A NEW TENSOR FACTORIZATION APPROACH FOR CONVOLUTIVE BLIND SOURCE SEPARATION IN TIME DOMAIN

Bahador Makkiabadi[1,2], *Foad Ghaderi*[1], *and Saeid Sanei*[1]

[1]Centre of Digital Signal Processing, School of Engineering, Cardiff University, Wales, UK
[2]Electrical Engineering Department, Islamic Azad University, Ashtian Branch, Ashtian, IRAN

ABSTRACT

In this paper a new tensor factorization based method is addressed to separate the speech signals from their convolutive mixtures. PARAFAC and majorization concepts have been used to estimate the model parameters which best fit the convolutive model. Having semi-diagonal covariance matrices for different source segments and also quasi static mixing channels are the requirements for our method. We evaluated the method using synthetically mixed real signals. The results show high ability of our method for separating the speech signals.

Index Terms— Blind Source Separation, Convoutive Mixture, Tensor Factorization, PARAFAC2, Majorization, Procrustes.

1. INTRODUCTION

Blind source separation (BSS) is a technique to estimate unknown source signals from their mixtures without any prior knowledge about the sources or the medium. In some applications, signals are mixed through a convolutive model. This makes BSS a difficult problem. BSS research started from the work of Hrault and Jutten [1] and continued by many researchers. There are three major approaches for solving the convolutive BSS problem; (i) time domain BSS has good results once the algorithm converges, but often they are computationally expensive and involve causality problem, (ii) frequency domain BSS, where the convolutive problem is transferred to frequency domain whereby, the convolution operation changes to multiplication. Then, instantaneous BSS is applied to each frequency bin. This method however is subject to permutation and scaling ambiguities and therefore, can be more complicated than the time domain BSS, (iii) third approach uses time-frequency domain. These methods estimate filter coefficients in the frequency domain and then apply the nonlinear functions to exploit time independency of the sources. This approach is free of permutation problem, but the switching between time and frequency domains is computationally expensive [2]. In this paper we deal with the first method and we develop our time domain BSS based on two well known concepts namely tensor factorization and majorization. To take advantage of tensor factorization we build up a tensor data from multichannel mixture matrix simply by temporal segmentation. Then, we define a model for tensor data based on the well known parallel factor analysis (PARAFAC) [3], more specifically PARAFAC2 [4] concept. Then, we use majorization and PARAFAC fitting optimization algorithms to estimate the parameters of the model. After convergence the estimated sources and the mixing systems at different lags will be at hand. The remainder of the paper is structured as follows. In Section 2 the tensor factorization methods (PARAFAC-PARAFAC2) will be discussed. In Section 3 our time domain convolutive BSS method and also the majorization concept are introduced. In Section 4 the results of applying the method to simulated data are provided. Finally, Section 5 concludes the paper.

2. OPTIMIZATION METHODS

In this section we introduce the tensor factorization and majorization concepts which are used in our time domain convolutive separation method. First, we introduce PARAFAC based tensor factorization metods.

2.1. PARAFAC and PARAFAC2

For mixture signals in tensor form X, the PARAFAC [5] model, which is used to decompose trilinear data sets with a unique solution, is given below:

$$X_{ijk} = \sum_{r=1}^{R} A_{ir} F_{jr} C_{kr} + E_{ijk} \tag{1}$$

where X_{ijk} represents the i, j, k-th element in the three-way data set, R is the number of components in common in the three modes, A_{ir}, F_{jr} and C_{kr} are respectively the elements in A, F and C used to obtain the X_{ijk} elements, and E_{ijk} is the residual term. Using matrix notations the above equation can be presented as:

$$X_k = F D_k A^T + E_k \tag{2}$$

for $k = 1, ..., K$, where $(.)^T$ refers to transpose operation and X_k represents the transposed kth frontal slice of the three-way array X, A and F are the component matrices in the first and second modes, respectively. D_k is a diagonal matrix, whose diagonal elements correspond to the kth row of the third component matrix C. Finally, E_k contains the error terms corresponding to the entries in the kth frontal slice. PARAFAC direct fitting algorithm includes an alternating least squares (ALS) optimization method for obtaining A, F, and D_k for all $k = 1, ..., K$, and consequently finding the three matrices A, F and C of equation (1) respectively [5]. Trilinear ALS fitting method to estimate the factors for PARAFAC can be summarized as follows:

$$\begin{aligned} A &= X_{(1)}((C \odot F)^{\dagger})^T \\ F &= X_{(2)}((C \odot A)^{\dagger})^T \\ C &= X_{(3)}((F \odot A)^{\dagger})^T \end{aligned} \tag{3}$$

Where $(.)^{\dagger}$ stands for Moore-Penrose pseudo inverse operation, $\odot$ is Khatri$-$Rao product and $X_{(n)}$ is the unfolded version of tensor X on mode n.

PARAFAC2 as an extension of PARAFAC is designed to deal with non-trilinear data sets, while keeping uniqueness in the solutions, as the PARAFAC model does. To do so, PARAFAC2 allows a certain degree of freedom on one of the modes [4] . A similar equation in matrix notation for PARAFAC2 is given as:

$$X_k = F_k D_k A^T + E_k \tag{4}$$

subject to $F_k^T F_k = \Phi, k = 1, ...K$, where F_k is the component matrix in the second mode corresponding to the kth frontal slice, Φ is required to be invariant for all slices $k = 1, ..., K$. To keep the uniqueness in the solutions all cross-product matrices $F_k F_k^T$ are forced to be constant over k, i.e. $F_1 F_1^T = F_2 F_2^T = ... = F_k F_k^T$. In equation (4) we observe that unlike in a PARAFAC model, in PARAFAC2 model the component matrix in the second mode can vary across slices. Having constant covariance matrices for all $F_k^T F_k$ we can assume that $F_k = P_k H$ for a columnwise orthonormal $K \times R$ matrix P_k and an $R \times R$ matrix H, for $k = 1, ..., K$.

Considering these new variables the PARAFAC2 model can be written as:

$$X_k = P_k H D_k A^T + E_k \tag{5}$$

The direct method for fitting PARAFAC2 model has been proposed by Kiers [6]. If we compare (5) with $X_k = S_k A^T + E_k$ which is standard formulation of linear mixture signal X_k at kth segment, S_k is the source matrix and A is the mixing matrix. It seems PARAFAC2 model tries to decompose each S_k by one orthonormal matrix P_k, one diagonal matrix D_k and an arbitrary matrix H as:

$$S_k = P_k H D_k \tag{6}$$

Recently, PARAFAC2 has been used as a BSS tool in biomedical and communication applications [7][8].

3. TIME DOMAIN SEPARATION OF CONVOLUTIVE MIXTURES

A number of papers and reviews on convolutive BSS (CBSS) as addressed in [9], have been published recently. In many practical situations the signals and their reflections reach the sensors with different time delays. The corresponding delay between source j and sensor i, in terms of numbers of samples, is directly proportional to the sampling frequency and conversely to the speed of sound in the medium, i.e. $\delta_{ij} \propto d_{ij} \times f_s/c$, where d_{ij}, f_s, and c are respectively, the distance between source j and sensor i, the sampling frequency, and the speed of sound. A general matrix formulation of the CBSS for mixing the source signals can be given as:

$$x_i(t) = \sum_{j=1}^{N_s} \sum_{\tau=0}^{M-1} s_j(t-\tau) a_{ij}(\tau) + v_i(t) \tag{7}$$

for $i = 1, \cdots, N_x$ where N_s and N_x are the number of sources and sensors respectively, $a_{ij}(\tau)$ are the elements of mixing matrix A at different time lags τ. In time domain the above convolutive mixing operator can be formulated using matrix notations as follows:

$$X = \sum_{\tau=0}^{M-1} \Theta_\tau(S) A_\tau^T + V, \quad \Theta_\tau(S) = \Xi_\tau S \tag{8}$$

where $\Theta_\tau(.)$ is a shift operator which can be implemented by pre-multiplication of shift matrix Ξ_τ[10]. As we mentioned before, to build up a tensor from our measurements we use temporal segmentation. The segment size must be much greater than maximum number of lags (M). In matrix notation we have:

$$X_k = \sum_{\tau=0}^{M-1} \Xi_\tau S_k A_\tau^T + V_k \tag{9}$$

for all $k = 1, \ldots, K$ where K is the number of segments. Then, we define our source model similar to that using PARAFAC2 represented in (6) by replacing $S_k = P_k H D_k$ in (9) using the same constraints. Therefore,

$$X_k = \sum_{\tau=0}^{M-1} \Xi_\tau P_k H D_k A_\tau^T + V_k, \; s.t. \; P_k^T P_k = I_R \tag{10}$$

where I_R is $R \times R$ identity matrix and the other parameters like H, D_k, and A_τ are similar to (5). Obviously, the above model is a generalization of PARAFAC2 and PARAFAC2 model is a special case of this model for $\tau = 0$. We apply alternating least squares optimization to estimate the parameters of the model alternatingly. Our optimization method includes three separate processes to estimate ($P_k \; for \; k = 1, \ldots, K$), ($A_\tau for \; \tau = 0, \ldots, M-1$) and ($H, \; D_k \; for \; k = 1, \ldots, K$). For the first part we need to estimate the orthonormal P_k matrix with respect to all other fixed parameters to find best fit of each X_k, however because of having summation for different lags this problem cannot be considered as a quadratic problem and normal least square solutions are not helpful to estimate P_k. Kiers has proposed a general method to convert this type of problems to a quadratic problem and the solution of the new quadratic problem is a solution of the original problem too [11]. So, in order to estimate P_k we apply majorization method. In the next subsection general solution based on majorization is explained. For the second part we convert the model into linear model by matrix manipulation of the lags. Finally, for the third part of our ALS optimization, we use the PARAFAC optimization method given in (3). Before starting the optimization process we randomly initialize all the above parameters.

3.1. Majorization

The problem of minimizing the trace of a matrix is tackled by means of its majorization. This is done by using another function which has a simple quadratic shape whose minimum can be easily found, and minimization on the majorizing function also minimizes the original function. By applying this method a monotonically converging algorithm for minimizing the matrix trace function iteratively is obtained. Kiers introduced a method to minimize a general function of a $(n \times p)$ matrix Π as follows [11]:

$$J(\Pi) = \beta + trW\Pi + \sum_{\tau=0}^{M-1} tr\left(\Phi_\tau \Pi \Psi_\tau \Pi^T\right) \tag{11}$$

where W is a fixed $p \times n$ matrix, Φ_τ a fixed $n \times n$ matrix, Ψ_τ a fixed $p \times p$ matrix, for $\tau = 0, \ldots, M-1$, Π an unknown $n \times p$ matrix, and β a constant that does not depend on Π. The update of

Π for minimizing $J(\Pi)$ is given as [12]:

$$\Pi \leftarrow \Pi - \left(2\sum_{\tau=0}^{M-1}\alpha_\tau\right)^{-1}\left(W^T + \sum_{\tau=0}^{M-1}\Phi_\tau\Pi^T\Psi_\tau + \sum_{\tau=0}^{M-1}\Phi_\tau^T\Pi\Psi_\tau^T\right) \tag{12}$$

where α_τ is a scalar equal or greater than the product of the largest singular values of Φ_τ and Ψ_τ [11].

When there is an orthonormality constraint on Π the solution is somehow simpler and it can be shown by estimating F using:

$$F = \left(W + \sum_{\tau=0}^{M-1}\Phi_\tau^T\Pi\Psi_\tau^T + \sum_{\tau=0}^{M-1}\Phi_\tau\Pi^T\Psi_\tau - 2\alpha_\tau\Pi^T\right) \tag{13}$$

and then finding nearest orthonormal matrix to F as the estimation of Π. If by singular value decomposition (SVD) of F, we have $F = PDQ^T$ then, the estimation of Π in (12) will change to [12]:

$$\Pi \leftarrow QP^T \tag{14}$$

In the next subsection we use the above majorization concept to estimate the second part of our ALS optimization procedure.

3.2. Estimation of P_k using majorization

Let's estimate P_k for $k = 1, \ldots, K$ by keeping all other parameters of the model fixed. We define the optimization problem for each P_k separately as:

$$J(P_k) = ||X_k - \sum_{\tau=0}^{M-1}\Xi_\tau P_k H D_k A_\tau^T||^2 \tag{15}$$

By defining a new variable G_τ

$$G_\tau = HD_kA_\tau^T \tag{16}$$

The optimization problem based on majorization is performed as:

$$\begin{aligned} J(P_k) &= ||X_k - \sum_{\tau=0}^{M-1}\Xi_\tau P_k G_\tau||^2 \\ J(P_k) &= tr\left(X_k^TX_k\right) - \left(2\sum_{\tau=0}^{M-1} tr\left(G_\tau X_k^T \Xi_\tau\right)\right)P_k + \\ &\sum_{\tau=0}^{M-1}\sum_{\gamma=0}^{M-1} tr\left(G_\tau G_\gamma^T P_k \Xi_\gamma^T \Xi_\tau P_k^T\right) \\ &= tr\left(X_k^TX_k\right) + tr\left(-2\sum_{\tau=0}^{M-1}\left(G_\tau X_k^T\Xi_\tau\right)\right)P_k + \\ &\sum_{\tau=0}^{M-1}\sum_{\gamma=0}^{M-1} tr\left(G_\tau G_\gamma^T P_k \Xi_\gamma^T \Xi_\tau P_k^T\right) \end{aligned} \tag{17}$$

Having orthonormality constraint on P_ks and comparing this minimization problem with (11), we can easily update each P_k using (13) and (14).

3.3. Estimating A_τ

Assume P_k, D_k for $k = 1, \ldots, K$, and H are known. Then to estimate A_τ we can convert $\sum_{\tau=0}^{M-1}$ to matrix multiplication as follows:

$$X_k = [\Xi_0 P_k H D_k, \Xi_1 P_k H D_k, \cdots, \Xi_{M-1} P_k H D_k]\begin{bmatrix} A_0^T \\ A_1^T \\ \vdots \\ A_{M-1}^T \end{bmatrix} \tag{18}$$

Now, let's define new variables Z_k and A as:

$$\begin{aligned} Z_k &= [\Xi_0 P_k H D_k, \Xi_1 P_k H D_k, \cdots, \Xi_{T-1} P_k H D_k] \\ \mathrm{A} &= \begin{bmatrix} A_0^T \\ A_1^T \\ \vdots \\ A_{M-1}^T \end{bmatrix} \end{aligned} \tag{19}$$

For different k values we have $X_k = Z_k\mathrm{A}$.

By stacking $X_1, \ldots, X_K$ and $Z_1, \ldots, Z_K$ in two new matrices we have a set of linear equations. The mixing matrix for different lags, A, can be estimated using pseudo inverse operation as follows:

$$\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_K \end{pmatrix} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_K \end{pmatrix}\mathrm{A} \tag{20}$$

$$\mathrm{A} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_K \end{pmatrix}^\dagger \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_K \end{pmatrix} \tag{21}$$

Then after rearranging A we have all A_τ of the model.

3.4. Estimation of H and D_k using PARAFAC

For this case we need to estimate P_kHD_k as part of the main model which is independent of τ, for each k separately and then we are able to apply PARAFAC method for estimating H and D_k. Similar to the solution given in [13] for arbitrary K_τ, L, and Q_τ :

$$vec\left(K_\tau L Q_\tau^T\right) = \left(K_\tau \otimes Q_\tau\right) vec\left(L\right) \tag{22}$$

where vec denotes matrix to vector converter operator and $\otimes$ is Kronecker product operator. We define new variables L_k, K_τ, and Q_τ as:

$$L_k = P_kHD_k, K_\tau = \Xi_\tau, Q_\tau = A_\tau \tag{23}$$

and rewrite our model as:

$$\begin{aligned} X_k &= \textstyle\sum_{\tau=0}^{M-1}\Xi_\tau P_k H D_k A_\tau^T \\ &= \textstyle\sum_{\tau=0}^{M-1} K_\tau L_k Q_\tau^T \end{aligned} \tag{24}$$

Then, the vector formulation for each slab X_k is obtained as follows:

$$vec\left(X_k\right) = \sum_{\tau=0}^{M-1}\left(K_\tau \otimes Q_\tau\right) vec\left(L_k\right) \tag{25}$$

Then, L_k can be easily estimated by:

Correlation	Original male	Original female
Separated male	0.703	0.041
Separated female	0.0508	0.796

Table 1. Correlation between original and separated sounds using proposed method.

$$vec\,(L_k) = \left(\sum_{\tau=0}^{M-1} (\Xi_\tau \otimes A_\tau)\right)^\dagger vec\,(X_k) \qquad (26)$$

After estimation of all $L_k, k = 1, ..., K$ to estimate H and D_k we can rewrite $L_k = P_k H D_k$ as $L_k = P_k H D_k A^T$ $s.t.$ $A = I_R$ and obviously we can fit a PARAFAC2 model to all L_k. However, from majorization outputs we have P_k at hand and because of this we must fit a PARAFAC method with identical mixing matrix to all $W_k = P_k^T L_k = H D_k I_R$. Using (3) we are able to fit the parameters by:

$$\begin{aligned} H &= W_{(2)}((C \odot I_R)^\dagger)^T \\ C &= W_{(3)}((H \odot I_R)^\dagger)^T \end{aligned} \qquad (27)$$

In this optimization we take the advantage of tensor factorization concept to estimate D_ks using the information about all the segments for $k = 1, ..., K$.

4. SIMULATED RESULTS

In this section we evaluated our proposed separation method to separate two sources from their convolutive mixtures and compare the results with those of Parra's method [14]. One male and one female speech signals sampled at 8000 Hz are chosen for our simulation. Maximum number of lags to build up their convolutive mixtures is selected as 8 ($M = 8$) and the mixing matrices for different lags are random. To build up the tensor data from the mixtures we used temporal segmentation with segment size of 108 with 8 samples overlap. All parameters of the model are randomly initialized and the algorithm converged after 148 iterations and convergence curve is shown at Figure 1. The sources can be estimated by stacking non-overlapped part of $\hat{S}_k = P_k H D_k$ matrices.

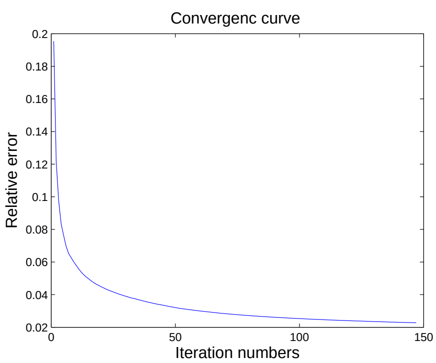

Fig. 1. Convergence curve for 148 iterations.

The separated sources are subject to delays less than 8 samples. Table 1 shows the correlation measured between the separated and original sources.

Correlation	Original male	Original female
Separated male	0.570	0.337
Separated female	0.091	0.4826

Table 2. Correlation between original and separated sounds using Parra's method.

Also, we applied the well known Parra's frequency domain convolutive BSS method [14] to the same convolutive mixture signals to compare the results. Table 2 shows the correlation measured between the separated and original sources using Parra's method.

Unlike in the frequency domain methods there is no permutation within a block of data. Figure 2 shows the normalized original signals, the normalized separated signals using the Parra method and our proposed method, and the mixture signals. The delay between male and female sounds with respect to their original sources measured by correlation between the normalized signals at different lags. The measured delay for male sound was 3 samples and for female 1 sample.

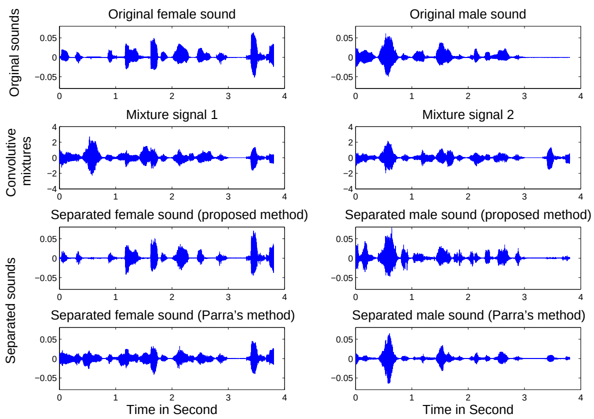

Fig. 2. Original female and male sounds on top, convolutive mixtures in the middle, and separated sources for both methods at the bottom plots.

5. CONCLUSIONS

In this paper a new PARAFAC2 based method is proposed to separate sources from their convolutive mixtures in time domain. We defined a generalized PARAFAC2 structure to model convolutive mixture signals within a tensor model and then we tried to optimize all the model parameters using ALS method. The majorization technique of [12] has been followed for solving minimization of the resulting trace function incorporating Procrustes concepts. The separated signals by this method are subject to delay and scaling for each source. To evaluate the performance of the system we used random value mixing channels for different lags. The results show the high performance of the method compared to Parra's CBSS method to achieve correlation values of greater than 70 percent between separated and original signals.

6. REFERENCES

[1] C. Jutten and J. Hrault, "Blind separation of sources, part 1: An adaptive algorithm based on neuromimetic architecture," *Elsevier, Signal Processing*, vol. 24, no. 1, pp. 1–10, 1991.

[2] S. Makino, H. Sawada, R. Makui, and S. Araki, "Blind source separation of convolutive mixtures of speech in frequency domain," *IEICE Trans. Fundamentals*, vol. 88, no. 7, pp. 1640 – 1654, 2005.

[3] R. A. Harshman, "Foundations of the PARAFAC procedure: models and conditions for an 'explanatory' multimodal factor analysis," *UCLA Working Papers Phonet*, pp. 1 – 84, 1970.

[4] R. A. Harshman, "PARAFAC2: Mathematical and technical notes," *UCLA Working Papers in Phonetics*, vol. 22, pp. 30–44, 1972.

[5] R. Bro, "PARAFAC. tutorial and applications," *Chemometrics and Intelligent Laboratory Systems*, vol. 38, no. 2, pp. 149 – 171, 1997.

[6] H. A. L. Kiers, J. M. F. Ten Berge, and R. Bro, "PARAFAC2 - Part I. A direct fitting algorithm for the PARAFAC2 model," *Journal of Chemometrics*, vol. 13, pp. 275–294, 1999.

[7] B. Makkiabadi, D. Jarchi, and S. Sanei, "Simultaneous localization and separation of biomedical signals by tensor factorization," *Proc. of IEEE Workshop on Statistical Signal Processing*, Sept. 2009.

[8] B. Makkiabadi, A. Sarrafzadeh, D. Jarchi, A. Abolghasemi, and S. Sanei, "Semi-blind signal separation and channel estimation in mimo communication systems by tensor factorization," *Proc. of IEEE Workshop on Statistical Signal Processing*, Sept. 2009.

[9] M. S. Pedersen, J. Larsen, Kjems U., and L. C. Parra, "A survey of convolutive blind source separation methods," *Springer Handbook of Speech Processing*, 2007.

[10] L. K. Huang and A. Manikas, "Blind single-user array receiver for mai cancellation in multipath fading ds-cdma channels," *EUSIPCO Proc. of*, vol. 2, pp. 647–650, Sept. 2000.

[11] H. A. L. Kiers, "Majorization as a tool for optimizing a class of matrix functions," *Psychometrika*, vol. 55, no. 3, pp. 417–428, 1990.

[12] H. A. L. Kiers, "Setting up alternating least squares and iterative majorization algorithms for solving various matrix optimization problems," *Computational Statistics & Data Analysis*, vol. 41, no. 1, pp. 157–170, 2002.

[13] H. A. L. Kiers and J. Berge, "Alternating least squares algorithms for simultaneous components analysis with equal component weight matrices in two or more populations," *Psychometrika*, vol. 54, no. 3, pp. 467–473, 1989.

[14] L. Parra and C. Spence, "Convolutive blind separation of nonstationary sources," *IEEE Trans. on Speech and Audio Processing*, vol. 8, no. 3, pp. 320–327, May 2000.