# Blind separation of mixtures of piecewise AR(1) processes and model mismatch

Petr Tichavský[1], Ondřej Šembera[1] and Zbyněk Koldovský[1,2]

[1] Institute of Information Theory and Automation of the CAS,
Prague, Czech Republic,
tichavsk@utia.cas.cz, sembeond@fjfi.cvut.cz,
[2] Faculty of Mechatronics, Informatics, and Interdisciplinary Studies,
Technical University of Liberec, Liberec, Czech Republic

**Abstract.** Modeling real-world acoustic signals and namely speech signals as piecewise stationary random processes is a possible approach to blind separation of linear mixtures of such signals. In this paper, the piecewise AR(1) modeling is studied and is compared to the more common piecewise AR(0) modeling, which is known under the names Block Gaussian SEParation (BGSEP) and Block Gaussian Likelihood (BGL). The separation based on the AR(0) modeling uses an approximate joint diagonalization (AJD) of covariance matrices of the mixture with lag 0, computed at epochs (intervals) of stationarity of the separated signals. The separation based on the AR(1) modeling uses the covariances of lag 0 and covariances of lag 1 jointly. For this model, we derive an approximate Cramér-Rao lower bound on the separation accuracy for estimation based on the full set of the statistics (covariance matrices of lag 0 and lag 1) and covariance matrices with lag 0 only. The bounds show the condition when AR(1) modeling leads to significantly improved separation accuracy.

**Keywords:** Autoregressive processes, Cramér-Rao Bound, Blind source separation

## 1 Introduction

Blind source separation has found applications namely in biomedical signal processing, for separating signals of interest from unwanted parasitic signals and noises, and in acoustical signal processing [6]. Modeling real-world acoustic signals and namely speech signals as piecewise stationary random processes is a possible approach to blind separation of linear mixtures of such signals. It appears that many times (depending on properties of the separated signals), methods utilizing nonstationarity of the separated signals outperform the more classical

[2] This work was supported by The Czech Science Foundation through Project No. 14-13713S.

methods based on non-Gaussianity of the separated signals, or perform equally well but with much lower computational complexity [5]. The methods using signal nonstationarity divide the received signals (mixtures) to epochs, in each epoch the signals are modeled as stationary, but properties of the signals (namely their power) are assumed to change significantly in different epochs [1, 7–9].

The separation is called determined or overdetermined, if the number of the available mixtures is higher or equal to the number of the sources, and is called underdetermined otherwise. The latter case is studied in [4]. In this paper we focus on the squared mixtures, where the number of the mixtures is equal to the number of the sources.

The simplest nonstationarity-based separation methods use only covariance matrices with lag 0. The mixing/demixing matrix can be found through an approximate joint diagonalization of these matrices [1]. This method can be statistically efficient (attaining a Cramér-Rao lower bound, CRB) [2, 8], if the separated data obey the assumed model, i.e. when the signals are i.i.d. in all epochs. Real-world signals such as speech signals rarely obey the condition. Our experiments with natural speech signals sampled at 16 kHz show that the correlation between two consecutive samples of the signals is typically $0.75 - 0.95$. This fact indicates that the separation methods using only the covariance matrices with lag 0 may not be optimal, and more accurate modeling of the separated signals may increase accuracy of the separation.

A method called Block-AutoRegressive Blind Identification (BARBI) [11] uses an autoregressive model of a general order $n$ in each epoch of the source signals. We refer to the method as BARBI($n$). The number of the estimated parameters grows with increasing model order $n$ and the method seems to suffer of overfitting, if $n > 2$. In this paper we provide a theoretical justification for improved performance of BARBI(1) compared to BARBI(0) through the CRB analysis.

## 2 Data Model

Consider linear instantaneous square mixing model

$$X_t = AZ_t, \tag{1}$$

where $Z_t$ denotes a single time instance of the input signals, $A \in \mathbb{R}^{d\times d}$ is a mixing matrix and $X_t \in \mathbb{R}$ is a time instance of the resulting mixtures. The input signals are modeled by mutually independent piecewise stationary processes. We divide the data into $M$ epochs of the length $T$ and assume that on each epoch the $i-$th signal $z_{it}$ takes a form of order one autoregressive process

$$z_{it} = -\rho_{im} z_{it-1} + \sigma_{im} w_{it}, \tag{2}$$

for $t = (m-1)T + 1, \dots mT$, where $w_{it}$ is a Gaussian white noise with zero mean and unit variance, $\rho_{im}$ is an autoregressive coefficient corresponding to the $i-$th input signal and the $m-$th epoch, and white noise sequences satisfy independence relation

$$\mathbf{Cov}\left[w_{it}, w_{jt'}\right] = \delta_{ij}\delta_{tt'}.$$

The covariance of the input vector $Z_t$ in the $m-$th epoch with lag 0 is then given by

$$D_m = \mathbf{Cov}\left[Z_t, Z_t\right] = \mathrm{diag}(d_{1m}, d_{2m}, \ldots d_{dm}), \tag{3}$$

where

$$d_{im} = \frac{\sigma_{im}^2}{1-\rho_{im}^2}.$$

Covariance of the mixture $X_t$ with lag 1 in the $m-$th epoch is

$$\mathbf{Cov}\left[Z_t, Z_{t+1}\right] = D_m Q_m \tag{4}$$

where $Q_m = -\,\mathrm{diag}(\rho_{1m}, \rho_{2m}, \ldots \rho_{dm})$ is a diagonal matrix of the $m-$th epoch autoregressive coefficients. The covariance matrices of the mixture $X_t$ in the $m-$th epoch with lag 0 and with lag 1

$$R_m = A D_m A^T, \qquad S_m = A D_m Q_m A^T \tag{5}$$

are estimated from the data as

$$\hat{R}_m = \frac{1}{T} \sum_{t=(m-1)T+1}^{mT} X_t X_t^T, \qquad \hat{S}_m = \frac{1}{T-1} \sum_{t=(m-1)T+1}^{mT-1} X_{t+1} X_t^T\ . \tag{6}$$

The vector of the unknown parameters is

$$\theta = [\mathrm{vec}(A)^T; \mathrm{vec}(D)^T; \mathrm{vec}(Q)^T]^T \tag{7}$$

where $D$ and $Q$ are $d \times M$ matrices with elements $d_{im}$ and $\rho_{im}$, $i = 1, \ldots, d$, $m = 1, \ldots, M$, respectively. Matrix $A$ is the main parameter of interest and $D$, $Q$ are nuisance parameters. Since each change in scale of the signals can be compensated by adequate change of the mixing matrix, the parameter $D$ is constrained by the condition $\sum_m d_{im} = 1$ for all $i = 1, \ldots, d$. It means that the sum of the variances of each signal over all epochs is 1. Indeed, there are inequality constraints $0 \leq d_{im}$ and $-1 < \rho_{im} < 1$ so that all signals in all epochs are *stable* AR processes.

## 3 Cramér-Rao Bound

The Cramér-Rao Bound is defined as an inverse of the Fisher information matrix. We shall assume, for simplicity, that the available data are Gaussian. It holds that for normally distributed data with a mean $\mu(\theta)$ and covariance matrix $C(\theta)$ the Fisher information matrix has elements

$$F_{\theta_i\theta_j} = \left(\frac{\partial\mu}{\partial\theta_i}\right)^{\top} C^{-1} \left(\frac{\partial\mu}{\partial\theta_j}\right) + \frac{1}{2}\,\mathrm{tr}\left(C^{-1}\frac{\partial C}{\partial\theta_i}C^{-1}\frac{\partial C}{\partial\theta_j}\right)\ . \tag{8}$$

In our case, the data have zero mean and only its covariance matrix $C$ depends on the estimated parameter. In particular,

$$C = \mathrm{Cov}([X_1, \ldots, X_{MT}]) = \mathrm{blockdiag}(C_1, \ldots, C_M) \tag{9}$$

where

$$C_m = \mathrm{btoeplitz}(AD_m A^T, AD_m Q_m A^T, \ldots, AD_m Q_m^{(T-1)} A^T) \ . \tag{10}$$

$C_m$ is the covariance matrix of the data in the $m-$th epoch, it is a symmetric block-Toeplitz matrix with the displayed first block-row.

Now, the CRB on $\mathrm{vec}(A)$ is given as the left-upper corner submatrix of $F^{-1}$ of the size $d^2 \times d^2$, obeys

$$\mathrm{CRB}(\mathrm{vec}(A)) = \mathrm{CRB}_A = (A^{-1} \otimes I)\mathrm{CRB}_I(A^{-T} \otimes I) \ . \tag{11}$$

$\mathrm{CRB}_I$ is esentially (i.e. after a suitable re-ordering its columns and rows) block diagonal, with diagonal blocks of size $1 \times 1$ and $2 \times 2$,

$$\mathrm{CRB}_I(A_{kk}) = \frac{1}{T} \tag{12}$$

for $k = 1, \ldots, d$, and

$$\mathrm{CRB}_I([A_{k\ell}, A_{\ell k}]) = \frac{1}{MT} \frac{1}{\phi_{k\ell}\phi_{\ell k} - 1} \begin{bmatrix} \phi_{k\ell} & -1 \\ -1 & \phi_{\ell k} \end{bmatrix} \tag{13}$$

for $k, \ell = 1, \ldots, d$, $k \neq \ell$, where [11]

$$\phi_{k\ell} = \frac{1}{M} \sum_{m=1}^{M} \frac{d_{km}}{d_{\ell m}} \frac{1 - 2\rho_{km}\rho_{\ell m} + \rho_{\ell m}^2}{1 - \rho_{\ell m}^2} \ . \tag{14}$$

Note that in the special case of all autoregressive parameters identical, $\rho_{km} = \rho$ for $k = 1, \ldots, d$, $m = 1, \ldots, M$, the resultant CRB expressions are independent of $\rho$.

### 3.1 CRB for estimates based on the statistics

In this subsection, we investigate the maximum possible accuracy of the separation using only the statistics $\{\hat{R}_m\}$ and $\{\hat{R}_m, \hat{S}_m\}$, respectively. Thanks to the central limit theorem it holds that for $T \to \infty$ these statistics have asymptotically normal distribution with the asymptotic mean equal to the theoretical covariances $\{R_m\}$ and $\{R_m, S_m\}$, respectively, and have asymptotical covariance of errors proportional to $\frac{1}{T}$. The CRB for the estimates based on the statistics means computing the information content about the estimated parameter $\theta$ in the "concentrated" data $\{\hat{R}_m\}$ and $\{\hat{R}_m, \hat{S}_m\}$, assuming that the noise in the "concentrated" data is exactly zero mean and and has exactly Gaussian distribution with the covariance structure that follows from analysis of the true statistics.

For computing the CRB as the inverse of the Fisher information matrix we will use the general formula (8) again. In this case, both the mean $\mu$ and the covariance matrix $C$ depend on the estimated parameter. A detailed computation, not included here due to lack of space, shows that only the former term in (8) is dominant, has asymptotic order $O(T)$ for large $T$, and the latter term is negligible, having the order $O(1)$ only. The $O(1)$ terms will be neglected with respect to the leading term proportional to $T$. The asymptotic CRB is inversely proportional to $T$.

The concentrated data using the covariances with only lag 0 denoted $\hat{Y}_0$ are composed of $\mathcal{L}(\hat{R}_m)$ for $m = 1, \ldots, M$, where $\mathcal{L}(R)$ is the vector of elements of a lower triangular part of a matrix $R$,

$$\mathcal{L}(R) = [R_{11}, R_{21} \ldots R_{d1}, R_{22}, R_{32} \ldots R_{d2}, R_{33} \ldots R_{dd}]^T \ . \tag{15}$$

The concentrated data using the covariances with lag 0 and 1 denoted $\hat{Y}_{0+1}$ will be composed of $\mathcal{L}(\hat{R}_m)$ and $\mathcal{L}((\hat{S}_m + \hat{S}_m^T)/2)$ for $m = 1, \ldots, M$. Note that while $S_m$ is symmetric, its sample estimate $\hat{S}_m$ may not be symmetric and thus we symmetrize it.

The covariance matrices of $\hat{Y}_0$ and $\hat{Y}_{0+1}$ can be computed as functions of parameter $\theta$ in (7). Again, they are block diagonal, having $M$ blocks, because data in individual epochs and also the sample covariance matrices in them are mutually statistically independent. A straightforward but lengthy computation leads to the result that the asymptotic CRB for estimates based on the statistics $\hat{Y}_{0+1}$, denoted $\mathrm{CRB}^{(0+1)}(A)$ are identical to those in (13). It follows that $\hat{Y}_{0+1}$ is asymptotically sufficient. CRB for estimates based on $\hat{Y}_0$, denoted $\mathrm{CRB}^{(0)}(A)$, is higher, sometimes significantly. In particular,

$$\mathrm{CRB}^{(0)}(\mathrm{vec}\, A)) = \mathrm{CRB}_A^{(0)} = (A^{-1} \otimes I)\mathrm{CRB}_I^{(0)}(A^{-T} \otimes I) \tag{16}$$

where $\mathrm{CRB}_I^{(0)}$ is block diagonal and independent of $A$ again, and

$$\mathrm{CRB}_I^{(0)}([A_{k\ell}, A_{\ell k}]) = \frac{1}{MT} \frac{1}{\varphi_{k\ell}\varphi_{\ell k} - \omega_{k\ell}^2} \begin{bmatrix} \varphi_{k\ell} & -\omega_{k\ell} \\ -\omega_{k\ell} & \varphi_{\ell k} \end{bmatrix} \tag{17}$$

for $k, \ell = 1, \ldots, d$, $k \neq \ell$, with

$$\varphi_{k\ell} = \frac{1}{M} \sum_{m=1}^{M} \frac{d_{km}}{d_{\ell m}} \frac{1 - \rho_{km}\rho_{\ell m}}{1 + \rho_{\ell m}^2}, \qquad \omega_{k\ell} = \frac{1}{M} \sum_{m=1}^{M} \frac{1 - \rho_{km}\rho_{\ell m}}{1 + \rho_{\ell m}^2} \ . \tag{18}$$

In the special case $\rho_{in} = \rho$ for all $i = 1, \ldots, d$, $m = 1, \ldots, M$, it holds

$$\mathrm{CRB}^{(0)}(A) = \mathrm{CRB}(A) \frac{1 + \rho^2}{1 - \rho^2} \ . \tag{19}$$

If $\rho$ is close to $\pm 1$, the difference is significant.

## 4 Estimating $A$

From (10) it follows that $A^{-1}$ can be sought as a matrix that jointly diagonalizes the matrices $\hat{R}_m$ and $\hat{S}_m$, $m = 1, \ldots, M$. The ordinary (unweighted) approximate joint diagonalization algorithms as UWEDGE [1] produce consistent but not optimal estimates of $A$. The asymptotically optimum estimate of $\theta$ can be found by minimizing the expression

$$\hat{\theta} = \mathrm{argmin}_\theta (\hat{Y}_{0+1} - Y_{0+1}(\theta))^T [\mathrm{Cov}(\hat{Y}_{0+1})]^{-1} (\hat{Y}_{0+1} - Y_{0+1}(\theta)) \ . \qquad (20)$$

The matrix $C_{0+1}(\theta) = \mathrm{Cov}(\hat{Y}_{0+1})$ is a function of the unknown parameter $\theta$. In practice, $C_{0+1}(\theta)$ can be replaced by $C_{0+1}(\hat{\theta}_c)$, where $\hat{\theta}_c$ is a consistent estimate of $\theta$, to achieve an asymptotically optimum estimate. Note that $C_{0+1}(\theta)$ is nearly block diagonal if its columns and rows are appropriately sorted and $A \approx I$. The weighted AJD algorithm WEDGE[1], and also BARBI[11] estimate a demixing matrix $V$. Let $V^{[i]}$ be an estimate of $V = A^{-1}$ from the $i-$th iteration. Then, the partially demixed covariance matrices are given as $\hat{R}_m^{[i]} = V^{[i]} \hat{R}_m V^{[i]T}$ and $\hat{S}_m^{[i]} = V^{[i]} \hat{S}_m V^{[i]T}$. These matrices are used to estimate parameters of the separated signals, i.e. $d_{jm}^{[i]} = (\hat{R}_m^{[i]})_{jj}$ and $\rho_{jm}^{[i]} = -(\hat{S}_m^{[i]})_{jj}/(\hat{R}_m^{[i]})_{jj}$ where $(X)_{jj}$ means the $(j,j)-$th element of matrix $X$.

The main iteration of WEDGE is

$$V^{[i+1]} = [A^{[i]}]^{-1} V^{[i]},$$

where the diagonal elements of $A^{[i]}$ are set to 1, and the off-diagonal elements of $A^{[i]}$ obey the $2 \times 2$ linear systems

$$\begin{bmatrix} A_{k\ell}^{[i]} \\ A_{\ell k}^{[i]} \end{bmatrix} = \left\{ \sum_{m=1}^{M} \begin{bmatrix} \hat{p}_{\ell\ell m}^T W_{k\ell m} \hat{p}_{\ell\ell m} & \hat{p}_{kkm}^T W_{k\ell m} \hat{p}_{\ell\ell m} \\ \hat{p}_{kkm}^T W_{k\ell m} \hat{p}_{\ell\ell m} & \hat{p}_{kkm}^T W_{k\ell m} \hat{p}_{kkm} \end{bmatrix} \right\}^{-1} \sum_{m=1}^{M} \begin{bmatrix} \hat{p}_{\ell\ell m}^T W_{k\ell m} \hat{p}_{k\ell m} \\ \hat{p}_{kkm}^T W_{k\ell m} \hat{p}_{k\ell m} \end{bmatrix}, \qquad (21)$$

where $\hat{p}_{k\ell m} = [(\hat{R}_m^{[i]})_{k\ell}, (\hat{S}_m^{[i]})_{k\ell}]$ and $W_{k\ell m}$ should be proportional to the inverse of a $2 \times 2$ covariance matrix of $\hat{p}_{k\ell m}$ for $k, \ell = 1, \ldots, d$, $k \neq \ell$. We use the choice

$$W_{k\ell m}^{-1} = \frac{d_{km} d_{\ell m}}{1 - \rho_{km}\rho_{\ell m}} \begin{bmatrix} 1 + \rho_{km}\rho_{\ell m} & -\rho_{km} - \rho_{\ell m} \\ -\rho_{km} - \rho_{\ell m} & (1 + (\rho_{km} + \rho_{\ell m})^2 - \rho_{km}^2 \rho_{\ell m}^2)/2 \end{bmatrix} \ . \qquad (22)$$

In BARBI, the relation (21) is replaced by

$$\begin{bmatrix} \widehat{A}_{k\ell}^{[i]} \\ \widehat{A}_{\ell k}^{[i]} \end{bmatrix} = \left\{ \sum_{m=1}^{M} \begin{bmatrix} \hat{p}_{\ell\ell m}^T q_{km} & \hat{p}_{kkm}^T q_{km} \\ \hat{p}_{kkm}^T q_{km} & \hat{p}_{kkm}^T q_{\ell m} \end{bmatrix} \right\}^{-1} \sum_{m=1}^{M} \begin{bmatrix} q_{km}^T \hat{p}_{k\ell m} \\ q_{\ell m}^T \hat{p}_{k\ell m} \end{bmatrix}, \qquad (23)$$

where $q_{km} = W_{k\ell m} p_{\ell\ell m}$, and in the case of the AR order 1 it reads

$$q_{km} = \frac{1}{2 d_{km}(1 - \rho_{km}^2)} \begin{bmatrix} 1 + \rho_{km}^2 \\ -2\rho_{km} \end{bmatrix} \ . \qquad (24)$$

## 5 Simulations

In the first simulation we consider a mixture of three piecewise AR(1) signals. The signals are composed of $M = 10$ epochs, each of the length $T = 100$. The signals have the same AR coefficient $\rho$ in all epochs. The variances of the signals are increasing, $1, 2, ..., 10$, decreasing $10, 9, \ldots, 1$ and constant $5, \ldots, 5$, respectively, in the 10 epochs. We mix the signals using a random orthogonal (for simplicity) mixing matrix and demix them by BARBI(0) and BARBI(1) algorithms. The resultant average interference-to-signal ratios (ISR) obtained in 100 independent trials and corresponding CRB and $\text{CRB}^{(0)}$ are plotted as function of $\rho$ in Fig.1. We can see that BARBI(1) is nearly statistically efficient

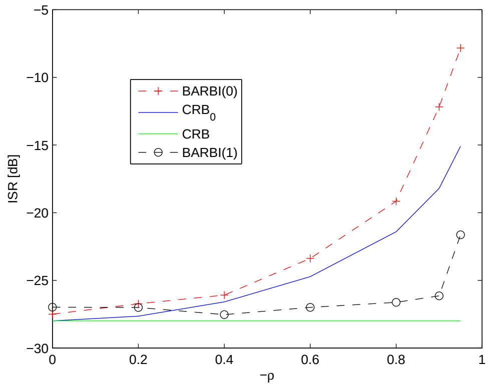


**Fig. 1.** Average ISR for separation of a mixture of artificial piecewise AR(1) signals achieved by BARBI(0) and BARBI(1) and corresponding CRBs versus the AR coefficient.

unless $\rho$ is in a vicinity of $-1$. BARBI(0) does not achieve the $\text{CRB}^{(0)}$ except for $\rho$ close to zero, but it follows the trend of $\text{CRB}^{(0)}$.

In the second simulation, we consider a mixture of 16 natural speech signals sampled at 16 kHz of the total length of 8.375 s, taken from the database in [4]. The average correlation between two consecutive samples in these signals is from 0.65 to 0.95, and the overall average is 0.81. Average ISR achieved by BARBI(0) and BARBI(1) versus the number of epochs is shown in Fig. 2.

## 6 Conclusions

We have proved that in blind separation of natural signals, piecewise AR(1) modeling represented by the algorithm BARBI(1) gives significantly improved separation accuracy if the sample lag-1 correlation of the original signals is close to 1. We plan to extend these results to underdetermined mixtures.

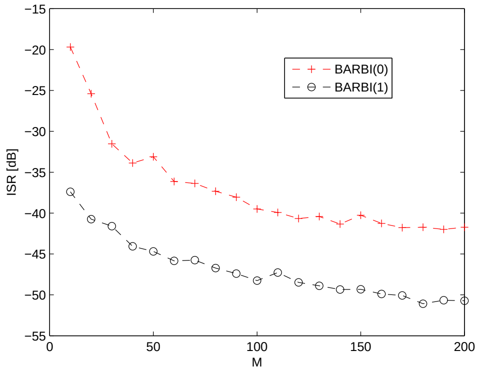


**Fig. 2.** Average ISR for separation of a mixture 16 natural speech signals achieved by BARBI(0) and BARBI(1) versus the number of epochs.

## References

1. Tichavský P., Yeredor A.: Fast approximate joint diagonalization incorporating weight matrices, IEEE Tr. Signal Process. 57, 878–891 (2009)
2. Doron E., Yeredor A., Tichavský P.: Cramér-Rao-induced bound for blind separation of stationary parametric Gaussian sources , IEEE Signal Process. Letters 14, 417–420 (2007)
3. Chabriel G., Kleinsteuber M., Moreau E., Shen H., Tichavský P., Yeredor A.: Joint matrices decompositions and blind source separation, IEEE Signal Processing Magazine 31, 34–43 (2014)
4. Tichavský P., Koldovský Z.: Weight adjusted tensor method for blind separation of underdetermined mixtures of nonstationary sources, IEEE Tr. Signal Process. 59, 1037–1047 (2011)
5. Tichavský P., Koldovský Z.: Fast and accurate methods of independent component analysis: A survey, Kybernetika 47, 426–438 (2011)
6. Koldovský Z., Tichavský P.: Time-domain blind separation of audio sources on the basis of a complete ICA decomposition of an observation space, IEEE Tr. Audio Speech and Language Process. 19, 406–416 (2011)
7. Pham D.-T., Garat P.: Blind separation of mixtures of independent sources through a quasi maximum likelihood approach, IEEE Tr. Signal Process. 45, 1712–1725 (1997)
8. Pham D.-T.: Blind separation of instantaneous mixture of sources via the Gaussian mutual information criterion, Signal Process. 81, 855-870 (2001)
9. Dégerine S., Zaïdi A.: Separation of an instantaneous mixture of Gaussian autoregressive sources by the exact maximum likelihood approach, IEEE Tr. Signal Process. 52, 1492-1512 (2004)
10. Koldovský Z. and Tichavský P.: A comparison of independent component and independent subspace analysis algorithms, Proc. EUSIPCO 2009, Glasgow, Scotland, 1447–1451 (2009).
11. Tichavský P., Yeredor A., and Koldovský Z.: A fast asymptotically efficient algorithm for blind separation of a linear mixture of block-wise stationary autoregressive processes, Proc. ICASSP 2009, Taipei, Taiwan, 3133–3136 (2009).