SPARSE BLIND DECONVOLUTION ACCOUNTING FOR TIME-SHIFT AMBIGUITY

Christian Labat, Jérôme Idier

IRCCyN (CNRS UMR 6597)
ECN, 1 rue de la Noë, BP 92101, F44321 Nantes Cedex 3, France
`firstname.lastname@irccyn.ec-nantes.fr`

ABSTRACT

Our contribution deals with blind deconvolution of sparse spike trains. More precisely, we examine the problem in the Markov chain Monte-Carlo (MCMC) framework, where the unknown spike train is modeled as a Bernoulli-Gaussian process. In this context, we point out that time-shift and scale ambiguities jeopardize the robustness of basic MCMC methods, in quite a similar manner to the label switching effect studied by Stephens (2000) in mixture model identification. Finally, we propose proper modifications of the MCMC approach, in the same spirit as Stephens' contribution.

1. INTRODUCTION

The problem of the restoration of sparse spike trains distorted by a linear system and additive noise arises in many fields such as seismic exploration [1–3], and non-destructive evaluation [4]. It is classically dealt with a discrete-time noisy convolution model for the observations:

$$z = h \star x + \epsilon, \tag{1}$$

where h is the impulse response (IR) of the system (assumed finite here), x is the sparse spike train to be restored and ϵ is a stationary white Gaussian noise.

The deconvolution problem is said *blind* when the IR h is unknown. In the present study, we will assume that parameters such as the noise variance are also unknown. It is clear however that some statistical information must be available, at least to distinguish the input signal from the IR. Here, we adopt a Bernoulli-Gaussian process (BG) for x, following [1] and many posterior contributions such as [2–4]. Moreover, we adopt a Markov chain Monte-Carlo (MCMC) approach, akin to that of [2].

Blind deconvolution is a fundamentally information-deficient issue: since $h \star x$ is equal to $(f \star h) \star (f^{-1} \star x)$ for any invertible filter f, the solution of a blind deconvolution problem is not unique. Here, the BG prior helps to raise the main ambiguities, but there remain the following ones:

- Scale ambiguity: $h \star x = (ah) \star (x/a)$, $\forall a \neq 0$.
- Time-shift ambiguity: $h \star x = (d_\tau \star h) \star (d_{-\tau} \star x)$, $\forall \tau \in \mathbb{Z}$, where d_τ is the time delay filter of τ samples.

Such ambiguities must be taken into account. Otherwise, classical estimators for h and x such as posterior expectations (as typically approximated by MCMC computations) become meaningless, as averaged quantities over the variations of a and τ.

Whereas the scale ambiguity is rather easily raised by an arbitrary scaling, the time-shift ambiguity is more difficult to handle within the MCMC framework. In [2], it is simply circumvented by constraining the maximum of h to a prescribed position. However, as discussed in Section 5, such a solution is not always satisfying (see also Section 7).

The potential effect of time-shifts can be compared to the *label-switching* effect dealt by Stephens in [5] in the case of mixture model identification. It is our goal here to analyze the time-shift effect and to compensate for it, in the same spirit as Stephens' contribution.

The formulation of the blind deconvolution problem is presented in Section 2. Section 3 introduces the fully Bayesian framework. A Gibbs sampling scheme quite similar to that of [2] is proposed in Section 4. Sections 5 and 6 contain our main contributions: Section 5 examinates the time-shift problem, and Section 6 proposes an hybrid Gibbs sampler to compensate for it. The proposed method is compared to the method of [2] in Section 7, and conclusive remarks are made in Section 8.

2. PROBLEM STATEMENT

Let $z = \{z_1, z_2, \ldots, z_N\}$, $h = \{h_0, h_1 \ldots, h_P\}$, and $x = \{x_1, x_2 \ldots, x_M\}$. Here, we adopt a "zero boundary" condition: the input coefficients x_m are assumed to vanish for all $m < 1$ and $m > M$), so that $N = M + P$. The length P of the unknown wavelet is assumed to be available in the sequel.

3. BAYESIAN APPROACH

3.1. Prior laws

The unknown input signal x is modeled as a BG sequence:

$$q_m \sim \text{Bi}(\lambda), \quad (x_m \mid q_m) \sim \mathcal{N}(0, q_m \sigma_1^2), \tag{2}$$

where $\text{Bi}(\lambda)$ is the Bernoulli law of parameter λ, so that $p(q_m = 1) = \lambda$. For the sake of simple notations, "p" will indifferently denote probabilities, probability densities, and products of the two. Moreover, random variables will not be distinguished from their realizations. With such simplified notations, we have

$$p(q, x \mid \lambda, \sigma_1^2) = p(q \mid \lambda) \, p(x \mid q, \sigma_1^2)$$
$$= \lambda^L \, (1 - \lambda)^{M-L} \prod_{m, q_m=1} g(x_m; \sigma_1^2) \prod_{m, q_m=0} \delta(x_m),$$

where $L = \sum_{m=1}^{M} q_m$ and $g(\cdot \, ; \mathbf{A})$ stands for the zero-mean Gaussian density of covariance $\mathbf{A}$.

We also assume that $\boldsymbol{\epsilon}$ is a Gaussian white noise with unknown variance σ^2:

$$\boldsymbol{\epsilon} \sim \mathcal{N}\left(\mathbf{0}, \, \sigma^2 \mathbf{I}_N\right), \tag{3}$$

where $\mathbf{I}_N$ is the identity matrix of size N.

Within our fully Bayesian framework, prior laws are also needed for $\boldsymbol{h}$ and for the hyperparameters. Our choices are much similar to those found in [2]:

- $\boldsymbol{h}$ has a Gaussian prior of known variance:

$$\boldsymbol{h} \sim \mathcal{N}\left(\mathbf{0}, \, \sigma_h^2 \mathbf{I}_{P+1}\right). \tag{4}$$

- The noise variance σ^2 follows an inverse gamma prior $\mathrm{IG}(\nu, \eta)$ with known parameters $\nu, \eta > 0$.

- λ is assigned a beta law $\mathrm{Be}(a, b)$ with known parameters $a, b > 0$.

Finally, we assume that σ_1^2 is a known constant. For instance, we can take $\sigma_1^2 = 1$. This assumption is not restrictive because of the scale ambiguity

$$\boldsymbol{h} \star \boldsymbol{x} = (\pm \sigma_1^{-1} \boldsymbol{h}) \star (\pm \sigma_1 \boldsymbol{x}), \quad \forall \sigma_1 > 0.$$

3.2. Joint posterior law

Let $\boldsymbol{\theta} = (\boldsymbol{q}, \boldsymbol{x}, \boldsymbol{h}, \lambda, \sigma^2)$. Given our previous assumptions, the joint *posterior* probability reads [2]:

$$p(\boldsymbol{\theta} \,|\, \boldsymbol{z}) = g(\boldsymbol{z} - \boldsymbol{h} \star \boldsymbol{x} \, ; \, \sigma^2 \mathbf{I}_N) \, p(\boldsymbol{\theta}) / p(\boldsymbol{z}), \tag{5}$$

with $p(\boldsymbol{\theta}) = p(\boldsymbol{q} \,|\, \lambda) \, p(\boldsymbol{x} \,|\, \boldsymbol{q}) \, p(\lambda) \, p(\boldsymbol{h}) \, p(\sigma^2)$.

4. GIBBS SAMPLING

Using conditional posterior laws, the Gibbs sampler generates a Markov chain of random samples $\boldsymbol{\theta}^{(k)}$ whose equilibrium law coincides with the joint posterior law [6]. For the sake of simplicity, the iteration number k is omitted. The following conditional posterior laws can be deduced from (5) (see also [2]):

①　Let $\boldsymbol{r} = \{\boldsymbol{z}, \boldsymbol{\theta}\} \setminus \{q_m, x_m\}$. Then, $(x_m \,|\, \boldsymbol{r})$ follows a BG law with a nonzero mean:

$$(q_m \,|\, \boldsymbol{r}) \sim \mathrm{Bi}(\lambda_{1,m}), \quad (x_m \,|\, q_m, \boldsymbol{r}) \sim \mathcal{N}(\mu_{1,m}, q_m \sigma_{1,m}^2)$$

with

$$\lambda_{1,m} = \frac{\widetilde{\lambda}_{1,m}}{\widetilde{\lambda}_{1,m} + 1 - \lambda}, \quad \widetilde{\lambda}_{1,m} = \lambda \frac{\sigma_{1,m}}{\sigma_1} \exp\left(\frac{\mu_{1,m}^2}{2\sigma_{1,m}^2}\right),$$

$$\sigma_{1,m}^2 = \frac{\sigma^2 \sigma_1^2}{\sigma^2 + \sigma_1^2 \|\boldsymbol{h}\|^2}, \quad \mu_{1,m} = \frac{\sigma_{1,m}^2}{\sigma^2} \sum_{i=0}^{P} h_i e_{m+i},$$

where $e_n = (\boldsymbol{z} - \boldsymbol{h} \star \boldsymbol{x})_n + h_{n-m} x_m$.

②　Let $\boldsymbol{r} = \{\boldsymbol{z}, \boldsymbol{\theta}\} \setminus \{\boldsymbol{h}\}$. Then $(\boldsymbol{h} \,|\, \boldsymbol{r}) \sim \mathcal{N}(\boldsymbol{m}, \mathbf{R})$ with

$$\mathbf{R} = \left(\sigma^{-2} \mathbf{X}^{\mathrm{t}} \mathbf{X} + \sigma_h^{-2} \mathbf{I}_{P+1}\right)^{-1}, \quad \boldsymbol{m} = \sigma^{-2} \mathbf{R} \mathbf{X}^{\mathrm{t}} \boldsymbol{z}, \tag{6}$$

where $\mathbf{X}$ is the Toeplitz matrix of size $N \times (P+1)$ with first row $[x_1 \, \mathbf{0}_P]$ and first column $\left[\boldsymbol{x}^{\mathrm{t}} \, \mathbf{0}_P\right]^{\mathrm{t}}$.

③　Conditionally to all other variables, the law of σ^2 is $\mathrm{IG}(N/2 + \nu, \, \|\boldsymbol{z} - \boldsymbol{h} \star \boldsymbol{x}\|^2 \,/2 + \eta)$.

④　Conditionally to all other variables, the law of λ is $\mathrm{Be}\,(a + L, \, b + M - L)$, with $L = \sum_{m=1}^{M} q_m$.

The Gibbs sampler iterates steps ①-④, all of which corresponding to classical sampling operations.

5. DEALING WITH TIME-SHIFT AND SCALE AMBIGUITIES

5.1. Principle

In [2], it is proposed to raise the time-shift ambiguity by constraining the maximizer of $\boldsymbol{h}$ to a given position i^*:

$$\text{Time constraint [2]:} \qquad |h_{i^*}| = \max_i |h_i| \tag{7}$$

In practice, the sampling procedure is only slightly modified: Step ② is repeated until constraint (7) is met. The following constraint is also introduced to deal with scale ambiguity:

$$\text{Scale constraint [2]:} \qquad h_0 = 1. \tag{8}$$

The resulting posterior mean estimates can then be approximated by averages over the last $K - D$ samples:

$$\hat{\boldsymbol{\theta}} = \frac{1}{K - D} \sum_{k=D+1}^{K} \boldsymbol{\theta}^{(k)}$$

Enforcing conditions (7)-(8) is typical of the *identifiability constraint* approach, whose limitations are underlined in [5] in the context of label-switching.

In particular, the scale constraint $h_0 = 1$ is not always appropriate: if the true value of h_0 vanishes (or nearly so), such a constraint will be artificial and unsuited to impose a common dynamic to the samples $(\boldsymbol{h}^{(k)})$.

Similarly, condition (7) will be ineffective when the maximum magnitude $\max_i |h_i|$ of the true response is reached at several positions. There is another more specific drawback in imposing (7): a slight error in positioning i^* may yield severely degraded estimation results, as checked in Section 7. The reason is the following: in contrast with the perfect label-switching effect, where all label permutations are equally likely, different values of time-shifts do not yield perfectly equivalent solutions, because both $\boldsymbol{h}$ and $\boldsymbol{x}$ are defined as *finite length vectors*. In particular, the shape of the IR will no longer fit the time window $\{0, 1, \ldots, P\}$ after an arbitrary time-shift.

Following [5], we propose to get rid of constraints on the sampler. We rather cope with time ambiguities by shifting the samples $(\boldsymbol{h}^{(k)}, \boldsymbol{x}^{(k)})$ w.r.t. the time index, prior to computing averages. Scale ambiguities are dealt simultaneously, in the same spirit.

5.2. Scaling-shifting algorithm

Once a series of random samples $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(K)}$ is available, the question is to compute appropriate averages to estimate the unknown quantities. Usual averages are suited for scalar parameters λ and σ^2. Only the series $(\boldsymbol{h}^{(k)})$ and $(\boldsymbol{x}^{(k)})$ may be affected by time-shifts and scale fluctuations.

To compensate for label switching in mixture model identification, Stephens proposes a relabelling algorithm. The principle is to find the best permutation for each sample, in the sense of a well-chosen loss function to be minimized [5].

In the same spirit, we propose a scaling-shifting algorithm to remove the time and scale ambiguities in the series $(\boldsymbol{x}^{(k)})$ and $(\boldsymbol{h}^{(k)})$. Let the following loss function:

$$\mathcal{R}(\boldsymbol{h}, \boldsymbol{a}, \boldsymbol{\tau}) = \frac{1}{K-D} \sum_{k=D+1}^{K} \|\boldsymbol{h} - a_k \, d_{\tau_k} \star \boldsymbol{h}^{(k)}\|^2,$$

to be minimized w.r.t. $\boldsymbol{h}, \boldsymbol{a}, \boldsymbol{\tau}$. We are led to a combinatorial problem, for which we propose a suboptimal solution inspired from [5, Algorithm 4.1]. Each iteration of the following two-step scheme reduces $\mathcal{R}(\boldsymbol{h}, \boldsymbol{a}, \boldsymbol{\tau})$ until a fixed point is reached:

1. For all $k = D+1, \ldots, K$, choose $(\tilde{a}_k, \tilde{\tau}_k)$ as the minimizer of $\|\tilde{\boldsymbol{h}} - a_k \, d_{\tau_k} \star \boldsymbol{h}^{(k)}\|^2$ when $\tilde{\boldsymbol{h}}$ is held constant.

2. Choose $\tilde{\boldsymbol{h}}$ as the minimizer of $\mathcal{R}(\boldsymbol{h}, \tilde{\boldsymbol{a}}, \tilde{\boldsymbol{\tau}})$, *i.e.*,

$$\tilde{\boldsymbol{h}} = \frac{1}{K-D} \sum_{k=D+1}^{K} \tilde{a}_k \, d_{\tilde{\tau}_k} \star \boldsymbol{h}^{(k)}.$$

Step 1 yields $(\tilde{a}_k, \tilde{\tau}_k)$ as a matched filtering solution with reference to $\tilde{\boldsymbol{h}}$, while Step 2 updates $\tilde{\boldsymbol{h}}$ as the average of the scaled, shifted versions of $\boldsymbol{h}^{(k)}$. We propose to initialize the procedure by $\tilde{\boldsymbol{h}} = \boldsymbol{h}^{(K)}$. Convergence is observed after a few iterations.

The procedure applies to $(\boldsymbol{h}^{(k)})$, since our loss function is defined as a function of $\boldsymbol{h}$ only, but it also provides corrections for $(\boldsymbol{x}^{(k)})$, from which an estimated input signal is computed:

$$\tilde{\boldsymbol{x}} = \frac{1}{K-D} \sum_{k=D+1}^{K} \frac{1}{\tilde{a}_k} \, d_{-\tilde{\tau}_k} \star \boldsymbol{x}^{(k)}. \tag{9}$$

6. HYBRID GIBBS SAMPLING

6.1. Metropolis-Hastings within Gibbs sampling

In practice, removing constraints (7)-(8) in favor of our scaling-shifting scheme is not sufficient to provide a fully satisfying procedure. There is still an issue to deal with: the Gibbs sampling procedure of Section 4 scarcely produces any time-shifts. As a consequence, the estimates produced after the scaling-shifting procedure will be strongly influenced by the initialization of the Gibbs procedure: typically, the position of the maximum value of $\boldsymbol{h}^{(0)}$ nearly plays the role of i^* in constraint (7).

It is actually quite easy to cope with such a deficiency by slightly modifying the Gibbs sampler of Section 4 in order to stimulate time-shifts. Our proposition only affects Step ②. Instead of merely resampling $\boldsymbol{h}$ according to its posterior law, we envisage a Metropolis-Hastings (MH) procedure involving both $\boldsymbol{h}$ and $(\boldsymbol{q}, \boldsymbol{x})$. To simplify notations in the whole section, we drop dependence of probability terms on current parameter values $\lambda^{(k)}, \sigma^{(k)}$, and we use $\boldsymbol{\theta}$ as a shorthand for $(\boldsymbol{q}, \boldsymbol{x}, \boldsymbol{h})$. Let us define the proposal kernel of the MH step as $\pi(\boldsymbol{\theta}' \,|\, \boldsymbol{z}, \boldsymbol{\theta}) = p(\boldsymbol{h}' \,|\, \boldsymbol{q}', \boldsymbol{x}', \boldsymbol{z}) \, \pi(\boldsymbol{q}', \boldsymbol{x}' \,|\, \boldsymbol{q}, \boldsymbol{x})$, where

$$\pi(\boldsymbol{q}', \boldsymbol{x}' \,|\, \boldsymbol{q}, \boldsymbol{x}) = \begin{cases} 1 - 2\alpha & \text{if } (\boldsymbol{q}', \boldsymbol{x}') = (\boldsymbol{q}, \boldsymbol{x}), \\ \alpha & \text{if } (\boldsymbol{q}', \boldsymbol{x}') = (\mathbf{C}\boldsymbol{q}, \mathbf{C}^{-1}\boldsymbol{x}), \\ \alpha & \text{if } (\boldsymbol{q}', \boldsymbol{x}') = (\mathbf{C}^{-1}\boldsymbol{q}, \mathbf{C}\boldsymbol{x}), \end{cases}$$

with $\mathbf{C}$ defined as the right circular shift operator, and $\alpha \in (0, 1/2)$. With probability $1 - 2\alpha$, the MH procedure boils down to Step ②. Otherwise, a time-shift of ± 1 is proposed. Circular shifting provides a good trade-off between easy implementation and a fair acceptation probability. According to our "MH within Gibbs" sampler, Step ② is replaced by:

☐2 Propose $\boldsymbol{\theta}'$ according to $\pi(\boldsymbol{\theta}' \,|\, \boldsymbol{z}, \boldsymbol{\theta}^{(k)})$. Accept $\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}'$ with probability $\min\{1, \rho(\boldsymbol{z}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\theta}')\}$, where

$$\rho(\boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{\theta}') = \frac{\pi(\boldsymbol{\theta} \,|\, \boldsymbol{z}, \boldsymbol{\theta}')}{\pi(\boldsymbol{\theta}' \,|\, \boldsymbol{z}, \boldsymbol{\theta})} \frac{p(\boldsymbol{\theta}' \,|\, \boldsymbol{z})}{p(\boldsymbol{\theta} \,|\, \boldsymbol{z})}. \tag{10}$$

Otherwise, let $\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)}$.

Let us establish a simple expression for $\rho(\boldsymbol{z}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\theta}')$.

6.2. Acceptation probability

According to $\pi(\boldsymbol{q}', \boldsymbol{x}' \,|\, \boldsymbol{q}, \boldsymbol{x}) = \pi(\boldsymbol{q}, \boldsymbol{x} \,|\, \boldsymbol{q}', \boldsymbol{x}')$, (10) reads

$$\begin{aligned} \rho(\boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{\theta}') &= \frac{p(\boldsymbol{h} \,|\, \boldsymbol{q}, \boldsymbol{x}, \boldsymbol{z})}{p(\boldsymbol{h}' \,|\, \boldsymbol{q}', \boldsymbol{x}', \boldsymbol{z})} \frac{p(\boldsymbol{h}' \,|\, \boldsymbol{q}', \boldsymbol{x}', \boldsymbol{z})}{p(\boldsymbol{h} \,|\, \boldsymbol{q}, \boldsymbol{x}, \boldsymbol{z})} \frac{p(\boldsymbol{q}', \boldsymbol{x}' \,|\, \boldsymbol{z})}{p(\boldsymbol{q}, \boldsymbol{x} \,|\, \boldsymbol{z})} \\ &= \frac{p(\boldsymbol{q}', \boldsymbol{x}' \,|\, \boldsymbol{z})}{p(\boldsymbol{q}, \boldsymbol{x} \,|\, \boldsymbol{z})}. \end{aligned}$$

Then, according to $p(\boldsymbol{q}', \boldsymbol{x}') = p(\boldsymbol{q}, \boldsymbol{x})$, (10) also reads

$$\rho(\boldsymbol{\theta}, \boldsymbol{\theta}', \boldsymbol{z}) = \frac{p(\boldsymbol{z} \,|\, \boldsymbol{q}', \boldsymbol{x}')}{p(\boldsymbol{z} \,|\, \boldsymbol{q}, \boldsymbol{x})} = \frac{p(\boldsymbol{z} \,|\, \boldsymbol{x}')}{p(\boldsymbol{z} \,|\, \boldsymbol{x})}, \tag{11}$$

which shows that $\rho(\boldsymbol{\theta}, \boldsymbol{\theta}', \boldsymbol{z})$ depends neither on $\boldsymbol{h}$ nor on $\boldsymbol{h}'$. Moreover, $\boldsymbol{x}' = \boldsymbol{x}$ implies $\rho(\boldsymbol{\theta}, \boldsymbol{\theta}', \boldsymbol{z}) = 1$, as expected.

In all cases, it is easy to establish that $\rho(\boldsymbol{\theta}, \boldsymbol{\theta}', \boldsymbol{z})$ is the ratio of two Gaussian densities. More precisely, it can be deduced from (1), (4) and (3) that $(\boldsymbol{z} \,|\, \boldsymbol{x}) \sim \mathcal{N}(\mathbf{0}, \mathbf{P}^{-1})$, with

$$\mathbf{P} = (\sigma^2 \mathbf{I}_N + \sigma_h^2 \mathbf{X}\mathbf{X}^{\mathrm{t}})^{-1} = \sigma^{-2}\mathbf{I}_N - \sigma^{-4}\mathbf{X}\mathbf{R}\mathbf{X}^{\mathrm{t}}.$$

given the matrix inversion lemma. Hence,

$$\begin{aligned} 2 \ln \rho(\boldsymbol{\theta}, \boldsymbol{\theta}', \boldsymbol{z}) &= \boldsymbol{z}^{\mathrm{t}}(\mathbf{P} - \mathbf{P}')\boldsymbol{z} + \ln |\mathbf{P}^{-1}\mathbf{P}'| \\ &= (\boldsymbol{m}')^{\mathrm{t}}(\mathbf{R}')^{-1}\boldsymbol{m}' - \boldsymbol{m}^{\mathrm{t}}\mathbf{R}^{-1}\boldsymbol{m} + \ln |\mathbf{R}^{-1}\mathbf{R}'|, \end{aligned}$$

where we make use of the matrix inversion lemma again. The latter expression can be evaluated in $O(P^2)$ operations given that matrices $\mathbf{R}, \mathbf{R}'$ are Toeplitz.

7. SIMULATION RESULTS

A simulation is proposed to compare the original MCMC method of [2] with our modified approach incorporating the MH step and the scaling-shifting procedure. It is based on an example found in [2]. An IR of order $P = 40$ is defined by $\boldsymbol{h} = \boldsymbol{h}^*/h_0^*$, with $h_i^* = \sin(\pi(i+1)/6.4)\exp(-0.12\,|i-9|)$. The input signal $\boldsymbol{x}$ is generated from a BG law (2) with $\lambda = 0.1$ and $\sigma_1^2 = 4$. The observed signal is obtained from (1) for $N = 2000$ and $\sigma^2 = 1$, which corresponds to a signal-to-noise ratio of 18.6dB. The parameter values for the priors are taken from [2]:

$$\boldsymbol{h} \sim \mathcal{N}(\mathbf{0}, 100\,\mathbf{I}_{P+1}), \quad \lambda \sim \text{Be}(10, 50), \quad \sigma^2 \sim \text{IG}(1, 0.3),$$

as well as the initial values: $\sigma^{(0)} = 1$, $\boldsymbol{x}^{(0)} = \mathbf{0}$, $\lambda^{(0)} = 0.1$. Both sampling schemes were carried out for 4000 iterations,

and only the last 1000 samples were considered to compute estimates.

The robustness of the method presented in [2] is tested by assuming a slight error in the position of the maximizer $i^* = \arg\max_i |h_i|$: $i^* = 6$ is considered instead of $i^* = 9$. Accordingly, $\boldsymbol{h}^{(0)}$ is chosen as a Kronecker function at $i = 6$, instead of $i = 9$. Figure 1(a) shows that the left part of the resulting estimate $\hat{\boldsymbol{h}}$ is slightly altered, but the consequence on the estimated input $\hat{\boldsymbol{x}}$ is a more severe degradation as shown in Figure 1(b). The corresponding estimated value of λ is $\hat{\lambda} = 0.18$, which significantly differs from the true $\lambda = 0.1$, while the noise variance is rather well estimated $\hat{\sigma}^2 = 1.13$.
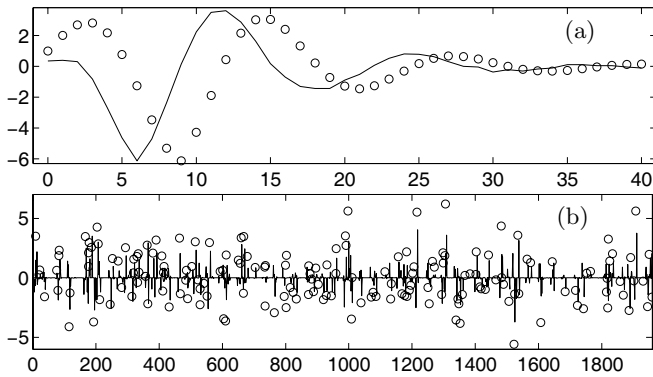

Fig. 1. Application of the standard MCMC method found in [2], tested with an error of -3 time samples on the position of $i^* = \arg\max_i |h_i|$. The circles indicate true values. (a) Estimated IR $\hat{\boldsymbol{h}}$ after proper scaling. (b) Estimated BG $\hat{\boldsymbol{x}}$ signal after proper scaling.

When tested in the same conditions, our modified MCMC method clearly shows a better robustness according to Figure 2. The estimated value of λ is now $\tilde{\lambda} = 0.099$ while the noise variance is still well estimated $\tilde{\sigma}^2 = 1.07$. Table 1 provides additional information about the frequency of accepted shifts within the MH step of Subsection 6.1. During the first 200 iterations, nearly 20% of right shifts are accepted, because the algorithm compensates for the initialization of the IR as
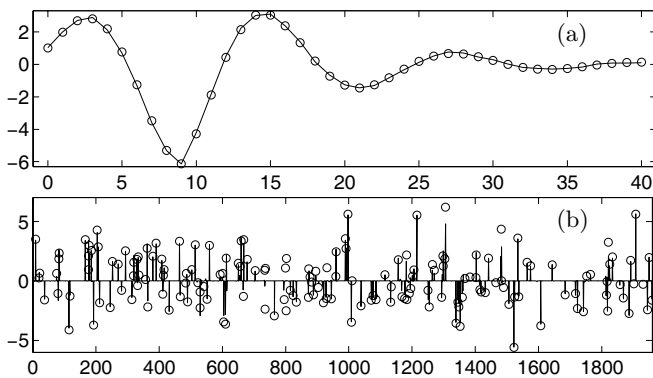

Fig. 2. Proposed modified MCMC method in the same conditions as in Figure 1, with $\alpha = 0.1$. The circles indicate true values. (a) Estimated IR $\tilde{\boldsymbol{h}}$ after proper scaling. (b) Estimated BG $\tilde{\boldsymbol{x}}$ signal after proper scaling.

a Kronecker function at the left of the time domain. Then the proportion of accepted left and right shifts gets balanced, at about 7%. After the burn-in, the proportion is below 4%, which does not mean that our MH procedure is not efficient, but rather that the correct position of the IR is significantly more probable than the others.

Sample index $k \in$	$[1, 200]$	$[201, 3000]$	$[3001, 4000]$
# of proposed right shifts	22	267	76
# of accepted right shifts	4	20	3
# of proposed left shifts	9	277	98
# of accepted left shifts	1	19	3

Table 1. Number of proposed and accepted shifts at different stages of the MH algorithm.

8. DISCUSSION

In this paper, we have pointed out that time and scale ambiguities jeopardize the robustness of basic MCMC methods applied to sparse blind deconvolution problem. We have established a formal link between this issue and the label switching effect studied by Stephens in [5]. We have proposed to introduce some modifications in the light of Stephens' contribution. The proposed method is based on an "MH within Gibbs" sampler, and estimation of the unknowns are only obtained after an operation of scaling-shifting on the generated samples. The additional cost is negligible compared to a more standard application of the MCMC approach, as found in [2], while it is no more necessary to assume that the position of the maximum value of the IR is known in advance.

Our main perspective is to consider an IR $\boldsymbol{h}$ of unknown length. In our opinion, the present contribution is a prerequisite step towards estimating the length of $\boldsymbol{h}$, since it provides a robust way of letting the IR make best use of the alloted time window. Different window lengths could then be explored, for instance using a reversible jump procedure.

9. REFERENCES

[1] J. M. Mendel, *Optimal Seismic Deconvolution*, Academic Press, New York, NY, 1983.

[2] Q. Cheng, R. Chen, and T.-H. Li, "Simultaneous wavelet estimation and deconvolution of reflection seismic signals", *IEEE Trans. Geosci. Remote Sensing*, vol. 34, pp. 377–384, Mar. 1996.

[3] O. Rosec, J. Boucher, B. Nziri, and T. Chonavel, "Blind marine seismic deconvolution using statistical mcmc methods", *IEEE oceanic engineering*, vol. 28, pp. 502–512, July 2003.

[4] F. Champagnat and J. Idier, "Deconvolution of sparse spike trains accounting for wavelet phase shifts and colored noise", in *Proc. IEEE ICASSP*, Minneapolis, MN, 1993, pp. 452–455.

[5] M. Stephens, "Dealing with label-switching in mixture models", *J. R. Statist. Soc. B*, vol. 62, pp. 795–809, 2000.

[6] C. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer Texts in Statistics. Springer Verlag, New York, NY, 2nd edition, 2004.