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