Toward Efficient Process Monitoring Using Spatiotemporal PCA

Yunhui Li, *Member, IEEE*, Xianchao Xiu, *Member, IEEE*, and Wanquan Liu, *Senior Member, IEEE*

***Abstract*—Principal component analysis (PCA) has shown its high efficiency in process monitoring (PM). However, most of the existing PCA-based PM approaches only consider the spatial prior and ignore the temporal prior. Therefore, in this brief, we propose a novel PM framework using spatiotemporal PCA (STPCA), which incorporates both spatial and temporal priors. Technically, the spatial prior is integrated to preserve the cause-effect relationship of process variables, and the temporal prior is embedded to maintain the geometric structure of process samples. Moreover, an efficient optimization algorithm is developed using the alternating direction method of multipliers (ADMM) in a symmetric Gauss-Seidel (sGS) manner. Finally, the improved monitoring performance is verified on the benchmark Tennessee Eastman (TE) process. In particular, compared with PCA, the fault detection rate of fault IDV(20) is increased by 9.88%. This suggests that the proposed framework is promising for PM.**

***Index Terms*—Optimization algorithm, principal component analysis (PCA), process monitoring (PM), spatiotemporal prior.**

I. Introduction

DURING the last few decades, process monitoring (PM) has gained extensive attention [1]. For example, the PM for antenna systems is very important with crucial impacts on ensuring product quality, maintaining operation efficiency, and improving industrial safety [2], [3]. In general, PM approaches can be divided into three categories: model-based, data-driven, and knowledge-based. With the rapid development of sensor technology and computing capabilities, data-driven approaches have achieved tremendous success in both academia and industry. Compared with the other two categories, data-driven PM becomes more suitable for large-scale complicated industrial processes because it only relies on data itself without providing additional information about the control system, such as physical laws and chemical models. The most commonly used data-driven PM approaches are based on, for example, multivariate analysis (MVA), signal processing, and neural networks [4]–[7]. As a basic MVA-based PM approach, principal component analysis (PCA) aims to seek a set of principal components (PCs) to characterize the given process information as much as possible. Since the mathematical model is simple and the optimization implementation is easy, PCA has been widely studied and the performance is proved to be inspiring. In fact, PCA-based PM contains two stages: offline modeling and online monitoring, where the offline modeling stage determines the control limit and the online monitoring stage estimates the test statistic. If the achieved test statistic violates the control limit, a fault has been detected, otherwise, no fault has occurred.

In literature, Lee *et al.* [8] developed a multi-mode kernel PCA-based strategy for monitoring the batch process; Harmouche *et al.* [9] considered a KullbackâĂŞLeibler divergence-based PCA approach for incipient detection; Liu *et al.* [10] proposed a PCA variant by adding the sparsity on the loading matrix to improve process variable interpretation. However, due to sensor failures and malicious attacks, process structure may be destroyed in the optimization procedure. From the aspect of feature dimension reduction, many researchers have suggested embedding some prior knowledge; see [11]–[13]. For example, Liu *et al.* [14] introduced a graph Laplacian regularization term into PCA to capture the cause-effect relationship of process variables and verified the efficiency of representing the spatial graph prior. Xiu *et al.* [15] considered Laplacian regularized robust PCA to reduce the effects of noise and outliers. Recently, Zhai *et al.* [16] studied why the graph Laplacian matrix can describe the spatial prior and contribute to better monitoring performance. Since only the spatial graph prior is exploited, the above-mentioned approaches can be denoted as spatial PCA (SPCA). Unfortunately, the geometric structure of process samples (temporal prior) has never been investigated for PCA-based PM. Therefore, a natural question comes: is it possible for us to construct a PCA variant that can incorporate the spatial prior and temporal prior.

Motivated by the above observations and other applications in computer vision [17], [18], we propose a novel spatiotemporal PCA (STPCA) framework, which, to the best of our knowledge, is the first work to integrate both the spatial prior and the temporal prior into PCA-based PM. By introducing graph Laplacian regularization terms, STPCA not only can preserve the global cause-effect relationship of process variables but also can capture the local geometric information of process samples. To derive a convergent algorithm, we then develop an alternating direction method of multipliers (ADMM) algorithm based on the symmetric Gauss-Seidel (sGS) technique. Since all the resulting subproblems admit closed-form solutions, the proposed algorithm works very

Manuscript received 18 December 2021; revised 14 February 2022 and 23 March 2022; accepted 26 April 2022. Date of publication 29 April 2022; date of current version 9 February 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 12001019, and in part by the China Postdoctoral Science Foundation under Grant 2021M702078. This brief was recommended by Associate Editor J. Liu. *(Corresponding author: Xianchao Xiu.)*

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

Wanquan Liu is with the School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou 510275, China (e-mail: liuwq63@mail.sysu.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCSII.2022.3171205.

Digital Object Identifier 10.1109/TCSII.2022.3171205

efficiently. The contributions of this brief can be summarized as follows.

1) It introduces a spatiotemporal PCA framework to better preserve the global and local process structure.
2) It designs an efficient and convergent optimization algorithm to solve the proposed model.
3) It demonstrates the effectiveness on the Tennessee Eastman (TE) benchmark process.

II. Spatiotemporal PCA

Let $\mathbf{X} \in \mathbb{R}^{n\times p}$ be the given process data matrix, where n is the number of samples and p the number of variables. A robust version of PCA [19] can be described as

$$\begin{aligned} \min_{\mathbf{A},\mathbf{E}} \ & \|\mathbf{A}\|_* + \lambda\|\mathbf{E}\|_1 \\ \text{s.t.} \ & \mathbf{X} = \mathbf{A} + \mathbf{E}, \end{aligned} \tag{1}$$

where $\|\mathbf{A}\|_*$ is the sum of all singular values of $\mathbf{A}$, $\|\mathbf{E}\|_1$ is the sum of absolute values of all elements of $\mathbf{E}$, and λ is the tuning parameter. Note that PCA has an assumption that the residual is small, independent, and Gaussian. However, when the residual follows a non-Gaussian distribution, PCA cannot work well for monitoring industrial processes. Different from PCA, problem (1) decomposes $\mathbf{X}$ into a low-rank embedding $\mathbf{A}$ and a sparse error $\mathbf{E}$. By introducing $\|\mathbf{E}\|_1$, PCA is able to filter out noise and outliers, thus enhancing the robustness. Obviously, when λ tends to zero, model (1) will reduce to the classical PCA, so it can be regarded as an extension of PCA. Although robust PCA has been comprehensively applied in computer vision, it is rarely applied in the field of PM.

For data representation, both the spatial graph prior and the temporal graph prior should be enforced into PCA to help preserve the process structure. This motivates us to construct the spatiotemporal PCA (STPCA), which can be given by the following optimization problem

$$\begin{aligned} \min_{\mathbf{A},\mathbf{E}} \ & \|\mathbf{A}\|_* + \lambda\|\mathbf{E}\|_1 + \mu_1 \mathrm{Tr}(\mathbf{A}\mathbf{L}_\mathrm{S}\mathbf{A}^\top) + \mu_2 \mathrm{Tr}(\mathbf{A}^\top\mathbf{L}_\mathrm{T}\mathbf{A}) \\ \text{s.t.} \ & \mathbf{X} = \mathbf{A} + \mathbf{E}, \end{aligned} \tag{2}$$

where $\mathbf{L}_\mathrm{S} \in \mathbb{R}^{p\times p}, \mathbf{L}_\mathrm{T} \in \mathbb{R}^{n\times n}$ are the spatial Laplacian matrix and the temporal Laplacian matrix, respectively, and μ_1, μ_2 are the regularization parameters to balance the prior information. To be specific, by encoding $\mathrm{Tr}(\mathbf{A}\mathbf{L}_\mathrm{S}\mathbf{A}^\top)$, the column structure of matrix $\mathbf{A}$ can be captured, then STPCA preserves the cause-effect relationship of process variables. Moreover, by encoding $\mathrm{Tr}(\mathbf{A}^\top\mathbf{L}_\mathrm{T}\mathbf{A})$, the row structure of matrix $\mathbf{A}$ can be captured, then STPCA preserves the geometric property of process samples. For more illustrations about the proposed framework, see Fig. 1.

To end this section, we will explain why $\mathrm{Tr}(\mathbf{A}^\top\mathbf{L}_\mathrm{T}\mathbf{A})$ can represent the temporal prior. Recall $\mathbf{L}_\mathrm{T} = \mathbf{D}_\mathrm{T} - \mathbf{W}_\mathrm{T}$, where $\mathbf{W}_\mathrm{T}$ is made up of $\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/\sigma)$ for samples $\mathbf{x}_i, \mathbf{x}_j$ and σ is the parameter to be tuned. The diagonal matrix $\mathbf{D}_\mathrm{T}$ is composed by the sum of each row of $\mathbf{W}_\mathrm{T}$. In mathematics, $\mathrm{Tr}(\mathbf{A}^\top\mathbf{L}_\mathrm{T}\mathbf{A})$ can be reformulated as

$$\begin{aligned} \mathrm{Tr}(\mathbf{A}^\top\mathbf{L}_\mathrm{T}\mathbf{A}) &= \mathrm{Tr}(\mathbf{A}^\top\mathbf{D}_\mathrm{T}\mathbf{A}) - \mathrm{Tr}(\mathbf{A}^\top\mathbf{W}_\mathrm{T}\mathbf{A}) \\ &= \sum_{i=1}^{n} \mathbf{a}_i\mathbf{a}_i^\top (d_\mathrm{T})_{ii} - \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{a}_i\mathbf{a}_j^\top (w_\mathrm{T})_{ij} \\ &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \|\mathbf{a}_i - \mathbf{a}_j\|^2 (w_\mathrm{T})_{ij}, \end{aligned} \tag{3}$$

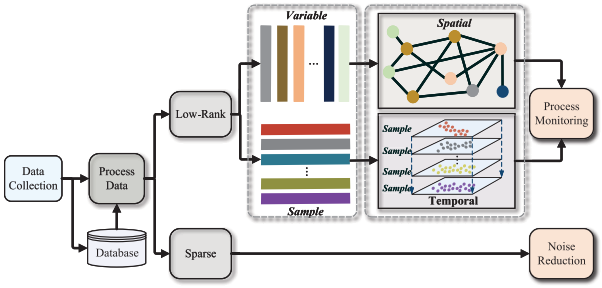

Fig. 1. Illustration of the proposed STPCA framework.

where $(d_\mathrm{T})_{ii}$ is the iith element of $\mathbf{D}_\mathrm{T}$ and $(w_\mathrm{T})_{ij}$ is the ijth element of $\mathbf{W}_\mathrm{T}$. If two samples $\mathbf{a}_i, \mathbf{a}_j$ are close in the process space, then their mappings are also close to each other in the low-dimensional manifold space.

III. Optimization Algorithm

Although there exist numerous solvers to deal with PCA-related problems [20]–[22], ADMM is one of the most popular and impressive methods. Unfortunately, variables in the objective function of problem (2) are not separable, so ADMM cannot be directly adopted here. Therefore, we first introduce two auxiliary variables $\mathbf{S}, \mathbf{T} \in \mathbb{R}^{n\times p}$, and then reformulate (2) as the following equivalent optimization problem

$$\begin{aligned} \min_{\mathbf{A},\mathbf{E},\mathbf{S},\mathbf{T}} \ & \|\mathbf{A}\|_* + \lambda\|\mathbf{E}\|_1 \\ & + \ \mu_1 \mathrm{Tr}(\mathbf{S}\mathbf{L}_\mathrm{S}\mathbf{S}^\top) + \mu_2 \mathrm{Tr}(\mathbf{T}^\top\mathbf{L}_\mathrm{T}\mathbf{T}) \\ \text{s.t.} \ & \mathbf{X} = \mathbf{A} + \mathbf{E}, \mathbf{A} = \mathbf{S}, \mathbf{A} = \mathbf{T}. \end{aligned} \tag{4}$$

The augmented Lagrangian function associated with the above problem (4) can be defined as the form of

$$\begin{aligned} &\mathcal{L}_\beta(\mathbf{A}, \mathbf{E}, \mathbf{S}, \mathbf{T}, \mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3) \\ &= \|\mathbf{A}\|_* + \lambda\|\mathbf{E}\|_1 + \mu_1 \mathrm{Tr}(\mathbf{S}\mathbf{L}_\mathrm{S}\mathbf{S}^\top) \\ &\quad + \mu_2 \mathrm{Tr}(\mathbf{T}^\top\mathbf{L}_\mathrm{T}\mathbf{T}) - \mathrm{Tr}(\mathbf{W}_1^\top(\mathbf{X} - \mathbf{A} - \mathbf{E})) \\ &\quad - \mathrm{Tr}(\mathbf{W}_2^\top(\mathbf{S} - \mathbf{A})) - \mathrm{Tr}(\mathbf{W}_3^\top(\mathbf{T} - \mathbf{A})) \\ &\quad + \frac{\beta}{2}(\|\mathbf{X} - \mathbf{A} - \mathbf{E}\|_\mathrm{F}^2 + \|\mathbf{S} - \mathbf{A}\|_\mathrm{F}^2 + \|\mathbf{T} - \mathbf{A}\|_\mathrm{F}^2), \end{aligned} \tag{5}$$

where $\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3 \in \mathbb{R}^{n\times p}$ are the Lagrangian multipliers and β is the penalty parameter.

Technically, all variables are now separable, and problem (4) can be alternately updated by minimizing the augmented Lagrangian function as $\mathbf{E} \to \mathbf{A} \to \mathbf{S} \to \mathbf{T}$. However, another critical issue arises: when there exist three or more variables, ADMM cannot guarantee global convergence. In this regard, sGS-ADMM was proposed with the help of symmetric Gauss-Seidel [21]. Different from the classical ADMM, this sGS-ADMM first performs a backward GS sweep and then performs a forward GS sweep: $\mathbf{S} \to \mathbf{T} \to \mathbf{E} \to \mathbf{A} \to \mathbf{T} \to \mathbf{S}$. Because the convexity and smoothness of problem (4) are fully exploited, the convergence can be guaranteed. Therefore, we will apply sGS-ADMM to solve the proposed STPCA in (4) and provide detailed implementations in the following.

1) *Updating* $\mathbf{S}$*:* When $\mathbf{A}$, $\mathbf{E}$, and $\mathbf{T}$ are fixed, the problem for updating $\mathbf{S}$ can be simplified to

$$\min_{\mathbf{S}} \ \mu_1 \mathrm{Tr}(\mathbf{S}\mathbf{L}_\mathrm{S}\mathbf{S}^\top) + \frac{\beta}{2}\|\mathbf{S} - \mathbf{A}^k - \mathbf{W}_2^k/\beta\|_\mathrm{F}^2. \tag{6}$$

By taking derivative, it has the following closed-form solution

$$\mathbf{S}^{k+\frac{1}{2}} = (\beta\mathbf{A}^k + \mathbf{W}_2^k)(2\mu_1\mathbf{L}_\mathrm{S} + \beta\mathbf{I}_p)^{-1}. \tag{7}$$

2) *Updating* $\mathbf{T}$*:* When $\mathbf{A}$, $\mathbf{E}$, and $\mathbf{S}$ are fixed, the problem for updating $\mathbf{T}$ can be equivalent to

$$\min_{\mathbf{T}} \ \mu_2\mathrm{Tr}(\mathbf{T}\mathbf{L}_\mathrm{T}\mathbf{T}^\top) + \frac{\beta}{2}\|\mathbf{T} - \mathbf{A}^k - \mathbf{W}_3^k/\beta\|_\mathrm{F}^2. \tag{8}$$

Similarly, the closed-form solution is given by

$$\mathbf{T}^{k+\frac{1}{2}} = (2\mu_2\mathbf{L}_\mathrm{T} + \beta\mathbf{I}_n)^{-1}(\beta\mathbf{A}^k + \mathbf{W}_3^k). \tag{9}$$

3) *Updating* $\mathbf{E}$*:* When $\mathbf{A}$, $\mathbf{S}$, and $\mathbf{T}$ are fixed, the problem for updating $\mathbf{E}$ can be reformulated as

$$\min_{\mathbf{E}} \ \lambda\|\mathbf{E}\|_1 + \frac{\beta}{2}\|\mathbf{X} - \mathbf{A}^k - \mathbf{E} - \mathbf{W}_1^k/\beta\|_\mathrm{F}^2, \tag{10}$$

which has the following closed-form solution

$$\mathbf{E}^{k+1} = \mathcal{S}_{\lambda/\beta}(\mathbf{X} - \mathbf{A}^k - \mathbf{W}_1^k/\beta). \tag{11}$$

Here, $\mathcal{S}_{\lambda/\beta}(\cdot)$ is the soft thresholding operator [19] defined as $\mathcal{S}_\tau(z) = \mathrm{sgn}(z)\max\{|z| - \tau, 0\}$.

4) *Updating* $\mathbf{A}$*:* When $\mathbf{E}$, $\mathbf{S}$, and $\mathbf{T}$ are fixed, the problem for updating $\mathbf{A}$ can be rewritten as

$$\min_{\mathbf{A}} \ \|\mathbf{A}\|_* + \frac{\beta}{2}\|\mathbf{A} - \hat{\mathbf{A}}^k\|_\mathrm{F}^2, \tag{12}$$

where $\hat{\mathbf{A}}^k = \mathbf{X} - \mathbf{E}^{k+1} + \mathbf{S}^{k+\frac{1}{2}} + \mathbf{T}^{k+\frac{1}{2}} - \mathbf{W}_1^k/\beta - \mathbf{W}_2^k/\beta - \mathbf{W}_3^k/\beta$. Then, the closed-form solution is obtained by

$$\mathbf{A}^{k+1} = \mathbf{U}\mathcal{S}_{1/\beta}(\Sigma)\mathbf{V}^\top, \tag{13}$$

where $\mathbf{U}\Sigma\mathbf{V}^\top$ is the singular value decomposition (SVD) of $\hat{\mathbf{A}}^k$; see [23] for singular value thresholding (SVT).

The detailed optimization procedure of solving the proposed STPCA are provided in Algorithm 1. It should be pointed out that $\mathbf{S}$ and $\mathbf{T}$ are updated twice symmetrically.

5) *Convergence and Complexity Analysis:* In the proposed Algorithm 1, the iterative scheme stops if k reaches up to 200, or $\max\{\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4\} \le 10^{-3}$ where

$$\frac{\|\mathbf{A}^{k+1} - \mathbf{A}^k\|_\mathrm{F}}{\|\mathbf{A}^k\|_\mathrm{F}} \le \varepsilon_1; \ \frac{\|\mathbf{E}^{k+1} - \mathbf{E}^k\|_\mathrm{F}}{\|\mathbf{E}^k\|_\mathrm{F}} \le \varepsilon_2;$$
$$\frac{\|\mathbf{S}^{k+1} - \mathbf{S}^k\|_\mathrm{F}}{\|\mathbf{S}^k\|_\mathrm{F}} \le \varepsilon_3; \ \frac{\|\mathbf{T}^{k+1} - \mathbf{T}^k\|_\mathrm{F}}{\|\mathbf{T}^k\|_\mathrm{F}} \le \varepsilon_4. \tag{14}$$

Since Algorithm 1 is a direct application of the existing sGS-ADMM framework, it is natural to conclude that the generated sequence $\{(\mathbf{A}^k, \mathbf{E}^k, \mathbf{S}^k, \mathbf{T}^k, \mathbf{W}_1^k, \mathbf{W}_2^k, \mathbf{W}_3^k)\}$ converges to the global solution of problem (2) after finite iterations. In addition, the computational complexity of each iteration of Algorithm 1 is $\mathcal{O}(2np^2 + 2n^2p + 2n^3 + 3p^3) \sim \mathcal{O}(n^3)$. Moreover, it can be further reduced because (1) the inverses of some matrices only require to be calculated once in the iterations; (2) the SVD can be efficiently computed using the rank prediction strategy as discussed in [12].

Algorithm 1 Optimization Algorithm

Input: Process $\mathbf{X} \in \mathbb{R}^{n\times p}$, parameters $\lambda, \mu_1, \mu_2, \beta$
Initialize: Compute $\mathbf{L}_\mathrm{S} \in \mathbb{R}^{p\times p}, \mathbf{L}_\mathrm{T} \in \mathbb{R}^{n\times n}$
While not converged **do**

1: **(Backward GS sweep)** Update $\mathbf{S}$ by
$$\mathbf{S}^{k+\frac{1}{2}} = (\beta\mathbf{A}^k + \mathbf{W}_2^k)(2\mu_1\mathbf{L}_\mathrm{S} + \beta\mathbf{I}_p)^{-1}$$

2: **(Backward GS sweep)** Update $\mathbf{T}$ by
$$\mathbf{T}^{k+\frac{1}{2}} = (2\mu_2\mathbf{L}_\mathrm{T} + \beta\mathbf{I}_n)^{-1}(\beta\mathbf{A}^k + \mathbf{W}_3^k)$$

3: **(Backward GS sweep)** Update $\mathbf{E}$ by
$$\mathbf{E}^{k+1} = S_{\lambda/\beta}(\mathbf{X} - \mathbf{A}^k - \mathbf{W}_1^k/\beta)$$

4: **(Backward GS sweep)** Update $\mathbf{A}$ by
$$\mathbf{A}^{k+1} = \mathbf{U}\mathcal{S}_{1/\beta}(\Sigma)\mathbf{V}^\top$$

5: **(Forward GS sweep)** Update $\mathbf{T}$ by
$$\mathbf{T}^{k+1} = (2\mu_2\mathbf{L}_\mathrm{T} + \beta\mathbf{I}_n)^{-1}(\beta\mathbf{A}^{k+1} + \mathbf{W}_3^k)$$

6: **(Forward GS sweep)** Update $\mathbf{S}$ by
$$\mathbf{S}^{k+1} = (\beta\mathbf{A}^{k+1} + \mathbf{W}_2^k)(2\mu_1\mathbf{L}_\mathrm{S} + \beta\mathbf{I}_p)^{-1}$$

7: Update Lagrangian multipliers $\mathbf{W}_1^{k+1}, \mathbf{W}_2^{k+1}, \mathbf{W}_3^{k+1}$ by
$$\mathbf{W}_1^{k+1} = \mathbf{W}_1^k - \beta(\mathbf{X} - \mathbf{A}^{k+1} - \mathbf{E}^{k+1})$$
$$\mathbf{W}_2^{k+1} = \mathbf{W}_2^k - \beta(\mathbf{S}^{k+1} - \mathbf{A}^{k+1})$$
$$\mathbf{W}_3^{k+1} = \mathbf{W}_3^k - \beta(\mathbf{T}^{k+1} - \mathbf{A}^{k+1})$$

8: Check convergence
End while
Output: $\mathbf{A}, \mathbf{E}$

IV. Simulation Studies

This section will demonstrate the effectiveness and efficacy of the proposed STPCA over PCA, SPCA (spatial PCA), and TPCA (temporal PCA), on the TE process. In fact, SPCA can include these approaches [14]–[16] in PM, and TPCA can include these ideas [17], [18] in computer vision.

- **PCA**: $\mu_1 = 0, \ \mu_2 = 0$;
- **SPCA**: $\mu_1 \neq 0, \ \mu_2 = 0$;
- **TPCA**: $\mu_1 = 0, \ \mu_2 \neq 0$;
- **STPCA**: $\mu_1 \neq 0, \ \mu_2 \neq 0$.

A. Data Preparation

The TE process is a benchmark database, which has been broadly used to test different PM approaches; see Fig. 2 for the detailed flowchart, and more illustrations can be found in [24]. In the TE process, a total of 21 fault data sets are collected, each of which has a modeling set with 480 samples and a monitoring set with 960 samples. Moreover, a fault has been introduced to the monitoring set at the 161st sample. Each data set has 52 variables, however in this brief we remove 19 analysis variables sampled less frequently and only choose the other 33 variables. Since faults IDV(3), IDV(9), IDV(15), and IDV(21) are relatively difficult for data-driven PM monitoring approaches, thus these faults are not considered here.

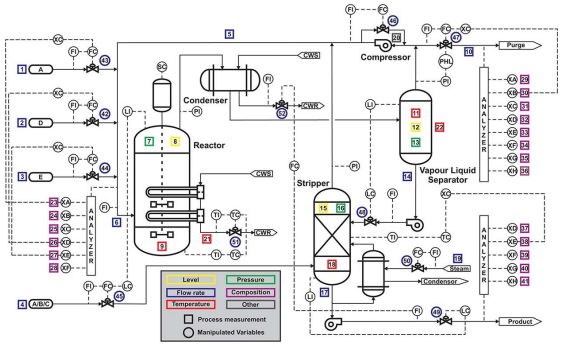

Fig. 2. A Flowchart of the TE process provided by [24].

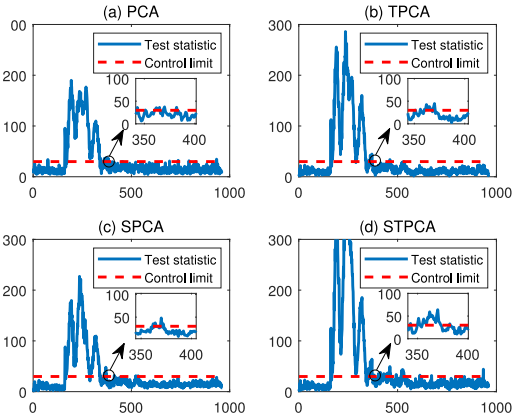

Fig. 3. Monitoring performance for fault IDV(5) in the TE process: (a) PCA, (b) SPCA, (c) TPCA, and (d) STPCA.

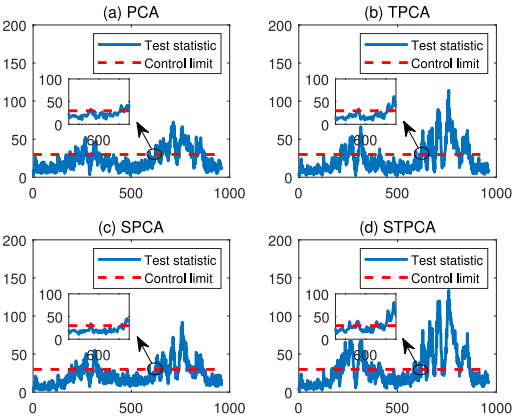

Fig. 4. Monitoring performance for fault IDV(10) in the TE process: (a) PCA, (b) SPCA, (c) TPCA, and (d) STPCA.

It is admitted that choosing a proper regularization parameter is not a trivial task. For the compared SPCA, TPCA, and STPCA, parameters λ, μ_1, μ_2 are selected using five-fold cross-validation and then fixed in the optimization procedure. To measure the monitoring performance, two probability indicators, i.e., fault detection rate (FDR) and false alarm rate (FAR), are defined as FDR $=$ prob($\mathrm{T}^2 > J_{th,\mathrm{T}^2} \mid f \neq 0$) and FAR $=$ prob($\mathrm{T}^2 > J_{th,\mathrm{T}^2} \mid f = 0$). A higher FDR value or lower FAR value implies better monitoring performance. The FDR value is 100% if all faulty samples are detected, while the FAR value is 0% if all fault-free samples are not alarmed.

B. Monitoring Strategy

Following a similar line as the arguments in [1], the whole monitoring strategy contains offline modeling and online monitoring, which can be summarized as follows.

1) Data Normalization: To begin with, the modeling data $\mathbf{X}$ and monitoring data $\mathbf{Y}$ should be normalized to eliminate the influence of measurement units.

2) Noise Reduction: After solving the proposed STPCA in (2), the modeling data $\mathbf{X}$ is decomposed into a low-rank component $\mathbf{A}$ plus a sparse component $\mathbf{E}$, where $\mathbf{A}$ reflects the clean process information and $\mathbf{E}$ reflects the random noise. Unlike the classical PCA, SVD is performed on matrix $\mathbf{A}$ as

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^\top, \tag{15}$$

where Σ is the singular matrix and $\mathbf{V}$ is the loading matrix.

3) Control Limit Determination: The corresponding control limit for T^2 statistic can be estimated by

$$J_{th,\mathrm{T}^2} = \frac{r(n^2-1)}{n(n-1)} F_\alpha(r, n-r), \tag{16}$$

where the F distribution depends on the degrees of freedom $r, n-r$ and the significance level α.

4) Test Statistic Calculation: Denote $\mathbf{W} = \Sigma^\top\Sigma$ be the variance matrix of PCs, then the T^2 statistic can be applied to monitor the process, which is defined as

$$\mathrm{T}_i^2 = \mathbf{y}_i\mathbf{V}\mathbf{W}^{-1}\mathbf{V}^\top\mathbf{y}_i^\top, \tag{17}$$

where $\mathbf{y}_i$ is the ith sample of monitoring data $\mathbf{Y}$.

5) Online Monitoring: Once the control limit is determined, the detection logic can be checked by

$$\begin{cases} \mathrm{T}^2 > J_{th,\mathrm{T}^2} & \Rightarrow \text{ faulty}, \\ \mathrm{T}^2 \leq J_{th,\mathrm{T}^2} & \Rightarrow \text{ fault-free}. \end{cases} \tag{18}$$

C. Monitoring Performance

The monitoring performance for faults IDV(5) and IDV(10) are shown in Fig. 3 and Fig. 4, respectively. For fault IDV(5), the fault can be detected at the 161st sample but cannot be detected continuously. In comparison, the proposed STPCA can still achieve outstanding monitoring results; see the numbers that violated the control limit between 350-400 samples. For fault IDV(10), all mentioned PCA-based approaches fail to detect it at the 161st sample. However, after a few samples, the fault can be detected to a certain extent, but the monitoring performance is unstable. In contrast, STPCA can achieve better monitoring results; see samples around 600 in Fig. 4. Actually, this superiority can be attributed to the spatiotemporal prior, which has strong potential in PM.

Table I reports the achieved FDR and FAR values for the selected faults in the TE process. Moreover, the monitoring results of the proposed STPCA are indicated in bold. It can be easily concluded that faults IDV(1), IDV(2), IDV(7), IDV(8), IDV(12), IDV(13), and IDV(14) are relatively easy to detect for all approaches since their FDR values are more than 90%. In particular, the FDR value of fault IDV(7) is 100%, which means that this fault can be successfully detected. Compared with the classical PCA, all its variants, i.e., SPCA, TPCA, and STPCA, obtain larger FDR values, which illustrates that the graph prior is beneficial for PM as it preserves the latent structure of processes. Furthermore, the performance of the proposed STPCA is better than or as good as others. Even for

TABLE I
FDR(%) AND FAR(%) FOR THE SELECTED FAULTS IN THE TE PROCESS

Fault No.	PCA		SPCA		TPCA		STPCA	
	FDR	FAR	FDR	FAR	FDR	FAR	FDR	FAR
IDV(1)	99.13	0.63	99.25	0.63	99.13	0.00	**99.25**	**0.00**
IDV(2)	98.38	0.63	98.38	0.63	98.38	0.63	**98.38**	**0.63**
IDV(4)	20.88	1.25	27.63	0.63	29.75	0.63	**37.63**	**0.00**
IDV(5)	24.13	1.88	26.75	0.63	25.50	0.63	**32.50**	**0.00**
IDV(6)	99.13	1.25	99.13	1.25	99.13	0.63	**99.13**	**0.63**
IDV(7)	100	1.25	100	0.63	100	0.63	**100**	**0.00**
IDV(8)	96.88	0.63	96.88	0.63	97.50	0.63	**97.50**	**0.00**
IDV(10)	29.63	0.63	32.63	0.00	33.75	0.00	**37.88**	**0.00**
IDV(11)	40.63	2.50	43.75	1.88	42.88	1.25	**49.63**	**1.25**
IDV(12)	98.38	1.25	98.75	0.63	99.13	0.63	**99.50**	**0.63**
IDV(13)	93.63	0.00	93.63	0.00	93.63	0.00	**93.63**	**0.00**
IDV(14)	99.25	1.25	99.25	0.63	99.50	0.63	**99.75**	**0.63**
IDV(16)	13.50	1.88	14.13	1.25	15.25	1.88	**16.50**	**1.25**
IDV(17)	76.25	2.50	77.50	1.88	77.13	1.25	**83.13**	**0.63**
IDV(18)	89.25	2.50	89.88	1.25	89.88	1.88	**90.25**	**1.25**
IDV(19)	14.13	0.63	16.75	0.00	1563	0.63	**17.50**	**0.00**
IDV(20)	31.75	1.25	37.63	1.25	36.50	0.63	**41.63**	**0.63**

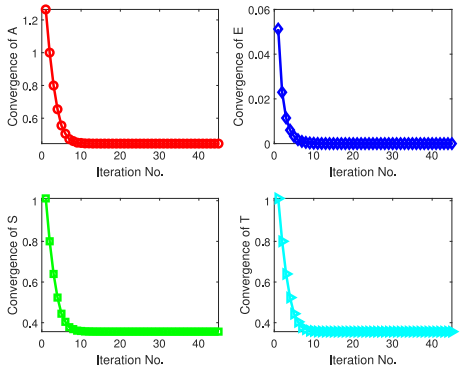

Fig. 5. Illustration of convergence for relative differences in (14).

some difficult situations, such as fault IDV(20), the increase is still positive, which validates that both spatial and temporal graph priors are helpful for monitoring processes. The reason behind is that the local geometric manifold structure can be captured and preserved, thereby making the monitoring more robust to corruptions than SPCA and TPCA.

D. Convergence Verification

Fig. 5 shows the relative differences in (14) of variables for fault IDV(10). It can be found that with the increase of iteration number, the relative differences decrease dramatically, which verifies the convergence of the proposed algorithm.

V. CONCLUSION

In this brief, we have proposed and studied a novel spatiotemporal PCA framework, in which the spatial Laplacian is added to preserve the cause-effect relationship of process variables and the temporal Laplacian is added to maintain the geometric structure of process samples. Further, an efficient optimization algorithm based on the sGS-ADMM has been developed and analyzed in detail. Numerical studies on the TE process have suggested that the proposed framework is able to improve the monitoring performance significantly in comparison with the standard PCA.

In the future, we will be devoted to developing fast solvers, including unrolled learning methods [25] and online optimization techniques, and integrating the spatiotemporal prior with other data-driven PM approaches.

REFERENCES

[1] S. Ding, *Data-Driven Design of Fault Diagnosis and Fault-Tolerant Control Systems*. London, U.K.: Springer-Verlag, 2014.
[2] M. Alibakhshikenari *et al.*, "Dual-polarized highly folded bowtie antenna with slotted self-grounded structure for sub-6 GHz 5G applications," *IEEE Trans. Antennas Propag.*, vol. 70, no. 4, pp. 3028–3033, Apr. 2022.
[3] I. Nadeem *et al.*, "A comprehensive survey on 'circular polarized antennas' for existing and emerging wireless communication technologies," *J. Phys. D, Appl. Phys.*, vol. 55, Jan. 2022, Art. no. 33002.
[4] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3757–3767, Jun. 2015.
[5] 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.
[6] L. Li, S. X. Ding, Y. Na, and J. Qiu, "An asynchronized observer based fault detection approach for uncertain switching systems with mode estimation," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 2, pp. 514–518, Feb. 2022.
[7] K. Ren, Z. Zhang, and C. Xia, "Event-based fault diagnosis of networked discrete event systems," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 3, pp. 1787–1791, Mar. 2022.
[8] J.-M. Lee, C. Yoo, and I.-B. Lee, "Fault detection of batch processes using multiway kernel principal component analysis," *Comput. Chem. Eng.*, vol. 28, no. 9, pp. 1837–1847, 2004.
[9] J. Harmouche, C. Delpha, and D. Diallo, "Incipient fault detection and diagnosis based on Kullback–Leibler divergence using principal component analysis: Part I," *Signal Process.*, vol. 94, pp. 278–287, Jan. 2014.
[10] Y. Liu, G. Zhang, and B. Xu, "Compressive sparse principal component analysis for process supervisory monitoring and fault detection," *J. Process Control*, vol. 50, pp. 1–10, Feb. 2017.
[11] X. Chang, F. Nie, S. Wang, Y. Yang, X. Zhou, and C. Zhang, "Compound rank-k projections for bilinear analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 7, pp. 1502–1513, Jul. 2016.
[12] M. Luo, F. Nie, X. Chang, Y. Yang, A. G. Hauptmann, and Q. Zheng, "Adaptive unsupervised feature selection with structure regularization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 944–956, Apr. 2018.
[13] M. Luo, X. Chang, L. Nie, Y. Yang, A. G. Hauptmann, and Q. Zheng, "An adaptive semisupervised feature analysis for video semantic recognition," *IEEE Trans. Cybern.*, vol. 48, no. 2, pp. 648–660, Feb. 2018.
[14] 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.
[15] X. Xiu, Y. Yang, L. Kong, and W. Liu, "Laplacian regularized robust principal component analysis for process monitoring," *J. Process Control*, vol. 92, pp. 212–219, Aug. 2020.
[16] R. Zhai, J. Zeng, and Z. Ge, "Structured principal component analysis model with variable correlation constraint," *IEEE Trans. Control Syst. Technol.*, vol. 30, no. 2, pp. 558–569, Mar. 2022.
[17] S. Li, K. Li, and Y. Fu, "Temporal subspace clustering for human motion segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 4453–4461.
[18] H. Cai, V. W. Zheng, and K. C.-C. Chang, "A comprehensive survey of graph embedding: Problems, techniques, and applications," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 9, pp. 1616–1637, Sep. 2018.
[19] E. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1–37, 2011.
[20] K. Kavaklı, H. Urey, and K. Aksit, "Learned holographic light transport: Invited," *Appl. Opt.*, vol. 61, no. 5, pp. B50–B55, 2022.
[21] L. Chen, D. Sun, and K.-C. Toh, "An efficient inexact symmetric Gauss–Seidel based majorized ADMM for high-dimensional convex composite conic programming," *Math. Program.*, vol. 161, nos. 1–2, pp. 237–270, 2017.
[22] D. Han, "A survey on some recent developments of alternating direction method of multipliers," *J. Oper. Res. Soc. China*, vol. 10, pp. 1–52, Jan. 2022.
[23] J. Cai, E. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
[24] 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.
[25] F. S. Oktem, O. F. Kar, C. D. Bezek, and F. Kamalabadi, "High-resolution multi-spectral imaging with diffractive lenses and learned reconstruction," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 489–504, 2021, doi: 10.1109/TCI.2021.3075349.