FAULT DETECTION AND ISOLATION USING CORRESPONDENCE ANALYSIS

Detroja K. P.[1]**, Patwardhan S. C.**[2]**, Gudi R. D.**[2*]

[1] *Systems and Control Engineering*
[2] *Associate Professor, Department of Chemical Engineering*
Indian Institute of Technology, Bombay, Powai, Mumbai - 400076, India
[*] *E-mail : ravigudi@che.iitb.ac.in*

Abstract: In this paper, a new approach to Fault Detection and Diagnosis that is based on Correspondence Analysis is proposed. Correspondence analysis is a powerful multivariate technique based on the generalized singular value decomposition. The advantage of using correspondence analysis is that it depicts rows as well as columns as points in the dual lower dimensional vector space. Correspondence analysis has been shown to capture association between various features and events quite effectively. In this paper, the correspondence analysis approach is used for Fault Detection and Diagnosis (FDD) and is validated on representative process systems.

Keywords: Correspondence Analysis, Generalized Singular Value Decomposition (GSVD), Fault Detection & Isolation (FDI)

1. INTRODUCTION

Online process monitoring for fault detection and diagnosis (FDD) is very important for ensuring plant safety and maintaining product quality. The area of FDD has therefore been very active in recent years. Both, model based and process history based methods have been proposed (Kresta *et al.* 1991, Chiang *et al.* 2000, Venkat *et al.* 2003) with a fair amount of success.

In a typical process plant, hundreds of variables are measured every few seconds. These measurements bring in useful signatures about the normalcy or otherwise of plant operation. While model based methods can be used to detect and isolate signals indicating abnormal operation, such quantitative cause-effect models may be difficult to develop. Historical data based methods attempt to extract maximum information out of such datasets with minimal physical knowledge of plant. Because of high dimensionality and correlation amongst variables, multivariate statistical tools, which take correlation amongst variables into account, are better suited for this task. Principal components analysis (PCA) is one of the most popular multivariate statistical methods used for process monitoring and data compression. PCA computes new orthogonal principal directions which are linear combination of the actual variables. This is done by singular value decomposition of a suitably scaled data matrix (X) and retaining those principal components that have significant singular values. PCA has been used for fault detection using statistical control limits Q (Squared Prediction Error) and/ or T^2 statistics (Goulding *et al.* 2000). Once a fault is detected using either Q or T^2 statistics, contributions plot have been used for fault isolation.

The main limitation of PCA is that it assumes normality and independence of the data (Luo *et al.* 1999). In most process plants, because of the dynamic nature of the process, these conditions are very difficult to meet. Other drawbacks of PCA are that (*i*) it is scale dependent and (*ii*) it can not handle dynamic data in its traditional form. To overcome these shortcomings of PCA, many variants of PCA have been proposed in the literature (Ku *et al.* 1995, Bakshi 1998). Also, PCA tries to achieve data compression only across the column space and rows are assumed to be statistically independent. In this paper, we propose the use of Correspondence Analysis for the task of FDI. Correspondence analysis is a powerful multivariate statistical tool, which is also based on generalized singular value decomposition. Correspondence analysis is a dual analysis, as it displays column as well as row points in the dual lower dimensional space. Due to its capability to compress as well as classify the data set, correspondence analysis has been extensively used in ecological problems, study of vegetation habit of species and social networks, etc. Here we discuss how correspondence analysis can be used to classify process data for fault detection as well as isolation simultaneously. The approach is explained through two case studies, viz (*i*) simulation of a CSTR for solution copolymerization of methyl-methacrylate (MMA) and vinyl acetate (VA), developed by Congalidis *et al.* (1986) & (*ii*) Simulation and experimental validation on a quadruple-tank process operating under closed loop.

In the following section, some basic terms used in correspondence analysis are explained. We then present a brief introduction to correspondence analysis followed by how it can be used for statistical process monitoring, fault detection and isolation. Finally the application of this method on the CSTR and the quadruple tank process is discussed along with results.

2. BRIEF PRELIMINARIES

2.1 Weighted Euclidean Space

Correspondence analysis is based on Euclidean distance between points in the weighted Euclidean space. In general, the multidimensional weighted Euclidean space is defined by the scalar product

$$x^T D_q y \triangleq \sum_j q_j x_j y_j \qquad (1)$$

where, q_1, q_2 ... q_j are positive real numbers defining the relative weights assigned to the J respective dimensions. The squared distance between two points x and y in this space is thus the weighted sum of squared differences in coordinates

$$d^2(x,y) \triangleq (x-y)^T D_q (x-y) \qquad (2)$$

This type of distance function is often referred to as a diagonal metric. One of the most common examples of a weighted Euclidean distance is the chi-square (χ^2) statistic.

2.2 Assigning masses to vectors

Different observations are often assigned weights in many statistical methods. In the context of a set of observations, assigning of different masses to the vectors amounts to attaching different degrees of importance to the positions of the points in space. The centroid of points $x_1, x_2, ..., x_I$ with different masses $w_1, w_2, ..., w_I$ is the weighted average point.

$$\overline{x} \triangleq \frac{\sum_i w_i x_i}{\sum_i w_i} \qquad (3)$$

Hence, $\overline{x}$ also tends towards the direction of the points with higher mass. Inertia of a point x_i can be expressed as weighted average

$$in(x_i) = \sum_i w_i d_i^2 \qquad (4)$$

where d_i is the squared distance between x_i and $\overline{x}$.

3. CORRESPONDENCE ANALYSIS

Correspondence analysis takes a single block data $X_{m\times n}$. Each row of matrix X is considered to be a point in an $n-$dimensional space and similarly each column is a point in the $m-$dimensional space. The objective of correspondence analysis is to identify the subspaces of lower dimensionality which best contain the set of points. Thus, in correspondence analysis we try to find a sub-space which comes closest to the set of points. If we look at other SVD based multivariate statistical methods, the objective is always to find lower dimensional subspace which "best" contains set of points. The definition of this optimal subspace, i.e. objective function for optimization, changes from method to method. In PCA, the objective is to maximize the variance explained in the X block. In correspondence analysis however, the objective function is based on the weighted distance between a point and a given subspace. The optimal subspace is defined as the subspace with minimum weighted average distance from the set of points. The measure of closeness is based on the squared distances rather than the distances themselves. This can be viewed as taking measure of the sum of squared errors instead of the absolute errors. Like other methods, such as Fisher Discriminant Analysis (FDA) (Chiang *et al.* 2000), which are considered more appropriate than PCA for fault diagnosis, it can be shown that correspondence analysis is also more suited for associating faults and dynamic correlations in the data, when compared with PCA.

In correspondence analysis first, the correspondence matrix P is constructed as the matrix of elements of X divided by the grand total of X.

$$P \triangleq \frac{1}{g} X \tag{5}$$

where, $g = \mathbf{1}^T X \mathbf{1}$ is grand total of elements of X. Similarly, row and column sums of P are, $r = P\mathbf{1}$ and $c = P^T\mathbf{1}$. In the above expression the symbol $\mathbf{1}$ denotes a vector of appropriate dimension, which has all the elements as 1.

The matrices of row profiles (R) and column profiles (C) of P are defined as the vectors of rows and columns of P divided by their respective sums.

$$\begin{aligned} R &\triangleq D_r^{-1} P \\ C &\triangleq D_c^{-1} P^T \end{aligned} \tag{6}$$

where, $D_r = diag(r)$ and $D_c = diag(c)$. It can be easily shown that r and c would now denote the column and row centroid respectively. The row and column profiles define two clouds of points in respective $n-$ and $m-$dimensional weighted Euclidean space.

Let generalized SVD of $P - rc^T$ be written (Greenacre, 1984) as,

$$P - rc^T = A D_\mu B^T \tag{7}$$

where, $A^T D_r^{-1} A = B^T D_c^{-1} B = I$ and D_μ is the diagonal matrix consisting of generalized singular vaules μ_i in descending order. Then A and B define the principal axes of the column and row clouds respectively.

The respective co-ordinates of the row and column profiles with respect to their own principal axes are related to the principal axes of the other clouds of profiles by simple rescaling.

The principal co-ordinates of row profiles can be written as,

$$F = \left(D_r^{-1} P - \mathbf{1} c^T\right) D_c^{-1} B \tag{8}$$

and likewise, the principal co-ordinates of column profiles can be written as,

$$G = \left(D_c^{-1} P^T - \mathbf{1} r^T\right) D_r^{-1} A \tag{9}$$

Thus, if the principal co-ordinates of row profiles is computed, the principal co-ordinates of column profiles can be derived without any further computations on column profiles. F and G can also be computed as eigen vector matrices of the matrices RC and CR respectively.

3.1 Interpretation and analysis of results

Once the new principal co-ordinates (F and G) are computed, the next step is to interpret and analyze the results that are generated from correspondence analysis. Some metrics and their interpretation can be described as follows.

Singular values and inertia: The sum of the squared singular values gives the total inertia of the cloud. The inertia explained by each principal co-ordinate can then be computed by

$$In(i^{th}\ co-ordinate) = \frac{\mu_i^2}{\sum_j \mu_j^2} \tag{10}$$

where, $j = 1...n$. Similarly, cumulative inertia explained up to the i^{th} principal co-ordinate is the sum of inertias explained up to that principal co-ordinate.

Number of principal co-ordinates to be retained: There is no fixed criterion to determine how many principal co-ordinates should be retained. Although mathematical criterion do exist for selecting number of principal co-ordinate, for better graphical interpretation of results it is generally limited to 2 or at the most 3. In most practical applications, the first two principal co-ordinates explain more than 80% of total inertia.

Row and column scores: The row and column scores are the new coordinates of each row and column profile points on the principal co-ordinates retained.

Absolute contribution by points: This is the metric that indicates the points that have contributed most in ordination to each of principal co-ordinates.

Relative contribution of each axes: This metric measures how well a particular row or column profile point is represented by a particular principal co-ordinate.

3.2 Interpretation from graphical results

The most important part in correspondence analysis is graphical representation of the row and column profile points in the dual lower dimensional space. In most cases the first two principal co-ordinates are plotted against each other and sometimes, one could consider up to the first three principal co-ordinates. The two dimensional plane thus created, can be used to plot the row profile points and/ or column profile points on a single figure. This plot can be used for analysis and interpretation. If the points from the same cloud are close to each other, they can be considered to have similar profile. In order to analyze points in different clouds, the angle between the points is used and not the distance. If the angle between two points in different clouds is acute, then the two are correlated. If the angle is obtuse, the two are correlated but negatively and if the angle is right angle, the points do not interact, or there is no association between the points.

4. APPLICATION OF CORRESPONDENCE ANALYSIS TO FAULT DETECTION

Correspondence analysis can be used for statistical process monitoring as well as fault detection and isolation. In statistical process monitoring, correlation amongst variables is taken into account and statistical limits are drawn for normal operating regions. If the correlation structure is broken, these statistical limits are violated and can be indicative of a fault. Once the fault is detected, the next step is fault isolation. Fault isolation is basically a classification problem and depending on the input-output data, one would be able to classify what would be the root cause.

Correspondence analysis is carried out in dual, i.e. on row as well as on column profiles. This can be used to advantage for performing both fault detection and isolation simultaneously. For process monitoring using correspondence analysis, the data matrix X is formed such that each column constitutes an input/ output variable and each row constitutes time sample measurement of these variables. Each row profile then can be indicative of operating condition of plant. Column profiles on the other hand give information about how variables are related to a particular operating point.

4.1 Data collection

In all statistical process monitoring techniques, data collection is the most important aspect. The information extracted from any analysis depends on how good or rich the data is in terms of significant events of interest. The proposed correspondence analysis method also rely on the historical data available. The results obtained using this method, can also be used to isolate different region of operations e.g. normal operation and different fault scenarios.

4.2 Model Building

Once the data is collected in matrix X, the next step is to compute the correspondence matrix P. Having generated the correspondence matrix P from the data matrix X, further calculations of principal co-ordinates are straight forward (Equation 7, 8, 9). For the applications considered here, we retain the first two principal components for better understanding and insight through graphical display. Analyzing the results without the help of graphical display is also possible using quantitative metrics discussed earlier.

4.3 Online implementation

The key issue here is implementing this method online, under the regular plant operation. For online process monitoring, the new measurement has to be projected onto the lower dimension space obtained via the correspondence analysis. Since, the number of variables (number of columns) being measured do not change during the plant operation, the column cloud remains unaltered. The new measurement should be scaled by its sum to give the new row profile point x_{new}. Its lower dimensional approximation can then be obtained from the following equation.

$$F_{new}(j) = \frac{1}{\mu_j} \sum_{i=1}^{n} x_{new}(i)\, G(i,j) \qquad (11)$$

where, $j = 1, 2$ assuming that first two principal axes are retained.

5. CASE STUDIES

For validation of above methodology, we have carried out two case studies. We first present simulation as well as experimental results on the quadruple tank setup followed by the case study involving copolymerization reactor.

5.1 The Quadruple tank setup

The Quadruple tank process is a multivariable laboratory process with an adjustable zero (Johansson, 2000). Simulations were carried out using the first principles model deployed in closed loop using multi-loop PI-controllers. The data was collected for normal operating condition as well as biases in each of the sensors. The data matrix formed contained two inputs and two outputs as column profile points. Representative points for normal and sensor biases were taken as row profile points.

Figure 1 shows the result obtained from correspondence analysis method. As can be seen in figure, the column profile points for the inputs, namely U_1, U_2 are far away from each other as

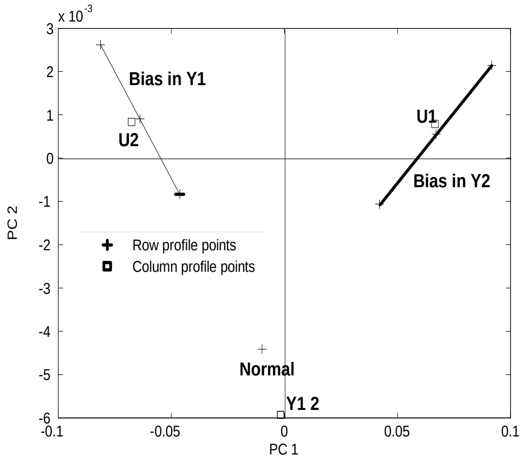

Fig. 1. Correspondence analysis result on Quadruple tank process simulation

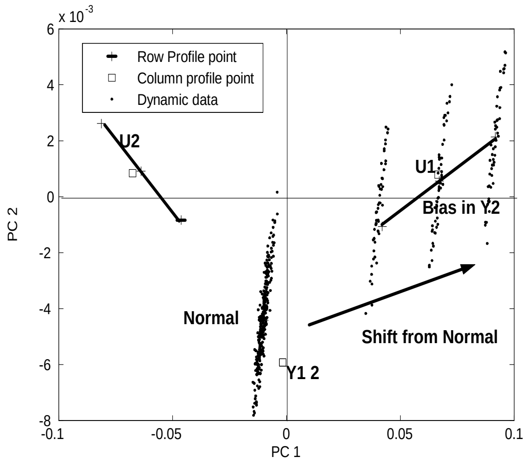

Fig. 2. Online implementation of proposed method on Quadruple tank proces ssimulation

well as Y_1 and Y_2. This indicates that the column profiles for the inputs are quite dissimilar where as column profiles for outputs are almost similar. This is because in close loop condition, the variability of outputs is transformed into the variability in inputs. The row profile points show that process deviates in different direction for faults in different sensors. For this particular example, retaining only first two principle co-ordinates gave full representation of the inertia of the profile points. Absolute contribution to first principal axes is mainly through input column profiles and to that of second principal axes is through output column profiles. This can also be observed from Figure 1. Figure 2 shows the online implementation of the proposed method. The new row profile points lie in the vicinity of normal operating point. When a bias in sensor 2 was introduced, the row profile points can be seen to deviate from normal operating point towards "Bias in Y_2" cloud.

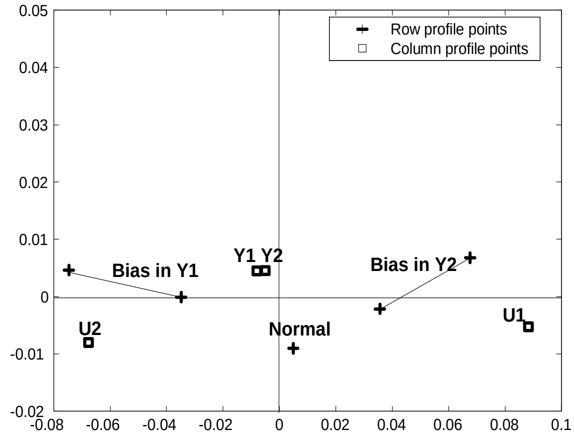

Fig. 3. Correspondence analysis result on Quadruple tank Experimental setup

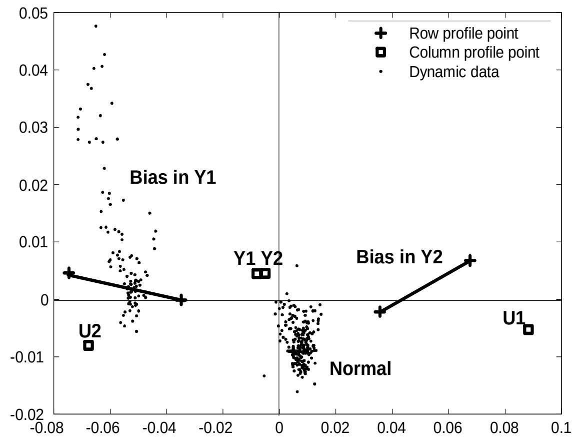

Fig. 4. Online implementation of proposed method on Quadruple tank Experimental setup

To validate the proposed method, we also carried out experimentation on the quadruple tank setup housed in the Process Automation Lab at the Department of Chemical Engineering, Indian Institute of Technology, Bombay. Similar results as obtained from the simulations were also observed from the experimental results. Figure 3 and 4 respectively show the results obtained from correspondence analysis and its online implementation on the experimental setup. It can be seen that sensor bias in the output Y_1 get quickly classified into the "Bias in Y_1" cloud and hence gets isolated.

5.2 Copolymerization Reactor simulation

A 4×5 transfer function matrix model for a CSTR used for solution copolymerization of methyl-methacrylate (MMA) and vinyl acetate (VA) (Congalidis *et al.*, 1986) was simulated under closed loop condition. The pairing of controllers was chosen based on RGA analysis, resulting in

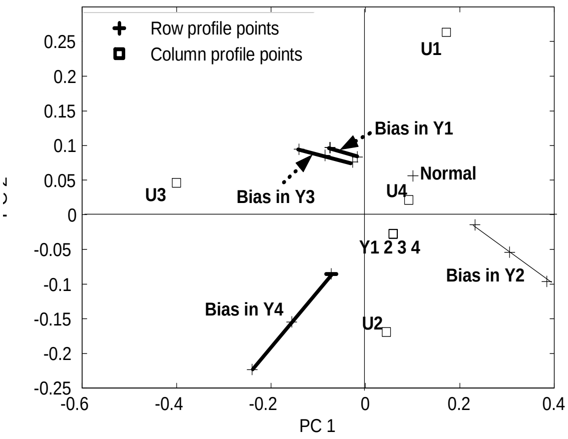

Fig. 5. Correspondence analysis result for CSTR simulation case study

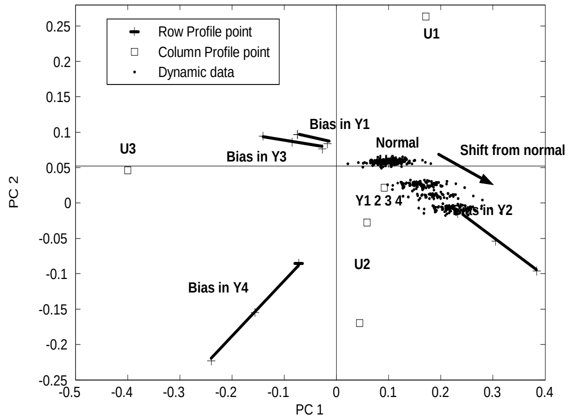

Fig. 6. Online implementation of proposed method for Fault Detection and Isolation

effectively a 4×4 system. The data was collected for simulations of normal operating condition and bias in all sensors.

The data matrix formed contained four inputs and four outputs as columns profile points. From the simulations, representative points for normal operating condition and different fault scenarios are taken as row profile points. Figure 5 shows the results obtained from correspondence analysis. In this case also, column profiles of outputs are similar, whereas column profiles for inputs are entirely different for each variable. The row profiles shows how the normal operating point is placed relative to the points depicting sensor biases. As can be seen from figure 5, points corresponding to bias in Y_1 and bias in Y_3 are very close to each other. Hence, isolation of these faults would be expected to be difficult under dynamic conditions. Figure 6 shows results obtained from online implementation of proposed method. When a bias in sensor 2 was introduced, it can be seen that the row profile points started drifting towards "Bias in Y_2" points.

6. CONCLUSION

A new method based on correspondence analysis is proposed for FDI. The ability of correspondence analysis to depict the row profile points and column profile points in the dual lower dimensional space can be readily used for FDI. A method has also been proposed for online implementation of this method. The method has been validated through simulations and experimental setup.

7. REFERENCES

Bakshi B. R. (1998), Multiscale PCA with Application to Multivariate Statistical Process Monitoring, *AIChE Journal*, **44(7)**, 1596-1610.

Chiang L. H., Russel E. L. & Braatz R. D. (2000), Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis, *Chemometrics & Intelligent Laboratory Systems*, **50**, 243-252.

Congalidis J. P., Richards J. R. & Ray W. H. (1986), Modeling and control of copolymerization reactor, *Proceedings of American Control Conference*, **3**, 1779-1793.

Goulding P. R., Lennox B., Sandoz D. J., Smith K. & Marianovik O. (2000), Fault detection in continuous processes using multivariate statistical methods, *International Journal of Systems and Science*, **31(11)**, 1459-1471.

Greenacre M. J. (1984), *Theory and Application of Correspondence Analysis*, Academic Press, Inc., London

Johansson K. H. (2000), A Quadruple-tank process: A Multivariable laboratory process with an adjustable zero., *IEEE transactions on Control Systems Technology*, **8(3)**, 456-465.

Kresta J. V., MacGregor J. F. & Marlin T. E. (1991), Multivariate statistical monitoring of process operating performance. *Canadian Journal of Chemical Engineering*, **69**, 35-47.

Ku W., Storer R. H. & Georgakis C. (1995), Disturbance Detection & Isolation by dynamic principal components analysis, *Chemometrics & Intelligent Laboratory Systems*, **30**, 179-196.

Luo R., Misra M. & Himmelblau D. M. (1999), Sensor Fault Detection via Multiscale analysis & Dynamic PCA, *Industrial and Engineering Chemistry Research*, **38**, 1489-1495.

Venkatasubramanian V., Rengaswamy R., Yin K. & Kavuri S. N. (2003), A review of process fault detection and diagnosis Part I: Quantitative model-based methods, *Computers and Chemical Engineering*, **27(3)**, 293-311.