

BLIND SEPARATION OF NON-STATIONARY CONVOLUTIVELY MIXED SIGNALS IN THE TIME DOMAIN

Iain Russell *, Jiangtao Xi *, Alfred Mertins **, and Joe Chicharo *

* School of Elec., Comp., and Tele. Eng., University of Wollongong,
Wollongong, N.S.W. 2522, Australia, Email: {iainr,jiangtao,chicharo}@uow.edu.au

** Signal Processing Group, Institute of Physics, University of Oldenburg,
26111 Oldenburg, Germany, Email: alfred.mertins@uni-oldenburg.de

Abstract— This paper proposes a new algorithm for solving the Blind Signal Separation (BSS) problem for convolutive mixing completely in the time domain. The closed form expressions used for first and second order optimization techniques derived in [1] are extended to accommodate the more practical convolutive mixing scenario. Traditionally convolutive BSS problems are solved in the frequency domain [2], [3], [4] but this requires additional solving of the inherent frequency permutation problem. We demonstrate the performance of the algorithm using two optimization methods with a convolutive synthetic mixing system and real speech data.

I. INTRODUCTION

Blind Signal Separation (BSS) [5], [6] has been a topic which attracted many researchers in recent years. With the advent of more powerful processors and the ability to realize more complex algorithms BSS has found useful applications in the areas of audio processing such as speech recognition, audio interfaces, and hands free telephony in reverberant environments. In view of the exponential growth of mobile users in the wireless-communications world together with the limited capacity of resources available for data transmission, modern communication systems increasingly require training-less adaptation, to save on bandwidth capacity or to accommodate unpredictable channel changes. Future systems must utilize spatial diversity multiple access techniques that obtain their channel information exclusively from the received signal. These systems fit the instantaneous and convolutive BSS models. Blind algorithms are useful here as they can be self-recovering and do not require *a priori* knowledge of any training sequence [7]. For example communication systems such as GSM can devote up to 22% of their transmission time to pilot tones which could be otherwise used for data transmission [8]. BSS has also found a fruitful application in multimedia modelling and recent work on modelling combined text/image data for the purpose of cross-media retrieval has been made using ICA [9].

There is an abundance of various methods used to solve BSS problems and these are often application dependent, however; this paper investigates an algorithm which demonstrates the convolutive mixing model which is relevant to the applications mentioned

above and provides a method that avoids the frequency permutation problem in the time domain. The most prevalent of the aforementioned applications suitable for this particular BSS criterion is in the area of speech processing as it exploits the non-stationarity assumption of the algorithm.

This paper extends approaches in [1] to the convolutive mixing cases. Section II gives a brief description of modelling BSS in a convolutive mixing environment. In Section III the approaches in [1] are briefly reviewed. The extended approach in convolutive mixing cases is given in Section IV. Section V presents the simulation results giving the performance of two optimization methods: Gradient, and Newton optimization with speech data. Finally, a conclusion is provided in Section VI.

The following notations are used in this paper. Vectors and matrices are printed in boldface with matrices being in capitals. Matrix and vector transpose, complex conjugation, and Hermitian transpose are denoted by $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H \triangleq ((\cdot)^*)^T$, respectively. \dagger is the pseudo-inverse while $E(\cdot)$ means the expectation operation. $\|\cdot\|_F$ is the Frobenius norm of a matrix. With $\mathbf{a} = \text{diag}(\mathbf{A})$ we obtain a vector whose elements are the diagonal elements of \mathbf{A} and $\text{diag}(\mathbf{a})$ is a square diagonal matrix which contains the elements of \mathbf{a} . $\text{ddiag}(\mathbf{A})$ is a diagonal matrix where its diagonal elements are the same as the diagonal elements of \mathbf{A} and

$$\text{off}(\mathbf{A}) \triangleq \mathbf{A} - \text{ddiag}(\mathbf{A}). \quad (1)$$

II. BSS MODEL

The main issue of BSS is that neither the signal sources nor the mixing system are known *a priori*. The only assumption made is that the unknown signal sources are statistically independent. Assume there are N statistically independent sources, $\mathbf{s}(t) = [s_1(t), \dots, s_N(t)]^T$. These sources are mixed in a medium providing M sensor or observed signals, $\mathbf{x}(t) = [x_1(t), \dots, x_M(t)]^T$, given by:

$$\mathbf{x}(t) = \mathbf{H}(t) * \mathbf{s}(t) \quad (2)$$

where $\mathbf{H}(t)$ is a $M \times N$ mixing matrix with its element $\mathbf{h}_{ij}(t)$ being the impulse response from j th source

signal to i th measurement. $*$ defines the convolution of corresponding elements of $\mathbf{H}(t)$ and $\mathbf{s}(t)$ following the same rules for matrix multiplication.

Assuming that the mixing channels can be modelled as FIR filters with length P , Equation (2) can be rewritten as:

$$\mathbf{x}(t) = \sum_{\tau=0}^{P-1} \mathbf{H}(\tau) \mathbf{s}(t-\tau). \quad (3)$$

The M observed signals $\mathbf{x}(t)$ are coupled to the N reconstructed signals $\hat{\mathbf{s}}(t)$ via the de-mixing system. The de-mixing system has a similar structure to the mixing system. It contains $N \times M$ FIR filters of length Q , where $Q \geq P$. The de-mixing system can also be expressed as an $N \times M$ matrix $\mathbf{W}(t)$, with its element $w_{ij}(t)$ being the impulse response from j th measurement to i th output. The reconstructed signal can be obtained as:

$$\hat{\mathbf{s}}(t) = \sum_{\tau=0}^{Q-1} \mathbf{W}(\tau) \mathbf{x}(t-\tau) \quad (4)$$

where $\hat{\mathbf{s}}(t) = [\hat{s}_1(t), \dots, \hat{s}_N(t)]^T$. A straight forward approach for BSS is to identify the unknown system first and then to apply the inverse of the identified system to the measurement signals in order to restore the signal sources. This approach can lead to problems of instability. Therefore it is desired that the de-mixing system be estimated based on the observations of mixed signals.

The simplest case is the instantaneous mixing in which matrix $\mathbf{H}(t)$ is a constant matrix with all elements being scalar values. In practical applications such as hands free telephony or mobile communications where multi-path propagation is evident mixing is convolutive, in which situation BSS is much more difficult due to the added complexity of the mixing system. The frequency domain approaches are considered to be effective to separate signal sources in convolutive cases, but another difficult issue, the inherent permutation and scaling ambiguity in each individual frequency bin, arises which makes the reconstruction of signal sources almost impossible [10]. Therefore it is worthwhile to develop an effective approach in the time domain. Joho and Rahbar [1] proposed a BSS approach based on joint diagonalization of the output signal correlation matrix using gradient and Newton optimization methods. However the approaches in [1] are limited to the instantaneous mixing cases whilst in the time domain.

III. OPTIMIZATION OF INSTANTANEOUS BSS

This section gives a brief review of the algorithms proposed in [1]. Assuming that the sources are statistically independent and non-stationary, observing the signals over K different time slots, we define the following noise free instantaneous BSS problem. In the instantaneous mixing cases both the mixing and de-mixing matrices are constant, that is, $\mathbf{H}(t) = \mathbf{H}$

and $\mathbf{W}(t) = \mathbf{W}$. In this case the reconstructed signal vector can be expressed as:

$$\hat{\mathbf{s}}(t) = \mathbf{W}\mathbf{x}(t). \quad (5)$$

The instantaneous correlation matrix of $\hat{\mathbf{s}}(t)$ at time frame k can be obtained as:

$$\mathbf{R}_{\hat{s}\hat{s},k} = \mathbf{W}\mathbf{R}_{xx,k}\mathbf{W}^H \quad (6)$$

$$\mathbf{R}_{xx,k} = E\{\mathbf{x}(k)\mathbf{x}^H(k)\}. \quad (7)$$

For a given set of K observed correlation matrices, $\{\mathbf{R}_{xx,k}\}_{k=1}^K$, the aim is to find a matrix \mathbf{W} that minimizes the following cost function:

$$\mathcal{J}_1 \triangleq \sum_{k=1}^K \beta_k \|off(\mathbf{W}\mathbf{R}_{xx,k}\mathbf{W}^H)\|_F^2 \quad (8)$$

where $\{\beta_k\}$ are positive weighting *normalization* factors such that the cost function is independent of the absolute norms and are given as:

$$\beta_k = (\sum_{k=1}^K \|\mathbf{R}_{xx,k}\|_F^2)^{-1}. \quad (9)$$

Perfect joint diagonalization is possible under the condition that $\{\mathbf{R}_{xx,k}\} = \{\mathbf{H}\Lambda_{ss,k}\mathbf{H}^H\}$ where $\{\Lambda_{ss,k}\}$ are diagonal matrices due to the assumption of the mutually independent unknown sources. This means that full diagonalization is possible, and when this is achieved, the cost function is zero at its global minimum. This constrained non-linear multivariate optimization problem can be solved using various techniques including gradient-based steepest descent and Newton optimization routines. However, the performance of these two techniques depends on the initial guess of the global minimum, which in turn relies heavily on an initialization of the unknown system that is near the global trough. If this is not the case then the solution may be sub-optimal as the algorithm gets trapped in one of the local multi-minima points.

To prevent a trivial solution where $\mathbf{W} = \mathbf{0}$ would minimize Equation (8), some constraints need to be placed on the unknown system \mathbf{W} to prevent this. One possible constraint is that \mathbf{W} is unitary. This can be implemented as a penalty term such as given below:

$$\mathcal{J}_2 \triangleq \|\mathbf{W}\mathbf{W}^H - \mathbf{I}\|_F^2 \quad (10)$$

or as a hard constraint that is incorporated into the adaptation step in the optimization routine. For problems where the unknown system is constrained to be unitary, Manton presented a routine for computing the Newton step on the manifold of unitary matrices referred to as the *complex Stiefel manifold*. For further information on derivation and implementation of this hard constraint refer to [1] and references therein.

The closed form analytical expressions for first and second order information used for gradient and Hessian expressions in optimization routines are taken from Joho and Rahbar [1] and will be referred to

when generating results for convergence. Both the Steepest gradient descent (SGD) and Newton methods are implemented following the same frameworks used by Joho and Rahbar. The primary weakness of these optimization methods is that although they do converge relatively quickly there is no guarantee for convergence to a global minimum which provides the only true solution. This is exceptionally noticeable when judging the audible separation of speech signals. To demonstrate the algorithm we assume a good initial starting point for the unknown separation system to be identified by setting the initial starting point of the unknown system in the region of the global trough of the multivariate objective function.

IV. OPTIMIZATION OF CONVOLUTIVE BSS IN THE TIME DOMAIN

As mentioned previously and as with most BSS algorithms that assume convolutive mixing, solving many BSS problems in the frequency domain for individual frequency bins can exploit the same algorithm derivation as the instantaneous BSS algorithms in the time domain. However the inherent *frequency permutation problem* remains a major challenge and will always need to be addressed. The tradeoff is that by formulating algorithms in the frequency domain we can perform less computations and processing time falls, but we still must fix the permutations for individual frequency bins so that they are all aligned correctly. The main contribution in this paper is to provide a way to utilize the existing algorithm developed for instantaneous BSS but avoid the permutation problem.

Now we extend the above approach to the convolutive cases. We still assume that the de-mixing systems are defined by Equation (4), which consists of $N \times M$ FIR filters with length Q. We want to get a similar expression to those in the instantaneous cases. It can be shown that Equation (4) can be written as the following matrix form:

$$\hat{\mathbf{s}}(n) = \mathcal{W}\mathcal{X}(n) \quad (11)$$

where \mathcal{W} is a $(N \times QM)$ matrix given by:

$$\mathcal{W} = [\mathbf{W}(0), \mathbf{W}(1), \dots, \mathbf{W}(Q-1)] \quad (12)$$

and $\mathcal{X}(n)$ is a $(QM \times 1)$ vector defined as:

$$\mathcal{X}(n) = \begin{bmatrix} \mathbf{x}(n) \\ \mathbf{x}(n-1) \\ \vdots \\ \mathbf{x}(n-(Q-1)) \end{bmatrix}. \quad (13)$$

Then the output correlation matrix at time k can be derived as:

$$\mathbf{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}},k}(0) = \mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X},k}(0)\mathcal{W}^H \quad (14)$$

where,

$$\mathbf{R}_{\mathcal{X}\mathcal{X},k}(0) = E\{\mathcal{X}(k)\mathcal{X}^H(k)\}. \quad (15)$$

Correlation matrices for the recovered sources for all necessary time lags can also be obtained as:

$$\begin{aligned} \mathbf{R}_{\hat{\mathbf{s}}\hat{\mathbf{s}},k}(\tau) &= \mathcal{W}E\{\mathcal{X}(k)\mathcal{X}^H(k+\tau)\}\mathcal{W}^H \\ &= \mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X},k}(\tau)\mathcal{W}^H \end{aligned} \quad (16)$$

Using the joint-diagonalization criterion in [1] for the instantaneous modelling of the BSS problem we can formulate a similar expression for convolutive mixing in the time domain. Consider the correlation matrices with all different time lags we should have the following cost function:

$$\mathcal{J}_3 \triangleq \sum_{\tau=-\tau_{min}}^{\tau_{max}} \sum_{k=1}^K \beta_k \|off(\mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X},k}(\tau)\mathcal{W}^H)\|_F^2. \quad (17)$$

The only difference between \mathcal{J}_1 and \mathcal{J}_3 is that we now take into account all the different time lags τ for the correlation matrices for each respective time epoch k where the SOS are changing. Also β_k is now defined as,

$$\beta_k = \left(\sum_{\tau=-\tau_{min}}^{\tau_{max}} \sum_{k=1}^K \|\mathbf{R}_{\mathcal{X}\mathcal{X},k}(\tau)\|_F^2 \right)^{-1}, \quad (18)$$

and we note the new structure of \mathcal{W} . In the ideal case where we know the exact system \mathcal{W}_{ideal} , all off-diagonal elements would equal zero and the value of the objective function would reach its global minimum where $\mathcal{J}_3 = 0$. Each value of k represents a different time window frame where the Second Order Statistics (SOS) are considered stationary over that particular time frame. In adjacent non-overlapping time frames k , the SOS are changing due to the non-stationarity assumption. As this is a non-linear constrained optimization problem with NQM unknown parameters we can rewrite it as,

$$\begin{aligned} \mathcal{W}_{opt} = \arg \min_{\mathcal{W}} \mathcal{J}_3(\mathcal{W}) \\ s/t \quad \|ddiag(\mathcal{W}\mathcal{W}^H - \mathbf{I})\|_F^2 = 0 \end{aligned} \quad (19)$$

Due to the structure of the matrices and with the technique of matrix multiplication to perform convolution in the time domain, optimization algorithms similar to those performed in the instantaneous climate can be utilized. Notice also that in the instantaneous version the constraint used to prevent the trivial solution $\mathbf{W} = 0$ was a unitary one. In the convolutive case a different constraint is used where the row vectors of \mathcal{W} are normalized to have length one. Again referring to the SGD and Newton algorithms closed form analytical expressions of the gradient and Hessian deduced by Joho and Rahbar [1] are extended slightly to accommodate the time domain convolutive climate of the new algorithm. These expressions are shown in Table 1. $\mathbf{R}_{\mathcal{X}\mathcal{X},k}(\tau)$ will be denoted as $\mathbf{R}_{\mathcal{X}\mathcal{X},k}^\tau$. With these expressions the SGD and Newton methods are summarized in the Tables 2 and 3 respectively and are mainly similar to the method proposed in [1], however, they work for convolutive mixing. Table 2 is relatively

Cost function - \mathcal{J}_W	
$\mathcal{J}_W \triangleq \sum_{\tau=-\tau_{min}}^{\tau_{max}} \sum_{k=1}^K \ off(\mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X},k}^\tau \mathcal{W}^H)\ _F^2$	Gradient - \mathbf{G}_W
$\mathbf{G}_W = 2 \sum_{\tau=-\tau_{min}}^{\tau_{max}} \sum_{k=1}^K \{off(\mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X},k}^\tau \mathcal{W}^H)\mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau H}$	$+ off(\mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau H} \mathcal{W}^H)\mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X},k}^\tau\}$
	Hessian - \mathbf{H}_W
$\mathbf{H}_W = 2 \sum_{\tau=-\tau_{min}}^{\tau_{max}} \sum_{k=1}^K \{(\mathbf{R}_{\mathcal{X}\mathcal{X},k}^\tau)^* \otimes off(\mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X},k}^\tau \mathcal{W}^H)$	
$+ (\mathbf{R}_{\mathcal{X}\mathcal{X},k}^\tau \otimes off(\mathcal{W}\mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau H} \mathcal{W}^H))$	
$+ (\mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau T} \mathcal{W}^T \otimes \mathbf{I})\mathbf{P}_{off}(\mathcal{W}^* \mathbf{R}_{\mathcal{X}\mathcal{X},k}^\tau)^* \otimes \mathbf{I})$	
$+ (\mathbf{R}_{\mathcal{X}\mathcal{X},k}^\tau)^* \mathcal{W}^T \otimes \mathbf{I})\mathbf{P}_{off}(\mathcal{W}^* \mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau T} \otimes \mathbf{I})$	
$+ (\mathbf{R}_{\mathcal{X}\mathcal{X},k}^\tau \mathcal{W}^H \otimes \mathbf{I})\mathbf{P}_{vec}\mathbf{P}_{off}(\mathcal{W}^* \mathbf{R}_{\mathcal{X}\mathcal{X},k}^\tau)^* \otimes \mathbf{I})$	
$+ (\mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau H} \mathcal{W}^H \otimes \mathbf{I})\mathbf{P}_{off}\mathbf{P}_{vec}(\mathcal{W}^* \mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau T} \otimes \mathbf{I})\}$	
	Row-normalized Constraint
$\mathcal{J}_4 = \ ddiag(\mathcal{W}\mathcal{W}^H - \mathbf{I})\ _F^2$	Constraint Gradient
$\mathbf{G}_4 = 4ddiag(\mathcal{W}\mathcal{W}^H - \mathbf{I})\mathcal{W}$	

TABLE I

CLOSED FORM ANALYTICAL EXPRESSIONS FOR THE GRADIENT AND HESSIAN OF THE COST FUNCTION AND CONSTRAINTS

Initialization ($k = 0$) : \mathcal{W}_0
For $k = 1, 2, \dots$
$\Delta\mathcal{W}_k = -\mu(\mathbf{G}_W + \alpha\mathbf{G}_4)$
$\mathcal{W}_{k+1} = \mathcal{W}_k + \Delta\mathcal{W}_k$

TABLE II

STEEPEST-DESCENT ALGORITHM FOR THE JOINT-DIAGONALIZATION TASK WITH A WEIGHTED CONSTRAINT

easy to interpret as it is a simple iterative update or learning rule with a fixed step size. As an alternative to a constant step-size μ the natural gradient method proposed by Amari [11] could be used instead of the absolute gradient. Table 3 gives the general Newton update with penalty terms incorporated to ensure that the Hessian of the constraint \mathbf{H}_4 and the gradient of the constraint \mathbf{G}_4 are accounted for in the optimization process.

V. SIMULATION RESULTS

To investigate the performance of the extended instantaneous BSS algorithm to the convolutive case in the time domain the SGD and Newton algorithm implementations in [1] were slightly altered to the learning rules given in Tables 2 and 3 respectively. As the constraint no longer requires the unknown system

Initialization ($k = 0$) : \mathcal{W}_0
For $k = 1, 2, \dots$
$\Delta\mathcal{W}_k = (\mathbf{H}_W + \alpha\mathbf{H}_4)^{-1}(\mathbf{G}_W + \alpha\mathbf{G}_4)$

TABLE III

NEWTON ALGORITHM FOR THE JOINT-DIAGONALIZATION TASK WITH A WEIGHTED CONSTRAINT

\mathcal{W} to be unitary the constraint was changed to that given in Equation (19). The technique of weighted penalty functions was used to ensure the constraints preventing the trivial solution were met. No longer performing the optimization on the Stiefel manifold as in [1] the SGD and Newton algorithms were changed to better reflect the row normalization constraint for the convolutive case. A two-input-two-output (TITO) two tap FIR known mixing system was chosen and is given below in the z domain.

$$\mathbf{H}(z) = \begin{bmatrix} 1+z^{-1} & -1+z^{-1} \\ -1+z^{-1} & 1+z^{-1} \end{bmatrix}. \quad (20)$$

A first order system was chosen to demonstrate the convergence properties of the SGD compared to the Newton algorithm. The corresponding known unmixing system which would separate mixed signals which are produced by convolving the source signals with the TITO mixing system $\mathbf{H}(z)$ given above is

$$\mathbf{W}_{ideal}(z) = \begin{bmatrix} 1+z^{-1} & 1-z^{-1} \\ 1-z^{-1} & 1+z^{-1} \end{bmatrix}. \quad (21)$$

This is the exact known inverse multiple-input-multiple-output (MIMO) FIR system of the same order. The convolution of these two systems in cascade would ensure the global system $\mathbf{G}(z) = \mathbf{W}_{ideal}(z)\mathbf{H}(z)$ would be a delayed version of the identity, i.e. $z^{-1}\mathbf{I}$. Using matrix multiplication to perform convolution in the time domain, Equation (12) can be used to represent the equivalent structure of Equation (21),

$$\mathcal{W}_{ideal} = \begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \end{bmatrix}. \quad (22)$$

Through empirical analysis we set the parameters $\mu = 0.6$ and $\alpha = 0.2$ and solve the constrained optimization problem given in Equation (19) using the SGD and Newton methods. With the Newton method it should be noted that in Table 3 a closed form expression for the Hessian of the constraint has not been given. For this part numerical differentiation using finite differences was used to obtain \mathbf{H}_4 . A set of $K = 15$ real diagonal square uncorrelated matrices for the unknown source input signals were randomly generated. Using convolution in the time domain a corresponding set of correlation matrices $\mathbf{R}_{\mathcal{X}\mathcal{X},k}^\tau$ for each respective time instant $k = 1, \dots, 15$ at multiple time lags τ were generated for the observed signals. Each optimization algorithm was run ten independent times and convergence graphs were observed and are shown in Figure 1. The various slopes of the different convergence curves of the gradient method depends entirely on the ten different sets of randomly generated diagonal input matrices. Poor initial values for the unknown system lead to convergence to local minima as opposed to the desired global minimum. The initialization of the SGD and Newton algorithms plays an important role in the convergence to either a local or global minimum. Initial values for the estimated un-mixing system \mathcal{W} were randomly generated by adding Gaussian random

variables with standard deviation $\sigma = 0.1$ to the coefficients of the true system. As a possible alternative strategy, a global optimization routine *glcCluster* from TOMLAB [12], a robust global optimization software package, can be used where no initial value for the unknown system is needed. This particular solver uses a global search to approximately obtain the set of all global solutions and then uses a local search method which utilizes the derivative expressions to obtain more accuracy on each global solution. This method will be further analyzed as a future alternative to obtaining additional information on the initial system value.

After convergence of the objective function to an order of magnitude approximately equal to 10^{-34} the unknown de-mixing FIR filter system \mathcal{W} in cascade with the known mixing system $\mathbf{H}(z)$ resulted in a global system which was equivalent to a scaled and permuted version of the true global system $z^{-1}\mathbf{I}$ as can be seen by the following example,

$$\begin{aligned}\mathbf{G}(0) &= \begin{bmatrix} -0.17 & 0.17 \\ 0.19 & -0.19 \end{bmatrix} \times 10^{-14}, \\ \mathbf{G}(1) &= \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \\ \mathbf{G}(2) &= \begin{bmatrix} -0.23 & -0.23 \\ -0.14 & -0.14 \end{bmatrix} \times 10^{-14}.\end{aligned}\quad (23)$$

A first order system has been identified up to an arbitrary global permutation and scaling factor. The TITO system identified above using the optimization algorithms has only 8 unknown variables to identify. We now examine a MIMO FIR mixing system with a higher dimension. Again we have chosen an analytical MIMO multivariate system whose exact FIR inverse is known. The 3rd order mixing system is given below in the z domain

$$\mathbf{H}_{11}(z) = -4 - 4z^{-1} + z^{-2} + z^{-3}, \quad (24)$$

$$\mathbf{H}_{12}(z) = -7 - 7z^{-1} + z^{-3}, \quad (25)$$

$$\mathbf{H}_{21}(z) = 7 - 7z^{-1} + z^{-3}, \quad (26)$$

$$\mathbf{H}_{22}(z) = 9 - 9z^{-1} - z^{-2} + z^{-3}. \quad (27)$$

The corresponding known inverse FIR system of the same order is given below also in the z domain as

$$\mathbf{W}_{11}^{ideal}(z) = \frac{1}{13}\mathbf{H}_{22}(z), \quad (28)$$

$$\mathbf{W}_{12}^{ideal}(z) = -\frac{1}{13}\mathbf{H}_{12}(z), \quad (29)$$

$$\mathbf{W}_{21}^{ideal}(z) = -\frac{1}{13}\mathbf{H}_{21}(z), \quad (30)$$

$$\mathbf{W}_{22}^{ideal}(z) = \frac{1}{13}\mathbf{H}_{11}(z). \quad (31)$$

The convolution of the mixing and un-mixing MIMO FIR systems given in Equations (24-31) gives the identity matrix \mathbf{I} exactly. Again a comparison of

the convergence behavior for both SGD and Newton methods was made using the same methods described for the first order mixing system. Due to the lack of information about the error surface of the multivariate objective function and how it varies with a higher number of dimensions, convergence of the SGD and Newton methods is not as good. In Figure 2 we can see that the SGD method only converges to a local minima of about 10^{-5} when ideally it should be close to 10^{-34} . The Newton method also shows instability. The limiting factor here is that even with relatively good initial values we cannot predict how the error surface of the objective function behaves without more complex analysis on the error surface of the multivariate function. This does not mean that the algorithm does not work for higher order systems but rather more information such as error surface characteristics should be known for a better initialization of the SGD and Newton algorithms and their step-size values given in Tables 2 and 3.

To test the performance of the algorithm on real speech data two independent segments of speech were used as input signals to the MIMO FIR mixing system given in Equation (20). These signals were both 4 seconds long and sampled at 8kHz. The signals were convolutoively mixed with the synthetic mixing system to obtain 2 mixed signals. With the assumption that speech is quasi-stationary over a period of approximately 20ms, the observed mixed signals were buffered and segmented into 401 frames each having 160 samples in length. The non-stationarity assumption assumes that the SOS in each frame does not change. The correlation matrices $\mathbf{R}_{\mathcal{X}\mathcal{X},k}^{\tau}$ can be found via Equations (15,16) for $K = 401$ frames of the two mixed signals. This allows the method of joint diagonalization by minimizing the off-diagonal elements of the correlation matrices of the recovered signals at each respective time lag τ as defined in Equations (17,19). Figure 3 shows the input, mixed and recovered speech signals. A good qualitative recovery is confirmed by subjective listening to the recovered audio signals and inspection of graphs (e) and (f) in Figure 3.

VI. CONCLUSION

A new method for convolutive BSS in the time domain using an existing instantaneous BSS framework has been presented. This method avoids the inherent permutation problem when dealing with solving the convolutive BSS problem in the frequency domain. Optimization algorithms including SGD and Newton methods have been compared for convolutive mixing environments. Future work will be directed at implementing the simulations with recorded data such as speech in real reverberant environments where the orders of the mixing and un-mixing MIMO FIR systems are very high.

REFERENCES

- [1] M. Joho and K. Rahbar, "Joint diagonalization of correlation matrices by using Newton methods with application to blind signal separation," (Rosslyn, VA), pp. 403–407, SAM, Aug 4-6 2002.
- [2] K. Rahbar and J. Reilly, "A New Frequency Domain Method for Blind Source Separation of Convulsive Audio Mixtures," *Submitted to IEEE Trans. on Speech and Audio Processing*, January 2003.
- [3] S. Ikeda and N. Murata, "A method of ICA in Time-Frequency Domain," (Aussois), pp. 365–361, January 1999.
- [4] N. Murata, "An Approach to Blind Source Separation of Speech Signals," *Proceedings of the 8th International Conference on Artificial Neural Networks*, vol. 2, pp. 761–766, September 1998.
- [5] K. Pope and R. Bogner, "Blind signal separation I: Linear, instantaneous combinations," *Digital Signal Processing*, vol. 6, pp. 5–16, Jan. 1996.
- [6] K. Pope and R. Bogner, "Blind signal separation II: Linear, convulsive combinations," *Digital Signal Processing*, vol. 6, pp. 17–28, Jan. 1996.
- [7] M. Feng and K. Kammeyer, "Blind Source separation for Communication Signals Using Antenna Arrays," (Florence, Italy), accepted by ICUPC, Oct. 1998.
- [8] T. Petermann, D. Boss, and K. Kammeyer, "Blind GSM Channel Estimation Under Channel Coding Conditions," (Phoenix, USA), pp. 180–185, December 1999.
- [9] J. Larsen, L. Hansen, T. Kolenda, and F. Nielsen, "Independent Component Analysis in Multimedia Modelling," (Nara, Japan), pp. 687–696, ICA, April 1-4 2003.
- [10] M. Ikram and D. Morgan, "Exploring permutation inconsistency in Blind Separation of Speech Signals in a Reverberant Environment," (Instanbul, Turkey), pp. 1041–1044, ICASSP, June 2000.
- [11] S. Amari, S. Douglas, A.Cichocki, and H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," (Paris, France), pp. 101–104, Proceedings of the IEEE Workshop on Signal Processing Advances in Wireless Communications, 1997.
- [12] K. Holmström, "User's Guide for TOMLAB v4.0." URL: <http://tomlab.biz/docs/tomlabv4.pdf>, Sept. 2 2002.

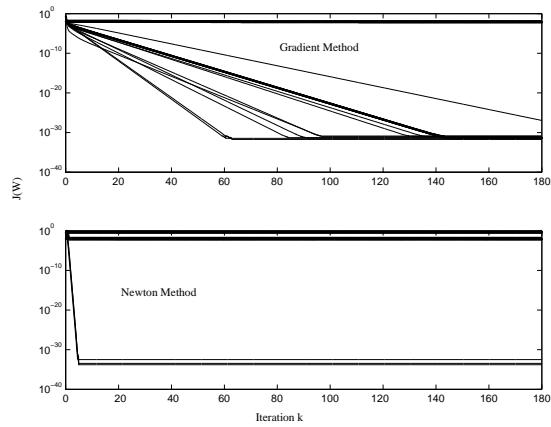


Fig. 1. Convergence of SGD and Newton algorithms for a 1st order MIMO FIR unmixing system.

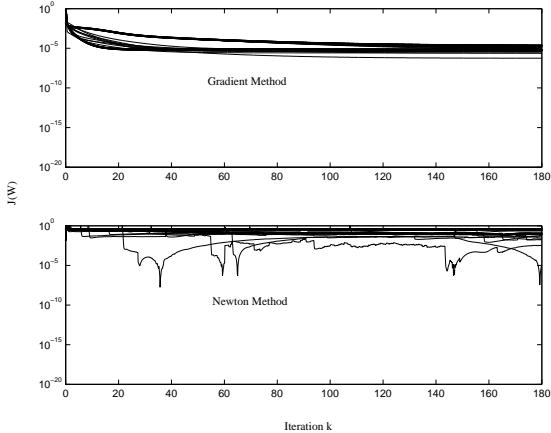


Fig. 2. Convergence of SGD and Newton algorithms for a 3rd order MIMO FIR unmixing system.

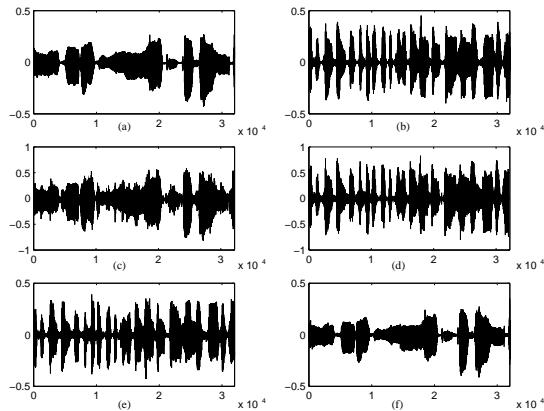


Fig. 3. (a) and (b) are the two original signals, (c) and (d) are the convolutely mixed signals, (e) and (f) are the permuted separated results.