A NEW MCMC ALGORITHM FOR BLIND BERNOULLI-GAUSSIAN DECONVOLUTION

Di Ge, Jérôme Idier, Eric Le Carpentier

IRCCyN (CNRS UMR 6597)
1, rue de la Noë, BP 92101, F44321 Nantes Cedex 3, France
(Di.Ge|Jerome.Idier|Eric.Le-Carpentier)@irccyn.ec-nantes.fr

ABSTRACT

This paper proposes a new algorithm for Bernoulli-Gaussian (BG) blind deconvolution in the Markov chain Monte Carlo (MCMC) framework. To tackle such a problem, the classical Gibbs sampler is usually adopted, as proposed by Cheng *et al.* [1]. However, as already pointed out by Bourguignon and Carfantan [2], it fails to explore the state space efficiently. In principle, a more efficient exploration technique could be obtained by integrating the Gaussian amplitudes out of the target distribution. Unfortunately, some of the sampling steps then become intractable. Therefore, our solution mixes steps in which the amplitudes are integrated out with others where they are not. The *invariant condition* is shown to hold, and simulations indicate that it behaves much more satisfactorily than the reference Gibbs sampler.

Index Terms— Blind deconvolution, Bernoulli-Gaussian model, Markov chain Monte Carlo methods

1. INTRODUCTION

The problem of the restoration of a sparse spike train distorted by a linear system and noise arises in many fields such as seismic exploration [1, 3]. It is classically dealt with using a discrete-time noisy convolution model for the observed vector $\boldsymbol{z} = [z_1, \ldots, z_N]^{\mathrm{t}}$:

$$z_n = \sum_{k=0}^{P} h_k x_{n-k} + \varepsilon_n, \forall n = 1, \ldots, N. \qquad (1)$$

$\boldsymbol{h} = [h_0, \ldots, h_P]^{\mathrm{t}}$ denotes the impulse response (IR) of the system (assumed finite here). For the sake of simplicity, the "zero boundary" condition is imposed on the sparse spike train to be restored, such that $\boldsymbol{x} = [x_1, \ldots, x_M]^{\mathrm{t}}$, $M = N - P$. $\boldsymbol{\varepsilon}$ is a stationary white Gaussian noise. The deconvolution problem is said *blind* when $\boldsymbol{h}$ is unknown.

In the present study, we adopt a BG model for the spike train $\boldsymbol{x}$, following [3] and many posterior contributions such as [1, 4]. A BG signal is defined in two stages involving a Bernoulli sequence $\boldsymbol{q} = [q_1, \ldots, q_M]^{\mathrm{t}}$, such that: $\forall m = 1, \ldots, M$.

$$q_m \sim Bi(\lambda), \quad (x_m \mid q_m) \sim \mathcal{N}(0, q_m \sigma_x^2). \qquad (2)$$

The sparse nature of the spikes is governed by the Bernoulli law, while amplitudes are assumed iid zero-mean Gaussian. The MCMC approach [5, 6] is a powerful numerical tool, appropriate to solve complex inference problems such as blind deconvolution. In the field of blind BG deconvolution, Cheng *et al.* pioneered the introduction of MCMC methods [1]. They proposed to rely on a Gibbs sampler, for which their algorithm constitutes a simple and canonical example of the application. However, our simulation results indicate that it lacks reliability: from different initial conditions, significantly different estimations are obtained, even after considerable iterations. This conclusion agrees with that of Bourguignon *et al.* in the context of BG spectral analysis [2].

The recent contribution of [7] already identified a convergence issue linked to time-shift ambiguities, and proposed an efficient way to solve it. There exists another source of inefficiency, unrelated to the time-shift ambiguity: instead of exploring the 2^M configurations at an acceptable speed, the Gibbs sampler tends to get stuck in one particular configuration of $\boldsymbol{q}$ or another, as shown in Sect. 2.2. In fact, in a multi-modal space a Markov chain equilibrates rapidly within a mode (a configuration of $\boldsymbol{q}$), but takes a long time to move from mode to mode. Guan [8] studied a comparable slow convergence phenomenon of a *local* chain in comparison with his *small world* MCMC approach.

The new Gibbs-type sampler proposed here explores the posterior distribution *marginally* to the amplitudes $\boldsymbol{x}$, such that it probes the Bernoulli sequence $\boldsymbol{q}$ more freely in its space than Cheng *et al.*'s Gibbs sampler. However, a plain Gibbs sampler of the marginal posterior distribution involves hardly implementable sampling steps. In particular, it is all but simple to simulate $\boldsymbol{h}$ conditional on $\boldsymbol{z}$ and $\boldsymbol{q}$ and marginally w.r.t. $\boldsymbol{x}$. Our scheme solves this problem by incorporating steps where the sampling of $\boldsymbol{x}$ is still involved, while $\boldsymbol{q}$ is sampled marginally w.r.t. $\boldsymbol{x}$. Therefore, it is a *Gibbs-type sampler*, though fully valid from the mathematical viewpoint, since the invariant condition is shown to hold. A comparable idea is found in statistical signal segmentation [9] where some hyperparameters are partially marginalized, though without mathematical justification. Finally, for the sake of conciseness, it will remain implicit that the time-shift ambiguities are dealt with according to [7] in Cheng *et al.*'s Gibbs sampler, as well as in the proposed sampler.

In Sect. 2, the blind BG deconvolution problem is formulated. The Gibbs sampler of the joint posterior distribution [1] is also presented, and an example illustrates its inefficiency. Our Gibbs-type sampler is introduced in Sect. 3 and a simulated example gives an insight into the way it escapes from suboptimal configurations. The invariant condition is shown to hold, and an adapted implementation inspired from [4] is proposed. Finally, simulation results are presented in Sect. 4

2. PROBLEM FORMULATION

2.1 Statistical model

Akin to [1], the following assumptions are made:

- $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2 \mathbf{I})$ is independent of $\boldsymbol{x}$ and $\boldsymbol{h}$;
- $\boldsymbol{x}$ is a BG process defined by (2) with $\sigma_{\boldsymbol{x}} = 1$;
- $\boldsymbol{h} \sim \mathcal{N}(0, \sigma_h^2 \mathbf{I}_{P+1})$;

$\sigma_{\boldsymbol{x}}$ is arbitrarily set to one in order to remove the scale ambiguity inherent to the blind deconvolution problem. According to the Monte Carlo principle, a posterior mean estimator of $\Theta = \{\boldsymbol{q}, \boldsymbol{x}, \boldsymbol{h}, \lambda, \sigma_\varepsilon, \sigma_{\boldsymbol{h}}\}$ given $\boldsymbol{z}$ can be approximated by:

$$\widehat{\Theta} = \frac{1}{I-J} \sum_{k=J+1}^{I} \Theta^{(k)}, \tag{3}$$

where the sum extends over the last $I-J$ samples. In the MCMC framework, the samples are generated recursively, so that the asymptotic distribution of $\Theta^{(k)}$ is the joint posterior distribution [6]:

$$\begin{aligned} \rho(\Theta \,|\, \boldsymbol{z}) \propto\ & g(\boldsymbol{z} - \boldsymbol{x} \star \boldsymbol{h}; \sigma_\varepsilon^2 \mathbf{I}_N)\, g(\boldsymbol{x}; \sigma_{\boldsymbol{x}}^2 \mathrm{diag}\{\boldsymbol{q}\}) \\ & g(\boldsymbol{h}; \sigma_h^2 \mathbf{I}_{P+1})\, p(\boldsymbol{q}; \lambda)\, p(\sigma_{\boldsymbol{h}})\, p(\lambda)\, p(\sigma_\varepsilon) \end{aligned} \tag{4}$$

where $\mathrm{diag}\{\boldsymbol{q}\}$ denotes a diagonal matrix whose diagonal is $\boldsymbol{q}$, $g(\cdot; \mathbf{R})$ the centered Gaussian pdf of covariance $\mathbf{R}$, and $\star$ the convolution operator as defined in Eq. (1).

2.2 Classical Gibbs Sampling

Cheng *et al.*'s Gibbs sampler proceeds along Tab. 1.

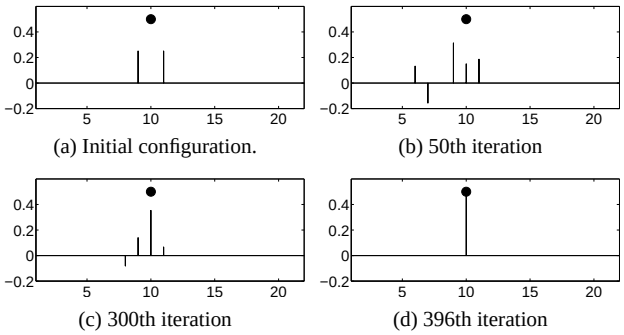
① Let $\boldsymbol{y} = (\boldsymbol{q}, \boldsymbol{x})$. For each $i = 1 \ldots, M$,
(a) draw $q_i^{(k+1)} \,\vert\, \boldsymbol{y}_{1:i-1}^{(k+1)}, \boldsymbol{y}_{i+1:M}^{(k)}, \boldsymbol{h}^{(k)}, \sigma_\varepsilon^{(k)}, \lambda^{(k)}, \boldsymbol{z}$
(b) draw $x_i^{(k+1)} \,\vert\, \boldsymbol{y}_{1:i-1}^{(k+1)}, \boldsymbol{y}_{i+1:M}^{(k)}, q_i^{(k+1)}, \boldsymbol{h}^{(k)}, \sigma_\varepsilon^{(k)}, \boldsymbol{z}$
② draw $\boldsymbol{h}^{(k+1)} \,\vert\, \boldsymbol{x}^{(k+1)}, \sigma_{\boldsymbol{h}}^{(k)}, \sigma_\varepsilon^{(k)}, \boldsymbol{z}$
③ draw $\sigma_\varepsilon^{(k+1)} \,\vert\, \boldsymbol{x}^{(k+1)}, \boldsymbol{h}^{(k+1)}, \boldsymbol{z}$
④ draw $\lambda^{(k+1)} \,\vert\, \boldsymbol{q}^{(k+1)}$
⑤ draw $\sigma_{\boldsymbol{h}}^{(k+1)} \,\vert\, \boldsymbol{h}^{(k+1)}$

Table 1. Cheng *et al.*'s Gibbs sampler (see [1] for implementation details).

The five simulation steps are iterated until convergence towards the posterior distribution, and $\widehat{\Theta}$ is finally built according to (3). This scheme presents several drawbacks despite its simplicity. Labat and Idier [7] have already pointed out that time-shift ambiguities could lead to unreliable estimates, depending on the initialization of $\boldsymbol{h}$. Here, we rather focus on the fact that step 1 explores the state space of $(\boldsymbol{q}, \boldsymbol{x})$ with a low efficiency. More precisely, the correspondingly sampled chains hardly escape from local maximizers of the posterior likelihood, due to highly dependent consecutive samples. Fig. 1 illustrates this phenomenon, in which simulation data are generated from a single spike, convolved with the IR defined by

$$h_i = \cos\left((i-10)\frac{\pi}{4}\right) \exp\left(-|0.225i - 2|^{1.5}\right),\ i = 0, \ldots, 20,$$

and plotted on Fig. 2(b). A suboptimal configuration is chosen as the initial state. The chain takes several hundreds of iterations to visit the optimal configuration for the first time.



(a) Initial configuration. (b) 50th iteration

(c) 300th iteration (d) 396th iteration

Fig. 1. Sampled BG sequences obtained by Cheng *et al.*'s sampler on a simple example. From a suboptimal configuration chosen as initial state, the Markov chain spends several hundreds of iterations before visiting the solution, *i.e.*, a unique spike in position 10 (marked as a bullet).

As a consequence, Cheng *et al.*'s method tends to produce unreliable estimated values, as shown in Fig. 2. The same IR is adopted here while $\boldsymbol{x}$ is the well-known Mendel's test sequence [3]. The data are corrupted by a Gaussian noise with $\sigma_\varepsilon^2 = 4 \times 10^{-6}$, corresponding to SNR $= 12.81$ dB. Non-informative conjugate prior laws on the parameters are adopted:

$$\sigma_\varepsilon^2 \sim IG(1,1), \quad \sigma_{\boldsymbol{h}}^2 \sim IG(1,1), \quad \lambda \sim Be(1,1).$$

Initialization of the Markov chain $\Theta^{(0)}$ is done according to:

$$\sigma_\varepsilon^{(0)} = 10^{-2},\ \boldsymbol{q}^{(0)} = \mathbf{0},\ \boldsymbol{h}^{(0)} = \frac{\|\boldsymbol{z}\|_1}{M} \delta(n-11),\ \lambda^{(0)} = 0.1.$$

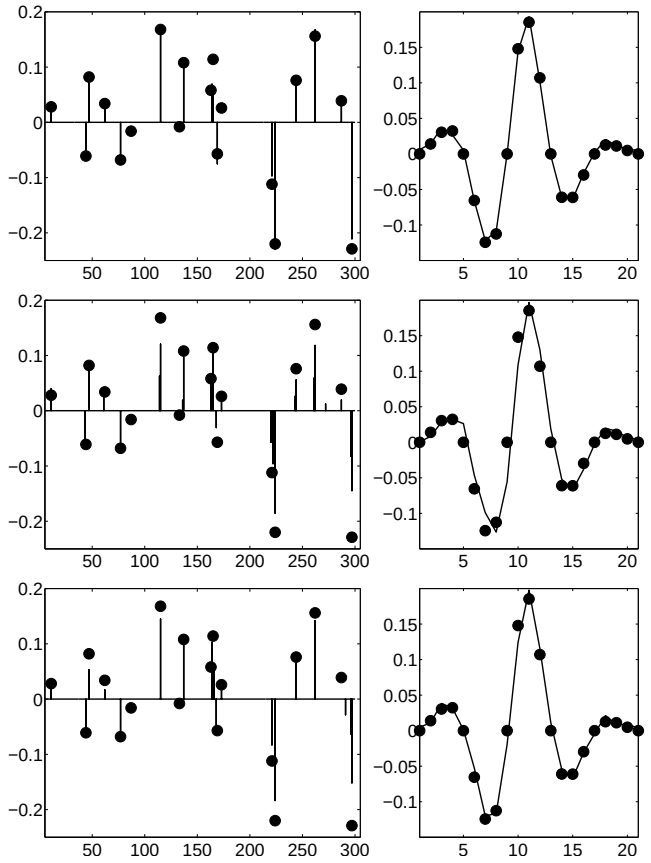
The three estimation results in Fig. 2 are obtained from the same simulated data $\boldsymbol{z}$, and the same initialization, the only difference being the random seed value. 1000 samples are produced for each Markov chain and the last 250 are averaged to compute the estimation. Substantial variations exist from one estimated result to the other, especially in the number and the positions of spikes $\widehat{\boldsymbol{x}}$. Actually, it can be checked that each sequence $\{\boldsymbol{q}^{(k)}\}$ tends to become constant for very long period of time, without fully exploring the state space. It should be stressed that these results are typical, *i.e.*, not selected on purpose.

The inefficiency of Cheng *et al.*'s Gibbs sampling has already been noticed by Bourguignon *et al.* [2]. Their solution involves proposals to shift several adjacent BG components, very similar to the idea introduced by Chi and Mendel [10] in the context of deterministic posterior likelihood maximization. Our approach, however, aims to tackle the problem indirectly by marginalizing the amplitudes $\boldsymbol{x}$ out of the target posterior distribution, and by so doing releases the Markov chain from local maxima of the likelihood.

3. TOWARDS A MORE EFFICIENT SAMPLER

3.1 Partial marginalization of $\boldsymbol{x}$

The motivation to exclude $\boldsymbol{x}$ from the Gibbs sampler is twofold. Firstly, the determination of $\boldsymbol{x}$ conditional on $(\widetilde{\Theta}, \boldsymbol{z})$ with $\widetilde{\Theta} = (\boldsymbol{q}, \boldsymbol{h}, \sigma_\varepsilon, \sigma_h, \lambda)$ is a linear estimation, easy to solve afterwards. Secondly, a Gibbs sampler with $\rho(\widetilde{\Theta} \,|\, \boldsymbol{z})$ as its



(a) estimation of $\boldsymbol{x}$ (b) estimation of $\boldsymbol{h}$

Fig. 2. Three different solutions of Cheng *et al.*'s method (under the improved version of [7]), obtained by changing the random seed value only. The actual values are marked as bullets.

target distribution is likely to be more efficient than Cheng *et al.*'s sampler, particularly w.r.t. to the Bernoulli sequence $\boldsymbol{q}$. Theoretical foundations are available in [11] and [5, Chapter 6.7] regarding the convergence rate of the so called *collapsed* Gibbs sampler, that converges to the marginalized distribution $\rho(\widetilde{\Theta}\,|\,\boldsymbol{z}) = \int \rho(\Theta\,|\,\boldsymbol{z})d\boldsymbol{x}$ by analytically integrating $\boldsymbol{x}$ out of the scheme. According to [5, Chapter 6.7], the collapsed Gibbs sampler produces a substantial gain in terms of convergence rate if the integrated variables form highly dependent pairs with others in the Markov chain (see also [12] and references therein). This is exactly the case of the pair $\{\boldsymbol{x},\boldsymbol{q}\}$ in the spike train deconvolution problem.

3.2 Proposed Gibbs-type sampler

Unfortunately, marginalizing $\boldsymbol{x}$ leads to practical difficulties. In particular, the conditional sampling of $\boldsymbol{h}$ given $\widetilde{\Theta}\backslash\boldsymbol{h}$ becomes extremely difficult when $\boldsymbol{x}$ is integrated out, since $\rho(\boldsymbol{h}\,|\,\boldsymbol{q},\sigma_{\boldsymbol{h}},\sigma_\varepsilon,\boldsymbol{z})$ is a multivariate, non Gaussian law with a complex structure. Instead of a plain Gibbs sampler on $\widetilde{\Theta}$, the sampling scheme of Tab. 2 is proposed to circumvent the direct conditional sampling of $\boldsymbol{h},\sigma_\varepsilon \in \widetilde{\Theta}$. The four steps are iterated until convergence towards the posterior distribution $\rho(\widetilde{\Theta}\,|\,\boldsymbol{z})$. Compared with the classical sampler, the main difference appears in step 1, where $\boldsymbol{x}$ has been analytically

① for each $i = 1\ldots,M$,
draw $q_i^{(k+1)}\,|\,\boldsymbol{q}_{1:i-1}^{(k+1)},\boldsymbol{q}_{i+1:M}^{(k)},\boldsymbol{h}^{(k)},\sigma_\varepsilon^{(k)},\lambda^{(k)},\boldsymbol{z}$
② (a) draw $\boldsymbol{x}\,|\,\boldsymbol{q}^{(k+1)},\boldsymbol{h}^{(k)},\sigma_\varepsilon^{(k)},\boldsymbol{z}$
(b) draw $\boldsymbol{h}^{(k+1)}\,|\,\boldsymbol{x},\sigma_{\boldsymbol{h}}^{(k)},\sigma_\varepsilon^{(k)},\boldsymbol{z}$
(c) draw $\sigma_\varepsilon^{(k+1)}\,|\,\boldsymbol{x},\boldsymbol{h}^{(k+1)},\boldsymbol{z}$
③ draw $\lambda^{(k+1)}\,|\,\boldsymbol{q}^{(k+1)}$
④ draw $\sigma_{\boldsymbol{h}}^{(k+1)}\,|\,\boldsymbol{h}^{(k+1)}$

Table 2. The proposed Gibbs-type sampler.

integrated out. Subsect. 3.3 provides an efficient way to implement step 1. On the other hand, in step 2, $\boldsymbol{x}$ only plays the role of an auxiliary quantity, since it does not belong to the Markov chain anymore. Steps 2(b), (c) are structurally identical to steps 2, 3 in Tab. 1. In order to justify that the sampling scheme of Tab. 2 is mathematically valid, we let $\widetilde{\Theta}' = (\boldsymbol{q},\boldsymbol{h}',\sigma_\varepsilon',\sigma_h,\lambda)$ and check the *invariant condition*, *i.e.*, that

$$\int_{\boldsymbol{h},\sigma_\varepsilon} \kappa_2(\boldsymbol{h}',\sigma_\varepsilon'\,|\,\widetilde{\Theta},\boldsymbol{z})\,\rho(\widetilde{\Theta}\,|\,\boldsymbol{z})\,d\boldsymbol{h}d\sigma_\varepsilon = \rho(\widetilde{\Theta}'\,|\,\boldsymbol{z}) \qquad (5)$$

holds for step 2, where ρ and κ_2 denote the invariant target distribution and the transition kernel corresponding to the step 2, respectively. We actually have

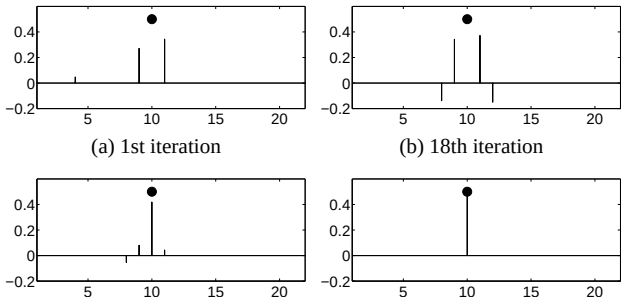
$$\kappa_2(\boldsymbol{h}',\sigma_\varepsilon'\,|\,\widetilde{\Theta},\boldsymbol{z}) = \int_{\boldsymbol{x}} \rho(\boldsymbol{x}\,|\,\widetilde{\Theta},\boldsymbol{z})\,\rho(\boldsymbol{h}'\,|\,\boldsymbol{x},\widetilde{\Theta}\backslash\boldsymbol{h},\boldsymbol{z})\,\rho(\sigma_\varepsilon'\,|\,\boldsymbol{x},\widetilde{\Theta}'\backslash\sigma_\varepsilon',\boldsymbol{z})d\boldsymbol{x}. \quad (6)$$

We emphasize that Eq. (6) is not obtained by Monte Carlo integration. It is rather the exact expression of the transition kernel of steps 2 to update the pair $\{\boldsymbol{h},\sigma_\varepsilon\}$ while $\boldsymbol{x}$ is no longer retained in the chain. It follows that the left handside of Eq. (5) reads

$$\begin{aligned}
&\int \rho(\boldsymbol{x},\widetilde{\Theta}\,|\,\boldsymbol{z})\rho(\boldsymbol{h}'\,|\,\boldsymbol{x},\widetilde{\Theta}\backslash\boldsymbol{h},\boldsymbol{z})\rho(\sigma_\varepsilon'\,|\,\boldsymbol{x},\widetilde{\Theta}'\backslash\sigma_\varepsilon',\boldsymbol{z})d\boldsymbol{h}d\sigma_\varepsilon d\boldsymbol{x}\\
&= \int \rho(\boldsymbol{x},\widetilde{\Theta}\backslash\boldsymbol{h}|\boldsymbol{z})\rho(\boldsymbol{h}'|\boldsymbol{x},\widetilde{\Theta}\backslash\boldsymbol{h},\boldsymbol{z})\rho(\sigma_\varepsilon'|\boldsymbol{x},\widetilde{\Theta}'\backslash\sigma_\varepsilon',\boldsymbol{z})d\sigma_\varepsilon d\boldsymbol{x}\\
&= \int \rho(\boldsymbol{h}',\boldsymbol{x},\widetilde{\Theta}\backslash\boldsymbol{h}\,|\,\boldsymbol{z})\rho(\sigma_\varepsilon'\,|\,\boldsymbol{x},\widetilde{\Theta}'\backslash\sigma_\varepsilon',\boldsymbol{z})d\sigma_\varepsilon d\boldsymbol{x}\\
&= \int \rho(\boldsymbol{x},\widetilde{\Theta}'\backslash\sigma_\varepsilon'\,|\,\boldsymbol{z})\rho(\sigma_\varepsilon'\,|\,\boldsymbol{x},\widetilde{\Theta}'\backslash\sigma_\varepsilon',\boldsymbol{z})d\boldsymbol{x}\\
&= \int \rho(\boldsymbol{x},\widetilde{\Theta}'\,|\,\boldsymbol{z})d\boldsymbol{x} = \rho(\widetilde{\Theta}'\,|\,\boldsymbol{z}).
\end{aligned}$$

It is thus proven that the partial marginalization technique can be applied to generate a Markov chain $\{\widetilde{\Theta}^{(k)}\}$ that converges to its equilibrium distribution $\rho(\widetilde{\Theta}\,|\,\boldsymbol{z})$.

To compare with the example of Fig. 1, our Gibbs-type scheme escapes from a local maximum configuration within an acceptable number of iterations, as Fig. 3 illustrates. The real configuration is reached after 20 iterations. Moreover, it is observed from the configurations obtained at the 18th and 19th iterations that our scheme is able to radically modify the amplitude vector $\boldsymbol{x}$ in one single step.



(a) 1st iteration (b) 18th iteration

(c) 19th iteration (d) 20th iteration

Fig. 3. Sampled BG sequences obtained by our Gibbs-type scheme in the example of Fig. 1. The Markov chain escapes rapidly the initial configuration neighborhood.

Finally, it is worthwhile to mention that an efficient sampler should meet the two conflicting criteria [5, Chapter 7]:

- drawing each component conditional on others should be computationally simple (complexity per iteration);
- the induced Markov chain should converge reasonably fast to its equilibrium law (number of iterations).

The next section deals with the issue of the complexity per iteration, whereas the overall computational costs are compared in Sect. 4 using simulation tests.

3.3 Marginal posterior distribution

In this subsection, an efficient implementation of step 1 of Tab. 2 is proposed, inspired from the recursive method in [4] to evaluate $\rho(q_i = 0,1|\widetilde{\Theta}\backslash q_i; \boldsymbol{z})$ sequentially. While the latter is based on the storage and update of an N_e^2 ($N_\mathrm{e} = \sum_i q_i$) matrix, we finally propose an even less burdensome strategy based on the handling of its Cholesky factor only.

Let $\mathbf{H}$ denote the $N \times M$ Toeplitz matrix such that (1) also reads $\boldsymbol{z} = \mathbf{H}\boldsymbol{x} + \boldsymbol{\varepsilon}$. Then, the conditional posterior distribution of q_i takes the following form:

$$\begin{aligned}\rho(q_i|\widetilde{\Theta}\backslash q_i; \boldsymbol{z}) &\propto |\mathbf{B}|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{z}^\mathrm{t}\mathbf{B}^{-1}\boldsymbol{z}\right)\left(\frac{\lambda}{1-\lambda}\right)^{q_i} \\ &\propto \exp\left(-f(q_i)/2\right)\end{aligned}$$

where

$$\begin{aligned}\mathbf{B} &= \mathbf{H}\mathrm{diag}\{\boldsymbol{q}\}\mathbf{H}^\mathrm{t} + \sigma_\varepsilon^2 \mathbf{I}_N, \\ f(q_i) &= \boldsymbol{z}^\mathrm{t}\mathbf{B}^{-1}\boldsymbol{z} + \log|\mathbf{B}| + 2q_i \log(1/\lambda - 1).\end{aligned}$$

After normalization, the marginal probability of q_i reads:

$$\rho(q_i = 1\,|\,\widetilde{\Theta}\backslash q_i, \boldsymbol{z}) = (1 + \exp\left(-(f(0) - f(1))/2\right))^{-1},$$

which reduces to the evaluation of $f(0) - f(1)$, using the following steps [4]:

$$\begin{aligned}\widetilde{\tau}_i &= \delta_i + \mu \boldsymbol{h}_i^\mathrm{t}\widetilde{\mathbf{B}}_0^{-1}\boldsymbol{h}_i \\ \widetilde{\mathbf{B}}_i^{-1} &= \widetilde{\mathbf{B}}_0^{-1} - \mu\widetilde{\mathbf{B}}_0^{-1}\boldsymbol{h}_i\widetilde{\tau}_i^{-1}\boldsymbol{h}_i^\mathrm{t}\widetilde{\mathbf{B}}_0^{-1} \\ |\widetilde{\mathbf{B}}_i^{-1}| &= |\widetilde{\mathbf{B}}_0^{-1}|\delta_i\widetilde{\tau}_i\end{aligned}$$

where $\widetilde{\mathbf{B}} = \mathbf{B}/\sigma_\varepsilon^2$, $\mu = \sigma_\varepsilon^{-2}$ and $\delta_i = \pm 1$ depending on whether 1 is added or removed at q_i; $\boldsymbol{h}_i$ denotes the ith column of matrix $\mathbf{H}$ while $\widetilde{\mathbf{B}}_i$ and $\widetilde{\mathbf{B}}_0$ differ only at q_i.

Further simplifications are also introduced in [4] by exploiting the sparse nature of $\widetilde{\mathbf{B}}$ and then applying the matrix inversion lemma. Noticing that the rank of $\mathbf{H}\mathrm{diag}\{\boldsymbol{q}\}\mathbf{H}^\mathrm{t}$ is only N_e, $\widetilde{\mathbf{B}}$ takes the alternate form $\widetilde{\mathbf{B}} = \mu\mathbf{G}\mathbf{G}^\mathrm{t} + \mathbf{I}$, where $\mathbf{G} = \mathbf{H}\mathbf{D}$ is full rank, $\mathbf{D}$ being made of the nonzero columns of $\mathrm{diag}\{\boldsymbol{q}\}$. Then, we have $\widetilde{\mathbf{B}}^{-1} = \mathbf{I} - \mu\mathbf{G}\widetilde{\mathbf{C}}^{-1}\mathbf{G}^\mathrm{t}$, where $\widetilde{\mathbf{C}} = \mu\mathbf{G}^\mathrm{t}\mathbf{G} + \mathbf{I}$ is $N_\mathrm{e} \times N_\mathrm{e}$. Thus, only $\widetilde{\mathbf{C}}^{-1}$ needs to be stored and updated instead of $\widetilde{\mathbf{B}}^{-1}$, which amounts to $\mathscr{O}(N_\mathrm{e}^2)$ operations instead of $\mathscr{O}(N^2)$. In the case of adding a pulse at q_i, the formula for the update of $\widetilde{\mathbf{C}}_i^{-1}$ is

$$\widetilde{\mathbf{C}}_i^{-1} = \begin{bmatrix} \widetilde{\mathbf{C}}_0^{-1} + \widetilde{\mathbf{b}}\widetilde{\tau}_i\widetilde{\mathbf{b}}^\mathrm{t} & \widetilde{\mathbf{b}} \\ \widetilde{\mathbf{b}}^\mathrm{t} & \widetilde{\tau}_i^{-1} \end{bmatrix} \tag{7}$$

where $\widetilde{\mathbf{b}} = -\mu\widetilde{\tau}_i^{-1}\widetilde{\mathbf{C}}_0^{-1}\mathbf{G}_0^\mathrm{t}\boldsymbol{h}_i$. On the other hand, it can be shown that $f(1-q_i) - f(q_i)$ takes the following form:

$$\begin{aligned}&f(1-q_i) - f(q_i) = \\ &\log(\delta_i\widetilde{\tau}_i) - \mu\widetilde{\tau}_i^{-1}\sigma_\varepsilon^{-2}(\boldsymbol{z}^\mathrm{t}\widetilde{\mathbf{B}}_0^{-1}\boldsymbol{h}_i)^2 + 2\delta_i\log\left(\frac{1}{\lambda} - 1\right)\end{aligned} \tag{8}$$

where

$$\begin{aligned}\widetilde{\tau}_i &= \delta_i + \mu\,\|\boldsymbol{h}_i\|^2 - \mu^2\|\widetilde{\mathbf{R}}_0\mathbf{G}_0^\mathrm{t}\boldsymbol{h}_i\|^2, \\ \boldsymbol{z}^\mathrm{t}\widetilde{\mathbf{B}}_0^{-1}\boldsymbol{h}_i &= \boldsymbol{z}^\mathrm{t}\boldsymbol{h}_i - \mu(\widetilde{\mathbf{R}}_0\mathbf{G}_0^\mathrm{t}\boldsymbol{z})^\mathrm{t}(\widetilde{\mathbf{R}}_0\mathbf{G}_0^\mathrm{t}\boldsymbol{h}_i).\end{aligned}$$

Last but not least, the computation and memory requirements can be still lowered, by updating and storing the upper triangular *Cholesky* factor $\widetilde{\mathbf{R}}$ of $\widetilde{\mathbf{C}}^{-1}$, instead of $\widetilde{\mathbf{C}}^{-1}$. Since $(\boldsymbol{x}\,|\,\widetilde{\Theta}, \boldsymbol{z})$ is Gaussian with covariance $\widetilde{\mathbf{C}}^{-1}$, its Cholesky factor allows to solve step 2(a) in $\mathscr{O}(N_\mathrm{e}^2)$ operations instead of $\mathscr{O}(N_\mathrm{e}^3)$. In the mean time, $\widetilde{\mathbf{R}}$ can be updated in $\mathscr{O}(N_\mathrm{e}^2)$ by using a *Cholesky update* method, given that (7) also reads:

$$\widetilde{\mathbf{C}}_i^{-1} = \begin{bmatrix} \widetilde{\mathbf{C}}_0^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \widetilde{\mathbf{b}}\sqrt{\widetilde{\tau}_i} \\ 1/\sqrt{\widetilde{\tau}_i} \end{bmatrix}\begin{bmatrix} \widetilde{\mathbf{b}}\sqrt{\widetilde{\tau}_i} \\ 1/\sqrt{\widetilde{\tau}_i} \end{bmatrix}^\mathrm{t}$$

Finally, step 1 of Tab. 2 can be summarized as follows:

- Evaluate $\rho(q_i = 0,1\,|\,\widetilde{\Theta}\backslash q_i; \boldsymbol{z})$ using (8), and sample $q_i^{(k+1)}$ accordingly;
- Update $\widetilde{\mathbf{R}}$ if needed, *i.e.*, if $q_i^{(k+1)} \neq q_i^{(k)}$.

4. SIMULATION RESULTS

A test scenario is designed to compare our proposed method with that of Cheng *et al.*'s sampler [1], in terms of robustness w.r.t. different random initial conditions, while the time-shift ambiguities are dealt with in both methods according to [7].

Brooks *et al.*'s convergence diagnostic [13] is based upon parallel chains. Let $\{\Phi_{jt}, j = 1,\ldots,m; t = 1,\ldots,n\}$ denote samples from m independent Markov chains of equal length n, and $\overline{\Phi}_{j.}$ (and $\overline{\Phi}_{..}$) the local mean of the j^{th} chain (*resp.* the global mean). Let the intra-chain and inter-chain variances be defined as covariace matrix averages:

$$\begin{aligned}\mathbf{V}_\mathrm{intra} &= \frac{1}{m(n-1)}\sum_{j=1}^{m}\sum_{t=1}^{n}(\Phi_{jt} - \overline{\Phi}_{j.})(\Phi_{jt} - \overline{\Phi}_{j.})^\mathrm{t} \\ \mathbf{V}_\mathrm{inter} &= \frac{1}{m-1}\sum_{j=1}^{m}(\overline{\Phi}_{j.} - \overline{\Phi}_{..})(\overline{\Phi}_{j.} - \overline{\Phi}_{..})^\mathrm{t}.\end{aligned}$$

$\mathbf{V}_{\text{intra}}$ and $\mathbf{V}_{\text{inter}}$ allow to characterize the convergence behavior. Brooks and Gelman [13] proposed to evaluate:

$$\widehat{R} = \frac{n-1}{n} + \left(\frac{m+1}{m}\right) \lambda(\mathbf{V}_{\text{intra}}^{-1}\mathbf{V}_{\text{inter}})$$

where $\lambda(\cdot)$ denotes the largest eigenvalue, and to wait until $\widehat{R}$ is close to one (*e.g.*, $\widehat{R} < 1.2$).

Let us reexamine the example of Fig. 2. To concentrate on the convergence quality of the Bernoulli sequence, we take $\Phi_{jt} = q_{jt}$ for 10 independent Markov chains. Fig. 4 shows that while $\widehat{R}$ is near but still above 1.2 after 10000 iterations for the reference Gibbs sampler, this threshold is reached after about 1000 iterations for the proposed sampler.

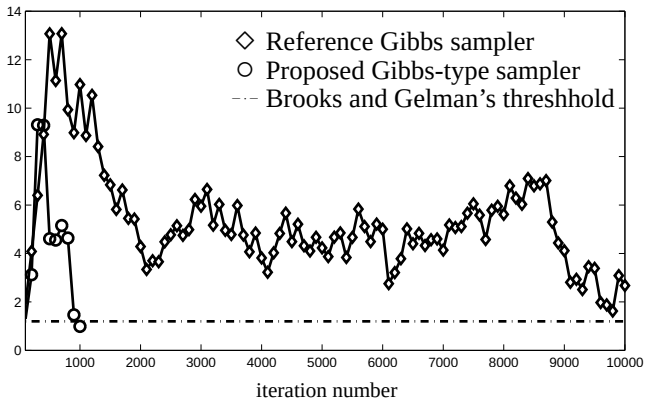


Fig. 4. Evolution of $\widehat{R}$ for both methods in the case of SNR $=$ 12.81 dB. $\widehat{R}$ has been computed every 100 samples.

To further illustrate the robustness of the proposed sampler, deconvolution results are shown in Fig. 5 to compare with that of the reference sampler in Fig. 2 under identical initial conditions. Only one of the results is reported, as they are undistinguishable from each other (perfectly consistent estimations). We noticed that instead of calculating the posterior Gaussian mean based on $\{\widehat{\boldsymbol{q}}, \widehat{\boldsymbol{h}}, \boldsymbol{z}\}$, the estimator $\widehat{\boldsymbol{x}}$ can also be obtained by averaging the auxiliarily sampled $\boldsymbol{x}^{(k)}$ in step 2(a). In fact, as $\{\widetilde{\Theta}^{(k)}\}$ is shown to converge to the distribution $\rho(\widetilde{\Theta} \,|\, \boldsymbol{z})$ in Sect. 3.2, step 2(a) samples $\boldsymbol{x}$ conditional on $\widetilde{\Theta}$ and $\boldsymbol{z}$ such that taken marginally $\boldsymbol{x}^{(k)}$ follows the distribution $\rho(\boldsymbol{x} \,|\, \boldsymbol{z})$.

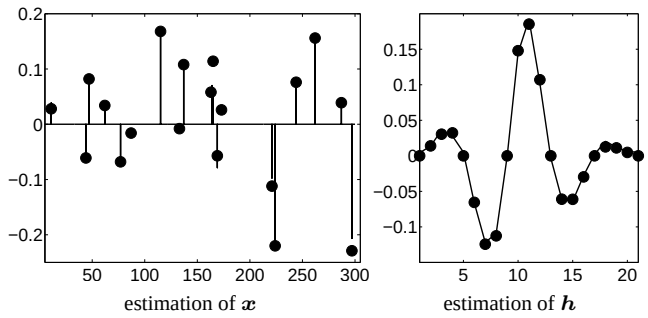


Fig. 5. Estimation result using the proposed sampler, under the same conditions as in Fig. 2. The 10 independent Markov chains yield undistinguishable estimates.

Most importantly, it should also be noted that 1000 iterations of the Gibbs-type sampling take slightly less time than 2000 iterations of the reference Gibbs sampler [7]. We are thus able to conclude that our Gibbs-type sampler achieves a far better compromise between the two criteria [5] than the classical Gibbs sampler [1, 7]: in the displayed example, the time required to reach convergence is reduced by a factor of at least 5.

REFERENCES

[1] 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.

[2] S. Bourguignon and H. Carfantan, "Bernoulli-Gaussian spectral analysis of unevenly spaced astrophysical data", in *IEEE Workshop Stat. Sig. Proc.*, Bordeaux, France, July 2005, pp. 811–816.

[3] J. J. Kormylo and J. M. Mendel, "Maximum-likelihood seismic deconvolution", *IEEE Trans. Geosci. Remote Sensing*, vol. GE-21, no. 1, pp. 72–82, Jan. 1983.

[4] F. Champagnat, Y. Goussard, and J. Idier, "Unsupervised deconvolution of sparse spike trains using stochastic approximation", *IEEE Trans. Signal Processing*, vol. 44, no. 12, pp. 2988–2998, Dec. 1996.

[5] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer Series in Statistics. Springer Verlag, New York, NY, 2001.

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

[7] C. Labat and J. Idier, "Sparse blind deconvolution accounting for time-shift ambiguity", in *Proc. IEEE ICASSP*, Toulouse, France, May 2006, pp. 616–619.

[8] Y. Guan and S. M. Krone, "Small-world MCMC and convergence to multi-modal distributions: From slow mixing to fast mixing", *J. Appl. Prob.*, vol. 17, no. 1, pp. 284–304, Mar. 2007.

[9] N. Dobigeon, J.-Y. Tourneret, and J. D. Scargle, "Joint segmentation of multivariate astronomical times series: Bayesian sampling with a hierarchical model", *IEEE Trans. Signal Processing*, vol. 55, no. 2, pp. 414–423, Feb. 2007.

[10] C. Y. Chi and J. M. Mendel, "Improved maximum-likelihood detection and estimation of Bernoulli-Gaussian processes", *IEEE Trans. Inf. Theory*, vol. 30, pp. 429–435, 1984.

[11] J. S. Liu, W. H. Wong, and A. Kong, "Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes", *Biometrika*, vol. 81, pp. 27–40, 1994.

[12] O. Cappé, A. Doucet, M. Lavielle, and E. Moulines, "Simulation-based methods for blind maximum-likelihood filter identification", *Signal Processing*, vol. 73, no. 1-2, pp. 3–25, Feb. 1999.

[13] S. P. Brooks and A. Gelman, "General methods for monitoring convergence of iterative simulations", *J. Comput. Graph. Statist.*, vol. 7, no. 4, pp. 434–455, 1998.