SPARSITY-BASED SINGLE-CHANNEL BLIND SEPARATION OF SUPERIMPOSED AR PROCESSES

Ron Shiff and Arie Yeredor

Dept. of Elec. Eng. - Systems, Tel-Aviv University
shiffaha@post.tau.ac.il, arie@eng.tau.ac.il

ABSTRACT

We address the blind separation of two autoregressive (AR) processes from a single mixture thereof, when their respective driving-noise ("innovation") sequences are known to be temporally sparse. Unlike other single-channel separation schemes, which use dictionary-learning, our method essentially estimates the sparsifying transformation of each source directly from the observed mixture (by estimating the respective AR parameters), and therefore does not require a training stage. We cast the problem as a constrained, non-convex ℓ_1-norm minimization and propose an iterative solution scheme, which iterates between linear-programming-based estimation of the respective driving-sequences given estimates of the AR parameters, and gradient-based refinement of the estimated AR parameters given the estimated driving sequences. Near-perfect separation is demonstrated using a simulated example.

1. INTRODUCTION

The exploitation of sparsity in many signal processing tasks, such as Blind Source Separation (BSS), Sparse Component Analysis (SCA), Compressed Sensing, Error Correction and Spectrum Estimation, has recently seen increased interest in the signal processing community. One of the principal reasons is that the sparsity-model assumption, either in the form of explicit time-domain sparsity or in some hidden, underlying sparse representation, is often well-justified in practice. Furthermore, when the relevant sparsity-model assumption is indeed justified, sparsity-based tools are capable of delivering significant performance improvement over classical tools which ignore the sparsity.

In this work we address the problem of single-channel blind separation of the sum of two autoregressive (AR) sources. Our key to separation is the underlying assumption, that the driving-noise (sometimes called the "innovation") sequences of these AR processes are temporally sparse, i.e., mainly consist of sporadic spikes (of unknown locations and amplitudes). Such a model can be justified, for example, when the processes are voiced speech segments, where the "driving noise" resembles a "spikes train" generated at the vocal chords (e.g., [5]). Other possible examples are seismic measurements, electrocardiograms, or, more generally, processes which consist of several superimposed, differently-scaled and shifted replica of damped sinusoids (each representing the impulse-response of the all-poles system associated with the AR process generation).

The "blindness" implies that neither the respective AR-parameters of the two sources, nor the locations and amplitudes of the spikes in their respective innovation sequences, are known in advance. The only available data is the observed sum of the two processes (and, of course, the assumed sparsity of the innovation sequences). The goal is to recover the two sources, and, as a possibly important by-product, to provide estimates of their AR parameters and innovation sequences. Note that classical sparsity-based approaches to single-channel source separation usually rely on a predetermined sparse representation (e.g., the short-time Fourier transform, or some wavelet transform [4], [8]), or employ a "training" stage, in which statistical or structural properties of the sources are learned (in the form of an over-complete dictionary learning, e.g., [10]). In this context, our approach can be seen as being based on specially parameterized sparse decompositions, in which each source has its own sparsifying transformation (the inverse of its all-poles generating system), and the unknown (AR) parameters of these transformation are estimated directly from the observed signal, rather than being fixed in advance or learned in a training stage.

A common approach to sparsity-based estimation is the formulation of a sparsity-based criterion, whose minimization with respect to the unknown parameters and signals would yield the desired estimates. While the "natural" sparsity measure is the ℓ_0-norm (counting the number of non-zero elements), the minimization thereof often becomes computationally-prohibitive and extremely sensitive to even the slightest noise. A well-established, computationally more permissive alternative to ℓ_0-norm minimization, is the ℓ_1-norm minimization, which was shown to yield consistent estimates (equivalent to ℓ_0-norm minimization under some mild conditions) in various contexts - see, e.g., [6, 7, 1, 2].

Therefore, our separation approach in this paper is based on ℓ_1-norm minimization of the implied innovation sequences with respect to the separated sources and to their AR parameters. However, despite the convexity of the ℓ_1-norm, the resulting constrained minimization problem is generally non-convex. Our proposed solution is based on an alternating-directions iterative approach, alternating between ℓ_1 minimization of the implied innovation sequences given the AR parameters, and gradient-based refinement of the AR parameters' estimates given the innovation sequences.

2. PROBLEM FORMULATION

Consider the mixture

$$x[n] = x_1[n] + x_2[n] \quad \forall n, \tag{1}$$

 ISSN 2076-1465

where $x_1[n]$ and $x_2[n]$ are two AR processes,

$$x_k[n] = -\sum_{\ell=1}^{P_k} a_{k,\ell} x_k[n-\ell] + s_k[n] \quad k=1,2 \ \forall n, \quad (2)$$

P_1 and P_2 denote the respective AR-orders (assumed to be known), $\{a_{1,\ell}\}_{\ell=1}^{P_1}$, $\{a_{2,\ell}\}_{\ell=1}^{P_2}$ are the respective (unknown, real-valued) AR-parameters, and $s_1[n]$, $s_2[n]$ are the respective (unknown, real-valued) innovation (or "driving-noise") sequences. The innovation sequences $s_1[n]$ and $s_2[n]$ are assumed to be sparse.

Given N samples $x[n]$, $n = 0, \dots, N-1$, we wish to estimate $x_1[n]$ and $x_2[n]$ (over the same interval, $n = 0, \dots, N-1$). As by-products we would also obtain estimates of the AR parameters and of the innovation sequences.

In the sequel we shall denote by $X(z)$, $X_k(z)$ and $S_k(z)$ the Z-transforms of the respective sequences $x[n]$, $x_k[n]$ and $s_k[n]$ ($k = 1,2$) taken over the observation interval $n = 0, \dots, N-1$. Likewise, we shall denote by $A_k(z)$ ($k = 1,2$) the Z-transform of the respective AR-coefficients, $A_k(z) = 1 + \sum_{\ell=1}^{P_k} a_{k,\ell} z^{-\ell}$. We assume that $A_1(z)$ and $A_2(z)$ do not have common roots. We further define $A(z) \triangleq A_1(z)A_2(z)$, and denote the coefficients of this polynomial as $\{a_\ell\}_{\ell=0}^{P}$, with $P \triangleq P_1 + P_2$.

3. THE CLASSICAL APPROACH

For comparison (and also for use as an initial guess) in the sequel, let us briefly describe the "classical" approach, which ignores the sparsity assumption and regards the sources as general AR processes. Assuming that the driving-sequences $s_1[n]$ and $s_2[n]$ are mutually uncorrelated, spectrally-white random processes, with variances σ_1^2 and σ_2^2, respectively, $x[n]$ is obviously an autoregressive - moving-average (ARMA) process of orders $(P, \max\{P_1, P_2\})$, whose Z-spectrum is given by

$$S_{xx}(z) = \frac{\sigma_1^2 A_2(z)A_2(1/z) + \sigma_2^2 A_1(z)A_1(1/z)}{A(z)A(1/z)}. \quad (3)$$

Following estimation of the correlation sequence $R_{xx}[\ell]$ for $|\ell| \le P + \max\{P_1, P_2\}$, the coefficients of $A(z)$ can be estimated using the *modified Yule-Walker* (MYW) equations (see, e.g., [9]). Next, the estimated correlation sequence is convolved with the polynomial coefficients of the estimated $\hat{A}(z)\hat{A}(1/z)$, yielding an estimate of the numerator polynomial (for $|\ell| \le \max\{P_1, P_2\}$), denoted $\hat{B}(z)$. Now, in order to associate each of the poles (or complex-conjugate pole-pairs) of $\hat{A}(z)$ with either $A_1(z)$ or $A_2(z)$, we can try all possible partitions: for each candidate partition (implying a choice of $\hat{A}_1(z)$ and $\hat{A}_2(z)$), we would obtain (by a linear least-squares solution) values $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ optimizing the fit of

$$\hat{\sigma}_1^2 \hat{A}_2(z)\hat{A}_2(1/z) + \hat{\sigma}_2^2 \hat{A}_1(z)\hat{A}_1(1/z) \approx \hat{B}(z), \quad (4)$$

(in the sense of minimizing the sum of squared differences between the polynomial coefficients). We would then select the partition which yields the closest fit (with positive $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$), thereby obtaining the estimates of $\hat{A}_1(z)$ and $\hat{A}_2(z)$, which in turn provide estimates of the respective AR coefficients.

Once the AR coefficients are estimated, the sources can be estimated by Wiener filtering of $x[n]$, e.g., in Z-transform domain,

$$\hat{X}_1(z) = \frac{\hat{S}_{x_1x}(z)}{\hat{S}_{xx}(z)} X(z) = \frac{\hat{S}_{x_1x_1}(z)}{\hat{S}_{x_1x_1}(z) + \hat{S}_{x_2x_2}(z)} X(z) \quad (5)$$

with $\hat{S}_{x_kx_k}(z) = \hat{\sigma}_k^2 / \hat{A}_k(z)\hat{A}_k(1/z)$, $k = 1,2$. Note that under mild ergodicity conditions the AR coefficients' estimates are consistent, and, therefore, if the innovation sequences $s_1[n]$ and $s_2[n]$ are Gaussian, then asymptotically, the Wiener-filtering separation would be optimal (in the mean square error sense). In our case, however, the innovation sequences are clearly not stationary white Gaussian processes, which therefore leaves much room for improvement by exploitation of their sparsity.

4. SEPARATION USING ℓ_1-NORM MINIMIZATION

Applying Z-transform to (1), (2) over the interval $n = 0, \dots, N-1$ and neglecting end-effects, we get

$$X(z) = \frac{S_1(z)}{A_1(z)} + \frac{S_2(z)}{A_2(z)} = \frac{S_1(z)A_2(z) + S_2(z)A_1(z)}{A(z)}, \quad (6)$$

and therefore

$$X(z)A(z) = S_1(z)A_2(z) + S_2(z)A_1(z). \quad (7)$$

Thus, we can consider finding P_1-th and P_2-th order causal, monic FIR filters $h_1[n]$, $h_2[n]$ (resp.) and sequences $\hat{s}_1[n]$, $\hat{s}_2[n]$, which satisfy the convolution relation:

$$x[n] * h_1[n] * h_2[n] = \hat{s}_1[n] * h_2[n] + \hat{s}_2[n] * h_1[n], \quad (8)$$

such that $\hat{s}_1[n]$ and $\hat{s}_2[n]$ are "as sparse as possible". More explicitly, using the definition $h[n] \triangleq \sum_{\ell=0}^{P_1} h_1[\ell]h_2[n-\ell]$, (8) reads

$$\sum_{\ell=0}^{P} h[\ell]x[n-\ell] = \sum_{\ell=0}^{P_2} h_2[\ell]\hat{s}_1[n-\ell] + \sum_{\ell=0}^{P_1} h_1[\ell]\hat{s}_2[n-\ell]. \quad (9)$$

$h_1[n]$, $h_2[n]$ would then serve as estimates of $a_{1,n}$ and $a_{2,n}$ (resp.), and $\hat{s}_1[n]$, $\hat{s}_2[n]$ will be the estimated innovation sequences, which may all be substituted in (2) so as to yield the separated signals.

Thus, using the ℓ_1-norm as a measure of sparsity, we need to solve the following optimization problem:

$$\min_{\mathbf{h}_1, \mathbf{h}_2, \hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2} \|\hat{\mathbf{s}}_1\|_1 + \|\hat{\mathbf{s}}_2\|_1$$
$$\text{s.t.} : \begin{cases} h_1[0] = h_2[0] = 1 \\ x[n] * h[n] = \hat{s}_1[n] * h_2[n] + \hat{s}_2[n] * h_1[n], \end{cases} \quad (10)$$

where for shorthand we used $\mathbf{h}_k \triangleq [h_k[0] \ \cdots \ h_k[P_k]]^T$, $\hat{\mathbf{s}}_k \triangleq [\hat{s}_k[0] \ \cdots \ \hat{s}_k[N-1]]^T$ ($k = 1,2$), and where $\|\cdot\|_1$ denotes the ℓ_1-norm.

Although this problem is generally non-convex, we propose a method for finding a (possibly local) minimum, which, as we shall demonstrate, can yield estimates of the sources and AR parameters which are considerably more accurate than those attained by the "classical" approach above.

An outline of the proposed algorithm is given below, followed by details regarding implementation of the non-trivial steps.

Algorithm Outline

1. Find an initial guess for the FIR filters coefficients $\mathbf{h}_1$, $\mathbf{h}_2$, possibly taking the estimated AR parameters $\hat{a}_{1,n}$ and $\hat{a}_{2,n}$ (resp.) obtained by the "classical" approach above;
2. Given $\mathbf{h}_1$, $\mathbf{h}_2$ (therefore also $\mathbf{h}$), solve

$$\mathcal{O}_1: \quad \mathbf{s}_1^*, \mathbf{s}_2^* = \underset{\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2}{\operatorname{argmin}} \{\|\hat{\mathbf{s}}_1\|_1 + \|\hat{\mathbf{s}}_2\|_1\}$$
$$\text{s.t.}: \; x[n] * h[n] = \hat{s}_1[n] * h_2[n] + \hat{s}_2[n] * h_1[n]$$

3. Find the derivatives of $f \triangleq \|\mathbf{s}_1^*\|_1 + \|\mathbf{s}_2^*\|_1$ with respect to $\mathbf{h}_1$, $\mathbf{h}_2$;
4. Update $\mathbf{h}_k$ using the gradient method, $\mathbf{h}_k = \mathbf{h}_k - \eta \nabla_{\mathbf{h}_k} f$, where

$$\nabla_{\mathbf{h_k}} f \triangleq \left[\frac{\partial f}{\partial h_k[0]} \cdots \frac{\partial f}{\partial h_k[P_k]} \right]^T$$

   ($k = 1, 2$) and η is a small step-size;
5. Repeat 2-4 until convergence of $\mathbf{h}_1$, $\mathbf{h}_2$;
6. Generate the sources' estimates $\hat{x}_1[n]$, $\hat{x}_2[n]$ (for $n = 0, \ldots, N-1$) as:

$$\hat{x}_k[n] = -\sum_{\ell=1}^{P_k} h_k[\ell] \hat{x}_k[n-\ell] + s_k^*[n], \quad k = 1, 2$$

   (using zero initial conditions).

4.1 Solving the optimization $\mathcal{O}_1$ in Step 2

Problem $\mathcal{O}_1$ (in Step 2 above) is a convex minimization problem, which can be cast as a standard linear program as follows. First, we express $\mathcal{O}_1$ in matrix-vector form:

$$\min_{\hat{\mathbf{s}}} \|\hat{\mathbf{s}}\|_1 \quad \text{s.t} \quad \mathbf{H}\hat{\mathbf{s}} = \mathbf{X}\mathbf{h}, \tag{11}$$

where $\hat{\mathbf{s}} \triangleq [\hat{\mathbf{s}}_1^T \;\; \hat{\mathbf{s}}_2^T]^T$, $\mathbf{h} = [h[0] \; \cdots \; h[P]]^T$, and the matrices $\mathbf{H}$ and $\mathbf{X}$ are structured as follows: Neglecting end-effects, $\mathbf{H}$ is an $N \times 2N$ matrix, $\mathbf{H} = [\mathbf{H}_2 \;\; \mathbf{H}_1]$ with:

$$\mathbf{H}_k = \begin{bmatrix} 1 & & & & & \\ h_k[1] & 1 & & & & \\ \vdots & & \ddots & & & \\ h_k[P_k] & \ddots & \ddots & 1 & & \\ & \ddots & & & \ddots & \\ & & h_k[P_k] & \cdots & h_k[1] & 1 \end{bmatrix} \tag{12}$$

(an $N \times N$ matrix) for $k = 1, 2$ (the empty entries are zeros), and $\mathbf{X}$ is an $N \times (P+1)$ matrix,

$$\mathbf{X} = \begin{bmatrix} x[0] & & & \\ x[1] & x[0] & & \\ \vdots & & \ddots & \\ x[P] & \cdots & \cdots & x[0] \\ x[P+1] & \ddots & \ddots & x[1] \\ \vdots & & & \vdots \\ x[N-1] & \cdots & \cdots & x[N-1-P] \end{bmatrix}. \tag{13}$$

It can be shown [3] that a solution to (11) is obtained from the solution of the following linear program (LP):

$$\min_{\mathbf{z}} \mathbf{1}^T \mathbf{z} \quad \text{s.t} \quad \mathbf{z} \succeq \mathbf{0}, \;\; \mathbf{H}[\mathbf{I} \;\; -\mathbf{I}]\mathbf{z} = \mathbf{X}\mathbf{h} \tag{14}$$

where $\mathbf{1}$ and $\mathbf{0}$ are $4N \times 1$ all-ones and all-zeros vectors, $\mathbf{I}$ is the $2N \times 2N$ identity matrix and $\succeq$ stands for an elementwise $\geq$ relation. The solution $\mathbf{s}^*$ of (11) is related to the solution $\mathbf{z}^*$ of (14) via $\mathbf{s}^* = [\mathbf{I} \;\; -\mathbf{I}]\mathbf{z}^*$, with the minimizing norm given by $\|\mathbf{s}^*\|_1 = \|\mathbf{z}^*\|_1 = \mathbf{1}^T \mathbf{z}^*$.

4.2 Calculating the Gradient in Step 3

In order to calculate the gradients of $\|\mathbf{s}^*\|_1$ (the solution of (11)) with respect to $\mathbf{h}_1$ and $\mathbf{h}_2$, we first obtain the derivatives of a general LP problem with respect to its constraints, as derived by Pearlmutter *et al.* in [10], and then apply the chain-rule with reverse accumulation to obtain the derivatives with respect to the filter coefficients.

Let us consider the general LP problem:

$$\min_{\mathbf{z}} \mathbf{w}^T \mathbf{z} \quad \text{s.t} \quad \mathbf{A}\mathbf{z} \preceq \mathbf{a} \;\; , \;\; \mathbf{B}\mathbf{z} = \mathbf{b} \tag{15}$$

With $\mathbf{w}, \mathbf{z} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{k \times n}$, $\mathbf{B} \in \mathbb{R}^{m \times n}$.

Each row in $\mathbf{A}$ and in $\mathbf{B}$ (with their matching elements in $\mathbf{a}$ and $\mathbf{b}$) defines a constraint. A constraint is called "active" if the solution lies on its respective boundary. Given feasibility, boundedness and uniqueness, the LP solution satisfies n independent linear equations, which correspond to the n active constraints [3]. In [10], the authors defined a sparse $n \times (k+m)$ matrix $\mathbf{P}$, which is an all-zeros matrix with a single 1 in each row, transforming the constraint matrix $\mathbf{C} \triangleq \left[\mathbf{A}^T \; \mathbf{B}^T\right]^T$ and constraints vector $\mathbf{c} \triangleq \left[\mathbf{a}^T \; \mathbf{b}^T\right]^T$ into a matrix $\mathbf{C_A}$ and a vector $\mathbf{c_A}$, which contain only the active constraints (it is assumed that the information regarding the indices of the active constraints is provided by the LP solver). Therefore, defining

$$\mathbf{C_A} \triangleq \mathbf{P}\mathbf{C}, \quad \mathbf{c_A} \triangleq \mathbf{P}\mathbf{c}, \tag{16}$$

the minimizing solution $\mathbf{z}$ of (15) can be readily expressed as the solution to the $n \times n$ set of linear active constraints,

$$\mathbf{z} = \mathbf{C_A}^{-1} \mathbf{c_A}. \tag{17}$$

4.2.1 Differentiation of the minimized criterion with respect to the constraint matrices and vectors

For convenience of notations we shall adopt the notation from [10] for the derivative of $\mathbf{w}^T\mathbf{z}$ with respect to an arbitrary matrix $\mathbf{Y}$:

$$\grave{\mathbf{Y}} \triangleq \frac{\partial(\mathbf{w}^T\mathbf{z})}{\partial \mathbf{Y}}, \tag{18}$$

where $\partial f(\mathbf{Y})/\partial \mathbf{Y}$ denotes a matrix of the same size as $\mathbf{Y}$, whose (i,j)-th element equals $\partial f(\mathbf{Y})/\partial Y_{i,j}$.

It is relatively straightforward to show (see [10] for details) that the derivatives $\grave{\mathbf{c}}_\mathbf{A}$ and $\grave{\mathbf{C}}_\mathbf{A}$ of the minimized criterion $\mathbf{w}^T\mathbf{z} = \text{Trace}(\mathbf{z}\mathbf{w}^\text{T})$ with respect to the active constraints vector $\mathbf{c}_\mathbf{A}$ and matrix $\mathbf{C}_\mathbf{A}$ (resp.) are given by

$$\grave{\mathbf{c}}_\mathbf{A} \triangleq \frac{\partial(\mathbf{w}^T\mathbf{z})}{\partial \mathbf{c}_\mathbf{A}} = (\mathbf{w}^T\mathbf{C}_\mathbf{A}{}^{-1})^T \quad , \quad \grave{\mathbf{C}}_\mathbf{A} = -\grave{\mathbf{c}}_\mathbf{A}\mathbf{z}^T. \tag{19}$$

Since the derivative of the minimized criterion with respect to the non-active constraints is zero, we conclude that

$$\grave{\mathbf{c}} = \mathbf{P}^T\grave{\mathbf{c}}_\mathbf{A} \quad , \quad \grave{\mathbf{C}} = \mathbf{P}^T\grave{\mathbf{C}}_\mathbf{A}. \tag{20}$$

4.2.2 Differentiation of the minimization criterion with respect to the filters' coefficients

We now return to our original problem (14), which, in terms of the general LP problem (15), has

$$\mathbf{w} = \mathbf{1},\ \mathbf{A} = -\mathbf{I},\ \mathbf{a} = \mathbf{0},\ \mathbf{B} = \mathbf{H}[\mathbf{I}\ \ -\mathbf{I}],\ \mathbf{b} = \mathbf{X}\mathbf{h}. \tag{21}$$

$\mathbf{C}$ and $\mathbf{c}$ are therefore given by

$$\mathbf{C} = \begin{bmatrix} -\mathbf{I} \\ \mathbf{H}[\mathbf{I}\ \ -\mathbf{I}] \end{bmatrix} \quad , \quad \mathbf{c} = \begin{bmatrix} \mathbf{0} \\ \mathbf{X}\mathbf{h} \end{bmatrix}, \tag{22}$$

and $\mathbf{C}_\mathbf{A}$ and $\mathbf{c}_\mathbf{A}$ are obtained from (16) once $\mathbf{P}$ is provided by the LP solver.

Since both $\mathbf{C}_\mathbf{A}$ and $\mathbf{c}_\mathbf{A}$ are functions of the filter coefficients, we apply the chain-rule with reverse accumulation in order to obtain the derivatives of the minimized ℓ_1-norm criterion with respect to the filter coefficients. Let $\grave{h}_k[n]$ denote the derivative of the minimization criterion with respect to the k-th sources' n-th filter coefficients ($k = 1, 2$). The derivative can be calculated by combining the two following terms:

- $\grave{h}^I_k[n]$, the term resulting from the relation to the constraint matrix $\mathbf{H}[\mathbf{I}\ \ -\mathbf{I}]$; and
- $\grave{h}^{II}_k[n]$, the term resulting from the relation to the constraint vector $\mathbf{X}\mathbf{h}$,

so that

$$\grave{h}_k[n] \triangleq \frac{\partial \|\hat{\mathbf{s}}\|_1}{\partial h_k[n]} = \grave{h}^I_k[n] + \grave{h}^{II}_k[n]. \tag{23}$$

Defining $\mathbf{y} \stackrel{\triangle}{=} \mathbf{X}\mathbf{h}$, we have, from (22), $\grave{\mathbf{y}} = [\mathbf{0}\ \ \mathbf{I}]\,\grave{\mathbf{c}}$, and

$$\grave{\mathbf{H}} = [\mathbf{0}\ \ \mathbf{I}]\,\grave{\mathbf{C}} \begin{bmatrix} \mathbf{I} \\ -\mathbf{I} \end{bmatrix}. \tag{24}$$

Now, beginning with the derivatives with respect to $\mathbf{h}_1$, we have

$$\begin{aligned}
\grave{h}^I_1[n] &= \sum_{i=1}^{3N}\sum_{j=1}^{2N} \frac{\partial\|\hat{\mathbf{s}}\|_1}{\partial C_{i,j}} \frac{\partial C_{i,j}}{\partial h_1[n]} \\
&= \sum_{i=1}^{3N}\sum_{j=1}^{2N} \grave{C}_{i,j}\frac{\partial C_{i,j}}{\partial h_1[n]} = \sum_{i=1}^{N}\sum_{j=1}^{2N} \grave{H}_{i,j}\frac{\partial H_{i,j}}{\partial h_1[n]} \\
&= \sum_{i=n+1}^{N}\sum_{j=1}^{2N} \grave{H}_{i,j}(\delta[j-(N+i-n)]) \\
&= \sum_{i=n+1}^{N} \grave{H}_{i,N+i-n}
\end{aligned} \tag{25}$$

and

$$\begin{aligned}
\grave{h}^{II}_1[n] &= \sum_{i=1}^{3N} \frac{\partial\|\hat{\mathbf{s}}\|_1}{\partial c_i}\frac{\partial c_i}{\partial h_1[n]} \\
&= \sum_{i=1}^{3N} \grave{c}_i \frac{\partial c_i}{\partial h_1[n]} = \sum_{i=1}^{N} \grave{y}_i \frac{\partial y_i}{\partial h_1[n]} \\
&= \sum_{i=1}^{N} \grave{y}_i \left(\mathbf{X}\tilde{\mathbf{h}}^n_2\right)_i = \grave{\mathbf{y}}^T\mathbf{X}\tilde{\mathbf{h}}^n_2,
\end{aligned} \tag{26}$$

where we have used

$$\frac{\partial y_i}{\partial h_1[n]} = \left(\mathbf{X}\tilde{\mathbf{h}}^n_2\right)_i \tag{27}$$

with

$$\tilde{\mathbf{h}}^n_k \stackrel{\triangle}{=} [\underbrace{0,...,0}_{n}, \mathbf{h}^T_{\bar{k}}, \underbrace{0,...,0}_{P_k - n}]^T \quad k = 1, 2, \tag{28}$$

$\bar{k}$ denoting the other filter's index, which is 2 if $k = 1$ and 1 if $k = 2$.

A similar derivation for the derivative with respect to the second process' filter coefficients $h_2[n]$ yields:

$$\grave{h}^I_2[n] = \sum_{i=n+1}^{N} \grave{H}_{i,i-n} \quad . \quad \grave{h}^{II}_2[n] = \grave{\mathbf{y}}^T\mathbf{X}\tilde{\mathbf{h}}^n_1. \tag{29}$$

5. SIMULATION RESULTS

To demonstrate the attainable separation, we applied the proposed algorithm to a mixture generated as follows. First, we generated each innovation process $s_k[n]$ ($k = 1, 2$) as a product of an independent, identically distributed (iid) Bernoulli process (taking the values $1, 0$ with probabilities $\rho_k, 1-\rho_k$, resp.) with an iid zero-mean Gaussian process with variance σ^2_k. We used $\rho_1 = \rho_2 = 0.1$ and $\sigma^2_1 = \sigma^2_2 = 1$. The generated sequences are shown in Figure 1. Then, these sequences were used for generating the sources $x_k[n]$ ($k = 1, 2$) by applying the difference equation (2) with AR orders $P_1 = P_2 = 2$, with AR coefficients set such that $A_1(z)$ has its poles at $0.9 \cdot e^{\pm j\pi/6}$ and $A_2(z)$ has its poles at $0.98 \cdot e^{\pm j\pi/3}$. The resulting signals and their sum $x[n]$

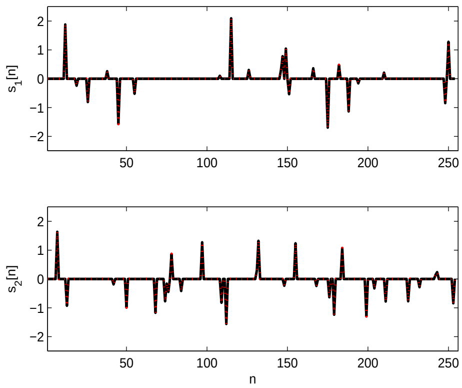

Figure 1: The innovation sequences (solid) and their sparsity-based estimates (dotted) (the two plots are nearly indistinguishable).

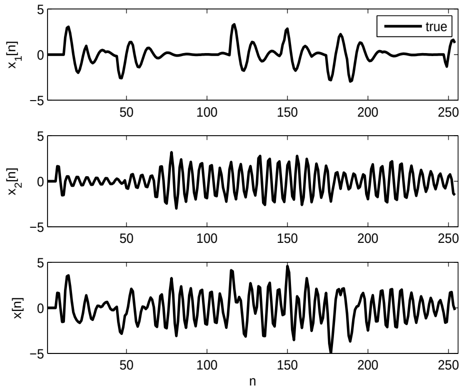

Figure 2: The generated AR sources and their mixture.

(from which they need to be estimated) are shown in Figure 2. The observation interval's length is $N = 256$.

We first applied the "classical" approach described in Section 3, and obtained initial estimates of the AR parameters (as well as estimates of the separated sources). The AR estimates were used as the initial guess for our sparsity-based algorithm. As shown in Table 1, the accuracy in estimating the poles was significantly improved by this algorithm.

Table 1: AR parameters and poles estimation

		True	MYW est.	ℓ_1-norm est.
$x_1[n]$	a_1	-1.5588	-1.6275	-1.5589
	a_2	0.8100	0.8705	0.8101
	$\|p_{1,2}\|$	0.9000	0.9330	0.9000
	$\angle p_{1,2}$	$\pm\frac{\pi}{6}$	$\pm 0.9763\frac{\pi}{6}$	$\pm 1.0001\frac{\pi}{6}$
$x_2[n]$	a_1	-0.9800	-0.9868	-0.9799
	a_2	0.9604	0.9691	0.9605
	$\|p_{1,2}\|$	0.9800	0.9844	0.9801
	$\angle p_{1,2}$	$\pm\frac{\pi}{3}$	$\pm 0.9987\frac{\pi}{3}$	$\pm 1.0001\frac{\pi}{3}$

The errors in the separated signals are shown in Figure 3, comparing our sparsity-based separation error to the error attained by the "classical" Wiener separation (based on the MYW AR parameters estimate). The time-averaged square error was 0.0879 for the Wiener separation, and $9.15 \cdot 10^{-9}$ for the sparsity-based separation (for both signals; note that since the sum of outputs must equal $x[n]$, the estimation error signals are "mirror images" of each other, so their averaged squared values

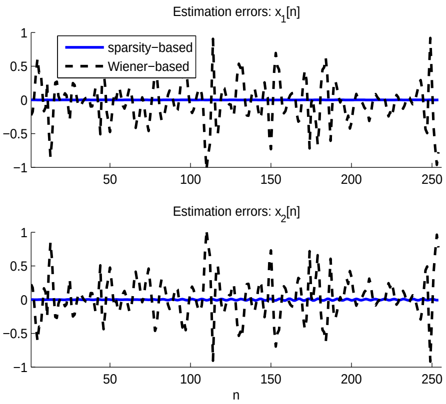

Figure 3: The estimation errors in $x_1[n]$ and in $x_2[n]$ attained by the Wiener-based and sparsity-based approaches.

are the same).

6. CONCLUSION

We demonstrated, as a "proof of concept", the ability to obtain good separation results for a single mixture of AR sources of unknown parameters, based solely on the assumption that their innovation sequences are sparse. Such AR processes are generally not sparse in the time-domain, frequency-domain or in classical time-frequency or wavelet-decomposition domains. Nevertheless, they each admit a sparsifying linear transformation (by FIR filtering), however the parameters (filter coefficients) of these transformations are unknown in a blind, untrained scenario. Our approach enables joint estimation of these parameters together with the ℓ_1-norm based separation process, without a need for a training stage.

Statistical characterization of the performance, as well as computational and convergence issues and additive-noise effects, are still under study.

REFERENCES

[1] Yeredor A. Multipath Mitigation Estimation Using ℓ_1 Minimization. *Proceedings of EUSIPCO 2008*, 2008.

[2] Yeredor A. On the Consistency of ℓ_1-norm based AR Parameters Estimation in a Sparse Multipth Environment. *Proceedings of ICASSP 2009*, 2009.

[3] D. Bertsimas and J. N. Tsitsiklis. *Introduction to Linear Optimization.* Dynamic Ideas, 1997.

[4] P. Bofill and M. Zibulevsky. Underdetermined Blind Source Separation Using Sparse Representations. *Signal Processing*, 81(11):2353–2362, 2001.

[5] Hansen J.H.L. Deller, J.R. and J.G. Proakis. *Discrete Time Processing of Speech Signals (2d ed.)*. IEEE Press, 2000.

[6] Donoho D.L and Elad M. Optimally Sparse Represntations from Overcomplete Dictionaries via ℓ_1-norm Minimization. *Proc. Nat. Acad. Sci. USA*, 100(5):2197–2202, Mar. 2002.

[7] Donoho D.L and Tsaig Y. Breakdown Equivalance Between the Minimal ℓ_1-norm Solution and the Sparsest Solution. *Signal Processing*, 86(3):533–548, Mar. 2006.

[8] Y. Li, S.-I. Amari, A. Cichocki, D.W.C. Ho, and S. Xie. Underdetermined Blind Source Separation Based on Sparse Representation. *IEEE Trans. on Signal Processing*, 54(2):423–437, 2006.

[9] Hayes M.H. *Statistical Digital Signal Processing and Modeling.* John Wiley Sons, Inc., 1996.

[10] B. A. Pearlmutter and R. K. Olsson. Linear program differentiation for single-channel speech separation. *Proc. MLSP'06*, Sep 2006.