Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Sound and Vibration 272 (2004) 317–339
JOURNAL OF SOUND AND VIBRATION
www.elsevier.com/locate/jsvi

Blind source separation of noisy harmonic signals for rotating machine diagnosis

C. Servière*, P. Fabry

Laboratoire des Images et des Signaux (LIS), ENSIEG, BP 46, 38402 Saint Martin d'Hères cedex, France

Received 10 July 2002; accepted 17 March 2003

Abstract

Blind source separation (BSS) consists of recovering signals from different physical sources from several observed combinations independently of the propagation medium. BSS is also a promising tool for non-destructive machine condition monitoring by vibration analysis, as it is intended to retrieve the signature of a single rotating machine from combinations of several working machines. In this way, BSS can be seen as a pre-processing step that improves the diagnosis.

BSS methods generally assume observations that are either noise-free or corrupted with spatially distinct white noises. In the latter case, principal component analysis (PCA) is applied as a first step to filter out the noise and whiten the observations. Obviously, the efficiency of the whole separation procedure depends on the accuracy of the first step (PCA).

However, in the real world, signals of rotating machine vibration may be severely corrupted with spatially correlated noises and therefore the signal subspace will not be correctly estimated with PCA. The purpose of this paper is to propose a 'robust-to-noise' technique for the separation of rotating machine signals. The sources are assumed here to be periodic and so can be modelled as the sum of sinusoids of harmonic frequencies. A new estimator of the signal subspace and the whitening matrix is introduced which exploits the model of sinusoidal sources and uses spectral matrices of delayed observations to eliminate the influence of the noise. After whitening, the second step of source separation remains unchanged. Finally, performance of the algorithm is investigated with artificial data and experimental rotating machine vibration data.

1. Introduction

Blind source separation (BSS) consists in recovering signals of different physical sources from a finite set of observations recorded by sensors. Under the only hypothesis of mutually independent

*Corresponding author. Tel.: +33-4-76-896411; fax: +33-4-76-826384.
E-mail address: christine.serviere@lis.inpg.fr (C. Servière).

0022-460X/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0022-460X(03)00774-0

sources, BSS extracts the contributions of the sources independently of the propagation medium. These methods have been successfully used in many fields such as medicine, telecommunications or audio processing [1].

BSS is also a promising tool for non-destructive machine condition monitoring by vibration analysis, as it is aimed at the retrieval of the signature of a single rotating machine from mixtures of several working machines [2–7]. Specific methods of fault detection could then be applied to the signatures of each machine to be diagnosed. In this way, BSS can be seen as a pre-processing step that improves the diagnosis.

The crucial point in BSS methods remains that the observations are generally assumed to be noise-free or corrupted with spatially white noises [8]. Usually, noisy BSS techniques apply principal components analysis (PCA) as a first step, to filter out the noise and whiten the observations (if more observations than sources are recorded) [9]. In short, PCA exploits the only hypothesis of decorrelation between sources and projects the observations on an orthonormal base of the signal subspace. The estimation of the signal subspace requires either knowledge of the covariance noise matrix, or the hypothesis of spatially white noises. After this step, projections of the observations on the signal subspace are uncorrelated but are still mixtures of the sources. Further separation is then achieved using the stronger hypothesis of mutually independent sources to recover the sources [8]. Obviously, the efficiency of the whole process depends on the accuracy of the first step (PCA). In other words, without any a priori knowledge on the additive noises, the performance of BSS relies on the availability of the assumption of spatially white noises.

However, in the real world, signals of rotating machine vibration may be severely corrupted with spatially correlated noises coming from a noisy environment and therefore the signal subspace will not be correctly estimated [10]. In such an experimental context, PCA and thus BSS lose efficiency.

The purpose of this paper is to propose a robust-to-noise technique for the separation of rotating machine signals. The sources are assumed here to be periodic and so can be modelled as the sum of sinusoids of harmonic frequencies; for example, harmonics of the rotation frequencies [10,11]. No assumption is made regarding the statistical property of the noise component. The only hypothesis is that the autocorrelation lengths of the sources are larger than the autocorrelation lengths of all the noises. As the standard PCA cannot provide a correct estimate of the signal subspace in a noisy situation, a new estimator of the signal subspace and the whitening matrix is introduced. It exploits the model of harmonic sources and uses spectral matrices of delayed observations to eliminate the noise influence. After whitening, the second step of source separation remains unchanged. The whole method is semi-blind in the sense that the propagation medium and the noise spectral matrices are unknown.

Finally, performance of the algorithm is investigated with artificial data and experimental rotating machine vibration data.

2. Principle of BSS

2.1. Introduction

BSS consists in recovering signals of different physical sources from the observation of several combinations of them. Typically, the observations are obtained as the output of a set of sensors, where each sensor receives a different combination of source signals. The adjective "blind" indicates that the source signals are not observed and also that no information is available about the combinations and the noises. This approach is usually used when it is too difficult to model the transfer from the sources to the sensors. The lack of knowledge about the combinations and the sources is compensated by the hypothesis of mutually independent sources. Nevertheless, signals of different physical sources generally satisfy this condition. This assumption allows exploitation of the spatial diversity provided by many sensors and is the fundamental basis of BSS.

The general model of BSS is shown in Fig. 1. p statistically independent signals (called sources) $\mathbf{S}(n) = [s_1(n), \ldots, s_p(n)]$ are received on m sensors ($m \geqslant p$) through an unknown propagation medium at the discrete time n. The observation vector $\mathbf{X}(n) = [x_1(n), \ldots, x_m(n)]$ is therefore a mixture of the p sources. In the most general case the mixture $\mathbf{X}(n)$ is described by the unknown mixing matrix $\mathbf{A}(n)$ and is corrupted with additive noise $\mathbf{N}(n) = [n_1(n), \ldots, n_m(n)]$. Noise is also supposed to be independent of the sources. Both the sources and the noises are required to be stationary and have a zero mean.

Most research is done for BSS with a linear instantaneous mixture model (see Refs. [8,9,12]) whereas the unknown mixing matrix $\mathbf{A}(n)$ is supposed to be time-independent. This model is only applicable if the propagation delay through the medium is very short (in relation to the sampling period). It is commonly established in fields such as telecommunications or biomedicine. However, mechanical systems are one example in which the instantaneous assumption does not always hold true due to the transfer delays of the vibrations through the structures. Usually, the model used for this application is a linear time dependent mixture, called a convolutive mixture [4,5,7,10]. Supposedly, this is related to the rigidity and size of the structure under investigation [13]. Large and elastic structures will account for severe filtering and delays in the transmission, whereas small and rigid structures may show approximately instantaneous mixtures [13].

BSS methods consist of finding an estimate $\mathbf{R}(n)$ of the sources $\mathbf{S}(n)$ as an unknown linear transformation of the observations $\mathbf{X}(n)$. This separating function is adapted from the data and attempts to equalize the effect of the propagation medium. In the noise free model, it is known to lead to components of $\mathbf{R}(n)$ as independent as possible [8,14], after a theorem of Darmois. Many approaches explore the different formulations of statistical independence, cross-cumulant cancellation, entropy minimization, likelihood approaches, etc., for a complete survey on statistical principles for BSS see Refs. [15–17]. In the noisy case, PCA is first used to eliminate a major part of the noise [9].

Several approaches can be found in the literature to realize BSS with a convolutive mixing model, using either a temporal model [18–20], or a frequency mixing model [5,10,21,22]. Both are detailed below in Sections 2.2 and 2.3. Later, inherent ambiguities to the BSS problem are stressed.

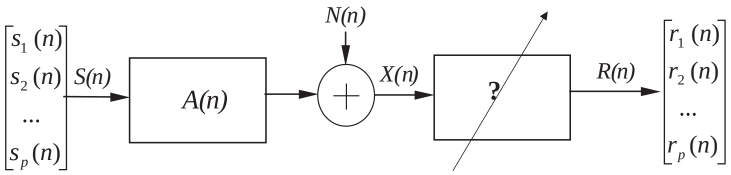

Fig. 1. BSS general scheme.

2.2. The temporal convolutive mixing model

The general model of a convolutive mixture can be represented as shown in Fig. 2.

Each $A_{ij}(z)$ represents the linear transfer function from the ith source to the jth sensor and is given by $A_{ij}(z) = \sum_{l=0}^{\infty} A_{ij}(l) \cdot z^{-l}$. The whole mixing system can be summed up as

$$\mathbf{A}(z) = \begin{bmatrix} A_{11}(z) & \dots & A_{1m}(z) \\ \vdots & \ddots & \vdots \\ A_{p1}(z) & \dots & A_{pm}(z) \end{bmatrix}. \tag{1}$$

So, the observation vector $\mathbf{X}(n)$ can be written as a convolution between the sources $\mathbf{S}(n)$ and the mixing process. Using the z transform, the ith component of $\mathbf{X}(z)$, $x_i(z)$, is described by

$$x_i(z) = \sum_{k=1}^{p} A_{ki}(z) s_k(z) + n_i(z), \quad i = 1, \dots, p. \tag{2}$$

System (2) is usually noted with a matrix formulation:

$$\mathbf{X}(n) = [\mathbf{A}(z)] \cdot \mathbf{S}(n) + \mathbf{N}(n), \tag{3}$$

where z^{-1} is both the backward-shift operator i.e., $z^{-1} \cdot \mathbf{S}(n) = \mathbf{S}(n-1)$ as well as the complex variable in the z transform. So, the aim of separation is to estimate a stable inverse system of $\mathbf{A}(z)$ i.e., a filter $\mathbf{C}(z)$ such as

$$\mathbf{R}(n) = [\mathbf{C}(z)] \cdot \mathbf{X}(n) = \mathbf{S}(n) + [\mathbf{C}(z)] \cdot \mathbf{N}(n). \tag{4}$$

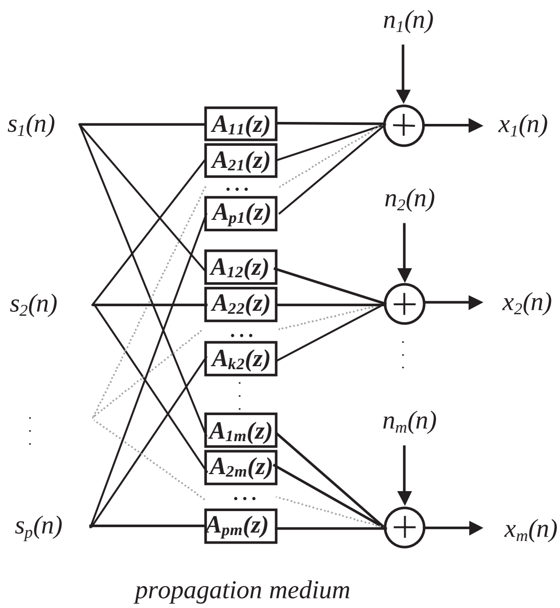

Fig. 2. Convolutive mixing model.

2.3. The frequency convolutive model

In frequency domain, the convolutive mixture is reduced to an instantaneous complex mixture for each frequency bin. Let $\mathbf{X}(n,f)$ (respectively, $\mathbf{S}(n,f)$ and $\mathbf{N}(n,f)$) be the L-points discrete Fourier transform (DFT) of the data block of the data vector $[\mathbf{X}(n), \ldots, \mathbf{X}(n+L-1)]$ (respectively $[\mathbf{S}(n), \ldots, \mathbf{S}(n+L-1)]$ and $[\mathbf{N}(n), \ldots, \mathbf{N}(n+L-1)]$). $\mathbf{X}(n,f)$ is a vector that denotes the sliding DFT along the time axis starting at time n for each signal component. With these notations, the mixing model can be written as

$$\mathbf{X}(n,f) = \mathbf{A}(f)\mathbf{S}(n,f) + \mathbf{N}(n,f), \quad f = 0, \ldots, L-1. \tag{5}$$

The transfer matrix $\mathbf{A}(f)$ $(m \times p)$ characterizes the linear propagation from p sources to m sensors and must be pseudo-invertible to recover the p sources at frequency bin f. Then, the aim of BSS is to estimate a separating matrix $\mathbf{C}(f)$ whose the p outputs are

$$\mathbf{R}(n,f) = \mathbf{C}(f) \cdot \mathbf{X}(n,f) = \mathbf{S}(n,f) + \mathbf{C}(f) \cdot \mathbf{N}(n,f). \tag{6}$$

So, $\mathbf{R}(n,f)$ represents an estimate of the vector $\mathbf{S}(n,f)$. The ideal result is obtained with $\mathbf{C}(f)$ equal to the pseudo-inverse of $\mathbf{A}(f)$.

2.4. Indeterminacies and uniqueness

Intrinsically, two ambiguities are inherent to the BSS problem. First, the mathematical independence is insensitive to a source permutation. For instance, consider the previous frequency mixing model (5). If $\mathbf{P}(f)$ is assumed to be a permutation matrix, the mixture model can be re-written in the noise free case ($\mathbf{N}(n) = 0$) as

$$\mathbf{X}(n,f) = [\mathbf{A}(f) \cdot \mathbf{P}(f)^{-1}] \cdot \mathbf{P}(f) \cdot \mathbf{S}(n,f), \tag{7}$$

where the elements of $\mathbf{P}(f) \cdot \mathbf{S}(f)$ are the permuted original sources and the mixing matrix $\mathbf{A}(f) \cdot \mathbf{P}(f)^{-1}$ is a new mixing matrix estimated by the BSS algorithm. It implies that it is impossible to know the original labelling of the sources.

Given that the mathematical independence is insensitive to a scaling factor applied on the sources, the second indeterminacy is that it is impossible to uniquely identify the sources. Hence, from Eq. (5) the observations can be expressed as

$$\mathbf{X}(n,f) = [\mathbf{A}(f) \cdot \mathbf{D}(f)^{-1}] \cdot \mathbf{D}(f) \cdot \mathbf{S}(n,f), \tag{8}$$

where $\mathbf{D}(f)$ is any diagonal matrix.

So, any scalar multiplier in one of the sources $S_i(n,f)$ can be cancelled by dividing the corresponding column of the matrix $\mathbf{A}(f)$.

In short, in the frequency domain, the assumption of independent components allows the matrix $\mathbf{C}(f)$ to be estimated such that the product $\mathbf{C}(f) \cdot \mathbf{A}(f)$ is equal to a diagonal matrix $\mathbf{D}(f)$ up to a permutation matrix $\mathbf{P}(f)$:

$$\mathbf{C}(f) \cdot \mathbf{A}(f) = \mathbf{D}(f) \cdot \mathbf{P}(f). \tag{9}$$

In time-domain, the scaling indeterminacy is a filtering indeterminacy and Eq. (3) does not define the filters $\mathbf{A}(z)$ uniquely (i.e., they are not identifiable) but up to a linear filtering $\mathbf{D}(z)$ and a sources permutation.

The non-uniqueness of the BSS results can set many problems to the user, especially in monitoring or diagnosis purposes. It is possible, however, to reduce the shape indeterminacy in the model by setting a constraint either on the matrices $\mathbf{A}(z)$ (the diagonal terms of $\mathbf{A}(z)$ are usually supposed to be unity) [5,16,18] or on $\mathbf{S}(n,f)$ (its components are generally assumed to have unit variance) [5,8]. However, the filter indeterminacy can be partially circumvented by restoring the contribution of each source on each sensor with a correlation measure between the sources $\mathbf{R}(n)$ (or $\mathbf{R}(n,f)$) estimated with BSS and the observations $\mathbf{X}(n)$ (or $\mathbf{X}(n,f)$) [4,5].

However, if the above-mentioned order ambiguity is not disturbing in the time domain BSS, it is the main obstacle when applying BSS in the frequency domain as a different permutation of different frequency components leads to mixed outputs and to degraded separation results. To remove the permutation indeterminacy, several approaches exist in the literature [21–26]. For instance, some algorithms aim to order the output channel according to the maximal correlation between the frequency components [21–24] or their envelopes [25], or according to their fourth order cumulants [26]. Moreover, a specific approach is proposed for rotating machine data in Refs. [7,27].

3. BSS of noisy convolutive mixtures

3.1. Introduction

Many approaches can be found in the literature to realize BSS with a convolutive mixing model, using either a temporal model (Section 2.2) [18–20], or a frequency mixing model (Section 2.3) [5,10,21,22]. An overview is proposed in Ref. [28]. First, it is shown in Section 3.2 how the second order statistics of data (which only exploit hypothesis of decorrelation of sources) are not generally sufficient to solve the BSS problem. Indeed, they need extra information contained in the hypothesis of sources independence to achieve their objective, as detailed later in Section 3.6.

The estimation of the noise-free model seems to be a challenging task in itself, and thus the noise is most often omitted. Many functions, sometimes called contrast functions, have been proposed for instantaneous mixture (or ICA) in the noise-free case [15–17,29]. They all measure the independence between the components of the BSS system outputs $\mathbf{R}(n)$ or $\mathbf{R}(n,f)$, as the separating matrices $\mathbf{C}(z)$ or $\mathbf{C}(f)$ must perform signals that are as independent as possible in the noise-free model. Whatever the envisaged method, results seem to be very accurate.

Few methods exist for noisy instantaneous BSS. Practically all methods taking noise explicitly into account assume that the additive noise is either spatially white [8,9,30] or sometimes Gaussian [31–33]. Often, PCA is widely used as a pre-processing for its denoising capabilities [9,30]. Its basic goal is first to reduce the dimension of the data. Assuming strictly more sensors than sources ($m > p$), a projection of the data into the p dimensional PCA subspace (also called the signal subspace) removes noise. The estimation of the signal subspace requires either knowledge of the covariance noise matrix, which is difficult to obtain in practice, or the hypothesis of spatially white noises [9,30]. The previous contrast functions can afterwards be applied to the de-noised observations.

BSS with a convolutive mixing model can be achieved using either a temporal or a frequency model. Some contrast functions can be extended to the convolutive case, expressed in the time-domain [18–20]. However, the resulting optimizations are not easy, as there usually exist local maxima [34]. The optimization is generally performed in an adaptive way and therefore no noise can be taken into account [5,18–20]. Yet, as convolutive mixtures can be reduced to several instantaneous complex mixtures after DFT (Section 2.3), PCA can be used in each frequency bin to de-noise the observations.

The paper focuses on the application of BSS of rotating machine signals by vibration analysis. In an industrial environment, a rotating machine signal may be severely corrupted by other interfering sources and noise, which are often much louder. Negative signal-to-noise ratios occur frequently [7]. Consequently, in this paper, emphasis is placed on the convolutive model expressed in the frequency domain while additive noises can be taken into account with the wide hypothesis ($m > p$). First, the principle of PCA is recalled in Section 3.3 and next its application for the de-noising of spatially white noises in Section 3.4. However vibration signals are generally corrupted with spatially correlated noises [35]. Hence, a modified PCA is developed for spatially correlated noises in Section 3.5, adding the model of periodic sources. While the standard PCA cannot provide a correct estimate of the signal subspace in this noisy situation, a new estimator of the whitening matrix and the signal subspace is introduced. It exploits the model of harmonic sources and uses spectral matrices of delayed observations to eliminate the noise influence. Finally, the principle of independence measurement and some principal contrast functions are briefly reviewed.

3.2. Second order inadequacy

Initially, a constraint of unit variance is commonly set on the sources $\mathbf{S}(n,f)$ to eliminate the scaling indeterminacy [8,9,17]. The general model then becomes

$$\mathbf{X}(n,f) = \mathbf{A}(f) \cdot \mathbf{S}(n,f) + \mathbf{N}(n,f) \quad \text{with } \mathrm{E}\{\mathbf{S}(n,f)\mathbf{S}(n,f)^{+}\} = \mathbf{I}_p, \tag{10}$$

where $^{+}$ denotes the transconjugate and $\mathbf{I}_p$, the identity matrix $p \times p$. After a singular value decomposition (SVD), the mixing matrix $\mathbf{A}(f)$ can be expressed as a product of three matrices:

$$\mathbf{A}(f) = \mathbf{V}(f) \cdot \mathbf{\Delta}(f)^{1/2} \cdot \mathbf{\Pi}(f), \tag{11}$$

where $\mathbf{V}(f)$ and $\mathbf{\Pi}(f)$ are two unitary matrices respectively $m \times m$ and $p \times p$. $\mathbf{\Delta}(f)^{1/2}$ is a $m \times p$ diagonal matrix with elements $\sqrt{\lambda_i}$, $i = 1, \ldots, p$. The eigenvalues are assumed to be ranged in decreasing order.

To understand the inadequacy of second order statistics, the spectral matrix of the noise-free observations $\mathbf{Y}(n,f)$ is considered:

$$\mathbf{Y}(n,f) = \mathbf{A}(f) \cdot \mathbf{S}(n,f). \tag{12}$$

Given that the spectral matrix of $\mathbf{Y}(n,f)$ is equal to

$$\begin{aligned} \mathbf{R_Y}(f) &= \mathrm{E}[\mathbf{Y}(n,f)\mathbf{Y}(n,f)^{+}] = \mathrm{E}[(\mathbf{A}(f)\mathbf{S}(n,f))(\mathbf{A}(f)\mathbf{S}(n,f))^{+}] \\ &= \mathbf{A}(f)\mathbf{A}(f)^{+} = \mathbf{V}(f)\mathbf{\Delta}(f)\mathbf{V}(f)^{+} \end{aligned} \tag{13}$$

it may be seen that $\mathbf{R_Y}(f)$ does not depend on the rotation matrix $\mathbf{\Pi}(f)$, which is necessary to identify the matrix $\mathbf{A}(f)$. In short, the BSS problem cannot be solved only using second order statistics because independence is a stronger condition than uncorrelation. To estimate the missing unitary matrix $\mathbf{\Pi}(f)$, an independence criterion is needed using the spatial diversity provided by the mixing model (see Refs. [8,9,15,17]).

3.3. Principle of PCA

PCA is a standard technique to reduce the dimension of data. It is also a first step for BSS techniques to spatially whiten the observations and filter out additive noises if the observations number m is assumed to be greater than the sources number p [9,30,36]. Note the general mixing model:

$$\mathbf{X}(n,f) = \mathbf{A}(f) \cdot \mathbf{S}(n,f) + \mathbf{N}(n,f) \quad \text{with } \mathrm{E}\{\mathbf{S}(n,f)\mathbf{S}(n,f)^{+}\} = \mathbf{I}_p$$
$$\text{and} \quad \mathbf{A}(f) = \mathbf{V}(f) \cdot \mathbf{\Delta}(f)^{1/2} \cdot \mathbf{\Pi}(f).$$

PCA projects the m dimensional noisy observation $\mathbf{X}(n,f)$ on an orthonormal base of the signal subspace of dimension p with a $(p \times p)$ dimensional whitening matrix $\mathbf{W}(f)$. The signal subspace is spanned by the p first eigenvectors of the spectral matrix of the noise-free observations $\mathbf{R_Y}(f)$ i.e., the p first columns of matrix $\mathbf{V}(f)$ [30,36]. The whitening matrix $\mathbf{W}(f)$ verifies relation (14):

$$\mathbf{W}(f) \cdot \mathbf{R_Y}(f) \cdot \mathbf{W}(f)^{+} = \mathbf{I}_p \tag{14}$$

with $\mathbf{R_Y}(f) = \mathbf{V}(f)\mathbf{\Delta}(f)\mathbf{V}(f)^{+}$. Therefore the solution for $\mathbf{W}(f)$ is given as

$$\mathbf{W}(f) = \mathbf{\Delta}_s(f)^{-1/2}\,\mathbf{V}_s(f)^{+} \tag{15}$$

where $\mathbf{\Delta}_s(f)$ is the square submatrix containing the first p diagonal elements of $\mathbf{\Delta}(f)$ and $\mathbf{V}_s(f)$, the rectangular submatrix containing the first p columns of $\mathbf{V}(f)$.

While $\mathbf{\Delta}_s(f)$ and $\mathbf{V}_s(f)$ are given by the eigenvalues and the eigenvectors of $\mathbf{R_Y}(f)$, PCA exploits only the second order statistics contained in the spectral matrix of the noise-free observations to identify the whitening matrix $\mathbf{W}(f)$ at each frequency bin f.

The observations $\mathbf{X}(n,f)$ are then projected on the subspace and whitened with the help of the whitening matrix. The resulting data $\mathbf{Z}(n,f)$ are written as a p dimensional vector $\mathbf{Z}(n,f)$:

$$\mathbf{Z}(n,f) = \mathbf{W}(f) \cdot \mathbf{X}(n,f) = \mathbf{\Pi}(f) \cdot \mathbf{S}(n,f) + \mathbf{W}(f) \cdot \mathbf{N}(n,f). \tag{16}$$

By a projection on the signal subspace, all the contributions of the noise that are contained in the noise space are taken out. This allows an important noise reduction when the number of sensors is much greater than the sources number [36].

In these conditions, the noise residues are generally neglected after PCA and the data are assumed to be [9,17]:

$$\mathbf{Z}(n,f) = \mathbf{\Pi}(f) \cdot \mathbf{S}(n,f). \tag{17}$$

The problem is then reduced to the general noise-free case. The data are still linked to the components of the source vector $\mathbf{S}(n,f)$ by a remaining unitary transformation matrix $\mathbf{\Pi}(f)$ which is identified by a measure of independence (detailed in Section 3.6).

3.4. PCA with noisy observations

It has been shown in Section 3.3 that PCA involves the eigenvalues and eigenvectors of $\mathbf{R_Y}(f)$ to compute the whitening matrix (i.e., $\mathbf{V}_s(f)$ and $\mathbf{\Delta}_s(f)$). Unfortunately, in a noisy context, the spectral matrix of the noise-free observations $\mathbf{R_Y}(f)$ is not easily accessible and the goal of PCA is to compute $\mathbf{R_Y}(f)$, under some conditions, using only the noisy observations $\mathbf{X}(n,f)$.

Taking into account that the spectral matrix of $\mathbf{X}(f)$ is equal to

$$\mathbf{R}_X(f) = \mathbf{R_Y}(f) + \mathbf{R}_N(f) = \mathbf{V}(f)\mathbf{\Delta}(f)\mathbf{V}(f)^+ + \mathbf{R}_N(f), \tag{18}$$

where $\mathbf{R}_N(f)$ is the noise spectral matrix, it is clear that the problem can be solved if a priori information on $\mathbf{R}_N(f)$ is known. In that particular case, the eigen value decomposition (EVD) of the matrix $(\mathbf{R}_X(f) - \mathbf{R}_N(f))$ provides the whitening matrix.

When the noises are spatially white, the noise spectral matrix is of the form

$$\mathbf{R}_N(f) = \sigma_N^2 \mathbf{I_m} \tag{19}$$

and the EVD of $\mathbf{R}_X(f)$ provides the same eigenvectors as for $\mathbf{R_Y}(f)$ (19), as shown below:

$$\mathbf{R}_X(f) = \mathbf{V}(f) \cdot \underbrace{(\mathbf{\Delta}(f) + \sigma_N^2 \mathbf{I_m})}_{\mathbf{\Omega}(f)} \cdot \mathbf{V}(f)^+. \tag{20}$$

σ_N^2 can be estimated from the $(m-p)$ last eigenvalues of $\mathbf{\Omega}(f)$ and subtracted to the p first eigenvalues to get an estimate of $\mathbf{\Delta}_s(f)$.

With experimental signals, $\mathbf{R}_N(f)$ is hardly ever on the form of Eq. (19) and factorization (20) is no longer possible [35]. Consequently, in the presence of spatially correlated noises, the first p eigenvectors of $\mathbf{R}_X(f)$ no longer span the signal subspace but a p dimensional subspace in the m dimensional space of observations. The whitening matrix is not correctly estimated because of errors in the estimation of both $\mathbf{V}_s(f)$ and $\mathbf{\Delta}_s(f)$. The PCA loses efficiency while the noisy observations are projected on a wrong subspace. Therefore the model after PCA step is no longer equal to

$$\mathbf{Z}(f) \neq \mathbf{\Pi}(f) \cdot \mathbf{S}(f) \tag{21}$$

and the second step of estimation of the remaining matrix $\mathbf{\Pi}(f)$ also fails.

In the case of mechanical signals from rotating machines, a priori information on the noise spectral matrix is not available. So, a model of periodic sources is added in the next section to replace this lack of information. The model of periodic sources is then exploited to get a new estimator of the whitening matrix $\mathbf{W}(f)$, which is robust for spatially correlated noises.

3.5. PCA with periodic sources

A direct estimator of the noise-free spectral matrix $\mathbf{R_Y}(f)$ is proposed in this section, computed from spectral matrices of delayed observations. The sources are first assumed here to be signals from rotating systems and so are periodic. Their autocorrelations are also periodic and the autocorrelation lengths are therefore infinite. For this reason, the autocorrelation lengths of the sources are assumed to be larger than the correlation lengths of all the noises. Thus an estimator of the noise-free spectral matrix $\mathbf{R_Y}(f)$ is derived from spectral matrices of delayed observations, as they eliminate the noise influence.

3.5.1. Modeling the observations issued of rotating machines

Particular attention is paid to the types of signals commonly encountered in machine condition monitoring applications. Since many machines contain rotating parts, the vibration fields they generate are often periodic (or quasi-periodic) [4,5,7,10,11,37] and their spectra contain principally spectral lines. Based on the principle that a line is a sinusoidal component with steady frequency, at least at the scale of the temporal sequence examined, mixed with noise [37], each temporal source $s_i(n)$ is modelled here as a sum of P sinusoids of constant amplitudes A_{ik}, phases ϕ_{ik} and frequencies f_{ik}. Some frequencies can be for instance harmonic to the rotation frequency:

$$s_i(n) = \sum_{k=1}^{P} A_{ik} \sin(2\pi f_{ik} n + \phi_{ik}). \tag{22}$$

The observations from rotating machines are assumed to be convolutive mixtures of these periodic signals (22), corrupted with additive noises that may be spatially and temporally correlated. Both the sources and the noises are required to be stationary and have zero mean. Although it is a simple model, it will be seen later that it provides good results even on experimental signals (Section 4.2) [35]. The periodic signals may contain closely spaced frequencies belonging to various sources. In one frequency bin f, several physical sources (and so different sinusoids of frequencies f_{ik}) can be mixed. Hence, the mixing model is changed into:

$$\mathbf{X}(n,f) = \mathbf{A}(f) \cdot \mathbf{S}(n,f) + \mathbf{N}(n,f) \quad \text{with } \mathrm{E}\{\mathbf{S}(n,f)\mathbf{S}(n,f)^{+}\} = \mathbf{I}_p. \tag{23}$$

The number p of sources then becomes the number of different frequencies that are mixed in the frequency bin f. If $S_i(n,f)$ is the ith component of source vector $\mathbf{S}(n,f)$, $S_i(n,f)$ is the L-points DFT of a sinusoid $S_i(n)$ of constant amplitude A_{ik}, phase ϕ_{ik} and frequency f_{ik}: $S_i(n) = A_{ik} \sin(2\pi f_{ik} n + \phi_{ik})$. The L-points DFT $S_i(n,f)$ is computed on the data block $[S_i(n), \ldots, S_i(n + L - 1)]$. The sliding window along the signal causes a phase shift since the time origin n is changed for each segment $[S_i(n), \ldots, S_i(n + L - 1)]$.

$$\begin{aligned} S_i(n,f) &= \frac{1}{L} \sum_{l=0}^{L-1} A_{ik} \sin(2\pi f_{ik} n + \phi_{ik}) \exp\left(-2\pi \mathrm{j} \frac{nl}{L}\right) \\ &\approx \frac{A_{ik}}{2} \exp[\mathrm{j}(\phi_{ik} + 2\pi f_{ik} n + \pi(f_{ik} - f)(L-1))] \frac{\sin(\pi(f_{ik} - f)L)}{\sin(\pi(f_{ik} - f))}. \end{aligned} \tag{24}$$

The constraint of unit variance of the source $S_i(n,f)$ eliminates the scaling indeterminacy and requires the terms of $S_i(n,f)$ to be complex of unit modulus, i.e., only phase terms. Therefore $S_i(n,f)$ is reduced to

$$S_i(n,f) = \exp[\mathrm{j}(\phi_{ik} + 2\pi f_{ik} n + \pi(f_{ik} - f)(L-1))]. \tag{25}$$

3.5.2. *Definition of delayed spectral matrix*

The delayed spectral matrix at delay τ is defined as the interspectral matrix between the observations $\mathbf{X}(n)$ and the delayed observations $\mathbf{X}(n + \tau)$. A similar definition is proposed in Ref. [37] for a mono-dimensional signal and is called the autocoherent spectrum. It was developed to improve the extraction of pure sinusoidal components for vibration analysis of a diesel engine [37].

Let $\mathbf{X}(n,f)$ be the L-points discrete Fourier transform (DFT) of the data block of the data vector $[\mathbf{X}(n), \ldots, \mathbf{X}(n+L-1)]$ and let $\mathbf{X}(n+\tau,f)$ be the L-points discrete Fourier transform of the data block $[\mathbf{X}(n+\tau)\ldots\mathbf{X}(n+L-1+\tau)]$.

The delayed spectral matrix, noted $\mathbf{R}_X^\tau(f)$, is defined as

$$\mathbf{R}_X^\tau(f) = \mathrm{E}\{\mathbf{X}(n,f)\mathbf{X}(n+\tau,f)^+\}. \tag{26}$$

An estimator of this delayed spectral matrix, $\hat{\mathbf{R}}_X^\tau(f)$, is obtained from data by replacing the statistical average in Eq. (26) with averages on time segments. It is deduced from an estimator of the spectral matrix $\mathbf{R}_X(f)$. For the latter, its estimate $\widehat{\mathbf{R}}_X(f)$ is usually computed with the Welch method [38]. If the data record is sectioned into K adjacent time segments of L samples each, $\widehat{\mathbf{R}}_X(f)$ is written as:

$$\hat{\mathbf{R}}_X(f) = \frac{1}{K}\sum_{k=0}^{K-1} \mathbf{X}(kL,f)\mathbf{X}(kL,f)^+. \tag{27}$$

An estimator of the delayed spectral matrix arrives naturally from Eq. (27) and is defined by

$$\hat{\mathbf{R}}_X^\tau(f) = \frac{1}{K}\sum_{k=0}^{K-1} \mathbf{X}(kL,f)\mathbf{X}(kL+\tau,f)^+. \tag{28}$$

The statistical properties of estimator (28) are detailed in Ref. [37]. In practice, for a noisy sinusoidal component with steady frequency, the sinusoid appears as a coherent signal in the autocoherent spectrum, with a steady phase relationship, while the incoherent noise level is lessened [37]. It will yield de-noised spectral matrices of noisy periodic sources.

3.5.3. Estimation of the whitening matrix for periodic sources

The PCA problem, previously detailed in Section 3.4, is now considered further. The delayed spectral matrix can be employed in PCA to build the noise-free spectral matrix $\mathbf{R}_\mathbf{Y}(f)$ under some conditions. Since the sources are assumed to be periodic, the autocorrelation lengths of the sources can be supposed to be larger than the correlation lengths of all the noises. Let τ_N be the greater correlation or cross-correlation of the m noises. If $\tau \geqslant \tau_N + L$, at each frequency bin, the matrix $\mathbf{R}_X^\tau(f)$ contains only information about the sources as $\mathbf{R}_N^\tau(f)$ is equal to zero:

$$\mathbf{R}_X^\tau(f) = \mathbf{A}(f)\cdot\mathrm{E}\{\mathbf{S}(n,f)\mathbf{S}(n+\tau,f)^+\}\cdot\mathbf{A}(f)^+. \tag{29}$$

Note that p sinusoids of frequencies f_i are present in the analyzed frequency f. Given Eq. (25) for $S_i(n,f)$, $\mathbf{R}_X^\tau(f)$ can be written as

$$\mathbf{R}_X^\tau(f) = \mathbf{A}(f)\cdot\underbrace{\mathrm{E}\{\mathbf{S}(n,f)\mathbf{S}(n+\tau,f)^+\}}_{\mathbf{R}_S^\tau(f)}\cdot\mathbf{A}(f)^+ \tag{30}$$

with

$$\mathbf{R}_S^\tau(f) = \begin{pmatrix} \mathrm{e}^{-\mathrm{j}2\pi f_1\tau} & & 0 \\ & \ddots & \\ 0 & & \mathrm{e}^{-\mathrm{j}2\pi f_p\tau} \end{pmatrix},$$

where $\mathbf{R}_S^\tau(f)$ is a diagonal matrix with only phase terms, due to the constraint of normalization of the sources variance.

If the mixing matrix $\mathbf{A}(f)$ is replaced with its SVD (Eq. (11)) in Eq. (30):

$$\mathbf{R}_X^\tau(f) = \mathbf{V}(f)\cdot\mathbf{\Delta}(f)^{1/2}\cdot\mathbf{\Pi}(f)\cdot\mathbf{R}_S^\tau(f)\cdot\mathbf{\Pi}(f)^+\cdot(\mathbf{\Delta}(f)^{1/2})^+\cdot\mathbf{V}(f)^+. \quad (31)$$

The aim of this section is to recover the spectral matrix of noise-free observations $\mathbf{R}_\mathbf{Y}(f)$, ($\mathbf{R}_\mathbf{Y}(f) = \mathbf{V}(f)\mathbf{\Delta}(f)\mathbf{V}(f)^+$). The great interest of the delayed interspectral matrix $\mathbf{R}_X^\tau(f)$, versus the classical spectral matrix $\mathbf{R}_X(f)$, lies in the fact that the noise influence is eliminated, as the noise spectral matrix $\mathbf{R}_N(f)$ does not appear in Eq. (31) (contrary to Eq. (18)).

Unfortunately, the phase matrix $\mathbf{R}_S^\tau(f)$ is present in Eq. (31) and prevents the suppression of the unitary matrix $\mathbf{\Pi}(f)$ in $\mathbf{R}_X^\tau(f)$, as it happens in Eq. (18). Nevertheless, taking into account the phase relationship in $\mathbf{R}_S^\tau(f)$, several delayed spectral matrices are combined in order to remove $\mathbf{R}_S^\tau(f)$ (and so $\mathbf{\Pi}(f)$) and to recover the EVD of $\mathbf{R}_\mathbf{Y}(f)$. Note that if the delay τ is replaced with $(-\tau)$ in Eq. (31), the phase sign is also changed in the expression of $\mathbf{R}_S^\tau(f)$.

For that reason, $\mathbf{R}_X^{-\tau}(f)$ is given by

$$\mathbf{R}_X^{-\tau}(f) = \mathbf{V}(f)\cdot\mathbf{\Delta}(f)^{1/2}\cdot\mathbf{\Pi}(f)\cdot(\mathbf{R}_S^\tau(f))^{-1}\cdot\mathbf{\Pi}(f)^+\cdot(\mathbf{\Delta}(f)^{1/2})^+\cdot\mathbf{V}(f)^+. \quad (32)$$

In a similar way, a second delayed spectral matrix with delay (-2τ) is now computed:

$$\mathbf{R}_X^{-2\tau}(f) = \mathbf{V}(f)\cdot\mathbf{\Delta}(f)^{1/2}\cdot\mathbf{\Pi}(f)\cdot(\mathbf{R}_S^\tau(f)^2)^{-1}\cdot\mathbf{\Pi}(f)^+\cdot(\mathbf{\Delta}(f)^{1/2})^+\cdot\mathbf{V}(f)^+. \quad (33)$$

According to Eqs. (31)–(33), it appears that a judicious combination of the two matrices $\mathbf{R}_X^\tau(f)$ and $\mathbf{R}_X^{-2\tau}(f)$ allows the problem to be solved. Hence, two delays (τ and -2τ) are sufficient to eliminate the phase matrix $\mathbf{R}_S^\tau(f)$. Let $\mathbf{H}(f)$ be equal to

$$\mathbf{H}(f) = \mathbf{R}_X^\tau(f)\cdot(\mathbf{R}_X^\tau(f)^+)^{\neq}\mathbf{R}_X^{-2\tau}(f), \quad (34)$$

where $^{\neq}$ represents the pseudo-inverse matrix.

Detail now the computation of $(\mathbf{R}_X^\tau(f)^+)^{\neq}$:

$$\mathbf{R}_X^\tau(f)^+ = \mathbf{V}(f)\cdot\mathbf{\Delta}(f)^{1/2}\cdot\mathbf{\Pi}(f)\cdot(\mathbf{R}_S^\tau(f))^{-1}\cdot\mathbf{\Pi}(f)^+\cdot(\mathbf{\Delta}(f)^{1/2})^+\cdot\mathbf{V}(f)^+ \quad (35)$$

as $\mathbf{R}_S^\tau(f)^+ = (\mathbf{R}_S^\tau(f))^{-1}$.

$$(\mathbf{R}_X^\tau(f)^+)^{\neq} = \mathbf{V}(f)\cdot((\mathbf{\Delta}(f)^{1/2})^+)^{\neq}\cdot\mathbf{\Pi}(f)\cdot\mathbf{R}_S^\tau(f)\cdot\mathbf{\Pi}(f)^+\cdot(\mathbf{\Delta}(f)^{1/2})^{\neq}\cdot\mathbf{V}(f)^+. \quad (36)$$

Consequently, from Eqs. (31), (33) and (36), it can be proved that it is theoretically possible to re-determine the EVD of $\mathbf{R}_\mathbf{Y}(f)$. Due to the phase terms of the diagonal matrix $\mathbf{R}_S^\tau(f)$, Eq. (34) yields

$$\begin{aligned}\mathbf{H}(f) &= \mathbf{R}_X^\tau(f)\cdot(\mathbf{R}_X^\tau(f)^+)^{\neq}\mathbf{R}_X^{-2\tau}(f)\\ &= \mathbf{V}(f)\cdot\mathbf{\Delta}(f)^{1/2}\cdot\mathbf{\Pi}(f)\cdot\underbrace{\mathbf{R}_S^\tau(f)\cdot\mathbf{R}_S^\tau(f)(\mathbf{R}_S^\tau(f)^2)^{-1}}_{\mathbf{I}_p}\cdot\mathbf{\Pi}(f)^+\cdot\mathbf{\Delta}(f)^{1/2}\cdot\mathbf{V}(f)^+\\ &= \mathbf{V}(f)\cdot\mathbf{\Delta}(f)\cdot\mathbf{V}(f)^+\end{aligned} \quad (37)$$

and finally, matrix $\mathbf{H}(f)$ is equal to

$$\mathbf{H}(f) = \mathbf{R}_\mathbf{Y}(f). \quad (38)$$

The key point of the method lies in the phase relationship in the delayed spectral matrices of the sinusoidal sources:

$$\mathbf{R}_S^{\tau}(f)\cdot\mathbf{R}_S^{\tau}(f)\cdot\mathbf{R}_S^{-2\tau}(f)=\mathbf{R}_S^{\tau}(f)\cdot\mathbf{R}_S^{\tau}(f)(\mathbf{R}_S^{\tau}(f)^2)^{-1}=\mathbf{I}_p. \tag{39}$$

As the expression of matrix $\mathbf{H}(f)$ (Eq. (30)) can be estimated directly from the noisy data $\mathbf{X}(n,f)$, the un-noisy spectral matrix $\mathbf{R_Y}(f)$ is reconstructed exactly, whatever the chosen delay τ, provided that the noise components between $\mathbf{X}(n,f)$ and $\mathbf{X}(n+\tau,f)$ are decorrelated. Indeed, $\mathbf{H}(f)$ is estimated from noisy data through the estimation of the delayed spectral matrices $\mathbf{R}_X^{\tau}(f)$ and $\mathbf{R}_X^{-2\tau}(f)$, using the estimator described in Eq. (28).

Unfortunately, this estimator of $\mathbf{H}(f)$ (Eq. (37)) involves the inverse of a matrix $\mathbf{R}_X^{\tau}(f)$ that is theoretically of rank p but numerically of rank n. So, stability is improved using the pseudo-inverse algorithm so that only p eigenvalues are inversed. Choosing the number of eigenvalues p is the same as choosing the number of sources. The usual AIC and MDL criteria [39] perform an hypothesis testing on the eigenvalues of the spectral matrix but both methods assume that the noises are Gaussian and spatially white. It is then more robust to establish the source number directly from a delayed spectral matrix for example by using the eigenvectors [40].

The EVD of the estimate of matrix $\mathbf{H}(f)$ generates estimates of the eigenvalues matrix $\mathbf{\Delta}(f)$ and the eigenvectors matrix $\mathbf{V}(f)$ of $\mathbf{R_Y}(f)$. Therefore, correct estimates of the signal subspace and the whitening matrix $\mathbf{W}(f)$ are derived, even in severe spatially correlated noise conditions.

As a conclusion, a modified PCA for periodic sources is proposed in this section. Performance of the algorithm is investigated in Section 4 with simulation and experimental results for rotating machine data.

3.6. Measuring independence and contrast functions

BSS consists of recovering independent sources from convolutive mixtures of these sources. As described in previous sections, PCA is usually used as pre-processing to reduce the dimension of the data and for its de-noising capabilities. After the classical PCA step or the modified PCA for periodic sources (Section 3.5), the observations $\mathbf{X}(n,f)$ are projected on the signal subspace and whitened with the help of the whitening matrix $\mathbf{W}(f)$. The noise components are generally neglected and the sphered data $\mathbf{Z}(n,f)$ are assumed to be equal to

$$\mathbf{Z}(n,f)=\mathbf{\Pi}(f)\cdot\mathbf{S}(n,f). \tag{40}$$

After this step, the BSS problem is reduced to the estimation of the inverse of the remaining unitary matrix $\mathbf{\Pi}(f)$, $\mathbf{C}(f)$. Extra information contained in the hypothesis of independent sources is exploited to identify it. Indeed, it has been proved theoretically that the separation is achieved when the p components of the separating system $\mathbf{C}(f)\cdot\mathbf{Z}(f)$ are mutually and statistically independent [8,14], after a theorem of Darmois [14]. In practice, the p components of the separating system $\mathbf{C}(f)\cdot\mathbf{Z}(f)$ are required to be as independent as possible. Many approaches explore the different formulations of statistical independence, cross-cumulant cancellation, entropy minimization, likelihood approaches, etc.

For a complete survey on statistical principles for BSS see Refs. [15–17]. Usually this unknown rotation matrix $\mathbf{C}(f)$ is obtained by maximization or minimization of an objective function which is also called the contrast function [8] which implements an independence measurement.

The definition of independent random variables is noted below:

Definition 1. Random variables are said to be independent if their joint probability density function (pdf) is equal to the product of their respective marginal densities, i.e., assume a $m \times 1$ random vector $R = [R_1, R_2, \ldots, R_m]$ with a multivariate probability density function $p(r)$, then independence allows $p(r)$ to be factorized, i.e.,

$$p(r) = p_1(r_1) \cdot p_2(r_2) \cdots p_m(r_m). \tag{41}$$

Contrast functions for BSS were first introduced by Comon [8] and many different approaches exist in the literature. The principal contrast functions are only reviewed here. They can be derived from the Maximum Likelihood principle [15,41], the Kulback divergence [41] or the mutual information [12,41].

One way to simplify these contrasts is to introduce cumulants via polynomial density expansion [42]. The first example is given by Gaeta and Lacoume [31] with an approximation of the likelihood by a Gram–Charlier expansion. Comon [8] suggests an approximation of the mutual information by an Edgeworth expansion. Another way to derive some empirical contrast is the approximation of independence via cumulants. Indeed, two signals are statistically independent if all cross-cumulants (at any order k) are equal to zero. For computational reasons, the approximation of cumulants and moments is generally done up to the fourth order. The minimization of cross-cumulants leads to empirical contrasts like kurtosis cancellation [43] or some fourth order cumulants (called the JADE algorithm) [44].

All these contrast functions are closely related [15] and their performance is similar. In the next section, results on experimental signals were obtained with the well-known algorithm JADE, which maximizes the following contrast function [15]:

$$\Phi(f) = -\sum_{\substack{i,j,k,l=1 \\ i,j,k,l \neq i,j,k,k}}^{p} |Cum[R_i(n,f), R_j(n,f)^*, R_k(n,f), R_l(n,f)^*|^2. \tag{42}$$

Concerning statistical independence, the additional information exists only under the hypothesis of non-Gaussian sources. This explains why Gaussian variables are inadequate for solving the BSS problem. Fourier transform is often thought to converge towards Gaussianity, but it was shown in Ref. [21] that this assertion is not valid for spectral lines signals like those of rotating machines. More precisely, a distance to Gaussianity, called the spectral kurtosis, is defined for complex variables as the normalized fourth order cumulant:

$$K(f) = \frac{Cum[\mathbf{R}(n,f), \mathbf{R}(n,f)^*, \mathbf{R}(n,f), \mathbf{R}(n,f)^*]}{|Cum[\mathbf{R}(n,f), \mathbf{R}(n,f)^*|^2}. \tag{43}$$

Its estimation $\widehat{K}(f)$,

$$\widehat{K}(f) = \frac{\frac{1}{K}\sum_{n=0}^{K-1} |\mathbf{R}(n,f)|^4 - |\frac{1}{K}\sum_{n=0}^{K-1} \mathbf{R}(n,f)^2|^2}{(\frac{1}{K}\sum_{n=0}^{K-1} |\mathbf{R}(n,f)|^2)^2} - 2 \tag{44}$$

tends toward -1 at all the harmonic bins when the number of averages K is large enough. Hence, for these latter, any higher order based source separation algorithm for instantaneous complex mixtures can be employed in each frequency bin.

4. Applications

The purpose of this section is to provide an illustration of the capability of BSS algorithms to separate signals from noisy rotating machine vibration. First the performances of the proposed PCA versus classical PCA are compared for noisy convolutive mixtures of synthetic sinusoidal sources. Finally, the results of a complete implementation of the BSS algorithm (whitening and separation) on a real process are presented and discussed.

4.1. Performances of the modified PCA on simulations

In this section, the performance of the proposed PCA method versus standard PCA for noisy mixtures of sinusoidal sources and its robustness versus low signal to noise ratios is given.

The efficiency of PCA relies first on the estimation (called $\widehat{\mathbf{V}}_s(f)$) of an orthonormal base of the signal subspace $\mathbf{V}_s(f)$ and secondly on the estimation of a whitening matrix (Section 3.3). The accuracy of estimation of the signal subspace is measured by means of the distance between the exact signal subspace and the estimated one. From Ref. [45], the distance between two subspaces, E1 and E2, is calculated as the 2-norm of the difference between the orthogonal projections onto E1 and E2. Indeed, the orthogonal projection on a vectorial space is unique: it is characterized by $\mathbf{V}_s(f)\mathbf{V}_s(f)^+$ where $\mathbf{V}_s(f)$ is any orthonormal basis of this space. Therefore, the distance between the exact signal subspace and the estimated one is given by

$$d = \|\widehat{\mathbf{V}}_s(f)\widehat{\mathbf{V}}_s(f)^+ - \mathbf{V}_s(f)\mathbf{V}_s(f)^+\|. \tag{45}$$

A Monte-Carlo study on the distance to the exact signal subspace d was performed and illustrates the robustness of the method versus low signal to noise ratios. For that, the signal-to-noise ratio between source and noise components varied from -10 dB to 10 dB for all sensors. 200 examples of noisy convolutive mixtures of sinusoidal sources were generated, for each signal-to-noise ratio. Two sources were mixed and observed using 6 sensors. These observations were processed in the frequency bin of normalized frequency [0.148; 0.164] of centre frequency 0.156 and the performances were computed in this bin. The sinusoidal sources of respectively normalized frequency 0.15 and 0.16 and with constant amplitudes and phases were both present in this bin.

The mixtures were performed in the frequency domain (cf. Eq. (4)) with the help of a constant pseudo-invertible mixing matrix 2×6. The noise-free mixtures were identical for all the realizations and signal-to-noise ratios, and also in the signal subspace $\mathbf{V}_s(f)$. Spatially correlated and spectrally coloured noises were then added to the noise-free mixtures. They result from the filtering of 200 realizations of Gaussian white noise with fixed AR filters, for each signal-to-noise ratio.

This section is devoted to the comparison between the proposed PCA method and the standard PCA and shows off its robustness against low signal-to-noise ratios. For each ratio and each sample, delayed spectral matrices were estimated on 100 DFT data blocks of 64 samples with 32 samples overlapping. Due to the correlation lengths of the noises, the delay τ was chosen equal to 80. A base of the estimated signal subspace $\widehat{\mathbf{V}}_s(f)$ and a distance d (40) were calculated for the two methods.

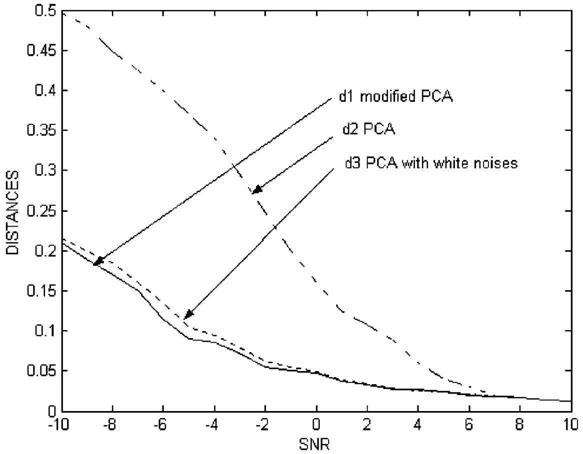

Fig. 3. Distance between subspaces.

Fig. 3 presents the average of the 200 computed distances for each signal-to-noise ratio where $d1$ and $d2$ denote the distance measured with the modified PCA and the classical PCA. The results indicate that the two methods are equivalent only for high signal-to-noise ratio (higher than 7 dB). Below 7 dB, the estimation of the signal subspace is greatly improved versus PCA in a context of spatially correlated noises ($d1$). The increase of performances (i.e., the difference between the curves $d1$ and $d2$) depends obviously on the value of the noise spectral matrix and on its distance to a diagonal matrix. The gap between the two curves also shows the limit of PCA in that context.

The performances of the two methods versus PCA are also compared in the presence of spatially white noises. In that case, the signal subspace is theoretically correct and the distance to the signal subspace, noted $d3$, is only due to estimation errors of the spectral matrix. The curves $d1$ and $d3$ are very close which proves the accuracy of the method proposed for spatially correlated noises. $d1$ is even inferior to $d3$ when the model of spatially white noises is verified. Therefore, the proposed PCA is a useful tool for the whitening of noisy sinusoidal sources.

4.2. Experimental results on rotating machine vibration data

BSS is a promising tool for non-destructive monitoring as it recovers the signature of a single rotating machine, from combinations of several operating machines. It can be seen as a pre-processing step that improves the diagnosis since specific methods of fault detection could then be applied to the signatures of each machine to be diagnosed. BSS methods were already tested on real signals from a mechanical bench in a noise-free context in Ref. [5]. Thus, it proves the capability of BSS to separate signals of rotating machine vibration. However, in a noisy environment, the mixtures are generally corrupted with spatially correlated noises and BSS methods lose efficiency.

The following results illustrate a complete implementation of the algorithm (whitening and separation steps) on a real noisy process. Two test beds were fixed to the same structure. Each test bed included a synchronous alternator, a motor and a pump. The two test beds were mechanically independent but electrically linked: the alternator rotation speeds were in a ratio of 0.995. Six accelerometers were placed on the structure and recorded mixtures of all the sources. In particular, they all received harmonics of the two rotation frequencies. For example, the spectra of two accelerometers are shown in Fig. 4 in the band $[0, 0.2]$ (normalized frequency). The bandwidth is equal to $(1/4096)$ and the harmonics of the rotation frequencies are not separated. It was also verified that the spectra are composed of spectral lines and corrupted with spatially correlated noises. Thus it appears that the hypotheses of the proposed method of PCA are correctly satisfied on these experimental signals.

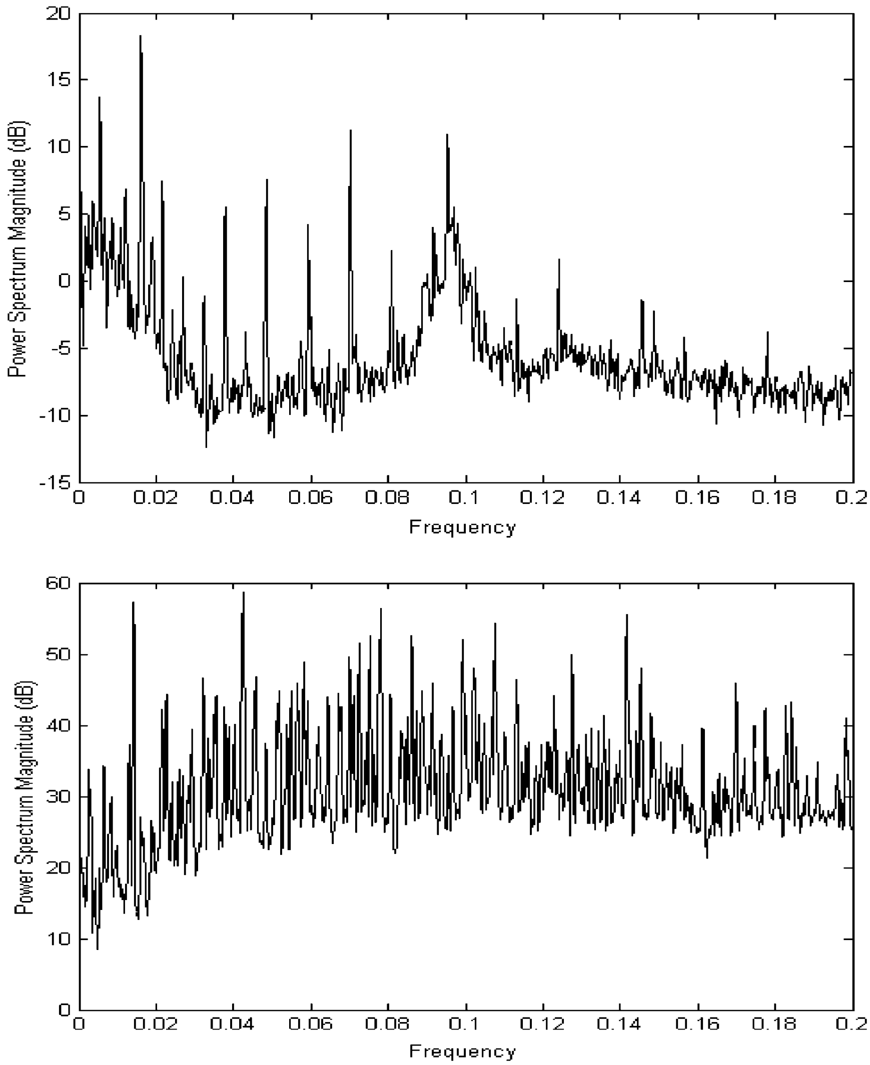

Fig. 4. PSD of two accelerometers.

The whitening and separation procedure was performed in frequency domain. 4096-points DFT were computed on 256 disjoined windows and next spectral matrices and delayed spectral matrices were estimated on these data blocks. Each frequency bin has been independently processed and the re-ordering step described in Refs. [5,21] was used to remove the permutation indeterminacy. The estimated sources are then temporal signals.

To illustrate the efficiency of the whitening and separation steps, the results obtained in one specific frequency bin of widthband ($1/4096 = 2.44 \times 10^{-4}$) are displayed, for instance in the band [0.0514, 0.0525] (normalized frequency). From this point of view, a DFT was computed on all the signals to focus on the spectral content in this particular frequency bin. First, the spectra of the six observations are represented in Fig. 5 in this band.

Two sinusoids clearly appear Fig. 5 at frequencies 0.0518 and 0.05205. They are present and mixed on several sensors. The first decorrelating step has been performed with standard PCA (Fig. 6) or with the proposed PCA method exploiting delayed spectral matrices (Section 3.5) (Fig. 7). In each case, a two-dimensional signal subspace was estimated and the observations were projected on these subspaces. The spectra of the two resulting signals (in fine and thick lines) are drawn in Fig. 6 for PCA and Fig. 7 for the proposed PCA. At this step, noise has been clearly reduced more with the second method over the whole band. It indicates that the proposed method gives satisfactory results to de-noise sinusoidal signals.

After this decorrelation step, the separation was achieved with a unique method and the remaining unitary matrix was performed by the cancellation of some fourth order cumulants [15,44] (the JADE algorithm). JADE maximizes the contrast function (Eq. (42)). The main reason for considering this criterion is its link to underlying eigenstructures which allows for an efficient optimization of it by the mean of joint diagonalization. The unitary matrix is the result of a joint diagonalization of a set of eigenmatrices issued from fourth order cumulants. These cumulants are

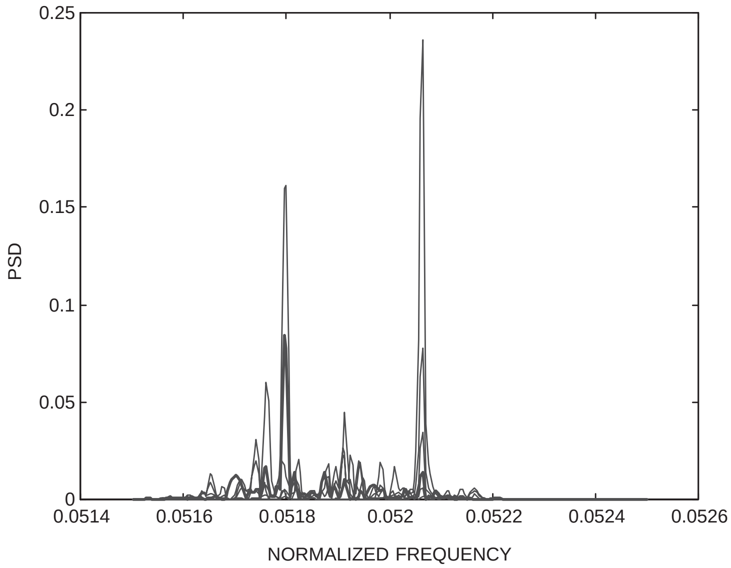

Fig. 5. PSD of the observations in the band [0.0514, 0.0525].

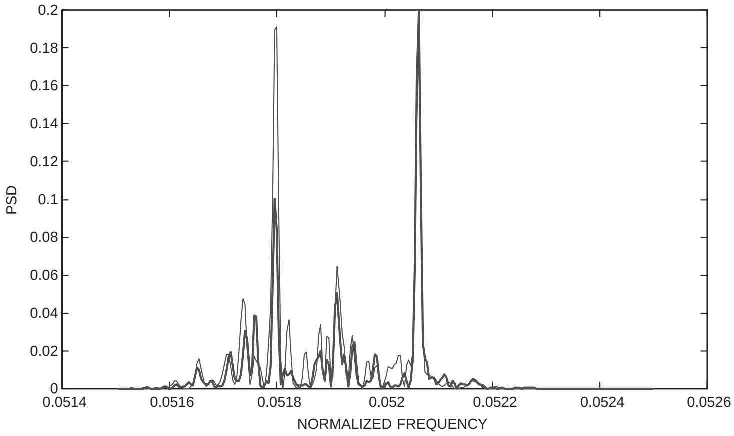

Fig. 6. Spectra of the two whitened signals with PCA (in fine and thick lines).

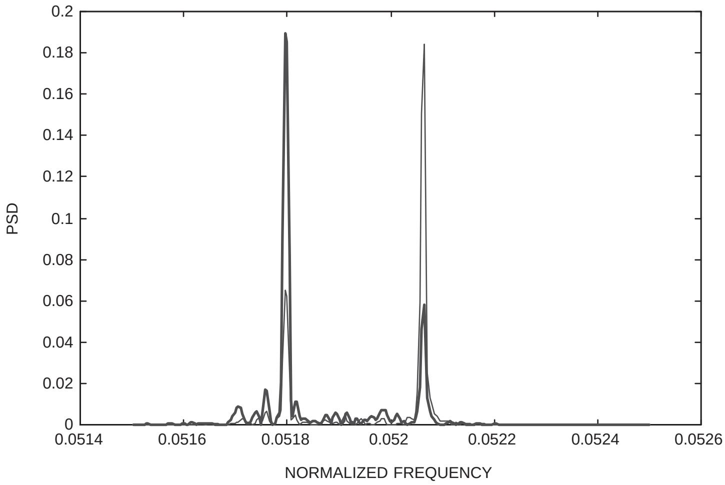

Fig. 7. Spectra of the two whitened signals with proposed PCA (in fine and thick lines).

estimated from the whitened process. The algorithm exploits the extended Jacobi technique and makes the maximization of Eq. (42) computationally attractive. It is detailed in Ref. [44] and the code can be found on the website http://www.tsi.enst.fr/~cardoso/.

The spectra of the two sources estimated with the two steps, PCA and JADE, are superposed in Fig. 8 in fine and thick lines. It reveals that the two sinusoids are still present in the first estimated source (in fine line) and clearly have not been separated. The second estimated source (in thick line) seems to contain only noise. Because of the presence of spatially correlated noises, the signal subspace and whitening matrix were poorly estimated with PCA. Indeed, noise is still evident after projection on a wrong signal subspace (Fig. 6). Therefore, the separation step completely failed.

Next, delayed spectral matrices are exploited initially (proposed PCA Section 3.5) and the JADE algorithm as a further step. The spectra of the two estimated sources are depicted in Fig. 9

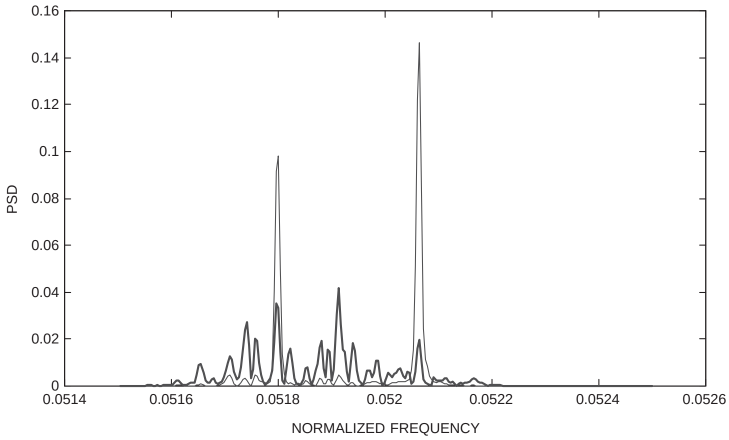

Fig. 8. Spectra of the two separated sources with standard PCA and JADE algorithm (in fine and thick lines).

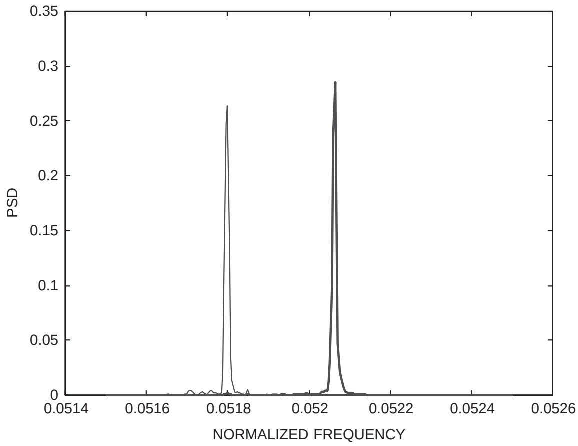

Fig. 9. Spectra of the two separated sources with proposed PCA and JADE algorithm (in fine and thick lines).

(in fine and thick lines). As the proposed PCA method succeeded well in whitening and de-noising the observations (see Fig. 7), the two sinusoids are now correctly separated at frequency 0.0518 (fine line) and 0.05205 (thick line). As the frequencies are in a ratio of 0.995, each sinusoid can be attributed to an alternator.

The results prove the significance of the first decorrelation step in the source separation technique. The accuracy of the signal subspace and the whitening matrix is fundamental for the effective estimate of the mixing matrix. The usual PCA assumes the hypothesis on the noises (spatial whiteness) but it is generally not verified by experimental noisy observations. In that case, a model of periodic sources is exploited to compute the whitening matrix and to show the efficiency of the proposed method (versus PCA) based on experimental noisy observations.

5. Conclusions

This paper involves a source separation technique in the context of spatially correlated noise. Source separation techniques generally use principal components analysis (PCA) as a first step, to whiten the observations. In this context the PCA fails because the noise-free spectral matrix is poorly estimated. As the usual PCA cannot provide a correct estimate of the signal subspace in this situation, a new estimator of the whitening matrix and the signal subspace is introduced. This new estimator exploits the model of periodic sources and uses spectral matrices of delayed observations. The proposed method is efficient for low signal-to-noise ratios and its performance is illustrated with simulations and rotating machine vibration data.

References

[1] Y. Deville, Towards industrial applications of blind source separation and independent component analysis, *Proceedings of the First Workshop on Blind Source Separation and Independent Component Analysis* (*ICA'99*), Aussois, France, 1999, pp. 19–24.

[2] P. Ruiz, J.L. Lacoume, Extraction of independent sources from correlated inputs, a solution based on cumulants, *Proceedings of the Workshop on Higher Order Spectral Analysis*, 1989, pp. 28–30.

[3] V. Capdevielle, C. Servière, J.L. Lacoume, Blind separation of wide band sources: application to rotating machine signals, *Proceedings of European Signal Processing Conference'96* (*EUSIPCO'96*), 1996, pp. 2085–2088.

[4] G. Gelle, M. Colas, G. Delaunay, Blind sources separation applied to rotating machines monitoring by acoustical and vibrations analysis, Mechanical System and Signal Processing 14 (2000) 427–442.

[5] G. Gelle, M. Colas, C. Servière, Blind source separation: a tool for rotating machine monitoring by vibrations analysis, Journal of Sound and Vibration 248 (5) (2001) 865–885.

[6] A. Ypma, A. Leshem, Blind separation of machine vibration with bilinear forms, *Proceedings of the Second Workshop on Blind Source Separation and Independent Component Analysis* (*ICA'00*), Helsinki, Finland, 2000, pp. 109–114.

[7] M. Knaak, M. Fausten, D. Filbert, Acoustical machine monitoring using blind source separation, *Proceedings of the Fourth International Conference: Acoustical and Vibratory Surveillance Methods and Diagnostic Techniques*, Compiègne, France, 2000, pp. 139–149.

[8] P. Comon, Independent component analysis, A new concept?, Signal Processing 36 (1994) 287–314.

[9] J. Karhunen, Neural approaches to independent component analysis and source separation, Neural Computation 99 (1998) 249–266.

[10] P. Fabry, C. Servière, Blind separation of noisy harmonic signals using orthogonal techniques for rotating machine diagnosis, *Proceedings of the Second Workshop on Blind Source Separation and Independent Component Analysis* (*ICA'00*), Helsinki, Finland, 2000, pp. 369–374.

[11] A. Ypma, P. Pajunen, Rotating machine vibration analysis with second order independent component analysis, *Proceedings of the First Workshop on Blind Source Separation and Independent Component Analysis* (*ICA'99*), Aussois, France, 1999, pp. 37–42.

[12] A. Bell, T. Sejnowski, An information-maximisation approach to blind separation and blind deconvolution, Neural Computation 7 (1995) 1129–1159.

[13] A. Ypma, A. Leshem, Blind separation of rotating machine vibration with bilinear forms, *Proceedings of the Second Workshop on Blind Source Separation and Independent Component Analysis* (*ICA'00*), Helsinki, Finland, 2000, pp. 405–410.

[14] X.R. Cao, R.-W. Liu, General approach to blind source separation, IEEE Transactions on Signal Processing 44 (3) (1996) 562–571.

[15] J.F. Cardoso, Blind signal separation: statistical principle, Proceedings of the IEEE 86 (1998) 2009–2025.

[16] J. Hérault, C. Jutten, Blind separation of sources I. An adaptive algorithm based on neuromimetic architecture, Signal Processing 24 (1) (1991) 1–10.

[17] A. Hyvarinen, Survey on independent component analysis, Neural Computing Surveys 2 (1999) 94–128.
[18] H.L. Nguyen Thi, C. Jutten, Blind source separation for convolutive mixture, Signal Processing 45 (1995) 209–229.
[19] N. Charkani, Y. Deville, Self adaptive separation of convolutively mixed signals with a recursive structure. Part 1: stability analysis and optimization of asymptotic behavior, Signal Processing 73 (1999) 225–254.
[20] S. Cruces, L. Castedo, Stability analysis of adaptive algorithms for blind sources separation of convolutive mixtures, Signal Processing 78 (1999) 265–275.
[21] V. Capdevielle, C. Servière, J.L. Lacoume, Blind separation of wide band sources in frequency domain, *Proceedings of ICASSP'95*, Detroit, USA, 1995.
[22] H.C. Wu, J.C. Principe, Simultaneous diagonalization in the frequency domain (SDIF) for source separation, *Proceedings of the First Workshop on Blind Source Separation and Independent Component Analysis* (*ICA'99*), Aussois, France, 1999, pp. 245–250.
[23] J. Anemuller, B. Kollmeier, Amplitude modulation decorrelation for convolutive blind source separation, *Proceedings of the Second Workshop on Blind Source Separation and Independent Component Analysis* (*ICA'00*), Helsinki, Finland, 2000, pp. 215–220.
[24] L. Parra, C. Spence, B. de Bries, Convolutive blind source separation based on multiple deconvolution, *Proceedings of the IEEE Workshop on Neural Processing*, Cambridge, UK, 1998.
[25] S. Murrata, S. Ikeda, A. Ziehe, An approach to blind source separation based on temporal structure of speech signals, BSIS, Technical Report, 1998, http:/www.bsp.brain.riken.go.jp
[26] C. Mejuto, A. Dapena, L. Castedo, Frequency-domain infomax for blind separation of convolutive mixtures, *Proceedings of the Second Workshop on Blind Source Separation and Independent Component Analysis* (*ICA'00*), Helsinki, Finland, 2000, pp. 315–320.
[27] M. Knaak, D. Filbert, Acoustical semi-blind source separation for machine monitoring, *Proceedings of the Third Workshop on Blind Source Separation and Independent Component Analysis* (*ICA'01*), San Diego, USA, 2001, pp. 361–366.
[28] K. Torkkola, in: S. Haykin (Ed.), *Blind Separation of Delayed and Convolved Mixtures in Unsupervised Adaptative Filtering*, Vol. 1, Wiley, New York, 2000.
[29] S. Haykin, in: S. Haykin (Ed.), *Blind Source Separation in Unsupervised Adaptative Filtering*, Vol. 1, Wiley, New York, 2000.
[30] C. Therrien, Discrete Random Signals and Statistical Signal Processing, Prentice-Hall, Englewood Cliffs, NJ, 1992.
[31] M. Gaeta, J.L. Lacoume, Source separation without a priori knowledge: the maximum likelihood solution, *Proceedings of EUSIPCO'90*, 1990, pp. 621–624.
[32] O. Bermond, J.F. Cardoso, Approximate likelihood for noisy mixtures, *Proceedings of the First Workshop on Blind Source Separation and Independent Component Analysis* (*ICA'99*), Aussois, France, 1999, pp. 325–330.
[33] A. Hyvarinen, Noisy independent component analysis by Gaussian moments, *Proceedings of the First Workshop on Blind Source Separation and Independent Component Analysis* (*ICA'99*), Aussois, France, 1999, pp. 473–478.
[34] K. Torkkola, Blind separation of convolved sources based on information maximization, *IEEE Workshop on Neural Networks for Signal Processing*, Kyoto, Japan, 1996.
[35] C. Servière, P. Fabry, Blind separation of noisy harmonic signals for rotating machine diagnosis, *Proceedings of the Fourth International Conference*: *Acoustical and Vibratory Surveillance Methods and Diagnostic Techniques*, Compiègne, France, 2000, pp. 129–138.
[36] J.L. Lacoume, A survey of source separation, *Proceedings of the First Workshop on Blind Source Separation and Independent Component Analysis* (*ICA'99*), Aussois, France, 1999, pp. 1–6.
[37] J.M. Danthez, C. Giroussens, R. Aquilina, The autocoherent spectrum: a useful spectral estimator for vibration analysis of rotating machinery. Accurate estimation and cancellation of pure tones, Mechanical System and Signal Processing 12 (1998) 753–768.
[38] P.D. Welch, The use of fast Fourier transform for the estimation of power spectra, IEEE Transactions on Audio Electroacoustics AU-15 (1970) 70–73.
[39] G. Schwarz, Estimating the dimension of a model, The Annals of Statistics 6 (2) (1978) 461–464.

[40] P. Fabry, C. Servière, J.-L. Lacoume, Improving signal subspace estimation and source number detection in the context of spatially correlated noises, *Proceedings of European Signal Processing Conference'98* (*EUSIPCO'98*), 1998.
[41] J.-F. Cardoso, Infomax and maximum likelihood for blind signal separation, *IEEE Signal Processing Letters* (4) (1998) 112–114.
[42] M.G. Kendall, A. Stuart, *The Advanced Theory of Statistics*, Vol. 1, Griffin, London, 1963.
[43] E. Moreau, O. Macchi, Higher order contrast for self-adaptive source separation, International Journal of Adaptive Control and Signal Processing 10 (1996) 19–46.
[44] J.-F. Cardoso, A. Souloumiac, Blind beamforming for non-Gaussian signals, IEE-Proceedings-F 40 (1993) 362–370.
[45] G.H. Golub, F. Van Loan, Matrix Computations, The Johns Hopkins University Press, Baltimore, MD, 1983.