Journal of Machine Learning Research 4 (2003) 1411-1446 Submitted 10/02; Published 12/03

Blind Source Recovery: A Framework in the State Space

Khurram Waheed WAHEED@ENGINEERING.SDSU.EDU
Department of Electrical and Computer Engineering
San Diego State University
San Diego, CA 91913-1309, USA

Fathi M. Salem SALEM@EGR.MSU.EDU
Department of Electrical and Computer Engineering
Michigan State University
East Lansing, MI 48824-1226, USA

Editors: Te-Won Lee, Jean-François Cardoso, Erkki Oja and Shun-ichi Amari

Abstract

Blind Source Recovery (BSR) denotes recovery of original sources/signals from environments that may include convolution, temporal variation, and even nonlinearity. It also infers the recovery of sources even in the absence of precise environment identifiability. This paper describes, in a comprehensive fashion, a generalized BSR formulation achieved by the application of stochastic optimization principles to the Kullback-Liebler divergence as a performance functional subject to the constraints of the general (i.e., nonlinear and time-varying) state space representation. This technique is used to derive update laws for nonlinear time-varying dynamical systems, which are subsequently specialized to time-invariant and linear systems. Further, the state space demixing network structures have been exploited to develop learning rules, capable of handling most filtering paradigms, which can be conveniently extended to nonlinear models. In the special cases, distinct linear state-space algorithms are presented for the minimum phase and non-minimum phase mixing environment models. Conventional (FIR/IIR) filtering models are subsequently derived from this general structure and are compared with material in the recent literature. Illustrative simulation examples are presented to demonstrate the *online* adaptation capabilities of the developed algorithms. Some of this reported work has also been implemented in dedicated hardware/software platforms.

Keywords: Blind Source Recovery, Feedforward Network, Feedback Network, Kullback-Liebler Divergence, State Space Representation.

1. Introduction

Blind Source Recovery (BSR) is informally described as follows: several unknown but stochastically independent temporal signals propagate through a natural or synthetic, dynamic mixing and filtering environment. By observing only the outputs of this environment, a system (e.g., a filter bank, a neural network, or a device) is constructed to counteract, to the extent

© 2003 Khurram Waheed and Fathi M. Salem

possible, the effect of the environment and adaptively recover the best estimate of the original signals. The adaptive approach is a form of unsupervised or autonomous learning.

Blind Source Recovery (BSR) is a challenging signal processing formulation that in the linear case includes the well-known adaptive filtering sub-problems of blind source separation and blind source deconvolution. Blind Source Separation (BSS) is the process of recovering a number of original stochastically independent sources/signals when only their *linear static* mixtures are available. Blind Source Deconvolution (BSD) deals with deconvolving the effects of a temporal or a spatial mixing *linear* filter on signals without a priori information about the filtering medium/network or the original sources/signals (Haykin 2000; Salam & Erten, 1999a; Cichocki & Amari, 2002).

For this unsupervised adaptive filtering task, only the property of signal independence is assumed. No additional a priori knowledge of the original signals/sources is required. BSR requires few assumptions and possesses the self-learning capabilities, which render such networks attractive from the viewpoint of real world applications where on-line training is often desired (Herault & Jutten, 1986; Torkkola, 1996; Hyvarinen, et al., 1997 & 2001; Erten & Salam, 1999). The challenges for the BSR reside in the development of sound mathematical analyses and a framework capable of handling a variety of diverse problems. While other approaches to BSR have used more conventional signal processing structures, we propose a comprehensive BSR framework based on multi-variable state-space representations, optimization theory, the calculus of variations and higher order statistics. For some results obtained using implementation of the BSR framework in real-time hardware, see (Gharbi & Salam, 1995; Erten & Salam, 1999).

As a background, interest in the field of blind source separation and deconvolution has grown dramatically during recent years, motivated to a large extent by its similarity to the mixed signal separation capability of the human brain. The brain makes use of unknown parallel nonlinear and complex dynamic signal processing with auto-learning and self-organization ability to perform similar tasks. The peripheral nervous system integrates complex external stimuli and endogenous information into packets, which are transformed, filtered and transmitted in a manner that is yet to be completely understood. This complex mixture of information is received by the central nervous system (brain), split again into original information, and relevant information relayed to various sections of cerebral cortex for further processing and action (Nicholls, 2001).

BSR is valuable in numerous applications that include telecommunication systems, sonar, acoustics and radar systems, mining and oil drilling systems, image enhancement, feature extraction and biomedical signal processing. Consider, for example, the audio and sonar applications where the original signals are sounds, and the mixed signals are the output of several microphones or sensors placed at different vantage points. A network would receive, via each sensor, a mixture of original sounds that usually undergo multi-path delays. The network's role in this scenario will be to dynamically reproduce, as closely as possible, the original signals. These separated signals can subsequently be channeled for further processing or transmission. Similar application scenarios can be described in situations involving measurement of neural, cardiac or other vital biological parameters, communication in noisy environments, engine or plant diagnostics, and cellular mobile communications to name a few (see, Lee et al., 1997; Lambert & Bell 1997; Girolami 1999; Haykin, 2000; Hyvarinen et al., 2001; Cichocki & Amari, 2002).

1.1. Modeling: the State Space Framework

Development of adequate models for the environment that include time delays or filtering, multi-path effects, time-varying parameters and/or nonlinear sensor dynamics is important as representative of desired practical applications. The choice of inadequate models of the

environment would result in highly sensitive and non-robust processing by a network. Indeed, in order to render the network operable in real world scenarios, robust operations must account for parameter variations, dynamic influences and signal delays that often result in asynchronous signal propagation. The environment needs to be modeled as a rich dynamic linear (or even nonlinear) system.

The state space formulation provides a general framework capable of dealing with a variety of situations. The multi-variable state space provides a compact and computationally attractive representation of multi-input multi-output (MIMO) filters. This rich state space representation allows for the derivation of generalized iterative update laws for the BSR. The state notion abridges weighted past as well as filtered versions of input signals and can be easily extended to include non-linear networks. There are several reasons for choosing this framework.

- State space models give an efficient internal description of a system. Further, this choice allows various equivalent state space realizations for a system, more importantly being the canonical observable and controllable forms. Transfer function models, although equivalent to linear state space models when initial conditions are zero, do not exploit any internal features that may be present in the real dynamic systems.
- The inverse for a state space representation is easily derived subject to the *invertibility* of the instantaneous relational mixing matrix between input-output – in case this matrix is not square; the condition reduces to the existence of the pseudo-inverse of this matrix. This feature quantifies and ensures *recoverability* of original sources provided the environment model is invertible.
- Parameterization methods are well known for specific classes of models. In particular, the state space model enables much more general description than standard finite/infinite impulse response (FIR/IIR) convolutive filtering. All known (dynamic) filtering models, like AR, MA, ARMA, ARMAX and Gamma filters, can be considered as special cases of flexible state space models.
- The linear state space formulation of a problem is conveniently extendable to include non-linear models and specific component dynamics. The richness of state space models also allows them to accommodate a direct representation for generalized non-linear model representations that include neural networks, genetic algorithms, etc.

This paper presents a generalized BSR framework based on the multi-variable canonical state space representation of both the mixing environment and the demixing system. The cases of minimum phase and non-minimum phase mixing environments have been dealt separately. The demixing network can have either a feedforward or a feedback state space configuration. Various filtering paradigms have been consequently derived as special cases from the proposed BSR framework. Simulation results verifying the theoretical developments have also been included.

1.2. Adaptive Framework for Blind Source Recovery

In the most general setting, the mixing/convolving environment may be represented by an unknown dynamic process $\bar{\mathcal{H}}$ with inputs being the n-d independent sources $\underline{\boldsymbol{s}}$ and the outputs being the m-d measurements $\underline{\boldsymbol{m}}$. In this extreme case, no structure is assumed about the model of the environment.

The environment can also be modeled as a dynamic system with a structure and fixed but unknown parameters. The processing network $\mathcal{H}$ must be constructed with the capability to recover or estimate the original sources. For the linear case, this is equivalent to computing the

"inverse" (or the "closest to an inverse") of the environment model without assuming any knowledge of the mixing environment or the distribution structure of the unknown sources.

It is possible that an augmented network be constructed so that the inverse of the environment is merely a subsystem of the network with learning. In this case, even if the environment is unstable (e.g., due to existence of non-minimum phase zeros in a linear state space model), the overall augmented network may represent a nonlinear adaptive dynamic system, which may converge to the desired parameters as a stable equilibrium point. Thus achieving the global task of blind identification, see (Salam & Erten, 1997 & 1999a; Zhang et al., 1999 & 2000b; Waheed & Salam 2001, 2002c).

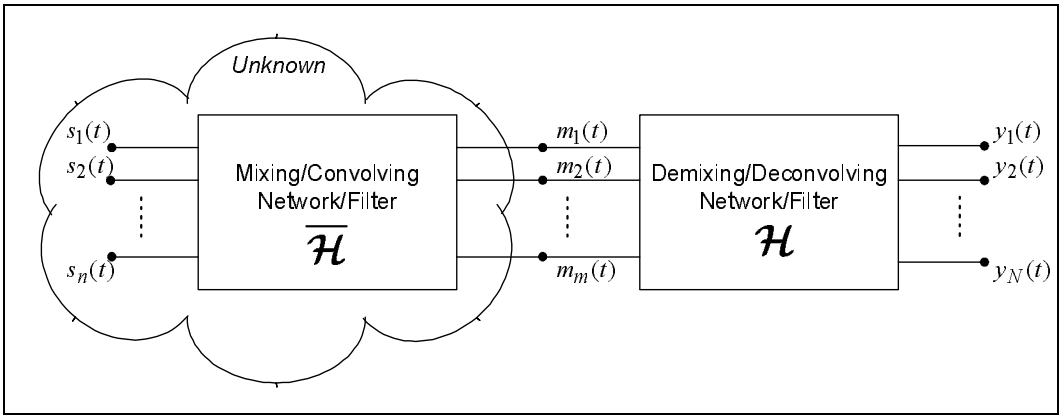


Figure 1. General framework for Blind Source Recovery

For the linear filtering environments, this problem may be presented as

$$\underline{Y} = \mathcal{H} * \underline{m} = \mathcal{H} * \overline{\mathcal{H}} * \underline{s} = P * D * \underline{s} \tag{1}$$

where

P – is a generalized permutation matrix

D – is a diagonal matrix of filters with each diagonal filter having only one non-zero tap.

$*$ – represents ordinary matrix multiplication for the static mixing case, while it represents polynomial matrix multiplication for the multi-tap deconvolution case.

For notational and mathematical convenience, we will derive all BSR algorithms for the case where $n = m = N$. This convenient choice allows for unambiguous mathematical manipulations in the derivation for all the algorithms. The under-complete mixture case of BSR, where $m > n$, can be easily adopted to the fore mentioned case, by appropriately discarding some measurements, or by using pre-processors, e.g., Principal Component Analysis (PCA), eigen decomposition etc. (see, Hyvarinen et al., 2001; Cichocki & Amari, 2002). On the other hand, the over-complete mixture case for $m < n$, in general, is not tractable and cannot be handled by the proposed algorithms in a straightforward fashion.

2. Formulation of the Optimization Structure

This section formally describes the BSR optimization framework, see (Salam & Erten, 1999a; Salam et al., 2001) for more details. We first discuss the performance functional used for the derivation of the BSR update laws based on the stochastic gradient update rules and constrained multivariable optimization. This performance functional is a well known information-theoretic "distance" measure, appropriate for quantifying the mutual *dependence* of signals in a mixture. Further, we present our optimization framework based on this performance functional and derive update algorithms first for a more general non-linear dynamic structure, which subsequently is specialized to the case of a linear dynamical state space model.

2.1. The Performance Functional

We employ the divergence of a random output vector y (Kullback-Liebler divergence) as our performance measure. In the continuous case (Kendall & Stuart, 1977) this relation is given by

$$L(y) = \int_{y \in Y} p_y(y) \ln \left| \frac{p_y(y)}{\prod_{i=1}^{n} p_{y_i}(y_i)} \right| dy \qquad (2)$$

where

$p_y(y)$ is the probability density function of the random output vector y

$p_{y_i}(y_i)$ is the probability density function of the i^{th} component of the output vector y

This functional $L(y)$ is a "distance" measure with the following properties

i) $L(y) \geq 0$

ii) $L(y) = 0$ iff $p_y(y) = \prod_{i=1}^{n} p_{y_i}(y_i)$

This measure can provide an estimate of the degree of dependence among the various components of the recovered output signal vector and is an appropriate functional to be used in the optimization framework of the problem.

For the discrete case, the functional becomes

$$L(y) = \sum_{y \in Y} p_y(y) \ln \left| \frac{p_y(y)}{\prod_{i=1}^{n} p_{y_i}(y_i)} \right|$$

The functional can be further simplified assuming the statistical properties of the output signals to be ergodic:

$$L(y) = \sum_k L(y(k)) = \sum_k p_y(y(k)) \ln \left| \frac{p_y(y(k))}{\prod_{i=1}^{n} p_{y_i}(y_i(k))} \right| \tag{3}$$

Further simplification can be achieved using the assumption that as the algorithm approaches convergence the components of the output vector $y(k)$ will become statistically less dependent, ultimately approaching independence as close as possible. Therefore the above functional can be re-written in its compact mutual information form using the entropy of signals, namely,

$$L(y(k)) = -H(y(k)) + \sum_{i=1}^{n} H_i(y_{i(k)}) \tag{4}$$

where

$H(y(k))$: the Entropy of the signal vector $y(k)$, given by

$$H(y) = -E\left[\ln\left|p_y(y)\right|\right] = \begin{cases} -\int_{y \in Y} p_y(y) \ln\left|p_y(y)\right| dy & \text{Continuous Case} \\ -\sum_{y \in Y} p_y(y) \ln\left|p_y(y)\right| & \text{Discrete Case} \end{cases}$$

$H_i(y_{i(k)})$: the Marginal entropy of a component signal $y_{i(k)}$

2.2. Algorithms for the Nonlinear Dynamic Case

Assume that the environment can be modeled as the following nonlinear discrete-time dynamic forward model

$$X_e(k+1) = f_e(X_e(k), s(k), h_1) \tag{5}$$

$$m(k) = g_e(X_e(k), s(k), h_2) \tag{6}$$

where

$s(k)$: n-d vector of original source signals

$m(k)$: m-d vector of measurements

$X_e(k)$: M_e -d state vector for the environment

h_1 : constant parameter vector (or matrix) of dynamic state equation

h_2 : constant parameter vector (or matrix) of output equation

$f_e(.)$ and $g_e(.)$: differentiable nonlinear functions that specify the structure of the environment.

Further it is assumed that existence and uniqueness of solutions are satisfied for any given initial conditions $X_e(t_o)$ and sources $s(k)$, thus a Lipschitz condition on $f_e(.)$ is satisfied (Khalil, 2002)

The processing demixing network model may be represented by a dynamic feedforward or feedback network. Focusing on a feedforward network model, we assume the network to be represented by

$$X(k+1) = f(X(k), m(k), w_1) \tag{7}$$

$$y(k) = g(X(k), m(k), w_2) \tag{8}$$

where

$m(k)$: m-d vector of measurements

$y(k)$: N-d vector of network output

$X(k)$: M-d state vector for the processing network (Note, that M and M_e may be different)

w_1 : parameters of the network state equation

w_2 : parameters of the network output equation

$f(.)$ and $g(.)$: differentiable nonlinear functions defining the structure of the demixing network

The assumption of existence and uniqueness of solutions of the nonlinear difference equations is also assumed for the network model for any given initial conditions $X(t_o)$ and measurement vector $m(k)$.

In order to derive the update law, we abuse the notation for the sake of convenience so that $y(k)$ in (4) is represented as y_k and $L(y(k))$ is generalized to $L^k(y_k)$ so that the functional may also be a function of the time index k. Thus, we formulate the following constrained optimization problem (Salam & Erten 1999a; Salam et al., 2001)

Minimize

$$J_o(w_1, w_2) = \sum_{k=k_0}^{k_1} L^k(y_k) \tag{9}$$

subject to

$$X_{k+1} = f^k(X_k, m_k, w_1)$$

$$y_k = g^k(X_k, m_k, w_2)$$

with the initial condition X_{k_o}, where $[k_0, k_1)$ is the discrete-time frame used for the computation of the cost functional. Thus, the augmented cost functional to be optimized becomes

$$J(w_1, w_2) = \sum_{k=k_0}^{k_1} L^k(y_k) + \lambda_{k+1}^T (f^k(X_k, m_k, w_1) - X_{k+1}) \tag{10}$$

where λ_k is the Lagrange variable (Lewis & Syrmos, 1995). Define the Hamiltonian as

$$\mathrm{H}^k = L^k(y_k) + \lambda_{k+1}^T f^k(X_k, m_k, w_1) \tag{11}$$

Consequently, the *necessary* conditions for optimality are

$$X_{k+1} = \frac{\partial \mathrm{H}^k}{\partial \lambda_{k+1}} = f^k(X_k, m_k, w_1) \tag{12}$$

$$\lambda_k = \frac{\partial \mathrm{H}^k}{\partial X_k} = (f_{X_k}^k)^T \lambda_{k+1} + \frac{\partial L^k}{\partial X_k} \tag{13}$$

and the change in "weight" parameters become

$$\Delta w_1 = -\eta \frac{\partial \mathrm{H}^k}{\partial w_1} = -\eta (f_{w_1}^k)^T \lambda_{k+1} \tag{14}$$

$$\Delta w_2 = -\eta \frac{\partial \mathrm{H}^k}{\partial w_2} = -\eta \frac{\partial L^k}{\partial w_2} \tag{15}$$

where

$f_a^k \triangleq \frac{\partial f^k}{\partial a}$: represents the partial derivative w.r.t. the parameter a, in the limit.

η : represents a positive learning rate which may be variable.

2.3. Algorithms for the Linear Dynamic Case

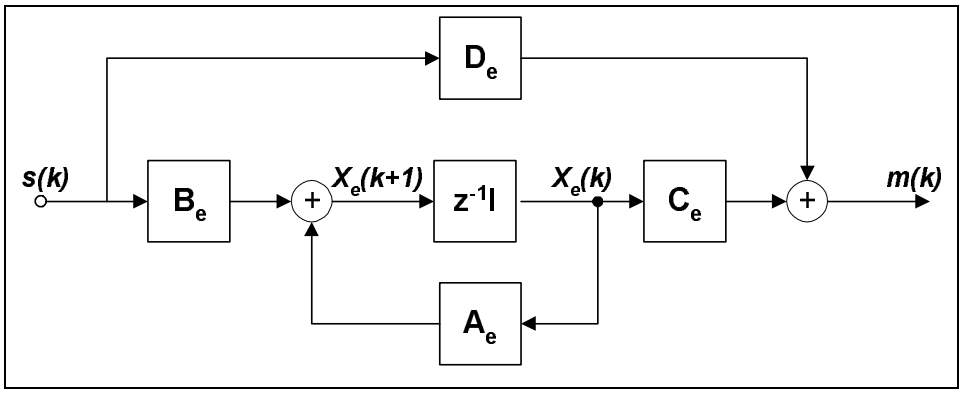


Figure 2. A State Space Mixing Environment

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

$$X_e(k+1) = A_e X_e(k) + B_e s(k) \tag{16}$$

$$m(k) = C_e X_e(k) + D_e s(k) \tag{17}$$

In this case the feedforward separating network will attain the state space form

$$X(k+1) = A\, X(k) + B\, m(k) \tag{18}$$

$$y(k) = C\ X(k) + D\ m(k) \tag{19}$$

The existence of an explicit solution in this case has been shown by (Salam, 1993; Salam & Erten, 1997; Salam et al., 1997a). This existence of solution ensures that the network has the capacity to compensate for the environment and consequently recover the original signals, see Figures 2 and 3.

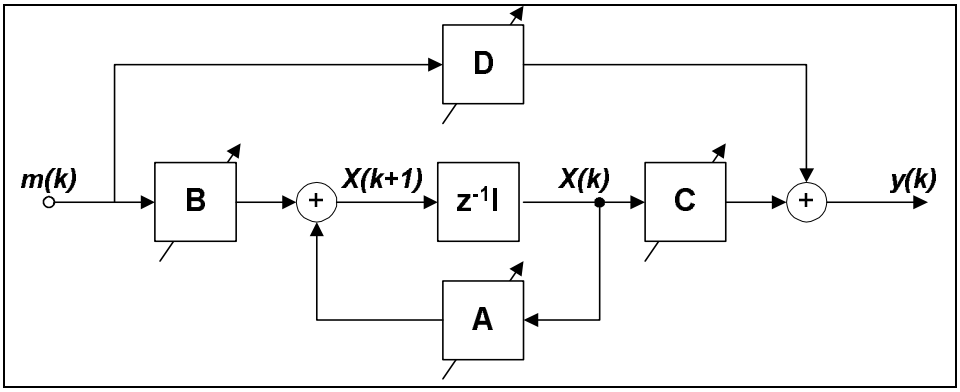


Figure 3. A State Space Demixing Network

Specializing the general BSR algorithm to the linear network case, one forms the (Kullback-Liebler divergence) performance ("distance") measure, see (9) as

$$J_o(A,B,C,D) = \sum_{k=k_0}^{k_1} L^k(y_k) \tag{20}$$

and the augmented cost functional to be optimized becomes

$$J(A,B,C,D) = \sum_{k=k_0}^{k_1} L^k(y_k) + \lambda_{k+1}^T(A\ X_k + B\ m_k - X_{k+1}) \tag{21}$$

Again, define the Hamiltonian as

$$\mathrm{H}^k = L^k(y_k) + \lambda_{k+1}^T(A\ X_k + B\ m_k) \tag{22}$$

For the linear time-invariant case, the ordinary stochastic gradient update laws are given as (Gharbi & Salam, 1997; Salam & Erten, 1999a; Salam et al., 2001)

$$X_{k+1} = \frac{\partial \mathrm{H}^k}{\partial \lambda_{k+1}} = A\ X_k + B\ m_k \tag{23}$$

$$\lambda_k = \frac{\partial \mathrm{H}^k}{\partial X_k} = A_k^T \lambda_{k+1} + C_k^T \frac{\partial L^k}{\partial y_k} \tag{24}$$

$$\Delta A = -\eta \frac{\partial \mathrm{H}^k}{\partial A} = -\eta \lambda_{k+1} X_k^T \tag{25}$$

$$\Delta B = -\eta \frac{\partial \mathrm{H}^k}{\partial B} = -\eta \lambda_{k+1} m_k^T \tag{26}$$

$$\Delta C = -\eta \frac{\partial \mathrm{H}^k}{\partial C} = -\eta \frac{\partial L^k}{\partial C} = -\eta \varphi(y) X^T \tag{27}$$

$$\Delta D = -\eta \frac{\partial \mathrm{H}^k}{\partial D} = -\eta \frac{\partial L^k}{\partial D} = \eta([D]^{-T} - \varphi(y) m^T) \tag{28}$$

The above derived update laws form a comprehensive theoretical stochastic gradient algorithm and provide the dynamics of the network states X_k, the generated co-states λ_k and all the update laws of the parametric matrices of the state space. The invertibility of the state space as shown in (Salam, 1993; Salam & Erten, 1997) is guaranteed if (and only if) the matrix D is invertible. In the above derived laws:

η : a positive learning rate of the algorithm, which may be variable.

$[D]^{-T}$: represents the transpose of the inverse of the matrix D if it is a square matrix or the transpose of its pseudo-inverse in case D is not a square matrix (Salam & Erten, 1997).

$\varphi(y)$: represents a vector of the usual nonlinearity or score function (Gharbi & Salam, 1997; Salam & Erten, 1999; Waheed & Salam 2002, 2002a, 2003) which acts individually on each component of the output vector y, with each component, say $\varphi(y_i)$, is given as

$$\varphi(y_i) = -\frac{\partial \log p(y_i)}{\partial y_i} = -\frac{\partial p(y_i) \big/ \partial y_i}{p(y_i)} \tag{29}$$

where

$p(y_i)$: is the (estimate) of the probability density function of each source (Kendall & Stuart, 1977).

The update law in (27) and (28), also see (Gharbi & Salam, 1997; Salam & Erten, 1999; Waheed & Salam 2002d, 2003), is similar to the standard gradient descent results (Amari et al., 1997; Zhang & Cichocki 2000b), indicating its optimality for the Euclidean parametric structure. The update law provided above although non-causal for the update of parametric matrices A and B, can be easily implemented using frame-buffering and memory storage (Haykin, 1996 & 2000). An inherent delay or latency in the recovered signals is related to the length of the buffer and the dimension of the state space.

3. Extensions to the Natural Gradient

In this section, we extend the linear state space algorithms for the problem of Blind Source Recovery (BSR) derived in the previous section using the natural gradient (Amari 1985, 1996, 1998; Amari & Nagaoka, 1999; Cardoso 1997). We present specialized algorithms for the class of minimum phase mixing environments both in feedforward and feedback state space configuration (Salam et al., 2001; Waheed & Salam, 2001). For the non-minimum phase mixing environment, the requisite demixing system becomes unstable due to the presence of poles outside the unit circle. These unstable poles are required to cancel out the non-minimum phase transmission zeros of the environment. In order to avoid instability due to the existence of these poles outside the

unit circle, the natural gradient algorithm may be derived with the constraint that the demixing system is a double sided FIR filter, i.e., instead of trying to determine the IIR inverse of the environment, we will approximate the inverse using an all zero non-causal filter (Amari et al., 1997; Waheed & Salam, 2002d). The double-sided filters have been proven to adequately approximate IIR filters at least in the magnitude terms with a certain associated delay (Benveniste et al., 1980).

The minimum phase algorithms have lesser computational requirements and exhibit asymptotically better convergence characteristics, compared to the algorithm for the non-minimum phase mixing environments, when applied to the same class of minimum phase mixing systems. The minimum phase algorithm also exhibits more robust performance with respect to the choice of the order of demixing/deconvolving filter per channel, and the multi-source distribution composition of mixtures that may include at most one pure gaussian source. The non-minimum phase algorithm gives good performance for the non-gaussian mixtures and/or the non-gaussian contents of a complex multi-source distribution mixture. The gaussian source although separated from other sources may still be auto-convolved (Note that the sum of co-centric copies of a gaussian distributions is also gaussian). Also notice that the presented algorithms in this paper focus primarily on adaptive determination of zeros. The poles are kept fixed either at zero or at randomly selected stable locations. Otherwise, they are assumed known or estimated using other methods, e.g., adaptive state estimation techniques such as the kalman filter (Zhang et al., 2000, 2000b) etc.

3.1. Theorem 3.1: Feedforward Minimum Phase Demixing Network

Assume the (mixing) environment is modeled by a multiple-input, multiple-output (MIMO) minimum phase transfer function. Then, the update laws for the zeros of the feedforward state space demixing network using the natural gradient are given by

$$\Delta C(k) = \eta\Big((I_N - \varphi(y(k))y^T(k))C(k) - \varphi(y(k))X^T(k)\Big) \tag{30}$$

$$\Delta D(k) = \eta(I_N - \varphi(y(k))y^T(k))D(k) \tag{31}$$

where $\varphi(y)$ is a nonlinear score function given by (29)

Proof

For the feedforward demixing network, its linear state space representation is assumed to be as in (18)-(19). We present a formal formulation for deriving the feedforward update laws for the problem using the output equation (19) and the natural gradient. This leads to modified update laws for the matrices C and D. Further, these new update laws based on the natural gradient have better convergence performance compared to those based on the stochastic gradient (27) and (28). Note that in the following derivation, instantaneous time index k has been dropped for convenience

Defining augmented vectors $\tilde{y}$ and $\tilde{x}$, and the matrix $\tilde{W}$ as

$$\tilde{y} = \begin{bmatrix} y \\ X \end{bmatrix} \qquad \tilde{x} = \begin{bmatrix} m \\ X \end{bmatrix} \qquad \tilde{W} = \begin{bmatrix} D & C \\ 0 & I_{(L-1)m} \end{bmatrix}$$

so the augmented output equation becomes

$$\tilde{y} = \tilde{W}\,\tilde{x} \tag{32}$$

The update law for this augmented parameter matrix $\tilde{W}$ is similar in form to (28) or the stochastic gradient law for the static mixing case. Thus the update is

$$\Delta\tilde{W} = \eta\left[\tilde{W}^{-T} - \varphi(\tilde{y})\tilde{x}^T\right] \tag{33}$$

where

$$\tilde{W}^T = \begin{bmatrix} D^T & 0 \\ C^T & I_{(L-1)m} \end{bmatrix}$$

Consequently for the general case where D may not be square, its inverse (assuming the pseudo-inverse for D to exist) is given by

$$\tilde{W}^{-T} = \begin{bmatrix} D(D^TD)^{-1} & 0 \\ -C^TD(D^TD)^{-1} & I_{(L-1)m} \end{bmatrix}$$

Factoring out the augmented weight term $\tilde{W}^{-T}$, (33) can be written as

$$\Delta\tilde{W} = \eta\left[I - \varphi(\tilde{y})\tilde{x}^T\tilde{W}^T\right]\tilde{W}^{-T} \tag{34}$$

Post-multiplying by the matrix $\tilde{W}^T\tilde{W}$, the update law becomes

$$\Delta\tilde{W} = \eta\left[I - \varphi(\tilde{y})\tilde{x}^T\tilde{W}^T\right]\tilde{W}$$

Using (32), we can rewrite the above update as

$$\Delta\tilde{W} = \eta\left[I - \varphi(\tilde{y})\tilde{y}^T\right]\tilde{W} \tag{35}$$

Writing in terms of the original state space variables the update law (35) is given by

$$\Delta\begin{bmatrix} D & C \\ 0 & I_{(L-1)m} \end{bmatrix} = \eta\left[I - \varphi\begin{bmatrix} y \\ X \end{bmatrix}\begin{bmatrix} y^T & X^T \end{bmatrix}\right]\begin{bmatrix} D & C \\ 0 & I_{(L-1)m} \end{bmatrix} \tag{36}$$

Considering the update laws for matrices C and D only, we have

$$\begin{bmatrix} \Delta D & \Delta C \end{bmatrix} = \eta[D \quad C] - \eta\begin{bmatrix} \varphi(y)\,y^T & \varphi(y)X^T \end{bmatrix}\begin{bmatrix} D & C \\ 0 & I_{(L-1)m} \end{bmatrix}$$

Therefore the final instantaneous update laws for the matrices C and D are

$$\Delta C(k) = \eta\left((I_N - \varphi(y(k))y^T(k))C(k) - \varphi(y(k))X^T(k)\right) \tag{37}$$

$$\Delta D(k) = \eta(I_N - \varphi(y(k))y^T(k))D(k) \tag{38}$$

The update law in (37) and (38) is related to the earlier derived update law (27) and (28) by the relation.

$$\tilde{\nabla} l = \nabla l \begin{bmatrix} D^T D & D^T C \\ C^T D & I_{(L-1)m} + C^T C \end{bmatrix} = \nabla l \begin{bmatrix} D & C \\ 0 & I_{(L-1)m} \end{bmatrix}^T \begin{bmatrix} D & C \\ 0 & I_{(L-1)m} \end{bmatrix} \tag{39}$$

where

$$\nabla l = -\eta \begin{bmatrix} \frac{\partial L^k}{\partial D} & \frac{\partial L^k}{\partial C} \end{bmatrix} \tag{40}$$

denotes the update according to the ordinary stochastic gradient, the conditioning matrix in (39) is symmetric and (since D is nonsingular) is positive definite.

Auxiliary Conditions for Convergence

Examining the update of the remaining terms in equation (36), we have (in a stochastic sense)

$$0_{(L-1)m \times N} = \varphi(X) y^T D \;\Rightarrow\; \varphi(X) y^T = 0_{(L-1)m \times N} \tag{41}$$

Also

$$\Delta I_{(L-1)m} = 0 = I_{(L-1)m} - \varphi(X) y^T C - \varphi(X) X^T$$

Rearranging terms, we have

$$\varphi(X) y^T C + \varphi(X) X^T = I_{(L-1)m}$$

Using the relation (41), the condition reduces to

$$\varphi(X) X^T = I_{(L-1)m} \tag{42}$$

Remark 3.1: The resulting update laws for the natural gradient update derived in (Zhang & Cichocki, 2000b) are in exact agreement with the update laws (37)-(38), (see Salam et al., 1997 and Salam, et al., 2001). However, the derivation approaches and methodologies are different.

Remark 3.2: The new auxiliary relations (41) and (42) define supplementary stochastic conditions on the nonlinear "correlation" between the outputs and the states of the demixing network. In this feedforward minimum phase structure case, these relations do not explicitly appear in the update laws and thus can not be used directly to simplify the update laws (37)-(38). (In contrast, similar auxiliary conditions derived for the feedback structure *do* simply the resulting update laws, see Section 3.2.) However, for effective convergence of the algorithm, these conditions need to be satisfied (in a stochastic sense), i.e.,

$\mathrm{E}\left[\varphi(X) y^T\right] = 0_{(L-1)m \times N}$, and

$\mathrm{E}\left[\varphi(X) X^T\right] = I_{(L-1)m}$

The first equality signifies that the output and the states of the network are "uncorrelated" as defined. This condition supplements the condition $\mathrm{E}\left[\varphi(y) X^T\right] \to 0_{(L-1)m \times N}$, which must be satisfied in (37) at convergence. Viewing the states as prior signal information, the causal algorithm necessarily includes only causal correlation while the auxiliary condition (41) signifies the noncausal correlation. The second condition signifies that the states of the demixing network

need to be a white process. It is noted that, in the special case of FIR filtering where the states are basically delayed versions of previous measurements, i.e., $X_i(k) = m_{k-i}$, the second condition (42) implies that whitening the measurement signals prior to applying the algorithm would enhance the convergence of the update law. These conditions are in addition to the stability conditions for the algorithm (see, Amari & Cardoso, 1997; Cardoso, 1998; Salam & Erten 1999a).

3.1.1. Special Cases: Filtering Structures

One of the advantages of using the state space model is that most signal processing filter structures form special cases where the constituent matrices of the dynamic state equation take specific forms. As an example, we discuss the implementation for both the Infinite Impulse Response (IIR) and the Finite Impulse Response (FIR) filters (Salam & Erten, 1999a; Salam et al., 2001).

IIR Filtering

Consider the case where the matrices A and B are set in the canonical form I (or controller form). In this case A and B are conveniently represented as

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & \cdots & A_{1(L-1)} \\ I_m & O_m & \cdots & \cdots & O_m \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ O_m & O_m & \cdots & I_m & O_m \end{bmatrix} \text{ and } B = \begin{bmatrix} I_m \\ O_m \\ \vdots \\ O_m \end{bmatrix} \tag{43}$$

where

A : matrix of dimension $(L\text{-}1)m \times (L\text{-}1)m$.

A_{1j} : Block sub-matrix of dimension $m \times m$, may be simplified to a diagonal matrix

I_m : Identity matrix of dimension $m \times m$

O_m : Zero matrix of dimension $m \times m$

B: matrix of dimension $(L\text{-}1)m \times m$

The state matrix is given by

$$X_k = X(k) = \begin{bmatrix} X_1(k) \\ X_2(k) \\ \vdots \\ X_{L-1}(k) \end{bmatrix}$$

where

$X(k)$: is a $(L\text{-}1)m \times 1$ dimensional state vector for the filter, and each

$X_j(k)$: is an $m \times 1$ dimensional component state vector

For this model the state model reduces to the following set of equations representing an IIR filtering structure.

$$
\begin{aligned}
X_1(k+1) &= \sum_{j=1}^{L-1} A_{1j} X_j(k) + m(k) \\
X_2(k+1) &= X_1(k) \\
&\vdots \\
X_{L-1}(k+1) &= X_{L-2}(k) \\
y(k) &= \sum_{j=1}^{L-1} C_j X_j(k) + Dm(k)
\end{aligned}
\tag{44}
$$

For an IIR filter represented in the state space canonical form I, there is no update law required for B, while for the matrix A we need to update only the first block row and therefore the update law (25) reduces to

$$
\Delta A_{1j} = -\eta \frac{\partial \mathrm{H}^k}{\partial A_{1j}} = -\eta (f^k_{A_{1j}})^T \lambda_1(k+1) = -\eta \lambda_1(k+1) X_j^T(k) \tag{45}
$$

The special structure in matrix A also affects the update law for the co-state equations (propagating in future time) which reduces to

$$
\begin{aligned}
\lambda_1(k) &= \lambda_2(k+1) + C_1^T \frac{\partial L^k}{\partial y_k}(k) \\
\lambda_2(k) &= \lambda_3(k+1) + C_2^T \frac{\partial L^k}{\partial y_k}(k) \\
&\vdots \\
\lambda_{L-1}(k) &= C_{L-1}^T \frac{\partial L^k}{\partial y_k}(k)
\end{aligned}
$$

Solving specifically for time k and then using time shift for time k+1, we obtain the recursive form for the co-state update as

$$
\begin{aligned}
\lambda_1(k) &= \sum_{i=1}^{L-1} C_i^T \frac{\partial L^k}{\partial y_k}(k+i-1) \\
\lambda_1(k+1) &= \sum_{i=1}^{L-1} C_i^T \frac{\partial L^k}{\partial y_k}(k+i)
\end{aligned}
\tag{46}
$$

Using (46), we can thus simplify (45) for the update of the block sub-matrices in A as

$$
\Delta A_{1j} = -\eta \sum_{i=1}^{L-1} C_i^T \frac{\partial L^k}{\partial y_k}(k+i) X_j^T \tag{47}
$$

which is non-causal but can be implemented using the usual time delayed version of the algorithm implemented using buffer storage memory (Haykin, 1996 & 2000).

FIR Filtering

In this case, the first row of block sub-matrices A_{1j} in (43) is zero, therefore the above filtering model (44) reduces to the FIR filtering structure. In this case the state space model reduces to

$$
\begin{aligned}
X_1(k+1) &= m(k) \\
X_2(k+1) &= X_1(k) \\
&\vdots \\
X_{L-1}(k+1) &= X_{L-2}(k) \\
y(k) &= \sum_{j=1}^{L-1} C_j X_j(k) + Dm(k)
\end{aligned}
\tag{48}
$$

where the state dynamic equations reduce merely to the delays of the mixture inputs i.e.

$$
\begin{aligned}
X_1(k) &= m(k-1) \\
X_2(k) &= m(k-2) \\
&\vdots \\
X_{L-1}(k) &= m(k-L+1)
\end{aligned}
\tag{49}
$$

In this case only matrices C and D need to be updated, as in the MIMO controller form both matrices A and B contain only fixed block identity and zero sub-matrices and are conveniently absorbed in (49).

The double-sided FIR filter version can be handled using a time delayed version of the above state space. This results in the corresponding increased requirement of buffered storage, but with the advantage of being able to converge to "stable" solutions even for non-minimum phase environments (Amari et al., 1997, Waheed & Salam, 2002d).

One possible method to deal with the problem of the filter approximation length and resulting delay is in the frequency domain implementations of the algorithm, where the converged solution represents a 1024 or higher tap double sided FIR filter in equivalent frequency domain. The result can be converted to its time domain equivalent by appropriately using the inverse FFT followed by chopping or windowing techniques to optimally contain maximum possible filter power spectral density (PSD) while minimizing algorithmic delay and hence the buffering memory requirement (Lambert, 1996; Lee et al., 1997 & 1997a; Lambert & Bell, 1997).

Next, we propose an alternate structure (Salam & Erten, 1997; Salam & Waheed, 2001) where we model the demixing network to be comprised of a feedback transfer function, see Figure 4. This feedback path is composed of tunable poles and zeros while the feedforward path is assumed to have fixed parameters. Using only the update laws for adaptively determining the zeros of this feedback path can potentially alter the poles and zeros of the overall demixing structure.

3.2. Theorem 3.2: Feedback Minimum Phase Demixing

Assume the MIMO (mixing) environment modeled by a minimum-phase linear dynamic system or transfer function. Then the update laws for the zeros of the feedback state space demixing network using the natural gradient are given by

$$\Delta D = \eta\left[(I_N + D)(\varphi(y)y^T - I_N)\right] \tag{50}$$

$$\Delta C = \eta[(I_N + D)\varphi(y)X^T] \tag{51}$$

where $\varphi(y)$ *is the appropriate score function given by (29)*

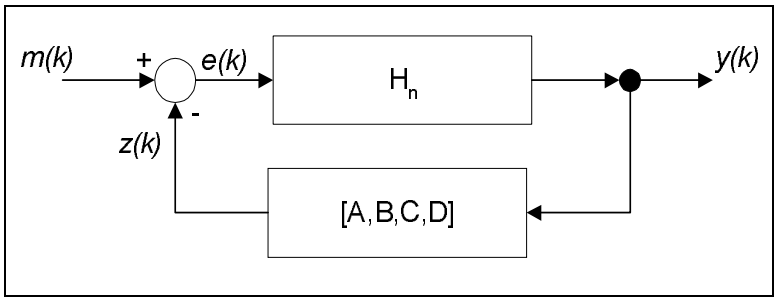


Figure 4. A State Space Feedback Demixing Structure

Proof

The network equations for the feedback path are (see Figure 4)

$$z(k) = C\,X(k) + D\,y(k) \tag{52}$$

Defining

$$e(k) = m(k) - z(k) \tag{53}$$

the output of the feedback structure is given by

$$y(k) = H_n * e(k)$$

Assuming for simplicity $H_n = I$, we have

$$y(k) = I * (m(k) - z(k)) = m(k) - z(k)$$

rearranging terms, we get

$$(I_N + D)\,y(k) \;+\; C\,X(k) = m(k) \tag{54}$$

In matrix form, we augment (54) to become

$$\begin{bmatrix} I_N + D & C \\ 0 & I_{(L-1)m} \end{bmatrix} \begin{bmatrix} y \\ X \end{bmatrix} = \begin{bmatrix} m \\ X \end{bmatrix}$$

or

$$\begin{bmatrix} y \\ X \end{bmatrix} = \begin{bmatrix} I_N + D & C \\ 0 & I_{(L-1)m} \end{bmatrix}^{-1} \begin{bmatrix} m \\ X \end{bmatrix} \tag{55}$$

Defining the augmented vectors $\tilde{y}$ and $\tilde{x}$, and the matrix $\tilde{W}$ as

$$\widetilde{X} = \begin{bmatrix} m \\ X \end{bmatrix},\ \widetilde{Y} = \begin{bmatrix} y \\ X \end{bmatrix},\ \text{and } \widetilde{W} = \begin{bmatrix} I_N + D & C \\ 0 & I_{(L-1)m} \end{bmatrix}$$

we have

$$\widetilde{Y} = \widetilde{W}^{-1}\widetilde{X} := \widetilde{\widetilde{W}}\widetilde{X} \text{ where } \widetilde{\widetilde{W}} = \widetilde{W}^{-1} \tag{56}$$

Using the natural gradient, the update law for $\tilde{\tilde{W}}$ is

$$\Delta\tilde{\tilde{W}} = \eta\left[\tilde{\tilde{W}}^{-T} - \varphi(\tilde{Y})\tilde{X}^T\right]\tilde{\tilde{W}}^T\tilde{\tilde{W}}$$

$$\Delta\tilde{\tilde{W}} = \eta\left[I - \varphi(\tilde{Y})\tilde{X}^T\tilde{\tilde{W}}^T\right]\tilde{\tilde{W}} \tag{57}$$

Differentiating $\tilde{\tilde{W}}\tilde{W} = I$ (w.r.t. time), we get

$$\tilde{\tilde{W}}^{/}\tilde{W} + \tilde{\tilde{W}}\tilde{W}^{/} = 0$$

rearranging terms, we get

$$\tilde{\tilde{W}}^{/} = -\tilde{\tilde{W}}\tilde{W}^{/}\tilde{W}^{-1} = -\tilde{W}^{-1}\tilde{W}^{/}\tilde{W}^{-1} \tag{58}$$

Using 1st order Euler approximation for the derivatives in (58), we have

$$\Delta\tilde{\tilde{W}} = -\tilde{W}^{-1}(\Delta\tilde{W})\tilde{W}^{-1} \tag{59}$$

Also note that from (56)

$$\tilde{Y}^T = \tilde{X}^T\tilde{\tilde{W}}^T \tag{60}$$

Using (59) and (60), the update law in (57) becomes

$$-\tilde{W}^{-1}(\Delta\tilde{W})\tilde{W}^{-1} = \eta\left[I - \varphi(\tilde{Y})\tilde{Y}^T\right]\tilde{\tilde{W}}$$

or

$$-\tilde{W}^{-1}(\Delta\tilde{W}) = \eta\left[I - \varphi(\tilde{Y})\tilde{Y}^T\right]$$

since $\tilde{\tilde{W}}\tilde{W} = I$.

Arranging terms

$$\Delta\tilde{W} = -\eta\,\tilde{W}\left[I - \varphi(\tilde{Y})\tilde{Y}^T\right] = \eta\,\tilde{W}\left[\varphi(\tilde{Y})\tilde{Y}^T - I\right] \tag{61}$$

Inserting the definition of $\tilde{W}$ and $\tilde{Y}$, (61) becomes

$$\Delta\begin{bmatrix} I_N + D & C \\ 0 & I_{(L-1)m} \end{bmatrix} = \eta\begin{bmatrix} I_N + D & C \\ 0 & I_{(L-1)m} \end{bmatrix} \times \begin{bmatrix} \varphi(y)y^T - I_N & \varphi(y)X^T \\ \varphi(X)y^T & \varphi(X)X^T - I_{(L-1)m} \end{bmatrix}$$

Therefore the update laws for the matrices C and D are extracted from equating the first rows block wise to get

$$\Delta(I_N + D) = \Delta D = \eta\left[(I_N + D)(\varphi(y)y^T - I_N) + C\varphi(X)y^T\right] \tag{62}$$

$$\Delta C = \eta[(I_N + D)\varphi(y)X^T + C(\varphi(X)X^T - I_{(L-1)m})] \tag{63}$$

Auxiliary Conditions for Convergence

By equating the second row block wise, one gets (in a stochastic sense)

$$0_{(L-1)m\times N} = \varphi(X)_{(L-1)m\times 1}\, y^T{}_{1\times N} \tag{64}$$

$$\Delta I_{(L-1)m} = 0 = \varphi(X)X^T - I_{(L-1)m}, \text{ or}$$

$$\varphi(X)X^T = I_{(L-1)m} \tag{65}$$

Remark 3.3: Observe that (64) and (65) are identical to (41) and (42). These relations form *supplementary conditions* for the convergence of the proposed feedback algorithm. In this case, however, the states of the feedback structure are a function of the previous network outputs, i.e., $X_i(k) = \varsigma(y_{k-i})$. Therefore, the auxiliary conditions in the feedback case define nonlinear correlation conditions on the demixing outputs that signify that the expectation of the demixing network outputs will be (nonlinearly) decorrelated leading to the components being stochastically independent.

Unlike the feedforward update laws, however, these conditions appear explicitly in the derived update laws (62) and (63). Therefore, enforcing the derived auxiliary conditions in the update laws (62) and (63) not only simplifies them computationally, but also improves the convergence dynamic properties of the proposed algorithm. The final update law for the feedback structure is given by

$$\Delta D = \eta\left[(I_N + D)(\varphi(y)y^T - I_N)\right] \tag{66}$$

$$\Delta C = \eta[(I_N + D)\varphi(y)X^T] \tag{67}$$

The above natural gradient update laws for feedback BSR are different from earlier proposed feedback (or recurrent) update laws in (Cichocki et al., 1996 and Lee et al., 1996). The algorithm (66) and (67) exhibits better stability and faster convergence. See (Waheed et al., 2003a) for a comparison.

3.3. Theorem 3.3: Feedforward Non-minimum Phase Demixing

Assume the MIMO (mixing) environment modeled by a non-minimum phase transfer function. Then, the update law for the zeros of the state space FIR demixing network using the natural gradient is given by

$$\Delta C_i = \eta(k)\left[C_i - \varphi(y(k))u(k-i)^T\right], i = 1, 2, \cdots, L-1 \tag{68}$$

$$\Delta D = \eta(k)\left[D - \varphi(y(k))u(k)^T\right] \tag{69}$$

where the state space matrix C is defined as

$$C = \begin{bmatrix} C_1 & C_2 & \cdots & C_{L-1} \end{bmatrix} \tag{70}$$

C_i - *being the MIMO FIR filter coefficients corresponding to delay* z^{-i}

$u(k)$ – *represents an information back-propagation filter, given as*

$$u(k) = \sum_{i=1}^{L-1} C_i^T y(k+i) + D^T y(k)$$

or equivalently in the adjoint state space as

$$\lambda(k-1) = A^T \lambda(k) + C^T y(k)$$

$$u(k) = B^T \lambda(k) + D^T y(k)$$

where T represents the matrix Hermitian transpose operator, $\varphi(y)$ *is the appropriate score function given by (29), and* A *and* B *are given by (43), with* $A_{1j} = 0$.

Proof

On the outset, we state that the algorithm requires a frame of 3L-2 samples of the input signal to initiate the computations of the update of the parameter matrices in (68)-(69). Note also that the time index of the parameter update (68)-(69) is not necessarily the same time index of the signals. In practice, it may be chosen so that the time index of the parameters is a delayed version of the signal time index so that no issues of non-causality would arise. Also in practice, see Simulation III in Section 4.2, the initialization of the update law (68)-(69) uses an identity (or a nonsingular matrix) for the center tap matrix (specifically, $j = \lfloor (L-1)/2 \rfloor$) and small random or zero values for the remaining tap matrices.

Let the mixing environment to assume the same state space representation as in (16)-(17), while the demixing network assumes the state space representation (18)-(19), which is repeated below for convenience

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

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

As mentioned before, for the non-minimum phase mixing case, we constrain the demixing network to be a double-sided FIR filter so as to approximate the intended unstable IIR inverse. Using the MIMO controller canonical form, the matrices A and B of dimensions $(L\text{-}1)m \times (L\text{-}1)m$ and $(L\text{-}1)m \times m$, respectively take the form

$$A = \begin{bmatrix} O_m & O_m & \cdots & O_m & O_m \\ I_m & O_m & \cdots & O_m & O_m \\ O_m & I_m & \cdots & O_m & O_m \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ O_m & O_m & \cdots & I_m & O_m \end{bmatrix}, \quad B = \begin{bmatrix} I_m \\ O_m \\ O_m \\ \vdots \\ O_m \end{bmatrix}$$

where

I_m: Identity matrix of dimension $m \times m$

O_m: Zero matrix of dimension $m \times m$

For compactness, we can represent (18)-(19) using the compact (mixed time-frequency) transfer function notation of the MIMO FIR filter as

$$y_k = \left[\bar{W}(z)\right] m_k \tag{71}$$

where, $\bar{W}(z)$ is the transfer function of the demixing MIMO FIR filter of the form,

$$\bar{W}(z) = W_0 + W_1 z^{-1} + \ldots + W_{L-1} z^{-L+1} \tag{72}$$

Further we can define a matrix $\bar{W}$, comprised of the coefficients of the filter $\bar{W}(z)$. Note that each constituent sub-matrix W_i contains co-efficients pertaining to the i^{th} filter lag acting on the time instant k-i. This matrix $\bar{W}$ is related to the state space matrices C and D as given below.

$$\bar{W} \triangleq \begin{bmatrix} W_0 & W_1 & \cdots & W_{L-1} \end{bmatrix} = \begin{bmatrix} D & C \end{bmatrix}$$

where the state space matrices C and D in terms of the sub-matrices W_i are represented as

$$\begin{aligned} D &= W_0 \\ C &= [W_1 \quad W_2 \quad \cdots \quad W_{L-1}] \end{aligned} \tag{73}$$

For the MIMO transfer function $\bar{W}(z)$, the conjugate (or Hermitian) transpose of the transfer function can be defined in a fashion similar to the one proposed by (Lambert, 1996), i.e.,

$$\begin{aligned} \left[\bar{W}(z)\right]^T &= \left[W_0 + W_1 z^{-1} + \ldots + W_{L-1} z^{-L+1}\right]^T \\ &= W_0^T + W_1^T z^1 + \ldots + W_{L-1}^T z^{L-1} \\ &= \bar{W}^T(z^{-1}) \end{aligned} \tag{74}$$

where, T is the Hermitian transpose operator. Notice that again we can define a matrix of coefficients $\bar{W}^T$ comprised of the transpose of the constituent sub-matrices acting in a time reversed fashion, i.e., the i^{th} lag of this time reversed filter acts on the future (non-causal) time instant k+i, i.e.,

$$\bar{W}^T \triangleq \begin{bmatrix} W_0^T & W_1^T & \cdots & W_{L-1}^T \end{bmatrix} \tag{75}$$

Using the definition of $\bar{W}$, we can express the network output equation (71) in the time domain as

$$y_k = \bar{W}\underline{m}_k \equiv D m_k + \sum_{i=1}^{L-1} C_i m_{k-i} \tag{76}$$

where $\underline{m}_k$ represents an $Lm \times 1$ matrix of observations at the k^{th} iteration, which includes the k^{th} and previous L-1 m-d measurement (or observation) vectors, i.e.,

$$\underline{m}_k \triangleq \begin{bmatrix} m_k^T & m_{k-1}^T & \cdots & m_{k-L+1}^T \end{bmatrix}^T$$

y_k: represents the output vector at the k^{th} iteration

For the derivation of the update laws, we minimize the Kullback-Liebler divergence with the constraints of an FIR representation as described above. Using the maximization of entropy approach, the Hamiltonian can be defined (Gharbi & Salam, 1997; Sabala et al., 1998; Salam & Erten, 1999a; Amari, 1998) as

$$H^k = -\log \det\left|W_j\right| - \sum_i \log\left(q_i\left(y_i, W(z)\right)\right) \tag{77}$$

where $q_i\left(y_i, W(z)\right)$ is the online estimated marginal distribution of $y_i(k)$ and $j \in [0, L-1]$.

Using the ordinary stochastic gradient update laws (27) and (28), also see (Gharbi & Salam, 1997; Salam & Erten 1999a; Amari 1998), the update law for the sub-matrices W_i is given by

$$\Delta W_i = -\eta_k \frac{\partial H^k}{\partial W_i} = \eta_k \left[\left(W_j^{-1}\right)^T \delta_{ji} - \varphi(y_k) m_{k-i}^T \right] \tag{78}$$

where δ_{ji} equals 1 only when $i = j$, and $j \in [0, L-1]$. It should be noted that W_j is required here to be nonsingular. For the non-minimum phase case, j is appropriately chosen to be an intermediate tap, e.g. $j = \lfloor (L-1)/2 \rfloor$, in order to render an appropriate double-sided FIR filter. In any event, it will be seen below that using the natural gradient, the resulting parameter update laws (87) or (88) do not depend on the explicit non-singularity of the j-th tap matrix and will include all tap matrices. The Stochastic Gradient (SG) law for the complete MIMO filter matrix is given by

$$\left[\Delta \bar{W}(z)\right]_{SG} = \sum_{i=0}^{L-1} \Delta W_i z^{-i} = \eta_k \left[\left(W_j^{-1}\right)^T \delta_{ji} - \sum_{i=0}^{L-1} \varphi(y_k) m_{k-i}^T z^{-i} \right] \tag{79}$$

Using the definition of the natural gradient in the systems space (Amari 1998), we can modify (79) to have the update of the demixing filter parameters according to the Riemannian contra-variant gradient as

$$\begin{aligned} \Delta \bar{W}(z) &= \left[\Delta \bar{W}(z)\right]_{SG} \bar{W}^T(z^{-1}) \bar{W}(z) \\ &= \eta_k \left[\left(W_j^{-1}\right)^T \delta_{ji} - \sum_{i=0}^{L-1} \varphi(y_k) m_{k-i}^T z^{-i} \right] \bar{W}^T(z^{-1}) \bar{W}(z) \\ &= \eta_k \left[\left(W_j^{-1}\right)^T \delta_{ji} \bar{W}^T(z^{-1}) \bar{W}(z) - \sum_{i=0}^{L-1} \varphi(y_k) m_{k-i}^T z^{-i} \bar{W}^T(z^{-1}) \bar{W}(z) \right] \end{aligned} \tag{80}$$

where, the first term on the right hand side can be simplified as

$$\begin{aligned} \left(W_j^{-1}\right)^T \delta_{ji} \bar{W}^T(z^{-1}) \bar{W}(z) &= \left(W_j^{-1}\right)^T \left(\delta_{ji} \bar{W}^T(z)\right) \bar{W}(z) \\ &= \left(W_j^{-1}\right)^T W_j^T \bar{W}(z) = I^T \bar{W}(z) = \bar{W}(z) \end{aligned} \tag{81}$$

while the second term can be simplified as

$$\begin{aligned}\sum_{i=0}^{L-1}\varphi(y_k)m_{k-i}^T z^{-i}\bar{W}^T(z^{-1})\bar{W}(z) &= \varphi(y_k)\sum_{i=0}^{L-1}\left(\left[\bar{W}(z)\right]m_{k-i}\right)^T\bar{W}(z)z^{-i}\\ &= \varphi(y_k)\sum_{i=0}^{L-1}y_{k-i}^T\bar{W}(z)z^{-i}\\ &= \varphi(y_k)\sum_{i=0}^{L-1}\left(\left[\bar{W}^T(z^{-1})\right]y_{k-i}\right)^T z^{-i}\\ &= \varphi(y_k)\sum_{i=0}^{L-1}u_{k-i}^T\, z^{-i}\end{aligned} \tag{82}$$

where, we used equality (71) and u_k^T is the transpose of the non-causal back-propagation filter output, defined as

$$u_k \triangleq \left[\bar{W}^T(z^{-1})\right]y_k \equiv D^T y_k + \sum_{i=1}^{L-1}C_i^T y_{k+i} \tag{83}$$

Let $\underline{y}_k$ represent an $NL \times 1$ matrix of outputs at the k^{th} iteration, which includes the k^{th} and the future L-1 N-d output vectors, i.e.,

$$\underline{y}_k \triangleq \begin{bmatrix} y_k^T & y_{k+1}^T & \cdots & y_{k+L-1}^T \end{bmatrix}^T$$

then the back-propagation filter can also be expressed in the compact form

$$u(k) = \bar{W}^T\underline{y}_k \tag{84}$$

Using (81) and (82) in (80), we get the following update laws for the MIMO FIR transfer Function

$$\Delta\bar{W}(z) = \eta_k\left[\bar{W}(z) - \sum_{i=0}^{L-1}\varphi(y_k)u_{k-i}^T z^{-i}\right] \tag{85}$$

The update law (85) is non-causal as computation of $u_{k-i}^T; i = 0,1,2,\ldots,L-1$ requires up to $L-1$ future outputs to be available. Practically the update law can be implemented (or made causal) by introducing a delay of $L-1$ iterations between the update of the parameters and the computation of the output at the k^{th} iteration (this in fact amount of using a frame-buffering in order to initialize the update law as is used in practice). Thus, a causal version of (85) becomes

$$\left[\Delta\bar{W}(z)\right]_C = \eta_k\left[\bar{W}(z) - \sum_{i=0}^{L-1}\varphi(y_{k-L+1})u_{k-L-i+1}^T z^{-i}\right] \tag{86}$$

Writing down the update laws in the time domain for the i^{th} component coefficient matrix W_i, we have

$$\left[\Delta W_i\right]_C = \eta_k\left[W_i - \varphi(y_{k-L+1})u_{k-L-i+1}^T\right] \tag{87}$$

Defining the new time index $\bar{k} = k - L + 1$, we have

$$[\Delta W_i]_C = \eta_k \left[W_i - \varphi\left(y\left(\bar{k}\right)\right) u^T\left(\bar{k} - i\right) \right] \tag{88}$$

where the computation of the network output $y\left(\bar{k}\right)$ and the back-propogation filter output $u\left(\bar{k}\right)$ is computed as in the following expressions

$$y\left(\bar{k}\right) \triangleq y\left(k - L + 1\right) = \bar{W}\underline{m}_{k-L+1} = \sum_{i=1}^{L-1} C_i m(k - L + 1 - i) + Dm(k - L + 1) \tag{89}$$

$$u\left(\bar{k}\right) \triangleq u\left(k - L + 1\right) = \bar{W}^T \underline{y}_{k-L+1} = \sum_{i=1}^{L-1} C_i^T y(k - L + 1 + i) + D^T y(k - L + 1) \tag{90}$$

Therefore the explicit causal update laws for the state space matrices C and D, using the time index $\bar{k}$ are

$$\Delta C_i = \eta_k \left[C_i - \varphi(y(\bar{k}))u(\bar{k} - i)^T \right], i = 1, 2, \cdots, L - 1 \tag{91}$$

$$\Delta D = \eta_k \left[D - \varphi(y(\bar{k}))u(\bar{k})^T \right] \tag{92}$$

In the state space regime, the information back-propagation filter assumes the form

$$\lambda(\bar{k} - 1) = A^T \lambda(\bar{k}) + C^T y(\bar{k}) \tag{93}$$

$$u(\bar{k}) = B^T \lambda(\bar{k}) + D^T y(\bar{k}) \tag{94}$$

The non-minimum natural gradient algorithms (Amari et al., 1997a; Zhang et al., 1999 & 2000a; Waheed & Salam 2001, 2002, 2002d) require both forward and backward in time propagation by the adaptive MIMO FIR filter. One would require a buffer of $3L$-2 input samples to initiate the adaptation according to the update laws.

The non-minimum phase state space BSR update laws, being more general, also encompass the domain of the minimum phase mixing network case. This can be verified conveniently by using definition (83) in the update law for the matrix D. Note, without any loss of generality, the delay introduced in the update law (92) to make it causal has been ignored for clarity.

$$\begin{aligned}
\Delta D &= \eta_k \left[D - \varphi(y(k))u(k)^T \right] \\
&= \eta_k \left[D - \varphi(y(k)) \left[D^T y(k) + \sum_{i=1}^{L-1} C_i^T y(k + i) \right]^T \right] \\
&= \eta_k \left[D - \varphi(y(k)) \left(y(k)^T D + \sum_{i=1}^{L-1} y(k + i)^T C_i \right) \right] \\
&= \eta_k \left[\left(I - \varphi(y(k))y(k)^T \right) D - \varphi(y(k)) \sum_{i=1}^{L-1} y(k + i)^T C_i \right]
\end{aligned} \tag{95}$$

where the first term on the right hand side of equation (95) comprises the instantaneous minimum phase update of the matrix D (see equation (38)), while the second term constitutes the "non-

causal" update term corresponding to the back-propagating processing required for the double sided FIR demixing filter in the case of non-minimum phase environments. Similarly, the update law (91) can be expanded using (76)-(83) to show the explicit relationship to the minimum phase feedforward update law (37).

Remark 3.4: It should be noted that Equation (80) holds for the doubly infinite filter space. For an FIR Space domain, it does not satisfy a Lie group in the sense of the formulation in (Zhang et al., 2000a). In that event, the matrix product in (80) is defined differently and consequently a similar, but different, update law results.

4. Simulation Results

MATLAB simulation results for the three proposed algorithms are now presented. In all three cases, the environment comprises of a 3 × 3 IIR mixing convolving filter with the following model (see, Zhang & Cichocki, 2000b; Waheed & Salam, 2001).

$$\sum_{j=0}^{m-1} \mathcal{A}_i m(k-i) = \sum_{i=0}^{n-1} \mathcal{B}_i s(k-i) + v(k) \tag{96}$$

where

$\mathcal{A}_i, \mathcal{B}_i$: are the co-efficient matrices for the i^{th} tap of the autoregressive and moving average sections of the filter respectively, and

$v(k)$: is the additive gaussian measurement noise.

m, n : are the lengths of the auto-regressive and moving average lags in the filter.

The demixing network for each simulation can be initialized in a number of ways, depending on the available environment information, which may be collected via auxiliary means (Zhang et al., 2000, 2000b). This may include complete, partial, or no knowledge of the poles of the mixing environment (Waheed & Salam 2001, 2002). The primary advantage of incorporating knowledge of poles in a BSR algorithm is to obtain a more compact demixing network representation. All the three cases have been verified via simulation and the algorithms were able to converge satisfactorily using *online* update algorithms. However, due to limited space only typical results are being presented in this paper (see Salam et al., 2001; Waheed & Salam 2001, 2002, 2002a, 2002d for more details). Unless otherwise specified, the three original sources have sub-gaussian, gaussian and super-gaussian densities respectively.

The convergence performance of the BSR algorithms, for the linear and convolutive mixing, is measured using the multi-channel inter-symbol interference (ISI) benchmark. ISI is a measure of the global transfer function diagonalization and permutation as achieved by the demixing network and is defined as

$$ISI_k = \sum_{i=1}^{N} \frac{\left\| \sum_j \sum_p |G_{pij}| - \max_{p,j} |G_{pij}| \right\|}{\max_{p,j} |G_{pij}|} + \sum_{j=1}^{N} \frac{\left\| \sum_i \sum_p |G_{pij}| - \max_{p,i} |G_{pij}| \right\|}{\max_{p,i} |G_{pij}|} \tag{97}$$

where

$G(z) = H(z) * \bar{H}(z)$ - 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.

As discussed earlier, the optimal score function for the derived update laws depends 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 score function (Waheed & Salam 2002, 2002a, 2003), which relies on the batch kurtosis of the output of the demixing system. This score function given by (98) converges to the optimal non-linearity for the demixing system as the network's outputs approach stochastic independence.

$$\varphi(y) = \begin{cases} y - \alpha \tanh(\beta y) & \kappa_4(y) \le -\gamma \\ y & \text{for } |\kappa_4(y)| < \gamma \\ \alpha \tanh(\beta y) & \kappa_4(y) \ge \gamma \end{cases} \tag{98}$$

where

$\kappa_4(y)$: represents batch kurtosis of the output signals.

γ : represents a positive (density classification) threshold.

4.1. BSR from Minimum Phase Mixing Environments.

In this section, we present the BSR results for a minimum phase IIR mixing model of the form (96), where $m = n = 3$ and

$$\mathcal{A}_0 = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \end{bmatrix}, \mathcal{A}_1 = \begin{bmatrix} 0.5 & 0.8 & -0.7 \\ 0.8 & 0.3 & -0.2 \\ -0.1 & -0.5 & 0.4 \end{bmatrix}, \mathcal{A}_2 = \begin{bmatrix} 0.06 & 0.4 & -0.5 \\ 0.16 & -0.1 & -0.4 \\ -0.3 & -0.06 & 0.3 \end{bmatrix}$$

$$\mathcal{B}_0 = \begin{bmatrix} 1 & 0.6 & 0.8 \\ 0.3 & 1 & 0.1 \\ 0.6 & -0.8 & 1 \end{bmatrix}, \mathcal{B}_1 = \begin{bmatrix} 0.5 & 0.5 & 0.6 \\ -0.3 & 0.2 & -0.3 \\ -0.2 & -0.43 & 0.6 \end{bmatrix}, \mathcal{B}_2 = \begin{bmatrix} .125 & 0.06 & 0.2 \\ -0.1 & 0 & 0.4 \\ 0.08 & -0.13 & 0.3 \end{bmatrix}$$

In the transfer function domain, this MIMO transfer function will constitute a 3 × 3 matrix with each element being an order 18 IIR filter (provided each element has been scaled to account for all the environment transmission poles). The transmission pole-zero map, the impulse response of the environment transfer function, the theoretical environment inverse, the corresponding typical demixing network solution estimate (after convergence), and the overall global transfer function using the minimum phase BSR algorithm (see Section 3.1) are respectively shown in Figure 5.

4.1.1. Simulation I: Feedforward Recovery of Minimum Phase Mixing

Using the proposed feedforward structure (Section 3.1), three different simulation scenarios are presented, where the poles of the demixing network have been fixed (a) at the solution, (b) all set to zero, or (c) set randomly but stable. In all cases, the matrix C is initialized to have small random elements, the matrix D is initialized to be the identity matrix, while matrices A and B are correspondingly set to be in the Canonical (controllable) Form I. The number of states in the

demixing state space network was correspondingly chosen to be 18, 60, and 27. For all three cases, the pole-zero map of the final demixing network and the online MISI convergence characteristics are presented in Figure 6 after 30,000 online iterations. It can be observed that in all the three cases the global transfer function is diagonalized, although the demixing networks have quite different pole-zero maps, indicating multiple possible recovery solutions to the same problem. A quantitative comparison of recovered results has been summarized in Table 1.

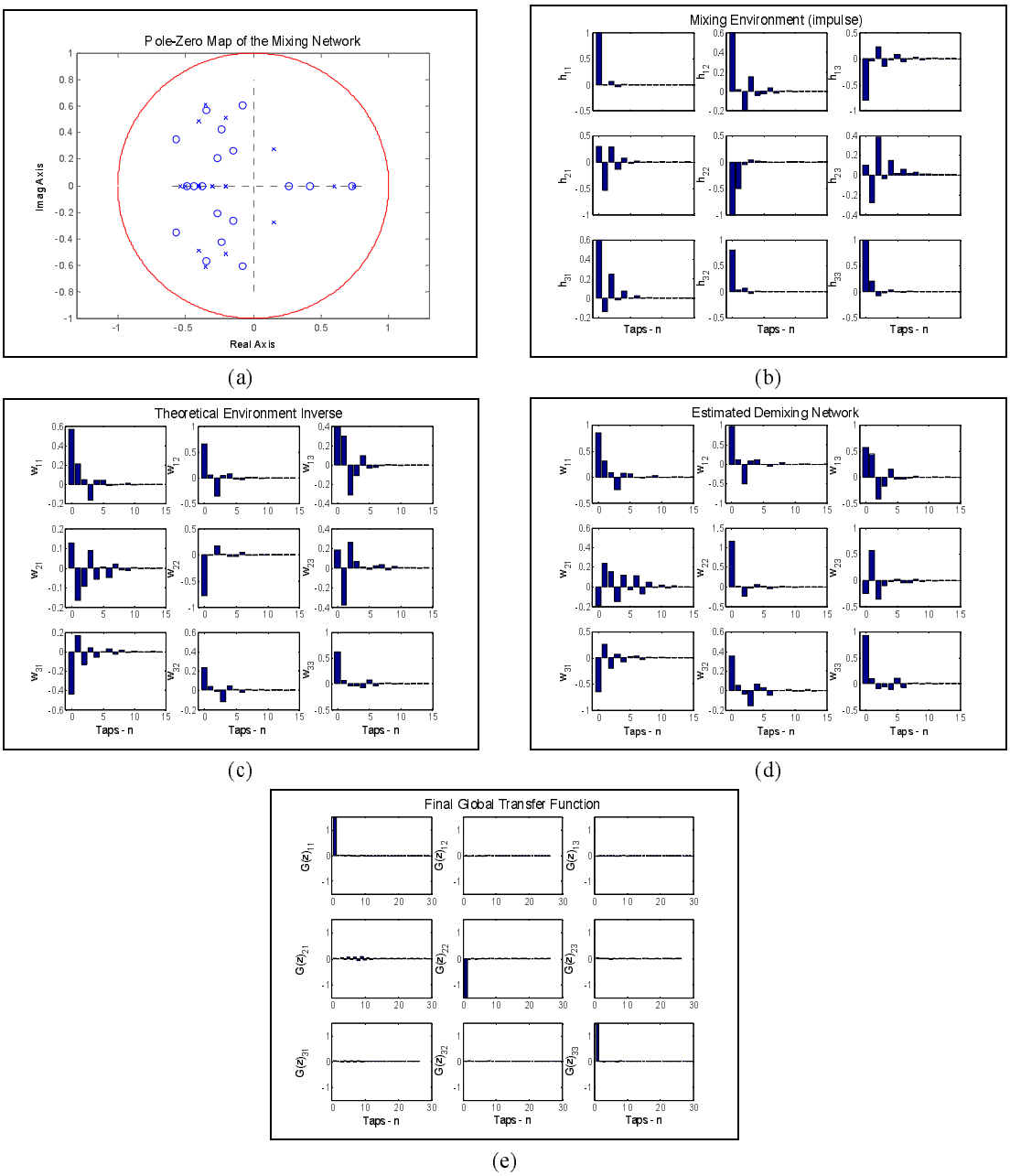


Figure 5. (a) Transmission Zeros and Poles of the Minimum Phase Mixing Environment, (b) Impuse response of the Environment Model, (c) Theoretical Environment Inverse, (d) Typical Estimated Demixing Network using Minimum Phase BSR algorithms, (e) Final Global Transfer Function

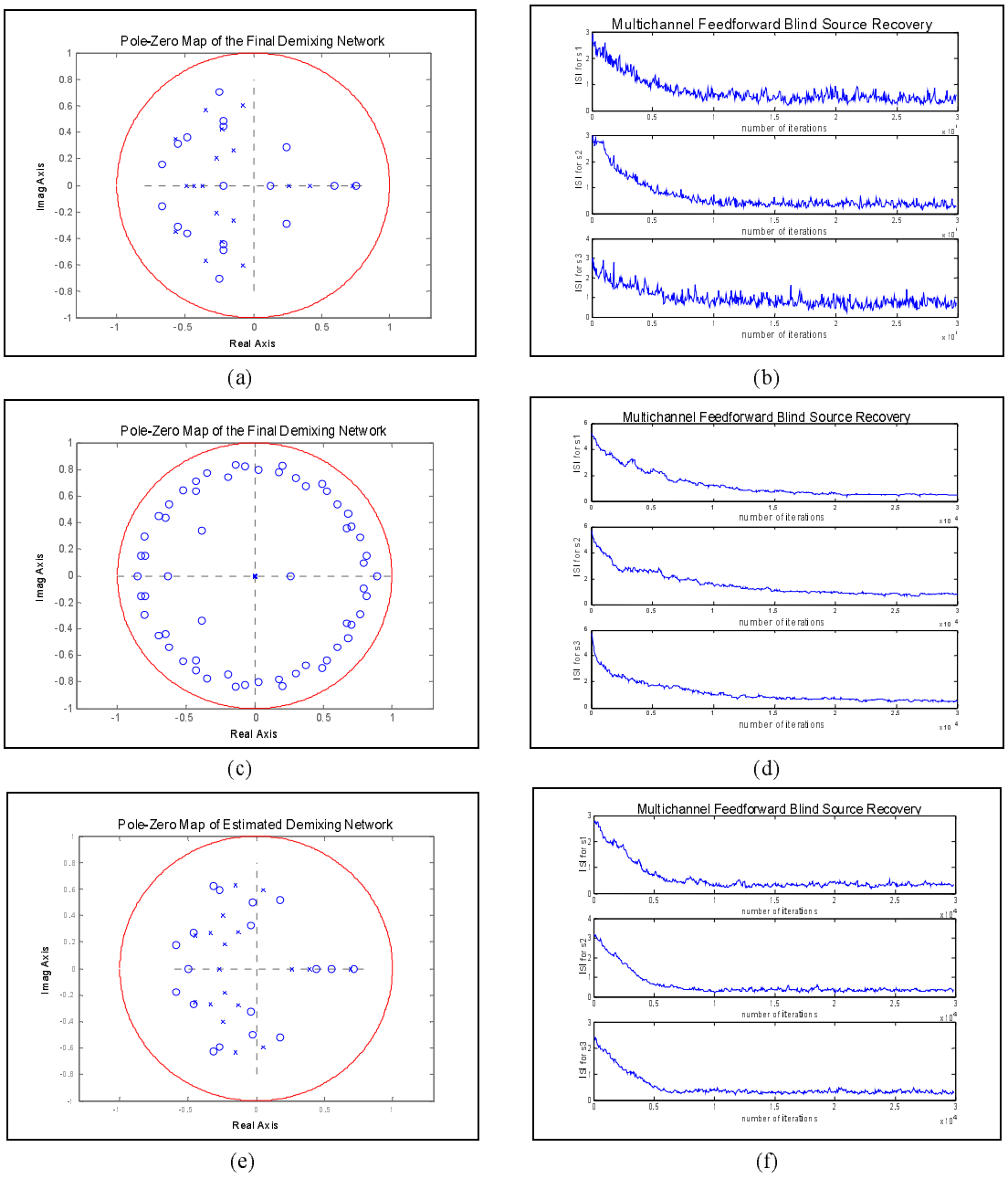


Figure 6. Transmission pole-zero map and Convergence of MISI performance index for the minimum phase feedforward network, when (a) all poles are set to the theoretical solution, (b) all poles are fixed at zero, (c) all poles are fixed but randomly set to be stable.

4.1.2. Simulation II: Feedback Recovery of Minimum Phase Mixing

For the proposed feedback structure (Section 3.2), similar three different simulation scenarios are presented, where the demixing network poles have been (a) fixed at the solution, (b) all set to

zero, or (c) set to be random but stable. The state space matrices and the number of states for the feedback network were chosen similar to the feedforward case. Note that again three different global diagonalizing demixing network pole-zero maps are obtained. The qualitative simulation result summary is presented in Figure 7. See Table 1 for a quantitative comparison.

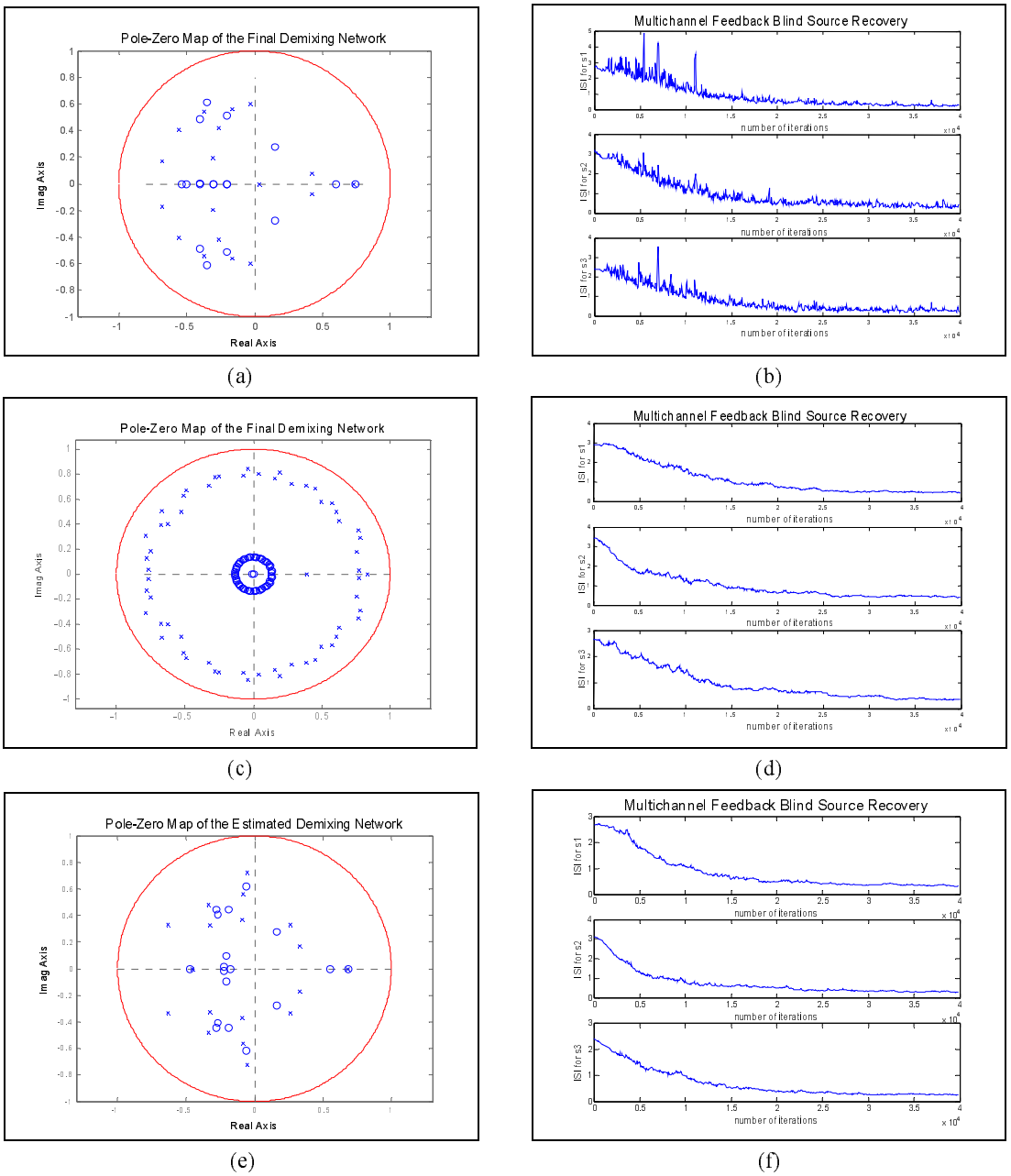


Figure 7. Feedback transmission pole-zero map and Convergence of MISI performance index for the minimum phase feedback network, when (a) all poles are set to the theoretical solution, (b) all poles are fixed at zero, (c) all poles are fixed but randomly set to be stable.

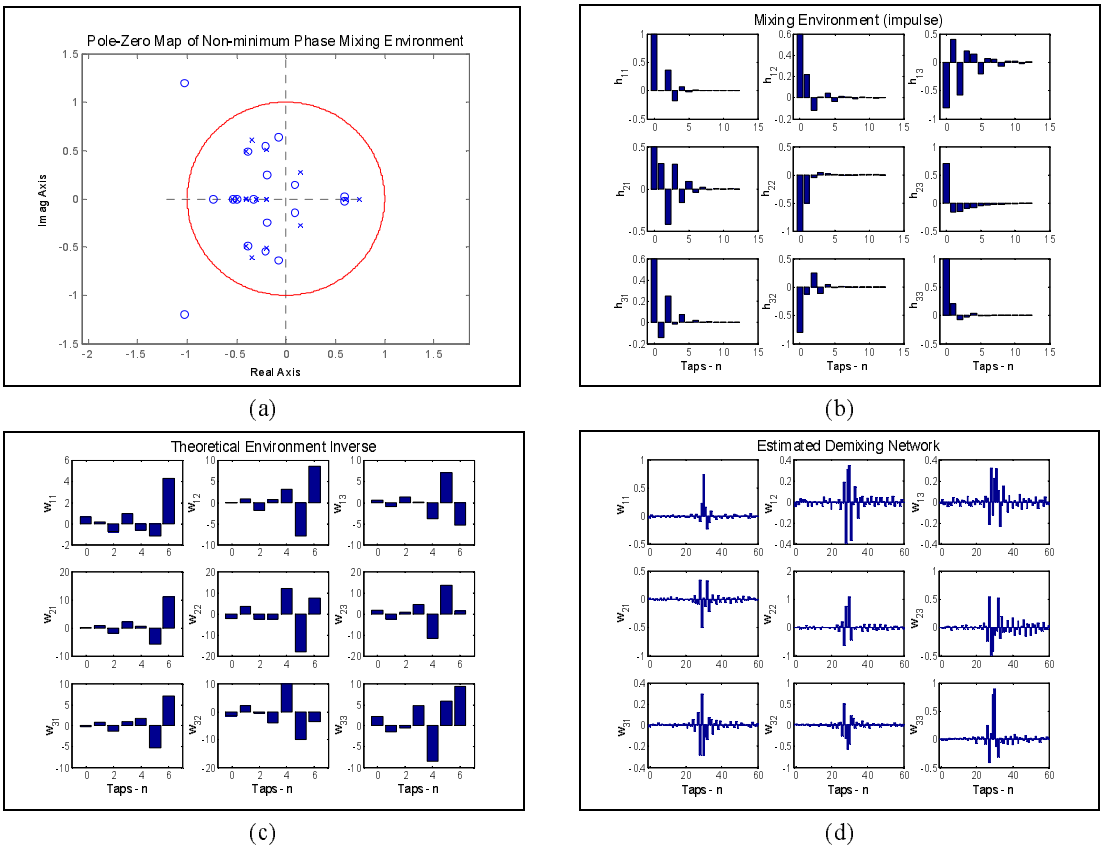
4.2. Simulation III: Feedforward Recovery of Non-minimum Phase Mixing

For the case of non-minimum phase IIR mixing environment model (96), we again choose $m = n = 3$ and the coefficient matrices are given by

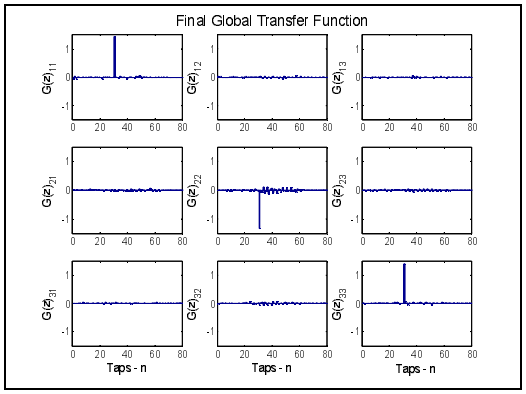
$$\mathcal{A}_0 = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \end{bmatrix}, \mathcal{A}_1 = \begin{bmatrix} 0.5 & 0.8 & -0.7 \\ 0.8 & 0.3 & -0.2 \\ -0.1 & -0.5 & 0.4 \end{bmatrix}, \mathcal{A}_2 = \begin{bmatrix} 0.06 & 0.4 & -0.5 \\ 0.16 & -0.1 & -0.4 \\ -0.3 & -0.06 & 0.3 \end{bmatrix}$$

$$\mathcal{B}_0 = \begin{bmatrix} 1 & 0.6 & 0.8 \\ 0.5 & 1 & 0.7 \\ 0.6 & 0.8 & 1 \end{bmatrix}, \mathcal{B}_1 = \begin{bmatrix} 0.5 & 0.7 & 0.16 \\ 0.7 & 0.2 & -0.3 \\ -0.2 & 0.53 & 0.6 \end{bmatrix}, \mathcal{B}_2 = \begin{bmatrix} .425 & 0.3 & 0.7 \\ -0.1 & 0 & -0.4 \\ 0.08 & -0.13 & 0.3 \end{bmatrix}$$

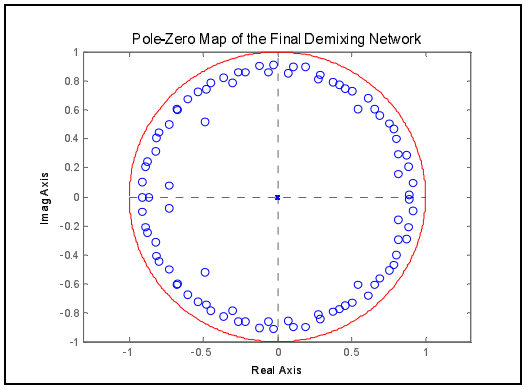
The theoretical inverse of this environment model will be an unstable IIR filter with a minimum of 20 states. This instability stems from having two poles of the intended demixing network to be outside the unit circle. However, the demixing network is setup to be comprised of a doubly-finite 3×3 FIR demixing filter with 31 taps per filter (i.e., a total of 90 states) and supplied with mixtures of multiple source distributions. The augmented matrix $\bar{W} = [D\ C]$ is initialized to be full rank, while A and B are in Canonical form I. The update results after 40,000 iterations are depicted in Figure 8. A close comparison of the theoretical environment inverse and the adaptively estimated inverse reveals that the theoretical result forms a part of the much larger (in number of taps), estimated inverse.



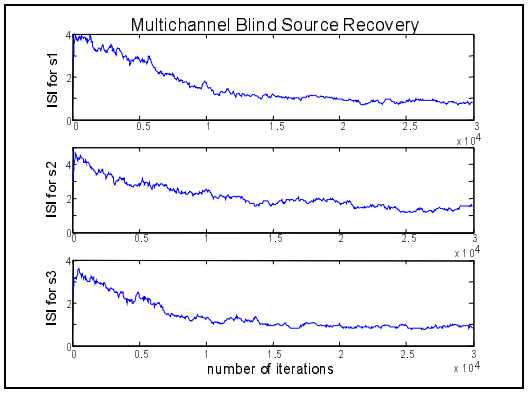
(a) (b) (c) (d)



(e)



(f)



(g)

Figure 8. State Space BSR for a non-minimum phase system: (a) Transmission poles-zero map for the mixing environment, (b) Environment Impulse Response, (c) Theoretical Environment Inverse, (d) Estimated Demixing Network, (e) Global Transfer Function Achieved, (f) Pole-zero map of Estimated Demixing Network (g) Convergence of MISI Performance Index

A quantitative performance comparison of the (blindly) recovered communication signals using the presented BSR algorithms is provided in Table 1. For the table below, we have defined

$$ISI_i(\text{in dB}) = 20\log 10(ISI_i) \tag{99}$$

where ISI_i is the intersymbol-interference in the i^{th} signal computed using the relation (97).

		Minimum Phase Feedforward Demixing			Minimum Phase Feedback Demixing			Non-minimum Phase Demixing
		Poles known	*Poles fixed at zero*	*Stable Poles fixed randomly*	*Poles known*	*Poles fixed at zero*	*Stable Poles fixed randomly*	
ISI	AM Signal	-22.55 dB	-21.77 dB	-26.02 dB	-27.67 dB	-16.47 dB	-23.42 dB	-17.67 dB
	PSK Signal	-26.66 dB	-18.48 dB	-23.52 dB	-23.52 dB	-18.85 dB	-24.85 dB	-11.43 dB
	White Noise	-21.09 dB	-21.85 dB	-21.57 dB	-26.03 dB	-18.36 dB	-23.96 dB	-11.39 dB

Table 1. Quantitative Signal Recovery Comparison of the proposed algorithms

5. Conclusion

This paper presents a generalized Blind Source Recovery (BSR) framework based on the theory of optimization under constraints of a state space representation with the Kullback-Liebler divergence as the performance functional. The framework is then specialized to the state space blind source recovery for both static and dynamic environments, which may possess either a minimum phase or a non-minimum phase structure. For the case of minimum phase environments, two possible state-space demixing networks in feedforward and feedback configuration are proposed. For non-minimum phase (convolutive) mixing, to avoid any instability, the demixing network is set as a non-causal all-zero feedforward state space network. The update laws for all the cases have been derived using the natural (or the Riemannian contra-variant) gradient.

The algorithms have been extensively simulated for both synthetic and physical data. The simulation results for all the cases have been provided using mixtures of multiple source densities, including one gaussian source. The algorithms exhibit success in recovery of sources in all cases. Furthermore, for the case of minimum phase environments, various possible scenarios of mixing network zeros have been explored. A comparative table for a set of communication signals has also been provided as a quantitative guideline for the achievable separation/recovery results using the proposed online adaptive algorithms.

Acknowledgements

This work was supported in part by the National Science Foundation under NSF grant EEC-9700741. The authors would also like to thank the anonymous reviewers for their insightful comments.

References

S. Amari. Differential-Geometrical Methods in Statistics. Lecture Notes in Statistics 28. New York: Springer-Verlag, 1985.

S. Amari. Neural Learning in Structured Parameter Spaces—Natural Riemannian gradient. Advances in Neural Processing Systems, 9. Cambridge, MA: MIT Press, 1996.

S. Amari, & J. F. Cardoso. Blind Source Separation—Semi-Parametric Statistical Approach; IEEE Trans. on Signal Processing. Vol. 45, No. 11, pp. 2692-2700, 1997

S. Amari, S. C. Douglas, A. Cichocki, & H. H. Yang. Multichannel Blind Deconvolution and Equalization using the Natural Gradient; Proceeding IEEE Workshop on Signal Processing, Adv. In Wireless Communications, pp. 101-104, Paris, France, April, 1997

S. Amari, T.-P. Chen, and A. Cichocki. Stability Analysis of Adaptive Blind Source Separation. Neural Networks, 10 (8), 1345-1352, 1997a.

S. Amari. Natural Gradient Works Efficiently in Learning. Neural Computation 10, pp. 251-276, MIT Press, 1998.

S. Amari and H. Nagaoka. Methods of Information Geometry. AMS and Oxford University Press, 1999.

A. Cichocki, S. I. Amari and J. Cao. Blind Separation of Delayed and Convolved Sources with Self Adaptive Learning Rate. Int'l Symposium on Nonlinear Theory and Applications 1996, NOLTA-96. Kochi, Japan, pp. 229-232, 1996.

A. J. Bell and T. J. Sejnowski. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. Neural Computation, Vol. 7, 1129–1159, 1995.

A. Benveniste, M. Goursat & G. Ruget. Robust Identification of a Non-Minimum Phase System: Blind Adjustment of a Linear Equalizer in Data Communications. IEEE Transactions on Automatic Control, vol. 25, pp. 385-399, 1980.

J. F. Cardoso. Blind Signal Processing: Statistical Principles. Proceedings of the IEEE, Vol. 90, no. 8, pp. 2009-20026, Oct. 98, Special issue on Blind Identification and Estimation, R.-W. Liu and L. Tong editors, 1998.

J. F. Cardoso, & B. Laheld. Equivariant Adaptive Source Separation. IEEE Trans. on Signal Processing, 44, 3017–3030, 1996.

A. Cichocki, S. Amari, J. Cao. Neural Network Models for Blind Separation of Time Delayed and Convolved Signals. IECE Trans. Fundamentals, E82-A, 1997.

A. Cichocki and S. Amari. Adaptive Blind Signal and Image Processing. John Wiley & Sons, ISBN: 0-471-60791-6, 2002.

S. C. Douglas and S. Haykin. On the Relationship between Blind Deconvolution and Blind Source Separation. Proc. 31st Asilomar Conf. on Signals, Systems, and Computers, 1998. Pacific Grove, CA, vol. 2, pp. 1591-1595, 1997.

G. Erten and F. M. Salam. Voice Output Extraction by Signal Separation, Proceedings of the 1998 IEEE International Symposium on Circuits and Systems, 1998, ISCAS 1998. Volume: 3, 1998, Vol.3, pp. 5 – 8, 1998.

G. Erten and F. M. Salam. Real Time Separation of Audio Signals using Digital Signal Processors; Proceedings of the 40th Midwest Symposium on Circuits and Systems, 1997, Volume: 2, pp. 1237 - 1240, 1998a.

G. Erten and F. M. Salam. Voice Extraction by on-line Signal Separation and Recovery. IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, Volume: 46, Issue: 7, July 1999 Page(s): 915 -922, 1999

A. B. A Gharbi and F. M. Salam. Separation of Mixed Signals in Dynamic Environments: Formulation and some Implementation. Proc. of 37th Midwest Symposium on Circuits and Systems. 1994: Lafayette, LA. p. 587-590, 1994.

A. B. A. Gharbi and F. M. Salam. Implementation and Test Results of a Chip for the Separation of Mixed Signals. IEEE Transactions on Circuits and Systems, vol. 42, no.11, pp.748-751, 1995.

A. B. A. Gharbi and F. M. Salam. Algorithms for Blind Signal Separation and Recovery in Static and Dynamic Environments. Proceedings of 1997 International Symposium on Circuits and Systems (ISCAS-1997), June 9-12, Hong Kong, pp. 713-716, 1997.

M. Girolami. Advances in Independent Component Analysis (Perspectives in Neural Computing). Springer Verlag; ISBN: 1852332638, 2000.

S. Haykin. Adaptive Filter theory. 3rd Edition, Prentice-Hall, 1996.

S. Haykin, Editor. Unsupervised Adaptive Filtering. Vol. I & II, John Wiley & Sons, ISBN: 0-471-29412-8, 0-471-37941-7, 2000.

J. Herault and C. Jutten. Space or Time Adaptive Signal Processing by Neural Network Models. Neural Networks for Computing, AIP Conference Proceedings, 151, New York, 1986. American Institute for physics, 1986.

A. Hyvarinen and E. Oja. A Fast Fixed-point Algorithm for Independent Component Analysis. Neural Computation, 9:1483-1492, 1997.

A. Hyvarinen, J. Karhunen, E. Oja. Independent Component Analysis. Wiley Inter-science Series, ISBN: 0-471-40540-X, 2001.

J. Karhunen. Neural Approaches to Independent Component Analysis and Source Separation. Proc. of 4th European Symposium on Artificial Neural Networks (ESANN-96), Bruges, Belgium, pp. 249-266, 1996.

M. Kendall and A. Stuart. The Advanced Theory of Statistics; Charles Griffin & Co. Ltd., Vol. I, 1977.

H. K. Khalil. Nonlinear Systems. 3rd Edition, Prentice Hall, ISBN 0-13-067389-7, 2002.

R. Lambert. Multichannel Blind Deconvolution: FIR Matrix Algebra and Separation of Multipath Mixtures. Ph.D. Thesis, University of Southern California, Department of Electrical Engineering, 1996.

R. Lambert and A. Bell. Blind Separation of Multiple Speakers in a Multipath Environment. ICASSP, pp. 423-426, Munich, Germany, April 21-26, 1997.

T. W. Lee and T. Sejnowski. Independent Component Analysis for Sub-gaussian and Super-gaussian Mixtures. 4th Joint Symposium on Neural Computation, 7, pp. 132-140, Institute for Neural Computation, 1997.

T. W. Lee, A. J. Bell, and R. Lambert. Blind Separation of Delayed and Convolved Sources. Advances in Neural Information processing Systems. Vol 9, pp. 758-764, 1997.

T. W. Lee, A. J. Bell, and R. Orglmeister. Blind Source Separation of Real-World Signals. IEEE Proc. ICNN, Houston, 2129-2135, 1997.

T. W. Lee, M. Girolami, A. Bell, & T. J. Sejnowski. A Unifying Information-theoretic Framework for Independent Component Analysis. Int'l journal on Mathematical and Computer Modeling, vol. 39, pp.1-21, 2000.

F. L. Lewis and V. L. Syrmos. Optimal Control. 2nd Edition, Wiley Interscience Series, ISBN: 0-471-03378-2, 1995.

J. Nicholls. From Neuron to Brain; 4th Ed, Published 2001; Sinauer Associates Incorporated; ISBN: 0878934391, 2001

I. Sabala, A. Cichocki and S. Amari. Relationships between Instantaneous Blind Source Separation and Multichannel Blind Deconvolution. Proc. Int'l Joint Conference on Neural Networks, Alaska USA, pp. 148-152, 1998.

F. M. Salam. An Adaptive Network for Blind Separation of Independent Signals. Proc. of International Symposium on Circuits and Systems, ISCAS-93, Chicago, IL, USA pp. 431–434, 1993.

F. M. Salam and G. Erten. Blind Signal Separation and Recovery in Dynamic Environments. Proceedings of 3rd IEEE Workshop on Nonlinear Signal and Image Processing, Mackinac Island, MI (Online proceedings available at http://www.ecn.purdue.edu/NSIP/), 1997

F. M. Salam, A. B. Gharbi, and G. Erten. Formulation and Algorithms for Blind Signal Recovery. Proceedings of the 40th Midwest Symposium on Circuits and Systems, MWSCAS-97, Volume: 2, Page(s): 1233 -1236, 1997

F. M. Salam and G. Erten. Exact Entropy Series representation for Blind Source Separation. Systems, Man, and Cybernetics, Proceedings of 1999. IEEE SMC-99 Conference 1999, Volume: 1 pp. 553 - 558, 1999.

F. M. Salam and G. Erten. The State Space Framework for Blind Dynamic Signal Extraction and Recovery. Proceedings of the 1999 IEEE International Symposium on Circuits and Systems, ISCAS-99, Volume: 5, 1999, pp. 66 -69 vol. 5, 1999a.

F. M. Salam, A. B. A. Gharbi and G. Erten. Formulation and Algorithms for Blind Signal Recovery. Proceedings of the 40th Midwest Symposium on Circuits and Systems, 1997, Volume: 2, 1998 pp. 1233 -1236, 1998

F. M. Salam; G. Erten and K. Waheed. Blind Source Recovery: Algorithms for Static and Dynamic Environments. Proc. of INNS-IEEE International Joint Conference on Neural Networks, IJCNN, 2001, Vol. 2, Page(s) 902-907, 2001.

F. M. Salam and K. Waheed. State-space Feedforward and Feedback Structures for Blind Source Recovery. Proc. 3rd International Conference on Independent Component Analysis and Blind Signal Separation, San Diego, California, Page(s) 248-253, 2001.

K. Torkkola. Blind Separation of Convolved Sources based on Information Maximization. IEEE Workshop on Neural Networks for Signal Processing, Kyoto, Japan, pp. 423-432, 1996.

K. Waheed and F. M. Salam. Blind Source Recovery: Some Implementation and Performance Issues, Proceedings of the 44th IEEE Midwest Symposium on Circuits and Systems, Dayton, Ohio, vol. 2, pp. 694-697, 2001.

K. Waheed and F. M. Salam. State-Space Blind Source Recovery for Mixtures of Multiple Source Distributions. Proc. of IEEE International Symposium on Circuits and Systems, ISCAS-2002 Scottsdale, Arizona, pp. 197-200, 2002.

K. Waheed and F. M. Salam. Blind Source Recovery using an Adaptive Generalized Gaussian Score Function. 45th IEEE International Midwest Symposium on Circuits and Systems, Tulsa, Oklahoma, vol. 2, pp. 418 – 422, 2002a.

K. Waheed and F. M. Salam. A Data-Derived Quadratic Independence Measure for Adaptive Blind Source Recovery in Practical Applications. 45th IEEE International Midwest Symposium on Circuits and Systems, Tulsa, Oklahoma, Vol. 3, pp. 473-476, 2002b.

K. Waheed and F. M. Salam. Cascaded Structures for Blind Source Recovery; 45th IEEE Int'l Midwest Symposium on Circuits and Systems, Tulsa, Oklahoma, vol. 3, pp. 656-659, 2002c.

K. Waheed and F. M. Salam. State Space Blind Source Recovery of Non-Minimum Phase Environments. 45th IEEE International Midwest Symposium on Circuits and Systems, Tulsa, Oklahoma, vol. 2, pp. 422-425, 2002d.

K. Waheed and F. M. Salam. New Hyperbolic Source Density Models for Blind Source Recovery Score Functions. IEEE International Symposium on Circuits and Systems, Bangkok, Thailand, vol. 3, pp. 32-35, 2003.

K. Waheed, K. Desai and F. M. Salem. Blind Multi User Detection in DS-CDMA Systems using the Natural Gradient based Symbol Recovery Structures. 4th Int'l Symp. on Independent Component Analysis and Blind Source Separation 2003, ICA-2003. Nara, Japan, pp. 727-732, 2003a.

H. H. Yang, S. Amari, A. Cichocki. Information Back-propagation for Blind Separation of Sources from Non-linear Mixtures. Proc. of ICNN-97, Houston, pp. 2141-2146, 1997.

L. Q. Zhang, A. Cichocki and S. Amari. Multichannel Blind Deconvolution of Nonminimum Phase systems using Information Backpropagation. Proc. of 5th Int'l conf. on Neural Information Processing (ICONIP-99), November 16-20, 1999 - Perth, Australia, pp. 210-216, 1999

L. Q. Zhang, A. Cichocki & S. Amari. Kalman Filter and State-Space Approach to Multichannel Blind Deconvolution. IEEE Workshop on Neural Networks for Signal Processing, NNSP '2000, Sydney, Australia, December 11-13, 2000, pp. 425-434, 2000.

L. Q. Zhang, S. Amari and A. Cichocki. Semiparametric Approach to Multichannel Blind Deconvolution of Non-minimum Phase Systems. Advances in Neural Information Processing Systems 12. MIT Press, pp. 363-369, 2000a.

L. Q. Zhang. and A. Cichocki. Blind Deconvolution of Dynamical Systems: A State Space Approach. Journal of Signal Processing, March 2000, Vol.4, No.2 pp.111-130, 2000b.