A BLIND SOURCE SEPARATION ALGORITHM USING WEIGHTED TIME DELAYS

Bert-Uwe Köhler and Reinhold Orglmeister

Institute of Electronics, Berlin University of Technology,
Einsteinufer 17/EN3, 10587 Berlin, Germany
E-mail: koehler@tubife1.ee.tu-berlin.de

ABSTRACT

A new algorithm for the blind separation of mixed non-white signals is proposed. After removing the second order cross-correlation between the signals for zero time delays (sphering), the signals are linearly transformed in order to reveal common components that are still existent within the signals. The linear transform is accomplished by filtering the sphered signals using a non-recursive filter which is equivalent to the summation of weighted time delayed signal values. This leads to a non-zero cross-correlation that is used to eventually separate the signals. In this paper it is shown how the filter coefficients are determined and how this principle is applied to blind source separation.

1. INTRODUCTION

Recently, independent component analysis (ICA) has become an attractive tool for blind source separation (BSS) [1, 2, 3, 4, 5, 6, 7, 8, 9]. To assume the source signals as statistically independent makes the blind source separation problem mathematically tractable. Due to their close relationship, in this paper independent component analysis and blind source separation are considered as equivalent.

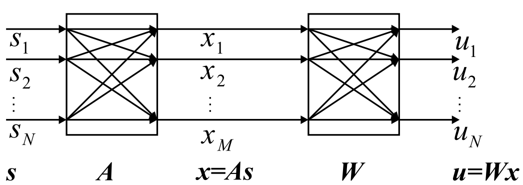

Figure 1: ICA model: The independent sources $s_1 \dots s_N$ are linearly mixed by the mixing matrix $\mathbf{A}$ and unmixed by the separation matrix $\mathbf{W}$. The set of output signals $u_1 \dots u_N$ is ideally a scaled and permuted version of the original source signals. The signals are assumed to have zero mean.

The model of the blind source separation problem is depicted in figure 1. Statistically independent sources $\mathbf{s}$ are mixed in a mixing system $\mathbf{A}$. The separation system $\mathbf{W}$ is computed using the criterion of statistical independence between the output signals $\mathbf{u}$. These outputs are the estimates of the original sources $\mathbf{s}$.

In the instantaneous mixing case the mixing and the unmixing system are matrices with scalar elements. It is then possible to formulate the blind source separation problem as

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{s} \qquad \text{and} \quad \mathbf{u} = \mathbf{W} \cdot \mathbf{x}. \qquad (1)$$

Ideally, the separation system is the inverse of the mixing system, that is $\mathbf{W} = \mathbf{A}^{-1}$. However, by only using the criterion of statistical independence it is impossible to recover the correct scaling and the original order of the sources [1, 2].

In general, there are two major strategies for the development of ICA algorithms: using higher order statistics (HOS), e.g. [1, 2, 3, 4, 5], or using the spectral differences between the different source signals, e.g [6, 7, 8, 9]. Both strategies are complementary in some sense. They differ in their pre-assumptions. The HOS approach assumes non-gaussian probability density functions. The spectral approach assumes non-white source signals with different spectral characteristics. For many natural signals the application of both strategies is possible.

The proposed algorithm follows the second strategy, i.e. the spectral differences between the source signals are exploited. Possible ways for the spectral approach are described in [6, 7, 8, 9]. In general these algorithms pick up common signal parts by computing the cross-correlation at several time delays or perform an equivalent procedure in a transform domain. A cost function depending on these cross-correlations is then minimized. However, these algorithms suffer from certain disadvantages, such as no signal adaptivity [6, 7], strong pre-assumptions [8] or the use of potentially unstable filters.

In this paper we propose a new separation method for non-white signals based on the spectral differences between the sources and non-recursive filtering. The proposed method overcomes the aforementioned disadvantages of [6, 7, 8, 9]. The low computational load and the possible parallel processing of the signals lead to a fast separation and make real-time applications feasible.

We have organized our paper as follows. In section two we introduce our new method. In section three simulation results are shown. They are discussed in section four and conclusions are drawn in section five.

2. METHODS

2.1. Data Pre-Processing by Sphering

Sphering is a method for the removal of second order cross-correlation between signals. Time delays are left unconsidered. Sphering is accomplished by

$$\mathbf{x}_s = \mathbf{M}\mathbf{x} = E[\mathbf{x}\mathbf{x}^T]^{-1/2}\mathbf{x}, \tag{2}$$

where $\mathbf{M}$ is the sphering matrix and $E[\cdot]$ denotes the expected value. In [1] it is shown that the mixing system $\mathbf{A}$ is factorizable into the inverse of the sphering matrix $\mathbf{M}^{-1}$ and an orthogonal rotation matrix $\mathbf{O}$, i.e.

$$\mathbf{A} = \mathbf{M}^{-1} \cdot \mathbf{O}. \tag{3}$$

Having identified the sphering matrix $\mathbf{M}$ the rotation matrix $\mathbf{O}$ must be estimated. To accomplish this the sphered signals are linearly transformed. This transformation is needed to create new equations and hence new conditions for the computation of the rotation matrix $\mathbf{O}$. The linear transformation is a generalization of simple time delays as described in [6].

2.2. The Application of a Linear Transform for Blind Source Separation

Mathematically, statistical independence of the sources $\mathbf{s} = [s_1, s_2, \ldots, s_N]$ is equivalent to the separability of the multivariate probability density function of the sources into the product of the marginal densities

$$p_{\mathbf{s}}(\mathbf{s}) = \prod_{i=1}^{N} p_{s_i}(s_i). \tag{4}$$

As shown in [9, 10], a linear transform defined by

$$\mathbf{T}(\mathbf{s}) = [T_1(s_1), T_2(s_2), \ldots, T_N(s_N)]^T \tag{5}$$

leaves the property of statistical independence unaffected. This is easy to prove since the Jacobian $\mathbf{J}$ of the transform $\mathbf{T}(\mathbf{s})$ is a diagonal matrix. Hence, the transformed probability density function is still separable, i.e.

$$p_{\mathbf{z}}(\mathbf{z}) = \frac{p_{\mathbf{s}}(\mathbf{s})}{|\mathbf{J}|} = \prod_{i=1}^{N} \frac{p_{s_i}(s_i)}{|dT_i(s_i)/ds_i|} = \prod_{i=1}^{N} p_{z_i}(z_i), \tag{6}$$

where $\mathbf{z} = \mathbf{T}(\mathbf{s})$ and $z_i = T_i(s_i)$. In contrast to the impact of linear transforms on the statistical independence between the signals, linear transforms influence the spectral composition of the signals.

If the linear transform is defined as $T_1 = T_2 = \ldots = T_N = T$, the transformation of the sphered signals $\mathbf{x}_s$ can be written as

$$\mathbf{T}(\mathbf{x}_s) = \mathbf{T}(\mathbf{O} \cdot \mathbf{s}) = \mathbf{O} \cdot \mathbf{T}(\mathbf{s}). \tag{7}$$

Computing the covariance matrix of the transformed and sphered data, one obtains

$$\begin{aligned} E[\mathbf{T}(\mathbf{x}_s)\mathbf{T}(\mathbf{x}_s^T)] &= E[\mathbf{T}(\mathbf{O}\mathbf{s})\mathbf{T}(\mathbf{O}\mathbf{s})^T] \\ &= \mathbf{O}E[\mathbf{T}(\mathbf{s})\mathbf{T}(\mathbf{s}^T)]\mathbf{O}^T \\ &= \mathbf{O}\mathbf{D}_T\mathbf{O}^T. \end{aligned} \tag{8}$$

According to equation 6 the covariance matrix of the transformed source signals $\mathbf{z}$ is diagonal, i.e.

$$E[\mathbf{z}\mathbf{z}^T] = E[\mathbf{T}(\mathbf{s})\mathbf{T}(\mathbf{s})^T] = \mathbf{D}_{\mathbf{T}}, \tag{9}$$

where

$$\mathbf{D}_T = E[diag(T(s_1)^2, T(s_2)^2, \ldots, T(s_N)^2)]. \tag{10}$$

Hence, equation 8 describes an eigenvalue problem for the matrix $E[\mathbf{T}(\mathbf{x}_s)\mathbf{T}(\mathbf{x}_s^T)]$ where the eigenvectors form the orthogonal rotation matrix $\mathbf{O}$.

2.3. Blind Source Separation Using Weighted Time Delays

In order to obtain new equations for the computation of the rotation matrix $\mathbf{O}$ the linear transform must be chosen such that $E[\mathbf{T}(\mathbf{x}_s)\mathbf{T}(\mathbf{x}_s^T)]$ in equation 8 is non-diagonal.

After the pre-processing by sphering the cross-correlation between the signals is zero. That is, dependencies between the signals are not visible to the cross-correlation coefficient. Hence, it is reasonable to design linear transforms that reveal the remaining, i.e. hidden dependencies between the signals in such a way that the cross-correlation coefficient is non-zero after the transformation. Considering that only second order statistics is used the transforms have to make use of the spectral properties of the signals which is equivalent to processing the correlation function at several delays. To achieve this objective the algorithms [6, 7] use different time delays or wavelet scales as linear transforms to solve this problem.These methods are not really adapted to the actual signals under consideration. For that reason they are not suited to reveal an optimal amount of hidden dependencies between the signals. A step forward is the matched filter algorithm described [8]. The disadvantage of the matched filter approach is that signal patterns must be known *a-priori.* To make full use of the spectral content of the signals it is reasonable to employ spectral estimation methods. Such an approach is described in [9] where the spectrum is estimated by autoregressive models and, using the spectral estimate, the signals are transformed by recursive filtering. The disadvantage of this approach is that these filters can be potentially unstable, in particular if the coefficients are estimated using higher order statistics or in practical situations where, due to finite word length effects, the poles may drift to locations outside the unit circle of the z-plane. It is therefore highly desirable to use filters with a finite impulse response (FIR) that extract the spectral information that is necessary for the computation of the rotation matrix $\mathbf{O}$.

The algorithm proposed in this paper is designed to overcome the disadvantages of the older algorithms. It is explained for two signals but can easily be expanded to the general case of N signals.

As linear transform the algorithm uses the sum of weighted time delayed signal values. This approach is equivalent to the application of a non-recursive (FIR) filter on the signals. The objective is to reveal common components between the two sphered signals, i.e. to get a non-zero crosscorrelation.

For the two-dimensional case the transformed signals may be written as

$$\begin{aligned} T[x_{s1}] &= T[o_{11}s_1] + [o_{12}s_2] \qquad \text{and} &(11)\\ T[x_{s2}] &= T[x_{s1}] + \\ &+ T[(o_{21} - o_{11}) + s_1(o_{22} - o_{12})s_2], &(12) \end{aligned}$$

where the o_{ij} are the coefficients from the rotation matrix $\mathbf{O}$. After a transformation the cross-correlation coefficient will be at maximum if it is possible to find a FIR filter (= linear transform) that maximizes $T[x_{s1}]$ in equation 12 and minimizes the error term $T[(o_{21} - o_{11})s_1 + (o_{22} - o_{12})s_2]$ in equation 12. In case of more than $N = 2$ signals the error term varies between the different signals. Hence, it is reasonable to find a FIR filter that maximizes the term $T[x_{s1}]$ in equation 12. This can be accomplished by a filter that maximizes the power of $T[x_{s1}]$ under the constrained that the filter coefficients are normalized.

The filtering operation can be parametrized by

$$y_i(n) = \sum_{k=0}^{P} a_k \cdot x_{si}(n-k) = a(n) * x_{si}(n), \qquad (13)$$

where the a_k are the coefficients of the FIR filter and x_{si} is one of the sphered signals. Writing equation 13 in matrix/vector notation one obtains

$$\mathbf{y} = \mathbf{X}\mathbf{a}, \qquad (14)$$

where matrix $\mathbf{X}$ contains the time delayed versions of the used sphered signal x_{si}, i.e.

$$\mathbf{X} = \begin{pmatrix} x_{si}(0) & 0 & 0 & \cdots & 0 \\ x_{si}(1) & x_{si}(0) & 0 & \cdots & 0 \\ \ddots & \ddots & \ddots & \cdots & 0 \\ x_{si}(P) & x_{si}(P-1) & x_{si}(P-2) & \cdots & x_{si}(0) \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}, \qquad (15)$$

$\mathbf{a} = [a_0, a_1, a_2, \ldots, a_P]^T$ is the coefficient vector of the FIR filter and $\mathbf{y} = [y_i(1), y_i(2), \ldots, y_i(n), \ldots]^T$ is the output signal.

Maximizing the power of the output signal and normalizing the the filter coefficients to $||\mathbf{a}|| = 1$ leads to the optimization problem

$$K = \mathbf{y}^T\mathbf{y} + \lambda(\mathbf{a}^T\mathbf{a} - 1) \rightarrow \max. \qquad (16)$$

Expanding the output signal $\mathbf{y}$ in equation 16 one obtains

$$K = \mathbf{a}^T\mathbf{X}^T\mathbf{X}\mathbf{a} + \lambda(\mathbf{a}^T\mathbf{a} - 1) \rightarrow \max. \qquad (17)$$

The gradient with respect to the filter coefficients is given by

$$\nabla_{\mathbf{a}}K = 2\mathbf{X}^T\mathbf{X}\mathbf{a} + 2\lambda\mathbf{a}. \qquad (18)$$

Applying the necessary condition for a local maximum $\nabla_{\mathbf{a}}K = 0$ one obtains the eigenvalue problem

$$\mathbf{X}^T\mathbf{X}\mathbf{a} = \lambda\mathbf{a}. \qquad (19)$$

Note that $\mathbf{X}^T\mathbf{X}$ is proportional to the correlation matrix of x_{si}. Disregarding a normalisation constant and using $||\mathbf{a}|| = \mathbf{a}^T\mathbf{a} = 1$ the power of the output is given by

$$\mathbf{y}^T\mathbf{y} = \mathbf{a}^T\mathbf{X}^T\mathbf{X}\mathbf{a} = \mathbf{a}^T\lambda\mathbf{a} = \lambda\mathbf{a}^T\mathbf{a} = \lambda. \qquad (20)$$

Hence, the eigenvector belonging to the greates eigenvalue is the wanted coefficient vector of the FIR filter. After filtering common components are amplified and components not present in both signals are (relatively) attenuated. Hidden dependencies between both signals are revealed. The rotation matrix $\mathbf{O}$ is eventually computed by solving the eigenvalue problem in equation 8.

In the general case of N signals the coefficients of the FIR filter are determined for one signal and afterwards all signals are filtered by the computed FIR filter. Corresponding to equation 8 the correlation is then measured by the covariance matrix. In order to use all available information this procedure is repeated for all N signals $x_{s1}, x_{s2}, \ldots, x_{sN}$, i.e. N FIR filters need to be determined and after filtering N covariance matrices need to be computed. In general these covariance matrices do not have a diagonal structure. The matrix $\mathbf{O}$ that diagonalizes all the matrices simultaneously is the wanted orthogonal rotation matrix of the blind source separation problem.

2.4. Joint Diagonalisation

The application of more than one linear transform, i.e. filter, gives more than one eigenvalue problem. A possible rotation matrix $\mathbf{O}$ must solve all these eigenvalue problems at the same time.

There are several ways to jointly solve these eigenvue problems. One possible way is to combine the set of different eigenvalue problems into only one eigenvalue problem. For two different linear transformations, i.e. FIR filters $\mathbf{T}$ and $\tilde{\mathbf{T}}$ one obtains according to equation 8 the eigenvalue problems

$$\mathbf{C} = E[\mathbf{T}(\mathbf{x}_s)\mathbf{T}(\mathbf{x}_s^T)] = \mathbf{O}\mathbf{D}_T\mathbf{O}^T \quad (21)$$

$$\tilde{\mathbf{C}} = E[\tilde{\mathbf{T}}(\mathbf{x}_s)\tilde{\mathbf{T}}(\mathbf{x}_s^T)] = \mathbf{O}\mathbf{D}_{\tilde{T}}\mathbf{O}^T. \quad (22)$$

These two equations can be combined into one joint eigenvalue problem by

$$\mathbf{C}\tilde{\mathbf{C}} = \mathbf{O}\mathbf{D}_T\mathbf{D}_{\tilde{T}}\mathbf{O}^T. \quad (23)$$

The extension of equation 23 for more than two linear transformations is simple. However, this method accumulates estimation errors and heavily increases the condition number of the matrix product. In most cases it provides only poor results.

Much more satisfying results can be achieved by using the joint approximate diagonalisation algorithm proposed in [5]. For that reason we use this latter diagonalisation method for our algorithm.

3. SIMULATION RESULTS

The general performance of blind source separation algorithms using second order statistics was investigated in [11]. In [11] it has been shown that the performance of such algorithms heavily depends on the spectral overlap of the source signals. In contrast to the spectral dependence the performance does not depend on the shape of the probability density functions of the sources. We have made extensive simulations on artificially created signals and artificial mixtures that confirmed these theoretical results [10]. Applications to real EEG data are also shown in [10]. An example of the performance of the algorithm is shown in figure 2. Due to the simple structure of the algorithm,

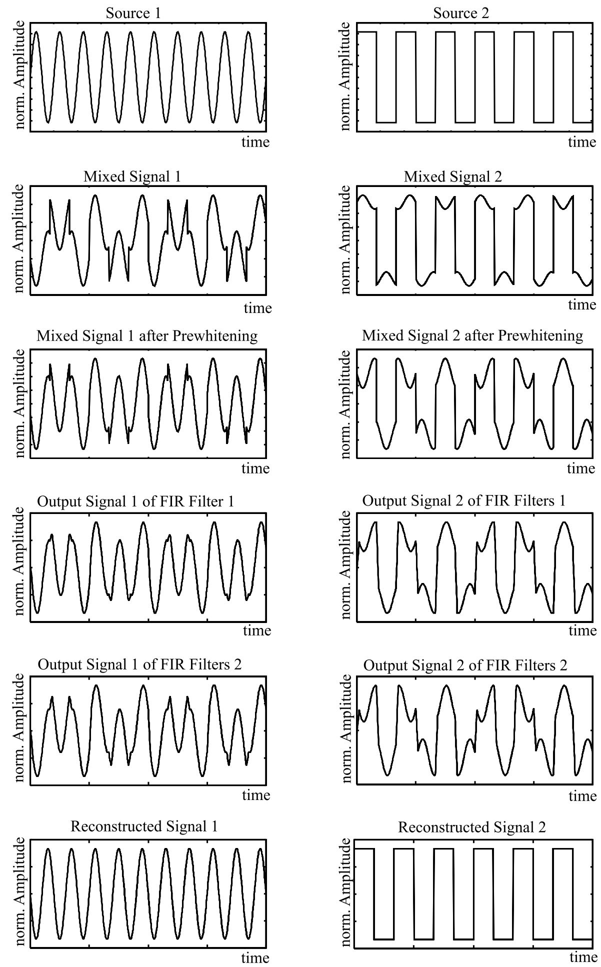

Figure 2: Signals at the several stages of the algorithm

the computation time is very low compared to algorithms using higher order statistics such as [2, 5, 3] and slightly above the simple second order algorithms such as [6, 7]. Similar to other algorithms, the performance of the proposed method gracefully degrades if noise is present.

4. DISCUSSION

The proposed algorithm can not overcome the limitations of using second order statistics as described in

[11], i.e. the necessary spectral differences between the source signals. However, it makes use of the spectral differences in an optimal way. This leads to a performance that is comparable to established algorithms, e.g. [2, 5, 3]. In contrast to the algorithms [6, 7] it is signal adaptive. Furthermore, the number of matrices to be diagonalized is limited to the number of signals. In contrast to [6] the matrices are always symmetric, i.e. real valued solutions are guaranteed. The advantage of the proposed algorithm over the autoregressive model algorithm [9] is that the filter is now non-recursive and hence always stable.

5. CONCLUSIONS

A new method for blind source separation of non-white signals was proposed. It was shown that the concept of the linear transformation can be implemented using a non-recursive filter, i.e. a summation of weighted time delayed signal values. It is possible to compute an optimal set of filter coefficients (weights), which makes the algorithm, in contrast to former methods, signal adaptive and always stable. The performance of the algorithm was shown by an example. Due to the relatively low computational load the algorithm can potentially be applied in real-time applications.

In this paper the proposed method is described as a batch algorithm. However, the idea of maximizing the power under the constraint of normalized coefficients can probably be used in recursive algorithms which potentially can be implemented as online algorithms. This will be a topic of our future work.

6. REFERENCES

[1] J.F. Cardoso: Source Separation Using Higher-Order Moments. Proc. ICASSP '89, pp. 2109-2112, Glasgow, 1989

[2] A.J. Bell and T.J. Sejnowski: An Information-Maximization Approach to Blind Separation and Blind Deconvolution. Neural Computation, 7:1129-1159, July 1995

[3] A. Hyvärinen and E. Oja: A Fast Fixed-Point Algorithm for Independent Component Analysis. Neural Computation 9, pp. 1483-1492, 1997.

[4] S. Amari and A. Cichocki and H. Yang: A New Learning Algorithm for Blind Signal Separation. Advances in Neural Information Processing Systems, 8,1996

[5] J.F. Cardoso and A. Souloumiac: Blind Beamforming for Non Gaussian Signals. IEE-Procedings-F., Vol. 140, No. 6, pp. 362-370, December 1993

[6] L. Molgedey and H.G. Schuster: Separation of a Mixture of Independent Signals Using Time Delayed Correlations. Physical Review Letter, 72(23), June 1994

[7] B.-U. Köhler and R. Orglmeister: Independent Component Analysis of Electroencephalographic Data Using Wavelet Decomposition. Proc. VIII Mediterranian Conference on Medical and Biological Engineering and Computing, Lemesos, Cyprus, 1998

[8] B.-U. Koehler and R. Orglmeister and B. Brehmeier-Flick: Independent Component Analysis of Electroencephalographic Using Data Matched Filters. EMBS98, Hongkong, October/November 1998.

[9] B.-U. Koehler and R. Orglmeister: Independent Component Analysis Using Autoregressive Models. Proc. First International workshop on Independent Component Analysis and Blind Separation of Signals - ICA'99. , pp. 359-363; Jan.1999. Aussois, France.

[10] B.-U. Koehler: Realzeitfähige Blinde Quellentrennung am Beispiel Elektroenzephalographischer Signale. PhD Thesis. Berlin University of Technology, 1999. to be published.

[11] A. Belouchrani and K.A. Meraim and J.F. Cardoso: A blind source separation technique based on second order statistics. IEEE Transactions of Signal Processing, 45(2), pp. 434-444)