Optimal Multichannel Blind Deconvolution for Parameterized Channels and Known Source Densities (EDICS:SSP-SSEP)

Robert M. Taylor Jr.*, *Student Member, IEEE,*, L. Mili *Senior Member, IEEE* and Amir I. Zaghloul *Fellow, IEEE*

***Abstract*— In this study we consider the problem of how to design an optimal multichannel blind deconvolution (MBD) algorithm in the case where the probability density functions of the source signals are known. We assume existence of a parametric channel model that accurately characterizes the propagation environment. Through three major steps we derive a blind channel parameter estimator that is used to jointly compute the separation system and recover all the source signals. First, we replace the normally assumed nonparametric channel model with a physical model. Next, we introduce a symbolic pseudoinverse for our separation model to replace the ubiquitous inverse filter separation model. Thirdly, we introduce a minimum divergence estimator formulation to replace the commonly used minimum entropy formulation. We prove that the new estimator formed in this way is asymptotically consistent and Fisher-efficient. Through simulation we show the superior performance of our algorithm compared with existing techniques based on entropy minimization and inverse filter separation.**

***Index Terms*— optimal multichannel blind deconvolution, blind source separation, parametric channel model, known source densities**

I. Introduction

THE problem of blindly separating multiple source signals impinging on an array of receivers spans numerous fields including multi-antenna wireless communications, sensor networks, sonar, radar, speech, and biomedical sensing. The goal of multichannel blind deconvolution (MBD) is to process an array of observations consisting of mixed and potentially delayed and convolved source signals in such a way as to extract every source signal without knowing the channel or the transmitted waveforms. The impact of such a capability is profound. Yet, to date, there is no known algorithm that provides a stable *consistent* and *Fisher-efficient* estimate for general convolutive mixture models. For these reasons and others MBD fails for many practical scenarios. This is likely the reason why we have not seen much of an insertion of this kind of technology in products today despite the abundance of scholarly attention in the recent past (see e.g. [1]–[3])

Mr. Robert M. Taylor Jr. is with The MITRE Corporation, 7515 Colshire Drive, McLean, VA 22102. (Tel: 703-983-5211, Fax: 703-983-6708 email: rtaylor@mitre.org)

Dr. Lamine Mili is with the Electrical and Computer Engineering Department, Alexandria Research Institute of Virginia Tech, 206 North Washington Street, Suite 400 Alexandria, VA 22314, (Tel: 703-535-3453, Fax: 703-518-8085, Email: lmili@vt.edu)

Dr. Amir I. Zaghloul is with the Electrical and Computer Engineering Department, Alexandria Research Institute of Virginia Tech, 206 North Washington Street, Suite 400 Alexandria, VA 22314, (Tel: 703-535-3465, Fax: 703-518-8085, Email: amirz@vt.edu)

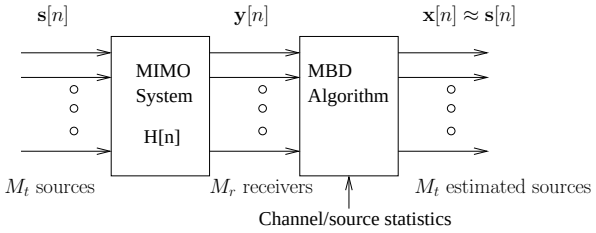

Fig. 1. High level block diagram of information flow from source signals, to observations, to estimated source signals. The multichannel blind deconvolution algorithm attempts to "invert" the multi-input multi-output channel given only partial knowledge of channel or sources.

In MBD problems the channel is a multi-input multi-output (MIMO) system that is generally linear time-varying, but it can be modeled as a linear time-invariant (LTI) sytem over a small enough time window. The MIMO LTI system $H[n]$ is a polynomial matrix which acts as a multichannel convolution mapping from the source signal space with M_t-vector process $\mathbf{s}[n]$ to the observation space with M_r-vector process $\mathbf{y}[n]$ as

$$\mathbf{y}[n] = \sum_{l=-\infty}^{\infty} H[l]\mathbf{s}[n-l] \tag{1}$$

where M_t and M_r are the number of source signals and received signals respectively (see Fig. 1). The job of the MBD algorithm is to recover an approximation to the source vector process $\mathbf{s}[n]$ given only the observation and known statistics of the channel and/or source. Every MBD algorithm consists of four essential components:

- **[C1]** Signal and System Model
- **[C2]** Separation Model
- **[C3]** Separation Criterion
- **[C4]** Optimization and Initialization Method

All separation criteria use some form of prior information such as non-gaussianity of the source signals (e.g. [6]–[8]), non-stationarity of second order statistics (e.g. [9], [10]), or time dependence (e.g. [11], [12]). In this study we will restrict our attention to separation criteria based on exploiting the non-gaussianity of the sources.

There are varying levels of information that may be available for the design of an MBD algorithm (see Fig. 2). To our knowledge, nearly all MBD techniques based on maximizing nongaussianity assume the least amount of knowledge and operate at Level 3 (which is the most challenging). With this grand pursuit comes many potential pitfalls. Most MBD

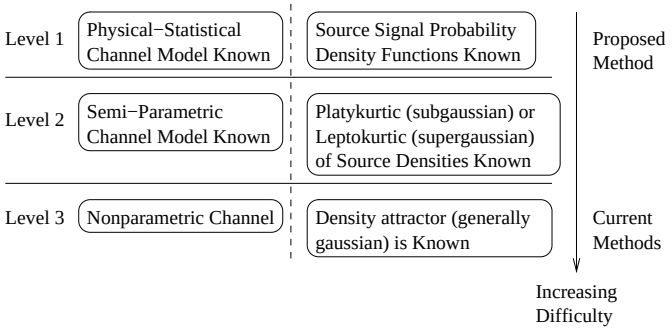

Fig. 2. Levels of information that may be used in design of a multichannel blind deconvolution algorithm.

methods utilize an inverse filter separation model (see e.g. [1]–[8] and the reference therein) which has the form:

$$\mathbf{x}[n] = \sum_{l} W[l]\mathbf{y}[n-l] \tag{2}$$

for all $n = n_0, n_0+1, ..., n_0+N-1$. The variable $W[l]$ is the *multichannel equalization filter* that must be estimated and $\mathbf{x}[n]$ is the estimate for source signal vector $\mathbf{s}[n]$. The problem here is that every element of the polynomial matrix $W[l]$ must be estimated since there is no knowledge of physics or phenomenology to create a *reduced* parameterized system. This can lead to prohibitively slow convergence for many applications (eg. packet communications, fast moving emitters, resource limited sensor networks). To complicate issues even more, most nonparametric channel models are constructed to be minimum phase even though most channels observed in nature are not. Nonminimum phase channels lead to an unstable inverse filter. Additionally, when the source densities are unknown the only admittable cost functions are those that maximize nongaussianity (assuming the gaussian density is the stable attractor density). This leads to minimum entropy formulations which generally yield statistically inconsistent and inefficient (in Fisher sense) estimators. The cost function used by Amari [7], Douglas [5], and many others has the form

$$\mathcal{J}(W(z,n)) = -\sum_{i=1}^{M_t} \log f_{X_i}(x_i[n]) \tag{3}$$
$$-\frac{1}{2\pi\jmath}\oint \log|\det W(z,n)|z^{-1}dz$$

where $f_{X_j}(x)$ is the pdf of the j^{th} estimated source signal and $W(z,n)$ is the z-transform of the multichannel equalization filter $W[l]$ at time instance n. Amari [13] showed that stochastic approximation based on (3) can only be made Fisher efficient in certain cases.

The focus of this paper is to address MBD design at Level 1. The requirement that the source signal probability densities be known is not unreasonable since for many applications of interest the mechanisms for their generation are known. For example, in communication systems, knowledge of the modulation format and pulse-shaping filter is sufficient to generate the pdf of the source (see eg. Proakis [16]). As another example, speech signals are frequently modeled with a Laplace distribution (see eg. Rabiner et al. [17]). The other Level 1 assumption that the channel be parametric is also quite acceptable since most sensor and communication systems possess a well understood statistical-physical model of the environment derived from the underlying physics. To restrict the scope of this paper, we assume that we know how many source signals are present. We assume that the source signals are nongaussian strict sense stationary ergodic processes within a given observation window. (They can be nonstationary from window to window). We introduce a parametric symbolic pseudoinverse for our separation system to replace the typical inverse filter system. We also will replace the minimum entropy formulation which assumes source densities are unknown with a minimum divergence formulation which assumes the source densites are known. Table I shows the fundamental differences between current MBD methods and the proposed method.

We derive an optimal MBD algorithm to simultaneously estimate the multi-channel impulse response and recover the sources up to an unknown amplitude and source index permutation. Most MBD algorithms also contain an ambiguity in the delay and sign of the estimated sources. Our proposed batch estimator finds the channel parameter vector that minimizes the Kullback-Leibler divergence between the probability density of the estimated sources and the probability density of the true sources. Borrowing from concepts in information geometry, our update equation moves the estimate along the minimum path of the induced Riemannian manifold. See [14] for a thorough introduction to information geometry. Specifically, the list of contributions are as follows:

1) We develop a closed form parameterized separation model that asymptotically approaches the true inverse MIMO LTI system.
2) We derive a minimum divergence estimator that is asymptotically Fisher efficient and asymptotically locally consistent.
3) We show that the estimator satisfies these optimality properties by proving uniqueness of the global minimum, convexity of the objective function about the global minimum, and equivalence of the objective function to the maximum likelihood estimator.
4) We derive the Riemannian metric for the sum of Kullback-Liebler divergence functions. (This is used to compute the natural gradient direction for faster convergence than conventional gradient.)
5) We show through simulation that our estimator vastly outperforms minimum entropy inverse filter MBD algorithms.

The advantages of proposed method are that it: 1) provides aymptotically optimal separation of sources when source densities and channel structure are known, 2) works with minimum or nonminimum phase systems. 3) converges rapidly for relatively small parameter spaces, and 4) encompasses any kind of channel model. The disdvantages of proposed method are that it is computationally intensive and does not scale well to large number of source signals.

TABLE I
FUNDAMENTAL DIFFERENCES BETWEEN CURRENT MBD METHODS AND THE PROPOSED METHOD.

Components of MBD Algorithms	Current Approaches	Proposed Approach
Signal and System Model	ad hoc nonparametric model	Parameterized statistical-physical model
Separation Model	Inverse filter	Parameterized symbolic pseudoinverse
Separation Criteria	Minimize entropy (maximize non-gaussianity)	Minimize divergence to known source density
Optimization and Initialization	Cannot incorporate any channel knowledge	Incorporates all channel knowledge and priors

The remainder of the paper is organized as follows. In Section II we derive parametric models for the received signal, channel, separation system, estimated source signal, and estimated source probability density function. In Section III we describe our new minimum divergence estimator. We analyze the asymptotic optimality of the new estimator in Section IV. We show simulation results in Section V and conclude in Section VI.

II. PARAMETRIC SIGNAL AND SYSTEM MODELS

A. Received Signal and Channel Model

Here, we briefly derive the received signal model for the output of the MIMO LTI system. For M_t source signals $\{s_j(t), \forall j = 1, ..., M_t\}$ with pdf $f_{S_j}(x)$ passing through a linear time-invariant channel $h_{ij}(\tau; \Theta)$ with parameterization Θ and impinging on a M_r-element receiver array, the i^{th} received signal $\{y_i(t), \forall i = 1, ..., M_r\}$ is

$$y_i(t) = \sum_{j=1}^{M_t} \int_{\tau=-\infty}^{\infty} h_{ij}(\tau; \Theta) s_j(t-\tau) d\tau \quad . \tag{4}$$

In pursuit of a discrete-time equivalent representation we use the sampling theorem to write the source signals as $s_j(t) = \sum_{m=-\infty}^{\infty} s_j[m] \text{sinc}(\frac{1}{T}(t - mT))$, plug this into (4), and sample at $t = nT$ (T=sampling period) to get:

$$y_i(nT) = \sum_{j=1}^{M_t} \sum_{m} s_j[m] \underbrace{\int_{\tau} h_{ij}(\tau; \Theta) \text{sinc}(n - m - \frac{\tau}{T}) d\tau}_{h_{ij}[n, n-m]} \tag{5}$$

Letting $l = n - m$ in (5) allows us to write the discrete-time input-output convolutive mixture model $\forall i = 1, 2, ..., M_r$ as

$$y_i[n] = \sum_{j=1}^{M_t} \sum_{l=-\infty}^{\infty} h_{ij}[l; \Theta] s_j[n-l] \tag{6}$$

which is true $\forall n = n_0, n_0 + 1, ..., n_0 + N - 1$. The time-invariant discrete-time channel impulse response is

$$h_{ij}[l; \Theta] = \int_{\tau=-\infty}^{\infty} h_{ij}(\tau, \Theta) \text{sinc}(l - \frac{\tau}{T}) d\tau \ \ \forall l \in \mathbb{I} \tag{7}$$

The channel matrix polynomial $h_{ij}[l; \Theta]$ as written this way is a noncausal infinite impulse response (IIR) filter. For practical purposes we must truncate the IIR filter to length L to make it FIR and delay by the appropriate number of samples L_1 to enforce causality. We therefore use the approximation $\forall l = 0, ..., L-1$:

$$h_{ij}[l; \Theta] \approx \int_{\tau=-\infty}^{\infty} h_{ij}(\tau, \Theta) \text{sinc}(l - L_1 - \frac{\tau}{T}) d\tau \quad . \tag{8}$$

The channel model in (8) is very general and applies to any environment in which a receiver is collecting samples over a coherence time for which the channel is quasi-static. Later on in Section V we will examine the real-valued *line-of-sight channel* in which the term $h_{ij}(\tau, \Theta)$ inside the integral has the form:

$$h_{ij}(\tau, \Theta) = \alpha_j \delta(\tau - \tau_{ij}) \tag{9}$$

where α_j is the received signal amplitude for source j, $\delta(\tau)$ is the Dirac delta function, and τ_{ij} is the relative time delay between receiver i and the array origin for source signal j.

B. Separation Model

The received signal model (6) can be written in matrix vector form as

$$\underbrace{\begin{bmatrix} \mathbf{y}_1[n_0] \\ \vdots \\ \mathbf{y}_{M_r}[n_0] \end{bmatrix}}_{\mathbf{y}[n_0] \in \mathbb{R}^{NM_r \times 1}} = \underbrace{\begin{bmatrix} C_{11} & \cdots & C_{1M_t} \\ \vdots & & \vdots \\ C_{M_r1} & \cdots & C_{M_rM_t} \end{bmatrix}}_{\mathcal{C} \in \mathbb{R}^{NM_r \times (N+L-1)M_t}} \underbrace{\begin{bmatrix} \mathbf{s}_1[n_0] \\ \vdots \\ \mathbf{s}_{M_t}[n_0] \end{bmatrix}}_{\mathbf{s}[n_0] \in \mathbb{R}^{(N+L-1)M_t \times 1}} \tag{10}$$

$$\approx \underbrace{\begin{bmatrix} \bar{C}_{11}(\Theta^*) & \cdots & \bar{C}_{1M_t}(\Theta^*) \\ \vdots & & \vdots \\ \bar{C}_{M_r1}(\Theta^*) & \cdots & \bar{C}_{M_rM_t}(\Theta^*) \end{bmatrix}}_{\bar{\mathcal{C}}(\Theta^*) \in \mathbb{R}^{NM_r \times NM_t}} \underbrace{\begin{bmatrix} \mathbf{x}_1[n_0] \\ \vdots \\ \mathbf{x}_{M_t}[n_0] \end{bmatrix}}_{\mathbf{x}[n_0] \in \mathbb{R}^{NM_t \times 1}}$$

where $\mathbf{y}_i[n_0] = [y_i[n_0], y_i[n_0+1], \cdots, y_i[n_0+N-1]]^T$, $\mathbf{s}_j[n_0] = [s_j[n_0], s_j[n_0+1], \cdots, x_j[n_0+N-1]]^T$, $\mathbf{x}_j[n_0]$ is the approximation of $\mathbf{s}_j[n_0]$, Θ^* is the true parameter vector for the generally unknown channel parameter vector Θ, and C_{ij} is the "fat form" convolution matrix between source j and receiver i written as:

$$C_{ij} = \begin{bmatrix} h_{ij}[L-1] & \cdots & h_{ij}[0] & & \\ \ddots & & \ddots & & \\ & h_{ij}[L-1] & \cdots & h_{ij}[0] & \end{bmatrix} \quad . \tag{11}$$

When $N >> L$, $x_j[n]$ is a suitable approximation to $s_j[n]$. Note that the actual MIMO LTI system expressed in the first line of (10) is a non-invertible linear mapping unless $M_rN \geq M_t(N + L - 1)$. Furthermore, the inverse mapping is not parameterizable in closed form using the actual system mapping. However, if we replace C_{ij} by the circular (parameterizable) convolution matrix $\bar{C}_{ij}(\Theta)$, we obtain an approximation to the true mapping that admits a closed form parameterizable separation model. Furthermore, with the approximate system mapping $\bar{\mathcal{C}}(\Theta)$, invertibility only requires that $M_r \geq M_t$.

The eigendecomposition $\bar{C}_{ij}(\Theta) = V\Lambda_{ij}(\Theta)V^H$ allows us to write

$$\bar{\mathcal{C}}(\Theta) = \underbrace{\text{diag}(V, \cdots, V)}_{M_rN \times M_rN} \times \underbrace{\begin{bmatrix} \Lambda_{11}(\Theta) & \cdots & \Lambda_{1M_t}(\Theta) \\ \vdots & & \vdots \\ \Lambda_{M_r1}(\Theta) & \cdots & \Lambda_{M_rM_t}(\Theta) \end{bmatrix}}_{M_rN \times M_tN} \underbrace{\text{diag}(V^H, \cdots, V^H)}_{M_tN \times M_tN} \tag{12}$$

Note that

$$V = \left\{ v_{kl} : v_{kl} = \frac{1}{\sqrt{N}} \exp(\jmath 2\pi(k-1)(l-1)/N) \right\} \tag{13}$$

is the $N \times N$ scaled inverse discrete fourier transform (IDFT) matrix. Consequently V^H is the scaled (forward) DFT. Note also that

$$\Lambda_{ij}(\Theta) = \text{diag}\left(\tilde{h}_{ij}[k;\Theta] = \sum_{l=0}^{N-1} h_{ij}[l;\Theta] \exp(-\frac{\jmath 2\pi kl}{N}) \right) \tag{14}$$

is the diagonal matrix of the N-point discrete fourier transform (DFT) of $h_{ij}[l;\Theta]$ since $\bar{C}_{ij}(\Theta)$ is a circular convolution matrix.

We can pseudoinvert the approximate system in (10) and write our separation model as:

$$\mathbf{x}[n_0;\Theta^*] = \mathcal{C}^\dagger(\Theta^*)\mathbf{y}[n_0] \tag{15}$$

where † denotes the Moore-Penrose pseudoinverse. Since we don't generally know the true channel parameter Θ^*, we will write the estimated source signal in (15) as $\mathbf{x}[n_0;\hat{\Theta}]$ where $\hat{\Theta}$ is the parameter vector estimate we will be solving for. From (12) and using the orthogonality of V we can show that $\bar{\mathcal{C}}^\dagger(\Theta)$ can be written as (16) (bottom of page) where we have suppressed the Θ dependence on Λ_{ij} for notational compactness. In this way the separation model is parameterized the same as the channel model. So by estimating the parameter vector Θ we can simultaneously compute the inverse system mapping to recover the source signals and determine channel parameters of interest (e.g. angle of arrival of sources).

C. Estimated Source Signal and Probability Density Model

We now desire to move from the matrix-vector form of the estimated sources in (15) back into the scalar time series expressions for each of the M_t source signals. Define $\tilde{y}_i[k]$ and $\tilde{h}_{ij}[k;\Theta]$ as the N-point DFT of the observation $y_i[n]$ starting at sample n_0 and channel impulse response $h_{ij}[n;\Theta]$ respectively. Let $d[k;\Theta]$ and $A[k;\Theta]$ respectively denote the determinant and adjoint matrix of

$$\Xi[k;\Theta] = \left\{ \xi_{pj}[k;\Theta] = \sum_{q=1}^{M_r} \tilde{h}^*_{qp}[k;\Theta]\tilde{h}_{qj}[k;\Theta] \right\} \tag{17}$$

such that $\forall k = 0, 1, ..., N-1$

$$d[k;\Theta] = \det(\Xi[k;\Theta]) \tag{18}$$

and

$$A[k;\Theta] = \{a_{jp}[k;\Theta]\} = \text{adjoint}(\Xi[k;\Theta]) \quad . \tag{19}$$

From (13), (14), (16), (18) and (19), we write the j^{th} **source signal estimate** as:

$$x_j[n;\Theta] = \frac{1}{N} \sum_{k=0}^{N-1} \frac{e^{\jmath 2\pi kn/N}}{d[k;\Theta]} \sum_{p=1}^{M_t} a_{jp}[k;\Theta] \sum_{i=1}^{M_r} \tilde{h}^*_{ip}[k;\Theta]\tilde{y}_i[k] \tag{20}$$

$\forall j = 1, 2, ..., M_t$ and $\forall n = n_0, n_0+1, ..., n_0+N-1$. Since this is the scalar form of (15), this separation model equation maps us from the observation $y_i[n]$ contained in $\tilde{y}_i[n]$ to the estimated source signal $x_j[n]$ where the only unknown is the channel parameter Θ.

Proposition 1 (Convergence of Estimated Sources): For $M_r \geq M_t$ as $N \to \infty$, then $x_j[n;\hat{\Theta} = \Theta^*] \to s_j[n]$.

Proof: From (10), as $N \to \infty$ $\bar{\mathcal{C}}(\Theta^*) \to \mathcal{C}$ since L is fixed and finite. When $M_r \geq M_t$, $\bar{\mathcal{C}}(\Theta^*)$ has full rank and the Moore-Penrose pseudoinverse in (16) is unique. Since (20) is just the scalar form of (15), this unique inverse mapping causes $\lim_{N\to\infty} x_j[n;\hat{\Theta} = \Theta^*] = s_j[n]$. □

The probability density estimate for the j^{th} estimated source signal in (20) is written $\forall x \in \mathcal{X}$ as

$$f^{(N)}_{X_j}(x|\Theta, \mathbf{y}[n_0]) = \frac{1}{N} \sum_{n=n_0}^{n_0+N-1} \psi(x - x_j[n;\Theta]) \quad . \tag{21}$$

where $\psi(x)$ is an appropriately chosen basis function. We standardize the random process $x_j[n;\Theta]$ before plugging into (21) to ensure it has zero mean and unit variance. It is worth pointing out here that the known pdf of the source signals can be described using an infinite sum of the same basis functions as

$$f_{S_j}(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \psi(x - s_j[n]) \quad . \tag{22}$$

We generally choose our basis function to be $\psi[x] = \kappa \exp(-\frac{x^2}{2\sigma^2})$ where σ parameterizes the variance of the kernel and κ is a normalization constant.

$$\bar{\mathcal{C}}^\dagger(\Theta) = \underbrace{\begin{bmatrix} V & & \\ & \ddots & \\ & & V \end{bmatrix}}_{M_tN \times M_tN} \underbrace{\begin{bmatrix} \sum_{i=1}^{M_r} \Lambda^H_{i1}\Lambda_{i1} & \cdots & \sum_{i=1}^{M_r} \Lambda^H_{i1}\Lambda_{iM_t} \\ \vdots & & \vdots \\ \sum_{i=1}^{M_r} \Lambda^H_{iM_t}\Lambda_{i1} & \cdots & \sum_{i=1}^{M_r} \Lambda^H_{iM_t}\Lambda_{iM_t} \end{bmatrix}^{-1}}_{M_tN \times M_tN} \underbrace{\begin{bmatrix} \Lambda^H_{11} & \cdots & \Lambda^H_{M_r1} \\ \vdots & & \vdots \\ \Lambda^H_{1M_t} & \cdots & \Lambda^H_{M_rM_t} \end{bmatrix}}_{M_tN \times M_rN} \underbrace{\begin{bmatrix} V^H & & \\ & \ddots & \\ & & V^H \end{bmatrix}}_{M_rN \times M_rN} \tag{16}$$

III. Minimum Divergence Estimator

Instead of minimizing the entropy of the estimated sources as in (3), we desire to minimize the "distance" between our estimated source pdf and the known source pdf. For reasons we will show later, we select our cost function $\mathcal{J}(\Theta)$ to be the Kullback-Leibler divergence between the joint pdf of the source densities and the joint pdf of the estimated source densities written as

$$\mathcal{J}(\Theta, \mathbf{y}[n_0]) = D^{KL}(f_{S_1,...,S_{M_t}}(x_1, ..., x_{M_t})|| \quad f^{(N)}_{X_1,...,X_{M_t}}(x_1, ..., x_{M_t}|\Theta)) \tag{23}$$

where $D^{KL}(p(x)||q(x)) = \sum_{x\in\mathcal{X}} p(x) \log \frac{p(x)}{q(x)}$. For the case of M_t spatially independent source signals we have

$$\mathcal{J}(\Theta, \mathbf{y}[n_0]) = \sum_{j=1}^{M_t} D^{KL}(f_{S_j}(x)||f^{(N)}_{X_j}(x|\Theta, \mathbf{y}[n_0])) \tag{24}$$

which is the objective function we will use throughout the remainder of the paper. We define our minimum contrast estimator to be:

$$\hat{\Theta} = \arg_\Theta \min \mathcal{J}(\Theta, \mathbf{y}[n_0]) \ . \tag{25}$$

Note that our estimator only has to search over all possible channel parameter vectors Θ instead of over all possible polynomial matrices $W[l]$. This reduction in the dimension of the search space enables much faster convergence of the estimator.

Amari proved in [13] that the steepest descent direction of a cost function $\mathcal{J}(\Theta, \mathbf{y}[n_0])$ in a Riemannian space is given by its *natural gradient* defined as:

$$\tilde{\nabla}\mathcal{J}(\Theta, \mathbf{y}[n_0]) = G^{-1}(\Theta)\nabla\mathcal{J}(\Theta, \mathbf{y}[n_0]) \tag{26}$$

where $G(\Theta)$ is the Riemannian metric tensor at the point Θ,

$$\nabla\mathcal{J}(\Theta, \mathbf{y}[n_0]) = [\frac{\partial}{\partial\theta_1}\mathcal{J}(\Theta, \mathbf{y}[n_0]), \cdots, \frac{\partial}{\partial\theta_{M_p}}\mathcal{J}(\Theta, \mathbf{y}[n_0])] \tag{27}$$

is the conventional gradient of the cost function defined in (24), and M_p is the dimension of the parameter vector Θ. The natural gradient exploits the geometric structure of the statistical manifold and provides faster convergence than the conventional gradient.

We will use this same technique but with a slight modification. We choose to use the following descent algorithm to optimize (25):

$$\hat{\Theta}^{(r+1)} = \hat{\Theta}^{(r)} - \mu \frac{\tilde{\nabla}\mathcal{J}(\hat{\Theta}^{(r)}, \mathbf{y}[n_0])}{||\tilde{\nabla}\mathcal{J}(\hat{\Theta}^{(r)}, \mathbf{y}[n_0])||_2} \tag{28}$$

where $||\cdot||_2$ is the L_2-norm and μ is a fixed step size parameter. If any *a priori* information about the channel exists (e.g. possible cone of angles over which signals may arrive, channel estimate from training data, etc.), this information can be used to assign $\Theta^{(0)}$ for initialization. Fig. 3 illustrates the separation and optimization approach we use in our estimator. The estimator update equation in (28) adjusts the channel parameter vector in a way so as to reduce the Kullback-Leibler divergence between the pdf of the estimated sources and the pdf of the true sources at each iteration.

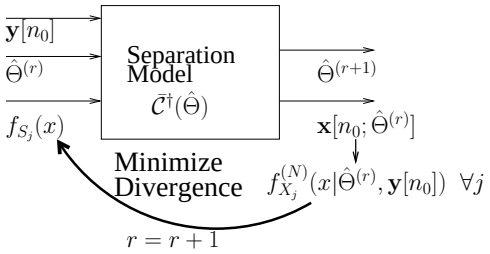

Fig. 3. Illustration of the separation and optimization components for one iteration of the batch estimator given in (28). The inputs in order are the observations starting at time sample n_0, the channel parameter vector at iteration r, and the known source density $\forall j$. The outputs in order are the channel parameter vector at iteration $r + 1$ and the estimated source signals at time sample n_0.

Eguchi [15] showed that any divergence function $D(\cdot||\cdot)$ induces a unique *Riemannian metric* $G(\Theta) = \{g_{ij}(\Theta)\}$ on the statistical manifold at the point Θ given by:

$$g_{ij}(\Theta) = -\frac{\partial}{\partial\theta'_i}\frac{\partial}{\partial\theta''_j} D(f_X(x|\Theta')||f_X(x|\Theta''))\Big|_{\Theta'=\Theta''=\Theta} \tag{29}$$

For the following theorem and proof we will shorten the notation on the marginal and joint pdfs by dropping the subscript on the pdf and replacing $f_{X_p}(x)$ and $f_{X_1X_2\cdots X_{M_t}}(x_1, x_2, ..., x_{M_t})$ with $f(x_p)$ and $f(x_1, x_2, ...x_{M_t})$ respectively for notational convenience.

Theorem 1 (Riemannian Metric for (24)): The Riemannian metric elements for the contrast function given in (24) is

$$g_{ij}(\Theta) = \sum_{p=1}^{M_t} E_{f(x_p|\Theta)}\left[\frac{\partial}{\partial\theta_i}\log f(x_p|\Theta)\frac{\partial}{\partial\theta_j}\log f(x_p|\Theta)\right] \tag{30}$$

Proof: See Appendix I. □

Following the chain rule of differentiation, we compute the terms in (27) $\forall m$ as:

$$\frac{\partial}{\partial\theta_m}\mathcal{J}(\Theta, \mathbf{y}[n_0]) = \sum_{j=1}^{M_t}\frac{\partial}{\partial\theta_m} D^{KL}(f_{S_j}(x)||f^{(N)}_{X_j}(x|\Theta, \mathbf{y}[n_0])) \tag{31}$$

which uses $\forall m, j$

$$\frac{\partial}{\partial\theta_m} D^{KL}(f_{S_j}(x)||f^{(N)}_{X_j}(x|\Theta, \mathbf{y}[n_0])) = -\sum_{x\in\mathcal{X}}\frac{\partial}{\partial\theta_m} f^{(N)}_{X_j}(x|\Theta, \mathbf{y}[n_0])\frac{f_{S_j}(x)}{f^{(N)}_{X_j}(x|\Theta, \mathbf{y}[n_0])} \tag{32}$$

which uses $\forall m, j, x$

$$\frac{\partial}{\partial\theta_m} f^{(N)}_{X_j}(x|\Theta, \mathbf{y}[n_0]) = \frac{\kappa}{N\sigma^2}\sum_{n=0}^{N-1}(x - x_j[n; \Theta])\frac{\partial}{\partial\theta_m}x_j[n; \Theta]\exp(-\frac{1}{2\sigma^2}(x - x_j[n; \Theta])^2) \tag{33}$$

which uses $\forall m, j, n$

$$\frac{\partial}{\partial\theta_m}x_j[n;\Theta] = \frac{1}{N}\sum_{k=0}^{N-1} e^{\frac{\jmath 2\pi kn}{N}} \times \quad (34)$$
$$\left[-\frac{\frac{\partial}{\partial\theta_m}d[k;\Theta]}{d^2[k;\Theta]}\sum_q a_{jq}[k;\Theta]\sum_i \tilde{h}^*_{iq}[k;\Theta]\tilde{y}_i[k] + \right.$$
$$\frac{1}{d[k;\Theta]}\sum_q \frac{\partial}{\partial\theta_m}a_{jq}[k;\Theta]\sum_i \tilde{h}^*_{iq}[k;\Theta]\tilde{y}_i[k] +$$
$$\left.\frac{1}{d[k;\Theta]}\sum_q a_{jq}[k;\Theta]\sum_i \frac{\partial}{\partial\theta_m}\tilde{h}^*_{iq}[k;\Theta]\tilde{y}_i[k]\right] .$$

Our new algorithm is based on finding the minimum divergence between the parameterized estimated source signal densities and the true source densities. We call it the minimum divergence parameterized multichannel blind deconvolution (MDPMBD) algorithm and summarize it here:

MDPMBD Algorithm

1) Select channel model $h_{ij}[n;\Theta]$ (see e.g. (54))
2) Set estimator update counter $r = 0$, $n_0 = 0$, and initialize channel parameter estimate $\hat{\Theta}^{(r)}$ with any prior knowledge.
3) Compute $\tilde{h}_{ij}[k;\hat{\Theta}^{(r)}]$ (the N-point discrete fourier transform of channnel model evaluated at $\hat{\Theta}^{(r)}$.
4) Collect N-sample observation $y_i[n]$ $\forall i = 1, ..., M_r$ and $\forall n = n_0, n_0+1, ..., n_0+N-1$ and compute its N-point discrete fourier transform to form $\tilde{y}_i[k]$.
5) Form matrix $\Xi[k;\hat{\Theta}^{(r)}]$ as in (17) $\forall k = 0, 1, ..., N-1$ and compute its adjoint matrix $A[k;\hat{\Theta}^{(r)}]$ and determinant $d[k;\hat{\Theta}^{(r)}]$ as in (19) and (18) respectively.
6) Compute source signals estimate $x_j[n;\hat{\Theta}^{(r)}]$ from (20) $\forall j = 1, ...M_t$ and $\forall n = n_0, n_0+1, ..., n_0+N-1$.
7) Compute pdf estimate $f^{(N)}_{X_j}(x|\hat{\Theta}^{(r)}, \mathbf{y}[n_0])$ from (21).
8) Compute conventional gradient vector $\nabla\mathcal{J}(\Theta, \mathbf{y}[n_0])$ in (27) from (31)-(33).
9) Compute Riemannian metric $G(\hat{\Theta}^{(r)})$ in (30) and natural gradient $\tilde{\nabla}\mathcal{J}(\Theta, \mathbf{y}[n_0])$ in (26).
10) Update channel parameter vector as in (28).
11) Set r=r+1, $n_0 = n_0 +$ (window step size) and go to step 3.

IV. Asymptotic Optimality of Algorithm

In this section we prove that the estimator formed from (24) is locally asymptotically consistent and Fisher efficient. For the proof of consistency we first prove the global minimizer is unique and then that the objective function is convex about the global minimum.

Theorem 2 (Uniqueness of Solution): As $N \to \infty$ there is one and only one global minimizer of (24).

Proof: We need to show that

$$\hat{\Theta} = \Theta^* \Leftrightarrow \mathcal{J}(\hat{\Theta} = \Theta^*) = 0 \quad (35)$$

We start by proving the forward implication. From Proposition 1, we know that as $N \to \infty$, $x_j[n;\Theta^*] \to s_j[n]$ $\forall n = n_0, ..., n_0+N-1$. This means that $\forall j = 1, ..., M_t$

$$\lim_{N\to\infty} f^{(N)}_{X_j}(x|\Theta^*) = \lim_{N\to\infty}\frac{1}{N}\sum_{n=n_0}^{n_0+N-1}\psi[x - s_j[n]] = f_{S_j}(x) \quad (36)$$

since from (22) we formed $f_{S_j}(x)$ as an infinite basis expansion using $\psi(x)$ basis functions. Therefore,

$$\mathcal{J}(\Theta^*) = \sum_{j=1}^{M_t} D^{KL}(f_{S_j}(x)||f^{(N)}_{X_j}(x|\Theta^*)) \quad (37)$$
$$\stackrel{N\to\infty}{=} \sum_{j=1}^{M_t} D^{KL}(f_{S_j}(x)||f_{S_j}(x)) = 0$$

which proves the left to right implication. Now we must prove the right to left implication which we do through the contrapositive

$$\hat{\Theta} \neq \Theta^* \Rightarrow \mathcal{J}(\hat{\Theta}) > 0 \quad (38)$$

To this we simply turn to the definition of a divergence function (which the Kullback-Leibler divergence satisfies). A divergence function is any smooth function $D(\cdot||\cdot) : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ satisfying for any points $p, q \in \mathcal{M}$

$$D(p||q) \geq 0, \text{ and } D(p||q) = 0 \text{ iff } p = q \ . \quad (39)$$

For $\hat{\Theta} \neq \Theta^*$ we have $\lim_{N\to\infty} f^{(N)}_{X_j}(x|\hat{\Theta}) \neq f_{S_j}(x)$. Therefore, under this condition the divergence function is strictly positive and the reverse implication is satisfied. □

This completes the first part of the consistency proof. Now, we must show that the objective function surface is convex about the global minimum in order for the optimization of (28) to converge for the appropriate fixed step size μ.

Theorem 3 (Convexity about Global Minimum): As $N \to \infty$ the objective function in (24) is locally convex about the global minimizer Θ^*.

Proof: Since our objective function is twice-differentiable we can use positive definiteness of the Hessian matrix to prove convexity. We start with

$$\frac{\partial^2}{\partial\theta_m\partial\theta_{m'}}\mathcal{J}(\Theta) = \sum_{j=1}^{M_t}\frac{\partial^2}{\partial\theta_m\partial\theta_{m'}}D^{KL}(f_{S_j}(x)||f^{(N)}_{X_j}(x|\Theta)) \quad (40)$$

where

$$\frac{\partial^2}{\partial\theta_m\partial\theta_{m'}}D^{KL}(f_{S_j}(x)||f^{(N)}_{X_j}(x|\Theta)) = \quad (41)$$
$$-\sum_{x\in\mathcal{X}}\frac{\partial^2}{\partial\theta_m\partial\theta_{m'}}f^{(N)}_{X_j}(x|\Theta))\frac{f_{S_j}(x)}{f^{(N)}_{X_j}(x|\Theta)} +$$
$$\sum_{x\in\mathcal{X}}\frac{\partial}{\partial\theta_m}f^{(N)}_{X_j}(x|\Theta)\frac{f_{S_j}(x)}{[f^{(N)}_{X_j}(x|\Theta)]^2}\frac{\partial}{\partial\theta_{m'}}f^{(N)}_{X_j}(x|\Theta)$$
$$\stackrel{\Theta=\Theta*}{=} E_{f^{(N)}_{X_j}(x|\Theta^*)}\left[\frac{\partial\log f^{(N)}_{X_j}(x|\Theta^*)}{\partial\theta_m}\frac{\partial\log f^{(N)}_{X_j}(x|\Theta^*)}{\partial\theta_{m'}}\right] .$$

From the second line to third line we have used the fact that $\lim_{N\to\infty} f^{(N)}_{X_j}(x|\Theta^*) = f_{S_j}(x)$. In the last line of (41) we observe that we're left with the (m, m') component of the Fisher information matrix–which is always positive definite.

The Hessian components of the objective function in (40) are simply a sum of M_t of these Fisher information matrix components. Since the sum of any number of positive definite matrices is also positive definite, the Hessian in (40) is also positive definite. Therefore, the objective function in (24) is convex about the point Θ^*. □

It is interesting to note that the Hessian of our cost function is equivalent to the Riemannian metric in (30). Now we proceed to show that the estimator is asymptotically Fisher-efficient (achieves the Cramer-Rao bound asymptotically).

Theorem 4 (Fisher Efficiency of Estimator): The estimator formed from (24) is asymptotically Fisher-efficient.

Proof: We prove the estimator is asymptotically efficient by proving that (24) is equivalent to the maximum likelihood objective function.

$$\begin{aligned}
\hat{\Theta}^{MLE} &= \arg_\Theta \max \log f_{\mathbf{Y}_1 \cdots \mathbf{Y}_{M_r}}(\mathbf{y}_1, ..., \mathbf{y}_{M_r}|\Theta) \\
&= \arg_\Theta \max \log f^{(N)}_{\mathbf{X}_1, \cdots \mathbf{X}_{M_t}}(\mathbf{x}_1, ..., \mathbf{x}_{M_t}|\Theta) \\
&\overset{\substack{spatial \\ i.i.d.}}{=} \arg_\Theta \max \sum_{j=1}^{M_t} \log f^{(N)}_{\mathbf{X}_j}(\mathbf{x}_j|\Theta) \\
&\overset{\substack{temporal \\ i.i.d.}}{=} \arg_\Theta \max \sum_{j=1}^{M_t} \log \prod_{i=1}^{N} f^{(N)}_{X_{ji}}(x_{ji}|\Theta) \\
&\overset{N\to\infty}{=} \arg_\Theta \max \sum_{j=1}^{M_t} E_{f_{S_j}(x)}[\log f^{(N)}_{X_j}(x|\Theta)] \\
&= \arg_\Theta \min \sum_{j=1}^{M_t} \left\{ E_{f_{S_j}(x)} \left[\log \frac{f_{S_j}(x)}{f^{(N)}_{X_j}(x|\Theta)} \right] \right. \\
&\quad \left. + E_{f_{S_j}(x)} \left[-\log f_{S_j}(x) \right] \right\} \\
&= \arg_\Theta \min \sum_{j=1}^{M_t} \left\{ E_{f_{S_j}(x)} \left[\log \frac{f_{S_j}(x)}{f^{(N)}_{X_j}(x|\Theta)} \right] + \mathcal{H}(f_{S_j}) \right\} \\
&= \arg_\Theta \min \sum_{j=1}^{M_t} D^{KL}(f_{S_j}(x) || f^{(N)}_{X_j}(x|\Theta))
\end{aligned} \tag{42}$$

Here, we have used the one-to-one correspondence between $\mathbf{x}[n_0]$ and $\mathbf{y}[n_0]$ from (10) in step two, spatial independence of sources in step three, temporal independence of sequences in step four, strong law of large numbers in step five, and the definition of entropy for $\mathcal{H}(f_S) = -E_{f_S}[\log f_S(x)]$ in the penultimate step. □

Note that we used temporal independence in the proof of Theorem 4 even though we previously stated that this was not a necessary condition. We get around this technicality in the proof by assuming that we can downsample the infinitely long sequence until the samples are temporally independent.

V. Simulation Results

In this section we present the results from computer simulation of the observed signals and the estimator algorithm. To motivate the significance of the proposed estimator, we compare its performance to the algorithm of Amari et. al. in [4], [7].

A. Comparison Algorithm

Amari's multichannel blind deconvolution algorithm [7] will serve as our benchmark algorithm for comparison. It is derived from a entropy minimization formulation as in (3), has very good performance under special circumstances, and is computationally efficient. For this reason and others, it has emerged as a top MBD algorithm for researchers in the field (see e.g. [5]).

The estimator update for the polynomial matrix separation system is:

$$W^{(r+1)}[l] = W^{(r)}[l] + \mu_r \{ W^{(r)}[l] - \mathbf{q}(\mathbf{x}[r-L])\mathbf{u}^H[r-l] \} \tag{43}$$

where

$$\mathbf{u}[r] = \sum_{q=0}^{L} [W^{(r)}]^H[L-q]\mathbf{x}[r-q] \tag{44}$$

and where $q_j(x)$ is a nonlinear (nonpolynomial) function to be specified by the user. The estimated source signals at update step r are formed from the inverse filter separation as in (2):

$$\mathbf{x}[r] = \sum_{l=0}^{L} W^{(r)}[l]\mathbf{y}[r-l] \tag{45}$$

Amari (and most others using a minimum entropy objective function) incorporate knowledge of the source densities through the judicious choice of the nonlinear function $q_j(x)$ in (43). The optimal choice when the source densities are known is given in [7] as:

$$q_j(x) = -\partial \log(f_{S_j}(x))/\partial x \tag{46}$$

where $f_{S_j}(x)$ is the known pdf of the j^{th} source signal. To be fair we must initialize Amari's algorithm with the same amount of prior information we use in the proposed method. To incorporate a parameterized channel with an initial parameter estimate into an inverse filter method formulation we will make use of the following relation:

$$\sum_{i=1}^{M_r} \sum_{l=0}^{L-1} W[l]H[n-l] = \begin{bmatrix} \delta[n] & & \\ & \ddots & \\ & & \delta[n] \end{bmatrix} \tag{47}$$

If we limit our attention to the $M_r = M_t$ case we can solve for the inverse filter polynomial matrix initial estimate from the system of equations produced by the discrete fourier transform of (47) to yield:

$$W^{(0)}[n] = \frac{1}{N} \sum_{k=0}^{N-1} e^{\jmath \frac{2\pi k}{N}(n + \lceil \frac{L}{2} \rceil)} \tilde{H}^{-1}[k; \hat{\Theta}^{(0)}] \tag{48}$$

where $\tilde{H}[k;\Theta] = \{\tilde{h}_{ij}[k;\Theta] \ \ \forall i,j\}$.

B. Experimental Setup

We test our algorithms using 512-sample windows of two spatially independent source signals. We use two different examples to show the performance of the MDPMBD algorithm. In the unrealistic example in Section V-B.1, we use an i.i.d. gaussian mixture model for the source signals and a parameterized two-point minimum-phase FIR filter. This

example is created to allow Amari's algorithm to successfully converge. It has no connection to any known channel but is chosen for illustrative purposes. In the realistic example in Section V-B.2, we use two different speech waveforms sampled at 8kHz for the source signals and pass them through a line-of-sight (LOS) channel. Conceptually, this example that is closer to what we would observe in nature. We use (6) to create the observations. We use the MDPMBD algorithm given at the end of Section III to compute the estimate of the channel parameter and source signals. We make use of the gradients in Appendix II to calculate the components of (35). We then use (43) - (48) to implement Amari's algorithm. To measure the performance of the proposed and comparison algorithm, we will use the average signal-to-interference ratio (SIR) defined as:

$$\text{SIR}_{\text{avg}} = \frac{1}{M_t}\sum_{j=1}^{M_t}\frac{\sum_{n=n_0}^{n_0+N-1}|x_j[n;\hat{\Theta}]|^2}{\sum_{n=n_0}^{n_0+N-1}|x_j[n;\hat{\Theta}]-s_j[n]|^2} \tag{49}$$

where $x_j[n;\Theta]$ and $s_j[n]$ are the standardized estimated and true source signals respectively. (We must also properly align $x_j[n;\Theta]$ and $s_j[n]$ before computing the denominator.) This appears to be the best way to compare the performance of an algorithm based on inverse filter separation to our algorithm based on parametric symbolic pseudoinverse separation.

1) Unrealistic Example: Here we define the channel model to be

$$h_{ij}[n;\Theta] = \delta[n] + (-1)^i\theta_j\delta[n-1] \quad \theta_j \in (0,1) \tag{50}$$

$\forall j = 1,2,\ \ \forall i = 1,2,$ and $\forall n = n_0, ..., n_0+N-1$ and

$$\frac{\partial}{\partial\theta_m}\tilde{h}_{ij}[k;\Theta] = \begin{cases} (-1)^i e^{\jmath\frac{2\pi k}{N}}, & m = j \\ 0, & \text{otherwise} \end{cases}. \tag{51}$$

The source signals are i.i.d. and have common pdf given by the gaussian mixture

$$f_{S_j}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x_j-\eta_1)^2}{2\sigma^2}}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x_j-\eta_2)^2}{2\sigma^2}} \tag{52}$$

where we chose $\sigma = 0.4$, $\eta_1 = 1$, and $\eta_2 = -1$ for our simulation. This means the nonlinear function $q_j(x)$ from (46) will be $q_j(x) = \frac{2x}{\sigma^2}$. We choose $\Theta^* = [0.2, 0.6]^T$ and initialize the algorithm with $\hat{\Theta}^{(0)} = [0.3, 0.7]^T$. In Fig. 4 we plot the average SIR from (49) for both the Amari algorithm and the proposed algorithm as a function of the time sample index. We compare and show the performance for two different sets of step sizes. Generally, the larger the step size, the faster the algorithm converges but the more it oscillates about the solution. Picking a step size that is too large can also lead to convergence failure.

2) Realistic Example: We now consider a number of observable signals arriving in planar wavefronts upon an arbitrary collection of sensors at known positions $\mathbf{P} = \{\mathbf{p}_i = [p_{x_i}, p_{y_i}]^T \ \ \forall i\}$ as in Fig. 5. When the sensors are spaced far enough apart, each receiver sees a different delay for each source. Now, we define the 2-D directional unit vector for source j as $[-\cos\theta_j, -\sin\theta_j]^T$ and project it onto the position vector $\mathbf{p}_i = [p_{x_i}, p_{y_i}]^T$ for sensor i. This gives the distance between the sensor and the origin along the direction of arrival

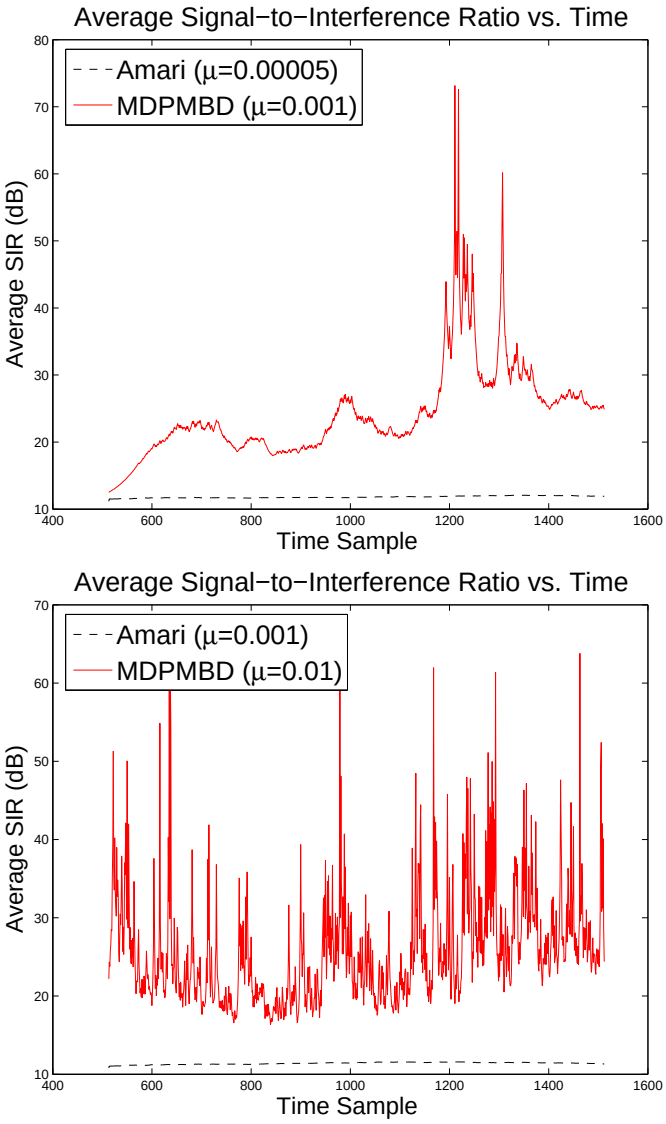

Fig. 4. Unrealistic Example Performance: Average SIR as computed in (49) for relatively small step sizes in the top figure and relatively large step sizes in the bottom figure. (Note that the convergence time and variability of the SIR in Amari's algorithm is unnoticable since the difference in dB between the algorithms is so large.)

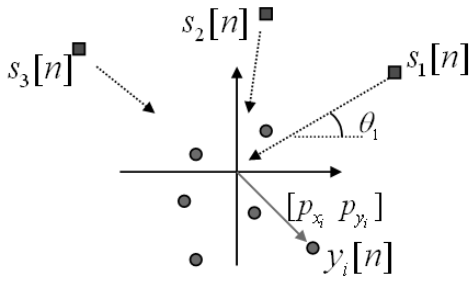

Fig. 5. In this line-of-sight channel model we illustrate a sensor array with source signals $s_j[n]$ impinging at angles of arrival θ_j to create a received signal $y_i[n]$ at sensor position $[p_{x_i}, p_{y_i}]^T$.

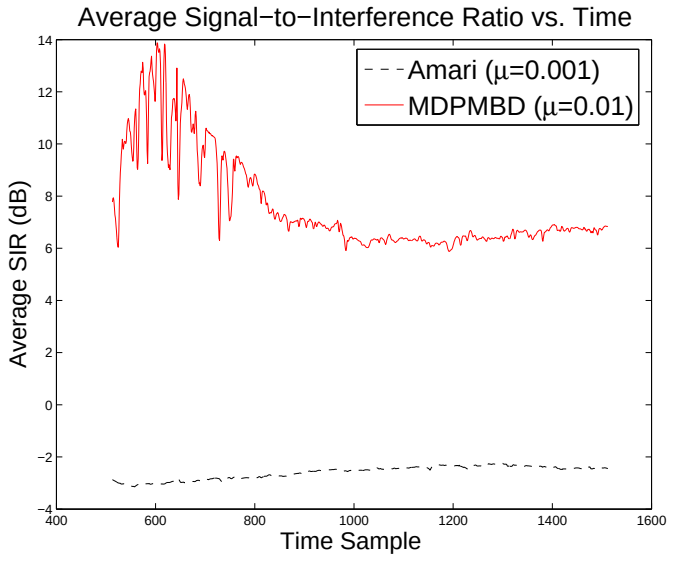

Fig. 6. Realistic Example Performance: We plot average SIR versus time sample for the two algorithms as before. Notice that the average SIR for Amari's algorithm is below 0 dB which means that the algorithm is failing to separate the sources.

(DOA) of the j-th source signal. The relative time delay is then simply this distance divided by the speed c of propagating wavefront to give

$$\tau_{ij} = -\frac{1}{c}(p_{x_i}\cos(\theta_j) + p_{y_i}\sin(\theta_j)) \tag{53}$$

For a real-valued signal in a LOS channel, plugging (53) into (8) gives the far-field **real-valued line-of-sight channel model**:

$$h_{ij}[l;\Theta] = \alpha_j \text{sinc}(l + \frac{1}{cT}(p_{x_i}\cos(\theta_j) + p_{y_i}\sin(\theta_j))) \tag{54}$$

$\forall j = 1, 2,\ \ \forall i = 1, ..., 8$, and $\forall n = n_0, ..., n_0 + N - 1$ and

$$\frac{\partial}{\partial\theta_m}\tilde{h}_{ij}[k;\theta_m] = \sum_{n=0}^{N-1} e^{-\frac{\jmath 2\pi kn}{N}}\left(\frac{\cos(\eta_{ijn})\gamma_{ij}}{\eta_{ijn}} - \frac{\sin(\eta_{ijn})\gamma_{ij}}{\eta_{ijn}^2}\right) \tag{55}$$

where $\eta_{ijn} = \pi(n - L_1) + \frac{\pi}{cT}(p_{x_i}\cos(\theta_j) + p_{y_i}\sin(\theta_j))$ and $\gamma_{ij} = \frac{\partial}{\partial\theta_j}\eta_{ijn}$. We have α_j=amplitude of source j, c=propagation speed, θ_j=source j direction of arrival, T=sampling period, $[p_{x_i}\ \ p_{y_i}]$=2-D position of sensor i, and $\Theta = [\theta_1, \theta_2, ..., \theta_{M_t}]$ are the DOAs of the source signals. We let the source signals for this simulation be two acoustic waveforms impinging on a 8-element uniformly spaced circular array. We use $\Theta^* = [45^o, 60^o]^T$ and $\Theta^{(0)} = [55^o, 70^o]^T$ in the simulation. In Fig. 6 we plot average SIR versus time sample for the two algorithms as before. Amari's algorithm fails to converge primarily due to the fact that the polynomial matrix with terms from (54) is not minimum phase. Notice that the average SIR for Amari's algorithm is below 0 dB which means that the algorithm is failing to separate the sources.

VI. Conclusion

In this paper we examined the problem of multichannel blind deconvolution (MBD) for the specific case in which the probability density functions of the source signals are known. We focused attention on parametric channel models to reduce the dimension of the search space compared with nonparametric channels (in which case every element of the separation system polynomial matrix must be estimated). We derived a new MBD algorithm and proved its asymptotic optimality in the sense of being statistically consistent and Fisher efficient. We compared the performance of our new estimator to a popular minimum entropy inverse filter technique used for MBD. For the contrived example where the minimum entropy inverse filter method actually achieves source separation, the average signal-to-interference ratio of the proposed estimator exceeded it by 15-20 dB. For the more realistic physics-based line-of-sight acoustic example, the minimum entropy inverse filter method completely failed, while the proposed estimator achieved an average SIR around 7 dB.

We would like to extend the algorithm to do multichannel blind deconvolution on complex-valued wireless communication signals operating in both line-of-sight and non-line-of-sight multipath models. We would like to derive a recursive form of the estimator instead of the current batch formulation. We need to examine effects of channel and source modeling errors and devise methods and cost functions to make the estimator robust against these effects. Finally we would like to analyze what happens when the number of sources is unknown. Or, analogously, what happens when the algorithm is designed to handle more source signals than actually appear?

Appendix I
Proof of Theorem 1

To derive the Riemannian metric for (24), we use (29) on (23) and assert spatial independence. For notational compactness we define $\sum_{x_1,...,x_{M_t}} \equiv \sum_{x_1\in\mathcal{X}}\cdots\sum_{x_{M_t}\in\mathcal{X}}$

$$\begin{aligned} g_{ij}(\Theta) &= -\frac{\partial}{\partial\theta_i'}\frac{\partial}{\partial\theta_j''}\sum_{x_1,...,x_{M_t}} f(x_1,...,x_{M_t}|\Theta')\times \\ &\quad \log\frac{f(x_1,...,x_{M_t}|\Theta')}{f(x_1,...,x_{M_t}|\Theta'')}\Big|_{\Theta'=\Theta''=\Theta} \\ &= \frac{-\partial}{\partial\theta_j''}\Bigg[\sum_{x_1,...,x_{M_t}}\frac{\partial}{\partial\theta_i'}f(x_1,...,x_{M_t}|\Theta')\times \\ &\quad \log\frac{f(x_1,...,x_{M_t}|\Theta')}{f(x_1,...,x_{M_t}|\Theta'')} + \sum_{x_1,...,x_{M_t}} f(x_1,...,x_{M_t}|\Theta')\times \\ &\quad \frac{f(x_1,...,x_{M_t}|\Theta'')}{f(x_1,...,x_{M_t}|\Theta')}\frac{\frac{\partial}{\partial\theta_i'}f(x_1,...,x_{M_t}|\Theta')}{f(x_1,...,x_{M_t}|\Theta'')}\Bigg]\Bigg|_{\Theta'=\Theta''=\Theta} \\ &= \sum_{x_1,...,x_{M_t}}\left(\frac{\partial}{\partial\theta_i}\log f(x_1,...,x_{M_t}|\Theta)\right)\times \\ &\quad \left(\frac{\partial}{\partial\theta_j}\log f(x_1,...,x_{M_t}|\Theta)\right) f(x_1,...,x_{M_t}|\Theta) \end{aligned}$$

$$\overset{\text{spatial}}{\underset{i.i.d.}{=}} \sum_{x_1,\ldots,x_{M_t}} \left(\sum_{p=1}^{M_t} \frac{\frac{\partial}{\partial\theta_i} f(x_p|\Theta)}{f(x_p|\Theta)}\right) \times \left(\sum_{q=1}^{M_t} \frac{\frac{\partial}{\partial\theta_j} f(x_q|\Theta)}{f(x_q|\Theta)}\right) \prod_{r=1}^{M_t} f(x_r|\Theta)$$

$$= \sum_{x_1,\ldots,x_{M_t}} \sum_{p=1}^{M_t} \frac{\frac{\partial}{\partial\theta_i} f(x_p|\Theta) \frac{\partial}{\partial\theta_j} f(x_p|\Theta)}{f^2(x_p|\Theta)} \prod_{r=1}^{M_t} f(x_r|\Theta) + \sum_{x_1,\ldots,x_{M_t}} \sum_{p=1}^{M_t} \sum_{q\neq p} \frac{\frac{\partial}{\partial\theta_i} f(x_p|\Theta) \frac{\partial}{\partial\theta_j} f(x_p|\Theta)}{f(x_p|\Theta) f(x_q|\Theta)} \prod_{r=1}^{M_t} f(x_r|\Theta)$$

$$= \sum_{p=1}^{M_t} E_{f(x_p|\Theta)} \left[\frac{\partial}{\partial\theta_i} \log f(x_p|\Theta) \frac{\partial}{\partial\theta_j} \log f(x_p|\Theta)\right]$$

This completes the proof of Theorem 1. In the second and penultimate step we are able to cancel the second term altogether using the fact that $\sum_{x\in\mathcal{X}} \frac{\partial}{\partial\theta} f(x|\Theta) = 0$. Also, throughout the proof, we use $\frac{\partial}{\partial\theta} \log f(x|\Theta) = \frac{\frac{\partial}{\partial\theta} f(x|\Theta)}{f(x|\Theta)}$. We see that $G(\Theta) = \{g_{ij}(\Theta)\}$ is the sum of Fisher information matrices.

Appendix II
Gradient Calculations for Simulation Examples

For the two examples given in Section V-B we have the following gradients for the terms in (33)

$$\begin{aligned}
\frac{\partial}{\partial\theta_1} a_{11}[k;\Theta] &= \frac{\partial}{\partial\theta_2} a_{22}[k;\Theta] = 0 \qquad (56)\\
\frac{\partial}{\partial\theta_1} a_{12}[k;\Theta] &= -\sum_{i=1}^{M_r} \frac{\partial}{\partial\theta_1} \tilde{h}^*_{i1}[k;\theta_1] \tilde{h}_{i2}[k;\theta_2]\\
\frac{\partial}{\partial\theta_1} a_{21}[k;\Theta] &= \left[\frac{\partial}{\partial\theta_1} a_{12}[k;\Theta]\right]^*\\
\frac{\partial}{\partial\theta_1} a_{22}[k;\Theta] &= 2\Re\{\sum_{i=1}^{M_r} \frac{\partial}{\partial\theta_1} \tilde{h}^*_{i1}[k;\theta_1] \tilde{h}_{i1}[k;\theta_1]\}\\
\frac{\partial}{\partial\theta_2} a_{11}[k;\Theta] &= 2\Re\{\sum_{i=1}^{M_r} \frac{\partial}{\partial\theta_2} \tilde{h}^*_{i2}[k;\theta_2] \tilde{h}_{i2}[k;\theta_2]\}\\
\frac{\partial}{\partial\theta_2} a_{12}[k;\Theta] &= -\sum_{i=1}^{M_r} \frac{\partial}{\partial\theta_1} \tilde{h}_{i1}[k;\theta_1] \tilde{h}^*_{i2}[k;\theta_2]\\
\frac{\partial}{\partial\theta_2} a_{21}[k;\Theta] &= \left[\frac{\partial}{\partial\theta_2} a_{12}[k;\Theta]\right]^*
\end{aligned}$$

$$\frac{\partial}{\partial\theta_1} d[k;\Theta] = \sum_{i=1}^{M_r} \frac{\partial}{\partial\theta_1} |\tilde{h}_{i1}[k;\theta_1]|^2 \sum_{i=1}^{M_r} |\tilde{h}_{i2}[k;\theta_2]|^2 \qquad (57)$$
$$-2\Re\left[\sum_{i=1}^{M_r} \frac{\partial}{\partial\theta_1} \tilde{h}^*_{i1}[k;\theta_1] \tilde{h}_{i2}[k;\theta_2] \sum_{i=1}^{M_r} \tilde{h}^*_{i2}[k;\theta_2] \tilde{h}_{i1}[k;\theta_1]\right]$$

$$\frac{\partial}{\partial\theta_2} d[k;\Theta] = \sum_{i=1}^{M_r} \frac{\partial}{\partial\theta_2} |\tilde{h}_{i2}[k;\theta_2]|^2 \sum_{i=1}^{M_r} |\tilde{h}_{i1}[k;\theta_1]|^2$$
$$-2\Re\left[\sum_{i=1}^{M_r} \frac{\partial}{\partial\theta_2} \tilde{h}^*_{i2}[k;\theta_2] \tilde{h}_{i1}[k;\theta_1] \sum_{i=1}^{M_r} \tilde{h}^*_{i1}[k;\theta_1] \tilde{h}_{i2}[k;\theta_2]\right]$$

Acknowledgment

The authors would like to thank the reviewers of this paper for their time and consideration.

References

[1] S. Haykin, editor, *Unsupervised Adaptive Filtering Volume I: Blind Source Separation*, John Wiley and Sons, 2000.
[2] A. Hyvärinen, J. Karhunen, E. Oja, *Independent Component Analysis*, John Wiley and Sons, 2001.
[3] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*, John Wiley and Sons, 2002.
[4] S. Amari, S. Douglas, A. Cichocki, H. Yang "Novel on-line adaptive learning algorithms for blind deconvolution using the natural gradient approach," *IEEE 11th IFAC Symposium on System Identification*, July 1997.
[5] S. Douglas, H. Sawada, and S. Makino, "Natural gradient multichannel blind deconvolution and speech separation using causal FIR filters," *IEEE Transactions on Speech and Audio Processing*, Vol. 13, No. 1, January 2005.
[6] Tugnait, J.K., "Identification and deconvolution of multichannel linear non-Gaussian processes using higher order statistics and inverse filter criteria" *Signal Processing, IEEE Transactions on*, Volume: 45, Issue: 3 , March 1997.
[7] S. Amari, S. Douglas, A. Cichocki, H. Yang "Multichannel blind deconvolution and equalization using the natural gradient," *IEEE Signal Processing Advances in Wireless Communications*, 1997.
[8] Inouye, Y.; Tanebe, K., "Super-exponential algorithms for multichannel blind deconvolution" *Signal Processing, IEEE Transactions on*, Volume: 48, Issue: 3 , March 2000.
[9] K. Rahbar, J. Reilly, J. Manton, "Blind identification of MIMO FIR systems driven by quasi-stationary sources using second order statistics: a frequency domain approach," *IEEE Trans. on Signal Processing*, Vol. 52, Issue: 2, Feb. 2004.
[10] L. Parra and C. Spence, "Convolutive blind separation of nonstationary source," *IEEE Trans. Speech Audio Processing*, vol. 8, no. 3, pp. 320-327, May 2000.
[11] K. Rahbar and J. Reilly, "Blind source separation of convolved sources by joint approximate diagonalization of cross-spectral density matrices", *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, Volume: 5 , 7-11 May 2001.
[12] C. Ma; Z. Ding; S. Yau, "A two-stage algorithm for MIMO blind deconvolution of nonstationary colored signals" *Signal Processing, IEEE Transactions on*, Volume: 48 , Issue: 4 , April 2000.
[13] S. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, 10, 251-276, 1998.
[14] S. Amari and H. Nagaoka, *Methods of Information Geometry*, American Mathematical Society, 2000.
[15] S. Eguchi, "Second order efficiency of minimum contrast estimators in a curved exponential family," *Annals of Statistics*, Vol. 11, No. 3, 793-803, 1983.
[16] J. Proakis, *Digital Communications: Fourth Edition*, McGraw-Hill, 2001.
[17] L. Rabiner and R. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall Inc., 1978.
[18] T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley and Sons, Inc. 1991.

Robert M. Taylor Jr. received the B.S. degree in Electrical Engineering in 1996 and the M.S. degree in Electrical Engineering in 1997 from the Georgia Institute of Technology, Atlanta, GA. After graduation, he was a research engineer with Johns Hopkins University Applied Physics Laboratory, Laurel, MD until 2001. Since then he has been a senior signal processing engineer at MITRE Corporation, McLean, VA. He enrolled at Virginia Tech in 2003 where he is currently pursuing a Ph.D. in Electrical Engineering. His research interests include statistical signal processing, information theory, digital communications, and information geometry.

Lamine Mili (SM90), received the electrical engineering diploma from the Swiss Federal Institute of Technology, Lausanne, in 1976, and the Ph.D. degree from the University of Liege, Belgium, in 1987. Currently, he is a Professor of electrical engineering at the Alexandria Research Institute of Virginia Tech. His research interests include robust statistics, robust estimation and detection, clutter mitigation in radar systems, image and speech processing, long memory processes, system identification, risk management of complex systems, nonlinear dynamics and control, and power system analysis and control. Dr. Mili is a senior member of the Power Engineering Society of IEEE, the recipient of a 1990 NSF Research Initiation Award, and of an 1992 NSF Young Investigator Award.

Amir I. Zaghloul (F'92) received the Ph.D. and M.A.Sc. degrees from the University of Waterloo, Canada in 1973 and 1970, respectively, and the B.Sc. degree (Honors) from Cairo University, Egypt in 1965, all in electrical engineering. In 2001 he joined Virginia Polytechnic Institute and State University (Virginia Tech) as Professor in the Bradley Department of Electrical and Computer Engineering. Prior to Virginia Tech, he was at COMSAT Laboratories for 24 years performing and directing R&D efforts on satellite communications and antennas, where he received several research and patent awards, including the Exceptional Patent Award. He held positions at the University of Waterloo, Canada (1968-1978), University of Toronto, Canada (1973-74), Aalborg University, Denmark (1976) and Johns Hopkins University, Maryland (1984-2001). He is a Fellow of the IEEE and the recipient of the 1986 Wheeler Prize Award for Best Application Paper in the IEEE Transactions on Antennas and Propagation. He is also an Associate Fellow for The American Institute of Aeronautics and Astronautics (AIAA), a Member of Commissions A & B of the International Union of Radio Science (URSI), and member of the IEEE Committee on Communications and Information Policy (CCIP).

Dr. Zaghloul is the general chair of the "IEEE International Symposium on Antennas and Propagation and USNC/URSI Meeting," Washington, D.C., July 2005.