A GEOMETRICALLY CONSTRAINED MULTIMODAL TIME DOMAIN APPROACH FOR CONVOLUTIVE BLIND SOURCE SEPARATION

Bahador Makkiabadi, Delaram Jarchi, Vahid Abolghasemi, and Saeid Sanei

NICE Group, Faculty of Engineering and Physical Sciences, University of Surrey, UK

ABSTRACT

A novel time domain constrained multimodal approach for convolutive blind source separation is presented which incorporates geometrical 3-D cordinates of both the speakers and the microphones. The semi-blind separation is performed in time domain and the constraints are incorporated through an alternative least squares optimization. Orthogonal source model and gradient based optimization concepts have been used to construct and estimate the model parameters which fits the convolutive mixture signals. Moreover, the majorization concept has been used to incorporate the geometrical information for estimating the mixing channels for different time lags. The separation results show a considerable improvement over time domain convolutive blind source separation systems. Having diagonal or quasi diagonal covariance matrices for different source segments and also having independent profiles for different sources (which implies nonstationarity of the sources) are the requirements for our method. We evaluated the method using synthetically mixed real signals. The results show high capability of the method for separating speech signals.

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 and this makes the BSS a difficult problem. A number of reviews on convolutive BSS (CBSS) as addressed in [1], have been published recently. There are three major approaches for solving the convolutive BSS problem; (i) time domain BSS, (ii) frequency domain BSS, where the convolutive problem is transferred to frequency domain whereby the convolution operation changes to multiplication and (iii) the approach which uses time-frequency domain in the sense of doing adaptation in both time and frequency domains. This method, however, is computationally inefficient since frequent switching between time and frequency domains becomes necessary [2]. A time domain CBSS approach has been recently developed using tensor factorization and majorization concepts [3]. This method divides the mixture signals into different time segments and then defines a tensor model for the segmented source signals. It then uses majorization concept for estimating the orthogonal part of source tensor model and parallel factor analysis (PARAFAC) for the other parts of tensor model. The majorization process, which is used by this method to estimate the orthogonal part of tensor model, is computationally expensive. Moreover, the estimated sources by this method are normally colored (filtered) version of the original sources and respectively the estimated mixing channels (for different lags) are not sometimes physically meaningful for typical applications like separation of speech signals recorded in a room by solving the so called cocktail party problem. On the other hand, there are some multi modal research works which deal with CBSS problem in frequency domain and take the geometrical information of the speakers and microphones (provided by 3-D video based tracker) into account to improve the performance of separation process [4],[5],[6]. In this paper the proposed time domain approach of previous work [3] is computationally improved by substituting majorization based optimization with a faster gradient based approach. Moreover, in order to improve the quality of separated sources and faster convergence the geometrical information has been incorporated with a majorization based method to estimate the mixing channels for different lags. Here, it is assumed that the sources are independent or more specifically the covariance matrix of the source signals and all their reasonable size segments are diagonal. Consider the following instantaneous mixing system:

$$x_i(t) = \sum_{j=1}^{N_s} a_{ij} s_j(t) + v_i(t),\ i = 1, \cdots, N_x \ \ t = 0, \cdots, N-1 \tag{1}$$

where N is the number of time samples, N_s and N_x are respectively the number of sources and sensors, a_{ij} are the elements of mixing matrix $\mathbf{A}$, and $x_i(t)$, $s_j(t)$, and $v_i(t)$ are ith sensor, jth source, and ith noise signals at time instant t. Using matrix notations the above formulation can be represented as follows:

$$\mathbf{X} = \mathbf{S}\mathbf{A}^T + \mathbf{V} \tag{2}$$

where $\mathbf{X}\in\mathbb{R}^{N\times N_x}$, $\mathbf{S} \in \mathbb{R}^{N\times N_s}$, and $\mathbf{V} \in \mathbb{R}^{N\times N_x}$ denote respectively the matrices of observed signals, source signals, and noise. $\mathbf{A}\in\mathbb{R}^{N_x\times N_s}$ is the mixing matrix. Recovering sources from the acquired mixtures has been investigated by incorporating different assumptions about the sources or mixing systems. The approach proposed here relies on orthogonality of the sources for different time segments. A simple temporal segmentation procedure has been developed to divide the signal $\mathbf{X}$ to K segments with/without overlap and with segment size of N_k. Having columnwise orthogonal $\mathbf{S}_k$s is an important criterion which must be considered. So, after temporal segmentation of $\mathbf{X}$ the main model changes to:

$$\begin{aligned} &\mathbf{X}_k = \mathbf{S}_k\mathbf{A}^T + \mathbf{V}_k; \forall\, k = 1, \ldots, K \\ &\mathbf{S}_k^T\mathbf{S}_k = \mathbf{D}_k^2 \end{aligned} \tag{3}$$

where $\mathbf{X}_k \in \mathbb{R}^{N_k\times N_x}$ and $\mathbf{S}_k \in \mathbb{R}^{N_k\times N_s}$ are mixture and source signals and $\mathbf{D}_k$ is diagonal/semi-diagonal for each segment k. For

simplicity, we ignore the noise term $\mathbf{V}_k$ and, also based on orthogonality of $\mathbf{S}_k$, each orthogonal $\mathbf{S}_k$ can be decomposed into one orthonormal matrix $\mathbf{P}_k$ and one diagonal matrix $\mathbf{D}_k$, which absorbs the norm of different sources at each segment k. This decomposition can be considered as a specific case of Polar decomposition [7] which decomposes each $\mathbf{S}_k$ by product of one orthonormal $\mathbf{P}_k$ and one positive semidefinite (PSD) matrix $\mathbf{D}_k$, here the $\mathbf{D}_k$ is considered as a diagonal/semi-diagonal PSD matrix. So, based on the above decomposition the source model can be rewritten as:

$$\begin{aligned} &\mathbf{S}_k = \mathbf{P}_k\mathbf{D}_k; \forall\ k = 1, \ldots, K \\ &\mathbf{P}_k^T\mathbf{P}_k = \mathbf{I}_{N_s} \end{aligned} \tag{4}$$

where $\mathbf{I}_{N_s} \in \mathbb{R}^{N_s \times N_s}$ is an identity matrix. Actually the above formulation tries to define a structured model for source signals $\mathbf{S}_k$. Above source model is independent of the mixing system and is valid for convolutive mixing systems as well. This source model has been used to define a structured model for convolutive mixture signals and ultimately separation of the sources and estimation of the mixing channels for different lags. A majorization based method is developed for semi-blind estimation of the mixing gains using the existing geometrical information of speakers and microphones.

The remainder of the paper is structured as follows. In Section 2 the problem formulation is described. In Section 3 estimation of the model parameters is provided. In Section 4 the results of applying the method to simulated data are provided. Finally Section 5 concludes the paper.

2. CONVOLUTIVE MIXING PROBLEM FORMULATION

Let's investigate the CBSS problem based on the orthogonal model (4). In many practical situations the signals and their reflections reach the sensors with different time delays. In a homogeneous medium such as air, the corresponding delay of direct path between source j and sensor i, in terms of number of samples, is directly proportional to the sampling frequency and conversely to the speed of sound in the medium, i.e. $\tau_{ij} \propto \frac{d_{ij}f_s}{C}$, 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. Similarly, the attenuation is related to the square of distances as $a_{ij} \propto \frac{k_{ij}}{d_{ij}^2}$ where k_{ij} is dependent on the directionality patterns of jth source and ith sensor. In a real case (e.g. a real room) there are always indirect paths between the sources and sensors due to the wall reflections which are not easily measurable. However, the direct path information can be achievable using the geometries of the sensors and sources using video. A general formulation of the CBSS for each time segment of k (ignoring the noise part) can be written as:

$$x_{ki}(t) = \sum_{j=1}^{N_s}\sum_{\tau=0}^{M-1} s_{kj}(t-\tau)a_{ij}(\tau); \forall\ i = 1, \ldots, N_x \tag{5}$$

where $a_{ij}(\tau)$ are the elements of mixing matrix $\mathbf{A}_\tau$ at different time lags τ and M is the maximum number of lags. From geometrical information a few largest $a_{ij}(\tau)$ related to direct paths are available as $a_{ij}(\tau_{ij})$ for $i = 1, ..., N_x, j = 1, ..., N_s$. The above convolutive mixing model can be formulated using matrix notations as follows:

$$\mathbf{X}_k = \sum_{\tau=0}^{M-1} \mathbf{\Xi}_\tau \mathbf{S}_k \mathbf{A}_\tau^T; \forall\ k = 1, \ldots, K \tag{6}$$

where $\mathbf{\Xi}_\tau$ denotes a shift matrix as shifting operator applied to $\mathbf{S}_k$ [8]. Regarding (4) and after substituting $\mathbf{S}_k$ with its orthogonal model the final convolutive model of the mixture signals $\mathbf{X}_k$ can be shown as:

$$\begin{aligned} &\mathbf{X}_k = \sum_{\tau=0}^{M-1} \mathbf{\Xi}_\tau \mathbf{P}_k \mathbf{D}_k \mathbf{A}_\tau^T; \\ &\mathbf{P}_k^T\mathbf{P}_k = \mathbf{I}_{N_s} \end{aligned} \tag{7}$$

Define the overall cost function J for our optimization problem as:

$$\begin{aligned} &J(\mathbf{P}_k, \mathbf{D}_k, \mathbf{A}_\tau) = \sum_{k=1}^{K} ||\mathbf{X}_k - \sum_{\tau=0}^{M-1} \mathbf{\Xi}_\tau \mathbf{P}_k \mathbf{D}_k \mathbf{A}_\tau^T||^2 \\ &\text{subject to } \mathbf{P}_k^T\mathbf{P}_k = I_{N_s} \end{aligned} \tag{8}$$

Two sets of parameters $(\mathbf{P}_1, ..., \mathbf{P}_K)$ and $(\mathbf{D}_1, ..., \mathbf{D}_K)$ vary for different ks whereas, $(\mathbf{A}_0, \mathbf{A}_1, ..., \mathbf{A}_\tau)$ are fixed for all ks. In order to approach to a unique solution (subject to estimation of colored sources and permutation ambiguities) of the above problem one extra constraint is imposed on those parameters which are not fixed for all segments. Orthogonality of the source profiles is a constraint that is imposed on $\mathbf{D}_k$s for all the segments. This constraint physically means that the activity of the sources are relatively sparse along the segments rather than being sparse for each time sample. In this work no constraint is imposed on the mixing channels $\mathbf{A}_\tau$. However, having fixed $\mathbf{A}_\tau$s for all segments can be considered as a weak constraint on $\mathbf{A}_\tau$s. In order to fit the model of mixtures (7), alternating optimizations, are developed for estimation of the three sets of parameters $(\mathbf{A}_0, \mathbf{A}_1, ..., \mathbf{A}_\tau)$, $(\mathbf{P}_1, ..., \mathbf{P}_K)$, and $(\mathbf{D}_1, ..., \mathbf{D}_K)$.

3. ESTIMATION OF THE MODEL PARAMETERS

The parameters of problem (8) can be estimated using three alternating minimizations for estimation of the three sets of existing parameters separately. The following procedures introduce the minimizing processes for estimation of $(\mathbf{A}_0, ..., \mathbf{A}_\tau)$ and each set of $(\mathbf{D}_1, ..., \mathbf{D}_K)$ and $(\mathbf{P}_1, ..., \mathbf{P}_K)$ parameters.

3.1. Estimation of $\mathbf{A}_\tau$s

Assume $\mathbf{P}_k$ and $\mathbf{D}_k$ for $k = 1, \ldots, K$ are known. Then, to estimate $\mathbf{A}_\tau$ the term $\sum_{\tau=0}^{M-1}$ can be converted to matrix multiplication as:

$$\mathbf{X}_k = [\mathbf{\Xi}_0\mathbf{P}_k\mathbf{D}_k, \mathbf{\Xi}_1\mathbf{P}_k\mathbf{D}_k, \cdots, \mathbf{\Xi}_{M-1}\mathbf{P}_k\mathbf{D}_k] \begin{bmatrix} \mathbf{A}_0^T \\ \mathbf{A}_1^T \\ \vdots \\ \mathbf{A}_{M-1}^T \end{bmatrix} \tag{9}$$

After defining new variables $\mathbf{Z}_k$ and $\mathbf{A}$ as

$$\begin{aligned} &\mathbf{Z}_k = [\mathbf{\Xi}_0\mathbf{P}_k\mathbf{D}_k, \mathbf{\Xi}_1\mathbf{P}_k\mathbf{D}_k, \cdots, \mathbf{\Xi}_{M-1}\mathbf{P}_k\mathbf{D}_k] \\ &\mathbf{A} = \begin{bmatrix} \mathbf{A}_0^T \\ \mathbf{A}_1^T \\ \vdots \\ \mathbf{A}_{M-1}^T \end{bmatrix} \end{aligned} \tag{10}$$

every $\mathbf{X}_k$ can be modeled as $\mathbf{X}_k = \mathbf{Z}_k\mathbf{A}$. By stacking $\mathbf{X}_1, \ldots, \mathbf{X}_K$ and $\mathbf{Z}_1, \ldots, \mathbf{Z}_K$ in two new super-matrices a single linear equation can be written as:

$$\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_K \end{pmatrix} = \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \vdots \\ \mathbf{Z}_K \end{pmatrix} \mathbf{A} \tag{11}$$

The super mixing matrix for different lags, A, can be estimated as follows:

$$\mathbf{A} = \mathbf{Z}^{\dagger}\mathbf{X} \tag{12}$$

where $\mathbf{X}$ and $\mathbf{Z}$ are super-matrices made by stacking $\mathbf{X}_k$s and $\mathbf{Z}_k$s respectively and † refers to the pseudo-inverse operation. After rearranging $\mathbf{A}$, estimation of $\mathbf{A}_\tau$ for each τ will be available. This concept has been used to estimate the mixing gains blindly [3]. When some geometrical data are available then, it can be assumed that mixing gains are partially known (at least the gains for the direct paths between speakers and microphones). So, a semi-blind process can benefit from the existing geometrical information. Next subsection shows how majorization concept can be used to develop an iterative method which takes the existing information of channels into account to set a semi-blind process for estimation of $\mathbf{A}$.

3.1.1. Semi-Blind Geometrically Constrained estimation of $\mathbf{A}_\tau$

In the presence of geometrical information (11) can be converted to a minimization problem using a Lagrangian penalty term λ as:

$$J(\mathbf{A}) = ||\mathbf{X} - \mathbf{Z}\mathbf{A}||_F^2 + \lambda||\tilde{\mathbf{A}} - \mathbf{A}||_F^2 \tag{13}$$

where $\tilde{\mathbf{A}}$ includes the existing geometrical information. In order to make a balance between the error of constraint part and major part of cost function the above minimization problem can be changed to:

$$J(\mathbf{A}) = (1-\lambda)||\mathbf{X} - \mathbf{Z}\mathbf{A}||_F^2 + \lambda||\mathbf{Z}\tilde{\mathbf{A}} - \mathbf{Z}\mathbf{A}||_F^2 \tag{14}$$

where $||\mathbf{X}||_F^2 = tr(\mathbf{X}^T\mathbf{X})$ using trace function. The majorization concept can be employed to minimize (14). This can be approached by means of majorizing this function by one having a simple quadratic shape whose minimum is easily found. Moreover, it is shown that the majorization based process further minimizes the main function and makes the algorithm to monotonically converge iteratively [9]. Using trace function the minimization problem changes to:

$$J(\mathbf{A}) = \beta - 2tr\mathbf{W}\mathbf{A} + \sum_{m=1}^{2} tr\left(\mathbf{\Phi}_m\mathbf{A}^T\mathbf{A}\right) \tag{15}$$

where $\mathbf{W} = ((1-\lambda)^2\mathbf{X}^T\mathbf{Z} + \lambda^2\tilde{\mathbf{A}}^T\mathbf{Z}^T\mathbf{Z})$, $\mathbf{\Phi}_1 = (1-\lambda)^2\mathbf{Z}^T\mathbf{Z}, \mathbf{\Phi}_2 = \lambda^2\mathbf{Z}^T\mathbf{Z}$, and $\beta = ||(1-\lambda)\mathbf{X}||_F^2 + ||\lambda\mathbf{Z}\tilde{\mathbf{A}}||_F^2$ a constant that does not depend on $\mathbf{A}$. The update of $\mathbf{A}$ for minimizing $J(\mathbf{A})$ is given as [9]:

$$\mathbf{A} \leftarrow \mathbf{A} - \left(\sum_{m=1}^{2}\alpha_m\right)^{-1}\left(\sum_{m=1}^{2}\Phi_m\mathbf{A} - \mathbf{W}^T\right) \tag{16}$$

where α_m is a scalar equal or greater than the largest singular value of $\mathbf{\Phi}_m$ [9].

3.2. Estimation of $\mathbf{P}_k$s

At this stage it is assumed that $\mathbf{A}_\tau$s and $\mathbf{D}_k$s are available for all k and τ and estimation of all $\mathbf{P}_k$s is required. Based on the model $\mathbf{X}_k = \sum_{\tau=0}^{M-1}\mathbf{\Xi}_\tau\mathbf{P}_k\mathbf{D}_k\mathbf{A}_\tau^T$ it is necessary to find orthonormal $\mathbf{P}_k$s to fit the model at each segment k. This problem can be solved for each k separately. So, after defining a new variable $\mathbf{G}_\tau = \mathbf{D}_k\mathbf{A}_\tau^T$ a local minimization problem for each k can be defined as:

$$\begin{aligned} J(\mathbf{P}_k) &= ||\mathbf{X}_k - \sum_{\tau=0}^{M-1}\mathbf{\Xi}_\tau\mathbf{P}_k\mathbf{G}_\tau||^2 \\ &\text{subject to } \mathbf{P}_k^T\mathbf{P}_k = I_{N_s} \end{aligned} \tag{17}$$

without having orthogonality constraint on $\mathbf{P}_k$ there is a closed solution for $\mathbf{P}_k$ as:

$$vec\,(\mathbf{P}_k) = \left(\sum_{\tau=0}^{M-1}\mathbf{G}_\tau^T \otimes \mathbf{\Xi}_\tau\right)^{\dagger} vec\,(\mathbf{X}_k) \tag{18}$$

where $vec(.)$ is matrix to vector converter operator and $\otimes$ denotes Kronecker product. Because of high dimensionality of $\mathbf{\Xi}_k$ this solution is computationally expensive and also does not support the orthogonality constraint of $\mathbf{P}_k$. Moreover, in our blind process, the exact $\mathbf{G}_\tau$s are not available and the above exact solution may force the algorithm to converge to a local minimum. In order to overcome these problems an iterative approach has been developed. One standard iterative solution of unconstrained version of (17) is proposed in [10]. Using this iterative concept the solution of constrained problem can be proposed as:

$$\begin{aligned} \mathbf{Q} =& \mathbf{P}_k + \frac{\mu}{M}\left(\sum_{i=0}^{M-1}\mathbf{\Xi}_i^T\left(\mathbf{X_k} - \sum_{\tau=0}^{M-1}\mathbf{\Xi}_\tau\mathbf{P}_k\mathbf{G}_\tau\right)\mathbf{G}_i^T\right) \\ \mathbf{P}_k \leftarrow& \mathbf{U}\mathbf{V}^T \end{aligned} \tag{19}$$

where $\mathbf{U}$ and $\mathbf{V}$ include orthonormal left and right singular vectors of $\mathbf{Q}$ using singular value decomposition (SVD) as $\mathbf{Q} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ and $\mu \leq \left(\sum_{\tau=0}^{M-1}||\mathbf{G}_\tau||^2\right)^{-1}$. In above formulation $\mathbf{Q}$ is the solution of unconstrained version of (17) and it is computed using iterative gradient minimization. However, updation of $\mathbf{P}_k$ as $\mathbf{P}_k \leftarrow \mathbf{U}\mathbf{V}^T$ imposes the orthogonality constraint. There is another standard iterative solution for the above constrained problem using majorization concept which is computationally more intensive than the proposed gradient based method [9],[3].

3.3. Estimation of $\mathbf{D}_k$s

Estimating $\mathbf{D}_k$s as part of the main model can be performed for each k separately. Similar to the solution given in (18) the unconstrained estimation of $\mathbf{D}_k$ called $\mathbf{M}_k$ can be shown by:

$$vec\,(\mathbf{M}_k) = \left(\sum_{\tau=0}^{M-1}\mathbf{A}_\tau \otimes \mathbf{\Xi}_\tau\mathbf{P}_k\right)^{\dagger} vec\,(\mathbf{X}_k) \tag{20}$$

where the diagonal elements of $\mathbf{M}_k$ are the estimation of diagonal elements of $\mathbf{D}_k$. This solution, because of having smaller matrices, is not computationally as expensive as (18) and non-iterative solution can be employed. Moreover, in order to have more relatively unique estimations an orthogonality constraint is imposed between

the vectors including the diagonal elements of all $\mathbf{D}_k$s. Figure 1 shows typical profiles (absolute value of diagonal elements of $\mathbf{D}_k$s) of speech signals. Actually, the orthogonality is applied to the activity of the sources along time segments called their profiles. For this, the diagonal elements of $\mathbf{M}_k$s for all $k = 1, ..., K$ must be stacked in matrix $\mathbf{C} \in \mathbb{R}^{K \times N_s}$ and then each row of the orthogonalized version of $\mathbf{C}$ will be the final estimation of diagonal elements $\mathbf{D}_k$s as below:

$$\mathbf{D}_k = diag(\mathbf{m}_k \mathbf{R} \mathbf{\Sigma}^{-1} \mathbf{R}^T) \quad (21)$$

where $diag(\mathbf{x})$ makes a diagonal matrix with diagonal elements equal to $\mathbf{x}$ elements, $\mathbf{m}_k \in \mathbb{R}^{1 \times N_s}$ is a horizontal vector includes the diagonal elements of $\mathbf{M}_k$, and $\mathbf{R}$, $\mathbf{\Sigma}$ include the right singular vectors and singular values of $\mathbf{C}$ respectively. An alternative way for estimating $\mathbf{D}_k$s also can be implemented using Khatri-Rao product to estimate $\mathbf{m}_k$, only the diagonal elements of $\mathbf{M}_k$, directly. However, Kronecker based method has better speed of overall convergence.

The final algorithm for alternating minimization process for estimation of all the parameters is shown in Algorithm 1. In the next section this algorithm will be applied to some nonstationary signals such as speech signals. These signals can be considered mutually orthogonal for certain size segments. Moreover, their profiles are normally independent of each other which provides orthogonality of profile signals as the second requirement for the proposed method.

Algorithm 1 Semi-blind source separation parameter estimation using alternating minimization

Step1 : Initialize all of the model parameters.
Step2 : Estimation of $\mathbf{P}_k$ using (19) for all $k = 1, ..., K$.
Step3 : Estimation of $\mathbf{A}_\tau$s using (16).
Step4 : Estimation of $\mathbf{D}_k$s using (20) and (21).
Step5 : Check the convergence rate $\sigma = ||J_{new} - J_{old}||/||J_{old}||$ if $\sigma > \epsilon$, **go to Step2** till convergence

4. SIMULATION RESULTS

In this section the proposed method is evaluated for separation of speech sources from their convolutive mixtures in a simulated room with dimensions (2, 2, 2). Three speech signals (from one female and two males) are chosen to be mixed convolutively. The coordinates of microphones m1, m2, m3 are respectively (0, 0, 0), (0, .50 ,0), (0, .50, .50), and the coordinates of sources s1, s2, s3 are (.47, .01, .01), (.47, .49, .01), (.47, .49, .49), respectively. Reflection coefficient of walls are chosen as [0.9 0.9 0.7 0.7 0.6 0.6]. The audio signals are sampled at 8 kHz. The received signals from the microphone array are computed using RoomSim software [11]. The reverberation time is measured about 18 millisecond (140 taps). To build up the segmented data from the mixtures a temporal segmentation scenario has been used with segment size of $N_k = 300$ without overlap and maximum number of lags to build up the tensor model is selected as 140 ($M = 140$). The initial values for mixing gains $\mathbf{A}_\tau$s are considered as zero except for τ_{ij} as $a_{ij} \propto \frac{k_{ij}}{d_{ij}^2}$ for $i, j = 1, ..., N_s$ which can be computed using Roomsim (when all reflection coefficients of walls are considered as zero). All other parameters of the model are randomly initialized. λ is initialized by 0.7 and it is gradually decreased during the iterations by $\lambda = 0.7e^{-.05j}$ where j is the iteration counter. The error decreased monotonically and the optimization converged after 94 iterations. The original and estimated profiles for different sources are shown in Figure 1. It can be seen that the estimated profiles have closely followed the original ones. Also, a completely blind version of the above algorithm is applied to

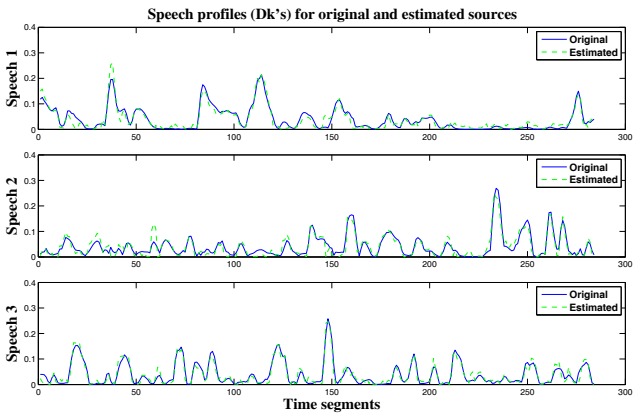

Fig. 1: Original and separated profiles ($\mathbf{D}_k$'s) of the source signals .

the same convolutive mixture signals to compare the results. Comparing with the completely blind algorithm, the estimated mixing channels (impulse responses between speakers and microphones) are more correlated with the actual ones. Figure 2 shows the actual (in simulated room), shortest path, blindly estimated, and semi-blindly estimated impulse response between the third source (s3) and second microphone (m2). The separated sources can be estimated by

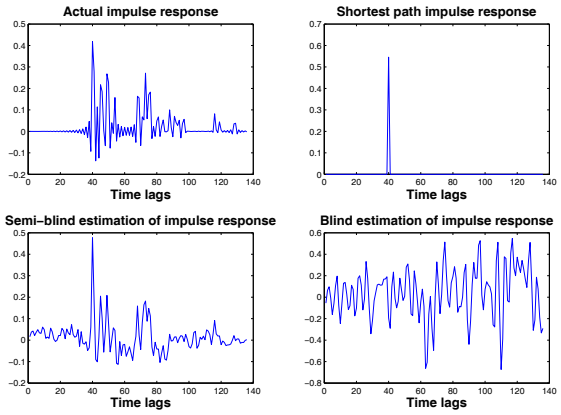

Fig. 2: Actual on top left, shortest path on top right, semi-blindly estimated on bottom left, and on bottom right blindly estimated impulse response between s3 and m2.

stacking $\hat{\mathbf{S}}_k = \mathbf{P}_k \mathbf{D}_k$ matrices. Because of blind estimation of channels $\mathbf{A}_\tau$s (except for points available from the geometrical information) there are different scaling ambiguities for different lags and this causes the separated sources to be the filtered versions of the original sources. Because of having filtered version of the sources,

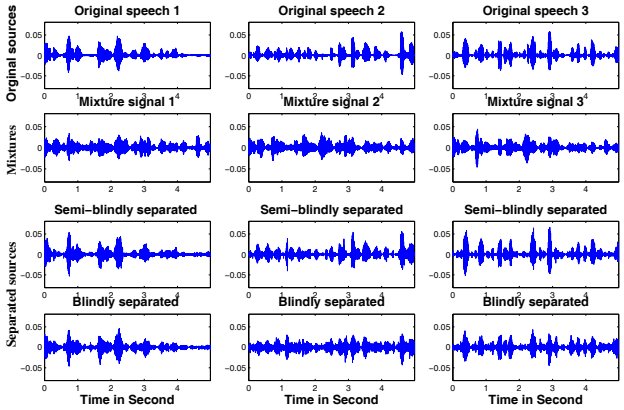

Fig. 3: Original signals on top, convolutive mixtures in the middle, and separated sources for both methods at the bottom plots.

Table 1: Correlation between original and separated signals using the proposed method.

Correlation	Separated 1	Separated 2	Separated 3
Original 1	0.682	0.029	-0.045
Original 2	0.0131	0.768	0.026
Original 3	0.033	0.022	0.7373

measuring the signal to interference ratio (SIR) for lag zero may not be an accurate performance measure (specially for completely blind version of the algorithm). So, in order to measure the performance, the cross correlation of lagged versions of each normalized estimated source with all normalized original sources has been measured. Table 1 shows the maximum cross correlation (for all lags between -M to M) measured between the separated and original sources using the proposed semi-blind method. Table 2 shows the maximum correlation measured between the separated and original sources using the completely blind version of the above method. Figure 3 shows the normalized original signals, the normalized separated signals using the completely blind method, and our proposed semi-blind method, and the mixture signals.

5. CONCLUSIONS

This paper improved the recently developed time domain CBSS using an orthogonal signal model which is defined for source signals and then used to define a signal model for convolutive mixture signals. A computationally efficient gradient based approach is developed to estimate the orthogonal part of model ($\mathbf{P}_k$s). Moreover, the majorization concept is employed to impose the existing geometrical information and developing a semi-blind time-domain CBSS. Although, the estimated channels by the proposed method are not unique but compared to the blind algorithm they are more correlated to actual channels. To evaluate the performance of the system the mixing channels of a simulated room are used to mix the speech signals. The results show the high performance of the method compared with those of completely blind CBSS method (using (12), rather than (16), to estimate the channels) to achieve higher correlation values for the same speaker and lower correlation values for different speakers between the separated and original signals.

Table 2: Correlation between original and separated signals using blind method.

Correlation	Separated 1	Separated 2	Separated 3
Original 1	0.388	-0.066	0.057
Original 2	-0.029	-0.256	-0.069
Original 3	-0.046	0.100	-0.337

6. REFERENCES

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

[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] B. Makkiabadi, F. Ghaderi, and S. Sanei, "A new tensor factorization approach for convolutive blind source separation in time domain," in *EUSIPCO proc.*, (Aalborg Denmark), Aug. 2010.

[4] S. Sanei, S. M. Naqvi, J. Chambers, and Y. Hicks, "A geometrically constrained multimodal approach for convolutive blind source separation," in *Acoustics, Speech and Signal Processing, IEEE Int. Conf. on*, vol. 3, pp. III–969 –III–972, april 2007.

[5] R. Mukai, H. Sawada, S. Araki, and S. Makino, "Robust real-time blind source separation for moving speakers in a room," in *Acoustics, Speech, and Signal Processing, Proc. IEEE Int. Conf. on*, vol. 5, pp. 469–472, april 2003.

[6] S. M. Naqvi, M. Yu, and J. Chambers, "A multimodal approach to blind source separation of moving sources," *Selected Topics in Signal Processing, IEEE Journ. of*, vol. 4, pp. 895 –910, oct. 2010.

[7] G. H. Golub and C. F. Van Loan, *Matrix computations*. Johns Hopkins University Press, Baltimore, 1983.

[8] 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, Sep 2000.

[9] H. A. L. Kiers, "Majorization as a tool for optimizing a class of matrix functions," *Psychometrika*, vol. 55, pp. 417–428, Sep 1990.

[10] F. Ding and T. Chen, "Gradient based iterative algorithms for solving a class of matrix equations," *IEEE Trans., Automatic Control*, vol. 50, pp. 1216 – 1221, Aug. 2005.

[11] *Roomsim Toolbox*. http://sourceforge.net/projects/roomsim.