Proceedings of the 40th Chinese Control Conference
July 26-28, 2021, Shanghai, China

Fault Detection Based on Canonical Correlation Analysis with Rank Constrained Optimization

Yuhan Zhang[1], Xianchao Xiu [1], Ying Yang[1], Wanquan Liu[2]

1. State Key Laboratory for Turbulence and Complex Systems, Department of Mechanics and Engineering Science, College of Engineering, Peking University, Beijing 100871, P. R. China
E-mail: yuhanzhang@pku.edu.cn, xcxiu@bjtu.edu.cn, yy@pku.edu.cn

2. School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou 510275, P. R. China
E-mail: wanquanliu@qq.com

Abstract: Canonical correlation analysis (CCA) has attracted increasing attention in the field of fault detection because it provides an effective way to explore the relationship between the input and output data. This paper develops a novel rank constrained CCA (RCCCA) framework, which is the first approach that takes the rank prior information into consideration. Technically, the rank constrained optimization is able to capture the global structures of variables, and thus improve the performance of fault detection. In order to solve RCCCA, an alternating minimization algorithm is designed, which aims to preserve the maximum correlation with the low-rank learning. A fault detection residual is then generated, and the test statistic is constructed to determine whether a fault occurs. The RCCCA-based fault detection is finally tested on a numerical example and the Tennessee Eastman benchmark process. Monitoring results indicate the efficiency and feasibility of the proposed method.

Key Words: Fault detection, canonical correlation analysis, rank constrained optimization, alternating minimization algorithm, Tennessee Eastman process

1 Introduction

Fault detection (FD) is widely used to meet the high requirements for product quality, maintenance efficiency and production costs. With the development of modern computing and sensor technology, data-driven FD has drawn more and more attention [1]. In general, it can be divided into three categories: multivariate analysis, signal processing, deep learning. It's no exaggeration to say that multivariate analysis (MVA) has become the dominant FD technique because of its simple design, easy implementation and high efficiency. In the last decades, a large number of MVA-based methods have been developed, of which the most popular ones include principal component analysis (PCA) [2], partial least squares (PLS) [3], independent component analysis (ICA) [4], canonical correlation analysis (CCA) [5]. Different from PCA, PLS and ICA, CCA integrates the multi-view information of process data in order to provide a better monitoring performance. More illustrations of CCA-based FD can be found in recent papers [6–9].

However, in practice, the FD process data are usually not only multi-view but also redundant. The word "redundant" indicates that not all the process variables in the original space are identically informative for representing the canonical variables. In order to obtain a consistent representation with encouraging discriminability, extensive contributions have been made in the research community [10]. One of the most impressive works is sparse CCA (SCCA), which embeds sparse structures to the learning process. It is verified that SCCA can discard certain unimportant variables and thus improve the interpretability of the canonical variables [11]. For detailed statistical properties and optimization algorithms about SCCA, please refer [12]. Assume that if one canonical variable is randomly discarded, it may cause changes in linear dependency, which has great damage to the underlying structures and brings significant negative effects on FD. There comes a question: how to discover a discriminative representation by incorporating prior information from multiple views with correlation analysis.

In fact, this prior information can be described by low-rank learning; see, e.g., [13, 14]. However, to the authors' best knowledge, the integration of CCA and rank optimization has not been fully studied yet. Only one related work [15] is found, which adds nuclear norm (sum of singular values) regularization terms to the objective, abbreviated as LRCCA. Even though LRCCA yields better performances in image inpainting, it is found that there exist some limitations. In the statistical aspect, LRCCA only considers a special case when the coefficients for nuclear norm are equal. Nevertheless, this assumption cannot be satisfied in real-world FD. In the optimization aspect, LRCCA relaxes the rank constraint into the nuclear norm. Although the objective is convex for each variable, it is not the intuitive description of prior global structures. In the algorithmic aspect, LRCCA lacks efficient solvers, which cannot guarantee consistent performance.

Motivated by the above discussions, this paper focuses on developing a novel CCA-based FD method. In particular, the rank constrained optimization is adopted to facilitate the interpretability of inherent relationships and control the learned canonical variables. Compared to the existing work, the main contributions of this paper are summarized as:

1) A new CCA-based FD method is proposed by introducing rank constraints to learn the canonical variables, which is the first work to establish a rank constrained CCA framework for FD.
2) An effective optimization approach is designed by combing the alternating minimization algorithm and gradient projection techniques.
3) The superiority of the proposed FD strategy is illustrated by case studies on a numerical example and the Tennessee Eastman process.

This work is supported by National Natural Science Foundation (NNSF) of China under Grand 12001019 and Grant 61633001.

The remainder of this paper is organized as follows. In Section 2, an online detection strategy using the proposed method is briefly discussed. An efficient optimization algorithm is developed in Section 3. The advantages of the proposed method are demonstrated by simulated and real-world processes in Section 4. Finally, conclusions and future work are given in Section 5.

2 Detection Strategy

This section describes the basic idea of offline training and online testing strategy by using the proposed RCCCA method. To some degree, CCA-type FD can be viewed as an extension of PCA-based FD, because CCA takes into account the relationship between input and output variables. The difficulties mainly lie in: how to construct an optimization model, how to design a residual signal, how to define a test statistic, how to determine a corresponding control limit, and how to make a decision.

Assume that the input training data and output training data are, respectively, given by

$$\begin{aligned} \mathbf{X} &= [\mathbf{x}_1\ \mathbf{x}_2\ \cdots\ \mathbf{x}_n]^T \in \mathbb{R}^{n\times p}\ (p \text{ variables}), \\ \mathbf{Y} &= [\mathbf{y}_1\ \mathbf{y}_2\ \cdots\ \mathbf{y}_n]^T \in \mathbb{R}^{n\times q}\ (q \text{ variables}). \end{aligned} \tag{1}$$

By exploiting the latent structures, a novel RCCCA model can be built in the form of

$$\begin{aligned} \min_{\mathbf{A},\mathbf{B}}\quad & -\mathrm{Tr}(\mathbf{A}^T\mathbf{X}^T\mathbf{Y}\mathbf{B}) \\ \text{s.t.}\quad & \mathbf{A}^T\mathbf{X}^T\mathbf{X}\mathbf{A} = \mathbf{I},\ \mathbf{B}^T\mathbf{Y}^T\mathbf{Y}\mathbf{B} = \mathbf{I}, \\ & \mathrm{rank}(\mathbf{A}) \le r_1,\ \mathrm{rank}(\mathbf{B}) \le r_2, \end{aligned} \tag{2}$$

where r_1, r_2 are the given low-rank levels. Optimization problem (2) is nonconvex even NP-hard due to the presence of the rank and manifold constraints. Although one can approximate it by convex relaxations, they are not intuitive from engineering. To this end, an efficient alternating minimization algorithm (AMA) will be given in the next section to seek the optimal solution.

Let $\mathbf{x} \in \mathbb{R}^p$ and $\mathbf{y} \in \mathbb{R}^q$ be the input testing sample and output testing sample, respectively. Following the arguments in [16], the residual vector can be defined as

$$\mathbf{r} = \mathbf{A}^T\mathbf{x} - \mathbf{\Sigma}\mathbf{B}^T\mathbf{y}, \tag{3}$$

where $\mathbf{\Sigma}$ is the correlation matrix of $\mathbf{A}$ and $\mathbf{B}$. In fact, it shares the same idea as in [5]. For data-driven FD methods, the T^2 statistic and squared prediction error (SPE) are most commonly used statistics. As is illustrated in [17], if there is no available prior information about the faults, the T^2 statistic delivers the best fault detectability. Therefore, the corresponding T^2 test statistic is then generated in the following way:

$$T^2 = \mathbf{r}^T(\mathbf{I} - \mathbf{\Sigma}^2)^{-1}\mathbf{r}. \tag{4}$$

Accordingly, the control limit can be determined by standard $\mathcal{X}^2$ distribution with a significance level α:

$$J_{th} = \mathcal{X}^2_\alpha(p), \tag{5}$$

which only relies on parameters p and α. In the experiment, α is usually set as 0.01 or 0.05.

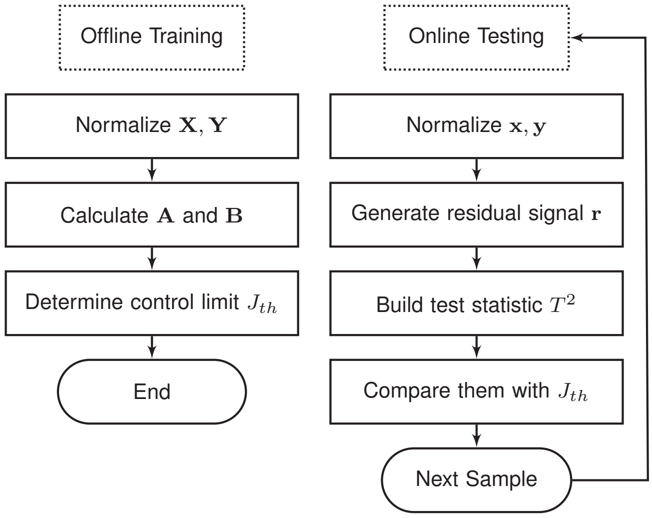

Fig. 1: Flow chart of the detection strategy.

Finally, a proper decision logic should be well addressed to detect the fault. If the T^2 test statistic violates the control limit, a fault occurs, otherwise fault-free. Therefore, the detection logic can be constructed as

$$\begin{cases} T^2 > J_{th} & \Rightarrow \ \text{faulty}, \\ T^2 \le J_{th} & \Rightarrow \ \text{fault-free}. \end{cases} \tag{6}$$

To end this section, a step-by-step detection strategy is provided in Fig. 1.

3 Optimization Algorithm

This section explains how to solve the proposed RCCCA model. By introducing two auxiliary variables $\mathbf{C}, \mathbf{D}$, problem (2) is first written as

$$\begin{aligned} \min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}}\quad & -\mathrm{Tr}(\mathbf{C}^T\mathbf{D}) \\ \text{s.t.}\quad & \mathrm{rank}(\mathbf{A}) \le r_1,\ \mathrm{rank}(\mathbf{B}) \le r_2, \\ & \mathbf{C}^T\mathbf{C} = \mathbf{I},\ \mathbf{D}^T\mathbf{D} = \mathbf{I}, \\ & \mathbf{X}\mathbf{A} = \mathbf{C},\ \mathbf{Y}\mathbf{B} = \mathbf{D}. \end{aligned} \tag{7}$$

To describe the iterates of the alternating minimization algorithm (AMA), the above optimization problem is then reformulated as follows,

$$\begin{aligned} \min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}}\quad & -\mathrm{Tr}(\mathbf{C}^T\mathbf{D}) + \frac{\beta_1}{2}\|\mathbf{X}\mathbf{A} - \mathbf{C}\|_F^2 + \frac{\beta_2}{2}\|\mathbf{Y}\mathbf{B} - \mathbf{D}\|_F^2 \\ \text{s.t.}\quad & \mathrm{rank}(\mathbf{A}) \le r_1,\ \mathrm{rank}(\mathbf{B}) \le r_2, \\ & \mathbf{C}^T\mathbf{C} = \mathbf{I},\ \mathbf{D}^T\mathbf{D} = \mathbf{I}, \end{aligned} \tag{8}$$

where $\beta_1, \beta_2 > 0$ are the penalty parameters to control the regularization terms.

For convenience, the four nonconvex constraints are denoted as

$$\begin{aligned} \mathcal{S}_1 &= \{\mathbf{A} \mid \mathrm{rank}(\mathbf{A}) \le r_1\}, \\ \mathcal{S}_2 &= \{\mathbf{B} \mid \mathrm{rank}(\mathbf{B}) \le r_2\}, \\ \mathcal{M}_1 &= \{\mathbf{C} \mid \mathbf{C}^T\mathbf{C} = \mathbf{I}\}, \\ \mathcal{M}_2 &= \{\mathbf{D} \mid \mathbf{D}^T\mathbf{D} = \mathbf{I}\}, \end{aligned} \tag{9}$$

and the objective function is represented as $F(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$. The AMA for solving RCCCA consists of the following four

Algorithm 1 Alternating minimization algorithm for solving the proposed RCCCA

Input: Given data $\mathbf{X}, \mathbf{Y}$, parameters $\beta_1, \beta_2 > 0$, and low-rank level $r_1, r_2 > 0$

Initialize: $(\mathbf{A}^0, \mathbf{B}^0, \mathbf{C}^0, \mathbf{D}^0)$

While not converged **do**

1: Update $\mathbf{A}^{k+1} = \underset{\mathbf{A}\in\mathcal{S}_1}{\operatorname{argmin}} F(\mathbf{A}, \mathbf{B}^k, \mathbf{C}^k, \mathbf{D}^k)$

2: Update $\mathbf{B}^{k+1} = \underset{\mathbf{B}\in\mathcal{S}_2}{\operatorname{argmin}} F(\mathbf{A}^{k+1}, \mathbf{B}, \mathbf{C}^k, \mathbf{D}^k)$

3: Update $\mathbf{C}^{k+1} = \underset{\mathbf{C}\in\mathcal{M}_1}{\operatorname{argmin}} F(\mathbf{A}^{k+1}, \mathbf{B}^{k+1}, \mathbf{C}, \mathbf{D}^k)$

4: Update $\mathbf{D}^{k+1} = \underset{\mathbf{D}\in\mathcal{M}_2}{\operatorname{argmin}} F(\mathbf{A}^{k+1}, \mathbf{B}^{k+1}, \mathbf{C}^{k+1}, \mathbf{D})$

End while

Output: $(\mathbf{A}^{k+1}, \mathbf{B}^{k+1}, \mathbf{C}^{k+1}, \mathbf{D}^{k+1})$

steps, which is described in Algorithm 1. The minimization for $\mathbf{A}$ and $\mathbf{C}$ will be presented in detail, and the algorithms work similarly for $\mathbf{B}$ and $\mathbf{D}$.

3.1 Updating A

Consider the case that $\mathbf{B}^k$, $\mathbf{C}^k$, $\mathbf{D}^k$ are determined, the minimization for $\mathbf{A}$ becomes

$$\min_{\mathbf{A}\in\mathcal{S}_1} \frac{1}{2}\|\mathbf{X}\mathbf{A} - \mathbf{C}^k\|_F^2. \tag{10}$$

It is a projection onto the low-rank constraint $\mathcal{S}_1$. Inspired by [18, 19], it can be solved by a two-stage method: 1) Compute the gradient descent at the point $\mathbf{A}^k$ with step-size $\eta > 0$; 2) Apply hard thresholding of the associated singular vector to obtain its approximate solution. Especially, the hard thresholding for matrix cases is defined as follows.

Definition 3.1 *For a given matrix $\mathbf{Z}$, the singular value decomposition (SVD) is given by*

$$\mathbf{Z} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \tag{11}$$

where $\mathbf{\Sigma} = diag(\sigma_1, \ldots, \sigma_n)$ and $\sigma_1 \geq \ldots \geq \sigma_n$ are positive. Moreover, $\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices. Therefore, for low-rank level r, the matrix hard thresholding operator $\mathcal{H}_r$ can be defined as

$$\mathcal{H}_r(\mathbf{Z}) := \mathbf{U}\mathcal{H}_r(\mathbf{\Sigma})\mathbf{V}^T, \tag{12}$$

where $\mathcal{H}_r(\mathbf{\Sigma}) = diag(\sigma_1, \ldots, \sigma_r)$.

From the above definition, the matrix hard thresholding can be regarded as a truncation of the singular values with the top r entries preserved. In some sense, this matrix hard thresholding operator is a straightforward extension of the hard thresholding operator for vector cases. The detailed description for solving $\mathbf{A}$-subproblem is presented in Algorithm 2, which involves a gradient descent method and hard singular value thresholding.

3.2 Updating C

For the case that $\mathbf{A}^{k+1}$, $\mathbf{B}^{k+1}$, $\mathbf{D}^k$ are chosen, the minimization with respect to $\mathbf{C}$ can be obtained by solving

$$\min_{\mathbf{C}\in\mathcal{M}_1} \frac{1}{2}\|\mathbf{C} - (\mathbf{X}\mathbf{A}^{k+1} + \mathbf{D}^k/\beta_1)\|_F^2. \tag{13}$$

Even though the objective is convex, it is not easy to be estimated because of the manifold constraint $\mathcal{M}_1$. In order to

Algorithm 2 Solution of $\mathbf{A}$-subproblem

Input: Given data $\mathbf{X}$, $\mathbf{C}$, parameter $\eta \in (0, 1)$

Initialize: $\mathbf{A}^0 = \mathbf{0}$

While not converged **do**

1: Compute $\bar{\mathbf{A}}^{k+1} = \mathbf{A}^k - \eta\mathbf{X}^T(\mathbf{X}\mathbf{A}^k - \mathbf{C})$

2: Do $\mathbf{A}^{k+1} = \mathcal{H}_{r_1}(\bar{\mathbf{A}}^{k+1})$ with top r_1 singular entries preserved

End while

Output: $\mathbf{A}^{k+1}$

Algorithm 3 Solution of $\mathbf{C}$-subproblem

Input: Given data $\mathbf{X}$, $\mathbf{A}$, $\mathbf{D}$

Initialize: $\mathbf{C}^0 = \mathbf{0}$

While not converged **do**

1: Compute SVD of $\mathbf{X}\mathbf{A}^{k+1} + \mathbf{D}^k/\beta_1 = \bar{\mathbf{U}}\bar{\mathbf{\Sigma}}\bar{\mathbf{V}}^T$

2: Set $\mathbf{C}^{k+1} = \bar{\mathbf{U}}\bar{\mathbf{V}}^T$

End while

Output: $\mathbf{C}^{k+1}$

analyze this problem, the matrix norm is first expanded as

$$\begin{aligned}&\frac{1}{2}\mathrm{Tr}(\mathbf{C}^T\mathbf{C}) - \mathrm{Tr}(\mathbf{C}^T(\mathbf{X}\mathbf{A}^{k+1} + \mathbf{D}^k/\beta_1))\\ &+ \frac{1}{2}\mathrm{Tr}((\mathbf{X}\mathbf{A}^{k+1} + \mathbf{D}^k/\beta_1)^T(\mathbf{X}\mathbf{A}^{k+1} + \mathbf{D}^k/\beta_1)).\end{aligned} \tag{14}$$

Since the first term is constant and the last term is not relevant to $\mathbf{C}$, only the middle term needs to be maximized. Let the SVD of $\mathbf{X}\mathbf{A}^{k+1}+\mathbf{D}^k/\beta_1$ be $\bar{\mathbf{U}}\bar{\mathbf{\Sigma}}\bar{\mathbf{V}}^T$. According to [20], the optimal solution can be achieved by $\mathbf{C}^{k+1} = \bar{\mathbf{U}}\bar{\mathbf{V}}^T$, and the scheme is given in Algorithm 3.

4 Simulations and Applications

This section illustrates the advantages of RCCCA over the state-of-the-art CCA-based methods, i.e., CCA [5], SCCA [11] and LRCCA [15], on a simulated example and the Tennessee Eastman (TE) process. Some extensions of CCA, such as kernel CCA [21], deep CCA [22], multimode CCA [8], distributed CCA [6], are not compared because the current work only focuses on the sparse variants and aims to verify the necessity of rank constrained optimization.

4.1 Simulated Example

This section tests on a simulation numerical example, including input variables $\mathbf{x} = [x_1\ x_2]^T$ and output variables $\mathbf{y} = [y_1\ y_2\ y_3\ y_4]^T$, which are generated from

$$[\mathbf{x}\ \mathbf{y}]^T = N(0, \sigma) + [\mathbf{e_1}\ \mathbf{e_2}]^T, \tag{15}$$

where

$$\sigma = \begin{bmatrix} 1 & 0 & 0.9 & 0.5 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0.9 & 0 & 1 & 0.4 & 0 & 0 \\ 0.5 & 0 & 0.4 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \tag{16}$$

and $\mathbf{e_1}, \mathbf{e_2}$ represent the Gaussian distributed process noises. In this study, r_1, r_2 are chosen as 2. For offline training, a total of 200 samples under normal operating conditions are generated. Two faulty data are constructed for online testing:

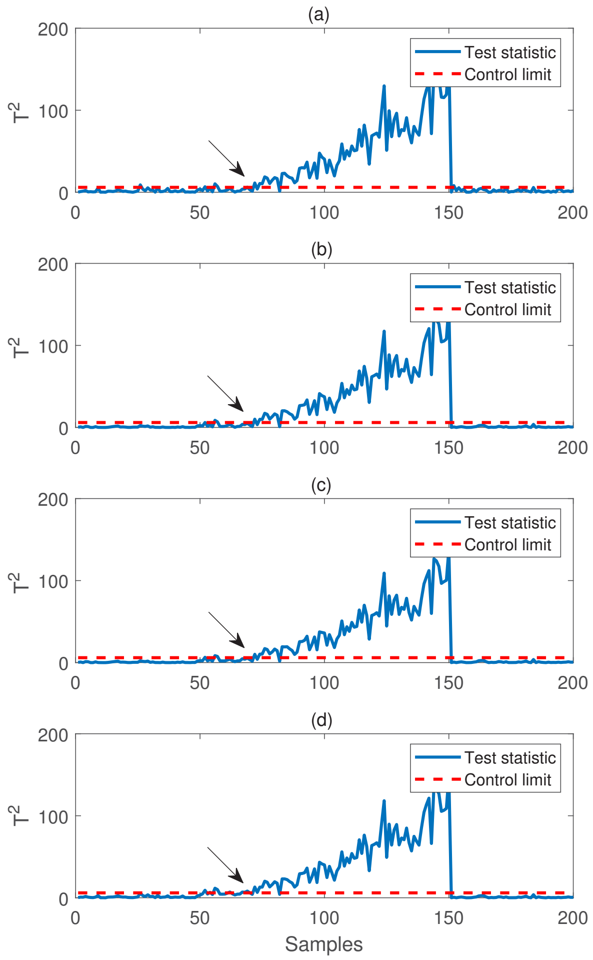

Fig. 2: Monitoring results for Type I: (a) CCA, (b) SCCA, (c) LRCCA, (d) RCCCA.

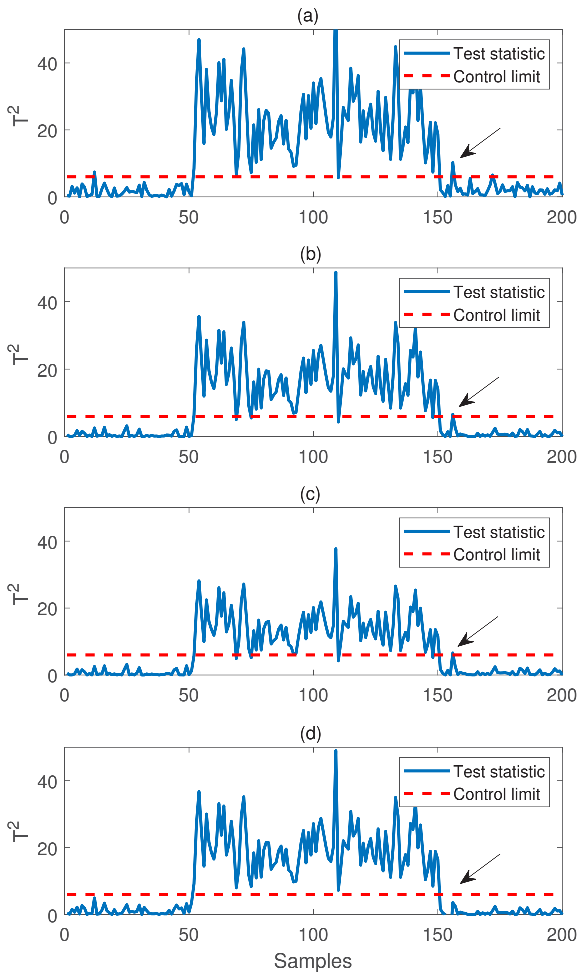

Fig. 3: Monitoring results for Type II: (a) CCA, (b) SCCA, (c) LRCCA, (d) RCCCA.

1) Type I: A ramp change of $0.05 \times (t - 50)$ is introduced into x_1 from the 51st sample to the 150th sample;
2) Type II: A step change of 4.5 is added to x_2 from the 51st sample to the 150th sample.

The first fault simulates a ramp change of x_1, which is related to the coupled system $\mathbf{y}$. The monitoring results are given in Fig. 2. It can be observed that even though a fault is introduced in the 51st sample, all four methods fail to detect it immediately. After the 80th sample, the test statistic becomes sensitive. For 70-80 samples, CCA, SCCA and LRCCA can not detect the fault. However, the proposed RCCCA can recognize this fault and raise an alarm successfully, which illustrates the superiority of the proposed method.

The second fault occurs in x_2, which is unrelated to the coupled system $\mathbf{y}$. Fig. 3 shows the monitoring results. It is indicated that all the four methods perform competitively with similar detection alarms. However, for CCA, SCCA and LRCCA, there exist some false alarms both before the 50th sample and after the 150th sample. By embedding prior information, false alarms are thus eliminated for the proposed RCCCA.

4.2 Case Study on the TE Process

The Tennessee Eastman (TE) process is a famous benchmark dataset for validating different FD techniques. The descriptions of variables and faults are provided in [23]. In this study, 11 manipulated variables are chosen as $\mathbf{X}$ and 22 measured variables are selected as $\mathbf{Y}$. Moreover, the fault-free data are used for offline training, and the 21 faults data are applied for online testing. According to [3], IDV (3), IDV (9), IDV (15), and IDV (21) are relatively hard for detecting by data-driven FD methods, thus omitted here. Throughout the experiment, r_1, r_2 are tuned by 10-fold cross validation.

To appreciate the comparable performance achieved by RCCCA, the monitoring results are illustrated via two typical faults, i.e., IDV(4) and IDV(10). Fault IDV(4) involves a step-change in the reactor cooling-water inlet temperature, and Fig. 4 presents the monitoring performance. It can be seen that all four methods can obtain satisfactory monitoring results for this fault because the fault magnitude is large and easy to be detected. However, for false alarms, RCCA has fewer numbers that violated the control limit (highlighted by the arrows), which illustrates that the proposed RCCA possesses superiority to CCA, SCCA and LRCCA.

Fault IDV(10) is a random variation in the feed C temper-

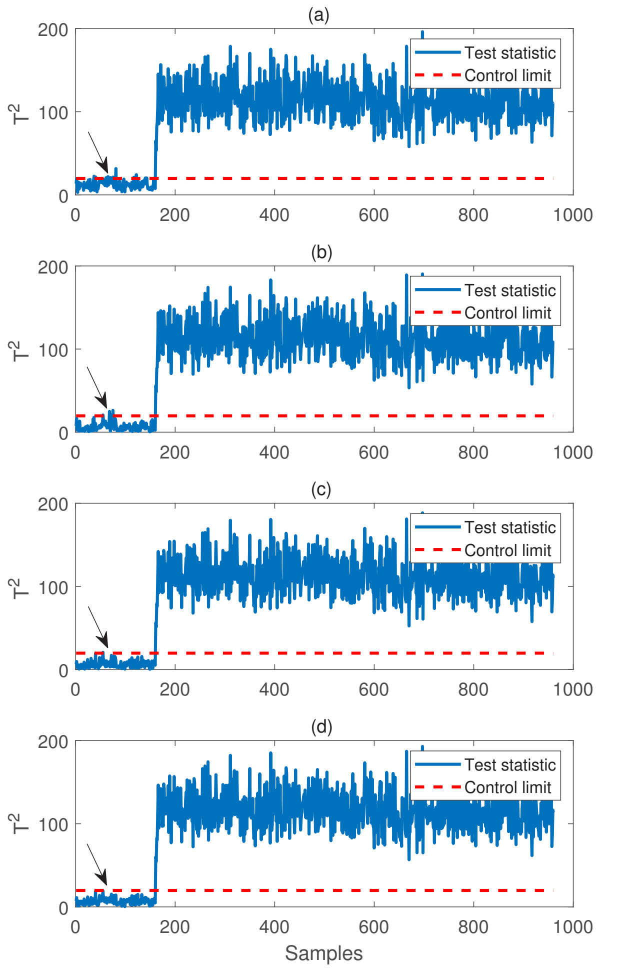

Fig. 4: Monitoring results for fault IDV(4): (a) CCA, (b) SCCA, (c) LRCCA, (d) RCCCA.

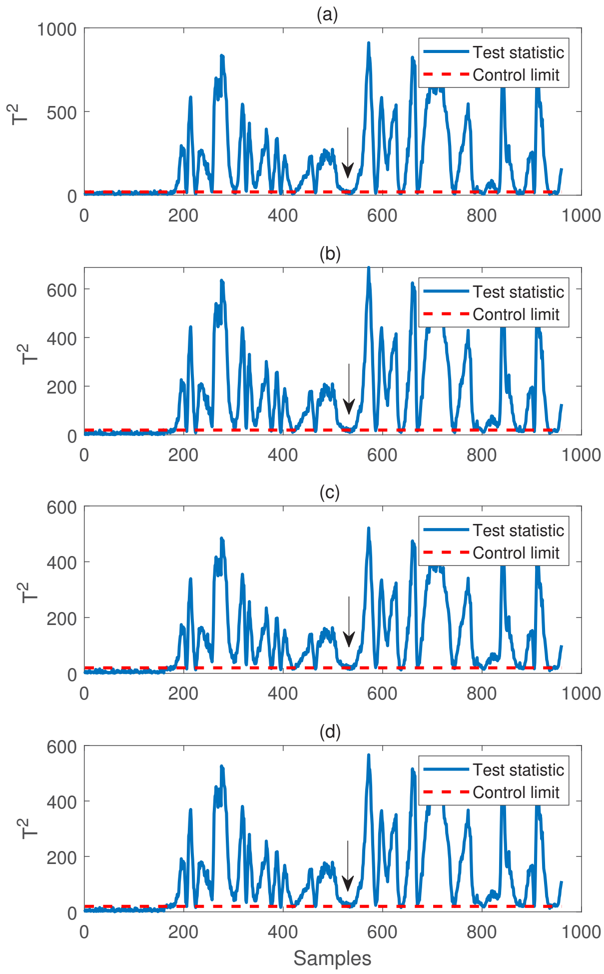

Fig. 5: Monitoring results for fault IDV(10): (a) CCA, (b) SCCA, (c) LRCCA, (d) RCCCA.

ature in stream 4. The variation in the feed C temperature causes a change in the conditions of the stripper and then the condenser. Fig. 5 provides the monitoring T^2 test statistic for all four methods. It shows that RCCA-based FD outperforms the other three methods, which indicates the efficiency of the proposed rank constrained monitoring scheme.

In order to evaluate different FD methods, two common measurement indices are adopted, i.e., fault detection rate (FDR) and false alarm rate (FAR). Table 1 reports the monitoring performance for the selected faults in the TE process. In particular, the best results are marked in bold for convenient of comparing. For all the selected faults, the sparse or low-rank variants (i.e., SCCA, LRCCA, RCCCA) offer better performance than CCA. It is convinced that prior information is helpful for FD. Note that LRCCA can not provide consistent superiority to SCCA because the same parameters for low-rank regularization terms are not appropriate. No matter for easy faults or difficult faults, the proposed RCCCA always performs better than or as good as CCA, SCCA and LRCCA. This considerate fault detection performance can be attributed to the introduction of low-rank embedding, which preserves the structured correlation relationship between process variables.

5 Conclusions

In this paper, a novel FD architecture that incorporates rank constrained optimization and CCA is proposed. Moreover, an efficient optimization algorithm for solving RCCCA is provided. The feasibility and efficiency of the proposed monitoring scheme are demonstrated by case studies on a numerical example and the benchmark TE process. In comparison with some existing state-of-the-art methods, the RCCCA involves the global information of a process and thus achieves superior performance in terms of FDR and FAR. For these reasons, it is believed that the proposed rank constrained CCA is encouraging for FD.

Although RCCCA achieves satisfactory performance, several interesting and important questions remain open. On one hand, the convergence and complexity of the proposed algorithm should be discussed theoretically. On the other hand, the fault isolation and diagnosis of the proposed detection strategy need to be further investigated.

References

[1] S. X. Ding, *Data-driven Design of Fault Diagnosis and Fault-Tolerant Control Systems*. Springer, 2014.

Table 1: Detection results in terms of FDR and FAR.

Fault No.	CCA		SCCA		LRCCA		RCCCA	
	FDR	FAR	FDR	FAR	FDR	FAR	FDR	FAR
IDV(1)	99.75%	**0.00%**	99.75%	**0.00%**	**99.88%**	**0.00%**	**99.88%**	**0.00%**
IDV(2)	96.50%	0.63%	98.38%	0.63%	98.38%	**0.00%**	**99.47%**	**0.00%**
IDV(4)	**100%**	1.88%	**100%**	0.63%	**100%**	0.63%	**100%**	**0.00%**
IDV(5)	**100%**	3.75%	**100%**	2.50%	**100%**	2.50%	**100%**	**0.63%**
IDV(6)	**100%**	4.38%	**100%**	3.75%	**100%**	3.75%	**100%**	**2.50%**
IDV(7)	**100%**	3.75%	**100%**	2.50%	**100%**	1.88%	**100%**	**0.63%**
IDV(8)	96.50%	1.88%	98.50%	**0.00%**	98.88%	0.63%	**99.25%**	**0.00%**
IDV(10)	86.88%	1.25%	88.88%	**0.00%**	89.38%	**0.00%**	**92.63%**	**0.00%**
IDV(11)	76.50%	0.63%	79.13%	**0.00%**	82.75%	0.63%	**84.38%**	**0.00%**
IDV(12)	99.00%	1.25%	99.50%	**0.00%**	99.50%	0.63%	**99.75%**	**0.00%**
IDV(13)	95.75%	0.63%	96.25%	**0.00%**	**96.50%**	**0.00%**	**96.50%**	**0.00%**
IDV(14)	**100%**	1.88%	**100%**	1.25%	**100%**	0.63%	**100%**	**0.00%**
IDV(16)	93.00%	7.50%	95.47%	3.75%	95.25%	2.50%	**96.88%**	**1.25%**
IDV(17)	94.13%	3.13%	94.25%	1.88%	94.63%	0.63%	**95.75%**	**0.00%**
IDV(18)	90.88%	1.88%	95.50%	0.63%	94.88%	0.63%	**95.13%**	**0.00%**
IDV(19)	92.00%	1.25%	93.88%	1.25%	93.50%	**0.63%**	**94.25%**	**0.63%**
IDV(20)	86.88%	0.63%	87.75%	**0.00%**	87.75%	0.63%	**90.38%**	**0.00%**

[2] Y. Liu, J. Zeng, L. Xie, S. Luo, and H. Su, "Structured joint sparse principal component analysis for fault detection and isolation," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2721–2731, 2018.

[3] L. H. Chiang, E. L. Russell, and R. D. Braatz, *Fault Detection and Diagnosis in Industrial Systems*. Springer Science & Business Media, 2000.

[4] S. Yin, S. X. Ding, A. Haghani, H. Hao, and P. Zhang, "A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process," *Journal of Process Control*, vol. 22, no. 9, pp. 1567–1581, 2012.

[5] Z. Chen, S. X. Ding, T. Peng, C. Yang, and W. Gui, "Fault detection for non-Gaussian processes using generalized canonical correlation analysis and randomized algorithms," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1559–1567, 2017.

[6] Z. Chen, Y. Cao, S. X. Ding, K. Zhang, T. Koenings, T. Peng, C. Yang, and W. Gui, "A distributed canonical correlation analysis-based fault detection method for plant-wide process monitoring," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2710–2720, 2019.

[7] Z. Chen, C. Liu, S. Ding, T. Peng, C. Yang, W. Gui, and Y. Shardt, "A just-in-time-learning aided canonical correlation analysis method for multimode process monitoring and fault detection," *IEEE Transactions on Industrial Electronics*, doi: 10.1109/TIE.2020.2989708.

[8] Q. Jiang and X. Yan, "Multimode process monitoring using variational bayesian inference and canonical correlation analysis," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 4, pp. 1814–1824, 2019.

[9] B. Song, H. Shi, S. Tan, and Y. Tao, "Multi-subspace orthogonal canonical correlation analysis for quality related plant wide process monitoring," *IEEE Transactions on Industrial Informatics*, doi: 10.1109/TII.2020.3015034.

[10] X. Yang, L. Weifeng, W. Liu, and D. Tao, "A survey on canonical correlation analysis," *IEEE Transactions on Knowledge and Data Engineering*, doi: 10.1109/TKDE.2019.2958342.

[11] X. Xiu, Y. Yang, L. Kong, and W. Liu, "Data-driven process monitoring using structured joint sparse canonical correlation analysis," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 68, no. 1, pp. 361–365, 2021.

[12] C. Gao, Z. Ma, and H. H. Zhou, "Sparse CCA: Adaptive estimation and computational barriers," *The Annals of Statistics*, vol. 45, no. 5, pp. 2074–2101, 2017.

[13] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

[14] X. Li, N. Xiu, and S. Zhou, "Matrix optimization over low-rank spectral sets: Stationary points and local and global minimizers," *Journal of Optimization Theory and Applications*, vol. 184, no. 3, pp. 895–930, 2020.

[15] X. Jin, Y. Su, L. Zou, Y. Wang, P. Jing, and Z. J. Wang, "Sparsity-based image inpainting detection via canonical correlation analysis with low-rank constraints," *IEEE Access*, vol. 6, pp. 49 967–49 978, 2018.

[16] Z. Chen, S. X. Ding, K. Zhang, Z. Li, and Z. Hu, "Canonical correlation analysis-based fault detection methods with application to alumina evaporation process," *Control Engineering Practice*, vol. 46, pp. 51–58, 2016.

[17] S. X. Ding, *Model-Based Fault Diagnosis TechniquesDesign Schemes, Algorithms and Tools*. Berlin: Springer-Verlag, 2008.

[18] X. Yuan, P. Li, and T. Zhang, "Gradient hard thresholding pursuit," *Journal of Machine Learning Research*, vol. 18, no. 1, pp. 6027–6069, 2017.

[19] J. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.

[20] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *Journal of Computational and Graphical Statistics*, vol. 15, no. 2, pp. 265–286, 2006.

[21] W. Zheng, X. Zhou, C. Zou, and L. Zhao, "Facial expression recognition using kernel canonical correlation analysis (KCCA)," *IEEE Transactions on Neural Networks*, vol. 17, no. 1, pp. 233–238, 2006.

[22] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, "Deep canonical correlation analysis," *International Conference on Machine Learning*, pp. 1247–1255, 2013.

[23] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," *Computers & Chemical Engineering*, vol. 17, no. 3, pp. 245–255, 1993.