Robust Multi-Scale Strategy for Multivariate Process Monitoring

D. Wang, Y.Y. Joe* and J. A. Romagnoli
Dept. of Chemical Engineering, The University of Sydney, NSW 2006, Australia
*Dept. of Electrical & Computer Engineering, The National University of Singapore, 10 Kent Ridge Cres., Singapore 119260

Abstract

An improved multi-scale principal component analysis method for process monitoring is proposed. The novelty of the approach is to integrate multi-scale PCA with the robustness to the typical normality assumption of noisy data. This method, using an M-estimator based on the generalized T distribution, adaptively transforms the data in the score space at each scale in order to eliminate the effects of the outliers in the original data. The robust estimation of the covariance or correlation matrix at each scale is obtained by the proposed approach so that the accurate MSPCA model can be obtained for the process monitoring purposes. The performance of the proposed approach in process monitoring is illustrated and compared with the conventional MSPCA approach through chemical engineering example.

Keywords: PCA, Wavelets, Robust estimation, Fault detection, Process monitoring.

1. Introduction

Data-driven process monitoring based on multivariate statistic techniques is widely used in chemical industries. The measurement variables are usually highly correlated and the real dimensionality of the process variables is considerably less than that represented by the number of process variables collected. Monitoring this “data-rich” process inevitably needs dimensionality reduction techniques to grasp the driven force embedded in these measurements. By converting the large amount of data collected from the process into a few meaningful measures, one can assist the industrial operators in determining the status of the operations and in detecting and diagnosing the faults. Principal components analysis (PCA) is such a dimensionality reduction technique and it is heavily used in modeling the multivariate process for monitoring purpose (Kresta, et. al., 1991).

The recent extension of PCA is Multi-scale PCA, which combines the ability of PCA to decorrelate the variables by extracting a linear relationship with that of wavelet analysis to extract deterministic features and approximately decorrelate auto-correlated measurements (Bakshi, 1998). Despite the advantage of separating the noise from the deterministic signal, there are still unsolved issues within the application of multi-scale PCA approach. One key aspect is the robustness of the approach when dealing with real data. As we know, either in conventional or MSPCA, it is usually assumed that the data, in single or multi-scales, are normally distributed. Practical examination tells us that the real plant data seldom satisfies
to this crucial assumption. The data are usually unpredictable, having heavier tails than the normal ones, especially when they contain anomalous outliers. This will inevitably result in the loss of performance leading, in some cases, to wrong modeling, which will, in turn, lead to missed detections under faulty conditions or false alarms under normal operation. Preprocessing of data is not a good idea, since it may destroy the correlation structure of the multivariate data and result in the loss of information.

In this paper, we will integrate the robust consideration into MSPCA with the aim to improve MSPCA considering the outliers in the training data. In this method, after the wavelet decomposition of the data, the PCA at each scale is modified to eliminate the effects of outliers in the data, considering the fact that the outliers will manifest themselves at different scales. A winsorization procedure is employed in the score space at each scale using robust estimation theory. The robust estimator used in this work is based on the Generalized T distribution and adaptively transforms the data in the score space (at a particular scale) to eliminate the effects of the outliers in the original data. A robust estimation of the covariance or correlation matrix is obtained by the proposed approach so that the accurate multi-scale PCA models can be obtained for the purpose of process monitoring. It will be shown that the proposed approach leads to a flexible strategy and is able to adjust to different data characteristics/distributions.

2. Multi-scale Principal Component Analysis

2.1 Principal Component Analysis (PCA)

PCA decomposes the observation \( X (n \times m) \) as

\[
X = TP^T = \sum_{i=1}^{m} t_i p_i^T
\]

where \( p_i \) and \( t_i \) can be calculated by finding the eigenvalues and their companion eigenvectors of covariance or correlation matrix \( S \) of data \( X \),

\[
S = P \Lambda P^T
\]

\[
T = XP
\]

where: \( \Lambda \) is the diagonal matrix containing the ordered eigenvalues of \( S \) and \( P \) is the corresponding eigenvector matrix. In general, if the process variables are collinear, the first \( k \) principal components can be used to explain sufficiently the variability in the whole data set with less information loss, and the determination of the number \( k \) can be obtained via several techniques such as screen test and cross-validation. It then follows that

\[
X = T_k P_k^T + E = \sum_{i=1}^{k} t_i p_i^T + E
\]

\[
\hat{X} = T_k P_k^T = \sum_{i=1}^{k} t_i p_i^T
\]

2.2 Wavelets and Multi-resolution Analysis

Wavelet transform provides a signal analysis tool by providing a mapping from the time domain to the time-scale domain. Using wavelets, the original signal can be decomposed into its contributions in different regions of the time-scale space by projection on the
corresponding wavelet basis functions. According to the theory of multi-resolution analysis (Mallat, 1989), any signal \( f(t) \in L^2(R) \) can be approximated by successively projecting it down onto scaling functions and wavelets functions. \( \Phi_{m,n}(t) \), \( \Psi_{m,n}(t) \) are translated and dilated versions of the scaling function \( \Phi(t) \) and the wavelet function \( \Psi(t) \), respectively,

\[
\Phi_{m,n}(t) = 2^{-m/2} \Phi(2^{-m} t - n) \\
\Psi_{m,n}(t) = 2^{-m/2} \Psi(2^{-m} t - n)
\]  

(5)

The projections onto the scaling function are known as approximation coefficients, which are smoother versions of the original signal and the degree of smoothness increases as the scale increases. The projections onto the wavelet functions are known as the wavelets coefficients, which capture the details of the signal lost when moving from an approximation at one scale to the next coarser scale. The coefficients can be used to represent the contributions of signal at each scale.

2.3 Multi-scale Principal Component Analysis (MSPCA)

The MSPCA methodology consists of decomposing each variable on a selected family of wavelets. The PCA model is then determined independently for the coefficients at each scale. The models at important scales are then combined in an efficient scale-recursive manner to yield the multi-scale model. For multivariate process monitoring by MSPCA, the region of normal operation is determined at each scale from data representing normal operation. For new data, the important scales are determined as those where the current coefficient violates the detection limits. The actual state of the process is confirmed by checking whether the signal reconstructed from the selected coefficients violates the detection limits of the PCA model for the significant scales. This approach is equivalent to adaptively filtering each value of scores and residuals by a filter of dyadic length that is best suited for separating the deterministic change from the normal process variation. The detection limits for the scores and residuals also adapt to the nature of the signal. Furthermore, retaining only those coefficients that violate the detection limits at each scale integrates the task of process monitoring and extraction of features relevant to abnormal operation, without any prefiltering.

3. Robust PCA Based on M-estimate Winsorization

After implementing PCA, the outliers presented in the original data can manifest themselves in the score space. By recurrently winsorizing the scores and replacing them with suitable values, it is possible to detect multivariate outliers and replace them by values which conform to the correlation structure in the data. A more accurate PCA model can result from the reconstructed data. The concept of winsorization is briefly explained first and its application to robust PCA is then investigated.

3.1 Winsorization

Consider the linear regression problem

\[
y = f(X, \theta) + \varepsilon
\]  

(6)
where: $y = (y_1, y_2, ..., y_n)'$ is a $n \times 1$ vector of dependent variables, $X = (x_1, x_2, ..., x_n)'$ is a $n \times m$ matrix of independent variables, and $\theta$ is a $p \times 1$ vector of parameters, $\varepsilon$ is a $n \times 1$ vector of model error or residual. An estimation of parameter $\theta$ ($\hat{\theta}$) can be obtained by minimizing the function

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^{n} \rho \left( \frac{y_i - f_i(x_i, \theta)}{s} \right)$$

(7)

where: $s$ is an estimation of the scale of the distribution of residuals and $\rho$ is objective function to be minimized.

With the parameter $\hat{\theta}$ estimated, the prediction or estimation of the dependent variable $y_i$ ($i = 1, ..., n$) is given by

$$\hat{y}_i = f_i(x_i, \hat{\theta})$$

(8)

and the residual is given by

$$r_i = y_i - \hat{y}_i$$

(9)

In the winsorization process, the variable $y_i$ is transformed using pseudo observation according to specified M-estimates, which characterizes the residual distribution (Huber, 1981). The normal assumption of residual data will result in poor performance of winsorization. In this work, we fit the residual data with a more flexible distribution, i.e. the generalized T distribution, which can accommodate the shapes of most distributions one meets in practice, and then winsorize the variable $y_i$ using its corresponding influence function.

### 3.2 Robust PCA Based on GT Winsorization

The proposed robust estimator for PCA modeling in this work is based on the assumption that the data in the score space follow the generalized T distribution (GT) (Butler, et. al., 1990), which has the flexibility to accommodate various distributional shapes:

$$f_{GT}(u; \sigma, p, q) = \frac{1}{\sigma q^p B(q/p, q)} \left( \frac{1 + \left( \frac{|u|}{\sigma q^p} \right)^{q/p}}{q^p} \right)^{-q/p}$$

(10)

where: $\sigma$, $p$, $q$ are distributional parameters, $\sigma$ corresponds to the standard deviation, $p$ and $q$ are parameters corresponding to the shape of distribution. This density is symmetric about zero, uni-modal, and suitable to describe the error characteristics in most cases. By choosing different values of $p$ and $q$, the GT distribution will accommodate the real shape of the error distribution. The tail behavior and other characteristics of the distribution depend upon these two distributional parameters, which will be estimated from the data (Wang et. al., 2003).

The robustness of the estimator based on a GT distribution can be discussed by investigating its $\psi$ or influence function. This $\psi$ function, corresponding to the objective function $\rho(u; \sigma, p, q) = -\log f_{GT}(u; \sigma, p, q)$ is given by...
The technique of winsorization can be used in PCA to eliminate the effects of outliers in the following way. The data value $y$ in score space can be transformed into a new value $y^{w}$ by winsorization. By doing this, the large values exhibited as outliers in the original data set will be brought closer to the other observations after they are transformed from the score space back to the original data space. A new PCA model is obtained using the new data set. This process is carried out iteratively until there is not much change in the loading vectors.

4. Robust MPCA and Process Monitoring

4.1 Robust MPCA

The proposed approach in this work is to integrate robust PCA based on GT distribution into multi-scale PCA method with the aim of eliminating the effects of outliers in the training data and obtaining more accurate PCA model for process monitoring. Process monitoring with a less accurate PCA model will result in poor performance and can increase the rate of false alarms or misdetection. The improvement can be obtained by implementing robust method to PCA at each scale, which results in a RMSPCA procedure:

- Compute wavelet decomposition for each column in training data matrix $X$.
- At each scale, compute PCA loading and score for the data matrix composed of wavelet coefficients.
- At each scale, apply robust PCA procedures based on GT winsorization described in section 3.2.
- Determine the PCA model at each scale by selecting the numbers of loadings retained.
- Select scores and wavelet coefficients which are larger than appropriate threshold at each scale.
- Calculate PCA for all the scales with the covariance matrix obtained by combining the covariance matrix of the scales at which the coefficients or scores violate the threshold.

Compared with MPCA, RMAPCA includes a robust method to PCA at each scale so that more accurate model will result when the assumption of normality is not satisfied in the data set.

4.2 Statistical Process Monitoring by RMSPCA

Given the RMSPCA model, the principal component loadings and detection limits (e.g. 95% or 99%) for the scores and residuals are computed from data representing normal operation. The moving window is used for on-line monitoring. The new data are decomposed by the same procedures as in the modeling phases and the process monitoring in the score space is achieved by computing the $T^2$ value as the sum of the squares of the selected scores scaled by the respective eigenvalue computed from the data representing the normal operation, i.e.:

$$T^2 = \sum_{j=1}^{a} \frac{T^2_j}{\lambda_j}$$  \hspace{1cm} (12)
where: $T_i^2$ is the $T^2$ value for the $i$th row of measurements, $a$ is the PC number retained and $\lambda_j$ is the eigenvalue of the $j$th score. The score for the new data at each scale is compared with the detection limit at each scale and the contribution at the scale is taken into account if the violation happens. The data is reconstructed by those contributions and the overall score of the new data is obtained by combining all the score contributions when the limits are violated. Given the overall detection limit the detection limits at each scale for on-line process monitoring need to be adjusted to account for the over-completeness of the on-line wavelet decomposition. This over-completeness can be explained as follows. For example, if 99% limit is given for the original data, and if the data are decomposed into four detail signals and one scaled signal by wavelets, application of the 99% limit at each scale will result in an effective confidence of only 95% for the reconstructed data, since the decomposed part violating the detection limits at each scale need not be at the same location. This requires that the detection limit at each scale be increased to maintain the detection confidence limit for the reconstructed signal. The adjusted limit can be calculated as follows (Bakshi, 1998):

$$C_L = 100 - \frac{1}{L+1}(100 - C) \quad (13)$$

5. Case study

The proposed RMSPCA strategy is applied to a simulated case study of a pilot-scale setting containing two CSTRs, a mixer and a number of heat exchangers (Figure 1). The total number of measurement variables in this study is 25, including 14 temperature variables, 7 flowrate variables and 4 level variables. Training data are generated when the plant is running in normal operation, and then they are contaminated with $t$ distributed noise and outliers (Figure 2). The calculation results show that different PCA models are obtained by the different approaches. For the same PC number retained, the different degrees of variance are explained in the two approaches. Further investigations show that even for the case where the same retained PC number explains the similar variance in both approaches, the detection limit at each scale is different, with the bounds in MSPCA being usually increased.
larger than that in RMSPCA. This is due to the effects of outliers. In MSPCA, for the given $C_L$ percentage at a scale, the bound is increased due to the larger values which are the manifestation of the outliers, whereas in RMSPCA, these values are squeezed via the winsorization process. This will result in the fault detection by RMSPCA being more sensitive than that of MSPCA. Figure 3 shows the on-line detection results of both RMSPCA and MSPCA, where the validation data is composed of 286 samples with first 100 being normal data, the next 100 being from the training data and the rest being normal data. It can be seen that RMSPCA clearly claims that faults exist in the process, while MSPCA is not able to detect the malfunction. Figure 4 shows the detection results when there are 30% mean shift in the validation data. RMSPCA can detect the process upset, while MSPCA is ambiguous.

6. Conclusions

A robust multi-scale PCA modeling method based on winsorization in score space using adaptive robust estimator was developed and presented. The effects of outliers in the data can be eliminated by the method while the effectiveness as well as the robustness is retained by using GT-like estimator. The proposed method can obtain a more accurate model when the normal assumption is interrupted in the training data. Monitoring by RMSPCA has better performance than that of the conventional MSPCA. The performance of the proposed method is validated using a chemical engineering example. The use of the approach in process monitoring is promising.

References
