ORIENTED PCA AND BLIND SIGNAL SEPARATION

K. I. Diamantaras

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

Th. Papadimitriou

Department of Int. Economic Relat. and Devel.,
Democritus University of Thrace
Komotini 69100, Greece
papadimi@uom.gr

ABSTRACT

Oriented PCA (OPCA) extends standard Principal Component Analysis by maximizing the power ratio of a pair of signals rather than the power of a single signal. We show that OPCA in combination with almost arbitrary temporal filtering can be used for the blind separation of linear instantaneous mixtures. Although the method works for almost any filter, the design of the optimal temporal filter is also discussed for filters of length two. Compared to other Second Order Statistics (SOS) methods this approach avoids the spatial prewhitening step. Further, it is a fast converging iterative approach which achieves better performance by combining several time lags for the estimation of the mixing parameters.

1. INTRODUCTION

Blind Source Separation (BSS) has been receiving a lot of attention in the last decade due to its usefulness in a variety of applications including digital communications, signal processing, and medical science. BSS is the task of recovering n unknown signals from their linear mixtures observed at m sensors. The term "blind" refers to the fact that the underlying mixing operator is unknown as well. BSS is part of a large family of Blind Problems including blind deconvolution, blind system identification, blind channel equalization, etc. Here we are interested in the special case where the linear mixing operator is memoryless. This problem is known 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].

The use of second order statistics (SOS) for blind problems was first noted by Tong, e.a. [12] in the early '90's. The problem studied in this classic paper was blind equalization. Although blind equalization is related to BSS it was not until 1997 that second order methods were proposed for BSS by Belouchrani e.a. [1]. We must note that second order methods do not actually replace higher order ones since each approach is based on different assumptions. For example, second order methods assume that the sources be temporally colored whereas higher order methods assume white sources. Another difference is that higher order methods do not apply on Gaussian signals but second order methods do not have any such constraint.

An adaptive second order Hebbian-like approach for BSS has been proposed in [7, 8]. This approach is based on neural SVD analyzers known as Asymmetric PCA models. One such model, the Cross-Coupled Hebbian rule proposed in [13], has been applied for the blind extraction of sources such as images or speech signals using a pre-selected time lag l. In [14] 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. Moreover, we are able to derive the optimal length-2 prefilter, although the design of the optimal prefilter, in general, is an open issue.

In this paper we demonstrate that Oriented PCA, an extension of PCA can be used for tackling the instantaneous BSS problem. The data fed to the OPCA method must be temporally prefiltered but almost any filter is sufficient except for the trivial impulse or delayed impulse function. Since OPCA relates to the Generalized Eigenvalue Decomposition (GED) method it is a well studied problem for which there are a number of fast and well known algorithms implemented in standard libraries such as MATLAB, etc. The advantages of the OPCA approach are summarized as follows: (a) using a single time lag it is a batch approach without slow recursive iterations (b) it can combine more than two time lags for better performance in an iterative algorithm with fast convergence and (c) it does not require spatial prewhitening.

2. PCA AND ORIENTED PCA (OPCA)

Oriented PCA (OPCA) is a term introduced by Kung and Diamantaras, (see for example [15][chapter 7]) as a generalization of PCA. It corresponds to the generalized eigenvalue decomposition (GED) of a pair of covariance matrices in the same way that PCA corresponds to the eigenvalue decomposition of a single covariance matrix. The cost function maximized by OPCA is the signal-to-signal ratio (SSR) between a pair of n-dimensional signals $\mathbf{u}$, $\mathbf{v}$

$$J_{OPCA}(\mathbf{w}) = \frac{E(\mathbf{w}^T\mathbf{u})^2}{E(\mathbf{w}^T\mathbf{v})^2} = \frac{\mathbf{w}^T\mathbf{R}_u\mathbf{w}}{\mathbf{w}^T\mathbf{R}_v\mathbf{w}}$$

where $\mathbf{R}_u = E\{\mathbf{u}\mathbf{u}^T\}$, $\mathbf{R}_v = E\{\mathbf{v}\mathbf{v}^T\}$. The maximizer $\mathbf{w} = \mathbf{e}_1$ of J_{OPCA} is called *principal oriented component* and it is the generalized eigenvector of the matrix pencil $[\mathbf{R}_u, \mathbf{R}_v]$ corresponding to the maximum generalized eigenvalue λ_1. Since $\mathbf{R}_u$ and $\mathbf{R}_v$ are symmetric all the generalized eigenvalues are real and thus they can be arranged in decreasing order, as with ordinary PCA. The rest of the generalized eigenvectors $\mathbf{e}_2$, $\mathbf{e}_3$, ..., $\mathbf{e}_n$, will be called second, third, ..., n-th oriented principal components. They all maximize the same cost function J_{OPCA} subject to the orthogonality constraint

$$\mathbf{e}_i^T\mathbf{R}_u\mathbf{e}_j = \mathbf{e}_i^T\mathbf{R}_v\mathbf{e}_j = 0, i \neq j$$

The term “oriented” is justified by the fact that $\mathbf{e}_1$ is similar to the ordinary principal component for $\mathbf{u}$ except that it is oriented towards the least principal direction of $\mathbf{v}$. In other words, the distribution of $\mathbf{v}$ “steers” $\mathbf{e}_1$ towards the direction of the least energy of $\mathbf{v}$. Clearly if the distribution of $\mathbf{v}$ is isotropic in all directions, ie. $\mathbf{v}$ is white noise, then the steering is absent and the oriented PC is identical to the ordinary PC. This is easily verified mathemetically by setting $\mathbf{R}_v = \mathbf{I}$ in J_{OPCA} and recognizing it as the cost function maximized by PCA.

3. BSS PROBLEM DESCRIPTION AND ASSUMPTIONS

The description of the instantaneous BSS problem and the adopted assumptions are outlined next. We consider n observed signals x_1, ..., x_n, resulting from the linear mixing of n source signals s_1, ..., s_n. Defining the stochastic observation and source vectors $\mathbf{x}(k) = [x_1(k), \cdots, x_n(k)]^T$ and $\mathbf{s}(k) = [s_1(k), \cdots, s_n(k)]^T$ we have:

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

For simplicity, we assume that the linear mixing operator $\mathbf{A}$ is a square and invertible matrix. 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.

It is important to note that the order and the scale of the individual sources are unobservable.

In the following discussion we shall use the following notation

$$\mathbf{R}_z(l) \stackrel{\triangle}{=} E\{\mathbf{z}(k)\mathbf{z}(k-l)^T\}$$

to denote the covariance of any signal $\mathbf{z}$ with time-lag l. This notation is useful because the assumptions described next involve time-lagged second order statistics. This set of assumptions is typical to most SOS methods (see [1, 16]):

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

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

A2. The sources are temporally colored. In particular, there exist $M > 1$ positive time lags l_1, ..., l_M such that:

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

Define $l_0 = 0$.

A3. The sources have distinct covariance functions, i.e. their spectral densities are not identical:

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

4. SOLVING BSS USING OPCA

Subsequently, we shall relate the instantaneous BSS problem with the OPC Analysis of the following pair of signals (i) the observed signal $\mathbf{x}$ and (ii) almost any filtered version of it. Note that the 0-lag covariance matrix of $\mathbf{x}(k)$ is

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

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

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

for some lags l_0, ..., l_M. The 0-lag covariance matrix of $\mathbf{y}$ is written as

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

Assume that $\mathbf{R}_x(l_p - l_q)$ satisfy assumptions [A2], [A3] for all p, q. From Eq. (1) it follows that

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

so

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

with

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

Since $\mathbf{A}$ is square and invertible we can write

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

where $\mathbf{A}^{-T} \triangleq {\mathbf{A}^{-1}}^T$. Note that, by assumptions [A1], [A2], $\mathbf{D}$ is diagonal. 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 eigenvalues 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, i.e. each row and each column contains exactly one no-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.

4.1. Designing 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

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

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.

4.2. 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

$$\begin{aligned} \mathbf{D} = (1+\alpha^2+\beta^2)\mathbf{I} &+ 2\alpha\mathbf{R}_s(l_\alpha) + 2\beta\mathbf{R}_s(l_\beta) \\ &+ 2\alpha\beta\mathbf{R}_s(l_\alpha - l_\beta) \end{aligned} \tag{19}$$

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 (16) 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 $\mathbf{h} = [h_0, h_1] = [1, 1]$, an initial estimate of the source signals $\hat{\mathbf{s}}$ can be obtained (eq. (12)). The missing correlation matrices of eq. (19) can be calculated using the estimated source signals $\hat{\mathbf{s}}$. Following the reasoning of subsection **4.1** 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. (16)). 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.

5. RESULTS

We shall compare the proposed method against the AMUSE approach [17]. AMUSE is a non-iterative (batch) approach which uses second order statistics. OPCA using the optimal filter of length 2 is also a batch method. Consequently, the two methods have comparable speeds. There are however, two major differences between these methods: (a) AMUSE is limited to only one time lag and (b) it uses spatial prewhitening. The first difference, as we shall see, has a severe effect on performance.

The AMUSE method, was compared with the proposed iterative approach, described in section **4.2**. First blind separation is performed using the OPCA method with a temporal pre-filter of length 2 and then the calculated results are introduced in a four iteration process, performing OPCA using filter of length 3.

The experiment setup is described next. The source and observation data dimension is $n = 4$ and the mixing matrix is chosen randomly to be

$$\mathbf{A} = \begin{bmatrix} 0.4437 & 0.4422 & -0.5027 & -0.5972 \\ 0.3517 & -0.6333 & 0.8880 & 0.9198 \\ 0.3547 & -0.3207 & -0.8829 & 0.8573 \\ 0.7458 & -0.7583 & 0.6749 & -0.7159 \end{bmatrix}$$

The sources are four multilevel PAM signals filtered by an ARMA coloring filter of length 20. We take $N = 2000$ samples per signal. In the AMUSE method the observations are spatially prewhitened by the transform $\mathbf{x}_w(k) = \mathbf{R}_x(0)^{-1/2}\mathbf{x}(k)$.

If $\mathbf{Q}$ is the estimated separation matrix, as mentioned in the section **4** then $\mathbf{A}^T\mathbf{Q} = \mathbf{P}$ with $\mathbf{P}$ being a scaled permutation matrix. Performance is calculated using the matrix $\mathbf{P}$. First, the columns of $\mathbf{P}$ are arranged in order to achieve an approximation of a diagonal matrix i.e. the maximum absolute value of every column must be in the matrix diagonal. Then, every column is divided by its diagonal element. The transformed matrix $\mathbf{P}'$ is an approximation of the identity matrix $\mathbf{I}$. The perfect estimation will yield that $\mathbf{P}' = \mathbf{I}$. Therefore the estimation quality is measured using the sum of the absolute value of the off-diagonals elements of matrix $\mathbf{P}'$:

$$EQ = \sum_{i \neq j} |P'(i,j)| \tag{20}$$

It must be noted that as a consequence to the pre-whitening process in the AMUSE case, $\mathbf{P} = \mathbf{W}\mathbf{A}^T\mathbf{Q}$, where $\mathbf{W}$ is the whitening matrix.

Comparison between the two methods was made using the difference of their estimation quality.

$$C = EQ_{AMUSE} - EQ_{OPCA} \tag{21}$$

The comparison formula has positive values whenever the OPCA has better performance than AMUSE, and negative values in the opposite case.

We conducted tests using 50 datasets. Various levels of noise were injected in order to report the robustness of the algorithm. Noise is calculated in dB using SNR. OPCA results are extracted from every step of the algorithm. The mean values over the 50 datasets are presented in table 1.

It is easily witnessed that the results extracted from the first iteration of OPCA and AMUSE are almost equal, meaning that OPCA with filter of length 2 performs as the AMUSE method. This is obvious since both methods use the same amount of information in their estimation. On the contrary, in every case we tested OPCA using filter of length 3 with the AMUSE method the increase of the estimation accuracy was very important. The accuracy improvement varies from 0.15 to 0.35. The use of more correlation matrices in the case of OPCA using filter with length 3, increases the information input in the estimation process, improving

SNR	Iter 1	Iter 2	Iter 3	Iter 4	Iter 5
100	0.0269	0.1875	0.2238	0.2410	0.2325
50	-0.0112	0.3217	0.3531	0.3563	0.3564
30	-0.0056	0.3723	0.3774	0.3769	0.3757
20	-0.0312	0.2196	0.2163	0.2276	0.2196
10	0.0027	0.3063	0.1644	0.1754	0.1524

Table 1. Comparison between the AMUSE and the OPCA blind separation technique. OPCA results are extracted in every iteration. Iteration 1 corresponds to the batch OPCA technique using filter with length 2. Comparison is made using datasets with different noise levels.

the separation quality. Finaly, it is important to outline the convergence quickness. In every experiment we performed convergence was reached by the third iteration.

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 [16]. 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 proposed approach is iterative and improves the performance by combining several time lags for the estimation of the mixing parameters.

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] 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.

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

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

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

[16] 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.

[17] L. Tong, R.-W. Liu, V.C. Soon, and Y.-F. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Trans. Circuits and Systems*, vol. 38, pp. 499–509, 1991.