International Conference on
Information, Communications and Signal Processing
ICICS '97
Singapore, 9-12 September 1997

3E2.5

Sequential Approach to Blind Source Separation Using Second Order Statistics

Chunqi Chang[1], Sze Fong Yau[2], Paul Kwok[1], F K Lam[1], Francis H Y Chan[1]

[1]Dept of Electrical and Electronic Engineering
The University of Hong Kong
Pokfulam Road, Hong Kong
{cqchang, fklam, kwok, fhychan}@eee.hku.hk

[2]Dept of Electrical and Electronic Engineering
Hong Kong University of Science & Technology
Clear Water Bay, Kowloon, Hong Kong
Email: eeyau@ee.ust.hk

Abstract

A general result on identifiability for the blind source separation problem, based on second order statistics only, is presented in this paper. The separation principle using second order statistics only is first proposed. This is followed by a discussion on a number of algorithms to separate the sources one by one.

1. Introduction

The blind source separation problem [1] is described as

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

where
(i) $\mathbf{s}(n)$ is a vector of unknown input signals which are zero mean stationary processes with unit variance;
(ii) $\mathbf{A}$ is the unknown mixing matrix which is of full column rank;
(iii) $\mathbf{x}(n)$ is the vector of measured signals, called mixtures.

The so-called blind source separation problem is in fact a blind source extraction problem. The task is to extract the source signals $\mathbf{s}(n)$ from $\mathbf{x}(n)$. Our objective is to design an *extractor* **b** such that

$$z(n) = \mathbf{b}^{\mathrm{T}}\mathbf{x}(n) \tag{2}$$

is a scaled version of one of the sources. Then we design another extractor to extract another source. The procedure is iterated to get as many sources as we want. This is referred to a sequential approach [2] in which we extract the sources one by one. The sequential approach is more effective when there are many sources but we want to extract only some of them. It also leads to a simpler implementation.

Usually blind source separation is done by exploiting higher order statistics. On the other hand, the use of second order statistics leads to reduced amount of computation and the ability to extract Gaussian sources. Furthermore, in practice, empirical estimates of second order statistics are more stable than that of higher order statistics. It is well known that without using temporal information, second order statistics itself is not enough for blind source separation, but recently it has been proved that under some assumptions the problem can be solved by using second order statistics only [1, 3].

The work in this paper is based on the sequential approach. The extractor is designed using only the second order statistics of $\mathbf{x}(n)$.

2. Extractability

For blind source separation problem, there are three basic issues: extractability, extraction principle, and extraction algorithm. In this section we consider the first issue, extractability. The other two issues will be considered in the subsequent sections.

The issue of extractability considers the problem that under what conditions the sources can be extracted? For non-Gaussian sources (or at most one is Gaussian), this problem has already been solved by exploiting higher order statistics of the signals [4]. Here we try to solve the problem by using second order statistics only.

If $E\left\{s_i(n)s_j(n-k)\right\} = 0$ for any $i \neq j$ and $k \in \Omega$, the sources $\{s_i(n),\ i = 1, 2, \ldots, m\}$ are said to be *uncorrelated on* Ω, where Ω is an assembly which contains 0 and is included in the assembly of integer, $\boldsymbol{I}$. If $\Omega = \boldsymbol{I}$, the sources are also simply called *uncorrelated*.

Assume that the signal model is described by (1), where **A** is of full column rank. Denote

$$\Omega = \left\{0, k_1, k_2, \cdots, k_L\right\}$$

$$\Omega_0 = \Omega - \left\{0\right\} = \left\{k_1, k_2, \cdots, k_L\right\}$$

The autocorrelation function of source $s_i(n)$ on Ω is

$$r_i(\Omega)=\begin{bmatrix} r_i(0) & r_i(k_1) & r_i(k_2) & \cdots & r_i(k_L)\end{bmatrix}^T$$

where $r_i(k)=E\{s_i(n)s_i(n-k)\}$.

Since the sources are of unit variance, $r_i(0)=1$, we can consider the autocorrelation function on Ω_0 only, which is denoted as

$$\mathbf{r}_i=r_i(\Omega_0)=\begin{bmatrix} r_i(k_1) & r_i(k_2) & \cdots & r_i(k_L)\end{bmatrix}^T$$

Let $\mathbf{x}(n)=\mathbf{A}\mathbf{s}(n)$, besides the three assumptions in Section 1, further assume that

(iv) The sources are uncorrelated on Ω.

Then, for the extractability of the source separation problem, we have the following theorem.

Theorem 1: For a source whose autocorrelation function on Ω is different from that of the others, it can be extracted by using only second order statistics of the mixtures $\mathbf{x}(n)$.

Proof: Here we give a constructive proof. Without losing generality, denote $s_1(n)$ the source with the property specified in the theorem.

(1) Standardization.

Find a matrix $\mathbf{W}$ for which $\mathbf{U}_0=\mathbf{WA}$ is orthogonal. Note that this procedure can be easily done by various of ways [2]. Let

$$\mathbf{x}_0(n)=\mathbf{W}\mathbf{x}(n)=\mathbf{U}_0\mathbf{s}(n) \tag{3}$$

(2) Partition.

For $k=k_1$, let

$$\begin{aligned}\mathbf{R}_1&=E\{\mathbf{x}_0(n)\mathbf{x}_0(n-k_1)\}\\&=\mathbf{U}_0\Lambda_1\mathbf{U}_0^T=\sum_{m=1}^{M}r_m(k_1)\mathbf{u}_{0,m}\mathbf{u}_{0,m}^T\end{aligned} \tag{4}$$

where $\Lambda_1=diag\{r_1(k_1)\ \cdots\ r_M(k_1)\}$.

Let $\Gamma_1\subseteq\{1,2,\cdots,M\}$, $\forall\ i\in\Gamma_1$, $r_i(k_1)=r_1(k_1)$; and $\forall\ i\notin\Gamma_1$, $r_i(k_1)\neq r_1(k_1)$. Then the eigen decomposition of $\mathbf{R}_1$ can be written as

$$\mathbf{R}_1=r_1(k_1)\tilde{\mathbf{U}}_1\tilde{\mathbf{U}}_1^T+\tilde{\mathbf{U}}_2\Lambda_{1,2}\tilde{\mathbf{U}}_2^T \tag{5}$$

Under this decomposition, we have

$$\mathbf{x}_1(n)=\tilde{\mathbf{U}}_1^T\mathbf{x}_0(n)=\mathbf{U}_1\mathbf{s}_1(n) \tag{6}$$

where $\mathbf{s}_1(n)=\mathbf{s}_{\Gamma_1}(n)$ is a subset of $\mathbf{s}(n)$, and $\mathbf{U}_1$ is an orthogonal matrix.

If $\Gamma_1=\{1\}$, then $\mathbf{x}_1(n)=\pm s_1(n)$, the source has already been extracted. Otherwise, go to next step.

(3) Iteration

For $k=k_2$, let

$$\begin{aligned}\mathbf{R}_2&=E\{\mathbf{x}_1(n)\mathbf{x}_1(n-k_2)\}\\&=\mathbf{U}_1\Lambda_2\mathbf{U}_1^T=\sum_{m\in\Gamma_1}r_m(k_2)\mathbf{u}_{1,m}\mathbf{u}_{1,m}^T\end{aligned} \tag{7}$$

Similarly as in step (2), we have

$$\mathbf{x}_2(n)=\tilde{\mathbf{U}}_2^T\mathbf{x}_1(n)=\mathbf{U}_2\mathbf{s}_2(n) \tag{8}$$

where $\mathbf{s}_2(n)=\mathbf{s}_{\Gamma_2}(n)$ is a subset of $\mathbf{s}(n)$, and $\mathbf{U}_2$ is an orthogonal matrix.

If $\mathbf{x}_2(n)=\pm s_1(n)$, stop. Otherwise, repeat this procedure for $k=k_3,\cdots$. Note that there must be a $l\leq L$, for which $\mathbf{x}_l(n)=\pm s_1(n)$. Otherwise, $\exists\ i\neq 1$, $\mathbf{r}_i=\mathbf{r}_1$, which is contradict to the assumption made in the theorem. Therefore, we has proved that the source can be extracted.

3. Extraction Principle

If a source is proved to be extractable, then the subsequent problem is how to extract it. The problem can further be divided into two sub-problems: extraction principle which considers the principle under which the sources can be extracted, and extraction algorithm which is based on the extraction principle. In this section we consider the problem of extraction principle.

It can be proved that if the sources are uncorrelated on Ω, then there exists $\mathbf{c}=\begin{bmatrix} c_1 & \cdots & c_L\end{bmatrix}$ for which the entries of $\mathbf{r}_c=\sum_{l=1}^{L}c_l\mathbf{r}(k_l)$, where the ith entry of $\mathbf{r}(k_l)$ is

$E\{x_i(n)x_i(n-k_l)\}$, are different from each other. With $z(n)$ defined in (2), we have the following theorem.

Theorem 2: If $\mathbf{r}_c$ has only one maximum (minimum) element, then the maximizer (minimizer) of $F_c(\mathbf{b}) = \sum_{l=1}^{L} c_l E\{z(n)z(n-k_l)\}$, subject to $E\{z^2\} = 1$, is an extractor.

Proof: Let $\mathbf{d} = \mathbf{A}^T\mathbf{b}$, and $J_c(\mathbf{d}) = F_c(\mathbf{b}) = \sum_{m=1}^{M} d_m^2 r_{c,m}$, where d_m is the mth entry of $\mathbf{d}$, and $r_{c,m}$ is the mth entry of $\mathbf{r}_c$. Since $\mathbf{A}$ is of full column rank, the extremization problem described in the theorem is equivalent to

$$\max_{\mathbf{d}^T\mathbf{d}=1} / \min J_c(\mathbf{d}) \tag{9}$$

Note that $\sum_{m=1}^{M} d_m^2 r_{c,m} \le r_c^{\max} \sum_{m=1}^{M} d_m^2 = r_c^{\max}$, where $r_c^{\max}$ is the maximum value of the entries of $\mathbf{r}_c$. Denote m the index of the single entry which is equal to $r_c^{\max}$. The maximun value of $J_c(\mathbf{d})$ is reached if and only if $d_m = \pm 1$ and $d_n = 0$ for any $n \ne m$. For this maximizer $\mathbf{d}$ we have $z(n) = \pm s_m(n)$ which is just one of the sources. Similarily, for the minimizer, we can also get this result. Therefore the theorem is proved.

Remark 1: If all the sources have different spectra, then the assumption in Theorem 2 is reasonable, thus the extremizer will extract one of the sources. When some of the sources have identical spectra, the assumption will be invalid if the elements of $\mathbf{r}_c$ corresponding to these sources are just the maximum/minimum elements, and in this case the maximizer/minimizer will extract a signal which is a combination of these sources rather than a pure source. Also note that if **c** is not properly chosen, even the source with autocorrelation function on Ω different from that of the others cannot be extracted correctly by this separation principle. The choice of **c** is problem dependent. In practice, for computational efficiency, **c** should be chosen with the smallest number of nonzero components possible. If the sources are uncorrelated on I, we can also use an AR filter to make **c** with an infinite number of nonzero components.

4. Algorithms to Extract One Source

The algorithms depend on the separation principle stated in Theorem 2. The extremization of $F_c(\mathbf{b})$ subject to $E\{z^2\} = 1$, or its equivalent, extremization of $J_c(\mathbf{d})$ subject to $\mathbf{d}^T\mathbf{d} = 1$, is a constrained optimization problem. To solve the constrained optimization problem we can use the Projected Gradient method [5]. It is shown that the convergence rate of Projected Gradient method is very slow [5]. Here we give a more efficient method.

First, we solve the extremization problem (9) by a special gradient based algorithm described below.

$$\mathbf{d}^{(k+1)} = \frac{1}{2}\frac{\partial J_c(\mathbf{d})}{\partial \mathbf{d}}\bigg|_{\mathbf{d}=\mathbf{d}^{(k)}} = diag(\mathbf{r}_c)\mathbf{d}^{(k)} \tag{10}$$

$$\mathbf{d}^{(k+1)} = \frac{\mathbf{d}^{(k+1)}}{\left\|\mathbf{d}^{(k+1)}\right\|} \tag{11}$$

Assume that $|r_{c,\max}| \ne |r_{c,\min}|$. Denote by m_0 the index for which $|r_{c,m_0}| = \max(|r_{c,\max}|, |r_{c,\min}|)$. Then we have

$$\frac{d_m^{(k+1)}}{d_{m_0}^{(k+1)}} = \frac{r_{c,m}}{r_{c,m_0}}\frac{d_m^{(k)}}{d_{m_0}^{(k)}} = \left(\frac{r_{c,m}}{r_{c,m_0}}\right)^k \frac{d_m^{(0)}}{d_{m_0}^{(0)}} \tag{12}$$

It means that the algorithm will converge exponentially to $\operatorname{sgn}(d_{m_0})\mathbf{e}_{m_0}$ or a limit circle $(\mathbf{e}_{m_0}, -\mathbf{e}_{m_0})$. In both cases the extracted signal is one of the sources.

Then we can derive the equivalent algorithm to extremize $F_c(\mathbf{b})$ subject to $E\{z^2\} = 1$. Due to space limitation, the detailed derivation is omitted here. Basically the algorithm can be implemented in two forms.

Algorithm 1

$$\mathbf{b}^{(k+1)} = \mathbf{R}_0^{-1}\mathbf{R}_c\mathbf{b}^{(k)} \tag{13}$$

$$\mathbf{b}^{(k+1)} = \frac{\mathbf{b}^{(k+1)}}{\sqrt{(\mathbf{b}^{(k+1)})^T\mathbf{R}_0\mathbf{b}^{(k+1)}}} \tag{14}$$

Algorithm 2

$$z(n) = (\mathbf{b}^{(k)})^T\mathbf{x}(n) \tag{15}$$

$$\mathbf{b}^{(k+1)} = \mathbf{R}_0^{-1}\sum_m c_m E\{\mathbf{x}(n)z(n-k_m)\} \tag{16}$$

$$\mathbf{b}^{(k+1)} = \frac{\mathbf{b}^{(k+1)}}{\sqrt{(\mathbf{b}^{(k+1)})^T\mathbf{R}_0\mathbf{b}^{(k+1)}}} \tag{17}$$

where

$$\mathbf{R}_0 = E\{\mathbf{x}(n)\mathbf{x}^T(n)\} \quad (18)$$

$$\mathbf{R}_c = \sum_m c_m E\{\mathbf{x}(n)\mathbf{x}^T(n-k_m)\} \quad (19)$$

Remark 2:

(1) If $\mathbf{R}_0 = \mathbf{I}$, then the algorithm will be very simple. Thus in practice we let $\tilde{\mathbf{x}}(n) = \mathbf{T}\mathbf{x}(n)$ in order that $E\{\tilde{\mathbf{x}}(n)\tilde{\mathbf{x}}^T(n)\} = \mathbf{I}$. Note that this procedure is called prewhittening which can be done by various of ways such as eigen decomposition, PCA (Principle Component Analysis), Choleskey Factorization, and the Cardoso's adaptive approach [6]. Then we can use the following simplified algorithm to extract one of the sources.

Algorithm 3

$$z(n) = (\mathbf{b}^{(k)})^T \tilde{\mathbf{y}}(n) \quad (20)$$

$$\mathbf{b}^{(k+1)} = \sum_m a_m E\{\tilde{\mathbf{y}}(n)z(n-k_m)\} \quad (21)$$

$$\mathbf{b}^{(k+1)} = \frac{\mathbf{b}^{(k+1)}}{\|\mathbf{b}^{(k+1)}\|} \quad (22)$$

(2) The algorithms can be straightforwardly modified to recursive/adaptive form by stochastic approximation [7]. The derivation is omitted here due to space limitation.

5. Extraction of the Other Sources

If we assume that all the sources have different autocorrelation functions on Ω_0 and $\mathbf{c}$ is properly set, we have derived several methods to extract the other sources. One method is just use the algorithm again on the prewhittened data but in each iteration, we project $\mathbf{b}$ to the null space spanned by all the previously obtained extractors. Then we obtain a new extractor and a new source. The second method is to modify the data so that they are free of the extracted sources. Then the source extraction algorithm is used again on the modified data to get a new source. Data modification can be implemented by Delfosse and Loubaton's deflation technique [2] if the data are prewhittened. We also have another simple method which can work on non prewhittened data.

Assume that the extracted source $z(n) = s_m(n)$, then

$$\mathbf{h} = E\{\mathbf{x}(n)z(n)\} = E\{\mathbf{A}\mathbf{s}(n)s_m(n)\} = A_m \quad (23)$$

where A_m is the mth column of the mixing matrix $\mathbf{A}$.

Then we let

$$\tilde{\mathbf{x}}(n) = \mathbf{x}(n) - \mathbf{h}z(n) \quad (24)$$

Note that

$$\tilde{\mathbf{x}}(n) = \mathbf{A}\mathbf{s}(n) - A_m s_m(n) = \tilde{\mathbf{A}}\tilde{\mathbf{s}}(n) \quad (25)$$

where

$$\tilde{\mathbf{A}} = \begin{bmatrix} A_1 & \cdots & A_{m-1} & A_{m+1} & \cdots & A_M \end{bmatrix}$$
$$\tilde{\mathbf{s}} = \begin{bmatrix} s_1 & \cdots & s_{m-1} & s_{m+1} & \cdots & s_M \end{bmatrix}^T.$$

Run the one-source extraction algorithms again on the data $\tilde{\mathbf{x}}(n)$ in which the extracted source is already subtracted away, then we can get another source $\tilde{z}$ different from z.

When there are some sources whose autocorrelation functions on Ω_0 are identical or $\mathbf{c}$ is not properly chosen, the iterative procedure may not always extract the correct source in each iteration. Fortunately this does not make the procedure completely invalid. For the extracted signals at all the iteration stages, any signal that is uncorrelated with the others on Ω is a correct source.

6. Simulation

In the simulation, four audio signals as shown in Fig 1 are used as the input sources. The sampling rate is 8000Hz, and the signal length is 4.375 seconds. We mix the sources by applying a random 4 by 4 matrix to obtain 4 mixtures. As shown in Fig 2, the signals used for source extraction are the Gaussian noise contaminated mixtures with 20dB SNR. Then we use Algorithm 1 in combination with the source substraction technique described by (23) (24) (25) to extract all the sources, and the extracted signals are shown in Fig 3. For the purpose of comparision, higher order statistics based blind signal separation algorithm described in [8] is also used to extract the sources, and the results are shown in Fig 4. In the algorithms we use the first 1000 data (0.125 seconds) to estimate the required statistics. As one can see, Fig 3 is very similar to Fig 1 except the order and the sign of the signals, which means that our second order statistics based sequential approach works well in this case. Comparing Fig 3 with Fig 4, we find that the two methods, second order statistics based and higher order statistics based, have similar performance. Listening test shows that the result of our method is a little better.

7. Concluding Remarks

In this paper we have proposed a sequential approach to blind source separation using second order statistics only. Algorithms were designed successfully to solve the problem. There are two key issues in this work.

The first issue is the use of second order statistics only. Using second order statistics only is of great interest since it has some advantages over higher order statistics. In this paper we first presented a general result on the identifiablity using second order statistics only, as stated in Theorem 1. A new result from Theorem 1 is that when there are some sources that cannot be extracted, we can still correctly extract the others.

Another issue is the sequential approach. Comparing to simultaneous extraction of all the sources, extracting the sources one by one has some advantages. Sequential approach to blind source separation using higher order statistics has been done before by others [2]. Our work is the first that makes use of second order statistics only.

Good performance of our approach is verified by computer simulation.

References

1. L. Tong, R.W. Liu, V.C. Soon, and Y.F. Huang, "Indeterminacy and identifiability of blind deconvolution", IEEE Trans on Circuits Syst., pp. 499-509, May 1991.
2. N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: a deflation approach", Signal Processing, vol. 45, no. 1, pp59-83, 1995.
3. A. Belouchrani, K.A. Meraim, J.F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics", IEEE Trans. On Signal Processing, vol. 45, no. 2, pp. 434-444, Feb. 1997.
4. P. Comon, "Independent component analysis, a new concept?", Signal Processing, vol. 36(3):287-314, Apr. 1994.
5. J.C. Dunn, "On the convergence of projected gradient processes to singular critical points", Journal of Optimization Theory and Applications, vol. 55, no. 2, pp. 203-216, Nov. 1987.
6. J.F. Cardoso and B. Laheld, "Equivariant adaptive source separation", IEEE Trans. Signal Processing, Dec. 1996.
7. H.J. Kushner and D.S. Clark, Stochastic Approximation Methods for Constrained and Unconstrained Systems, Springer, New York, 1978.
8. E. Oja and A. Hyvarinen, "Blind signal separation by neural networks", Proc. ICONIP96, pp7-14, Sep 1996.

Acknowledgement

This work is supported in part by grants from Research Grants Council of Hong Kong, grant numbers HKU553/96M, HKU7036/97E, and HKUST776/96E.

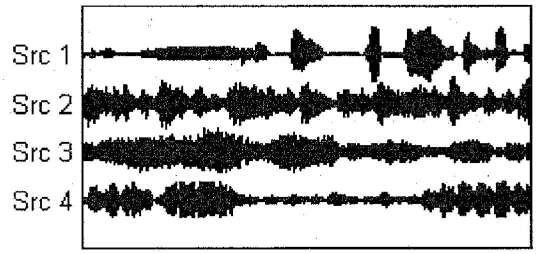

Fig. 1 The Original Sources

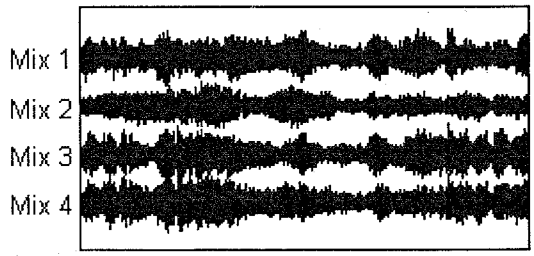

Fig. 2 The Mixtures

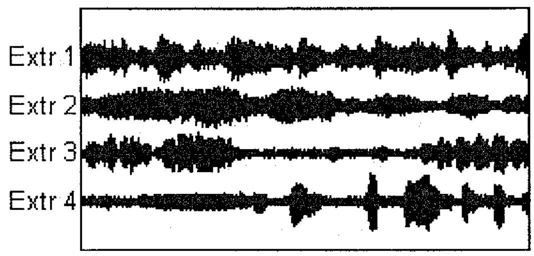

Fig. 3 Extracted Sources by our approach

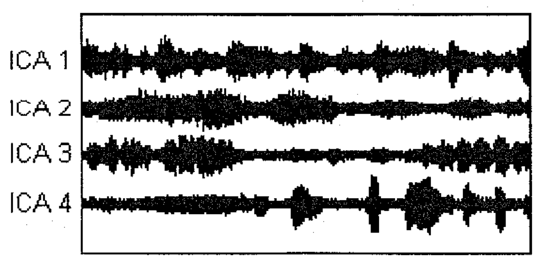

Fig. 4 Extracted Sources by HOS based method