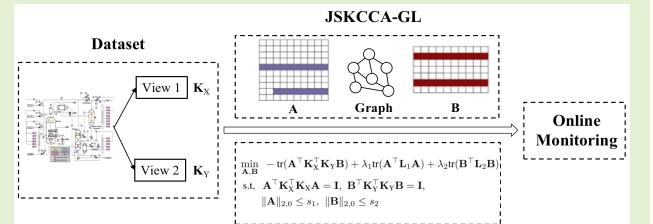


Learning Sparse Kernel CCA With Graph Priors for Nonlinear Process Monitoring

Xianchao Xiu^{ID}, Member, IEEE, and Yunhui Li^{ID}, Member, IEEE

Abstract—Process monitoring (PM) is important for improving product quality and ensuring plant safety in industrial systems. Recently, canonical correlation analysis (CCA)-based PM has shown excellent performance. The core idea is to first seek the relationship between two sets of process variables via CCA, then construct a residual generator and determine test statistics, and finally achieve the best monitoring. However, the global and local structure of process variables is not fully exploited. In this article, we propose a new nonlinear PM approach called joint sparse constrained kernel CCA and graph learning (JSKCCA-GL) by incorporating joint sparse kernel CCA (JSKCCA) and graph priors. Technically, imposing the $\ell_{2,0}$ -norm joint sparse constraints can capture the global row-wise structure, and embedding the graph Laplacian priors can preserve the local structured variable correlation information. Furthermore, an efficient alternating minimization algorithm is developed by integrating least squares and iterative hard thresholding (IHT). The superiority of the proposed JSKCCA-GL is extensively verified on the benchmark Tennessee Eastman process (TEP) and a practical cylinder-piston assembly (CPA) of marine diesel engines. In particular, the fault detection rate (FDR) values for unknown-type faults in the TEP have been improved by around 16.10% on average.

Index Terms—Graph priors, kernel canonical correlation analysis (KCCA), process monitoring (PM), sparse optimization.



I. INTRODUCTION

IN THE past decade, process monitoring (PM) has received broad research attention in process control. Roughly speaking, the PM techniques can be categorized into model-based PM [1] and data-driven PM [2]. However, model-based PM approaches always require precise system models, which makes practical applications difficult [3]. With the development of data acquisition and computing, data-driven PM has become popular. A large number of data-driven approaches using a multivariate statistical analysis (MVA) have emerged, including a principal component analysis (PCA), partial least squares (PLS), independent component analysis (ICA), Fisher discriminant analysis (FDA), and canonical correlation analysis (CCA) (see [4], [5], [6], [7], [8], [9], for example). Unlike other PM approaches, the CCA considers the relationship between inputs and outputs in order to achieve excellent monitoring performance by exploring their different but

complementary information. Now, CCA-based PM approaches have been widely used in both academia and industry.

Modern industrial processes are often nonlinear, which may lead to poor performance of classical CCA, because it assumes a linear relationship between process variables. However, most of the existing contributions for PM are related to linear CCA, such as distributed CCA [10] and multimode CCA [11]. Thus, it is necessary to extend linear CCA to the nonlinear case, as it allows dealing with nonlinear processes [12]. Due to the simple mathematical formulation and easy computing implementation, kernel CCA (KCCA) [13] is often considered in image processing [14] and signal processing [15] when compared with locally weighted CCA [16] and variational CCA [17]. In fact, KCCA first projects the process data onto a higher dimensional space by nonlinear mapping functions and then exploits the canonical correlations on that projected dimensional space. Therefore, KCCA has the ability to capture the nonlinear relationship between process variables. However, the application for PM has not been investigated sufficiently in the literature [18].

As demonstrated in [19], selecting highly sensitive process variables can improve the monitoring performance, since some variables are redundant and cannot provide useful information. To alleviate the issue, a sparse variant of KCCA, called joint sparse KCCA (JSKCCA), was constructed by imposing $\ell_{2,1}$ -norm regularization terms on process

Manuscript received 10 January 2023; accepted 13 February 2023. Date of publication 22 February 2023; date of current version 31 March 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 12001019. The associate editor coordinating the review of this article and approving it for publication was Dr. Priyadip Ray. (*Corresponding author: Yunhui Li.*)

The authors are with the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China (e-mail: xcxiu@shu.edu.cn; liyunhui@shu.edu.cn).

Digital Object Identifier 10.1109/JSEN.2023.3245832

variables [20], [21], [22]. As a consequence, JSKCCA not only enjoys the advantages of sparse approaches but also enjoys the advantages of kernel approaches. It is also validated that the $\ell_{2,1}$ -norm can reduce the effect of outliers and dropout useless variables. However, in some complicated industry processes, JSKCCA may fail to reveal the underlying relationship between process variables. Motivated by recent advances in manifold learning [23], [24], [25], graph priors can be applied to preserve the local prior information embedded in high-dimensional settings [26]. Enforcing graph Laplacian priors provides significant improvements when projecting the original data onto a well-designed space [27]. From the perspective of representation, the combination of learning techniques and data geometric structure is essential. Based on the above analysis, there is a need to construct a new KCCA-based PM approach that can extract useful process variables and capture the local structure between process variables.

In this article, we propose and study an efficient nonlinear PM approach, called joint sparse constrained kernel CCA and graph learning (JSKCCA-GL), to improve the monitoring performance. Specially, the main contributions of this article can be summarized as follows.

- 1) We introduce $\ell_{2,0}$ -norm joint sparse constraints to determine the number of extracted process variables, which is more flexible than the existing $\ell_{2,1}$ -norm regularization. To the best of our knowledge, this is the first work that combines $\ell_{2,0}$ -norm with KCCA.
- 2) We integrate graph learning priors to preserve the local geometry between process variables. Therefore, the proposed JSKCCA-GL can capture both global and local structured variable correlation information.
- 3) We develop a hybrid efficient optimization algorithm by combining least squares and iterative hard thresholding (IHT). Although the convergence is hard to prove mathematically, we verify it through simulations.
- 4) We propose a two-stage monitoring strategy and conduct comparisons with several state-of-the-art approaches on the benchmark Tennessee Eastman process (TEP) and a practical process to validate the superiority of the proposed JSKCCA-GL.

The remainder of this article is structured as follows. Section II introduces some notations and JSKCCA. Section III presents the proposed JSKCCA-GL with an efficient optimization algorithm. Section IV verifies the advantages with numerical comparisons. Section V concludes this article.

II. PRELIMINARIES

A. Notations

Let \mathbb{R}^p and $\mathbb{R}^{p \times n}$ denote the sets of all p -dimensional vectors and $p \times n$ matrices. For $\mathbf{X} \in \mathbb{R}^{p \times n}$, let x_{ij} denote its ij th entry, and \mathbf{x}_i represents its i th row. The ℓ_1 -norm is the sum of absolute values of all components. Following this notation, the $\ell_{2,1}$ -norm is given by $\|\mathbf{X}\|_{2,1} = \sum_{i=1}^p \|\mathbf{x}_i\|_2$. This is a generalized ℓ_1 -norm, called joint sparse or row-wise sparse. Different from the aforementioned $\ell_{2,1}$ -norm, the $\ell_{2,0}$ -norm of matrix \mathbf{X} is denoted by $\|\mathbf{X}\|_{2,0} = |\{i : \|\mathbf{x}_i\|_2 \neq 0\}|$,

which counts the number of nonzero rows of \mathbf{X} . Moreover, the transpose and inverse are written as \mathbf{X}^\top and \mathbf{X}^{-1} , respectively. In addition, for \mathbf{X} and \mathbf{Y} with the same size, their trace is $\text{tr}(\mathbf{X}^\top \mathbf{Y}) = \langle \mathbf{X}, \mathbf{Y} \rangle = \sum_{i=1}^p \sum_{j=1}^n x_{ij} y_{ij}$.

B. Joint Sparse KCCA

Let \mathbf{X} and \mathbf{Y} be the input process data and output process data of n samples, respectively, given by

$$\begin{aligned} \mathbf{X} &= \left[\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_n^T \right] \\ &= \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix} \in \mathbb{R}^{p \times n} \quad (p \text{ variables}) \end{aligned} \quad (1)$$

and

$$\begin{aligned} \mathbf{Y} &= \left[\mathbf{y}_1^T, \mathbf{y}_2^T, \dots, \mathbf{y}_n^T \right] \\ &= \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{q1} & y_{q2} & \cdots & y_{qn} \end{bmatrix} \in \mathbb{R}^{q \times n} \quad (q \text{ variables}). \end{aligned} \quad (2)$$

Suppose that $\varphi : \mathbb{R}^p \rightarrow \mathbb{R}^{\mathcal{N}_X}$ ($\mathcal{N}_X > p$) is a nonlinear mapping function; then, $\Phi_X = [\varphi(\mathbf{x}_1), \dots, \varphi(\mathbf{x}_n)] \in \mathbb{R}^{\mathcal{N}_X \times n}$ is performed by a kernel function κ_X , which satisfies $\kappa_X(\mathbf{x}_i, \mathbf{x}_j) = \langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j) \rangle$. Following the same pattern, \mathbf{Y} can also be mapped into $\Phi_Y \in \mathbb{R}^{\mathcal{N}_Y \times n}$. Therefore, KCCA aims to seek two linear transformations, such that the projected variables achieve the maximum correlation [13]. The mathematical model can be characterized by

$$\begin{aligned} \min_{\mathbf{A}, \mathbf{B}} \quad & -\text{tr}(\mathbf{A}^\top \mathbf{K}_X^\top \mathbf{K}_Y \mathbf{B}) \\ \text{s.t.} \quad & \mathbf{A}^\top \mathbf{K}_X^\top \mathbf{K}_X \mathbf{A} = \mathbf{I}, \quad \mathbf{B}^\top \mathbf{K}_Y^\top \mathbf{K}_Y \mathbf{B} = \mathbf{I} \end{aligned} \quad (3)$$

where $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times r}$ are linear transformations, \mathbf{I} is the r -dimensional identity matrix, and \mathbf{K}_X and \mathbf{K}_Y are kernel matrices composed by $\mathbf{K}_X = \langle \Phi_X, \Phi_X \rangle = [\kappa_X(\mathbf{x}_i, \mathbf{x}_j)]_{i,j=1}^n$, and $\mathbf{K}_Y = \langle \Phi_Y, \Phi_Y \rangle = [\kappa_Y(\mathbf{y}_i, \mathbf{y}_j)]_{i,j=1}^n$.

However, the KCCA formulation in (3) lacks sparseness in transformations \mathbf{A} and \mathbf{B} , which may derive a meaningless solution or bring a misunderstanding interpretation [28]. Then, the following JSKCCA can be considered:

$$\begin{aligned} \min_{\mathbf{A}, \mathbf{B}} \quad & -\text{tr}(\mathbf{A}^\top \mathbf{K}_X^\top \mathbf{K}_Y \mathbf{B}) + \lambda_1 \|\mathbf{A}\|_{2,1} + \lambda_2 \|\mathbf{B}\|_{2,1} \\ \text{s.t.} \quad & \mathbf{A}^\top \mathbf{K}_X^\top \mathbf{K}_X \mathbf{A} = \mathbf{I}, \quad \mathbf{B}^\top \mathbf{K}_Y^\top \mathbf{K}_Y \mathbf{B} = \mathbf{I}. \end{aligned} \quad (4)$$

Here, $\lambda_1, \lambda_2 > 0$ are two sparse parameters. Through choosing different λ_1, λ_2 values, JSKCCA in (4) can throw away certain rows (related to process variables) in matrices \mathbf{A} and \mathbf{B} to improve the interpretation. It is proven in [29] that $\ell_{2,1}$ -norm is more effective than ℓ_1 -norm when processing structured data. This is because the prior information can be exploited, and a better representation is obtained. However, the assumption of models (3) and (4) that the process data are composed of several linear subspaces cannot be ensured in real scenarios. To overcome this drawback, a novel KCCA optimization

Algorithm 1 Least Squares-Based Method for (5)

Input: Process data \mathbf{X}, \mathbf{Y} , regularization parameters λ_1, λ_2 , joint sparse levels s_1, s_2

While not converged **do**

- 1: Construct and center $\mathbf{K}_X, \mathbf{K}_Y$
- 2: Calculate \mathbf{T}_X and \mathbf{T}_Y via (10)
- 3: Solve \mathbf{A} -subproblem according to (12)
- 4: Solve \mathbf{B} -subproblem according to (13)
- 5: Check convergence

End while

Output: Transformations \mathbf{A}, \mathbf{B}

model along with an efficient algorithm will be discussed in Section III.

III. PROPOSED METHOD

A. Model Construction

To fully consider the prior information of process variables, we construct the following JSKCCA-GL model:

$$\begin{aligned} \min_{\mathbf{A}, \mathbf{B}} \quad & -\text{tr}(\mathbf{A}^\top \mathbf{K}_X^\top \mathbf{K}_Y \mathbf{B}) + \lambda_1 \text{tr}(\mathbf{A}^\top \mathbf{L}_1 \mathbf{A}) + \lambda_2 \text{tr}(\mathbf{B}^\top \mathbf{L}_2 \mathbf{B}) \\ \text{s.t.} \quad & \mathbf{A}^\top \mathbf{K}_X^\top \mathbf{K}_X \mathbf{A} = \mathbf{I}, \quad \mathbf{B}^\top \mathbf{K}_Y^\top \mathbf{K}_Y \mathbf{B} = \mathbf{I} \\ & \|\mathbf{A}\|_{2,0} \leq s_1, \quad \|\mathbf{B}\|_{2,0} \leq s_2 \end{aligned} \quad (5)$$

where $s_1, s_2 > 0$ are joint sparse levels, and $\mathbf{L}_1, \mathbf{L}_2 \in \mathbb{R}^{n \times n}$ are graph Laplacian matrices. In fact, KCCA can provide an effective way to explore the correlations of given data, and joint sparse can select the important process variables, while graph priors can preserve the geometric structure.

We would like to point out that, unlike other joint sparse regularized KCCA models, $\ell_{2,0}$ -norm constrained optimization can accurately determine the selected process variables [30], [31] and can be solved by fast solvers [32], [33].

B. Optimization Algorithm

In this section, an optimization algorithm for dealing with the proposed JSKCCA-GL will be discussed in detail. Obviously, problem (5) is not trivial to be solved, because all the constraints are nonconvex, and the objective function involves two variables. Motivated by the observations in recent work [34], [35], it first computes the classical KCCA model in (3) to get a good initial point, and then adds joint sparse constraints and graph Laplacian regularization terms to obtain the solution of JSKCCA-GL in (5). Overall, the iterative framework can be established in Algorithm 1.

First, denote

$$k_1 = \text{rank}(\mathbf{K}_X), \quad k_2 = \text{rank}(\mathbf{K}_Y), \quad m = \text{rank}(\mathbf{K}_X^\top \mathbf{K}_Y). \quad (6)$$

Compute eigenvalue decompositions of \mathbf{K}_X as follows:

$$\begin{aligned} \mathbf{K}_X &= \mathbf{U} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \mathbf{U}^\top \\ &= [\mathbf{U}_1 \quad \mathbf{U}_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} [\mathbf{U}_1 \quad \mathbf{U}_2]^\top \\ &= \mathbf{U}_1 \Sigma_1 \mathbf{U}_1^\top \end{aligned} \quad (7)$$

Algorithm 2 IHT Method for (12)

Input: Data \mathbf{T}_X , regularization parameter λ_1 , joint sparse level s_1 , step-size η

While not converged **do**

- 1: Calculate
- $$\mathbf{A}^{k+1} = \mathbf{A}^k - \eta (\mathbf{K}_X^\top (\mathbf{K}_X \mathbf{A}^k - \mathbf{T}_X) + 2\lambda_1 \mathbf{L}_1 \mathbf{A}^k)$$

- 2: Truncate \mathbf{A}^{k+1} with top s_1 rows preserved

- 3: Check convergence

End while

Output: Transformation \mathbf{A}

where $\mathbf{U} \in \mathbb{R}^{n \times n}$, $\mathbf{U}_1 \in \mathbb{R}^{n \times k_1}$, $\mathbf{U}_2 \in \mathbb{R}^{n \times (n-k_1)}$, and $\Sigma_1 \in \mathbb{R}^{k_1 \times k_1}$. Following a similar idea, $\mathbf{K}_Y = \mathbf{V}_1 \Sigma_2 \mathbf{V}_1^\top$, where $\mathbf{V}_1 \in \mathbb{R}^{n \times k_2}$, and $\Sigma_2 \in \mathbb{R}^{k_2 \times k_2}$.

Consider $\mathbf{U}_1^\top \mathbf{V}_1 = \mathbf{P}_1 \Sigma \mathbf{P}_2^\top$ as the singular value decomposition (SVD) of $\mathbf{U}_1^\top \mathbf{V}_1$, where $\mathbf{P}_1 \in \mathbb{R}^{k_1 \times k_1}$ and $\mathbf{P}_2 \in \mathbb{R}^{k_2 \times k_2}$ are orthogonal matrices, and $\Sigma \in \mathbb{R}^{k_1 \times k_2}$ is a diagonal matrix. Thus, the solution of the KCCA model in (3) can be given by

$$\begin{aligned} \mathbf{A} &= \mathbf{U}_1 \Sigma_1^{-1} \mathbf{P}_1(:, 1:r) + \mathbf{U}_2 \mathbf{E} \\ \mathbf{B} &= \mathbf{V}_1 \Sigma_2^{-1} \mathbf{P}_2(:, 1:r) + \mathbf{V}_2 \mathbf{F} \end{aligned} \quad (8)$$

in which $\mathbf{E} \in \mathbb{R}^{(n-k_1) \times r}$ and $\mathbf{F} \in \mathbb{R}^{(n-k_2) \times r}$. However, the second terms $\mathbf{U}_2 \mathbf{E}$ and $\mathbf{V}_2 \mathbf{F}$ in (8) can be neglected, since they do not contribute to canonical correlations. After some simple algebraic manipulations, it is easy to derive that

$$\begin{aligned} \mathbf{U}_1 \Sigma_1^2 \mathbf{U}_1 \mathbf{A} &= \mathbf{U}_1 \Sigma_1 \mathbf{P}_1(:, 1:r) \\ \mathbf{V}_1 \Sigma_2^2 \mathbf{V}_1 \mathbf{B} &= \mathbf{V}_1 \Sigma_2 \mathbf{P}_2(:, 1:r). \end{aligned} \quad (9)$$

By defining

$$\mathbf{T}_X = \mathbf{U}_1 \mathbf{P}_1(:, 1:r), \quad \mathbf{T}_Y = \mathbf{V}_1 \mathbf{P}_2(:, 1:r) \quad (10)$$

it is concluded that $\mathbf{K}_X \mathbf{K}_X^\top \mathbf{A} = \mathbf{K}_X \mathbf{T}_X$, and $\mathbf{K}_Y \mathbf{K}_Y^\top \mathbf{B} = \mathbf{K}_Y \mathbf{T}_Y$.

Now, it is possible for us to solve the KCCA model in (3) by considering an alternating minimization of the following two subproblems:

$$\begin{aligned} \mathbf{A} &= \arg \min_{\mathbf{A}} \|\mathbf{K}_X \mathbf{A} - \mathbf{T}_X\|_F^2 \\ \mathbf{B} &= \arg \min_{\mathbf{B}} \|\mathbf{K}_Y \mathbf{B} - \mathbf{T}_Y\|_F^2. \end{aligned} \quad (11)$$

It is noted that this step depends on the relationship between KCCA and least squares, which has been proposed in [35]. It also has nice convergence properties from any initial point if there exist no additional constraints or regularization terms.

Next, after embedding joint sparse constraints and graph Laplacian regularization terms into (3), one can obtain problem (5). Accordingly, the solution can be derived by the following.

- 1) **A Subproblem:**

$$\begin{aligned} \min_{\mathbf{A}} \quad & \frac{1}{2} \|\mathbf{K}_X \mathbf{A} - \mathbf{T}_X\|_F^2 + \lambda_1 \text{tr}(\mathbf{A}^\top \mathbf{L}_1 \mathbf{A}) \\ \text{s.t.} \quad & \|\mathbf{A}\|_{2,0} \leq s_1. \end{aligned} \quad (12)$$

2) **B Subproblem:**

$$\begin{aligned} \min_{\mathbf{B}} \quad & \frac{1}{2} \|\mathbf{K}_Y \mathbf{B} - \mathbf{T}_Y\|_F^2 + \lambda_2 \text{tr}(\mathbf{B}^\top \mathbf{L}_2 \mathbf{B}) \\ \text{s.t.} \quad & \|\mathbf{B}\|_{2,0} \leq s_2. \end{aligned} \quad (13)$$

Due to the existence of $\ell_{2,0}$ -norm, joint sparse constrained optimization problems are NP-hard in general [36]. Therefore, the main computational load of Algorithm 1 is solving (12) and (13). Next, taking problem (12) as an example, the algorithm works similarly for problem (13). Inspired by Blumensath [37], an efficient method called IHT can be adopted to pursue the solution of problem (12). It first calculates the first-order gradient descent of the objective at point \mathbf{A}^k with step size $\eta > 0$, and then approximates the solution by simply truncating the top s_1 rows. Algorithm 2 provides the iterative scheme for solving **A** subproblem.

Remark 3.1: Compared with [35], the proposed algorithm considers a different optimization problem. In fact, JSKCCA-GL integrates the $\ell_{2,0}$ -norm joint sparse constraints and graph Laplacian regularization terms into KCCA, while [35] only discusses the ℓ_1 -norm regularized KCCA. More importantly, the interest in this work about JSKCCA-GL is providing a potential solution for monitoring nonlinear processes.

For sparse constrained optimization problems (12) and (13), it is relatively difficult to achieve the global optimum [38]. As demonstrated in [32], even for local minima, some additional conditions should be required, such as the restricted isometric property (RIP) [39]. In addition, alternating minimization algorithms will only converge if optimal solutions to subproblems can be guaranteed [40], [41]. Clearly, Algorithm 1 does not satisfy this condition. Therefore, the discussion on convergence and complexity will be left to our future research.

IV. NUMERICAL STUDIES

This section demonstrates the superiority of the proposed JSKCCA-GL over some state-of-the-art approaches, including CCA [5], SJSCCA [42], KCCA [43], and JSKCCA [22], on the TEP and a practical cylinder-piston assembly (CPA) of marine diesel engines. Although some sparse or nonlinear variants of CCA have been proposed, these models are not considered here, because this article only focuses on kernel-based approaches.

Section V-A presents the monitoring strategy. Section V-B describes the experimental setup. Sections V-C and V-D show application results for the TEP and CPA, respectively. Section V-E illustrates the convergence by a numerical example.

A. Monitoring Strategy

This section presents a monitoring procedure based on the proposed JSKCCA-GL, including offline modeling and online monitoring. To some extent, CCA-based PM can be regarded as an extension of PCA-based and PLS-based ones. Therefore, we have to consider how to design a residual generator, define test statistics, determine control limits, and make a decision.

For kernel-based approaches, normalization is the first step to eliminate engineering scales [44]. After normalization, the training datasets have zero mean and unit variance. Based on this, kernel matrices \mathbf{K}_X and \mathbf{K}_Y are computed and centered, for example

$$\mathbf{K}_X = \mathbf{1}_n \mathbf{K}_X \mathbf{1}_n^\top - \mathbf{K}_X \mathbf{1}_n \mathbf{1}_n^\top + \mathbf{1}_n \mathbf{K}_X \mathbf{1}_n \quad (14)$$

where

$$\mathbf{1}_n = \frac{1}{n} \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix} \in \mathbb{R}^{n \times n}. \quad (15)$$

When optimization problem (5) is solved by Algorithms 1 and 2, the sparse transformation matrices \mathbf{A} and \mathbf{B} are, thus, obtained. Subsequently, test statistics and control limits should be well defined. In CCA-based monitoring approaches, the Hotelling T^2 statistics are often employed for PM, which relies on the fact that the T^2 test statistics have the best monitoring if there exists no information about the fault [5]. Suppose $\mathbf{x} \in \mathbb{R}^p$ and $\mathbf{y} \in \mathbb{R}^q$ are the testing datasets, the residual signal \mathbf{r} can be constructed by

$$\mathbf{r} = \mathbf{A}^\top \mathbf{K}(\mathbf{x}) - \Lambda \mathbf{B}^\top \mathbf{K}(\mathbf{y}) \quad (16)$$

where $\mathbf{K}(\mathbf{x})$ and $\mathbf{K}(\mathbf{y})$ are the kernel vectors of \mathbf{x} and \mathbf{y} , for example, $\mathbf{K}(\mathbf{x}) = [k(\mathbf{x}_1, \mathbf{x}), \dots, k(\mathbf{x}_n, \mathbf{x})] \in \mathbb{R}^n$, and Λ is the covariance matrix of \mathbf{A} and \mathbf{B} . In addition, the T^2 test statistics can be estimated by

$$T^2 = \mathbf{r}^\top (\mathbf{I} - \Lambda^2)^{-1} \mathbf{r}. \quad (17)$$

For nonlinear processes, how to determine the control limits is still an open question. As suggested in [16] and [45], the χ^2 distribution is valid for KCCA-based PM, and its performance is promising. Therefore, in this study, the corresponding control limits can be determined by

$$J_{th,T^2} = \chi_{\alpha}^2(r) \quad (18)$$

which depends on the freedom r and significance level α .

If the T^2 statistics violate the control limits, there happens a fault; otherwise, there is no fault. Thus, the following logic can be adopted:

$$\begin{cases} T^2 > J_{th,T^2} \Rightarrow \text{faulty} \\ T^2 \leq J_{th,T^2} \Rightarrow \text{fault-free}. \end{cases} \quad (19)$$

To conclude this section, a two-stage monitoring framework based on JSKCCA-GL is provided below.

1) Offline Modeling:

- a) Normalize training datasets \mathbf{X} and \mathbf{Y} .
- b) Center kernel matrices \mathbf{K}_X and \mathbf{K}_Y .
- c) Solve JSKCCA-GL using Algorithms 1 and 2.
- d) Compute control limits J_{th,T^2} using (18).

2) Online Monitoring:

- a) Normalize testing datasets \mathbf{x} and \mathbf{y} .
- b) Center kernel vectors $\mathbf{K}(\mathbf{x})$ and $\mathbf{K}(\mathbf{y})$.
- c) Calculate T^2 test statistics based on (17).
- d) Make a decision according to (19).
- e) Next sample.

B. Experimental Setup

1) *Kernel Selection*: For all kernel-based approaches, the following Gaussian kernel function is chosen, i.e.,

$$\kappa_X(\mathbf{x}_i, \mathbf{x}_j) = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{c}} \quad (20)$$

where c is the kernel bandwidth parameter. According to [46], it is set by $c = \tau p \sigma^2$, where p is the number of variables, σ^2 is the variance of the monitoring variables, and τ is set to 10 empirically. The same kernel function is applied to dataset \mathbf{Y} . Recall that the datasets \mathbf{X} and \mathbf{Y} have been normalized before preceding; thus, the variance σ^2 is equivalent to 1.

2) *Laplacian Matrices*: For datasets \mathbf{X} and \mathbf{Y} , Laplacian matrices \mathbf{L}_1 and \mathbf{L}_2 can be constructed based on [27]. Next, we will take how to construct \mathbf{L}_1 as an example. First, compose a graph as $\mathcal{G} = (\mathcal{V}, \mathcal{E}; \mathbf{W})$ with \mathcal{V} being the vertex, \mathcal{E} being the edges, and \mathbf{W} being the edge weights. For vertices i and j , if $\mathbf{x}_i \in \text{KNN}(\mathbf{x}_j)$ or $\mathbf{x}_j \in \text{KNN}(\mathbf{x}_i)$, then $w_{ij} = 1$, otherwise 0. Here, KNN denotes the k nearest neighbor. Let \mathbf{D} denote the diagonal matrix, and its diagonal elements represent the degrees of each vertex. Therefore, the graph Laplacian matrix of \mathbf{X} can be described as $\mathbf{L}_1 = \mathbf{D} - \mathbf{W}$, and the graph Laplacian matrix of \mathbf{Y} can be constructed similarly.

3) *Regularization Parameters*: For the proposed JSKCCA-GL, the regularization parameters λ_1 and λ_2 can be determined by using the fivefold cross-validation technique, including the following: 1) define the candidate set, say, $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$; 2) divide the training datasets into five disjoint parts with equal lengths; 3) select one part as the testing set and others as the training set; 4) repeat the above procedure five times; and 5) choose the one that delivers the minimum objective. Another parameter that must be chosen with care is the number of joint sparse, as this parameter significantly impacts the PM accuracy. If they are too large, the constraints will be meaningless. If they are too small, the monitoring performance will be very poor. As a matter of experience, joint sparse parameters s_1 and s_2 are initialized with small values and then gradually increased by a certain proportion.

4) *Stopping Criteria*: For all CCA-based approaches, the maximum number of iterations is set as 500. Moreover, the relative differences of variables, i.e., $\|\mathbf{A}^{k+1} - \mathbf{A}^k\|_F / \|\mathbf{A}^k\|_F$, $\|\mathbf{B}^{k+1} - \mathbf{B}^k\|_F / \|\mathbf{B}^k\|_F$, are less than 10^{-3} .

5) *Monitoring Indicators*: To measure the performance of PM, fault detection rate (FDR) and false alarm rate (FAR) are often adopted [2]. Fig. 1 provides the illustration of fault detection and false alarm, where a fault is introduced at the 161st sample. Note that a sample that occurs when there is a fault is considered as a fault detection sample, while a sample that triggers an alarm when there is no fault is considered as a false alarm sample. From the statistical perspective, FDR and FAR are given by $\text{FDR} = \text{prob}(T^2 > J_{th,T^2} | f \neq 0)$, $\text{FAR} = \text{prob}(T^2 > J_{th,T^2} | f = 0)$.

C. Application to the TEP

1) *Dataset and Preparation*: The TEP is a benchmark industrial process dataset, which has been widely used to validate different PM techniques. It contains five units, i.e., a reactor,

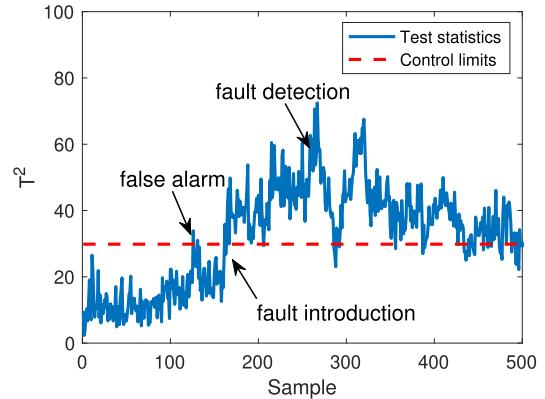


Fig. 1. Illustration of detection indices.

a condenser, a separator, a stripper, and a compressor; see Fig. 2 for a diagram of the TEP. More detailed information can be found in [47].

In the process, there exist one fault-free dataset and 21 fault datasets. For each fault dataset, it collects 960 samples and introduces a fault at the 161st sample. It means that there occurs a fault from the 161st sample to the end. In this study, the fault-free dataset is employed for offline modeling, while fault datasets are utilized for online monitoring. According to [11], variables XMV(1–11) are chosen as \mathbf{X} , and variables XMEAS(1–22) are chosen as \mathbf{Y} .

2) *Monitoring Performance*: The monitoring results of CCA, SJSCCA, KCCA, JSKCCA, and JSKCCA-GL are listed in Table I. Moreover, the best results are indicated in bold. Note that the FDR values of faults IDV(03), IDV(09), and IDV(15) are relatively low, because the means or variances of these faults have small changes. From Table I, it can be concluded that the kernel-based CCA variants, including KCCA, JSKCCA, and JSKCCA-GL, outperform CCA and SJSCCA in most cases. For unknown types of faults IDV(16)–IDV(20), KCCA can achieve an average FDR gain of 16.10% and an average FAR reduction of 2.63% compared with CCA. It demonstrates that the nonlinear PM is useful for the data without knowing fault types. Furthermore, compared with other KCCA-based approaches, the proposed JSKCCA-GL always performs the best, even for some problematic faults. In particular, for fault IDV(16), the FDR value of JSKCCA-GL is 6.13% higher than that of JSKCCA and 8.88% higher than that of KCCA. This considerable performance comes from the introduction of joint sparse constraints and graph priors, which help to preserve the global and local structure of process data.

In order to observe the performance achieved by the proposed JSKCCA-GL, the visual results are illustrated by faults IDV(07), IDV(10), and IDV(16). Moreover, some obvious improvements are provided in inset plots.

- 1) IDV(07) involves a step change, which is relatively simple to be detected, because the magnitude of this fault is large. The monitoring performance is displayed in Fig. 3. All CCA-based PM approaches can detect the fault at the 161st sample and show good monitoring results. From inset plots in Fig. 3, it can be concluded that kernel-based approaches have fewer false alarms than linear ones.

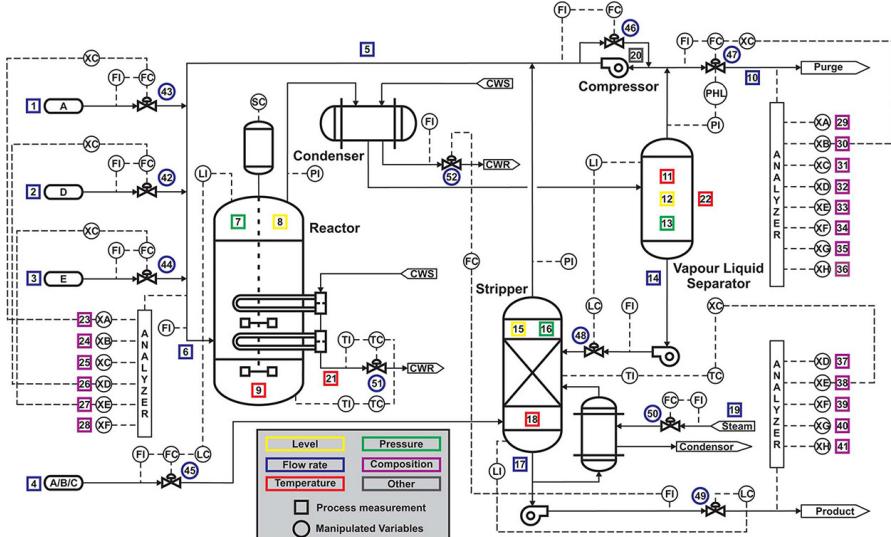


Fig. 2. Diagram of the TEP.

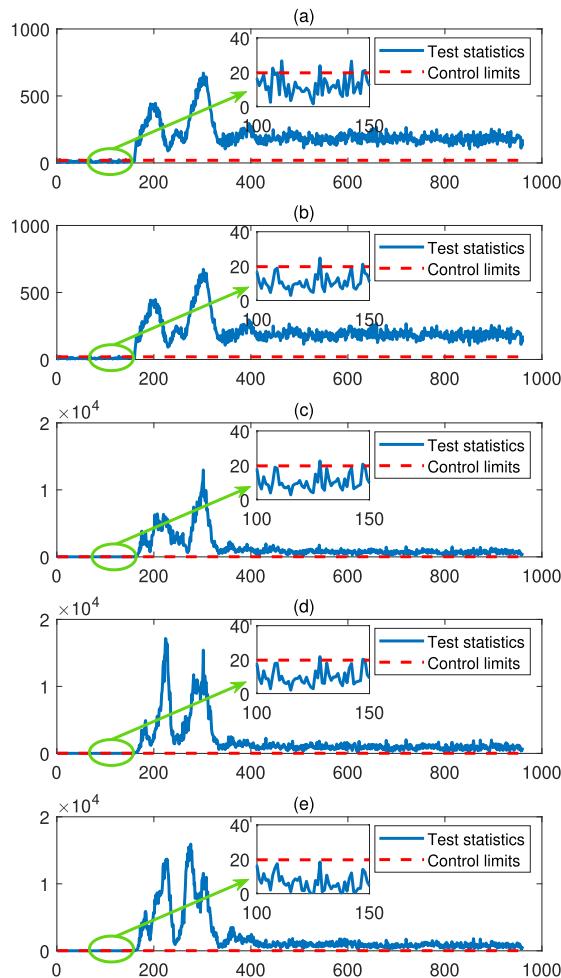


Fig. 3. Visual performance for IDV(07). (a) CCA. (b) SJSCCA. (c) KCCA. (d) JSKCCA. (e) JSKCCA-GL.

- 2) IDV(10) occurs a random variation in the feed C temperature in stream 4. The test statistics for all five approaches are illustrated in Fig. 4. In comparison, KCCA, JSKCCA, and JSKCCA-GL have excellent monitoring performance, i.e., violated numbers between

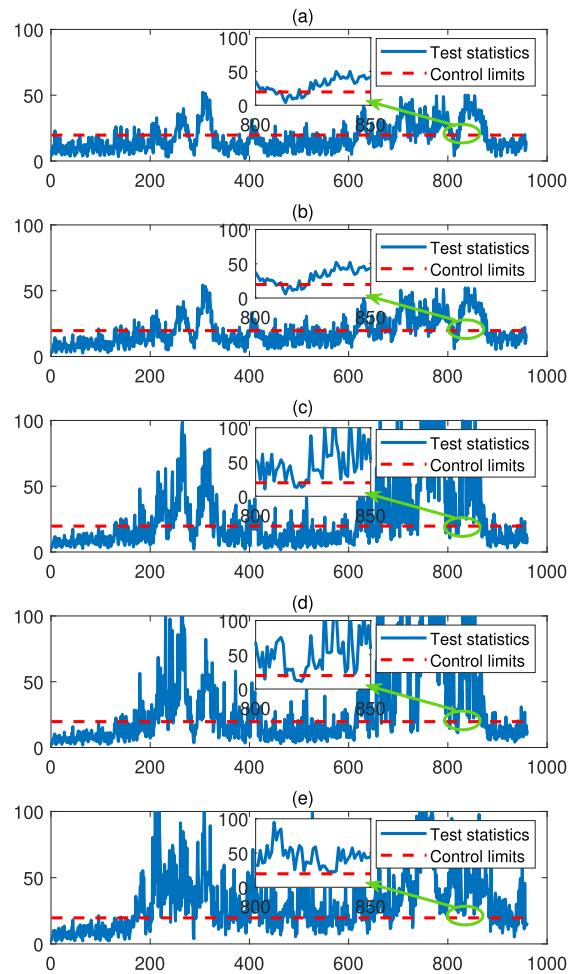


Fig. 4. Visual performance for IDV(10). **(a)** CCA. **(b)** SJSCCA. **(c)** KCCA. **(d)** JSKCCA. **(e)** JSKCCA-GL.

800 and 850 samples; see inset plots in Fig. 4. It reflects that nonlinear PM is promising for industrial processes.

3) IDV(16) is an unknown fault, which is relatively difficult to be detected in practice. The visual results are presented in Fig. 5. By carefully observing the values after the 161st sample, JSKCCA-GL obtains more violated

TABLE I
MONITORING RESULTS FOR SELECTED FAULTS IN THE TEP

Fault No.	CCA		SJSCCA		KCCA		JSKCCA		JSKCCA-GL	
	FDR	FAR	FDR	FAR	FDR	FAR	FDR	FAR	FDR	FAR
IDV(01)	99.25%	0.00%	99.38%	0.00%	99.38%	0.00%	99.50%	0.00%	99.75%	0.00%
IDV(02)	97.63%	1.25%	98.25%	0.63%	98.38%	0.63%	98.63%	0.63%	99.47%	0.00%
IDV(03)	4.63%	1.88%	8.63%	0.63%	6.75%	1.25%	8.75%	0.63%	9.13%	0.63%
IDV(04)	100%	1.88%	100%	0.00%	100%	0.00%	100%	0.00%	100%	0.00%
IDV(05)	33.38%	1.25%	43.75%	0.00%	52.63%	0.00%	58.25%	0.00%	62.78%	0.00%
IDV(06)	100%	0.63%	100%	0.63%	100%	0.63%	100%	0.63%	100%	0.63%
IDV(07)	95.50%	1.88%	99.75%	1.25%	99.75%	0.63%	99.75%	0.63%	99.88%	0.00%
IDV(08)	97.38%	2.50%	98.13%	0.00%	97.50%	0.63%	98.38%	0.00%	99.25%	0.00%
IDV(09)	4.00%	5.00%	4.88%	2.50%	5.25%	2.50%	5.75%	1.88%	6.13%	1.88%
IDV(10)	35.50%	2.50%	36.25%	0.00%	37.63%	0.00%	39.58%	0.00%	42.75%	0.00%
IDV(11)	67.38%	1.88%	75.00%	0.00%	76.13%	0.00%	79.50%	0.00%	80.38%	0.00%
IDV(12)	93.38%	1.88%	99.38%	0.00%	99.38%	0.00%	99.50%	0.00%	99.75%	0.00%
IDV(13)	93.88%	0.00%	94.75%	0.00%	95.25%	0.00%	95.38%	0.00%	95.88%	0.00%
IDV(14)	90.38%	1.88%	93.88%	0.00%	93.63%	0.00%	95.75%	0.00%	99.25%	0.00%
IDV(15)	3.75%	2.50%	5.63%	1.63%	4.63%	1.88%	5.25%	1.88%	6.13%	1.25%
IDV(16)	19.63%	8.75%	38.25%	1.25%	43.50%	2.50%	46.25%	2.50%	52.38%	0.63%
IDV(17)	57.50%	1.88%	65.13%	0.00%	82.38%	0.00%	92.50%	0.00%	95.88%	0.00%
IDV(18)	87.88%	2.50%	90.63%	1.25%	91.75%	0.63%	92.25%	0.63%	94.63%	0.00%
IDV(19)	14.50%	0.63%	28.47%	0.63%	36.38%	0.00%	38.38%	0.00%	41.75%	0.00%
IDV(20)	54.75%	0.00%	60.13%	0.00%	60.75%	0.00%	61.88%	0.00%	62.63%	0.00%
IDV(21)	32.13%	0.63%	38.13%	0.00%	37.75%	0.00%	39.25%	0.00%	42.88%	0.00%

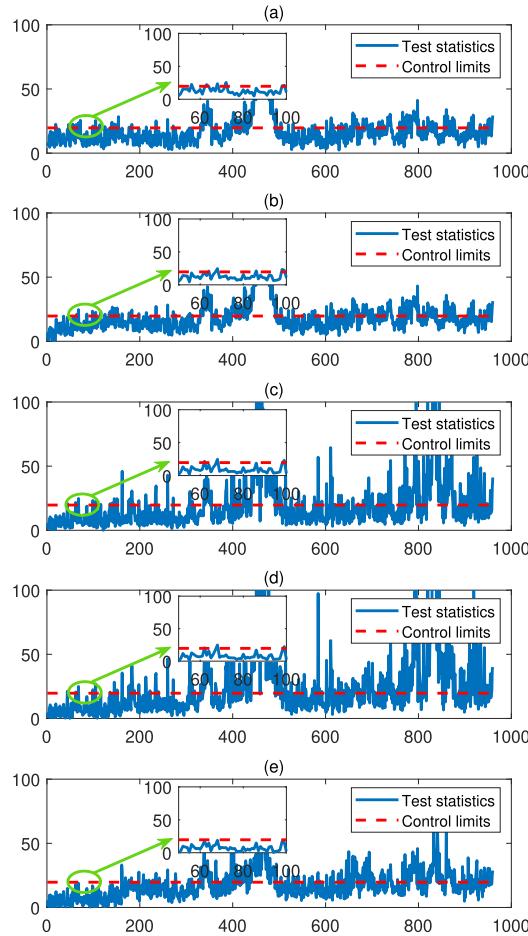


Fig. 5. Visual performance for IDV(16). (a) CCA. (b) SJSCCA. (c) KCCA. (d) JSKCCA. (e) JSKCCA-GL.

samples than other approaches. Moreover, from inset plots in Fig. 5, it can be seen that false alarms of JSKCCA-GL are reduced, which verifies that graph priors are also effective for PM.

TABLE II
SELECTED VARIABLES IN THE CPA

Variable No.	Description
01–05	exhaust gas temperature
06–10	cooling oil inlet and outlet temperature difference
11–15	oil inlet pressure
16–20	JCW inlet and outlet temperature difference
21–25	JCW inlet pressure

TABLE III
MONITORING RESULTS FOR THE CPA

	CCA	SJSCCA	KCCA	JSKCCA	JSKCCA-GL
FDR	18.75%	39.00%	63.75%	66.25%	69.50%
FAR	1.88%	0.63%	0.63%	0.63%	0.63%

D. Application to the CPA

1) *Dataset and Preparation*: The CPA is an important part of diesel engines, and it holds great pressure during the normal operation. The CPA dataset is collected in a real-world two-stroke low-speed marine diesel engine with five cylinders [48].

In the process, 1000 samples are chosen for offline modeling, and a step change is introduced at the 201st sample for online monitoring. It means that it has a fault from the 201st sample to the end. Note that the magnitude of a fault will affect the monitoring performance. If it is significant, all approaches can detect this fault successfully, and if it is small, all approaches fail to detect this fault. Therefore, we choose a step change of 5 to test different PM approaches. For each cylinder, five variables are selected, so there exist 25 variables in total; see Table II for the detailed descriptions. Here, JCW denotes jacket cooling water. In this study, variable Nos. 01–10 are set to **X**, and variable Nos. 11–25 are set to **Y**.

2) *Monitoring Performance*: The monitoring results for the CPA dataset are established in Table III, and the corresponding visual results are shown in Fig. 6. Obviously, all five

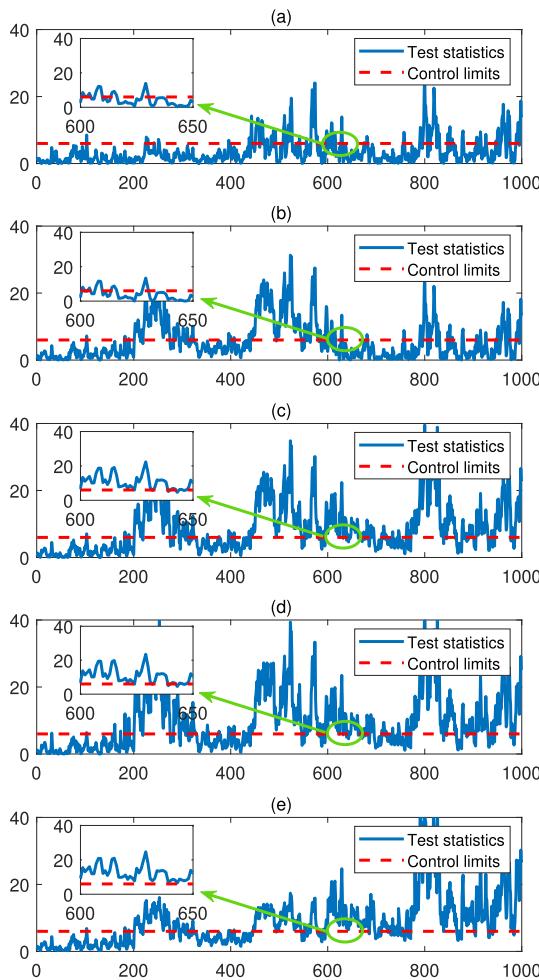


Fig. 6. Visual performance for CPA. (a) CCA. (b) SJSCCA. (c) KCCA. (d) JSKCCA. (e) JSKCCA-GL.

approaches can detect the fault immediately after the fault is introduced at the 201st sample. However, it can be seen from the inset plots in Fig. 6 that JSKCCA-GL achieves a better monitoring performance than others, because most of the fault samples that violated the control limits are detected. This convinces that JSKCCA-GL can monitor the fault consistently. Meanwhile, JSKCCA-GL has a very similar percentage of false alarms. All these results suggest that the proposed approach has an outstanding performance for monitoring CPA.

Through the above numerical studies on the benchmark TEP dataset and a practical CPA dataset, it is concluded that the proposed JSKCCA-GL can achieve a more satisfactory PM ability than linear and other kinds of kernel approaches. It is further shown that the introduced joint sparse and graph priors are beneficial to preserve the variable structure.

E. Convergence Analysis

At the end of this section, the convergence will be illustrated numerically through a simple example. To validate the convergence property, the relative differences of variables **A** and **B** for IDV(07) are displayed in Fig. 7. It is not difficult to conclude that both of the relative differences decrease rapidly when the number of iterations increases. After finite iterations, the relative differences become less than 10^{-3} , which indicates

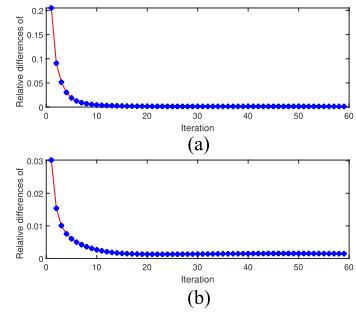


Fig. 7. Convergence illustration. (a) Relative differences of **A**. (b) Relative differences of **B**.

that the proposed algorithm converges. For PM, a convergent algorithm means that the monitoring performance is consistent and stable.

V. CONCLUSION

In this article, we have proposed a new nonlinear monitoring approach. Compared with other existing work, the proposed JSKCCA-GL incorporates joint sparse constraints with $\ell_{2,0}$ -norm and graph priors for structure preservation. Moreover, we have intuitively developed an efficient optimization algorithm by combining least squares and an IHT technique. In addition, we have verified the monitoring performance of the proposed JSKCCA-GL on the benchmark TEP dataset and a practical CPA dataset in comparison with the related existing approaches with better performance.

Although the proposed method is efficient in the field of PM, several problems remain open. First, it is urgent to analyze the convergence and complexity theoretically. Second, it is possible to construct a deep CCA framework to learn a better nonlinear projection. Finally, it is necessary to investigate the fault isolation and identification once the fault is detected.

REFERENCES

- [1] S. X. Ding, *Model-Based Fault Diagnosis Techniques—Design Schemes, Algorithms and Tools*. Berlin, Germany: Springer-Verlag, 2008.
- [2] S. X. Ding, *Data-Driven Design of Fault Diagnosis and Fault-Tolerant Control Systems*. London, U.K.: Springer-Verlag, 2014.
- [3] C. M. Furse, M. Kafal, R. Razzaghi, and Y.-J. Shin, “Fault diagnosis for electrical systems and power networks: A review,” *IEEE Sensors J.*, vol. 21, no. 2, pp. 888–906, Jan. 2021.
- [4] S. Yin, S. 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,” *J. Process Control*, vol. 22, no. 9, pp. 1567–1581, Oct. 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 Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1559–1567, Feb. 2018.
- [6] M. Rezamand, M. Kordestani, R. Carriveau, D. S. Ting, and M. Saif, “A new hybrid fault detection method for wind turbine blades using recursive PCA and wavelet-based PDF,” *IEEE Sensors J.*, vol. 20, no. 4, pp. 2023–2033, Feb. 2020.
- [7] L. Liu, L. Chen, Z. Wang, and D. Liu, “Early fault detection of planetary gearbox based on acoustic emission and improved variational mode decomposition,” *IEEE Sensors J.*, vol. 21, no. 2, pp. 1735–1745, Jan. 2021.
- [8] L. Ma, M. Wang, and K. Peng, “A novel bidirectional gated recurrent unit-based soft sensor modeling framework for quality prediction in manufacturing processes,” *IEEE Sensors J.*, vol. 22, no. 19, pp. 18610–18619, Oct. 2022.
- [9] H. Chen, B. Jiang, S. X. Ding, and B. Huang, “Data-driven fault diagnosis for traction systems in high-speed trains: A survey, challenges, and perspectives,” *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 3, pp. 1700–1716, Mar. 2022.

- [10] Z. Chen et al., "A distributed canonical correlation analysis-based fault detection method for plant-wide process monitoring," *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 2710–2720, May 2019.
- [11] 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.
- [12] X. Yang, W. Liu, W. Liu, and D. Tao, "A survey on canonical correlation analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 6, pp. 2349–2368, Dec. 2019.
- [13] P. L. Lai and C. Fyfe, "Kernel and nonlinear canonical correlation analysis," *Int. J. Neural Syst.*, vol. 10, pp. 365–377, Oct. 2000.
- [14] W. Zheng, X. Zhou, C. Zou, and L. Zhao, "Facial expression recognition using kernel canonical correlation analysis (KCCA)," *IEEE Trans. Neural Netw.*, vol. 17, no. 1, pp. 233–238, Jan. 2006.
- [15] S. Van Vaerenbergh, J. Via, and I. Santamaria, "Blind identification of SIMO Wiener systems based on kernel canonical correlation analysis," *IEEE Trans. Signal Process.*, vol. 61, no. 9, pp. 2219–2230, May 2013.
- [16] Q. Jiang and X. Yan, "Locally weighted canonical correlation analysis for nonlinear process monitoring," *Ind. Eng. Chem. Res.*, vol. 57, no. 41, pp. 13783–13792, Oct. 2018.
- [17] Y. Liu, B. Liu, X. Zhao, and M. Xie, "A mixture of variational canonical correlation analysis for nonlinear and quality-relevant process monitoring," *IEEE Trans. Ind. Electron.*, vol. 65, no. 8, pp. 6478–6486, Aug. 2018.
- [18] P.-E. P. Odiwei and Y. Cao, "Nonlinear dynamic process monitoring using canonical variate analysis and kernel density estimations," *IEEE Trans. Ind. Informat.*, vol. 6, no. 1, pp. 36–45, Feb. 2010.
- [19] J. Yi, L. Wu, W. Zhou, H. He, and L. Yao, "A sparse dimensionality reduction approach based on false nearest neighbors for nonlinear fault detection," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 51, no. 8, pp. 4980–4992, Aug. 2021.
- [20] Z. Zhang, M. Zhao, and T. W. S. Chow, "Binary-and multi-class group sparse canonical correlation analysis for feature extraction and classification," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 10, pp. 2192–2205, Oct. 2013.
- [21] J. Fang, D. Lin, S. C. Schulz, Z. Xu, V. D. Calhoun, and Y.-P. Wang, "Joint sparse canonical correlation analysis for detecting differential imaging genetics modules," *Bioinformatics*, vol. 32, no. 22, pp. 3480–3488, 2016.
- [22] K. Yoshida, J. Yoshimoto, and K. Doya, "Sparse kernel canonical correlation analysis for discovery of nonlinear interactions in high-dimensional data," *BMC Bioinf.*, vol. 18, no. 1, pp. 1–11, Dec. 2017.
- [23] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
- [24] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.
- [25] P. Absil, R. Mahony, and R. Sepulchre, *Optim. Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton Univ. Press, 2009.
- [26] S.-K. Kim and C. K. Ahn, "Robust invariant manifold-based output voltage-tracking controller for DC/DC boost power conversion systems," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 51, no. 3, pp. 1582–1589, Mar. 2021.
- [27] 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.
- [28] Z. Shi, C. Zhou, Y. Gu, N. A. Goodman, and F. Qu, "Source estimation using coprime array: A sparse reconstruction perspective," *IEEE Sensors J.*, vol. 17, no. 3, pp. 755–765, Feb. 2017.
- [29] S. Zhang, C. Gao, F. Chen, S. Luo, and N. Sang, "Group sparse-based mid-level representation for action recognition," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 47, no. 4, pp. 660–672, Apr. 2017.
- [30] F. Nie, X. Dong, L. Tian, R. Wang, and X. Li, "Unsupervised feature selection with constrained $\ell_{2,0}$ -norm and optimized graph," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 4, pp. 1702–1713, Apr. 2022.
- [31] X. Xiu, Z. Miao, and W. Liu, "A sparsity-aware fault diagnosis framework focusing on accurate isolation," *IEEE Trans. Ind. Informat.*, vol. 19, no. 2, pp. 1356–1365, Feb. 2023.
- [32] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, Nov. 2009.
- [33] S. Zhou, N. Xiu, and H. Qi, "Global and quadratic convergence of Newton hard-thresholding pursuit," *J. Mach. Learn. Res.*, vol. 22, no. 12, pp. 1–45, 2021.
- [34] D. Chu, L.-Z. Liao, M. K. Ng, and X. Zhang, "Sparse canonical correlation analysis: New formulation and algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 3050–3065, Dec. 2013.
- [35] D. Chu, L. Liao, M. Ng, and X. Zhang, "Sparse kernel canonical correlation analysis," in *Proc. Int. MultiConference Eng. Comput. Scientists*, 2013, pp. 322–327.
- [36] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, 1995.
- [37] T. Blumensath, "Compressed sensing with nonlinear observations and related nonlinear optimization problems," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3466–3474, Jun. 2013.
- [38] A. Beck and Y. C. Eldar, "Sparsity constrained nonlinear optimization: Optimality conditions and algorithms," *SIAM J. Optim.*, vol. 23, no. 3, pp. 1480–1509, 2013.
- [39] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Nov. 2005.
- [40] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka–Łojasiewicz inequality," *Math. Oper. Res.*, vol. 35, no. 2, pp. 438–457, May 2010.
- [41] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Math. Program.*, vol. 146, nos. 1–2, pp. 459–494, Jan. 2014.
- [42] X. Xiu, Y. Yang, L. Kong, and W. Liu, "Data-driven process monitoring using structured joint sparse canonical correlation analysis," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 68, no. 1, pp. 361–365, Jan. 2021.
- [43] Q. Liu, Q. Zhu, S. J. Qin, and T. Chai, "Dynamic concurrent kernel CCA for strip-thickness relevant fault diagnosis of continuous annealing processes," *J. Process Control*, vol. 67, pp. 12–22, Jul. 2017.
- [44] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998.
- [45] Q. Chen and Y. Wang, "Key-performance-indicator-related state monitoring based on kernel canonical correlation analysis," *Control Eng. Pract.*, vol. 107, Feb. 2021, Art. no. 104692.
- [46] J.-M. Lee, C. K. Yoo, S. W. Choi, P. A. Vanrolleghem, and I.-B. Lee, "Nonlinear process monitoring using kernel principal component analysis," *Chem. Eng. Sci.*, vol. 59, no. 1, pp. 223–234, Jan. 2004.
- [47] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," *Comput. Chem. Eng.*, vol. 17, no. 3, pp. 245–255, Mar. 1993.
- [48] X. Xiu, J. Fan, Y. Yang, and W. Liu, "Fault detection using structured joint sparse nonnegative matrix factorization," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.



Xianchao Xiu (Member, IEEE) received the Ph.D. degree in operations research from Beijing Jiaotong University, Beijing, China, in 2019.

From June 2019 to May 2021, he has worked as a Postdoctoral Researcher with Peking University, Beijing. He is a Faculty Member with the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, China. His current research interests include large-scale sparse optimization, signal processing, deep learning, and data-driven process monitoring.



Yunhui Li (Member, IEEE) received the Ph.D. degree in control science and engineering from the Harbin Institute of Technology, Harbin, China, in 2019.

From October 2019 to July 2020, he has worked as an Assistant Researcher with the Shanghai Institute of Mechanical and Electrical Engineering, Shanghai, China. He is currently a Postdoctoral Researcher with the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai. His current research interests include state estimation, intelligent information fusion and their applications in navigation technology, and fault detection.