ESANN'2006 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2006, d-side publi., ISBN 2-930307-06-4.

Source separation with priors on the power spectrum of the sources

Jorge Igual[1], Raul Llinares[1] and Andres Camacho[1] *

1- Polytechnic University of Valencia - Dept of Communications, GATACA
Pza. Ferrandiz y Carbonell, 1, C.P. 03804, Alcoy - SPAIN

Abstract. A general approach introducing priors on the correlation function or equivalently power spectrum of the sources in the Blind Source Separation problem is presented. This prior modifies or constrains the contrast function that measures the independence of the recovered signals depending on its characteristics. Considering the case where the priors correspond to the sources that we are interested in recovering, the deflation approach is stated. This formulation is especially useful for those large-dimension problems where the ancillary sources are not needed to be estimated. We show its application to the biomedical problem of extracting the atrial activity from atrial fibrillation episodes, where discriminant information about the frequency content of the atrial activity with respect to the other components is available in advance.

1 Introduction

Blind Source Separation BSS consists of recovering the source signals from the observations obtained by mixtures of them. It is called blind because of nothing is assumed about the sources or the coefficients of the mixture but the statistical independence of the sources. However some assumptions are implicit in the model, such as: the sort of mixture (linear or non linear, instantaneous or convolutive, noisy or noise free), the dimensions of the problem (the size of the mixing matrix, i.e., the number of sources), at most one Gaussian source for algorithms based only on higher order statistics, non identical power spectra for algorithms based on second order statistics, i.e., exploiting the time structure, or the usual assumption of zero mean unit variance sources for simplification and fixing some indeterminacies.

In this paper we will follow the linear noise-free instantaneous mixing model:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \tag{1}$$

where $\mathbf{x}(t) = \left[x_1(t), \ldots x_m(t)\right]^T$ is the vector of observed signals, i.e., the mixtures, $\mathbf{s}(t) = \left[s_1(t), \ldots, s_n(t)\right]^T$ the source vector and $\mathbf{A}_{mxn}$ the mixing matrix. The aim is recovering the sources from the only assumption of the statistical independence of the sources $p(\mathbf{s}) = \prod_{i=1}^{n} p(s_i)$:

$$\mathbf{y}(t) = \mathbf{B}\mathbf{x}(t) \tag{2}$$

* This work is partially funded by Polytechnic University of Valencia and GVA. We also acknowledge BET group for the signals used in the results Section.

where $\mathbf{y}(t) = [y_1(t), \ldots, y_n(t)]^T$ are the recovered signals and **B** the demixing matrix.

When the mixing process (1) is not explicit, we call (2) the Independent Component Analysis ICA of $\mathbf{x}(t)$. In the square case $m = n$, the recovered sources are the original ones up to a permutation, sign and amplitude indeterminacy, $\mathbf{y} = \mathbf{Ps}$, where **P** is a matrix that has one and only one nonzero entry in each row and column.

When additional knowledge is included in the model, the problem is no more called *blind* and sometimes some restrictions can be relaxed, such as $m \geq n$. When information about the form of the densities is available, new approaches and *ad hoc* adaptations of classic algorithms can be obtained, e.g., when the sources are sparse [2]. We must note that some gradient based algorithms use component wise nonlinearities that include implicit priors about the pdf and the kurtosis of the sources, such as the Infomax algorithm [1]. But in this case we do not call them "priors"; if the algorithm fails, we simply say that the nonlinear function is not properly selected instead of talking about an incorrect prior. In fact, the sparseness is considered a prior in the literature but not the sign of the kurtosis in the selection of the nonlinearity, just because the former is formulated explicitly in the statement of the problem. Other usual source priors include its temporal structure [3] or the relaxation of the independence hypothesis [4]. Concerning the mixing matrix, priors can be modeled in a Bayesian approach [5], [6], or as constraints, being assumed some of the entries, e.g., due to the available information about the positions of sensors and sources, or being parameterized like in array signal processing.

We focus in this paper in the case where there is prior information about the power spectrum of some sources, at least one. We find many real applications such as in communications or biomedicine where this knowledge is available but not used by BSS algorithms. In addition, in many of these applications, we are interested in recovering only some few sources, so an algorithm that first extracts the interesting sources imposing the prior knowledge is necessary.

In Section 2, a general approach including the information about the frequency content of the sources is set. Depending on the prior, it includes the modification of the objective function that measures the independence or the restriction of the possible solutions. In Section 3, we present a biomedical example where the prior consists on that the power of the interesting source is concentrated in some frequencies, showing the results in Section 4.

2 Priors on the power spectrum of the sources

In basic ICA, solving (2) requires the use of higher order statistics in a direct way, such as cumulants, or in a more subtle way such as nonlinear functions. This is because only the decorrelation of the observations is not enough, remaining an orthogonal separating matrix for being estimated.

BSS algorithms are usually carried out in two steps. First, a PCA stage, consisting of whitening spatially the observations and reducing the dimension of the problem, $\mathbf{z}(t) = \mathbf{Vx}(t)$, where **V** is the *nxm* whitening matrix, so $E\{\mathbf{z}(t)\mathbf{z}(t)^T\} = \mathbf{I}$. Second, the remaining *nxn* orthogonal matrix **W**, $\mathbf{s}(t) = \mathbf{Wz}(t)$, with $\mathbf{WW}^T = \mathbf{I}$, is calculated. The estimated mixing matrix is $\mathbf{A} = \mathbf{V}^{+}\mathbf{W}^{T}$, where superscript $^{+}$ denotes the pseudoinverve matrix.

In this approach, nothing is said about the value of the covariance of the sources for non-zero lags. In other words, BSS algorithms not including the temporal information consider the data as mixtures of the samples of the independent random variables collected in the source vector **s**. This also means that scrambling the observations does not alter the solution and time index can be dropped because of the ordering of the observations is not important.

It is immediate to obtain the spatially whitened covariance matrices for different lags τ (we consider only real wide-sense stationary processes):

$$\mathbf{R}_{\mathbf{z}}(\tau) = E\left\{\mathbf{z}(t)\mathbf{z}(t-\tau)^{T}\right\} = \mathbf{W}^{T}\mathbf{R}_{\mathbf{s}}(\tau)\mathbf{W} \qquad (3)$$

Algorithms based on (3) exploits the fact that $\forall\,\tau\ \mathbf{R}_{\mathbf{s}}(\tau) = E\left\{\mathbf{s}(t)\mathbf{s}(t-\tau)^{T}\right\} = \mathbf{D}^{\tau}$, where $\mathbf{D}^{\tau_0}$ is a diagonal matrix whose element $d_{ii}^{\tau_0}$ is the covariance of the source i at lag τ_0. In this case, higher order statistics are not required and consequent restriction of at most one Gaussian component disappears. On the other hand, the use of the power spectrum information of the sources requires that components must have different covariances. In conclusion, a *dual* property between solutions based on higher order statistics and second order time covariances arises: identical distributions but non Gaussian sources opposite to non identical spectra but Gaussian sources.

We are interested in the case where some prior about the characteristics of the covariance function of the sources is available. As we mentioned before, this extra information is only related sometimes to some few of the sources, usually the time series that we are interested in recovering, so a deflation approach to source separation is preferred. This approach will be very useful also in the case where the dimension of the problem is very high, stopping the algorithm after the required sources are estimated. Obviously, the prior knowledge will depend on the kind of signals that are involved in our application.

The idea is to model this prior on the power spectrum with a function that will constraint or modify the contrast function used to measure the independence of the sources. Considering the whitened signals $\mathbf{z}(t)$, we must obtain the *nx1* unit-norm vector $\mathbf{w}_1$ that recovers one of the interesting sources $y_1(t) = \mathbf{w}_1^{T}\mathbf{z}(t)$.

The deflation approach can be easily extended to recover other components searching for another vector $\mathbf{w}_i$ at every iteration that maximizes the contrast function and is orthogonal to the $\mathbf{w}_k, k = 1,\ldots,i-1$ vectors obtained for the previous extracted components. At each one source extraction stage, the proper prior must be considered if there exists. If this process is repeated up to the last component $y_n(t)$, then we have obtained the full demixing matrix **W**.

We will denote by $J(\mathbf{w})$ the contrast function that measures the independence of the sources and that must be maximized. Any of the different contrast functions available in the literature is admissible. With respect to the prior, from a mathematical point of view, it could be modeled as a function to be maximized (minimized) or as an equality or inequality constraint. In the first case, because the minimization problem can be easily converted to a maximization one changing the sign, the contrast function $J(\mathbf{w})$ will be modified including the prior, obtaining a new contrast function $J_1(\mathbf{w})$. The problem is stated as:

$$\mathbf{w}_1 = \arg\max_{\mathbf{w}} J_1(\mathbf{w}), \quad \textit{subject to } \mathbf{w}_1^T \mathbf{w}_1 = 1 \tag{4}$$

In the second case, for priors that can be expressed as a set of M equality and inequality constraints, the Lagrangian and penalty functions are the most common ways of replacing the constrained optimization problem with an unconstrained one:

$$\mathbf{w}_1 = \arg\max_{\mathbf{w}} J(\mathbf{w}), \textit{subject to } H_k(\mathbf{w}_1) = 0,\, k = 1\ldots K, H_m(\mathbf{w}_1) \geq 0,\, m = K+1\ldots M \tag{5}$$

Note that $H_k(\mathbf{w}_1) = \mathbf{w}_1^T \mathbf{w} - 1 = 0, \quad 1 \leq k \leq K$. Remember that $\mathbf{w}_1$ is orthonormal because of the independence of the sources and the sphering of the data during the PCA step, $\mathbf{R}_\mathbf{z}(0) = \mathbf{I} = \mathbf{R}_\mathbf{s}(0)$. It means that the condition $\|\mathbf{w}_1\| = 1$ is the same as a prior on the correlation function $R_{y_1}(0) = 1$. This restriction can be also carried out as a normalization step $\mathbf{w}_1 \leftarrow \mathbf{w}_1 / \|\mathbf{w}_1\|$.

3 A biomedical application: extraction of the atrial activity in atrial fibrillation recordings

Atrial fibrillation AF is one of the most common cardiac arrhythmia, with a high morbity studied since some decades ago [7]. During AF episodes, the contractions of the ventricles are usually irregular and may average 100-150 per minute, decreasing the amount of pumped blood. If it happens, the body begins to compensate by retaining fluid, leading to the accumulation of fluid in the legs or lungs (edema). In addition, quivering of the atria in AF causes blood inside the atria to stagnate and to form blood clots producing embolism.

The recorded ECG can be considered as the superposition of the atrial activity AA (the interesting source), the higher power ventricular activity VA and other ancillary signals. BSS has been applied successfully to the AA extraction problem [8] thanks to the fulfillment of the independence assumption between AA and VA. The AA is identified after separation because the AA power is concentrated around a peak in the frequency range [4-10] Hz, depending on the patient. This spectral feature is particular of the AA, contrary to the other components with no peak in this band. However, this prior information has never been used by the other BSS algorithms proposed for the AA extraction [9], nor other non BSS solutions [10].

The source separation solution including this prior requires its mathematical modeling. The restriction consists on the maximization of the integral of the power spectrum in the proper range of frequencies:

$$\hat{J}(y, \mathbf{w}) = \int_{f_1}^{f_2} \phi_y(f) df \tag{6}$$

where $\phi_y(f)$ is the power spectrum of the recovered signal $y(t)$. It is maximum for the AA component $y_{AA}(t)$; the interval of integration $[f_1, f_2]$ depends on the prior information about the patient. We will assume that the range of frequencies is [4-10] Hz, i.e., there is no prior knowledge about the kind of neither AA nor the patient.

As we can see, (6) corresponds to a case of equation (4) where the prior is modeled by a function that must be also maximized in addition to the contrast function used for the blind solution. Hence, a new contrast function is obtained $J_1(\mathbf{w}) = J(\mathbf{w}) \cdot \hat{J}(\mathbf{w})$. The AA source is:

$$y_{AA}(t) = \mathbf{w}_{AA}^T \mathbf{z}(t) \tag{7}$$

If we use the absolute value of the kurtosis [11] as the BSS contrast function $J(\mathbf{w}) = |kurt(y)|$, $\mathbf{w}_{AA}$ is calculated as:

$$\mathbf{w}_{AA} = \arg\max_{\mathbf{w}} \left(|kurt(y)| \cdot \int_4^{10} \phi_y(f)df \right) \quad subject\ to\ \mathbf{w}_{AA}^T \mathbf{w}_{AA} = 1 \tag{8}$$

We use the periodogram as the method to estimate the spectrum.

$$\hat{\phi}_y(e^{j\omega}) = \frac{1}{N} \left| Y(e^{j\omega}) \right|^2 = \frac{1}{N} \boldsymbol{\omega}^{*T} \vec{\mathbf{y}} \vec{\mathbf{y}}^T \boldsymbol{\omega} \tag{9}$$

with N the data size, $\boldsymbol{\omega} = [1, e^{j\omega}, \ldots, e^{j\omega(N-1)}]^T$ and $\vec{\mathbf{y}} = [y(0), y(1), \ldots, y(N-1)]^T$.

Obviously, the integral in (8) is calculated numerically and the frequencies normalized by the sampling frequency. Using (7) and (9), we can express (8) as:

$$\mathbf{w}_{AA} = \arg\max_{\mathbf{w}} \left(\left| E\left\{ (\mathbf{w}^T \mathbf{z})^4 \right\} - 3\left(E\left\{ (\mathbf{w}^T \mathbf{z})^2 \right\} \right)^2 \right| \cdot \frac{1}{N} \sum_{i=I_1}^{I_2} \mathbf{w}^T \mathbf{Z} \boldsymbol{\omega}_i \boldsymbol{\omega}_i^{*T} \mathbf{Z}^T \mathbf{w} \right) \tag{10}$$

with $\mathbf{w}^T \mathbf{w} = 1$, $[I_1, I_2]$ the corresponding interval of digital frequencies and $\mathbf{Z} = [\mathbf{z}(0), \mathbf{z}(1), \ldots, \mathbf{z}(N-1)]$. Different optimization techniques can be used to solve (10). One of them is the FastICA algorithm constrained by the prior:

$$\begin{aligned} &\mathbf{w} \leftarrow E\left\{ \mathbf{z}(\mathbf{w}^T \mathbf{z})^3 \right\} - 3\mathbf{w} + \mathbf{C}\mathbf{w} + \mathbf{C}^T \mathbf{w}, \quad \mathbf{C} = \sum_{i=I_1}^{I_2} \mathbf{Z} \boldsymbol{\omega}_i \boldsymbol{\omega}_i^{*T} \mathbf{Z}^T \\ &\mathbf{w} \leftarrow \mathbf{w} / \|\mathbf{w}\| \end{aligned} \tag{11}$$

4 Results

We applied the constrained FastICA algorithm (CFastICA) defined by equation (11) to five patients of AF. The experiments were carried out to many different 10 seconds intervals of the recordings. The first recovered source was always the AA, contrary to BSS algorithms non including the prior. In Fig. 1 we see that with the FastICA algorithm [11], it should have been necessary to extract eight components to including the interesting one, the AA. The advantages are clear: time of computation is reduced, the cost function is optimized, the postprocessing task to identify the AA is not necessary, the problem of propagating errors of deflation algorithms is eliminated and further AA analysis can be automatized for large databases because the permutation of the recovered components to allocate in the same position the AA is not required (in our multiple experiments with different recordings, the position in which AA is extracted ranges between the 5th and 8th place for any BSS non constrained algorithm for the case of 8 components). We show the sources #1 up to #4 recovered by CFastICA and the power spectrum of the sources #1 up to #3.

5 Conclusions

We have presented a general deflation approach to include priors on the power spectrum or covariances of the sources. We have shown that this is equivalent to modify the BSS original contrast function or to constrain it, depending on the way we model it, that can vary for every application. We applied it to a biomedical problem, where a well established in the literature prior condition about the power spectrum of the interesting signal was included. This approach eliminates the indeterminacy of the permutation, exhausts the prior information about the sources usually not used by blind classical techniques and can be very useful in problems where the number of sources is very high and we are only interested in some few of them.

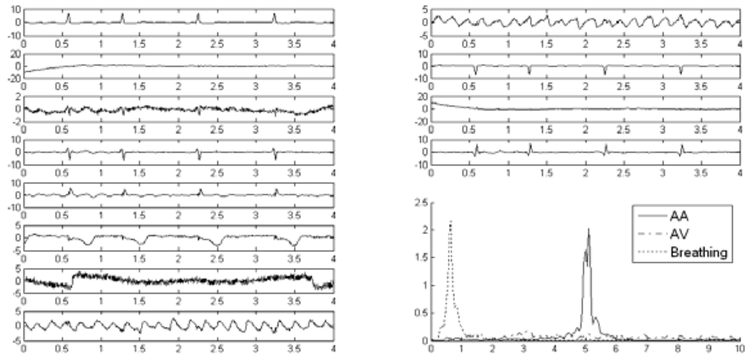

Fig.1. Left, estimated sources by FastICA; AA is the last one. Right up, sources #1-4 estimated by CFastICA; AA is the first one. Right bottom, power spectrum of the sources #1-3.

References

[1] A.J. Bell, T.J. Sejnowsky, An information-maximization approach to blind separation and blind deconvolution, *Neural computation* 7: 1129-1159, MIT Press, 1995.

[2] A. Hyvarinen, Sparse code shrinkage: denoising of nongaussian data by maximum likelihood estimation, *Neural computation* 11: 1739-1768, MIT Press, 1999.

[3] A. Belouchrani, K. Abed-Meraim, J.F. Cardoso, A Blind Source Separation technique using second-order statistics, *IEEE transactions on Signal Processing,* 45: 434-444.

[4] A. Hyvarinen, P. Hoyer, M. Inki, Topographic independent component analysis, *Neural computation* 13, MIT Press, 2001.

[5] J. Igual, A. Camacho, L. Vergara, Independent component analysis with prior information about the mixing matrix, *Neurocomputing* 50: 419-438 , Elsevier, 2003.

[6] K.H. Knuth, A Bayesian approach to source separation. In J.F. Cardoso, C. Jutten and P. Loubaton, editors, proceedings of the *1st Int workshop on ICA and Signal Separation* (ICA 1999), pp. 283-288.

[7] W.B. Kannel, R. Abbott, D. Savage, P. McNamara, Epidemiologic features of chronic atrial fibrillation: the Framingham study, *N Eng J Med* 306: 1018-1022, 1982.

[8] J.J. Rieta, F. Castells, J. Igual, C. Sanchez, ICA applied to atrial fibrillation analysis. In proceedings of the *3rd Int workshop on ICA and Signal Separation* (ICA 2003), Nara, Japan, 2003.

[9] Castells F., Igual J., Rieta J.J, Sanchez J., Millet J., Atrial fibrillation analysis based on ICA including statistical and temporal source information. *ICASSP 2003* Hong Kong, pp. 59-64, 2003.

[10] M. Stridh, L. Sornmo, Spatiotemporal QRST cancellation techniques for analysis of atrial fibrillation, *IEEE Trans. Biomedical Eng.* 48: 105-111, 2001.

[11] A. Hyvarynen, E.Oja, A fast fixed-point algorithm for ICA, *Neural computation* 9:1483-1492, 1997.