processes

MDPI

Article

A Real-Time Configuration Approach for an Observer-Based Residual Generator of Fault Detection Systems

Hao Zhao [1], Hao Luo [1,*] and Tianyu Liu [2]

[1] Department of Control Science and Engineering, Harbin Institute of Technology, Harbin 150001, China; zhaohao316@hit.edu.cn
[2] Institute of Automatic Control and Complex Systems, University of Duisburg-Essen, 47057 Duisburg, Germany; tianyu.liu@uni-due.de
* Correspondence: hao.luo@hit.edu.cn

Abstract: This paper is concerned with the real-time configuration of fault detection systems by exploiting an gradient optimization scheme. It is known that industrial processes may often encounter some uncertainties or changes of operating points and environment, which would lead to an unsatisfactory fault detection result. To handle this problem, a real-time (or online) configuration strategy is introduced, which plays an important role in ensuring the efficiency of the fault detection method without a high industrial cost. In this paper, a gradient-based iterative optimization scheme is taken into account for the real-time configuration implementation. By utilizing the gradient-based iterative algorithm to minimize the K-gap between the residual generator and the current system, the parameters of the residual generator can be configured from the online input/output data. Based on this, real-time configuration of the residual generator parameters is achieved and, correspondingly, the fault detection performance is guaranteed. Then, a three-tank system, which is relatively common and important in chemical industrial systems, is studied and explored to verify the effectiveness and superiority of the gradient optimization configuration strategy proposed in this work.

Keywords: real-time configuration; observer-based residual generator; fault detection; gradient optimization

check for updates

Citation: Zhao, H.; Luo, H.; Liu, T. A Real-Time Configuration Approach for an Observer-Based Residual Generator of Fault Detection Systems. *Processes* **2022**, *10*, 276. https://doi.org/10.3390/pr10020276

Academic Editor: Jie Zhang

Received: 20 Decembyer 2021
Accepted: 25 January 2022
Published: 30 January 2022

1. Introduction

With the rapid development of industrial technology, increasing attention has been paid to certain safety and reliability problems, prompting further research on fault detection. Among the research of fault detection problems, model-based methods have been intensively studied, and considerable results have been reported during the past few decades [1–9]. To mention a few, by transforming the residual generator design problem into an optimization problem, an optimal periodic fault detection approach is obtained for linear discrete-time periodic systems in the light of robustness and sensitivity [10]. The definitions of finite horizon H_∞/H_∞ and H_-/H_∞ fault detection performance are first established for linear discrete time-varying systems, based on which the fault detection issue is dealt with by designing some observer gains [11]. The sensor stuck faults are considered for a class of stochastic systems, and a fault detection method is proposed in the stochastic framework to guarantee the effectiveness for arbitrary small sensor stuck faults [12]. Considering the linear system with elliptical uncertainties, a general parametrization is employed for less conservativeness and then a fault detection filter is designed to maximize the sensitivity of fault detection with a certain disturbance attenuation demand [13]. To make a balance between the sensitivity and robustness, a mixed H_-/H_∞ performance index is taken into account, and sufficient criteria are presented to achieve the fault detection observer design for a class of piecewise linear systems with weighted H_-/H_∞ performance [14]. On the basis of linear system study, the model-based fault detection schemes have been extended to the issues of nonlinear fields. For example, the observer-based fault detection design is achieved for general nonlinear systems by investigating the analysis and integrated design scheme [15]. By approximating a nonlinear system as a T-S fuzzy model, the observer-based fault detection for nonlinear systems with disturbances is investigated based on the L_2 stability theory [16]. Via the logic-dynamic method, the fault detection issue for dynamic systems with non-differentiable nonlinearities is solved with the aid of the linear technique [17]. By considering the stochastic property of noises and process disturbance, a distributed fault detection and isolation scheme is proposed in a Plug-and-Plug scenario [18]. A systematic study is carried out for the fault detection of nonlinear systems by designing linear residual generators, which is further utilized in some practical applications [19]. The model-based method plays an important role in the field of fault detection, while it will unavoidably incur a high cost in acquiring the accurate model information. Therefore, the data-driven fault detection as an alternative method has drawn increasing attention in both academia and industry.

Recently, data-driven fault detection issues have been investigated thoroughly, due to their advantages in saving the costly modeling process and making great use of process data information compared with the model-based method. Over the last few years, many studies have been done for the data-driven fault detection issues of industrial process. For instance, the observer-based fault detection is constructed by exploiting the data-driven image and kernel representations [20]. By using the residual generators derived from the process data, a data-driven fault detection approach is devised for wind turbines with measurement noises, unknown disturbances, and nonlinearities [21]. Data-driven fault detection and isolation filters are constructed for sensor and actuator faults by taking advantages of available system data, and meanwhile an estimation approach is established and extended to an offline tuning strategy to compensate the estimation errors under the uncertainties and Markov parameters [22]. In the data-driven framework, a fault detection scheme is proposed to detect small sensor faults and a fault isolation algorithm is developed to distinguish different faults [23]. By identifying a data-driven SKR with the projecting technique, a robust residual generation is derived and, accordingly, the robust data-driven fault detection strategy is obtained for rolling mill processes with unknown eccentricity [24]. The quantitative diagnosability analysis is addressed for dynamic systems by virtue of the data-driven evaluation [25]. By employing a radically data-driven strategy, the fault detection and diagnosis are developed for wind turbines to enhance the reliability [26]. A q-step residual design approach is constructed for the data-driven fault detection of linear systems to ensure the stability and performance demand [27]. The distributed data-driven optimal fault detection is studied in large-scale systems by utilizing the average consensus algorithm [28]. By proposing a prediction model on the output of nonlinear dynamic systems, a detection method is devised according to the comparison between the measurement output and the prediction to determine a residual, and further an isolation scheme is constructed to clarify the fault location for the underlying system [29]. Considering the fact that incipient faults are not easy to discover in electrical drives because of their inapparent symptoms, a data-driven fault detection and diagnosis method is presented by applying the principal component analysis approach which improves the accuracy of fault detection for electrical drives without available system parameters or models [30].

Note that the practical industrial process inevitably suffers the changes of process environment and operation conditions. In this case, the predesigned residual generators may not provide satisfactory fault detection performance. To guarantee the ability and efficiency of the data-driven fault detection without sacrificing the industrial cost, the online configuration or updating becomes an important technology in solving such issues. Up to now, some methods use online configuration and updating of data-driven fault detection including adaptive algorithms, iterative optimization, etc. The adaptive residual generator combined with the data-driven scheme is designed and implemented to recursively estimate the corresponding parameters and improve the robustness against the undesired changes for discrete linear systems [31]. By virtue of a data-driven subspace-based predictor, an adaptive updating strategy is proposed for the fault detection filter of solar power generation systems with uncertainties [32]. Via adopting an autoregressive exogenous model to represent the dynamic process, an adaptive data-driven method is developed for fault detection of dynamic process with process drift [33]. The adaptive algorithm provides an effective way for the online configuration of data-driven residual generators, while each new measurement is utilized for the parameter estimation, and the parameter updating in the adaptive configuration happens at every sampling instant. Moreover, the configured system matrix of the adaptive method may be sensitive to small changes in the parameters. Compared with the adaptive configuration, the residual generator parameters based on the iterative optimization method remain unchanged between two iterations, which would greatly reduce the iteration number and meanwhile guarantee the fault detection performance [34,35]. However, the existing results about the online optimization configuration for the observer-based residual generators are quite limited, which motivates the current study.

Based on the observations above, this paper is aimed to investigate the real-time configuration for fault detection systems via the gradient optimization method. Considering the unavoidable changes of the industrial process and operating environment, it is necessary to establish a real-time configuration method to properly configure the residual generator parameters to guarantee fault detection ability. To achieve the real-time configuration, a gradient-based iterative optimization strategy is proposed by minimizing the K-gap metric between the residual generator and the current system. In this way, the residual generator can be updated from the available input/output (I/O) data without the identification of system matrices. A novel optimization algorithm for the real-time configuration of residual generators is developed, by which the validity of the fault detection process is guaranteed. Furthermore, a three-tank system plant is taken into account to illustrate the usefulness and advantages of the proposed approach, which can be seen as the prototype for many industrial processes, such as chemical process industries.

The structure of this paper is organized as below. In Section 2, the system descriptions and necessary preliminaries are provided. Section 3 presents the gradient optimization to achieve the online configuration for fault detection systems. In Section 4, a simulation example is used to demonstrate the effectiveness of the obtained method. Finally, the conclusions of this work are drawn in Section 5.

Notation 1. *Throughout this paper, the notations are generally standard. $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$, respectively, denote the n-dimensional Euclidean space and the set of all $m \times n$ real matrices. $\mathcal{H}_\infty$ is the set of all stable transfer functions and $\mathcal{H}_2$ is the subspace of all signals with bounded energy equal to 0 for any $t < 0$. $\mathbb{RH}_\infty$ defines the set of all real-rational transfer functions of stable systems. $\|\cdot\|_2$ and $\|\cdot\|_\infty$ stand for the $\mathcal{L}_2$-norm and the $\mathcal{H}_\infty$-norm, respectively. $vec(A)$ indicates the vectorization of matrix A. $\bar{\sigma}(A)$ and $\overline{eig}(A)$ represent the maximum singular value and the maximum eigenvalue of matrix A, respectively. A^+ denotes the pseudoinverse of matrix A. $diag\{\cdots\}$ defines a diagonal matrix.*

2. Preliminaries

2.1. System Descriptions

Consider the discrete-time linear time-invariant (LTI) system represented by

$$x(k+1) = Ax(k) + Bu(k), \quad (1)$$
$$y(k) = Cx(k) + Du(k), \quad (2)$$

where $x(k) \in \mathbb{R}^n$ is the state vector, $u(k) \in \mathbb{R}^l$ is the input signal, and $y(k) \in \mathbb{R}^m$ is the system output. A, B, C, D are system matrices with proper dimensions.

2.2. Stable Kernel Representation

For the nominal system (1) and (2), its transfer function representation is given by

$$y(z) = G(z)u(z). \tag{3}$$

Then, the stable kernel representation of $G(z)$ is described as below.

Consider a proper real-rational transfer function matrix $G(z)$ with the following left and right coprime factorizations:

$$G(z) = \hat{M}^{-1}(z)\hat{N}(z) = N(z)M^{-1}(z), \tag{4}$$

where $\hat{M}(z) \in \mathbb{RH}_\infty^{m\times m}, \hat{N}(z) \in \mathbb{RH}_\infty^{m\times l}, M(z) \in \mathbb{RH}_\infty^{l\times l}, N(z) \in \mathbb{RH}_\infty^{m\times l}$, and $\left(\hat{M}(z), \hat{N}(z)\right)$, $(M(z), N(z))$ are, respectively, left and right coprime pairs over $\mathbb{RH}_\infty$. It means that there exist $\hat{X}(z) \in \mathbb{RH}_\infty^{m\times m}, \hat{Y}(z) \in \mathbb{RH}_\infty^{l\times m}, X(z) \in \mathbb{RH}_\infty^{l\times l}, Y(z) \in \mathbb{RH}_\infty^{l\times m}$ such that the following equations hold:

$$\begin{aligned} \begin{bmatrix} \hat{M}(z) & \hat{N}(z) \end{bmatrix} \begin{bmatrix} \hat{X}(z) \\ \hat{Y}(z) \end{bmatrix} &= I_{m\times m}, \\ \begin{bmatrix} X(z) & Y(z) \end{bmatrix} \begin{bmatrix} M(z) \\ N(z) \end{bmatrix} &= I_{l\times l}. \end{aligned}$$

Further, if $\left(\hat{M}(z), \hat{N}(z)\right)$ and $(M(z), N(z))$ are satisfied with

$$\begin{aligned} \begin{bmatrix} \hat{M}(z) & \hat{N}(z) \end{bmatrix} \begin{bmatrix} \hat{M}(z) & \hat{N}(z) \end{bmatrix}^T &= I_{m\times m}, \\ \begin{bmatrix} M(z) \\ N(z) \end{bmatrix}^T \begin{bmatrix} M(z) \\ N(z) \end{bmatrix} &= I_{l\times l}, \end{aligned}$$

then they are called the normalized left and right coprime pairs, respectively.

Definition 1 ([35])**.** *Given a discrete-time LTI system $G(z)$ in Equation (3), a stable linear system $\mathcal{K}$ is called the stable kernel representation (SKR) of $G(z)$, if for any $u(z)$ and its response $y(z)$, the following equation holds:*

$$\mathcal{K}\begin{bmatrix} u(z) \\ y(z) \end{bmatrix} = 0.$$

Suppose that $r(z)$ is the residual signal of the underlying system. According to the description of the left and right coprime factorizations, it is clear that $\hat{M}(z)$ and $\hat{N}(z)$ correspond to the transfer matrices from the residual signal to the output signal and the input signal, respectively. Then, the following equation holds in the fault- and noise-free case:

$$r(z) = \begin{bmatrix} -\hat{N}(z) & \hat{M}(z) \end{bmatrix} \begin{bmatrix} u(z) \\ y(z) \end{bmatrix}. \tag{5}$$

Accordingly, a SKR of system $G(z)$ can be formed by the transfer matrices as below,

$$\mathcal{K} = \begin{bmatrix} -\hat{N}(z) & \hat{M}(z) \end{bmatrix}.$$

2.3. K-Gap Metric

To achieve the optimization objective, it is necessary to introduce a means to measure the distance between two kernel subspaces. As the K-gap metric has become a powerful tool in dealing with the measurement problems, this paper will adopt the K-gap metric technique for the optimization process. Before mentioning the K-gap metric, the gap metric concept is first restated for clarification. For this purpose, the graph definition is introduced and represented by

$$\mathcal{G} = \left\{ \zeta = \begin{bmatrix} u \\ y \end{bmatrix} = \begin{bmatrix} M \\ N \end{bmatrix} v, v \in \mathcal{H}_2 \right\}.$$

Note that the graph $\mathcal{G}$ is a subspace in $\mathcal{H}_2$ constructed by all the pairs (u, y), and it is closed [36,37]. Denote $G_1 = N_1 M_1^{-1}, G_2 = N_2 M_2^{-1}$ as the normalized right coprime factorizations of G_1, G_2, respectively. Let $\mathcal{G}_1, \mathcal{G}_2$ be the corresponding graphs. The direct gap from $\mathcal{G}_1$ to $\mathcal{G}_2$ is defined as $\vec{\delta}(\mathcal{G}_1, \mathcal{G}_2)$, which is formulated by

$$\vec{\delta}(\mathcal{G}_1, \mathcal{G}_2) = \sup_{\zeta_1 \in \mathcal{G}_1} \inf_{\zeta_2 \in \mathcal{G}_2} \frac{\|\zeta_1 - \zeta_2\|_2}{\|\zeta_1\|_2}. \tag{6}$$

According to the work in [36], the calculation on the direct gap (6) can be solved by

$$\vec{\delta}(\mathcal{G}_1, \mathcal{G}_2) = \inf_{Q \in \mathcal{H}_\infty} \left\| \begin{bmatrix} M_1 \\ N_1 \end{bmatrix} - \begin{bmatrix} M_2 \\ N_2 \end{bmatrix} Q \right\|_\infty.$$

Based on above, the definition of gap metric between $\mathcal{G}_1$ and $\mathcal{G}_2$ is derived as

$$\delta(\mathcal{G}_1, \mathcal{G}_2) = \max\left\{ \vec{\delta}(\mathcal{G}_1, \mathcal{G}_2), \vec{\delta}(\mathcal{G}_2, \mathcal{G}_1) \right\}.$$

Considering that the gap metric is based on the image subspace, the K-gap metric defined on the kernel subspace is further proposed. The corresponding graph is defined as

$$\mathcal{K} = \left\{ \begin{bmatrix} u \\ y \end{bmatrix} : \begin{bmatrix} -\hat{N} & \hat{M} \end{bmatrix} \begin{bmatrix} u \\ y \end{bmatrix} = 0, \begin{bmatrix} u \\ y \end{bmatrix} \in \mathcal{H}_2 \right\},$$

which indicates the kernel subspace and is a closed subspace in $\mathcal{H}_2$. Similarly, the directed K-gap from graph $\mathcal{K}_1$ to graph $\mathcal{K}_2$ is expressed as follows.

Definition 2 ([37])**.** *Suppose that $(\hat{M}_1(z), \hat{N}_1(z))$, $(\hat{M}_2(z), \hat{N}_2(z))$ are the left coprime factorizations of $G_1(z), G_2(z)$, respectively, and*

$$\mathcal{K}_i = \left\{ \varsigma_i = \begin{bmatrix} u_i \\ y_i \end{bmatrix} : \begin{bmatrix} -\hat{N}_i & \hat{M}_i \end{bmatrix} \begin{bmatrix} u_i \\ y_i \end{bmatrix} = 0, \begin{bmatrix} u_i \\ y_i \end{bmatrix} \in \mathcal{H}_2 \right\}, i = 1, 2.$$

The directed K-gap from $\mathcal{K}_1$ to $\mathcal{K}_2$ is defined by

$$\vec{\delta}_k(\mathcal{K}_1, \mathcal{K}_2) = \sup_{\varsigma_1 \in \mathcal{K}_1} \inf_{\varsigma_2 \in \mathcal{K}_2} \frac{\|\varsigma_1 - \varsigma_2\|_2}{\|\varsigma_1\|_2}.$$

Subsequently, the K-gap metric between $\mathcal{K}_1$ and $\mathcal{K}_2$ is given by

$$\delta_k(\mathcal{K}_1, \mathcal{K}_2) = \max\left\{ \vec{\delta}_k(\mathcal{K}_1, \mathcal{K}_2), \vec{\delta}_k(\mathcal{K}_2, \mathcal{K}_1) \right\}.$$

Moreover, a computation strategy of the K-gap metric is also established in [37], which is recalled in the following lemma.

Lemma 1 ([37])**.** *Consider $\mathcal{K}_i, i = 1, 2$ defined in Definition 2 with normalized left coprime factorizations $(\hat{M}_i(z), \hat{N}_i(z)), i = 1, 2$. The direct K-gap can be computed by*

$$\vec{\delta}_k(\mathcal{K}_1, \mathcal{K}_2) = \inf_{Q \in \mathcal{H}_\infty} \left\| \begin{bmatrix} -\hat{N}_1 & \hat{M}_1 \end{bmatrix} - Q \begin{bmatrix} -\hat{N}_2 & \hat{M}_2 \end{bmatrix} \right\|_\infty.$$

It can be yielded from Lemma 1 that

$$0 \leq \vec{\delta}_k(\mathcal{K}_1, \mathcal{K}_2) \leq 1,$$

and when $\delta_k(\mathcal{K}_1, \mathcal{K}_2) < 1$,

$$\vec{\delta}_k(\mathcal{K}_1, \mathcal{K}_2) = \vec{\delta}_k(\mathcal{K}_2, \mathcal{K}_1) = \delta_k(\mathcal{K}_1, \mathcal{K}_2).$$

2.4. Data-Driven Framework

As the input and output data are crucial to the fault detection realization, the data model is introduced here for latter development. Taking a data vector $\lambda(k) \in \mathcal{R}^{\kappa}$ into account, the related notations are defined as below.

$$\begin{aligned}
\lambda_s(k) &= \begin{bmatrix} \lambda(k-s) \\ \lambda(k-s+1) \\ \vdots \\ \lambda(k) \end{bmatrix} \in \mathcal{R}^{(s+1)\kappa}, \\
\Lambda_k &= \begin{bmatrix} \lambda(k) & \cdots & \lambda(k+N-1) \end{bmatrix} \in \mathcal{R}^{\kappa \times N}, \\
\Lambda_{k,s} &= \begin{bmatrix} \lambda_s(k) & \cdots & \lambda_s(k+N-1) \end{bmatrix} = \begin{bmatrix} \Lambda_{k-s} \\ \vdots \\ \Lambda_k \end{bmatrix} \in \mathcal{R}^{(s+1)\kappa \times N},
\end{aligned}$$

where s, N are positive integers and $s+1$ is the length of the stacked data vector. Based on the data structure, the data-driven realization of the SKR is defined as follows.

Definition 3 ([20]). *$\mathcal{K}_{d,s}$ is called a data-driven realization of the SKR for system $G(z)$, if for all $k \geq 0$, the following equation is satisfied:*

$$\mathcal{K}_{d,s} \begin{bmatrix} u_s(k) \\ y_s(k) \end{bmatrix} = \begin{bmatrix} \mathcal{K}_{u,s} & \mathcal{K}_{y,s} \end{bmatrix} \begin{bmatrix} u_s(k) \\ y_s(k) \end{bmatrix} = 0.$$

Note that if the data-driven SKR is satisfied with $\mathcal{K}_{d,s}\mathcal{K}_{d,s}^T = I$, then it is called normalized [38]. By applying the singular value decomposition, it holds that

$$\mathcal{K}_{d,s} = U_{sys} \begin{bmatrix} \Sigma_{sys,1} & 0 \end{bmatrix} \begin{bmatrix} V_{sys,1}^T \\ V_{sys,2}^T \end{bmatrix}$$

and thus, the normalized data-driven SKR for system $G(z)$ is derived as

$$\bar{\mathcal{K}}_{d,s} = V_{sys,1}^T. \tag{7}$$

Next, the data-driven realization of the K-gap metric is presented on the basic of the normalized data-driven SKR.

Lemma 2 ([39]). *Suppose that $\bar{\mathcal{K}}_{1,d,s}, \bar{\mathcal{K}}_{2,d,s}$ are the normalized data-driven SKRs of SKRs $\mathcal{K}_1, \mathcal{K}_2$. The data-driven realization of the K-gap metric can be calculated by*

$$\delta_{k_{d,s}}(\mathcal{K}_1, \mathcal{K}_2) = \bar{\sigma}\left(\left[I - \bar{\mathcal{K}}_{2,d,s}^T \bar{\mathcal{K}}_{2,d,s} \right] \bar{\mathcal{K}}_{1,d,s}^T \bar{\mathcal{K}}_{1,d,s} \right). \tag{8}$$

3. Main Results

Considering that practical circumstance and industrial environment may change, the fault detection by the offline designed residual generator cannot satisfy the complicated industrial demand. In this section, a novel real-time configuration scheme for residual generators is proposed, which is essentially an optimization algorithm based on the K-gap metric. To be specific, an observer-based residual generator is constructed and the gradient optimization algorithm is used to update its parameters. Thus, the real-time configuration is achieved for the observer-based residual generators, the framework of which is displayed in Figure 1 for clarification.

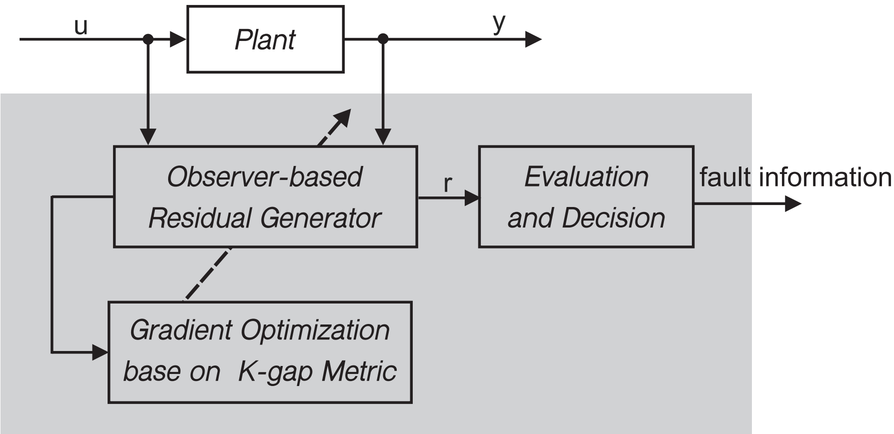

Figure 1. Framework of the online configuration based on K-gap metric.

3.1. The Observer-Based General Generator

For the discrete-time LTI system (1), a full-order state observer is constructed with the minimal state-space representation as

$$x_o(k+1) = A_o x_o(k) + B_o u(k) + L_o y(k), \tag{9}$$
$$r(k) = C_o x_o(k) + D_o u(k) + y(k), \tag{10}$$

where $x_o(k) \in \mathbb{R}^n$ indicates the state of the full-order observer and $r(k) \in \mathbb{R}^m$ stands for the residual vector. A_o, B_o, C_o, D_o, L_o are observer matrices with proper dimensions.

For observer (9) and (10), a similarity transformation is performed by $x_o = T_v x_v$, which yields

$$x_v(k+1) = A_v x_v(k) + B_v u(k) + L_v y(k), \tag{11}$$
$$r(k) = C_v x_v(k) + D_v u(k) + y(k), \tag{12}$$

where $A_v = T_v^{-1} A_o T_v, B_v = T_v^{-1} B_o, L_v = T_v^{-1} L_o, C_v = C_o T_v, D_v = D_o$. According to the authors of [40], the controllability Gramian matrix of systems (11) and (12) is equivalent to the identity matrix, i.e.,

$$A_v A_v^T + B_v B_v^T = I_n, \tag{13}$$

which gives rise to a column-orthogonal matrix $\begin{bmatrix} B_v^T \\ A_v^T \end{bmatrix}$. Consequently, there exists an invertible matrix Ψ such that

$$\begin{bmatrix} B_v^T \\ A_v^T \end{bmatrix} = \Psi \begin{bmatrix} 0 \\ I_n \end{bmatrix}.$$

Referring to the literature [41], the parameterization based on the input normal form can be described by

$$\begin{bmatrix} B_v^T(\theta_{AB}) \\ A_v^T(\theta_{AB}) \end{bmatrix} = \Psi_1(\theta_{AB}(1)) \cdots \Psi_{nl}(\theta_{AB}(nl)) \begin{bmatrix} 0 \\ I_n \end{bmatrix},$$

in which $\theta_{AB} \in \mathbb{R}^{nl}$ and its entries take values in the range $(-1, 1)$. By introducing the following form for each parameter $\theta_{AB}(i), i = 1, \cdots, nl$

$$U(\theta_{AB}(i)) = \begin{bmatrix} -\theta_{AB}(i) & \sqrt{1-\theta_{AB}^2(i)} \\ \sqrt{1-\theta_{AB}^2(i)} & \theta_{AB}(i) \end{bmatrix},$$

the matrices $\Psi_i(\theta_{AB}(i)), i = 1, \cdots, nl$ are represented as

$$\Psi_1(\theta_{AB}(1)) = \begin{bmatrix} I_{n-1} & 0 & 0 \\ 0 & U(\theta_{AB}(1)) & 0 \\ 0 & 0 & I_{l-1} \end{bmatrix},$$

$$\vdots$$

$$\Psi_{nl}(\theta_{AB}(nl)) = \begin{bmatrix} I_{l-1} & 0 & 0 \\ 0 & U(\theta_{AB}(nl)) & 0 \\ 0 & 0 & I_{n-1} \end{bmatrix}.$$

As the parameterization of the input normal form is on the basis of the asymptotic stability, no additional restriction on the parameter space is needed, which is a significant advantage of this transformation.

Then, the procedure to obtain the SKR corresponding to the observer (11) is given as below. It is obvious from Formula (12) that

$$r(k-s) = C_v x_v(k-s) + D_v u(k-s) + y(k-s).$$

Subsequently,

$$\begin{aligned}
r(k-s+1) =& C_v x_v(k-s+1) + D_v u(k-s+1) + y(k-s+1) \\
=& C_v[A_v x_v(k-s) + B_v u(k-s) + L_v y(k-s)] + D_v u(k-s+1) + y(k-s+1) \\
=& C_v A_v x_v(k-s) + C_v B_v u(k-s) + D_v u(k-s+1) + C_v L_v y(k-s) + y(k-s+1), \\
r(k-s+2) =& C_v x_v(k-s+2) + D_v u(k-s+2) + y(k-s+2) \\
=& C_v[A_v x_v(k-s+1) + B_v u(k-s+1) + L_v y(k-s+1)] \\
&+ D_v u(k-s+2) + y(k-s+2) \\
=& C_v A_v[A_v x_v(k-s) + B_v u(k-s) + L_v y(k-s)] + C_v B_v u(k-s+1) \\
&+ C_v L_v y(k-s+1) + D_v u(k-s+2) + y(k-s+2) \\
=& C_v A_v^2 x_v(k-s) + C_v A_v B_v u(k-s) + C_v B_v u(k-s+1) + D_v u(k-s+2) \\
&+ C_v A_v L_v y(k-s) + C_v L_v y(k-s+1) + y(k-s+2), \\
r(k-s+3) =& C_v x_v(k-s+3) + D_v u(k-s+3) + y(k-s+3) \\
=& C_v[A_v x_v(k-s+2) + B_v u(k-s+2) + L_v y(k-s+2)] \\
&+ D_v u(k-s+3) + y(k-s+3) \\
=& C_v A_v[A_v x_v(k-s+1) + B_v u(k-s+1) + L_v y(k-s+1)] + C_v B_v u(k-s+2) \\
&+ C_v L_v y(k-s+2) + D_v u(k-s+3) + y(k-s+3) \\
=& C_v A_v^2[A_v x_v(k-s) + B_v u(k-s) + L_v y(k-s)] + C_v A_v B_v u(k-s+1) \\
&+ C_v A_v L_v y(k-s+1) + C_v B_v u(k-s+2) + C_v L_v y(k-s+2) \\
&+ D_v u(k-s+3) + y(k-s+3) \\
=& C_v A_v^3 x_v(k-s) + C_v A_v^2 B_v u(k-s) + C_v A_v B_v u(k-s+1) + C_v B_v u(k-s+2) \\
&+ D_v u(k-s+3) + C_v A_v^2 L_v y(k-s) + C_v A_v L_v y(k-s+1) \\
&+ C_v L_v y(k-s+2) + y(k-s+3), \\
&\vdots
\end{aligned}$$

which implies

$$\begin{aligned} r(k) =& C_v A_v^s x_v(k-s) + C_v A_v^{s-1} B_v u(k-s) + \cdots + C_v B_v u(k-1) + D_v u(k) \\ &+ C_v A_v^{s-1} L_v y(k-s) + \cdots + C_v L_v y(k-1) + y(k). \end{aligned}$$

Then, it is easy to obtain that

$$r_s(k) = \Gamma_s x_v(k-s) + H_{u,s} u_s(k) + H_{y,s} y_s(k), \tag{14}$$

where

$$\Gamma_s = \begin{bmatrix} C_v \\ C_v A_v \\ \vdots \\ C_v A_v^s \end{bmatrix}, H_{u,s} = \begin{bmatrix} D_v & \cdots & 0 & 0 \\ C_v B_v & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ C_v A_v^{s-1} B_v & \cdots & C_v B_v & D_v \end{bmatrix},$$
$$H_{y,s} = \begin{bmatrix} I & \cdots & 0 & 0 \\ C_v L_v & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ C_v A_v^{s-1} L_v & \cdots & C_v L_v & I \end{bmatrix}.$$

Then, the following equation is obtained:

$$R_{k,s} = \Gamma_s X_{v,k-s} + H_{u,s} U_{k,s} + H_{y,s} Y_{k,s}.$$

As $C_v A_v^s \to 0, s \to \infty$, according to Definition 3, the SKR corresponding to the state observer can be obtained by the equation

$$\begin{aligned} \mathcal{K}_{o,s} =& \begin{bmatrix} H_{u,s} & H_{y,s} \end{bmatrix} \\ =& \begin{bmatrix} D_v & \cdots & 0 & 0 & I & \cdots & 0 & 0 \\ C_v B_v & \cdots & 0 & 0 & C_v L_v & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ C_v A_v^{s-1} B_v & \cdots & C_v B_v & D_v & C_v A_v^{s-1} L_v & \cdots & C_v L_v & I \end{bmatrix} \end{aligned} \tag{15}$$

with removing the front finite rows to eliminate the influence of past data. By applying the singular value decomposition to $\mathcal{K}_{o,s}$, it holds that

$$\mathcal{K}_{o,s} = U \begin{bmatrix} \Sigma_1 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$$

and thus, the normalized SKR is

$$\bar{\mathcal{K}}_{o,s} = V_1^T. \tag{16}$$

3.2. Online Gradient Optimization

Based on the above observer-based residual generator, this subsection will establish an gradient optimization method for the real-time configuration by taking advantage of the K-gap metric between the observer and system plant.

Define

$$\theta = \begin{bmatrix} \theta_{AB} \\ \theta_{C_v} \\ \theta_{D_v} \\ \theta_{L_v} \end{bmatrix} = \begin{bmatrix} \theta_{AB} \\ vec(C_v) \\ vec(D_v) \\ vec(L_v) \end{bmatrix},$$

and denote θ_i $(i = 1, 2, \cdots, \tau = nl + 2mn + ml)$ as the i-th term of θ.

The optimization problem is described by

$$\begin{cases} \text{minimize } J = \delta_{k_{d,s}}(\mathcal{K}_{o,s}, \mathcal{K}_{d,s}), \\ \text{subject to } \theta \in \mathcal{S}, \end{cases}$$

where $\mathcal{K}_{d,s}, \mathcal{K}_{o,s}$ are, respectively, the SKRs of the system (1) and the observer (11), and $\mathcal{S}$ is the set in which the entries of any vector belong to $(-1, 1)$. To achieve the minimization of cost function J, the Taylor expansion of cost function J at the j-th iteration is considered

$$\begin{aligned} J^{(j)}_{\theta^{(j)}} =& J^{(j)}_{\theta^{(j-1)}} + \left(\frac{\partial J^{(j)}_{\theta^{(j)}}}{\partial \theta^{(j)}} \bigg|_{\theta^{(j-1)}} \right)^T \left(\theta^{(j)} - \theta^{(j-1)}\right) \\ &+ \frac{1}{2}\left(\theta^{(j)} - \theta^{(j-1)}\right)^T \left(\frac{\partial^2 J^{(j)}_{\theta^{(j)}}}{\partial \theta^{(j)} \partial \left(\theta^{(j)}\right)^T} \bigg|_{\theta^{(j-1)}} \right) \left(\theta^{(j)} - \theta^{(j-1)}\right) + o^3\left(\theta^{(j)} - \theta^{(j-1)}\right), \end{aligned}$$

where $\frac{\partial J^{(j)}_{\theta^{(j)}}}{\partial \theta^{(j)}}$ and $\frac{\partial^2 J^{(j)}_{\theta^{(j)}}}{\partial \theta^{(j)} \partial \left(\theta^{(j)}\right)^T}$ denote the gradient and the Hessian matrix of cost function J, respectively. $o^3(\cdot)$ stands for the infinitesimal of order higher than 3. Regardless of the high-order infinitesimal, a necessary condition to minimize the function $J^{(j)}_{\theta^{(j)}}$ is to command

$$\left(\frac{\partial J^{(j)}_{\theta^{(j)}}}{\partial \theta^{(j)}} \bigg|_{\theta^{(j-1)}} \right) + \left(\frac{\partial^2 J^{(j)}_{\theta^{(j)}}}{\partial \theta^{(j)} \partial \left(\theta^{(j)}\right)^T} \bigg|_{\theta^{(j-1)}} \right) \left(\theta^{(j)} - \theta^{(j-1)}\right) = 0. \tag{17}$$

Rewriting the expression (17) gives the iteration procedure of updating the parameter θ as

$$\theta^{(j)} = \theta^{(j-1)} - \left(\frac{\partial^2 J^{(j)}_{\theta^{(j)}}}{\partial \theta^{(j)} \partial \left(\theta^{(j)}\right)^T} \bigg|_{\theta^{(j-1)}} \right)^{-1} \left(\frac{\partial J^{(j)}_{\theta^{(j)}}}{\partial \theta^{(j)}} \bigg|_{\theta^{(j-1)}} \right), \tag{18}$$

which is the Gauss–Newton iteration widely used in the literature. However, this method requires the invertibility of $\left(\frac{\partial^2 J^{(j)}_{\theta^{(j)}}}{\partial \theta^{(j)} \partial \left(\theta^{(j)}\right)^T} \bigg|_{\theta^{(j-1)}} \right)$ and may lead to a heavy computational burden. To improve its applicability in numerical computation, an iteration procedure known as the steepest-descent algorithm is introduced

$$\theta^{(j)} = \theta^{(j-1)} - \Delta^{(j)} \left(\frac{\partial J^{(j)}_{\theta^{(j)}}}{\partial \theta^{(j)}} \bigg|_{\theta^{(j-1)}} \right), \tag{19}$$

where $\Delta^{(j)} > 0$ is a diagonal matrix meaning the step length of the j-th iteration.

Clearly, the key question of the optimization is to calculate the gradient $\frac{\partial J}{\partial \theta} = \frac{\partial \delta_{k_{d,s}}(\mathcal{K}_{o,s}, \mathcal{K}_{d,s})}{\partial \theta}$.

It is obtained from Lemma 2 that

$$
\begin{aligned}
&\frac{\partial\delta_{k_{d,s}}(\mathcal{K}_{o,s},\mathcal{K}_{d,s})}{\partial\theta}\\
=&\frac{\partial\bar{\sigma}\left(\left[I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s}\right]\bar{\mathcal{K}}_{o,s}^T\bar{\mathcal{K}}_{o,s}\right)}{\partial\theta}\\
=&\frac{\partial\left\{\sqrt{\overline{eig}\left[\left(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s}\right)(\bar{\mathcal{K}}_{o,s}^T\bar{\mathcal{K}}_{o,s})(\bar{\mathcal{K}}_{o,s}^T\bar{\mathcal{K}}_{o,s})\left(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s}\right)\right]}\right\}}{\partial\theta}\\
=&\frac{\partial\left\{\sqrt{\overline{eig}\left[\left(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s}\right)(\bar{\mathcal{K}}_{o,s}^T\bar{\mathcal{K}}_{o,s})\left(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s}\right)\right]}\right\}}{\partial\theta}\\
=&\frac{1}{2\delta_{k_{d,s}}(\mathcal{K}_{o,s},\mathcal{K}_{d,s})}\frac{\partial\left\{\overline{eig}\left[\left(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s}\right)(\bar{\mathcal{K}}_{o,s}^T\bar{\mathcal{K}}_{o,s})\left(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s}\right)\right]\right\}}{\partial\theta}\\
=&\frac{1}{2\delta_{k_{d,s}}(\mathcal{K}_{o,s},\mathcal{K}_{d,s})}\begin{bmatrix}\frac{\partial\{\overline{eig}[(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s})(\bar{\mathcal{K}}_{o,s}^T\bar{\mathcal{K}}_{o,s})(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s})]\}}{\partial\theta_1}\\ \frac{\partial\{\overline{eig}[(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s})(\bar{\mathcal{K}}_{o,s}^T\bar{\mathcal{K}}_{o,s})(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s})]\}}{\partial\theta_2}\\ \vdots\\ \frac{\partial\{\overline{eig}[(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s})(\bar{\mathcal{K}}_{o,s}^T\bar{\mathcal{K}}_{o,s})(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s})]\}}{\partial\theta_i}\\ \vdots\\ \frac{\partial\{\overline{eig}[(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s})(\bar{\mathcal{K}}_{o,s}^T\bar{\mathcal{K}}_{o,s})(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s})]\}}{\partial\theta_\tau}\end{bmatrix}.
\end{aligned}
$$

Denote ξ as the eigenvector corresponding to the maximum eigenvalue of matrix $\left(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s}\right)(\bar{\mathcal{K}}_{o,s}^T\bar{\mathcal{K}}_{o,s})\left(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s}\right)$ and focus on the partial derivative on the i-th variable, i.e.,

$$
\begin{aligned}
&\frac{\partial\left\{\overline{eig}\left[\left(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s}\right)(\bar{\mathcal{K}}_{o,s}^T\bar{\mathcal{K}}_{o,s})\left(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s}\right)\right]\right\}}{\partial\theta_i}\\
=&\frac{\xi^T\partial\left\{\left(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s}\right)(\bar{\mathcal{K}}_{o,s}^T\bar{\mathcal{K}}_{o,s})\left(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s}\right)\right\}\xi}{\partial\theta_i}\\
=&\xi^T\frac{\partial\left\{\left(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s}\right)(\bar{\mathcal{K}}_{o,s}^T\bar{\mathcal{K}}_{o,s})\left(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s}\right)\right\}}{\partial\theta_i}\xi\\
=&\xi^T\left(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s}\right)\frac{\partial(\bar{\mathcal{K}}_{o,s}^T\bar{\mathcal{K}}_{o,s})}{\partial\theta_i}\left(I-\bar{\mathcal{K}}_{d,s}^T\bar{\mathcal{K}}_{d,s}\right)\xi.
\end{aligned}
$$

Making use of the result of singular value decomposition in (16) gives

$$\bar{\mathcal{K}}_{o,s}^T\bar{\mathcal{K}}_{o,s}=V_1V_1^T=v_1v_1^T+v_2v_2^T+\cdots+v_\mu v_\mu^T,$$

where $V_1=\begin{bmatrix}v_1 & v_2 & \cdots & v_\mu\end{bmatrix}$ and μ means the row number of matrix $\bar{\mathcal{K}}_{o,s}$. Therefore,

$$\frac{\partial(\bar{\mathcal{K}}_{o,s}^T\bar{\mathcal{K}}_{o,s})}{\partial\theta_i}=\frac{\partial\left(v_1v_1^T+v_2v_2^T+\cdots+v_\mu v_\mu^T\right)}{\partial\theta_i}=\sum_{\alpha=1}^{\mu}\frac{\partial(v_\alpha v_\alpha^T)}{\partial\theta_i},$$

in which

$$
\begin{aligned}
\frac{\partial(v_\alpha v_\alpha^T)}{\partial\theta_i}=&\frac{\partial v_\alpha}{\partial\theta_i}v_\alpha^T+v_\alpha\frac{\partial(v_\alpha^T)}{\partial\theta_i}\\
=&\frac{\partial v_\alpha}{\partial\theta_i}v_\alpha^T+v_\alpha\left(\frac{\partial v_\alpha}{\partial\theta_i}\right)^T.
\end{aligned}
$$

Notice that

$$\mathcal{K}_{o,s}^T\mathcal{K}_{o,s} = \begin{bmatrix} V_1 & V_2 \end{bmatrix}\begin{bmatrix} \Sigma_1^T \\ 0 \end{bmatrix}U^TU\begin{bmatrix} \Sigma_1 & 0 \end{bmatrix}\begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$$
$$= \begin{bmatrix} V_1 & V_2 \end{bmatrix}\begin{bmatrix} \Sigma_1^2 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix},$$

where $\Sigma_1 = diag\{\sigma_1, \sigma_2, \cdots, \sigma_\mu\}$. As a result, v_α is the eigenvector corresponding to the i-th eigenvalue of matrix $\mathcal{K}_{o,s}^T\mathcal{K}_{o,s}$. For any $\alpha = 1, 2, \cdots, \mu, i = 1, 2, \cdots, \tau$,

$$\frac{\partial v_\alpha}{\partial \theta_i} = \frac{(\sigma_\alpha^2 I - \mathcal{K}_{o,s}^T\mathcal{K}_{o,s})^+\partial(\mathcal{K}_{o,s}^T\mathcal{K}_{o,s})v_\alpha}{\partial\theta_i}$$
$$= \left(\sigma_\alpha^2 I - \mathcal{K}_{o,s}^T\mathcal{K}_{o,s}\right)^+\frac{\partial(\mathcal{K}_{o,s}^T\mathcal{K}_{o,s})}{\partial\theta_i}v_\alpha.$$

It can be easily obtained that

$$\frac{\partial(\mathcal{K}_{o,s}^T\mathcal{K}_{o,s})}{\partial\theta_i} = \frac{\partial(\mathcal{K}_{o,s}^T)}{\partial\theta_i}\mathcal{K}_{o,s} + \mathcal{K}_{o,s}^T\frac{\partial\mathcal{K}_{o,s}}{\partial\theta_i}$$
$$= \left(\frac{\partial\mathcal{K}_{o,s}}{\partial\theta_i}\right)^T\mathcal{K}_{o,s} + \mathcal{K}_{o,s}^T\frac{\partial\mathcal{K}_{o,s}}{\partial\theta_i}.$$

Substituting the Formula (15) into the partial derivative leads to

$$\frac{\partial\mathcal{K}_{o,s}}{\partial\theta_i} = \frac{\partial\begin{bmatrix} D_v & \cdots & 0 & 0 & I & \cdots & 0 & 0 \\ C_vB_v & \cdots & 0 & 0 & C_vL_v & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ C_vA_v^{s-1}B_v & \cdots & C_vB_v & D_v & C_vA_v^{s-1}L_v & \cdots & C_vL_v & I \end{bmatrix}}{\partial\theta_i}.$$

As θ_i may be one element of any vector of $\theta_{AB}, \theta_{C_v}, \theta_{D_v}, \theta_{L_v}$, the deduction will be divided into four cases in the following.

First, when the parameter θ_i is one element of vector θ_{AB},

$$\frac{\partial\mathcal{K}_{o,s}}{\partial\theta_{AB}(i)} = \begin{bmatrix} \frac{\partial H_{u,s}}{\partial\theta_{AB}(i)} & \frac{\partial H_{y,s}}{\partial\theta_{AB}(i)} \end{bmatrix},$$

where

$$\frac{\partial H_{u,s}}{\partial\theta_{AB}(i)} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ C_v\frac{\partial B_v}{\partial\theta_{AB}(i)} & 0 & \cdots & 0 \\ C_v\left(\frac{\partial A_v}{\partial\theta_{AB}(i)}B_v + A_v\frac{\partial B_v}{\partial\theta_{AB}(i)}\right) & C_v\frac{\partial B_v}{\partial\theta_{AB}(i)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C_v\left(\frac{\partial A_v^{s-1}}{\partial\theta_{AB}(i)}B_v + A_v^{s-1}\frac{\partial B_v}{\partial\theta_{AB}(i)}\right) & C_v\left(\frac{\partial A_v^{s-2}}{\partial\theta_{AB}(i)}B_v + A_v^{s-2}\frac{\partial B_v}{\partial\theta_{AB}(i)}\right) & \cdots & C_v\frac{\partial B_v}{\partial\theta_{AB}(i)} \end{bmatrix},$$

$$\frac{\partial H_{y,s}}{\partial\theta_{AB}(i)} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ C_v\frac{\partial A_v}{\partial\theta_{AB}(i)}L_v & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C_v\frac{\partial A_v^{s-1}}{\partial\theta_{AB}(i)}L_v & C_v\frac{\partial A_v^{s-2}}{\partial\theta_{AB}(i)}L_v & \cdots & 0 \end{bmatrix},$$

and

$$\frac{\partial\begin{bmatrix} B_\nu^T(\theta_{AB}) \\ A_\nu^T(\theta_{AB}) \end{bmatrix}}{\partial\theta_{AB}(i)} = \begin{bmatrix} \frac{\partial B_\nu^T(\theta_{AB})}{\partial\theta_{AB}(i)} \\ \frac{\partial A_\nu^T(\theta_{AB})}{\partial\theta_{AB}(i)} \end{bmatrix} = \Psi_1(\theta_{AB}(1))\cdots\frac{\partial\Psi_i(\theta_{AB}(i))}{\partial\theta_{AB}(i)}\cdots\Psi_{nl}(\theta_{AB}(nl))\begin{bmatrix} 0 \\ I_n \end{bmatrix},$$

$$\frac{\partial\Psi_i(\theta_{AB}(i))}{\partial\theta_{AB}(i)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{\partial U(\theta_{AB}(i))}{\partial\theta_{AB}(i)} & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\frac{\partial U(\theta_{AB}(i))}{\partial\theta_{AB}(i)} = \begin{bmatrix} -1 & \frac{-\theta_{AB}(i)}{\sqrt{1-\theta_{AB}^2(i)}} \\ \frac{-\theta_{AB}(i)}{\sqrt{1-\theta_{AB}^2(i)}} & 1 \end{bmatrix}.$$

Second, when the parameter θ_i is one element of vector θ_{C_ν}, assume $\theta_i = c_{pq}$, $p \in \{1,2,\cdots,m\}, q \in \{1,2,\cdots,n\}$, where c_{pq} corresponds to the p-th row and the q-th column of matrix C_ν,

$$\frac{\partial\begin{bmatrix} D_\nu & \cdots & 0 & 0 & I & \cdots & 0 & 0 \\ C_\nu B_\nu & \cdots & 0 & 0 & C_\nu L_\nu & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ C_\nu A_\nu^{s-1}B_\nu & \cdots & C_\nu B_\nu & D_\nu & C_\nu A_\nu^{s-1}L_\nu & \cdots & C_\nu L_\nu & I \end{bmatrix}}{\partial c_{pq}}$$

$$= \begin{bmatrix} 0 & 0 & \cdots & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \frac{\partial C_\nu}{\partial c_{pq}}B_\nu & 0 & \cdots & \vdots & \vdots & \frac{\partial C_\nu}{\partial c_{pq}}L_\nu & 0 & \cdots & 0 & 0 \\ \frac{\partial C_\nu}{\partial c_{pq}}A_\nu B_\nu & \frac{\partial C_\nu}{\partial c_{pq}}B_\nu & \cdots & \vdots & \vdots & \frac{\partial C_\nu}{\partial c_{pq}}A_\nu L_\nu & \frac{\partial C_\nu}{\partial c_{pq}}L_\nu & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \cdots & \ddots & \vdots \\ \frac{\partial C_\nu}{\partial c_{pq}}A_\nu^{s-1}B_\nu & \frac{\partial C_\nu}{\partial c_{pq}}A_\nu^{s-2}B_\nu & \ldots & \frac{\partial C_\nu}{\partial c_{pq}}B_\nu & 0 & \frac{\partial C_\nu}{\partial c_{pq}}A_\nu^{s-1}L_\nu & \frac{\partial C_\nu}{\partial c_{pq}}A_\nu^{s-2}L_\nu & \cdots & \frac{\partial C_\nu}{\partial c_{pq}}L_\nu & 0 \end{bmatrix}$$

with

$$\frac{\partial C_\nu}{\partial c_{pq}} = \frac{\partial\begin{bmatrix} c_{11} & \cdots & c_{1q} & \cdots & c_{1n} \\ \vdots & & \vdots & & \vdots \\ c_{p1} & \cdots & c_{pq} & \cdots & c_{pn} \\ \vdots & & \vdots & & \vdots \\ c_{m1} & \cdots & c_{mq} & \cdots & c_{mn} \end{bmatrix}}{\partial c_{pq}}$$

$$= \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}.$$

Third, when the parameter θ_i is one element of vector θ_{D_ν}, assume $\theta_i = d_{pq}$, $p \in \{1,2,\cdots,m\}$, $q \in \{1,2,\cdots,l\}$, where d_{pq} corresponds to the p-th row and the q-th column of matrix D_ν,

$$
\frac{\partial \begin{bmatrix} D_v & \cdots & 0 & 0 & I & \cdots & 0 & 0 \\ C_v B_v & \cdots & 0 & 0 & C_v L_v & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ C_v A_v^{s-1} B_v & \cdots & C_v B_v & D_v & C_v A_v^{s-1} L_v & \cdots & C_v L_v & I \end{bmatrix}}{\partial d_{pq}}
$$
$$
= \begin{bmatrix} \frac{\partial D_v}{\partial d_{pq}} & 0 & 0 & \cdots & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \frac{\partial D_v}{\partial d_{pq}} & 0 & \cdots & \vdots & 0 & 0 & \cdots & \cdots & \vdots \\ 0 & 0 & \frac{\partial D_v}{\partial d_{pq}} & \cdots & \vdots & 0 & 0 & 0 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & \frac{\partial D_v}{\partial d_{pq}} & 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}
$$

with

$$
\frac{\partial D_v}{\partial d_{pq}} = \frac{\partial \begin{bmatrix} d_{11} & \cdots & d_{1q} & \cdots & d_{1l} \\ \vdots & & \vdots & & \vdots \\ d_{p1} & \cdots & d_{pq} & \cdots & d_{pl} \\ \vdots & & \vdots & & \vdots \\ d_{m1} & \cdots & d_{mq} & \cdots & d_{ml} \end{bmatrix}}{\partial d_{pq}}
$$
$$
= \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}.
$$

Fourth, when the parameter θ_i is one element of vector θ_{L_v}, assume $\theta_i = l_{pq}$, $p \in \{1, 2, \cdots, n\}, q \in \{1, 2, \cdots, m\}$, where l_{pq} corresponds to the p-th row and the q-th column of matrix L_v,

$$
\frac{\partial \begin{bmatrix} D_v & \cdots & 0 & 0 & I & \cdots & 0 & 0 \\ C_v B_v & \cdots & 0 & 0 & C_v L_v & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ C_v A_v^{s-1} B_v & \cdots & C_v B_v & D_v & C_v A_v^{s-1} L_v & \cdots & C_v L_v & I \end{bmatrix}}{\partial l_{pq}}
$$
$$
= \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \vdots & C_v \frac{\partial L_v}{\partial l_{pq}} & 0 & \cdots & \cdots & \vdots \\ 0 & 0 & \cdots & \vdots & C_v A_v \frac{\partial L_v}{\partial l_{pq}} & C_v \frac{\partial L_v}{\partial l_{pq}} & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & C_v A_v^{s-1} \frac{\partial L_v}{\partial l_{pq}} & \cdots & \cdots & C_v \frac{\partial L_v}{\partial l_{pq}} & 0 \end{bmatrix}
$$

with

$$\frac{\partial L_v}{\partial l_{pq}} = \frac{\partial \begin{bmatrix} l_{11} & \cdots & l_{1q} & \cdots & l_{1m} \\ \vdots & & \vdots & & \vdots \\ l_{p1} & \cdots & l_{pq} & \cdots & l_{pm} \\ \vdots & & \vdots & & \vdots \\ l_{n1} & \cdots & l_{nq} & \cdots & l_{nm} \end{bmatrix}}{\partial l_{pq}} = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}.$$

Up to now, the optimization problem is solved and the parameter optimization is derived by the above deduction. To summarize, this study aims to achieve the online configuration of an observer-based residual generator for fault detection systems. To achieve this, the observer is first transformed into the input normal form, which guarantees the asymptotic stability. Based on the input normal form, the residual generator is established, and the purpose is to online configure its parameters to cope with the operation condition changes or uncertainties of the system. Then, the optimization configuration strategy is taken into account and the optimization problem is proposed by utilizing the K-gap metric concept. By employing the gradient descent method, the parameter configuration is finally obtained for the online configuration realization.

Remark 1. *Note that the optimization problem of this paper is established based on the K-gap metric concept. The purpose is to minimize the K-gap metric* $\delta_{k_{d,s}}(\mathcal{K}_{o,s}, \mathcal{K}_{d,s})$ *between the observer and system plant, which characterizes the distance between two kernel subspaces. Referring to the work in [37], a definition of cluster is described as below. For* $\delta_f \in (0,1)$*, the set*

$$\mathcal{C}_f \subseteq \{\mathcal{K} : \delta_k\left(\mathcal{K}, \mathcal{K}_f\right) \leq \delta_f\}$$

is called $\mathcal{C}_f$ *cluster with the cluster center* $\mathcal{K}_f$ *and cluster radius* δ_f*. From this point of view, by minimizing the K-gap metric* $\delta_{k_{d,s}}(\mathcal{K}_{o,s}, \mathcal{K}_{d,s})$*, the obtained* $\mathcal{K}_{o,s}$ *falls into the cluster with the cluster center* $\mathcal{K}_{d,s}$ *and a certain cluster radius. A smaller cluster radius contributes to a more similar level of* $\mathcal{K}_{o,s}$ *to* $\mathcal{K}_{d,s}$*. Therefore, the residual generator based on the optimized* $\mathcal{K}_{o,s}$ *can guarantee the reliability and efficiency of the fault detection system.*

3.3. Online Configuration Realization of Fault Detection

This subsection will present an online algorithm to achieve the fault detection goal by employing the newly proposed gradient configuration scheme. The detailed procedure of online configuring the observer-based residual generator is described in Algorithm 1 as follows.

To achieve the fault detection, the evaluation function is set to be $\mathcal{J}(k) = \|r(k)\|_2^2$ with the threshold $\mathcal{J}_{th}$ chosen as below,

$$\mathcal{J}_{th} = \sup_{f=0,d,\Delta} \mathcal{J}(k),$$

where d and Δ represent the disturbance and model uncertainties, respectively. Based on the online residual generator obtained by Algorithm 1, calculate the evaluation function $\mathcal{J}(k)$. The decision logic is described by

$$\begin{cases} \mathcal{J}(k) > \mathcal{J}_{th} \Rightarrow \text{faulty}, \\ \mathcal{J}(k) \leq \mathcal{J}_{th} \Rightarrow \text{fault-free}. \end{cases} \tag{20}$$

Algorithm 1 Online Configuration of Observer-Based Residual Generator

Step 1:	Collect the I/O data $u(k), y(k)$ at each k and select an iteration interval W
Step 2:	Execute Step 3–6 for the j-th iteration every W
Step 3:	According to Definition 3, compute the SKR $\mathcal{K}_{d,s}$ and its normalized result $\bar{\mathcal{K}}_{d,s}$
Step 4:	Compute the SKR $\mathcal{K}_{o,s}$ and its normalized result $\bar{\mathcal{K}}_{o,s}$ corresponding to the observer by (15) and (16)
Step 5:	Apply Lemma 2 to obtain the K-gap metric $\delta_{k_{d,s}}(\mathcal{K}_{o,s}, \mathcal{K}_{d,s})$ between $\mathcal{K}_{o,s}$ and $\mathcal{K}_{d,s}$
Step 6:	Given a scalar ε, if $\delta_{k_{d,s}}(\mathcal{K}_{o,s}, \mathcal{K}_{d,s}) > \varepsilon$, compute the gradient and update the parameters of residual generator according to (19), increase j by 1 and return Step 2, otherwise the configuration ends

4. Simulation Experiment

In this section, a simulation experiment is carried out on a three-tank system plant to show the effectiveness and advantages of the proposed real-time configuration scheme. Figure 2 shows the schematic diagram of the three-tank system, which is composed of three water tanks and some connecting pipes. In Figure 2, h_1, h_2, and h_3 refer to the water levels of the three tanks, which are measurable through sensors and deemed as the output signals. Q_1 and Q_2 stand for the incoming mass flow rates of Pump 1 and Pump 2 and are used as the input signals. Besides, PV_1, PV_2, PV_3, LV_1, LV_2, LV_3 are the adjustable ball valves to administrate the opening and closing of these pipes.

The system plant can be represented by the nonlinear dynamics

$$\begin{cases} \mathcal{A}\dot{h}_1 = Q_1 - \alpha_1 s sgn(h_1 - h_3)\sqrt{2g|h_1 - h_3|}, \\ \mathcal{A}\dot{h}_2 = Q_2 + \alpha_3 s sgn(h_3 - h_2)\sqrt{2g|h_3 - h_2|} - \alpha_2 s\sqrt{2gh_2}, \\ \mathcal{A}\dot{h}_3 = \alpha_1 s sgn(h_1 - h_3)\sqrt{2g|h_1 - h_3|} - \alpha_3 s sgn(h_3 - h_2)\sqrt{2g|h_3 - h_2|}, \end{cases} \tag{21}$$

in which $\mathcal{A} = 154\ \text{cm}^2$ and $s = 0.5\ \text{cm}^2$ denote the cross section area of the tanks and pipes, respectively, and $\alpha_1 = 0.46, \alpha_2 = 0.60, \alpha_3 = 0.45$ successively indicate the coefficients of flow for the three pipes. In addition, the maximum height of the tanks is chosen as $H_{max} = 62$ cm, and the maximum flow rates of pumps 1 and 2 are set to be $Q_{1max} = Q_{2max} = 100\ \text{cm}^3/\text{s}$. For certain operation points, the nonlinear representation (21) can be reformulated into the LTI system (1) by utilizing the linearization technique.

In this simulation, the operating time is set to be 25,000 s and the sampling period is chosen as 1s. At first, the operation point of the three-tank system is considered to be $h_1 = 45$ cm, $h_2 = 15$ cm, and $h_3 = 30$ cm, for which a residual generator is predesigned and applied in the fault detection process. Due to the demand of practical industry, it is supposed that the operation point changes to $h_1 = 50$ cm, $h_2 = 46$ cm and $h_3 = 48$ cm at 6000 s. The process data are collected and displayed in Figure 3.

Figure 4 displays the evolution curves of residual signals and evaluation function and Figure 5 exhibits the K-gap metric between system plant and observer. Before the change of the operation point at 6000 s, the evaluation function value is lower than the threshold and the K-gap metric value is 0.0007072, which means the kernel subspaces of system plant and observer are sufficiently close and thus the predesigned residual generator is proper. From 6000 s, the value of evaluation function increases and obviously exceeds the threshold, which causes a false alarm under the circumstance of no fault. Meanwhile, the K-gap metric between system plant and observer becomes enormous. Clearly, the predesigned residual generator is no more applicable for the system with a changed operation point. To handle this problem and demonstrate the validity of the proposed real-time configuration method, the real-time configuration algorithm, i.e., Algorithm 1 is implemented at 9000 s. It can be observed from Figures 4 and 5 that from 9000 s, the K-gap metric value begins to decline as the real-time configuration implementation, and the values of residual signal and evaluation function, begin to decrease and gradually converge. At ~14,700 s, the K-gap metric settles at ~0.001 and the kernel subspace of observer is adequately approximate to the kernel subspace of system plant. In addition, the value of evaluation function becomes less than the threshold and the false alarm disappears, which reflects that the real-time optimized residual generator is effective for the current operation point. Therefore, the optimization goal is realized and the real-time configuration of the residual generator is achieved for the system. In addition, the optimized parameters of θ_{AB}, θ_C are, respectively, given in Figures 6 and 7 to show the parameter configuration process.

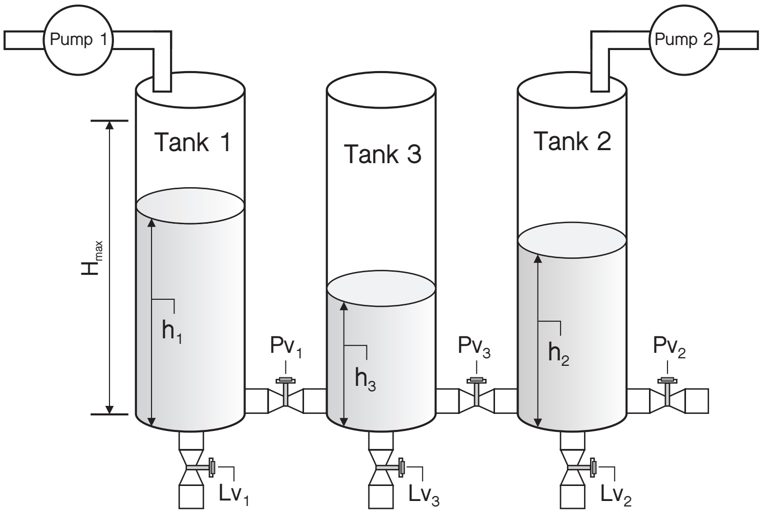

Figure 2. Schematic diagram of three-tank system.

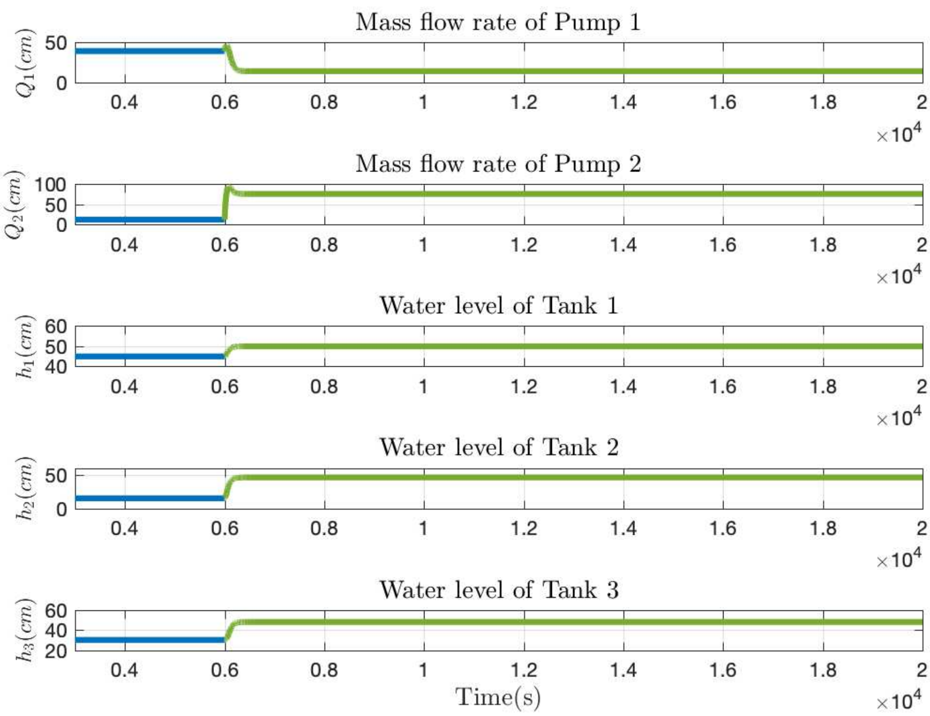

Figure 3. Input and output data.

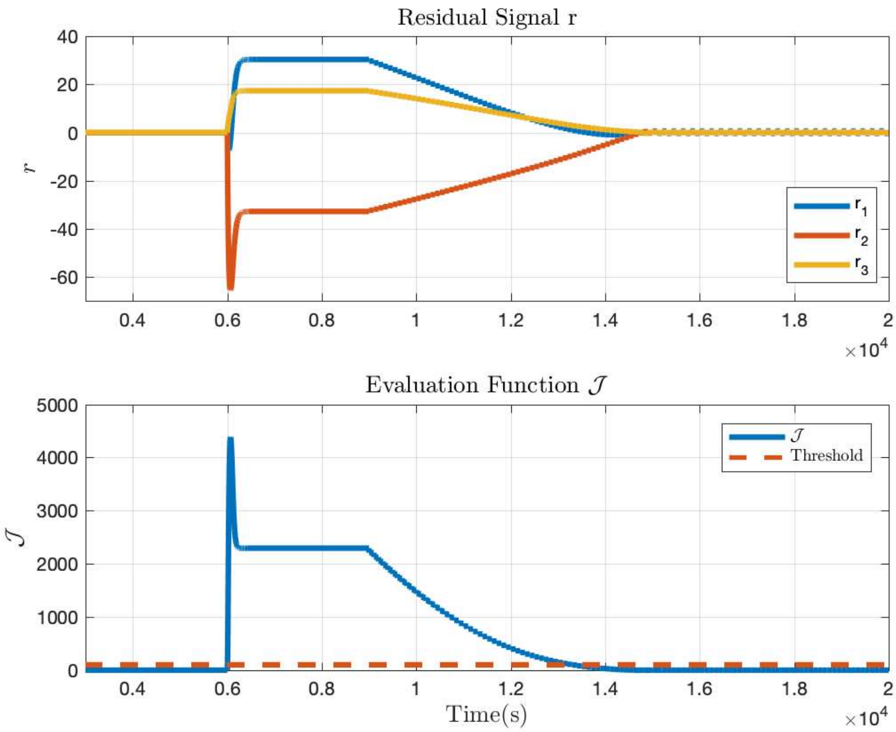

Figure 4. Residual signal and evaluation function under the optimization configuration.

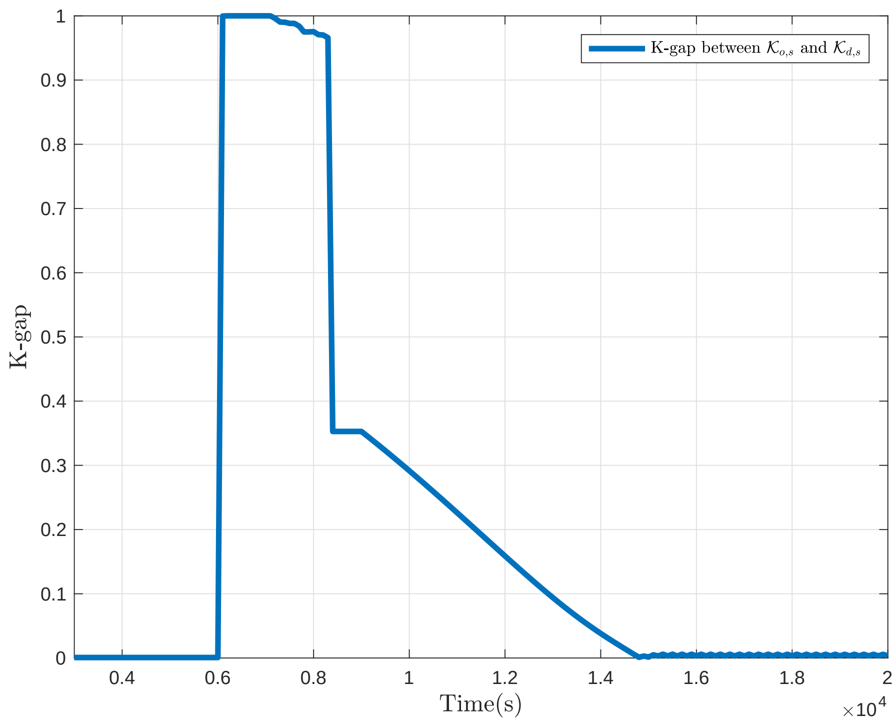

Figure 5. K-gap metric under the online optimization configuration.

Now, the online configured residual generator is applied to the fault detection implementation to verify its usefulness. Here, two types of fault will be considered and assumed to happen at 23,000 s, respectively. First, consider that a leakage fault happens in the tank 2 with a 20% leakage level at 23,000 s. Figure 8 shows the fault detection result by the online configured residual generator. It can be seen from the figure that the value of the evaluation function exceeds the threshold at 23,005 s, which means that the leakage fault is detected. Then, the usefulness of the fault detection based on the online configuration scheme is illustrated. Moreover, another type of fault, i.e., a drift fault with slope 0.01 occurring in the sensor of water level of tank 2, is considered at 23,000 s. Figure 9 shows the corresponding fault detection result, from which one can see that the drift fault is detected by the proposed optimization configuration algorithm at 23,042 s. According to the above fault detection results, it is clear that the optimization configuration method of this paper can detect the fault accurately and timely. As a consequence, the real-time configuration method proposed in this paper can effectively deal with the system under the influence of operation point changes or uncertainties, and the effectiveness of the method in fault detection implementation is demonstrated.

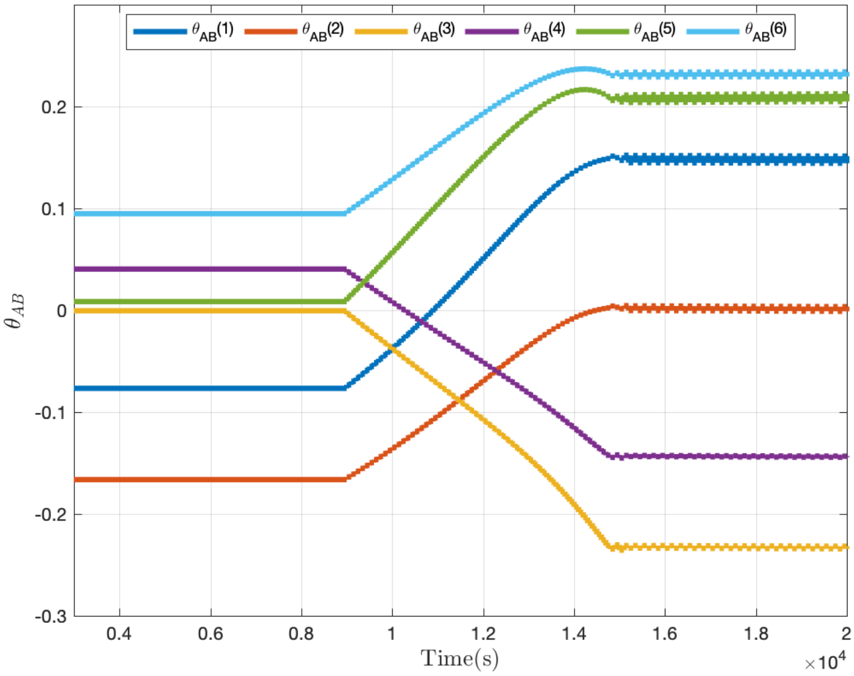

Figure 6. Optimized parameters of θ_{AB}.

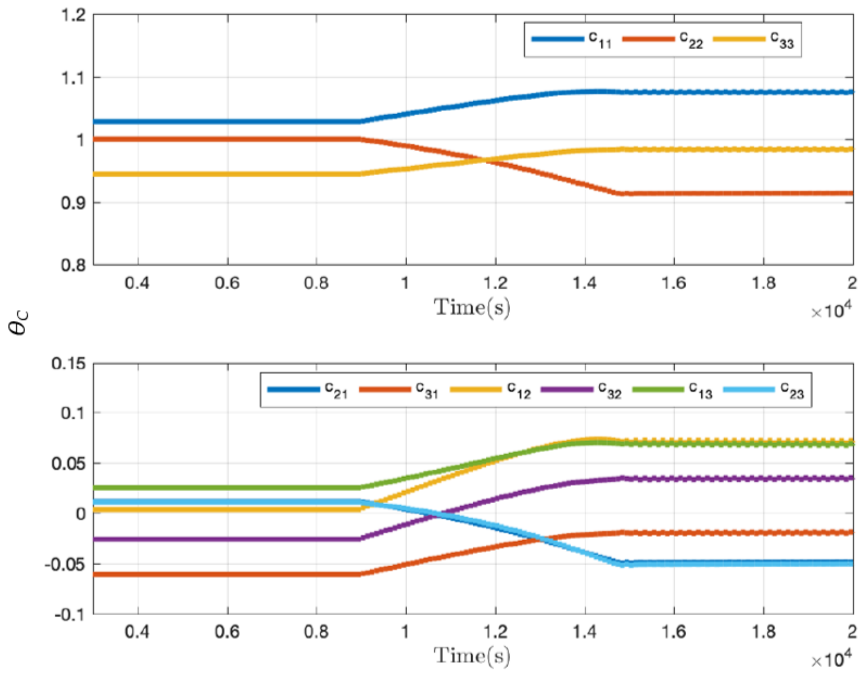

Figure 7. Optimized parameters of θ_C.

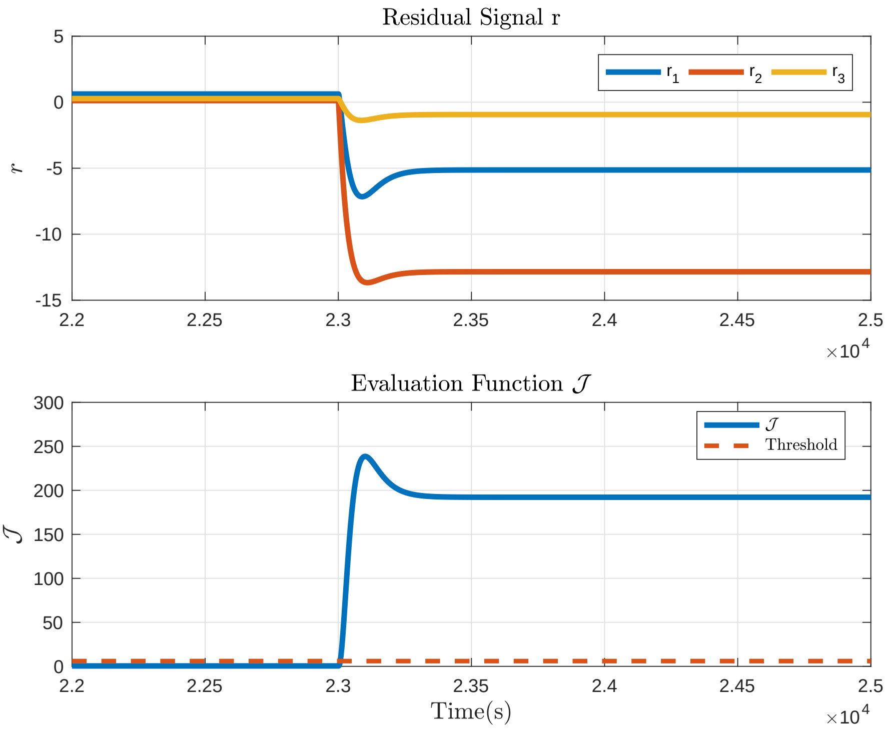

Figure 8. Fault detection by the optimization configuration under leakage fault.

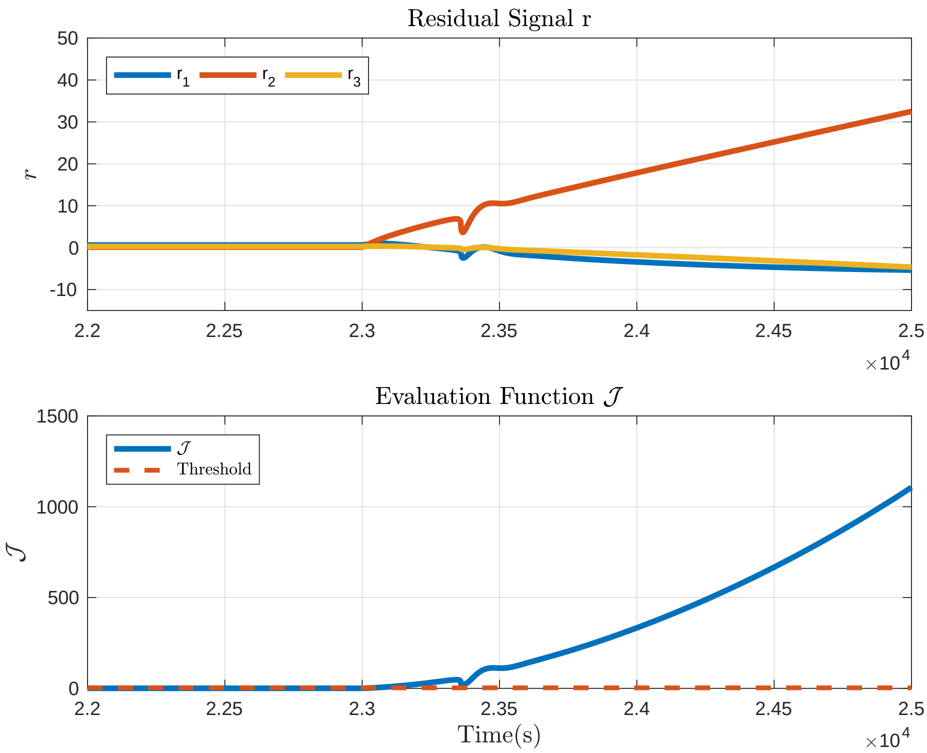

Figure 9. Fault detection by the optimization configuration under sensor drift fault.

5. Conclusions

Considering that the changes of operating conditions, practical environment, and some uncertainties may occur in industrial processes, this paper is dedicated to the real-time configuration design for fault detection systems. As an important means of real-time configuration, the gradient optimization scheme is considered and adopted in this work for the real-time configuration implementation. A novel optimization algorithm is developed by virtue of the gradient-based technique, in which the K-gap metric between the residual generator and the current system is minimized. Then, the residual generator parameters can be updated based on the I/O data, and the real-time configuration for fault detection systems is realized to satisfy a particular demand. Finally, the usefulness and merits of the proposed approach are demonstrated through the benchmark case on a three-tank model. One main advantage of this work is that by virtue of the K-gap metric technique together with the gradient-based method, the online-configured residual generator parameters are reliable to guarantee the fault detection performance for industrial systems with changeable operating points. Besides, the input/output data information are sufficiently exploited for the fault detection implementation, which avoids the difficulties of the system identification in practice. As the real-time configuration method is carried out based on the process data, the computation amount would be the main concern of the implementation, especially for large-scale industrial systems, which inspires us to conduct further research in this field. Note that one of the main contributions of this paper is to introduce the K-gap idea into the real-time configuration implementation, which exploits the essential characteristic of fault detection systems. In our future work, the K-gap-based optimization configuration approach will be further extended to systems with nonlinearities and fault-tolerant control issues.

Author Contributions: Conceptualization, H.Z.; Formal analysis, T.L.; Funding acquisition, H.L.; Methodology, H.Z.; Software, H.Z.; Supervision, H.L.; Writing—original draft, H.Z.; Writing—review & editing, H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by National Natural Science Foundation of China under Grant U20A20186, 62073104.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fekih, A.; Xu, H.; Chowdhury, F. Neural networks based system identification techniques for model based fault detection of nonlinear systems. *Int. J. Innov. Comput. Inf. Control* **2007**, *3*, 1073–1085.
2. Wang, X.L.; Yang, G.H. Event-triggered fault detection for discrete-time T-S fuzzy systems. *ISA Trans.* **2018**, *76*, 18–30. [CrossRef] [PubMed]
3. Son, J.; Du, Y. Model-based stochastic fault detection and diagnosis of Lithium-Ion batteries. *Processes* **2019**, *7*, 38. [CrossRef]
4. Mazzoletti, M.A.; Bossio, G.R.; De Angelo, C.H.; Espinoza-Trejo, D.R. A Model-based strategy for interturn short-circuit fault diagnosis in PMSM. *IEEE Trans. Ind. Electron.* **2017**, *64*, 7218–7228. [CrossRef]
5. Huang, S.P.; Xiang, Z.R.; Karimi, H.R. Mixed L_{-}/L_{1} fault detection filter design for fuzzy positive linear systems with time-varying delays. *IET Control Theory A* **2014**, *8*, 1023–1031. [CrossRef]
6. Poon, J.; Jain, P.; Konstantakopoulos, I.; Spanos, C.; Panda, S.; Sanders, S. Model-based fault detection and identification for switching power converters. *IEEE Trans. Power Electron.* **2017**, *32*, 1419–1430. [CrossRef]
7. Yang, C.; Fang, H. A new nonlinear model-based fault detection method using Mann-Whitney test. *IEEE Trans. Ind. Electron.* **2020**, *67*, 10856–10864. [CrossRef]
8. Gao, Y.; Xiao, F.; Liu, J.; Wang, R. Distributed soft fault detection for interval type-2 fuzzy-model-based stochastic systems with wireless sensor networks. *IEEE Trans. Ind. Inform.* **2019**, *15*, 334–347. [CrossRef]
9. Zhou, S.; Bai, J.; Wu, F. Decentralized fault detection and fault-tolerant control for nonlinear interconnected systems. *Processes* **2021**, *9*, 591. [CrossRef]
10. Zhang, P.; Ding, S.X.; Wang, G.Z.; Zhou, D.H. Fault detection of linear discrete-time periodic systems. *IEEE Trans. Automat. Control* **2005**, *50*, 239–244. [CrossRef]
11. Zhong, M.; Ding, S.X.; Ding, E.L. Optimal fault detection for linear discrete time-varying systems. *Automatica* **2010**, *46*, 1395–1400. [CrossRef]
12. Li, X.; Yang, G.H. Fault detection for linear stochastic systems with sensor stuck faults. *Optim. Contr. Appl. Met.* **2012**, *33*, 61–80. [CrossRef]
13. Su, Q.Y.; Li, J. Fault detection for a class of uncertain linear systems. *Math. Probl. Eng.* **2013**, *33*, 856914. [CrossRef]

14. Fan, C.; Lam, J.; Xie, X. Fault detection observer design for periodic piecewise linear systems. *Int. J. Syst. Sci.* **2020**, *51*, 1622–1636. [CrossRef]
15. Li, L.; Ding, S.X.; Qiu, J.; Yang, Y.; Xu, D. Fuzzy observer-based fault detection design approach for nonlinear processes. *IEEE Trans. Syst. Man Cybern. Syst.* **2017**, *47*, 1941–1952. [CrossRef]
16. Li, L.; Ding, S.X.; Yang, Y.; Zhang, Y. Robust fuzzy observer-based fault detection for nonlinear systems with disturbances. *Neurocomputing* **2016**, *174*, 767–772. [CrossRef]
17. Zhirabok, A.; Shumsky, A.; Solyanik, S.; Suvorov, A. Fault detection in nonlinear systems via linear methods. *Int. J. Appl. Math. Comput. Sci.* **2017**, *27*, 261–272. [CrossRef]
18. Boem, F.; Riverso, S.; Ferrari-Trecate, G.; Parisini, T. Plug-and-Play fault detection and isolation for large-scale nonlinear systems with stochastic uncertainties. *IEEE Trans. Automat. Control* **2019**, *64*, 4–19. [CrossRef]
19. Venkateswaran, S.; Liu, Q.C.; Wilhite, B.A.; Kravaris, C. Design of linear residual generators for fault detection and isolation in nonlinear systems. *Int. J. Control* **2020**, 1–17. [CrossRef]
20. Ding, S.X.; Yang, Y.; Zhang, Y.; Li, L. Data-driven realizations of kernel and image representations and their application to fault detection and control system design. *Automatica* **2014**, *50*, 2615–2623. [CrossRef]
21. Yin, S.; Wang, G.; Karimi, H.R. Data-driven design of robust fault detection system for wind turbines. *Mechatronics* **2014**, *24*, 298–306. [CrossRef]
22. Naderi, E.; Khorasani, K. A data-driven approach to actuator and sensor fault detection, isolation and estimation in discrete-time linear systems. *Automatica* **2017**, *85*, 165–178. [CrossRef]
23. Tariq, M.F.; Khan, A.Q.; Abid, M.; Mustafa, G. Data-driven robust fault detection and isolation of three-phase induction motor. *IEEE Trans. Ind. Electron.* **2019**, *66*, 4707–4715. [CrossRef]
24. Luo, H.; Li, K.; Kaynak, O.; Yin, S.; Huo, M.; Zhao, H. A robust data-driven fault detection approach for rolling mills with unknown roll eccentricity. *IEEE Trans. Control Syst. Technol.* **2020**, *28*, 2641–2648. [CrossRef]
25. Fu, F.; Wang, D.; Li, L.; Li, W.; Wu, Z. Data-driven method for the quantitative fault diagnosability analysis of dynamic systems. *IET Control Theory A* **2019**, *13*, 1197–1203. [CrossRef]
26. Yu, D.; Chen, Z.M.; Xiahou, K.S.; Li, M.S.; Ji, T.Y.; Wu, Q.H. A radically data-driven method for fault detection and diagnosis in wind turbines. *Int. J. Electr. Power Energy Syst.* **2018**, *99*, 577–584. [CrossRef]
27. Wang, X.; Yang, G.; Zhang, D. Data-driven fault detection for linear systems: A q-step residual iteration approach. *Int. J. Robust Nonlinear* **2020**, *30*, 5341–5355. [CrossRef]
28. Li, L.; Ding, S.X.; Peng, X. Distributed data-driven optimal fault detection for large-scale systems. *J. Process Control* **2020**, *96*, 94–103. [CrossRef]
29. Kallas, M.; Mourot, G.; Maquin, D.; Ragot, J. Data-driven approach for fault detection and isolation in nonlinear system. *Int. J. Adapt. Control Signal Process.* **2018**, *32*, 1569–1590. [CrossRef]
30. Chen, H.T.; Jiang, B.; Chen, W.; Yi, H. Data-driven detection and diagnosis of incipient faults in electrical drives of high-speed trains. *IEEE Trans. Ind. Electron.* **2019**, *66*, 4716–4725. [CrossRef]
31. Ding, S.X.; Yin, S.; Zhang, P.; Ding, E.L.; Naik, A. An approach to data-driven adaptive residual generator design and implementation. *IFAC Proc.* **2009**, *42*, 941–946. [CrossRef]
32. Chen, J.M.; Yang, F.W. Data-driven subspace-based adaptive fault detection for solar power generation systems. *IET Control Theory A* **2013**, *7*, 1498–1508. [CrossRef]
33. Chen, Z.; Peng, T.; Yang, C.; Li, F.; He, Z. An adaptive data-driven fault detection method for monitoring dynamic process. In Proceedings of the IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society, Washington DC, USA, 21–23 October 2018; pp. 5353–5358.
34. Luo, H. *Plug-and-Play Monitoring and Performance Optimization for Industrial Automation Processes*; Springer Vieweg: Wiesbaden, Germany, 2017.
35. Ding, S.X. *Data-Driven Design of Fault Diagnosis and Fault-Tolerant Control Systems*; Springer: New York, NY, USA, 2014.
36. Vinnicombe, G. *Uncertainty and Feedback: Hinf Loop-Shaping and the V-Gap Metric*; World Science: New York, NY, USA, 2000.
37. Li, L.; Ding, S.X. Gap metric techniques and their application to fault detection performance analysis and fault isolation schemes. *Automatica* **2020**, *118*, 109029. [CrossRef]
38. Koenings, T.; Krueger, M.; Luo, H.; Ding, S.X. A data-driven computation method for the gap metric and the optimal stability margin. *IEEE Trans. Automat. Control* **2018**, *63*, 805–810. [CrossRef]
39. Li, H.; Yang, Y.; Zhao, Z.; Zhou, J.; Liu, R. Fault detection via data-driven K-gap metric with application to ship propulsion systems. In Proceedings of the 37th Chinese Control Conference (CCC), Guangzhou, China, 27–30 July 2018; pp. 6023–6027.
40. Hanzon, B.; Olivi, M.; Peeters, R.L.M. Balanced realizations of discrete-time stable all-pass systems and the tangential Schur algorithm. *Linear Algebra Its Appl.* **2006**, *418*, 793–820. [CrossRef]
41. Verhaegen, M.; Verdult, V. *Filtering and System Identification: A Least Squares Approach*; Cambridge University Press: Cambridge, UK, 2012.