2003 IEEE XIII Workshop on
Neural Networks for Signal Processing

TEMPORAL FILTERING AND ORIENTED PCA NEURAL NETWORKS FOR BLIND SOURCE SEPARATION

K. I. Diamantaras
Department of Informatics
TEI of Thessaloniki
Sindos 54101, Greece
kdiamant@it.teithe.gr

Th. Papadimitriou
Dept. of Intern. Economic Relat. and Develop.
Democritus University of Thrace
Komotini 69100, Greece
papadimi@ierd.duth.gr

Abstract.

PCA-related (Principal Component Analysis) neural models have been shown to solve the instantaneous BSS (Blind Source Separation) problem for temporally colored sources. In this paper we show that arbitrary temporal filtering combined with models associated to the extension of standard PCA known as Oriented PCA (OPCA) provide a solution to the problem that is based on second order statistics and requires no prewhitening of the observation signals. Furthermore the issue of the optimal temporal filter is addressed for filters of length 2 and 3 although the design of the universally optimal filter is still an open question. Earlier neural OPCA networks are used to demonstrate the validity of the method on artificially generated datasets.

INTRODUCTION

The task of recovering m unknown signals solely from their linear mixtures observed at n sensors is known as Blind Source Separation (BSS). The problem has been studied extensively in the last decade due to its application in number of areas including biomedical signal processing, digital communications, speech processing, etc. The fact that the underlying mixing operator is unknown explains the use of the term "blind". Here we are interested in the special case of memoryless linear mixtures known as instantaneous BSS. Methods for this problem can be divided into methods using second-order [2, 7, 8, 18, 4] or higher-order statistics [3, 1, 16, 13, 17, 5, 15]. Further information on these methods and a coherent treatment of BSS, in general, can be found in [14].

Non-adaptive second order methods for BSS has been studied extensively. A well known example is the SOBI method introduced by Belouchrani e.a.

0-7803-8178-5/03/$17.00 © 2003 IEEE

[2]. SOBI is based on the joint diagonalization of an arbitrary number of covariance matrices for different time lags. Since there is no closed-form solution for joint diagonalization it is achieved by an iterative optimization procedure. Recently, Generalized Eigenvalue Decomposition on a pair of covariance matrices has been used by Chang e.a. [4]. The method however, is limited to the use of only two matrices. Standard PCA combined with temporal pre-filtering and spatial pre-whitening has been proposed in [9, 10]. This approach can be combined with neural PCA models which are well studied and have robust numerical behavior. Other neural second order approaches for BSS have also been proposed in the past. Hebbian SVD analyzers known as Asymmetric PCA models have been used in [7, 8]. Furthermore, Cross-Coupled Hebbian models [11], have been applied for the blind extraction of sources using a pre-selected time lag l.

In this paper, we show that cascading a temporal filter applied on the observation data and an Oriented PCA neural model solves the BSS problem without spatial pre-whitening (see Fig.1). Almost any filter is sufficient, except for the delayed impulse response. In general, the optimal filter that maximizes the eigenvalue spread is not known. Here we study the special cases of filters of length $L = 2$ and $L = 3$. A closed-form solution is provided for $L = 2$ and an iterative one for $L = 3$.

The instantaneous BSS problem can be described as follows. We consider n observed signals x_1, ..., x_n, resulting from the linear combination of m sources s_1, ..., s_m ($m \leq n$). Defining $\mathbf{x}(k) = [x_1(k), \cdots, x_n(k)]^T$ and $\mathbf{s}(k) = [s_1(k), \cdots, s_m(k)]^T$ we have:

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

In general, the number of observations may be greater than the number of sources, in which case the linear mixing operator $\mathbf{A}$ is a "tall" matrix with full column rank. However, for simplicity, in the rest of the paper we assume that $\mathbf{A}$ is a square, invertible matrix ($m = n$).

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 [2, 10]):

A1. Sources are pairwise uncorrelated, at least wide sense stationary with zero mean and unit variance. Defining the time-lagged covariance $\mathbf{R}_s(l) = E\{\mathbf{s}(k)\mathbf{s}(k-l)^T\}$ this assumption implies that the zero-lag covariance is the identity matrix:

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

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

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

Define $l_0 = 0$.

A3. The source covariance functions are distinct:

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

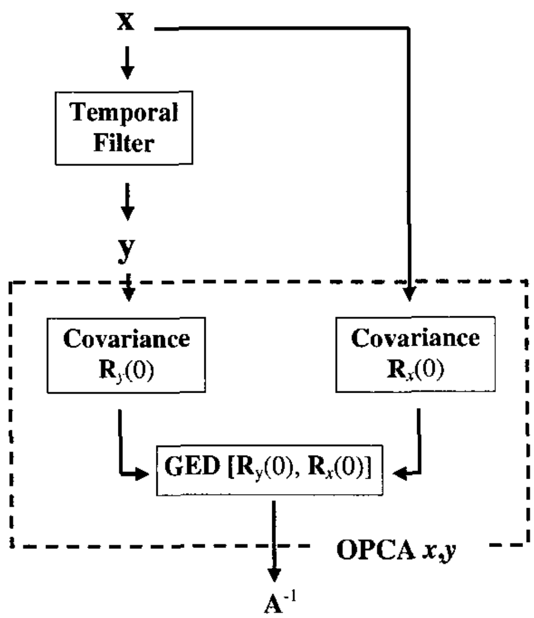

Figure 1: Graphical represenation of OPCA BSS.

TEMPORAL FILTERING AND OPCA SOLVES BSS

Oriented PCA (OPCA) [12] is an extension of PCA that involves two signals $\mathbf{u}(k)$ and $\mathbf{v}(k)$ and seeks for the optimal 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: $\mathbf{e}_i^T\mathbf{R}_u\mathbf{e}_j = \mathbf{e}_i^T\mathbf{R}_u\mathbf{e}_j = 0$, $i \neq j$. These directions are called Oriented Principal Eigenvectors.

Clearly OPCA is a second-order method. In fact it reduces to standard PCA if the second signal is spatially white, i.e. $\mathbf{R}_v = \mathbf{I}$. PCA is related to the Eigenvalue Decomposition of the mixtures covariance matrix. In an analogous way, OPCA is related to the Generalized Eigenvalue Decomposition (GED) of the matrix pencil $[\mathbf{R}_u, \mathbf{R}_v]$.

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

$$\mathbf{y}(k) = \sum_{m=0}^{M} h_m \mathbf{x}(k - l_m) \tag{4}$$

where l_0, ..., l_M, are the lags appearing in [A2], [A3].

The 0-lag covariance matrix of $\mathbf{y}$ is

$$\mathbf{R}_y(0) = E\{\mathbf{y}(k)\mathbf{y}(k)^T\} = \sum_{p,q=0}^{M} h_p h_q \mathbf{R}_x(l_p - l_q) \tag{5}$$

Note that the 0-lag covariance matrix of $\mathbf{x}(k)$ is

$$\mathbf{R}_x(0) = \mathbf{A}\mathbf{R}_s(0)\mathbf{A}^T = \mathbf{A}\mathbf{A}^T \tag{6}$$

From Eq. (1) it follows that

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

so

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

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

Note that, by [A1], [A2], all $\mathbf{R}_s$ matrices are diagonal and so $\mathbf{D}$ is diagonal as well. Using (8) and (6) and taking advantage of the fact that $\mathbf{A}$ is square and invertible we have

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

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.

Provided that the eignevalues are distinct (this is true in general),the eigenvectors are unique upto a permutation and scale. 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

$$\hat{\mathbf{s}}(k) = \mathbf{Q}^T\mathbf{x}(k) = \mathbf{P}^T\mathbf{A}^{-1}\mathbf{A}\mathbf{s}(k) = \mathbf{P}^T\mathbf{s}(k) \tag{11}$$

Thus, in the absence of noise, the estimated sources are equal to the true ones except for the (unobservable) arbitrary order and scale.

The optimal filter of length two

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

$$\begin{aligned} \mathbf{D} &= (1+\alpha^2)\mathbf{I} + \alpha\mathbf{R}_s(l) + \alpha\mathbf{R}_s(-l) \\ &= (1+\alpha^2)\mathbf{I} + 2\alpha\mathbf{R}_s(l) \end{aligned} \tag{12}$$

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

$$d_i = 1 + \alpha^2 + 2\alpha r_{ii}(l), \quad i = 1, \cdots, n \tag{13}$$

Using (13) we can compute the correlation matrix of the input signal $\mathbf{R}_s(l)$:

$$r_{ii}(l) = \frac{d_i - 1 - \alpha^2}{2\alpha} \tag{14}$$

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

$$J(\alpha) = \min_i \Big[\min_{j \neq i} \frac{(d_i - d_j)^2}{\max_k d_k^2}\Big]. \tag{15}$$

Using (13) this metric can be formulated in terms of the input correlation function $\mathbf{R}_s(l)$

$$J(\alpha) = \min_i \Big[\min_{j \neq i} \frac{4\alpha^2 (r_{ii}(l) - r_{jj}(l))^2}{\max_k (1 + \alpha^2 + 2\alpha r_{kk}(l))^2}\Big]. \tag{16}$$

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:

$$\begin{aligned} \frac{\partial J}{\partial \alpha} &= 4\Delta r^2 \Big[\frac{2\alpha}{(1 + \alpha^2 + 2\alpha r_{max})^2} \\ &\quad -2\frac{\alpha^2 (2\alpha + 2r_{max} - 2r'\alpha)}{(1 + \alpha^2 + 2\alpha r_{max})^3}\Big] \\ &= 8\Delta r^2 \frac{\alpha(1 - \alpha^2 - 2r'\alpha)}{(1 + \alpha^2 + 2\alpha r_{max})^3} = 0 \end{aligned} \tag{17}$$

where $r' = \partial r_{max}/\partial\alpha$. Since $J(0) = 0$, the solution $\alpha = 0$ to Eq. (17) 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.

Using filters of length three

Unlike filters of length 2 the research for the optimal filter of length 3 is a more demanding task. Consider two lags l_α and l_β. The filter of length 3 can be expressed as $\mathbf{h} = [h_0, h_1, h_2] = [1, \alpha, \beta]$, where α and β are free parameters. Then

$$\mathbf{D} = (1 + \alpha^2 + \beta^2)\mathbf{I} + 2\alpha\mathbf{R}_s(l_\alpha) + 2\beta\mathbf{R}_s(l_\beta) + 2\alpha\beta\mathbf{R}_s(l_\alpha - l_\beta) \quad (18)$$

It can be witnessed that $\mathbf{D}$ involves three unknown correlation matrices $\mathbf{R}_s(l_\alpha)$, $\mathbf{R}_s(l_\beta)$, $\mathbf{R}_s(l_\alpha - l_\beta)$. We expect the performance to be improved compared to the case of filters of length 2 where only two matrices are involved. However, the analytical optimization of eigenvalue spreading cost J in (15) is not possible because too many unknowns are involved. For this case we proposed to use an iterative process which delivers filters of length three with improved performance in each iteration.

Using the process described in the previous section for $\mathrm{h} = [h_0, h_1] = [1, 1]$, an initial estimate of the source signals $\hat{\mathbf{s}}$ can be obtained (eq. (11)). The missing correlation matrices of eq. (18) can be calculated using the estimated source signals $\hat{\mathbf{s}}$. Following the reasoning of the previous subsection the optimality is related with the eigenvalue spread. As a consequence the search of the optimal filter of length three is transformed in the search for the filter that spreads the eigenvalues as much as possible. The criterion used is the one used before (eq. (15)). The search is exhaustive $\forall\alpha, \beta \in [h_{min}, h_{max}]$. The algorithm can be described in brief as follows:

- Estimate $\hat{\mathbf{s}}^{(1)}$ using $\mathbf{h} = [1, 1, 0]$.
- For i=1 to Maximum Iteration
  - Estimate correlation matrices using $\hat{\mathbf{s}}^{(i)}$.
  - Calculate $\mathbf{D}$ $\forall\alpha, \beta \in [h_{min}, h_{max}]$
  - Keep $[\alpha, \beta]$ minimizing $J(\alpha, \beta)$
  - Compute new source estimates $\hat{\mathbf{s}}^{(i+1)}$.

In the experiments we performed $h_{min} = -5$, $h_{max} = 5$, while the increasing step was 0.20.

OPCA NEURAL NETWORKS FOR BSS

Three neural models for OPCA have been proposed by Diamantaras and Kung [12]. Here we shall use the third model originally proposed in [6]. 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:

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

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

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 [12][chapter 7] for details).

SIMULATIONS

In the experiments described below, we chose the filter $h = [1, -1]$ with a single lag $l = 1$ and the filter $h = [1, -1, -1.2]$ with two lags $l_1 = 1$ and $l_2 = 2$. In the subsequent experiments we used $N = 2000$ samples from a set of four randomly generated multi-level PAM sources. The sources were colored by different random FIR filters of length 20. The colored sources were mixed using the following 4×4, linear memoryless operator:

$$\mathbf{A} = \begin{bmatrix} 0.4782 & -0.6880 & 0.6825 & -0.3145 \\ 0.2681 & 0.4163 & 0.3496 & 0.3319 \\ 0.3139 & 0.2472 & -0.4142 & 0.3158 \\ -0.5384 & 0.7213 & -0.6041 & -0.9309 \end{bmatrix}.$$

The network convergence for the two prefilters of length $L = 2$ and $L = 3$ is shown in Figure 2 and 3, respectively. Every sub-figure corresponds to a row of the estimated absolute matrix $\mathbf{B} = \mathbf{WA}$. Once the iteration stops, we locate the absolute maximum element $b_{i,j*}$ of the i-th row of $\mathbf{B}$. We then go back and for every iteration k we normalize each element $b_{i,j}(k)$ of the i-th row with $b_{i,j*}(k)$. Thus we are able to see the relative strength of all the components with respect to the strongest one. It can be seen that the choice of filter length = 3 yields smaller final convergence error in some components.

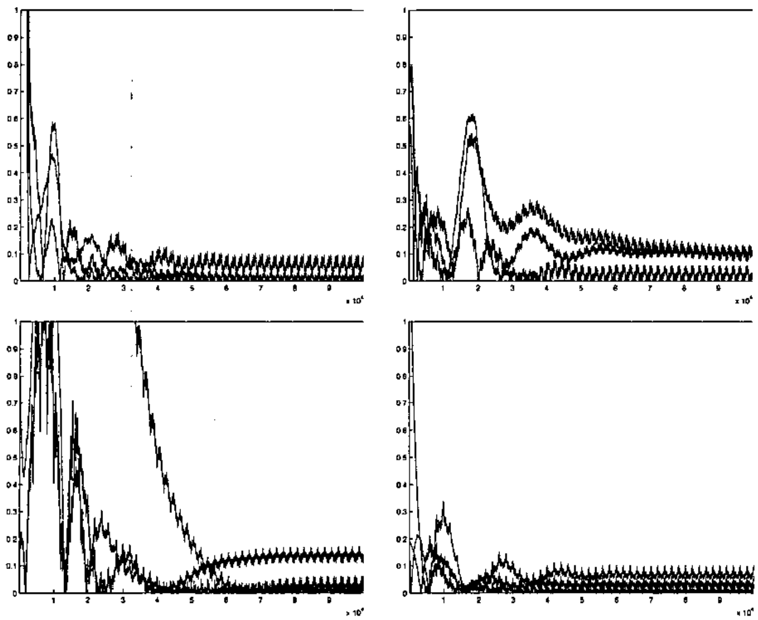

Figure 2: Converegence of the four OPCA components for a prefilter of length 2.

CONCLUSIONS

Second-order methods have been known to be related to the instantaneous BSS problem. Most earlier approaches have either required spatial prewhitening or have been limited to the use of two covariance matrices. The OPCA approach proposed here has the advantage that no preprocessing is required since spatial prewhitening is implicitly incorporated in the signal-to-signal ratio criterion which is optimized by OPCA. Furthermore, the use of multi-tap temporal prefilters allows for the implication of more than two matrices in the final estimation. This yields better and/or more robust results. Finally, the method can be implemented using a simple neural approach, making it more intuitively appealing compared to other neural approaches.

REFERENCES

[1] 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.

[2] 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.

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

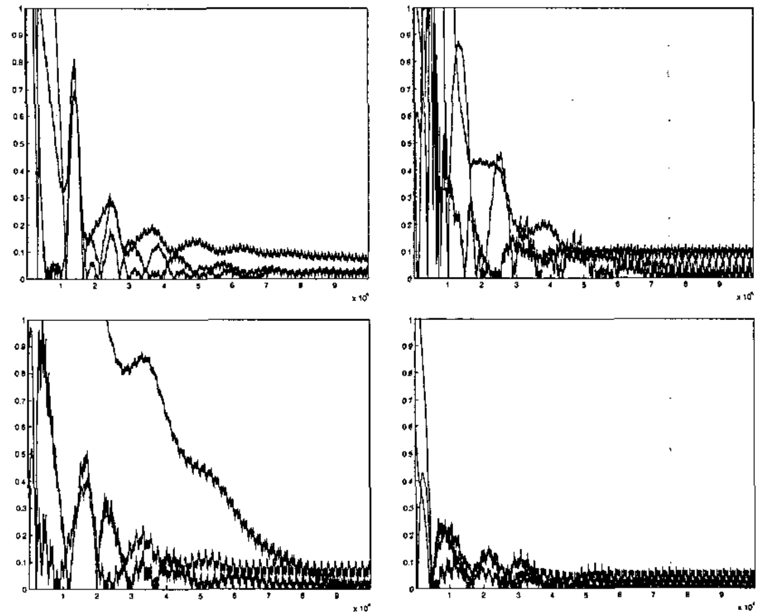

Figure 3: Converegence of the four OPCA components for a prefilter of length 3.

[4] C. Chang, Z. Ding, S. Yau and F. Chan, "A Matrix-Pencil Approach to Blind Separation of Colored Non-Stationaty Signals," **IEEE Trans. Signal Processing**, vol. 48, no. 3, pp. 900–907, 2000.

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

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

[7] K. 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. 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] K. Diamantaras, "PCA Neural Models and Blind Signal Separation," in **In proc. International Joint Conference on Neural Networks (IJCNN-2001)**, Washington DC, July 2001.

[10] K. Diamantaras, "Neural Networks and Principal Component Analysis," in Y. H. Hu and J. N. Hwang (eds.), **Handbook of Neural Network Signal Processing**, N.Y.: CRC Press, chap. 8, 2002.

[11] K. Diamantaras and S. Kung, "Cross-correlation Neural Network Models," **IEEE Trans. on Signal Processing**, vol. 42, no. 11, pp. 3218–3223, 1994.

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

[13] S. V. Gerven, D. V. 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.

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

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

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

[17] H. L. N. Thi and C. Jutten, "Blind Source Separation for convolutive mixtures," **Signal Processing**, vol. 45, pp. 209–229, 1995.

[18] L. Tong, G. Xu and T. Kailath, "Blind Identification and Equalization Based on Second-Order Statistics: A Time Domain Approach," **IEEE Trans. Information Theory**, vol. 40, no. 2, pp. 340–349, 1994.