NNSP'2003 proceedings - Neural Networks for Signal Processing
Toulouse (France), 17-19 September 2003, IEEE, pp. 359-368

IMPROVING INDEPENDENT COMPONENT ANALYSIS PERFORMANCES BY VARIABLE SELECTION

F. Vrins[1], J. A. Lee[1], M. Verleysen[1], V. Vigneron[2], C. Jutten[2]
[1] Department of Microelectronics, Machine Learning Group
UCL-DICE Place du levant 3, 1348 Louvain-la-Neuve, Belgique
[2] Images and signals laboratory
LIS-INPG Avenue Félix Viallet 46, 38031 Grenoble cedex, France
E-mail: {vrins,lee,verleysen}@dice.ucl.ac.be, {vvigne,jutten}@lis.inpg.fr

Abstract. Blind Source Separation (BSS) consists in recovering unobserved signals from observed mixtures of them. In most cases the whole set of mixtures is used for the separation, possibly after a dimension reduction by PCA. This paper aims to show that in many applications the quality of the separation can be improved by first selecting a subset of some mixtures among the available ones, possibly by an information content criterion, and performing PCA and BSS afterwards. The benefit of this procedure is shown on simulated electrocardiographic data by extracting the fetal electrocardiogram signal from mixtures recorded on the abdomen of a pregnant woman.

INTRODUCTION

Most Blind Source Separation (BSS) algorithms require a number of observed mixtures n equal to or larger than the number of sources m. In real situations however, it is rather difficult to estimate m precisely. Consequently, a usual way to circumvent the problem consists in measuring much more mixtures than necessary $n \gg \hat{m}$ ($\hat{m}$ being the estimated value of m).

Unfortunately, the number of separated signals is equal to the number of observed mixtures n and not to $\hat{m}$. When n is large, this can produce convergence problems or very high computational cost (especially for algorithms based on joint-diagonalization *e.g.* Jade, which require the computing of the fourth order cross-cumulants [2]). The method commonly used to satisfy both arguments is to take many observations ($n \gg \hat{m}$) and project them by Principal Component Analysis (PCA) on the $\hat{m}$-dimensional subspace before performing the separation, so that only $\hat{m}$ signals will be separated.

[1] M.V. is a Senior Research Fellow of the Belgian National Fund for Scientific Research.

0-7803-8178-5/03/$17.00 © 2003 IEEE

However, Stone and Porrill [13] have shown that directly using PCA to reduce the dimension of signals mixtures (from n to $\hat{m}$) used as input to ICA can compromise the ability of ICA to extract source signals.

We propose to use a slightly different procedure, sPCA (for *selected* PCA): it consists in adequately selecting n' mixtures among the n available ones before performing the projection with PCA from n' to $\hat{m}$, with $n > n' > \hat{m}$.

Note that in both cases, we are still confronted to the difficulty of choosing the dimension of the projected subspace ($\hat{m}$).

In the following of this paper, we first introduce the bases of ICA and its preprocessing (PCA and whitening), as well as several criteria to evaluate the projection and separation performances. We then apply ICA on a simple example, to emphasize the importance of the (number of) mixtures taken into account. Next, the core of the paper is the presentation of a variable selection method based on mutual information criteria and its application to the extraction of the fetal electrocardiogram.

INDEPENDENT COMPONENT ANALYSIS

The problem of Blind Source Separation (BSS) consists in recovering m unobserved signals $S_1, \ldots, S_m$ called *sources*, from a set of n observed mixtures signals $X_1, \ldots, X_n$. Usually, it is assumed that the number of sources is lower or equal to the number of mixtures. In practice, it is necessary to make assumptions on both sources and mixtures in order to be able to solve this problem. In the simpler model, these assumptions are: i) the sources are statistically independent and at most one source is Gaussian and ii) the mixtures are linear and instantaneous, without noise. Then, the relation between the sources and the mixtures can be written as follows:

$$X = AS \ , \tag{1}$$

with $X = [X_1, \ldots, X_n]^T$, $S = [S_1, \ldots, S_m]^T$ and where A is the unknown n-by-m mixing matrix.

Using the independence assumption leads to Independent Component Analysis (ICA) methods. It can be shown [3] that, under mild assumptions, in linear mixtures, ICA achieves BSS. In fact, using ICA, one can estimate a m-by-n separating matrix B such that

$$BA = PD \ , \tag{2}$$

where P denotes a m-by-m permutation matrix and D a diagonal one of the same size.

Usually, statistical independence of the sources is expressed as follows:

$$p(S_1, S_2, \ldots, S_m) = \prod_{i=1}^{m} p(S_i) \ , \tag{3}$$

NNSP'2003 proceedings - Neural Networks for Signal Processing
Toulouse (France), 17-19 September 2003, IEEE, pp. 359-368

where $p(S_i)$ is the marginal probability density function (pdf) of the i-th source S_i and $p(S_1, S_2, \ldots, S_m)$ the joint pdf. A more convenient criterion for measuring independence will be proposed in equation (6).

The assumptions on the mixtures correspond to the *linear* BSS. ICA methods can be extended to more complex (and realistic) mixtures, for example by allowing additive noise [6], delays [15] or nonlinearities [14] in the mixtures.

PREPROCESSING TO ICA

Two major processings are commonly used before ICA: dimension reduction and whitening. Both can be achieved by PCA.

Actually, a projection by PCA decorrelates the mixtures. For Gaussian sources, the decorrelation (or whitening) permits to reach the independence but not to recover the sources. It means that Gaussian sources cannot be separated by using independence. Decorrelation is not sufficient for ensuring independence, which motivates the use of ICA afterwards. Intuitively, PCA achieves 'half the job' of ICA. Sphering is similar to whitening, excepted that the decorrelated mixtures are also normalized.

Principal Component Analysis aims to decorrelate variables or signals, in order to find orthogonal directions with maximal variance [8].

The first step of PCA consists in removing the sample mean of each signal (s is the number of samples): $X_j \leftarrow X_j - \frac{1}{s} \sum_{i=1}^{s} X_j^i$.

The second step consists in applying a linear transformation on X. This transformation rotates the coordinate system in such a way that the first new axis points in the direction of maximal variance, the second axis, orthogonal to the first one, collects the largest part of the remaining variance, and so on.

The new axes are determined by a spectral decomposition of the sample covariance matrix $C_X = (XX^T)/s = V \Lambda V^T$, where V is an orthonormal matrix and Λ a diagonal one. As C is symmetric and semipositive definite, all eigenvalues λ_i (the diagonal entries of Λ) are real and non-negative [11]. The variance along each of the new axes V_i is simply given by its associated eigenvalue λ_i. In the same way, a projection on the m eigenvectors ($m < n$) associated with the m largest eigenvalues conserves a portion ρ_m of the total variance which can be written as

$$\rho_m = \frac{\sum_{j=1}^{m} \lambda_j}{\sum_{j=1}^{n} \lambda_j} \quad . \tag{4}$$

The projection gives the decorrelated signals Y according to $Y = V_{1:m}^T X$, where $V_{(1:m)}$ gathers the m eigenvectors associated to the m largest eigenvalues. It can be verified that the projected signals are decorrelated [7].

Sphered signals can be obtained with a slight modification of PCA. The projection is given by $Y = \sqrt{\Lambda_{(1:m,1:m)}^{-1}} V_{1:m}^T X$, yielding decorrelated signals with unit variance.

NNSP'2003 proceedings - Neural Networks for Signal Processing
Toulouse (France), 17-19 September 2003, IEEE, pp. 359-368

PROJECTION AND SEPARATION QUALITY CRITERIA

Projection (after PCA) and separation (after ICA) performances can be assessed by several criteria.

Projection. After dimension reduction by PCA, the quality of the projection can be measured by equation (4). This criterion is always non-negative and equals one when all non-zero eigenvalues are kept.

Separation. In the lack of any information about the mixing matrix or the sources, the best way to assess the separation quality would be to check whether equation (3) holds. Unfortunately, the evaluation of the joint pdf is a very tedious problem. If the mixing matrix A is known, the performance of a separation matrix B could be estimated by the Signal-Interference Ratio (SIR, see [12]), defined as:

$$\mathrm{SIR} = \frac{1}{n} \sum_{i,j=1}^{n} \frac{|[BA]_{ij}|}{\max_{1 \leq k \leq n}(|[BA]_{ik}|)} - 1 \ . \tag{5}$$

This criterion measures the interferences between the recovered sources and the true sources. A null SIR means that each recovered signal is strictly equal to one source. If the sources are (partially) known, other criterions may be based on correlation measures between true sources and recovered ones.

IMPACT OF THE NUMBER OF MIXTURES ON THE PCA PROJECTION

The size of the separating matrix

From equation (2) it follows that B is a m-by-n matrix. Hence, the dimension reduction achieved by PCA essentially determines the number of rows in B.

On the other hand, a selection among the n available mixtures decreases the number of columns of B. Even if, from a numerical point of view, the estimation of a small-sized matrix is easier, people usually project the whole set of mixtures without any prior selection. This paper tries to answer to the two following questions: i) how to choose the mixtures that, once projected by PCA, will improve the quality of the separation? ii) is the whole set of mixtures the best solution?

Analysis of the PCA variance ratio ρ_m on a simple example

Consider the situation of four independent uniform random sources ($m = 4$), linearly and simultaneously mixed without noise. This situation corresponds to the classical ICA assumptions. In this section, we respectively compare the

NNSP'2003 proceedings - Neural Networks for Signal Processing
Toulouse (France), 17-19 September 2003, IEEE, pp. 359-368

situation where the number of sources is correctly estimated ($\hat{m} = m = 4$) and the case where it is not (in our example: $\hat{m} = 3$)[1]. We have applied the Jade algorithm in order to use the SIR criterion, but FastICA gave the same results about the correlations.

We first analyze the situation where $\hat{m} = m$. The left column of the figure 1 shows the quality of the source separation (*locally* with correlations and *globally* with SIR values) versus the number of selected mixtures n: the quality increases while $n < m$. To evaluate the local separation, we have plotted the evolution of the absolute value of correlations between each sources S_i ($i = 1, \ldots, 4$) and *their best* recovered signals Z_j ($j = 1, \ldots, 4$)[2]. The SIR graph shows the difference between the SIR computed over the n first observations and the SIR (SIRtot) computed over the whole mixture set. ρ_m remains equal to one. When $n = m$, the PCA projection does not reduce the dimension, and for $n > m$, even if the number of mixtures is higher than the number of sources, the dimension of the space of the observed signals stops growing, because of the redundancies induced by the ideality of the mixture. In this case, an a priori selection of the first n' signals among n (with $n' > m$) before projection has no consequence, because the whole dimension of the initial set is kept after projection ($\rho_m = 1$).

The right column of figure 1 shows the same example but with a wrong estimation of the number of sources (3 instead of 4). We see that taking the whole set of the observations does not lead to the best possible separation. The analysis of the ρ_m curve is interesting. For $n < \hat{m}$, the initial space and the projection space have the same dimension. But for $n > \hat{m}$, increasing the number of observations modifies the initial space (because $\hat{m} < m$), which explains that the ρ_m curve decreases. The oscillations of this curve for observations with $n > m$ prove that all observations do not have the same effect over the projection quality: there are observations that complicate the projection on the sources space, and others that improve it. We can also see that at each increase (resp. decrease) of ρ_m curve is associated an improvement (resp. deterioration) of the separation of at least one signals.

In this section, we have considered a single imperfection of the model: a wrong estimation of the number of sources. We have detailed that the selection of some mixtures is better from the projection and separation point of view. In addition, the mixture could have (even slightly) non-linear components and/or noise, and the question of reducing the number of observed signals n is also critical. It could be interesting to remove several observed signals, which could have a bad influence on the convergence of the ICA algorithms. In the above example, the selection method used was the random choice.

---

[1] Of course, in this ideal unnoisy case, the number of sources is trivially equal to the number of non-zero eigenvalues of the covariance matrix. Nevertheless the goal of this comparison is to show the consequences of a bad estimation of m on the projection and separation performances; therefore in this example we assume that m could be wrongly estimated, leading to $\hat{m} < m$.

[2] The *best* recovered signal associated with one source is the one that maximizes the absolute value of their correlations.

NNSP'2003 proceedings - Neural Networks for Signal Processing
Toulouse (France), 17-19 September 2003, IEEE, pp. 359-368

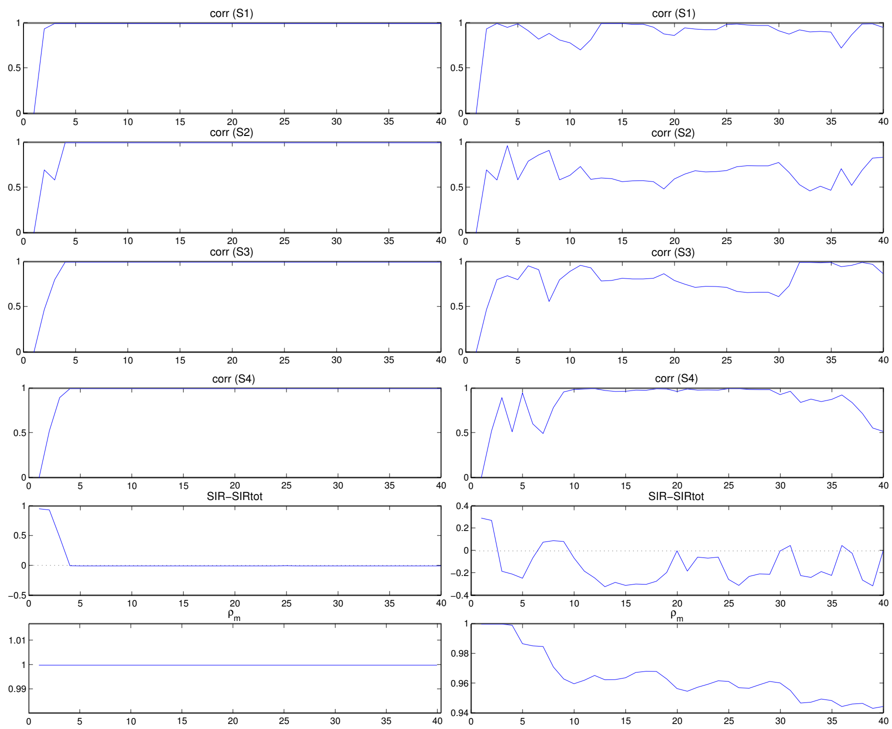

Figure 1: Evolution of the local and global separation quality and of projection *vs* the number of selected mixtures n. Left: correct estimation of the number of sources ($\hat{m} = m = 4$), right: bad estimation of the number of sources ($\hat{m} = 3, m = 4$).

We propose in the next section a selection method based on the mutual information.

VARIABLE SELECTION ALGORITHM

The proposed algorithm is based on the mutual information (I) between two signals. The mutual information is defined as the Kullback-Leibler divergence [5] between the joint pdf and the product of the marginal pdf (see for example Thomas and Cover [4]):

$$I(X,Y) = \int_{X,Y} p(x,y) log \frac{p(x,y)}{p(x)p(y)} dxdy \ . \tag{6}$$

Three major properties of I are: i) it is non-negative, ii) $I(X,Y) = 0$ if and only if X and Y are independent, and iii) $I(X,Y)$ is maximum for $X = Y$.

The aim of the selection algorithm is to select n' signals $(U_1, \ldots, U_{n'})$ among a set of n mixtures $(X_1, \ldots, X_n)$. The first signal U_1 must be chosen

NNSP'2003 proceedings - Neural Networks for Signal Processing
Toulouse (France), 17-19 September 2003, IEEE, pp. 359-368

by another method (*e.g.* randomly). At each step of the algorithm, we choose the mixture which is as independent as possible from the already selected mixtures U_j, $j = 1, \ldots, z-1$, *i.e.* which minimizes the sum of the mutual informations with the U_j's; in other words, X_k is the z^{th} selected signals ($U_z \doteq X_k$) if the following cost function $f_z(i)$ is minimized for $i = k$:

$$f_z(i) = \sum_{j=1}^{z-1} I(X_i, U_j) \ . \tag{7}$$

After the selection of X_k, it is removed from the initial set to avoid an eventual second selection. The selected subset will contain signals which are mutually "quite different", because of the minimisation of the mutual information. From another point of view, selecting signals which minimize the mutual information between them is a good preprocessing to ICA, because they have a low dependence level! We stop the algorithm when n' signals are selected ($z = n'$).

In the following, we call sPCA the classic PCA preceded by this selection algorithm.

APPLICATION: THE FETAL ECG EXTRACTION

We compare PCA and sPCA in a real situation: the extraction of the fetal electrocardiogram. In this application, it is difficult to estimate the number of sources m. This is one of the reasons why we work with simulated signals: to know exactly this number. Another reason is to be able to evaluate the performances of the separation: we can use a correlation-based criterion.

The problem

Description. Consider hundred electrical signals coming from sensors placed on a pregnant woman's abdomen. The electrodes on the maternal surface pick up the mother electrocardiogram (MECG) and, at lower level, the fetus one (FECG). Electrodes are also sensitive to other signals (*e.g.* electromyographic ones). The observed signals are thus a combination of all these sources. The aim is to recover the electrocardiogram of the fetus (*i.e.* not only the heart rate frequency and variability, which can be achieved for example like in [1], but the whole PQRST wave). This is a BSS problem, where the sources (mother's heart, fetal heart, breath, ...) may be supposed statistically independent. With this technique, one can obtain FECG signal during pregnancy in a non-invasive manner (*e.g.* [10], [16]).

Simulation of ECG signals. The signals of the electrodes are the sums of the electrical fields generated by several independent simulated sources: the maternal heart, the fetal heart, the uterus and the diaphragm. For each of these components, the equations of the electrical model are derived.

NNSP'2003 proceedings - Neural Networks for Signal Processing
Toulouse (France), 17-19 September 2003, IEEE, pp. 359-368

The shape of the maternal abdomen was considered to be formed by a parabolic function rotated around a middle axis. The difficulty of the task is that an accurate model of the heart is needed because the model has to provide the orientation of the heart dipole. This is achieved by using a template ECG data file recorded at a sampling rate of 500 Hz (the heart rate was 70 beats/mn). The signal was resampled so that a fetal heart rate was built at about twice the maternal one. Similar manipulations follow for the uterus and the diaphragm. The bioelectrical properties of the uterus are derived from [9]. The bioelectrical model of the diaphragm consists of 6 dipoles, symmetrically located over the uterus. The exact position as well as the amplitude can change over time. The calculation of the resulting (from superimposition of FECG, MECG, uterus and diaphragm signals) electrical field at the surface uses physics equations, which will not be detailed here, allows to simulate signals observed on electrodes according to their locations.

Results

Here we compare results obtained by simulation using both schemes: *projection-separation* (PCA) and *selection-projection-separation* (sPCA).

PCA. We have analyzed the ρ_m curve for a random growing set of mixtures and we show that the best separation of the FECG signals is not achieved for the projection of the whole set of observations (see figure 2). The *bar*-graph shows the first eigenvalues of the covariance matrix of the measured signals: we can observe two dominants components.

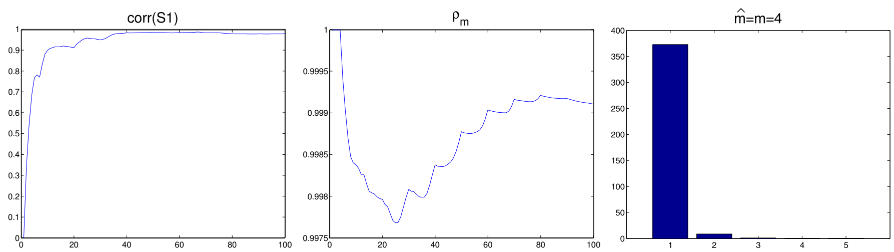

Figure 2: Evolution of the FECG separation *vs* the number of observed mixtures, with $\hat{m} = m = 4$.

sPCA. In this method, we apply our variable selection algorithm before performing projection by PCA. This algorithm requires an initialization: we have to choice the first signal of the selected subset. In the application of the FECG extraction, we have chosen this signal as the one measuring (as close as possible) the mother's ECG. All the following selected signals will be "quite different" from it, *e.g.* those which contain an important component due to the fetal heart beats. Figure 3 shows the results of this method after the selection of 15 signals ($n' = 15$) by the variable selection algorithm detailed in the previous section. The correlation between the FECG and the associated separated signal increases quickly (and reaches a value which is higher than the global maximum of the curve in figure 2) and the ρ_m decreases more slowly (see axes scales) with n. We can conclude that the quality of projection *and* separation are increased by this selection. This method requires to adjust a supplementary parameter: n'. Indeed, the graph of correlation in figure 3 shows that choosing $n' = 5$ is the optimal value n'_{opt} of n', from the separation of the FECG point of view.

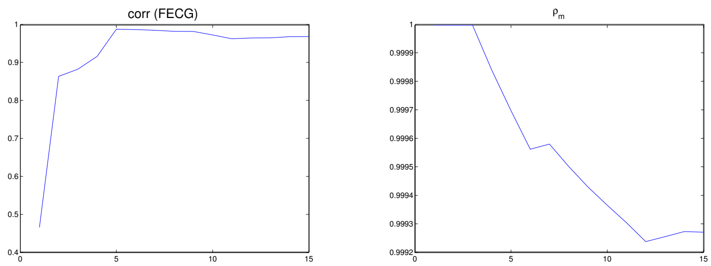

Figure 3: Evolution of the FECG separation *vs* the number of observed mixtures for and of the ρ_m with a selection preprocess (sPCA): n'=15, $n'_{opt} = 5$.

Fig. 4 shows the measured signals ($n = 5$) selected by the mutual information criterion (left) and the separated signals after ICA (right). The first selected signal is the reference and corresponds to MECG. We can clearly observe that the second one has an important component of FECG contribution. The first separated signal is the MECG and the fourth one is the FECG.

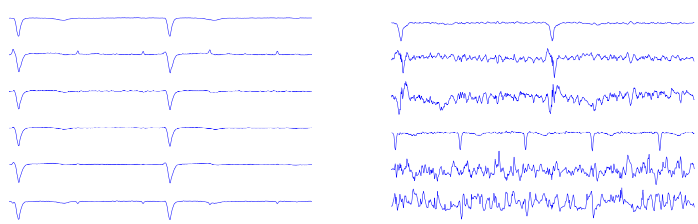

Figure 4: Left: measured signals selected by the mutual information criterion; right: separated signals.

CONCLUSION

Principal Component Analysis is a usual ICA preprocessing to project mixtures over a low-dimensional subset, with a dimension equal to the supposed number of sources. We have shown how sPCA (PCA preceded by variable selection using mutual information criterion) can improve the quality of the projection with respect to the classical PCA, but also the quality of the separation (ICA) of the sources. We have illustrated this improvement on a simulated problem of fetal ECG extraction from a set of signals measured on the mother's abdomen. One difficulty remains: the choice of the number n'

NNSP'2003 proceedings - Neural Networks for Signal Processing
Toulouse (France), 17-19 September 2003, IEEE, pp. 359-368

of signals to select.

In the simulated case, we can choose $n' = n'_{opt}$ were n'_{opt} is the number of selected electrodes which maximizes the correlation curve between the FECG and the associated recovered signal. However, n'_{opt} is not known in a real applications. Further work will adress the problem of choosing the n' parameter used in the proposed method.

REFERENCES

[1] A. K. Barros and N. Ohnishi, "Fetal heart rate variability extraction by frequency tracking," in **Proceedings of ICA'01**, 2001.

[2] J. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," **IEEE- proceedings F**, vol. 140, no. 6, pp. 362–370, 1993.

[3] P. Comon, "Independent Component Analysis, a new concept?" **Signal Processing**, vol. 36, no. 3, pp. 287–314, 1994.

[4] T. M. Cover and J. A.Thomas, **Elements of information theory**, John Wiley and sons, 1991.

[5] G. Deco and D. Obradovic, **An information-theoretic approach to neural computing**, Springer Verlag, 1996.

[6] A. Hyvärinen, "Gaussian Moments for Noisy Independent Component Analysis." **IEEE Signal Processing Letters**, vol. 6, no. 6, pp. 145–147, 1999.

[7] A. Hyvärinen, J. Karhunen and O. Oja, **Independent Component Analysis**, New York: Wiley-Interscience, 2000.

[8] I. T. Jollife, **Principal Component Analysis**, New York: Springer-Verlag, 1986.

[9] W. Lammers, "Circulating excitations and re-entry in the pregnant uterus," **Pflglers Arch-Eur J. Phys**, vol. 433, pp. 287–293, 1997.

[10] L. D. Lathauwer, B. D. Moor and J. Vandewalle, "Fetal electrocardiogram extraction by blind source subspace separation." **IEEE Trans. Biomedical Engineering**, vol. 47, no. 5, pp. 567–572, 2000.

[11] M. Marcus and H. Minc, **Introduction to linear algebra**, New York: Dover, 1988.

[12] K.-R. Muller, P. Philips and A. Ziehe, "JADETD : Combining higher-order statistics and temporal information for blind source separation (with noise)," **Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)**, pp. 87–92, 1999.

[13] J. V. Stone and J. Porrill, "Undercomplete independent component analysis for signal separation and dimension reduction." Techn. report, **Psychology Department, Sheffield University**, Sheffield, S10 2UR, England, Oct. 1997.

[14] A. Taleb and C. Jutten, "Source separation in post-nonlinear mixtures." **IEEE Trans. on Signal Srocessing**, vol. 47, no. 10, pp. 2807–2820, 1999.

[15] K. Torkkola, "Blind separation of delayed sources based on information maximization," **Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP'96)**, pp. 3509–3512, 1996.

[16] V. Vigneron, A. Paraschiv-Ionescu, A. Azancor, C. Jutten and O. Sibony, "Fetal electrocardiogram extraction based on non-stationary ICA and wavelet denoising." Techn. report, **LIS-INPG**, Grenoble, France, March 2003.