Journal of Computational Mathematics
Vol.28, No.3, 2010, 371–385.

http://www.global-sci.org/jcm
doi:10.4208/jcm.2009.10-m2900

A TIME DOMAIN BLIND DECORRELATION METHOD OF CONVOLUTIVE MIXTURES BASED ON AN IIR MODEL*

Jie Liu
Department of Mathematics, National University of Singapore, Singapore
Email: matlj@nus.edu.sg
Jack Xin
Department of Mathematics, University of California Irvine, USA
Email: jxin@math.uci.edu
Yingyong Qi
Department of Mathematics, University of California Irvine, USA
and School of Information Engineering, Shangdong University at Weihai, China
Email: yqi@uci.edu

Abstract

We study a time domain decorrelation method of source signal separation from convolutive sound mixtures based on an infinite impulse response (IIR) model. The IIR model uses fewer parameters to capture the physical mixing process and is useful for finding low dimensional separating solutions. We present inversion formulas to decorrelate the mixture signals and derive filter equations involving second order time lagged statistics of mixtures. We then formulate an l_1 constrained minimization problem and solve it by an iterative method. Numerical experiments on recorded sound mixtures show that our method is capable of sound separation in low dimensional parameter spaces with good perceptual quality and low correlation coefficient comparable to the known infomax method.

Mathematics subject classification: 94A12, 65H10, 65C60.
Key words: Blind Decorrelation, Convolutive Mixtures, IIR Modeling, l_1 Constrained Minimization

1. Introduction

Blind source separation(BSS) methods aim to extract the original source signals from their mixtures based on the statistical independence of the source signals without knowledge of the mixing environment, see [3, 7, 10]. Realistic sound signals are often mixed through a media channel, so the received sound mixtures are linear convolutions of the unknown sources and the channel transmission functions. In other words, the observed signals are unknown weighted sums of the signals and their delays. The length of delays or convolution is physically on the order of thousands or more, and results in a complex high dimensional optimization problem. Separating convolutive mixtures is a challenging problem, especially in realistic settings [5, 6, 9, 11, 13].

Let us consider the mixing of two sources, with one source representing the foreground and

* Received September 6, 2008 / Revised version received March 10, 2009 / Accepted June 8, 2009 / Published online February 1, 2010 /

the other the background source with possibly diffuse spectra. A standard mixing model is:

$$y^1(t) = \sum_{k=1}^{L_m} a_k^{11} s^1(t+1-k) + \sum_{k=1}^{L_m} a_k^{12} s^2(t+1-k), \tag{1.1}$$

$$y^2(t) = \sum_{k=1}^{L_m} a_k^{21} s^1(t+1-k) + \sum_{k=1}^{L_m} a_k^{22} s^2(t+1-k), \tag{1.2}$$

where $s^i(t)$'s $(i = 1, 2)$ are the two source signals, $y^i(t)$'s are the received mixtures, and the mixing length L_m is large enough to approximate the physical mixing process [5,20]. The $s^i(t)$'s are zero if $t < 0$. If L_m is finite, the summation contains finitely many terms in (1.1)–(1.2), and the model is called a finite impulse response (FIR) model. The BSS problem is to recover the sources and filter coefficients a_k^{ij}'s from $y^i(t)$'s, assuming the statistical independence of the source signals $s^i(t)$'s at $t > 0$. Statistical independence approach is similar in spirit to feature based filtering and decomposition in image analysis [4, 17, 18], and is applicable to image separation as well [6]. Convolutive mixtures occur naturally for sounds.

As observed in [14, 20] and verified by direct calculation, the mixtures y^1 and y^2 can be orthogonalized (decorrelated) by an explicit transform. Define w^1 and w^2 as:

$$w^1(t) = \sum_{k=1}^{L_m} a_k^{22} y^1(t+1-k) - \sum_{k=1}^{L_m} a_k^{12} y^2(t+1-k), \tag{1.3}$$

$$w^2(t) = \sum_{k=1}^{L_m} -a_k^{21} y^1(t+1-k) + \sum_{k=1}^{L_m} a_k^{11} y^2(t+1-k), \tag{1.4}$$

then $w^1(t)$ and $w^2(t)$ are independent of each other and contain only the information of independent sources s^1 and s^2 respectively. The proof of the independence of w^1 and w^2 will be a special case of what we will present in the next section (setting $B^1(z) = 1 = B^2(z)$ in (2.4)–(2.5)). The proof also implies that if s^i's are uncorrelated ($E[s^i(t)\, s^j(t-n)] = 0$, $i \neq j$, for any n), then the w^i's in (1.3)–(1.4) are uncorrelated as well.

In [14], a system of algebraic equations of a_k^{ij}'s follows from the uncorrelation of w^i's, and an optimization problem is formulated to compute a_k^{ij}'s. However, the objective function is quartically nonlinear and the support of the a_k^{ij} in k may be very large, rendering computation expensive. To actually approach the physical impulse response, L_m can be as large as $O(10^3)$ or more. Let us denote this physical limit by L_p. On the other hand, numerical experiments [14] indicate that there are lower dimensional solutions $\{a_k^{ij}, k = 1, 2, ..., L_m\}$, $L_m \ll L_p$, that suffice for a rather good separation. For example, in computing separation for three room recordings, $L_m = 50$ is found to be effective [14]. Low dimensional separating solutions of similar dimensions are also reported in [9] for an infomax method.

In [14], l_1 norm is employed as a constraint to select solutions with sparse structures as a way towards finding stable and low dimensional solutions. The sparsity from minimizing l_1 norm has been extensively studied recently in the context of compressive sensing and basis pursuits (see [2, 8, 19, 21] and references therein). Use of l_1 norm as a constraint is due to the scale invariance of BSS problem and the need to minimize correlation (or independence). Resulting sparseness appears new. In this paper, we study low dimensional BSS solutions by recasting

(1.1)–(1.2) into the form:

$$\sum_{k=1}^{q} b_k^1\, y^1(t+1-k) = \sum_{k=1}^{p} a_k^{11} s^1(t+1-k) + \sum_{k=1}^{p} a_k^{12} s^2(t+1-k), \tag{1.5}$$

$$\sum_{k=1}^{q} b_k^2\, y^2(t+1-k) = \sum_{k=1}^{p} a_k^{21} s^1(t+1-k) + \sum_{k=1}^{p} a_k^{22} s^2(t+1-k), \tag{1.6}$$

where b^i $(i = 1, 2)$ model the resonance frequencies of the mixing environment, or the poles of a rational filter. One can formally recover (1.1)–(1.2) with $L_m = \infty$ by solving (1.5)–(1.6) for $y^i(t)$ $(i = 1, 2)$ and expanding solutions into infinite series. Hence (1.5)-(1.6) is an infinite impulse response (IIR) model of mixing. The potential advantage is that IIR model is more compact in approximating the physical mixing process and may be effective at a smaller value of $q+p$ than a typical length of an FIR filter, thereby offering a natural low dimensional setting for computation. A study of separating complex synthetic mixture of independent identically distributed quadrature-amplitude-modulated signals using IIR model and infomax method is in [1] where $q = 3$ and $p = 2$.

The rest of the paper is organized as follows. In section 2, inversion formulas of IIR models are shown to map mixture signals to independent (decorrelated) outputs. In section 3, demixing filter equations are derived from the inversion formulas. In section 4, an l_1 constrained minimization problem is formulated for solving the filter equations along with a Levenberg-Marquardt iterative method for computing a minimizing sequence. In section 5, experimental results are shown for separating synthetic and recorded mixtures in IIR model with low dimensional parameter spaces. Concluding remarks will be presented in section 6.

2. Inversion and Decorrelation

Let us first derive the IIR model (1.5)-(1.6) from the infinite dimensional ($L_m = \infty$) version of (1.1)-(1.2). Denoting $a^{ij} = (a_1^{ij}, a_2^{ij}, ...)^\top$, $y^i = (y^i(1), y^i(2), ...)^\top$ and $s^i = (s^i(1), s^i(2), ...)^\top$, we write (1.1)–(1.2) as

$$\begin{pmatrix} y^1 \\ y^2 \end{pmatrix} = \begin{pmatrix} a^{11} * s^1 + a^{12} * s^2 \\ a^{21} * s^1 + a^{22} * s^2 \end{pmatrix} \tag{2.1}$$

$$:= \begin{pmatrix} a^{11}*, & a^{12}* \\ a^{21}*, & a^{22}* \end{pmatrix} \begin{pmatrix} s^1 \\ s^2 \end{pmatrix}, \tag{2.2}$$

where $*$ is linear convolution. For example,

$$(a^{11} * s^1)(t) = \sum_{k=1}^{t} a_k^{11} s^1(t+1-k).$$

To facilitate our later discussion, we adopt the matrix notation in (2.2) which is equivalent to (2.1).

Let the formal $z-$transform be:

$$A^{ij}(z) = \sum_{k=1}^{\infty} a_k^{ij}\, z^{1-k} \tag{2.3}$$

with i, j equal to $1, 2$. Now assume that $A^{ij}(z)$ are modeled by rational functions of z:

$$A^{11}(z) = \frac{C^{22}(z)}{B^1(z)}, \quad A^{12}(z) = \frac{-C^{12}(z)}{B^1(z)}, \tag{2.4}$$

$$A^{21}(z) = \frac{-C^{21}(z)}{B^2(z)}, \quad A^{22}(z) = \frac{C^{11}(z)}{B^2(z)}, \tag{2.5}$$

where

$$C^{ij}(z) = \sum_{k=1}^{q} c_k^{ij} z^{1-k}, \quad \text{and} \quad B^i(z) = \sum_{k=1}^{p} b_k^i z^{1-k}. \tag{2.6}$$

An infinite sequence $\{a_k^{ij}, k = 1, 2, ..., \}$ can be represented by rational functions with finite q and p. This allows the description of the infinite *long* $\{a_k^{ij}, k = 1, 2, ..., \infty\}$ filters by the *short* $\{c_k^{ij}, k = 1, ..., q\}$ and $\{b_k^i, k = 1, ..., p\}$ filters. Taking z-transform on both sides of (2.1), (2.1) or (2.2) becomes

$$\begin{pmatrix} Y_1(z) \\ Y_2(z) \end{pmatrix} = \begin{pmatrix} \frac{C^{22}(z)}{B^1(z)}, & \frac{-C^{12}(z)}{B^1(z)} \\ \frac{-C^{21}(z)}{B^2(z)}, & \frac{C^{11}(z)}{B^2(z)} \end{pmatrix} \begin{pmatrix} S_1(z) \\ S_2(z) \end{pmatrix}$$

Left multiplying the matrix $\operatorname{diag}(B^1(z), B^2(z))$ on the above equation and going back to the time domain, we see that (2.2) is put in the form

$$\begin{pmatrix} b^1*, & 0 \\ 0, & b^2* \end{pmatrix} \begin{pmatrix} y^1 \\ y^2 \end{pmatrix} = \begin{pmatrix} c^{22}*, & -c^{12}* \\ -c^{21}*, & c^{11}* \end{pmatrix} \begin{pmatrix} s^1 \\ s^2 \end{pmatrix} \tag{2.7}$$

where $b^i = (b_1^i, ..., b_p^i)^\top$ and $c^{ij} = (c_1^{ij}, ..., c_q^{ij})^\top$ are vectors of *finite* dimensions. We have derived the equivalent and more compact IIR model from the infinite dimensional version of the original mixing model (1.1)-(1.2).

Now we show that the orthogonalization of y^1 and y^2 holds for the IIR model. Define

$$\begin{pmatrix} v^1 \\ v^2 \end{pmatrix} = \begin{pmatrix} c^{11}*, & c^{12}* \\ c^{21}*, & c^{22}* \end{pmatrix} \begin{pmatrix} b^1*, & 0 \\ 0, & b^2* \end{pmatrix} \begin{pmatrix} y^1 \\ y^2 \end{pmatrix} \tag{2.8}$$

$$= \begin{pmatrix} c^{11} * b^1*, & c^{12} * b^2* \\ c^{21} * b^1*, & c^{22} * b^2* \end{pmatrix} \begin{pmatrix} y^1 \\ y^2 \end{pmatrix} \tag{2.9}$$

$$:= \begin{pmatrix} d^{11}*, & d^{12}* \\ d^{21}*, & d^{22}* \end{pmatrix} \begin{pmatrix} y^1 \\ y^2 \end{pmatrix}. \tag{2.10}$$

In the last step, we have introduced

$$d^{ij} = c^{ij} * b^j, \quad \text{namely} \quad d_k^{ij} = \sum_m c_m^{ij} b_{k+1-m}^j. \tag{2.11}$$

where we have zero padded c_k^{ij} and b_k^i when necessary. Now, plugging (2.7) into (2.8), and using the fact that $c^{ij} * c^{mn} = c^{mn} * c^{ij}$, we obtain

$$\begin{aligned} \begin{pmatrix} v^1 \\ v^2 \end{pmatrix} &= \begin{pmatrix} c^{11}*, & c^{12}* \\ c^{21}*, & c^{22}* \end{pmatrix} \begin{pmatrix} c^{22}*, & -c^{12}* \\ -c^{21}*, & c^{11}* \end{pmatrix} \begin{pmatrix} s^1 \\ s^2 \end{pmatrix} \\ &= \begin{pmatrix} c^{11} * c^{22} * -c^{12} * c^{21}*, & 0 \\ 0, & c^{11} * c^{22} * -c^{12} * c^{21}* \end{pmatrix} \begin{pmatrix} s^1 \\ s^2 \end{pmatrix}. \end{aligned} \tag{2.12}$$

It follows that if s^1 and s^2 are independent (or uncorrelated), so are v^1 and v^2. We summarize the above results in:

Theorem 2.1 (independence/uncorrelation) *Given the received signals y^i's in* (1.1)–(1.2), *if source signals s^i's are independent (or uncorrelated), then:* (1) *the v^i's in* (2.10) *are independent (or uncorrelated);* (2) *the w^i's in* (1.3)-(1.4) *are independent (or uncorrelated).*

Part (2) is a special case of part (1) with $B^i(z) = 1$ in (2.4)–(2.5) and the same steps leading to (2.10), where $d^{ii} = a^{ii}$ and $d^{ij} = -a^{ij}$ with $i, j = 1, 2$ and $i \neq j$.

Corollary 1.1 (dimensional reduction) *Consider* (1.1)–(1.2) *where the physical support of $a^{ij} = \{a^{ij}_k, k = 1, 2, ...\}$, $i, j = 1, 2$, may be either large (thousands) or infinite in k. Suppose that the z-transforms of a^{ij}'s* (2.3) *are in the form of rational functions* (2.4)–(2.6). *Then for source separation, it suffices to find $d^{ij} = c^{ij} * b^j$ in* (2.10)–(2.11), *whose support may be much shorter than that of the a^{ij}'s.*

3. Filter Equations

Let us derive equations for the filter coefficients $\{c^{ij}, b^m; i, j, m = 1, 2\}$. By Theorem 2.1, we have

$$E(v^1(t)v^2(t-n)) = 0, \qquad \forall\, n. \tag{3.1}$$

Now, we write (2.10) more explicitly as

$$v^1(t) = \sum_{k=1}^{p+q-1} d^{11}_k y^1(t+1-k) + \sum_{k=1}^{p+q-1} d^{12}_k y^2(t+1-k) \tag{3.2}$$

$$v^2(t) = \sum_{k=1}^{p+q-1} d^{21}_k y^1(t+1-k) + \sum_{k=1}^{p+q-1} d^{22}_k y^2(t+1-k) \tag{3.3}$$

where p is the length of the denominator filter coefficients $\{b^i_k, k = 1, ..., p\}$ and q is the length of the numerator filter coefficients $\{c^{ij}_k, k = 1, ..., q\}$ in (2.4)–(2.5). Substituting (3.2) and (3.3) into (3.1), we have

$$\sum_{k,m=1}^{p+q-1} d^{11}_k d^{21}_m E\Big(y^1(t{+}1{-}k)y^1(t{+}1{-}m{-}n)\Big) + d^{12}_k d^{21}_m E\Big(y^2(t{+}1{-}k)y^1(t{+}1{-}m{-}n)\Big)$$
$$+ d^{11}_k d^{22}_m E\Big(y^1(t{+}1{-}k)y^2(t{+}1{-}m{-}n)\Big) + d^{12}_k d^{22}_m E\Big(y^2(t{+}1{-}k)y^2(t{+}1{-}m{-}n)\Big) = 0,$$

which can be written as

$$\sum_{k,m=1}^{p+q-1} d^{11}_k d^{21}_m L^{11}_{k,m,n} + d^{12}_k d^{21}_m L^{21}_{k,m,n} + d^{11}_k d^{22}_m L^{12}_{k,m,n} + d^{12}_k d^{22}_m L^{22}_{k,m,n} = 0 \tag{3.4}$$

for any n, where

$$L^{ij}_{k,m,n} = E\Big(y^i(t+1-k)y^j(t+1-m-n)\Big). \tag{3.5}$$

If we introduce the $(p+q-1) \times (p+q-1)$ matrices $L^{ij}_n = (L^{ij}_{k,m,n})$, (3.4) can be rewritten as

$$\left(d^{11}; d^{12}\right) \begin{pmatrix} L^{11}_n, & L^{12}_n \\ L^{21}_n, & L^{22}_n \end{pmatrix} \begin{pmatrix} d^{21} \\ d^{22} \end{pmatrix} = 0 \tag{3.6}$$

for any n, where we assume d^{ij} are column vectors and $(\cdot;\cdot)$ means the concatenation of two column vectors. We let n vary from $-N$ to N with N large enough so that there are enough equations to solve for $\{c_k^{ij}\}$ and $\{b_k^m\}$. Note that (3.6) indicates that

$$(d^{11}; d^{12}) = (c^{11} * b^1; c^{12} * b^2) \quad \text{and} \quad (d^{21}; d^{22}) = (c^{21} * b^1; c^{22} * b^2)$$

are determined up to a constant. This fact can also be directly read off from (2.10). In our computation, we fix this ambiguity by normalizing

$$b_1^1 = 1 = b_1^2 \tag{3.7}$$

and also requiring

$$\|(c^{11}; c^{12})\|_1 = 1 \quad \text{and} \quad \|(c^{21}; c^{22})\|_1 = 1 \tag{3.8}$$

where $\|\cdot\|_1$ is the ℓ_1 norm of a vector. Equations (3.4) are in fact equations for $\{c_k^{ij}\}$ and $\{b_k^m\}$ in view of (2.11). We may solve for $\{c_k^{ij}\}$ and $\{b_k^m\}$'s from (3.4) under the constraints (3.7) and (3.8). The $\{a_k^{ij}\}$ in (1.1)–(1.2) are recovered from (2.4) and (2.5).

To summarize, we shall find $\{c_k^{ij}\}$ and $\{b_k^i\}$ from (3.6) with n varying from $-N$ to N, where $\{d_k^{ij}\}$ are expressed in terms of $\{c_k^{ij}\}$ and $\{b_k^m\}$ in (2.11). Moreover, we require $\{c_k^{ij}\}$ and $\{b_k^i\}$ to satisfy the constraint (3.7)–(3.8).

Once we have $\{c_k^{ij}\}$ and $\{b_k^m\}$, we may calculate w^1 and w^2 from (1.3)–(1.4) with $\{a_k^{ij}\}$ obtained from (2.4)–(2.5). Or we may calculate v^1 and v^2 from (2.9). Either way, we obtain signals that contain information of only one of the sources. For hearing purpose, w^1 and w^2 or v^1 and v^2 are good enough since they are a convolution of s^1 or s^2 alone. If s^1 and s^2 are desired, they can be recovered with a FFT-based deconvolution from w^1, w^2 or v^1, v^2.

4. Algorithm

Instead of directly solving (3.6) subject to constraint (3.7)–(3.8), we minimize a relaxed unconstrained cost function

$$\begin{aligned}&\{c^{ij}, \tilde{b}^m; i,j,m = 1,2\}\\ =&\text{argmin} \sum_{n=-N}^{N} |f_n|^2 + \sigma^2(\|(c^{11}; c^{12})\|_1^2 - 1)^2 + \sigma^2(\|(c^{21}; c^{22})\|_1^2 - 1)^2\end{aligned} \tag{4.1}$$

where

$$b^m = (1; \tilde{b}^m),\ f_n = \left(d^{11}; d^{12}\right) \begin{pmatrix} L_n^{11}, & L_n^{12} \\ L_n^{21}, & L_n^{22} \end{pmatrix} \begin{pmatrix} d^{21} \\ d^{22} \end{pmatrix},$$

d^{ij} and c^{ij}, b^i are related by (2.11). The constraints are handled by the last two penalty terms, and l_1-norm is chosen to improve sparsity of solutions. The advantage of l_1 norm over l_2 norm as relaxed constaints is studied in more detail in [14].

The objective function in (4.1) is expressed as a sum of the squares so that the Levenberg-Marquardt (LM) method applies to search for a minimizer. The LM method minimizes a function of the form:

$$g(x) = \frac{1}{2} \sum_{i=1}^{m} (f_i(x))^2,$$

or a sum of squares of functions of $x = (x_1, ..., x_n)$. If we use steepest descent, the searching direction is $-\nabla g = -J^\top f$ where $f = (f_1, ..., f_m)^\top$, and $J = (\partial f_i/\partial x_j)$ is the Jacobian. If we apply the standard Gauss-Newton method, then

$$x^{k+1} = x^k - (J^\top J)^{-1} J^\top f.$$

When $m = n$, J is a square matrix, this is the Newton method for solving $f(x) = 0$. The LM method is an interpolation between steepest descent and Gauss-Newton, namely

$$x^{k+1} = x^k - (J^\top J + \mu_k I)^{-1} J^\top f \tag{4.2}$$

with $\mu_k > 0$. When $\mu_k = 0$, it becomes Gauss-Newton; and when μ_k is large, it moves a small step along a direction very close to the steepest descent direction. The matrix being inverted is always nonsingular, even when $J^\top J$ is singular. The LM method has a strategy to choose μ_k to guarantee the reduction of g. For more exposition of the LM method, see [15,16] among others. The web site `http://www.ics.forth.gr/~lourakis/levmar/` provides a public C/C++ code by M. Lourakis, and a link to the Matlab code by H. B. Nielsen.

5. Computational Results

The computations reported here are for 2 sources and 2 receivers. The data are either synthetic convolutive mixtures or recorded mixtures in an office size room or conference room, as listed in Table 5.1. They will be called case (1) and case (2)-1,(2)-2,(2)-3 in the following discussion. All the three real room recorded data in case (2) are public data of Salk Institute [12].

Figure 5.2 and Table 5.2 plot the pairs of mixtures of case (1) and (2) respectively. The associated computation results are listed in Figure 5.3, Figure 5.4, Table 5.3 and Table 5.4.

The L_n's in the objective function are computed from (3.5) with $n = -N, ..., N$, while k and m range from 1 to $p + q - 1$. For case (1), we take $p = 9$, $q = 2$ and $N = 30$. For the three examples in case (2), we take $p = 20$, $q = 31$ and $N = 120$. There are a total of $2N + 1$ equations in (3.4) or (3.6) with $4q + 2(p - 1)$ unknowns.

A segment of $y_i(t)$ of length L is used to approximate the expectation. To reduce the statistical error, we have used all the available data stream to estimate the $L^{ij}_{k,m,n}$ in (3.5) in all the above cases. The maximum of $L^{ij}_{k,m,n}$ from data are typically on the order of $O(10^{-2})$. To avoid loss of significant digits, we multiply $L^{ij}_{k,m,n}$ by 100 so that the adjusted maximum of $L^{ij}_{k,m,n}$ is order 1. The value of σ in (4.1) is fixed at 0.002 for all the cases. If σ is too small, the constraint is too weak to be effective. If σ is too large, the minimizing sequence evolves too slowly. Once we found the above σ after a few tests, it works for all different mixtures. In principle, σ depends on the *size* of the optimization problem.

In all the cases, we take the initial value to be $(1, 0, ..., 0)$ for c^{ii} and b^i, $i = 1, 2$, and $(0, 0, ..., 0)$ for c^{ij} when $i \neq j$. The b^i_1 is fixed to be 1. When the LM method is applied to the objective function (4.1) that contains l_1 norm, the derivative terms such as

$$\nabla_{c_j} \|c\|_{l_1} = \nabla_{c_j} |c_j|$$

are numerically approximated by

$$\nabla_{c_j} |c_j| = \frac{c_j}{\epsilon + |c_j|}$$

with $\epsilon = 10^{-16}$.

The LM method is implemented in Matlab [15, 16], with the stopping parameters for iterations set to be opts=[10^{-3}, 10^{-7}, 10^{-12}, 1000, 10^{-15}]. These numbers have the following meaning: 10^{-3} and 10^{-15} are related to the initial value and lower bound of μ_k in LM method (see (4.2)). LM iteration will stop when the l_∞ norm of the gradient is less than 10^{-7}, or when

$$\|\{c^{ij}, b^m\}^{k+1} - \{c^{ij}, b^m\}^k\|_{l_2} \leq 10^{-12}\left(10^{-12} + \|\{c^{ij}, b^m\}^k\|_{l_2}\right),$$

or when the number of iterations exceeds 1000. The superscript k refers to the k-th iterate. In all cases above, LM method reduces the gradient of the objective function below the preset value 10^{-7}.

The computation time depends on the following factors: (I) how much data are used to estimate the $L^{ij}_{k,m,n}$ in (3.5); (II) size of p, q and N; (III) stopping criteria for LM or the number of iterations in LM. The LM iterations all converge with the stopping criteria described above. The CPU time for the three examples in case (2) are 40, 44, 102 seconds respectively. The computation is done with Matlab on a Compaq laptop with 1.6G Hz AMD 64 bit dual core CPU (in single thread mode).

Results of case (1) show that using the IIR model, we can greatly reduce the number of unknowns we need to solve for. Figure 5.1 shows the impulse response a^{ij}'s ((1.1)–(1.2)) that are used to generate the synthetic mixtures. The support of each a^{ij} is about 100. If we follow [14] and use (1.3)–(1.4) as the demixing equations, the number of unknowns we need to solve for is around 100×4=400. However, using the IIR model (2.4)–(2.5), we can reduce

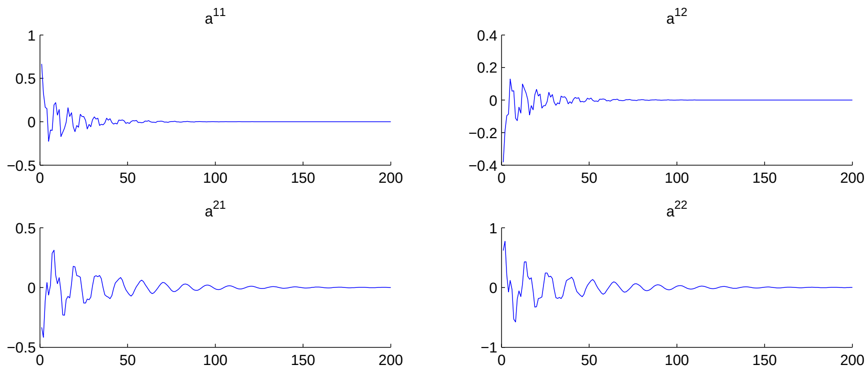


Fig. 5.1. Synthetic impulse response for case (1) which is the a^{ij}_k in (1.1)–(1.2).

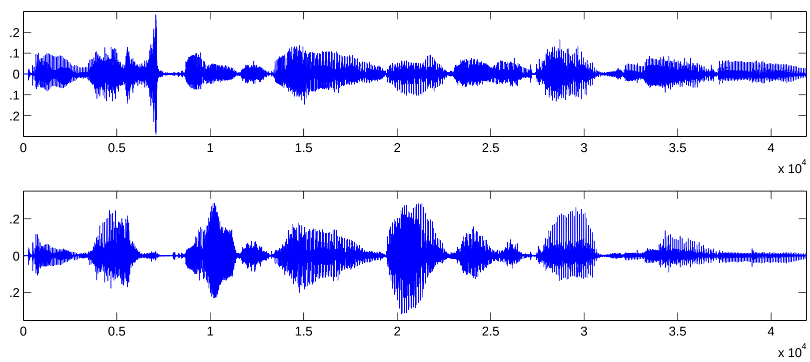


Fig. 5.2. Case (1): synthetic convolutive mixtures.

the total number of unknowns that we need to solve for to about 10×4=40. In this synthetic example, the support of c^{ij} in (2.4)–(2.5) is 1 and

$$(c^{11}, c^{12}, c^{21}, c^{22}) = (0.617, 0.383, 0.333, 0.667).$$

The exact b_k^i's are plotted in Figure 5.4 together with their approximations that we have found numerically. It is clear that we can recover the b_k^i's rather accurately. So are the c^{ij}'s. From the two plots in Figure 5.3, we can clearly see the separation of the two sources.

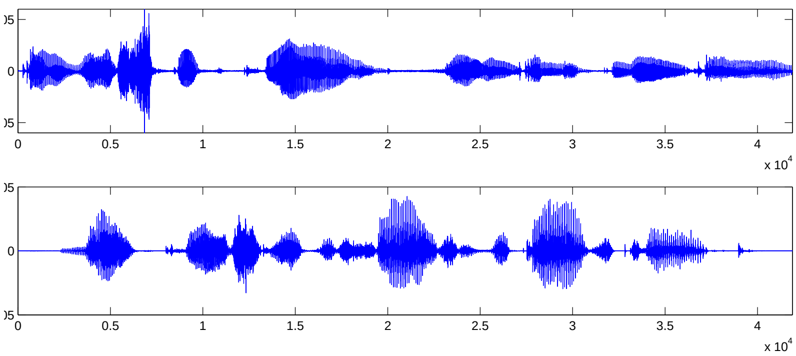


Fig. 5.3. Plots of the v^i $(i = 1, 2)$ for case (1).

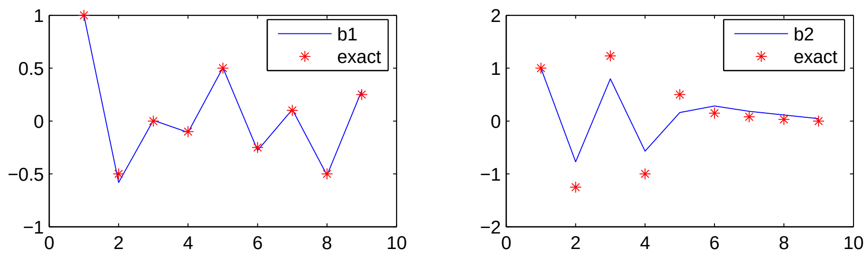


Fig. 5.4. Numerical results for case (1): b_k^i in (2.4)–(2.5).

For the three examples of realistic recorded mixtures in case (2), our separation results and those posted on [12] are perceptually quite close. For case (2)-1 and case (2)-2, the separation effect is clear. For case (2)-3, distinct improvement can be heard.

Ideally, we hope that the perceptually optimal $\{c^{ij}, b^m\}$ is a global minimizer of the objective function (call it F) in (4.1) so that we can "locate" it by minimizing F. However, numerical experiments indicated that the landscape of F is complicated, and F may have multiple local minimizers. A minimizer may not always sound better than a near minimizer. Partly this is because given any two clean sources of finite length, we do not expect their cross correlations with *different* time lags to be *all* zero as in (3.1). Hence the desired (perceptually optimal) $\{c^{ij}, b^m\}$ will not render the objective function in (4.1) equal to zero, even though they make it small. Algebraically, it is possible that there are other choices of $\{c^{ij}, b^m\}$ making F even smaller. If this happens, the LM solver searching for a minimizer may not find the optimal $\{c^{ij}, b^m\}$, though it always helps to separate the mixtures.

As a quantitative measure of separation (decorrelation), we compute the maximal correlation

Table 5.1: Description of room recorded data from [12], at sampling rate 16 kHz.

case #	description
(2)-1	A speaker has been recorded with two distance talking microphones in a normal office room with loud music in the background. The distance between the speaker, cassette player and the microphones is about 60cm in a square ordering.
(2)-2	Two Speakers have been recorded speaking simultaneously. Speaker 1 says the digits from one to ten in English and speaker 2 counts at the same time the digits in Spanish. The recording has been done in a normal office room. The distance between the speakers and the microphones is about 60cm in a square ordering.
(2)-3	Two Speakers have been recorded speaking simultaneously. This time the recording was in a conference room (5.5m by 8m). The conference room had some air-conditioning noise. Both speakers are reading a section from the newspaper for 16sec. The mics were placed 120 cm away from the speakers.

Table 5.2: Plot of the mixture signals for cases (2).

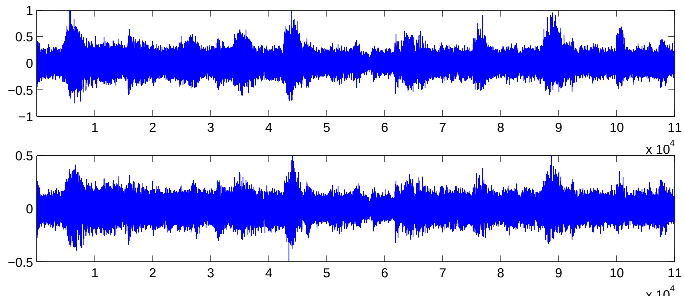
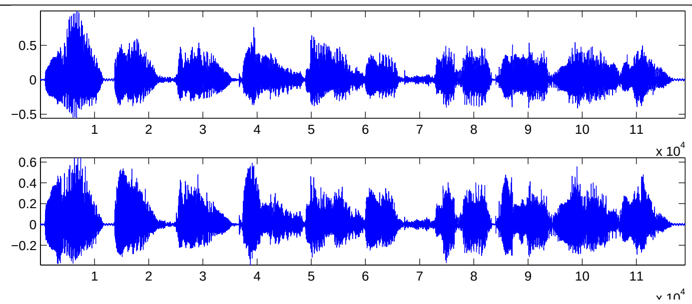
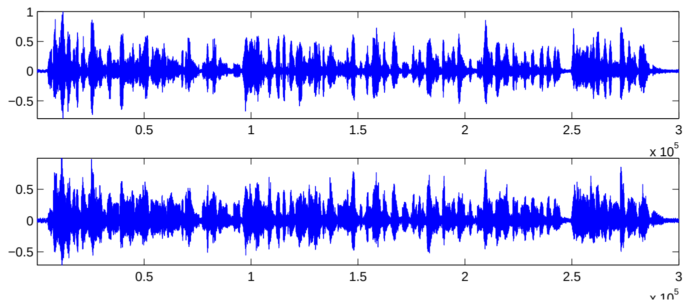
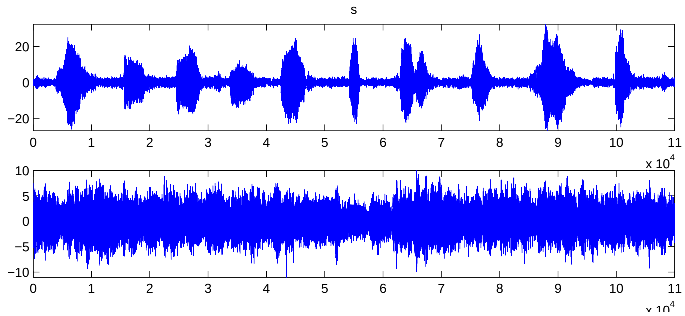
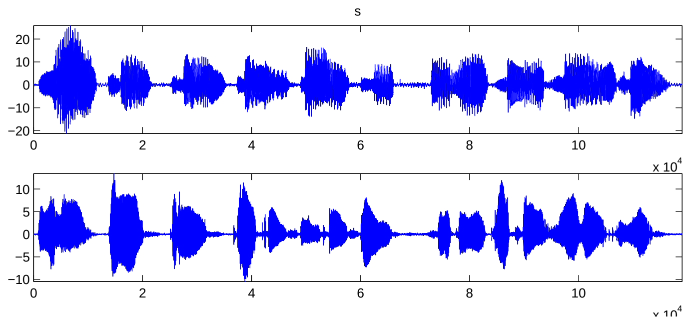
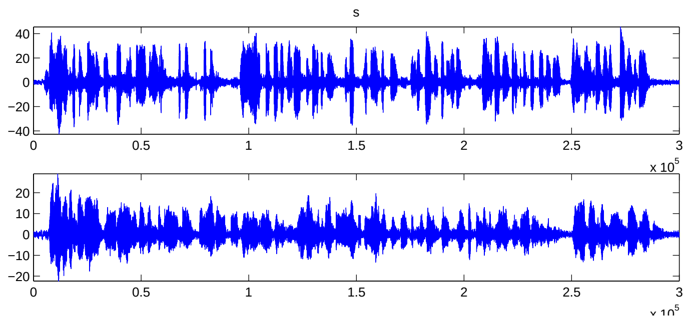
case #	Mixture
(2)-1	
(2)-2	
(2)-3	

Table 5.3: Plot of the computed output signals for cases (2).

case #	calculated s^i using (2.12) once v^i is obtained from (2.9)
(2)-1	
(2)-2	
(2)-3	

coefficient over multiple time lags:

$$\bar{\rho}(a,b) = \max_{k \in \{-K,\ldots,K\}} |\rho(a(t), b(t+k))| \tag{5.1}$$

where ρ is the correlation coefficient defined by

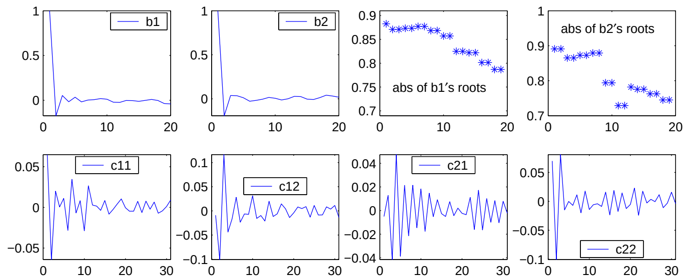
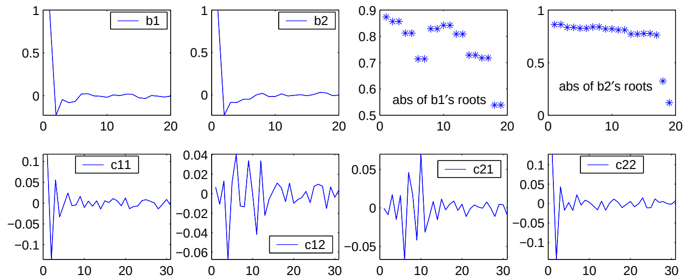
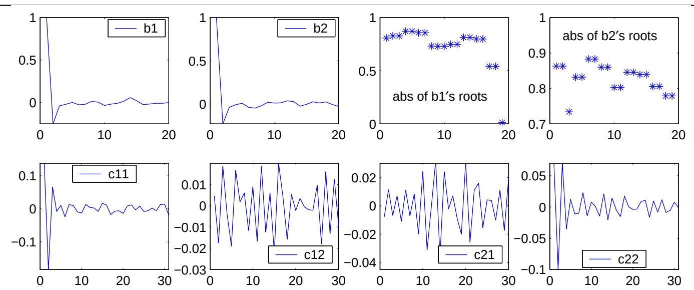
$$\rho(a(t), b(t)) = \frac{\operatorname{cov}(a(t), b(t))}{\sqrt{\operatorname{cov}(a(t), a(t))\operatorname{cov}(b(t), b(t))}} \tag{5.2}$$

with

$$\operatorname{cov}(a(t), b(t)) = L^{-1}\sum_{t=1}^{L} a(t)b(t) - L^{-2}\sum_{t=1}^{L} a(t)\sum_{t=1}^{L} b(t)$$

being the estimation of covariance of a and b when there are L samples. The values of L for computing ρ equals to the size of the total available data stream, which can be read from the plots in Table 5.2. The value K in (5.1) is 20 in our calculations. The $\bar{\rho}$ is computed for the

Table 5.4: Plots of the computed b^i and c^{ij} for cases (2), and absolute values of the roots of $B^i(z) = 0$, $i = 1, 2$.

case #	b^i and c^{ij} obtained from solving (4.1)
(2)-1	
(2)-2	
(2)-3	

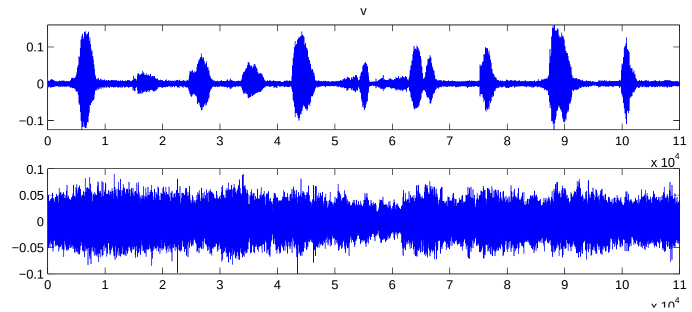
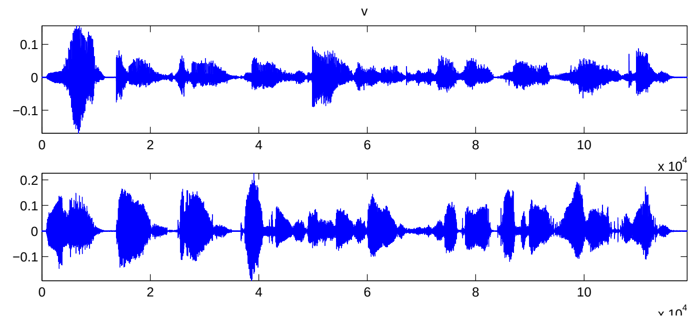
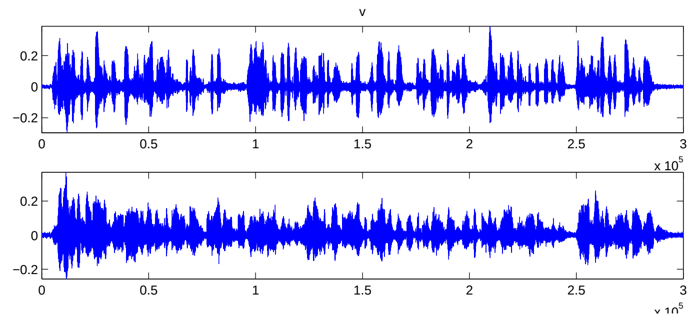
mixtures, and the separated v and s. For comparison, we also computed $\bar{\rho}$ from the separated signals of [12]. The results are listed in Table 5.5 which shows that the $\bar{\rho}$ values of the mixtures are much larger than those of the separated signals. Table 5.5 also implies that our method is comparable to [12]. Perceptually, we find so as well. FIR mixing model is used in [12] however, and the number of unknown filter coefficients can be as large as thousands.

Finally, much less data may still yield quite good separation. In Table 5.6, thirty LM iterations and 12800 samples (or 0.8 second of data stream at 16 kHz sampling frequency) are

Table 5.5: Comparison of the correlation coefficients $\bar{\rho}(\cdot,\cdot)$. The s^1_{Lee} and s^2_{Lee} are computed from the results of [12].

	$\bar{\rho}(y^1, y^2)$	$\bar{\rho}(v^1, v^2)$	$\bar{\rho}(s^1, s^2)$	$\bar{\rho}(s^1_{Lee}, s^2_{Lee})$
case (2)-1	8.31e-1	1.24e-2	1.39e-2	3.51e-2
case (2)-2	7.75e-1	7.87e-3	2.06e-2	1.12e-2
case (2)-3	5.46e-1	4.25e-3	8.78e-3	7.93e-3

Table 5.6: Plots of the v^i $(i = 1, 2)$ for cases (2) after 30 LM iterations. A piece of 0.8 sec of the data stream (starting from 1.0 second after the initial time to skip the initial silence) is used to estimate the demixing filter coefficients, with $p = 20$, $q = 31$, $N = 40$.

case #	calculated v^i using (3.2)–(3.3).
(2)-1	
(2)-2	
(2)-3	

sufficient for a quality separation of the recorded mixtures. To avoid the initial silent period in the recorded data, we use the data stream from 1.0 sec to 1.8 sec. Moreover, we also take $p = 20$, $q = 31$ and $N = 40$, so the number of equations $(2N + 1 = 81)$ is less than the number of unknowns $(4q + 2(p - 1) = 162)$. The LM method (4.2) can handle such a degenerate case because $\mu_k > 0$ and the matrix to be inverted is nonsingular. To further save the computation time, we directly use $v^i(t)$ obtained from (2.9) as the separation results, which is perceptually good enough. We do not go further to find s_i from each v_i. The results in Table 5.6 are plots of those v_i's, which are perceptually reasonable. The total computation time for Table 5.6 is 4.5, 5.3, and 4.9 seconds for case (2)-1, case (2)-2 and case (2)-3 respectively. For case (2)-1, convergence occurs under 30 iterations.

6. Conclusions

We developed a time domain decorrelation algorithm of convolutive mixtures based on an infinite impulse response model. We formulated an l_1 constrained minimization of cross correlations with time lags for solving the demixing filter equations. The method produces estimates of low dimensional demixing/mixing filter coefficients that are effective for separating room recorded mixtures of music and speech. Future work will investigate related stochastic learning algorithms to achieve fast and low complexity approximations of the filter equations with less demand on data input.

Acknowledgments. The work was partially supported by NSF grants DMS-0712881, NIH grant 2R44DC006734; the CORCLR (Academic Senate Council on Research, Computing and Library Resources) faculty research grant MI-2006-07-6, and a Pilot award of the Center for Hearing Research at UC Irvine.

References

[1] S. Amari, S. Douglas, A. Cichocki and H. Yang, Multichannel blind deconvolution and equalization using the natural gradient, *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications*, Paris, France, 1997, 101-104.

[2] E. Candès, J. Romberg and T. Tao, Stable signal recovery from incomplete and inaccurate measurements, *Commun. Pur. Appl. Math*, **59**:8 (2006), 1207-1223.

[3] J. Cardoso, Blind signal separation: Statistical principles, *Proc. IEEE*, **9**:10 (2008), 2009-2025.

[4] T. Chan and J. Shen, Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods, SIAM, Philadelphia, 2005.

[5] S. Choi, A. Cichocki, H. Park and S. Lee, Blind Source Separation and Independent Component Analysis: A Review, *Neural Information Processing -Letters and Reviews*, **6**:1 (2005), 1-57.

[6] A. Cichocki and S. Amari, Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications, John Wiley and Sons, 2005.

[7] P. Comon, Independent component analysis: A new concept, *Signal Process.*, **36** (1994), 287-314.

[8] D. Donoho, For most large undertermined systems of equations, the minimal l_1-norm solution is also the sparsest solution, *Comm. Pur. Appl. Math*, **59**:6 (2006), 797-829.

[9] S. Douglas, H. Sawada and S. Makino, Natural Gradient Multichannel Blind Deconvolution and Speech Separation Using Causal FIR Filters, *IEEE T. Speech Audi. P.*, **13**:1 (2005), 92-104.

[10] C. Jutten and J. Herault, Blind separation of sources, Part I: an adaptive algorithm based on neuromimetic architecture, *Signal Process.*, **24** (1991), 1-10.

[11] T-W Lee, Independent Component Analysis: Theory and Applications, Kluwer Academic Publishers, 1998.

[12] T.-W. Lee, *Blind Source Separation: Audio Examples*, `http://inc2.ucsd.edu/~tewon` or `http://www.cnl.salk.edu/~tewon/Blind/blind_audio.html`.

[13] J. Liu, J. Xin and Y. Qi, A dynamic algorithm for blind separation of convolutive sound mixtures, *Neurocomputing*, **72** (2008), 521-532.

[14] J. Liu, J. Xin, Y. Qi and F.-G. Zeng, A time domain algorithm for blind separation of convolutive sound mixtures and l_1 constrained minimization of cross correlations with time lags, *Comm. Math Sci*, **7**:1 (2009), 109-128.

[15] M. Lourakis, *A brief description of the Levenberg-Marquardt algorithm implemented by* `levmar`, `http://www.ics.forth.gr/~lourakis/levmar/levmar.pdf`

[16] K. Madsen, H. B. Nielsen and O. Tingleff, “Methods for non-linear least squares problems”, 2nd edition, `http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/3215/pdf/imm3215.pdf`

[17] Y. Meyer, Oscillating Patterns in Image Processing and Nonlinear Evolution Equations, AMS University Lecture Series, Vol. 22, 2001, Providence, Rhode Island.

[18] S. Osher and L. Rudin, Feature-Oriented Image Enhancement Using Shock Filters, *SIAM J. Numer. Anal.*, **27**:4 (1990), 919-940.

[19] F. Santosa and W. Symes, Linear inversion of band-limited reflection seismograms, *SIAM J. Sci. Comput.*, **7** (1986), 1307-1330.

[20] E. Weinstein, M. Feder and A. V. Oppenheim, Multi-channel signal separation by decorrelation, *IEEE T. Speech Audi. P.*, **1**:4 (1993), 405-413.

[21] W. Yin, S. Osher, D. Goldfarb and J. Darbon, Bregman Iterative Algorithms for l_1-Minimization with Applications for Compressed Sensing, *SIAM Journal on Imaging Sciences*, **1**:1 (2008), 143-168.