BAYESIAN SEPARATION AND RECOVERY OF CONVOLUTIVELY MIXED AUTOREGRESSIVE SOURCES

Simon J. Godsill - Christophe Andrieu

Signal Processing Group, University of Cambridge
Department of Engineering, Trumpington Street
CB2 1PZ Cambridge, UK
Email: sjg@eng.cam.ac.uk - ca226@eng.cam.ac.uk

ABSTRACT

In this paper we address the problem of the separation and recovery of convolutively mixed autoregressive processes in a Bayesian framework. Solving this problem requires the ability to solve integration and/or optimization problems of complicated posterior distributions. We thus propose efficient stochastic algorithms based on Markov chain Monte Carlo (MCMC) methods. We present three algorithms. The first one is a classical Gibbs sampler that generates samples from the posterior distribution. The two other algorithms are stochastic optimization algorithms that allow to optimize either the marginal distribution of the sources, or the marginal distribution of the parameters of the sources and mixing filters, conditional upon the observation. Simulations are presented.

1 INTRODUCTION

The paper is organized as follows. In Section 2 we present the model for the data. In Section 3 we propose a Bayesian model to solve the problem. In Section 4 we propose a MCMC algorithm that allows us to obtain samples from the posterior distribution. Section 5 is devoted to two optimization algorithms that estimate two MMAP: the MAP of the marginal posterior distributions of the sources conditional upon the observations, and the posterior distribution of the parameters conditional upon the observations, that is in the later case that the sources are integrated out. In Section 6 we present simulation results and in Section 7 we draw some conclusions and discuss our contributions.

2 MODEL OF THE DATA AND OBJECTIVES

The problem addressed is the problem of source separation, the n sources being modelled as autoregressive processes.

2.1 Model for the sources

Source i is modelled for $t = 1, \ldots, T$ as:

$$s_t^{\langle i \rangle} = \mathbf{a}_{\langle i \rangle}^{\mathrm{T}} \mathbf{s}_{k-1:k-l_{\langle i \rangle}}^{\langle i \rangle} + \sigma_{\langle i \rangle} e_t^{\langle i \rangle} \tag{1}$$

where $\mathbf{s}_{a:b}^{\langle i \rangle} \triangleq \begin{pmatrix} s_a^{\langle i \rangle} & \cdots & \cdots & s_b^{\langle i \rangle} \end{pmatrix}^{\mathrm{T}}$, $\mathbf{a}_{\langle i \rangle} \triangleq \mathbf{a}_{1:l_{\langle i \rangle}}^{\langle i \rangle}$ and $l_{\langle i \rangle}$ is the length of the i^{th} AR model. We assume that $\mathbf{s}_0^{\langle i \rangle} \triangleq \mathbf{0}_{l_{\langle i \rangle} \times 1}$ for $i = 1, \ldots, n$. $\left(e_t^{\langle i \rangle}\right)_{t=1,\ldots,T}$ is a zero mean normalized i.i.d. Gaussian sequence, *i.e.* for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, $e_t^{\langle i \rangle} \sim \mathcal{N}(0,1)$. $\left(\sigma_{\langle i \rangle}^2\right)_{i=1,\ldots,n}$ are the variances of the dynamic noise for each source. Further on we will denote $v_t^{\langle i \rangle} \triangleq \sigma_{\langle i \rangle} e_t^{\langle i \rangle}$ and we assume that the excitations of the different dynamic noises are independent.

C. Andrieu is sponsored by Olivetti Oracle Research Laboratory (ORL), Cambridge UK.

2.2 Mixing noisy model

The mixing model is assumed to be a multidimensional time invariant FIR filter. More precisely we assume that the sources are mixed and corrupted by an additive Gaussian i.i.d. noise sequence: at the j^{th} sensor, and for $j = 1, \ldots, m$

$$y_k^{\langle j \rangle} = \sum_{i=1}^{n} \mathbf{h}_{\langle i,j \rangle}^{\mathrm{T}} \mathbf{s}_{t:t-L_{\langle i,j \rangle}+1}^{\langle i \rangle} + \sigma_{\langle j \rangle}^{*} \varepsilon_k^{\langle j \rangle} \tag{2}$$

where $L_{\langle i,j \rangle}$ is the length of the filter from source i to sensor j. $\left(\varepsilon_t^{\langle i \rangle}\right)_{t=1,\ldots,T}$ is a zero mean normalized i.i.d. Gaussian sequence, *i.e.* for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, $\varepsilon_t^{\langle i \rangle} \sim \mathcal{N}(0,1)$. $\left(\sigma_{\langle i \rangle}^{2*}\right)_{i=1,\ldots,m}$ are the variance of the observation noise for each sensor. They are assumed independent of the excitations of the AR models. Note that we assume $h_t^{\langle i,j \rangle} = 1$.

2.3 State-space representation of the data

In order to design efficient algorithms we will see in Section 4 that it is important to rewrite the model of the data using a state space represention:

$$\mathbf{s}_{t:t-\lambda_{\langle i \rangle}+1}^{\langle i \rangle} = \mathbf{A}_{\langle i \rangle} \mathbf{s}_{t-1:t-\lambda_{\langle i \rangle}}^{\langle i \rangle} + \mathbf{B}_{\langle i \rangle} e_t^{\langle i \rangle} \tag{3}$$

where $\lambda_{\langle i \rangle} \triangleq \max\left\{ l_{\langle i \rangle}, \max_j \left\{ L_{\langle i,j \rangle} \right\} \right\}$. $\mathbf{A}_{\langle i \rangle}$ and $\mathbf{B}_{\langle i \rangle}$ are defined as:

$$\begin{aligned} \mathbf{A}_{\langle i \rangle} &\triangleq \begin{bmatrix} \mathbf{a}_{\langle i \rangle}^{\mathrm{T}} & \mathbf{0}_{1 \times (\lambda_{\langle i \rangle} - l_{\langle i \rangle})} \\ \mathbf{I}_{\lambda_{\langle i \rangle}-1} & \mathbf{0}_{(\lambda_{\langle i \rangle}-1) \times 1} \end{bmatrix} \\ \mathbf{B}_{\langle i \rangle} &\triangleq \begin{bmatrix} \sigma_{\langle i \rangle} & \mathbf{0}_{(\lambda_{\langle i \rangle}-1) \times 1} \end{bmatrix}^{\mathrm{T}} \end{aligned} \tag{4}$$

Then defining

$$\begin{aligned} \mathbf{A} &\triangleq diag\begin{bmatrix} \mathbf{A}_{\langle 1 \rangle} & \cdots & \mathbf{A}_{\langle n \rangle} \end{bmatrix} \\ \mathbf{B} &\triangleq diag\begin{bmatrix} \mathbf{B}_{\langle 1 \rangle} & \cdots & \mathbf{B}_{\langle m \rangle} \end{bmatrix} \end{aligned} \tag{5}$$

one can write the evolution equation:

$$\widetilde{\mathbf{s}}_t^{\langle 1:n\rangle} = \mathbf{A}\widetilde{\mathbf{s}}_{t-1}^{\langle 1:n\rangle} + \mathbf{B}e_t^{\langle 1:n\rangle} \tag{6}$$

The observation equation takes the following form:

$$\mathbf{y}_t^{\langle 1:m\rangle} = \mathbf{C}\widetilde{\mathbf{s}}_t^{\langle 1:n\rangle} + \mathbf{D}\boldsymbol{\varepsilon}_t^{\langle 1:m\rangle} \tag{7}$$

where $\mathbf{C}$ is equal to

$$\begin{bmatrix} \mathbf{h}_{\langle 1,1\rangle}^{\mathrm{T}} & \mathbf{0}_{1\times(\lambda_1-L_{1,1})} & \cdots & \mathbf{h}_{\langle n,1\rangle}^{\mathrm{T}} & \mathbf{0}_{1\times(\lambda_1-L_{n,1})} \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ \mathbf{h}_{\langle 1,m\rangle}^{\mathrm{T}} & \mathbf{0}_{1\times(\lambda_n-L_{1,m})} & \cdots & \mathbf{h}_{\langle n,m\rangle}^{\mathrm{T}} & \mathbf{0}_{1\times(\lambda_n-L_{n,m})} \end{bmatrix} \tag{8}$$

and

$$\begin{aligned} \widetilde{\mathbf{s}}_t^{\langle 1:n\rangle} &\triangleq \begin{pmatrix} \mathbf{s}_{t:t-\lambda_{\langle 1\rangle}+1}^{\langle 1\rangle\mathrm{T}} & \cdots & \mathbf{s}_{t:t-\lambda_{\langle n\rangle}+1}^{\langle n\rangle\mathrm{T}} \end{pmatrix}^{\mathrm{T}} \\ \mathbf{D} &\triangleq diag\begin{bmatrix} \sigma_{\langle 1\rangle}^* & \cdots & \sigma_{\langle m\rangle}^* \end{bmatrix} \end{aligned} \tag{9}$$

2.4 Objectives

Our aim is, given the number of sources n, $\mathbf{l}_{\langle 1:n\rangle}$ and $\mathbf{L}_{\langle 1:n,1:m\rangle}$ to estimate the parameters $\mathbf{a}_{\langle 1:n\rangle}$ and the sequences $\mathbf{v}_{1:T}^{\langle 1:n\rangle}$ from the observations $\mathbf{y}_{1:T}^{\langle 1:m\rangle}$. To achieve this, we adopt in this paper a Bayesian approach.

3 BAYESIAN MODEL

In a Bayesian framework, priors are needed for all the parameters of the model. We assume the following probabilistic structure of the *a priori* density of the parameters. We note

$$\boldsymbol{\theta} \triangleq \left\{\mathbf{a}_{\langle 1:n\rangle}, \mathbf{h}_{\langle 1:n,1:m\rangle}, \sigma_{\langle 1:n\rangle}^2, \sigma_{\langle 1:m\rangle}^{2*}\right\} \tag{10}$$

then

$$p(\boldsymbol{\theta}) = \prod_{i=1}^{n} p\left(\mathbf{a}_{\langle i\rangle}\middle|\,\sigma_{\langle i\rangle}^2\right) p\left(\sigma_{\langle i\rangle}^2\right) \prod_{j=1}^{m} p\left(\mathbf{h}_{\langle i,j\rangle}\middle|\,\sigma_{\langle i\rangle}^{2*}\right) p\left(\sigma_{\langle i\rangle}^{2*}\right) \tag{11}$$

that is we introduce independence among the sources and among the mixing filters. We now describe the different prior distributions of the parameters of the model.

3.1 Priors for $\left(\mathbf{a}_{\langle i\rangle}, \sigma_{\langle i\rangle}^2\right)$

We assume that for $i = 1, \ldots, n$

$$\begin{aligned} \sigma_{\langle i\rangle}^2 &\sim \mathcal{IG}\left(\frac{\nu_0^{\langle i\rangle}}{2}, \frac{\gamma_0^{\langle i\rangle}}{2}\right) \\ \mathbf{a}^{\langle i\rangle}\Big|\,\sigma_{\langle i\rangle}^2 &\sim \mathcal{N}\left(\mathbf{0}_{l_{\langle i\rangle}\times 1}, \sigma_{\langle i\rangle}^2\alpha_{\langle i\rangle}\mathbf{I}_{l_{\langle i\rangle}}\right) \end{aligned} \tag{12}$$

where $\mathcal{IG}$ is the inverted-gamma distribution [3] and $\alpha_{\langle i\rangle} > 0$. Note that as $\nu_0^{\langle i\rangle}, \gamma_0^{\langle i\rangle} \to 0$ one obtains Jeffreys' uninformative prior. The prior on $\mathbf{a}^{\langle i\rangle}$ becomes uninformative as $\alpha_{\langle i\rangle} \to +\infty$. Note however that if $l_{\langle i\rangle}$ is unknown, one should avoid such an improper prior that might lead to Lindley's paradox [3].

3.2 Priors for $\left(\mathbf{h}_{\langle i,j\rangle}, \sigma_{\langle i\rangle}^{2*}\right)$

We assume that for $j = 1, \ldots, m$ and $i = 1, \ldots, n$

$$\begin{aligned} \sigma_{\langle i\rangle}^{2*} &\sim \mathcal{IG}\left(\frac{\nu_0^{*\langle i\rangle}}{2}, \frac{\gamma_0^{*\langle i\rangle}}{2}\right) \\ \mathbf{h}_{\langle i,j\rangle}\Big|\,\sigma_{\langle i\rangle}^{2*} &\sim \mathcal{N}\left(\mathbf{0}_{L_{\langle i,j\rangle}\times 1}, \sigma_{\langle i\rangle}^{2*}\beta_{\langle i,j\rangle}\mathbf{I}_{L_{\langle i,j\rangle}}\right) \end{aligned} \tag{13}$$

where $\beta_{\langle i,j\rangle} > 0$. The remarks given in the preceding subsection concerning the hyperparameters of the model apply here. This allows us to define a uninformative prior distribution for these parameters.

3.3 Bayesian objectives

In a Bayesian framework our aim will be to estimate the following posterior distribution:

$$p\left(\boldsymbol{\theta}, \mathbf{s}_{1:T}^{\langle 1:n\rangle}\middle|\,\mathbf{y}_{1:T}^{\langle 1:m\rangle}\right) \propto p\left(\mathbf{y}_{1:T}^{\langle 1:m\rangle}\middle|\,\boldsymbol{\theta}, \mathbf{s}_{1:T}^{\langle 1:n\rangle}\right) p\left(\boldsymbol{\theta}, \mathbf{s}_{1:T}^{\langle 1:n\rangle}\right) \tag{14}$$

Neither this distribution, nor features such as the MAP estimate of the sources or the MMSE estimate of the parameters can be obtained in closed-form. This is why numerical methods are required either to optimize or integrate. Note that the use of reversible jump MCMC algorithms introduced in [6] could allow us to treat the case when the dimensions of the AR models and of the mixing filters are unknown.

4 MCMC ALGORITHM

MCMC methods consists of running an ergodic Markov chain (MC) $\left(\mathbf{x}^{(i)}\right)_{i\in\mathbb{N}}$ admitting as equilibrium distribution $\pi(\cdot)$ the a required distribution, the posterior distribution of the parameters of the model in a Bayesian framework. The samples obtained from the ergodic MC allow us to estimate, when they exist, quantities such as $\int f(\mathbf{x})\,\pi(\mathbf{x})\,d\mathbf{x}$ by ergodic averaging. We will illustrate in Section 5 that they also allows us to design optimization algorithms to find the maxima of $\pi(\cdot)$. We now describe a Gibbs sampler algorithm that generates an ergodic MC $\left(\boldsymbol{\theta}^{(i)}, \mathbf{s}_{1:T}^{\langle 1:n\rangle(i)}\right)_{i\in\mathbb{N}}$ with asymptotic distribution $p\left(\boldsymbol{\theta}, \mathbf{s}_{1:T}^{\langle 1:n\rangle}\middle|\,\mathbf{y}_{1:T}^{\langle 1:m\rangle}\right)$.

4.1 Algorithm

In this section we present a Gibbs sampler that allows us to asymptotically obtain samples from the posterior distribution $p\left(\boldsymbol{\theta}, \mathbf{s}_{1:T}^{\langle 1:n\rangle}\middle|\,\mathbf{y}_{1:T}^{\langle 1:m\rangle}\right)$.

Homogeneous algorithm

1. Initialization set $\boldsymbol{\theta}^{(i)}$ to a random value.
2. Iteration i
   - Sample the dynamic noise $\mathbf{v}_{1:T}^{\langle 1:n\rangle(i)}$ and the variances of the corresponding Gaussian noises:

$$\mathbf{v}_{1:T}^{\langle 1:n\rangle}\Big|\left(\mathbf{h}_{\langle 1:n,1:m\rangle}^{(i-1)}, \boldsymbol{\sigma}_{\langle 1:m\rangle}^{(i-1)2*}, \mathbf{a}_{\langle 1:T\rangle}^{(i-1)}, \boldsymbol{\sigma}_{\langle 1:n\rangle}^{(i-1)2}, \mathbf{y}_{1:T}^{\langle 1:m\rangle}\right) \tag{15}$$

   - For $k = 1, \ldots, n$ evaluate $\mathbf{s}_{1:T}^{\langle k\rangle(\star)}$ from (3).
   - For $k = 1, \ldots, n$ sample the variances of the dynamic noises $\sigma_{\langle k\rangle}^2\big|\,\mathbf{s}_{1:T}^{\langle k\rangle(i)}$.
   - For $k = 1, \ldots, n$ sample the coefficients of the autoregressive models: $\mathbf{a}_{\langle k\rangle}\Big|\left(\sigma_{\langle k\rangle}^{(i)2}, \mathbf{s}_{1:T}^{\langle k\rangle(i)}\right)$.

- For $k = 1, \ldots, n$ evaluate $\mathbf{s}_{1:T}^{\langle k \rangle (i)}$ from (3).
- For $k = 1, \ldots, m$ simulate $\sigma_{\langle k \rangle}^{2*} \left| \left(\mathbf{y}_{1:T}^{\langle k \rangle}, \mathbf{s}_{1:T}^{\langle k \rangle (i)} \right) \right.$ (see (20)).
- For $k = 1, \ldots, m$ simulate $\widetilde{\mathbf{h}}_{\langle k \rangle} \left| \left(\mathbf{y}_{1:T}^{\langle k \rangle}, \sigma_{\langle k \rangle}^{*(i)2}, \mathbf{s}_{1:T}^{\langle 1:n \rangle (i)} \right) \right.$ (see (22)).

3. $i \leftarrow i + 1$.
4. Go to 2.

■

By construction the transition kernel defined by one iteration of the algorithm admits $p\left(\boldsymbol{\theta}, \mathbf{s}_{1:T}^{\langle 1:n \rangle} \middle| \mathbf{y}_{1:T}^{\langle 1:m \rangle} \right)$ as invariant distribution. One can easily check that this MC is irreducible and aperiodic, and thus converges towards the required posterior distribution [7].

4.2 Implementation

4.2.1 Sampling $\mathbf{v}_{1:N}^{\langle 1:n \rangle}$

This is done by performing the efficient Kalman disturbance smoother described in [4], which uses the the state-space representation introduced in subsection 2.3.

4.2.2 Sampling $\sigma_{\langle k \rangle}^2$

Classic algebraic manipulations lead to

$$\sigma_{\langle j \rangle}^2 \left| \mathbf{s}_{1:T}^{\langle k \rangle (i)} \right. \sim \mathcal{IG}\left(\frac{\nu_{\langle j \rangle} + T}{2}, \frac{\gamma_{\langle j \rangle} + \left\| \mathbf{s}_{1:T}^{\langle i \rangle} \right\|_{\mathbf{P}_{\langle j \rangle}}^2}{2} \right) \tag{16}$$

where $\|\mathbf{x}\|_{\mathbf{A}}^2 \triangleq \mathbf{x}^{\mathrm{T}} \mathbf{A} \mathbf{x}$ and

$$\begin{aligned} \mathbf{P}_{\langle k \rangle} &\triangleq \mathbf{I}_T - \mathbf{S}_{\langle j \rangle}^{\mathrm{T}} \overline{\mathbf{S}}_{\langle j \rangle} \mathbf{S}_{\langle j \rangle} \\ \overline{\mathbf{S}}_{\langle j \rangle}^{-1} &\triangleq \mathbf{S}_{\langle j \rangle}^{\mathrm{T}} \mathbf{S}_{\langle j \rangle} + \alpha_{\langle j \rangle}^{-1} \mathbf{I}_{l_{\langle j \rangle}} \end{aligned} \tag{17}$$

and $\mathbf{S}_{\langle k \rangle}$ is the Toeplitz matrix with first column $\mathbf{s}_{0:T-1}$ and first line $\mathbf{s}_{0:1-l_{\langle k \rangle}}$. It is easy to obtain samples from this distribution using standard techniques [5].

4.2.3 Sampling $\mathbf{a}_{\langle k \rangle}$

Directly from Bayes' theorem one can find that:

$$\mathbf{a}_{\langle j \rangle} \left| \left(\sigma_{\langle j \rangle}^2, \mathbf{s}_{1:T}^{\langle k \rangle}, \mathbf{y}_{1:T}^{\langle 1:m \rangle} \right) \right. \sim \mathcal{N}\left(\mathbf{m}_{\langle j \rangle}, \sigma_{\langle j \rangle}^2 \mathbf{M}_{\langle j \rangle} \right) \tag{18}$$

with

$$\mathbf{m}_{\langle j \rangle} \triangleq \overline{\mathbf{S}}_{\langle j \rangle} \mathbf{S}_{\langle j \rangle}^{\mathrm{T}} \mathbf{s}_{1:T}^{\langle k \rangle} \tag{19}$$

4.2.4 Sampling the variance of the observation noise

Similarly

$$\sigma_{\langle j \rangle}^{2*} \left| \left(\mathbf{y}_{1:T}^{\langle k \rangle}, \mathbf{s}_{1:T}^{\langle k \rangle (i)} \right) \right. \sim \mathcal{IG}\left(\frac{T + \nu_0^{*\langle k \rangle}}{2}, \frac{\gamma_0^{*\langle k \rangle} + \left\| \mathbf{y}_{1:T}^{\langle k \rangle} - \mathbf{s}_{1:T}^{\langle k \rangle} \right\|_{\mathbf{P}_{\langle k \rangle}^*}^2}{2} \right) \tag{20}$$

where

$$\begin{aligned} \mathbf{P}_{\langle j \rangle}^* &\triangleq \mathbf{I}_T - \mathbf{R}_{\langle j \rangle}^{\mathrm{T}} \overline{\mathbf{R}}_{\langle j \rangle} \mathbf{R}_{\langle j \rangle} \\ \overline{\mathbf{R}}_{\langle j \rangle}^{-1} &\triangleq \mathbf{R}_{\langle j \rangle}^{\mathrm{T}} \mathbf{R}_{\langle j \rangle} + diag\left(\beta_{\langle i,j \rangle}^{-1} \mathbf{I}_{L_{\langle i,j \rangle}} \right) \end{aligned} \tag{21}$$

where $\mathbf{R}_{\langle j \rangle}$ is composed of the Toeplitz matrices with first column $\mathbf{x}_{0:T-1}$ and first line $\mathbf{x}_{0:-L_{\langle ij \rangle}}$, aligned from the left to the right.

4.2.5 Sampling $\mathbf{h}_{\langle i,j \rangle}$

For a given j one easily obtains

$$\begin{aligned} \mathbf{h}_{\langle 1:n,j \rangle} &\sim \mathcal{N}\left(\boldsymbol{\mu}_{\langle j \rangle}, \overline{\mathbf{R}}_{\langle j \rangle} \right) \\ \boldsymbol{\mu}_{\langle j \rangle} &\triangleq \overline{\mathbf{R}}_{\langle j \rangle} \mathbf{R}_{\langle j \rangle} \left(\mathbf{y}_{1:T}^{\langle j \rangle} - \mathbf{s}_{1:T}^{\langle j \rangle} \right) \end{aligned} \tag{22}$$

5 OPTIMIZATION ALGORITHMS

In this section we first present an algorithm to optimize $p\left(\boldsymbol{\theta} \middle| \mathbf{y}_{1:T}^{\langle 1:m \rangle} \right)$ which is a stochastic version of the algorithm presented in [9]. Then we show how it is possible to integrate the nuisance parameters out, here the mixing filters and the variance of the observation noise and of the dynamic noises to obtain an analytic expression of $p\left(\mathbf{s}_{1:T}^{\langle 1:n \rangle} \middle| \mathbf{y}_{1:T}^{\langle 1:m \rangle} \right)$ up to a normalizing constant. Then we use this expression to design an efficient stochastic optimisation algorithm that allows us to obtain the MMAP (Marginal Maximum A Posteriori) estimate of $\mathbf{s}_{1:N}^{\langle 1:n \rangle}$.

5.1 First algorithm

We consider an increasing positive sequence $(\gamma_i)_{i \in \mathbb{N}}$, then we can define the following optimization algorithm:

Simulated annealing algorithm

1. Initialization set $\boldsymbol{\theta}^{(i)}$ to a random value.
   Iteration i
2. Sample the dynamic noise $\mathbf{v}_{1:T}^{\langle 1:n \rangle (\star)}$:
$$\mathbf{v}_{1:T}^{\langle 1:n \rangle} \left| \left(\mathbf{h}_{\langle 1:n,1:m \rangle}^{(i-1)}, \boldsymbol{\sigma}_{\langle 1:m \rangle}^{(i-1)2*}, \mathbf{a}_{\langle 1:T \rangle}^{(i-1)}, \boldsymbol{\sigma}_{\langle 1:n \rangle}^{(i-1)2}, \mathbf{y}_{1:T}^{\langle 1:m \rangle} \right) \right. \tag{23}$$
3. Evaluate $\mathbf{s}_{1:N}^{\langle 1:n \rangle}$.
   Sample $\boldsymbol{\theta}^{(\star)}$, as in subsection 4.1.
4. Optimization step:
   - Evaluate
$$\alpha_{SA}^{(i)} = \min\left\{ 1, \left(\frac{p\left(\boldsymbol{\theta}^{(\star)} \middle| \mathbf{y}_{1:T}^{\langle 1:m \rangle} \right)}{p\left(\boldsymbol{\theta}^{(\star)(i-1)} \middle| \mathbf{y}_{1:T}^{\langle 1:m \rangle} \right)} \right)^{\gamma_i - 1} \right\} \tag{24}$$
   - If $\left(u \sim \mathcal{U}_{(0,1)} \right) \leq \alpha_{SA}^{(i)}$ then $\boldsymbol{\theta}^{(i)} \leftarrow \boldsymbol{\theta}^{(\star)}$ $\mathbf{s}_{1:T}^{\langle 1:n \rangle (i)} \leftarrow \mathbf{s}_{1:T}^{\langle 1:n \rangle (\star)}$
     else $\boldsymbol{\theta}^{(i)} \leftarrow \boldsymbol{\theta}^{(i-1)}$, $\mathbf{s}_{1:T}^{\langle 1:n \rangle (i)} \leftarrow \mathbf{s}_{1:T}^{\langle 1:n \rangle (i-1)}$
     EndIf
5. $i \leftarrow i + 1$.
6. Go to 2.

■

5.2 Integration of the parameters of the model

then one integrate $\mathbf{a}_{\langle 1:n \rangle}, \boldsymbol{\sigma}_{\langle 1:n \rangle}^2$ out, that is obtain an analytical expression of $p\left(\mathbf{s}_{1:T}^{\langle 1:n \rangle} \middle| \mathbf{y}_{1:T}^{\langle 1:m \rangle} \right)$ up to a normalizing constant.

$$\begin{aligned} p\left(\mathbf{s}_{1:T}^{\langle 1:n \rangle} \middle| \mathbf{y}_{1:T}^{\langle 1:m \rangle} \right) &\propto \prod_{i=1}^n \left| \overline{\mathbf{S}}_{\langle i \rangle} \right|^{\frac{1}{2}} \left[\gamma_0^{\langle i \rangle} + \mathbf{s}_{1:T}^{\langle i \rangle t} \mathbf{P}_{\langle i \rangle} \mathbf{s}_{1:T}^{\langle i \rangle} \right]^{-\frac{(T + \nu_0)}{2}} \\ &\times \prod_{j=1}^m \left| \overline{\mathbf{R}}_{\langle j \rangle} \right|^{\frac{1}{2}} \left[\gamma_0^{*\langle j \rangle} + \left\| \left(\mathbf{y}_{1:T}^{\langle j \rangle} - \mathbf{s}_{1:T}^{\langle j \rangle} \right) \right\|_{\mathbf{P}_{\langle j \rangle}^*}^2 \right]^{-\frac{\left(T + \nu_0^* \right)}{2}} \end{aligned} \tag{25}$$

The interpretation of this cost function is rather simple: the first term says how the sources fit the autoregressive model and the second term tells us how the sources fit the observed data.

5.3 Second algorithm

We present here an efficient stochastic algorithm that allows to optimise the probability density $p\left(\mathbf{s}_{1:T}^{\langle 1:n\rangle}\middle|\mathbf{y}_{1:T}^{\langle 1:m\rangle}\right)$, in other words to find the most probable sources. This algorithm uses the nuisance parameters in order to perform easy/efficient simulation of the parameters $\mathbf{s}_{1:T}^{\langle 1:n\rangle}$.

Simulated annealing algorithm

1. Initialization set $\boldsymbol{\theta}^{(i)}$ to a random value.
   Iteration i
2. Sample the dynamic noise $\mathbf{v}_{1:T}^{\langle 1:n\rangle(\star)}$:

$$\mathbf{v}_{1:T}^{\langle 1:n\rangle}\left|\left(\mathbf{h}_{\langle 1:n,1:m\rangle}^{(i-1)},\boldsymbol{\sigma}_{\langle 1:m\rangle}^{(i-1)2*},\mathbf{a}_{\langle 1:T\rangle}^{(i-1)},\boldsymbol{\sigma}_{\langle 1:n\rangle}^{(i-1)2},\mathbf{y}_{1:T}^{\langle 1:m\rangle}\right)\right. \tag{26}$$

3. Evaluate $\mathbf{s}_{1:N}^{\langle 1:n\rangle(\star)}$.
4. Optimization step:
   - Evaluate

$$\alpha_{SA}^{(i)}=\min\left\{1,\left(\frac{p\left(\mathbf{s}_{1:T}^{\langle 1:n\rangle(\star)}\middle|\mathbf{y}_{1:T}^{\langle 1:m\rangle}\right)}{p\left(\mathbf{s}_{1:T}^{\langle 1:n\rangle(i-1)}\middle|\mathbf{y}_{1:T}^{\langle 1:m\rangle}\right)}\right)^{\gamma_i-1}\right\} \tag{27}$$

   - If $\left(u\sim\mathcal{U}_{(0,1)}\right)\leq\alpha_{SA}^{(i)}$ then
     Sample $\boldsymbol{\theta}^{(\star)}$, as in subsection 4.1. $\boldsymbol{\theta}^{(i)}\leftarrow\boldsymbol{\theta}^{(\star)}$, $\mathbf{s}_{1:T}^{\langle 1:n\rangle(i)}\leftarrow\mathbf{s}_{1:T}^{\langle 1:n\rangle(\star)}$
     else $\boldsymbol{\theta}^{(i)}\leftarrow\boldsymbol{\theta}^{(i-1)},\mathbf{s}_{1:T}^{\langle 1:n\rangle(i)}\leftarrow\mathbf{s}_{1:T}^{\langle 1:n\rangle(i-1)}$
     EndIf
5. $i\leftarrow i+1$.
6. Go to 2.

6 SIMULATION RESULTS

We considered in simulation the case $n=m=2, l_{\langle i\rangle}=5$ and $L_{\langle ij\rangle}=10$ with noises equal to $\sigma_{\langle i\rangle}=.1$ and $\sigma_{\langle i\rangle}^{2*}=1.0$ for $i,j=1,2$ and 500 observations. We used the second optimization scheme for 500 iterations and a linear cooling schedule. On Fig. 1 we present the original spectra, and the recovered spectra. On Fig 2 we present the correlation of the real sources, the observation and the outputs of our algorithm.

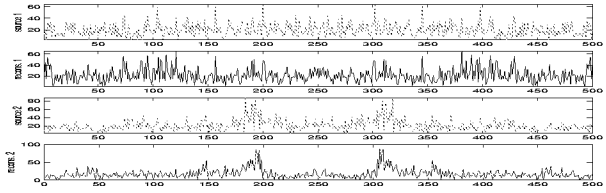

Figure 1: Spectra of the real sources, and recovered sources

7 CONCLUSION

In this paper we have introduced a Bayesian model to address the problem of separation and recovery of convolutively mixed autoregressive sources, corrupted by an additive white Gaussian noise. We present three stochastic algorithms, relying on MCMC techniques, that allow us to obtain samples from the posterior distribution of the model, and we show how they can be used to optimize

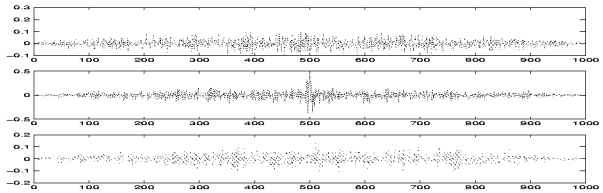

Figure 2: Correlation of the orginal sources, the observations and recovered sources

the posterior distribution of the sources conditional upon the observations or to optimize the set of parameters of the model conditional upon the observations, the sources being integrated out. Simulation results showed that in general source separation can be achieved in very general cases, as suggested in [9]. The main interest of our approach is that it will allow us to take into account more complex and reliable model for speech, such as autoregressive processes excited by impulsive noise, and to estimate all the unknown parameters, without any *ad hoc* tuning. Furthermore the number of sources can be more than the number of sensors. Note that the approach we developed also allows to address the problem of model selection of the order of the mixing filters and autoregressive processes.

8 REFERENCES

[1] C. Andrieu A. Doucet and S.J. Godsill, "Bayesian Blind Marginal Separation of Convolutively Mixed Discrete Sources", *Proc. Conf. IEEE NNSP'98*, Cambridge UK, Sept. 1998.

[2] C. Andrieu, A. Doucet, P. Duvaut, W.J. Fitzgerald and S.J. Godsill, *Simulation-Based Computational Methods for Statistical Signal Processing*, Tutorial Icassp'98, Seattle, May 98.

[3] J.M. Bernardo and A.F.M. Smith, *Bayesian Theory*, Wiley, 1994.

[4] P. DeJong and N. Shephard, "The Simulation Smoother for Time Series Models", *Biometrika*, vol. 82, no. 2, 1995, pp. 339-350.

[5] L. Devroye, *Non-Uniform Random Variate Generation*, Springer, New-York, 1986.

[6] P.J. Green, "Reversible Jump MCMC Computation and Bayesian Model Determination", *Biometrika*, vol. 82, pp. 711-732, 1995.

[7] A.F.M. Smith and G.O. Roberts, "Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods", *J. Roy. Stat. Soc. B*, vol. 55, no. 1, pp. 3-23, 1993.

[8] H. Broman, U. Lindgren, H. Sahlin and P. Stoica, "Source Separation: A TITO system identification approach", *Tech. rep.* CTH-TE-33, Department of Applied Electronics, Chalmers University of Technology, Sept. 1995, *Submitted to IEEE Trans. SP*.

[9] E. Weinstein, A.V. Oppenheim, M. Feder and J.R. Buck, "Iterative and Sequential Algorithms for Multisensor Signal Enhancement", *IEEE. Trans. SP*, Vol. 42, No. 4, April 1994.