BLIND DECONVOLUTION AND ICA WITH A BANDED MIXING MATRIX

Sam T. Kaplan and Tadeusz J. Ulrych

Department of Earth and Ocean Sciences, University of British Columbia,
6339 Stores Road, Vancouver, British Columbia, Canada V6T 1Z4

ABSTRACT

Convolution is a linear operation, and, consequently, can be formulated as a linear system of equations. If only the output of the system (the convolved signal) is known, then the problem is blind so that given one equation, two unknowns are sought. Here, the blind deconvolution problem is solved using independent component analysis (ICA). To facilitate this, several time lagged versions of the convolved signal are extracted and used to construct realizations of a random vector. For ICA, this random vector is the, so called, mixture vector, created by the matrix-vector multiplication of the two unknowns, the mixing matrix and the source vector. Due to the properties of convolution, the mixing matrix is banded with its nonzero elements containing the convolution's filter. This banded property is incorporated into the ICA algorithm as prior information, giving rise to a banded ICA algorithm (B-ICA) which is, in turn, used in a new blind deconvolution method. B-ICA produces as many independent components as the dimension of the filter; whereas for blind deconvolution, only one signal is sought (the deconvolved signal). Fortunately, the convolutional model provides additional information which enables one best independent component to be extracted from the pool of candidate solutions. This, in turn, yields estimates of both the filter and the deconvolved signal.

1. INTRODUCTION

Consider two time sequences, $h(t)$ and $\rho(t)$, and their convolution,

$$\begin{aligned} \chi(t) &= h(t) * \rho(t) \\ &= \int_{-\infty}^{\infty} h(t-\tau)\,\rho(\tau)d\tau. \end{aligned} \tag{1}$$

Neglecting noise, Eq. (1) is often used to model seismic data where $\chi(t)$ is a seismic trace generated by convolving a wavelet (the filter), $h(t)$, with the Earth's reflectivity, $\rho(t)$. This linear representation of the truth is useful but, for real data, introduces an equation with two unknowns (the wavelet and the reflectivity). Blind deconvolution solves for the two unknowns given only the trace. In this paper, independent component analysis (ICA) is used in a blind deconvolution algorithm. Formulated as an ICA problem, convolution gives rise to a banded mixing matrix which is incorporated into ICA as prior information yielding a banded ICA algorithm (B-ICA) and, consequently, a new method for blind deconvolution.

Wiggins [1] introduced a blind deconvolution algorithm called minimum entropy deconvolution where the statistics of the reflectivity were constrained using the varimax criteria (a measure of kurtosis). This algorithm, for a short time, was popular and has received much attention [2, 3, 4, 5]. Similar methods were, independently, derived by Shalvi and Weinstein [6]. More recently, Kaaresen and Taxt [7] derived an algorithm which explicitly incorporates the sparseness of the reflectivity by using a spike train as a model where the location, amplitude and number of spikes are considered. With the exception of Kaaresen and Taxt [7], all of these methods employ higher order statistics. As such, while the method presented in this paper is derived from ICA, it has roots reaching a wider scope of literature.

2. DISCRETE CONVOLUTION AND THE ICA MODEL

As is evident in Eq. (1), convolution is, of course, linear, and can be expressed as a linear system of equations. This lends itself to an ICA formulation of blind deconvolution which given only one time sequence, $\chi(t)$, allows for the reconstruction of both $h(t)$ and $\rho(t)$.

The convolutional model (Eq. (1)) is easily modified for discretely sampled signals such that

$$\chi(t_i) = \sum_j h(t_{i-j})\,\rho(t_j),$$

or equivalently,

$$\mathbf{A}\mathbf{s} = \mathbf{x} \tag{2}$$

where

$$\mathbf{s}^T = \begin{bmatrix} \rho(t_1) & \rho(t_2) & \cdots & \rho(t_n) \end{bmatrix},$$

$$\mathbf{x}^T = \begin{bmatrix} \chi(t_1) & \chi(t_2) & \cdots & \chi(t_n) \end{bmatrix}$$

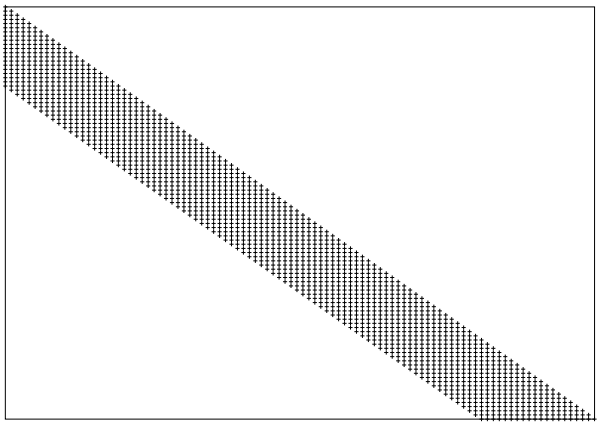

Fig. 1. A banded mixing matrix, $\mathbf{A}$, for convolution ($n = 100$ and $nw = 20$).

and $\mathbf{A}$ is an $n \times n$ banded matrix with columns constructed from delayed versions of the wavelet, $\mathbf{h}$, such that

$$\begin{aligned} \mathbf{A} &= \left[\begin{array}{ccccc} \mathbf{a}_1 & | & \mathbf{a}_2 & | & \cdots & | & \mathbf{a}_n \end{array} \right] \\ &= \left[\begin{array}{ccccc} \mathbf{N}_1\mathbf{h} & | & \mathbf{N}_2\mathbf{h} & | & \cdots & | & \mathbf{N}_n\mathbf{h} \end{array} \right] \end{aligned} \quad (3)$$

and

$$\mathbf{h}^T = \left[\begin{array}{cccc} h(t_1) & h(t_2) & \cdots & h(t_{nw}) \end{array} \right].$$

$\mathbf{N}_i$ are zero padding matrices [8, p. 107] where the i^{th} element of $\mathbf{a}_i$ is $h(t_1)$, element $(i+1)$ is $h(t_2)$ and so on. Fig. 1 illustrates $\mathbf{A}$, showing the nonzero elements of the matrix when $n = 100$ and $nw = 20$.

Eq. (2) is recognized as the ICA model where $\mathbf{s}$ are sources, $\mathbf{x}$ are mixtures, $\mathbf{A}$ is the mixing matrix and $\mathbf{s}$ and $\mathbf{x}$ are random vectors. However, the convolutional model in Eq. (2) provides only one realization of each $\mathbf{s}$ and $\mathbf{x}$. Obviously this is inadequate to characterize the corresponding statistics, and hence, is inadequate for ICA. However, the available information can be reorganized in a clever way, providing several realizations. The trick is to consider time delayed versions of $\rho(t)$ and $\chi(t)$ [9, p. 360]. In particular, let

$$\mathbf{s}^T = \left[\begin{array}{cccc} z^{n-1}\rho(t) & z^{n-2}\rho(t) & \cdots & z\rho(t) \quad \rho(t) \end{array} \right]$$

and

$$\mathbf{x}^T = \left[\begin{array}{cccc} z^{n-1}\chi(t) & z^{n-2}\chi(t) & \cdots & z\chi(t) \quad \chi(t) \end{array} \right]$$

where z is the unit delay operator. Hence, organizing the realizations of $\mathbf{s}$ into the columns of a matrix gives

$$\mathbf{s} = \left[\begin{array}{cccccc} 0 & 0 & 0 & \cdots & 0 & \rho(t_1) \\ 0 & 0 & 0 & \cdots & \rho(t_1) & \rho(t_2) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \rho(t_1) & \rho(t_2) & \cdots & \rho(t_{n-2}) & \rho(t_{n-1}) \\ \rho(t_1) & \rho(t_2) & \rho(t_3) & \cdots & \rho(t_{n-1}) & \rho(t_n) \end{array} \right] \quad (4)$$

and similarly,

$$\mathbf{x} = \left[\begin{array}{cccccc} 0 & 0 & 0 & \cdots & 0 & \chi(t_1) \\ 0 & 0 & 0 & \cdots & \chi(t_1) & \chi(t_2) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \chi(t_1) & \chi(t_2) & \cdots & \chi(t_{n-2}) & \chi(t_{n-1}) \\ \chi(t_1) & \chi(t_2) & \chi(t_3) & \cdots & \chi(t_{n-1}) & \chi(t_n) \end{array} \right] \quad (5)$$

where the j^{th} realization of $\mathbf{x}$ is the convolution of the j^{th} realization of $\mathbf{s}$ with the filter, $\mathbf{h}$. In other words,

$$\mathbf{x}(t_j) = \mathbf{h} * \mathbf{s}(t_j).$$

Thus, blind deconvolution is posed in a manner that can be solved using ICA. In particular, ICA computes some approximation to the rows of $\mathbf{s}$, each containing a portion of the reflectivity. However, ICA does not directly recover $\mathbf{h}$; but rather, a mapping between the mixtures and the independent components. Additionally, recall that ICA relies on computing the statistics of the independent components. Clearly the first few rows of $\mathbf{s}$ and $\mathbf{x}$ provide few nonzero realizations, and thus, do little to define the statistics of their corresponding random variables.

3. BANDED ICA

Let the following notation be used to describe ICA:

$$\mathbf{As} = \mathbf{x} \quad \mathbf{\Gamma As} = \mathbf{\Gamma x} = \mathbf{Ws} = \mathbf{z} \quad \mathbf{y} = \mathbf{Bx} \quad \mathbf{y} = \mathbf{Qz} \quad (6)$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{W}$ and $\mathbf{Q}$ are $m \times m$ matrices, $\mathbf{\Gamma}$ is chosen such that $\mathbf{z}$ is white, and $\mathbf{Q}$ is chosen such that the elements of $\mathbf{y}$, y_i, are independent components. That is, $\mathbf{Q}$ is chosen such that $y_i \propto s_j$ for some j where s_j is the j^{th} element of $\mathbf{s}$. Additionally, define a new matrix, $\mathbf{P}$, such that

$$\mathbf{Py} = \mathbf{x} \quad (7)$$

and

$$\mathbf{P} = \left[\begin{array}{ccccc} \mathbf{p}_1 & | & \mathbf{p}_2 & | & \cdots & | & \mathbf{p}_m \end{array} \right].$$

By definition, $\mathbf{y}$ is a scaled and permuted version of $\mathbf{s}$; hence, $\mathbf{P}$ and $\mathbf{A}$ provide similar mappings. Here the ICA algorithm is modified such that instead of finding rows of $\mathbf{Q}$, it finds columns of $\mathbf{P}$. This, conveniently, allows for application of the given prior knowledge to ICA. Namely, the banded nature of $\mathbf{A}$ can be applied to $\mathbf{P}$, leading to the new B-ICA algorithm and a solution to the blind deconvolution problem.

A popular method for ICA involves finding a minimum of some cost function, $\phi(\mathbf{q}_i)$, which measures the entropy of an independent component, $y_i = \mathbf{q}_i^T\mathbf{z}$, where $\mathbf{q}_i^T$ is the i^{th} row of $\mathbf{Q}$. Hence, ICA finds the rows of $\mathbf{Q}$, $\mathbf{q}_i$. As is well known, a relationship between $\mathbf{q}_i$ and $\mathbf{p}_i$ is readily

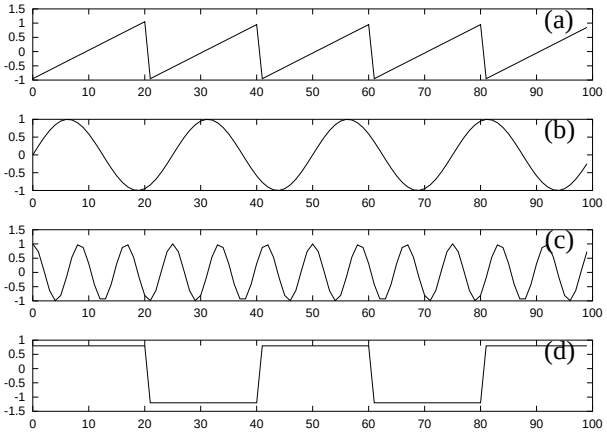

Fig. 2. An example of B-ICA The four sources (a)-(d) are mixed using Eq. (11) producing Fig. 3.

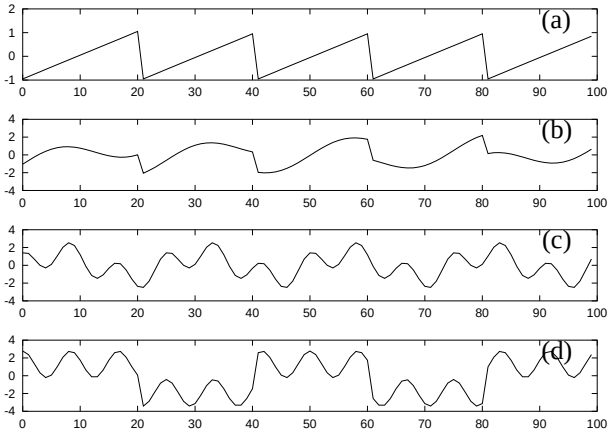

Fig. 3. An example of B-ICA. The four mixtures (a)-(d) are produced by mixing the sources in Fig. 2 according to the mixing matrix in Eq. (11).

found. Noting that the independent components, y_i, are zero mean and uncorrelated random variables with unit variance, and that $\mathbf{y} = \mathbf{Q}\mathbf{z}$ (Eq. (6)) gives

$$\mathrm{E}\left(\mathbf{y}\mathbf{y}^T\right) = \mathrm{E}\left(\mathbf{Q}\mathbf{z}\mathbf{z}^T\mathbf{Q}^T\right) = \mathbf{Q}\mathrm{E}\left(\mathbf{z}\mathbf{z}^T\right)\mathbf{Q}^T = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}$$

where $\mathbf{I}$ is the identity matrix. Therefore, assuming that $\mathbf{Q}^{-1}$ exists,

$$\begin{aligned}\mathbf{Q}^{-1}\mathbf{Q}\mathbf{Q}^T &= \mathbf{Q}^{-1} \\ \mathbf{Q}^T &= \mathbf{Q}^{-1}.\end{aligned}$$

Further, using Eq. (6),

$$\mathbf{z} = \mathbf{Q}^{-1}\mathbf{y} = \mathbf{Q}^T\mathbf{y} = \mathbf{\Gamma}\mathbf{x}.$$

Therefore,

$$\mathbf{x} = \mathbf{\Gamma}^{-1}\mathbf{Q}^T\mathbf{y}. \tag{8}$$

Eqs. (7) and (8) allow for the explicit formulation of $\mathbf{P}$ in terms of $\mathbf{Q}$ such that

$$\mathbf{P} = \mathbf{\Gamma}^{-1}\mathbf{Q}^T \qquad \mathbf{Q}^T = \mathbf{\Gamma}\mathbf{P}.$$

Hence, $\mathbf{q}_i = \mathbf{\Gamma}\mathbf{p}_i$, and

$$y_i = \mathbf{q}_i^T\mathbf{z} = \left(\mathbf{\Gamma}\mathbf{p}_i\right)^T\mathbf{z} = \left(\mathbf{\Gamma}\mathbf{N}_i\mathbf{h}\right)^T\mathbf{z} = \mathbf{h}^T\mathbf{N}_i^T\mathbf{\Gamma}^T\mathbf{z} = \mathbf{h}^T\tilde{\mathbf{x}} \tag{9}$$

where $\tilde{\mathbf{x}} = \mathbf{N}_i^T\mathbf{\Gamma}^T\mathbf{z}$ and $\mathbf{N}_i$ is a zero padding matrix which maps $\mathbf{h}$ to the i^{th} column of $\mathbf{P}$, $\mathbf{p}_i$, which, in turn, corresponds to a particular column of the mixing matrix. In other words, $\mathbf{N}_i$ is the prior information. It enforces the banded property of the mixing matrix, $\mathbf{A}$, by explicitly choosing the number and location of the zero entries in $\mathbf{p}_i$, and in doing so, forces $\mathbf{p}_i$ to correspond to columns of $\mathbf{A}$ which have an equivalent sparse structure.

In Eq. (9), $\tilde{\mathbf{x}}$ can be thought of as a new set of mixtures (nw in total) with corresponding independent components, $\tilde{\mathbf{y}}$, such that

$$\tilde{\mathbf{y}} = \tilde{\mathbf{B}}\tilde{\mathbf{x}}$$

where

$$\tilde{\mathbf{B}}^T = \left[\begin{array}{ccccc} \tilde{\mathbf{h}}_1 & | & \tilde{\mathbf{h}}_2 & | & \cdots & | & \tilde{\mathbf{h}}_{nw} \end{array}\right]$$

is an $nw \times nw$ matrix; thus, assuming an ICA model, Eq. (9) can be generalized such that

$$\tilde{\mathbf{A}}\tilde{\mathbf{s}} = \tilde{\mathbf{x}} \quad \tilde{\mathbf{\Gamma}}\tilde{\mathbf{A}}\tilde{\mathbf{s}} = \tilde{\mathbf{\Gamma}}\tilde{\mathbf{x}} = \tilde{\mathbf{W}}\tilde{\mathbf{s}} = \tilde{\mathbf{z}} \quad \tilde{\mathbf{y}} = \tilde{\mathbf{B}}\tilde{\mathbf{x}} \quad \tilde{\mathbf{y}} = \tilde{\mathbf{Q}}\tilde{\mathbf{z}} \tag{10}$$

where $\tilde{\mathbf{B}} = \tilde{\mathbf{Q}}\tilde{\mathbf{\Gamma}}$. Hence, given $\tilde{\mathbf{x}}$, the ICA algorithm is used to find $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{B}}$ where one element of $\tilde{\mathbf{y}}$ is the desired independent component, and one row of $\tilde{\mathbf{B}}$ is the nonzero elements of one column of $\mathbf{P}$. In other words, $\tilde{\mathbf{h}}_i \propto \mathbf{h}$ for some i.

The above algorithm can be further generalized so that $\mathbf{p}_i = \mathbf{N}_i\mathbf{h}_i$, where $\mathbf{h}_i$ are all of dimension nw, giving a more general form of Eq. (3) such that

$$\mathbf{A} = \left[\begin{array}{ccccc} \mathbf{N}_1\mathbf{h}_1 & | & \mathbf{N}_2\mathbf{h}_2 & | & \cdots & | & \mathbf{N}_m\mathbf{h}_m \end{array}\right].$$

For example, consider the mixing matrix,

$$\mathbf{A} = \begin{bmatrix} 1.0 & 0 & 0 & 0 \\ 1.1 & 1.2 & 0 & 0 \\ 0 & 1.3 & 1.4 & 0 \\ 0 & 0 & 1.5 & 1.6 \end{bmatrix} \tag{11}$$

which provides a mapping between the sources, $\mathbf{s}$, in Fig. 2 and the mixtures, $\mathbf{x}$, in Fig. 3. New mixtures, $\tilde{\mathbf{x}}$, are computed according to Eq. (9) such that $\tilde{\mathbf{x}} = \mathbf{N}_2^T\mathbf{\Gamma}^T\mathbf{z}$. Fig. 4a

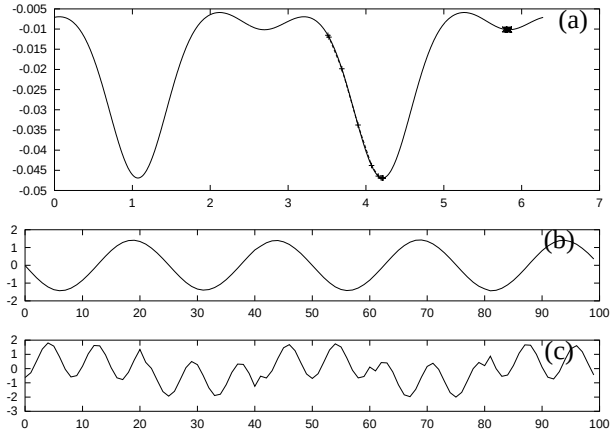

Fig. 4. An example of B-ICA. (a) The cost function computed from the mixtures in Fig. 3 using the prior information in $\mathbf{N}_2$ and plotted for $||\tilde{\mathbf{q}}||_2 = 1$. Superimposed on the cost function are the optimization paths which the algorithm followed to find the local minima. (b)-(c) The independent components $\tilde{\mathbf{y}}$ corresponding to the local minima in (a). Notice that the independent component in (b) is representative of the second source (Fig. 2b).

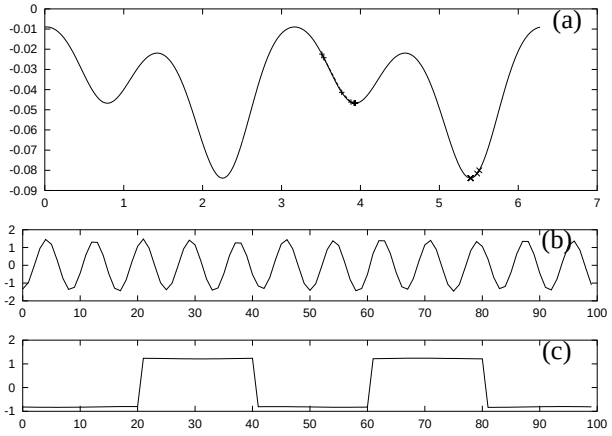

Fig. 5. A second example of B-ICA. (a) The cost function computed from the mixtures in Fig. 3 using the prior information in $\mathbf{N}_3$ and plotted for $||\tilde{\mathbf{q}}||_2 = 1$. Superimposed on the cost function are the optimization paths which the algorithm followed to find the local minima. (b)-(c) The independent components $\tilde{\mathbf{y}}$ corresponding to the local minima in (a). Notice that the independent components are representative of the sources in Figs. 2c-d.

plots the cost function, $\phi(\tilde{\mathbf{q}}_i)$, for $||\tilde{\mathbf{q}}_1||_2 = 1$, where $\tilde{\mathbf{q}}_i^T$ is the i^{th} row of $\tilde{\mathbf{Q}}$. The cost function is computed using the Gram-Charlier approximation to negentropy, $J(y_i)$, such that [10]

$$\phi(\tilde{\mathbf{p}}_i) = -J(\tilde{y}_i) \approx -\frac{\kappa_3^2(\tilde{y}_i)}{12} - \frac{\kappa_4^2(\tilde{y}_i)}{48}$$

where $\kappa_3(\tilde{y}_i)$ and $\kappa_4(\tilde{y}_i)$ are, respectively, the skewness and kurtosis of the i^{th} independent component in $\tilde{\mathbf{y}}$. As expected, there are four local minima corresponding to two independent components in $\tilde{\mathbf{y}}$. These independent components, $\tilde{y}_1$ and $\tilde{y}_2$, are plotted in Figs. 4b-c respectively. Clearly Fig. 4b corresponds to the source in Fig. 2b. For this example the prior information is $\mathbf{N}_2$ which constrains the ICA algorithm to find $\mathbf{a}_2$; consequently, it finds an independent component proportional to the second element of $\mathbf{s}$, s_2. Therefore, Fig. 4b is the expected result.

As a second example let $\tilde{\mathbf{x}} = \mathbf{N}_3^T \mathbf{\Gamma}^T \mathbf{z}$. The corresponding independent components are plotted in Fig. 5. Through examination of Eq. (11), it is clear that this prior information, $\mathbf{N}_3$, allows for both the third and fourth columns of $\mathbf{A}$. Both have the same nonzero elements, so both obey the prior information imposed by $\mathbf{N}_3$. As such, both the third and fourth elements of $\mathbf{s}$, s_3 and s_4, are represented in Figs. 5b-c respectively. The corresponding cost function is plotted in Fig. 5a.

While B-ICA allows for application of the given prior knowledge, it still presents a difficulty in the ambiguity of the result. Namely that the algorithm produces as many independent components as the dimension of $\mathbf{h}$ (or $\mathbf{h}_i$). As such, there is left the task of choosing one independent component and its corresponding row of $\tilde{\mathbf{B}}$. Fortunately, as will be shown, when B-ICA is used for blind deconvolution, a solution presents itself.

4. B-ICA FOR BLIND DECONVOLUTION

In Section 2 the convolutional model was formulated as an ICA problem with a banded mixing matrix. Here, B-ICA, presented in Section 3, is used to solve for the filter, $\mathbf{h}$. Unfortunately, the $\mathbf{x}$ and $\mathbf{s}$ proposed in Eqs. (4) and (5) are inadequate in that the first few mixtures and sources provide few nonzero realizations, doing little to constrain the statistics of their corresponding random variables. Therefore the algorithm must be modified to compensate for this lack of information. Further, as illustrated in Section 3, B-ICA provides as many independent components as the dimension of $\mathbf{h}$. The best solution must be chosen from the pool of candidate solutions, yielding one approximation of both $\mathbf{h}$ and $\rho(t)$.

To compensate for this lack of information in the first few mixtures and sources, an approximative convolutional model is used, such that

$$\mathbf{s}^T = \left[\begin{array}{cccc} z^{m-1}\rho(t) & z^{m-2}\rho(t) & \cdots & z\rho(t) \quad \rho(t) \end{array} \right] \tag{12}$$

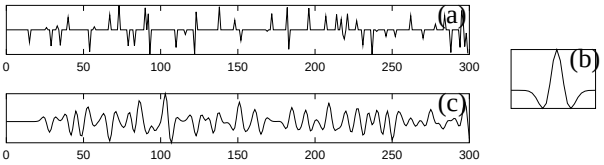

Fig. 6. B-ICA used for blind deconvolution. (a) A sparse spike train convolved with (b) the twenty-five point filter, $\mathbf{h}$, produces (c) the signal, $\chi(t)$. Given the data in (c), B-ICA finds the information, $\tilde{\mathbf{B}}$, presented in Fig. 7.

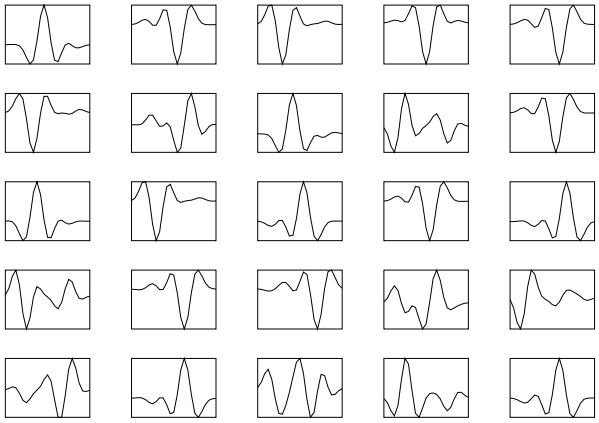

Fig. 7. B-ICA used for blind deconvolution. Plotted are the twenty-five rows of $\tilde{\mathbf{B}}$ computed using B-ICA and the data in Fig. 6c.

and $m < n$. Similarly from Eq. (4),

$$\mathbf{x}^T = \left[\begin{array}{cccc} z^{m-1}\chi(t) & z^{m-2}\chi(t) & \cdots & z\chi(t) \quad \chi(t) \end{array} \right]. \tag{13}$$

Hence, each random variable has a number of nonzero realizations to constrain their statistics. Given Eqs. (12) and (13), the mapping between sources, $\mathbf{s}$, and mixtures, $\mathbf{x}$, imposed by the convolutional mixing matrix (Eq. (3)) is not exact. In particular, through inspection of Eqs. (12) and (13), it is obvious that given $\mathbf{A}$ and $\mathbf{s}$, $x_i(t_j)$ is incorrectly mapped for

$$(i \in \{1 \ldots nw\}) \cap (j \in \{(m-1) \ldots n\}).$$

However, for the remainder of $\mathbf{x}$ the mapping is correct which, as will be illustrated, given only knowledge of $\chi(t)$ (i.e. $\mathbf{x}$ in Eq. (13)), allows ICA to find a wavelet following the true convolutional model.

Consider the synthetic time sequence, $\chi(t)$, in Fig. 6c which is the convolution of the sparse spike train, $\rho(t)$, in Fig. 6a with the twenty-five point filter, $\mathbf{h}$, (a Ricker wavelet) in Fig. 6b. Given only $\chi(t)$, B-ICA produces $\tilde{\mathbf{B}}$, the rows of which are plotted in Fig. 7. The matrix, $\tilde{\mathbf{B}}$, is computed using the ICA model in Eq. (10) which is arrived at using the prior information in the zero padding matrix, $\mathbf{N}_{(m-nw-10)}$. The algorithm employed uses the nonpolynomial expansion of negentropy [11] such that

$$J(\tilde{y}_i) \approx \frac{1}{2} \sum_{i=1}^{l} c_i^2$$

where $c_i = \mathrm{E}(r_i(\tilde{y}_i))$. The optimization routine of Hyvärinen [12] is used where $J(\tilde{y}_i)$ is approximated with $l = 1$, and $r_1(\tilde{y}_i) = \exp\left(-\frac{\tilde{y}_i^2}{2}\right)$. Additionally, the convolutional model is approximated such that $m = 100$. A quick search through the panels reveals good approximations to the wavelet.

While the $\tilde{h}_i$ in Fig. 7 are an interesting result, their utility is not immediately obvious. In practice the filter is, of course, not known. Hence, simply presenting the choices in Fig. 7 is inadequate. Instead, there is a sensible way to check for the best result. In particular, coefficients, c_i, can be calculated such that

$$\psi(c_i) = ||\mathbf{x}_k - c_i \tilde{\mathbf{h}}_i * \tilde{\mathbf{y}}_i||_2^2 \tag{14}$$

is minimized where, here,

$$\mathbf{x}_k^T = \left[\begin{array}{cccc} x_k(t_1) & x_k(t_2) & \cdots & x_k(t_N) \end{array} \right]$$

are the realizations of the k^{th} mixture and

$$\tilde{\mathbf{y}}_i^T = \left[\begin{array}{cccc} \tilde{y}_i(t_1) & \tilde{y}_i(t_2) & \cdots & \tilde{y}_i(t_N) \end{array} \right]$$

are the realizations of the i^{th} independent component. It is easily verified that Eq. (14) has its extreme points when

$$c_i = c_i^{(*)} = \frac{\mathbf{x}_k^T \left(\tilde{\mathbf{h}}_i * \tilde{\mathbf{y}}_i \right)}{\left(\tilde{\mathbf{h}}_i * \tilde{\mathbf{y}}_i \right)^T \left(\tilde{\mathbf{h}}_i * \tilde{\mathbf{y}}_i \right)}.$$

The best solution, $(\tilde{\mathbf{y}}_*, \tilde{\mathbf{h}}_*)$, is chosen such that

$$\psi\left(c_*^{(*)}\right) = \min_i \left\{ \psi\left(c_i^{(*)}\right) \right\} \quad , \quad i = 1 \ldots nw.$$

For example, consider the synthetic time sequence, $\chi(t)$, in Fig. 8c generated by convolving the spike train, $\rho(t)$, in Fig. 8a with the thirty-five point filter (a Berlage wavelet), $\mathbf{h}$, in Fig. 8b. B-ICA is used to compute $\tilde{\mathbf{B}}$ such that the prior information is the zero padding matrix, $\mathbf{N}_{(m-nw-10)}$, $m = 75$ and, again, the algorithm in [12] is used with $l = 1$ and $r_1(\tilde{y}_i) = \exp\left(-\frac{\tilde{y}_i^2}{2}\right)$. As a result, thirty-five wavelets are recovered, $\tilde{\mathbf{h}}_1 \ldots \tilde{\mathbf{h}}_{35}$. The best result $(\tilde{\mathbf{h}}_*, \tilde{y}_*(t))$ is extracted from the pool of thirty-five candidate solutions according to the criteria in Eq. (14) (with $k = m - nw - 10$) and is plotted in Figs. 8d-e. Fig. 8d is, as expected, an approximation to the sparse spike series in Fig. 8a with a linear phase shift. The shift is due to how the mixtures, $\mathbf{x}$, are

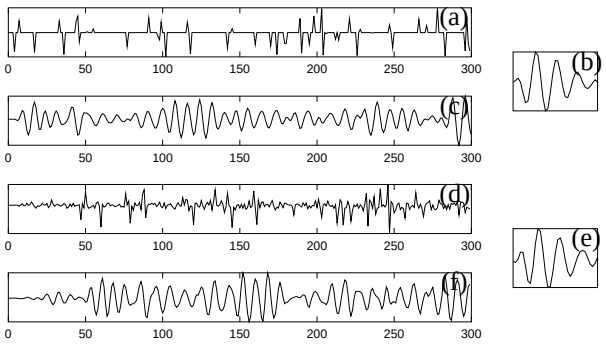

Fig. 8. B-ICA used for blind deconvolution. (a) A sparse spike train convolved with (b) the thirty-five point filter, $\mathbf{h}$, produces (c) the signal, $\chi(t)$. (e) The recovered filter, $\tilde{\mathbf{h}}_*$, and (d) the independent component using B-ICA and the criteria in Eq. (14). (f) The convolution of (d) and (e).

organized (see Eq. (4)). Finally, the convolution of the independent component in Fig. 8d with the recovered wavelet in Fig. 8e is plotted in Fig. 8f. Clearly the algorithm has done a reasonable job in recovering both the wavelet and the reflectivity.

5. SUMMARY

This paper adapted the ICA algorithm to include prior information on the mixing matrix. In particular, its banded nature was accounted for by specifying its nonzero elements. Additionally, the deconvolution problem was coaxed into an ICA formulation with a banded mixing matrix computed from the filter, $\mathbf{h}$. Initially, this resulted in an ICA model with only one realization of both the source and mixture vectors. To compensate for this lack of information, time delayed versions of $\rho(t)$ and $\chi(t)$ were considered. This resulted in a new blind deconvolution method, utilizing B-ICA, which in turn gave rise to a second complication. Namely that the first few mixtures and sources had few nonzero realizations, and so, had insufficient information to constrain their statistics. The solution was to use an approximate convolutional model which proved to be sufficient. B-ICA created a further complication. It produced as many candidate solutions as the dimension of $\mathbf{h}$. This problem was overcome by using the extra information provided by the convolutional model.

While the examples presented in this paper are noise free, real data is inherently noisy. The effect of noise can certainly be mitigated by linear and/or nonlinear processing. Incorporating noise directly into B-ICA is another appealing option [9] which deserves investigation.

6. REFERENCES

[1] R. Wiggins, "Minimum entropy deconvolution," *Geoexploration*, vol. 16, pp. 21–95, 1978.

[2] David Donoho, "On minimum entropy deconvolution," in *Applied Time Series Analysis II*, D. Findley, Ed. Academic Press Inc., 1981.

[3] M. Ooe and T. J. Ulrych, "Minimum entropy deconvolution with an exponential transformation," *Geophysical Prospecting*, vol. 27, pp. 458–473, 1979.

[4] Mauricio D. Sacchi, Danilo R. Velis, and Alberto H. Cominguez, "Minimum entropy deconvolution with frequency-domain constraints," *Geophysics*, vol. 59, no. 6, pp. 938–945, June 1994.

[5] A. T. Walden, "Non-Gaussian reflectivity, entropy, and deconvolution," *Geophysics*, vol. 50, no. 12, pp. 2862–2888, 1985.

[6] Ofir Shalvi and Ehud Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Transactions on Information Theory*, vol. 36, no. 2, March 1990.

[7] Kjetil F. Kaaresen and Tofinn Taxt, "Multichannel blind deconvolution of seismic signals," *Geophysics*, vol. 63, no. 6, pp. 2093–2107, November-December 1998.

[8] Jon Claerbout, *Earth Soundings Analysis: Processing versus Inversion*, Blackwell Scientific Publications, Inc., 1992.

[9] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja, *Independent Component Analysis*, Adaptive and Learning Systems for Signal Processing, Communications, and Control. John Wiley & Sons, Inc., 2001.

[10] M. C. Jones and Robin Sibson, "What is projection pursuit?," *Journal of the Royal Society. Series A (General)*, vol. 150, no. 1, pp. 1–37, 1987.

[11] Aapo Hyvärinen, "New approximation of differential entropy for independent component analysis and projection pursuit," in *Advances in Neural Information Procesing Systems 10*, Michael I. Jordan, Michael J. Kearns, and Sara A. Solla, Eds. 1998, pp. 273–279, MIT Press.

[12] Aapo Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *Neural Networks, IEEE Transactions on*, vol. 10, no. 3, pp. 626–634, May 1999.