Blind signal processing algorithms

Athanasios Margaris and Efthimios Kotsialos
Department of Applied Informatics, University of Macedonia, 540 06 Greece

Abstract: The objective of this survey paper is the presentation of the main aspects associated with blind signal processing. The paper describes some of the most important blind signal processing methodologies and briefly presents the fundamental algorithms used for their implementation.

Keywords: blind signal processing; blind source separation; blind identification; blind equalization.

Reference to this paper should be made as follows: Margaris, A. and Kotsialos, E. (2005) 'Blind Signal Processing Algorithms', *Proceedings of the* 12^{th} *International Workshop on Systems, Signals, and Image Processing*, Vol. x, No. x, pp.xxx–xxx.

Biographical notes: Athanasios Margaris received his PhD in Computer Science from the University of Macedonia. He is now a lecturer of informatics in Technological Educational Institute of Thessaloniki. His research interests include modelling of dynamical systems using artificial neural networks, parallel implementations of neural networks and blind signal processing applications.

Efthimios Kotsialos is a PHD Candidate in the Department of Applied Informatics of the University of Macedonia, Thessaloniki, Greece. His research interests are in the domain of $3D$ continuous dynamical systems and neural network modelling of chaotic dynamics, and he is currently involved in a research project concerning blind signal processing algorithms.

1 INTRODUCTION

Blind signal processing is a set of algorithms that are applied on a recorded signal without any prior knowledge about the underlying system. This family of algorithms plays an important role in many applications, such as communications, engineering tasks, and data analysis. There are many different types of blind signal processing; among the most important of them are the following:

Blind source separation: in this technique, a set of unknown source signals $s_1(t), s_2(t), \ldots, s_m(t)$ that are considered mutually independent of each other, are linearly mixed in an unknown environment to produce an n-dimensional observation vector $\vec{x}(t) = A\vec{s}(t)$ where A is an $m \times m$ unknown nonsingular mixing matrix. Given the observation vector $\vec{x}(t)$, the requirement is to recover the source signals $s_1(t), s_2(t), \ldots, s_m(t)$ in an unsupervised manner [1].

There are many different blind source separation methods according to the assumptions made about the sources, the filter and the additive noise, and a full taxonomy of them is not complete at this point. In a short description, sources may be statistically independent, may have different second order spectra, may be stationary or (quasi)-cyclostationary, may be non Gaussian, while in other cases they are considered independent and identically distributed. The basic hypothesis of the system is that in the most of cases it is modelled as a linear time invariant filter, while the assumptions for the additive noise describe a Gaussian or non Gaussian noise, a temporally white noise, a spectrally uncorrelated noise, or a spectrally correlated noise with known spatial coherence [2].

Blind identification: consider an invertible matrix A, and a source vector of independent components, $\vec{s}$. If there is another matrix, B, such that the vector $\vec{x} = A\vec{s}$ is characterized by the same distribution of a vector $B\vec{y}$, where $\vec{y}$ is a second vector of independent components, then the matrix A and the vector $\vec{s}$ can not be identified in a unique way. The blind identification is associated with treatment of this situation. Typical blind identification methods are described in [3] and [4].

Blind equalization: as it is well known from the basic signal processing theory, a signal $s(k)$ that represents a digital message is subjected to a set of transformations, such as encoding and modulation, before its transmission from a single sender to a single receiver. Upon receipt, the signal is subjected to the inverse transformations of demodulation and decoding. The effect of the transmission channel to the received signal can be removed by applying a second order discrete time filter, whose impulse response can be estimated via a training operation and by using a training sequence known both to the sender and receiver. However, in broadcast applications where a single sender interacts with multiple receivers, these training sequences are not available. In this case the equalization technique has to be performed without the knowledge of the training sequences described above [5]. This special equalization type is known as blind equalization or, in the domain of telecommunications, as blind deconvolution.

In the case of multichannel signals, the equalization is a preprocessing step before the actual signal separation - the equalization itself is not a signal separation technique. More specifically, while in the initial phase, there are temporal and spatial mixtures for a set of signals, after the equalization, each mixture is a convolution of only one of the initial signals.

Blind signal extraction: the objective here is the extraction of a small number of specific signals from a large set of recorded signals. These signals can be extracted with a certain order, according to the values of their stochastic properties, such as the generalized normalized kurtosis. This method is more general than Principal Component Analysis (PCA), and is strongly related to the method of blind deconvolution.

2 BASIC CONCEPTS

A necessary preprocessing step before the application of the algorithms described above is the whitening or 'sphering' of the observation data. In this stage, a whitening matrix, W, is applied on the observation vector, thereby converting the mixing matrix into a rotation matrix, which is easier to identify. In a mathematical notation, the whitening constraint is described by the equation $E\{H_w(\vec{y})\} = 0$, where $H_w(\vec{y}) = \vec{y}\vec{y}^T - I$. By assuming a square mixing matrix of order n, the whitening constraint imposes $n(n+1)/2$ constraints, leaving thus $n(n-1)/2$ unknown parameters to be determined by other than second order information.

In order to achieve blind source separation, we device real functions of the probability distributions known as contrast functions. These are considered as objective functions in the sense that the source separation is achieved when they reach their minimum value. There are many contrast functions, such as the maximum likelihood (ML) contrast $\Phi_{ML}[\vec{y}] = K[\vec{y}|\vec{s}]$ and the orthogonal mutual information contrast $\Phi^0_{MI}[\vec{y}] = \sum_i H[y_i]$. In the above notation, $K[\vec{x}|\vec{y}]$ is the Kullback divergence between the distributions of the random vectors $\vec{x}$ and $\vec{y}$ - in general, this distance between two probability density functions $f(s)$ and $g(s)$ is given by the equation

$$K(f|g) \triangleq \int_S f(s) log\left(\frac{f(s)}{g(s)}\right) ds \qquad (1)$$

while $H[\vec{y}]$ is the Shannon entropy. However, since these quantities are very difficult to calculate analytically, higher order approximations of those information theoretic contrasts are used, that can be estimated directly from the experimental data. A typical example of such an approximation, is the approximation of the likelihood contrast - this measures the mismatch between the output distribution and a model source distribution - by the quadratic mismatch between the cumulants of the second and fourth order $\Phi_2[y]$ and $\Phi_4[y]$. The main advantage of the fourth order cumulants is that they can be optimized by means of the iterative Jacobi algorithm [1].

Even though a valid contrast function reaches its minimum value at the separation point when the model is valid, in practice, contrasts are only estimated from a finite data set. This estimation introduces stochastic errors depending on the available samples and also on the contrast function. This means that a statistical characterization of the minima of the contrast functions is necessary; this characterization is based to the usage of the so-called estimating functions and by means of quantities such as natural gradient or relative gradient. More information about these techniques can be found in the literature [1][2].

3 BLIND SOURCE SEPARATION

Blind source separation can be achieved by many different methods; those can be either convolutive or instantaneous. In the case of the convolutive model, we assume the existence of P random (in general) source signals, corrupted by additive noise that are propagated though a linear deterministic channel and received by a set of K sensors. On the other hand, in the static model, the propagation channel is instantaneous (and thus we neglect any time delays that may occur in the mixing) and a single propagation link joins each source to every sensor. Regarding the number of sources and sensors, we can distinguish between three system types, namely the SISO (Single Input, Single Output) systems for values $P = 1$, $K = 1$, the SIMO (Single Input, Multiple Outputs) systems for values $P = 1$, $K > 1$, and the MIMO (Multiple Inputs, Multiple Outputs) systems, with values $P > 1$, $K > 1$. Regarding the methods used for blind source separation, the most important of them are the following:

The principle of Infomax: this method is based on the maximization of the output entropy of a neural network with nonlinear outputs [6]. This network reads a set of input vectors, $\vec{x}$, and produces outputs in the form $\vec{y}_i = f_i(w_i^T \vec{x}) + \vec{n}$ where f is a nonlinear scalar function, w_i the weight vectors of the neural network, and $\vec{n}$ is an additive Gaussian white noise. The key point of this method is that if the nonlinear transport function of the neural network matches the probability density function of the input vectors and the joint probability of the neuron outputs

$$H(\vec{y}) = H(f_1(w_1^T\vec{x}), f_2(w_2^T\vec{x}), \ldots, f_n(w_n^T\vec{x})) \quad (2)$$

gets its maximum value, then, the mutual information of the network output $I(\vec{y}) = I(y_1, y_2, \ldots, y_n)$ is minimized and the output signals are assumed to be independent. Bell and Sejnowski [7] proved that most real word signals are super Gaussian and satisfy the statistical independence criterion when the transfer function of the neural network is the sigmoidal function or the function of the hyperbolic tangent.

The improvement of the Infomax algorithm to work with general non-Gaussian signals (and not only super Gaussian as the original version) and the development of the extended Infomax algorithm has been done by Te-Won Lee in collaboration with Mark Girolami [8]. This extended algorithm is used in a lot of applications such that in the processing of EEG signals. These signals originating from the brain are quite weak at the scalp in the microvolt range and there are larger artifact components from eye movements and muscles. It has been shown that the isolation and elimination of these artifacts without altering the brain signals can be ideally performed by the extended Infomax algorithm, since the recorded signals are different linear mixtures of the brain signals and the artifacts. There are many other applications of this algorithm such that the analysis of extremely large data sets from functional magnetic resonance imaging (fMRI) experiments [9].

The JADE algorithm (Joint Approximate Diagonalization of Eigenmatrices) [10][11]: this method is based on the usage of the forth order cumulant tensor; it tries to maximize

$$J(A) = \sum_i \|diag(AF(M_i)A^T)\|^2 \quad (3)$$

where A is the whitened mixing matrix and M_i are the eigenmatrices of the fourth order cumulant tensor (the eigenvalues of those eigenmatrices are the kurtosis values of the independent component, which can be calculated in this way). The starting point of the JADE algorithm is that the requirement of the most BSS algorithms to calculate the distributions of the independent components, can be fulfilled by optimizing the cumulant approximations of data. The advantage of the JADE approach is that the gradient descent algorithm is not used, and therefore there are not problems of convergence. On the other hand, the main disadvantage is the storage of $O(N^4)$ cumulant matrices for the calculation of the complete set of fourth cumulants.

The last interesting source separation algorithm presented here, is the FastICA or fixed point algorithm [12]. This family of algorithms is associated with the projection pursuit style of methods and it is based on the fact that the independent components are bound to the projections whose distributions have the greatest possible distance from the Gaussian distribution. There are two approaches for this class of algorithms: the symmetric approach that uses a modified update rule that enables simultaneous separation of all independent components, and the deflation approach that finds the independent components once at a time. Regarding the contrast functions used by these techniques (kurtosis), the hyperbolic tangent as well as exponential or cubic functions can be used.

The main advantage of the FastICA algorithm is that it uses a fixed point iteration schema that has been found (in independent experiments) to be 10-100 times faster than conventional gradient descend methods for ICA. Another advantage of this algorithm is that it can be used to perform projection pursuit as well, thus providing a general-purpose data analysis method that can be used both in an exploratory fashion and for estimation of independent components.

Besides the approaches described above, there are many other approaches, which cannot be described here due to space limitations. The most important of them are the following: (a) The TDSEP (temporal decorrelation source separation) algorithm [13], exploiting the temporal structure of signals, in order to compute the time-delayed second order correlation for the source separation. The best results are achieved if the autocorrelations are as different as possible. (b) Blind separation of disjoint orthogonal signals [14], that uses only 2 mixtures of N sources, but the sources have to be pair-wise disjointly orthogonal. The algorithm is based on the Short Time Fourier Transform. (c) Principal component analysis (PCA)-known sometimes as KL-transform- which uses second-order methods in order to reconstruct the signal in the mean square error sense [1]. In this method, the PCA basis vectors are mutually orthogonal.

4 BLIND IDENTIFICATION

The most interesting algorithm for blind identification is the SOBI (Second Order Blind Identifiability) algorithm [15] that is based to the joint diagonalization of several covariance matrices. Even through in the blind context, a full identification of the mixing matrix is impossible because the exchange of a fixed scalar factor between a given source signal and the corresponding column of A does not affect the observations, we can assume without loss of generality that the source signals have unit variance, so that the dynamic range of the sources is accounted for by the magnitude of the corresponding columns of A. Based on this assumption, the SOBI algorithm is defined by the following implementation:

1. Estimate the sample covariance $R(0)$ from T data samples. Denote by $\lambda_1, \lambda_2, \ldots, \lambda_n$ the n largest eigenvalues and by $h_1, h_2, \ldots, h_n$ the corresponding eigenvectors of $R(0)$

2. Under the white noise assumption, an estimate σ^2 of the noise variance is the average of $m - n$ smallest eigenvalues of $R(0)$. The whitened signals are $\vec{z}(t) =$

$[z_1(t), z_2(t), \ldots, z_n(t)]^T$, which are computed by

$$z_i(t) = \frac{1}{(\lambda_i - \sigma^2)}(h_i * \vec{x}(t)) \quad (4)$$

for $1 \leq i \leq n$

3. Form sample estimates $S(\tau)$ by computing the average covariance matrices of $z(t)$ for a fixed set of time lags $\tau \in \{\tau_j | j = 1, 2, , K\}$

4. Obtain the unitary matrix U as joint diagonalizer of the set $\{S(\tau_j) | j = 1, 2, , K\}$

5. Estimate the source signals as $\vec{s}(t) = U^H W \vec{x}(t)$ and/or the mixing matrix A as $A = W^\sharp U$ where the superscript $\sharp$ denotes the Moore-Penrose pseudoinverse while the super-script H denotes Hermitian transposition.

The SOBI algorithm is a blind source separation technique relying only on second-order statistics of the received signals that allows - in contrast to higher order cumulant techniques - the separation of Gaussian sources. A typical application of this algorithm is to perform single-trial classification of EEG data for the development of Brain Computer Interface (BCI) which can support communication abilities for physical disabled parients [16]. This application is based to the ability of SOBI to separate functionally distinct neuronal signals from each other and to recover components that were physiologically and neuroanatomically interpretable. These experiments showed that the convergence of the SOBI algorithm was achieved after only a few tens of iterations, thus making real time applications a future possibility.

5 BLIND EQUALIZATION

In blind equalization, the observed discrete time signal is generated from an unknown source signal by a convolution model that mixes together delayed versions of the source signal. In order to estimate the deconvolution filter, we make the assumption that the source signal values $\vec{s}(t)$ at different times t, are nongaussian, statistically independent and identically distributed. The probability distribution of the source signal may be known or unknown. Blind equalization is used in many applications, such as wireless telecommunications, sonar and radar systems, audio and acoustics, image enhancement, and biomedical signal processing.

An interesting family of algorithms for blind equalization are the Bussgang methods [17] that use a non-causal FIR filter, whose weight values are functions of time and they can be adopted using the least squares algorithm. The filter output $y(t)$ is passed to a nonlinear function $g(x)$ satisfying the condition

$$E\{y(t)y(t-k)\} \approx E\{y(t)g(y(t-k))\} \quad (5)$$

The choice of the nonlinearity g(x) leads to different Bussgang type algorithms such as the Goddard algorithm, which tries to minimize the non-convex cost function

$$J_p(t) = E\{[|y(t)|^p - \gamma_p]^2\} \quad (6)$$

In this notation, p is a positive integer, while the parameter γ depends on the statistics of the source signal. It can be proven that for the parameter value $p = 2$, the above algorithm - identified now by the name constant modulus algorithm (CMA)- and the cost function $J_p(t)$, are related with the minimization of the kurtosis. The main drawback of the Bassgang methods is that it is possible for the iterative equalization schema implemented by them to converge to wrong solutions resulting thus in poor performance of the equalizer.

The second family of blind equalization techniques includes the cumulant-based methods; these use higher order statistics of the observed signal, $\vec{x}(t)$. One interesting algorithm is that of Shavi and Weinstein [18]. This is a stochastic gradient algorithm that tries to maximize a constraint kurtosis-based criterion. In the case of a whitened complex valued and symmetric source signal $\vec{s}(t)$ satisfying the condition $E\{\vec{s}(t)^2\} = 0$, the Shavi and Weinstein algorithm is given by the equations

$$\vec{u}(t+1) = \vec{u}(t) + \mu * sign(k_s)[|\vec{z}(t)|^2 \vec{z}(t)]\vec{y}^*(t) \quad (7)$$
$$\vec{w}(t+1) = \vec{u}(t+1)/\|\vec{u}(t+1)\| \quad (8)$$

where $\vec{y}(t)$ is the whitened output vector, $\vec{w}(t)$ is the M-dimensional weight vector of the causal FIR deconvolution filter of length M, k_s is the kurtosis of the source signal $\vec{s}(t)$,$\vec{u}(t)$ is the unnormalized filter weight vector, $\|..\|$ is the vector's norm, μ is a parameter of the method, and $\vec{z}(t) = \vec{w}^T(t)\vec{y}(t)$ is the filter's output.

The last algorithm for blind deconvolution presented here is based on state-space models [19], and on the assumption that both mixing and demixing models are described by stable linear state-space systems. The state equation of such a system has the form

$$\vec{x}(k+1) = A\vec{x}(k) + B\vec{s}(k) \quad (9)$$

where $\vec{s}(t)$ is a source vector with independent and identically distributed sources, $\vec{x}(k)$ is the state vector, A is the state mixing matrix, and B is the input mixing matrix. Regarding the system's output, this can be written as $\vec{u}(k) = C\vec{x}(k) + D\vec{s}(k)$, where C is the output mixing matrix and D is the input-output mixing matrix. Finally, the transfer function of the system is given by the equation

$$H(z) = C(zI - A)^{-1}B + D \quad (10)$$

where z^{-1} is the delay operator.

The basic suggestion of this algorithm is that the demixing model should be described as a linear state-space system too, with transfer function $W(z) = C'(zI - A')^{-1}B' + D'$ and output $\vec{y}(k) = W(z)H(z)\vec{s}(k)$. Based on the above description, the blind deconvolution problem becomes an optimization problem with the mutual information $L(W)$ to be the risk function. In the last step of the algorithm, the parameters of the problem are updated by a learning algorithm which is a generalized form of the natural gradient learning algorithm in the state-space model mixture - the two algorithms become the same if an instantaneous mixing is used and the matrices A, B, C, and D become the null matrices.

The main advantage of the state space description of blind deconvolution is that it not only gives the internal description of a system, but it also provides various equivalent types of state space realizations for a system, such as balanced realizations and observable canonical forms.

The blind deconvolution or equalization problem can also be resolved by means of other methods that are based to eigenvector calculations, least square approaches, and the relative Newton method [20].

6 CONCLUSIONS

This paper presents standard algorithms for basic blind signal processing techniques, such as blind source separation, blind equalization (or deconvolution) and blind identification. More specifically, the blind source separation algorithms mentioned here are the Infomax method, the JADE approximation and the FastICA or fixed-point family of algorithms. Regarding blind identifiability, the SOBI algorithm is presented, while for the technique of blind equalization the Bussgang methods - and more specifically the Godard algorithm and the Constant Modulus Algorithm - as well as the algorithm of Shavi/Weinstetin and the state space deconvolution has been described. These algorithms use second order as well as higher order statistics and the model associated with them can be convolutive or instantaneous. More information about those algorithms - and many others that are not possible to be presented here - can be found in literature.

ACKNOWLEDGEMENTS

Work supported by the "EPEAEK Archimedes II" Programme, funded in part by the European Union (75%) and in part by the Greek Ministry of National Education and Religious Affairs (25%).

REFERENCES

1 Hyvarinen A., Karhunen J., Oja E. (2001) *Independent Component Analysis*, John Wiley & Sons.

2 Haykin S. (Ed.) (2000) *Unsupervised Adaptive Filtering, Volume I, Blind Source Separation*, John Wiley & Sons.

3 Comon P. (2004) *Blind Identification and Source Separation in* 2×2 *Under-determined Mixtures*, IEEE Trans. Signal Proc., Vol. 52, No.1, pp. 11-22, January 2004.

4 Hua Y., Am S., Xiang Y. (2003) *Blind Identification of FIR MIMO Channels by Decorrelating Subchannels*, IEEE Trans. Signal Proc., Vol. 51, No. 5, pp. 1143-1155, xxxx.

5 Farid A., Luo Z-Q., Ding Z. (2005) 'Blind Channel Equalization Based on Second Order Statistics', *Proceedings of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, March 18-23, 2005, pp. 433-436, Pennsylvania, USA.

6 Cardoso J.F (1997) *Informax and Maximum Likelihood for Blind Source Separation*, IEEE Signal Processing Letters, Vol. 4, No. 4, pp. 112-114, 1997.

7 Bell A.J., Sejnowski J.T. (1995) *An Information Maximization Approach to Blind Source Separation and Blind Deconvolution*, Neural Computation, Vol.7, pp.1129-1159, 1995.

8 Lee T.-W., Girolami M., Sejnowski (1999) *Independent Component Analysis Using an Extended Infomax Algorithm for Mixed sub-Gaussian and super-Gaussian Sources*, Neural Computation, Vol. 11, pp. 417441, 1999.

9 McKeown M. J., Hansen L. K., Sejnowski J. T. (2003) *Independent component analysis of functional MRI: what is signal and what is noise?*, Current Opinion in Neurobiology, Vol. 13, No. 5, pp. 620-629, 2003.

10 Cardoso J.F, Souloumiac A. (1993) *Blind Beamforming for non-Gaussian signals*, IEEE Proceedings, Part F, Vol. 140, No. 6, pp. 362-370, 1993.

11 Comon P. (1994) 'Tensor Diagonalization, a Useful Tool in Signal Processing', in Blanke M., Soderstrom M. (Eds.) *IFAC-SYSID 10th Sympo-sium on System Identification*, Denmark, Vol. 1, pp 77-82, 1994.

12 Hyvarinen A., Oja E. (1997) *A Fast Fixed Point Algorithm for Independent Component Analysis*, Neural Computation, Vol. 9, No. 7, pp. 1483-1492, 1997.

13 Ziehe A., Muller K.R. (1998) 'TDSEP - An Efficient Algorithm for Blind Separation Using Time Structure', in *Proc. of International Conference on Artificial Neural Networks*, pp. 675-680, Sweden, 1998.

14 Jourjine A., Rickard S., Yilmaz O (2000) 'Blind Separation of Disjoint Orthogonal Signals: Demixing N Sources from 2 Mixtures', *Proceedings of IEEE Conference on Acoustics, Speech, and Signal Processing*, Vol. 5, pp. 2985-2988, Istanbul, Turkey, June 2000.

15 Belouchrani A, Meraim K.A., Cardoso J.F., Moulines E (1997) *A Blind Source Separation Technique Based on Second Order Statistics*, IEEE Trans. Signal Proc. Vol. 45, No. 2, pp 434-444, 1997.

16 Wang Y., Sutherland M. T., Sanfratello L. L., Tang A. C. (2004), 'Single-Trial Classification of ERPs Using Second-Order Blind Identification (SOBI)' in *Proceedings of the Third Intnl Conference on Machine Learning and Cybernetics*, Shanghai, China, August 26-29, Vol. 7, pp.4246-4251, 2004.

17 Lambert R.H. (1996) *Miltichannel Blind Deconvolution: FIR Matrix Algebra and Separation of Multipath Mixtures*, PHD Thesis, University of Southern California, 1996.

18 Shalvi O., Weinstein E. (1990) *New Criteria for Blind Deconvolution of Nonmiminun Phase Systems (Channels)*, IEEE Trans. Inf. Theory, Vol.36, No.2, pp.312-321, 1990.

19 Zhang L, Cichocki A. (1998) 'Blind Deconvolution / Equalization Using State Space Models', in *Proceedings of the IEEE Workshop on NNSP*, pp. 123-131, 1998.

20 Luenberger D. G. (1984) *Linear and Non Linear Programming*, Second edition, Reading, Addison Wesley.