Structured Pattern Discovery Using Dictionary Learning for Incipient Fault Detection and Isolation

Yi Liu, Jiusun Zeng, Zidong Wang, *Fellow, IEEE*, Weiguo Sheng, Chuanhou Gao, *Senior Member, IEEE*, Qi Xie, and Lei Xie

***Abstract*—To address the challenges encountered by dictionary learning-based monitoring, this article presents a novel pattern discovery scheme for detection and isolation of incipient faults that involves structured sparse coding and sequential dictionary augmentations. Through learning a basic dictionary for normal pattern and augmenting the low-dimensional sparse dictionaries for analyzing different fault patterns, the process signals can be decomposed into fault-free and fault-related components. To guarantee the in-statistical-control status of the fault-free part and improve detection sensitivity, a ℓ_2-penalty is imposed on the sum of coefficient vectors to ensure that the monitoring statistic related to the fault-free part will not exceed the control limit. In addition, two Frobenius norm penalties are imposed on the zero centered coefficient matrix and atom matrix to improve the robustness of signal decomposition. Instead of imposing ℓ_1-sparsity constraint on the atoms, a hard sparsity constraint is used to correctly select fault-related feature variables, so that fault patterns can be better revealed. The informative dictionaries are then incorporated into the moving window-based monitoring strategy, yielding a fault detection and isolation scheme suitable for incipient faults. The superior performance of our proposed approach is validated by application studies involving a numerical example and two practical industrial processes.**

***Index Terms*—Dictionary learning-based monitoring, manifold structure preservation, pattern discovery, statistical properties embedding, structured sparse coding.**

I. Introduction

To maintain modern industrial processes in healthy and available states, it is urgent to develop advanced monitoring technologies. Reviewing the evolution of these techniques, data-driven methods based on classical statistical models, such as principal component analysis (PCA), partial least squares, and canonical correlation analysis [1], have gained significant popularity, due to low implementation cost, effective monitoring results, and extensive application fields [2]. To monitor changes in process states, these methods combine the internal criteria, such as maximum variance or independence [3], with actual scenarios and specific goals, and project the data into a preset low-dimensional space, obtaining concentrated crucial information. However, it is pointed out by [4] that the complexity and diversity of process data may not be fully covered by low-dimensional space. With the aim of avoiding the loss of key information in dimension reduction, a feasible solution is to project the data into a high-dimensional space, and make use of the increased degrees of freedom to accommodate more details [5], thus obtaining the potential patterns and relationships within the process data.

The positive factors in high-dimensional projection for driving monitoring performance improvement primarily include: 1) decoupling the intertwined connections, 2) highlighting the subtle variations, and 3) extracting in-depth characteristics [6], which vigorously advance the application and development of machine learning techniques targeted at high-dimensional space modeling in this domain, such as kernel methods [7], deep learning [8], and dictionary learning. As a powerful signal processing technique, dictionary learning has simple model structure exhibits good process monitoring performance. In practice, dictionary learning finds an appropriate high-dimensional sparse representation for the original data, and performing signal decomposition using matrix theory and sparsity-inducing techniques [9]. Through the linear combination of representation coefficients, the overcomplete dictionary atoms can precisely reconstruct normal data, whilst generating significant reconstruction errors for fault detection [10].

Received 1 December 2024; revised 22 February 2025 and 17 March 2025; accepted 29 April 2025. This work was supported in part by the "Pioneer and Leading Goose + X" S&T Project of Zhejiang Province under Grant 2025C01022, in part by the Baima Lake Laboratory Joint Fund of Zhejiang Provincial Natural Science Foundation of China under Grant LBMHZ25F03001, and in part by the Hangzhou Joint Fund of Zhejiang Provincial Natural Science Foundation under Grant LHZSZ24F020002. Paper no. TII-24-6432. *(Corresponding author: Jiusun Zeng.)*

Yi Liu and Weiguo Sheng are with the School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China (e-mail: liuyi@hznu.edu.cn; wsheng@hznu.edu.cn).

Jiusun Zeng and Qi Xie are with the School of Mathematics, Hangzhou Normal University, Hangzhou 311121, China (e-mail: jszeng@hznu.edu.cn; qixie@hznu.edu.cn).

Zidong Wang is with the Department of Computer Science, Brunel University London, UB8 3PH London, U.K. (e-mail: zidong.wang@brunel.ac.uk).

Chuanhou Gao is with the School of Mathematical Sciences, Zhejiang University, Hangzhou 310027, China (e-mail: gaochou@zju.edu.cn).

Lei Xie is with the Institute of Cyber Systems & Control, Zhejiang University, Hangzhou 310027, China (e-mail: leix@csc.zju.edu.cn).

Digital Object Identifier 10.1109/TII.2025.3567271

By virtue of the model structure and the direct association between atoms and latent data characteristics, process modeling and performance enhancement based on dictionary learning, thus, have attracted widespread research attention in recent years. For example, to construct appropriate process models using dictionary learning, Peng et al. [11] embedded the manifold structure constraint into coefficients, improving the clustering ability of multimode data. To incorporate the newly generated modal data, Huang et al. [12] used atom similarity to guide the continuous update of dictionary. Based on multilevel knowledge graph, dictionary learning can be used to construct distributed plant-wide monitoring methods [13]. Kernel methods are also employed to assist dictionary learning in achieving more powerful representation capabilities, enabling nonlinear monitoring based on higher dimensional reconstruction residuals [14]. For time-varying process characteristics, it is suggested that both normal and uncontrolled samples within a moving window be included in online dictionary learning framework [15]. For better detectability, Ren et al. [10] iteratively selected effective samples from training dataset as atoms, which maintain the category properties and internal structure within the data. Combining dictionary learning with variational autoencoder, Zemouri et al. [16] integrated a series of reconstruction errors for the detection of abnormal vibration signals of hydroelectric generators. For dictionary learning-based fault isolation, Ning et al. [17] suggested fixing fault directions as the identity matrix and using them to augment the basic dictionary, forming the method of sparse contribution plot (SpCP). As well, the other classical isolation methods, such as reconstruction-based contribution (RBC) and iterative reconstruction-based contribution (IRBC) [18], were introduced into the developed dictionary learning-based methods [14], [19]. In addition, Wang et al. [20] integrated deep learning techniques to shape dictionary learning models into the industrial fault classifiers.

Despite the research progress, some important issues still remain for dictionary learning-based monitoring methods: 1) the employed sparse coding techniques, such as ℓ_0- or ℓ_1-norm regularization, may suppress subtle abnormal variation of coefficients, which are critical indicators of early-stage/incipient faults; 2) improving detection sensitivity by attending the reconstruction error may negatively affect the ability of dictionary coefficients to distinguish these incipient faults, and vice versa; 3) fixing fault direction or performing iterative reconstruction for isolation disregards variable interactions, potentially damages the internal structure and properties of the data, leading to inaccurate isolation results. To address these issues, this article proposes a novel structured augmented sparse dictionary learning (SASDL), by introducing a basic dictionary for fault-free components and an additional sparse dictionary for fault-related components of the reconstructed data. To guarantee in-statistical-control status of the fault-free part, stationary statistical properties and manifold structure constraint are embedded into two regularization terms of basic dictionary coefficients, and a hard sparsity constraint is used to correctly select fault-related feature variables, so that fault patterns can be better revealed. In order to further improve the monitoring performance, the informative dictionaries are then used to construct moving window-based monitoring statistics and an effective isolation model.

Compared with several moving window-based incipient fault detection methods, which enhance the detection sensitivity by significantly accumulating process variations [21], locally estimating statistical characteristics [22], or measuring data distribution differences [23], the proposed approach accurately represents the test data through sequential augmented dictionary learning, while simultaneously explicitly capturing systematic anomalies by fault pattern discovery for the purpose of decoupling the entangled relationships concealed in the data. This endows the proposed method with improved fault detection and isolation capability. The innovations and contributions of this article can be summarized as follows.

1) Decomposition of process signals into fault-free and fault-related dictionary components for detection of incipient faults: By augmenting the basic dictionary with sequential low-dimensional sparse dictionaries, process signals can be effectively decomposed into fault-free and fault-related components, so that fault information will not be dispersed. Also, with window-based aggregation, the sensitivity of monitoring statistics to incipient faults can be improved.

2) Integration of in-statistical-control and hard sparsity constraints for fault isolation: By imposing statistical and geometric constraints on basic dictionary coefficients, the fault-free components are ensured to remain uncontaminated, allowing for the separability of out-of-statistical-control components in coefficients. In addition, by applying the hard sparsity constraint on the introduced dictionaries, it is possible to uncover latent patterns in the fault-related components, resulting in enhanced isolation accuracy.

3) Theoretical guarantee for superior detection and isolation performance: Section IV provides theoretical support for the superior monitoring performance, which offers a valuable reference for subsequent in-depth investigations.

II. Dictionary Learning Based Monitoring

To monitor industrial process states, n-dimensional signals $\mathbf{x}_i \in \mathbb{R}^n$ for $i = 1, \ldots, N$ can be represented as K-dimensional ($K \gg n$) coefficient vectors $\mathbf{s}_i^0 \in \mathbb{R}^K$ using an overcomplete dictionary matrix $\mathbf{D}_0 \in \mathbb{R}^{n\times K}$. This results in a sparse coding formulation through basic dictionary learning [24] as follows:

$$\begin{aligned} &\min_{\mathbf{D}_0,\mathbf{s}} \sum\nolimits_i \frac{1}{2}\left\|\mathbf{x}_i - \mathbf{D}_0\mathbf{s}_i^0\right\|^2 + \lambda_1\left\|\mathbf{s}_i^0\right\|_1 + \frac{\lambda_2}{2}\left\|\mathbf{s}_i^0\right\|^2 \\ &s.t.\ \left[\mathbf{d}_k^0\right]^T\mathbf{d}_k^0 \le 1 \end{aligned} \tag{1}$$

where $\mathbf{d}_k^0$ denotes the kth atom of $\mathbf{D}_0$. By selecting appropriate values for λ_1 and λ_2, the elastic-net regularization [25], which combines the ℓ_1-norm ($\|\cdot\|_1$) and the squared ℓ_2-norm ($\|\cdot\|^2$), adjusts the sparsity and magnitude of dictionary coefficients. This helps balance the reconstruction error while reducing the influence of outliers. The constraint on the atoms prevents degeneration, which could lead to small coefficients. To solve this optimization problem, dictionary update and sparse coding can be performed using methods such as FISTA [26], LASSO [27], or K-SVD [28].

Once the dictionary $\mathbf{D}_0$ in (1) is learned, the sparse coding of $\mathbf{s}_t^0$ for the tth test sample $\mathbf{x}_t$ and the associated reconstruction residual $\mathbf{r}_t$ can be used to construct statistics for monitoring

$$T^2 = \|\mathbf{s}_t^0\|^2 \text{ and } SPE = \|\mathbf{r}_t\|^2 \tag{2}$$

where the coefficients are referred to as primary coding or coefficients. For fault isolation, the basic dictionary is first augmented to $\overline{\mathbf{D}}_0 = [\mathbf{D}_0, \mathbf{I}]$, with the identity matrix $\mathbf{I} \in \mathbb{R}^{n\times n}$ serving as the fault/feature direction. Substituting $\mathbf{D}_0$ with $\overline{\mathbf{D}}_0$ in (1), the primary coding can be extended to $\overline{\mathbf{s}}_t^0 = [(\mathbf{s}_t^0)^T, \alpha_t^T]^T$, where $\alpha_t \in \mathbb{R}^n$ is named as the expansion coding/part, corresponding to the introduced identity matrix. As insinuated by Ning et al. [17], the nonzero elements in the expansion part are used to indicate the faulty variables.

The dictionary in (1) characterizing the global features, in collaboration with the primary and expansion coding, can be used to coarsely assess process states. Nevertheless, as is pointed out in Section I, the following issues still need to be highlighted: 1) An incipient fault tends to be dispersed due to the inseparability of out-of-statistical-control variation in primary coding, resulting in insufficient detection sensitivity. 2) For fault isolation, it is unwise to fix the feature direction, which may negatively affect the isolation accuracy for fault involving multiple variables. 3) To allow the expansion coding to parse an occurred fault, it is crucial to estimate the latent healthy process states, that is, to bring the T^2 statistics in (2) back to normal.

III. Structured Sparse Coding for Enhanced Monitoring With Augmented Dictionary Learning

This section outlines the model construction of SASDL, including signal decomposition, statistical properties, and online dictionary learning, followed by the monitoring strategy.

A. Model Construction

1) Process Signal Decomposition: According to [11], [17], [29], etc., the fundamental assumption underlying the dictionary learning-based monitoring is the decomposition of process signal as

$$\mathbf{x}_t = \mathbf{D}_0\mathbf{s}_t^0 + \mathbf{r}_t. \tag{3}$$

The fault in $\mathbf{x}_t$ is often divided and dispersed between $\mathbf{s}_t^0$ and $\mathbf{r}_t$, which degrades detection performance, particularly for incipiently developing faults. To mitigate the undesired dispersion, an additional sparse dictionary $\mathbf{D}_t \in \mathbb{R}^{n\times\tau}$ ($\tau \ll K$) is introduced to capture the out-of-statistical-control coefficient variation, since it is more suitable to match the fault with sparse features. With this in mind, the signal decomposition containing the suspicious variation $\widetilde{\alpha}_t \in \mathbb{R}^\tau$ can be defined by

$$\mathbf{x}_t = \underbrace{\mathbf{D}_0\mathbf{s}_t^0}_{\overline{\mathbf{x}}_t} + \underbrace{\mathbf{D}_t\widetilde{\alpha}_t + \widetilde{\mathbf{r}}_t}_{\widetilde{\mathbf{x}}_t} \tag{4}$$

where $\overline{\mathbf{x}}_t$ denotes the latent fault-free part and $\widetilde{\mathbf{x}}_t$ the fault-related part consisting of the data for the variation $\widetilde{\mathbf{x}}_t^\alpha = \mathbf{D}_t\widetilde{\alpha}_t$ and the residual $\widetilde{\mathbf{r}}_t$.

2) Statistical Property Constraint: In contrast to (3), the primary coding in (4) must be in-statistical-control to prevent the fault-free component from being contaminated by process anomalies. This implies that the constraint $\|\mathbf{s}_t^0\|^2 \le \upsilon$ needs to be satisfied, with υ being the control limit of T^2 statistics in (2). However, this constraint lacks necessary flexibility for accurately estimating the fault-free component. To enhance flexibility, $L-1$ successive primary codings are incorporated into the embedding problem as

$$\min_{\mathbf{s}} f\left(\mathbf{s}_{t-L+1}^0, \ldots, \mathbf{s}_t^0\right) + \left\|\sum\nolimits_{j=t-L+1}^{t} \mathbf{s}_j^0\right\|^2 \tag{5}$$

with $f(\cdot)$ being the constructed loss function.

3) Online Dictionary Learning: For online monitoring, L samples in accordance with (5) are assembled by the moving window $\mathbf{X}_t = [\mathbf{x}_{t-L+1}, \ldots, \mathbf{x}_t] \in \mathbb{R}^{n\times L}$. By combining the fine-grained signal decomposition in (4) with the statistical constraint in (5), a structured augmented sparse dictionary learning approach for improved monitoring is proposed as follows:

$$\begin{aligned}\min_{\mathbf{D}_t, \overline{\mathbf{S}}_t^0} \; & \frac{1}{2}\left\{\left\|\mathbf{X}_t - \overline{\mathbf{D}}_0^t\overline{\mathbf{S}}_t^0\right\|_F^2 + \left\|\mathbf{X}_t - \mathbf{D}_0\mathbf{S}_t^0\right\|_F^2 + \left\|\sum\nolimits_j \mathbf{s}_j^0\right\|^2\right\} \\ & + \lambda_1\left\|\overline{\mathbf{S}}_t^0\right\|_1 + \frac{\lambda_2}{2}\left\|\mathbf{S}_t^0 - \widetilde{\mathbf{S}}_t^0\right\|_F^2 + \frac{\lambda_3}{2}\left\|\mathbf{D}_t - \widetilde{\mathbf{D}}_t\right\|_F^2 \\ s.t. \; & \left[\mathbf{d}_k^t\right]^T\mathbf{d}_k^t \le 1, \; \left[\mathbf{d}_k^t\right]_p = 0, \; p \in \mathcal{P}_k\end{aligned} \tag{6}$$

where $\|\cdot\|_F$ is the Frobenius norm, and $[\cdot]_p$ denotes the pth element of a vector or the related column of a matrix. Compared to the augmentation in Section II, the current augmentation $\overline{\mathbf{D}}_0^t = [\mathbf{D}_0, \mathbf{D}_t]$ gives rise to a more sophisticated extended coding as $\overline{\mathbf{S}}_t^0 = [(\mathbf{S}_t^0)^T, \widetilde{\mathbf{A}}_t^T]^T$, with the primary part $\mathbf{S}_t^0 = [\mathbf{s}_{t-L+1}^0, \ldots, \mathbf{s}_t^0]$ corresponding to $\mathbf{D}_0$ and the expansion part $\widetilde{\mathbf{A}}_t = [\widetilde{\alpha}_{t-L+1}, \ldots, \widetilde{\alpha}_t]$ in alignment with $\mathbf{D}_t$. The first data-fidelity term in (6) relates to the signal decomposition in (4). The other aims to eliminate the issue of zero coefficients, which can lead to information deficiency, thereby collaborating with the first term to effectively parse the triggered fault.

To maintain the correlation structure of the fault-free part, a k-nearest neighbor (KNN)-based geometric constraint is also imposed by $\left\|\mathbf{S}_t^0 - \widetilde{\mathbf{S}}_t^0\right\|_F^2$, with the auxiliary coefficients $\widetilde{\mathbf{S}}_t^0$ calculated as in [19]. While the final regularization term with parameter λ_3 encourages sparse atoms to robustly match fault patterns by adjusting the difference between $\mathbf{D}_t$ and its mean matrix $\widetilde{\mathbf{D}}_t$, where $\widetilde{\mathbf{D}}_t = (\sum_k \mathbf{d}_k^t/\tau) \times \mathbf{1}^T$.

In addition, instead of imposing the ℓ_1-regularization, this article achieves the sparsity for each atom $\mathbf{d}_k^t$ using the index set $\mathcal{P}_k$, which is actually a hard sparsity constraint. This hard sparsity constraint offers flexibility in selecting feature variables to uncover fault patterns, facilitating the subsequent fault isolation task. It is worth to note that process noise or disturbance may prevent some unimportant features from being discarded by the zero-element index set. However, this issue can be effectively managed through the utilization of expansion coefficients with sparsity at the corresponding positions. Thus, both sparse dictionaries and expansion coefficients contribute to discovering the structured sparse patterns.

Unlike the modeling of (1), the motivation behind structured sparse coding in (6) includes: 1) incorporating prior knowledge

Algorithm 1: Optimization for SASDL Requires the Parameters λ_1, λ_2, λ_3, L, τ, and η to be Determined First.

step 1: Initialize $\mathbf{D}_t$, $\mathbf{S}_t^0$, and $\widetilde{\mathbf{A}}_t$ randomly, and calculate $\widetilde{\mathbf{S}}_t^0$;

step 2: Solve the LASSO-like problem for sparse coding column by column:

solve $\min_{\mathbf{c}} \frac{1}{2}\|\mathbf{y}_j - \mathbf{D}_j\mathbf{c}\|^2 + \lambda_1\|\mathbf{c}\|_1$ [31]

$(j = t - L + 1 : t)$

// for $\mathbf{s}_j^0$ using the inputs:

1 $\mathbf{y}_j = [\mathbf{x}_j^T, -\sum_{i\neq j}[\mathbf{s}_i^0]^T, (\mathbf{x}_j - \mathbf{D}_t\widetilde{\alpha}_j)^T, \sqrt{\lambda_2}[\widetilde{\mathbf{s}}_j^0]^T]^T$

2 $\mathbf{D}_j = [\mathbf{D}_0^T,\ \mathbf{I},\ \mathbf{D}_0^T, \sqrt{\lambda_2}\mathbf{I}]^T$

// for $\widetilde{\alpha}_j$ using the inputs:

3 $\mathbf{y}_j = \mathbf{x}_j - \mathbf{D}_0\mathbf{s}_j^0$ and $\mathbf{D}_j = \mathbf{D}_t$

step 3: Update $\mathbf{D}_t$ atom by atom $(k = 1 : \tau)$:

4 construct $\mathcal{P}_k$ based on CVCR (Cumulative Variance Contribution Rate):

// compute variances: $V_m = \mathrm{var}([\mathbf{X}_t - \mathbf{D}_0\mathbf{S}_t^0]^m)$ $(m = 1 : n)$;

5 determine $\mathcal{P}_k$ based on small CVCRs: $V_m/\sum_i V_i \leq \eta$;

// take derivatives and projections:

6 $\widetilde{\mathbf{d}}_k^t = \frac{[\mathbf{X}_t - \mathbf{D}_0\mathbf{S}_t^0 - \sum_{\bar{k}\neq k}\mathbf{d}_{\bar{k}}^t(\tilde{\mathbf{A}}_t^{\bar{k}})^T]\tilde{\mathbf{A}}_t^k - \lambda_3\sum_{\bar{k}\neq k}\mathbf{d}_{\bar{k}}^t(\mathbf{B}_t^{\bar{k}})^T\mathbf{B}_t^k}{\|\tilde{\mathbf{A}}_t^k\|^2 + \lambda_3\|\mathbf{B}_t^k\|^2}$

7 $\{[\widetilde{\mathbf{d}}_k^t]_{\mathcal{P}_k} = 0\} \overset{\mathcal{P}_k}{\rightarrow} \widehat{\mathbf{d}}_k^t$, $\mathbf{d}_k^t = \widehat{\mathbf{d}}_k^t/\max([\widehat{\mathbf{d}}_k^t]^T\widehat{\mathbf{d}}_k^t, 1)$

step 4: If not converges, return to step 2.

such as statistical properties and local geometric structure, 2) analyzing the components and variations in process signals, 3) capturing and interpreting abnormal information, and 4) discovering essential fault patterns. Specifically, in contrast to Ning et al. [17] and other approaches [11], [29], [30] that fix an identity matrix as feature direction, the proposed SASDL incorporates an optimized sparse dictionary matrix, providing an genuine carrier for analyzing process behaviors. Preserving data properties in the primary coefficients, combined with a moving window approach for fault information aggregation, enables effective signal decomposition and improves the detection and isolation of incipient faults.

To solve (6), a two-stage approach [27] with sparse coding and dictionary updating is considered. Let $([\,\cdot\,]^k)^T$ be the kth row and $\mathbf{B}_t = \mathbf{I} - \mathbf{1}\times\mathbf{1}^T/\tau$. Algorithm 1 outlines the key steps.

B. Monitoring Strategy

To enhance detection sensitivity and isolation accuracy for incipient faults, a moving window includes the latest sample and discards the oldest, feeding the data into (6) for signal decomposition and subsequent fault analysis.

1) Monitoring Statistics: Assume a fault only impacts the dictionary coefficients, the expansion matrix in (6) can be directly used to measure the deviation outside the model of (1). For the case that the coefficients are not affected by the fault, the residual in (4) should be simplified to $\mathbf{R}_t = \mathbf{X}_t - \mathbf{D}_0\mathbf{S}_t^0$. The statistics T^2 for primary variation and the statistics SPE for residual thus can be constructed as follows:

$$\begin{cases} T^2 = \left\|\sum_j \mathbf{D}_t\widetilde{\alpha}_j/L\right\|^2 \\ \mathrm{SPE} = \bar{\mathbf{r}}_t^T\boldsymbol{\Theta}\bar{\mathbf{r}}_t \end{cases} \tag{7}$$

where $\bar{\mathbf{r}}_t$ denotes the mean of $\mathbf{R}_t$ and $\boldsymbol{\Theta}$ the precision matrix of the training dataset. Through (7), a series of T^2 and SPE statistics can be obtained from the training data. In order to obtain the control limits, kernel density estimation (KDE) is used to estimate the probability densities of both statistics. Once the probability densities are obtained, the corresponding control limits can be determined by observing at which threshold the cumulative distribution function reaches the chosen significance level. For the construction in (7), beyond using a moving window to condense dispersed fault information, key differences from (2) include: 1) noninterference in statistics due to coefficients separability and 2) incorporation of important factors through sparse atoms and variable correlation via precision matrix.

2) Isolation Model: Since (6) explicitly includes the fault information, the subsequent isolation step could have been simply conducted by inspecting the nonzero elements in dictionary $\mathbf{D}_t$ or by comparing the reconstruction errors derived from residual $\mathbf{R}_t$. However, to suppress the process outliers or disturbances, a $\ell_{2,0}$-regularization based feature selection model that considers similar fault patterns between adjacent windows is used, as follows:

$$\min_{\mathbf{F}_t} \frac{1}{2}\left\|\mathbf{F}_t - \boldsymbol{\Delta}_t\right\|_F^2 + \beta_1\left\|\mathbf{F}_t\right\|_{2,0} + \frac{\beta_2}{2}\left\|\mathbf{F}_t - \mathbf{F}_{t-1}\right\|_F^2 \tag{8}$$

where $\mathbf{F}_t$ stands for the fault indicator matrix with row sparsity induced by the $\ell_{2,0}$-norm (the number of nonzero rows in a matrix). The information matrix $\boldsymbol{\Delta}_t$ equals to $\mathbf{D}_t\widetilde{\mathbf{A}}_t$ for the fault affecting T^2 statistics, and $\mathbf{R}_t$ for the fault affecting SPE statistics. By introducing an auxiliary variable for the $\ell_{2,0}$-norm, the problem in (8) can be reformulated into two simpler subproblems, which can be solved iteratively until convergence within the framework of alternating direction method of multipliers (ADMM). For more details, interested readers can refer to [32] that deals with a similar nonconvex problem using ADMM. By adjusting the parameters β_1 and β_2, the fault-free features associated with the zero rows in the solution of $\mathbf{F}_t$ are discarded, enabling the retained features to indicate the potential faulty variables.

For the purpose of evaluating fault severity, a fault score (Fscore) for each variable is defined as

$$\delta_i = \frac{\sum_j \left|[\mathbf{F}_t]_{ij}\right|}{\max\left(\mathbf{F}_t\right)}. \tag{9}$$

Correspondingly, a score vector $\delta_t = [\delta_1, \ldots, \delta_n]^T$ can be obtained through (9) for each test window $\mathbf{X}_t$.

C. Model Parameter Determination and Complexity Analysis

According to [28], dictionary dimensions $K \in [3n, 4n]$ and $\tau \leq \frac{n}{2}$ are recommended. The regularization parameters are selected using grid search or trial and error. For λ_1, it is suggested that 20% –30% of dictionary coefficients be nonzero. For λ_2, an online adjustment strategy is given as $\lambda_2^t \leftarrow$

TABLE I
COMPUTATIONAL COMPLEXITY OF WINDOW-BASED METHODS

Method	Complexity	Method	Complexity
MWPCA	$\mathcal{O}\left(nL+n^2\right)$	KL	$\mathcal{O}(N)$
MCC	$\mathcal{O}\left(n^2L+n^3\right)$	SGGL	$\mathcal{O}\left(n^2L+n^3\log(1/\varepsilon)\right)$
SMD	$\mathcal{O}\left(nL+n^2\right)$	SASDL	$\mathcal{O}\left(L\left(m_1K^3+m_2n\right)\right)$

Note: ϵ represents the predefined accuracy threshold for ADMM, while m_1 and m_2 denote the iteration numbers required to solve equations (6) and (8), respectively.

$\lambda_2^{t-1}e^{-\frac{1}{L}\|\mathbf{S}_{t-1}^0-\widetilde{\mathbf{S}}_{t-1}^0\|_F^2}$. Smaller λ_3 values are used for simple faults, and vice versa. Based on experiments, $\eta \in [0.5, 0.8]$ can better capture the fault information. For isolation, β_1 in (8) inversely affects the number of identified faulty variables, while β_2 preserves fault feature consistency for accurate isolation. Both can be easily tuned through practical experiments. On the other hand, an appropriate window length L is determined through trial and error. More specifically, the fault detection rate (FDR) and false alarm rate (FAR) of the T^2 and SPE statistics are calculated by increasing the window lengths (for example, from 2 to 200), and the window length that yields the best FDR and FAR is chosen.

To evaluate the computational complexity of the proposed method, five window-based monitoring methods are considered, including moving window RBC (MWRBC) [21], mean and covariance charts (MCC) [33], statistics Mahalanobis distance (SMD) [22], Kullback–Leibler divergence (KL) [23], and Sparse Gaussian Graphical Lasso (SGGL) [34]. The computational complexities of these methods are listed in Table I. Since $K \propto n$ implies $K^3 = \mathcal{O}(n^3)$, the worst-case complexity of SASDL is $\mathcal{O}(\max(m_1, m_2) \times Ln^3)$. With $L \leq 200$ and $\max(m_1, m_2) \leq 200$, SASDL maintains a well-regulated computational cost, compared to other state-of-the-art methods, indicating that it is computationally acceptable for practical applications.

IV. THEORETICAL ANALYSIS OF DETECTION AND ISOLATION

This section presents theoretical foundations supporting the enhanced monitoring performance of SASDL.

1) Uncontaminated primary coding: To obtain the uncontaminated primary coding, the necessary condition $T^2 = \|\mathbf{s}_i^0\|^2 \leq M \leq \upsilon$ must be satisfied, where M is an upper boundary, as guaranteed by *Theorem 1*.

Theorem 1: If the monitoring vectors are independent and approximately follow a zero-mean Gaussian distribution, the constraint $\|\mathbf{s}_i^0\|^2 \leq \nu$ holds for (5).

Proof: Expanding the embedding problem in (5) gives

$$\left\|\sum\nolimits_i \mathbf{s}_i^0\right\|^2 = \sum\nolimits_i \left\|\mathbf{s}_i^0\right\|^2 + 2\sum\nolimits_{i<j} \left\|\mathbf{s}_i^0\right\|\left\|\mathbf{s}_j^0\right\| \cos\theta_{ij} \leq M. \tag{10}$$

Minimizing (5) reduces the length of $\sum_i \mathbf{s}_i^0$, while also decreasing both the vector angles θ_{ij} and magnitudes $\|\mathbf{s}_i^0\|$. Since $\mathbf{s}_i^0$ and $\mathbf{s}_j^0$ are independent, their initial angles approach 90°. Optimizing (5) reduces the angles θ_{ij} from values near 90° to 0°, which may increase the interdependence between the vectors. As a result, the dot product $\sum_{i<j} \|\mathbf{s}_i^0\|\|\mathbf{s}_j^0\| \cos\theta_{ij} \geq 0$ is more likely to hold. This leads to the following deduction:

$$\sum\nolimits_i \left\|\mathbf{s}_i^0\right\|^2 \leq \left\|\sum\nolimits_i \mathbf{s}_i^0\right\|^2 \leq M \Rightarrow \|\mathbf{s}_i^0\|^2 \leq M. \tag{11}$$

Increasing L drives $\sum_i \mathbf{s}_i^0$ towards $\mathbf{0}$, assuming that monitoring vectors approximately follow a zero-mean Gaussian distribution. This results in a tight upper bound M for $\left\|\sum_i \mathbf{s}_i^0\right\|^2$, and $\|\mathbf{s}_i^0\|^2 \leq M \leq \upsilon$ holds. It should be noted that the assumptions of independence and Gaussian distribution may not hold strictly in practical scenarios. However, it still can provide good guidance for practical application as they can be satisfied approximately in practice.

2) Nonstrict detection condition: SASDL achieves a tighter lower bound for fault detection, increasing sensitivity to the faults in coefficients or residuals, as verified by *Theorem 2*.

Theorem 2: The statistics in (7) have relatively mild detection condition with an appropriate moving window length.

Proof: The in-statistical-control sample $\mathbf{x}_t$ satisfies two inequalities: $\|\mathbf{x}_t - \mathbf{D}_0\mathbf{s}_t^0\|^2 \leq e_1$ and $\|\mathbf{x}_t - \mathbf{D}_0\mathbf{s}_t^0 - \mathbf{D}_t\widetilde{\alpha}_t\|^2 \leq e_2$, with e_1 and e_2 being two upper boundaries. It follows that:

$$\begin{aligned} &\left\|\left(\mathbf{x}^t - \mathbf{D}_0\mathbf{s}_t^0 - \mathbf{D}_t\widetilde{\alpha}_t\right) + \mathbf{D}_t\widetilde{\alpha}_t\right\|^2 \leq e_1 \\ &\Rightarrow \|\mathbf{D}_t\widetilde{\alpha}_t\| - \left\|\mathbf{x}^t - \mathbf{D}_0\mathbf{s}_t^0 + \mathbf{D}_t\widetilde{\alpha}_t\right\| \leq \sqrt{e_1} \\ &\Rightarrow \|\mathbf{D}_t\widetilde{\alpha}_t\| \leq \sqrt{e_1} + \sqrt{e_2}. \end{aligned} \tag{12}$$

Renaming the coefficient variation and feature dictionary for a fault-free sample as: $\widetilde{\alpha}_t \to \widetilde{\alpha}_t^*$ and $\mathbf{D}_t \to \mathbf{D}_t^*$, the detectability condition of T^2 statistics for SASDL can be inferred as

$$\begin{aligned} &\left\|\Delta\mathbf{D}_t\widetilde{\alpha}_t + \mathbf{D}_t^*\widetilde{\alpha}_t^*\right\| \geq \left\|\Delta\mathbf{D}_t\widetilde{\alpha}_t\right\| - \left\|\mathbf{D}_t^*\widetilde{\alpha}_t^*\right\| > \sqrt{e_1} + \sqrt{e_2} \\ &\Rightarrow \left\|\Delta\mathbf{D}_t\widetilde{\alpha}_t\right\| > \left\|\mathbf{D}_t^*\widetilde{\alpha}_t^*\right\| + \sqrt{e_1} + \sqrt{e_2} \\ &\Rightarrow \left\|\sum\nolimits_j \Delta\mathbf{D}_t\widetilde{\alpha}_j/L\right\|^2 > 4\left(\sqrt{e_1} + \sqrt{e_2}\right)^2/L \\ &\Rightarrow \mathcal{T}_{\min}^{\mathrm{SASDL}} = 4\left(\sqrt{e_1} + \sqrt{e_2}\right)^2/L. \end{aligned} \tag{13}$$

Similarly, the lower boundary of detectable faults for (1) can be derived as $\mathcal{T}_{\min}^{\mathrm{DL}} = 4\upsilon$. It follows that $\mathcal{T}_{\min}^{\mathrm{DL}} > \mathcal{T}_{\min}^{\mathrm{SASDL}}$ for bounded noise, indicating that the T^2 statistics of SASDL are more sensitive than those of basic dictionary learning, even without a large moving window. Nevertheless, window selection still depends on fault magnitudes [21], and an appropriate window length should also satisfy the condition

$$L \geq \max_i \left(\frac{4\left(\sqrt{e_1} + \sqrt{e_2}\right)^2}{\left\|\mathbf{\Theta}^{1/2}\xi_i\right\|^2\|f_i\|^2}\right) \tag{14}$$

where ξ_i is the ith column of identity matrix $\mathbf{I}$, and the associated fault magnitude f_i can be estimated using least-squares as: $f_i = (\xi_i^T\mathbf{\Theta}\xi_i)^{-1}\xi_i\mathbf{\Theta}(\mathbf{x}_t - \mathbf{D}_0\mathbf{s}_t^0)$, with $\mathbf{x}_t$ being a faulty sample. Similar results can be obtained for SPE statistics.

3) Robust isolation mechanism: A isolation mechanism with bounded reconstruction error and sequential sparse dictionary learning effectively identifies faulty variables, as ensured by *Theorems 3* and *4*.

Theorem 3: If the reconstruction error for the fault-free data is bounded, the fault magnitude that can be isolated is proportional to this upper bound.

Proof: To explore the smallest fault magnitude that can be isolated, a fault instance $\mathbf{f} \in \mathbb{R}^n$ is introduced into the data as $\mathbf{X}_t = \overline{\mathbf{X}}_t + \mathbf{f} \times \mathbf{1}^T$, where $\overline{\mathbf{X}}_t = [\overline{\mathbf{x}}_{t-L+1}, \ldots, \overline{\mathbf{x}}_t]$ represents

the fault-free part. The property of $\overline{\mathbf{X}}_t$ can be inferred as

$$\big\| \underbrace{\overline{\mathbf{X}}_t - \mathbf{D}_0\mathbf{S}_t^0}_{\widehat{\mathbf{R}}_t} \big\|_F^2 = \sum\nolimits_i \big\| \underbrace{\overline{\mathbf{x}}_i - \mathbf{D}_0\mathbf{s}_i^0}_{\widehat{\mathbf{r}}_i} \big\|^2 \le Le_1. \tag{15}$$

This leads to the isolation condition for $\mathbf{f}$ as follows:

$$\big\|\overline{\mathbf{X}}_t + \mathbf{f}\times\mathbf{1}^T - \mathbf{D}_0\mathbf{S}_t^0\big\|_F^2 > Le_1. \tag{16}$$

Shrinking the left side of (16) and using the Holder's inequality yields the derivation as

$$\begin{aligned} & L\|\mathbf{f}\|^2 - 2\|\mathbf{f}\|\|\widehat{\mathbf{R}}_t\times\mathbf{1}\| > Le_1 \\ & \overset{(18)}{\Longrightarrow} \|\mathbf{f}\|^2 - 2\sqrt{e_1}\|\mathbf{f}\| - e_1 > 0 \\ & \Rightarrow \|\mathbf{f}\| > \left(1+\sqrt{2}\right)\sqrt{e_1}, \end{aligned} \tag{17}$$

where the upper boundary of $\|\widehat{\mathbf{R}}_t\times\mathbf{1}\|$ is given by:

$$\begin{aligned} & \big\|\widehat{\mathbf{R}}_t\times\mathbf{1}\big\|^2 = \big\|\sum\nolimits_i \widehat{\mathbf{r}}_i\big\|^2 \\ & \le \sum\nolimits_i \|\widehat{\mathbf{r}}_i\|^2 + 2\sum\nolimits_{i<j}\|\widehat{\mathbf{r}}_i\|\|\widehat{\mathbf{r}}_j\| \\ & \le \sum\nolimits_i e_1 + 2\sum\nolimits_{i<j} e_1 = L^2e_1 \\ & \Rightarrow \|\widehat{\mathbf{R}}_t\times\mathbf{1}\| \le L\sqrt{e_1}. \end{aligned} \tag{18}$$

Hence, (16–18) establish a tight lower boundary for isolation.

Theorem 4: The introduced sparse dictionary enables accurate fault isolation, as long as the fault magnitude for each relevant variable exceeds a specified threshold.

Proof: For the sake of simplicity, the optimization problem in (6) with regard to $\mathbf{D}_t$ can be reduced to:

$$\min_{\mathbf{D}_t} \big\| \underbrace{\mathbf{f}\times\mathbf{1}^T + \widehat{\mathbf{R}}_t}_{\mathbf{M}} - \mathbf{D}_t\widetilde{\mathbf{A}}_t\big\|_F^2. \tag{19}$$

To ensure that the fault information in $\mathbf{M}$ is prioritized for retention in $\mathbf{D}_t$ during optimization, the following inequalities must be satisfied:

$$\begin{aligned} & \big\|f_i\mathbf{1}^T + [\widehat{\mathbf{r}}_t^i]^T\big\|^2 > \big\|\widehat{\mathbf{r}}_t^i\big\|^2 \\ & \overset{i\in\mathcal{G}}{\Longrightarrow} f_i^2 - 2\big|[\widehat{\mathbf{r}}_t^i]^T\mathbf{1}f_i\big|/L > 0 \\ & \Rightarrow |f_i| > \sqrt{(L+1)e_1/2Ln} \end{aligned} \tag{20}$$

where f_i denotes the magnitude of a true fault variable within $\mathbf{f}$ and $[\widehat{\mathbf{r}}_t^i]^T$ is ith row of $\widehat{\mathbf{R}}_t^i$, while $i\in\mathcal{G}$ with $\mathcal{G}$ being the fault variable set. The boundary in (20) is derived as follows:

$$\begin{aligned} & \big|[\widehat{\mathbf{r}}_t^i]^T\mathbf{1}\big|^2 \le \sum\nolimits_j [\widehat{\mathbf{r}}_t^i]_j^2 \\ & + \frac{L(L-1)}{2}\max_j [\widehat{\mathbf{r}}_t^i]_j^2 \le \frac{(L^2+L)e_1}{2n} \\ & \Rightarrow \big|[\widehat{\mathbf{r}}_t^i]^T\mathbf{1}\big| \le \sqrt{(L^2+L)e_1/2n}. \end{aligned} \tag{21}$$

The condition for accurate isolation, thus, is that fault magnitude must exceed the threshold in (20).

In addition, for each atom $\mathbf{d}_k^t$, the objective function in (19) can be elaborated as

$$\min_{\mathbf{d}_k^t} \big\| \underbrace{\mathbf{M} - [\mathbf{D}_t]_{-k}\left[\widetilde{\mathbf{A}}_t\right]^{-k}}_{\overline{\mathbf{M}}} - \mathbf{d}_k^t\left(\widetilde{\alpha}_t^k\right)^T\big\|_F^2. \tag{22}$$

Solving the optimization problem in (19) one can get an M that is close to $\mathbf{D}_t\widetilde{\mathbf{A}}_t$. In this case, the elements of $\overline{\mathbf{M}}$ can be decomposed into two parts: the first relates to significant deviations associated with a subset of fault variables, while the other relates to minor deviations of other variables, which are likely to be excluded from $\mathbf{d}_k^t$ through sparsity-inducing techniques. By combining both parts, a complete isolation result can be obtained as: $\mathcal{G} = \bigcup_k \mathcal{G}_k$. Consequently, the proof of the theorem is completed.

V. Numerical Studies

A numerical example is used to verify the effectiveness of SASDL in fault detection and isolation, which involves 10 process variables in $\mathbf{x}$ and 2 Gaussian latent variables in $\mathbf{h}$

$$\begin{aligned} & \mathbf{x} = \mathbf{\Xi}\mathbf{h} + \mathbf{e} \\ & \mathbf{h} = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \sim \mathcal{N}\left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.98 & 0 \\ 0 & 1 \end{bmatrix} \right\} \\ & \mathbf{\Xi} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & -0.6 & -0.6 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0.8 & 0.8 \end{bmatrix}^T \end{aligned} \tag{23}$$

where the matrix $\mathbf{\Xi}$ describes the relationship between process variables and latent variables in a zero-mean Gaussian noise environment, and the noise covariance matrix $\mathbf{\Sigma}_e$ is a diagonal matrix, with the diagonal elements being 0.09, 0.09, 0.09, 0.09, 0.16, 0.16, 0.16, 0.16, 0.25, and 0.25.

To prepare the SASDL-based monitoring, a normal dataset $\mathbf{X}_0$ with 500 samples from (23) is first used to construct the basic dictionary. For comparison, this study considers PCA (2 PCs), basic-dictionary learning (Basic-DL), MWRBC, KL, SMD, MCC, and SGGL for fault detection, as well as MWRBC, SpCP, DL-RBC [19], DL-IRBC [14], SMD, and SGGL for fault isolation.

A. Fault Detection

To evaluate monitoring performance, two faulty datasets, $\mathbf{X}_1$ and $\mathbf{X}_2$, each with 500 samples, are considered. $\mathbf{X}_1$ deviates by 0.6 on variable x_1 from the 151st instance, while $\mathbf{X}_2$ deviates by $\Delta h_1 = 1.2$ on latent variable h_1 from the 201st instance. The fault-noise ratios for both deviations are below 2, indicating that the introduced faults are incipient [35].

Following the guidelines in Section III-C and the experiments in Section V-C2, the parameters for dictionary learning-based methods are set as $K = 30$, $\tau = 5$, $\lambda_1 = 0.15$, $\lambda_2 = 0.03$, $\lambda_3 = 0.1$, and $k = 7$ for KNN, with $\eta = 0.8$ for CVCR. The strategy in Section III-B is then applied, with the window length $L = 30$ chosen to balance Type I and Type II errors as well as computational cost, yielding the monitoring results as shown in Fig. 1. It can be seen from Fig. 1 that sporadic violations in PCA and basic-DL indicate possible faults. In contrast, MWRBC and

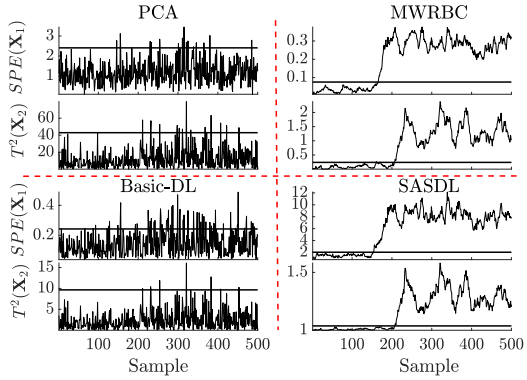

Fig. 1. Monitoring results using PCA, basic-DL, MWRBC, and SASDL. For simplicity, statistically insignificant alerts are omitted.

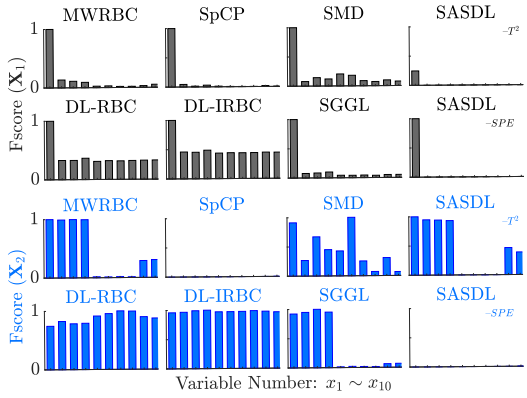

Fig. 2. Fault isolation results using MWRBC, SpCP, DL-RBC, DL-IRBC, SMD, SGGL, and SASDL for $\mathbf{X}_1$ and $\mathbf{X}_2$.

TABLE II
AVERAGE FDR AND FAR (%) OF DIFFERENT METHODS FOR $\mathbf{X}_1$ AND $\mathbf{X}_2$

Dataset	$\mathbf{X}_1$				$\mathbf{X}_2$			
Method	T^2/st. 1		SPE/st. 2		T^2/st. 1		SPE/st. 2	
	FDR	FAR	FDR	FAR	FDR	FAR	FDR	FAR
PCA	0.57	0.0	4.29	0.67	4.67	0.5	0.0	0.0
Basic-DL	0.86	0.0	15.71	5.33	2.67	0.0	0.0	0.0
MWRBC	1.43	0.0	95.71	0.0	**98.0**	1.5	0.0	0.0
KL	8.57	2.0	80.0	0.0	81.0	2.0	0.0	0.0
MCC	96.86	0.0	97.14	0.0	0.0	0.0	94.33	0.0
SMD	92.57	13.33	—	—	85.67	2.0	—	—
SGGL	90.86	0.0	—	—	89.43	0.0	—	—
SASDL	40.29	0.0	**99.14**	**0.0**	**98.0**	**0.0**	0.0	0.0

Note: st. 1 and st. 2 correspond to the statistics used by competitive methods, more specifically, they are constructed based on mean and covariance for MCC, Mahalanobis distance for SMD, and graph similarity for SGGL.

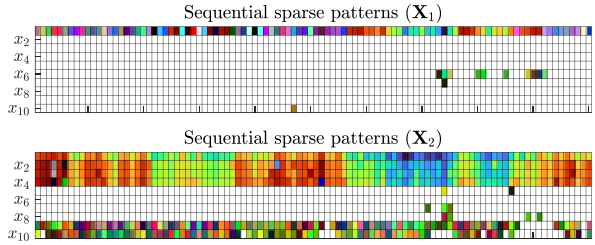

Fig. 3. Spare fault patterns in the introduced sparse dictionaries and expansion codings. The plots illustrate the patterns through $\mathbf{D}_t\widetilde{\alpha}_t$ for the final 100 windows, with white blocks signifying small absolute values or zeros.

SASDL confirm the faults with abundant alarms, suggesting that a moving window scheme enhances monitoring performance and enables successful incipient fault detection.

For a comprehensive comparison, all methods were tested over 10 experiments, with metrics such as FDR and FAR documented in Table II. Results indicate that SASDL outperforms PCA, basic-DL, and others, achieving a higher FDR of 99.14% for $\mathbf{X}_1$ and 98.0% for $\mathbf{X}_2$, along with a lower FAR of 0.0% . The better fault detection performance can be attributed to the effective signal decomposition capability of the proposed method, as supported by Theorem 1, which enables the extraction of uncontaminated normal components from primary coefficients, as well as the improved detection sensitivity, as ensured by Theorem 2, through the maintenance of coefficients being in-statistical-control states.

B. Fault Isolation

Once a fault is detected, monitoring moves to isolation. The model in Section III-B2 is probably applied twice for alarming statistics, with $\beta_1 = 3$ and $\beta_2 = 0.1$ set by trial and error. For fairness, dictionary learning methods take the window mean for isolation, with results shown in Fig. 2.

As shown in the top two rows of plots in Fig. 2, all methods assigned a higher fault score to x_1, indicating that it is successfully identified as faulty. A more careful inspection shows that the scores of x_2 to x_{10} for DL-RBC and DL-IRBC remain noticeable and can be mistakenly identified as faulty. This is expected, as incipient faults struggle to produce reconstruction errors that are distinguishable from normal variables. In contrast, MWRBC, SpCP, SMD, and SGGL yield reasonable isolation results. Unlike other methods, SASDL focuses on process signal decomposition, enabling the isolation model to achieve greater accuracy. The associated plots of SASDL show that the SPE fault score for x_1 is significantly higher than that of T^2, indicating that reconstruction error captures more fault information, which confirms SPE's higher FDR in Table II.

Applying isolation approaches to $\mathbf{X}_2$ yields the remaining plots in Fig. 2. This time, the higher fault scores reveal that MWRBC, SGGL, and SASDL accurately isolated six faulty variables ($x_1 \sim x_4$, x_9, x_{10}) induced by h_1. On the other hand, DL-RBC, DL-IRBC, SpCP, and SMD produced incorrect isolation results. Although MWRBC and SGGL also identified the six faulty variables, the fault scores for other variables are still noticeable. A closer examination shows that SASDL assigns fault scores to $x_9 \sim x_{10}$ near 0.6, as stipulated in (23), more accurately than MWRBC and SGGL. This indicates that SASDL better estimates fault magnitude by leveraging sparse dictionary learning to capture fault characteristics, achieving structured sparse patterns that alleviate the impact of fault propagation, as shown in Fig. 3.

One can see from Fig. 3 that sparse patterns can capture both the fault affecting reconstruction errors in $\mathbf{X}_1$ and the fault impacting dictionary coefficients in $\mathbf{X}_2$. Therefore, sparse

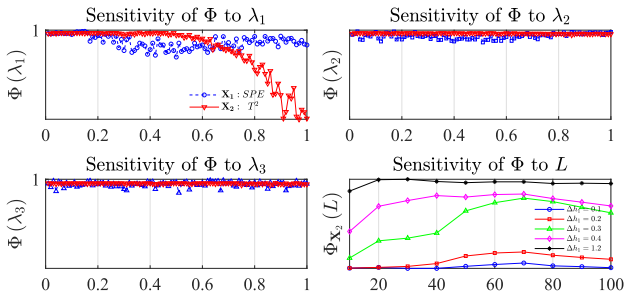

Fig. 4. SASDL-based monitoring metrics under different parameters, excluding those metrics based on T^2 statistics for $\mathbf{X}_1$ and SPE statistics for $\mathbf{X}_2$.

pattern discovery is effective for analyzing both types of faults. By uncovering fault patterns, the proposed SASDL offers a flexible approach to fault direction construction, addressing some limitations of other dictionary learning methods. It shows improved performance compared to traditional approaches, especially in detecting and isolating systematic faults involving multiple process variables. This is expected, as the enhanced fault isolation capability is grounded in the theoretical foundations, which shows that the sparse augmented dictionary efficiently consolidates dispersed fault information.

C. Experimental Validation

1) Validation of Assumptions in Theorems*: First* , in order to validate the assumptions in Theorem 1. The mutual information (MI) [36] between the reconstructed variables $\mathbf{D}_0\mathbf{S}_t^0$ and the ℓ_2-norm of $\frac{1}{L}\sum_{i=t-L+1}^{t}\mathbf{s}_i^0$ are calculated. A low MI, with $\max(\mathbf{MI}_t) = 0.18$ (on a [0,1] scale) for $\mathbf{D}_0\mathbf{S}_t^0$, validates the approximate independence. In addition, a small ℓ_2-norm of $\frac{1}{L}\sum_{i=t-L+1}^{t}\mathbf{s}_i^0$ (with $L = 30$) and a maximum of 0.015 confirm the near-zero mean in Theorem 1. *Second*, the reconstruction errors of 1000 simulated in-statistical-control samples satisfy $\left\|\mathbf{x}_t - \mathbf{D}_0\mathbf{s}_t^0\right\|_2 \le 0.83$ and $\left\|\mathbf{x}_t - \mathbf{D}_0\mathbf{s}_t^0 - \mathbf{D}_t\widetilde{\alpha}_t\right\|_2 \le 0.15$, supporting the preconditions of boundedness in Theorems 2 and 3. *Third*, to validate Theorem 4, varying degree of fault magnitude is introduced to $\mathbf{X}_1$. It is shown that no fault was detected for deviations below 0.2 using a window length of $L = 30$, while partial identification occurs in the range of 0.4–0.6. This suggests that accurate isolation is more likely when the fault magnitude exceeds a certain threshold.

2) Parameter Sensitivity Analysis: To assess the impact of parameter change on model performance, the metrics of $\Phi(\cdot) = \mathrm{FDR}(1 - \mathrm{FAR})$ are defined for the purpose of integrating Type I and Type II errors. Varying the parameters λ_1, λ_2, and λ_3 with a step of 0.01, and L with a step of 10 yields the results in Fig. 4. It shows that the proposed method is robust to regularization parameters, with λ_1 having the most significant impact as it directly influences data reconstruction in sparse coding, while λ_2 and λ_3 primarily regulate coefficient structure and atom similarities to ensure stable monitoring. In addition, experiments demonstrate that increasing the window length L enhances subtle fault detection and isolation but may degrade performance due to detection delay. Moreover, larger β_1 and β_2 in (8) reduce the number of isolated faulty variables, improving robustness but increasing the risk of losing crucial fault information. Compared to β_1, β_2 was found to be less sensitive as it maintains consistency in fault pattern discovery.

TABLE III
AVERAGE FDR AND FAR (%) IN ABLATION STUDIES FOR $\mathbf{X}_1$ AND $\mathbf{X}_2$

Dataset		$\mathbf{X}_1$				$\mathbf{X}_2$			
Module	Ablated term	T^2		SPE		T^2		SPE	
		FDR	FAR	FDR	FAR	FDR	FAR	FDR	FAR
I	$\|\sum_j \mathbf{s}_j^0\|^2$	8.57	0.0	**60.0**	1.33	**21.33**	0.0	0.0	0.0
II	$\|\mathbf{S}_t^0 - \widetilde{\mathbf{S}}_t^0\|_F^2$	17.43	**12.67**	95.1	**5.2**	96.67	**5.0**	61.6	1.5
III	$\|\mathbf{D}_t - \widetilde{\mathbf{D}}_t\|_F^2$	0.0	0.0	**60.57**	**2.67**	0.0	0.0	0.0	0.0

3) Ablation Study: To evaluate the contribution of each term in the proposed SASDL, ablation experiments are conducted, which focuses on statistical property embedding (Module I), coefficient structure (Module II), and atom similarity (Module III). The results are summarized in Table III.

The significant performance degradation in Table III confirms that Module I and Module III are critical for fault detection, as their removal substantially affects FDR across both T^2 and SPE statistics. In contrast, while eliminating Module II also reduces FDR, its primary role is to maintain stable monitoring, as indicated by the notable increase in FAR. Furthermore, isolation procedures reveal that removing Module III leads to inconsistent fault patterns, potentially causing erroneous results. These findings highlight the necessity of incorporating Module I and III for reliable detection and isolation, while Module II ensures monitoring stability.

VI. Application to Industrial Processes

This section evaluates the monitoring performance of the proposed SASDL in two industrial processes: a glass melter process (GMP) [37] and a distillation process (DP) [38].

A. Application to a Glass Melter Process

In the GMP, powdered waste and raw glass enter a vessel heated by four induction coils. Once melted, the mixture is poured out, repeating the cycle. Sensors track temperatures (T_1–T_8), coil powers (P_1–P_4), glass viscosity (μ), and coil voltage (V), collecting 14 variables every 5 min. For monitoring, two datasets are prepared: a reference dataset (780 samples, 65 h) for training and a faulty dataset (230 samples, 19 h) recording a developing vessel crack for testing.

For comparison, the methods in Section V are considered, yielding the monitoring results in Fig. 5 and Table IV. As seen in Fig. 5, the GMP experienced a developing fault, but basic-DL only flagged significant violations in the last 15 samples. In contrast, SASDL-SPE issued an early warning at the eleventh time instance, demonstrating its effectiveness in detecting developing or incipient faults. Table IV presents the FDRs of all methods, with higher FDRs further confirming SASDL's superior detection performance.

To isolate the melter fault, SASDL parameters are set as $\beta_1 = 10$, $\beta_2 = 0.01$ for sparse patterns and $\beta_1 = 2.8$, and $\beta_2 =$

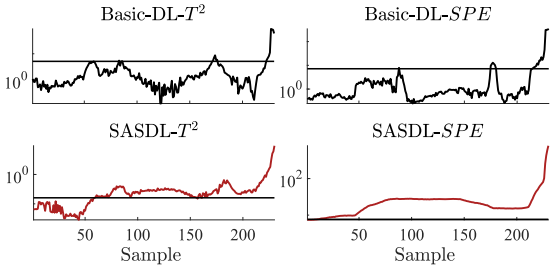

Fig. 5. Monitoring results using Basic-DL and SASDL for GMP with parameter settings: $K = 50$, $\tau = 5$, $L = 100$, $\eta = 0.6$, and $\{\lambda_1 = 0.3,\ \lambda_2 = 0.2,\ \lambda_3 = 0.5\}$.

TABLE IV
FDR (%) OF COMPETITIVE METHODS FOR GMP AND DP

Method	Glass melter process		Distillation process	
	T^2/st. 1	SPE/st. 2	T^2/st. 1	SPE/st. 2
PCA	5.22	13.48	10.67	2.22
Basic-DL	6.52	10.0	8.67	19.8
MWRBC	13.91	67.39	9.11	92.22
KL	58.70	57.39	8.89	63.11
SMD	74.35	—	95.78	—
MCC	70.43	73.04	85.11	82.0
SGGL	78.26	—	94.22	—
SASDL	**73.48**	**95.65**	67.51	**99.6**

Note: PCA-based methods use 3 PCs for GMP and 9 PCs for DP. In SGGL, $\lambda_1 = 0.01$ (sparsity) and $\lambda_2 = 0.4$ (graph similarity) for GMP, while $\lambda_1 = 0.2$ and $\lambda_2 = 4$ for DP.

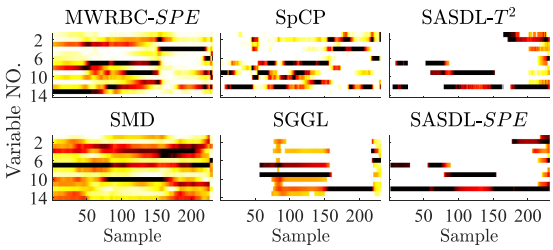

Fig. 6. Sample-by-sample fault scores using MWRBC, SpCP, SMD, SGGL, and SASDL for the GMP. DL-RBC and DL-IRBC are excluded as they provide no useful information for fault isolation.

TABLE V
IDENTIFIED FAULTY VARIABLES OF COMPETITIVE METHODS FOR GMP AND DP

Method	Glass Melter Process	Distillation Process
SpCP	$T_2, T_7, P_1 \sim P_4$	x_4, x_8
MWRBC	$T_2, T_3, T_7, T_8, P_1 \sim P_4, \mu$	$x_2, x_4, x_5, x_8, x_{10}, x_{16}$
SMD	$T_3 \sim T_5, T_7, P_2$	x_2, x_4
SGGL	$P_4, P_7, T_1, T_2, T_4, \mu$	x_2, x_4, x_5, x_8
SASDL	T_7, P_1, P_4	x_2, x_4, x_5, x_8

0.01 for residuals. The isolation results, including those of other methods, are shown in Fig. 6, which presents 2-D plots indicating two fault stages in the glass melter: an incipient stage (samples 1–160) and a late stage (remaining samples).

For a clearer comparison, Table V summarizes the isolation results after excluding variables with low fault scores due to process noise or disturbance, while later-stage faults are omitted due to widespread failures caused by the melter crack. Compared with competitive methods, which suffered from a "smearing effect" and misidentified more variables as faulty, SASDL focused on key problematic variables: T_7, P_1, and P_4 in the incipient stage. Further analysis revealed that an initial crack in the melter vessel was reflected in the readings of sensors at the seventh temperature position and the first and fourth coils. Thus, SASDL offered more effective incipient isolation.

TABLE VI
VARIABLES AND DESCRIPTIONS FOR A DISTILLATION PROCESS

Variable	Description	Unit	Variable	Description	Unit
x_1	Tray 14 temperature	°C	x_9	Reboiler vessel level	%
x_2	Top product level	%	x_{10}	Reboiler temperature	°C
x_3	Tray 2 temperature	°C	x_{11}	Botton draw	t/h
x_4	Reflux vessel level	%	x_{12}	Reflux flow	t/h
x_5	Butane product flow	t/h	x_{13}	Butane level in hexane	%
x_6	Tray 31 temperature	°C	x_{14}	Reboiler steam flow	t/h
x_7	Prone level in butane	%	x_{15}	Feed flow	t/h
x_8	Hexane level in butane	%	x_{16}	Feed temperature	°C

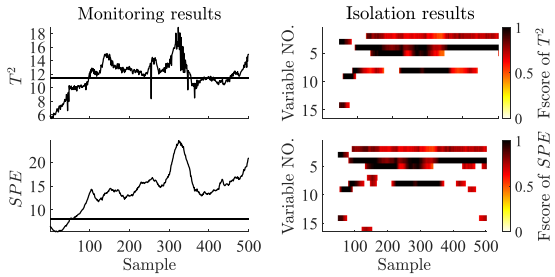

Fig. 7. Monitoring results and sample-by-sample fault scores using SASDL for the DP. The model parameters are set as: $K = 48$, $\tau = 5$, $L = 50$ $\lambda_1 = 0.15$, $\lambda_2 = 0.03$, $\lambda_3 = 0.1$, and $\eta = 0.8$.

B. Application to a Distillation Process

A DP purifies butane from a hydrocarbon mixture, involving 16 key variables in Table VI. To ensure stable operation, these variables are recorded every 30 s for process monitoring, helping maintain distillation tower temperatures and feed concentrations within predefined limits. For model training and testing, 1000 reference samples and 500 test samples are collected, with an incipient fault impacting the hexane level in butane from the 50th sample onward.

Similar to Section VI-A, the aforementioned approaches are used to monitor the DP, with the parameter configurations and monitoring results recorded in Fig. 7 and Table IV. For simplicity, the FARs of all methods, being less than 2% , are not included in Table IV. As shown in the left two plots of Fig. 7, numerous violations after the 50th sample are detected by both T^2 and SPE statistics for SASDL, achieving the highest FDR of 99.6% in Table IV. This again confirms its enhanced detection performance.

For fault isolation, setting $\beta_1 = 20$ and $\beta_2 = 0.01$ yields the results shown in Fig. 7 and Table V. It can be seen from the right two plots of Fig. 7 that SASDL identifies the faulty variables as x_2, x_4, x_5, and x_8, which aligns closely with the results of SGGL in Table V, which is in accordance with practical situations. In

contrast, other methods either identify just a part of the faulty variables (SpCP and SMD), or identified too many variables as faulty (MWRBC). This further demonstrates the superior performance of our proposed method.

VII. Conclusion

This article proposes a structured augmented sparse dictionary learning method for industrial process monitoring. By embedding statistical properties and local geometric structure into sparse coding, the process signal is decomposed into the fault-free and fault-related components using sequential low-dimensional sparse dictionaries. This enables a moving-window approach for fault detection and isolation, enhancing sensitivity by concentrating dispersed abnormal information and constructing noninterfering monitoring statistics. The resulting informative patterns allow feature selection algorithms, such as $\ell_{2,0}$-norm regularization, to accurately pinpoint faulty variables. Both theoretical analysis and practical applications confirm the method's superiority over competing techniques. Future work will aim to incorporate complex process characteristics and improve dynamic monitoring performance.

References

[1] M. Weese, W. Martinez, F. M. Megahed, and L. A. Jones-Farmer, "Statistical learning methods applied to process monitoring: An overview and perspective," *J. Qual. Technol.*, vol. 48, no. 1, pp. 4–24, 2016.

[2] S. A. A. Taqvi, H. Zabiri, L. D. Tufa, F. Uddin, S. A. Fatima, and A. S. Maulud, "A review on data-driven learning approaches for fault detection and diagnosis in chemical processes," *ChemBioEng Rev.*, vol. 8, no. 3, pp. 239–259, 2021.

[3] Y. Xu, S.-Q. Shen, Y.-L. He, and Q.-X. Zhu, "A novel hybrid method integrating ICA-PCA with relevant vector machine for multivariate process monitoring," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 4, pp. 1780–1787, Jul. 2019.

[4] S. Ayesha, M. K. Hanif, and R. Talib, "Overview and comparative study of dimensionality reduction techniques for high dimensional data," *Inf. Fusion*, vol. 59, pp. 44–58, 2020.

[5] J. Piironen, M. Paasiniemi, and A. Vehtari, "Projective inference in high-dimensional problems: Prediction and feature selection," *Electron. J. Statist.*, vol. 14, no. 1, pp. 2155–2197, 2020.

[6] Q. Tang, Y. Liu, Y. Chai, C. Huang, and B. Liu, "Dynamic process monitoring based on canonical global and local preserving projection analysis," *J. Process Control*, vol. 106, pp. 221–232, 2021.

[7] G. Wang, J. Yang, Y. Qian, J. Han, and J. Jiao, "KPCA-CCA-based quality-related fault detection and diagnosis method for nonlinear process monitoring," *IEEE Trans. Ind. Informat.*, vol. 19, no. 5, pp. 6492–6501, May 2023.

[8] F. Harrou et al. *Statistical Process Monitoring Using Advanced Data-Driven and Deep Learning Approaches: Theory and Practical Applications*. Amsterdam, The Netherlan: Elsevier, 2020.

[9] B. Dumitrescu and P. Irofti, *Dictionary Learning Algorithms and Applications*. Berlin, Germany: Springer, 2018.

[10] L. Ren and W. Lv, "Fault detection via sparse representation for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 27, no. 2, pp. 252–259, May 2014.

[11] X. Peng, Y. Tang, W. Du, and F. Qian, "Multimode process monitoring and fault detection: A sparse modeling and dictionary learning method," *IEEE Trans. Ind. Electron.*, vol. 64, no. 6, pp. 4866–4875, Jun. 2017.

[12] K. Huang et al., "Adaptive multimode process monitoring based on mode-matching and similarity-preserving dictionary learning," *IEEE Trans. Cybern.*, vol. 53, no. 6, pp. 3974–3987, Jun. 2023.

[13] H. Ren, Z. Chen, X. Liang, C. Yang, and W. Gui, "Association hierarchical representation learning for plant-wide process monitoring by using multilevel knowledge graph," *IEEE Trans. Artif. Intell.*, vol. 4, no. 4, pp. 636–649, Aug. 2023.

[14] K. Huang, H. Wen, H. Ji, L. Cen, X. Chen, and C. Yang, "Nonlinear process monitoring using kernel dictionary learning with application to aluminum electrolysis process," *Control Eng. Pract.*, vol. 89, pp. 94–102, 2019.

[15] K. Huang et al., "Adaptive process monitoring via online dictionary learning and its industrial application," *ISA Trans.*, vol. 114, pp. 399–412, 2021.

[16] R. Zemouri, R. Ibrahim, and A. Tahan, "Hydrogenerator early fault detection: Sparse dictionary learning jointly with the variational autoencoder," *Eng. Appl. Artif. Intell.*, vol. 120, 2023, Art. no. 105859.

[17] C. Ning, M. Chen, and D. Zhou, "Sparse contribution plot for fault diagnosis of multimodal chemical processes," *IFAC-PapersOnLine*, vol. 48, no. 21, pp. 619–626, 2015.

[18] C. F. Alcala and S. J. Qin, "Reconstruction-based contribution for process monitoring," *Automatica*, vol. 45, no. 7, pp. 1593–1600, 2009.

[19] K. Huang, H. Wen, H. Liu, C. Yang, and W. Gui, "A geometry constrained dictionary learning method for industrial process monitoring," *Inf. Sci.*, vol. 546, pp. 265–282, 2021.

[20] H. Wang, G. Dong, J. Chen, X. Hu, and Z. Zhu, "A novel dictionary learning named deep and shared dictionary learning for fault diagnosis," *Mech. Syst. Signal Process.*, vol. 182, 2023, Art. no. 109570.

[21] H. Ji, X. He, J. Shang, and D. Zhou, "Incipient sensor fault diagnosis using moving window reconstruction-based contribution," *Ind. Eng. Chem. Res.*, vol. 55, no. 10, pp. 2746–2759, 2016.

[22] H. Ji, "Statistics mahalanobis distance for incipient sensor fault detection and diagnosis," *Chem. Eng. Sci.*, vol. 230, Art. no. 116233, 2021.

[23] J. Zeng, U. Kruger, J. Geluk, X. Wang, and L. Xie, "Detecting abnormal situations using the kullback–leibler divergence," *Automatica*, vol. 50, no. 11, pp. 2777–2786, 2014.

[24] A. Taalimi, "Learning multimodal structures in computer vision," Ph.D. dissertation, University of Tennessee, Knoxville, 2017.

[25] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Stat. Soc. Ser. B: Stat. Methodol.*, vol. 67, no. 2, pp. 301–320, 2005.

[26] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[27] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, no. 1, pp. 19–60, 2010.

[28] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[29] K. Huang, H. Zhu, D. Wu, C. Yang, and W. Gui, "EaLDL: Element-aware lifelong dictionary learning for multimode process monitoring," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 2, pp. 3744–3757, Feb. 2025.

[30] K. Huang, Y. Wu, H. Wen, Y. Liu, C. Yang, and W. Gui, "Distributed dictionary learning for high-dimensional process monitoring," *Control Eng. Pract.*, vol. 98, 2020, Art. no. 104386.

[31] R. Tibshirani, "The lasso method for variable selection in the cox model," *Statist. Med.*, vol. 16, no. 4, pp. 385–395, 1997.

[32] W. Chen, J. Zeng, X. Xu, S. Luo, and C. Gao, "Structured sparsity modeling for improved multivariate statistical analysis based fault isolation," *J. Process Control*, vol. 98, pp. 66–78, 2021.

[33] K. Wang, A. B. Yeh, and B. Li, "Simultaneous monitoring of process mean vector and covariance matrix via penalized likelihood estimation," *Comput. Statist. Data Anal.*, vol. 78, pp. 206–217, 2014.

[34] J. Wang, Y. Liu, D. Zhang, L. Xie, and J. Zeng, "Structured discriminative gaussian graph learning for multimode process monitoring," *J. Chemometrics*, vol. 38, no. 3, 2024, Art. no. e3538.

[35] Z. He, Y. A. Shardt, D. Wang, B. Hou, H. Zhou, and J. Wang, "An incipient fault detection approach via detrending and denoising," *Control Eng. Pract.*, vol. 74, pp. 1–12, 2018.

[36] B. C. Ross, "Mutual information between discrete and continuous data sets," *PLoS One*, vol. 9, no. 2, 2014, Art. no. e87357.

[37] T. Feital, U. Kruger, L. Xie, U. Schubert, E. L. Lima, and J. C. Pinto, "A unified statistical framework for monitoring multivariate systems with unknown source and error signals," *Chemometrics Intell. Lab. Syst.*, vol. 104, no. 2, pp. 223–232, 2010.

[38] L. Xie, Z. Li, J. Zeng, and U. Kruger, "Block adaptive kernel principal component analysis for nonlinear process monitoring," *AIChE J.*, vol. 62, no. 12, pp. 4334–4345, 2016.

Yi Liu received the M.S. degree in military equipment from the Air Force Engineering University of PLA, Xi'an, China, in 2011, and the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2020.

He is currently working with Hangzhou Normal University, Hangzhou, China. His research interests include data-driven industrial process monitoring and structured deep graph representation learning.

Jiusun Zeng received the B.S. and Ph.D. degrees in mathematics from Zhejiang University, Hangzhou, China, in 2004 and 2009, respectively.

In 2022, he joined Hangzhou Normal University, Nanjing, China, where he is currently a Full Professor with the School of Mathematics. From 2009 to 2011, he was a Postdoctoral Research Associate with the Institute of Cyber Systems and Control, Zhejiang University. After that, he became a Faculty Member of China Jiliang University, Hangzhou, China. His research interests include Big Data analysis, multivariate statistical process monitoring, and fault diagnosis.

Zidong Wang (Fellow, IEEE) received the B.S. degree in mathematics from Suzhou University, Suzhou, China, in 1986, and the M.S. and Ph.D. degrees in applied mathematics and electrical engineering from the Nanjing University of Science and Technology, Nanjing, China, in 1990 and 1994, respectively.

He is currently a Professor of Dynamical Systems and Computing, with the Brunel University, London, U.K. He held Academic Positions in China, Germany, and the U.K. His research interests include dynamical systems, signal processing, bioinformatics, and control theory.

Weiguo Sheng received the M.Sc. degree in information technology from the University of Nottingham, U.K., in 2002, and the Ph.D. degree in computer science from Brunel University London, U.K., in 2005.

He is currently a Professor with Hangzhou Normal University, Hangzhou, China. He worked as a Researcher with the University of Kent, Canterbury, U.K., and Royal Holloway, University of London, London, U.K. His research interests include evolutionary computation, data mining/clustering, pattern recognition, and machine learning.

Chuanhou Gao (Senior Member, IEEE) received the B.Sc. degree in chemical engineering from the Zhejiang University of Technology, Hangzhou, China, in 1998, and the Ph.D. degree in operational research and cybernetics from Zhejiang University, Hangzhou, in 2004.

Since June 2006, he has joined the Department of Mathematics with Zhejiang University, where he is currently a Full Professor. From June 2004 until May 2006, he was a Postdoctor with the Department of Control Science and Engineering with Zhejiang University.

Qi Xie received the Ph.D. degree in applied mathematics from Zhejiang University, Hangzhou, China, in 2005.

He is currently a Professor with the Hangzhou Key Laboratory of Cryptography and Network Security, Hangzhou Normal University, Hangzhou, China. From 2009 to 2010, he was a Visiting Scholar with the University of Birmingham, Birmingham, U.K., and in 2012, he was with the City University of Hong Kong, Hong Kong. His research interests include applied cryptography, including digital signatures, authentication, and key agreement protocols.

Lei Xie received the B.S. and Ph.D. degrees in control science and engineering from Zhejiang University, Hangzhou, China, in 2000 and 2005, respectively.

He is currently a Professor with the College of Control Science and Engineering, Zhejiang University. From 2005 to 2006, he was a Postdoctoral Researcher with the Berlin University of Technology, Berlin, Germanyand from 2005 to 2008, an Assistant Professor with Zhejiang University. His research interests include applied statistics and system control theory.