

Chapter 12

New Robust Projection to Latent Structure



In many actual nonlinear systems, especially near the equilibrium point, linearity is the primary feature and nonlinearity is the secondary feature. For the system that deviates from the equilibrium point, the secondary nonlinearity or local structure feature can also be regarded as the small uncertainty part, just as the nonlinearity can be used to represent the uncertainty of a system (Wang et al. 2019). So this chapter also focuses on how to deal with the nonlinearity in PLS series method, but starts from an different view, i.e., robust PLS. Here the system nonlinearity is considered as uncertainty and a new robust L_1 -PLS is proposed.

The traditional PLS and its nonlinear improvement methods are usually to maximize the covariance between the input and output data, i.e., the square of L_2 norm. L_2 norm has the feature of clear physical meaning and convenient calculation, and its solution are unique unbiased and dense. While it is powerless for systems with rich local features such as nonlinear systems or uncertain systems. The proposed robust L_1 -PLS aims at the robustness of the feature extraction and the regression coefficients. This method maintains the signal relative size during the feature extraction. Moreover, it guarantees the features are robust to outliers in the global statistical view and sensitive to the local structure information.

12.1 Motivation of Robust L_1 -PLS

Many robust PLS methods have been developed to increase the robustness of traditional PLS method recently. Branden (2004) and Hubert (2008) replaced the empirical variance-covariance matrix in PLS by a robust covariance estimator, and used the minimum covariance determinant (MCD) estimator and the reweighted MCD estimator (RMCD) for low-dimensional data sets. Turkmen (2013) proposed the influence function analysis for the robust PLS estimator. Currently, the existing robust PLS methods use robust covariance estimation techniques with the identification of multivariate outliers to maintain robustness (Fortuna et al. 2007; Filzmoser 2016). These

methods actually perform with a potential assumption that the signal is subject to Gaussian distribution, which is not satisfied for many industrial processes. Usually the industrial data are full of lots of outliers and follow either heavy-tailed distribution (Doman'ski 2019) or multipeak distribution (Wang 2000). In other words, the statistical properties of this kind of data cannot be described by the robust covariance matrix estimation. Furthermore, outliers may contain very important information, so the outliers cannot be simply deleted or replaced (Liu et al. 2018). The data also have some nondominant local structure features besides the outliers. Robust covariance estimation methods also do not handle the small uncertainty correctly.

Recently, a robust PCA (RPCA) (Kwak 2008) and a robust sparse PCA (RSPCA) (Meng et al. 2012) were proposed, which the two methods maximized the L_1 norm rather than the square of L_2 norm of the input data. Experiments showed that they are efficient and robust for the data with inherent uncertainty and outliers. However, the two improved RPCA methods do not obtain any useful information from the output quality variables, so it is difficult to directly apply them to quality-relevant process monitoring and fault diagnosis (Zhou et al. 2018). The monitoring system will automatically alarm if a fault is detected whether it affects the product quality or not. Many alarms do not make sense for the final production quality.

It is known that the least absolute deviation (LAD) regression is often better than the least squares (LS) regression for non-Gaussian signals, especially those with a heavy-tailed distribution. While LAD regression is immune to outliers. Moreover, the solution of LAD regression is not unique, and it is necessary to introduce the optimal technique to obtain an optimal solution. So the LAD regression of high-dimensional system is a time-consuming task. To improve the efficiency of the LAD algorithm, the idea of partial least squares (PLS) regression is used to extend the conventional LAD regression to partial LAD regression. The PLS-based monitoring method decomposes the process space through the correlation between the quality and the process variables, which can reflect the quality-relevant product changes in the process variables (Wang et al. 2017; Zhou et al. 2018).

In order to enhance the robustness of the PLS method in a new way, this chapter proposes a novel dual robustness projection to latent structure regression method based on the L_1 norm, L_1 -PLS. The optimization objective during the principle components extraction in the PLS method is a square of L_2 norm, i.e., the least squares regression problem. L_1 -PLS use the L_1 norm maximization to replace the square of the L_2 norm maximization in the traditional PLS methods. The L_1 norm penalty terms are added to the direction vectors in the latent structure construction. Moreover, the partial LAD regression is used to obtain the regression coefficients. Therefore, the L_1 -PLS regression method achieves dual robust capabilities including robust principle components and regression coefficients. On the other hand, the L_1 norm optimization target also has the certain capability of local structural feature retention, compared with the L_2 norm optimization goal.

L_1 -PLS is distinguished from other existing robust PLS methods in several respects:

- (1) The noises, outliers, and local structure features generally enter the system through the direction vectors, and the L_1 norm can maintain the relative size of the original signal; its direction vectors are robust to outliers and contain more local structure features even if there is no preprocessing of outliers. This facilitates the L_1 norm to obtain the global and local features of the system at the same time without destroying the integrity of the samples;
- (2) The L_1 -PLS method with the L_1 norm penalty term to the direction vectors can obtain the sparse principle components, and filter out the disturbance variables or those sparse PCs that are robust to disturbance variables;
- (3) The regression coefficients are obtained by the partial LAD regression. The corresponding regression model is also robust to outliers or uncertainties, and the model has better predictive performance.

12.2 Introduction to RSPCA Method

Consider the input data $X = [x(1), \dots, x(n)] \in R^{m \times n}$, where $x = [x_i, \dots, x_m]$; m and n are the dimensionality of the input data and the size of the input matrix. The traditional PCA method aims to find the d ($d < m$) dimensional linear subspace with the largest input data variance. The objective function is as follows:

$$W^* = \arg \max \|W^T X\|_2^2, \text{ s.t. } W^T W = I_d, \quad (12.1)$$

where $W = [w_1^T, \dots, w_d^T]^T \in R^{m \times d}$ is weight matrix. $\|\cdot\|_2$ represents the L_2 norm of a matrix or vector.

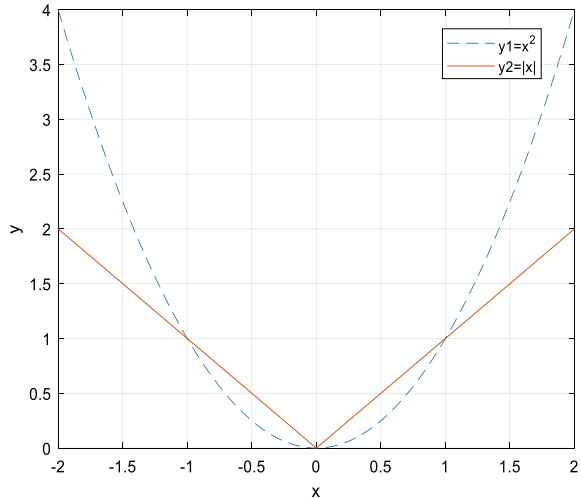
However, the principal components based on the PCA are usually a linear combination of the original variables usually with the non-zero weights. The non-zero weight results in that many irrelevant variables are included in the final model and cause unnecessary interference. Therefore, the sparse PCA (SPCA) method was proposed to achieve the sparse expression of the principal components as much as possible (Liu 2014). Its objective function is

$$W^* = \arg \max \|W^T X\|_2^2, \text{ s.t. } W^T W = I_d, \|W\|_1 < s, \quad (12.2)$$

where $\|\cdot\|_1$ is the L_1 norm of a matrix or vector. It is introduced as constraint or penalty term to enhance the sparsity of the principal components. s is the number of non-zero weights. The L_1 norm penalty term ($\|W\|_1 < s$) realizes the sparse expression of the direction vector.

Figure 12.1 shows the amplifying effect curve of L_1 norm and L_2 norm on noise. The blue dotted line is the square of the L_2 norm (for one-dimensional data, it is equivalent to the L_2 norm), and the red line is the L_1 norm. Obviously, the L_2 norm has an inhibitory effect on the data in $|x| \leq 1$ and has an enlarged effect on the data in $|x| > 1$. The L_1 norm maintains the relative size of the original data and has a

Fig. 12.1 The expanding effects of the L_1 norm and L_2 norm curve



relatively small expansion effect on all data. In order to further improve the robustness of SPCA, the RSPCA method is proposed to reduce the sensitivity of the principal components to outliers. The L_2 norm in the objective function is substituted by L_1 norm (Zou et al. 2006). The optimization function of RSPCA is given as follows:

$$w^* = \arg \max \|X^T w\|_1, \text{ s.t. } w^T w = 1, \|w\|_1 < s. \quad (12.3)$$

Here the optimization problem is a form of L_1 norm maximization with an L_1 norm penalty term simultaneously. In order to obtain the principal components of the RSPCA method, the optimal direction vector w^* is calculated by Algorithm 3.

The convergence of Algorithm 3 and the rationality of the obtained sparse direction vectors have been theoretically verified (Zou et al. 2006). However, Algorithm 3 indicates that the sparseness of the data needs to be given in prior during the calculation of the sparse direction vector. Generally speaking, the sparsity of input data is unknown and it contains uncertainty. More importantly, the RSPCA method cannot be directly applied to quality-related process monitoring. Therefore, this chapter introduces the L_1 norm into the PLS method.

12.3 Basic Principle of L_1 -PLS

The double robust projection to latent structure (L_1 -PLS) method is given based on the L_1 norm, aiming at improving the robustness of the traditional PLS method. The PLS method extracts principal components from the input space and output space, and the principal components should satisfy the following conditions: carry the maximum variation information (representation) of their respective variable spaces as much as

Algorithm 3 RSPCA algorithm for one sparse PC**Input:**Data matrix X , sparsity s .**Output:**The s sparse PC w^* .

- 1: Initialization $w(0) \in R^{1 \times m}$, set $w(0) = \frac{w(0)}{\|w(0)\|_2}$, and $k = 0$.
- 2: Let $v = (v_1, \dots, v_m)^T = \sum_{i=1}^n p_i(t) X_i$, where $p_i(k) = \begin{cases} 1, & w^T(k) X_i \geq 0 \\ -1, & w^T(k) X_i < 0 \end{cases}$ and X_i is the i th column of the matrix X . Let γ be the $(s + 1)$ largest element in $|v|$.
- 3: Let $\beta = (\beta_1, \dots, \beta_m)^T$, where $\beta_i = \text{sgn}(v_i)(|v_i| - \gamma)_+$, $i = 1, \dots, m$, and $(z)_+ = \begin{cases} z, & x > 0 \\ 0, & x \leq 0 \end{cases}$, $\text{sgn}(z) = \begin{cases} 1; & z > 0 \\ 0; & z = 0 \\ -1; & z < 0 \end{cases}$. Make $w(k + 1) = \frac{\beta}{\|\beta\|_2}$, and $k = k + 1$.
- 4: If $w(k) \neq w(k + 1)$, return to Step 2; otherwise continue to Step 5.
- 5: If there is i such that $w^T(k) X_i = 0$ and $\text{sgn}\left(\sum_{j=1}^m |w(k)_j X_{j,i}|\right) \neq 0$, then let $\frac{w^T(k) + \Delta w}{\|w^T(k) + \Delta w\|_2}$ and return to Step 2; otherwise continue to Step 6; Δw is a small non-zero random vector.
- 6: Set $w^* = w(k)$ and stop iteration.
- 7: **return** w^* ;

possible, and the degree of correlation between different variable spaces is as large as possible (correlation). Take the extraction of the first principal component as an example. The PLS method is expressed as follows:

$$\begin{aligned} E_0^T F_0 F_0^T E_0 w_1 &= \theta^2 w_1 \\ F_0^T E_0 E_0^T F_0 c_1 &= \theta^2 c_1, \end{aligned} \quad (12.4)$$

where w_1 and c_1 are the direction vector of the principle components t_1 and u_1 . The optimization problem (12.4) is transformed into finding the unit direction vectors w_1 and c_1 corresponding to the maximum eigenvalue θ^2 of matrices $E_0^T F_0 F_0^T E_0$ and $F_0^T E_0 E_0^T F_0$, respectively. It can be seen that the solution of (12.4) satisfies the requirements about the representation and correlation in PLS method.

Then, multiply both sides of the equation (12.4) by w_1^T and c_1^T , respectively, and obtain

$$\begin{aligned} w_1^T E_0^T F_0 F_0^T E_0 w_1 &= \theta^2, & \text{s.t. } w_1^T w_1 &= 1 \\ c_1^T F_0^T E_0 E_0^T F_0 c_1 &= \theta^2, & \text{s.t. } c_1^T c_1 &= 1. \end{aligned} \quad (12.5)$$

To simplify further, we can get

$$\begin{aligned} w_1^* &= \arg \max \|w_1^T E_0^T F_0\|_2^2, & \text{s.t. } w_1^T w_1 &= 1 \\ c_1^* &= \arg \max \|c_1^T F_0^T E_0\|_2^2, & \text{s.t. } c_1^T c_1 &= 1. \end{aligned} \quad (12.6)$$

The optimal problem of the traditional PLS (12.4) is expressed as L₂ norm optimization in (12.6). w_1^* and c_1^* are the optimal direction vectors.

It is known that the noise is flowed into the regression model through the direction vector (\mathbf{w}_1 and \mathbf{c}_1) in most cases, which affects the estimation of the regression parameters in the PLS method. Similar as the idea of equation (12.3), we replace the maximization of the L_2 norm in the objective function (12.6) with the maximization of L_1 norm. Moreover, the L_1 norm penalty term is added to the direction vector. Therefore, the objective function of the L_1 -PLS method based on the L_1 norm is given as follows:

$$\begin{aligned} \mathbf{w}_1^* &= \arg \max \left\| \mathbf{w}_1^T \mathbf{E}_0^T \mathbf{F}_0 \right\|_1, \quad \text{s.t.} \quad \mathbf{w}_1^T \mathbf{w}_1 = 1, \quad \|\mathbf{w}_1\|_1 < s_1 \\ \mathbf{c}_1^* &= \arg \max \left\| \mathbf{c}_1^T \mathbf{F}_0^T \mathbf{E}_0 \right\|_1, \quad \text{s.t.} \quad \mathbf{c}_1^T \mathbf{c}_1 = 1, \quad \|\mathbf{c}_1\|_1 < s_2, \end{aligned} \quad (12.7)$$

where s_1 and s_2 are the sparsity of input spatial data and output spatial data, respectively.

According to the above analysis, although the direction vectors (\mathbf{w}_1 and \mathbf{c}_1) in (12.4) contains the correlation between the input data \mathbf{E}_0 and the output data \mathbf{F}_0 , fortunately, they can be solved separately in (12.7). Therefore, Algorithm 3 also is suitable for the solution of (12.7) by replacing the corresponding input data matrix \mathbf{X} with $\mathbf{E}_0^T \mathbf{F}_0$ and $\mathbf{F}_0^T \mathbf{E}_0$, respectively. It is noted that the solution of \mathbf{w}_1 and \mathbf{c}_1 are independent but not jointed by Algorithm 3.

Once the optimal direction vectors \mathbf{w}_1 and \mathbf{c}_1 are obtained, the score vectors in the latent space, i.e., the first principle component pair, \mathbf{t}_1 and \mathbf{u}_1 can be calculated

$$\mathbf{t}_1 = \mathbf{E}_0 \mathbf{w}_1, \mathbf{u}_1 = \mathbf{F}_0 \mathbf{c}_1. \quad (12.8)$$

Next, the regression coefficients (loading vectors) of \mathbf{F}_0 and \mathbf{E}_0 to \mathbf{t}_1 will be established. In the traditional PLS model, the regression coefficients \mathbf{p}_1 and \mathbf{q}_1 are estimated by least squares, namely,

$$\begin{aligned} \mathbf{p}_1 &= \mathbf{E}_0^T \mathbf{t}_1 / \|\mathbf{t}_1\|^2 \\ \mathbf{q}_1 &= \mathbf{F}_0^T \mathbf{t}_1 / \|\mathbf{t}_1\|^2. \end{aligned} \quad (12.9)$$

Similarly, least squares estimation is also susceptible to outliers, and the least absolute deviation (LAD) method is introduced to deal with this problem. Therefore, in order to further improve the robustness, LAD regression is used to solve the regression coefficients in the L_1 -PLS algorithm, namely,

$$\begin{aligned} \mathbf{p}_1^* &= \arg \min \left\| \mathbf{E}_0 - \mathbf{t}_1 \mathbf{p}_1^T \right\|_1 \\ \mathbf{q}_1^* &= \arg \min \left\| \mathbf{F}_0 - \mathbf{t}_1 \mathbf{q}_1^T \right\|_1, \end{aligned} \quad (12.10)$$

where \mathbf{p}_1^* and \mathbf{q}_1^* are the optimal loading vectors of (12.10).

Obviously, the essence of (12.10) is also the form of L_1 norm. When there are few outliers, it is not necessary to use the norm to solve the regression coefficient. Due to the direction vector has been solved by maximizing the L_1 norm, the influence of

the outlier has been reduced, and as can be seen from Fig. 12.1. When the outlier is small, the L₂ norm and the L₁ norm have the same effect.

Calculate the residual matrix E_1 and F_1 :

$$E_1 = E_0 - t_1 p_1^T, F_1 = F_0 - t_1 q_1^T \quad (12.11)$$

Similar as the extraction of the first principal components pair, the other principal components are calculated iteratively by decomposing the residuals E_i and F_i ($i = 1, \dots, d - 1$). The extraction of principal components is stopped until the model determined by the extracted principal components satisfies the desired requirements.

The dual robustness of the L₁-PLS algorithm is reflected in the following two aspects:

1. Different from the PLS algorithm, Algorithm 3 is used to calculate the direction vector each time. By maximizing the L₁ norm in the objective function, and adding the L₁ norm penalty term to the direction vector, the robustness of the L₁-PLS algorithm is improved. This achieves the robustness of principal component extraction.
2. In the case of many outliers, the regression coefficients can be calculated using least absolute estimation, which can overcome the shortcomings of least squares estimation that is easily affected by outliers, and further enhance the robustness of the L₁-PLS algorithm.

12.4 L₁-PLS-Based Process Monitoring

It is found that only the calculation process of the direction vector w_1 and c_1 (12.7) or the regression coefficient p_1 and q_1 (12.10) is improved in the L₁-PLS method, and other steps are not affected. Therefore, the monitoring process based on the L₁-PLS method is the same as the PLS method. In the process monitoring based on the L₁-PLS method, the T² and T_e² statistics are still used to monitor the principal component subspace and the remaining subspace. Then, the L₁-PLS-based monitoring is described in detail in Algorithm 4 (offline process training) and Algorithm 5 (online process monitoring). The corresponding flowchart is shown in Fig. 12.2.

In Algorithms 4 and 5, Λ and Λ_e represent the sample covariance matrix. The non-parametric kernel density estimation (KDE) method (1.33) is used to estimate the corresponding control limits of T² and T_e².

There is still a key problem in the implementation of Algorithm 4: the sparsity degree s_1 and s_2 need to be given in prior. There are two common strategies to determine s_1 and s_2 . (1) The first one is the variable importance in prediction (VIP) method (Farrés et al. 2015). It judges whether the variable is an irrelevant variable based on the VIP score of the j th predicted value of the response variable. Usually, the “greater than ϵ ” criterion is used as the selection criterion. More precisely, the threshold ϵ

Algorithm 4 L_1 -PLS method for Offline process training

Input:

Normal data sets $\mathbf{X} = [x_1, \dots, x_m] \in R^{n \times m}$, $\mathbf{Y} = [y_1, \dots, y_l] \in R^{n \times l}$, sparsity s_1 and s_2 .

Output:

The control limits T_{lim}^2 and $T_{e,\text{lim}}^2$.

(1) Normalized \mathbf{X} and \mathbf{Y} as \mathbf{E}_0 and \mathbf{F}_0 ,

(2) For $i = 1, \dots, d$ (d is obtained by cross-validation):

(2.1) Apply Algorithm 3 to the projected matrices $\mathbf{E}_{i-1}^T \mathbf{F}_{i-1}$ and $\mathbf{F}_{i-1}^T \mathbf{E}_{i-1}$ to get the direction vectors \mathbf{w}_i and \mathbf{c}_i , respectively.

(2.2) Calculate the score vectors: $\mathbf{t}_i = \mathbf{E}_{i-1} \mathbf{w}_i$, $\mathbf{u}_i = \mathbf{F}_{i-1} \mathbf{c}_i$.

(2.3) Calculate the load vectors:

$$\mathbf{p}_1 = \mathbf{E}_0^T \mathbf{t}_1 / \|\mathbf{t}_1\|^2 \quad \text{or} \quad \mathbf{p}_1^* = \arg \min \left\| \mathbf{E}_0 - \mathbf{t}_1 \mathbf{p}_1^T \right\|_1$$

$$\mathbf{q}_1 = \mathbf{F}_0^T \mathbf{u}_1 / \|\mathbf{u}_1\|^2 \quad \text{or} \quad \mathbf{q}_1^* = \arg \min \left\| \mathbf{F}_0 - \mathbf{u}_1 \mathbf{q}_1^T \right\|_1$$

(2.4) Calculate the Residual matrix: $\mathbf{E}_i = \mathbf{E}_{i-1} - \mathbf{t}_i \mathbf{p}_i^T$, $\mathbf{F}_i = \mathbf{F}_{i-1} - \mathbf{u}_i \mathbf{q}_i^T$.

(3) Describe \mathbf{t}_i with the original matrix \mathbf{E}_0 : $\mathbf{T} = \mathbf{E}_0 \mathbf{R}$,

$$\mathbf{R} = [\mathbf{r}_1, \dots, \mathbf{r}_d], \text{ in which } \mathbf{r}_i = \prod_{j=1}^{i-1} (\mathbf{I}_n - \mathbf{w}_j \mathbf{p}_j^T) \mathbf{w}_i.$$

$$\hat{\mathbf{E}} = \mathbf{T} \mathbf{P}^T = \mathbf{E}_0 \mathbf{R} \mathbf{P}^T$$

$$\bar{\mathbf{E}} = \mathbf{E}_0 - \hat{\mathbf{E}} = \mathbf{E}_0 (\mathbf{I}_n - \mathbf{R} \mathbf{P}^T)$$

(4) For a normalized data sample \mathbf{x} , calculate its estimate, residual and the corresponding PC value.

$$\hat{\mathbf{x}} = \mathbf{R} \mathbf{P}^T \mathbf{x}$$

$$\mathbf{t} = \mathbf{R} \mathbf{x}$$

$$\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}} = (\mathbf{I} - \mathbf{R} \mathbf{P}^T) \mathbf{x}$$

(5) Calculate the statistics T^2 and T_e^2 :

$$T^2 := \mathbf{t} \mathbf{A}^{-1} \mathbf{t}^T = \mathbf{t} \left(\frac{1}{n-1} \mathbf{T}^T \mathbf{T} \right)^{-1} \mathbf{t}^T$$

$$T_e^2 := \mathbf{e} \mathbf{A}_e^{-1} \mathbf{e}^T = \mathbf{e} \left(\frac{1}{n-1} \bar{\mathbf{E}}^T \bar{\mathbf{E}} \right)^{-1} \mathbf{e}^T$$

return T_{lim}^2 and $T_{e,\text{lim}}^2$;

should be adjusted based on the distribution of the overall data in different situations.

(2) The second strategy is the selectivity ratio method (Branden and Hubert 2004). The variable selection ratio is calculated according to the ratio of the interpretation of the \mathbf{X} variable on the \mathbf{Y} target projection component to the residual variance. Then F test is performed to define the boundary between important variables and irrelevant variables. Since the VIP method is simple and easy to implement, the VIP method is selected to determine the sparsity s_1 and s_2 here.

Algorithm 5 L₁-PLS method for Online process monitoring**Input:**

New normalized data sets \mathbf{x}_{new} and \mathbf{y}_{new} .

Output:

Online process monitoring results.

(1) Calculate the new score vectors: $\mathbf{t}_{new} = \mathbf{x}_{new} \mathbf{R}$.

(2) Calculate the new prediction matrix and new residual:

$$\begin{aligned}\tilde{\mathbf{x}}_{new} &= \mathbf{t}_{new} \mathbf{P}^T = \mathbf{x}_{new} \mathbf{R} \mathbf{P}^T \\ \mathbf{e}_{new} &= \mathbf{x}_{new} - \tilde{\mathbf{x}}_{new} = \mathbf{x}_{new} (\mathbf{I}_n - \mathbf{R} \mathbf{P}^T).\end{aligned}$$

(3) Calculate the new statistics T_{new}^2 and $T_{e,new}^2$:

$$\begin{aligned}T_{new}^2 &= \mathbf{t}_{new} \mathbf{\Lambda}^{-1} \mathbf{t}_{new}^T = \mathbf{t}_{new} \left\{ \frac{1}{n-1} \mathbf{T}^T \mathbf{T} \right\}^{-1} \mathbf{t}_{new}^T \\ T_{e,new}^2 &:= \mathbf{e}_{new} \mathbf{A}_e^{-1} \mathbf{e}_{new}^T = \mathbf{e}_{new} \left\{ \frac{1}{n-1} \bar{\mathbf{E}}^T \bar{\mathbf{E}} \right\}^{-1} \mathbf{e}_{new}^T\end{aligned}$$

(4) Compare T_{new}^2 and $T_{e,new}^2$ with the corresponding control limits T_{lim}^2 and $T_{e,lim}^2$.

return Online process monitoring results.

It is worth noting that the role of sparsity is to achieve variable selection. If the established system model contains many irrelevant variables, giving the sparsity is helpful to limit the number of irrelevant variables, so as to realize L₁-sparse-PLS. However, if the sparsity of the input data is uncertain, the sparsity degree s_1 and s_2 can be set equal to the variable number in the input and output space, respectively, to eliminate the uncertainty caused by the sparsification. In this view, the proposed L₁-PLS method is uniformly called as L₁-(S)PLS method based on the different sparsity.

12.5 TE Simulation Analysis

In this simulation, the input variable X is composed of 31 variables [XMEAS(1:22)] and [XMV(1:11) (except XMV(5) and XMV(9))]. The output variable Y consists of the quality components G (XMEAS(35)) and H (XMEAS(36)). Two simulation examples are used to verify the effectiveness of the L₁-PLS method for fault detection.

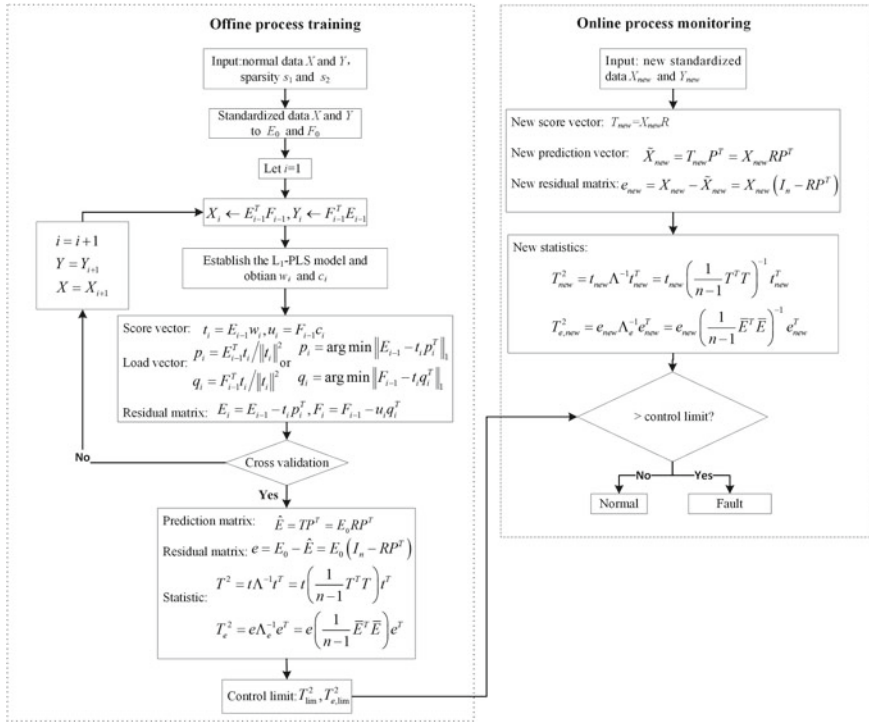


Fig. 12.2 The Flow chart of Algorithms 4 and 5

12.5.1 Robustness of Principal Components

The robustness of the L_1 -PLS method is mainly implemented on the direction vectors, which directly reflects the robustness of the PCs. The variation of the PC structure caused by outliers therefore is the focus of robustness analysis. Here results of PLS and RPLS methods are given for comparison. The input and output data ($X \in R^{960 \times 31}, Y \in R^{960 \times 2}$) are sampled from the TE process under the normal operation for training data. In order to test further the proposed L_1 -PLS, the outliers are added in the input space in the following form:

$$X(k) = X^*(k) + \Xi_j(k), \quad (12.12)$$

where $X^*(k)$ is the k th normal sample ($k = 1, 2, \dots, 960$) Ξ_j is the j -th randomly generated outlier that obey Gaussian distribution $\Xi_j \sim N(0, 2000)$. For ease of verification, three kinds of repeatable outliers that are generated using a specific random seed are added to the training set,

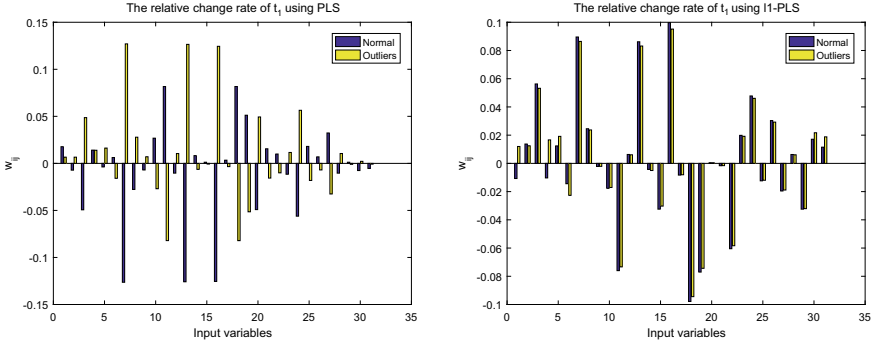


Fig. 12.3 The relative change rates of t_1 using PLS and L_1 -PLS

$$\begin{aligned}
 \mathcal{E}_1(12) &= [-71.294, 4.929, 35.199, -0.100]^T && \text{for } x_{14:17} \\
 \mathcal{E}_2(140) &= [4.164, -16.912, -66.307]^T && \text{for } x_{29:31} \\
 \mathcal{E}_3(200) &= [-1.960, 42.969, 77.737, -19.239, -72.776, 7.439]^T && \text{for } x_{1:6}.
 \end{aligned}$$

Outlier $\mathcal{E}_1(12)$ means that only the 14, 15, 16, and 17th variables at the 12th sample time $X(12)$ are abnormal, and the other variables at other sample times are still normal. The other two outliers have similar meanings.

The sparsity s_1 and s_2 in the L_1 -PLS method are set to 31 and 2. The sparsity is equal to the variable number of input and output space, respectively. In other words, the L_1 -PLS method can reflect the changes in all variables. The components numbers d are determined using cross-validation. They are 6, 6, and 2 for PLS, RPLS, and L_1 -PLS methods, respectively. The principle components are $t_i = \sum_{j=1}^n w_{ij}x_j, i = 1, \dots, d$, in which w_{ij} is the j th element of r_i . The coefficients w_{ij} are used to reflect whether the outliers affect the principle components. The relative rates of change (RRC) indices are defined as follows:

$$\begin{aligned}
 RRC_{1,i} &= \max\{|w_{ij,normal} - w_{ij,outliers}|\} \\
 RRC_{2,i} &= ||w_{i,normal} - w_{i,outliers}||_1,
 \end{aligned} \tag{12.13}$$

where $w_{i,normal} = [w_{ij}]_{normal}$ and $w_{i,outliers} = [w_{ij}]_{outliers}$ are the normalized coefficient vectors with normal samples and adding outliers samples for the i_{th} PC, respectively.

RRC_1 represents the maximum absolute deviation of the two coefficient sets, which indicates the worst changes of the normalized w_{ij} . RRC_2 represents the sum of the absolute deviations of the two coefficient sets, which indicates the overall change of the normalized w_{ij} .

The normalized coefficient w_{ij} values of the first two PCs (t_1 and t_2) of the PLS, RPLS and L_1 -PLS methods are shown in Figs. 12.3 and 12.4. The corresponding indices $RRC_i, i = 1, 2$ are given in Table 12.1 (a smaller value is better).

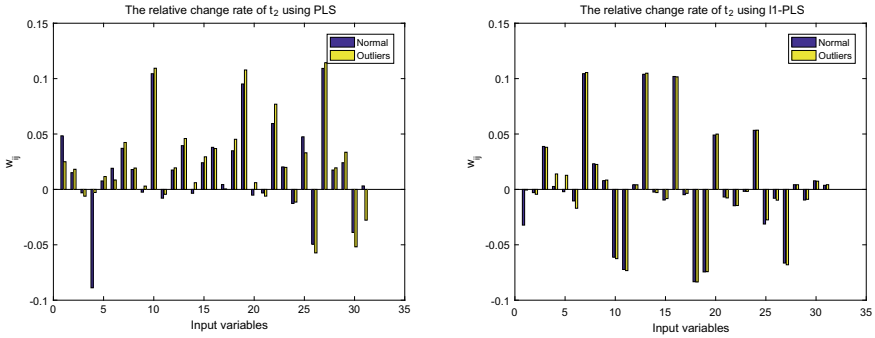


Fig. 12.4 The relative change rates of t_2 using PLS and L_1 -PLS

Table 12.1 RRC_i of t_1 and t_2 of the PLS, L_1 -PLS and L_1 -SPLS methods

	PLS		RPLS		L_1 -PLS	
	t_1	t_2	t_1	t_2	t_1	t_2
RRC_1	0.919	0.349	0.005	0.040	0.110	0.121
RRC_2	7.093	1.247	0.016	0.130	0.368	0.329

It can be seen from Figs. 12.3–12.4 and Table 12.1 that no matter which method is used, the outliers will always affect the structure of the PCs to some extent. In general, the outliers have a large adverse effect on the PCs extraction of the PLS method, and thus results in the largest change in its PC structures. With the robust covariance estimation method, the outliers have little effect on the PCs extraction of the RPLS method. L_1 -PLS method only uses the L_1 norm to be insensitive to outliers, without any outliers processing. Outliers that cause changes in the structure of its two PCs are nearly identical and within an acceptable range, whether in the RRC_1 or RRC_2 . The samples considered to be outliers may be a true reflection of the system state when the data set follows a heavy-tailed distribution (Doman’ski 2019). It is more important to retain all the samples to extract the PCs, although the outliers have a certain influence on the direction vectors.

By further analyzing the structure of t_1 and t_2 , it can be easily found that the extracted PCs by those methods are quite different. In order to better explain the structural differences of t_1 and t_2 in different methods, IDV(14) is taken as an example for in-depth analysis. The typical process variable monitoring results of IDV(14) are given in Fig. 12.5, in which, x_9 , x_{21} and x_{30} have similar monitoring results. Among the t_1 and t_2 , the sum of the absolute weights for x_9 , x_{21} and x_{30} of the PLS method (0.062) is more than twice that of the L_1 -PLS method (0.025).

These weight differences do not significantly affect the output prediction and the monitoring performance in the normal operation. But these differences are amplified in the fault modes. For example, consider the monitoring under the fault modes IDV(14) and IDV(17). The role of x_{21} and x_{30} (especially x_{30}) in the PLS method is

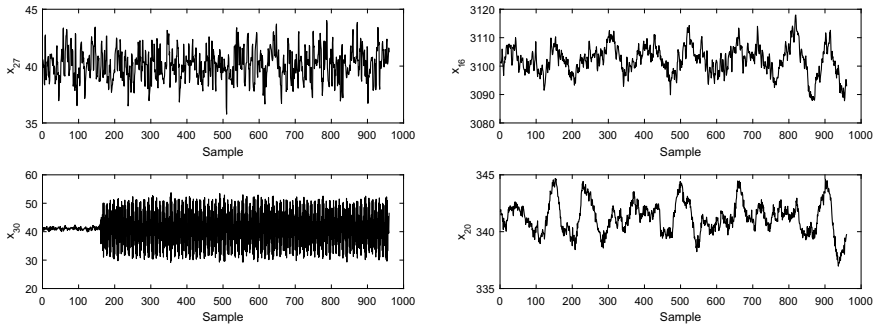


Fig. 12.5 Typical process variable monitoring results of IDV(14)

exaggerated, leading to incorrect predictions and quality-relevant monitoring results (see Figs. 12.6 and 12.7). Correspondingly, the L_1 norm can better maintain the relative size of those variables, therefore, the role of x_{21} and x_{30} in the extracted PCs is not exaggerated. In other words, the extracted PCs by the L_1 norm better capture the relationship between the input space and output space.

12.5.2 Robustness of Prediction and Monitoring Performance

The robustness of the principal components of the L_1 -PLS method is discussed in the previous section. But the number of principal components of the three methods is different, which only reflects one aspect of the robustness. Now, the robustness of prediction performance and monitoring is analyzed further, especially the prediction performance directly reflects the quality of the model. There are 21 types of faults in the TE process. The fault IDV(21) is a fault that the output drifts slowly, caused by the constant change of the steam valve position. So it does not reflect the robustness of the model. Therefore, the first 20 faults are analyzed in this simulation experiment. In this simulation, the sparsity in the L_1 -SPLS model is determined by the VIP method: input space $s_1 = 14$, output space $s_2 = 2$.

Experiment 1: Prediction Performance Analysis

In this experiment, the L_1 -PLS model shows good output prediction results for the 20 fault data sets. L_1 -PLS(outliers) and PLS(outliers) mean that the two models are trained by the normal operation data with adding outliers, described in previous Sect. 12.5.1. In order to illustrate the above conclusions more clearly, four faults IDV(7), IDV(14), IDV(17), and IDV(18) are selected to compare the prediction performance of the PLS model and the L_1 -PLS model. The output prediction results are good for all fault modes, but the four faults come from four different fault types, and the results of the L_1 -PLS model and the PLS model are quite different. Figures 12.6 and 12.7 give the output prediction results of the fault IDV(7), IDV(14), IDV(17),

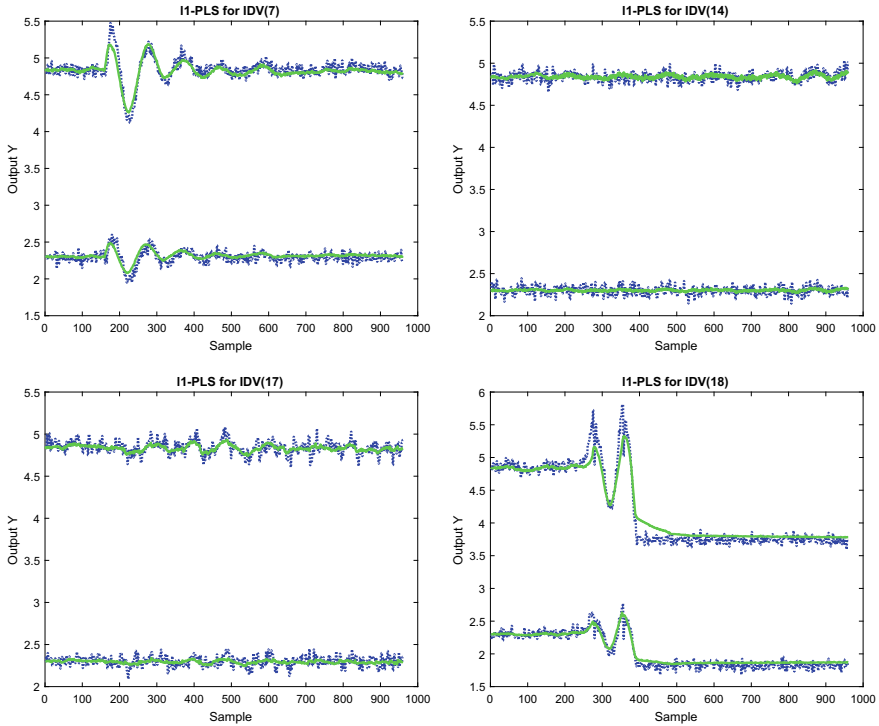


Fig. 12.6 Output predicted values for IDV(7), IDV(14), IDV(17), and IDV(18) using PLS(outliers)

and IDV(18). The horizontal axis represents data samples, and the vertical axis represents output values. The blue dashed line is the actual output value, and the green is the predicted output value.

In these prediction and monitoring diagrams, the first 160 samples are normal data, and the last 800 samples are data under different fault modes. The output prediction of fault IDV(7) shows a consistent conclusion under the step-change fault. The feedback controller or cascade controller reduces the impact of faults and abnormal values on product quality. For the other three types of fault IDV(14), IDV(17), and IDV(18), there are some differences in their output prediction results. When the system is under the normal operation, the PLS and L_1 -PLS models have the same good prediction results. However, after adding outliers, the PLS method cannot accurately predict the output (Fig. 12.6), while the L_1 -PLS method still quickly detects the output changes and makes correct predictions (Fig. 12.7). In particular, for faults IDV(17) and IDV(18), the PLS method gives a serious wrong predictions. Experiments show that the prediction performance of the L_1 -PLS method is better than PLS. Even if the data is contaminated by outliers, L_1 -PLS can still predict the output accurately. In other words, the L_1 -PLS model has stronger robust prediction performance.

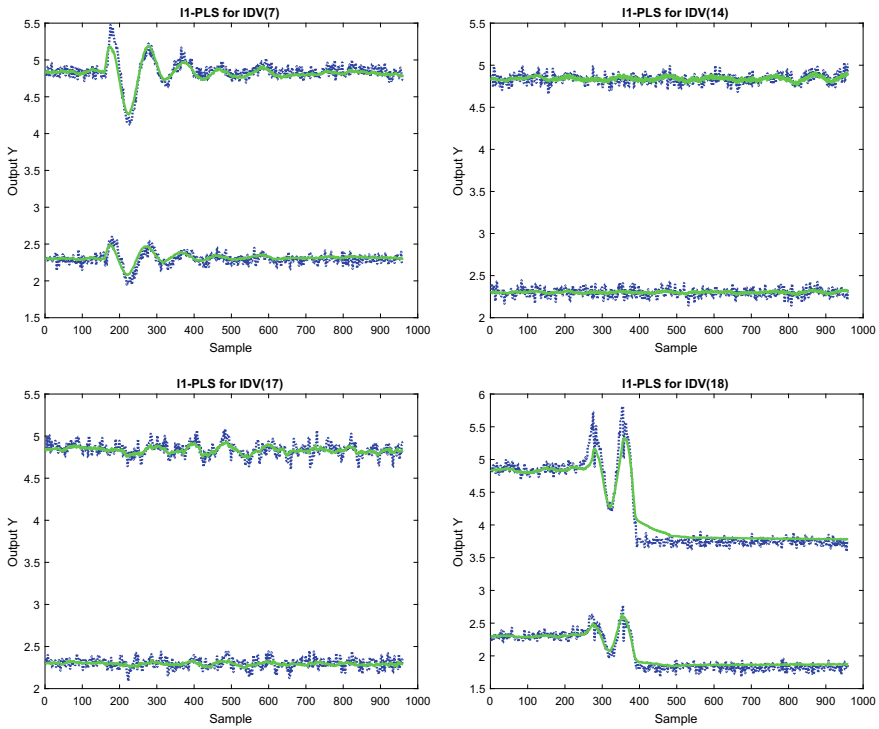


Fig. 12.7 Output predicted values for IDV(7), IDV(14), IDV(17), and IDV(18) using L_1 -PLS(outliers)

Experiment 2: Monitoring Performance Analysis

The robustness of monitoring performance is mainly verified by the accuracy of fault detection. The detection indices are FDR and FAR (4.1), the control limit is calculated with the confidence level 99.75% for both PLS and L_1 -PLS methods. The FAR results of the two models are basically same, this indicates that the proposed L_1 -PLS method does not increase the risk of false alarms, so it is not analyzed in this section. Table 12.2 lists the FDR results of the first 20 faults without adding outliers, corresponding to the models PLS, L_1 -PLS and L_1 -SPLS respectively. Table 12.3 shows the FDR results of 20 faults after adding outliers, corresponding to the models PLS (outliers), L_1 -PLS (outliers), and L_1 -SPLS (outliers).

For serious quality-related faults IDV(2), IDV(6), IDV(8), IDV(12), IDV(13), and IDV(18), the six models give consistent results. Therefore, these faults are not analyzed in this chapter. For other types of faults, their results are very different, including the quality-irrelevant faults, the quality-recoverable faults, and slight quality-related faults. The detailed analysis of the three situations is given below. In the monitoring figures of this section, the blue line represents the value of the statistic, where the

Table 12.2 FDRs of PLS, L₁-PLS, and L1-SPLS

IDV	PLS		L ₁ -PLS		L1-SPLS	
	T ²	T _e ²	T ²	T _e ²	T ²	T _e ²
1	99.63	99.75	60.00	99.75	31.38	99.75
2	98.50	98.25	98.25	98.38	98.25	98.38
3	1.00	1.38	0.75	1.75	0.50	1.75
4	19.13	100.00	0.88	100.00	0.00	100.00
5	22.00	100.00	18.38	100.00	17.13	100.00
6	99.25	100.00	98.38	100.00	98.13	100.00
7	100.00	100.00	68.75	100.00	31.38	100.00
8	96.00	97.88	89.00	97.88	88.50	97.88
9	0.50	1.13	0.25	1.38	0.38	1.38
10	26.38	84.25	19.13	85.38	15.63	85.38
11	26.63	76.50	1.13	77.88	0.88	77.88
12	97.50	99.88	84.00	99.88	84.00	99.88
13	94.88	95.13	82.13	95.25	82.25	95.25
14	91.50	100.00	0.38	100.00	0.00	100.00
15	1.25	2.63	1.00	3.75	0.63	3.75
16	20.13	42.75	9.00	46.13	7.00	46.13
17	77.38	96.75	10.00	97.00	1.63	97.00
18	89.38	90.13	88.75	90.13	88.75	90.13
19	0.50	34.50	0.13	37.88	0.00	37.88
20	30.50	90.50	20.75	90.38	19.00	90.38

upper curve is T², and the lower is T_e². The system alarms if the blue line exceeds the red control limit.

Case 1: Quality Irrelevant Fault

It can be found from Table 12.2 that very low alarm values are given for faults IDV(3), IDV(9), IDV(15), and IDV(19). However, the alarm values of the L₁-PLS and L₁-SPLS models are lower, which indicates that fewer false alarms will occur during the monitoring. It can also be seen from the corresponding Figs. 12.8, 12.9, 12.10, 12.11, and 12.12, the alarm points of the latter two models are much less. For faults IDV(4), IDV(11), and IDV(14), they are all related to the reactor cooling water and hardly affect the quality of output products. The PLS model gives a higher alarm value, which may lead to serious false alarms, while the L₁-PLS model effectively avoids these alarms and reduces the number of false alarms. In addition, the L₁-PLS model eliminates most of the false alarms in the monitoring Figs. 12.8, 12.9, 12.10, and the L1-SPLS model almost eliminates all false alarms.

When adding outliers, the PLS model provides the same wrong results for quality-irrelevant faults. The specific FDR values are shown in Table 12.3. However, the

Table 12.3 FDRs of PLS(outliers), L₁-PLS(outliers), and L₁-SPLS(outliers)

IDV	PLS(outliers)		L ₁ -PLS(outliers)		L ₁ -SPLS(outliers)	
	T ²	T _e ²	T ²	T _e ²	T ²	T _e ²
1	99.88	99.75	28.38	99.75	36.63	99.75
2	98.63	98.25	98.00	98.25	98.25	98.25
3	3.25	0.88	0.13	1.13	0.63	1.13
4	7.63	100.00	0.25	100.00	0.00	100.00
5	24.88	27.88	14.75	28.38	16.88	28.38
6	99.75	100.00	98.38	100.00	98.25	100.00
7	100.00	100.00	59.88	100.00	29.50	100.00
8	96.50	97.75	84.50	97.88	88.00	97.88
9	0.88	0.88	0.00	1.00	0.38	1.00
10	37.50	77.63	11.00	80.50	15.25	80.50
11	16.00	73.75	0.50	74.75	0.88	74.75
12	95.88	99.25	78.50	99.25	83.63	99.25
13	95.50	95.00	80.25	95.25	82.00	95.25
14	89.75	100.00	0.00	100.00	0.00	100.00
15	4.38	0.50	0.13	0.88	0.75	0.88
16	33.88	28.38	4.50	35.13	6.25	35.13
17	76.88	96.63	6.13	96.63	1.50	96.63
18	90.00	89.88	88.00	89.88	88.63	89.88
19	1.13	28.38	0.00	30.00	0.00	30.00
20	36.50	77.13	15.63	77.75	19.75	77.75

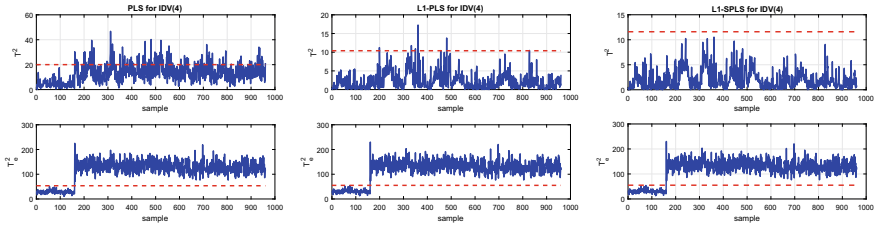


Fig. 12.8 PLS, L₁-PLS and L₁-SPLS monitoring results for IDV(4)

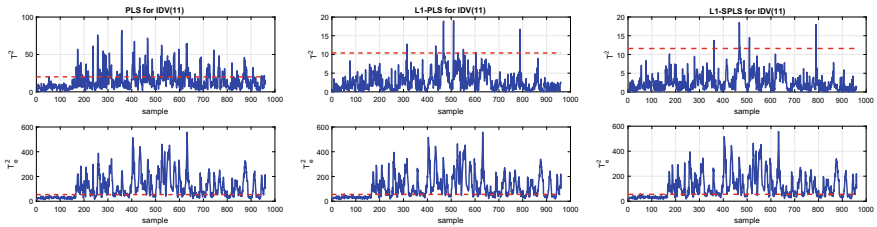


Fig. 12.9 PLS, L₁-PLS and L₁-SPLS monitoring results for IDV(11)

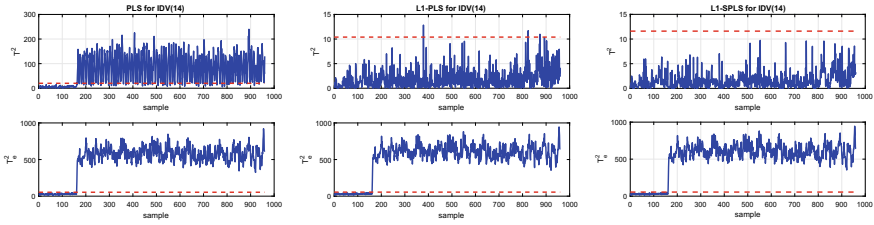


Fig. 12.10 PLS, L_1 -PLS and L_1 -SPLS monitoring results for IDV(14)

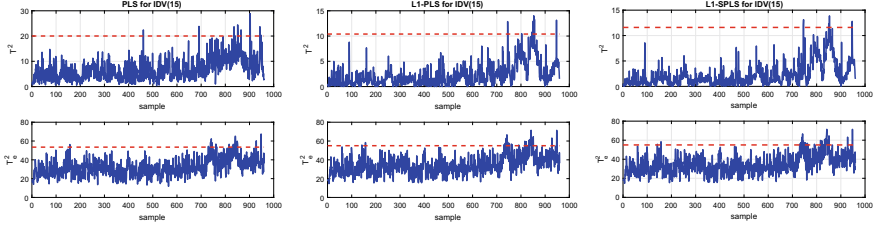


Fig. 12.11 PLS, L_1 -PLS and L_1 -SPLS monitoring results for IDV(15)

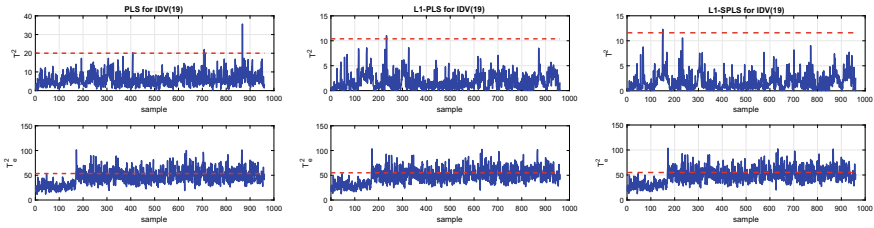


Fig. 12.12 PLS, L_1 -PLS and L_1 -SPLS monitoring results for IDV(19)

monitoring effect of the L_1 -PLS model is still very good, for fault IDV(9), IDV(14), and IDV(19). The detection rate has been reduced to 0, which means that false alarms are completely eliminated in these cases. Therefore, the L_1 -(S)PLS model will not interfere with the fault monitoring results after adding outliers. It should be noted that the monitoring performance of the L_1 -PLS model after adding outliers (Table 12.3) is better than the normal conditions (Table 12.2). The possible reason is outliers, and the total noise in the input data becomes larger. The L_1 -PLS method can filter out noise more effectively during the modeling. Therefore, the established model is more accurate and the monitoring performance is improved.

Case 2: Quality-Recoverable Fault

Faults IDV(1), IDV(5), and IDV(7) are quality-recoverable faults. The prediction value should tend to return to normal, but the statistic should be kept at a higher value. Figure 12.13 shows the monitoring results of the three models on the fault IDV(1). It can be seen that both the L_1 -PLS and L_1 -SPLS model methods give the

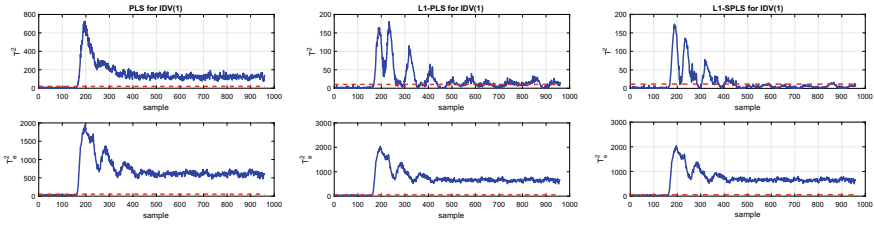


Fig. 12.13 PLS, L_1 -PLS and L_1 -SPLS monitoring results for IDV(1)

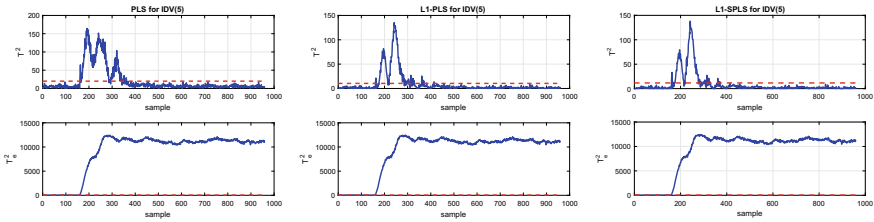


Fig. 12.14 PLS, L_1 -PLS and L_1 -SPLS monitoring results for IDV(5)

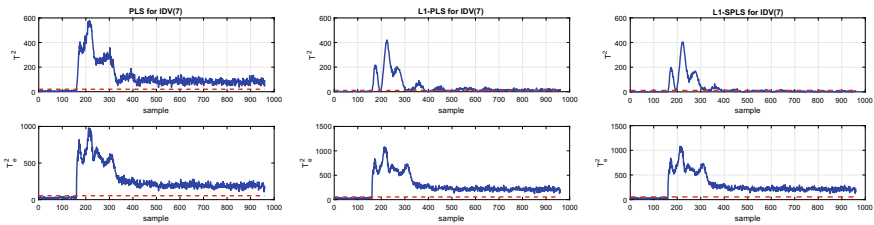


Fig. 12.15 PLS, L_1 -PLS and L_1 -SPLS monitoring results for IDV(7)

correct alarm results. In the PLS model, the value of the statistic exceeds the control limit, so a false alarm is generated in the process monitoring. For the fault IDV(5), it is also a process-related fault. It can be seen from Tables 12.2 and 12.3 that the fault detection rates of the L_1 -PLS and L_1 -SPLS models are lower than the PLS model, which means that the monitoring results are more accurate. Figures 12.14 and 12.16, respectively, show the monitoring diagrams of the three models for the fault IDV(5) in the normal case (without adding outliers) and with adding outliers. For fault IDV(7), the corresponding monitoring results are shown in Fig. 12.15. The PLS model gives completely wrong result, while the results of the other two models are more accurate.

The detection result for fault IDV(1) obtained by the L_1 -PLS (outliers) model seems to be better than the L_1 -PLS model, and the monitoring results are more reasonable. In addition, for the fault IDV (5), although the monitoring results of the L_1 -PLS and L_1 -SPLS (outliers) models may not be ideal, as shown in Fig. 12.14. The T_e^2 statistics of the L_1 -PLS and L_1 -SPLS models can detect the input space

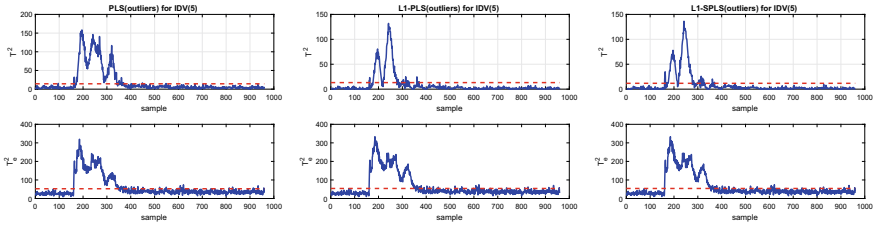


Fig. 12.16 PLS(outliers), L₁-PLS(outliers) and L₁-SPLS(outliers) monitoring results for IDV(5)

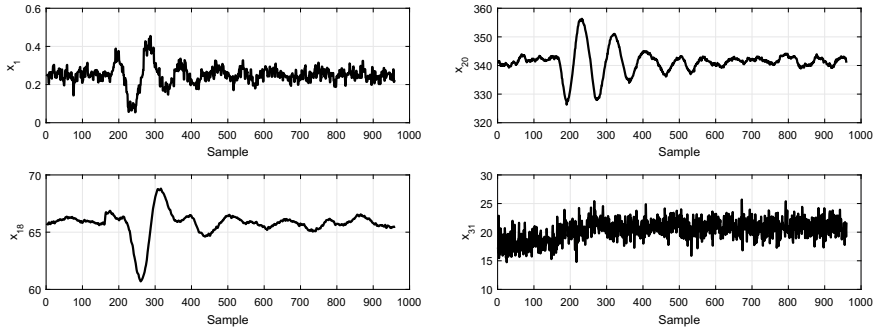


Fig. 12.17 Typical process variable monitoring results of IDV(5)

process-related faults. But the PLS (outliers), L₁-PLS (outliers) and L₁-SPLS (outliers) models gave wrong results (Fig. 12.16).

There are two possible reasons for this phenomenon. Firstly, the outliers were added directly without being regulated by the dynamic system, so its influence on the extraction of the principal components cannot be determined directly. Secondly, the typical process dynamics corresponding to fault IDV(5) is shown in Fig. 12.17. Only the variable 31 is a step change in all the monitored variables, and the rest gradually returns to the normal under the action of controller. In terms of the composition of the principal components, the contribution of variable 31 to the principal components is small. Therefore, its role is more in the residual space in the normal case (without adding outliers). After the outlier is added, its contribution to the principal component increases, which means its role in the residual space is weakened. It in turn causes the monitoring indicators in the residual space to return back to normal. On the other hand, the percentage of its contribution to the principal component is still small, so the monitoring indicators on the principal metric space also do not significantly reflect its characteristics.

Case 2: Slight Quality Related Fault

Fault IDV (16) and IDV (17) have a slight impact on quality, which means that they have almost no impact on output quality. Figure 12.18 shows the monitoring results of the three models after adding outliers. The fault monitoring results of PLS (outliers)

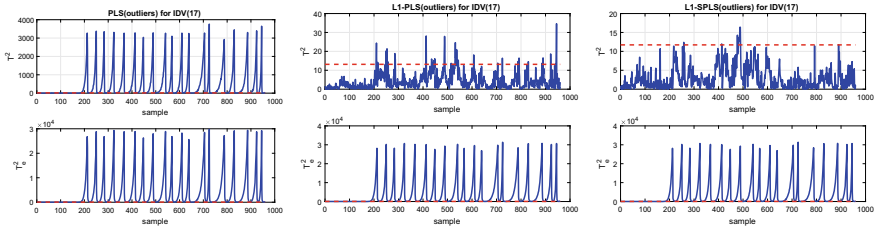


Fig. 12.18 PLS(outliers), L_1 -PLS(outliers), and L_1 -SPLS(outliers) monitoring results for IDV(17)

model is very bad, there have been many false positives. The L_1 -PLS (outliers) model and L_1 -SPLS (outliers) model effectively reduce these false alarms. It can also be seen from the corresponding FDR that the monitoring results of the L_1 -PLS (outliers) model and the L_1 -SPLS (outliers) model are more reasonable.

It can be seen from the above comparison results that even if outliers are added to the input data, the monitoring results of the L_1 -(S)PLS model have also been greatly improved. In other words, the L_1 -(S)PLS model improves the robustness performance and fault detection performance.

12.6 Conclusions

This chapter proposes a quality-related statistical monitoring method of double robust projection to latent structure (L_1 -PLS), which enhances the robustness of the PLS algorithm from two aspects. On the one hand, the L_1 -PLS method replaces the L_2 norm in the objective function with the L_1 norm, and adds the L_1 norm penalty term to the direction vector; On the other hand, the regression coefficient of the L_1 -PLS algorithm can also be obtained by the L_1 norm. Therefore, the L_1 -PLS algorithm has double robustness. Then a monitoring model based on the L_1 -PLS method is established, the robust performance and monitoring performance are verified on the TE process simulation platform. The results show that the L_1 -PLS method has better robustness and better performance in process monitoring and fault diagnosis.

References

- Branden KV, Hubert M (2004) Robustness properties of a robust partial least squares regression method. *Anal Chim Acta* 515:229–241
- Doman'ski PD (2019) Control quality assessment using fractal persistence measures. *ISA Transactions*
- Farrés M, Platikanov S, Tsakovski S, Tauler R (2015) Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. *J Chemom* 29(10):528–536

- Filzmoser P (2016) Identification of multivariate outliers: a performance study. *Austrian J Stat* 34:127–138
- Fortuna L, Graziani S, Rizzo A, Xibilia MG (2007) *Soft sensors for monitoring and control of industrial processes*. Springer, London
- Hubert M, Rousseeuw PJ, Van AS (2008) High-breakdown robust multivariate methods. *Stat Sci* 23:92–119
- Kwak N (2008) Principal component analysis based on l_1 -norm maximization. *IEEE Trans Pattern Anal Mach Intell* 30(9):1672–1680
- Liu JL (2014) Developing a soft sensor based on sparse partial least squares with variable selection. *J Process Control* 24:1046–1056
- Liu K, Chen YQ, Domański PD, Zhang X (2018) A novel method for control performance assessment with fractional order signal processing and its application to semiconductor manufacturing. *Algorithms* 11(7)
- Meng D, Zhao Q, Xu Z (2012) Improve robustness of sparse PCA by L_1 -norm maximization. *Pattern Recogn* 45:487–497
- Turkmen AS, Billor N (2013) Influence function analysis for the robust partial least squares (RoPLS) estimator. *Commun Stat-Theory Methods* 42:2818–2836
- Wang H (2000) *Bounded dynamic stochastic systems: modeling and control*. Springer, London
- Wang J, Zhong B, Zhou J (2017) Quality-relevant fault monitoring based on locality preserving partial least squares statistical models. *Ind Eng Chem Res* 56:7009–7020
- Wang Y, Karimi HR, Shen H, Fang Z, Liu M (2019) Fuzzy-model-based sliding mode control of nonlinear descriptor systems. *IEEE Trans Cybern* 49(9):3409–3419
- Zhou JL, Zhang SL, Zhang H, Wang J (2018) A quality-related statistical process monitoring method based on global plus local projection to latent structures. *Ind Eng Chem Res* 57:5323–5337
- Zou H, Hastie T, Tibshirani R (2006) Sparse principal component analysis. *J Comput Graph Stat* 15:265–286

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

