Blind Separation of Instantaneous Mixture of Sources via the Gaussian Mutual Information Criterion

Dinh Tuan Pham
Laboratory of Modeling and Computation, IMAG, C.N.R.S.
Univ. of Grenoble, B.P. 53X, 38041 Grenoble Cedex, France
e-mail: Dinh-Tuan.Pham@imag.fr

ABSTRACT

A method for blind separation of instantaneous mixture of colored sources, based on the minimization a Gaussian mutual information criterion, is proposed. It amounts to jointly approximately diagonalizing a set of estimated spectral density matrices. Separation is shown to be achievable (up to a scaling and a permutation) if no pair of sources can have proportional spectral densities. An effcient algorithm for the joint approximate diagonalization of positive matrix is described. Theoretical results on the asymptotic performance of the procedure are given and some simulations are performed showing good agreement with the theory. It is seen that nearly optimal performance can be attained by jointly diagonalizing only a few spectral matrices.

1 INTRODUCTION

Blind separation of sources have received much interest recently because of its many applications in signal processing. We consider here the simplest case in which linear instantaneous mixtures of independent sources are recorded and the goal is to recover the sources without relying on any specific assumption other than their mutual independence. In many earlier papers (see [3] for a review), only the marginal instantaneous distributions of the observations enter consideration and it is then well known that the use of second order statistics is insufficient. However, Pham and Garat [9] have considered the case of colored sources and derived a separation procedure based only on second order lagged moments. The use of such moments has also been exploited by Tong *et al.* [10] and Belouchrani *et al.* [1]. In this paper we propose a novel procedure derived from the Gaussian mutual information criterion which also relies on second order statistics only. It has some similarities to the SOBI method of [1] in that it consists essentially of diagonalizing jointly approximately a set of matrices. However it uses a *different* measure of deviation from diagonality, not requiring the constraint of orthogonality, as in the SOBI method. This is important since such constraint implies a pre-whitening step, which can have adverse effect on the overall performance of the method [2]. Further, our criterion results from the well understood concepts of mutual information and entropy and the matrices to be diagonalized arise naturally while in the SOBI method they are chosen in an ad hoc manner. It is then not surprising that our method can nearly achieve the Cramér-Rao (CR) bound, as seen in the simulations.

2 GAUSSIAN MUTUAL INFORMATION AND SOURCE SEPARATION

The concept of mutual information is well known and have been proposed as a criterion for of sources separation ([4], [6]), but it is normally defined for a set of random variables. To exploit the time dependence structure of the signal, Pham [7] has extended it to stationary random processes. It is known that the mutual information between K random vectors $Y_1, \ldots, Y_K$ equals $\sum_{i=1}^K H(Y_k) - H(Y_1, \ldots, Y_K)$ where $H(Y_k) = -\mathrm{E} \log f_{Y_k}(Y_k)$ is the (Shannon) entropy of Y_k, with density f_{Y_k}, and $H(Y_1, \ldots, Y_K)$ is the (joint) entropy of $Y_1, \ldots, Y_K$, that is the entropy of the vector obtained by stacking the components of $Y_1, \ldots, Y_K$. For a stationary process $\{Y(t),\ t \in \mathbb{Z}\}$, Pham [7] defines the entropy (power) $H[Y(\cdot)]$ as the limit as $T \to \infty$ of $H[Y(1), \ldots, Y(T)]/T$ and then defines the mutual information between K jointly stationary processes $Y_1(\cdot)$, ..., $Y_K(\cdot)$ by

$$I[Y_1(\cdot), \ldots, Y_K(\cdot)] = \sum_{i=1}^K H[Y_k(\cdot)] - H[Y_1(\cdot), \ldots, Y_K(\cdot)].$$

The property that the mutual information between random variables is non negative and can be zero only if they are independent (see for ex. [3], [6]) clearly extends to the case of random processes. However, the use of mutual information in this case leads to costly numerical procedures and also complex analysis. Therefore we introduce the concept of Gaussian entropy and Gaussian mutual information, denoted by H_g and I_g, defined in the same way but with the random vectors or processes involved being replaced by the Gaussian vectors or processes having the same covariance structure. It is known that for a K-vector stationary process $\mathbf{Y}(\cdot)$:

$$H_g[\mathbf{Y}(\cdot)] = \frac{1}{4\pi} \int_{-\pi}^{\pi} \log \det[4\pi^2 \mathbf{f}_{\mathbf{Y}}(\lambda)] \mathrm{d}\lambda + \frac{K}{2}$$

where $\mathbf{f}_{\mathbf{Y}}$ is its spectral density matrix [7]. Thus letting Y_1, ..., Y_K be the components of $\mathbf{Y}$:

$$I_g[Y_1(\cdot), \ldots, Y_K(\cdot)] =$$

$$\frac{1}{4\pi}\int_{-\pi}^{\pi}\{\log\det\operatorname{diag}[\mathbf{f}_{\mathbf{Y}}(\lambda)]-\log\det\mathbf{f}_{\mathbf{Y}}(\lambda)\}\mathrm{d}\lambda \quad (1)$$

where $\operatorname{diag}(\cdot)$ denotes the diagonal matrix with the same diagonal as that of the indicated matrix.

As the Gaussian mutual information involves only the covariance structure, it measures in fact only the correlation (and not dependence) between the source processes. But it is still enough for source separation because it also includes *lagged correlation* and the mixture model is *linear instantaneous.* Specifically we shall be concerned with the model

$$\mathbf{X}(t)=\mathbf{A}\mathbf{S}(t) \quad (2)$$

where $\mathbf{X}(t)$ is the vectors of observation, with components $X_1(t)$, ..., $X_K(t)$ and $\mathbf{S}(t)$ is the vector of sources, with components $S_1(t)$, ..., $S_K(t)$, and $\mathbf{A}$ is a $K\times K$ non singular matrix. In the blind context, a sensible separation procedure is to find a matrix $\mathbf{B}$ minimizing $I_g[Y_1(\cdot),\ldots,Y_K(\cdot)]$ where $Y_i(\cdot)$ are the components of $\mathbf{B}\mathbf{X}(\cdot)$ and represent the reconstructed sources. The following result shows that this criterion is a contrast, in the sense of [4].

Proposition 1 *Assume that the spectral densities f_{S_k} of S_k, $k=1,\ \ldots,\ K$, are almost everywhere positive and no pair of them can be proportional. Then $I_g[Y_1(\cdot),\ldots,Y_K(\cdot)]$, where the $Y_k(\cdot)$ are as above and assumed to be not identically zero, is minimized if and only if $\mathbf{B}\mathbf{A}$ is a product of a permutation and a diagonal matrix.*

3 THE SEPARATION METHOD

As the spectral density $f_{\mathbf{Y}}$ in (1) will have to be estimated by a local average, we begin by replacing it by a smoothed version $f_{\mathbf{Y}}^M=K_M\star f_{\mathbf{Y}}$ where K_M is a 2π-periodic non negative kernel tending the Dirac comb as $M\to\infty$ and $\star$ denotes the (circular) convolution. Then, replacing the integral by a discrete approximation, we obtain the criterion

$$I_g^M[Y_1(\cdot),\ldots,Y_K(\cdot)]=\frac{1}{2L}\sum_{l=1}^{L}\{\log\det\operatorname{diag}[\mathbf{f}_{\mathbf{Y}}^M(\lambda_l)]-\log\det\mathbf{f}_{\mathbf{Y}}^M(\lambda_l)\}, \quad (3)$$

which can be shown to retain the contrast property.

Proposition 2 *Assume that $f_{S_k}^M(\lambda_l)=(K_M\star f_{S_k})(\lambda_l)$, $l=1,\ldots,L$, are positive and there is no pair j,k for which the vectors*

$$[f_{S_j}^M(\lambda_1)\ \cdots\ f_{S_j}^M(\lambda_L)] \quad \textit{and} \quad [f_{S_k}^M(\lambda_1)\ \cdots\ f_{S_k}^M(\lambda_L)]$$

are proportional. Then $I_g^M[Y_1(\cdot),\ldots,Y_K(\cdot)]$ where the Y_k are as in Proposition 1, can be zero if and only if $\mathbf{B}\mathbf{A}$ is a product of a permutation and a diagonal matrix.

Since $\mathbf{f}_{\mathbf{Y}}^M=\mathbf{B}\mathbf{f}_{\mathbf{X}}^M\mathbf{B}^{\mathrm{T}}$ where $\mathbf{f}_{\mathbf{X}}^M=K_M\star\mathbf{f}_{\mathbf{X}}$, (3) can be consistently estimated by

$$\frac{1}{2L}\sum_{l=1}^{L}\{\log\det\operatorname{diag}[\mathbf{B}\hat{\mathbf{f}}_{\mathbf{X}}^M(\lambda_l)\mathbf{B}^{\mathrm{T}}]-\log\det[\mathbf{B}\hat{\mathbf{f}}_{\mathbf{X}}^M(\lambda_l)\mathbf{B}^{\mathrm{T}}]\}. \quad (4)$$

where $\hat{\mathbf{f}}_{\mathbf{X}}^M$ is a consistent estimate for $\mathbf{f}_{\mathbf{X}}^M$. Two classes of spectral estimators have been commonly used in literature:

$$\hat{\mathbf{f}}_{\mathbf{X}}^M(\lambda)=\frac{2\pi}{N}\sum_{n=0}^{N-1}K_M\Big(\lambda-n\frac{2\pi}{N}\Big)\mathbf{I}_N\Big(n\frac{2\pi}{N}\Big) \quad (5)$$

where $\mathbf{I}_N=(2\pi N)^{-1}[\sum_{t=1}^{N}e^{-i\lambda t}\mathbf{X}(t)][\sum_{t=1}^{N}e^{i\lambda t}\mathbf{X}(t)]^{\mathrm{T}}$ is the periodogram, N being the sample size, and

$$\hat{\mathbf{f}}_{\mathbf{X}}^M(\lambda)=\frac{1}{2\pi}\sum_{u=1-N}^{N-1}k_M(u)e^{-i\lambda u}\mathbf{R}_N(u) \quad (6)$$

where $\mathbf{R}_N(\tau)=\sum_{t=1}^{N}X(t)X(t+\tau)^{\mathrm{T}}$ is the sample covariance function and $k_M(u)=\int_0^{2\pi}K_M(\lambda)e^{iu\lambda}\mathrm{d}\lambda$.

3.1 Joint approximate diagonalization algorithm

The criterion (4) can be viewed as a measure of global deviation from diagonality of the matrices $\mathbf{B}\hat{\mathbf{f}}_{\mathbf{X}}^M(\lambda_l)\mathbf{B}^{\mathrm{T}}$, $l=1$, ..., L, as $\det\mathbf{M}\le\det\operatorname{diag}(\mathbf{M})$ for any positive matrix $\mathbf{M}$, with equality if and only if $\mathbf{M}$ is diagonal (by the Hadamard inequality, see [5], exercise 15.51). Thus minimizing (4) amounts to diagonalizing jointly approximately the matrices $\hat{\mathbf{f}}_{\mathbf{X}}^M(\lambda_l),\ldots,\hat{\mathbf{f}}_{\mathbf{X}}^M(\lambda_L)$. An efficient algorithm for this purpose has been derived by the author [8] and is briefly described here. The algorithm uses the classic Jacobi approach of making successive transformations, *non orthogonal* in general, on each pair of rows of $\mathbf{B}$. Let $\mathbf{B}_{i\cdot}$ and $\mathbf{B}_{j\cdot}$ denote a pair of rows of $\mathbf{B}$, the algorithm changes them to

$$\begin{bmatrix}\mathbf{B}_{i\cdot}\\ \mathbf{B}_{j\cdot}\end{bmatrix}-\frac{2}{1+\sqrt{1-4h_{ij}h_{ji}}}\begin{bmatrix}0 & h_{ij}\\ h_{ji} & 0\end{bmatrix}\begin{bmatrix}\mathbf{B}_{i\cdot}\\ \mathbf{B}_{j\cdot}\end{bmatrix},$$

where

$$\begin{bmatrix}h_{ij}\\ h_{ji}\end{bmatrix}=\begin{bmatrix}\omega_{ij} & 1\\ 1 & \omega_{ji}\end{bmatrix}^{-1}\begin{bmatrix}g_{ij}\\ g_{ji}\end{bmatrix}$$

with

$$g_{ij}=\frac{1}{L}\sum_{l=1}^{L}\frac{\Re[\mathbf{B}\hat{\mathbf{f}}_{\mathbf{X}}^M(\lambda_l)\mathbf{B}^{\mathrm{T}}]_{ij}}{[\mathbf{B}\hat{\mathbf{f}}_{\mathbf{X}}^M(\lambda_l)\mathbf{B}^{\mathrm{T}}]_{ii}},$$

$$\omega_{ij}=\frac{1}{L}\sum_{l=1}^{L}\frac{[\mathbf{B}\hat{\mathbf{f}}_{\mathbf{X}}^M(\lambda_l)\mathbf{B}^{\mathrm{T}}]_{jj}}{[\mathbf{B}\hat{\mathbf{f}}_{\mathbf{X}}^M(\lambda_l)\mathbf{B}^{\mathrm{T}}]_{ii}}.$$

This operation is applied to all pairs of rows (which constitutes a *sweep*, then repeated again until convergence.

3.2 On-line processing

Often $k_M(u)$ is chosen of the form $k(u/M)$ for some lag window generator k with support $[-1,1]$. Then formula (6) requires only the evaluation of $\mathbf{R}_N(0)$, ..., $\mathbf{R}_N(M-1)$ which can be done on-line. But there is a better method. Decompose k_M as $k_M(u)=\sum_{v=-\infty}^{\infty}k_M^{\star 1/2}(v-u)k_M^{\star 1/2}(v)$. Then it can be checked that (5) can be written as

$$\hat{\mathbf{f}}_{\mathbf{X}}^M(\lambda)=\frac{1}{2\pi N}\sum_{t=-\infty}^{\infty}(k_M^{\star 1/2}\star\mathbf{X}_\lambda)(t)(k_M^{\star 1/2}\star\bar{\mathbf{X}}^{\mathrm{T}})(t)$$

where $\mathbf{X}_\lambda(t) = e^{i\lambda t}\mathbf{X}(t)$ if $1 \le t \le N, = 0$ otherwise. This suggests estimating $\mathbf{f}_{\mathbf{X}}^M(\lambda)$ by the output of a smoothing filter applied to $(k_M^{\star 1/2} \star \mathbf{X}_\lambda)(t)(k_M^{\star 1/2} \star \bar{\mathbf{X}}^{\mathrm{T}})(t)$. Once the estimates $\hat{\mathbf{f}}_{\mathbf{X}}^M(\lambda_1), \ldots, \hat{\mathbf{f}}_{\mathbf{X}}^M(\lambda_L)$ have been obtained (on-line), their joint approximate diagonalization can be done by applying *only one sweep* of the above algorithm, starting with the most recent diagonalizing matrix, because this matrix should be already close to the solution.

4 ASYMPTOTIC PROPERTIES OF THE ESTIMATOR

We provide here some results on the asymptotic behavior of the estimator $\hat{\mathbf{B}}$ obtained from the minimization of (4), as the sample size N goes to infinity with M and $\lambda_1, \ldots, \lambda_L$ *fixed.*

Proposition 3 *There exists a random permutation matrix* $\hat{\mathbf{P}}$ *such that the matrix* $\hat{\mathbf{C}} = \hat{\mathbf{P}}\hat{\mathbf{B}}\mathbf{A}$ *has off diagonal elements converging almost surely to zero.*

It is natural to measure of the separation quality by the ratios $\hat{C}_{ij}/\hat{C}_{ii}$, $1 \le i \ne j \le K$, where $\hat{C}_{ij}$ are the elements of $\hat{\mathbf{C}}$. It can be shown that these elements satisfy the equations

$$\frac{1}{L}\sum_{l=1}^{L} \frac{\sum_{k,m=1}^{K} \hat{C}_{ik}\hat{\mathbf{f}}_{\mathbf{S},km}^M(\lambda_l)\hat{C}_{jm}}{\sum_{k,m=1}^{K} \hat{C}_{ik}\hat{\mathbf{f}}_{\mathbf{S},km}^M(\lambda_l)\hat{C}_{im}} = 0, \quad 1 \le i \ne j \le K,$$

where $\hat{\mathbf{f}}_{\mathbf{S},km}^M$ denote the elements of $\hat{\mathbf{f}}_{\mathbf{S}}^M = \mathbf{A}^{-1}\hat{\mathbf{f}}_{\mathbf{X}}(\mathbf{A}^{-1})^{\mathrm{T}}$. By Proposition 3 and the independence of the sources, it may be further shown that the $\hat{C}_{ij}/\hat{C}_{ii}$, $1 \le i \ne j \le K$ have the same asymptotic distribution as the C_{ji}^*/C_{ii}^*, solution of

$$\omega_{ji}\frac{C_{ij}^*}{C_{ii}^*} + \frac{C_{ji}^*}{C_{jj}^*} = -\frac{1}{L}\sum_{l=1}^{L} \frac{\hat{\mathbf{f}}_{\mathbf{S},ij}^M(\lambda_l)}{f_{S_i}^M(\lambda_l)}, \quad 1 \le i \ne j \le K \quad (7)$$

where $f_{S_i}^M = \hat{\mathbf{f}}_{\mathbf{S},ii}^M = K_M \star f_{S_i}$ and

$$\omega_{ji} = \frac{1}{L}\sum_{l=1}^{L} f_{S_j}^M(\lambda_l)/f_{S_i}^M(\lambda_l).$$

Moreover, the right hand side of (7) can be rewritten as

$$\frac{1}{N}\sum_{t=1}^{N}(h_{S_i}^M \star S_i^{[N]})(t)S_j(t) \quad (8)$$

where $h_{S_i}^M(j)$ are the Fourier coefficients of

$$H_{S_i}^M(\lambda) = \frac{1}{L}\sum_{l=1}^{L} K_M(\lambda - \lambda_l)/f_{S_i}^M(\lambda_l)$$

and $S_i^{[N]}(t) = S_i[1 + (t-1) \pmod N]$ in the case of formula (5) or $S_i^{[N]}(t) = S_i(t)$ if $1 \le t \le N, = 0$ otherwise, in the case of formula (6).

The vector with components the right hand side of (8) can be shown to be asymptotically normal with mean zero and covariance matrix block diagonal with diagonal blocks

$$\frac{\mathbf{G}_{(ij)}}{N} = \frac{2\pi}{N}\int f_{S_i}(\lambda)f_{S_j}(\lambda)\begin{bmatrix} H_{S_i}^M(\lambda) \\ H_{S_j}^M(\lambda)\end{bmatrix}[H_{S_i}^M(\lambda)\ H_{S_j}^M(\lambda)]\mathrm{d}\lambda.$$

One then deduces that the vector with components $\hat{C}_{ij}/\hat{C}_{ii}$, $1 \le i \ne j \le K$ is asymptotically normal with mean zero and covariance matrix block diagonal with diagonal blocks $\mathbf{\Omega}_{(ij)}^{-1}\mathbf{G}_{(ij)}\mathbf{\Omega}_{(ij)}^{-1}/N$, where $\mathbf{\Omega}_{(ij)}$ is the 2×2 matrix with ω_{ji}, ω_{ij} on the diagonal and 1 elsewhere.

5 SOME SIMULATION EXAMPLES

We present some simulations to assess the performance of our methods. We consider the case of two sources obeying an autoregressive (AR) model of order 2 with AR polynomials having complex roots and coefficients given below.

Exp.	source 1		source 2	
	AR poles	AR coeff.	AR poles	AR coeff.
1	$.8e^{\pm i3\pi/10}$	.9405, $-.64$	$.8e^{\pm i2\pi/4}$	0, $-.64$
2	$.85e^{\pm i2\pi/5}$	.5253, $-.7225$	$.9e^{\pm i2\pi/4}$	0, $-.81$

Experiment 1 corresponds to the easy case where spectral peaks of the sources are well separate while experiment 2 corresponds to the difficult case where they are closer and also more pronounced. To see the effect of M for resolving the spectral peaks, we have plotted on the same figures 1 and 2 the true spectra f_{S_i} and their expected estimates for two choices $M = 8$ and $M = 16$ and for the Parzen kernel

$$k_M(u) = \begin{cases} 1 - 6(u/M)^2(1 - |u/M|), & 0 \le |u| < M/2 \\ 2(1 - |u/M|)^3, & M/2 \le |u| < M \\ 0 & u \ge M \end{cases}$$

These figures show that our spectral estimators have a quite large bias. However, as will be seen below, our separation method still performs reasonably well.

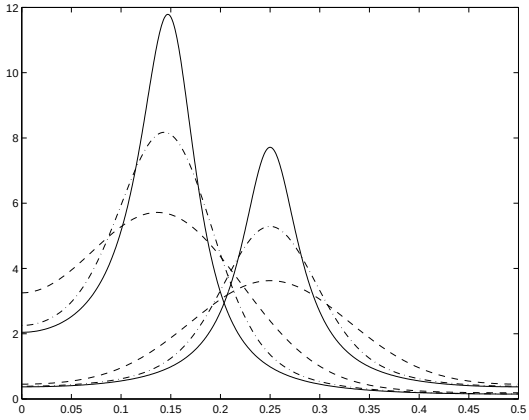

Figure 1: Spectra of the two sources in experiment 1 (solid) and their expected estimates using the Parzen kernel with parameter $M = 8$ (dash) and $M = 16$ (dot-dash)

The mixing matrix $\mathbf{A}$ is taken to be $\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$; but, as it can be easily shown, the performance of our method is independent of $\mathbf{A}$, only the number of iterations required by the algorithm to converge may be affected.

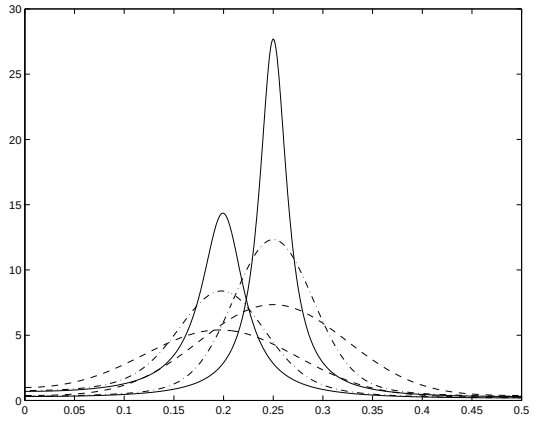

Figure 2: Spectra of the two sources in experiment 2 (solid) and their expected estimates using the Parzen kernel with parameter $M = 8$ (dash) and $M = 16$ (dot-dash)

The simulation results are reported in table 1. We have computed the empirical covariance matrix (without centering) of the random vector $[\hat{C}_{12}/\hat{C}_{11} \quad \hat{C}_{21}/\hat{C}_{22}/]^{\mathrm{T}}$ based on 1000 simulation trials. For comparison, we multiply these matrices by the sample size $N = 256$ and also list the corresponding theoretical asymptotic covariance matrix (computed with two choices of L) and the CR bounds. The average (across the trials) of the number of needed iterations is also reported.

It can be seen that the asymptotic covariance matrix is very close to the CR bound, especially for the higher value of M ($= 16$) and for the experiment 1. Even in worst case (experiment 2, $M = 8$) it is still not very far from the CR bound, although the spectral bias is quite large, as seen in figure 2. Also, the use of a large value of L ($= 512$) instead of $L = M$ is seen to have no appreciable. Note that by taking λ_l of the form $\pm(2l-1)\pi/L$, $l = 1, \ldots, L/2$, only $L/2$ matrices needed to be diagonalized, by symmetry. Turning to the empirical covariance matrices of the estimators, one can see that they are somewhat higher than the theoretical asymptotic values. This may be attributed to the finite sampling effect, since the CR bound, in nonlinear estimation problems, is a strict bound and can be attained only asymptotically. Still the performance of our estimator relative to the CR bound is quite respectable (at sample size 256). Finally, one can see that our joint diagonalization algorithm converges quite fast.

	Experiment 1	
	$M = 8$	$M = 16$
CR bound	$\begin{bmatrix} .4569 & -.1414 \\ -.1414 & .3531 \end{bmatrix}$	
Asymptotic ($L = 512$)	$\begin{bmatrix} .4670 & -.1425 \\ -.1425 & .3669 \end{bmatrix}$	$\begin{bmatrix} .4597 & -.1435 \\ -.1435 & .3554 \end{bmatrix}$
($L = M$)	$\begin{bmatrix} .4700 & -.1407 \\ -.1407 & .3641 \end{bmatrix}$	$\begin{bmatrix} .4597 & -.1435 \\ -.1435 & .3555 \end{bmatrix}$
Empirical ($L = M$)	$\begin{bmatrix} .4958 & -.1505 \\ -.1505 & .3691 \end{bmatrix}$	$\begin{bmatrix} .4845 & -.1568 \\ -.1568 & .3696 \end{bmatrix}$
# of iter.	3.4	4.3
	Experiment 2	
	$M = 8$	$M = 16$
CR bound	$\begin{bmatrix} .8045 & -.4680 \\ -.4680 & .8541 \end{bmatrix}$	
Asymptotic ($L = 512$)	$\begin{bmatrix} .9030 & -.4714 \\ -.4714 & .9288 \end{bmatrix}$	$\begin{bmatrix} .8359 & -.4907 \\ -.4907 & .8849 \end{bmatrix}$
($L = M$)	$\begin{bmatrix} .9332 & -.4854 \\ -.4854 & .9172 \end{bmatrix}$	$\begin{bmatrix} .8337 & -.4856 \\ -.4856 & .8786 \end{bmatrix}$
Empirical ($L = M$)	$\begin{bmatrix} 1.0140 & -.5195 \\ -.5195 & .9522 \end{bmatrix}$	$\begin{bmatrix} .9140 & -.5387 \\ -.5387 & .9488 \end{bmatrix}$
# of iter.	3.6	4.6

Table 1: The CR bound and asymptotic covariance matrix for $[\hat{C}_{12}/\hat{C}_{11} \quad \hat{C}_{21}/\hat{C}_{22}]^{\mathrm{T}}$ and its empirical estimates from 1000 simulation trials, all multiplied by the sample size $N = 256$. The average, across the trials, of the number of needed iterations is listed in the last row.

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*, **45**, 2, 434–444, 1997

[2] J.-F. Cardoso, "On the performance of orthogonal source separation technique". In *Proc. EUSIPCO*, Edinburg, Sept. 1994, 776–779.

[3] J.-F. Cardoso, "Blind Signal Separation: Statistical principles", *Proc. IEEE*, **90**, 2009–2026. Aug. 1998.

[4] P. Comon, "Independence components analysis, a new concept". *Signal Processing*, **36**, 3, 287–314, 1994.

[5] B. Noble and J. W. Daniel, *Applied linear Algebra.* New-Jersey: Prentice Hall, Englewood Cliff, 1977.

[6] D. T. Pham, "Blind Separation of Instantaneous Mixture of Sources via an Independent Component Analysis." *IEEE Trans. Signal Processing*, **44**, 11, 2768–2779, 1996.

[7] D. T. Pham, "Mutual information approach to blind separation for stationary sources." In *Proceeding of ICA'99 Conference*, 215 – 220. Aussois, January 1999.

[8] D. T. Pham, "Joint approximate diagonalization of positive definite Hermitian matrices. Submitted to SIAM Journal on Scientific Computing. Available as Technical report LMC/IMAG, http://www-lmc.imag.fr/lmc-sms/Dinh-Tuan.Pham/jadiag/jadiag.ps.gz, Apr. 1999.

[9] D. T. Pham and Ph. Garat, "Blind separation of mixtures of independent sources through a quasi maximum likelihood approach". *IEEE Trans. Signal Processing*, **45**, 7, 1712–1725, 1007, 1997

[10] L. Tong and V. Soon, Y. F. Huang and R. Liu, "AMUSE: A new blind identification algorithm". In *Proc. ISCAS*, 1990