2nd International Conference on Electrical and Electronics Engineering (ICEEE) and XI Conference on Electrical Engineering (CIE 2005)
Mexico City, Mexico. September 7-9, 2005

Fault Detection Using Dynamic Principal Component Analysis by Average Estimation

J. Mina and C. Verde
Instituto de Ingeniería, UNAM, DF, México
Fax (52) 55-562-33600 ext. 8052 E-mail: jminaa@iingen.unam.mx

Abstract — **This paper presents a fault detection method based on pattern classification in which the stationary property of the time series of a dynamic system is not required. In particular the dynamic principal component analysis, DPCA, method for feature extraction in diagnostics issues is extended to the case of nonstationary data. The idea is to improve the DPCA performance introducing on-line average estimations of the input and output of the dynamic systems. These new parameters of the data allow extending the method for no stationary time series reducing the false alarms during the detection stage. As study case the detection of faults in a flow control valve has been used, in which it is assumed that the control signal and stem displacement are measured signals. Simulator Data has been used to adjust the procedure and show the effectiveness of the proposed methodology.***

Keywords — **Dynamic Principal Component Analysis, Fault Detection, Classification in Diagnostics, No Stationary Time Series, Feature Extraction.**

I. INTRODUCTION

In many industrial process, the only process information is available from measurements generated by sensors and actuators. In this case the analytical model based fault diagnostics approaches [1] are useless and classification methods from the time series should be applied. One possibility is the use of algorithms based in pattern recognition, where it is necessary to get an implicit standard model based on time series which is compared on-line with the signals taken from the process and if they don't match with the standard, a fault symptom is obtained. A variety of approaches can be used to obtain the implicit standard models and the fault symptom [2]. In particular the multivariate statistical tool called Principal Component Analysis, PCA, has been recognized as a powerful tool for fault diagnosis issues, since it can be used to extract feature of a system based only on measured data. The procedure transforms, from a parallel correlation analysis, a set of multivariate observations to a lower dimension space, retaining the most of the variance of the original data [3]. Since the time series of multivariate dynamic systems exhibit an auto-correlation degree, the PCA was extended for data coming from a dynamic system taking into account series correlation of variables instead of only parallel correlation; this extension is called dynamic principal component analysis DPCA [4]. However, it is important to note that in both cases one assumes the stationarity property in signals to normalize the data and therefore, fault alarms are generated if the signals don't hold the property.

The false alarms problem of DPCA motivated this work in which one overcomes the stationary issue introducing on-line average estimations of the signals for the normalization step. One shows that these average estimations reduce the false alarms rate and do not require a precise description of the dynamic system. As a case of study it is analyzed a flow control valve which implicit model is obtained from the control input signal and stem displacement signal, it is also carried out a set of testing to demonstrate the effectiveness of the proposed methodology.

II. DYNAMIC PRINCIPAL COMPONENT ANALYSIS

Let $\boldsymbol{X}$ be a set of data composed of n_t observations from p variables, in which some of the variables are inputs and the rest outputs, described by the matrix

$$\boldsymbol{X} = \lfloor X_1 \quad X_2 \quad \cdots \quad X_p \rfloor_{(n_t \times p)} \tag{1}$$

It is assumed that each of the p time series are stationary, this is that their means and standard deviations stay practically constant for every time interval. However, because the data matrix (1) is generated from a dynamic system, there exists time dependence in each time series where a current value depends on past values. So, to include the serial correlation of data, it is constructed the so called *trajectory matrix* applying a 'time lag shift' of order w on each of the p columns of the matrix $\boldsymbol{X}$. This means

$$\boldsymbol{X}_i^{\,w} = \begin{bmatrix} X_i(1) & X_i(2) & \cdots & X_i(w) \\ X_i(2) & X_i(3) & \cdots & X_i(w+1) \\ \vdots & \vdots & \ddots & \vdots \\ X_i(n_t - w + 1) & X_i(n_t - w + 2) & \cdots & X_i(n_t) \end{bmatrix}_{(n \times w)}$$

$$\boldsymbol{X}^{w} = \left[\boldsymbol{X}_1^{\,w} \quad \boldsymbol{X}_2^{\,w} \quad \cdots \quad \boldsymbol{X}_p^{\,w}\right]_{(n \times m)} \tag{2}$$

where $n = n_t - w + 1$; $m = pw$. The choice of w is made establishing a compromise between information content and statistical confidence. It is common to select w in the same way as is defined the number of lags to use when constructing an auto-correlation function [5], $w = n_t/4$ [6].

* This paper includes results of the EOLI Project supported by the INCO program of the European Community (Contract number ICA4-CT-2002-10012) and the scholarship program of CONACYT and DGEP-UNAM.

IEEE Catalog Number: 05EX1097
ISBN: 0-7803-9230-2

In a multivariate approach some of the variables have different range of values or units, therefore it is convenient to perform a standardization of data around the corresponding means and the standard deviations to obtain a standardized data matrix $\boldsymbol{Xs}^w$ with zero mean and unit variance.

The means vector and standard deviations vector of the trajectory matrix are

$$\underline{\mu} = [\mu_1(.) \quad \mu_2(.) \quad \cdots \quad \mu_m(.)]_{(1\times m)}$$
$$\underline{\sigma} = [\sigma_1(.) \quad \sigma_2(.) \quad \cdots \quad \sigma_m(.)]_{(1\times m)}$$

where due to the stationarity property:

$$\mu_1(.) = \mu_2(.) = \ldots = \mu_m(.) \tag{3a}$$
$$\sigma_1(.) = \sigma_2(.) = \ldots = \sigma_m(.) \tag{3b}$$

Thus, the data standardization it is obtained through

$$\boldsymbol{xs}_{ij}^w = \frac{\boldsymbol{x}_{ij}^w - \mu_j}{\sigma_j} \quad \text{for} \quad \begin{matrix} i = 1,\ldots,n \\ j = 1,\ldots,m \end{matrix} \tag{4}$$

applying PCA on $\boldsymbol{Xs}^w$ is what constitutes DPCA.

III. Fault detection via DPCA

The principal components are linear transformation of the original variables $\boldsymbol{Xs}^w$, such principal components are a new set of variables which are uncorrelated to each other, this is

$$\boldsymbol{Y} = \boldsymbol{Xs}^w \boldsymbol{V}_t$$

The transformation matrix $\boldsymbol{V}_t$ is composed of an appropriate selection of eigenvectors associated to the correlation matrix $\boldsymbol{R}$ of the trajectory matrix, $\boldsymbol{X}^w$. From the principal components it is possible to define the univariate statistic $T_y^{\,2}$ (*Hotelling parameter*) for each observation in $\boldsymbol{Y}$

$$T_y^{\,2} = y \boldsymbol{S}_{\boldsymbol{Y}}^{-1} y'$$

$\boldsymbol{S}_{\boldsymbol{Y}}$ is the covariance matrix of $\boldsymbol{Y}$. Finally, it is defined a threshold of normal condition from the *probability density function* (*pdf*) of $T_y^{\,2}$. In [7] it is very well described the procedure to obtain this threshold named UCL (Upper Control Limit) in the area of Statistical Quality Control.

During the fault detection stage, it is taken on line an actual *m*-dimensional observation vector $\boldsymbol{xa}^w$ which is standardized around the means and standard deviations (3a) and (3b) of the reference data $\boldsymbol{X}^w$, transformed through $\boldsymbol{V}_t$, and mapped in the univariate parameter $T_{y_a}^{\,2}$; if the resulting value deviates from the threshold it is an indication of the presence of a fault.

This fault detection algorithm is capable of distinguish if an actual observation has, even if little, differences with regard to the mean and standard deviation of the reference data. However, given that $\boldsymbol{X}$ contains not only output signals but also input signals, and that the output signals are the response of the dynamic system to the input signals, when the magnitude of some of the input signals change or the operating point is changed, the stationarity property is broken, at least with regard to their means

$$\mu_1(.) \neq \mu_2(.) \neq \ldots \neq \mu_m(.)$$

so, even these changes are normal changes in the signals, the fault detection algorithm will interpret them as faults. To overcome these difficulties it is proposed the estimation of the output signals in order to get, on line, the averaging both of the input and output signals and use these averages in the standardization procedure. The objective is to fix the data at zero mean and unit variance even before magnitude changes in the input signal.

IV. Fault detection via DPCA with average estimation

Assuming a SISO system under normal condition, from the set of input data X_1 and output data X_2 and defining

$$\boldsymbol{X} = [X_1 \quad X_2] \tag{5}$$

a reference nominal threshold from the *pdf* of $T_{\boldsymbol{Y}}^{\,2}$, is obtained, off line, following the procedure of section II and III. Moreover, it is assumed a relationship $x_2 = f(x_1)$ between the input and the output signal in normal condition.

For the fault detection, on line, a two step algorithm will be carried on, the first one consists in estimating the average of both input and output signals, the second one in data standardization according to the on line estimated averages, and the application of the fault detection based in DPCA.

Step 1. Estimating x_2 from x_1 through *f(.)* the following vector is constructed

$$\hat{\boldsymbol{x}} = [x_1(.) \quad \hat{x}_2(.)] \tag{6}$$

which is composed of the actual measurement of the input signal and the output estimation. Applying the time lag shift of the same order w used for (5) in (6) it is obtained the following vector, for $k \geq w$

$$\hat{x}^{w} = [x_1(k-w+1) \ \cdots \ x_1(k) \ \ \hat{x}_2(k-w+1) \ \cdots \ \hat{x}_2(k)]_{(1\times m)}$$

Using the well known *single moving average* on $\hat{x}^{w}$ an estimation of the means vector is obtained

$$\underline{\hat{\mu}} = [\hat{\mu}_1(.) \ \ \hat{\mu}_2(.) \ \cdots \ \hat{\mu}_m(.)]_{(1\times m)} \qquad (7)$$

Step 2. Starting from the actual *2*-dimensional observation vector

$$\boldsymbol{xa} = [x_1(.) \ \ x_2(.)]$$

it is obtained the following time lag shift vector of actual observations

$$\boldsymbol{xa}^{w} = [x_1(k-w+1) \ \cdots \ x_1(k) \ \ x_2(k-w+1) \ \cdots \ x_2(k)]_{(1\times m)} \qquad (8)$$

Now, the standardization of (8) is carried out according to the standard deviations in (3b) and the means estimated in (7) obtaining $\boldsymbol{xas}^{w}$, where

$$\boldsymbol{xas}_j^{w} = \frac{\boldsymbol{xa}_j^{w} - \hat{\mu}_j}{\sigma_j} \qquad \text{for} \qquad j = 1,\ldots,m$$

$\boldsymbol{xas}^{w}$ will have approximately unit variance and zero mean under normal operating condition even before level changes in the input signal. Finally, the fault detection is carried on using $\boldsymbol{xas}^{w}$.

V. Fault Detection in a Flow Control Valve

A. *The Scheme for Fault Diagnosis*

The fault detection task can be carried out exploiting the variety of internal signals that exist into the different blocks which constitute an actuation system (positioner, signal conversion, modulator, final control element, etc.), see Fig. 1.

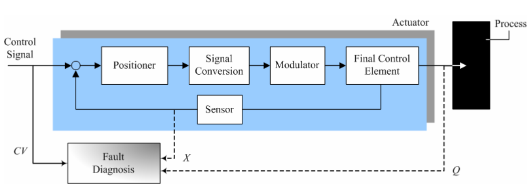

Fig. 1 Actuation System with Fault Diagnosis Module.

The fault diagnosis module will catch in every sampling period the control signal as input variable and other available signals as output variables. The idea is to take advantage of the correlation between time series signals.

B. *Valve Model*

For simulations it is used the DAMADICS (Development and Application of Methods for Actuator Diagnosis in Industrial Control Systems) benchmark [8]. This benchmark consists of the valve model in Simulink®, and other modules which comprise the DABLib (DAMADICS Actuator Benchmark Library), see Fig. 2.

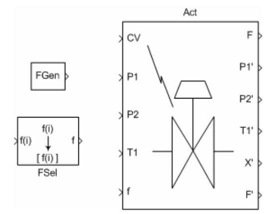

Fig. 2 DAMADICS Actuator Benchmark Library, DABLib.

C. *Validation of the Fault Detection Algorithm via DPCA*

It'll be evaluated the fault detection algorithm described above for the following cases: (a) normal operation of the valve before changes in the input signal; (b) fault condition, *f1-Valve Clogging* of magnitude 0.5 and 0.75.

In the first evaluation we are interested in testing how the algorithm responds with respect to step changes in the control signal *CV* and with the valve under normal operating condition; for this it was proposed positive and negative step changes non higher than the *maximum step size* permitted for a valve ($\cong 10\%$) [9], the sequence is as follows: positive step change from $(18s \le t \le 43s)$ and negative step change from $(93s \le t \le 143s)$.

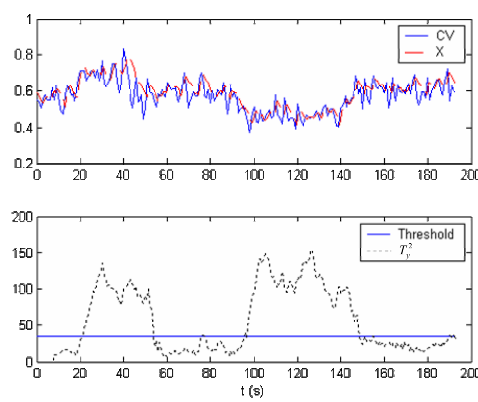

Fig. 3 Step changes in *CV* and response of the simple detection algorithm (it doesn't take into account changes in CV).

As it can be seen from Fig. 3 the simple detection algorithm generates false alarms exactly during the step changes in *CV* even with the valve operating under normal condition, this phenomenon is due to the break of stationarity property in both *CV* and *X* time series.

In Fig. 4 it is shown the response of the modified detection algorithm where it is noticeable the improvement in the detection task. Although the stem displacement estimation is not very good, this doesn't represent a significant problem given that we are interested just in obtaining an on line average of the output signal.

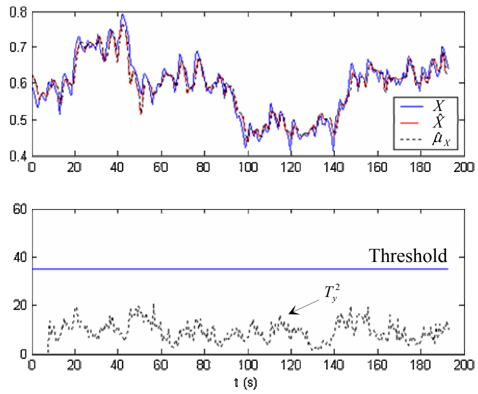

Fig. 4 Upper Plot: Plug displacement X, displacement estimation $\hat{X}$ and displacement estimation average $\hat{\mu}_X$. Lower Plot: Response of the modified detection algorithm under normal condition.

In the second part of the validation it will be evaluated the response of the modified algorithm under fault *f1-Valve Clogging* of magnitude 0.5 and 0.75. The fault appears at time 93 s. In the upper plot of Fig. 5 it is shown the behavior of displacement X under fault *f1* at different magnitudes. In the lower plot of Fig. 5 it is shown the detection capacity of the modified algorithm, where it was possible to detect both magnitude of faults.

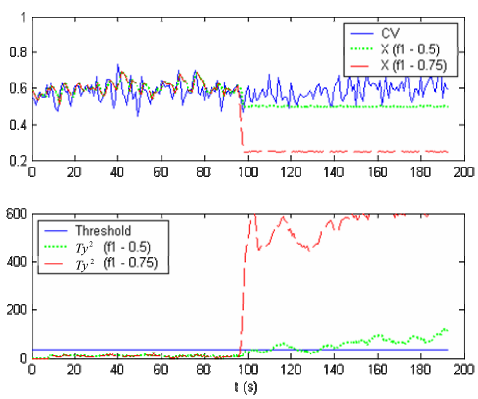

Fig. 5 Behavior of displacement X under *f1* at different magnitudes and fault detection

The last test consists of the presence of the fault *f1* of magnitude 0.5 at time 93 s, but before a step change in the input signal at 18 s.

As it is seen in Fig. 6 the modified detection algorithm is capable of make a distinction between normal changes in the signals and the authentic presence of a fault.

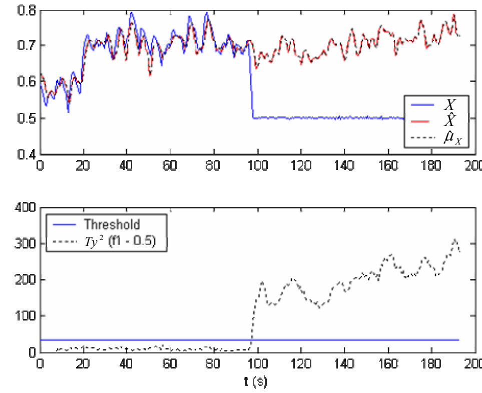

Fig. 6 Upper Plot: Plug displacement X, displacement estimation $\hat{X}$ and displacement estimation average $\hat{\mu}_X$. Lower Plot: Response of the modified detection algorithm under fault condition.

VI. Conclusion

Even when the DPCA can overcome the restriction of applying PCA on time correlated variables, however, due to the normal dynamic changes in the input and output signals the stationarity assumption is not satisfied. Here it has been proposed a modification to the algorithm for the DPCA based fault detection, in which it is carried out the appropriate standardization with regard to on-line averages of the input signal and the output signal estimation. This idea allows to deal with nonstationary signals and take advantage of the statistical multivariate approach of DPCA. It was shown through a series of tests the effectiveness of the fault detection modified algorithm to make a distinction between normal changes in signals and the variations due to the presence of a fault.

References

[1] R. Patton, P Frank R Clark. *Issues of fault Diagnosis for Dynamic Systems*. Springer, 2000.

[2] J. Korbicz, J. Koscielny et al. *Fault Diagnosis*, Springer, 2004.

[3] J. E. Jackson. *A Users Guide to Principal Components*. John Wiley and Sons, 1991.

[4] W. Ku, R. H. Storer and Ch. Georgakis, "Disturbance Detection and Isolation by Dynamic Principal Component Analysis", *Chemometrics and Intelligent Laboratory Systems*, vol. 30, pp. 179 – 196, 1995.

[5] Elsner and Tsonis, *Singular Spectrum Analysis: A New Tool in Time Series Analysis*. New York: Plenum Press, 1996, pp. 39 – 67.

[6] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, 3rd Ed. Prentice Hall, 1994.

[7] N. D. Tracy and J. C. Young, "Multivariate Control Charts for Individual Observations", *Journal of Quality Technology*, vol. 24, no. 2, pp. 88 – 95, 1992.

[8] M. Bartys and M. Syfert, *Using Damadics Actuator Benchmark Library (DABLib)*. In: *http ://diag. mchtr.pw, edu.pl /damadics/.*

[9] *Control Valve Dynamic Specification*, Version 3.0, 11/98, EnTech Control Engineering Inc.

IEEE Catalog Number: 05EX1097
ISBN: 0-7803-9230-2