Proceeding of the 2004 American Control Conference
Boston, Massachusetts June 30 - July 2, 2004

Improved Diagnosis of Sensor Faults Using Multivariate Statistics

Dirk Lieftucht, Uwe Kruger and George W. Irwin
Intelligent Systems and Control Research Group, Queen's University Belfast, BT9 5AH, U.K.
Correspondence should be addressed to : d.lieftucht@ee.qub.ac.uk

***Abstract*— This paper analyses a variable reconstruction technique for identifying a faulty sensor. The reconstruction is associated with the application of principal component analysis (PCA) and attempts to remove "fault information" from the sensor reading. It is shown that the reconstruction (i) affects the geometry of the PCA decomposition (ii) leads to changes in the covariance matrix of the sensor readings and (iii) alters the determination of PCA based monitoring statistics in terms of their confidence limits. These changes must be incorporated into the monitoring scheme, as false alarms may otherwise be encountered. Consequently, an improved reconstruction based fault diagnosis is proposed here.**

I. Introduction

Sensors provide vital information about the current state of the operation of a large-scale process, which is required for process control, optimization or monitoring. It is therefore essential to detect abnormal sensor readings, such as sensor bias, failure, drift or precision degradation.

MacGregor et al. [1] highlighted that the processes, commonly found in the chemical and manufacturing industry, frequently produced large sets of highly correlated process variables. This has led to the development of statistical modelling techniques that are collectively referred to as multivariate statistical process control (MSPC) [2], [3].

One of the most popular MSPC methods is PCA. Detecting abnormal sensor readings relies on PCA based monitoring statistics. Dunia et al. [4] proposed a reconstruction approach for identifying which sensor produced "faulty" readings. This utilized the linear relationships between the different sensors to predict, for example, the reading of one sensor from the readings of the others. Therefore, by predicting the sensor that produced an incorrect reading, the "fault information" can be removed from this sensor. The monitoring statistics were assumed to show an in-statistical-control situation after the reading of the "faulty" sensor had been reconstructed.

This paper provides a thorough geometric and mathematical analysis of this technique, which reveals that reconstructing a sensor reading leads to changes in the PCA decomposition that describe the relationships between the sensors. More precisely, the PCA model plane, describing the linear combinations among the sensors, remains unchanged although the base vectors, which span the model plane, change their orientation. Furthermore, the dimension of the residual subspace that, representing the direction of the residuals of the PCA model, reduces by one.

Since the model plane and the residual subspace are spanned by eigenvectors of the covariance matrix of the sensor readings, it is shown here that the above changes can be described by the influence of the reconstructed sensor readings upon this covariance matrix. The alteration of the covariance matrix also implies that either the PCA based monitoring statistics or their confidence limits are affected by this reconstruction. If this effect is not removed from these statistics, it is demonstrated that an "out-of-statistical-control" situation may remain even if the "fault information" is correctly removed from the sensor. An improved technique is then introduced, which incorporates the alterations caused by the reconstruction upon the covariance matrix.

The paper is divided into the following sections. A brief review of PCA is given prior to a discussion of how to detect faulty sensors and how to identify which sensor reading is "faulty" using the reconstruction technique. Then, a geometric analysis of the reconstruction technique is given, which is followed by a description of the influence of the reconstruction upon the covariance matrix and the monitoring statistics and their confidence limits. Finally, an application study is presented to (i) demonstrate the "out-of-statistical" situation may be noticeable after the "correct" sensor reading has been reconstructed and (ii) show that removing the influence of the reconstruction from the covariance matrix and the monitoring statistics overcomes this deficiency.

II. Principal Component Analysis

The application of PCA involves the construction of a reduced set of score variables that represents linear combinations of a set of sensor readings. The values of these variables are given by:

$$\mathbf{t} = \mathbf{P}^T \mathbf{z}, \tag{1}$$

where $\mathbf{t} \in \mathbb{R}^n$ is the vector of n score variables, $\mathbf{P} \in \mathbb{R}^{N \times n}$ is a transformation matrix with column vectors as the first n dominant eigenvectors of the correlation matrix $\mathbf{S}_{ZZ} = \frac{1}{K-1}\mathbf{Z}^T\mathbf{Z} \in \mathbb{R}^{N \times N}$. This is established using a reference data set, $\mathbf{Z}^T = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \ldots & \mathbf{z}_K \end{bmatrix}$ with K being the number of mean-centered and scaled sensor readings, $\mathbf{z} \in \mathbb{R}^N$ is the vector of the $(N > n)$ sensor readings. The mismatch error between the measured and predicted sensor readings is:

$$\mathbf{e} = \mathbf{z} - \mathbf{P}\mathbf{t} = \left[\mathbf{I}_N - \mathbf{P}\mathbf{P}^T\right]\mathbf{z}. \tag{2}$$

The application of PCA divides the sensor readings into a model plane and a residual subspace. The model plane is spanned by the first n dominant eigenvectors of $\mathbf{S}_{ZZ}$ and describes the linear combinations of the sensor readings. In contrast, the residual subspace is spanned by the remaining $(N-n)$ eigenvectors of $\mathbf{S}_{ZZ}$ and represents the mismatch error of the PCA prediction of the sensor readings.

III. Fault Detection Using Univariate Statistics

Two univariate statistics, denoted by T^2 and Q, can be established using the values of the score variables, $\mathbf{t}$, and the mismatch error $\mathbf{e}$:

$$\begin{aligned} T^2 &= \mathbf{t}^T\mathbf{\Lambda}\mathbf{t} = \sum_{i=1}^{n} \frac{t_i^2}{\lambda_i} \quad \text{and} \\ Q &= \mathbf{e}^T\mathbf{e} = \sum_{j=1}^{N} e_j^2. \end{aligned} \tag{3}$$

Here $\mathbf{\Lambda}$ is a diagonal matrix containing the n largest eigenvalues λ_i of $\mathbf{S}_{ZZ}$ in descending order, t_i is the i^{th} score and e_j is the j^{th} residual error. Each of the above statistics can be plotted against time and the confidence limit for the T^2 statistics is given by:

$$T^2_{n,K,\alpha} = \frac{n(K-1)}{K-n} F_{n,K-n,\alpha}, \tag{4}$$

where α is the confidence, typically 95% or 99%, and $F_{n,K-n,\alpha}$ is the value representing the confidence limit of a F-distribution. The confidence limit for the Q statistic can be calculated as follows:

$$Q_\alpha = \theta_1 \left(\frac{c_\alpha\sqrt{2\theta_2 h_0^2}}{\theta_1} + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} + 1 \right)^{\frac{1}{h_0}}, \tag{5}$$

where $\theta_1 = \sum_{i=n+1}^{m} \lambda_i$, $\theta_2 = \sum_{i=n+1}^{m} \lambda_i^2$, $\theta_3 = \sum_{i=n+1}^{m} \lambda_i^3$, $h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2}$ and the variable c_α is based on the inverse error function, evaluated for a confidence α [5].

IV. Fault Diagnosis Using Variable Reconstruction

The reconstruction of a "faulty" sensor reading is based on the prediction of the sensor readings, $\hat{\mathbf{z}}$ using the PCA model:

$$\widehat{\mathbf{z}} = \mathbf{P}\mathbf{t} = \mathbf{P}\mathbf{P}^T\mathbf{z} = \mathbf{C}\mathbf{z}, \tag{6}$$

where $\mathbf{C} = \mathbf{P}\mathbf{P}^T$. More precisely, given that the i^{th} sensor is "faulty", the prediction of the i^{th} sensor replaces the recorded reading of this sensor to filter out the "faulty" information. Using the elements of the i^{th} row of $\mathbf{C}$, c_{ij} the prediction of the i^{th} sensor reading is equal to:

$$\tilde{z}_i = \sum_{j=1\neq i}^{N} \frac{c_{ij}}{1-c_{ii}} z_j \tag{7}$$

where $\tilde{z}_i$ is the reconstructed reading of the i^{th} sensor z_i. Geometrically, the reconstruction process shifts the data point, representing the values of the sensor readings in an N dimensional space, along the direction of the i^{th} variable.

After $\tilde{z}_i$ replaces z_i to form the vector of sensor readings $\tilde{\mathbf{z}}$, the T^2 and Q statistics can be computed. This is followed by updating of the on-line monitoring charts. If the "faulty" sensor has been correctly identified, it is assumed that the monitoring statistics show an "in-statistical-control" situation. However, it is explained in this paper that the "out-of-statistical-control" situation may still remain even if the "faulty" sensor has been reconstructed.

V. Geometric Analysis of Reconstruction Based Fault Diagnosis

The geometrical properties of the reconstruction technique proposed by Dunia et al. [4] are now analysed. The distance of a shift of the data point to the model plane can be described by Theorem 1.

Theorem 1: The reconstruction of the i^{th} sensor reading results in a shift of the data point along the direction of the i^{th} variable in the N-dimensional space. This leads to a minimization of the distance between the model plane and the shifted data point.

A proof of Theorem 1 is given in Appendix A. Given that a data point $\mathbf{z}$ is described in a N dimensional space, the reconstruction of z_i to form $\tilde{\mathbf{z}}$ results in a projection of $\mathbf{z}$ onto a $(N-1)$-dimensional plane, as described in Theorem 2.

Theorem 2: The reconstruction of z_i is equivalent to the projection of $\mathbf{z}$ onto a $(N-1)$-dimensional plane (subspace) Π along the direction of the i^{th} variable. This plane is spanned by the following base vectors:

$$\Pi = \left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ \frac{c_{i1}}{1-c_{ii}} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ \frac{c_{i2}}{1-c_{ii}} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \frac{c_{i(i-1)}}{1-c_{ii}} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \frac{c_{i(i+1)}}{1-c_{ii}} \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \frac{c_{iN}}{1-c_{ii}} \\ 0 \\ \vdots \\ 1 \end{pmatrix} \right\}. \tag{8}$$

Note that this plane includes the model plane and the residual subspace. A proof of Theorem 2 is given in Appendix B.

VI. Reconstruction of the Covariance Matrix

The last section highlighted that reconstructing the i^{th} sensor reading gives rise to a description of each data point by a $(N-1)$-dimensional subspace rather than the original N-dimensional one.

Since the model plane and the residual subspace are spanned by the eigenvectors of $\mathbf{S}_{ZZ}$, it is imperative to analyze the influence of the variable reconstruction upon the covariance matrix. This allows the influence of the reconstruction upon the model plane and the residual subspace to be analyzed. The next subsection discusses how to determine the "reconstructed covariance matrix", which is followed by an evaluation of the changes in the model plane and the residual subspace.

A. Calculation of the Reconstructed Covariance Matrix

The reconstruction of the i^{th} sensor reading requires the i^{th} row and the i^{th} column of $\mathbf{S}_{ZZ}$ to be recalculated as follows:

$$ {}_i\tilde{s}_{ik} = \sum_{a=1\neq i}^{N} \frac{c_{ia}}{1-c_{ii}} s_{ka}. \tag{9} $$

In (9) ${}_i\tilde{s}_{ik}$ are the non-diagonal elements of the i^{th} row and the i^{th} column of reconstruction of $\tilde{\mathbf{S}}_{ZZ}$, ${}_i\tilde{\mathbf{S}}_{ZZ}$. The i^{th} diagonal element of ${}_i\tilde{\mathbf{S}}_{ZZ}$ is given by:

$$ \begin{aligned} {}_i\tilde{s}_{ii} &= \sum_{a=1\neq i}^{N} \frac{c_{ia}^2}{(1-c_{ii})^2} s_{aa} \\ &+ 2\sum_{a=1\neq i}^{N-1}\sum_{b=a+1\neq i}^{N} \frac{c_{ia}c_{ib}}{(1-c_{ii})^2} \end{aligned} \tag{10} $$

Remark 1: The rank of ${}_i\tilde{\mathbf{S}}_{ZZ}$ is $N-1$, as the i^{th} row or column is linearly dependent on the remaining $N-1$ rows or columns.

B. Changes in the Model Plane

Pearson [6] showed that the squared length of the residual vector, between a set of data points of dimension N and a given model plane of dimension n, is minimized if the model plane is spanned by the first-n dominant eigenvectors of the covariance matrix which is established on the basis of this set of data points.

Theorems 1 and 2 show that the projections of a set of data points onto Π leads to a minimum distance between the projected points and the model plane. This gives rise to the following Lemma:

Lemma 1: The reconstruction of the i^{th} sensor reading does not influence the orientation of the model plane.

The above lemma follows from [6].

Remark 2: The fact that the orientation of the model plane is not affected by the reconstruction procedure does not imply that the orientation of the first n dominant eigenvectors remain unchanged.

The above remark is a result of the changes that the reconstruction procedure imposes on $\mathbf{S}_{ZZ}$. Moreover, the dominant eigenvalues of ${}_i\tilde{\mathbf{S}}_{ZZ}$ may also change.

Remark 3: Since the dominant eigenvectors of ${}_i\tilde{\mathbf{S}}_{ZZ}$ may be different from those of $\mathbf{S}_{ZZ}$ implies that (i) the directions for which the score variables have maximum variance may change and (ii) the variance of each score variable may change.

C. Changes of the Residual Subspace

As shown in Section V, the reconstruction results in a shift of a data point along the i^{th} variable so that the squared length of the error vector is minimal (Theorem 1). Since the reconstruction procedure is, in fact, a projection of $\mathbf{z}$ onto Π, which is of dimension $(N-1)$ (Theorem 2), it follows that the dimension of the residual subspace is $(N-n-1)$, because the dimension of the model plane remains unchanged.

Since the model plane is assumed to describe the linear relationships between the sensors, the eigenvalues of the discarded eigenvectors represent the variance of the measurement uncertainty of the sensor reading. Given that ${}_i\tilde{\mathbf{S}}_{ZZ}$ has rank $(N-1)$, one eigenvalue is equal to zero. The measurement uncertainty of each sensor reading is assumed to be represented by normally distributed, independently and identically distributed (i.i.d.) sequences. This implies that the nonzero eigenvalues of the discarded eigenvectors are equivalent.

Remark 4: The squared length of the error vector is equal to the scaled sum of the "discarded" eigenvalues. This implies that the squared length of the error vector after reconstruction is smaller, or equal to the length of the error vector of the unreconstructed data point.

VII. Adaptation of the Confidence Limits

Subsection VI-C highlighted that a change in the model plane and residual subspace arise, as a consequence of the reconstruction procedure. These changes can be described by ${}_i\tilde{\mathbf{S}}_{ZZ}$ and may lead to a different set of eigenvalue and eigenvectors. Given the definition of the T^2 and Q statistics, which are determined using the eigenvector-eigenvalue decomposition of ${}_i\tilde{\mathbf{S}}_{ZZ}$, it is necessary to account for such changes, if the reconstruction procedure has been applied.

If the i^{th} sensor reading is to be reconstructed, the following steps must be carried out to incorporate the influence of the reconstruction upon the model plane and the residual subspace: (i) reconstruct the covariance matrix by applying (9) and (10), (ii) calculate the eigenvalues and eigenvectors of ${}_i\tilde{\mathbf{S}}_{ZZ}$, (iii) calculate the T^2 statistic using the retained eigenvectors and eigenvalues of ${}_i\tilde{\mathbf{S}}_{ZZ}$, i.e. $\tilde{\mathbf{t}} = \tilde{\mathbf{P}}^T\tilde{\mathbf{z}}$ and $\tilde{T}^2 = \tilde{\mathbf{t}}^T\tilde{\mathbf{\Lambda}}^{-1}\tilde{\mathbf{t}}$, (iv) compute the Q statistic using: $\tilde{\mathbf{e}} = \tilde{\mathbf{z}} - \tilde{\mathbf{P}}\tilde{\mathbf{t}}$ and (v) estimate the corrected confidence limits for the Q statistic by reapplying (5) using the "discarded" eigenvalues of ${}_i\tilde{\mathbf{S}}_{ZZ}$.

VIII. Example

A simulation study is used to demonstrate that the application of the conventional reconstruction technique by Dunia et al. [4] may present an "out-of-statistical-control" situation even if the "faulty" sensor is correctly identified. The improved reconstruction technique, introduced in Section 7 is shown to correct this anomalous situation.

The simulation data $\mathbf{Z}$ involves 10 sensors of which any 5 are linearly independent upon the remaining 5 ones. A reference data set containing 500 samples was generated as follows: 500 samples of 5 normally distributed random variables $\mathbf{x}$ of zero mean were simulated and the elements of $\mathbf{z}$ were determined as linear combinations of $\mathbf{x}$. After scaling $\mathbf{z}$ to unit variance, normally distributed i.i.d sequences of variance 0.05 were superimposed on $\mathbf{z}$ to represent the measurement uncertainty.

The retained eigenvectors, $\mathbf{P}$, and eigenvalues, λ, of $\mathbf{S}_{ZZ}$ were computed to be:

$$\lambda^T = \begin{pmatrix} 4.44 & 0.77 & 0.54 & 0.37 & 0.12 \end{pmatrix} \quad (11)$$

$$\mathbf{P} = \begin{bmatrix} -0.34 & 0.03 & -0.18 & 0.27 & -0.33 \\ -0.34 & 0.15 & 0.15 & -0.21 & 0.31 \\ -0.33 & 0.14 & -0.07 & -0.53 & 0.12 \\ -0.27 & -0.70 & -0.17 & 0.16 & 0.26 \\ -0.30 & 0.59 & 0.06 & 0.32 & -0.10 \\ -0.33 & -0.09 & -0.21 & 0.42 & 0.26 \\ -0.31 & -0.29 & 0.34 & -0.41 & -0.37 \\ -0.32 & -0.12 & 0.48 & 0.26 & -0.39 \\ -0.30 & 0.08 & -0.69 & -0.23 & -0.30 \\ -0.34 & 0.14 & 0.22 & -0.04 & 0.51 \end{bmatrix}.$$

The discarded eigenvalues are equal to 0.05. The second sensor reading of the reference data set was first reconstructed using the technique by Dunia et al. [4] then by, applying the improved reconstruction technique discussed in Section VII. The influence of the reconstruction upon the retained eigenvectors, $\tilde{\mathbf{P}}$, and eigenvalues, $\tilde{\lambda}$, is shown below.

$$\tilde{\lambda}^T = \begin{pmatrix} 4.53 & 1.01 & 0.57 & 0.42 & 0.17 \end{pmatrix} \quad (12)$$

$$\tilde{\mathbf{P}} = \begin{bmatrix} -0.34 & 0.01 & 0.20 & -0.18 & 0.38 \\ -0.33 & -0.18 & -0.13 & 0.10 & -0.36 \\ -0.32 & -0.22 & 0.01 & 0.48 & -0.21 \\ -0.29 & 0.75 & 0.04 & -0.03 & -0.14 \\ -0.28 & -0.52 & 0.12 & -0.41 & 0.09 \\ -0.34 & 0.24 & 0.26 & -0.37 & -0.16 \\ -0.31 & 0.07 & -0.49 & 0.42 & 0.31 \\ -0.31 & 0.02 & -0.47 & -0.29 & 0.41 \\ -0.29 & -0.04 & 0.61 & 0.40 & 0.29 \\ -0.34 & -0.11 & -0.15 & -0.10 & -0.53 \end{bmatrix}.$$

As expected, the effect of the reconstruction of the second sensor slightly increases the values of the retained eigenvalues and produces a different set of eigenvectors. The T^2 and Q statistics for the reconstructed data points $\tilde{\mathbf{z}}$ of the reference set are shown in Fig. 1 using the conventional reconstruction technique by Dunia et al. [4] and in Fig. 2 for the proposed improved reconstruction technique.

The T^2 and Q statistics that were obtained using the technique by Dunia et al. [4] represent an "out-of-statistical-control" situation for the T^2 statistic and suggest that

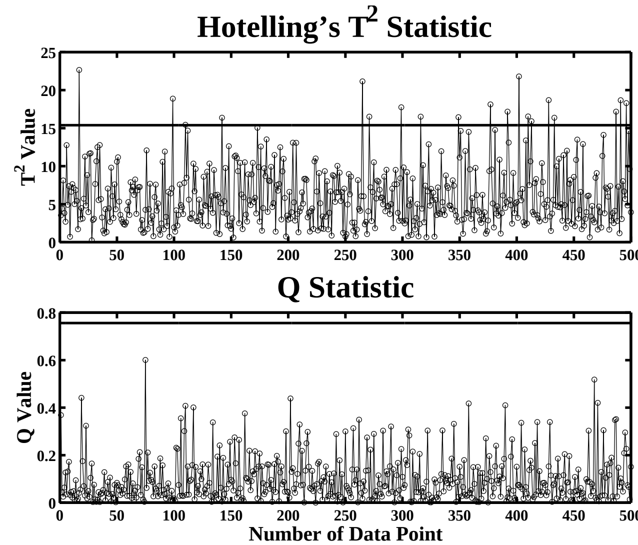

Fig. 1. T^2 and Q statistic after variable reconstruction based onto the technique proposed by Dunia et al. [4]

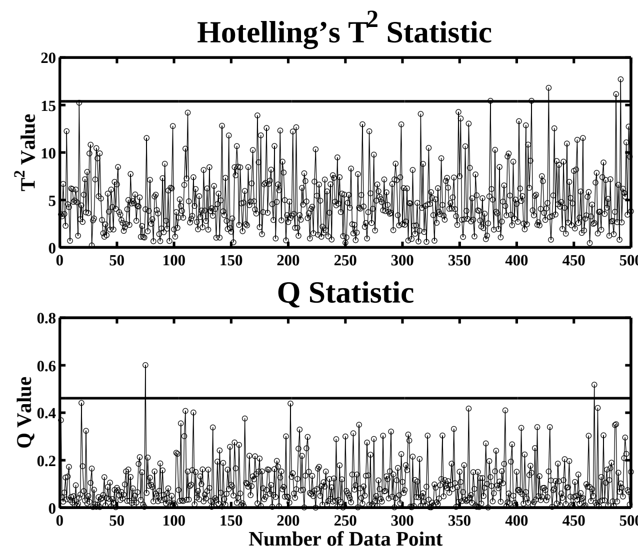

Fig. 2. T^2 and Q statistic with adapted confidence limits after a proposed variable reconstruction

the confidence limit of the Q statistic appears to be too large. Using the improved reconstruction technique, the "in-statistical-control" situation is noticeable for the T^2 statistic and the confidence limit of the Q statistic is also much smaller. Consequently, this reconstruction of the second variable has shown that the technique by Dunia et al. [4] may run into difficulties, whilst the proposed technique successfully removes the deficiency.

IX. Conclusions

The mathematical and geometric properties of the proposed variable reconstruction technique by Dunia et al. [4], for identifying sensor faults, were analyzed. This revealed that the reconstruction results in shifting a data point along the variable that is to be reconstructed in a multidimensional space of dimension N. It was further shown that the reconstruction is, in fact, a projection along the direction of the reconstructed variable onto a $(N-1)$-dimensional plane.

The reconstruction itself gives rise to changes in the model plane and the residual subspace, which can be described by incorporating the influence of the reconstruction procedure upon the covariance matrix of the sensor readings. However, the changes in this matrix also imply that the monitoring statistics are also affected by the reconstruction procedure. Since it is assumed that after a "faulty" sensor has been reconstructed, the monitoring statistics show the "in-statistical-control" situation, it is important to incorporate the effect of the reconstruction for determining these statistics. Otherwise an "out-of-statistical-control" situation may arise as was shown by a simulation study that involved a total of 10 sensors. This application study also demonstrated that by incorporating the influence of the effect of the reconstruction upon these statistics, the "in-statistical-control" situation prevailed.

APPENDIX

Following the reconstruction technique of Dunia et al. [4] the reconstruction of the i^{th} variable is analyzed and it is shown for the N dimensional case that a minimal distance is determined by applying this projection. Where the prediction error is equal to $\tilde{\mathbf{e}} = [\mathbf{I} - \mathbf{C}]\,\tilde{\mathbf{z}}$, and

$$\tilde{\mathbf{z}} = \begin{pmatrix} z_1 & \cdots & z_{i-1} & \sum_{a=1\neq i}^{N} \alpha_a z_a & z_{i+1} & \cdots & z_N \end{pmatrix}. \quad (13)$$

The squared distance of the prediction error using the reconstructed data point $\tilde{\mathbf{z}}$ is given by:

$$\begin{aligned} \tilde{\mathbf{e}}^T\tilde{\mathbf{e}} &= \tilde{\mathbf{z}}\,[\mathbf{I}-\mathbf{C}]\,\tilde{\mathbf{z}}^T \quad (14) \\ &= \left[\sum_{a=1\neq i}^{N} \alpha_a z_a\right]^2 (1-c_{ii}) - 2\sum_{a=1\neq i}^{N}\sum_{j=1\neq i}^{N} \alpha_a c_{ji} z_a z_j + u, \end{aligned}$$

where u is a remaining term which not depends on α and therefore is neglected. The derivatives for α_o with $o \in (1\ldots N) \wedge o \neq i$ are determined as

$$\frac{\partial f}{\partial \alpha_o} = 2\,(1-c_{ii}) z_o \sum_{a=1\neq i}^{N} \alpha_a z_a - 2 z_o \sum_{a=1\neq i}^{N} c_{ai} z_a \quad (15)$$

and set to be equal to zero:

$$(1-c_{ii}) \sum_{a=1\neq i}^{N} \alpha_a - \sum_{a=1\neq i}^{N} c_{ai} = 0. \quad (16)$$

This leads to

$$\alpha_a = \frac{c_{ia}}{1-c_{ii}} \text{ for } a = 1\ldots N \wedge a \neq i. \quad (17)$$

Based on this, the prediction error of the projected points has to be smaller or equal to the original prediction error.

To show that the reconstruction of the i^{th} sensor reading is, in fact, a projection of $\mathbf{z}$ onto a $(N-1)$-dimensional subspace, the computation of $\tilde{\mathbf{z}}$ can be revisited:

$$\tilde{\mathbf{z}} = \begin{pmatrix} z_1 \\ \vdots \\ z_{i-1} \\ \sum_{a=1\neq i}^{N} \frac{c_{ia}}{1-c_{ii}} z_a \\ z_{i+1} \\ \vdots \\ z_N \end{pmatrix}. \quad (18)$$

The projection plane is therefore described by the following $(N-1)$ base vectors:

$$\Pi = \left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ \frac{c_{i1}}{1-c_{ii}} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \frac{c_{i(i-1)}}{1-c_{ii}} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \frac{c_{i(i+1)}}{1-c_{ii}} \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \frac{c_{iN}}{1-c_{ii}} \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \right\}.$$

ACKNOWLEDGEMENT

Dirk Lieftucht would like to acknowledge the financial support from the Queen's University of Belfast and the European Social Fund.

REFERENCES

[1] J. F. MacGregor, T. E. Marlin, J. V. Kresta, and B. Skagerberg, "Multivariate statistical methods in process analysis and control," in *Proc. of the 4^{th} Int. Conf. on Chemical Process Control*. New York: AIChE Publication, No. P-67, 1991, pp. 79–99.

[2] M. J. Piovoso and K. A. Kosanovich, "Process data chemometric," *IEEE Trans. on Instrumentation and Measurement*, vol. 41, no. 2, pp. 262–268, 1992.

[3] J. F. MacGregor and T. Kourti, "Statistical process control of multivariate processes," *Control Engineering Practice*, vol. 3, no. 3, pp. 403–414, 1995.

[4] R. Dunia, S. J. Qin, T. F. Edgar, and T. J. McAvoy, "Identification of faulty sensors using principal component analysis," *AIChE Journal*, vol. 10, no. 10, pp. 2797–2812, October 1996.

[5] J. E. Jackson and G. S. Mudholkar, "Control procedures for residuals associated with principal component analysis," *Technometrics*, vol. 21, pp. 341–349, 1979.

[6] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Phil. Mag.*, vol. 2, no. 11, pp. 559–572, 1901.