Weight adjusted tensor method for blind separation of underdetermined mixtures of nonstationary sources: Abstract

Petr Tichavský[1] and **Zbyněk Koldovský**[1,2]

(The whole paper is submitted for publication elsewhere.)

Abstract

We propose a novel algorithm to blindly separate an instantaneous linear underdetermined mixture of nonstationary sources. The separation is based on the assumption that the sources are piecewise stationary with a different variance in each block. It proceeds in two steps: (1) estimating the mixing matrix, and (2) computing the optimum beamformer in each block to maximize the signal-to-interference ratio of each separated signal with respect to the remaining signals. In simulations, performance of the algorithm is successfully tested on blind separation of 16 speech signals from 9 linear instantaneous mixtures of these signals.

I. The proposed method

Assume that the received signal $\mathbf{X}$ can be written as

$$\mathbf{X} = \mathbf{A}\mathbf{S} \qquad (d \times N) \tag{1}$$

where the mixing matrix $\mathbf{A}$ has the size $d \times r$, $r > d$, and the source matrix $\mathbf{S}$ contains r independent sources, stored in rows. Next, assume that the signals can be partitioned into M blocks (epochs). In each block, the separated sources have zero mean and a fixed variance. Let D_{kf} – that is, the (k, f)th element of a matrix $\mathbf{D}$ – denote the variance of the fth signal in the kth block. The length of each block is $N_1 = N/M$ and we assume, for simplicity, that it is an integer. Let $\mathbf{R}_k$ denote the theoretical covariance matrix of the mixture in the kth block. Then

$$\mathbf{R}_k = \sum_{f=1}^{r} D_{kf} \mathbf{A}_{:,f} \mathbf{A}_{:,f}^T \tag{2}$$

where $\mathbf{A}_{:,f}$ is the fth column of the matrix $\mathbf{A}$. The set of the M covariance matrices represents a three way tensor of the dimension (M, d, d)

$$\mathcal{R}_{kij}(\boldsymbol{\theta}) = \sum_{f=1}^{r} D_{kf} A_{if} A_{jf} \ . \tag{3}$$

Here, $\boldsymbol{\theta}$ represents a parameter vector that consists of all elements of the matrices $\mathbf{A}$ and $\mathbf{D}$. The tensor is written as a sum of r rank-one tensors and by definition, the tensor rank is equal at most r. There are three modes, two of which are equal to $\mathbf{A}$, and the third mode, $\mathbf{D}$, has nonnegative elements.

Traditional tensor decomposition methods seek for the decomposition of the tensor by minimizing the mean square fit

$$\mathcal{Q}_1(\boldsymbol{\theta}) = \|\widehat{\mathcal{R}} - \mathcal{R}(\boldsymbol{\theta})\|^2 = \sum_{i,j,k} (\widehat{\mathcal{R}}_{kij} - \mathcal{R}_{kij}(\boldsymbol{\theta}))^2 \ . \tag{4}$$

[01]Institute of Information Theory and Automation, Pod vodárenskou věží 4, P.O.Box 18,182 08 Prague 8, Czech Republic. E-mail: tichavsk@utia.cas.cz

[2]Faculty of Mechatronic and Interdisciplinary Studies, Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic. E-mail: zbynek.koldovsky@tul.cz.

However, in our statistical model, the estimated tensor elements are mutually correlated, and the minimization of the criterion in (4) is not statistically optimum. We are able to prove that the asymptotically optimum criterion which should replace (4) is

$$\mathcal{Q}_2(\boldsymbol{\theta}) = \sum_{k=1}^{M} \mathrm{tr}[\widehat{\mathbf{R}}_k^{-1}(\widehat{\mathbf{R}}_k - \mathbf{R}_k(\boldsymbol{\theta}))\widehat{\mathbf{R}}_k^{-1}(\widehat{\mathbf{R}}_k - \mathbf{R}_k(\boldsymbol{\theta}))] \quad (5)$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix, assuming that the covariance matrices $\mathbf{R}_k$ are invertible. More generally, we can write

$$\mathcal{Q}_3(\boldsymbol{\theta}) = \sum_{k=1}^{M} \mathrm{tr}[\mathbf{C}_k(\widehat{\mathbf{R}}_k - \mathbf{R}_k(\boldsymbol{\theta}))\mathbf{C}_k(\widehat{\mathbf{R}}_k - \mathbf{R}_k(\boldsymbol{\theta}))] \quad (6)$$

where the matrices $\mathbf{C}_k$ are chosen as $\mathbf{C}_k = (\widehat{\mathbf{R}}_k + \varepsilon\mathbf{I})^{-1}$ with a suitable small positive constant ε to maintain regularity of the criterion under all circumstances. Note that in the special case $\mathbf{C}_1 = \ldots = \mathbf{C}_M = \mathbf{I}$ the criteria (4) and (6) coincide.

In order to maintain nonnegativity of the factor $\mathbf{D}$ in the tensor decomposition in (3), we suggest to follow the method from [2] and propose to minimize an augmented criterion

$$\mathcal{Q}_4(\boldsymbol{\theta}) = \mathcal{Q}_3(\boldsymbol{\theta}) + \alpha \sum_{f=1}^{r} \left[\sum_{k=1}^{M} \log D_{kf} + M \log\left(\sum_{k=1}^{d} A_{kf}^2 \right) \right] . \quad (7)$$

Here, α is a positive parameter which starts from an initial value (e.g., 1) and is decreased gradually towards zero during the optimization. The criterion in (7) has the advantage that if a column in mode $\mathbf{A}$ is multiplied by a constant and the corresponding column in $\mathbf{D}$ is divided by the same constant squared, the criterion is not affected.

Optimization of the criterion in (7) can proceed using the following steps.

1) Initialize the algorithm by the outcome of SOBIUM by Lathauwer [1].
2) Make the elements of the factor matrix $\mathbf{D}$ positive (take an absolute value or add a constant).
3) Iterate (until convergence is achieved) the Levenberg-Marquardt (LM) algorithm

$$\Delta\boldsymbol{\theta} = (\boldsymbol{\Psi} + \mu\mathbf{I})^{-1}\boldsymbol{\xi}$$

where $\boldsymbol{\Psi}$ is the Hessian of the criterion (7), $\boldsymbol{\xi}$ is its gradient, and μ is a positive parameter that is gradually modified by the technique. After each five steps of the LM algorithm, decrease ε and α to one half of their current respective values.

Once the model parameters $\mathbf{A}$ and $\mathbf{D}$ are estimated, the individual signals in $\mathbf{S}$ are estimated using a MVDR beamformer in each block separately.

II. Simulations

A. *Artificial data obeying the model*

Four artificial signals of the length $N = 10000$ partitioned into $M = 10$ epochs (blocks) of equal length were generated as independent zero mean Gaussian random variables with variances D_{mk}, where m is the index of the epoch and k is the index of the signal. The variances versus m and k were taken as variables in Fig. 1 . The variances were normalized so that $\sum_m D_{mk} = 1$ for all k. Note that three of the variances were set close to zero at some epoch, but it never happens that all but one signal is active, as some other separation methods assume.

The data were mixed in three observed channels via a mixing matrix with columns $\mathbf{A}_{:,k} = [1, \cos(\phi_k), \sin(\phi_k)]^T$ where $(\phi_1, \phi_2, \phi_3, \phi_4) = (0, \pi/4, \pi/2, 3\pi/4)$.

The data were generated repeatedly in 500 independent trials, mixed together and analyzed by the proposed algorithm. The outcome of the algorithm was sorted to fit the orders and the signs of the original

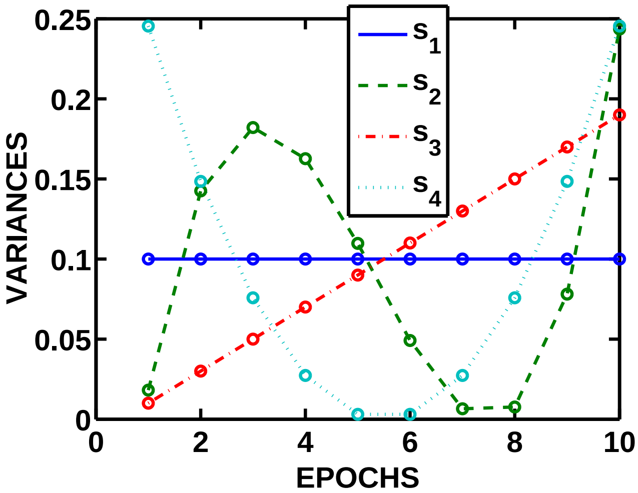

Fig. 1. Variances of the artificial signals in the ten epochs.

TABLE I
MEAN SQUARE ANGULAR ERROR OF COLUMNS OF $\mathbf{A}$ [DB]

component no.	1	2	3	4
UDSEP	-28.0	-29.4	-26.4	-34.8
SOBIUM	-24.8	-23.7	-21.4	-32.5

signals. Quality of the separation was measured by the mean square angular error (MSAE) of columns of the estimated mixing matrix from their true counterparts. The MSAE of the proposed algorithm and the MSAE of SOBIUM are summarized in Table 1. We note that the performance of UDSEP significantly (by 2.3 - 5.7 dB) exceeds the performance of SOBIUM.

B. Acoustic (speech) data separation

In this subsection, the set of 16 speech utterances of the length 8.375 s sampled at 16 kHz, normalized to have zero mean and unit variance, are taken as the original sources $\mathbf{S}$. The 9×16 mixing matrix is defined as by its columns of the form $\mathbf{A}_{:,k} = [1, \cos\phi_k, \cos 2\phi_k, \ldots, \cos 8\phi_k]^T$, where ϕ_k are auxiliary angles $\phi_k = k\pi/17$, $k = 1, \ldots, 16$. The SIR of the original signals in the mixture vary between -10 and -20 dB.

The mixture $\mathbf{X} = \mathbf{AS}$ is processed by UDSEP with $M = 90$ blocks. This means that the length of each block was 93 miliseconds. Variances of angular error of columns of the mixing matrix vary between -20 and -34 dB, while variances of SOBIUM vary between -4.4 dB and -27.7 dB. The latter variances are uniformly worse by at least 5.7 dB.

Output variance of the separated signals vary between 1.5 and 11.8 dB. The separated signals were understandable to a human ear. The mixture, the separated signals, and p-code of the separating procedure were posted on the Internet [3].

REFERENCES

[1] L. De Lathauwer, J. Castaing, "Blind identification of underdetermined mixtures by simultaneous matrix diagonalization", *IEEE Tr. Signal Processing*, Vol. 56, no.3, pp. 1096 - 1105.

[2] P. Paatero, "A weighted non-negative least squares algorithm for three-way 'PARAFAC' factor analysis", *Chemometrics and Intelligent Laboratory Systems*, vol. 38, pp. 223-242, 1997.

[3] P. Tichavský, "A Matlab p-code for UDSEP", http://si.utia.cas.cz/Tichavsky.html