BLIND SIGNAL SEPARATION USING ORIENTED PCA NEURAL MODELS

K. I. Diamantaras

Department of Informatics
TEI of Thessaloniki
Sindos 54101, Greece

Th. Papadimitriou

Department of Int. Economic Relat. and Devel.,
Democritus University of Thrace
Komotini 69100, Greece

ABSTRACT

Oriented PCA (OPCA) is a (second order) extension of standard Principal Component Analysis aiming at maximizing the power ratio of a pair of signals. It is shown that OPCA, preceded by almost arbitrary temporal filtering, can be used for blindly separating temporally colored signals from their linear instantaneous mixtures. The advantage over other second order techniques is the lack of the prewhitening (or sphereing) step. Although the design of the general optimal temporal pre-filter is an open problem, we show that the filters $[1, \pm 1]$ are the optimal ones for the special two-tap case. Neural OPCA models proposed earlier are used in simulations to separate a number of artificial sources demonstrating the validity of the method.

1. INTRODUCTION

The estimation of a set of source signals, given a set of linear mixtures has been receiving great attention in the last decade, due to its usefulness in a large number of applications. As no a priori information about the source signals and the mixture structure is needed, the task is referred as *blind source separation* (BSS). We are particularly interested in separating memoryless mixtures ie. for a special case referred to as instantaneous BSS. Methods for this problem can be divided into methods using second-order [1] or higher-order statistics [2], maximum-likelihood principle [3], Kullback-Liebler distance [4, 5, 6] PCA methods [7, 8], non-linear PCA [9], ICA methods [10]. Further information on these methods and a coherent treatment of BSS, in general, can be found in [11].

In [12] Diamantaras demonstrated that when the observed data are temporally prefiltered, standard PCA can be used for the solution of instantaneous mixture BSS. The method needed a step of spatial prewhitening (sphereing) over the observation data. In this paper, we show that without pre-whitening the problem is a typical Oriented PCA problem. The problem can be solved using standard neural networks models poposed in earlier works. Moreover, we are able to derive the optimal length-2 prefilter, although the design of the optimal prefilter, in general, is an open issue.

2. THE BLIND SIGNAL SEPARATION PROBLEM

In the instantaneous BSS problem the n observed signals x_1, ..., x_n, result from the linear combination of the sources s_1, ..., s_n, so defining $\mathbf{x}(k) = [x_1(k), \cdots, x_n(k)]^T$ and $\mathbf{s}(k) = [s_1(k), \cdots, s_n(k)]^T$ we have:

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k) \qquad (1)$$

We assume that the linear mixing operator $\mathbf{A}$ is a square, invertible matrix although, in general, the number of observations may be greater than the number of sources, in which case $\mathbf{A}$ is a "tall" matrix with full column rank.

Both the mixing operator $\mathbf{A}$ as well as the source sequence $\mathbf{s}(k)$ are assumed to be unknown. As a result, the order and the scale of the individual sources are unobservable. Additionally we adopt certain assumptions regarding the second order statistics of the sources as stated below (see also [1, 13]):

A1. Sources are pairwise uncorrelated, at least wide sense stationary with zero mean and unit variance:

$$\mathbf{R}_s(0) = \mathbf{I}\ . \qquad (2)$$

A2. There exist positive time lags l_1, ..., l_M such that:

$$\mathbf{R}_s(l_m) \triangleq \text{diagonal} \neq 0\ . \qquad (3)$$

Define $l_0 = 0$.

A3. Distinct source colors:

$$\forall l \neq 0 :\ \ r_{ii}(l) \neq r_{jj}(l),\ \text{if}\ i \neq j\ .$$

3. SOLVING BSS USING ORIENTED PCA (OPCA)

The term Oriented PCA (OPCA) [14] describes an extension of PCA involving two signals $\mathbf{u}(k)$, and $\mathbf{v}(k)$. The aim is to identify the so-called oriented principal directions $\mathbf{e}_1$, ..., $\mathbf{e}_n$, that maximize the signal-to-signal power ratio $E(\mathbf{e}_i^T\mathbf{u})^2/E(\mathbf{e}_i^T\mathbf{v})^2$ under the orthogonality constraint: $\mathbf{e}_i^T\mathbf{R}_u\mathbf{e}_j = \mathbf{e}_i^T\mathbf{R}_u\mathbf{e}_j = 0,\ i \neq j$. OPCA is a second-order statistics method, which reduces to standard PCA if the second signal is spatially white $\mathbf{R}_v = \mathbf{I}$. The solution of OPCA is shown to be the *Generalized Eigenvalue Decomposition* (GED) of the matrix pencil $[\mathbf{R}_u, \mathbf{R}_v]$.

Subsequently, we shall relate the instantaneous BSS problem with the OPC Analysis of the observed signal $\mathbf{x}$ and almost any filtered version of it. We shall use the following notation $\mathbf{R}_z(l) \triangleq E\{\mathbf{z}(k)\mathbf{z}(k-l)^T\}$ to describe the covariance of any signal $\mathbf{z}$ with time-lag l. Note that the 0-lag covariance matrix of $\mathbf{x}(k)$ is

$$\mathbf{R}_x(0) = \mathbf{A}\mathbf{R}_s(0)\mathbf{A}^T = \mathbf{A}\mathbf{A}^T \qquad (4)$$

Now, consider a scalar, linear temporal filter $\mathbf{h} = [h_0, \cdots, h_M]$ operating on $\mathbf{x}(k)$:

$$\mathbf{y}(k) = \sum_{m=0}^{M} h_m \mathbf{x}(k - l_m) \qquad (5)$$

where the lags l_0, ..., l_M, satisfy assumptions [A2], [A3]. The 0-lag covariance matrix of $\mathbf{y}$ is expressed as

$$\mathbf{R}_y(0) = E\{\mathbf{y}(k)\mathbf{y}(k)^T\} = \sum_{p,q=0}^{M} h_p h_q \mathbf{R}_x(l_p - l_q) \tag{6}$$

From Eq. (1) it follows that

$$\mathbf{R}_x(l_m) = \mathbf{A}\,\mathbf{R}_s(l_m)\,\mathbf{A}^T \tag{7}$$

so

$$\mathbf{R}_y(0) \quad = \quad \mathbf{A}\mathbf{D}\mathbf{A}^T \tag{8}$$

with

$$\mathbf{D} = \sum_{p,q=0}^{M} h_p h_q \mathbf{R}_s(l_p - l_q) \tag{9}$$

Note that, by assumptions [A1], [A2], $\mathbf{D}$ is diagonal. Provided that $\mathbf{A}$ is square and invertible we can write

$$\begin{aligned}\mathbf{R}_y(0)\mathbf{A}^{-T} &= \mathbf{A}\mathbf{D} \\ &= \mathbf{A}\mathbf{A}^T\mathbf{A}^{-T}\mathbf{D} \\ &= \mathbf{R}_x(0)\mathbf{A}^{-T}\mathbf{D}\end{aligned} \tag{10}$$

where $\mathbf{A}^{-T} \triangleq \mathbf{A}^{-1T}$. Eq. (10) expresses a Generalized Eigenvalue Decomposition problem for the matrix pencil $[\mathbf{R}_y(0),\ \mathbf{R}_x(0)]$. This is equivalent to the OPCA problem for the pair of signals $[\mathbf{y}(k),\ \mathbf{x}(k)]$. The generalized eigenvalues for this problem are the diagonal elements of $\mathbf{D}$. The columns of the matrix $\mathbf{A}^{-T}$ are the generalized eigenvectors.

The eigenvectors are unique upto a permutation and scale provided that the eignevalues are distinct (this is true in general). In this case, for any generalized eigenmatrix $\mathbf{Q}$ we have $\mathbf{Q} = \mathbf{A}^{-T}\mathbf{P}$ with $\mathbf{P}$ being a scaled permutation matrix, ie. each row and each column contains exactly one on-zero element. Then the sources can be estimated as

$$\hat{\mathbf{s}}(k) = \mathbf{Q}^T\mathbf{x}(k) \tag{11}$$

$$\hat{\mathbf{s}}(k) = \mathbf{P}^T\mathbf{A}^{-1}\mathbf{A}\mathbf{s}(k) = \mathbf{P}^T\mathbf{s}(k) \tag{12}$$

It follows that the estimated sources are equal to the true ones except for the (unobservable) arbitrary order and scale.

A similar approach using standard PCA and spatial prewhitening (sphering) has been proposed in [13]. The advantage of the OPCA method is that no particular preprocessing is needed and so this approach is simpler and more neural-like.

3.1. Designing the optimal filter with a single time lag

Let [A2] hold for just one time lag l, so we may use a two-tap filter $\mathbf{h} = [h_0,\ h_1] = [1,\ \alpha]$, where $h_1 = \alpha$ is a free parameter. Then

$$\begin{aligned}\mathbf{D} &= (1+\alpha^2)\mathbf{I} + \alpha\mathbf{R}_s(l) + \alpha\mathbf{R}_s(-l) \\ &= (1+\alpha^2)\mathbf{I} + 2\alpha\mathbf{R}_s(l)\end{aligned} \tag{13}$$

Denoting by d_i and $r_{ii}(l)$ the diagonal elements of $\mathbf{D}$ and $\mathbf{R}_s(l)$ respectively, we obtain

$$d_i = 1 + \alpha^2 + 2\alpha r_{ii}(l),\ \ i = 1, \cdots, n \tag{14}$$

Using (14) we can compute the correlation matrix of the input signal $\mathbf{R}_s(l)$:

$$r_{ii}(l) = \frac{d_i - 1 - \alpha^2}{2\alpha} \tag{15}$$

Once the correlation is obtained we can use it in order to design the optimal temporal filter $\mathbf{h}$. The optimality criterion will be related to the eigenvalue spread. It is desirable to spread the eigenvalues as much as possible for two reasons: (a) the convergence of any batch or neural generalized eigenvalue algorithm is typically faster when the eigenvalues are well separated, and (b) the perturbation of the eigenvalues due to noise can be better tolerated. Thus we need to define a suitable metric taking into account the relative size of the eigenvalues. We propose to use the following maximization criterion

$$J(\alpha) = \min_i \Big[\min_{j\neq i} \frac{(d_i - d_j)^2}{\max_k d_k^2}\Big]. \tag{16}$$

Using (14) this metric can be formulated in terms of the input correlation function $\mathbf{R}_s(l)$

$$J(\alpha) = \min_i \Big[\min_{j\neq i} \frac{4\alpha^2(r_{ii}(l) - r_{jj}(l))^2}{\max_k(1+\alpha^2+2\alpha r_{kk}(l))^2}\Big]. \tag{17}$$

Let $4\alpha^2\Delta r^2 = \min_i[\min_{j\neq i} 4\alpha^2(r_{ii}(l) - r_{jj}(l))^2]$ and let $r_{kk}(l) = r_{max}$ be the maximizer of the denominator $(1+\alpha^2+2\alpha r_{kk}(l))^2$. Then we can write

$$J(\alpha) = \frac{4\alpha^2\Delta r^2}{(1+\alpha^2+2\alpha r_{max})^2}\ .$$

The most robust filter is the one that maximizes $J(\alpha)$. Note that $J(\alpha) \geq 0$ and $\lim_{\alpha\to\pm\infty} J(\alpha) = 0$. Furthermore, J is bounded since $\max_k d_k^2 > 0$ and $|d_i| < \infty$ for all i. It follows that $J(\alpha)$ has at least one maximum, which is attained at a gradient zero-crossing:

$$\begin{aligned}\frac{\partial J}{\partial \alpha} &= 4\Delta r^2\Big[\frac{2\alpha}{(1+\alpha^2+2\alpha r_{max})^2} \\ &\quad -2\frac{\alpha^2(2\alpha + 2r_{max} - 2r'\alpha)}{(1+\alpha^2+2\alpha r_{max})^3}\Big] \\ &= 8\Delta r^2 \frac{\alpha(1-\alpha^2-2r'\alpha)}{(1+\alpha^2+2\alpha r_{max})^3} = 0\end{aligned} \tag{18}$$

where $r' = \partial r_{max}/\partial\alpha$. Since $J(0) = 0$, the solution $\alpha = 0$ to Eq. (18) does not correspond to a maximum. Furthermore, r_{max} takes values in the discrete set $\{r_{11}(l), \cdots, r_{nn}(l)\}$ therefore, it is not a continuous function of α and $r' = 0$ except for those points where a discontinuity appears. Assuming that $J(\alpha)$ is not maximized at such a discontinuity point, its maximum value must be attained for $(1-\alpha^2) = 0$, ie. for $\alpha = +1$ or -1.

4. OPCA NEURAL NETWORKS FOR BSS

Three neural models for OPCA have been proposed by Diamantaras and Kung [14]. Here we shall use the third model originally proposed in [15]. For a pair of signals $\mathbf{u}(k)$, $\mathbf{v}(k)$ a linear neuron model can extract the principal oriented eigenvector (i.e. the one associated with the largest eigenvalue). The learning rule is described by the following equations:

$$\Delta \mathbf{w}(k) = \beta[\mathbf{u}(k)a(k) - \xi(k)\mathbf{v}(k)b(k)] \quad (19)$$

$$\Delta \xi(k) = \beta'(\|\mathbf{w}(k)\|^2 - \xi(k)) \quad (20)$$

where β is a small positive learning rate parameter and β' is smaller than β (e.g. $\beta' = \beta/3$). The values $a(k)$ and $b(k)$ denote the neuron output activations when the inputs are $\mathbf{u}(k)$ and $\mathbf{v}(k)$, respectively:

$$a(k) = \mathbf{w}(k)^T \mathbf{u}(k)$$

$$b(k) = \mathbf{w}(k)^T \mathbf{v}(k)$$

Once the model has successfully extracted a vector $\mathbf{w}$ parallel to $\mathbf{e}_1$ the remaining components are obtained by employing successive deflation tranformations (see [14][chapter 7] for details).

5. RESULTS

In the experiments described below, we chose the filter $h = [1, -1]$ with a single lag $l = 1$. The estimation convergence for a set of 4 signals are shown in Figure 1.

Every sub-figure corresponds to a row of the estimated absolute matrix $\mathbf{WA}$. Once the iterations stop, all the row elements are normalized by dividing with the maximum row element. Therefore their value varies from 0 to 1. Moreover, the same technique is applied in every estimated $\mathbf{W_i}$ during the iterations, using always the element corresponding to the final maximum, as denominator.

The same experiment was conducted using the filter $h = [1, -3]$ with a single lag. Results are shown in 2.

It can be witnessed that in the case that the filter $h = [1, -1]$ was used, the estimation is of higher quality, while the convergence is slow. On the other hand the convergence using the filter $h = [1, -3]$, is faster, while the estimation accuracy is poor.

6. CONCLUSIONS

The instantaneous BSS problem is known to be related to second-order statistics methods. However, all earlier approaches have consistently used two steps: one preprocessing (sphering) step followed by a second-order analysis method such as SVD [1] or PCA [13]. The OPCA approach proposed in this paper has the advantage that no preprocessing step is required as sphering is implicitly incorporated in the signal-to-signal ratio criterion which is optimized by OPCA. Furthermore, the method can be implemented using a simple neural approach, making it more intuitively appealing compared to other neural approaches.

7. REFERENCES

[1] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A Blind Source Separation Technique Using Second-Order Statistics," *IEEE Trans. Signal Processing*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[2] J.-F. Cardoso, "Source separation using higher order moments," in *in Proc. IEEE ICASSP,*, Glasgow, U.K., 1989, vol. 4, pp. 2109–2112.

[3] J. Basak and S. Amari, "Blind separation of uniformly distributed signals: A general approach," *IEEE Trans. Neural Networks*, vol. 10, pp. 1173–1185, sep 1999.

[4] D.T. Pham, "Blind separation of instantaneous mixture of sources via an independent component analysis," *IEEE Trans. Signal Processing*, vol. 44, pp. 2768–2779, nov 1996.

[5] S. Van Gerven, D. Van Compernolle, L. Nguyen, and C. Jutten, "Blind separation of sources: A comparative study of 2nd and 4th order solution," in *in Signal Processing VII: Theories Appl. Proc. EUSIPCO*, Edinburgh, U.K., sep 1994, pp. 1153–1156.

[6] H. L. Nguyen Thi and C. Jutten, "Blind source separation for convolutive mixtures," *Signal Processing*, vol. 45, pp. 209–229, 1995.

[7] K.I. Diamantaras, "Asymmetric PCA Neural Networks for Adaptive Blind Source Separation," in *Proc. IEEE Workshop on Neural Networks for Signal Processing (NNSP'98)*, Cambridge, UK, 1998, pp. 103–112.

[8] K.I. Diamantaras, "Second Order Hebbian Neural Networks and Blind Source Separation," in *Proc. EUSIPCO'98 (European Signal Processing Conference)*, Rhodes, Greece, 1998, pp. 1317–1320.

[9] J. Karhunen and J. Joutsensalo, "Representation and separation of signals using nonlinear PCA type learning," *Neural Networks*, vol. 7, pp. 113–127, 1994.

[10] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, pp. 287–314, 1994.

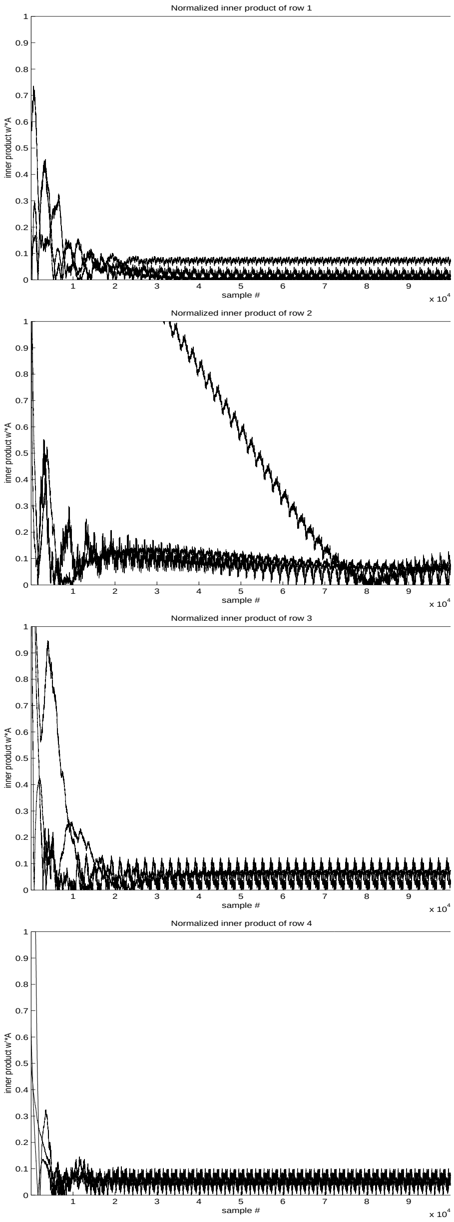
[11] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley, NY, 2001.

[12] K. Diamantaras, "Pca neural models and blind signal separation," in *In proc. International Joint Conference on Neural Networks (IJCNN-2001)*, Washington DC, July 2001.

[13] K.I. Diamantaras, "Neural Networks and Principal Component Analysis," in *Handbook of Neural Network Signal Processing*, Y. H. Hu and J. N. Hwang, Eds., chapter 8. CRC Press, N.Y., 2002.

[14] K.I. Diamantaras and S.Y. Kung, *Principal Component Neural Networks: Theory and Applications*, John Wiley, NY, 1996.

[15] K.I. Diamantaras, *Principal Component Learning Networks and Applications*, Ph.D. thesis, Princeton Univ., 1992.



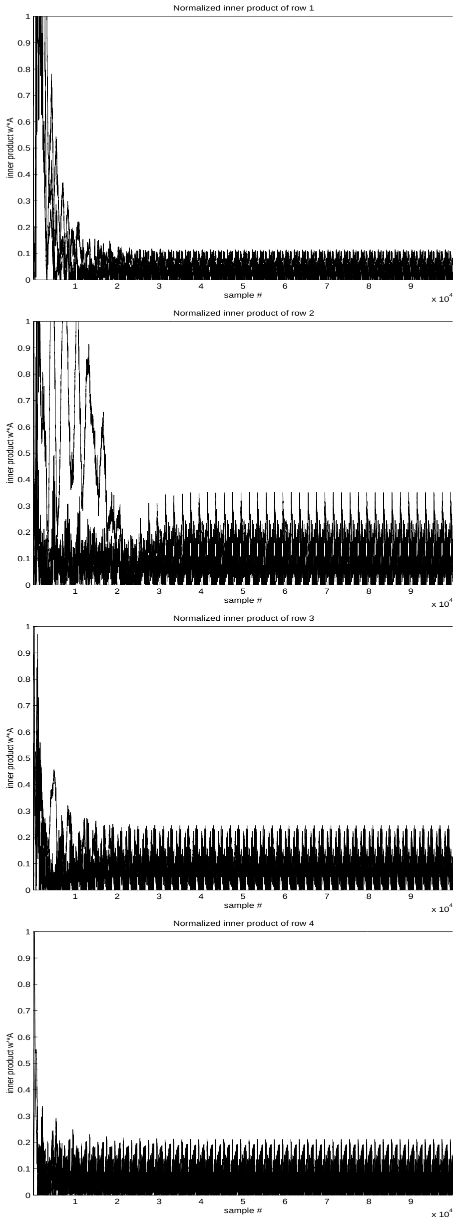


Fig. 1. Dynamic behavior of the four neurons in the sequential OPCA model. The plot depicts the normalized inner product of each neuron with the desired solution for 50 sweeps, using the filter $h = [1, -1]$. Each sweep contains $N = 2000$ data points.

Fig. 2. Dynamic behavior of the four neurons in the sequential OPCA model. The plot depicts the normalized inner product of each neuron with the desired solution for 50 sweeps, using the filter $h = [1, -3]$. Each sweep contains $N = 2000$ data points.