Incipient fault amplitude estimation using KL divergence with a probabilistic approach

Jinane Harmouche, Claude Delpha, Demba Diallo

Incipient Fault Amplitude Estimation using KL Divergence with a Probabilistic Approach

Jinane HARMOUCHE[a,b], Claude DELPHA[a], Demba DIALLO[b]

[a]*Laboratoire des Signaux et Systèmes (L2S), Univ. Paris Sud, CNRS, CentraleSupélec, 91192 Gif Sur Yvette, France*
[b]*Laboratoire Génie Electrique et Electronique de Paris (GEEPS), CNRS, CentraleSupélec, Univ. Paris Sud, UPMC, 91192 Gif Sur Yvette, France*

Abstract

The Kullback-Leibler (KL) divergence is at the centre of Information Theory and change detection. It is characterised with a high sensitivity to incipient faults that cause unpredictable small changes in the process measurements. This work yields an analytical model based on the KL divergence to estimate the incipient fault magnitude in multivariate processes. In practice, the divergence has no closed form and it must be numerically approximated. In the particular case of incipient fault, the numerical approximation of the divergence causes many false alarms and missed detections because of the slight effect of the incipient fault. In this paper, the ability and relevance to estimate the incipient fault amplitude using the numerical divergence is studied. The divergence is approximated through the calculation of discrete probabilities for faultless and faulty signals. The estimation results that are obtained by simulation induce an error lower than 1% on the fault amplitude.

Keywords: Fault Estimation, Kullback-Leibler Divergence, Principal Component Analysis

1. Introduction

The last three decades have shown an increased demand for improving the economy and safety of industrial processes. Health monitoring of such processes has been widely developed with studies of fault detection and diagnosis (FDD). Early detection and severity assessment of imperceptible faults are main functions of fault detection [1]. Measurements are basic representation of process behaviour, and faults in general manifest themselves as changes in their properties. The detection of a particular fault is based on checking whether the current measurements are statistically different from the *a priori* known faultless measurements. Detection indices with control charts are designed to this end [2–4]. The MEWMA (Multivariate Exponentially Weighted Moving Average) and the MCUSUM (Multivariate Cumulative Sum) are able to detect deviations related to the process mean vector [5]. The MEWMA-CM (MEWMA-Covariance Matrix) is used to detect changes in the process covariance matrix [6]. The simultaneous monitoring of the mean and the variance in an univariate framework has been presented in [7]. It was not extended to the multivariate framework due to the complexity of multivariate probability distributions. Statistical multivariate techniques, among

Email addresses: Jinane.Harmouche@supelec.fr (Jinane HARMOUCHE), claude.delpha@l2s.centralesupelec.fr (Claude DELPHA), demba.diallo@geeps.centralesupelec.fr (Demba DIALLO)

which the Principal Component Analysis (PCA) is a major component, are effective in the FDD of high dimensional processes [8]. PCA is optimal in terms of capturing variability in the data and constitutes a general framework for data representation and modelling [9]. It has been used for monitoring in a wide range of applications, including chemical processes [10], aerospace [11, 12], electronics [13], automotive [14], semi-conductors [15], and many others.

The KL divergence has been proposed in the PCA framework to be a general fault indicator which is characterised by high sensitivity with respect to incipient faults (the short duration change whose amplitude is less than 10% of the signal magnitude) [16]. It has been used as a distribution-free control chart that makes no assumption about the form of the process distribution. Consequently, its calculation requires the availability of a training sample of observations from which the reference (fault-free) empirical probability distribution can be computed. It showed superior efficiency in the detection of incipient faults, compared to the fault indices that are commonly used with PCA, namely the Hotelling $\mathbf{T^2}$ statistic and the squared prediction error (SPE). Beside the fault detection, the fault estimation problem has gained considerable attention in recent years. If a fault in the process measurements has been detected and the information contained into the data is important, it is necessary to retrieve the fault-free measurements from the faulty ones [17–19]. Sensor validation and correction is concerned with the problem of identifying the fault magnitude in order to retrieve the sensor response from faulty sensor data [20, 21]. In a system under fault tolerant control (FTC), whenever a fault is detected, the fault amplitude is estimated in order to compensate its effect through an appropriate reconfiguration of the controller module [22, 23]. The performance of the FTC system depends mainly on the estimation accuracy of the fault magnitude. As for fault detection, it is desirable for the fault estimation to be robust with respect to noises and unexpected uncertainties and perturbations.

Most fault estimation approaches are optimisation-based, and thus optimisation techniques are used to solve the fault estimation problem [24]. This paper looks into the problem of estimating faults using the proposed PCA-based KL divergence approach. The divergence that is numerically approximated to make the fault detection [16] will be used to estimate the incipient fault amplitude under the particular assumption of normal distribution. A theoretical analysis leading to an estimate of the incipient fault amplitude is described. The KL divergence has an analytical form in case of normal distributions. The numerical approximation of the divergence degrades the detection performance and affects the fault estimation accuracy especially for incipient faults. Therefore, the evaluation of the fault estimation accuracy is carried out in this paper. The probability density function (pdf) of the obtained fault amplitude estimate is calculated. The probabilistic model is validated and the relative error of estimation is assessed, through an AR process model.

2. Analytical approach to estimate fault amplitude

2.1. Main Notations

The following notations will be used in the overall paper.
Let's consider $X_{[N\times m]}$ the data matrix of m variables.

$X = (\boldsymbol{x_1}, ..., \boldsymbol{x_j}, ..., \boldsymbol{x_m}) = (x_{ij})_{i,j}$, where $\boldsymbol{x_j} = [x_{1j}...x_{Nj}]^{\mathrm{T}}$ is a vector of N observations acquired from the jth variable. For statistical significance, $m \geq 2$ and $N >> m$ [9]. Let $\bar{X}_{[N\times m]}$, where $\bar{X} = (\boldsymbol{\bar{x}_1}, ..., \boldsymbol{\bar{x}_j}, ..., \boldsymbol{\bar{x}_m})$, be the corresponding centered matrix.

S is the sample covariance matrix and l is the dimension of the principal subspace. Many criteria have been proposed in the literature to get the best choice of l. Authors in [25] compared 11 methods to determine l and concluded that minimizing the Variance of Reconstruction Error (VRE) is preferable.

$P_{[m\times m]}$, such as $P = (\boldsymbol{p_1}, ..., \boldsymbol{p_l}, ...\boldsymbol{p_m})$, is the matrix of eigenvectors of S associated to $\lambda_1, ..., \lambda_l, ..., \lambda_m$.

g denotes the fault amplitude and $\boldsymbol{x_j}$ is the faulty variable.

The star mark (*) refers to faultless and noise-free data and the superscript 'rf' refers to reference faultless data.

2.2. Assumptions

The analytical model of the KL divergence depending on the fault characteristics is obtained based on the following assumptions on the fault and data modeling:

1. Fault modeling: an incipient fault is often defined as a change or a degradation that develops slowly [26]. The fault model adopted here assumes that during the first stage of the incipient fault development, the fault amplitude (size or severity) is constant, see Fig.1. It is a gain fault characterised with a multiplicative factor with amplitude g that affects the last $(N - b)$ observations of the signal. b is the time occurrence of the fault.
2. Noise modeling: the process variables are affected with independent and identically distributed (i.i.d) Gaussian noise that represents measurement errors. The noise samples are considered to be drawn from a normal distribution with zero mean and variance σ_v^2. The noise variance is supposed to be not affected by the fault, since it is not a process noise but rather an environmental or measurements nuisance.
3. Assumption of normality for PCA data : the initial data distributions along the original axes are assumed as Gaussian. The principal components, which are linear combinations of the original variables, will be thus normally distributed. This assumption can usually take place, because basically PCA yields an optimal representation for approximately multivariate normal data. For this case, the principal subspace is spanned by the first l eigenvectors of the sample covariance matrix leading to the maximum variance representation of the dataset.

As a consequence, let $\boldsymbol{V}_{[N\times 1]}$ be a noise vector of N samples drawn from the distribution $\mathcal{N}(0, \sigma_v^2)$. We can write for $\boldsymbol{x_j}$:

$$\boldsymbol{x_j} = \boldsymbol{x_j^*} + \boldsymbol{F_j} + \boldsymbol{V} \quad (1)$$

where

$$\boldsymbol{F_j} = g \times [\ 0 \quad \ldots \quad 0 \quad x_{bj}^* \quad \ldots \quad x_{Nj}^*\ ]^{\mathrm{T}} \quad (2)$$

The theoretical study concerns incipient faults, and thus the fault characteristics are considered quite small according to the signal characteristics. g is a near-zero unknown constant and it introduces small amplitude variations on x_j.

The result is that these variations will not change the centre and the direction of the PCA's model. It has been shown in [27] that the direction of the first few principal components would not change following the occurrence of a small fault. The direction of the last principal components should be monitored in such case. In our case of incipient fault, the covariance matrix S can be written as:

$$S = P^* \Lambda P^{*\mathrm{T}} + \sigma_v^2 I_m \tag{3}$$

where

$$\Lambda = \Lambda^* + \Delta\Lambda \tag{4}$$

$\Lambda^* = diag(\lambda_1^*, ..., \lambda_l^*, 0, .., 0)$ is the $(m \times m)$ matrix of eigenvalues associated to eigenvectors $\boldsymbol{p_1^*}, ..., \boldsymbol{p_l^*}, \boldsymbol{p_{l+1}^*}, ..., \boldsymbol{p_m^*}$. $\Delta\Lambda = diag(\Delta\lambda_1, ..., \Delta\lambda_m)$ is the change due to the fault occurrence and I_m is the m-identity matrix. So, $\Delta\Lambda = 0$ when $g = 0$. The last $(m - l)$ eigenvalues correspond to the residual subspace. Since, in practice, the distributions of the last principal components may vanish ($\lambda_k^* = 0$) and their directions may change, the divergence is only concerned with the first l principal components, for which $\lambda_k^* \neq 0$ ($k = 1, ..., l$).

2.3. Definition

The KL divergence is an instance of f-divergence family, which has been used in many signal processing applications including anomaly detection [28, 29], pattern recognition [30] and fault diagnosis [31, 32]. For discrimination between two continuous probability distribution functions $f(x)$ and $h(x)$ of a random variable x, the Kullback-Leibler Information is defined in [36] as:

$$I(f \,||\, h) = \int f(x) \log \frac{f(x)}{h(x)} dx. \tag{5}$$

The Kullback-Leibler divergence is the symmetric version of the Information [34], and it is given by $D(f,h) = I(f \,||\, h) + I(h \,||\, f)$. If f and h are normal densities such that $f \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $h \sim \mathcal{N}(\mu_2, \sigma_2^2)$, where μ_1, μ_2 are the means and σ_1^2, σ_2^2 are the variances for f and h respectively, the divergence is reduced to [34]

$$D(f,h) = \frac{1}{2}\left[(\frac{1}{\sigma_2^2} + \frac{1}{\sigma_1^2})(\mu_1 - \mu_2)^2 + \frac{\sigma_1^2}{\sigma_2^2} + \frac{\sigma_2^2}{\sigma_1^2} - 2 \right]. \tag{6}$$

2.4. Fault amplitude estimation

From the assumption of normality, it follows that each of the first l principal components has a pdf denoted f_k such that $f_k \sim \mathcal{N}(0, \lambda_k + \sigma_v^2)$. It is proposed to compare f_k against its reference f_k^{rf}. $f_k^{rf} \sim \mathcal{N}(0, \lambda_k^* + \sigma_v^2)$. The fault does not affect the mean parameter of the distributions because the centre of the PCA's model is supposed unchanged after the fault occurrence. It follows from (4):

$$\lambda_k = \lambda_k^* + \Delta\lambda_k \tag{7}$$

Specializing (6) to the case considered as detailed in [16] gives:

$$D(f_k^{rf}, f_k) = \frac{1}{2}[\frac{\Delta\lambda_k^2}{(\lambda_k^* + \sigma_v^2)(\lambda_k^* + \sigma_v^2 + \Delta\lambda_k)}]. \quad (8)$$

With this symmetric version of the divergence, the fault estimation is unbiased.

The next step is then to write $\Delta\lambda_k$ in function of the fault amplitude g.

Suppose λ_k is a function of g and is infinitely differentiable in the neighborhood of zero ($g \approx 0$), the Taylor development of λ_k gives:

$$\lambda_k = \lambda_k^* + \frac{\partial\lambda_k}{\partial g}(0)g + \frac{1}{2}\frac{\partial^2\lambda_k}{\partial g^2}(0)g^2 + \frac{1}{3!}\frac{\partial^3\lambda_k}{\partial g^3}(0)g^3 + ... \quad (9)$$

It can be shown from [35] that writing S in function of the parameter g gives the nth-order eigenvalue derivative as:

$$\frac{\partial^n\lambda_k}{\partial g^n} = \boldsymbol{p_k^*}^{\mathrm{T}}\frac{\partial^n S}{\partial g^n}\boldsymbol{p_k^*} \quad (10)$$

where $\boldsymbol{p_k^*}$ is the eigenvector associated to λ_k^*. The covariance matrix S of X is given by :

$$S = \frac{1}{N-1}\bar{X}^{\mathrm{T}}\bar{X} = \frac{1}{N-1}\left(\boldsymbol{\bar{x}_r^{\mathrm{T}}\bar{x}_q}\right)_{r,q=\{1,...,m\}} \quad (11)$$

which is an unbiased estimate of the true covariance matrix in case of multinormally distributed data. Consider the fault modeling described by (1) and (2), it follows that:

$$\begin{array}{rcl} \boldsymbol{\bar{x}_j} & = & \boldsymbol{x_j} - \mu_j 1 \\ & = & (\boldsymbol{x_j^*} - \mu_j^* 1) + (\boldsymbol{F_j} - g \times \frac{1}{N}\sum_{i=b}^{N} x_{ij}^* 1) + \boldsymbol{V} \\ & = & \boldsymbol{\bar{x}_j^*} + \boldsymbol{\bar{F}_j} + \boldsymbol{V} \end{array} \quad (12)$$

where $\boldsymbol{\bar{F}_j} = \boldsymbol{F_j} - g \times \frac{1}{N}\sum_{i=b}^{N} x_{ij}^* 1$, 1 is a column vector of N ones. Based on (3) and (4), the derivation of S with respect to g can be made under the assumption that the noise is independent of the fault. if δ_r $(r = 1, ..., m)$ and τ are given by the following equations:

$$\begin{array}{c} \delta_r = \sum_{i=b}^{N}(x_{ir}^* - \mu_r^*)x_{ij}^* \quad \forall r \\ \tau = \sum_{q=b}^{N}\left(x_{qj}^* - \frac{1}{N}\sum_{i=b}^{N} x_{ij}^*\right)^2 \end{array} \quad (13)$$

we can then write, while substituting $\boldsymbol{\bar{x}_j}$ in S with its expression (12):

$$\frac{\partial\boldsymbol{\bar{x}_r^{\mathrm{T}}\bar{x}_q}}{\partial g} = 0, \quad \forall r, q \neq j \quad (14)$$

$$\frac{\partial\boldsymbol{\bar{x}_r^{\mathrm{T}}}(\boldsymbol{\bar{x}_j^*} + \boldsymbol{\bar{F}_j})}{\partial g} = \frac{\partial(\boldsymbol{\bar{x}_j^*} + \boldsymbol{\bar{F}_j})^{\mathrm{T}}\boldsymbol{\bar{x}_r}}{\partial g} = \delta_r \quad \forall r \neq j \quad (15)$$

$$\frac{\partial(\bar{\boldsymbol{x}}_{\boldsymbol{j}}^{*}+\bar{\boldsymbol{F}}_{\boldsymbol{j}})^{\mathrm{T}}(\bar{\boldsymbol{x}}_{\boldsymbol{j}}^{*}+\bar{\boldsymbol{F}}_{\boldsymbol{j}})}{\partial g}=2\delta_j+2g\tau. \tag{16}$$

δ_r $(r=1,...,m)$ and τ are functions of the original variables and can be computed from healthy data once for all. The first-order derivative of the covariance matrix is then given by:

$$\frac{\partial S}{\partial g}=\frac{1}{N-1}\begin{bmatrix} 0 & \dots & \delta_1 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ \delta_1 & \dots & 2\delta_j+2g\tau & \dots & \delta_m \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & \delta_m & \dots & 0 \end{bmatrix} \tag{17}$$

The second-order sensitivity of S with respect to the fault amplitude g is obtained by differentiating (17).

The higher-order sensitivities of S $(n>2)$ are all null, as for the eigenvalue derivatives. Writing the loading vector $\boldsymbol{p}_{\boldsymbol{k}}^{*}$ as $\boldsymbol{p}_{\boldsymbol{k}}^{*}=\begin{bmatrix} p_{1k} & \cdots & p_{mk}\end{bmatrix}^{\mathrm{T}}$, it follows that

$$\begin{cases} \dfrac{\partial\lambda_k}{\partial g}=\boldsymbol{p}_{\boldsymbol{k}}^{*\mathrm{T}}\dfrac{\partial S}{\partial g}\boldsymbol{p}_{\boldsymbol{k}}^{*}=\dfrac{2}{N-1}\left(p_{jk}\displaystyle\sum_{r=1}^{m}p_{rk}\delta_r+p_{jk}^2\tau g\right) \\ \dfrac{\partial^2\lambda_k}{\partial g^2}=\boldsymbol{p}_{\boldsymbol{k}}^{*\mathrm{T}}\dfrac{\partial^2 S}{\partial g^2}\boldsymbol{p}_{\boldsymbol{k}}^{*}=\dfrac{2}{N-1}p_{jk}^2\tau \end{cases}$$

and thus

$$\Delta\lambda_k=\lambda_k-\lambda_k^{*}=\frac{2}{N-1}p_{jk}\sum_{r=1}^{m}p_{rk}\delta_r g+\frac{3}{N-1}p_{jk}^2\tau g^2. \tag{18}$$

An estimate, denoted $\hat{g}$, of g is obtained based on (8) and (18). $\Delta\lambda_k^2$ is the estimated $\Delta\lambda_k$ squared. Let $\alpha_1=p_{jk}\sum_{r=1}^{m}p_{rk}\delta_r$ and $\alpha_2=3p_{jk}^2\tau$, the theoretical estimation of g that depends on the divergence value is finally given by

$$\hat{g}=\frac{-\alpha_1+\sqrt{\alpha_1^2+(N-1)\alpha_2(\lambda_k^{*}+\sigma_v^2)(D+\sqrt{(D^2+2D)})}}{\alpha_2} \tag{19}$$

where D is the shorthand of the divergence in (8). In practice, D is numerically approximated. The objective then is to evaluate the impact of the divergence approximation on the accuracy of the obtained model for incipient fault amplitude estimation.

3. Probability density function of the fault amplitude estimate : approximation with a Gamma distribution

The divergence is used to measure the difference between the two probability distributions corresponding to the faultless and faulty signals. As the divergence between two arbitrary probability distributions has no closed form, the integral function given by Eq.(5) should be numerically approximated. The common method to estimate the divergence value uses the interpretation of the Information in term of the likelihood ratio: the KL information from probability distribution f to h is the expected log-likelihood ratio $\log(f/h)$ under the distribution h. This induces two assumptions:

1. an observation set composed of N independent and identically distributed (i.i.d.) observations $\{z_i\}_1^N$ drawn from f is supposed available.

2. $h(z_i)$ can be calculated, and thus q is supposed to be known.

Under these assumptions, the Monte Carlo approximation consists in computing:

$$I_{MC}(f\|h) = \frac{1}{N}\sum_{i=1}^{N}\log\frac{f(z_i)}{h(z_i)} \tag{20}$$

However in our application, the probability distributions are unknown *a priori*. Nevertheless, two observation sets are available (the current and the reference), from which empirical probability distributions can be calculated.
An intuitive and fast way to approximate the divergence between two unknown probability distributions consists in the discrete form that uses probabilities from histograms calculation [36]. Consider an equipartition of the faultless signal into l disjoint intervals $\{[s_0, s_0+\Delta s), [s_0+\Delta s, s_0+2\Delta s), ..., [s_0+(l-1)\Delta s, s_l)\}$ where s_0 and s_l are the min and max values of the signal level. The probabilities $\{w_1, w_2, ..., w_l\}$ of the faultless signal levels are estimated as the proportion of the number of points within each interval to the whole number of points in the signal. The probabilities $\{u_1, u_2, ..., u_l\}$ of the faulty signal levels are calculated for the same set of intervals. Then D is approximated by:

$$\hat{D} = \sum_{i=1}^{l} w_i log\frac{w_i}{u_i} + \sum_{i=1}^{l} u_i log\frac{u_i}{w_i} = \sum_{i=1}^{l}(w_i - u_i)log\frac{w_i}{u_i} \tag{21}$$

To evaluate the accuracy of estimating g through (19), D in (19) is substituted by $\hat{D}$. If $\hat{D}$ fits a known distribution, the distribution of $\hat{g}$ can be calculated based on the following theorem [37]:
Let $\boldsymbol{X}$ have pdf $f_{\boldsymbol{X}}(x)$ and let $\boldsymbol{Y} = \psi(\boldsymbol{X})$, where ψ is a monotone function. Let $\mathcal{X} = \{x : f_{\boldsymbol{X}}(x) > 0\}$ and $\mathcal{Y} = \{y : y = \psi(x) \text{ for some } x \in \mathcal{X}\}$. Suppose that $f_{\boldsymbol{X}}(x)$ is continuous on $\mathcal{X}$ and that $\psi^{-1}(y)$ has a continuous derivation on $\mathcal{Y}$, then the pdf of $\boldsymbol{Y}$ is given by:

$$f_{\boldsymbol{Y}}(y) = \begin{cases} f_{\boldsymbol{X}}(\psi^{-1}(y)) \mid \dfrac{d}{dy}\psi^{-1}(y) \mid & y \in \mathcal{Y} \\ 0 & otherwise. \end{cases} \tag{22}$$

Consider $\psi(x) = \dfrac{-\alpha_1 + \sqrt{\alpha_1^2 + (N-1)\alpha_2(\lambda_k^* + \sigma_v^2)(x + \sqrt{(x^2+2x)})}}{\alpha_2}$. The variable x refers to $\hat{D}$.

The calculation of $\psi'(x)$ proves that ψ is monotone ($g'(x) > 0 \ \forall \ x \geq 0$). The inverse function of ψ is $\psi^{-1}(y) = \dfrac{1}{2}\dfrac{(a_1 y + a_2 y^2)^2}{a(a + a_1 y + a_2 y^2)}$, where the variable y refers to $\hat{g}$, $a = \lambda_k^* + \sigma_v^2$, $a_1 = 2\alpha_1/(N-1)$ and

$a_2 = \alpha_2/(N-1)$.
The derivation of $\psi^{-1}(y)$ denoted $Z(y) = \left(\psi^{-1}(y)\right)'$ is : $Z(y) = \dfrac{4a_1a_2^2y^4 + 2(a_1^2a_2 + aa_2^2)y^3 + 3aa_1a_2y^2 + 2aa_1^2y}{a(a + a_1y + a_2y^2)^2}$.

The simulations of the system represented by (24) and (25) showed that the probability density $f_{\boldsymbol{X}}(\hat{D})$ of the estimated divergence can fit with Gamma distributions. $f_{\boldsymbol{X}}(\hat{D}) \approx \dfrac{\hat{D}^{\alpha-1}}{\Gamma(\alpha)\theta^{\alpha}} exp(-\hat{D}/\theta)$.
As a consequence, the fault amplitude estimate $\hat{g}$ will also be Gamma distributed. Thus according to (22) the pdf of the fault amplitude estimate $\hat{g}$ is given by:

$$f_{\boldsymbol{Y}}(\hat{g}) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} \left(\psi^{-1}(\hat{g})\right)' \left(\psi^{-1}(\hat{g})\right)^{\alpha-1} \exp\left(\frac{-\psi^{-1}(\hat{g})}{\theta}\right) \tag{23}$$

where α and θ are obtained by a numerical fitting with the minimization of a quadratic error.

4. Simulation results

A typical application of the divergence model would be in structural health monitoring (SHM) systems, where the objective is to detect and identify incipient faults/damages in the structure using sensor data (typically vibration data) [38, 39]. The SHM relies on measurements acquired from a dense sensor network that provides sufficient analytical redundancy for diagnosis. It is of crucial importance that the acquired measurements are reliable for analysis and decision. Therefore, sensor faults should be detected and identified correctly so to avoid misinterpretation and confusion with structural faults. In such application, the sensor network can be modelled as a Gaussian process [40] (assumption 3), or any other type of distribution process as it will be illustrated in the following simulation results. The process dimension can be reduced by using PCA for example even if it is not optimal in the non Gaussian case. The divergence can be proposed to address sensitivity and robustness issues [41] in this application, while the analytical divergence model is able to estimate the severity of sensor faults, especially the gain faults.
The theoretical estimation will be evaluated here on a multivariate AR system inspired from [27]:

$$x(i) = \begin{bmatrix} 0.118 & -0.191 \\ 0.847 & 0.264 \end{bmatrix} x(i-1) + \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix} u(i-1)$$

$$y(i) = x(i) + v(i) \tag{24}$$

where u is the correlated input,

$$u(i) = \begin{bmatrix} 0.811 & -0.226 \\ 0.477 & 0.415 \end{bmatrix} u(i-1) + \begin{bmatrix} 0.193 & 0.689 \\ -0.320 & -0.749 \end{bmatrix} w(i-1). \tag{25}$$

w is a vector of 2 inputs $w = \begin{bmatrix} w_1 & w_2 \end{bmatrix}'$. Results will be shown in two cases: when the condition of multivariate normal distribution data is met and in case it is violated. In the first case, the inputs w_1 and w_2 are uncorrelated Gaussian signals with zero mean and unit variance. In the second one, denoted as the *Mixed* case in the following, w_1 is still Gaussian and w_2 is drawn from χ^2 distribution with two degrees of freedom. $u = \begin{bmatrix} u_1 & u_2 \end{bmatrix}^{\mathrm{T}}$ is the vector of measured inputs, and $y = \begin{bmatrix} y_1 & y_2 \end{bmatrix}^{\mathrm{T}}$ is the vector of outputs corrupted by uncorrelated Gaussian errors with zero mean and variance σ_v^2.

The vector of process variables will be formed with the measured inputs and outputs of the process at instant i, i.e. $\begin{bmatrix} y_1(i) & y_2(i) & u_1(i) & u_2(i) \end{bmatrix}^{\mathrm{T}}$.

PCA is applied on the corresponding covariance matrix; it gives 4 principal components with loading vectors $\{\boldsymbol{p_k}\}$ and variances $\lambda_k = \{40.26, 4.9, 1.14, 0.17\}$. The first principal component $\boldsymbol{t_1}$ accounts for 86.6% of variations, it will be used to estimate the fault affecting the output y_2. If $\bar{X}$ is the centered data matrix, then $\boldsymbol{t_1} = \bar{X}\boldsymbol{p_1}$.
The fault is modeled as $y_2(i) = (1+g)x_2(i) + v_2(i)$ where v_2 is the additive noise of variance σ_v^2. The added process noise allows a SNR of 25 dB which is a considerable noise level by reference to many industrial applications (like electrical systems). The last 20% samples of y_2 are affected by the fault (for which $g \neq 0$).
Considering small values of g, specifically $g = \{0.01, 0.015, 0.02, 0.025\}$ meaning variations of $\{1\%, 1.5\%, 2\%, 2.5\%\}$ of the signal amplitude, the pdfs of the estimated $\hat{g}$ obtained through the approximated divergence and described by (21) are displayed in Fig.2. The pdfs of the estimations are clearly centered at the actual fault amplitudes.
With the pdfs and considering a wide fault amplitude range from 0.001 (0.1%) to 0.3 (30%), we obtain Fig.3 that displays the actual and estimated fault amplitudes in the Gaussian and the *Mixed* cases. Fig.4 shows the relative error $E_r = (\hat{g} - g)/(1+g)$ on the estimated variable $\hat{y}_2(i) = (1+\hat{g})x_2(i) + v_2(i)$ of the faulty variable y_2.

In the Gaussian case the estimation relative error is less than 1%. However, even if the Gaussian assumption is no longer valid, the estimation relative error is still acceptable with a maximum value of approximately 3%. Thanks to this accurate estimation, the faultless observations thus can be reconstructed from the faulty ones which is interesting for monitoring purposes.

5. Conclusion

An analytical approach based on the KL divergence is proposed in order to estimate the incipient fault amplitude in highly dimensional processes. As the divergence has no closed form it has been approximated numerically. After the derivation of the analytical model of the fault amplitude estimate, its relevance has been studied with the probability density functions approximated as a gamma distribution. The estimated fault amplitude, when evaluated on a simulated AR process, has proven to be close to the actual value (relative error lower than 1% in the Gaussian case and 3% in the *Mixed* case for a fault amplitude in the $[0.001; 0.3]$ interval). With such an estimation, the faultless observations can be reconstructed from the faulty ones, which can be very useful for control and monitoring purposes.

References

[1] T. Stepinski, T. Uhl and W. Staszewski, *Advanced Structural Damage Detection: From Theory to Engineering Applications*, John Wiley & Sons, Ltd., 2013.

[2] Chih-Min Fan, Ruey-Shan Guo, S.-C. Chang and Chih-Shih Wei, ”SHEWMA: an end-of-line SPC scheme using wafer acceptance test data”, *IEEE Transactions on Semiconductor Manufacturing*, vol.13, no. 3, pp. 344-358, 2000.

[3] Wei Zhou, T.G. Habetler, and R.G. Harley, ”Bearing Fault Detection Via Stator Current Noise Cancellation and Statistical Control”, *IEEE Transactions on Industrial Electronics*, vol. 55, no. 12, pp. 4260-4269, 2008.

[4] N. Kaistha, ”Incipient fault detection and isolation in a PWR plant using principal component analysis”, *in: Proceedings of the American Control Conference*, vol. 3, pp. 12-24, June 2001.

[5] C. A. Lowry, W. H. Woodall, C. W. Champ and S. E. Rigdon, ”A Multivariate Exponentially Weighted Moving Average Control Chart”, *Technometrics*, pp. 46-53, vol. 34, no. 1, 1992.

[6] D. M. Hawkins and E. M. Maboudou-Tchao, ”Multivariate Exponentially Weighted Moving Covariance Matrix”, *Technometrics*, pp. 155-166, vol. 50, no. 2, 2008.

[7] S. W. Cheng and K. Thaga, ”Single variables control charts : an overview”, *Quality and Reliability Engineering International*, vol. 22, no. 7, pp. 811-820, 2006.

[8] U. Kruger and L. Xie, *Advances in Statistical Monitoring of Complex Multivariate Processes*, New York : Wiley, 2012.

[9] J.E. Jackson, *A User's Guide to Principal Components*, New York: Wiley, 1991.

[10] X. Deng, X. Tian and Sheng Chen, ”Modified kernel principal component analysis based on local structure analysis and its application to nonlinear process fault diagnosis”, *Chemometrics and Intelligent Laboratory Systems*, vol. 127, pp. 195-209, 2013,

[11] C. Sankavaram, B. Pattipati, K. Pattipati, Yilu Zhang, M. Howell and M. Salman, ”Data-driven fault diagnosis in a hybrid electric vehicle regenerative braking system”, *IEEE Aerospace Conference*, pp.1-11, 2012.

[12] Yu Gao, Tianshe Yang, Nan Xing and Minqiang Xu, ”Fault detection and diagnosis for spacecraft using principal component analysis and support vector machines”, *IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pp. 1984-1988, 2012.

[13] M. Hamadache and Dongik Lee, ”Principal Components Analysis based Fault Detection and Isolation for Electronic Throttle Control system”, *12th International Conference on Control, Automation and Systems (ICCAS)*, pp. 808-813, 2012.

[14] A.A.Silva, A.M. Bazzi and S. Gupta, ”Fault diagnosis in electric drives using machine learning approaches”, *IEEE International Electric Machines & Drives Conference (IEMDC)*, pp. 722-726, 2013.

[15] J. Yu, ”Fault detection using principal components-based gaussian mixture model for semiconductor manufacturing processes”, *IEEE Transactions on Semiconductor Manufacturing*, vol. 24, no. 3, pp. 471-486, 2011.

[16] J. Harmouche, C. Delpha and D. Diallo, ”Incipient fault detection and diagnosis based on Kullback-Leibler divergence using Principal Component Analysis: Part I”, *Elsevier Signal Processing*, vol. 94, pp. 278-287, 2014.

[17] J. Kullaa, ”Detection, identification, and quantification of sensor fault in a sensor network”, *Mechanical Systems and Signal Processing*, vol. 40, no. 1, pp. 1208-221, 2013.

[18] Shengwei Wang and Jingtan Cui, ”Sensor-fault detection, diagnosis and estimation for centrifugal chiller systems using principal-component analysis method”, *Applied Energy*, vol. 82, no. 3, pp. 197-213, 2005.

[19] S.A. Arogeti, Danwei Wang, Chang Boon Low and Ming Yu, ”Fault Detection Isolation and Estimation in a Vehicle Steering System”, *IEEE Transactions on Industrial Electronics*, vol. 59, no. 12, pp. 4810-4820, 2012.

[20] T. N. Pranatyasto and S. Joe Qin, ”Sensor validation and process fault diagnosis for FCC units under MPC feedback”, *Control Engineering Practice*, vol. 9, no. 8, pp. 877-888, 2001

[21] Z. Gao and S.X. Ding, ”Sensor fault reconstruction and sensor compensation for a class of nonlinear state-space systems via a descriptor system approach”, *Control Theory & Applications, IET*, vol. 1, no. 3, pp. 578-585, 2007.

[22] Ming Liu and Peng Shi, ”Sensor fault estimation and tolerant control for Ito stochastic systems with a descriptor sliding mode approach”, *Automatica*, vol. 49, no. 5, pp. 1242-1250, 2013

[23] Z. Gao and S.X. Ding, ”Fault estimation and fault-tolerant control for descriptor systems via proportional, multiple-integral and derivative observer design”, *Control Theory & Applications, IET*, vol. 1, no. 5, pp. 1208-1218, 2007.

[24] J. Stoustrup and H.H. Niemann, ”Fault estimation - a standard problem approach”, *International journal of robust and nonlinear control*, vol. 12, no. 8, pp. 649-673, 2002.

[25] S. Valle, W. Li and S. J. Qin, ”Comparison of multivariate statistical process control monitoring methods with applications to the Eastman challenge problem”, *Industrial & Engineering Chemistry Research*, vol. 38, no. 11, 4389-4401, 1999.

[26] J. -C. Trigeassou, *Electrical Machines Diagnosis*, John Wiley & Sons, 2013.

[27] M. Kano, S. Hasebe, I. Hashimoto and H. Ohno, ”A new multivariate statistical process monitoring method using principal component analysis”, *Computers & Chemical Engineering*, vol. 25, no. 7-8, pp. 1103-1113, 2001.

[28] J. Tajer, A. Makke, O. Salem and A. Mehaoua, ”A comparison between divergence measures for network anomaly detection”, *7th International Conference on Network and Service Management (CNSM)*, pp. 1-5, 2011.

[29] O. Salem, F. Nait-Abdesselam and A. Mehaoua, ”Anomaly detection in network traffic using Jensen-Shannon divergence”, *IEEE International Conference on Communications (ICC)*, pp. 5200-5204, 2012.

[30] Wen -Liang Hung and Miin -Shen Yang, ”On the J-divergence of intuitionistic fuzzy sets with its application to pattern recognition”, *Information Sciences*, vol. 178, no. 6, pp. 1641-1650, 2008.

[31] Andrew K.S. Jardine, Daming Lin, D. Banjevic, ”A review on machinery diagnostics and prognostics implementing condition-based maintenance Mechanical Systems and Signal Processing”, vol. 20, no. 7, pp. 1483-1510, 2006.

[32] D. Romano and M. Kinnaert, ”Robust fault detection and isolation based on the Kullback divergence”, *Fault Detection, Supervision and Safety of Technical Processes*, vol. 1, pp. 426-431, 2006.

[33] S. Kullback and R. A. leibler, ”On information and sufficiency”, *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79-86, 1951.

[34] M. Basseville, ”Distances Measures for Signal Processing and Pattern Recognition”, *Signal Processing*, vol. 18, no. 4, pp. 349-369, 1989.

[35] N. P. Van Der Aa, H. G. Ter Morsche and R. R. M. Mattheij, ”Computation of eigenvalue and eigenvector derivatives for a general complex-valued eigensystem”, *Electronic Journal of Linear Algebra ELA*, vol. 16, pp. 300-314, 2007.

[36] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, Wiley-Interscience, 2nd edition, 2006

[37] G. Casella and R.L. Berger. *Statistical Inference*, Pacific Grove, CA: Duxbury, 2002.

[38] J. Y. Park, M. B. Wakin, A. C. Gilbert, ”Modal Analysis With Compressive Measurements”, *IEEE Transactions on Signal Processing*, vol. 62, no. 7, pp. 1655-1670, 2014

[39] J. Kullaa, ”Distinguishing between sensor fault, structural damage, and environmental or operational effects in structural health monitoring”, *Mechanical Systems and Signal Processing (MSSP)*, vol. 25, no. 8, pp. 2976-2989, 2011.

[40] D. A. Tibaduiza, M. A. Torres-Arredondo, L. E. Mujica, J. Rodellar, C. P. Fritzen, ”Study of two unsupervised data driven statistical methodologies for detecting and classifying damages in structural health monitoring”, *Mechanical Syst. and Sig. Proc.*, vol. 41, pp. 467-484, 2013.

[41] J. Harmouche, C. Delpha and D. Diallo, ”Incipient fault detection and diagnosis based on Kullback-Leibler divergence using Principal Component Analysis: Part II”, *Elsevier Signal Processing*, vol. 109, pp. 334-344, 2015.

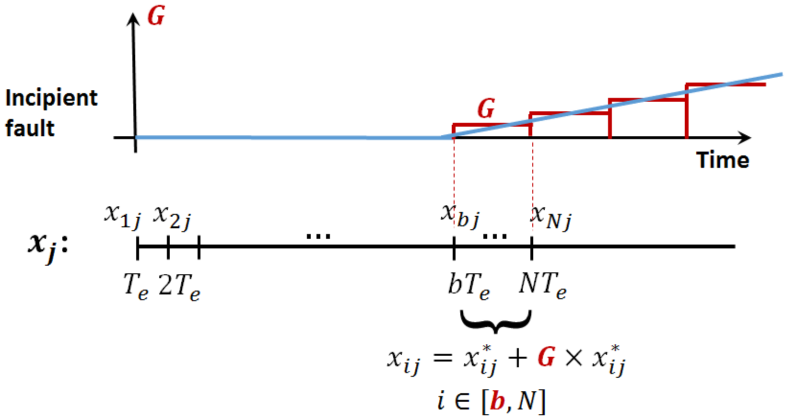

Figure 1: Incipient fault model

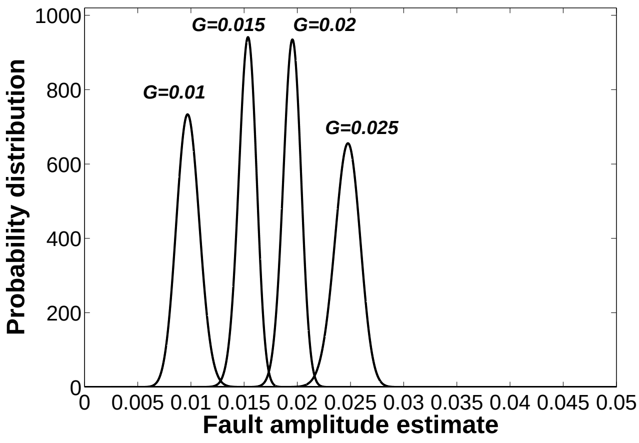

Figure 2: Probability density function of $\hat{g}$ in case the the condition of multivariate normal distribution data is met

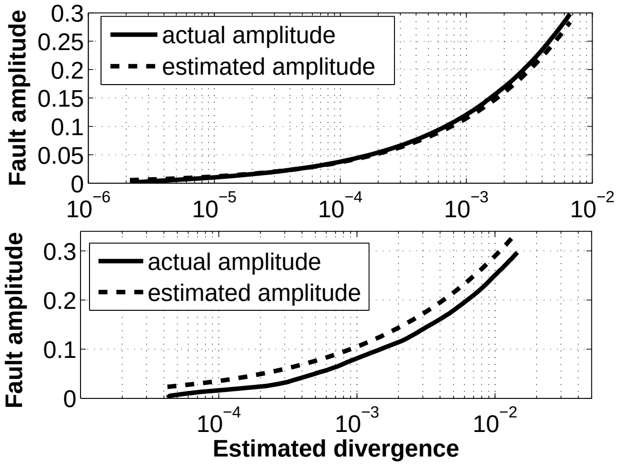

Figure 3: Real and estimated fault amplitude when the condition of multivariate normal distribution data is met (top) and violated (bottom)

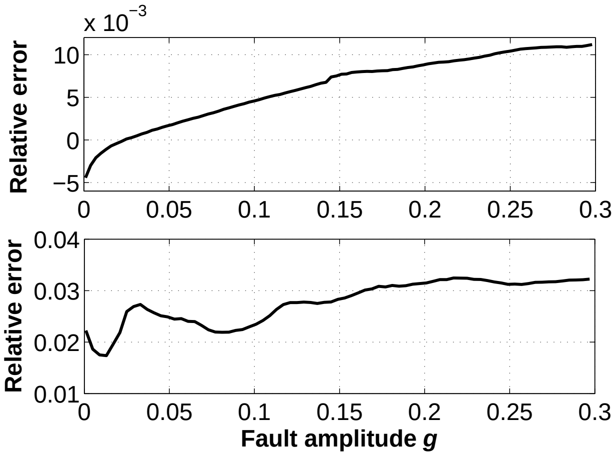

Figure 4: Estimation relative error when the condition of multivariate normal distribution data is met (top) and violated (bottom)