# Data-Driven Process Monitoring Using Structured Joint Sparse Canonical Correlation Analysis

Xianchao Xiu, Ying Yang, *Senior Member, IEEE*, Lingchen Kong, and Wanquan Liu *Senior Member, IEEE*

***Abstract*—In order to improve the performance of canonical correlation analysis (CCA) based methods for process monitoring, this brief proposes a novel process monitoring approach using the structured joint sparse canonical correlation analysis (SJSCCA). Technically, the graph Laplacian could incorporate structured variable correlation information and the joint sparsity could discard useless variables. The developed two-stage alternating direction method of multipliers is shown to be very efficient because each derived subproblem has a closed-form solution or can be solved by fast solvers. In order to detect abnormal situations, $T^2$ test statistic is adopted. The validity of SJSCCA is illustrated by the benchmark Tennessee Eastman process. The achieved results show that the proposed SJSCCA is able to improve the monitoring performance significantly in comparison with the existing state-of-the-art CCA-based methods.**

***Index Terms*—Process monitoring, canonical correlation analysis, graph Laplacian, joint sparsity, alternating direction method of multipliers (ADMM).**

## I. Introduction

PROCESS monitoring is becoming increasingly important to enhance industrial plant safety, reduce the production cost as well as improve the product quality. Due to the rapid development in the fields of data storage and sensor technologies, data-driven monitoring methods have attracted much attention in the last decades [2]. A number of techniques associated with multivariate statistic process control have been proposed, such as principal component analysis (PCA) [11], partial least square (PLS) [14], canonical correlation analysis (CCA) [1], [6], [7], and others [5], [13]. Different from PCA and PLS, CCA considers the relationship between input and output variables in order to make use of different but complementary information. Therefore, CCA provides a better understanding of process data and has the potential to improve the monitoring performance of fault detection.

Manuscript received March 30, 2020; accepted April 12, 2020. Date of publication April 15, 2020; date of current version December 21, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61633001, Grant 11671029, and Grant U1713223. This brief was recommended by Associate Editor X. Wang. *(Corresponding author: Ying Yang.)*

Xianchao Xiu and Ying Yang are with the Department of Mechanics and Engineering Science, Peking University, Beijing 100871, China (e-mail: xcxiu@bjtu.edu.cn; yy@pku.edu.cn).

Lingchen Kong is with the Department of Applied Mathematics, Beijing Jiaotong University, Beijing 100044, China (e-mail: lchkong@bjtu.edu.cn).

Wanquan Liu is with the Department of Computing, Curtin University, Perth, WA 6102, Australia (e-mail: w.liu@curtin.edu.au).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSII.2020.2988054

The standard CCA often works extremely slow or fails to extract meaningful features in high-dimensional data analysis. For example, when there exists collinearity between variables, the covariance matrix becomes ill-conditioned; when there exist a large number of variables that are not informative to faults, the interpretation becomes meaningless. To handle the limitations of CCA, sparse learning is an effective way [8], [17]. By incorporating sparse constraints onto variables, [15] extended it to sparse CCA (SCCA). Although, SCCA has been employed in the fields such as medicine [7] and image processing [16]. The application of SCCA in process monitoring has only been investigated by [12]. It is demonstrated that SCCA can improve the interpretability of sparse canonical vectors, discover important relationships among process variables, and thus achieve better fault detection performance. This convinces us to believe that SCCA shares both the advantages of sparse methods over dense methods, and the merits of CCA over other dimensionality reduction methods.

If the process data in the high-dimensional space are actually located on the union of some linear subspaces, using SCCA is easy to capture the low-dimensional structure. However, in practice, the assumption cannot be guaranteed [10]. In this brief, a novel data-driven process monitoring using structured joint sparse canonical correlation analysis (SJSCCA) is proposed. The objective function involves two regularization terms: the graph Laplacian and the joint sparsity. By imposing the graph Laplacian, SJSCCA not only learns the cause-effect relationship between process variables, but also discovers conditional independent process variables between individual operation units. Moreover, the joint sparsity could enforce that the variables corresponding to all zero rows will not be included in the projected subspace. This reduces the effect of outliers and discards useless features.

Compared to previous work, the main contributions of this brief are summarized in the following three aspects:

1) Instead of considering CCA and SCCA, a new SJSCCA is developed to enhance the performance of process monitoring. To the best of our knowledge, it is the first time to integrate the graph Laplacian and joint sparsity into a CCA framework.
2) An efficient optimization algorithm using two-stage alternating direction method of multipliers (ADMM) is developed to solve the proposed SJSCCA.

3) A comparative study on the Tennessee Eastman (TE) process demonstrates the superior monitoring performance of SJSCCA over other CCA-based methods.

The remainder of this brief is organized as follows. Section II introduces the proposed SJSCCA and optimization algorithm. Section III develops an online process monitoring strategy using the SJSCCA. Section IV illustrates its advantages by simulation examples on the TE process. Finally, Section V concludes this brief.

## II. Structured Joint Sparse CCA

### A. Model Structure

For two data matrices $\mathbf{X} \in \mathbb{R}^{n\times p}$ and $\mathbf{Y} \in \mathbb{R}^{n\times q}$, SCCA is formulated as follows

$$\begin{aligned} \min_{\mathbf{A},\mathbf{B}} \ & \frac{1}{2}\|\mathbf{XA}-\mathbf{YB}\|_F^2 + \lambda_1\|\mathbf{A}\|_1 + \lambda_2\|\mathbf{B}\|_1 \\ \text{s.t.} \ & \mathbf{A}^T\mathbf{X}^T\mathbf{XA} = \mathbf{I}_k, \ \mathbf{B}^T\mathbf{Y}^T\mathbf{YB} = \mathbf{I}_k, \end{aligned} \tag{1}$$

where $\|\cdot\|_1$ is the sum of absolute values of all entries, $\lambda_1, \lambda_2$ are the regularization parameters, and $k$ is the number of principle components. It is verified that, compared with CCA, SCCA is more powerful to identify subsets of features [7].

In most cases, prior structural knowledge is crucial for data analysis, which facilitates model's interpretability. In order to incorporate variable structure, the joint sparse CCA (JSCCA), as an extension of SCCA, is given by

$$\begin{aligned} \min_{\mathbf{A},\mathbf{B}} \ & \frac{1}{2}\|\mathbf{XA}-\mathbf{YB}\|_F^2 + \lambda_1\|\mathbf{A}\|_{2,1} + \lambda_2\|\mathbf{B}\|_{2,1} \\ \text{s.t.} \ & \mathbf{A}^T\mathbf{X}^T\mathbf{XA} = \mathbf{I}_k, \ \mathbf{B}^T\mathbf{Y}^T\mathbf{YB} = \mathbf{I}_k, \end{aligned} \tag{2}$$

where $\|\cdot\|_{2,1}$ denotes the sum of $\ell_2$ norm of all rows. JSCCA is able to achieve row-wise sparsity and make process monitoring much more stable. Even though JSCCA has been considered before [16], no one applies it to process monitoring.

From manifold point of view, it is reasonable to assume that if two variables are close in the original space, then the representations of these two variables in the projected space are also close. As a result, the graph Laplacian [11] can be applied to preserve local geometric and topological structures. This observation encourages us to propose the following structured joint sparse CCA (SJSCCA):

$$\begin{aligned} \min_{\mathbf{A},\mathbf{B}} \ & \frac{1}{2}\|\mathbf{XA}-\mathbf{YB}\|_F^2 + \lambda_1\|\mathbf{A}\|_{2,1} + \lambda_2\|\mathbf{B}\|_{2,1} \\ & + \mu_1\mathrm{tr}(\mathbf{A}^T\mathbf{L}_1\mathbf{A}) + \mu_2\mathrm{tr}(\mathbf{B}^T\mathbf{L}_2\mathbf{B}) \\ \text{s.t.} \ & \mathbf{A}^T\mathbf{X}^T\mathbf{XA} = \mathbf{I}_k, \ \mathbf{B}^T\mathbf{Y}^T\mathbf{YB} = \mathbf{I}_k, \end{aligned} \tag{3}$$

where $\mathbf{L}_1$ and $\mathbf{L}_2$ are graph Laplacian matrices corresponding to $\mathbf{X}$ and $\mathbf{Y}$, respectively.

Note that our proposed SJSCCA is different from [6], which integrates a variational Bayesian Gaussian mixture model with CCA. SJSCCA involves two types of regularization terms: the $\ell_{2,1}$ norm enforcing row-wise sparsity and the graph Laplacian taking the variable correlation information into consideration. Since the optimization problem (3) is not easy to solve, an efficient optimization algorithm with closed-form solutions will be developed in the next subsection.

**Algorithm 1** A Two-Stage Method for Solving (3)

**Input:** Datasets $\mathbf{X}$, $\mathbf{Y}$, graph Laplacian $\mathbf{L}_1$, $\mathbf{L}_2$, parameters $\lambda_1, \lambda_2, \mu_1, \mu_2 > 0$.
**Initialize:** $\mathbf{B}^0$.
**While** not converged **do**
1: Fix $\mathbf{B}^k$, compute $\mathbf{A}^{k+1}$;
2: Fix $\mathbf{A}^{k+1}$, compute $\mathbf{B}^{k+1}$;
**End While**
**Output:** $\mathbf{A}$, $\mathbf{B}$.

### B. Optimization Algorithm

If $\mathbf{B}$ is fixed, the resulting optimization problem is convex respect to $\mathbf{A}$, and in turn, if $\mathbf{A}$ is fixed, the optimization problem is also convex respect to $\mathbf{B}$. As a result, the minimization problem can be divided into two stages:

- Fix $\mathbf{B}$, problem (3) becomes

$$\begin{aligned} \min_{\mathbf{A}} \ & \frac{1}{2}\|\mathbf{XA}-\mathbf{YB}\|_F^2 + \lambda_1\|\mathbf{A}\|_{2,1} + \mu_1\mathrm{tr}(\mathbf{A}^T\mathbf{L}_1\mathbf{A}) \\ \text{s.t.} \ & \mathbf{A}^T\mathbf{X}^T\mathbf{XA} = \mathbf{I}_k. \end{aligned} \tag{4}$$

- Fix $\mathbf{A}$, problem (3) becomes

$$\begin{aligned} \min_{\mathbf{B}} \ & \frac{1}{2}\|\mathbf{XA}-\mathbf{YB}\|_F^2 + \lambda_2\|\mathbf{B}\|_{2,1} + \mu_2\mathrm{tr}(\mathbf{B}^T\mathbf{L}_2\mathbf{B}) \\ \text{s.t.} \ & \mathbf{B}^T\mathbf{Y}^T\mathbf{YB} = \mathbf{I}_k. \end{aligned} \tag{5}$$

This procedure can be summarized in Algorithm 1. The order of the updates can be reversed, i.e., it is possible first to optimize $\mathbf{B}$, followed by an optimization of $\mathbf{A}$.

In the next, it is demonstrated how to solve the minimization problem (4), and the algorithm works similarly for (5). To make use of its structure, two auxiliary variables $\mathbf{C}$, $\mathbf{D}$ are introduced. Thus, (4) can be reformulated as an equivalent optimization problem

$$\begin{aligned} \min_{\mathbf{A},\mathbf{C},\mathbf{D}} \ & \frac{1}{2}\|\mathbf{C}-\mathbf{YB}\|_F^2 + \lambda_1\|\mathbf{D}\|_{2,1} + \mu_1\mathrm{tr}(\mathbf{A}^T\mathbf{L}_1\mathbf{A}) \\ \text{s.t.} \ & \mathbf{C}^T\mathbf{C} = \mathbf{I}_r, \ \mathbf{XA} = \mathbf{C}, \ \mathbf{A} = \mathbf{D}. \end{aligned} \tag{6}$$

The augmented Lagrangian for (6) is then defined as

$$\begin{aligned} \mathcal{L}_\beta(\mathbf{C},\mathbf{D},\mathbf{A};\mathbf{W}_1,\mathbf{W}_2) = & \frac{1}{2}\|\mathbf{C}-\mathbf{YB}\|_F^2 + \lambda_1\|\mathbf{D}\|_{2,1} \\ & + \mu_1\mathrm{tr}(\mathbf{A}^T\mathbf{L}_1\mathbf{A}) - \langle\mathbf{W}_1, \mathbf{XA}-\mathbf{C}\rangle + \frac{\beta}{2}\|\mathbf{XA}-\mathbf{C}\|_F^2 \\ & - \langle\mathbf{W}_2, \mathbf{A}-\mathbf{D}\rangle + \frac{\beta}{2}\|\mathbf{A}-\mathbf{D}\|_F^2, \end{aligned}$$

where $\mathbf{W}_1, \mathbf{W}_2$ are the Lagrange multipliers, and $\beta > 0$ is the penalty parameter. The above expression requires minimizations over $\mathcal{M} = \{\mathbf{C} \mid \mathbf{C}^T\mathbf{C} = \mathbf{I}_k\}$.

It can be solved by alternatively minimizing one variable with the others fixed, which is the so-called ADMM [4]. Note that the subproblem of $\mathbf{C}$ is nonconvex, the subproblem of $\mathbf{D}$ is convex but nonsmooth, the subproblem of $\mathbf{A}$ is convex and smooth, thus the variables are updated as

$$\mathbf{C} \to \mathbf{D} \to \mathbf{A} \to \mathbf{W}_1 \to \mathbf{W}_2.$$

Now, it will illustrate how to solve the resulting subproblems.

**Algorithm 2** ADMM for Solving (6)

**Input:** Datasets $\mathbf{X}$, $\mathbf{Y}$, graph Laplacian $\mathbf{L}_1$, parameters $\lambda_1, \mu_1 > 0$, $\beta > 0$ and $\mathbf{B}$.
**Initialize:** Multipliers $\mathbf{W}_1^0 = \mathbf{W}_2^0 = 0$, and $\mathbf{A}^0$.
**While** not converged **do**
1: Compute $\mathbf{C}^{k+1}$ via (7);
2: Compute $\mathbf{D}^{k+1}$ via (8);
3: Compute $\mathbf{A}^{k+1}$ via (9);
4: Update $\mathbf{W}_1^{k+1} = \mathbf{W}_1^k - \beta(\mathbf{X}\mathbf{A}^{k+1} - \mathbf{C}^{k+1})$;
5: Update $\mathbf{W}_2^{k+1} = \mathbf{W}_2^k - \beta(\mathbf{A}^{k+1} - \mathbf{D}^{k+1})$;
**End While**
**Output:** $\mathbf{A}$.

- For $\mathbf{C}$-subproblem, after trivial manipulation, it can be written as
$$\min_{\mathbf{C}\in\mathcal{M}} \frac{1}{2}\|\mathbf{C} - \mathbf{Y}\mathbf{B}\|_F^2 + \frac{\beta}{2}\|\mathbf{X}\mathbf{A}^k - \mathbf{C} - \mathbf{W}_1^k/\beta\|_F^2.$$
According to [18], it is exactly a reduced rank Procrustes rotation problem. Let
$$\mathbf{C}^{k+\frac{1}{2}} = \frac{1}{1+\beta}\Big(\mathbf{Y}\mathbf{B} - \beta\mathbf{X}\mathbf{A}^k + \beta\mathbf{W}_1^k\Big),$$
and the SVD of $\mathbf{C}^{k+\frac{1}{2}}$ is $\mathbf{U}\Sigma\mathbf{V}^T$. Thus the solution is
$$\mathbf{C}^{k+1} = \mathbf{U}\mathbf{V}^T. \tag{7}$$
- For $\mathbf{D}$-subproblem, it can be simplified to
$$\min_{\mathbf{D}} \lambda_1\|\mathbf{D}\|_{2,1} + \frac{\beta}{2}\|\mathbf{A}^k - \mathbf{D} - \mathbf{W}_2^k/\beta\|_F^2.$$
Denote $\mathbf{D}^{k+\frac{1}{2}} = \mathbf{A}^k - \mathbf{W}_2^k$. Then the solution is given by
$$\mathbf{D}_i^{k+1} = \frac{\mathbf{D}_i^{k+\frac{1}{2}}}{\|\mathbf{D}_i^{k+\frac{1}{2}}\|_2} \circ \max\left\{0, \|\mathbf{D}_i^{k+\frac{1}{2}}\|_2 - \lambda_1/\beta\right\}, \tag{8}$$
where all operations are done componentwise. See [11] for illustrations.
- For $\mathbf{A}$-subproblem, it can be transformed as
$$\min_{\mathbf{A}} \mu_1 \mathrm{tr}(\mathbf{A}^T\mathbf{L}_1\mathbf{A}) + \frac{\beta}{2}\|\mathbf{X}\mathbf{A} - \mathbf{C}^{k+1} - \mathbf{W}_1^k/\beta\|_F^2 + \frac{\beta}{2}\|\mathbf{A} - \mathbf{D}^{k+1} - \mathbf{W}_2^k/\beta\|_F^2.$$
It is concluded that
$$\mathbf{A}^{k+1} = (2\mu_1\mathbf{L}_1 + \beta\mathbf{I} + \beta\mathbf{X}^T\mathbf{X})^{-1} (\beta\mathbf{C}^{k+1} + \mathbf{W}_1^k + \beta\mathbf{D}^{k+1} + \mathbf{W}_1^k). \tag{9}$$
Notice that the coefficient matrix $2\mu_1\mathbf{L}_1 + \beta\mathbf{I} + \beta\mathbf{X}^T\mathbf{X}$ is nonsingular whenever $\mu_1, \beta > 0$. In algorithms, it can be computed via the Cholesky decomposition or the conjugate gradient method. Moreover, the inverse only needs to be calculated once before iterations.

Under the framework of ADMM, the iterative scheme for solving (6) (equivalently (4)) can be described in Algorithm 2. Similarly, the algorithm for (5) can also be obtained. Updating $\mathbf{A}$ and $\mathbf{B}$ is equivalent to alternating projection of $\mathbf{A}$ and $\mathbf{B}$ onto manifolds $\mathbf{A}^T\mathbf{X}^T\mathbf{X}\mathbf{A} = \mathbf{I}_k$ and $\mathbf{B}^T\mathbf{Y}^T\mathbf{Y}\mathbf{B} = \mathbf{I}_k$. To end this section, the convergence result is established below.

*Theorem 1:* Suppose that $\{(\mathbf{A}^k, \mathbf{B}^k)\}$ is a sequence generated by the above algorithm. Then the sequence converges to a local minimizer of (3).

It follows a similar line of arguments as in [9], thus the detailed proof is omitted here for the sake of continuity.

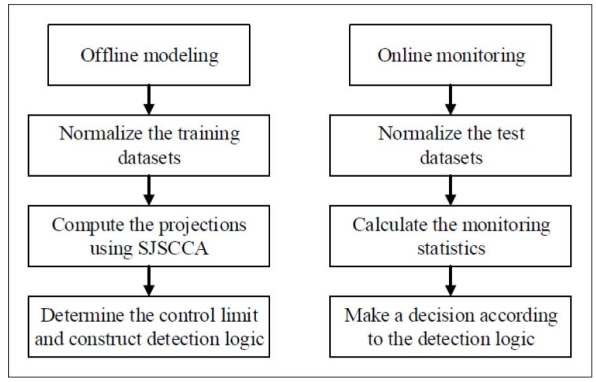


Fig. 1. Process monitoring procedure using SJSCCA.

## III. Process Monitoring Using the SJSCCA

In this section, an online process monitoring procedure is proposed based on the proposed SJSCCA. Let $\mathbf{X}_{train}$, $\mathbf{Y}_{train}$ be the observation data. Normalization is important to eliminate the effects of engineering units and measurement ranges. After conducting normalization, $\mathbf{X}_{train}$, $\mathbf{Y}_{train}$ have zero mean and unit variance. By solving the optimization problem (3), the transformation matrices $\mathbf{A}$ and $\mathbf{B}$ can be obtained.

According to [2], the $T^2$ test statistic is commonly adopted to detect abnormalities. The residual signal is defined as
$$\mathbf{r} = \mathbf{A}^T\mathbf{x}_{test} - \Lambda\mathbf{B}^T\mathbf{y}_{test},$$
where $\Lambda$ is the covariance matrix of $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{x}_{test} \in \mathbb{R}^p$ and $\mathbf{y}_{test} \in \mathbb{R}^q$ are the test sample vectors. Hence, the $T^2$ test statistic can be described as
$$T^2 = \mathbf{r}^T(\mathbf{I} - \Lambda^2)^{-1}\mathbf{r}.$$

Given a level of significance $\alpha$, the upper control limit for $T^2$ test statistic is
$$J_{th,T^2} = \frac{k(n^2-1)}{n(n-p)}F_\alpha(k, n-k),$$
where $F_\alpha$ represents the upper $\alpha$ percentile of $F$-distribution with $k$ and $n-k$ degrees of freedom.

For process monitoring purpose, it is possible to make a decision according to the following detection logic
$$T^2 \le J_{th,T^2} \Rightarrow \text{fault-free, otherwise faulty.}$$

To end this section, the offline modeling and online monitoring procedure are summarized in Fig. 1. When a fault happens, $T^2$ test statistic will exceed the corresponding control limit. After performing SJSCCA, the loading vectors only contain a few nonzero elements, hence it is convenient to find the root cause of the incipient fault.

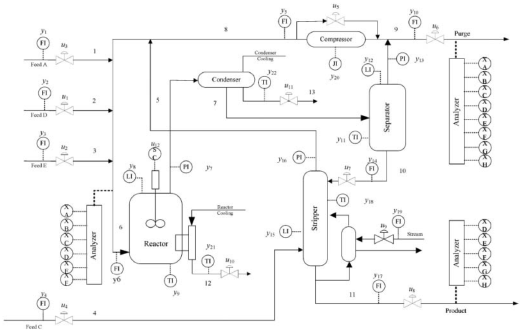


Fig. 2. The TE process and monitoring variables.

TABLE I
SELECTED VARIABLES IN THE TE PROCESS

| No. | Description | No. | Description |
|---|---|---|---|
| 1 | feed A | 18 | stripper temperature |
| 2 | feed D | 19 | stripper steam flow |
| 3 | feed E | 20 | compressor work |
| 4 | total feed | 21 | RCW outlet temperature |
| 5 | recycle flow | 22 | product separator temperature |
| 6 | reactor feed rate | 23 | feed D flow valve |
| 7 | reactor pressure | 24 | feed E flow valve |
| 8 | reactor level | 25 | feed A flow valve |
| 9 | reactor temperature | 26 | total feed flow valve |
| 10 | purge rate | 27 | compressor recycle valve |
| 11 | RCW outlet temperature | 28 | purge valve |
| 12 | product separator level | 29 | separator pot liquid flow valve |
| 13 | product separator pressure | 30 | stripper liquid product flow valve |
| 14 | product separator underflow | 31 | stripper steam valve |
| 15 | stripper level | 32 | RCW flow |
| 16 | stripper pressure | 33 | CCW flow |
| 17 | stripper underflow | | |

## IV. APPLICATION TO THE TE PROCESS

In this section, the benchmark TE process is tested to validate the priority of our proposed SJSCCA over CCA [1], SCCA [12] and JSCCA.

### A. Data Description

The TE process is a well-known process that is widely used to compare various fault detection and isolation techniques. There exist five major units, i.e., reactor, condenser, separator, stripper, compressor, and eight components, including four reactants, two products, a major byproduct, an inert component. See Fig. 2 for a flowchart of the TE process, and a detailed description is referred to [3].

The TE process records a total of 41 measured variables and 11 manipulated variables. In this brief, 22 measured variables and 11 manipulated variables are selected as listed in Table I. Here, RCW stands for reactor cooling water and CCW is an abbreviation of condenser cooling water. The TE process consists of 1 normal dataset and 21 faulty datasets. Each faulty dataset has 960 samples, and a fault is introduced to the process at the 161st sampling time point. Descriptions of the 21 faults are provided in Table II. It is noted that the normal dataset is used for offline modeling, and the faulty datasets are used for online monitoring.

### B. Setup

For the purpose of comparison, all the four CCA-based process monitoring methods are implemented by ourselves. Before normalization, the selected variables are divided into two parts: $y_1$ to $y_{22}$ (denoted as $\mathbf{X}_{train}$) and $u_1$ to $u_{11}$ (denoted as $\mathbf{Y}_{train}$). The control limit is determined with a significance level of 0.05, and the reduced rank $k$ is chosen as 10. For the proposed SJSCCA, the parameters $\lambda_1, \lambda_2, \mu_1, \mu_2$ are determined by 10-fold cross validation. Similar techniques are also applied to SCCA and JSCCA. In addition, the graph Laplacian matrices $\mathbf{L}_1$ and $\mathbf{L}_2$ are determined by setting the number of nearest neighbors to 10. Finally, all of test methods terminate based on the successive differences $\|\mathbf{A}^{k+1} - \mathbf{A}^k\|_F$ and $\|\mathbf{B}^{k+1} - \mathbf{B}^k\|_F$, and the maximum number of iterations is set as 200.

TABLE II
THE TE PROCESS FAULTS

| Fault No. | Description | Type |
|---|---|---|
| 1 | A/C feed ratio | step change |
| 2 | component B | step change |
| 3 | feed D temperature | step change |
| 4 | RCW inlet temperature | step change |
| 5 | CCW inlet temperature | step change |
| 6 | feed A loss | step change |
| 7 | C header pressure loss | step change |
| 8 | feed A−C components | random variation |
| 9 | feed D temperature | random variation |
| 10 | feed C temperature | random variation |
| 11 | RCW inlet temperature | random variation |
| 12 | CCW inlet temperature | random variation |
| 13 | reaction kinetics | slow drift |
| 14 | RCW valve | sticking |
| 15 | CCW valve | sticking |
| 16 | unknown fault | unknown |
| 17 | unknown fault | unknown |
| 18 | unknown fault | unknown |
| 19 | unknown fault | unknown |
| 20 | unknown fault | unknown |
| 21 | unknown fault | constant |

TABLE III
DETECTION RESULTS IN TERMS OF FDR(%) AND FAR(%)

| Fault No. | CCA | | SCCA | | JSCCA | | SJSCCA | |
|---|---|---|---|---|---|---|---|---|
| | FDR | FAR | FDR | FAR | FDR | FAR | FDR | FAR |
| 1 | 99.25 | **0.00** | 99.38 | **0.00** | 99.50 | **0.00** | **99.75** | **0.00** |
| 2 | 98.62 | 0.63 | **99.47** | **0.00** | **99.47** | **0.00** | **99.47** | **0.00** |
| 3 | 33.80 | 2.50 | 36.64 | 1.88 | 38.20 | **0.00** | **41.38** | **0.00** |
| 4 | **100** | 1.88 | **100** | 0.63 | **100** | **0.00** | **100** | **0.00** |
| 5 | 29.63 | 1.88 | 31.00 | 0.63 | 34.50 | **0.00** | **36.62** | **0.00** |
| 6 | 99.88 | 0.63 | 99.90 | **0.00** | **100** | **0.00** | **100** | **0.00** |
| 7 | **100** | 1.88 | **100** | 0.63 | **100** | 0.63 | **100** | **0.00** |
| 8 | 93.25 | 1.88 | 95.00 | 0.63 | 95.26 | **0.00** | **97.85** | **0.00** |
| 9 | 31.20 | 3.13 | 35.25 | 2.50 | 38.50 | **0.63** | **40.87** | **0.63** |
| 10 | 27.50 | 1.25 | 32.62 | **0.00** | 36.13 | **0.00** | **39.58** | **0.00** |
| 11 | 66.37 | 0.63 | 69.91 | **0.00** | 72.00 | **0.00** | **78.50** | **0.00** |
| 12 | 90.75 | 1.25 | 93.87 | 0.63 | 94.50 | 0.63 | **96.37** | **0.00** |
| 13 | 91.37 | 0.63 | 92.00 | 0.63 | 93.67 | **0.00** | **95.29** | **0.00** |
| 14 | 85.00 | 1.88 | 86.50 | **0.63** | 88.12 | **0.63** | **89.82** | **0.63** |
| 15 | 36.20 | 3.13 | 39.57 | 1.25 | 40.84 | 0.63 | **42.37** | **0.00** |
| 16 | 15.88 | 7.50 | 19.13 | 4.38 | 22.75 | 3.13 | **26.37** | **1.25** |
| 17 | 33.37 | 3.13 | 36.00 | 3.13 | 37.25 | 3.13 | **41.75** | **2.50** |
| 18 | 87.88 | 1.88 | 89.70 | 0.63 | 91.56 | 0.63 | **94.12** | **0.00** |
| 19 | 22.25 | **1.25** | 25.66 | **1.25** | 27.08 | **1.25** | **29.93** | **1.25** |
| 20 | 49.63 | 0.63 | 51.80 | **0.00** | **55.75** | **0.00** | **55.75** | **0.00** |
| 21 | 90.00 | 1.25 | 91.75 | 1.25 | 93.63 | **0.63** | **96.60** | **0.63** |
| Average | 65.80 | 1.85 | 67.86 | 0.98 | 69.46 | 0.57 | **71.54** | **0.33** |

### C. Monitoring Results

In order to evaluate the performance of process monitoring, the fault detection rate (FDR) and false alarm rate (FAR) [2]

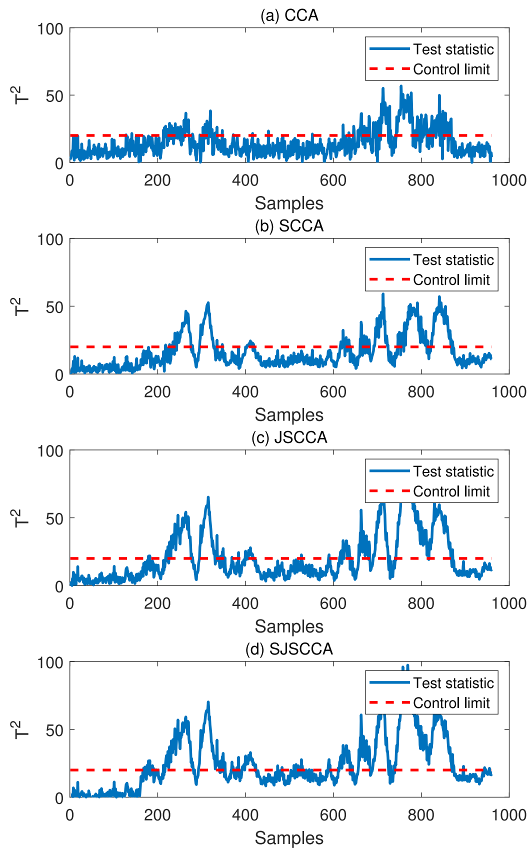


Fig. 3. Monitoring results for fault 10 in the TE process: (a) CCA, (b) SCCA, (c) JSCCA, (d) SJSCCA.

are adopted. The higher FDR and lower FAR indicate the better the monitoring performance. The monitoring results for the TE process are given in Table III. Moreover, the best method for each test case is highlighted in boldface.

Some observations can be made:

- The sparsity based CCA methods (i.e., SCCA, JSCCA, SJSCCA) are superior than CCA in terms of FDR and FAR, which shows that sparse process monitoring is promising. In average, the gains of FDR are 5.74%, 3.66%, 2.06%, and the reduces of FAR are 1.52%, 1.28%, 0.87%, for SJSCCA, JSCCA, SCCA, respectively.
- Compared with CCA and SCCA, JSCCA and SJSCCA could obtain higher FDR values and lower FAR values. The reason is the joint sparsity structure is explored, and outliers are removed. In particular, for fault 16, the FDRs by SJSCCA, JSCCA are 7.24%, 3.62% higher than that by SCCA, and 10.49%, 6.87% higher than that by CCA.
- For all the selected faults, the performances of our proposed SJSCCA are always the best. This good monitoring results come from the combination of the graph Laplacian and the joint sparsity. To appreciate the comparable performance achieved by SJSCCA, the monitoring results for fault 10 are presented in Fig. 3. SJSCCA gives much more samples violated the control limits, which illustrates that SJSCCA is sensitive to the fault. This shows that the monitoring performance of SJSCCA is convincing.

For these reasons, the proposed SJSCCA is a valuable approach for process monitoring.

## V. Conclusion

In this brief, a novel structured joint sparse canonical correlation analysis method is proposed for process monitoring. Compared with the existing CCA-based methods, the proposed SJSCCA embeds a graph Laplacian regularizer so as to capture the geometric structure information. In addition, the joint sparsity is employed to enhance robustness against outliers. More importantly, an efficient two-stage ADMM is established with convergence analysis. Numerical comparisons, on the TE process, are presented to demonstrate its effectiveness.

In future work, it is interested to incorporate the graph Laplacian and joint sparsity into nonlinear CCA.

## References

[1] Z. Chen, S. X. Ding, T. Peng, C. Yang, and W. Cui, "Fault detection for non-Gaussian processes using generalized canonical correlation analysis and randomized algorithms," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1559–1567, Feb. 2018.

[2] S. X. Ding, *Data-Driven Design of Fault Diagnosis and Fault-Tolerant Control Systems*, London, U.K.: Springer-Verlag, 2014.

[3] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," *Comput. Chem. Eng.*, vol. 17, no. 3, pp. 245–255, 1993.

[4] M. Fazel, T. Pong, D. Sun, and P. Tseng, "Hankel matrix rank minimization with applications to system identification and realization," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 3, pp. 946–977, 2013.

[5] M. Imani, E. R. Dougherty, and U. Braga-Neto, "Boolean Kalman filter and smoother under model uncertainty," *Automatica*, vol. 111, Jan. 2020, Art. no. 108609.

[6] Q. Jiang and X. Yan, "Multimode process monitoring using variational Bayesian inference and canonical correlation analysis," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 4, pp. 1814–1824, Oct. 2019.

[7] M. Kim, J. H. Won, J. Youn, and H. Park, "Joint-connectivity-based sparse canonical correlation analysis of imaging genetics for detecting biomarkers of Parkinson's disease," *IEEE Trans. Med. Imag.*, vol. 39, no. 1, pp. 23–34, Jan. 2020.

[8] B. N. G. Koneru and V. Vasudevan, "Sparse artificial neural networks using a novel smoothed LASSO penalization," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 66, no. 5, pp. 848–852, May 2019.

[9] A. Lewis and J. Malick, "Alternating projections on manifolds," *Math. Oper. Res.*, vol. 33, no. 1, pp. 216–234, 2008.

[10] H. Liang, Z. Zhang, and C. K. Ahn, "Event-triggered fault detection and isolation of discrete-time systems based on geometric technique," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 67, no. 2, pp. 335–339, Feb. 2020, doi: 10.1109/TCSII.2019.2907706.

[11] Y. Liu, J. Zeng, L. Xie, S. Luo, and H. Su, "Structured joint sparse principal component analysis for fault detection and isolation," *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 2721–2731, May 2019.

[12] Q. Lu, B. Jiang, R. Gopaluni, P. D. Loewen, and R. D. Braatz, "Sparse canonical variate analysis approach for process monitoring," *J. Process Control*, vol. 71, pp. 90–102, Nov. 2018.

[13] R. Raveendran, H. Kodamana, and B. Huang, "Process monitoring using a generalized probabilistic linear latent variable model," *Automatica*, vol. 96, pp. 73–83, Oct. 2018.

[14] G. Wang and S. Yin, "Quality-related fault detection approach based on orthogonal signal correction and modified PLS," *IEEE Trans. Control Syst. Technol.*, vol. 11, no. 2, pp. 398–405, Apr. 2015.

[15] D. M. Witten, R. Tibshirani, and T. Hastie, "A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis," *Biostatistics*, vol. 10, no. 3, pp. 515–534, 2009.

[16] M. Xu, Z. Zhu, X. Zhang, Y. Zhao, and X. Li, "Canonical correlation analysis with $\ell_{2,1}$-norm for multiview data representation," *IEEE Trans. Cybern.*, early access, Apr. 4, 2019, doi: 10.1109/TCYB.2019.2904753.

[17] R. Zhao and Q. Wang, "Learning separable dictionaries for sparse tensor representation: An online approach," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 66, no. 3, pp. 502–506, Mar. 2019.

[18] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *J. Comput. Graph. Stat.*, vol. 15, no. 2, pp. 265–286, 2006.