International Conference on Control, Engineering & Information Technology (CEIT'13)
Proceedings Engineering & Technology - Vol.1, pp. 162-167, 2013


# Fault detection and isolation with Interval Principal Component Analysis

Anissa Benaicha
Research Unit ATSI
National Engineering School of Monastir
Rue Ibn El Jazzar, 5019 Monastir, Tunisia
Email: benaicha.anissa@gmail.com

Gilles Mourot
and José Ragot
Research Center for Automatic of Nancy
UMR 7039-University of Nancy
CNRS, 2, avenue of the Forêt of Haye
54516 Vandœuvre-Lès Nancy
Email: gilles.mourot@ensem.inpl-nancy.fr
Email: josé.ragot@ensem.inpl-nancy.fr

Kamel Benothman
Research Unit LARA-Automatic
National Engineering School of Tunis
BP 37, 1002 Tunis Belvédère, Tunisia
Email: kamelbenothman@yahoo.fr

***Abstract*—Diagnosis method based on Principal Component Analysis (PCA) has been widely developed. However, this method deals only with data which are described by single-valued variables. The purpose of the present paper is to generalize the diagnosis method to interval PCA. The fault detection is performed using the new indicator $[SPE]$. To identify the faulty variables, this work proposes a new method based on the reconstruction principle to this indicator. It aims to solve an interval linear system to obtain the reconstructed variables. The analysis of the reconstruction conditions permits to determine the useful directions. Then, the residuals structuring according to these directions allows to identify the set of faulty variables. This new diagnosis method based on interval PCA model is validated by a simulation example.**

***Index Terms*—interval PCA model, fault detection and isolation, variable reconstruction, interval linear system.**

## I. Introduction

Principal Component Analysis PCA is a widely used technique for sensor fault detection and isolation [1], [17], [16] and more generally for the detection of aberrant information [18]. PCA allows to develop an implicit model of the system and reveals linear relationships between its variables without making explicit model; this model can then be used to monitor the system behavior or its components. The works in the field of PCA are usually carried out under single-valued variables (the data have punctual values). However, in real situations, the use of single-valued variables is the result of a simplification and can cause a severe loss of information [15].

The diagnosis technique using PCA may be extended to data described by multivalued variables allowing to take into account the concepts of imprecision, variation and data confidence intervals. Several methods have been proposed for the extension of PCA to a set of interval type data [3], [4]. This generalization must first, ensure the main functions of the traditional PCA that aims to reduce the dimension and to extract the main data structure. In addition, it must return the information variation or the inaccuracy introduced by these variables [4]. The results provided by interval PCA must coincide with those from classic PCA. However, until now, these methods are not applied to diagnosis procedures.

This work presents a new method for fault detection and isolation from an uncertain PCA model of interval type. This paper is organized as follows. Section 2 is a brief reminder of the interval PCA. Section 3 is devoted to the extension of the indicator *SPE* [5], used in classical PCA, to detect fault in interval PCA case. Section 4 focuses on the identification of faulty variables using the reconstruction principle to the proposed indicator $[SPE]$. The reconstruction method aims to solve an interval linear system. Under certain conditions, the structuring of the useful interval residuals can, then, be used to isolate the faulty variables. The last section illustrates the application of the methodological results on data from a linear system whose measurements are corrupted by bounded errors of interval type and affected by faults.

## II. Interval PCA model

Given the analytical approach proposed by Ragot el al. [11], we can extend the PCA tools developed for singleton type data to an interval PCA model.

Once the number $\ell$ ( $\ell < m$) of components to retain in the interval PCA model is determined (from a threshold on the eigenvalues magnitude), by considering the interval eignenvectors partitioning:

$$[\mathbf{P}] = \left[\ [\hat{\mathbf{P}}] \mid [\tilde{\mathbf{P}}]\ \right], \quad [\hat{\mathbf{P}}] \in \mathbb{R}^{m\times\ell} \quad \text{and} \quad [\tilde{\mathbf{P}}] \in \mathbb{R}^{m\times(m-\ell)} \tag{1}$$

we define the following matrices:

$$[\hat{\mathbf{C}}] = [\hat{\mathbf{P}}]\,[\hat{\mathbf{P}}]^{\mathrm{T}} \tag{2}$$

$$[\tilde{\mathbf{C}}] = \mathbf{I}_m - [\hat{\mathbf{C}}] \tag{3}$$

which form the interval PCA of the process.

The diagonal terms $[\mathbf{c}_{ii}], i = 1\dots m$ of the matrix $\left[\hat{\mathbf{C}}\right]$ (2) are given by:

$$[\mathbf{c}_{ii}] = \sum_{k=1}^{\ell} [\mathbf{p}_{ik}]\,[\mathbf{p}_{ik}] \tag{4}$$

Given the properties of interval arithmetic [14], we have:

$$[\mathbf{p}_{ih}]\,[\mathbf{p}_{ih}] \supseteq [\mathbf{p}_{ih}]^2 \tag{5}$$

As we seek to obtain the tightest enclosure of $[\mathbf{c}_{ii}]$, these terms are, then, given by:

$$[\mathbf{c}_{ii}] = \sum_{h=1}^{\ell} [\mathbf{p}_{ih}]^2 \quad (6)$$

*a) Remark:* The idempotence property of the matrices $[\hat{\mathbf{C}}]$ and $[\tilde{\mathbf{C}}]$ is not verified in the interval PCA, we note:

$$\left[\tilde{\mathbf{H}}\right] = \left[\tilde{\mathbf{C}}\right]\left[\tilde{\mathbf{C}}\right] \quad (7)$$

Thus, the two matrices (2) and (3) divide the data space into two interval subspaces: the principal subspace spanned by the $\ell$ first interval eigenvectors and the residual subspace, spanned by the $m-\ell$ last interval eigenvectors. From this decomposition using interval type expressions, we can, then, propose a diagnosis procedure i.e of measurement detection reflecting an abnormal system behavior.

## III. Fault detection

Similar to the traditional PCA, the faults presence is performed through the detection indicators. In this section, we propose to extend the detection indicator *SPE* [5], used in the classical PCA, to detect the faults presence from a PCA model type interval.

The projection of the singleton measurements vector $\mathbf{x}(k)$ onto the residual subspace defined by the matrix $[\tilde{\mathbf{C}}]$ allows to define the interval residual vector:

$$[\tilde{\mathbf{x}}](k) = [\tilde{\mathbf{C}}]\mathbf{x}(k) \quad (8)$$

*Remark:* The measurements are considered here as singleton. However, the proposed procedure in the remainder of this work immediately extends to measurements represented by interval.

In the presence of faults affecting a subset $\mathbf{F}$ of variables, the singleton measurement vector is written as:

$$\mathbf{x}(k) = \underbrace{\mathbf{x}^0(k) + \delta\mathbf{x}^*(k)}_{=\mathbf{x}^*(k)} + \Xi_{\mathbf{F}}\,\mathbf{f}(k) \quad (9)$$

where $\mathbf{x}^0(k)$ denotes the true measurement vector, $\delta\mathbf{x}^*(k)$ the variation vector due to noise measurement assumed to be white, $\mathbf{x}^*(k)$ the fault-free measurement vector, $\mathbf{f}(k) \in \mathbb{R}^f$ ($f \geq 1$) the fault magnitude vector, and $\Xi_{\mathbf{F}} \in \mathbb{R}^{m\times f}$ the fault direction matrix which is unknown. This orthonormal matrix is formed with 0 to indicate a fault-free variable (respectively with 1 for a faulty variable).

Refer to (9), the interval residual vector (8) is written:

$$[\tilde{\mathbf{x}}](k) = \left[\tilde{\mathbf{C}}\right]\mathbf{x}^*(k) + \left[\tilde{\mathbf{C}}\right]\Xi_{\mathbf{F}}\mathbf{f}(k) \quad (10)$$

The generalized detection indicator $[SPE]$ is defined at instant $k$ by:

$$[SPE](k) = [\tilde{\mathbf{x}}]^{\mathrm{T}}(k)\,[\tilde{\mathbf{x}}](k) \quad (11)$$

Substituting (8) in (11) and taking into account the definition (7), we have:

$$[SPE](k) = \mathbf{x}^{\mathrm{T}}(k)\,[\tilde{\mathbf{H}}]\,\mathbf{x}(k) \quad (12)$$

To optimize the width of the indicator $[SPE](k)$ (12), its expression is given by:

$$[SPE](k) = \sum_{j=1}^{m} \left([\tilde{x}_j](k)\right)^2 \quad (13)$$

where $[\tilde{x}_j](k)$ is the $i^{th}$ component of $[\tilde{\mathbf{x}}](k)$ (10).

Given the properties of interval arithmetic [14], the system is declared in failure mode at instant $k$, if the lower bound of $[SPE](k)$ (13) over the value 0, conjointly if its upper bound exceeds its detection threshold adapted by training on nominal data.

## IV. Fault isolation by reconstruction

The detection phase must be completed by a phase fault isolation inorder to identify the faulty variables. The variables reconstruction and then their projection onto the residual subspace allows this identification. The simultaneous reconstruction of a set of $\mathbf{R}$ variables consits in estimating the variables of this set from the remaining variables $\overline{\mathbf{R}}$ and the PCA model by minimizing their influence on the detection index.

In this section, we try to establish the relation between the reconstruction of a set of variables from an interval PCA model and a remaining variables ( of singleton type) by minimizing their influence on the proposed detection indicator $[SPE]$. The reconstructed variables are obtained by solving an interval linear system. Under reconstruction conditions, the structured residuals analysis permits the identification of faulty variables.

### A. Variables reconstruction method

Consider, first, the partitioning according to the singleton measurement vector $\mathbf{x}(k)$:

$$\mathbf{x}(k) = \begin{bmatrix} \mathbf{x}_{\mathbf{R}}(k) & \mathbf{x}_{\overline{\mathbf{R}}}(k) \end{bmatrix}^{\mathrm{T}} \quad (14)$$

with

$$\mathbf{x}_{\mathbf{R}}(k) = \Xi_{\mathbf{R}}^{\mathrm{T}}\,\mathbf{x}(k) \quad (15)$$

$$\mathbf{x}_{\overline{\mathbf{R}}}(k) = \Xi_{\overline{\mathbf{R}}}^{\mathrm{T}}\,\mathbf{x}(k) \quad (16)$$

where $\mathbf{x}_{\mathbf{R}}(k)$ corresponds to the $r$ components of $\mathbf{x}(k)$ to be reconstructed ($r \geq 1$) and $\mathbf{x}_{\overline{\mathbf{R}}}(k)$ are the $m-r$ remaining components. The two matrices $\Xi_{\mathbf{R}}$ and $\Xi_{\overline{\mathbf{R}}}$ select respectively these two groups of variables.

The simultaneous reconstruction formula is obtained by solving the following optimization problem like in [5]:

$$[\hat{\mathbf{x}}_{\mathbf{R}}](k) = \arg\min_{\mathbf{x}_{\mathbf{R}}(k)} [SPE_{\mathbf{R}}](k) \quad (17)$$

where the index $[SPE_{\mathbf{R}}]$ corresponds to the detection index $[SPE]$ (12) obtained after the reconstruction of the $r$ variables of the set $\mathbf{R}$.

Given the partitioning of the matrix $[\tilde{\mathbf{C}}]$:

$$[\tilde{\mathbf{C}}] = \begin{bmatrix} [\tilde{\mathbf{C}}_{\mathbf{RR}}] & [\tilde{\mathbf{C}}_{\mathbf{R}\overline{\mathbf{R}}}] \\ [\tilde{\mathbf{C}}_{\mathbf{R}\overline{\mathbf{R}}}]^{\mathrm{T}} & [\tilde{\mathbf{C}}_{\overline{\mathbf{R}}\overline{\mathbf{R}}}] \end{bmatrix} \quad (18)$$

with

$$\begin{array}{lcl} [\tilde{\mathbf{C}}_{\mathbf{RR}}] & = & \Xi_{\mathbf{R}}^{\mathrm{T}}\,[\tilde{\mathbf{C}}]\,\Xi_{\mathbf{R}} \in \mathbb{IR}^{r\times r} \\ [\tilde{\mathbf{C}}_{\mathbf{R}\bar{\mathbf{R}}}] & = & \Xi_{\mathbf{R}}^{\mathrm{T}}\,[\tilde{\mathbf{C}}]\,\Xi_{\bar{\mathbf{R}}} \in \mathbb{IR}^{r\times(m-r)} \\ [\tilde{\mathbf{C}}_{\bar{\mathbf{R}}\bar{\mathbf{R}}}] & = & \Xi_{\bar{\mathbf{R}}}^{\mathrm{T}}\,[\tilde{\mathbf{C}}]\,\Xi_{\bar{\mathbf{R}}} \in \mathbb{IR}^{(m-r)\times(m-r)} \end{array} \tag{19}$$

where $\mathbb{IR}$ denotes the set of closed bounded intervals of $\mathbb{R}$. and the matrix definition $[\tilde{\mathbf{H}}]$ (7), the index $[SPE_{\mathbf{R}}]$ is defined at instant $k$ by:

$$[SPE_{\mathbf{R}}](k) = \left[\begin{array}{cc} \mathbf{x_R}(k)^{\mathrm{T}} & \mathbf{x}_{\bar{\mathbf{R}}}(k)^{\mathrm{T}} \end{array}\right] \left[\begin{array}{cc} [\tilde{\mathbf{H}}_{\mathbf{RR}}] & [\tilde{\mathbf{H}}_{\mathbf{R}\bar{\mathbf{R}}}] \\ [\tilde{\mathbf{H}}_{\mathbf{R}\bar{\mathbf{R}}}]^{\mathrm{T}} & [\tilde{\mathbf{H}}_{\bar{\mathbf{R}}\bar{\mathbf{R}}}] \end{array}\right] \left[\begin{array}{c} \mathbf{x_R}(k) \\ \mathbf{x}_{\bar{\mathbf{R}}}(k) \end{array}\right] \tag{20}$$

with

$$\begin{array}{lcl} [\tilde{\mathbf{H}}_{\mathbf{RR}}] & = & \Xi_{\mathbf{R}}^{\mathrm{T}}\,[\tilde{\mathbf{H}}]\,\Xi_{\mathbf{R}} \\ [\tilde{\mathbf{H}}_{\mathbf{R}\bar{\mathbf{R}}}] & = & \Xi_{\mathbf{R}}^{\mathrm{T}}\,[\tilde{\mathbf{H}}]\,\Xi_{\bar{\mathbf{R}}} \\ [\tilde{\mathbf{H}}_{\bar{\mathbf{R}}\bar{\mathbf{R}}}] & = & \Xi_{\bar{\mathbf{R}}}^{\mathrm{T}}\,[\tilde{\mathbf{H}}]\,\Xi_{\bar{\mathbf{R}}} \end{array} \tag{21}$$

We develop the $[SPE_{\mathbf{R}}](k)$ expression (20), then its minimizing over $\mathbf{x_R}(k)$ yields:

$$[\tilde{\mathbf{H}}_{\mathbf{RR}}]\,[\hat{\mathbf{x}}_{\mathbf{R}}](k) + [\tilde{\mathbf{H}}_{\mathbf{R}\bar{\mathbf{R}}}]\,\mathbf{x}_{\bar{\mathbf{R}}}(k) = 0 \tag{22}$$

The expression (22) corresponds at instant $k$ to a system of linear interval equations of the form:

$$[\mathbf{A}]\,[\mathbf{y}]\,(k) = [\mathbf{b}]\,(k) \tag{23}$$

with
$[\mathbf{A}] = \left[\tilde{\mathbf{H}}_{\mathbf{RR}}\right]$, $[\mathbf{b}]\,(k) = -\left[\tilde{\mathbf{H}}_{\mathbf{R}\bar{\mathbf{R}}}\right]\mathbf{x}_{\bar{\mathbf{R}}}(k)$ and $[\mathbf{y}]\,(k) = [\hat{\mathbf{x}}_{\mathbf{R}}]\,(k)$ is the interval vector to be estimated.
Note that the matrix $[\mathbf{A}]$ is symmetric since $[\tilde{\mathbf{H}}_{\mathbf{RR}}]^{\mathrm{T}} = [\tilde{\mathbf{H}}_{\mathbf{RR}}]$.
There are various methods for solving interval linear system [14], [10], [8]. They tend to find a tight enclosure of the interval vector $[\mathbf{y}]\,(k)$ which contains the true solution set. [9] and [6] proved in the case where the matrix $[\mathbf{A}]$ is symmetric, the interest of using the approaches taking into account this property compared to the application of other methods in terms of optimized solution. In this context, there are: the $[\mathbf{L}]\,[\mathbf{D}]\,[\mathbf{L}]^{\mathrm{T}}$ decomposition and the Cholesky method adapted to interval data [2]. Recently, [7] proposes an iterative method that reduces significantly the overestimation. However, it does not necessarily converge to the optimal solution and requires a high computational cost especially for a large system. Then, the $[\mathbf{L}]\,[\mathbf{D}]\,[\mathbf{L}]^{\mathrm{T}}$ decomposition will be used here because it provides the tight enclosure of the solution and it is simple to apply.

*Remark:* From (22), we notice that the reconstruction of a set of $\mathbf{R}$ variables consists in estimating these variables from the set of remaining variables $\overline{\mathbf{R}}$ and the interval PCA model. Thus, if the set $\mathbf{R}$ is faulty, its reconstruction provides an independently estimated of faults.

### *B. Generation of interval structured residuals*

Once the vector $[\hat{\mathbf{x}}_{\mathbf{R}}](k)$ is determined, the reconstructed measurement vector is:

$$[\hat{\mathbf{z}}_{\mathbf{R}}](k) = \left[\begin{array}{cc} [\hat{\mathbf{x}}_{\mathbf{R}}](k) & \mathbf{x}_{\bar{\mathbf{R}}}(k) \end{array}\right]^{\mathrm{T}} \tag{24}$$

The structured residuals are defined with respect to the reconstruction directions $\Xi_{\mathbf{R}}$ as the projection of the reconstructed measurement vector $[\hat{\mathbf{z}}_{\mathbf{R}}](k)$ onto the residual subspace:

$$[\tilde{\mathbf{z}}_{\mathbf{R}}](k) = \left[\tilde{\mathbf{C}}\right]\,[\hat{\mathbf{z}}_{\mathbf{R}}]\,(k) \tag{25}$$

Given (24) and the matrix partitioning (18), the interval residual vector (25) is:

$$[\tilde{\mathbf{z}}_{\mathbf{R}}](k) = \left[\begin{array}{c} \left[\tilde{\mathbf{C}}_{\mathbf{RR}}\right][\hat{\mathbf{x}}_{\mathbf{R}}]\,(k) + \left[\tilde{\mathbf{C}}_{\mathbf{R}\bar{\mathbf{R}}}\right]\mathbf{x}_{\bar{\mathbf{R}}}(k) \\ \left[\tilde{\mathbf{C}}_{\mathbf{R}\bar{\mathbf{R}}}\right]^{\mathrm{T}}[\hat{\mathbf{x}}_{\mathbf{R}}]\,(k) + \left[\tilde{\mathbf{C}}_{\bar{\mathbf{R}}\bar{\mathbf{R}}}\right]\mathbf{x}_{\bar{\mathbf{R}}}(k) \end{array}\right] \tag{26}$$

The expression (26) is interesting because it allows to maintain the structural properties of the structured residuals. First, the envelopes of the $r$ first components of the structured residuals contain 0 and then the $m-r$ last components of this vector are independent of the reconstructed variables. This indicates, clearly, the interest of this expression for fault isolation as it is possible to generate a set of structured residuals insensitive to certain variables.

The analysis of the indicator $[SPE_{\mathbf{R}}](k)$ (20) computed from the structured residuals $[\tilde{\mathbf{z}}_{\mathbf{R}}](k)$ (25) allows the identification of the set of faulty variables. To obtain the tight enclosure of this indicator, it is calculated as follows:

$$[SPE_{\mathbf{R}}]\,(k) = \sum_{j=1}^{m}\left(\left[\tilde{z}_{\mathbf{R}}^{j}\right](k)\right)^{2} \tag{27}$$

where $\left[\tilde{z}_{\mathbf{R}}^{j}\right](k)$ is the $i^{th}$ component of $[\tilde{\mathbf{z}}_{\mathbf{R}}](k)$.

The faults influence is eliminated on this indicator (27) if the reconstruction direction $\Xi_{\mathbf{R}}$ corresponds to the faults direction $\Xi_{\mathbf{F}}$ i.e, if $\mathbf{R} = \mathbf{F}$.
As the diagnosis purpose is to determine the fault directions that are unknown a priori, the model structure analysis is needed to reduce the number of fault directions to be considered.

### *C. Reconstruction conditions*

The existence of the resolution methods requires that the matrix $[\mathbf{A}]$ is invertible. This condition implies that this matrix must satisfy [13]:

$$\sigma_{\max}\left(\left|(\mathbf{A}_c)^{-1}\right|\Delta\right) < 1 \tag{28}$$

where $\sigma_{\max}(*)$ is the maximum singular value of $*$, the matrices $\mathbf{A}_c$ and $\Delta$ represent respectively the midpoint and the radius matrices of $[\mathbf{A}]$.
The inequality (28) is equivalent to the following condition provided by [12]:

$$\sigma_{\max}(\Delta) < \sigma_{\min}(\mathbf{A}_c) \tag{29}$$

where $\sigma_{\min}(*)$ is the minimum singular value of $*$.
We define the following reconstruction ratio:

$$\mathbf{R_R} = \frac{\sigma_{\max}(\Delta)}{\sigma_{\min}(\mathbf{A}_c)} \tag{30}$$

From (29), plus the ratio $\mathbf{R_R}$ is less than 1, the more we guarantee a good reconstruction of the set of $\mathbf{R}$ variables, and

vice versa. In the particular case where the data are described by certain variables $\sigma_{\max}(\Delta) = 0$ and therefore $\mathbf{R_R} = 0$.
This ratio is used to reduce the number of combinations of the useful variables. The variable directions whose ratio (30) is greater than or close to 1 will be eliminated from the set $\mathbf{R}$ of variables to be reconstructed.
The second condition of reconstruction concerns the matrix $[\tilde{\Xi}_{\mathbf{R}}] = [\tilde{\mathbf{C}}]\,\Xi_{\mathbf{R}}$ which must be of full column rank $r$. This condition implies that:
- the columns of the matrix $[\tilde{\Xi}_{\mathbf{R}}]$ are not in this form:

$$[\tilde{\Xi}_{\mathbf{R}}] = \begin{bmatrix} [\tilde{\Xi}_{\mathbf{R}_1}] & a\,[\tilde{\xi}_r] & [\tilde{\xi}_r] \end{bmatrix} \tag{31}$$

where $a$ is a nonzero scalar.
which means that some or all of the columns of this matrix shouldn't are collinear (or close to collinearity). In this case, the matrix $[\tilde{\mathbf{H}}_{\mathbf{RR}}]$ is written:

$$[\tilde{\mathbf{H}}_{\mathbf{RR}}] = \begin{bmatrix} [\tilde{\Xi}_{\mathbf{R}_1}]^{\mathrm{T}}[\tilde{\Xi}_{\mathbf{R}_1}] & a\,[\tilde{\Xi}_{\mathbf{R}_1}]^{\mathrm{T}}[\tilde{\xi}_r] & [\tilde{\Xi}_{\mathbf{R}_1}]^{\mathrm{T}}[\tilde{\xi}_r] \\ a\,[\tilde{\Xi}_{\mathbf{R}_1}]^{\mathrm{T}}[\tilde{\xi}_r] & a^2\,[\tilde{\xi}_r]^{\mathrm{T}}[\tilde{\xi}_r] & a\,[\tilde{\xi}_r]^{T}[\tilde{\xi}_r] \\ [\tilde{\Xi}_{\mathbf{R}_1}]^{\mathrm{T}}[\tilde{\xi}_r] & a\,[\tilde{\xi}_r]^{\mathrm{T}}[\tilde{\xi}_r] & [\tilde{\xi}_r]^{\mathrm{T}}[\tilde{\xi}_r] \end{bmatrix} \tag{32}$$

The matrix (32) is not invertible and therefore the system (22) is not resolvable.
- The number $r$ of variables to be reconstructed must satisfy the following condition:

$$r \leq m - \ell \tag{33}$$

The model structure analysis in terms of isolation by reconstruction permits to reduce the number of scenarios to be considered and therefore to determine the isolable faults.

## V. Application

To illustrate the proposed methods presented above, we consider a system governed by 5 variables and described at different instants $k$ by the following equations:

$$\begin{cases} x_1^0(k) = 0.4\nu_1(k) - 1.2\sin(k/2)cos(k/4)exp(-k/2N) \\ x_2^0(k) = 0.8\nu_2(k) - 1.5sin(k/5)^2 \\ x_3^0(k) = x_1^0(k) + x_2^0(k) \\ x_4^0(k) = x_3^0(k) + x_2^0(k) \\ x_5^0(k) = 2x_1^0(k) + x_3^0(k) \\ \nu_1(k) \sim \eta(0,\, \sigma^2) \\ \nu_2(k) \sim \eta(0,\, \sigma^2) \end{cases} \tag{34}$$

This data set, generated from two variables from two normal distributions, show three linear analytical redundancy relations between the variables $x_i^0, i = 1, \ldots, 5$. To the generated data, which form the matrix $\mathbf{X}^0$, were superimposed the variation $\delta\mathbf{X}^*$. This variation are considered as realizations of centered random variables to simulate the presence of measurement noise. The system is simulated a first time for $N = 100$ observations. The application of the approximative method proposed by [11] permits to find the interval eigenvalues matrix $[\Lambda]$ and the interval eigenvectors matrix $[\mathbf{P}]$. Table I gives the interval eigenvalues thus obtained. Given the magnitude order of the

TABLE I
MATRIX $[\Lambda]$

| |
|---|
| [560.817 579.300] |
| [319.857 326.875] |
| [0.044 0.132] |
| [0.080 0.150] |
| [0.059 0.150] |

interval eigenvalues, the number of components to retain in the interval PCA model is $\ell = 2$. The matrix $[\Lambda]$ is then decomposed into two matrices $[\hat{\Lambda}]$ and $[\tilde{\Lambda}]$ and the matrix $[\mathbf{P}]$ is decomposed into $[\hat{\mathbf{P}}]$ and $[\tilde{\mathbf{P}}]$. The matrix $[\tilde{\mathbf{P}}]^{\mathrm{T}}$ is presented in table II. Their interval column vectors should be aproximated to the redundancy equations used to generate the data. Indeed, the redundancy equations (34) can explain the variables $x_1^0$, $x_2^0$ and $x_3^0$ as follows:

$$\begin{cases} x_1^0(k) = -\frac{1}{5}x_4^0(k) + \frac{2}{5}x_5^0(k) \\ x_2^0(k) = \frac{3}{5}x_4^0(k) - \frac{1}{5}x_5^0(k) \\ x_3^0(k) = \frac{2}{5}x_4^0(k) + \frac{1}{5}x_5^0(k) \end{cases} \tag{35}$$

The joint review of these relations and of the interval eigenvectors of table II shows the consistency of the estimates of these eigenvectors.

TABLE II
MATRIX $[\tilde{\mathbf{P}}]^{\mathrm{T}}$

| | | | | |
|---|---|---|---|---|
| 0 | 0 | 1 | [-0.405 -0.393] | [-0.204 -0.194] |
| 0 | 1 | 0 | [-0.607 -0.591] | [0.194 0.204] |
| 1 | 0 | 0 | [0.194 0.206] | [-0.404 -0.395] |

### A. Fault detection

Three faults are simultaneously added to the variables $x_1$, $x_2$ and $x_3$ from the samples 40 to 80. These faults are represented by a constant bias of magnitude equal to 10% of the variation range of each variable.

Figure 1 presents the evolution of the indicator $[SPE](k)$ (13) for 100 samples. Note that this indicator shows the presence of faults in the instants where the faults are introduced. Indeed, in this interval time, the lower bound of $[SPE](k)$ over 0 and its upper bound exceeds its detection threshold adapted by training on nominal data.

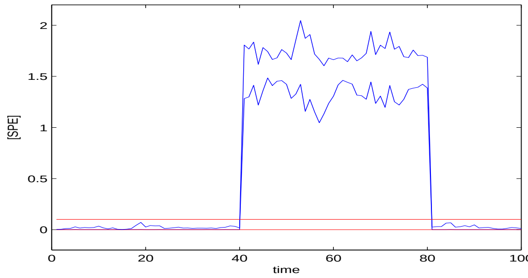


Fig. 1. Evolution of the detection index $[SPE](k)$

### B. Fault isolation by reconstruction

Refer to (33), the maximum number of simultaneous reconstructed variables is 3. The maximum number of reconstructions is 21. Table III presents the values of $\mathbf{R_R}$ (30) for the set of combinations of a variable i.e for $r = 1$. As the different values are less than 1, then the set of faults that appear on a single variable are isolable. The same computation is performed for $r = 2$. The results are explained in table IV. The analysis of this table reveals that the set of faults affecting two variables are isolable. For the set of directions of 3 variables, the ratio $\mathbf{R_R}$ is calculated. Table V presents their values according to possible combinations of 3 variables. The ratios $\mathbf{R}_{1,2,5}$, $\mathbf{R}_{2,3,4}$ and $\mathbf{R}_{3,4,5}$ are higher than 1. Thus, the couples of variables associated with them ($\{1,2,5\}$, $\{2,3,4\}$ and $\{3,4,5\}$) can not be reconstructed. The required number of reconstructions is reduced to 18.

After determining the useful reconstruction directions, the indicators $[SPE_\mathbf{R}]$ associated to these directions are constructed in order to isolate faulty variables.

As the number of faulty variables is a priori unknown, let's start with the calculation of indicators $[SPE_\mathbf{R}]$ for the set of combinations consisting of a variable. Figure 2 presents the evolution of these indicators. The first graph of this figure is relative to the residuals projection with reconstruction with the indicator $[SPE]$ without using the first variable. The second graph relative to the indicator $[SPE_2]$ corresponds to the indicator $[SPE]$ without using the variable $x_2$,... Note that the five graphs of this figure are sensitive to faults since the instant 40 to 80. The elimination of one variable does not eliminate the fault effect and therefore the fault is not simple. Now, we reconstruct two variables simultaneously. Figure 3 traces the evolution of indicators $[SPE_\mathbf{R}]$ obtained by variables reconstruction corresponding to the set of combinations of 2 faults. The various indicators presented in this figure are above their respective threshold during the time where the faults appear. Thus, the reconstruction of 2 variables doesn't eliminate the faults presence. Figure 4 shows the evolution of indicators $[SPE_\mathbf{R}]$ corresponding to possible combinations consisting of 3 faults. The first graph of this figure relative to the indicator $[SPE]_{1,2,4}$ (corresponding to the indicator $[SPE]$ after the reconstruction of variables $x_1$, $x_2$ and $x_4$) is sensitive to faults between the instants 40 and 80. The second graph plots the evolution of the indicator $[SPE]_{2,3,5}$. By comparing its lower bound to the value 0, we have a difficulty to identify the presence of faults. Despite, the comparison of the values of its upper bound to its threshold shows the faults presence between times 40 and 80. Thus, the use of the lower bound gives a much clearer indication and therefore removes any ambiguity as to the faults isolation. Only the simultaneous reconstruction of the contaminated variables $x_1$, $x_2$ and $x_3$ eliminates the effect of faults. The evolution of the resulting detection indicator $[SPE_{1,2,3}]$, introduced by the last graph of figure 4, confirms this result since it is insensitive to faults.

Thus, from this example, we correctly identify the set of faulty variables despite the presence of false alarms. In fact, the indicators $[SPE_\mathbf{R}]$ having a ratio $\mathbf{R_R}$ close to 1 are tainted by false alarms that can lead to false isolation. We have mentioned that the difficulty of identification by reconstruction in the interval case is directly related to this ratio. Thus, to avoid the risk of false isolation, the set $\mathbf{R}$ of variables to be reconstructed must have the ratio $\mathbf{R_R}$ well below 1.

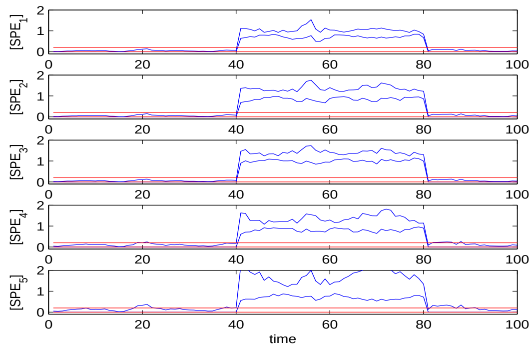


Fig. 2. Evolution of the indicators $[SPE_{\mathbf{R}}](k)$ for $r = 1$

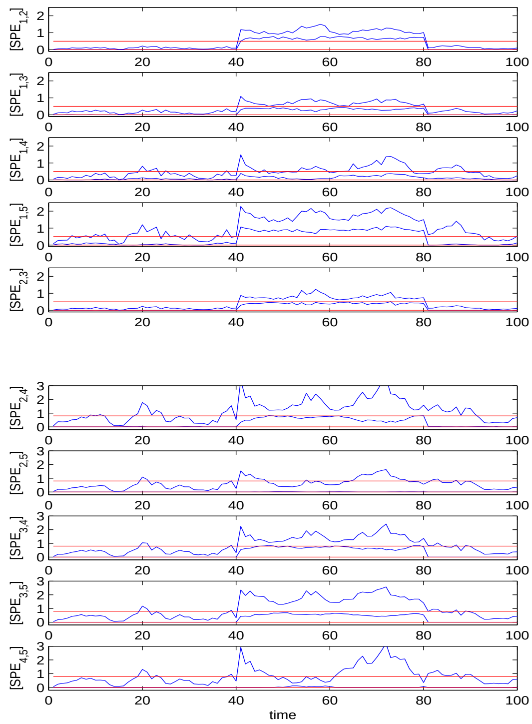


Fig. 3. Evolution of the indicators $[SPE_\mathbf{R}](k)$ for $r = 2$

## VI. Conclusion

This work is devoted to the generalization of diagnosis method to interval PCA. To detect the faults, we proposed to extend the indicator *SPE* used in the classical PCA. Concerning the identification of the set of faulty variables, a new method based on the reconstruction principle to the proposed

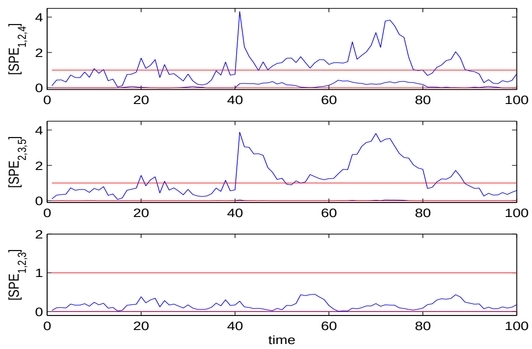


Fig. 4. Evolution of the indicators $[SPE_{\mathbf{R}}](k)$ for $r = 3$

indicator $[SPE]$ is proposed. The reconstructed variables are obtained by solving an interval linear system. The $[\mathbf{L}][\mathbf{D}][\mathbf{L}]^T$ decomposition is used because it provides the tight enclosure of the solution. The reconstruction conditions analysis permits to reduce the number of scenarios related to multiple faults to be considered. The residuals structuring according to these useful directions allows the identification of the implicated variables. The results of a simulation example showed the interest of the new method of diagnosis based on interval PCA. Indeed, it has allowed to identify correctly the set of faulty variables despite the presence of false alarms related to the reconstruction ratio $\mathbf{R_R}$.

TABLE III
VALUES OF $\mathbf{R_R}$ FOR $r = 1$

| **R** | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $\mathbf{R_R}$ | 0.021 | 0.024 | 0.024 | 0.068 | 0.088 |

TABLE IV
VALUES OF $\mathbf{R_R}$ FOR $r = 2$

| | | | | | |
|---|---|---|---|---|---|
| **R** | {1,2} | {1,3} | {1,4} | {1,5} | {2,3} |
| $\mathbf{R_R}$ | 0.056 | 0.047 | 0.136 | 0.571 | 0.059 |
| **R** | {2,4} | {2,5} | {3,4} | {3,5} | {4,5} |
| $\mathbf{R_R}$ | 0.339 | 0.233 | 0.2 | 0.286 | 0.254 |

TABLE V
VALUES OF $\mathbf{R_R}$ FOR $r = 3$

| **R** | {1,2,3} | {1,2,4} | {1,2,5} | {2,3,4} | {2,3,5} | {3,4,5} |
|---|---|---|---|---|---|---|
| $\mathbf{R_R}$ | 0.093 | 0.666 | 2.064 | 2.956 | 0.474 | 1.052 |

## REFERENCES


[1] C. F. Alcala, and S. J. Qin Reconstruction-based contribution for process monitoring. *Automatica*, volume 45, pages 1593–1600, 2009.

[2] G. Alefeld, and G. Mayer. The cholesky method for interval data. *Linear Algebra Application*, volume 194, pages 161–182, 1993.

[3] P. Cazes, A. Chouakria, E. Diday, and Y. Schektman. Extension de l'analyse en composantes principales à des données de type intervalle. *Revue de la statistique appliquée*, tome 45, nř3, pages 5–24, 1997.

[4] A. Chouakria. Extension des méthodes d'analyse factorielle à des données de type intervalle. *Thèse de Doctorat à l'Université Paris IXDauphine*, 06 Septembre 1998.

[5] R. Dunia, and S. J. Qin. A subspace approach to multidimensional fault identification and reconstruction. *American Institute of Chemical Engineers Journal*, volume 44, pages 1813–1831, 1998.

[6] G. H. Golub, and C. F. Van Loan. Matrix Computations. *London: The Johns Hopkins University Press*, 1996.

[7] M. Hladik. A contractor for the symmetric solution set. *World Academy of Science. Engineering and Technology*, 2010.

[8] E. Hansen, and W. G. Walster. Global optimization using interval analysis. *Revised and Expanded*. second edition, New York, 2004.

[9] C. Jansson, and H. Harburg. Interval linear systems with symmetric matrices, skew-symmetric matrices and dependencies in the right hand side. *Computing*, volume 46, pages 265–274, 1991.

[10] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter. Applied interval analysis. *Springer Verlag*, 2001.

[11] J. Ragot, G. Mourot, A. Ben Aicha, and K. Benothman. Application de l'ACP par intervalles au diagnostic. *QUALITA*, Mars 2011.

[12] S. M. Rump. Verification methods for dense and sparse systems of equations. In J. Herzberger, editor,*Topics in Validated Computations*, pages 63–135, Amsterdam, 1994.

[13] J. Rohn. A Handbook of results on interval linear problems. *Czech Academy of Sciences Prague*, Czech Republic, European Union www.cs.cas.cz/rohn, 2010.

[14] A. Neumaier. Interval methods for systems of equations. *Cambridge University Press*, 1990.

[15] F. Palumbo, and M. Gettler-Summa. Symbolic Interpretation of PCA on Interval Data. *Actes de la XL réunion scientifiques de la Societé Italienne de statistique*, pages 103–114, France, 26-28 Avril 2000.

[16] S. J. Qin. Statistical process monitoring: basics and beyond. *Journal of Chemometrics*, volume 17, pages 480–501, 2003.

[17] M. Tamura, and S. Tsujita A study on the number of principal components and sensitivity of fault detection using PCA. *Journal of Computers and chemical Engineering*, volume 31, pages 1035–1046, 2007.

[18] Y. Tharrault, G. Mourot, J. Ragot, and D. Maquin. Fault detection and isolation with robust principal component analysis. *International Journal of Applied Mathematics and Computer Science AMCS*, volume 18, 2008.