De Persis, C.; Isidori, A. A geometric approach to nonlinear fault detection and isolation. *IEEE Trans. Automat. Contr.* **2001**, *46*, 853–856.
5. Frank, P.M. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy a survey and some new results. *Automatica* **1990**, *26*, 459–474.
6. Guo, L.; Wang, H. Minimum entropy filtering for multivariate stochastic systems with non-Gaussian noises. *IEEE Trans. Automat. Contr.* **2006**, *51*, 695–700.
7. Shields, D.N.; Ashton, S.A.; Daley, S. Robust fault detection observers for nonlinear polynomial systems. *Int. J. Syst. Sci.* **2001**, *32*, 723–737.
8. Guo, L.; Wang, H.; Chai, T. Fault detection for non-linear non-Gaussian stochastic systems using entropy optimization principle. *Trans. Inst. Meas. Contr.* **2006**, *28*, 145–161.
9. Mao, Z.H.; Jiang, B.; Shi, P. H-infinity fault detection filter design for networked control systems modelled by discrete Markovian jump systems. *IET Proc. Contr. Theor. Appl.* **2007**, *1*, 1336–1343.
10. Jiang, B. Staroswiecki, M.; Cocquempot, V. Fault accommodation for a class of nonlinear systems. *IEEE Trans. Automat. Contr.* **2006**, *51*, 1578–1583.
11. Basseville, M.; Nikiforv, I. Fault isolation and diagnosis: Nuisance rejection and multiple hypothesis testing. *Annu. Rev. Contr.* **2002**, *26*, 189–202.
12. Bar-Shalom, Y.; Li, X.R. *Estimation and Tracking: Principles, Techniques, and Software*; Artech House: Norwood, MA, USA, 1996.
13. Zhang, J.; Cai, L.; Wang, H. Minimum entropy filtering for networked control systems via information theoretic learning approach. In Proceedings of the 2010 International Conference on Modelling, Identification and Control, Okayama, Japan, 17–19 July 2010; pp. 774–778.
14. Orchardand, M.; Vachtsevanos, G. A particle filtering approach for on-line fault diagnosis and failure prognosis. *Trans. Inst. Meas. Contr.* **2009**, *31*, 221–246.

15. Chen, C.; Zhang, B.; Vachtsevanos, G.; Orchard, M. Machine condition prediction based on adaptive neuro-fuzzy and high-order particle filtering. *IEEE Trans. Ind. Electron.* **2011**, *58*, 4353–4364.
16. Wang, H.; Afshar, P. ILC-based fixed-structure controller design for output PDF shaping in stochastic systems using LMI techniques. *IEEE Trans. Automat. Contr.* **2009**, *54*, 760–773.
17. Zhou, J.; Zhou, D.; Wang, H.; Guo, L.; Chai, T. Distribution function tracking filter design using hybrid characteristic functions. *Automatica* **2010**, *46*, 101–109.
18. Chen, C.; Brown, D.; Sconyers, C.; Zhang, B.; Vachtsevanos, G.; Orchard, M. An integrated architecture for fault diagnosis and failure prognosis of complex engineering systems. *Expert Syst. Appl.* **2012**, *39*, 9031–9040.
19. Yin, L.; Guo, L. Fault isolation for dynamic multivariate nonlinear non-gaussian stochastic systems using generalized entropy optimization principle. *Automatica* **2009**, *45*, 2612–2619.
20. Yin, L.; Guo, L. Fault detection for NARMAX stochastic systems using entropy optimization principle. In Proceedings of the 2009 Chinese Control and Decision Conference, Guilin, China, June 2009; pp. 870–875.
21. Guo, L.; Chen, W. Disturbance attenuation and rejection for systems with nonlinearity via DOBC approach. *Int. J. Robust Nonlinear Contr.* **2005**, *15*, 109–125.
22. Yang, F.; Wang, Z.; Hung, S. Robust kalman filtering for discrete time-varying uncertain systems with multiplicative noise. *IEEE Trans. Automat. Contr.* **2002**, *47*, 1179–1183.
23. Guo, L.; Wang, H.; Wang, A.P. Optimal probability density function control for NARMAX stochastic systems. *Automatica* **2008**, *44*, 1904–1911.
24. Yin, L.; Guo, L. Joint stochastic distribution tracking control for multivariate descriptor systems with non-gaussian variables. *Int. J. Syst. Sci.* **2012**, *43*, 192–200.
25. Guo, L.; Yin, L.; Wang, H. Robust PDF control with guaranteed stability for nonlinear stochastic systems under modeling errors. *IET Contr. Theor. Appl.* **2009**, *3*, 575–582.
26. Yin, L.; Guo, L.; Wang, H. Robust minimum entropy tracking control with guaranteed stability for nonlinear stochastic systems under modeling errors. In Proceedings of the 10th International Conference on Control, Automation, Robotics and Vision, Hanoi, Vietnam, 17–20 December 2008; pp. 1469–1474.
27. Papoulis, A. *Probability, Random Variables and Stochastic Processes*, 3rd ed.; McGraw-Hill: New York, NY, USA, 1991.
28. Patton, R.J.; Frank, P.; Clark, R. *Fault Diagnosis in Dynamic Systems: Theory and Application*; Prentice Hall: Englewood Cliff, NJ, USA, 1989.
29. Feng, X.B.; Loparo, K.A. Active probing for information in control system with quantized state measurements: A minimum entropy approach. *IEEE Trans. Automat. Contr.* **1997**, *42*, 216–238.
30. Wang, H. Minimum entropy control of non-Gaussian dynamic stochastic systems. *IEEE Trans. Automat. Contr.* **2002**, *47*, 398–403.
31. Saridis, G.N. Entropy formulation of optimal and adaptive control. *IEEE Trans. Automat. Contr.* **1988**, *33*, 713–720.
32. Wang, H. *Bounded Dynamic Stochastic Systems: Modelling and Control*; Springer-Verlag: London, UK, 2000.

33. Zhang, J.; Du, L.; Ren, M.; Hou, G. Minimum error entropy filter for fault detection of networked control systems. *Entropy* **2012**, *14*, 505–516.
34. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Chapman and Hall: London, UK, 1986.
35. Wang, H.; Lin, W. Applying observer based FDI techniques to detect faults in dynamic and bounded stochastic distributions. *Int. J. Contr.* **2000**, *73*, 1424–1436.
36. Yue, H.; Wang, H. Minimum entropy control of closed-loop tracking errors for dynamic stochastic systems. *IEEE Trans. Automat. Contr.* **2003**, *48*, 118–121.
37. Guo, L.; Zhang, Y.M.; Wang, H.; Fang, J.C. Observer-based optimal fault detection and diagnosis using conditional probability distributions. *IEEE Trans. Signal Process.* **2006**, *54*, 3712–3719.
38. Zhang, Y.M.; Guo, L.; Wang, H. Filter-based fault detection and diagnosis using output PDFs for stochastic systems with time delays. *Int. J. Adapt. Contr. Signal Process.* **2006**, *20*, 175–194.
39. Li, T.; Guo, L.; Wu, L.Y. Observer-based optimal fault detection using PDFs for time-delay stochastic systems. *Nonlinear Anal. R. World Appl.* **2008**, *9*, 2337–2349.
40. Guo, L.; Yin, L.; Wang, H.; Chai, T.Y. Entropy optimization filtering for fault isolation of nonlinear non-gaussian stochastic systems. *IEEE Trans. Automat. Contr.* **2009**, *54*, 804–810.

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).