

STATE SPACE BLIND SOURCE RECOVERY FOR MIXTURES OF MULTIPLE SOURCE DISTRIBUTIONS

Khurram Waheed and Fathi M. Salam

Circuits, Systems and Artificial Neural Networks Laboratory
 Department of Electrical and Computer Engineering
 Michigan State University
 East Lansing, MI 48824-1226

ABSTRACT

The paper discusses State Space Blind Source Recovery (BSR) for minimum phase and non-minimum phase mixtures of gaussian and non-gaussian distributions. The State Space Natural Gradient approach results in compact iterative update laws for BSR. Two separate state space algorithms for minimum phase and non-minimum phase mixing environments are presented. The advantages and disadvantages of both algorithms in the context of multiple source distribution mixtures are examined. The presented BSR algorithms require use of nonlinearities, which depend on the distribution of the unknown sources. We propose use of an adaptive non-linearity based on the batch kurtosis of the output. This renders the adaptive estimation of the demixing network to be completely blind.

1. INTRODUCTION

Blind Source Recovery (BSR) is an interesting autonomous (or unsupervised) learning problem that includes well-known adaptive signal processing problems of multi-channel Blind Source Separation (BSS) and Blind Source Deconvolution (BSD) with several potential engineering applications [2,3,6]. The BSR problem denotes recovering original sources from environments that may include convolution, transients, and even possible nonlinearity.

The state space notion provides a compact representation, capable of handling both time delayed and filtered versions of signals in an organized manner [2,3,6]. Unlike the transfer function models of standard dynamic filters, the use of the state-space can result in several generalized, equivalent and efficient internal descriptions of a system. This allows for recovery of original sources independent from (and even in the absence of) environment *identifiability*, i.e. determining the exact (or a specific function of) parameters of the environment. There exist many adaptive network solutions (representations), which succeed in recovering the original signals even in the absence of precise identifiability, termed as *recoverability* [3,6]. Existence and constructions of a theoretical solution to the BSR problem can be easily derived

using the state space, given a structure of the environment [1,3, and the references therein].

Most of the mixtures encountered in practical BSR problems are from sources with a variety of non-gaussian distributions. On the other hand, most noise phenomena or unidentified sources are assumed to possess gaussian distributions. This results in practical situations, where one has to cope with multiple source distribution mixtures including gaussian distributions. We describe the use of linear state space models for these problems.

2. ALGORITHMS FOR LINEAR DYNAMIC CASE

In the linear dynamic case, the environment model is assumed to be of the state space form

$$X_e(k+1) = A_e X_e(k) + B_e s(k) \quad (2.1)$$

$$m(k) = C_e X_e(k) + D_e s(k) \quad (2.2)$$

The proposed feedforward separating network has the state space form

$$X(k+1) = A X(k) + B m(k) \quad (2.3)$$

$$y(k) = C X(k) + D m(k) \quad (2.4)$$

The existence of explicit solutions in this case has been shown by Salam et. al. [1,3, and the references therein]. This existence of solutions ensure that the network has the capacity to compensate for the environment and consequently recover the original signals. *Recoverability* can thus be ensured.

2.1. Minimum Phase Systems

For the linear state-space demixing of minimum phase environments, the update laws have been derived in [3,6]

$$\Delta A = -\eta(k) \lambda_{k+1} X_k^T \quad (2.5)$$

$$\Delta B = -\eta(k) \lambda_{k+1} m_k^T \quad (2.6)$$

$$\Delta C = \eta(k) \left((I - \varphi(y_k) y_k^T) C - \varphi(y_k) X_k^T \right) \quad (2.7)$$

$$\Delta D = \eta(k) (I - \varphi(y_k) y_k^T) D \quad (2.8)$$

where

$$\lambda_k = A_k^T \lambda_{k+1} + C_k^T \frac{\partial L^k}{\partial y_k} \quad (2.9)$$

The above derived update laws form a comprehensive algorithm and provide the update laws for the states, the co-states and all the parametric matrices in the state space model. Note, that the states X_k are computed forward in time, while the co-states λ_k need to be computed backward in time. The invertibility of the state space is guaranteed if the matrix D is “invertible” [3]. In the above derived laws

$\eta(k)$ - represents the learning rate of the algorithm

$\varphi(y)$ - represents an element-wise non-linearity acting individually on each component of the output vector y , where the optimal non-linearity depends on the stochastic distribution of sources to be separated, defined as

$$\varphi(y) = -\frac{\partial p(y)/\partial y}{p(y)} = -\left[\begin{array}{cccc} p'_1(y_1) & p'_2(y_2) & \dots & p'_N(y_N) \\ p_1(y_1) & p_2(y_2) & \dots & p_N(y_N) \end{array} \right]^T$$

Note that the update laws provided above for determining the poles of the demixing system become non-causal for MIMO filtering structures. Practically this can be easily implemented using some time delay and buffer storage memory. A small *latency* in the BSR recovered signal is acceptable as long as the sample delay is fixed for all the recovered signals. Further, the update laws in (2.5) and (2.6) are only locally stable. The state estimation in the above update laws may be replaced by other state estimation methods, e.g., Kalman filter, RLS etc. [2].

On the other hand, the update laws (2.7) and (2.8) require only causal, time forward computations and may be used for determining the transmission zeros of the [2,3,4]. Alternately, using FIR approximation of the demixing system, all the poles are set to zero and the demixing is achieved entirely by using an all zero model. For such an update, the matrices A and B are defined to be in the (controller) canonical form I

$$A = \begin{bmatrix} Z & Z & \dots & Z & Z \\ I & Z & \dots & Z & Z \\ Z & I & \dots & Z & Z \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Z & Z & \dots & I & Z \end{bmatrix}, B = \begin{bmatrix} I \\ Z \\ Z \\ \vdots \\ Z \end{bmatrix} \quad (2.10)$$

where

Z – represents a zero matrix of appropriate size

I – represents an identity matrix of appropriate size

2.2. Non-minimum Phase Systems

For the non-minimum phase mixing environment, the demixing system becomes unstable due to the need of having poles outside the unit circle for cancellation of the non-minimum phase environment transmission zeros. In order to avoid this instability, the natural gradient algorithm may be derived with the constraint that the demixing system is a double-sided FIR filter. This algorithm requires both forward and backward in time propagation by the FIR filter rendering

it to be non-causal. This setup can also be practically realized by using a delayed version of the algorithm. The algorithmic delay and the interim data storage requirements can be minimized by operating both forward and backward in time filters on the same indexed batch of observations and outputs. The double-sided FIR filters can adequately approximate IIR filters at least in the magnitude terms [5, and the reference therein] leading to determination of a demixing system with minimal assumptions. With these constraints the equations derived for iteratively estimating the zeros of the double-sided FIR demixing system are [5]

$$\Delta C_i = \eta(k) [C_i - \varphi(k)u(k-i)] \quad (2.11)$$

$$\Delta D = \eta(k) [D - \varphi(k)u(k)] \quad (2.12)$$

where the matrix C is defined as

$$C = [C_1 \ C_2 \ \dots \ C_{M-1}] \quad (2.13)$$

and $u(k)$ represents a filter for information back propagation of output as

$$\lambda(k) = A' \lambda(k+1) + C' y(k) \quad (2.14)$$

$$u(k) = B' \lambda(k) + D' y(k) \quad (2.15)$$

All the poles of the system are thus fixed at zero which leads to the definition of matrices A and B to be in controller canonical form (2.10). The algorithm requires the augmented matrix $\hat{W} = [D, C]$ to be initialized as a non-singular matrix of polynomials [5].

3. SIMULATION RESULTS

We present two simulation results with 3×3 IIR mixing environments, one for minimum phase while the other for non-minimum phase mixing environment. The environment models for both systems were represented using MIMO controller canonical state space form [3,4,6]. The demixing system is also formulated as a state-space network. For both simulations, the poles of the demixing network are set to zero i.e., the structure for matrices A and B is fixed (see (2.10)), whereas matrices C and D are adaptively updated.

The convergence performance of the algorithm is measured using the multi-channel intersymbol interference benchmark, which is defined in [2,3] as

$$ISI_k = \sum_{i=1}^N \frac{\left| \sum_j \sum_p |G_{pij}| - \max_{p,j} |G_{pij}| \right|}{\max_{p,j} |G_{pij}|} + \sum_{j=1}^N \frac{\left| \sum_i \sum_p |G_{pij}| - \max_{p,i} |G_{pij}| \right|}{\max_{p,i} |G_{pij}|} \quad (2.16)$$

$G(z) = H(z) * \bar{H}(z)$ represents Global Transfer Function,

$\bar{H}(z) = [A_e, B_e, C_e, D_e]$ – Transfer Function of Environment

$H(z) = [A, B, C, D]$ – Transfer Function of Network,

3.1. Choice of Nonlinearity

As outlined above, the optimal nonlinearities for the update laws depend on the density function of the sources to be

separated, which upon successful convergence of the algorithm is similar to the density of the separated outputs. We have used an adaptive nonlinearity, which relies on the batch kurtosis of the output of the demixing system. This nonlinearity, defined below, converges to the optimal nonlinearity for the demixing system as the network's outputs approach stochastic independence.

$$\varphi(y) = \begin{cases} y - \alpha \tanh(\beta y) & \kappa_4(y) \leq -\gamma \\ y & \text{for } |\kappa_4(y)| < \gamma \\ \alpha \tanh(\beta y) & \kappa_4(y) \geq \gamma \end{cases} \quad (2.17)$$

where $\kappa_4(y)$ - batch kurtosis of the output signals

3.2. Simulation I: Minimum Phase Mixing

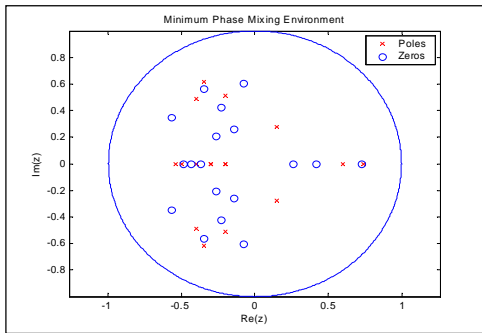
This simulation presents the results for a 3×3 minimum phase IIR filtering environment model

$$\sum_{j=0}^{m-1} A_j m(k-i) = \sum_{i=0}^{m-1} B_i s(k-i) + v(k) \quad (2.18)$$

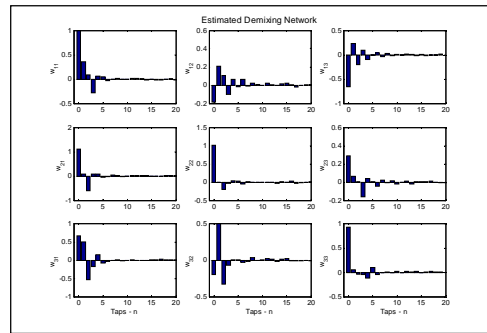
$$A_0 = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \end{bmatrix}, A_1 = \begin{bmatrix} 0.5 & 0.8 & -0.7 \\ 0.8 & 0.3 & -0.2 \\ -0.1 & -0.5 & 0.4 \end{bmatrix}, A_2 = \begin{bmatrix} 0.06 & 0.4 & -0.5 \\ 0.16 & -0.1 & -0.4 \\ -0.3 & -0.06 & 0.3 \end{bmatrix}$$

$$B_0 = \begin{bmatrix} 1 & 0.6 & 0.8 \\ 0.3 & 1 & 0.1 \\ 0.6 & -0.8 & 1 \end{bmatrix}, B_1 = \begin{bmatrix} 0.5 & 0.5 & 0.6 \\ -0.3 & 0.2 & -0.3 \\ -0.2 & -0.43 & 0.6 \end{bmatrix}, B_2 = \begin{bmatrix} .125 & 0.06 & 0.2 \\ -0.1 & 0 & 0.4 \\ 0.08 & -0.13 & 0.3 \end{bmatrix}$$

$v(k)$ - additive gaussian noise

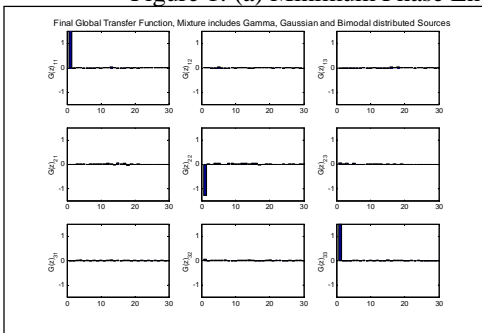


(a)

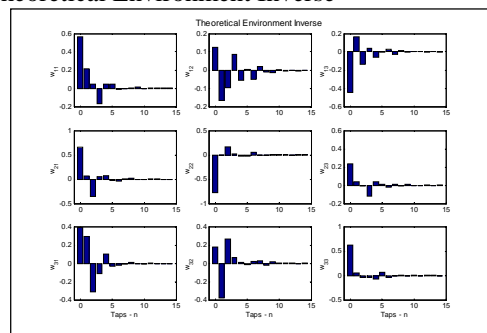


(b)

Figure 1: (a) Minimum Phase Environment, (b) Theoretical Environment Inverse



(a)



(b)

Figure 2: (a) Final Global Transfer Function (b) Estimated Demixing Network

The theoretical inverse of this IIR mixing environment will be an IIR filter of dimension 10 or larger. For the presented simulations all the poles were set to zero. The matrix C was initialized with very small random numbers with a variance of 0.1, while the matrix D was initialized to be non-singular. The number of taps per filter was chosen to be 31. Note for this equivariant minimum phase setup, any over-determination in the number of taps does not effect the performance of the algorithm. Presented in figures 1 & 2 are the instantaneous results after 20000 iterations with mixtures of multiple distributions, and use of adaptive non-linearity with the update laws (2.7) and (2.8).

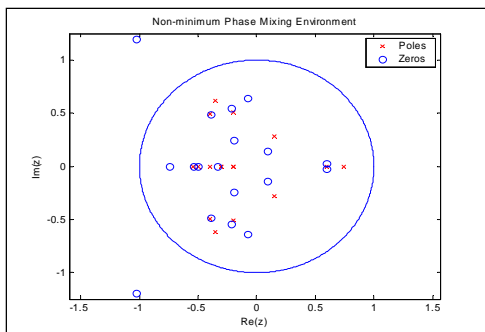
3.3. Simulation II: Non-minimum Phase Mixing

This simulation presents the results for a 3×3 non-minimum phase IIR filtering environment model as in (2.18), where

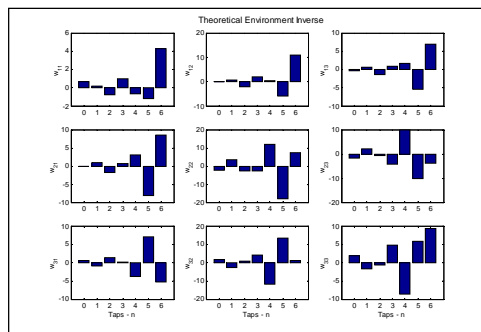
$$A_0 = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \end{bmatrix}, A_1 = \begin{bmatrix} 0.5 & 0.8 & -0.7 \\ 0.8 & 0.3 & -0.2 \\ -0.1 & -0.5 & 0.4 \end{bmatrix}, A_2 = \begin{bmatrix} 0.06 & 0.4 & -0.5 \\ 0.16 & -0.1 & -0.4 \\ -0.3 & -0.06 & 0.3 \end{bmatrix}$$

$$B_0 = \begin{bmatrix} 1 & 0.6 & 0.8 \\ 0.5 & 1 & 0.7 \\ 0.6 & 0.8 & 1 \end{bmatrix}, B_1 = \begin{bmatrix} 0.5 & 0.7 & 0.16 \\ 0.7 & 0.2 & -0.3 \\ -0.2 & 0.53 & 0.6 \end{bmatrix}, B_2 = \begin{bmatrix} .425 & 0.3 & 0.7 \\ -0.1 & 0 & -0.4 \\ 0.08 & -0.13 & 0.3 \end{bmatrix}$$

$v(k)$ - additive gaussian noise

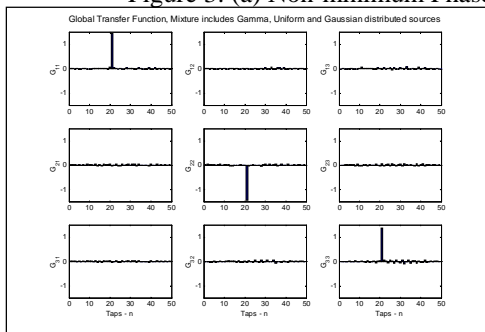


(a)

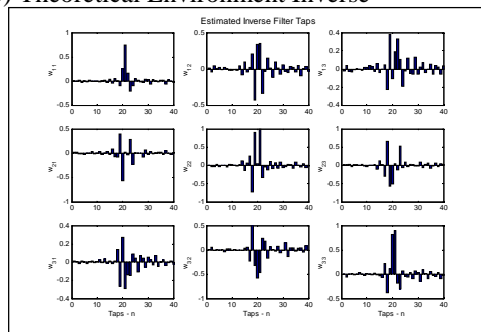


(b)

Figure 3: (a) Non-minimum Phase Environment, (b) Theoretical Environment Inverse



(a)



(b)

Figure 4: (a) Global Transfer Function, (b) Estimated Demixing Network

The theoretical inverse of this IIR mixing environment will also be an IIR filter of dimension 10 or larger. Also two poles of the intended demixing network need to be outside the unit circle, making it an unstable filter. However, the network is implemented as a doubly finite FIR inverse filter with 41 taps and supplied with mixtures of multiple source distributions. Matrix \hat{W} is initialized to have one unity tap in each diagonal filter, while other taps are set to small random numbers. Instantaneous update results after 40,000 iterations are shown in Figures 3 and 4. The convergence for non-minimum phase systems is however relatively slow with adaptive non-linearity.

4. CONCLUSIONS

This paper demonstrates the use of the State Space BSR algorithms for both minimum and non-minimum phase mixing environments. Both algorithms are able to converge for all evaluated mixtures of non-gaussian source distributions. The minimum phase algorithm provides equivariant performance, independent of over-estimation effects in the order of demixing model, even when one of the inputs is gaussian or “close to gaussian” (which may represent a combination of several unmodeled sources).

The non-minimum phase algorithm is also able to converge for all non-gaussian distributions independent of the order of selected filter order, provided it is larger than the minimum required filter order. However, if the order is incorrectly over-estimated, the algorithm exhibits excessive tap noise behavior for only gaussian type distributions while

it converges successfully for all other constituent non-gaussian source distributions in a mixture.

5. REFERENCES

- [1] F. M. Salam and G. Erten: “Blind Signal Separation and Recovery in dynamic Environments”, *3rd IEEE wkshp. On nonlinear signal and image processing 1997*, Mackinac Island, MI - September 8-10, 1997.
- [2] L. Zhang and A. Cichocki: “Blind deconvolution of dynamical systems: A state space approach”, *Journal of Signal Processing*, Vol. 4, No. 2, Mar. 2000, pp. 111-130.
- [3] F. M. Salam, G. Erten and K. Waheed: “Blind Source Recovery: Algorithms for Static and Dynamic Environments”; *Int'l Joint Conf. on Neural Networks*, July 14-19, 2001-Washington, D.C.; Vol. 2, Page(s) 902-907
- [4] K. Waheed and F. M. Salam: “Blind Source Recovery: Some Implementation and Performance Issues”; *44th IEEE Midwest Symp. On Circuits and Systems*; August 14-17 2001- Dayton Ohio. Vol. 2, Page(s) 580-583
- [5] K. Waheed and F. M. Salam: “Blind Source Recovery: A State Space Framework”; *Technical Memo No. ECE-CSANN'002, Dept of ECE, Michigan State University*, August 2001.
- [6] F. M. Salam and K. Waheed: “State Space Feedforward and Feedback Structures for Blind Source Recovery”, *3rd Int'l Conf. on ICA and BSS*, December 9-12, 2001 - San Diego, CA; Page(s) 248-253