Damage Detection in Rotating Machinery
Using Statistical Methods: PCA Analysis and Autocorrelation Matrix

Christophe Thiry, Ai-Min Yan, Jean-Claude Golinval
Université de Liège
Département d’Aérospatiale, Mécanique et Matériaux
Vibrations et Identification des Structures
Chemin des Chevreuils, 1 (B52) B-4000 Liège (Belgique)
E-mail: C.Thiry@ulg.ac.be, AM.Yan@ulg.ac.be, JC.Golinval@ulg.ac.be

Abstract

This paper presents different approaches of a statistical method called Principal Components Analysis (PCA), and which can be used for detecting early symptoms of potential faults in machine components. These methods are based on the comparison of the current machine working state with a reference working state.

PCA techniques are applied to the real case of a rotating machinery-testing bench with faulty roller bearings (defects on inner and outer races). Results are compared for the different statistical methods.

Keywords

1 Introduction

In the industrial world, production losses directly affect the bottom line as costly unexpected shutdowns of critical equipment may have catastrophic effects. Using appropriate predictive maintenance procedures allows to prevent unexpected mechanical failures and to keep overhead costs to a minimum.

Vibration monitoring is one of the main tools that allow to determine the mechanical health of various components of a machine in a non-intrusive manner. The mechanical components that are the most often encountered in rotating machinery are bearings - roller bearings in particular – and gears. Furthermore, these
components are generally the most loaded and consequently subject to early damages in the machine’s life.

The common feature of these critical components is the presence of repeated contacts between their internal elements e.g. the balls rolling on the inner and outer races of the bearing, or the teeth of two gear wheels, which roll on each other. A lack of lubrication or the smallest steel particle entering the contact zone induces strong concentrated stresses, which can cause damages such as surface scaling and notching of contact elements or cracks in gear teeth. Even if those damages are small, they produce abnormal vibrations from which fault features can be retrieved.

This paper presents different approaches of a statistical method called Principal Components Analysis (PCA), and which can be used for detecting early symptoms of potential faults in machine components. These methods are based on the comparison of the current machine working state with a reference working state. The working state of the machine is determined from vibration measurements at different locations and during a given operation cycle of the machine. PCA is based on the singular value decomposition (SVD) of the observation matrix. Subspaces associated to a selected set of singular values are constructed, and angles between subspaces are compared for different machine working states in order to detect possible faults.

The proposed methodology is applied to the real case of a rotating machinery-testing bench with faulty roller bearings (defects on inner and outer races). Results are compared for the different statistical methods.

2 Theory of principal component analysis

Instead of performing an exact modal identification to compute the trajectories covered by the measurements, it appears more efficient to directly identify the sensor principal components, also called principal directions.

Let \( Q(t) \) denote a discrete block time-history of \( n_s \) sampled responses at \( b \) time instants, assumed to be normalized:

\[
Q = \begin{pmatrix}
q_1(t_1) & \cdots & q_1(t_b) \\
\vdots & \ddots & \vdots \\
q_{n_s}(t_1) & \cdots & q_{n_s}(t_b)
\end{pmatrix}
\]

where \( n_s \), the number of sensor (generally, \( b >> n_s \)), is assumed to be greater than the number of involved structural modes in order to assess a redundancy of data.

The singular value decomposition of the observation matrix \( Q \) gives:

\[
Q = U \Sigma V^T
\]

where \( U \) is an orthonormal matrix \((n_s \times n_s)\) for which the columns lay in the geometrical subspace generated by the sensors. Each column of \( U \) is associated
with the \((b \times b)\) matrix \(V\) of time coefficients. The singular values, given by the \((n_s \times b)\) diagonal matrix \(\Sigma\) and sorted in descending order, can be related to the energies associated with the corresponding principal components of \(U\). This means that the structure mainly reacts in the directions of the principal components associated with the highest energies. Note that it is computationally more efficient to calculate the SVD of the correlation matrix:

\[
QQ^T = U\Sigma^2U^T
\]

Theoretically, only the first \(m+1\) eigenvalues of \(Q\) are non-zeros. Nevertheless, we know that test data contains sensor noise. Since noise has much lower energy than the structural modes, the components of \(U\) associated with eigenvalues presenting an order of magnitude much lower than the others have to be discarded from the principal component base.

In the linear case, the principal directions extracted from test data always lay in the subspace (or hyper-plane) generated by the participating modes (see figure 1 for a 3-D illustration). Mathematically speaking, this means that the so-called principal hyper-plane is invariant, even if the directions of the principal vectors are dependent of the structural excitation.

PCA may then be considered as a powerful and straightforward approach to compute a modal metric of test data and to detect potential structural damages by comparing reference and current structural states.

One way to compare hyper-planes is to use the concept of angles between two subspaces which is illustrated in figure 1.

This concept, introduced by Jordan [6], allows quantifying the spatial coherence between two time-history blocks of an oscillating system. Let \(A \in \mathbb{R}^{n \times p}\) and \(B \in \mathbb{R}^{n \times q}\) \((p \geq q\) each with linearly independent columns. First a QR factorization allows to compute the orthonormal bases of \(A\) and \(B\):

\[
A = Q_A R_A \quad Q_A \in \mathbb{R}^{n \times p} \\
B = Q_B R_B \quad Q_B \in \mathbb{R}^{n \times q}
\]
Thus, the singular values of $Q_A^T Q_B$ define the $q$ cosines of the principal angles $\theta_i$ between $A$ and $B$:

$$SVD( Q_A^T Q_B ) \rightarrow \text{diag}(\cos(\theta_i)) \quad i = 1...q$$

The largest angle allows quantifying how the subspaces $A$ and $B$ are globally different.

## 3 What about the number of sensors?

It may be shown [3] that the number of sensors $n_s$ has to be greater than the number of involved modes in order to assess a redundancy of data. Nevertheless, in many practical cases, the number of possible sensor locations on the structure is low, because of the presence of moving and rotating components. So, in such cases, the principal components extracted by means of the classical PCA method are not well representative of the system dynamics.

To solve this problem, a technique consists in applying the PCA method to the autocorrelation matrix of the observation matrix $Q$.

## 4 PCA of the autocorrelation matrix

### 4.1 Auto-Regressive model associated to observations

The auto-regressive (AR) model of a time data sequence $y(n)$ may be computed in the linear form:

$$\hat{y}(n) = \sum_{i=0}^{N-1} w_i \cdot y(n-i) + e(n)$$

where $\hat{y}(n)$ is the prediction of the signal, $e(n)$ is supposed to be a Gaussian white noise, $w_i$ are the $N$ constant weight coefficients associated to the AR model of the structure.

In the matrix form, we have:

$$\hat{y}(n) = W_N' \cdot Y_N + e(n)$$

with

$$W_N' = (w_0 \quad w_1 \quad \cdots \quad w_{N-1})$$

$$Y_N' = \begin{pmatrix} y(n) & y(n-1) & \cdots & y(n-N+1) \end{pmatrix}$$
4.2 Prediction error

The prediction error of the AR model defined by the squared difference between the observations \( y(n) \) and the estimations \( \hat{y}(n) \):

\[
\varepsilon^2(n) = \left[ y(n) - \hat{y}(n) \right]^2
\]

is usual to express the prediction error of the observations from a statistical point of view:

\[
\Delta = E\{\varepsilon^2(n)\}
\]

where \( E\{ \} \) is the mathematical expectation operator.

The prediction error may be expanded as:

\[
E\{\varepsilon^2(n)\} = E\{\left( y(n) - e(n) \right)^2 \} - 2 \cdot W_N \cdot E\{ Y_N \cdot \left( y(n) - e(n) \right) \} + W_N \cdot E\{ Y_N \cdot Y_N' \} \cdot W_N
\]

The first expectation is simply the mean square power, \( \sigma_e^2 \), of the difference between the observations and the white noise. The second expectation is called the cross-correlation vector and will be denoted by \( p_N \):

\[
p_N = \begin{bmatrix}
   E\{ y(n) \cdot (y(n) - e(n)) \} \\
   E\{ y(n-1) \cdot (y(n) - e(n)) \} \\
   \vdots \\
   E\{ y(n-N+1) \cdot (y(n) - e(n)) \}
\end{bmatrix} = \begin{bmatrix}
   \phi_{ye} (0) \\
   \phi_{ye} (1) \\
   \vdots \\
   \phi_{ye} (N-1)
\end{bmatrix}
\]

where the \( \phi_{\alpha \beta} (m) \) are the cross-correlation coefficients and can generally be estimated by the expression [1]:

\[
\phi_{uv} (m) \equiv \frac{1}{K} \sum_{i=0}^{K-1} u(n-i) \cdot v(n-i)
\]

The third expectation is the autocorrelation matrix, \( R_{yy} \), defined as:

\[
R_{yy} = \begin{bmatrix}
   E\{ Y_N \cdot Y_N' \}
\end{bmatrix}
\]
\[ R_{NN} = E \{ Y_N \cdot Y_N' \} = E \left\{ \begin{bmatrix} y(n) \\ y(n-1) \\ \vdots \\ y(n-N+1) \end{bmatrix} \cdot \begin{bmatrix} y(n) & y(n-1) & \cdots & y(n-N+1) \end{bmatrix} \right\} \]

\[ \approx \begin{bmatrix} \phi_{1y} (0) & \phi_{1y} (1) & \cdots & \phi_{1y} (N-1) \\ \phi_{2y} (1) & \phi_{2y} (0) & \cdots & \phi_{2y} (N-2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{Ny} (N-1) & \phi_{Ny} (N-2) & \cdots & \phi_{Ny} (0) \end{bmatrix} \]

From these definitions, the mean square prediction error \( \Delta \), which is a function of the predictor filter coefficients \( W_N \) is derived:

\[ \Delta(W_N) = \sigma_d^2 - 2 \cdot W_N' \cdot P_N + W_N' \cdot R_{NN} \cdot W_N \]

This expression corresponds to an error « surface » called the Mean Square Error surface (MSE). In the case of two coefficients \( (w_1, w_2) \), it leads to a surface in a three-dimensional space, as illustrated in figure 2.

figure 2 : Location of the minimum of the MSE surface

One of the most important properties of the MSE surface is that it has only one extremum, which is a minimum. At this point, the prediction of the AR model is the most representative of the system structural dynamics. The extremum is reached for the prediction coefficients \( W_N^* \) verifying the two relations listed below:
The solution obtained after substitutions can be written as the single matrix equation:

\[ R_{NN} \cdot W_N^* = p_N \]

### 4.3 Eigenvalues and eigenvectors of the autocorrelation matrix

Let us define a « translated » filter coordinate space by:

\[ V_N = W_N - W_N^* \]

Substituting this expression in the relation giving \( \Delta(W_N) \) and after some algebra [1], one obtains the prediction error in function of the variables \( V_N \) and of the minimum prediction error \( \Delta(W_N)_{\text{min}} \):

\[ V_N' \cdot R_{NN} \cdot V_N = \Delta(W_N^*) \]

In figures 3 a and b, the effect of the « translated » filter on the prediction error is shown. It produces \( MSE \) contours, which correspond to the iso-values of the prediction error, centered about the origin. It is also seen that the \( MSE \) contours have a set of principal axes, which, in general, are not aligned with the \( v_i \) coordinate axes.

Since the autocorrelation matrix \( R_{NN} \) is real and symmetric, it can be shown [1] that there exists a similarity transformation for \( R_{NN} \) in terms of its eigenvalues and eigenvectors. This transformation is given by:

\[ R_{NN} \cdot M_{NN} = \Lambda_{NN} \cdot M_{NN} \]

We also have the following equation:

\[ V_N = M_{NN} \cdot U_N \]

which allows the \( MSE \) contours to be aligned with the coordinate axes \( U_N \), as illustrated in figure 3 c.
4.4 Damage detection using the autocorrelation matrix

The AR model, associated with the observations \( y(t) \), is representative of the system structural dynamics. One can keep a close eye on the validity of the AR model during the life of the structure. The presence of a structural damage is detected by looking for any changes in the reconstruction parameters \( w_i \).

Instead of computing all the parameters \( w_i \), the method proposed in this work consists in the detection of any rotation of the principal coordinate axes of the autocorrelation matrix \( R_{NN} \). This matrix is dependent of the resonance frequencies and the structural modes of the structure; so, information coming from eigenvectors associated with the autocorrelation matrix \( R_{NN} \) can be used for structural damage detection.

5 Experimental setup

In this section, actual vibration data are collected to check the ability of the PCA methods to detect damage. The vibration data are generated with a “Machinery Fault Simulator” manufactured by SpectraQuest, Richmond, VA, 1997. The testing bench allows measurements of vibration data with the rotor operating under a large variety of fault conditions.
The “Machinery Fault Simulator” is able to simulate different types of faults. For example, bearing fault vibration is generated by replacing the rear bearing with another one which specified defects are provided by the manufacturer. Two kinds of bearing fault vibrations are produced: defect on bearing inner raceway (BPFI) and defect on bearing outer raceway (BPFO). In addition, the normal bearing vibration is also measured in order to have a reference state.

![figure 5](image1.png) a) Roller bearings ; b) Faulty bearing setting

The vibration signals are measured by means of four piezoelectric sensors located on the horizontal and the vertical axes – both perpendicular to the rotation axis – of each shaft bearing. The signals are transmitted to a PCB amplifier and a Siglab DSP Technology Inc. acquisition module for signal conditioning. The overall data acquisition system setup is shown in figure 6.

![figure 6](image2.png) a) Sensors setup ; b) Acquisition system

The rotation speed of the shaft is controlled by the simulator. The vibration signals under normal conditions and bearing inner and outer raceway defects are measured at five shaft rotation frequencies: 10, 20, 30, 40 and 52 Hz.

The frequency bandwidth is 0-500 Hz with a sampling frequency ($f_s$) of 1280 Hz. The number of samples ($N_b$) on the observation period ($T$) is equal to 4096.
6 Experimental results – Classical PCA method

In this section, the classical PCA method described previously is applied to the vibration signals measured on the testing bench in order to detect the roller bearing faults.

The observation matrix $Q$ is formed, each line corresponding to the response of each sensor. To represent the variation of the angles during the observation period, the $Q$ matrix is segmented into 200 parts corresponding to different time data blocks.

The first step of the procedure begins with the determination of the number of participating principal directions in the structural responses. By inspecting the principal singular values (also called inertia) of $\Sigma$, it can be shown, in figure 7, that the first 2 principal directions cover more than 90% of the vibration energy.

![Figure 7: Normalized principal singular values of the reference sampling data](image)

Figure 8 shows the subspace angles between the three considered working states of the machine (the reference state corresponding to the undamaged bearing, the BPFO state corresponding to a small defect located on the outer race of the bearing and the BPFI state corresponding to a defect located on the inner race of the bearing) and the reference working state. This result refers to the machinery rotating at a frequency of 40 Hz.

In each case, the angles are represented in function of the time block number. It can be observed that the reference subspace angle remains low (less than 5°), while the angle reaches 15° for the BPFO faulty bearing and 38° for the BPFI faulty bearing. Consequently, this result can be used for a detection of bearings faults.

To show the robustness of the technique, the angles between subspaces are computed for other machinery rotating frequencies. For example, a 20 Hz rotating frequency gives the results shown in figure 9.
However, even if the classical PCA method allows bearing damage detection, it has two important drawbacks. The first one is related to the low number of possible measurement points on a rotating machinery. The number of sensors is equal to the number of the available components of the PCA technique. So, there are not enough components to assess a redundancy of data and to allow the separation of the components linked to the noise perturbations. The other drawback is related to the high sensitivity of the classical PCA method to a variation of the rotation speed. Figure 10 shows the subspace angles between the reference state at 20 Hz and 40 Hz, and the reference state at 20 Hz. It may be seen that the method indicates a damage (angle > 60°) even if no defect is present.
As already explained, this variant of the method consists in applying PCA to the autocorrelation matrix associated with the observation matrix $Q$. In this case, the number of components becomes a parameter (the order of the technique) and does not depend on the number of sensors. So it’s possible to perform the autocorrelation PCA method with one sensor only. The order has to be high enough to clearly separate the features depending on the system and the features associated with perturbations i.e. white noise. In this example, a value of 48 is used for the order.

The results are shown in figures 11 to 13. As we can see, these results can be used for detecting bearing damages. Furthermore, the drawback of the classical PCA technique concerning the sensitivity of the rotation frequency disappears with the autocorrelation PCA method, as illustrated in figure 14.
figure 12 : Comparison of subspace angles for the three machinery working states (40 Hz)

figure 13 : Comparison of subspace angles for the three machinery working states (20 Hz)

figure 14 : Comparison of subspace angles for two rotation frequencies (20 Hz and 40 Hz)
8 Conclusion

In rotating machinery, the number of measurement locations is usually low. Furthermore, the rotation speed of the shafts cannot generally be kept constant. In these cases, the classical PCA method has two important drawbacks. Nevertheless, the autocorrelation PCA method gets round of these because the order $N$ (the number of singular values) becomes a parameter and does not depend on the number of sensors, and because the sensitivity of the results to a variation of the rotation frequency is lower.

PCA techniques are very effective to detect small defects on the inner or outer race of a roller bearing. Furthermore, the statistical techniques presented here are computationally cheap and, therefore, could be easily used for on-line maintenance strategy. They can also be used for any type of defects in rotating machinery, which considerably enlarges their application domain.

9 References


