Efficient and Robust Process Monitoring Using Structured Low-Rank Representation

Zhonghua Miao and Xianchao Xiu, *Member, IEEE*

***Abstract*—In this brief, a new multivariate statistical method based on structured low-rank representation (SLR) is proposed to detect minor faults for industrial process monitoring (PM). The core idea of the proposed SLR is to enhance the representation of minor faults by using the $\ell_{2,0}$-norm, and improve the robustness to noise by introducing a regularization term. Further, a learnable manifold constraint is incorporated to preserve the cause-effect relationship between monitoring variables. More importantly, a distributed optimization algorithm is developed with convergence analysis. Simulation examples are conducted to demonstrate the effectiveness and robustness of the proposed PM method.**

***Index Terms*—Process monitoring (PM), robust optimization, structured low-rank representation (SLR), $\ell_{2,0}$-norm.**

I. Introduction

PROCESS monitoring (PM) plays an extremely important role in guaranteeing safe production of modern industries. Due to the development of sensor technology and computing capabilities, collecting and analyzing the measured data have become easier, which makes data-driven PM methods receive considerable attention and gain significant progress over the past few decades [1]–[3]. However, there often exist a certain amount of corruptions, such as minor faults and noise, because sensor failures, software malfunctions, and recording mistakes occur in practical industrial processes [4]–[6]. This may trigger an inaccurate data-driven PM modeling framework and even derive the unreliable test statistic and control limit, leading to a false monitoring conclusion. Therefore, how to effectively deal with minor faults and noise has developed into a hot research topic of process control [7].

To improve the monitoring performance, many researchers have made plenty of contributions related to the data-driven PM modeling and monitoring strategies. From the perspective of multivariate statistical analysis, the following two categories are impressive. The first category is based on robust principal component analysis (PCA) [8]. For example, Yan *et al.* [9] extended robust PCA to the application of PM and proposed an online monitoring procedure. It has been verified to perform as well as PCA when the process data is noise-free and much better in the presence of noise. Pan *et al.* [10] introduced two structured sparse regularization terms into the objective of robust PCA, and thus successfully detected the involved minor faults. After that, Xiu *et al.* [11] suggested another robust PCA method by integrating a graph Laplacian constraint to reduce the negative influences of noise and preserve the relationship of process variables. The second category is related to a new dimension reduction technique called low-rank representation (LRR) [12]. In fact, LRR has been extensively used in social networks, document clustering, and face recognition, however, it has not been fully investigated in PM [13]. For example, Guo *et al.* [14] considered a spatiotemporal LRR variant for characterizing minor faults. The monitoring results show that LRR not only can exhibit the robustness consistent with robust PCA but also can reveal the global geometric structure of process data. Later, Yu and Zhao [15] constructed a generalized LRR-based PM framework with a total of three monitoring statistics. Very recently, Fu *et al.* [16] used an explicit projection to map the given data onto a low-dimensional space, thereby improving the representation of the global and local geometric structure.

Although the aforementioned LRR-based PM methods have achieved great success, there exist two shortcomings that may degrade the modeling performance. On one hand, minor faults often occur over a period of time, which brings disturbances in some specific variables. It is indicated that minor faults can be characterized by column-sparsity, i.e., $\ell_{2,1}$-norm (sum of ℓ_2-norm of all columns) [10]. From an optimization point of view, $\ell_{2,0}$-norm (number of nonzero columns) is the original description of column-sparsity and has a better ability to improve the interpretability and reliability of variable selection; see, e.g., [17], [18], for applications in circuits and systems. On the other hand, the proposed algorithm in [16] is based on the framework of alternating direction method of multipliers (ADMM), but the convergence property is not discussed. Actually, when it comes to an ADMM-based algorithm with more than three variables, there exists no theoretical guarantee of convergence. More illustrations can be found in [19], [20].

Motivated by the above observations, in this brief, we propose an efficient and robust LRR-based PM method, called structured low-rank representation (SLR), to filter out minor faults and noise of industrial processes. Specifically, the goals of the current work can be summarized as follows.

1) To construct a robust multivariate statistical PM method by filtering out minor faults.
2) To enhance the representation of process variables using nonconvex optimization.
3) To preserve the local geometric prior via incorporating manifold learning.

Manuscript received 12 March 2022; accepted 28 March 2022. Date of publication 31 March 2022; date of current version 1 August 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 12001019. This brief was recommended by Associate Editor J. Wang. *(Corresponding author: Xianchao Xiu.)*

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

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCSII.2022.3163734.

Digital Object Identifier 10.1109/TCSII.2022.3163734

1549-7747 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: SHANGHAI UNIVERSITY. Downloaded on April 17,2023 at 02:32:15 UTC from IEEE Xplore. Restrictions apply.

4) To develop a fast iterative algorithm along with detailed convergence guarantee.

II. Problem Formulation

Consider the process data $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p] \in \mathbb{R}^{n\times p}$, which has n samples and p variables. The classical LRR aims to explore a low-rank matrix $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_p] \in \mathbb{R}^{p\times p}$ to linearly approximate the low-dimensional subspace information embedded in the data [12]. Mathematically, LRR can be defined as the following optimization problem

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{E}} \ & \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_{2,1} \\ \text{s.t.} \ & \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}, \end{aligned} \tag{1}$$

where $\|\mathbf{Z}\|_*$ is the nuclear norm, which is defined as the sum of all singular values, $\|\mathbf{E}\|_{2,1}$ is the $\ell_{2,1}$-norm defined before, and $\lambda > 0$ is the regularization parameter. Compared with the classical PCA, this LRR model can control the magnitude of noise by choosing different λ, thus making the monitoring performance more robust.

In fact, the success of LRR mainly relies on the assumption that the process data can be approximated by low-dimensional subspaces, allowing the noise $\mathbf{E}$ to be randomly distributed. However, if there occur some minor faults in the process, the modeling performance degrades seriously since LRR cannot distinguish between minor faults and noise. To overcome this shortcoming, we construct the following structured low-rank representation (SLR) as the form of

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{E},\mathbf{N}} \ & \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_{2,0} + \mu\|\mathbf{N}\|_{\mathrm{F}}^2 \\ \text{s.t.} \ & \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E} + \mathbf{N}, \\ & \mathbf{Z} \in \mathcal{M}, \end{aligned} \tag{2}$$

where $\mathcal{M}$ is the manifold constraint, which can be learned from the process data [21]. Although similar model structures have been considered before, the main differences are

1) The $\ell_{2,0}$-norm is able to better characterize the minor faults that occur in the process, while the existing $\ell_{2,1}$-norm regularization technique in [10], [16] may generate some residuals in the low-rank matrix.
2) The matrix $\mathbf{N}$ is introduced to describe the dense Gaussian noise in the process data $\mathbf{X}$, and $\|\mathbf{N}\|_{\mathrm{F}}^2$ is adopted for denoising in the optimization with parameter $\mu > 0$.
3) The learnable manifold $\mathcal{M}$ is integrated to preserve the latent relationship of process variables. In contrast, only a specific manifold is employed in [16].

III. Optimization Algorithm

Obviously, problem (2) is nonconvex even NP-hard, which makes it difficult to obtain the optimal solution. Although there exist some solvers, no convergence result can be guaranteed. By the spirit of [19], [20], an iterative optimization algorithm can be applied by minimizing one variable with other variables fixed. To begin with, the augmented Lagrangian function of problem (2) is formulated as

$$\begin{aligned} \mathcal{L}_\beta(\mathbf{Z},\mathbf{E},\mathbf{N},\mathbf{Y}) &= \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_{2,0} + \mu\|\mathbf{N}\|_{\mathrm{F}}^2 \\ &\quad -\langle \mathbf{Y}, \mathbf{X}\mathbf{Z} + \mathbf{E} + \mathbf{N} - \mathbf{X}\rangle + \frac{\beta}{2}\|\mathbf{X}\mathbf{Z} + \mathbf{E} + \mathbf{N} - \mathbf{X}\|_{\mathrm{F}}^2, \end{aligned} \tag{3}$$

Algorithm 1 Optimization Algorithm

Input: Data $\mathbf{X}$, parameters λ, μ, β.
Output: $\mathbf{Z}^{k+1}$.
While not converged **do**

$$\begin{cases} \mathbf{Z}^{k+1} = \underset{\mathbf{Z}\in\mathcal{M}}{\operatorname{argmin}} \ \mathcal{L}_\beta(\mathbf{Z}, \mathbf{E}^k, \mathbf{N}^k, \mathbf{Y}^k); & \text{(4a)} \\ \mathbf{E}^{k+1} = \underset{\mathbf{E}}{\operatorname{argmin}} \ \mathcal{L}_\beta(\mathbf{Z}^{k+1}, \mathbf{E}, \mathbf{N}^k, \mathbf{Y}^k); & \text{(4b)} \\ \mathbf{N}^{k+1} = \underset{\mathbf{N}}{\operatorname{argmin}} \ \mathcal{L}_\beta(\mathbf{Z}^{k+1}, \mathbf{E}^{k+1}, \mathbf{N}, \mathbf{Y}^k); & \text{(4c)} \\ \mathbf{Y}^{k+1} = \mathbf{Y}^k - \beta(\mathbf{X}\mathbf{Z}^{k+1} + \mathbf{E}^{k+1} + \mathbf{N}^{k+1} - \mathbf{X}). & \text{(4d)} \end{cases}$$

End while

where $\mathbf{Y}$ is the Lagrange multiplier and $\beta > 0$ is the penalty parameter. According to the ADMM, the iterative scheme of solving problem (2) are provided in Algorithm 1.

A. Update $\mathbf{Z}$

The subproblem (4a) can be equivalently simplified into the following minimization

$$\min_{\mathbf{Z}\in\mathcal{M}} \ \underbrace{\frac{\beta}{2}\|\mathbf{X}\mathbf{Z} + \mathbf{E}^k + \mathbf{N}^k - \mathbf{X} - \mathbf{Y}^k/\beta\|_{\mathrm{F}}^2}_{g(\mathbf{Z})} + \|\mathbf{Z}\|_*. \tag{5}$$

Although the objective is convex, it does not admit a closed-form solution since $g(\mathbf{Z})$ is nonlinear in terms of $\mathbf{Z}$. Consequently, the linearized technique will be applied, which derives

$$\min_{\mathbf{Z}\in\mathcal{M}} \ \langle \nabla_{\mathbf{Z}} g(\mathbf{Z}^k), \mathbf{Z} - \mathbf{Z}^k\rangle + \frac{\eta}{2}\|\mathbf{Z} - \mathbf{Z}^k\|_{\mathrm{F}}^2 + \|\mathbf{Z}\|_*. \tag{6}$$

Here, $\nabla_{\mathbf{Z}} g(\mathbf{Z}^k)$ is the gradient of $g(\mathbf{Z})$ at $\mathbf{Z}^k$ and $\|\mathbf{Z} - \mathbf{Z}^k\|_{\mathrm{F}}^2$ is the proximal term with parameter $\eta > 0$. Thus, the closed-form solution can be estimated by

$$\mathbf{Z}^{k+\frac{1}{2}} = \Theta_{1/\eta}(\mathbf{Z}^k - \nabla_{\mathbf{Z}} g(\mathbf{Z}^k)/\eta), \tag{7}$$

and

$$\mathbf{Z}^{k+1} = \Pi_{\mathcal{M}}(\mathbf{Z}^{k+\frac{1}{2}}), \tag{8}$$

where $\Theta_{1/\eta}(\cdot)$ is the singular value thresholding (SVT) [22], which is used to deal with nuclear-norm related problems, and $\Pi_{\mathcal{M}}(\cdot)$ is the projection onto manifold $\mathcal{M}$, which can be easily computed by optimization techniques.

B. Update $\mathbf{E}$

After some basic algebraic manipulations, the subproblem (4b) can be transformed into

$$\min_{\mathbf{E}} \ \frac{\beta}{2}\|\mathbf{X}\mathbf{Z}^{k+1} + \mathbf{E} + \mathbf{N}^k - \mathbf{X} - \mathbf{Y}^k/\beta\|_{\mathrm{F}}^2 + \lambda\|\mathbf{E}\|_{2,0}. \tag{9}$$

Different from the existing $\ell_{2,1}$-norm minimization problems, the above model seems more difficult since $\ell_{2,0}$-norm is nonconvex. According to [13], the generalized hard thresholding operator can be adopted. Denote

$$\mathbf{E}^{k+\frac{1}{2}} = \mathbf{X} - \mathbf{X}\mathbf{Z}^{k+1} - \mathbf{N}^k + \mathbf{Y}^k/\beta, \tag{10}$$

then the closed-form solution can be described as

$$\mathbf{e}_i^{k+1} = \begin{cases} \mathbf{e}_i^{k+\frac{1}{2}}, & \text{if } \|\mathbf{e}_i^{k+\frac{1}{2}}\|_2 \geq \sqrt{2\lambda/\beta}, \\ \mathbf{0}, & \text{otherwise}, \end{cases} \tag{11}$$

Authorized licensed use limited to: SHANGHAI UNIVERSITY. Downloaded on April 17,2023 at 02:32:15 UTC from IEEE Xplore. Restrictions apply.

where $\mathbf{e}_i^{k+\frac{1}{2}}$ and $\mathbf{e}_i^{k+1}$ are the ith columns of the corresponding matrices $\mathbf{E}^{k+\frac{1}{2}}$ and $\mathbf{E}^{k+1}$, respectively. Therefore, the updated $\mathbf{E}^{k+1}$ is derived.

C. Update N

Once $\mathbf{Z}$, $\mathbf{E}$, and $\mathbf{Y}$ are determined, the subproblem (4c) can be optimized by

$$\min_{\mathbf{N}} \ \frac{\beta}{2}\|\mathbf{X}\mathbf{Z}^{k+1}+\mathbf{E}^{k+1}+\mathbf{N}-\mathbf{X}-\mathbf{Y}^k/\beta\|_{\mathrm{F}}^2+\mu\|\mathbf{N}\|_{\mathrm{F}}^2, \quad (12)$$

which has the following first-order optimality condition

$$0=\beta(\mathbf{X}\mathbf{Z}^{k+1}+\mathbf{E}^{k+1}+\mathbf{N}^{k+1}-\mathbf{X}-\mathbf{Y}^k/\beta)+2\mu\mathbf{N}^{k+1}. \quad (13)$$

Thus, the closed-form solution can be given by

$$\mathbf{N}^{k+1}=\frac{\beta}{2\mu+\beta}(\mathbf{X}-\mathbf{X}\mathbf{Z}^{k+1}-\mathbf{E}^{k+1}+\mathbf{Y}^k/\beta). \quad (14)$$

Remark 1: It is admitted that although the ADMM can be used to solve problem (2), a global optimal solution is relatively difficult to be obtained since the proposed algorithm involves the hard thresholding of $\|\mathbf{E}\|_{2,0}$ and the manifold projection of $\mathbf{Z} \in \mathcal{M}$ [19]. Therefore, in this brief, a local minimizer (stationary point to be precise) will be defined and pursued, as established in the next subsection.

D. Convergence Analysis

In order to establish the convergence property, the definition of a stationary point [23] of problem (2) is presented as

$$\begin{cases} 0\in\partial\|\mathbf{Z}^*\|_*-\mathbf{X}^\top\mathbf{Y}^*, \\ 0\in\lambda\partial\|\mathbf{E}^*\|_{2,0}-\mathbf{Y}^*, \\ 0=2\mu\mathbf{N}^*-\mathbf{Y}^*, \\ 0=\mathbf{X}\mathbf{Z}^*+\mathbf{E}^*+\mathbf{N}^*-\mathbf{X}. \end{cases} \quad (15)$$

Now, the convergence of the proposed Algorithm 1 will be discussed by the following theorem.

Theorem 1: Suppose that $\{(\mathbf{Z}^k,\mathbf{E}^k,\mathbf{N}^k,\mathbf{Y}^k)\}$ is a sequence generated by Algorithm 1. Then the sequence converges to a stationary point of problem (2).

Proof: First of all, let's consider

$$\begin{aligned} &\mathcal{L}_\beta(\mathbf{Z}^{k+1},\mathbf{E}^{k+1},\mathbf{F}^{k+1},\mathbf{Y}^{k+1}) \\ &\quad-\mathcal{L}_\beta(\mathbf{Z}^{k+1},\mathbf{E}^{k+1},\mathbf{F}^{k+1},\mathbf{Y}^k) \\ &=-\langle\mathbf{Y}^{k+1}-\mathbf{Y}^k,\mathbf{X}\mathbf{Z}^{k+1}+\mathbf{E}^{k+1}+\mathbf{N}^{k+1}-\mathbf{X}\rangle \\ &=\frac{1}{\beta}\|\mathbf{Y}^{k+1}-\mathbf{Y}^k\|_{\mathrm{F}}^2, \end{aligned} \quad (16)$$

where the last equality follows from (4d). This, together with relation (4c), derives that

$$0=2\mu\mathbf{N}^{k+1}+\mathbf{Y}^{k+1}. \quad (17)$$

Obviously, it can be concluded that

$$\|\mathbf{Y}^{k+1}-\mathbf{Y}^k\|_{\mathrm{F}}^2=4\mu^2\|\mathbf{N}^{k+1}-\mathbf{N}^k\|_{\mathrm{F}}^2, \quad (18)$$

which means that the dual variable $\mathbf{Y}$ can be bounded by the primal variable $\mathbf{N}$.

Next, note that $\mathcal{L}_\beta(\mathbf{Z}^{k+1},\mathbf{E}^{k+1},\mathbf{N},\mathbf{Y}^k)$ is strongly convex with modulus at least $2\mu+\beta$. Recall that $\mathbf{N}^{k+1}$ is a minimizer of $\mathbf{N}$-subproblem, it implies

$$\begin{aligned} &\mathcal{L}_\beta(\mathbf{Z}^{k+1},\mathbf{E}^{k+1},\mathbf{N}^{k+1},\mathbf{Y}^k)-\mathcal{L}_\beta(\mathbf{Z}^{k+1},\mathbf{E}^{k+1},\mathbf{N}^k,\mathbf{Y}^k) \\ &\le-\frac{2\mu+\beta}{2}\|\mathbf{N}^{k+1}-\mathbf{N}^k\|_{\mathrm{F}}^2. \end{aligned} \quad (19)$$

Moreover, since $\mathbf{E}^{k+1}$ is a local minimizer of $\mathbf{E}$-subproblem, i.e., $\mathcal{L}_\beta(\mathbf{Z}^{k+1},\mathbf{E},\mathbf{N}^k,\mathbf{Y}^k)$, it has

$$\mathcal{L}_\beta(\mathbf{Z}^{k+1},\mathbf{E}^{k+1},\mathbf{N}^k,\mathbf{Y}^k)-\mathcal{L}_\beta(\mathbf{Z}^{k+1},\mathbf{E}^k,\mathbf{N}^k,\mathbf{Y}^k)\le 0. \quad (20)$$

Similarly, $\mathbf{Z}^{k+1}$ is a local minimizer of $\mathcal{L}_\beta(\mathbf{Z},\mathbf{E}^k,\mathbf{N}^k,\mathbf{Y}^k)$, then it achieves

$$\mathcal{L}_\beta(\mathbf{Z}^{k+1},\mathbf{E}^k,\mathbf{N}^k,\mathbf{Y}^k)-\mathcal{L}_\beta(\mathbf{Z}^k,\mathbf{E}^k,\mathbf{N}^k,\mathbf{Y}^k)\le 0. \quad (21)$$

Thus, summing (17), (19), (20), (21) and involving (18), it is obtained that

$$\begin{aligned} &\mathcal{L}_\beta(\mathbf{Z}^{k+1},\mathbf{E}^{k+1},\mathbf{N}^{k+1},\mathbf{Y}^{k+1})-\mathcal{L}_\beta(\mathbf{Z}^k,\mathbf{E}^k,\mathbf{N}^k,\mathbf{Y}^k) \\ &\le-C\|\mathbf{N}^{k+1}-\mathbf{N}^k\|_{\mathrm{F}}^2, \end{aligned} \quad (22)$$

where $C=\frac{2\mu+\beta}{2}-\frac{4\mu^2}{\beta}$. It is obviously to see that if $\beta\ge 2\mu$, then $C\ge 0$ and thus the corresponding augmented Lagrangian sequence $\{\mathcal{L}_\beta(\mathbf{Z}^k,\mathbf{E}^k,\mathbf{N}^k,\mathbf{Y}^k)\}_{k=1}^\infty$ is monotone decreasing.

Denote $\mathcal{L}_\beta(\mathbf{Z}^1,\mathbf{E}^1,\mathbf{N}^1,\mathbf{Y}^1)$ be the upper bound. For $k\ge 1$, it has the following relation

$$\begin{aligned} \mathcal{L}_\beta(\mathbf{Z}^1,\mathbf{E}^1,\mathbf{N}^1,\mathbf{Y}^1)\ge{}&\|\mathbf{Z}^k\|_*+\lambda\|\mathbf{E}^k\|_{2,0}+(\mu-2\mu^2/\beta)\|\mathbf{N}^k\|_{\mathrm{F}}^2 \\ &+\frac{\beta}{2}\|\mathbf{X}\mathbf{Z}^k+\mathbf{E}+\mathbf{N}^k-\mathbf{X}-\mathbf{Y}^k/\beta\|_{\mathrm{F}}^2. \end{aligned} \quad (23)$$

In fact, if $\beta\ge 2/\mu$, then the constant $\mu-2\mu^2/\beta$ is positive. Hence, variables $\{\mathbf{Z}^k\}$, $\{\mathbf{E}^k\}$, and $\{\mathbf{N}^k\}$ are all bounded, and the variable $\{\mathbf{Y}^k\}$ is also bounded from (17). Therefore, the generated sequence $\{(\mathbf{Z}^k,\mathbf{E}^k,\mathbf{N}^k,\mathbf{Y}^k)\}_{k=1}^\infty$ is bounded.

Suppose that $\{(\mathbf{Z}^{k_i},\mathbf{E}^{k_i},\mathbf{N}^{k_i},\mathbf{Y}^{k_i})\}$ is a convergent subsequence and $(\mathbf{Z}^*,\mathbf{E}^*,\mathbf{N}^*,\mathbf{Y}^*)$ is a cluster point [23] of the subsequence. Then, it should satisfy

$$\lim_{i\to\infty}(\mathbf{Z}^{k_i},\mathbf{E}^{k_i},\mathbf{N}^{k_i},\mathbf{Y}^{k_i})=(\mathbf{Z}^*,\mathbf{E}^*,\mathbf{N}^*,\mathbf{Y}^*), \quad (24)$$

Summing both sides of (22) from $k=1$ to $k=k_i-1$, it has

$$\begin{aligned} &\mathcal{L}_\beta(\mathbf{Z}^{k_i},\mathbf{E}^{k_i},\mathbf{N}^{k_i},\mathbf{Y}^{k_i})-\mathcal{L}_\beta(\mathbf{Z}^1,\mathbf{E}^1,\mathbf{N}^1,\mathbf{Y}^1) \\ &\le-C\sum_{k=1}^{k_i-1}\|\mathbf{N}^{k+1}-\mathbf{N}^k\|_{\mathrm{F}}^2. \end{aligned} \quad (25)$$

Passing to the limit on both sides of (25), it is concluded that

$$\begin{aligned} &C\sum_{k=1}^{\infty}\|\mathbf{N}^{k+1}-\mathbf{N}^k\|_{\mathrm{F}}^2 \\ &\le\mathcal{L}_\beta(\mathbf{Z}^1,\mathbf{E}^1,\mathbf{N}^1,\mathbf{Y}^1)-\mathcal{L}_\beta(\mathbf{Z}^*,\mathbf{E}^*,\mathbf{N}^*,\mathbf{Y}^*)<\infty. \end{aligned} \quad (26)$$

This, together with the fact that $C\ge 0$, derives

$$\sum_{k=1}^{\infty}\|\mathbf{N}^{k+1}-\mathbf{N}^k\|_{\mathrm{F}}^2<\infty, \quad (27)$$

which implies

$$\mathbf{N}^{k+1}-\mathbf{N}^k\to 0. \quad (28)$$

Authorized licensed use limited to: SHANGHAI UNIVERSITY. Downloaded on April 17,2023 at 02:32:15 UTC from IEEE Xplore. Restrictions apply.

From (17), it then claims that

$$\mathbf{Y}^{k+1} - \mathbf{Y}^k \to 0, \tag{29}$$

and then recall (4d), it is evident that

$$\mathbf{Z}^{k+1} - \mathbf{Z}^k \to 0,\ \mathbf{E}^{k+1} - \mathbf{E}^k \to 0. \tag{30}$$

Finally, with the help of KL property [19], it is guaranteed that the generated sequence $\{(\mathbf{Z}^k, \mathbf{E}^k, \mathbf{N}^k, \mathbf{Y}^k)\}$ converges to a stationary point of problem (2). The proof is completed. ■

IV. Monitoring Strategy

In real-world industrial processes, there exist some minor faults and noise due to the failures of sensors or the errors of recording operations. As is verified in [10], if no preprocessing is performed, the PM performance will be greatly degraded. Therefore, it is necessary to filter out these minor faults and noise before monitoring. In this brief, after conducting matrix decomposition, the process data $\mathbf{X}$ can be divided into $\mathbf{XZ} + \mathbf{E} + \mathbf{N}$, of which $\mathbf{XZ}$ is the clean process information, $\mathbf{E}$ is the minor faults, and $\mathbf{N}$ is the Gaussian noise.

In the following, the SLR-based monitoring strategy will be provided in detail.

Step 1) Normalization: To eliminate the negative effects of measurement units, the process data $\mathbf{X}$ should be normalized, which is described as

$$x_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}, \tag{31}$$

where x_{ij} is the ijth element of $\mathbf{X}$, $\bar{x}_j$ is the mean of $\mathbf{x}_j$, and σ_j is the standard deviation of $\mathbf{x}_j$.

Step 2) Noise Reduction: Through exploring the special structures of minor faults $\mathbf{E}$ and noise $\mathbf{N}$, the clean information of process data $\mathbf{X}$ can be achieved from the proposed SLR in problem (2) as follows.

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{E},\mathbf{N}}\ & \|\mathbf{Z}\|_* + \lambda\|\mathbf{E}\|_{2,0} + \mu\|\mathbf{N}\|_{\mathrm{F}}^2 \\ \text{s.t.}\ & \mathbf{X} = \mathbf{XZ} + \mathbf{E} + \mathbf{N}, \\ & \mathbf{Z} \in \mathcal{M}. \end{aligned} \tag{32}$$

It can be efficiently optimized by Algorithm 1 and guaranteed the convergence by Theorem 1.

Step 3) Loading Matrix Computation: Compared with PCA, the loading matrix $\mathbf{P}$ can be obtained by computing the SVD on matrix $\mathbf{XZ}$, rather than $\mathbf{X}$, which is given by

$$\mathbf{XZ} = \mathbf{U}\Lambda\mathbf{P}^\top, \tag{33}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$ is the singular value matrix, of which r is the number of principal components (PCs).

Step 4) Control Limit Determination: According to [1], the control limits for T^2 test statistic and SPE test statistic are determined by the following estimations

$$\begin{aligned} J_{th,\mathrm{T}^2} &= \frac{r(n^2-1)}{n(n-1)}F_\alpha(r, n-r), \\ J_{th,\mathrm{SPE}} &= \theta_1\left(\frac{C_\alpha\sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0(h_0-1)}{\theta_1^2}\right)^{1/h_0}, \end{aligned} \tag{34}$$

with

$$\theta_i = \sum_{j=r+1}^{p}(\sigma_j^2)^i\ (i = 1, 2, 3),\ h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}, \tag{35}$$

where $F_\alpha(r, n-r)$ is the F-distribution with significance level α and degrees of freedom $r, n-r$.

Step 5) Test Statistic Calculation: For monitoring the process, T^2 test statistic and SPE test statistic are defined as

$$\begin{aligned} \mathrm{T}^2 &= \mathbf{dPW}^{-1}\mathbf{P}^\top\mathbf{d}^\top, \\ \mathrm{SPE} &= \mathbf{d}^\top(\mathbf{I} - \mathbf{PP}^\top)\mathbf{d}, \end{aligned} \tag{36}$$

where $\mathbf{d}$ is the new data sample and $\mathbf{W} = \Lambda^\top\Lambda$.

Step 6) Online Monitoring: If the resulting T^2 test statistic or SPE test statistic is larger than the predefined control limit, a fault is detected, otherwise, no fault is detected. Therefore, the monitoring results will be checked by the following logic

$$\begin{cases} \mathrm{T}^2 > J_{th,\mathrm{T}^2} \text{ and } \mathrm{SPE} > J_{th,\mathrm{SPE}} & \Rightarrow \text{ faulty}, \\ \mathrm{T}^2 \le J_{th,\mathrm{T}^2} \text{ or } \mathrm{SPE} \le J_{th,\mathrm{SPE}} & \Rightarrow \text{ fault-free}. \end{cases} \tag{37}$$

For online monitoring, Step 5) and Step 6) will be performed until all samples are detected.

V. Simulation Examples

A. Data Preparation

This section illustrates the effectiveness and robustness of the proposed PM method. For offline modeling, it first generates a total of 500 samples, each with 10 process variables. Specifically, the process sample is constructed by

$$\mathbf{x} = \mathbf{Sh} + \mathbf{e} + \mathbf{n}, \tag{38}$$

where $\mathbf{S}$ and $\mathbf{h}$ are defined as

$$\begin{aligned} \mathbf{S} &= \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & -0.6 & -0.6 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0.8 & 0.8 \end{bmatrix}^\top, \\ \mathbf{h} &= \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \sim \mathcal{N}\left\{\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.98 & 0 \\ 0 & 1 \end{bmatrix}\right\}, \end{aligned} \tag{39}$$

respectively, $\mathbf{e}$ denotes the minor faults that occur between the 101th to 500th samples with the magnitude of 0.1, and $\mathbf{n}$ denotes the added Gaussian noise with mean 0 and variance $\Sigma = \mathrm{diag}([0.09\ 0.09\ 0.09\ 0.09\ 0.16\ 0.16\ 0.16\ 0.16\ 0.25\ 0.25])$. For online monitoring, a total of 300 samples are then generated. Two faults, as described below, are introduced to the monitoring data between the 101th and 300th samples.

- Case I: Two sensor biases

$$\mathbf{x}_f = \mathbf{Sh} + \mathbf{e} + \mathbf{n} + [0\ 1.2\ 0\ 1.8\ 0\ 0\ 0\ 0\ 0\ 0]^\top. \tag{40}$$

- Case II: A step change

$$\mathbf{x}_f = \mathbf{S}(\mathbf{h} + [3\ 0]^\top) + \mathbf{e} + \mathbf{n}. \tag{41}$$

In this brief, the proposed method is compared with some state-of-the-art methods, including PCA [1], RPCA [10], robust sparse PCA (RSPCA) [24], LRR [12], and LRJE [16]. All algorithms are implemented based on the ADMM framework and choose the same stopping criterion. Because two PCs can capture at least 85% variance, r is determined to be 2 for all above-mentioned methods. Another critical problem is how to determine these regularization parameters. From experience, five-fold cross-validation techniques can be applied to seek proper candidates. Moreover, $\beta \ge 2\mu$ should be satisfied to guarantee the convergence of the proposed algorithm.

Authorized licensed use limited to: SHANGHAI UNIVERSITY. Downloaded on April 17,2023 at 02:32:15 UTC from IEEE Xplore. Restrictions apply.

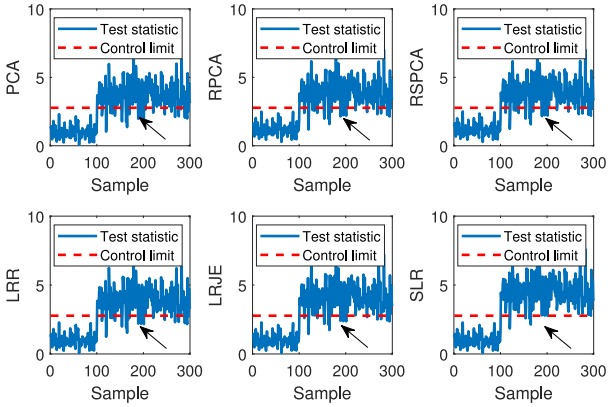

Fig. 1. Monitoring performance of SPE test statistic for Case I.

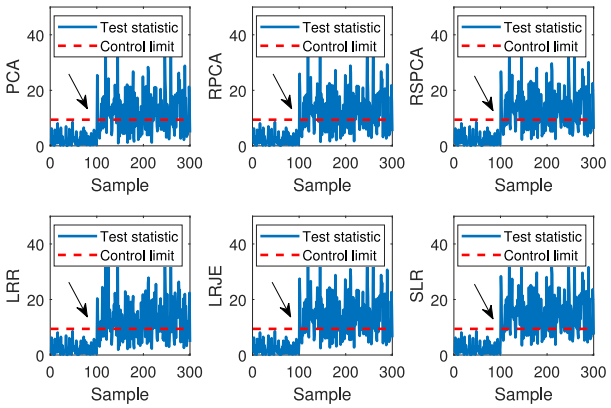

Fig. 2. Monitoring performance of T^2 test statistic for Case II.

B. Monitoring Performance

Since Case I involves two sensor biases, T^2 test statistic will be affected slightly but SPE test statistic will be affected largely. Hence, only the monitoring results of SPE test statistic are presented in Fig. 1. Although this fault can be successfully detected at the 101st sample, all six methods cannot detect it continuously. This suggests that minor faults and noise are difficult to detect. In contrast, the proposed SLR is able to obtain more samples that exceed the control limit, which indicates that its monitoring performance is better.

Case II simulates a step change that occurs in a hidden variable. Compared with Case I, the fault slightly affects SPE test statistic, but largely affects T^2 test statistic. Thus, only the monitoring results of T^2 test statistic are displayed in Fig. 2. It is concluded that Case II is more difficult to detect since all six methods fail to detect this fault at the 101st sample. Even so, the proposed SLR can detect more fault samples, which verifies that the proposed method is much more promising.

VI. Conclusion

This brief has proposed an efficient and robust data-driven PM to deal with minor faults and noise of industrial processes. The objective function contains the $\ell_{2,0}$-norm optimization to improve the global variable selection and the manifold learning to preserve the local variable relationship. In addition, an iterative minimization algorithm has been developed and the convergence property has also been established mathematically. Finally, case studies have illustrated its superiority on simulated data.

In the future, it is interesting to extend the proposed method to other practical applications such as face recognition and image denoising.

References

[1] S. X. Ding, *Data-Driven Design of Fault Diagnosis and Fault-Tolerant Control Systems*. London, U.K.: Springer, 2014.
[2] X. Chen, J. Wang, and S. X. Ding, "Complex system monitoring based on distributed least squares method," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 4, pp. 1892–1900, Oct. 2021.
[3] 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.
[4] Q. Sun and Z. Ge, "A survey on deep learning for data-driven soft sensors," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 5853–5866, Sep. 2021.
[5] Y. Wang and Z. Wang, "Data-driven model-free adaptive fault-tolerant control for a class of discrete-time systems," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 1, pp. 154–158, Jan. 2022.
[6] 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.
[7] S. Yin, B. Xiao, S. Ding, and D. Zhou, "A review on recent development of spacecraft attitude fault tolerant control system," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3311–3320, May 2016.
[8] N. Vaswani, T. Bouwmans, S. Javed, and P. Narayanamurthy, "Robust subspace learning: Robust PCA, robust subspace tracking, and robust subspace recovery," *IEEE Signal Process. Mag.*, vol. 35, no. 4, pp. 32–55, Jul. 2018.
[9] Z. Yan, C. Chen, Y. Yao, and C. Huang, "Robust multivariate statistical process monitoring via stable principal component pursuit," *Ind. Eng. Chem. Res.*, vol. 55, no. 14, pp. 4011–4021, 2016.
[10] Y. Pan, C. Yang, R. An, and Y. Sun, "Robust principal component pursuit for fault detection in a blast furnace process," *Ind. Eng. Chem. Res.*, vol. 57, no. 1, pp. 283–291, 2018.
[11] 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.
[12] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.
[13] Z. Hu, F. Nie, R. Wang, and X. Li, "Low rank regularization: A review," *Neural Netw.*, vol. 136, pp. 218–232, Apr. 2021.
[14] T. Guo, D. Zhou, J. Zhang, M. Chen, and X. Tai, "Fault detection based on robust characteristic dimensionality reduction," *Control. Eng. Pract.*, vol. 84, pp. 125–138, Mar. 2019.
[15] W. Yu and C. Zhao, "Low-rank characteristic and temporal correlation analytics for incipient industrial fault detection with missing data," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 6337–6346, Sep. 2021.
[16] Y. Fu, C. Luo, and Z. Bi, "Low-rank joint embedding and its application for robust process monitoring," *IEEE Trans. Instrum. Meas.*, vol. 70, Apr. 2021, Art. no. 3515313, doi: 10.1109/TIM.2021.3075017.
[17] T. Pang, F. Nie, J. Han, and X. Li, "Efficient feature selection via $\ell_{2,0}$-norm constrained sparse regression," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 5, pp. 880–893, May 2019.
[18] F.-Y. Wu, K. Yang, and Y. Hu, "Sparse estimator with ℓ_0-norm constraint Kernel maximum-correntropy-criterion," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 67, no. 2, pp. 400–404, Feb. 2020.
[19] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *J. Sci. Comput.*, vol. 78, no. 1, pp. 29–63, 2019.
[20] D.-R. 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.
[21] C. Li *et al.*, "Deep manifold structure transfer for action recognition," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4646–4658, Sep. 2019.
[22] 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.
[23] R. T. Rockafellar and R. Wets, *Variational Analysis*. Cham, Switzerland: Springer, 2009.
[24] A. Breloy, S. Kumar, Y. Sun, and D. Palomar, "Majorization-minimization on the Stiefel manifold with application to robust sparse PCA," *IEEE Trans. Signal Process.*, vol. 69, pp. 1507–1520, Feb. 2021.

Authorized licensed use limited to: SHANGHAI UNIVERSITY. Downloaded on April 17,2023 at 02:32:15 UTC from IEEE Xplore. Restrictions apply.