Journal of Process Control 92 (2020) 212–219

Contents lists available at ScienceDirect

Journal of Process Control

journal homepage: www.elsevier.com/locate/jprocont

Laplacian regularized robust principal component analysis for process monitoring

Xianchao Xiu [a], Ying Yang [a,*], Lingchen Kong [b], Wanquan Liu [c]

[a] Department of Mechanics and Engineering Science, Peking University, Beijing, China
[b] Department of Applied Mathematics, Beijing Jiaotong University, Beijing, China
[c] Department of Computing, Curtin University, Perth, WA, Australia

ARTICLE INFO

Article history:
Received 28 December 2019
Received in revised form 29 May 2020
Accepted 22 June 2020
Available online 30 June 2020

Keywords:
Process monitoring
Principal component analysis (PCA)
Robust
Hypergraph Laplacian
Alternating direction method of multipliers

ABSTRACT

Principal component analysis (PCA) is one of the most widely used techniques for process monitoring. However, it is highly sensitive to sparse errors because of the assumption that data only contains an underlying low-rank structure. To improve classical PCA in this regard, a novel Laplacian regularized robust principal component analysis (LRPCA) framework is proposed, where the "robust" comes from the introduction of a sparse term. By taking advantage of the hypergraph Laplacian, LRPCA not only can represent the global low-dimensional structures, but also capture the intrinsic non-linear geometric information. An efficient alternating direction method of multipliers is designed with convergence guarantee. The resulting subproblems either have closed-form solutions or can be solved by fast solvers. Numerical experiments, including a simulation example and the Tennessee Eastman process, are conducted to illustrate the improved process monitoring performance of the proposed LRPCA.

1. Introduction

Due to the development of modern industrial production, process monitoring becomes increasingly important to guarantee safety and reliability [1]. A number of multivariate statistical techniques have been proposed, such as principal component analysis (PCA) [2–5], partial least squares (PLS) [6,7], canonical correlation analysis (CCA) [8,9], Fisher discriminant analysis (FDA) [10], slow feature analysis (SFA) [11–13], probabilistic latent variable model [14,15], and so on. It would not be exaggerating to say that PCA is one of the most widely used tools in process monitoring. From a dimension reduction perspective, PCA can be described as a small set of orthogonal linear combinations of the original variables, named principal components (PCs), such that it can capture as much information as possible [16]. Therefore, the original data space can be divided into two orthogonal subspaces: a PC subspace and a residual subspace. The PC subspace preserves most of the information and the residual subspace contains the rest information. To monitor the process data, T^2 and *SPE* statistics [17] can be defined in the PC subspace or in the residual subspace. With a given significance level, the corresponding control limits are determined. Once any data causes fault detection indices out of the control limits, faults are thus detected.

In high-dimensional data analysis, classical PCA often breaks down because one can usually find meaningless solutions, which makes PCA-based process monitoring ambiguous [4,10]. To overcome this shortcoming, many researchers have made lots of contributions related to the theory as well as algorithms. Roughly speaking, they can be sorted into three categories: conventional contribution plot and reconstruction-based approaches [18], conventional approaches with process knowledge [19], sparse PCA (SPCA) and its extensions [2,3,20–22]. Compared with the first two categories, SPCA is much more effective and efficient. By imposing a sparse regularization on the PCA objective function, the extracted PCs become linear combinations of a small part of process variables, which means that faults are easier to understand [23]. Yan and Yao [22] applied SPCA to process monitoring, and achieved higher computational efficiency and good isolation results. Liu et al. [2] proposed adaptive SPCA with reconstruction-based contribution plot, and the obtained fault diagnosis was promising. Liu et al. [21] studied the joint sparsity regularized PCA by introducing the $\ell_{2,1}$ norm. It is demonstrated that joint sparsity prompts the variables corresponding to all zero rows not be included in the projected subspace. Therefore, the unimportant variables are removed [24].

If the data in a high-dimensional space actually lie on the union of several linear subspaces, PCA is easy to capture the low-dimensional structures embedded in the data. However, in real-world applications the assumption cannot be ensured. In this case, PCA may fail to discover the intrinsic geometric structures. In order to preserve local geometric structures embedded

* Corresponding author.
E-mail addresses: xcxiu@bjtu.edu.cn (X. Xiu), yy@pku.edu.cn (Y. Yang), lchkong@bjtu.edu.cn (L. Kong), W.Liu@curtin.edu.au (W. Liu).

https://doi.org/10.1016/j.jprocont.2020.06.011
0959-1524/

in a high-dimensional space, manifold learning can be considered [25]. Accordingly, Jiang et al. [20] developed a graph Laplacian regularized PCA, where the graph Laplacian able to capture the cause–effect relationship between process variables and discover conditional independent process variable sets between individual operation units. Very recently, by combining joint sparsity and graph Laplacian regularization, Liu et al. [3] proposed a structured joint sparse PCA (SJSPCA) for fault detection and isolation. SJSPCA shares both the advantages of sparse methods over dense methods, and the merits of manifold learning over alternative learning methods. However, all the above mentioned PCA-based approaches are a little bit more sensitive to outliers, which may result in unstable performance.

Robust PCA (RPCA) was proposed Hubert et al. [26], and can be tracked back to Stahel [27] and Donoho [28]. The interested readers can refer [29] for a recent survey. In fact, it has a deep link with low-rank recovery [30–32]. Although, RPCA has been employed in the fields such as medicine, economics and engineering, the application of RPCA in process monitoring has not been widely investigated yet. As far as we know, Yan et al. [33] was the first to incorporate RPCA and developed a two-stage monitoring strategy. The applications to the Tennessee Eastman (TE) process showed that RPCA can get a similar performance to PCA when the training data is clean and still preserved its effectiveness when the training data is contaminated by sparse noises. PCA-based methods are not robust to sparse noises and RPCA is unable to preserve local geometric structures. *An interesting question naturally comes out to us: can we derive a variant of PCA that is able to learn the prior information and is also robust to sparse noises.*

Therefore, in this paper, we propose a novel Laplacian regularized robust principal component analysis (LRPCA), and apply it to process monitoring. We would like to point out that the "robust" here is embedding a sparse error term in PCA, which is different from [33]. Meanwhile, a hypergraph Laplacian, instead of a normal graph Laplacian in [3], is used to describe similarity structures among data. The main contributions of this paper are summarized in the following three aspects:

- By introducing a sparse term and a hypergraph Laplacian term, we extend classical PCA to a novel Laplacian regularized robust PCA (LRPCA). To the best of author's knowledge, we are the first to model PCA in such a LRPCA framework.
- An efficient alternating direction method of multipliers (ADMM) is developed to optimize LRPCA. In theory, its local convergence is established. This gives a fast and convergent algorithm.
- Experiments on a simulation example and the TE practical process are conducted to illustrate that LRPCA outperforms the state-of-the-art methods, such as PCA, RPCA and SJSPCA.

The remainder of this paper is organized as follows. In Section 2, some developments of PCA are reviewed. In Section 3, the proposed LRPCA model, as well as optimization algorithm, are presented. In Section 4, an online process monitoring using LRPCA is developed. In Section 5, numerical results are reported to illustrate the advantages of the proposed LRPCA. Finally, Section 6 concludes this paper.

2. Principal component analysis

Let $\mathbf{X} \in \mathbb{R}^{n\times p}$ be a data matrix, where n is the number of samples and p is the number of variables. The classical PCA [34] can be presented as the following orthogonal minimization problem

$$\begin{aligned} \min_{\mathbf{A}} \ & \frac{1}{2}\|\mathbf{X}-\mathbf{X}\mathbf{A}\mathbf{A}^T\|_F^2 \\ \text{s.t.} \ & \mathbf{A}^T\mathbf{A}=\mathbf{I}_k, \end{aligned} \tag{1}$$

where k is the reduced rank. From the aspect of optimization, $\mathbf{A}$ implies that PCA can be interpreted as the best k-rank approximation to the data matrix $\mathbf{X}$. Moreover, $\mathbf{A}$ reveals some underlying affinity between variables.

By introducing an auxiliary variable $\mathbf{B}$, problem (1) can be reformulated as

$$\begin{aligned} \min_{\mathbf{A},\mathbf{B}} \ & \frac{1}{2}\|\mathbf{X}-\mathbf{X}\mathbf{B}\mathbf{A}^T\|_F^2 \\ \text{s.t.} \ & \mathbf{A}^T\mathbf{A}=\mathbf{I}_k,\ \ \mathbf{A}=\mathbf{B}. \end{aligned}$$

It is proved that when $n > p$ the equality constraint $\mathbf{A}=\mathbf{B}$ can be dropped and still acquire the loading vectors exactly [23]. In high-dimensional settings, interpreting the derived PCs is an important task. Sparse principal component analysis was proposed to enhance the interpretation, which is given by

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

where $\|\cdot\|_1$ is the sum of absolute values of all entries, and λ_1 is a penalty parameter to control the sparsity. As illustrated in [35], joint sparsity acts like the sparsity at the group level in order to discard useless variables. Thus, Liu et al. [21] incorporated joint sparsity with PCA, and considered

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

where $\|\cdot\|_{2,1}$ is defined as the sum of the ℓ_2 norm of all rows.

In practice, it is reasonable to assume that if two variables are close in the intrinsic space, then the representations of these two variables in a new space are also close to each other. This observation inspired Liu et al. [3] to propose the graph Laplacian regularized joint sparse PCA (SJSPCA), i.e.,

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

where $\mathbf{L}$ is a graph Laplacian matrix. This formulation shares the similar idea as that used in our proposed model, but with the purpose of obtaining more robust results and learning more information from process data. In contrast, we embed a sparse term and a hypergraph Laplacian term into a PCA framework.

3. Laplacian regularized robust PCA

3.1. Model formulation

Motivated by the recent development of RPCA, we inset a sparse error $\mathbf{E}$ into the objective function of (2), and establish the following optimization problem

$$\begin{aligned} \min_{\mathbf{A},\mathbf{B},\mathbf{E}} \ & \frac{1}{2}\|\mathbf{X}-\mathbf{X}\mathbf{B}\mathbf{A}^T-\mathbf{E}\|_F^2+\lambda_1\|\mathbf{B}\|_{2,1}+\lambda_2\|\mathbf{E}\|_1 \\ \text{s.t.} \ & \mathbf{A}^T\mathbf{A}=\mathbf{I}_k. \end{aligned} \tag{4}$$

Similar to (3), a graph Laplacian regularization can be put into PCA such that the cause–effect relationship among process variables is efficiently represented [3]. However, if the data relationship is complex enough, it cannot be simply described by a normal graph. Hypergraph is an extension of graph, in which an edge can connect more than two vertices. Before introducing our proposed model, we first give some basic notations of hypergraphs. Let $\mathcal{G}=(\mathcal{V},\mathcal{E})$ be a hypergraph, where $\mathcal{V}$ is the vertex set corresponding to n samples, and $\mathcal{E}$ is the hyperedge set corresponding to correlation between p variables. For a subset $e \in \mathcal{E}$, the edge weight is denoted by $\mathbf{W}(e)$ and the weight matrix $\mathbf{W}_{\mathcal{E}}$ consists of diagonal entries $\mathbf{W}(e)$. For a vertex $v \in \mathcal{V}$, the degree is defined as $d(v) = \sum_{e\in\mathcal{E}} \mathbf{W}(e)h(v,e)$ with

$$h(v,e) = \begin{cases} 1 & \text{if } v \in e; \\ 0 & \text{otherwise.} \end{cases}$$

$\mathbf{D}_{\mathcal{V}}$ denotes the diagonal matrix whose diagonal entries correspond to the degree of each vertex. In addition, for a hyperedge e, the degree is denoted by $d(e) = \sum_{v\in\mathcal{V}} h(v,e)$ and the diagonal degree matrix $\mathbf{D}_{\mathcal{E}}$ is thus obtained. From [36], the hypergraph Laplacian matrix can be given by

$$\mathbf{L}^h = \mathbf{D}_{\mathcal{V}} - \mathbf{H}\mathbf{W}_{\mathcal{E}}\mathbf{D}_{\mathcal{E}}^{-1}\mathbf{H}^T.$$

Consequently, the hypergraph Laplacian regularized robust PCA (LRPCA) considers the following minimization problem:

$$\begin{aligned} \min_{\mathbf{A},\mathbf{B},\mathbf{E}} \ & \frac{1}{2}\|\mathbf{X} - \mathbf{X}\mathbf{B}\mathbf{A}^T - \mathbf{E}\|_F^2 + \lambda_1\|\mathbf{B}\|_{2,1} + \lambda_2\|\mathbf{E}\|_1 + \lambda_3 \mathrm{tr}(\mathbf{B}^T\mathbf{L}^h\mathbf{B}) \\ \text{s.t.} \ & \mathbf{A}^T\mathbf{A} = \mathbf{I}_k. \end{aligned} \tag{5}$$

Note that our proposed LRPCA is more general than SJSPCA. It involves three regularization terms: the $\ell_{2,1}$ norm of $\mathbf{B}$, the ℓ_1 norm of $\mathbf{E}$ and the hypergraph Laplacian. The $\ell_{2,1}$ norm ensures row-wise sparsity, the ℓ_1 norm controls the noisy level, and the hypergraph Laplacian takes the variable correlation information into consideration. Since the optimization problem (5) is not easy to solve, an efficient optimization algorithm with closed-form solutions will be developed in the following.

3.2. Optimization algorithm

In this subsection, we present an ADMM for solving (5), which can be equivalently formulated as

$$\begin{aligned} \min_{\mathbf{A},\mathbf{C},\mathbf{D},\mathbf{E},\mathbf{B}} \ & \frac{1}{2}\|\mathbf{X} - \mathbf{X}\mathbf{C}\mathbf{A}^T - \mathbf{E}\|_F^2 + \lambda_1\|\mathbf{D}\|_{2,1} + \lambda_2\|\mathbf{E}\|_1 + \lambda_3 \mathrm{tr}(\mathbf{B}^T\mathbf{L}^h\mathbf{B}) \\ \text{s.t.} \ & \mathbf{A}^T\mathbf{A} = \mathbf{I}_k, \ \mathbf{B} = \mathbf{C}, \ \mathbf{B} = \mathbf{D}, \end{aligned} \tag{6}$$

where $\mathbf{C}$ and $\mathbf{D}$ are two introduced variables to make the objective function separable. To describe the iterates of the ADMM, we first write the augmented Lagrangian function of (6):

$$\begin{aligned} & \mathcal{L}_\beta(\mathbf{A},\mathbf{C},\mathbf{D},\mathbf{E},\mathbf{B};\mathbf{W}_1,\mathbf{W}_2) \\ = & \frac{1}{2}\|\mathbf{X} - \mathbf{X}\mathbf{C}\mathbf{A}^T - \mathbf{E}\|_F^2 + \lambda_1\|\mathbf{D}\|_{2,1} + \lambda_2\|\mathbf{E}\|_1 + \lambda_3 \mathrm{tr}(\mathbf{B}^T\mathbf{L}^h\mathbf{B}) \\ & - \langle \mathbf{W}_1, \mathbf{B} - \mathbf{C}\rangle + \frac{\beta}{2}\|\mathbf{B} - \mathbf{C}\|_F^2 - \langle \mathbf{W}_2, \mathbf{B} - \mathbf{D}\rangle + \frac{\beta}{2}\|\mathbf{B} - \mathbf{D}\|_F^2, \end{aligned}$$

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

It is difficult to simultaneously optimize all these variables. We therefore approximately solve it by alternatively minimizing one variable with the others fixed. This is the so-called ADMM [37]. Now, let us look at the resulting subproblems. Firstly, for $\mathbf{A}$-subproblem, after trivial manipulation, it can be written as

$$\mathbf{A}^{k+1} = \arg\min_{\mathbf{A}\in\mathcal{M}} \left\{ \frac{1}{2}\|\mathbf{X} - \mathbf{X}\mathbf{C}^k\mathbf{A}^T - \mathbf{E}^k\|_F^2 \right\}.$$

This is a reduced rank Procrustes rotation problem [23], and the solution is

$$\mathbf{A}^{k+1} = \mathbf{U}\mathbf{V}^T, \tag{7}$$

where $\mathbf{U}, \mathbf{V}$ is defined as $(\mathbf{X} - \mathbf{E}^k)^T\mathbf{X}\mathbf{C}^k = \mathbf{U}\Sigma\mathbf{V}^T$.

Next, for $\mathbf{C}$-subproblem, the closed-form solution is

$$\mathbf{C}^{k+1} = (\mathbf{X}^T\mathbf{X} + \beta\mathbf{I})^{-1}\left(\mathbf{X}^T\mathbf{X}\mathbf{A}^{k+1} - \mathbf{X}^T\mathbf{E}^k\mathbf{A}^{k+1} + \beta\mathbf{B}^k - \mathbf{W}_1^k\right). \tag{8}$$

Algorithm 1 ADMM for solving LRPCA model (6)

Input: Data $\mathbf{X}$, parameters $\lambda_1, \lambda_2, \lambda_3 > 0$, and $\beta > 0$.
Output: Loading matrix $\mathbf{B}$.
Initialize: Compute $\mathbf{L}^h$, $\mathbf{B}^0 = 0$, $\mathbf{W}_1^0 = \mathbf{W}_2^0 = 0$, $\varepsilon_1 = 10^{-3}$, $\varepsilon_2 = 10^{-2}$.
While not converged **do**
1: compute $\mathbf{A}^{k+1}$ according to (7);
2: compute $\mathbf{C}^{k+1}$ according to (8);
3: compute $\mathbf{D}^{k+1}$ according to (9);
4: compute $\mathbf{E}^{k+1}$ according to (10);
5: compute $\mathbf{B}^{k+1}$ according to (11);
6: update $\mathbf{W}_1^{k+1}$ via $\mathbf{W}_1^{k+1} = \mathbf{W}_1^k - \beta(\mathbf{B}^{k+1} - \mathbf{C}^{k+1})$;
7: update $\mathbf{W}_2^{k+1}$ via $\mathbf{W}_2^{k+1} = \mathbf{W}_2^k - \beta(\mathbf{B}^{k+1} - \mathbf{D}^{k+1})$.
End while

The coefficient matrix $\mathbf{X}^T\mathbf{X} + \beta\mathbf{I}$ is nonsingular whenever $\beta > 0$. It can be computed via the Cholesky decomposition or the conjugate gradient method.

Then, for $\mathbf{D}$-subproblem, it can be transformed as

$$\mathbf{D}^{k+1} = \arg\min_{\mathbf{D}} \left\{ \frac{\beta}{2}\|\mathbf{B}^k - \mathbf{D} - \mathbf{W}_2^k/\beta\|_F^2 + \lambda_1\|\mathbf{D}\|_{2,1} \right\}.$$

Denote $\mathbf{D}^{k+\frac{1}{2}} = \mathbf{B}^k - \mathbf{W}_2^k/\beta$, then the solution is given by

$$\mathbf{D}_i^{k+1} = \frac{\mathbf{D}_i^{k+\frac{1}{2}}}{\|\mathbf{D}_i^{k+\frac{1}{2}}\|_2} \circ \max\{0, \|\mathbf{D}_i^{k+\frac{1}{2}}\|_2 - \lambda_1/\beta\}, \tag{9}$$

where all operations are done componentwise. We refer [35] for illustrations.

In addition, for $\mathbf{E}$-subproblem, the solution is

$$\mathbf{E}^{k+1} = \mathrm{sign}(\mathbf{E}^{k+\frac{1}{2}}) \circ \max\left\{0, |\mathbf{E}^{k+\frac{1}{2}}| - \lambda_2\right\}, \tag{10}$$

where $\mathbf{E}^{k+\frac{1}{2}} = \mathbf{X} - \mathbf{X}\mathbf{C}^{k+1}(\mathbf{A}^{k+1})^T$, and $\mathrm{sign}(\cdot)$ is the sign function. See [38] for more details.

Finally, for $\mathbf{B}$-subproblem, it can be simplified to

$$\mathbf{B}^{k+1} = (2\lambda_3\mathbf{L}^h + 2\beta\mathbf{I})^{-1}\left(\beta\mathbf{C}^{k+1} + \mathbf{W}_1^k + \beta\mathbf{D}^{k+1} + \mathbf{W}_2^k\right). \tag{11}$$

Overall, the ADMM for solving (6) (equivalently (5)) is thus presented as in Algorithm 1. To end this section, we discuss its convergence result. The first-order optimality condition of (5) at the local minimizer $(\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{E}})$ is defined as

$$\begin{cases} 0 \in \Pi_{\mathcal{M}}(-(\mathbf{X} - \mathbf{X}\bar{\mathbf{B}}\bar{\mathbf{A}}^T - \bar{\mathbf{E}})^T\mathbf{X}\bar{\mathbf{B}}); \\ 0 \in -\mathbf{X}^T(\mathbf{X} - \mathbf{X}\bar{\mathbf{B}}\bar{\mathbf{A}}^T - \bar{\mathbf{E}})\bar{\mathbf{A}} + \lambda_1\partial\|\bar{\mathbf{B}}\|_{2,1} + 2\lambda_3\mathbf{L}^h\bar{\mathbf{B}}; \\ 0 \in -(\mathbf{X} - \mathbf{X}\bar{\mathbf{B}}\bar{\mathbf{A}}^T - \bar{\mathbf{E}}) + \lambda_2\partial\|\bar{\mathbf{E}}\|_1, \end{cases}$$

where $\Pi_{\mathcal{M}}$ is the projection onto $\mathcal{M}$. If $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{E}^*)$ satisfies the above relation in place of $(\bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\mathbf{E}})$, we say that $(\mathbf{A}^*, \mathbf{B}^*, \mathbf{E}^*)$ is a stationary point of (5).

Following a similar line of arguments as in [39], it is not hard to get the local convergence of Algorithm 1.

Theorem 3.1. *Suppose $\{(\mathbf{A}^k, \mathbf{C}^k, \mathbf{D}^k, \mathbf{E}^k, \mathbf{B}^k; \mathbf{W}_1^k, \mathbf{W}_2^k)\}$ is a sequence generated by Algorithm 1. Then it converges to a stationary point of (6) (also (5)).*

This theorem gives a theoretical guarantee of our proposed Algorithm 1, and we will verify the effectiveness by conducing some experiments.

4. Process monitoring using LRPCA

In this section, a statistical process modeling and process monitoring procedure are given based on our proposed LRPCA. We

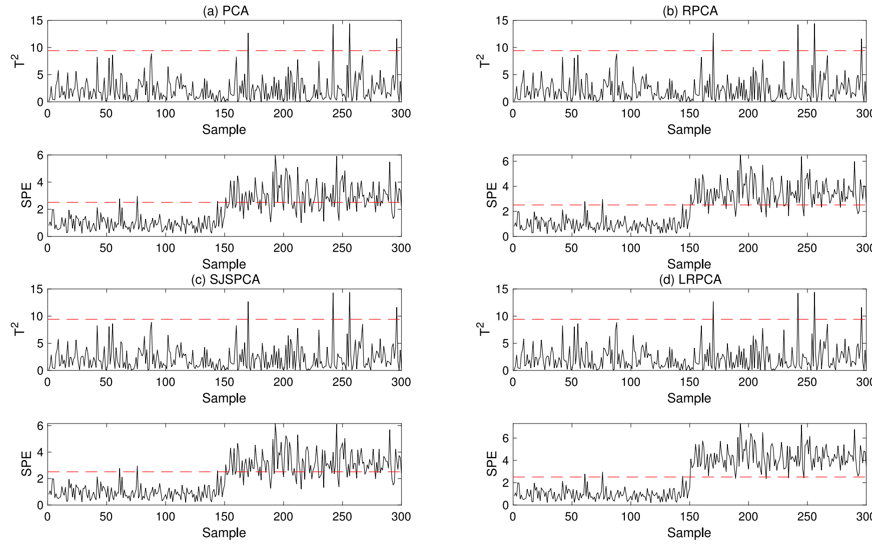

Fig. 1. Monitoring results for two sensor bias 1.2 and 1.8: (a) PCA, (b) RPCA, (c) SJSPCA, (d) LRPCA.

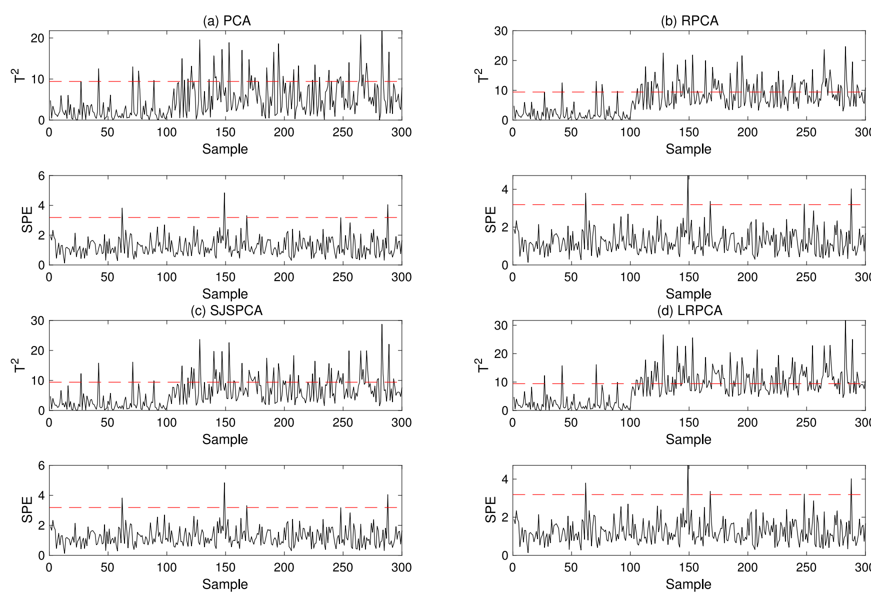

Fig. 2. Monitoring results for a step change of 2: (a) PCA, (b) RPCA, (c) SJSPCA, (d) LRPCA.

begin with normalizing the data for eliminating the effects of engineering units and measurement ranges. Given the observation data $\mathbf{X}_{obs}$, the normalization step is

$$x_{ij} = \frac{x_{obs,ij} - \bar{x}_{obs,j}}{\sigma_{obs,j}},$$

where $x_{obs,ij}$ denotes the ijth entry of the observation dat, $\bar{x}_{obs,j}$ is the mean value of the jth process variable, $\sigma_{obs,j}$ is the corresponding standard deviation, and x_{ij} is the ijth entry after normalization.

After solving optimization problem (5), the PCA loading matrix $\mathbf{B}$ can be computed. We further denote $\mathbf{B} = [\mathbf{B}_r, \mathbf{B}_d]$, where $\mathbf{B}_r$ is the loading vectors of the retained PCs and $\mathbf{B}_d$ is the loading vectors of the discarded PCs. The conventional T^2 and SPE statistics [1] can be defined as

$$T^2 = \mathbf{x}^T\mathbf{B}_r\Lambda^{-1}\mathbf{B}_r^T\mathbf{x},\ \ SPE = \mathbf{x}^T(\mathbf{I} - \mathbf{B}_r\mathbf{B}_r^T)^{-1}\mathbf{x}, \tag{12}$$

where $\Lambda = \text{diag}(\text{var}(PC_1), \text{var}(PC_2), \dots, \text{var}(PC_r))$, and $\mathbf{x}$ is the testing data. With a given significance level α, we set the control limits for T^2 and SPE statistics as follows

$$J_{th,T^2} = \frac{r(n^2-1)}{n(n-r)}F_\alpha(r, n-r),$$
$$J_{th,SPE} = \theta_1\left(\frac{C_\alpha\sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0(h_0-1)}{\theta_1^2}\right)^{1/h_0}, \tag{13}$$

with

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

Here, $F_\alpha(r, n-r)$ stands for F-distribution with r and $n-r$ degrees of freedom for the given significant level α. For fault detection purpose, we adopt the following logic

$$T^2 \le J_{th,T^2}\ \text{and}\ SPE \le J_{th,SPE} \Rightarrow \text{fault-free, otherwise faulty.} \tag{14}$$

To end this section, we summarize the process monitoring procedure in Algorithm 2. We would like to point out that the

Table 1
Selected variables in the TE process.

No.	Description	No.	Description
y_1	Feed A	y_{18}	Stripper temperature
y_2	Feed D	y_{19}	Stripper steam flow
y_3	Feed E	y_{20}	Compressor work
y_4	Total feed	y_{21}	RCW outlet temperature
y_5	Recycle flow	y_{22}	Product separator temperature
y_6	Reactor feed rate	u_1	Feed D flow valve
y_7	Reactor pressure	u_2	Feed E flow valve
y_8	Reactor level	u_3	Feed A flow valve
y_9	Reactor temperature	u_4	Total feed flow valve
y_{10}	Purge rate	u_5	Compressor recycle valve
y_{11}	RCW outlet temperature	u_6	Purge valve
y_{12}	Product separator level	u_7	Separator pot liquid flow valve
y_{13}	Product separator pressure	u_8	Stripper liquid product flow valve
y_{14}	Product separator underflow	u_9	Stripper steam valve
y_{15}	Stripper level	u_{10}	RCW flow
y_{16}	Stripper pressure	u_{11}	CCW flow
y_{17}	Stripper underflow		

Algorithm 2 Online process monitoring using LRPCA

Offline modeling:

1: Normalize the training data $\mathbf{X}$;
2: Compute LRPCA using Algorithm 1;
3: Determine the control limits based on (13);

Online detection:

1: Normalize the testing data $\mathbf{x}$;
2: Calculate the T^2 and SPE values via (12);
3: Make a decision according to the detection logic (14).

monitoring results can also be used for fault isolation. After conducting LRPCA, the loading vectors only contain a few nonzero elements, hence it is convenient to find the root cause of the incipient fault.

5. Numerical experiments

In this section, a simulation example and a TE practical study are presented to demonstrate the superiority of our proposed LR-PCA over the existing state-of-art methods, i.e., PCA [1], RPCA [33], and SJSPCA [3].

5.1. A simulation example

In this subsection, we test a simulation example, which is modified from [3]. We first generate the data with ten process variables, where the first four variables are the observations of the first hidden variable corrupted by Gaussian-distributed measurement noises, the next four variables are related to the second hidden variable, and the last two variables record the linear combinations of the two hidden variables corrupted by Gaussian noises. It follows:

$$\mathbf{x} = \mathbf{Sh} + \mathbf{e};$$

$$\mathbf{S} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -0.6 & -0.6 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0.8 & 0.8 \end{bmatrix}^T;$$

$$\mathbf{h} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \approx N\left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.98 & 0 \\ 0 & 1 \end{bmatrix} \right\},$$

where $\mathbf{e}$ is a sparse noise.

For the purpose of comparison, all the four PCA-based process monitoring methods are implemented by ourselves. The parameters $\lambda_1, \lambda_2, \lambda_3$ are determined by 10-fold cross-validation technique. Similar parameter selection strategy is also applied to RPCA and SJSPCA. A total of 500 samples under normal operation conditions are generated as the reference data set. In addition, the number of retained PCs depends on the contribution to variance accounting for more than 85%. In this study, it is set as 2. Two faulty data sets of 300 samples are generated:

- Two sensor bias 1.2 and 1.8 to the first variable after the 150th sample:

$$\mathbf{x}_{f1} = \mathbf{Sh} + \mathbf{e} + [0\ 1.2\ 0\ 1.8\ 0\ 0\ 0\ 0\ 0\ 0]^T.$$

- A step change of 2 to the first hidden variable after the 100th sample:

$$\mathbf{x}_{f2} = \mathbf{S}(\mathbf{h} + [2\ 0]^T) + \mathbf{e}.$$

Based on the process monitoring procedure, the T^2 and SPE statistics are constructed. Fig. 1 shows the monitoring results for two sensor bias of 1.2 and 1.8. It can be found that the T^2 statistics of all the four PCA-based methods are similar, while significant number of violations are observed after the 150th sample in the SPE statistics. Although the fault can be successfully detected by the four methods, different portion of the SPE values

Table 2
The TE process faults.

Fault No.	Description	Type
1	A/C feed ratio	Step change
2	Component B	Step change
3	Feed D temperature	Step change
4	RCW inlet temperature	Step change
5	CCW inlet temperature	Step change
6	Feed A loss	Step change
7	C header pressure loss	Step change
8	Feed A–C components	Random variation
9	Feed D temperature	Random variation
10	Feed C temperature	Random variation
11	RCW inlet temperature	Random variation
12	CCW inlet temperature	Random variation
13	Reaction kinetics	Slow drift
14	RCW valve	Sticking
15	CCW valve	Sticking
16	Unknown fault	Unknown
17	Unknown fault	Unknown
18	Unknown fault	Unknownown
19	Unknown fault	Unknownown
20	Unknown fault	Unknownown
21	Unknown fault	Constant

Table 3
Detection results in terms of FDR(%).

Fault No.	PCA		RPCA		SJSPCA		LRPCA	
	T^2	SPE	T^2	SPE	T^2	SPE	T^2	SPE
1	99.13	99.88	99.25	99.88	99.25	99.88	99.25	100
2	98.38	95.75	98.38	98.00	98.38	99.00	98.38	99.75
3	0.88	2.63	1.88	3.25	2.25	4.25	3.88	6.75
4	20.88	100	30.25	100	34.50	100	39.38	100
5	24.13	20.88	28.75	24.25	30.25	24.25	34.50	24.50
6	99.13	100	99.25	100	99.38	100	99.38	100
7	100	100	100	100	100	100	100	100
8	96.88	83.63	97.13	91.50	97.13	96.75	98.50	96.75
9	1.75	1.75	2.63	2.50	1.75	2.50	3.75	2.75
10	29.63	25.75	33.13	32.75	34.00	30.38	36.13	34.58
11	40.63	74.88	46.37	81.25	48.38	84.88	49.50	90.25
12	98.38	89.50	98.50	90.75	98.50	90.75	99.25	90.75
13	93.63	95.25	93.63	96.25	93.63	97.50	93.63	99.75
14	99.25	100	99.50	100	99.88	100	99.88	100
15	1.38	3.00	2.50	3.88	1.50	3.88	3.75	7.25
16	13.50	27.38	14.13	32.25	14.63	39.50	16.78	39.50
17	76.25	95.38	78.00	95.88	83.50	96.25	88.63	96.75
18	89.25	90.13	89.38	91.25	89.38	92.50	90.63	92.50
19	14.13	18.50	16.25	24.38	16.25	22.38	16.75	28.47
20	31.75	49.75	42.13	52.25	39.38	68.25	48.38	69.63
21	39.25	47.25	39.50	47.38	44.63	49.25	45.75	52.38
Average	55.63	60.14	57.64	62.82	58.41	64.98	**60.29**	**66.75**

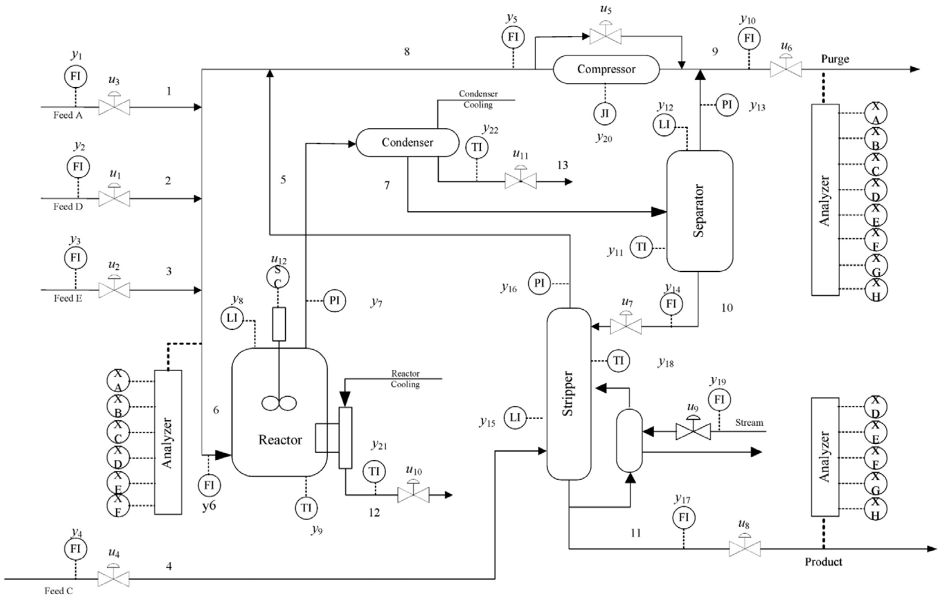

Fig. 3. The TE process and monitoring variables.

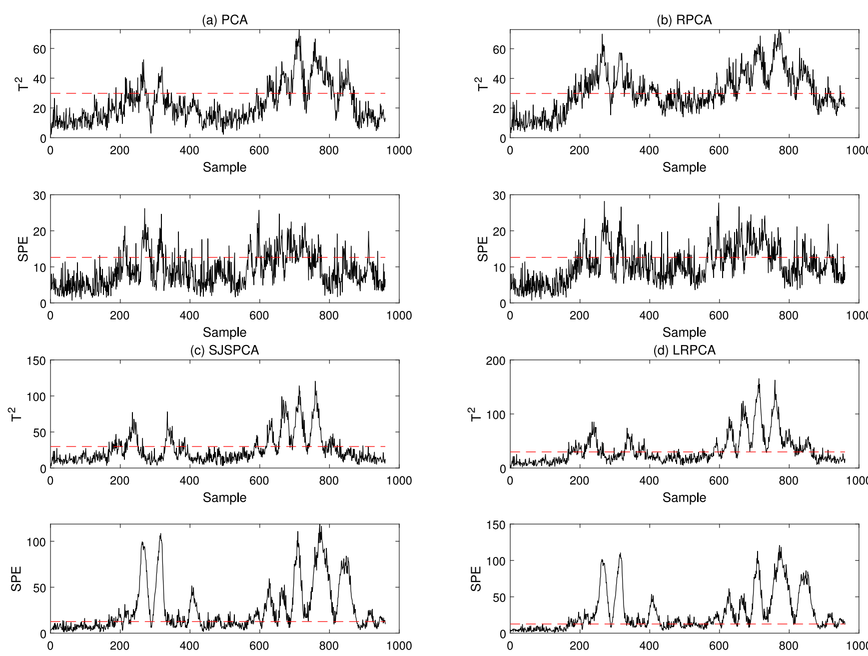

Fig. 4. Monitoring results for fault 10 in TE: (a) PCA, (b) RPCA, (c) SJSPCA, (d) LRPCA.

exceed the control limit. Compared with PCA and RPCA, the Laplacian regularized PCA methods (i.e., SJSPCA and LRPCA) are more preferable. This good performance owes to the graph Laplacian regularization, which could embed the prior information into optimization process. Moreover, LRPCA could remove much more outliers and enhance process monitoring performance because the sparse term **E** can efficiently characterize the gross noise.

Fig. 2 summarizes the monitoring results for a step change of 2. This time, the monitoring results of *SPE* statistics for all the four PCA-based methods are similar, and significant number of violations after the 100th sample in the T^2 statistics are observed. It is explained by the fact that the hidden variables are the main sources of variations in the T^2 statistics. Similar conclusions can be made. For example, compared with PCA, RPCA and SJSPCA, our proposed LRPCA gives much more samples violated the control limit, which illustrates that LRPCA is sensitive to this fault.

5.2. TE practical study

In this subsection, we test the TE process, which is developed based on a real industrial process. The TE process consists of five major units, i.e., reactor, condenser, separator, stripper, compressor. There exist eight components, including four reactants, two products, a major byproduct, and an inert component. A flowchart of the TE process is shown in Fig. 3, and more information can be found in [40].

During the operation of the TE process, a total of 41 measured variables and 12 manipulated variables are recorded. In

Table 4
Detection results in terms of FAR(%).

Fault No.	PCA		RPCA		SJSPCA		LRPCA	
	T^2	*SPE*	T^2	*SPE*	T^2	*SPE*	T^2	*SPE*
1	0.00	0.63	0.00	0.00	0.00	0.00	0.00	0.00
2	1.25	0.63	0.00	0.63	0.63	0.63	0.00	0.63
3	0.00	1.25	0.00	0.63	0.00	0.63	0.00	0.00
4	0.63	1.25	0.63	1.25	0.63	0.63	0.63	0.00
5	0.63	1.88	0.00	0.63	0.00	0.00	0.00	0.00
6	0.00	1.25	0.00	1.25	0.00	1.25	0.00	0.63
7	0.00	1.25	0.00	0.00	0.00	0.63	0.00	0.00
8	0.00	0.63	0.00	0.63	0.00	0.63	0.00	0.63
9	1.88	1.88	1.25	0.63	0.63	0.63	0.63	0.00
10	0.00	0.63	0.00	0.00	0.00	0.00	0.00	0.00
11	0.63	2.50	0.63	1.25	0.63	1.88	0.63	0.63
12	0.00	1.25	0.00	1.25	0.00	0.63	0.00	0.63
13	0.63	0.00	0.63	0.00	0.63	0.00	0.63	0.00
14	0.00	1.25	0.00	0.63	0.00	0.63	0.00	0.63
15	0.00	1.25	0.00	0.63	0.00	1.25	0.00	0.00
16	3.75	1.88	2.50	1.88	1.88	1.88	0.63	1.25
17	1.25	2.50	0.63	1.88	1.25	1.88	0.63	1.88
18	0.00	2.50	0.00	1.25	0.00	1.25	0.00	1.25
19	0.00	0.63	0.00	0.00	0.00	0.63	0.00	0.00
20	0.00	1.25	0.00	0.63	0.00	0.63	0.00	0.63
21	0.00	3.13	0.00	2.50	0.00	2.50	0.00	1.25
Average	0.51	1.40	0.30	0.84	0.29	0.90	**0.18**	**0.48**

this paper, we select 22 measured variables and 11 manipulated variables as listed in Table 1. Here, RCW stands for reactor cooling water and CCW is an abbreviation of condenser cooling water. The TE process consists of 1 normal data set and 21 faulty data sets. The normal operation data is usually used for model training, while each faulty data is injected a fault at the 161st sampling time point. Descriptions of the 21 faults are provided in Table 2.

In order to evaluate the performance of process monitoring, the fault detection rate (FDR) and false alarm rate (FAR) are adopted [1]. FDR is defined as the percentage of faulty data samples detected during the total faulty data. FAR is defined as the percentage of normal samples identified as faults during the normal operation. The higher FDR and lower FAR indicate the better the monitoring performance.

The monitoring results of PCA, RPCA, SJSPCA and our proposed LRPCA are given in Tables 3 and 4. The performances of LRPCA are always the best in terms of FDR and FAR values. In average, for FDR values, the gains of T^2 statistics are 4.66%, 2.65%, 1.88%, the gains of *SPE* statistics are 6.61%, 3.93%, 1.77% when compared with PCA, RPCA and SJSPCA, respectively. Similarly, for FAR values, the reduces of T^2 statistics are 0.33%, 0.12%, 0.11%, the reduces of *SPE* statistics are 0.92%, 0.36%, 0.42%. To appreciate the comparable performance achieved by LRPCA, the monitoring results for fault 10 are presented in Fig. 4. Fault 10 is a random variation in the feed C temperature in stream 4. The variation in the feed C temperature causes a change in the conditions of the stripper and then the condenser. It shows that the random variation in the feed C temperature can be detected by PCA, RPCA and SJSPCA but with a low sensitivity. However, by combining the robust idea and a hypergraph Laplacian regularizer, our proposed LRPCA could improve the detection in some degree.

6. Conclusion

In this paper, we propose a novel Laplacian regularized robust principal component analysis framework, in which we explicitly consider the noise reduction by introducing a sparse term. In particular, a hypergraph Laplacian regularizer is integrated into the RPCA objective function so as to capture the geometric structure information. An efficient ADMM is established with convergence analysis. Numerical comparisons between our proposed LRPCA and state-of-art methods, on a simulation example and the TE process, are presented to demonstrate its effectiveness. For these reasons, we believe that our proposed LRPCA is a valuable method for process monitoring.

CRediT authorship contribution statement

Xianchao Xiu: Conceptualization, Writing - original draft, Writing - review & editing, Data curation, Validation, Formal analysis, Investigation, Methodology, Software, Visualization, Resources. **Ying Yang:** Supervision, Conceptualization, Project administration, Funding acquisition. **Lingchen Kong:** Supervision, Conceptualization, Writing - original draft. **Wanquan Liu:** Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the anonymous referees for their comments and suggestions, which have greatly improved the paper. This work was supported in part by the National Natural Science Foundation of China (61633001, U1713223) and the 111 Project of China (B16002).

References

[1] S.X. Ding, Data-Driven Design of Fault Diagnosis and Fault-Tolerant Control Systems, Springer, 2014.
[2] K. Liu, Z. Fei, B. Yue, J. Liang, H. Lin, Adaptive sparse principal component analysis for enhanced process monitoring and fault isolation, Chemometr. Intell. Lab. Syst. 146 (2015) 426–436.
[3] Y. Liu, J. Zeng, L. Xie, S. Luo, H. Su, Structured joint sparse principal component analysis for fault detection and isolation, IEEE Trans. Ind. Inf. 15 (5) (2018) 2721–2731.
[4] L. Xie, X. Lin, J. Zeng, Shrinking principal component analysis for enhanced process monitoring and fault isolation, Ind. Eng. Chem. Res. 52 (49) (2013) 17475–17486.
[5] J. Zhu, Z. Ge, Z. Song, Distributed parallel PCA for modeling and monitoring of large-scale plant-wide processes with big data, IEEE Trans. Ind. Inf. 13 (4) (2017) 1877–1885.
[6] U. Kruger, Y. Zhou, X. Wang, D. Rooney, J. Thompson, Robust partial least squares regression: Part i, algorithmic developments, J. Chemometr.: J. Chemometr. Soc. 22 (1) (2008) 1–13.
[7] K. Peng, K. Zhang, B. You, J. Dong, Z. Wang, A quality-based nonlinear fault diagnosis framework focusing on industrial multimode batch processes, IEEE Trans. Ind. Electron. 63 (4) (2016) 2615–2624.
[8] B. Jiang, D. Huang, X. Zhu, F. Yang, R.D. Braatz, CaNonical variate analysis-based contributions for fault identification, J. Process Control 26 (2015) 17–25.
[9] Z. Chen, S.X. Ding, T. Peng, C. Yang, W. Gui, Fault detection for non-gaussian processes using generalized canonical correlation analysis and randomized algorithms, IEEE Trans. Ind. Electron. 65 (2) (2017) 1559–1567.
[10] Y. Hu, Y. Wang, C. Zhao, A sparse fault degradation oriented fisher discriminant analysis (FDFDA) algorithm for faulty variable isolation and its industrial application, Control Eng. Pract. 90 (2019) 311–320.
[11] C. Shang, B. Huang, F. Yang, D. Huang, Slow feature analysis for monitoring and diagnosis of control performance, J. Process Control 39 (2016) 21–34.
[12] W. Yu, C. Zhao, Recursive exponential slow feature analysis for fine-scale adaptive processes monitoring with comprehensive operation status identification, IEEE Trans. Ind. Inf. 15 (6) (2018) 3311–3323.
[13] Z. Chai, C. Zhao, Enhanced random forest with concurrent analysis of static and dynamic nodes for industrial fault classification, IEEE Trans. Ind. Inf. 16 (1) (2020) 54–66.
[14] Z. Ge, X. Chen, Dynamic probabilistic latent variable model for process data modeling and regression application, IEEE Trans. Control Syst. Technol. 27 (1) (2017) 323–331.
[15] R. Raveendran, H. Kodamana, B. Huang, Process monitoring using a generalized probabilistic linear latent variable model, Automatica 96 (2018) 73–83.

[16] N. Martin, H. Maes, Multivariate Analysis, Academic Press London, 1979.
[17] L.H. Chiang, E.L. Russell, R.D. Braatz, Fault Detection and Diagnosis in Industrial Systems, Springer Science & Business Media, 2000.
[18] J. Shang, M. Chen, H. Ji, D. Zhou, Isolating incipient sensor fault based on recursive transformed component statistical analysis, J. Process Control 64 (2018) 112–122.
[19] B. Cai, L. Huang, M. Xie, Bayesian networks in fault diagnosis, IEEE Trans. Ind. Inf. 13 (5) (2017) 2227–2240.
[20] R. Jiang, H. Fei, J. Huan, Anomaly localization for network data streams with graph joint sparse PCA, in: Proceedings of the 17th International Conference on Knowledge Discovery and Data Mining, ACM, 2011, pp. 886–894.
[21] Y. Liu, G. Zhang, B. Xu, Compressive sparse principal component analysis for process supervisory monitoring and fault detection, J. Process Control 50 (2017) 1–10.
[22] Z. Yan, Y. Yao, Variable selection method for fault isolation using least absolute shrinkage and selection operator (LASSO), Chemometr. Intell. Lab. Syst. 146 (2015) 136–146.
[23] H. Zou, T. Hastie, R. Tibshirani, Sparse principal component analysis, J. Comput. Graph. Statist. 15 (2) (2006) 265–286.
[24] S. Yi, Z. Lai, Z. He, Y. Cheung, Y. Liu, Joint sparse principal component analysis, Pattern Recognit. 61 (2017) 524–536.
[25] C. Li, B. Zhang, C. Chen, Q. Ye, J. Han, G. Guo, R. Ji, Deep manifold structure transfer for action recognition, IEEE Trans. Image Process. 28 (9) (2019) 4646–4658.
[26] M. Hubert, P.J. Rousseeuw, K. Vanden Branden, ROBPCA: a new approach to robust principal component analysis, Technometrics 47 (1) (2005) 64–79.
[27] W.A. Stahel, Robuste Schätzungen: Infinitesimale Optimalität und Schätzungen von Kovarianzmatrizen (Ph.D. thesis), ETH Zurich, 1981.
[28] D.L. Donoho, Breakdown properties of multivariate location estimators, Technical Report, Harvard University, Boston, 1982.
[29] M. Hubert, T. Reynkens, E. Schmitt, T. Verdonck, Sparse PCA for high-dimensional data with outliers, Technometrics 58 (4) (2016) 424–434.
[30] E.J. Candès, X. Li, Y. Ma, J. Wright, Robust principal component analysis?, J. ACM 58 (3) (2011) 1–37.
[31] B. Recht, M. Fazel, P.A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, SIAM Rev. 52 (3) (2010) 471–501.
[32] T. Bouwmans, A. Sobral, S. Javed, S.K. Jung, E.-H. Zahzah, Decomposition into low-rank plus additive matrices for background/foreground separation: A review for a comparative evaluation with a large-scale dataset, Comp. Sci. Rev. 23 (2017) 1–71.
[33] Z. Yan, C. Chen, Y. Yao, C.-C. Huang, Robust multivariate statistical process monitoring via stable principal component pursuit, Ind. Eng. Chem. Res. 55 (14) (2016) 4011–4021.
[34] K. Pearson, Liii. On lines and planes of closest fit to systems of points in space, Lond. Edinb. Dublin Philos. Mag. J. Sci. 2 (11) (1901) 559–572.
[35] J. Liu, S. Ji, J. Ye, Multi-task feature learning via efficient $\ell_{2,1}$-norm minimization, in: Proceedings of the Twenty-fifth Conference on Uncertainty in Artificial Intelligence, AUAI Press, 2009, pp. 339–348.
[36] D. Zhou, J. Huang, B. Schölkopf, Learning with hypergraphs: Clustering, classification, and embedding, Adv. Neural Inf. Process. Syst. (2007) 1601–1608.
[37] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, et al., Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends® Mach. Learn. 3 (1) (2011) 1–122.
[38] D.L. Donoho, De-noising by soft-thresholding, IEEE Trans. Inform. Theory 41 (3) (1995) 613–627.
[39] X. Xiu, W. Liu, L. Li, L. Kong, Alternating direction method of multipliers for nonconvex fused regression problems, Comput. Statist. Data Anal. 136 (2019) 59–71.
[40] J.J. Downs, E.F. Vogel, A plant-wide industrial process control problem, Comput. Chem. Eng. 17 (3) (1993) 245–255.