Fault Detection Using Structured Joint Sparse Nonnegative Matrix Factorization

Xianchao Xiu, *Member, IEEE*, Jun Fan, Ying Yang, *Senior Member, IEEE*, and Wanquan Liu, *Senior Member, IEEE*

***Abstract*— Nonnegative matrix factorization (NMF) is an efficient dimension reduction technique, which has been extensively used in the fields, such as image processing, automatic control, and machine learning. The application to fault detection (FD) is still not investigated sufficiently. To improve the performance of NMF-based FD approaches, this article proposes a novel FD approach using the structured joint sparse NMF (SJSNMF) for non-Gaussian processes. The basic idea of SJSNMF is to incorporate the graph Laplacian to preserve the relationship between process variables and operation units and introduce the joint sparsity to exploit row-wise sparsity of the latent variables. Technically, an optimization algorithm based on the alternating direction method of multipliers (ADMM) is established. To detect the fault, two test statistical metrics are adopted and the kernel density estimation (KDE) is calculated to estimate the control limit. The effectiveness of the proposed SJSNMF is verified on the benchmark Tennessee Eastman process (TEP) and the cylinder–piston assembly of diesel engines.**

***Index Terms*— Fault detection (FD), joint sparsity, kernel density estimation (KDE), non-Gaussian processes, nonnegative matrix factorization (NMF).**

I. Introduction

FAULT detection (FD) has attracted much attention so as to ensure the safe and reliable operation of industrial processes. In general, the FD methods can be sorted into model-, signal-, and data-driven ones [1]–[4]. Due to the rapid development of data acquisition and processing technology, data-driven FD becomes dominant. Multivariate statistical process monitoring, as a class of data-driven methods, has been successfully applied to improve the performance of FD. The most popular ones include principal component analysis (PCA) [5], [6], partial least squares (PLS) [7], [8], independent component analysis (ICA) [9], [10], Fisher discriminant analysis (FDA) [11], [12], canonical correlation analysis (CCA) [13], [14], and nonnegative matrix factorization (NMF) [15], [16]. The successful applications of PCA-based FD largely depend on the assumption that the latent variables and the residual signals obey the Gaussian distribution. However, in actual processes, this assumption cannot be guaranteed because there exists non-Gaussian distribution in the latent variables. Unlike PCA and other FD methods, NMF does not need any assumption except for nonnegativity. Therefore, it provides a new way for dealing with FD.

In fact, NMF can be tracked back to [17], [18], which is designed to exploit a low-rank matrix approximation with nonnegativity. At present, NMF has been widely used in image processing [19], document clustering [20], and so on (see [21] for a recent review of models, algorithms, and applications). However, the application of NMF for FD has not been fully studied yet as described next. Li *et al.* [15] was the first to apply NMF-based statistical monitoring strategy for FD in non-Gaussian processes. To detect faulty variables, the T^2 (also N^2) and squared prediction error (SPE) statistical metrics were built and the kernel density estimation (KDE) was applied to estimate the control limit. It was verified that NMF could achieve better performance than PCA and ICA on the benchmark Tennessee Eastman process (TEP). After that, several variants of NMF have been developed by adding constraints to the classical NMF. Yang *et al.* [22] considered a discriminative NMF formulation based on the class information, which enhanced the discriminant ability of the learned basis matrix. The effectiveness was illustrated on a diesel engine. Later, Wang *et al.* [23] constructed an adaptive partition NMF framework by integrating nonfixed subblock NMF models. It contained global FD, data-adaptive partition, and block FD. Recently, Ren *et al.* [16] incorporated deep learning methods with NMF and proposed deep NMF. Different from other nonlinear NMF-based approaches, deep NMF could learn the nonlinear function for input data automatically via a deep autoencoder framework.

However, all the abovementioned generations of NMF share some of the following shortcomings for FD applications. One shortcoming is that most of these generations ignore the sparse information of process variables. It is also demonstrated that using all process variables may select highly sensitive variables [24], [25]. Thus, a sparse variant of NMF can be considered to extract the discriminative basis for data representation. Instead of using ℓ_1-norm-induced sparsity, the joint sparsity is more effective to learn global structures of process data [26]. This is because the joint sparsity can reduce the influence of outliers and remove undesirable variables by forcing the variables associated with all zero rows to be excluded from the

Manuscript received March 1, 2021; accepted March 10, 2021. Date of publication March 18, 2021; date of current version March 26, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61633001, Grant 12001019, and Grant 11801130. The Associate Editor coordinating the review process was Dr. Gaigai Cai. *(Corresponding author: Ying Yang.)*

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

Jun Fan is with the Institute of Mathematics, Hebei University of Technology, Tianjin 300401, China (e-mail: junfan@hebut.edu.cn).

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

Digital Object Identifier 10.1109/TIM.2021.3067218

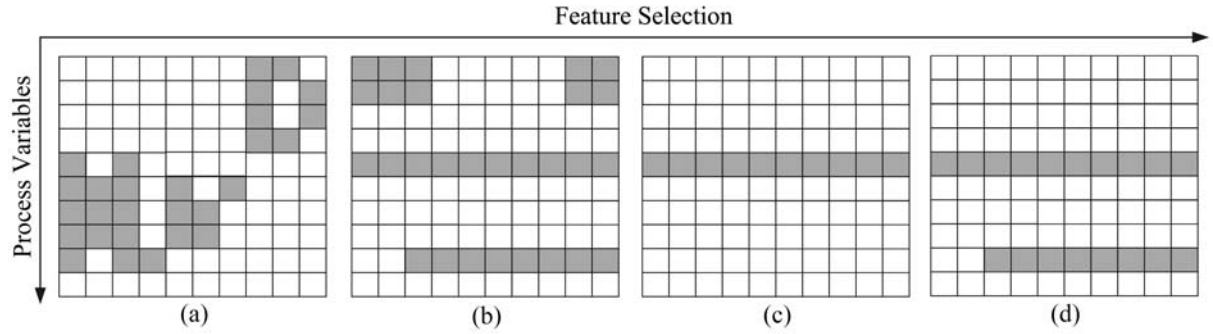


Fig. 1. Matrices obtained from different kinds of NMF approaches. (a) ℓ_1-norm regularized NMF. (b) $\ell_{2,1}$-norm regularized NMF. (c) $\ell_{2,0}$-norm constrained NMF with $s = 1$. (d) $\ell_{2,0}$-norm constrained NMF with $s = 2$.

projection space. The other shortcoming is that the underlying cause–effect relationship between process variables cannot be preserved. If all the process data are composed of several linear spaces, NMF can easily extract the underlying structures [5], [27]. However, modern process data are often highly complex, which leads to the classical NMF that cannot be well. From the perspective of representation, the combination of NMF and data geometry is necessary. Therefore, there is a need to derive a kind of NMF, which can not only extract meaningful process variables but also preserve the cause–effect relationship.

In this article, a novel NMF framework based on the graph Laplacian regularization and joint sparse constraint is proposed. For the convenience of description, it is abbreviated to structured joint sparse NMF (SJSNMF). The graph Laplacian can combine the structured variable correlation information to learn the cause–effect relationship between process variables. The joint sparse constraint can prompt row-wise sparsity to improve the interpretability of process variables. Overall, the graph Laplacian regularization is applied to preserve local structures and the joint sparse constraint is embedded to retain global structures. Both of them contribute to a better representation of the process data. Different from $\ell_{2,1}$-norm regularization, $\ell_{2,0}$-norm constraint is adopted in the proposed SJSNMF. Fig. 1 compares the matrices obtained from different NMF-type approaches, where white blocks represent zero elements and gray blocks represent nonzero elements. In FD applications, each row corresponds to a process variable. The zero rows indicate that these variables are unimportant or not related to the fault and thus can be dropped out. By imposing $\ell_{2,0}$-norm constraint and choosing different sparsity level s, the proposed approach can accurately determine the selected process variables.

The main contributions of this article can be summarized in the following three aspects.

1) A generalized NMF-based FD model is constructed to extract low-dimensional representation.
2) An efficient minimization algorithm is proposed to solve the proposed SJSNMF.
3) An online monitoring procedure is developed by using two test statistical metrics and KDE.

The rest of this article is outlined as follows. Section II reviews the classical NMF and its representative variants. Section III designs an efficient algorithm to solve the proposed SJSNMF. Section IV presents an offline training and online testing procedure based on SJSNMF. Section V reports the numerical results on the benchmark data sets by comparing with PCA, ICA, and other NMF-based approaches. Finally, Section VI concludes this article.

II. Related Works

This section begins with some basic notations, followed by NMF and its variants.

A. Notation

Let $\mathbb{R}^m$ and $\mathbb{R}^{m\times N}$ be the sets of vectors and matrices, where m denotes the number of variables and N denotes the number of samples. For a matrix $\mathbf{X} \in \mathbb{R}^{m\times N}$, $\mathbf{x}_i$ denotes its ith row and x_{ij} denotes its ijth element. The $\ell_{2,1}$-norm as an extension of ℓ_1-norm is given by

$$\|\mathbf{X}\|_{2,1} = \sum_{i=1}^{m} \|\mathbf{x}_i\|_2. \tag{1}$$

The $\ell_{2,0}$-norm, which comes from the ℓ_0-norm (number of nonzero rows), can be denoted as $\|\mathbf{X}\|_{2,0}$, that is

$$\|\mathbf{X}\|_{2,0} = |\{i : \|\mathbf{x}_i\|_2 \neq 0\}| \tag{2}$$

where $|\cdot|$ represents the cardinality. In addition, $\mathbf{X}^{\mathrm{T}}$ and $\mathbf{X}^{-1}$ are denoted as its transpose matrix and inverse matrix, respectively. Finally, for two matrices $\mathbf{X}$ and $\mathbf{Y}$, the inner product can be given by

$$\langle \mathbf{X}, \mathbf{Y} \rangle = \mathrm{tr}(\mathbf{X}^{\mathrm{T}}\mathbf{Y}) = \sum_{i=1}^{m}\sum_{j=1}^{N} x_{ij}y_{ij}. \tag{3}$$

B. Nonnegative Matrix Factorization

Given the nonnegative process data matrix $\mathbf{X}$ of m process variables and N samples, that is

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mN} \end{bmatrix} \in \mathbb{R}^{m\times N} \tag{4}$$

where each column represents a sample of the data. NMF aims to seek two nonnegative low-rank matrices $\mathbf{W} \in \mathbb{R}^{m\times k}$ and $\mathbf{H} \in \mathbb{R}^{k\times N}$ such that

$$\mathbf{X} \approx \mathbf{WH} \tag{5}$$

in which k is the reduced dimension satisfying $(m+N)k < mN$. Obviously, any sample of $\mathbf{X}$ can be represented by a nonnegative linear combination of row vectors in matrix $\mathbf{W}$. In this regard, $\mathbf{W}$ and $\mathbf{H}$ are often explained as a basis matrix and a coefficient matrix, respectively.

To find an approximate factorization, a proper loss function should be defined to evaluate the quality of the approximation. Lee and Seung [17] considered the Euclidean distance square to measure the approximation between $\mathbf{WH}$ and $\mathbf{X}$, as shown in the following mathematical formulation:

$$\begin{aligned} &\min_{\mathbf{W},\mathbf{H}} \ \frac{1}{2}\|\mathbf{X}-\mathbf{WH}\|_F^2 \\ &\text{s.t.} \ \ \mathbf{W}\geq 0, \quad \mathbf{H}\geq 0 \end{aligned} \tag{6}$$

where $\|\cdot\|_F$ is the Frobenius norm and ≥ 0 means that each element is nonnegative.

In the past decades, a number of optimization methods have been proposed for solving NMF-related problems, such as multiplicative updates (MUs) [28], projected gradient descent (PGD) [29], alternative nonnegative least squares (ANLS) [30], and proximal alternating nonnegative least squares (PANLS) [31]. However, these algorithms converge slowly. Boyd *et al.* [32] proposed to solve NMF-type problems by the alternating direction method of multipliers (ADMM). The convergence was analyzed in [33]. Compared with MU, PGD, ANLS, and PANLS, the ADMM always demonstrates encouraging fast convergence and good numerical performance. This owes to the fact that the resulting subproblems can be solved by fast solvers.

C. Variants of NMF

Apart from classical NMF model (6), various extensions have been considered to enhance the decomposition performance by adding regularization terms or constraints to NMF. Among them, the most popular one is sparsity-constrained NMF (SNMF)

$$\begin{aligned} &\min_{\mathbf{W},\mathbf{H}} \ \frac{1}{2}\|\mathbf{X}-\mathbf{WH}\|_F^2 \\ &\text{s.t.} \ \ \mathbf{W}\geq 0, \quad \mathbf{H}\geq 0, \ \ \|\mathbf{H}\|_0 \leq s \end{aligned} \tag{7}$$

in which s is the sparsity level to control the number of variables. Compared with the existing sparsity regularized NMF models, this kind of SNMF is easier to tune the number of selected variables [34], [35].

Another popular one is graph regularized NMF (GNMF). As is discussed before, NMF may fail to explore the intrinsic geometric structures of process data. To handle this situation, Cai *et al.* [36] proposed GNMF by introducing the regularization term $\text{tr}(\mathbf{H}^{\mathrm{T}}\mathbf{LH})$. The corresponding optimization model is given by

$$\begin{aligned} &\min_{\mathbf{W},\mathbf{H}} \ \frac{1}{2}\|\mathbf{X}-\mathbf{WH}\|_F^2+\lambda\text{tr}(\mathbf{H}^{\mathrm{T}}\mathbf{LH}) \\ &\text{s.t.} \ \ \mathbf{W}\geq 0, \quad \mathbf{H}\geq 0. \end{aligned} \tag{8}$$

Here, $\mathbf{L}$ is the graph Laplacian matrix learned from $\mathbf{X}$, and $\lambda > 0$ is a regularized parameter for balancing the graph Laplacian and the reconstruction fidelity. It is well proved that GNMF can obtain a better representation that keeps the local geometry. More specifically, the nearest neighbor graph is applied to model the local structures, and then, the graph is merged into NMF to preserve the prior information.

Algorithm 1 Structured Joint Sparse NMF

Input: Given process date $\mathbf{X}$, graph Laplacian $\mathbf{L}$, parameter $\lambda > 0$, penalty parameters $\beta_1, \beta_2 > 0$.
Initialize: $(\mathbf{W}_0, \mathbf{H}_0, \mathbf{U}_0, \mathbf{Y}_0, \mathbf{A}_0, \mathbf{B}_0)$, set $k = 0$.
While not converged **do**
1: Update $\mathbf{W}_{k+1}$ by (14) and (15).
2: Update $\mathbf{H}_{k+1}$ by (17).
3: Update $\mathbf{U}_{k+1}$ by (19) and (20).
4: Update $\mathbf{Y}_{k+1}$ by (22).
5: Update $\mathbf{A}_{k+1} = \mathbf{A}_k - \beta_1(\mathbf{W}_{k+1}\mathbf{H}_{k+1} - \mathbf{Y}_{k+1})$.
6: Update $\mathbf{B}_{k+1} = \mathbf{B}_k - \beta_2(\mathbf{H}_{k+1} - \mathbf{U}_{k+1})$.
End While
Output: $(\mathbf{W}_{k+1}, \mathbf{H}_{k+1})$.

III. Structured Joint Sparsity NMF

This section presents the proposed SJSNMF-based FD approach. On the one hand, a novel NMF model is introduced by incorporating the graph Laplacian regularization and joint sparse constraint. On the other hand, an optimization algorithm is developed to solve the proposed approach.

A. Model Construction

Although SNMF has been successfully applied in many areas, it ignores the similarities between process variables. In this case, joint sparsity can be replaced to improve the representation, such as joint sparse PCA [5] and joint sparse CCA [37]. In NMF literature, efforts to incorporate joint sparse structures are quite limited (see [38]–[40]). However, all of them only focus on joint sparsity regularized NMF. Inspired by the above related works, a novel NMF model is constructed as follows:

$$\begin{aligned} &\min_{\mathbf{W},\mathbf{H}} \ \frac{1}{2}\|\mathbf{X}-\mathbf{WH}\|_F^2+\lambda\text{tr}(\mathbf{H}^{\mathrm{T}}\mathbf{LH}) \\ &\text{s.t.} \ \ \mathbf{W}\geq 0, \ \ \mathbf{H}\geq 0, \ \ \|\mathbf{H}\|_{2,0} \leq s. \end{aligned} \tag{9}$$

In fact, it integrates the joint sparse ($\ell_{2,0}$-norm) constraint and the graph Laplacian regularization, which can be seen an extension of models (7) and (8).

B. Optimization Algorithm

Due to the joint sparse constraint $\|\mathbf{H}\|_{2,0} \leq s$, it is difficult to obtain the global solution. Therefore, an approximate solution is pursued. To apply the ADMM, two auxiliary variables $\mathbf{Y}$ and $\mathbf{U}$ are first introduced and (9) can be reformulated as

$$\begin{aligned} &\min_{\mathbf{W},\mathbf{H},\mathbf{U},\mathbf{Y}} \ \frac{1}{2}\|\mathbf{X}-\mathbf{Y}\|_F^2+\lambda\text{tr}(\mathbf{H}^{\mathrm{T}}\mathbf{LH}) \\ &\text{s.t.} \quad \mathbf{WH}=\mathbf{Y}, \quad \mathbf{H}=\mathbf{U}, \\ &\qquad \mathbf{W}\geq 0, \quad \mathbf{U}\geq 0, \ \ \|\mathbf{U}\|_{2,0} \leq s. \end{aligned} \tag{10}$$

The associated augmented Lagrangian is

$$\mathcal{L}(\mathbf{W},\mathbf{H},\mathbf{U},\mathbf{Y},\mathbf{A},\mathbf{B}) = \frac{1}{2}\|\mathbf{X}-\mathbf{Y}\|_F^2 + \lambda \mathrm{tr}(\mathbf{H}^{\mathrm{T}}\mathbf{L}\mathbf{H}) - \langle \mathbf{A}, \mathbf{W}\mathbf{H}-\mathbf{Y}\rangle + \frac{\beta_1}{2}\|\mathbf{W}\mathbf{H}-\mathbf{Y}\|_F^2 - \langle \mathbf{B}, \mathbf{H}-\mathbf{U}\rangle + \frac{\beta_2}{2}\|\mathbf{H}-\mathbf{U}\|_F^2 \quad (11)$$

where $\mathbf{A}$ and $\mathbf{B}$ are the Lagrange multipliers for equality constraints and $\beta_1, \beta_2 > 0$ are penalty parameters. The above minimization also requires constraints $\mathcal{M} = \{\mathbf{W} \mid \mathbf{W} \geq 0\}$ and $\mathcal{N} = \{\mathbf{U} \mid \mathbf{U} \geq 0, \|\mathbf{U}\|_{2,0} \leq s\}$. Under the framework of the ADMM, all the variables can be updated one-by-one in a Gauss–Seidel manner. The iterative scheme for solving (10) can be described as

$$\begin{cases} \mathbf{W}_{k+1} = \arg\min_{\mathbf{W}\in\mathcal{M}} \ \mathcal{L}(\mathbf{W}, \mathbf{H}_k, \mathbf{U}_k, \mathbf{Y}_k, \mathbf{A}_k, \mathbf{B}_k) \\ \mathbf{H}_{k+1} = \arg\min_{\mathbf{H}} \mathcal{L}(\mathbf{W}_{k+1}, \mathbf{H}, \mathbf{U}_k, \mathbf{Y}_k, \mathbf{A}_k, \mathbf{B}_k) \\ \mathbf{U}_{k+1} = \arg\min_{\mathbf{U}\in\mathcal{N}} \ \mathcal{L}(\mathbf{W}_{k+1}, \mathbf{H}_{k+1}, \mathbf{U}, \mathbf{Y}_k, \mathbf{A}_k, \mathbf{B}_k) \\ \mathbf{Y}_{k+1} = \arg\min_{\mathbf{Y}} \ \mathcal{L}(\mathbf{W}_{k+1}, \mathbf{H}_{k+1}, \mathbf{U}_{k+1}, \mathbf{Y}, \mathbf{A}_k, \mathbf{B}_k) \\ \mathbf{A}_{k+1} = \mathbf{A}_k - \beta_1(\mathbf{W}_{k+1}\mathbf{H}_{k+1} - \mathbf{Y}_{k+1}) \\ \mathbf{B}_{k+1} = \mathbf{B}_k - \beta_2(\mathbf{H}_{k+1} - \mathbf{U}_{k+1}). \end{cases} \quad (12)$$

Next, the solutions with respect to each variable will be discussed in detail.

1) Updating $\mathbf{W}$ can be obtained by minimizing

$$\min_{\mathbf{W}\in\mathcal{M}} \ \frac{1}{2}\|\mathbf{W}\mathbf{H}_k - \mathbf{Y}_k - \mathbf{A}_k/\beta_1\|_F^2. \quad (13)$$

It can be solved by projected gradient descent, which includes

$$\mathbf{W}_{k+1/2} = (\mathbf{Y}_k + \mathbf{A}_k/\beta_1)\mathbf{H}_k^{\mathrm{T}}(\mathbf{H}_k\mathbf{H}_k^{\mathrm{T}})^{-1} \quad (14)$$

and

$$\mathbf{W}_{k+1} = \max\{\mathbf{W}_{k+1/2}, 0\}. \quad (15)$$

2) Updating $\mathbf{H}$ can be simplified to

$$\min_{\mathbf{H}} \ \lambda \mathrm{tr}(\mathbf{H}^{\mathrm{T}}\mathbf{L}\mathbf{H}) + \frac{\beta_1}{2}\|\mathbf{W}_{k+1}\mathbf{H} - \mathbf{Y}_k - \mathbf{A}_k/\beta_1\|_F^2 + \frac{\beta_2}{2}\|\mathbf{H} - \mathbf{U}_k - \mathbf{B}_k/\beta_2\|_F^2. \quad (16)$$

By simple calculation, one can obtain the minimizer $\mathbf{H}_{k+1}$ by solving the following linear equation:

$$(2\lambda\mathbf{L} + \beta_1\mathbf{W}_{k+1}^{\mathrm{T}}\mathbf{W}_{k+1} + \beta_2\mathbf{I})\mathbf{H}_{k+1} = \beta_1\mathbf{W}_{k+1}^{\mathrm{T}}\mathbf{Y}_k + \beta_2\mathbf{U}_k + \mathbf{W}_{k+1}^{\mathrm{T}}\mathbf{A}_k + \mathbf{B}_k. \quad (17)$$

3) Updating $\mathbf{U}$ can be expressed by

$$\min_{\mathbf{U}\in\mathcal{N}} \ \frac{1}{2}\|\mathbf{U} - (\mathbf{H}_{k+1} - \mathbf{B}_k/\beta_2)\|_F^2. \quad (18)$$

It is a projection onto $\mathcal{N}$. Based on Proposition 3.1 [41], one can obtain $\mathbf{U}_{k+1}$ via

$$\mathbf{U}_{k+1/2} = \max\{\mathbf{H}_{k+1} - \mathbf{B}_k/\beta_2, 0\} \quad (19)$$

and

$$\mathbf{U}_{k+1} = \Pi(\mathbf{U}_{k+1/2}) \quad (20)$$

where $\Pi(\cdot)$ is a simple row-wise truncation operation, meaning that if $\|\mathbf{U}_{k+1/2}\|_{2,0} \leq s$, then $\mathbf{U}_{k+1} = \mathbf{U}_{k+1/2}$; otherwise, $\mathbf{U}_{k+1}$ is taken as a s-sparse (in row-level) matrix whose nonzero rows are selected from the s largest rows according to the norm $\|\cdot\|_2$ of $\mathbf{U}_{k+1/2}$.

4) Updating $\mathbf{Y}$ can be written as

$$\min_{\mathbf{Y}} \ \frac{1}{2}\|\mathbf{Y} - \mathbf{X}\|_F^2 + \frac{\beta_1}{2}\|\mathbf{Y} - \mathbf{W}_{k+1}\mathbf{H}_{k+1} + \mathbf{A}_k/\beta_1\|_F^2 \quad (21)$$

which derives the following solution:

$$\mathbf{Y}_{k+1} = \frac{1}{1+\beta_1}(\mathbf{X} + \beta_1\mathbf{W}_{k+1}\mathbf{H}_{k+1} - \mathbf{A}_k). \quad (22)$$

Finally, the proposed algorithm for solving problem (10) [also (9)] can be summarized in Algorithm 1.

IV. Monitoring Strategy Using SJSNMF

This section presents an online SJSNMF-based monitoring strategy. Different from PCA and ICA, normalization is not necessary since NMF requires each element to be nonnegative. After solving optimization problem (9), $\mathbf{W}$ and $\mathbf{H}$ are obtained. As is pointed in [15], the coefficient matrix $\mathbf{H}$ reflects the status of industrial processes. From the objective function of (9), the reconstruction of $\mathbf{H}$ is calculated by

$$\hat{\mathbf{H}} = (\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{X}. \quad (23)$$

Hence, the reconstruction of $\mathbf{X}$ is formulated as

$$\hat{\mathbf{X}} = \mathbf{W}\hat{\mathbf{H}}. \quad (24)$$

Two test statistical metrics, i.e., T^2 statistic and SPE statistic, are commonly employed to monitor industrial processes [42]. For NMF-based FD, the corresponding T^2 statistic and SPE statistic are constructed as follows:

$$T^2 = \hat{\mathbf{H}}^{\mathrm{T}}\hat{\mathbf{H}}$$
$$\mathrm{SPE} = (\mathbf{X} - \hat{\mathbf{X}})^{\mathrm{T}}(\mathbf{X} - \hat{\mathbf{X}}). \quad (25)$$

Because the assumption of Gaussian distribution is not satisfied, the control limits cannot be directly approximated by a specific distribution. According to [16], the KDE is used to estimate the control limits for T^2 statistic and SPE statistic. Suppose that x_i $(i = 1, 2, \ldots, N)$ is an independent and identically distributed sample set. Let $p(x)$ be the density function, which is given by

$$p(x) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right) \quad (26)$$

where h denotes the bandwidth and $K(\cdot)$ denotes the kernel function satisfying

$$\int_{-\infty}^{\infty} K(x)\mathrm{d}x = 1, \quad K(x) \geq 0. \quad (27)$$

Accordingly, for given confidence limit α, the control limit can be estimated by

$$P(J < J_{\mathrm{th}}) = \int_{-\infty}^{\mathrm{th}} p(J)\mathrm{d}J = \alpha. \quad (28)$$

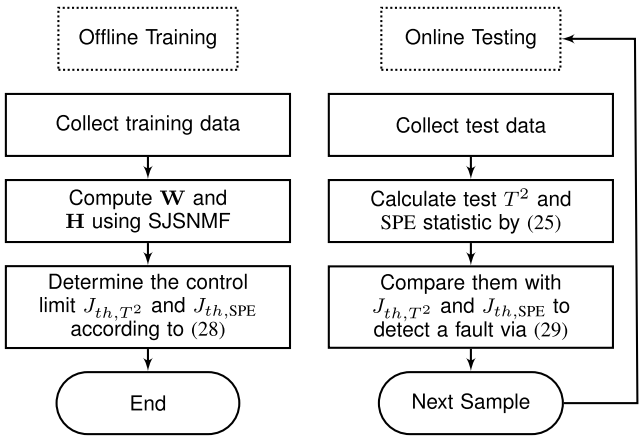


Fig. 2. Framework of monitoring strategy.

The corresponding control limits for T^2 statistic and SPE statistic are obtained and denoted as J_{th,T^2} and $J_{\mathrm{th,SPE}}$, respectively. To detect the fault, an FD decision logic should be well addressed. If the test statistic violates the control limit, a fault occurs, otherwise fault-free. Therefore, the detection logic can be defined as

$$\begin{cases} T^2 < J_{\mathrm{th},T^2} \text{ and } \mathrm{SPE} < J_{\mathrm{th,SPE}} \Rightarrow \text{fault-free} \\ T^2 \geq J_{\mathrm{th},T^2} \text{ or } \mathrm{SPE} \geq J_{\mathrm{th,SPE}} \Rightarrow \text{faulty}. \end{cases} \tag{29}$$

At the end of this section, the offline training and online testing framework is established in Fig. 2, which contains two stages. In the first stage, the control limit is determined. In the second stage, the test statistic is estimated to determine whether a fault occurs.

V. Application Studies

In this section, a comprehensive performance comparison is conducted to illustrate the advantages of the proposed SJSNMF on the benchmark TEP and the cylinder–piston assembly of diesel engines. Three NMF-based approaches, namely, NMF [15], SNMF [39], and GNMF [36], are compared. It is worth pointing out that some nonlinear versions of NMF, such as kernel NMF [43] and deep NMF [16], are not considered here, because most of them cost plenty of time to be computed. Besides, two state-of-the-art data-driven approaches, i.e., PCA and ICA, are reported.

A. Setup

For PCA, the contribution of retained principal components accounts for more than 85%. For ICA- and NMF-based approaches, the KDE is adopted to estimate the control limits for T^2 statistic and SPE statistic. The type of kernel function has little effect on the control limit. Normally, the Gaussian-type kernel function is adopted. However, the parameter h should be tuned carefully. Inspired by [44], the optimal bandwidth h is set as $h = 1.06\sigma\ \mathrm{n}^{-1/5}$ with σ being the standard deviation of $\mathbf{X}$. In addition, the cumulative percentage α is set as 0.99.

For the proposed SJSNMF, the parameter λ is determined by the fivefold cross-validation technique, which contains the following four steps: 1) determine the candidate set $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$; 2) separate the fault-free data set into five disjoint but same-length parts; 3) choose one part for testing and the rest for training; and 4) repeat the above procedures and select the part that gives the smallest value. For joint sparsity level s, there is no standard of choice. In principle, it is chosen as the desired faulty variables. If s is too small, the performance will degrade. If s is too large, the variable selection will be meaningless. In all experiments, s is initialized with a smaller value and then increased until the best FD performance is achieved. The graph Laplacian matrix $\mathbf{L}$ is chosen by setting $k = 10$ according to [5]. The maximum number of iterations is set as 500, and the criterion of stopping is defined as

$$\mathrm{RelErr} = \max\left\{ \frac{\|\mathbf{W}_{k+1} - \mathbf{W}_k\|_F}{\|\mathbf{W}_k\|_F + 1}, \frac{\|\mathbf{H}_{k+1} - \mathbf{H}_k\|_F}{\|\mathbf{H}_k\|_F + 1} \right\} \tag{30}$$

with RelErr being 10^{-3}.

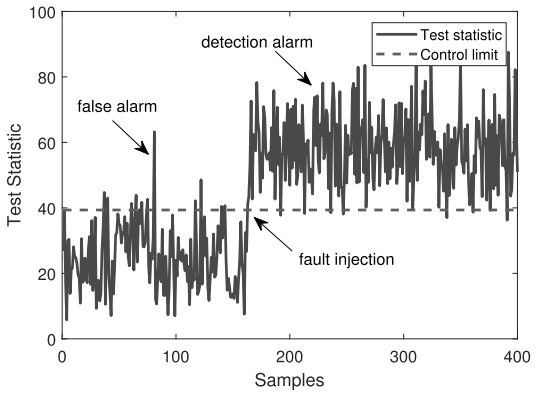


Fig. 3. Illustration of evaluation index.

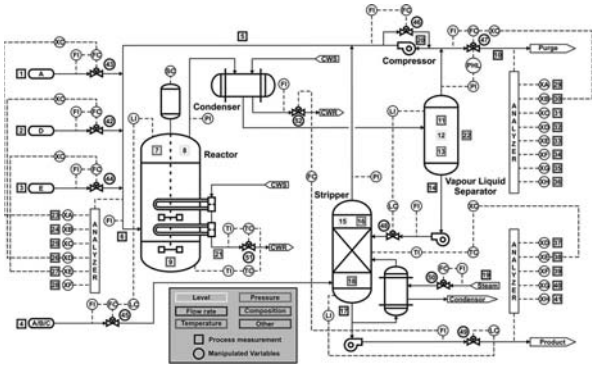


Fig. 4. Flowchart of the TEP.

To measure the performance of FD, two indicators are widely used: detection alarm and false alarm [42]. In Fig. 3, a fault is injected at the 161st sample, and thus, a detection alarm is considered as an event that occurs when there exists a fault, while a false alarm is deemed as an event that triggers an alarm when there exists no fault. Therefore, the FD rate (FDR) and false alarm rate (FAR) can be defined as follows:

$$\begin{aligned} \mathrm{FDR} &= \frac{\text{number of samples } (T^2 \geq J_{\mathrm{th},T^2} \mid f \neq 0)}{\text{total of samples } (f \neq 0)} \\ \mathrm{FAR} &= \frac{\text{number of samples } (T^2 \geq J_{\mathrm{th},T^2} \mid f = 0)}{\text{total of samples } (f = 0)}. \end{aligned} \tag{31}$$

It is easy to see that FDR denotes the percentage of faulty samples detected in the faulty cases, and FAR means the

TABLE I
FDR VALUES OF PCA, ICA, NMF, SNMF, GNMF, AND SJSNMF

Fault No.	PCA		ICA		NMF		SNMF		GNMF		SJSNMF	
	T^2	SPE	T^2	SPE	T^2	SPE	T^2	SPE	T^2	SPE	T^2	SPE
IDV(1)	99.13%	99.25%	99.25%	99.88%	99.25%	99.13%	99.38%	98.25%	99.38%	99.05%	**99.63%**	**99.88%**
IDV(2)	98.38%	95.75%	98.50%	98.00%	98.38%	97.05%	98.75%	97.50%	98.38%	97.88%	**98.38%**	**99.13%**
IDV(3)	0.88%	2.63%	1.75%	1.13%	4.25%	3.88%	5.63%	5.25%	5.75%	4.38%	**5.13%**	**6.63%**
IDV(4)	20.88%	100%	76.25%	90.88%	86.50%	97.63%	95.75%	99.50%	94.88%	99.13%	**100%**	**100%**
IDV(5)	24.13%	20.88%	20.25%	21.75%	24.13%	22.38%	24.63%	23.75%	24.88%	23.05%	**24.63%**	**30.25%**
IDV(6)	99.13%	100%	99.25%	100%	99.75%	100%	99.50%	100%	99.13%	100%	**99.88%**	**100%**
IDV(7)	100%	100%	100%	99.88%	100%	100%	100%	100%	100%	100%	**100%**	**100%**
IDV(8)	96.88%	83.63%	91.75%	93.63%	92.50%	88.13%	93.63%	90.38%	93.50%	89.25%	**94.75%**	**96.50%**
IDV(9)	1.75%	1.75%	2.63%	1.13%	3.63%	2.88%	5.75%	4.13%	5.25%	3.38%	**6.13%**	**4.75%**
IDV(10)	29.63%	25.75%	30.13%	35.25%	31.00%	38.25%	33.13%	36.50%	34.38%	34.75%	**36.88%**	**48.63%**
IDV(11)	20.63%	74.88%	53.63%	63.38%	52.25%	70.63%	54.63%	74.13%	53.38%	72.13%	**56.50%**	**75.88%**
IDV(12)	92.38%	98.50%	93.88%	98.25%	92.05%	95.75%	93.13%	96.88%	92.13%	95.25%	**94.63%**	**98.50%**
IDV(13)	93.63%	95.25%	95.38%	94.25%	93.63%	94.75%	94.25%	95.13%	94.38%	94.88%	**96.13%**	**97.75%**
IDV(14)	89.25%	98.88%	92.50%	99.88%	90.25%	92.38%	91.75%	95.88%	94.13%	93.25%	**97.63%**	**99.13%**
IDV(15)	1.38%	3.00%	3.75%	1.13%	4.38%	4.50%	5.63%	5.75%	5.25%	5.63%	**6.50%**	**8.38%**
IDV(16)	13.50%	27.38%	32.00%	36.25%	39.25%	38.38%	41.88%	39.15%	40.25%	43.38%	**42.05%**	**42.75%**
IDV(17)	46.25%	55.38%	57.00%	52.63%	45.38%	58.50%	52.13%	61.13%	54.25%	60.88%	**58.75%**	**69.13%**
IDV(18)	89.25%	90.13%	90.25%	90.00%	89.13%	90.00%	91.25%	92.63%	90.05%	91.75%	**92.13%**	**91.75%**
IDV(19)	1.88%	3.52%	4.25%	2.75%	2.38%	4.25%	2.75%	4.75%	3.13%	5.88%	**5.50%**	**7.25%**
IDV(20)	21.75%	29.75%	21.38%	37.38%	22.75%	33.38%	25.88%	36.05%	28.38%	35.63%	**29.13%**	**41.88%**
IDV(21)	39.25%	47.25%	37.50%	29.50%	39.25%	42.13%	41.50%	42.88%	41.05%	44.63%	**42.75%**	**49.50%**
Average	51.43%	59.70%	57.20%	59.38%	57.62%	60.67%	59.57%	61.89%	59.61%	61.63%	**61.30%**	**65.13%**

TABLE II
FAR VALUES OF PCA, ICA, NMF, SNMF, GNMF, AND SJSNMF

Fault No.	PCA		ICA		NMF		SNMF		GNMF		SJSNMF	
	T^2	SPE	T^2	SPE	T^2	SPE	T^2	SPE	T^2	SPE	T^2	SPE
IDV(1)	0.00%	0.63%	0.63%	0.00%	0.00%	0.63%	0.00%	0.00%	0.00%	0.63%	**0.00%**	**0.00%**
IDV(2)	1.25%	0.63%	0.63%	0.63%	1.25%	0.63%	0.63%	0.63%	0.63%	0.63%	**0.00%**	**0.63%**
IDV(3)	0.00%	1.25%	0.63%	2.50%	0.00%	1.25%	0.00%	0.63%	0.00%	0.00%	**0.00%**	**0.00%**
IDV(4)	0.63%	1.25%	0.00%	0.63%	0.63%	1.88%	0.63%	1.25%	0.63%	0.63%	**0.63%**	**0.00%**
IDV(5)	0.63%	1.25%	1.88%	0.63%	1.25%	1.25%	0.63%	0.00%	0.00%	0.63%	**0.00%**	**0.00%**
IDV(6)	0.00%	1.88%	0.00%	1.25%	0.00%	0.63%	0.00%	0.63%	0.00%	0.63%	**0.00%**	**0.63%**
IDV(7)	0.00%	1.25%	0.63%	1.88%	0.00%	1.88%	0.00%	0.00%	0.00%	0.63%	**0.00%**	**0.00%**
IDV(8)	0.00%	0.63%	0.00%	0.63%	0.63%	1.25%	0.00%	0.63%	0.00%	0.63%	**0.00%**	**0.63%**
IDV(9)	1.88%	1.88%	2.50%	0.63%	1.25%	1.88%	1.25%	0.63%	0.63%	1.25%	**0.00%**	**0.63%**
IDV(10)	0.00%	0.63%	0.00%	0.63%	0.00%	0.63%	0.00%	0.00%	0.00%	0.00%	**0.00%**	**0.00%**
IDV(11)	0.63%	2.50%	0.63%	1.25%	1.88%	2.50%	1.25%	1.25%	0.63%	1.25%	**0.63%**	**1.25%**
IDV(12)	0.00%	1.25%	0.63%	0.63%	0.00%	1.88%	0.00%	1.25%	0.00%	0.63%	**0.00%**	**0.63%**
IDV(13)	0.63%	0.00%	0.63%	0.00%	0.63%	0.00%	0.63%	0.00%	0.63%	0.00%	**0.63%**	**0.00%**
IDV(14)	0.00%	1.25%	1.88%	0.63%	0.00%	1.88%	0.00%	1.25%	0.00%	1.25%	**0.00%**	**0.63%**
IDV(15)	0.00%	1.25%	0.00%	2.50%	0.00%	2.50%	0.00%	1.25%	0.00%	1.25%	**0.00%**	**0.00%**
IDV(16)	3.75%	1.88%	4.38%	1.75%	3.75%	2.50%	2.50%	1.25%	1.88%	1.88%	**0.63%**	**1.25%**
IDV(17)	1.25%	2.50%	0.63%	1.25%	1.25%	2.50%	0.63%	2.50%	1.25%	2.50%	**0.63%**	**1.88%**
IDV(18)	0.00%	2.50%	0.00%	1.25%	0.00%	2.50%	0.00%	1.25%	0.00%	1.25%	**0.00%**	**0.63%**
IDV(19)	0.00%	0.63%	0.00%	0.00%	0.63%	0.63%	0.00%	0.63%	0.00%	0.63%	**0.00%**	**0.00%**
IDV(20)	0.00%	1.25%	1.25%	0.63%	0.00%	1.88%	0.00%	0.63%	0.00%	0.63%	**0.00%**	**0.63%**
IDV(21)	0.00%	3.13%	0.00%	2.50%	0.00%	3.13%	0.00%	2.50%	0.00%	2.50%	**0.00%**	**1.25%**
Average	0.21%	1.40%	0.81%	1.04%	0.63%	1.61%	0.39%	0.83%	0.30%	0.93%	**0.14%**	**0.51%**

percentage of fault-free samples identified as faults in the fault-free cases. When a fault occurs, higher FDR values or lower FAR values imply better monitoring performance. The FDR value is 1 if all the faulty samples are detected, while the FAR value is 0 if all the fault-free samples are undetected.

B. Application on the TEP

1) Data Set: The TEP, as a benchmark open-source data set, is developed according to a practical industry process [45]. Even though the TEP has been used for decades, it is still one of the most popular data sets for testing different kinds of FD techniques. Fig. 4 shows the flowchart of the TEP, including five significant units and eight components. The TEP has 11 manipulation variables, 22 process variables, and 19 analysis variables. Among them, 33 variables (11 manipulation variables and 22 process variables) are considered [37].

The TEP contains one fault-free data set and 21 faulty data sets. The fault-free data set is used for offline training, whereas the faulty data sets are collected for online testing. For faulty data sets, a fault is injected at the 161st sampling time point, which means that the first 160 samples are fault-free and the rest 800 samples are faulty. In this study, the number of retained principal components is set as 17 for PCA, and s is set as 10 for the proposed SJSNMF.

2) Monitoring Results: The computational results of PCA, ICA, NMF, SNMF, GNMF, and SJSNMF are reported in Tables I and II. Moreover, the values of the proposed SJSNMF are highlighted in boldface. In Table I, most of the FDR values are relatively large, except for faults IDV(3),

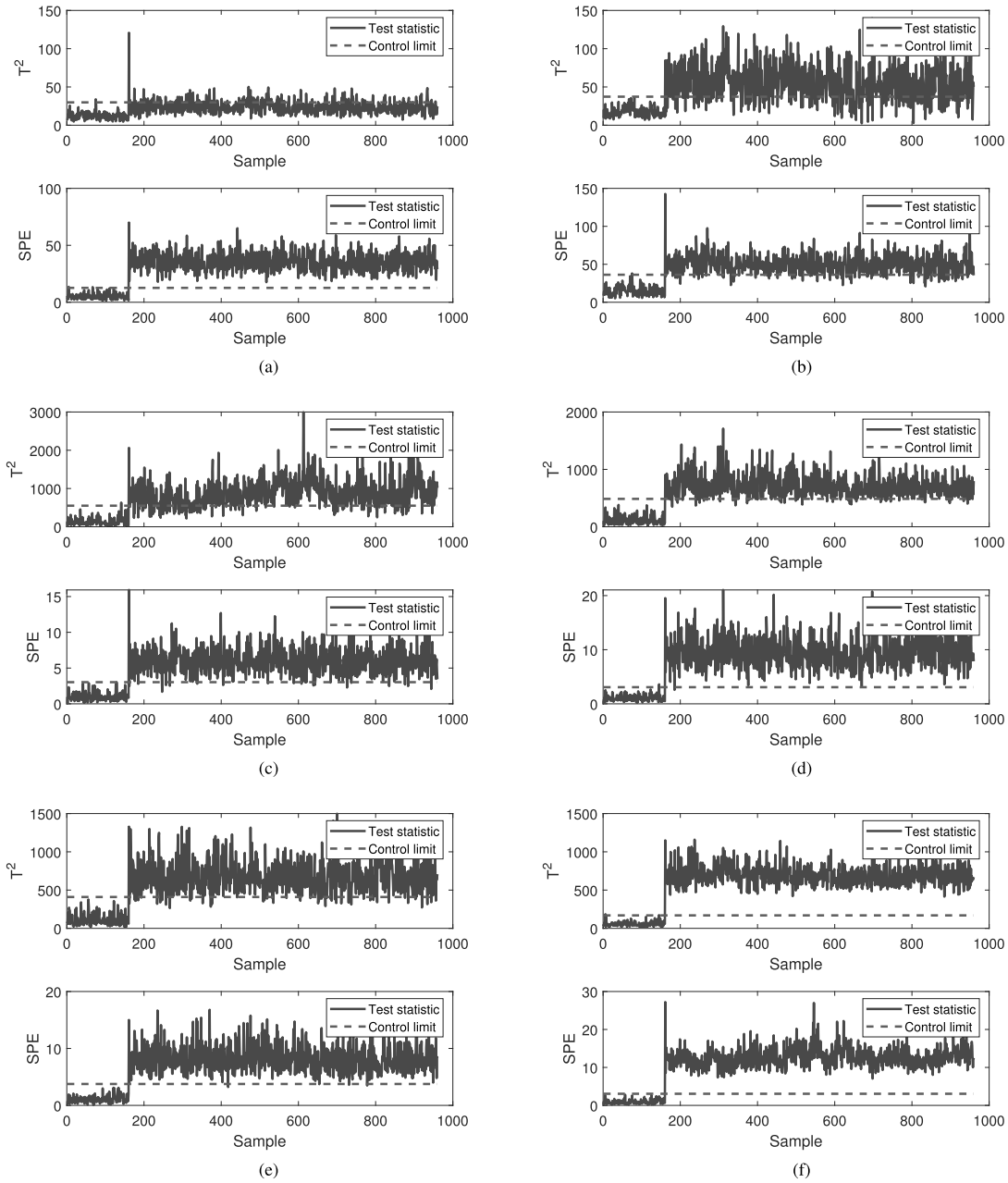


Fig. 5. Detection results of fault IDV(4) in the TEP. (a) PCA. (b) ICA. (c) NMF. (d) SNMF. (e) GNMF. (f) SJSNMF.

IDV(9), IDV(15), and IDV(19) because the mean or variance of process variables has a smaller change in these faults [46]. In Table II, the results of FAR are relatively low, which shows that all the approaches are beneficial for fault alarms.

According to Table I, some conclusions can be easily obtained.

1) Compared with PCA and ICA, NMF-based approaches (i.e., NMF, SNMF, GNMF, and SJSNMF) achieve better performance in terms of FDR. In average, the gains of T^2 statistical values are 6.14%, 8.14%, 8.18%, and 9.87% than PCA and 0.42%, 2.37%, 2.41%, and 4.10% than ICA. Consequently, the superiority of NMF-based FD is well documented.
2) Both SNMF and GNMF improve the values of FDR, which verifies that the sparsity and graph Laplacian are useful. The reason lies in exploring the prior structures of data and making the FD more stable. In particular, for fault IDV(4), the T^2 values of SNMF and GNMF are 9.25% and 8.38% higher than that of NMF.
3) For all the selected faults, the proposed SJSNMF works better than or as good as other NMF-based approaches. It indicates that the integration of graph Laplacian regularization and joint sparse constraint is promising. The graph Laplacian can preserve the structured correlation relationship between process variables, and the joint sparse constraint can retain the fault variables and excludes fault-free ones.

In order to visualize the monitoring results clearly, two different types of faults are chosen, i.e., IDV(4) and IDV(10). Fault IDV(4) involves a step change in the reactor cooling water inlet temperature, which is relatively easy to be detected. In Fig. 5, for SPE statistic, all the FD approaches have

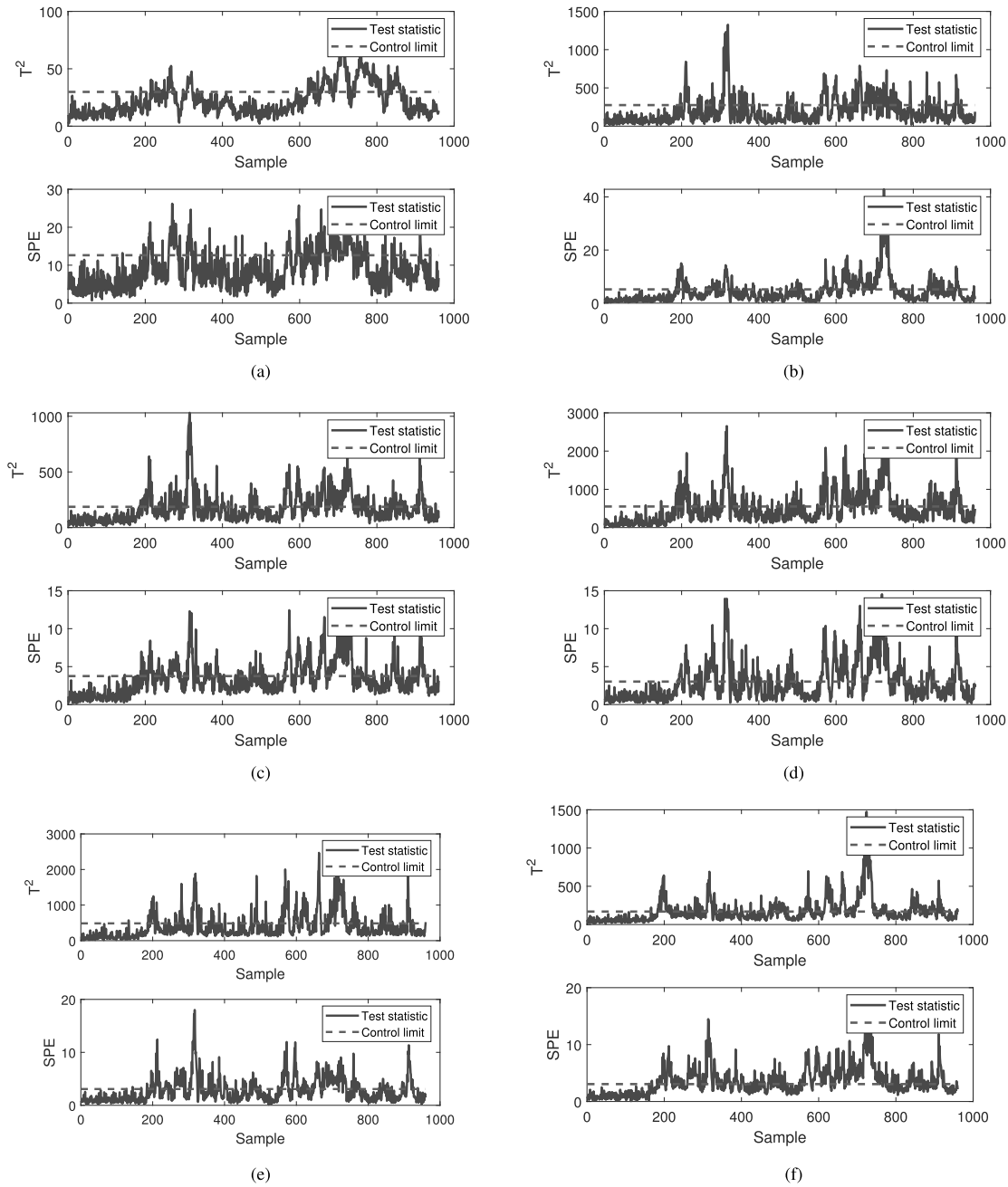


Fig. 6. Detection results of fault IDV(10) in the TEP. (a) PCA. (b) ICA. (c) NMF. (d) SNMF. (e) GNMF. (f) SJSNMF.

successfully detected the fault at around 161st sample, and the values of FDR are over 90%. It reflects that NMF-based approaches can also complete the detection task successfully. Moreover, for T^2 statistic, NMF-based approaches show obvious advantages. In particular, SJSNMF is able to detect the fault completely. Fault IDV(10) is a random variation in the feed C temperature in stream 4, which is relatively hard to be detected. Fig. 6 shows the monitoring test statistics. In comparison, SJSNMF has a better detection capacity than others for detection alarms. For example, by carefully observing the values between 600 and 800 samples for T^2 statistic and SPE statistic, the percentages that violated the control limit of SJSNMF are larger than those of PCA, ICA, and other NMF's variants. It confirms that SJSNMF has a more satisfactory detection performance.

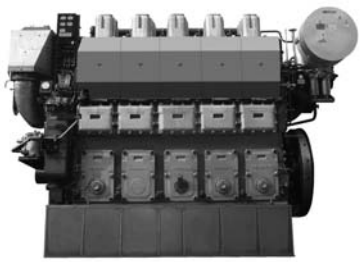


Fig. 7. Cylinder–piston assembly of diesel engines.

C. Application on the Cylinder–Piston Assembly of Diesel Engines

1) Data Set: Cylinder–piston assembly is a basic component of diesel engines, which bears the heaviest load in the

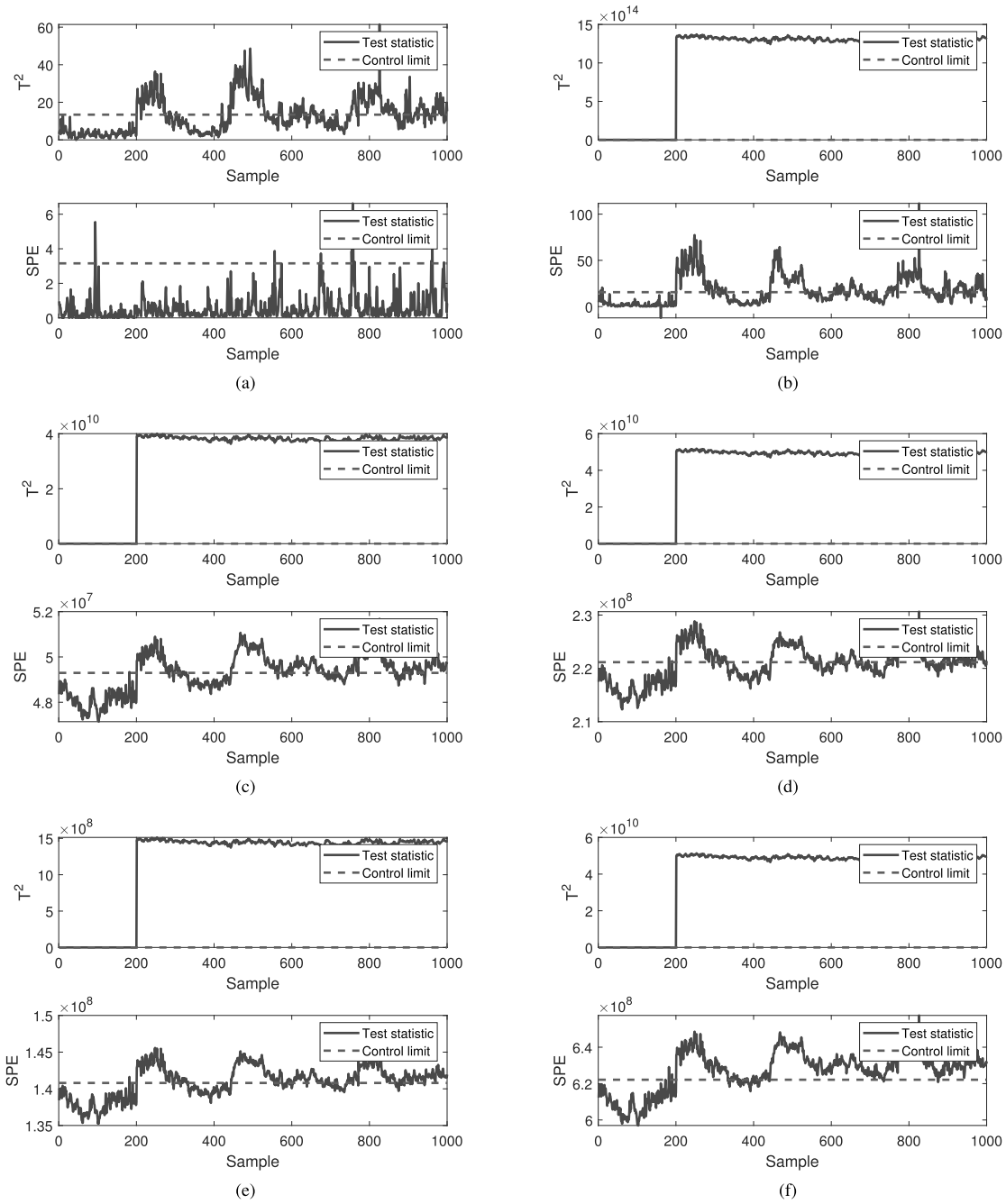


Fig. 8. Detection results of the cylinder–piston assembly. (a) PCA. (b) ICA. (c) NMF. (d) SNMF. (e) GNMF. (f) SJSNMF.

TABLE III
SELECTED VARIABLES IN THE CYLINDER–PISTON ASSEMBLY

No.	Description
1	cylinder exhaust gas temperature
2	cylinder cooling oil inlet and outlet temperature difference
3	cylinder cooling oil inlet pressure
4	cylinder jacket cooling water inlet and outlet temperature difference
5	cylinder jacket cooling water inlet pressure

process. This study collects the cylinder–piston assembly data from a two-stroke five-cylinder low-speed marine diesel engine [47] (refer to Fig. 7 for a visual illustration). For each cylinder, five measurement variables are selected, which are shown in Table III. In addition, a number of 1000 samples are chosen for training and the fault is injected at the 201st sample.

2) Monitoring Results: The monitoring results are presented in Fig. 8. It is not hard to conclude that ICA- and NMF-type approaches have very outstanding detection results than PCA because the fault can be immediately and consistently detected through the T^2 statistic. Moreover, for SPE statistic, SJSNMF has more numbers violated the control limit than those of other approaches. It appears that the proposed SJSNMF delivers a more encouraging FD ability, mainly for the reason that it automatically learns the latent structure information in the process monitoring data.

D. Convergence Analysis

In order to illustrate the convergence of SJSNMF, the relative differences of $\mathbf{W}$ and $\mathbf{H}$ on fault IDV(10) are shown

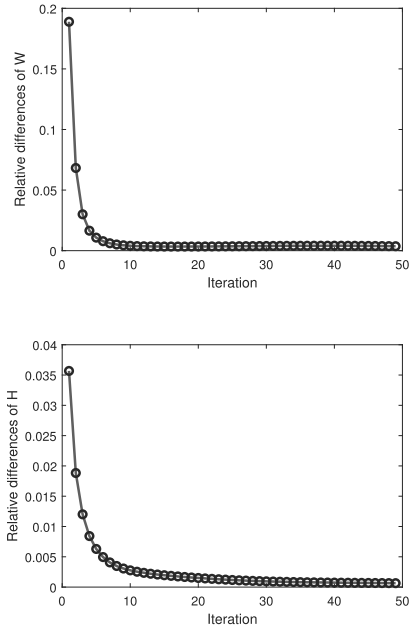


Fig. 9. Relative differences of **W** and **H**.

in Fig. 9. As is shown in the figure, both of the relative differences decrease rapidly when the number of iterations increases, which reflects that the proposed approach converges after finite iterations.

VI. Conclusion

FD is a hot topic in complicated industrial processes. In this article, a novel structured joint sparsity regularized NMF framework (SJSNMF) is proposed. The graph Laplacian is regularized to automatically preserve the local geometry prior. The joint sparsity is constrained to learn the global low-level sparse structure. Therefore, SJSNMF can both capture the intrinsic nonlinear geometric information and represent the global low-dimensional structures in process data. The extensive experimental results on two different data sets, including benchmark TEP and real-world cylinder–piston assembly of diesel engines, show the effectiveness of the proposed approach.

Naturally, several interesting and important issues need to be further explored. First, the current work is only developed intuitively based on FD induction, and hence, the computational complexity and convergence of the proposed algorithm should be investigated theoretically. Second, although the proposed SJSNMF achieves a significant improvement in terms of FDR and FAR, the choice of s has a great influence on the FD performance. Actually, how to exploit the physical properties of real-world processes to select a proper parameter is still an open problem. Finally, the extension to fault isolation and quality-related FD deserve considered.

Acknowledgment

The authors would like to thank the associate editor and three anonymous reviewers for their numerous insightful comments, which have greatly improved this article. They would also like to thank Dr. Yuhan Zhang from Peking University, Beijing, China, for gracefully providing the data of cylinder–piston assembly of diesel engines.

References

[1] 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.

[2] X. Jin, Y. Sun, Z. Que, Y. Wang, and T. W. S. Chow, "Anomaly detection and fault prognosis for bearings," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 9, pp. 2046–2054, Sep. 2016.

[3] J. Sun, C. Yan, and J. Wen, "Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 1, pp. 185–195, Jan. 2018.

[4] L. Song, H. Wang, and P. Chen, "Vibration-based intelligent fault diagnosis for roller bearings in low-speed rotating machinery," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 8, pp. 1887–1899, Aug. 2018.

[5] 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.

[6] Y. Tao, H. Shi, B. Song, and S. Tan, "A novel dynamic weight principal component analysis method and hierarchical monitoring strategy for process fault detection and diagnosis," *IEEE Trans. Ind. Electron.*, vol. 67, no. 9, pp. 7994–8004, Sep. 2020.

[7] S. Yin, S. X. Ding, A. Haghani, H. Hao, and P. Zhang, "A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process," *J. Process Control*, vol. 22, no. 9, pp. 1567–1581, Oct. 2012.

[8] D. Zhou, G. Li, and S. J. Qin, "Total projection to latent structures for process monitoring," *AIChE J.*, vol. 56, no. 1, pp. 168–178, 2010.

[9] X. Jin, J. Fan, and T. W. S. Chow, "Fault detection for rolling-element bearings using multivariate statistical process control methods," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 9, pp. 3128–3136, Sep. 2019.

[10] J. E. Garcia-Bracamonte, J. M. Ramirez-Cortes, P. Gomez-Gil, J. D. J. Rangel-Magdaleno, H. Peregrina-Barreto, and V. Alarcon-Aquino, "An approach on MCSA-based fault detection using independent component analysis and neural networks," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 5, pp. 1353–1361, May 2019.

[11] M. Van and H.-J. Kang, "Wavelet kernel local Fisher discriminant analysis with particle swarm optimization algorithm for bearing defect classification," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 12, pp. 3588–3600, Dec. 2015.

[12] J. Liu, C. Song, J. Zhao, and P. Ji, "Manifold-preserving sparse graph-based ensemble FDA for industrial label-noise fault classification," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 2621–2634, Jun. 2020.

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

[14] B. Song, H. Shi, S. Tan, and Y. Tao, "Multi-subspace orthogonal canonical correlation analysis for quality related plant wide process monitoring," *IEEE Trans. Ind. Informat.*, early access, Aug. 7, 2020, doi: 10.1109/TII.2020.3015034.

[15] X.-B. Li, Y.-P. Yang, and W.-D. Zhang, "Fault detection method for non-Gaussian processes based on non-negative matrix factorization," *Asia–Pacific J. Chem. Eng.*, vol. 8, no. 3, pp. 362–370, May 2013.

[16] Z. Ren, W. Zhang, and Z. Zhang, "A deep nonnegative matrix factorization approach via autoencoder for nonlinear fault detection," *IEEE Trans. Ind. Informat.*, vol. 16, no. 8, pp. 5042–5052, Aug. 2020.

[17] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

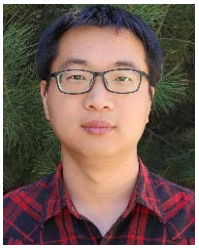
[18] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111–126, Jun. 1994.

[19] B. Qin, C. Hu, and S. Huang, "Target/Background classification regularized nonnegative matrix factorization for fluorescence unmixing," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 4, pp. 874–889, Apr. 2016.

[20] W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," in *Proc. 26th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR)*, 2003, pp. 267–273.

[21] Y.-X. Wang and Y.-J. Zhang, "Nonnegative matrix factorization: A comprehensive review," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 6, pp. 1336–1353, Jun. 2013.

[22] Y.-S. Yang, A.-B. Ming, Y.-Y. Zhang, and Y.-S. Zhu, "Discriminative non-negative matrix factorization (DNMF) and its application to the fault diagnosis of diesel engine," *Mech. Syst. Signal Process.*, vol. 95, pp. 158–171, Oct. 2017.
[23] Y. Wang, S.-M. Yuan, D. Ling, Y.-B. Zhao, X.-G. Gu, and B.-Y. Li, "Fault monitoring based on adaptive partition non-negative matrix factorization for non-Gaussian processes," *IEEE Access*, vol. 7, pp. 32783–32795, 2019.
[24] J. Yu and H. Liu, "Sparse coding shrinkage in intrinsic time-scale decomposition for weak fault feature extraction of bearings," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 7, pp. 1579–1592, Jul. 2018.
[25] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1693–1702, Jul. 2017.
[26] J. Kim, R. D. C. Monteiro, and H. Park, "Group sparsity in nonnegative matrix factorization," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2012, pp. 851–862.
[27] 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.
[28] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2001, pp. 556–562.
[29] C.-J. Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural Comput.*, vol. 19, no. 10, pp. 2756–2779, Oct. 2007.
[30] J. Kim and H. Park, "Fast nonnegative matrix factorization: An Active-Set-Like method and comparisons," *SIAM J. Sci. Comput.*, vol. 33, no. 6, pp. 3261–3281, Jan. 2011.
[31] C. Zhang, L. Jing, and N. Xiu, "A new active set method for non-negative matrix factorization," *SIAM J. Sci. Comput.*, vol. 36, no. 6, pp. A2633–A2653, Jan. 2014.
[32] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.
[33] D. Hajinezhad, T.-H. Chang, X. Wang, Q. Shi, and M. Hong, "Nonnegative matrix factorization using ADMM: Algorithm and convergence analysis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 4742–4746.
[34] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *J. Mach. Learn. Res.*, vol. 5, pp. 1457–1469, Dec. 2004.
[35] X. Yuan, P. Li, and T. Zhang, "Gradient hard thresholding pursuit," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 6027–6069, 2017.
[36] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.
[37] 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.
[38] C. G. Tsinos, A. A. Rontogiannis, and K. Berberidis, "Distributed blind hyperspectral unmixing via joint sparsity and low-rank constrained non-negative matrix factorization," *IEEE Trans. Comput. Imag.*, vol. 3, no. 2, pp. 160–174, Jun. 2017.
[39] X. Wang, Y. Zhong, L. Zhang, and Y. Xu, "Spatial group sparsity regularized nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6287–6304, Nov. 2017.
[40] S. Yang, C. Hou, C. Zhang, and Y. Wu, "Robust non-negative matrix factorization via joint sparse and graph regularization for transfer learning," *Neural Comput. Appl.*, vol. 23, no. 2, pp. 541–559, Aug. 2013.
[41] L.-L. Pan, N.-H. Xiu, and S.-L. Zhou, "On solutions of sparsity constrained optimization," *J. Oper. Res. Soc. China*, vol. 3, no. 4, pp. 421–439, Dec. 2015.
[42] S. X. Ding, *Data-Driven Design of Fault Diagnosis and Fault-Tolerant Control Systems*. London, U.K.: Springer-Verlag, 2014.
[43] L. Zhai, Y. Zhang, S. Guan, Y. Fu, and L. Feng, "Nonlinear process monitoring using kernel nonnegative matrix factorization," *Can. J. Chem. Eng.*, vol. 96, no. 2, pp. 554–563, Feb. 2018.
[44] A. R. Mugdadi and I. A. Ahmad, "A bandwidth selection for kernel density estimation of functions of random variables," *Comput. Statist. Data Anal.*, vol. 47, no. 1, pp. 49–62, Aug. 2004.
[45] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," *Comput. Chem. Eng.*, vol. 17, no. 3, pp. 245–255, Mar. 1993.
[46] L. H. Chiang, E. L. Russell, and R. D. Braatz, "Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis," *Chemometric Intell. Lab. Syst.*, vol. 50, no. 2, pp. 243–252, Mar. 2000.
[47] Y. Zhang, Y. Yang, H. Li, H. Han, and Z. He, "Layer-wise fault diagnosis based on key performance indicator for cylinder-piston assembly of diesel engine," in *Proc. IEEE 15th Int. Conf. Control Autom. (ICCA)*, Jul. 2019, pp. 307–312.

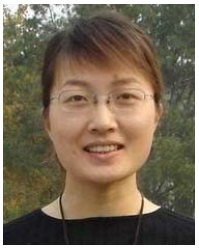


Xianchao Xiu (Member, IEEE) received the Ph.D. degree in operations research from Beijing Jiaotong University, Beijing, China, in 2019.

He is currently a Post-Doctoral Researcher with the State Key Laboratory for Turbulence and Complex Systems, Peking University, Beijing. His current research interests include large-scale sparse optimization, signal processing, machine learning and deep learning, and data-driven fault detection.

Jun Fan received the Ph.D. degree in operations research from Beijing Jiaotong University, Beijing, China, in 2017.

He is currently a Faculty Member with the Institute of Mathematics, Hebei University of Technology, Tianjin, China. His current research interests include mathematical programming and its applications, high-dimensional statistics, and machine learning.



Ying Yang received the Ph.D. degree in control theory from Peking University, Beijing, China, in 2002.

From January 2003 to November 2004, she has worked as a Post-Doctoral Researcher with Peking University. From 2005 to 2014, she was an Associate Professor with the Department of Mechanics and Engineering Science, College of Engineering, Peking University, where she has been a Full Professor since 2014. Her current research interests include robust and optimal control, nonlinear systems control, numerical analysis, fault detection, and fault-tolerant systems.

Wanquan Liu (Senior Member, IEEE) received the B.S. degree in applied mathematics from Qufu Normal University, Jining, China, in 1985, the M.S. degree in control theory and operation research from the Chinese Academy of Sciences, Beijing, China, in 1988, and the Ph.D. degree in electrical engineering from Shanghai Jiaotong University, Shanghai, China, in 1993.

He once held the Australian Research Council (ARC) Fellowship, the U2000 Fellowship, and the Japan Society for the Promotion of Science (JSPS) Fellowship; and attracted research funds from different resources over 2.4 million dollars. He is currently an Associate Professor with the Department of Computing, Curtin University, Perth, WA, Australia. His current research interests include large-scale pattern recognition, signal processing, machine learning, and control systems.

Dr. Liu is on editorial board of nine international journals.