



# **Fault detection and estimation using kernel principal component analysis**

Maya Kallas, Gilles Mourot, Kwami Anani, José Ragot, Didier Maquin

## **► To cite this version:**

Maya Kallas, Gilles Mourot, Kwami Anani, José Ragot, Didier Maquin. Fault detection and estimation using kernel principal component analysis. 20th IFAC World Congress, IFAC 2017, Jul 2017, Toulouse, France. pp.1025-1030, <10.1016/j.ifacol.2017.08.212>. <hal-01478133>

**HAL Id: hal-01478133**

**<https://hal.science/hal-01478133v1>**

Submitted on 26 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Fault detection and estimation using kernel principal component analysis

Maya Kallas\* Gilles Mourot\* Kwami Anani\* José Ragot\*  
Didier Maquin\*

\* *Centre de Recherche en Automatique de Nancy, CNRS UMR 7039,  
Université de Lorraine,  
2 Avenue de la forêt de Haye, 54518 Vandœuvre-Lès-Nancy  
{maya.kallas, gilles.mourot, kwami-dodzivi.anani, jose.ragot,  
didier.maquin} @ univ-lorraine.fr*

---

**Abstract:** The principal component analysis (PCA) is a linear technique widely used to retrieve a subspace that maximizes the variance of the data, making the presence of a fault easy to detect. Nevertheless, the real systems are nonlinear. To this end, we propose in this paper to use a kernel-based technique known as kernel principal component analysis (KPCA) for fault diagnosis. The main idea is to use a nonlinear transformation that projects data into a higher dimensional feature space, where conventional PCA is applied. Although detection can be defined in this space, the estimation of the fault requires the map back to the input space. In this sense, we derive an iterative pre-image technique. A study on possible initial points is done. Three initialization techniques based on different properties are presented. The relevance of the proposed technique is illustrated on simulated data.

*Keywords:* Fault diagnosis, nonlinear systems, kernel principal component analysis, optimization problem.

---

## 1. INTRODUCTION

In the last decade, the need of highly performance industrial processes emerges in complex systems and products. Therefore, it is important to supervise these complex systems for important gain in product quality. The diagnosis of a system is a main aspect of their supervision. The first important aspect in system diagnosis is the detection of functioning changes due for example to sensor or actuator or even system components failures.

System fault diagnosis consists of three main phases: fault detection, isolation but also estimation of the detected fault. Principal component analysis (PCA) is a widely used technique for process monitoring and fault diagnosis in Qin (2003). It seeks for the axes that maximize the variance of data, making it easy to evaluate the error by projecting the observation on the residual axes. The detection part is done through the computation of some indices, while the isolation is frequently done using the structured residuals. The main idea is to find a modification that makes the transformed residuals sensitive or insensitive to certain faults. A variable reconstruction technique is then applied to estimate the amplitude of the detected and isolated fault.

However, the PCA identifies only linear structure in a given dataset, as it is nothing but a linear projection. In order to extend this technique to deal with nonlinear structures, many studies have been proposed to define the nonlinearities generally using the kernel machines. In this particular case, it is an extension to the PCA with the

so-called kernel PCA (KPCA) detailed in Schölkopf et al. (1999).

The idea behind the use of the kernel machines is essentially due to the concept of the *kernel trick*, a “double-edged sword”. In fact, in the case of KPCA, the data are implicitly mapped from an input space into some high-dimensional nonlinear feature space, where the linear PCA can be applied. Therefore, in this feature space, we can detect the presence of the fault. Cho et al. (2005); Choi et al. (2005) proposed different methods based on the kernel PCA for fault detection. Bin Shams (2011); Kallas et al. (2014, 2015) recently proposed nonlinear PCA for fault diagnosis and more generally for process monitoring.

Despite the fact that the isolation of the fault with different techniques relies on the idea of structuring indicators of failure, the extension of this idea to the nonlinear case with the kernel machines remains difficult in the feature space. Alcalá and Qin (2010) proposed to extend the reconstruction technique to the nonlinear case in the feature space. However, it was necessary to retrieve a value in the input space, leading to solve an optimization problem in order to obtain the reconstructed observation. Due to the nonlinear transformation in a high dimensional space, estimating fault magnitude cannot be interpreted in the feature space. Thus most of the studies investigate only fault detection. In order to retrieve a meaningful result, it is necessary to map back the obtained feature to the input space, the space of observations. Due to the use of the kernels, the nonlinear transformation is not explicit, resulting in a mapping back which is complex. This inverse operation to return to the input space is called the *pre-*

*image problem* detailed in Honeine and Richard (2011); Kallas et al. (2013). Its resolution is required to reconstruct the variables. This problem is *ill-posed*, since the exact pre-image may not exist, and even if it exists, it might not be unique. Therefore, we seek an observation in the input space whose image in the feature space is as close as possible to the obtained feature in the virtual space. It can be defined by an optimization problem whose convergence properties should be studied.

In this paper, we propose to diagnose a nonlinear system using the kernel principal component analysis. To this end, the dataset is projected into a higher dimensional feature space, where the linear PCA is applied. The detection part is done directly in this feature space. However the estimation of the fault requires a map back to the input space. Therefore, we present an iterative fixed-point technique that needs an initialization. To this end, we propose three different initial point methods in order to evaluate the pre-image namely the amplitude of the fault. The first one considers only the direction of reconstruction, while the second uses the properties of a Gaussian function to define an attraction zone using the bandwidth. The third technique seeks an observation with a detection limit along the considered direction.

The rest of the paper is organized as follows: section 2 introduces the KPCA technique for fault detection, while section 3 presents a detailed method for fault estimation. Three initial points methods for the fixed-point iterative technique are detailed in section 4. Finally, section 5 illustrates the efficiency of the proposed technique on simulated example.

## 2. FAULT DETECTION FOR NONLINEAR SYSTEMS

### 2.1 Kernel PCA

We consider an input space  $\mathcal{X}$  with an inner product between the data. Let  $\kappa: \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$  be positive semi-definite kernel, namely  $\sum_{i,j} \beta_i \beta_j \kappa(\mathbf{x}_i, \mathbf{x}_j) \geq 0$  for all  $\beta_i, \beta_j \in \mathbb{R}$  and all  $\mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}$ . Based on the Moore-Aronszajn theorem in Aronszajn (1950), any positive definite kernel guarantees the existence of a unique (up to an isometry) feature space (or reproducing kernel Hilbert space)  $\mathcal{H}$  where  $\kappa$  defines an inner product. In other words, there exists a map  $\varphi: \mathcal{X} \mapsto \mathcal{H}$ , from the input space to the feature space, such that

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j) \rangle_{\mathcal{H}} \quad (1)$$

for any  $\mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}$ , where  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$  denotes the corresponding inner product in  $\mathcal{H}$ .

Using this concept, a given data matrix, defined by  $N$  observations describing  $m$  variables, namely,

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \vdots \\ \mathbf{x}_N^\top \end{bmatrix} \quad \mathbf{x}_i \in \mathbb{R}^m, \quad \mathbf{X} \in \mathbb{R}^{N \times m}$$

is mapped to  $\mathcal{H}$ , more precisely

$$\mathbf{x}_i \mapsto \phi_i = \varphi(\mathbf{x}_i), \quad \phi_i \in \mathbb{R}^h$$

In a matrix form, we can write

$$\Phi = \begin{bmatrix} \phi_1^\top \\ \vdots \\ \phi_N^\top \end{bmatrix}, \quad \Phi \in \mathbb{R}^{N \times h}$$

where  $h \gg m$  is the dimension of the feature space.

We evaluate the covariance matrix, in the feature space  $\mathcal{H}$  using:

$$\begin{aligned} \mathbf{S} &= \frac{1}{N-1} \sum_{i=1}^N \phi_i \phi_i^\top, \quad \mathbf{S} \in \mathbb{R}^{h \times h} \\ &= \frac{1}{N-1} \Phi^\top \Phi \end{aligned} \quad (2)$$

It is worth nothing that  $\varphi(\cdot)$  is not explicitly defined. With the use of the *kernel trick*, one can evaluate the Gram matrix  $\mathbf{K} = \Phi \Phi^\top$  using the kernel function  $\kappa(\mathbf{x}_i, \mathbf{x}_j)$ . We obtain:

$$\mathbf{K} = \begin{bmatrix} \phi_1^\top \phi_1 & \cdots & \phi_1^\top \phi_N \\ \vdots & \ddots & \vdots \\ \phi_N^\top \phi_1 & \cdots & \phi_N^\top \phi_N \end{bmatrix} \quad (3)$$

As in the case of the conventional PCA, KPCA seeks to solve the eigenvector equation in the feature space, by evaluating the eigenvectors  $\boldsymbol{\nu}_i$  where  $i = 1, \dots, N$  and eigenvalues  $\lambda_i$  of the covariance matrix  $\mathbf{S}$ . Let  $\boldsymbol{\alpha}_i$  and  $\lambda_i$  be an eigenvector and its corresponding eigenvalue of the matrix  $\mathbf{K}$  with  $\boldsymbol{\alpha}_i \propto \Phi \boldsymbol{\nu}_i$ . Nevertheless in the feature space, the eigenvectors with a unit length will be given by

$$\boldsymbol{\nu}_i = \frac{\Phi^\top \boldsymbol{\alpha}_i}{\sqrt{\lambda_i(N-1)}} \quad (4)$$

We can form a matrix  $\mathbf{V} = [\boldsymbol{\nu}_1 \cdots \boldsymbol{\nu}_\ell, \boldsymbol{\nu}_{\ell+1} \cdots \boldsymbol{\nu}_N]$  with the eigenvectors  $\boldsymbol{\nu}_i$  where  $\ell$  is the number of principal components (PC). It corresponds to the eigenvectors of the covariance matrix  $\mathbf{S}$  in the feature space. We denote  $\hat{\mathbf{V}} = [\boldsymbol{\nu}_1 \cdots \boldsymbol{\nu}_\ell]$  the principal subspace that spans the maximal variance between data, and  $\tilde{\mathbf{V}} = [\boldsymbol{\nu}_{\ell+1} \cdots \boldsymbol{\nu}_N]$  the residual subspace that contains the noises. The two subspaces are complementary. Using this configuration, we estimate the KPCA model. Therefore, the  $\ell$  value is determined based on a certain percentage of the overall eigenvalues. A threshold  $\delta^2$ , linked to the level of noise affecting the system, is then computed from a fault-free database where we consider few outliers. It will be used as a reference to detect the presence of a fault.

### 2.2 Fault detection

Once the KPCA model is estimated on healthy data, faults can be detected. Let us consider  $\mathbf{x}$  a new fault-free observation. Its image in the feature space is

$$\phi = \varphi(\mathbf{x}), \quad \phi \in \mathbb{R}^h$$

for which, the projection on the principal and residual spaces are given by

$$\begin{aligned} \hat{\mathbf{t}} &= \hat{\mathbf{V}}^\top \phi, \quad \in \mathbb{R}^\ell \\ \tilde{\mathbf{t}} &= \tilde{\mathbf{V}}^\top \phi, \quad \in \mathbb{R}^{h-\ell} \end{aligned}$$

We evaluate the statistics Square Prediction Error (*SPE*), where the dot product in  $\mathcal{H}$  is replaced by the kernel function as follows

$$\begin{aligned} SPE(\mathbf{x}) &= \tilde{\mathbf{t}}^\top \tilde{\mathbf{t}} \\ &= \kappa(\mathbf{x}, \mathbf{x}) - \kappa^\top(\mathbf{x}) \mathbf{C} \kappa(\mathbf{x}) \end{aligned} \quad (5)$$

with

$$\begin{aligned} \kappa(\mathbf{x}) &= [\kappa(\mathbf{x}_1, \mathbf{x}), \dots, \kappa(\mathbf{x}_N, \mathbf{x})]^\top \\ \mathbf{C} &= (N-1)^{-1} \hat{\mathbf{A}} \hat{\Lambda}^{-1} \hat{\mathbf{A}}^\top \end{aligned}$$

where  $\mathbf{A}$  is the matrix of the eigenvectors of the Gram matrix  $\mathbf{K}$  with  $\mathbf{A} = [\alpha_1, \dots, \alpha_N]$ , from which we choose the first  $\ell$  columns as  $\hat{\mathbf{A}}$  and  $\hat{\Lambda}$  is the diagonal matrix of the first  $\ell$  eigenvalues  $\lambda_i$ , for  $i = 1, \dots, \ell$ .

*Remark 1.* For statistic reasons, it is preferred to have centered data in the feature space. The centering is detailed in Chapter 5 of Shawe-Taylor and Cristianini (2004).

For the following, kernels are assumed centered.

The value obtained with the  $SPE$  is then compared to the threshold  $\delta^2$ . If it is greater than this value, we have detected a fault on the observation  $\mathbf{x}$ . Then, we need to isolate the faulty variable and estimate the amplitude of the fault in this observation. Let

$$\mathbf{x} = \mathbf{x}^* + \xi_j \mathbf{f} \quad (6)$$

be the measured observation, where  $\mathbf{x}^*$  is the fault-free value, and  $\xi_j = [0 \dots 1 \dots 0]^\top$  is a vector defining the  $j^{th}$  direction of the fault  $f$  with 1 on the  $j^{th}$  component and 0 elsewhere.

### 3. FAULT ISOLATION AND ESTIMATION

As aforementioned, it is important to estimate its value  $\hat{f}$  and recognize its direction  $\hat{\xi}$  once a fault is detected. The main idea is to find a magnitude  $\hat{f}_i$  for a given direction  $\xi_i$  for which  $SPE \leq \delta^2$ . Therefore, we seek to minimize the  $SPE$ . Let us consider  $\mathbf{z}_i$  the corrected measure with a reconstructed fault along the  $i^{th}$  direction, with

$$\mathbf{z}_i = \mathbf{x} - \xi_i f_i \quad (7)$$

The expression of the  $SPE(\mathbf{z}_i)$  is

$$SPE(\mathbf{z}_i) = \kappa(\mathbf{z}_i, \mathbf{z}_i) - \kappa^\top(\mathbf{z}_i) \mathbf{C} \kappa(\mathbf{z}_i)$$

Eventhough a fault can be detected in the feature space, the estimation of its amplitude can't be interpreted in this virtual space. Therefore, we need to map back to the input space, where the amplitude can be expressed. However, due to the use of the *kernel trick*, the transformation  $\varphi$  is not explicit. Thus, the map back is not direct. This is the pre-image problem, where we seek to find a solution that minimizes the  $SPE$ . Therefore, for each direction  $\xi_i$ , we seek a magnitude  $\hat{f}_i$ . Thus, we derive the  $SPE$  with respect to  $f_i$ , and we obtain

$$\begin{aligned} \frac{\partial SPE(\mathbf{z}_i)}{\partial f_i} &= \frac{\partial \kappa(\mathbf{z}_i, \mathbf{z}_i)}{\partial f_i} - 2\kappa^\top(\mathbf{z}_i) \mathbf{C} \frac{\partial \kappa(\mathbf{z}_i)}{\partial f_i} \\ \frac{\partial \kappa(\mathbf{z}_i)}{\partial f_i} &= \frac{\partial \kappa(\mathbf{z}_i)}{\partial \mathbf{z}_i^\top} \frac{\partial \mathbf{z}_i}{\partial f_i} = -\frac{\partial \kappa(\mathbf{z}_i)}{\partial \mathbf{z}_i^\top} \xi_i \end{aligned}$$

In this paper, we use the Gaussian kernel, the mostly used kernel in fault diagnosis, defined by

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{(\mathbf{x}_i - \mathbf{x}_j)^\top (\mathbf{x}_i - \mathbf{x}_j)}{2c}\right) \quad (8)$$

where  $\sqrt{c}$  is the bandwidth of the kernel. The derivative of  $SPE$  with respect to  $f_i$  is defined by

$$\frac{\partial SPE(\mathbf{z}_i)}{\partial f_i} = -\frac{2}{c} \kappa^\top(\mathbf{z}_i) \mathbf{C} \mathbf{D}_i(\mathbf{z}_i) \kappa(\mathbf{z}_i)$$

where  $\mathbf{z}_i = \mathbf{x} - \xi_i f_i$  and  $\mathbf{D}_i(\mathbf{z}_i)$  is a diagonal matrix with

$$\mathbf{D}_i(\mathbf{z}_i) = \begin{bmatrix} (\mathbf{z}_i - \mathbf{x}_1)^\top \xi_i & 0 & \dots & 0 \\ 0 & \ddots & \dots & 0 \\ 0 & \dots & (\mathbf{z}_i - \mathbf{x}_t)^\top \xi_i & 0 \\ 0 & \dots & \dots & 0 \\ 0 & \dots & \dots & (\mathbf{z}_i - \mathbf{x}_N)^\top \xi_i \end{bmatrix}$$

Therefore, we can write the gradient of  $SPE(\mathbf{z}_i)$  as

$$\frac{\partial SPE}{\partial f_i} = -\frac{2}{c} (\kappa^\top(\mathbf{z}_i) \mathbf{C} \mathbf{D}_i(\mathbf{x}) \kappa(\mathbf{z}_i) - f_i \kappa^\top(\mathbf{z}_i) \mathbf{C} \kappa(\mathbf{z}_i))$$

In order to estimate  $f_i$ , we set the gradient to zero and we obtain the following expression for the stationary condition

$$f_i \kappa^\top(\mathbf{z}_i) \mathbf{C} \kappa(\mathbf{z}_i) = \kappa^\top(\mathbf{z}_i) \mathbf{C} \mathbf{D}_i(\mathbf{x}) \kappa(\mathbf{z}_i)$$

Taking this expression, we can see that there is a necessary condition to estimate  $f_i$  defined by

$$\hat{\Lambda}^{-\frac{1}{2}} \hat{\mathbf{A}}^\top \kappa(\mathbf{z}_i) \neq \mathbf{0} \quad (9)$$

otherwise  $\mathbf{0} f_i = \mathbf{0}$  for every  $f_i$  along direction  $\xi_i$ .

Based on this expression, the estimation of  $f_i$  can be evaluated via a fixed-point iterative technique ( $f = g(f)$ ) for all possible directions  $\xi_i$ . The solution is thus given by

$$\hat{f}_i = \frac{\kappa^\top(\hat{\mathbf{z}}_i) \mathbf{C} \mathbf{D}_i(\mathbf{x}) \kappa(\hat{\mathbf{z}}_i)}{\kappa^\top(\hat{\mathbf{z}}_i) \mathbf{C} \kappa(\hat{\mathbf{z}}_i)} \quad (10)$$

From expression (10), it is important that  $\kappa(\hat{\mathbf{z}}_i)$  is not null, in order to obtain a value for  $f_i$ . After the convergence of  $f$  to a certain value, the  $SPE$  is reevaluated on the corrected observation  $\mathbf{z}_i$  defined by (7), where  $f_i$  and  $\xi_i$  are replaced with their estimated values  $\hat{f}_i$  and  $\hat{\xi}_i$ . If the  $SPE$  of this corrected observation is less than the threshold  $\delta^2$ , then a fault on the  $\hat{\xi}_i$  direction is isolated. However, it is worth nothing that the true direction  $\xi$  is not defined, since the diagnosis aims to localize the fault among the component of  $\mathbf{x}$ . Therefore, this procedure is done for all possible directions  $\xi_i$ .

However, for convergence purpose in some cases, it is necessary to start with a value of  $\mathbf{z}_i$  near the dataset, in other words, an observation within a certain neighborhood in order not to obtain  $\kappa(\mathbf{z}_i) \approx 0$ . Due to the fact that the used technique is iterative, we might obtain a local minimum. Therefore, the initial point for the iterative expression is important. The derivative of  $SPE$  evaluated on the chosen initial point should not be zero or very small leading to a slow convergence. We need to obtain a value of  $\hat{f}$  that minimizes the  $SPE$  and the value obtained should be less than  $\delta^2$ , in order to eliminate the influence of the fault. Thus, we propose in the following three different initializations.

### 4. INITIALIZATION OF FIXED-POINT ITERATIVE TECHNIQUE

As for all nonlinear optimization techniques, the method can converge toward a local minimum. The idea is to choose an initial point for which  $\kappa(\mathbf{z}_i)$  is not null. Therefore, this section presents three different initialization of

the search algorithm. The first one can be applied directly. It seeks the closest observation in a given direction. The second technique is based on the properties of a Gaussian function and its bandwidth. The third proposition searches for the decision limit along a certain direction.

#### 4.1 Closest observation in a direction

This solution is a simple technique that takes into consideration the closest observation according to the direction of the reconstruction. To this end, we write a minimization problem to seek for the nearest observation

$$t^* = \arg \min_t |(\mathbf{x} - \mathbf{x}_t)^\top \boldsymbol{\xi}_i|$$

where  $t^*$  corresponds to the observation number retrieved. This leads to an initial start for the algorithm with

$$\widehat{f}_i^0 = -(\mathbf{x} - \mathbf{x}_{t^*})^\top \boldsymbol{\xi}_i \quad (11)$$

However, this solution considers only one component, and in some cases the observation retained, taken with all its components, is far from the faulty one.

#### 4.2 Observation in the attraction zone of the dataset

Another solution can be proposed in the case where the faulty observation  $\mathbf{x}$  is far from the dataset, in other words  $\boldsymbol{\kappa}(\mathbf{x}) \approx 0$ . In the case of a Gaussian kernel defined by (8),  $\boldsymbol{\kappa}(\mathbf{x}_i, \mathbf{x}_j)$  is very small when the distance between  $\mathbf{x}_i$  and  $\mathbf{x}_j$  is greater than  $3\sqrt{c}$ . Due to the fact that  $\boldsymbol{\kappa}(\mathbf{z}_i)$  appears as a quadratic form  $\boldsymbol{\kappa}^\top(\mathbf{z}_i)\boldsymbol{\kappa}(\mathbf{z}_i)$ , the corresponding value tends to zero. To overcome this problem, in practice the radius  $3\sqrt{c}$  is limited to  $p\sqrt{c}$  where  $p < 3$  tending the exponential to a value higher than  $\exp(-\frac{9}{2})$ . Due to the fact that we seek some observations near the direction of reconstruction, we can therefore define a condition of belonging to an attraction zone of the kernel making the  $\boldsymbol{\kappa}(\mathbf{z}_i)$  not null. In other words, we consider a minimal amplitude of  $f_i$  that verifies the following condition for a certain observation  $\mathbf{x}_t$

$$\|\mathbf{z}_i - \mathbf{x}_t\|^2 = p^2 c$$

Replacing  $\mathbf{z}_i$  by its expression leads to

$$f_i^2 - 2f_i(\mathbf{x} - \mathbf{x}_t)^\top \boldsymbol{\xi}_i + \|\mathbf{x} - \mathbf{x}_t\|^2 - p^2 c = 0 \quad (12)$$

(12) is a second order equation in  $f_i$ , for which the discriminant is

$$\Delta = ((\mathbf{x} - \mathbf{x}_t)^\top \boldsymbol{\xi}_i)^2 - (\|\mathbf{x} - \mathbf{x}_t\|^2 - p^2 c)$$

Due to the fact that  $\boldsymbol{\kappa}(\mathbf{x}) \approx 0$ , and the expression of the Gaussian distribution,  $\|\mathbf{x} - \mathbf{x}_t\|^2 \geq p^2 c$ , if  $p = 3$ . The discriminant is positive or null if

$$|(\mathbf{x} - \mathbf{x}_t)^\top \boldsymbol{\xi}_i| \geq \sqrt{\|\mathbf{x} - \mathbf{x}_t\|^2 - p^2 c} \quad (13)$$

In other words, there is an intersection between the reconstruction direction and a circle of center  $\mathbf{x}_t$  with radius  $p\sqrt{c}$ . Therefore, the solutions are given by

$$\widehat{f}_{i1,2}^0 = (\mathbf{x} - \mathbf{x}_t)^\top \boldsymbol{\xi}_i \pm \sqrt{\Delta} \quad (14)$$

We thus search in the database all possible observations verifying (13) and we choose the one with the minimal amplitude  $f_i$  using (14). The searching algorithm for this proposed technique is presented in Algorithm 1.

This algorithm needs to be evaluated for each  $t = 1, \dots, N$ , and has some parameters to be chosen making it a technique with high computational cost. We propose

the next initialization method that considers the detection limit.

---

#### Algorithm 1 Search of observation in attraction zone

---

1: **Inputs:**

$\mathbf{x}_t, \quad t = 1, \dots, N$

$\mathbf{x}$ , with  $\boldsymbol{\kappa}(\mathbf{x}) = 0$

$\boldsymbol{\xi}_i$  a given direction

2: **Initialize:**

$j \leftarrow 1$

$m \leftarrow \|\mathbf{x} - \mathbf{x}_1\|$

$p \leftarrow a$ , with  $a < 3$  depending on the value of  $\lambda_\ell^{-1}$

3: **while**  $j \leq N$  **do**

4:   **if**  $|(\mathbf{x} - \mathbf{x}_j)^\top \boldsymbol{\xi}_i| \geq \sqrt{\|\mathbf{x} - \mathbf{x}_j\|^2 - p^2 c}$  **then**

5:      $\Delta \leftarrow ((\mathbf{x} - \mathbf{x}_j)^\top \boldsymbol{\xi}_i)^2 - (\|\mathbf{x} - \mathbf{x}_j\|^2 - p^2 c)$

6:     **if**  $(\mathbf{x} - \mathbf{x}_j)^\top \boldsymbol{\xi}_i \pm \sqrt{\Delta} < m$  **then**

7:        $m \leftarrow (\mathbf{x} - \mathbf{x}_j)^\top \boldsymbol{\xi}_i \pm \sqrt{\Delta}$

8:        $\mathbf{x}_{sol} \leftarrow \mathbf{x}_j$

9:     **end if**

10:   **end if**

11:    $j \leftarrow j + 1$

12: **end while**

13:  $\widehat{f}_{i1,2}^0 \leftarrow (\mathbf{x} - \mathbf{x}_{sol})^\top \boldsymbol{\xi}_i \pm \sqrt{\Delta}$

---

#### 4.3 Observation with detection limit according to the reconstruction direction

If we take a Gaussian kernel where an observation is assumed to be affected by a fault along direction  $\boldsymbol{\xi}_i$ , we can write the expression as

$$\boldsymbol{\kappa}(\mathbf{z}_i, \mathbf{x}_t) = \kappa_{-i}(\mathbf{x}, \mathbf{x}_t) \Delta_t$$

where

$$\kappa_{-i}(\mathbf{x}, \mathbf{x}_t) = \prod_{p=1, p \neq i}^m \exp \left\{ -\frac{([\mathbf{x}]_p - [\mathbf{x}_t]_p)^2}{2c} \right\}$$

$$\Delta_t = \exp \left\{ -\frac{1}{2c} ((\mathbf{x} - \mathbf{x}_t)^\top \boldsymbol{\xi}_i - f_i)^2 \right\}$$

where  $[\cdot]_p$  corresponds to the  $p^{th}$  component of the considered vector. Therefore, we can write in a matrix form

$$\boldsymbol{\kappa}(\mathbf{z}_i) = \boldsymbol{\Delta}_f \boldsymbol{\kappa}_{-i}(\mathbf{x})$$

with

$$\boldsymbol{\Delta}_f = \begin{bmatrix} \Delta_1 & 0 & \dots & \dots & 0 \\ 0 & \ddots & \dots & \dots & 0 \\ 0 & 0 & \Delta_t & \dots & 0 \\ 0 & 0 & \dots & \ddots & 0 \\ 0 & 0 & \dots & \dots & \Delta_N \end{bmatrix}$$

We seek the minimal fault amplitude for which  $SPE(\mathbf{z}_i) \geq \delta^2$ . Minimizing the  $SPE$  leads to maximizing  $\boldsymbol{\kappa}^\top(\mathbf{z}_i) \mathbf{C} \boldsymbol{\kappa}(\mathbf{z}_i)$  given by

$$\max (\boldsymbol{\kappa}_{-i}^\top(\mathbf{x}) \boldsymbol{\Delta}_f \mathbf{C} \boldsymbol{\Delta}_f \boldsymbol{\kappa}_{-i}(\mathbf{x})) \leq 1 - \delta^2 \quad (15)$$

Taking the maximum of the eigen decomposition of  $\mathbf{C}$ , the inequality is written as

$$\lambda_{max}(\mathbf{C}) \max (\boldsymbol{\kappa}_{-i}^\top(\mathbf{x}) \boldsymbol{\Delta}_f^2 \boldsymbol{\kappa}_{-i}(\mathbf{x})) \leq 1 - \delta^2$$

Let us define the different terms with

$$\lambda_{max}(\mathbf{C}) = \frac{1}{N-1} \lambda_{max}(\hat{\Lambda}^{-1}) = \frac{1}{(N-1)\lambda_\ell}$$

$$\max (\boldsymbol{\kappa}_{-i}^\top(\mathbf{x}) \boldsymbol{\Delta}_f^2 \boldsymbol{\kappa}_{-i}(\mathbf{x})) \leq N \max_t \kappa_{-i}^2(\mathbf{x}, \mathbf{x}_t) \Delta_t^2$$

Since  $\Delta_t^2$  is a decreasing exponential, it is maximum when  $\frac{1}{c} ((\mathbf{x} - \mathbf{x}_t)^\top \boldsymbol{\xi}_i - f_i)^2$  is minimum, and the term

$\kappa_{-i}(\mathbf{x}, \mathbf{x}_t)$  is maximum when the observation  $\mathbf{x}_t$  is as close as possible to  $\mathbf{x}$  in all directions except the  $i^{th}$ . Therefore, we seek the observation  $\mathbf{x}_{t^*}$  satisfying

$$t^* = \arg \min_t \left( \sum_{p=1, p \neq i}^m ([\mathbf{x}]_p - [\mathbf{x}_t]_p)^2 \right) |(\mathbf{x} - \mathbf{x}_t)^\top \boldsymbol{\xi}_i| \quad (16)$$

Let us denote  $d_i = (\mathbf{x} - \mathbf{x}_{t^*})^\top \boldsymbol{\xi}_i$ , (15) can be written as

$$\kappa_{-i}^2(\mathbf{x}, \mathbf{x}_{t^*}) \exp \left\{ -\frac{1}{c} (d_i - f_i)^2 \right\} \leq (1 - \delta^2) \frac{(N-1)\lambda_\ell}{N}$$

Taking the logarithm of this expression gives us

$$\frac{1}{c} (d_i - f_i)^2 \geq \log(\kappa_{-i}^2(\mathbf{x}, \mathbf{x}_{t^*})) + d$$

where  $d = \log \left( \frac{N}{(N-1)\lambda_\ell} \right) - \log(1 - \delta^2)$

Replacing  $\kappa_{-i}^2(\mathbf{x}, \mathbf{x}_{t^*})$  with its expression leads to

$$(d_i - f_i)^2 \geq c d - \underbrace{\sum_{p=1, p \neq i}^m ([\mathbf{x}]_p - [\mathbf{x}_{t^*}]_p)^2}_{=a}$$

Using this inequality, we seek to solve a second order equation in  $f_i$  defined by  $f_i^2 - 2d_i f_i + d_i^2 - a = 0$ . The discriminant of this equation is  $\Delta = a$ . It is positive or null if  $a$  is positive or null. In this case, the solutions are

$$\widehat{f_{i1,2}^0} = d_i \pm \sqrt{a} \quad (17)$$

Usually, the minimum value of these two solutions is considered as the starting point.

## 5. EXPERIMENTS

In this section, we use simulated data to test our approaches. The main idea is to illustrate on a simple academic example the results of the three techniques. To this end, we consider a 2D banana dataset. It is defined by a parabola having  $(\mathbf{x}, \mathbf{x}^2 + \epsilon)$  as coordinates, where  $\mathbf{x}$  is defined by 75 points uniformly distributed within the interval  $[-2, 1]$ , and  $\epsilon$  is normally distributed with a standard deviation of  $\nu = 0.1$ .

The limit of detection which correspond to  $SPE = \delta^2$  is evaluated with the Gaussian kernel (8) with a bandwidth of 0.35, and by taking 95% of the fault-free database. It is represented in Figure 1 with the green dashed line. In order to detect a fault, it should have sufficient amplitude to obtain an observation outside the boundary. The  $\ell$  retained PCs correspond to 98% of the sum of eigenvalues.

In order to visualize the possible cases, where the faulty observations can't be projected on the principal axes as in (9), we use the derivative of the  $SPE$  with respect to the fault  $f_i$ . Figure 1 illustrates this derivative, for which we choose the direction  $\mathbf{x}_1$ , on a grid that covers the dataset. The values are presented by vectors corresponding to their amount. As we can see, for some points, the derivative is null making it impossible to converge to the initial dataset where there is no fault affecting the observations.

Let us start with the simple initialization technique proposed in section 4.1. Figure 2 illustrates the choice of the initialization observation. We see that the faulty observation in red is in a zone where the derivative is null as

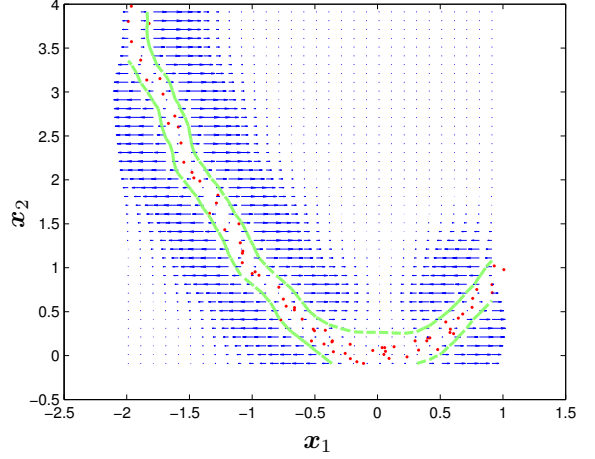


Fig. 1. Evaluation of the derivative of  $SPE$  with respect to  $f_i$  in each point in the grid containing the learning database and illustration of the threshold  $\delta^2$  (green dashed line) in the case of a Gaussian kernel

defined in Figure 1. The retained observation is the magenta one. We see that it is not near the faulty observation even though in the considered direction  $\mathbf{x}_1$ , it is the closest as defined in (11). The initial point is given by the + sign and the obtained solution is the cyan observation which is not near the database.

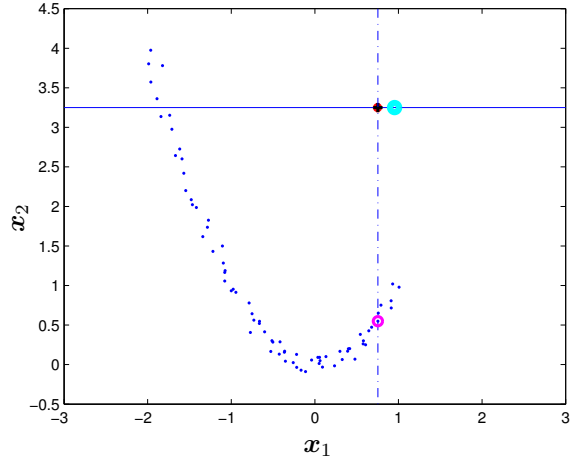


Fig. 2. Initialization using the technique of the closest observation in the direction  $\mathbf{x}_1$  defined in 4.1, where the red is a faulty observation, the magenta observation is the chosen one for the initialization and the cyan observation is the solution using the fixed-point iterative technique

The Figure 3 illustrates the initialization technique based on the search of an observation in the attraction zone of the Gaussian kernel detailed in section 4.2, for the same faulty observation. Using the direction of reconstruction  $\mathbf{x}_1$  as presented with the line, we can see the intersection with the circle of radius  $p\sqrt{c}$  given by the Algorithm 1. In our case,  $p = 2.5$ . Therefore, we seek an observation that satisfies both equation (12) and the initialization of  $f$  is evaluated by (14) for the retained magenta observation. The initial point is presented by the + sign at the intersection between the circle and the line defining the direction.

The solution of the fixed-point iterative technique is the cyan observation.

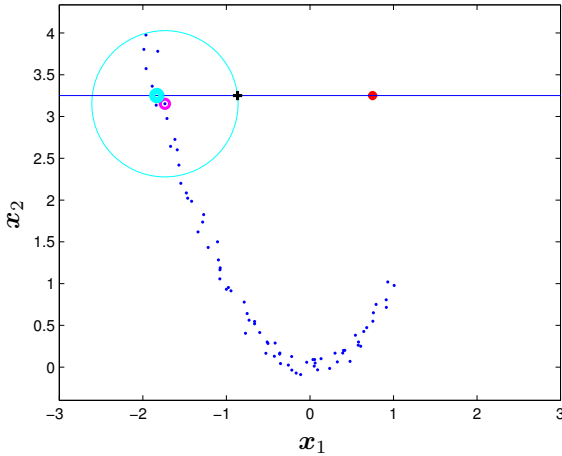


Fig. 3. Initialization using the technique of attraction defined in 4.2, where the red is a faulty observation, the magenta is the chosen observation for the initialization and the cyan observation is the solution using the fixed-point iterative technique in the  $x_1$  direction

The Figure 4 presents the initialization technique where we seek the observation near detection limit according to the reconstruction direction  $x_1$  defined in section 4.3. To this end, a search for the observation  $x_t$  maximizing the decreasing exponential  $\Delta_t^2$  detailed in (16) is done. It is illustrated with the magenta observation. The initial value for  $f$  is given by (17). It is presented by the + sign. As we can see, it is near the threshold illustrated in Figure 1.

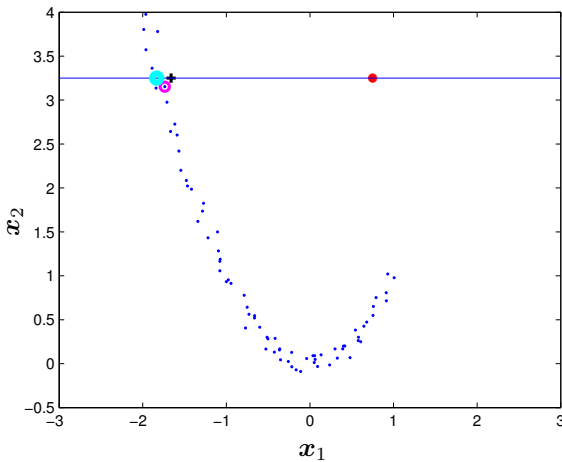


Fig. 4. Initialization using the technique of the observation with detection limit according to the reconstruction direction  $x_1$  defined in 4.3, where the red is a faulty observation, the magenta observation is retained for the initialization and the cyan observation is the solution using the fixed-point iterative technique

## 6. CONCLUSION

In this paper, the fault detection and estimation is presented. The fault is detected if the  $SPE$  is greater than a threshold. we proposed a study about the estimation of

the amplitude of the detected fault. To this end, a fixed-point iterative technique is used to resolve an optimization problem. The initialization of this technique is studied, where a simple method uses the closest observation. A second technique takes advantage of the bandwidth of the Gaussian kernel and uses it as an attraction zone. The third one searches for an observation near the decision limit. The relevance of the proposed techniques is illustrated on a simulated example.

As for future work, we intend to apply these techniques on more complicated examples with more dimensions. We want also to present other techniques to resolve the optimization problem in order to estimate the amplitude of the fault such as a gradient descent technique with an adaptive stepsize. We might consider a direct technique to be a solution for the optimization problem based on some neighbors in the dataset.

## REFERENCES

- Alcala, C. and Qin, S. (2010). Reconstruction-based contribution for process monitoring with kernel principal component analysis. *Industrial & Engineering Chemistry Research*, 49(17), 7849–7857.
- Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68, 337–404. doi:10.2307/1990404.
- Bin Shams, M. (2011). Fault identification using kernel principal component analysis. In *18th IFAC World Congress*, 4320–4325. Milano, Italy.
- Cho, J., Lee, J., Choi, S., Lee, D., and Lee, I. (2005). Fault identification for process monitoring using kernel principal component analysis. *Chemical Engineering Science*, 60(1), 279–288. doi:10.1016/j.ces.2004.08.007.
- Choi, S., Lee, C., Lee, J., Park, J., and Lee, I. (2005). Fault detection and identification of nonlinear processes based on kernel PCA. *Chemometrics and Intelligent Laboratory Systems*, 75(1), 55–67. doi: 10.1016/j.chemolab.2004.05.001.
- Honeine, P. and Richard, C. (2011). Preimage problem in kernel-based machine learning. *IEEE Signal Processing Magazine*, 28(2), 77–88.
- Kallas, M., Honeine, P., Richard, C., Francis, C., and Amoud, H. (2013). Non-negativity constraints on the pre-image for pattern recognition with kernel machines. *Pattern Recognition*, 46(11), 3066–3080.
- Kallas, M., Mourot, G., Maquin, D., and Ragot, J. (2014). Diagnosis of nonlinear systems using kernel principal component analysis. In *11th European Workshop on Advanced Control and Diagnosis, ACD*. Berlin, Germany.
- Kallas, M., Mourot, G., Maquin, D., and Ragot, J. (2015). Fault estimation of nonlinear processes using kernel principal component analysis. In *14th European Control Conference*. Linz, Austria.
- Qin, S. (2003). Statistical process monitoring : basis and beyond. *Journal of Chemometrics*, 17(8-9), 480–502.
- Schölkopf, B., Smola, A.J., and Müller, K.R. (1999). Kernel principal component analysis. In B. Schölkopf, C.J.C. Burges, and A.J. Smola (eds.), *Advances in Kernel Methods*, 327–352. MIT Press, Cambridge, MA, USA.
- Shawe-Taylor, J. and Cristianini, N. (2004). *Kernel Methods for Pattern Analysis*. Cambridge University Press.