A Partially Collapsed Gibbs Sampler with Accelerated Convergence for EEG Source Localization

Facundo Costa, Hadj Batatia, Thomas Oberlin, Jean-Yves Tourneret
University of Toulouse, INP/ENSEEIHT - IRIT, 2 rue Charles Camichel, BP 7122, 31071 Toulouse Cedex 7, France

***Abstract*—This paper addresses the problem of designing efficient sampling moves in order to accelerate the convergence of MCMC methods. The Partially collapsed Gibbs sampler (PCGS) takes advantage of variable reordering, marginalization and trimming to accelerate the convergence of the traditional Gibbs sampler. This work studies two specific moves which allow the convergence of the PCGS to be further improved. It considers a Bayesian model where structured sparsity is enforced using a multivariate Bernoulli Laplacian prior. The posterior distribution associated with this model depends on mixed discrete and continuous random vectors. Due to the discrete part of the posterior, the conventional PCGS gets easily stuck around local maxima. Two Metropolis-Hastings moves based on multiple dipole random shifts and inter-chain proposals are proposed to overcome this problem. The resulting PCGS is applied to EEG source localization. Experiments conducted with synthetic data illustrate the effectiveness of this PCGS with accelerated convergence.**

***Index Terms*—MCMC, partially collapsed Gibbs sampler, hierarchical Bayesian model, Metropolis-Hastings moves**

I. Introduction

Bayesian models and methods have become standard statistical tools to solve inference problems associated with many signal and image processing applications (e.g., see [1]). Unfortunately, it is often not possible to obtain closed-form expressions for the Bayesian estimators of the target parameters associated with the posterior distributions of interest. In these situations, several techniques are now available to estimate the parameters of interest. These techniques include variational Bayes [2] and Markov chain Monte Carlo (MCMC) methods [3]. Variational Bayes methods approximate the posterior as a product of separable distributions that depend on smaller subsets of parameters and calculate the MAP estimator of each subset separately. Conversely, MCMC methods generate samples that are asymptotically distributed according to the target posterior and use these generated samples to estimate the parameters of interest. One of the main disadvantages of MCMC techniques is that it can be difficult to determine the amount of iterations required to converge to the posterior distribution [3], which can result in a very large computational complexity. Some convergence assessments are available in the literature to determine whether the convergence has been obtained or not (e.g., see [4]). However, there are few rules allowing this convergence to be accelerated. For instance, by using techniques such as variable reordering, marginalization and trimming, it is possible to convert the sampler into a partially collapsed Gibbs sampler (PCGS) that has generally better convergence properties [5–7].

The main contribution of this paper is to study some specific schemes that allow the convergence of the PCGS to be accelerated. For this, we consider a Bayesian model introduced in [8] to solve an EEG ill-posed inverse problem. The model approximates an $\ell_{2,0}$ mixed norm regularization using a multivariate Bernoulli Laplacian prior. The posterior distribution associated with this model depends on mixed discrete and continuous random vectors. As a consequence, this posterior has a large number of local maxima, which can slow down considerably the convergence speed of the associated PCGS. In order to solve this problem, this paper considers two kinds of Metropolis-Hastings moves that help the sampler to escape from the local maxima without affecting the target distribution. Several chains with the same target distribution are run simultaneously. The first kind of move is applied within each MCMC chain allowing the exchange of source locations using multiple dipole shift proposals. The second kind operates on pairs of chains and allows the estimated source locations to be propagated between chains.

The remainder of the paper is organized as follows: Section II introduces the considered Bayesian model with its multivariate Bernoulli Laplacian prior and its posterior distribution. Section III presents the Metropolis-Hastings moves that are investigated in this work. The performance of these moves is studied in Section IV, which shows some simulation results obtained on realistic synthetic data. Conclusions are reported in Section V.

II. Bayesian EEG source localization

The EEG source localization problem consists in estimating the electrically active areas in the brain from EEG measurements [9]. Since the amount of dipoles used to represent the brain activity is typically much larger than the amount of electrodes, the problem is ill-posed. Thus, a regularization is classically used to provide a unique solution. The following sections summarize the observation model and the priors adopted to solve the EEG source localization problem.

A. Observation model

The EEG measurements can be classically expressed as [9]

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{E} \qquad (1)$$

where $\boldsymbol{X} \in \mathbb{R}^{N\times T}$ contains the amplitudes of the N dipoles considered for the corresponding T time samples, $\boldsymbol{Y} \in \mathbb{R}^{M\times T}$ contains the measurements of the M electrodes, $\boldsymbol{H} \in \mathbb{R}^{M\times N}$ is the head operator and $\mathbf{E} \in \mathbb{R}^{M\times T}$ is a noise term.

B. Likelihood

Considering an additive white Gaussian noise [9], the probability density function (pdf) of $\boldsymbol{Y}$ is

$$f(\boldsymbol{Y}|\boldsymbol{\theta}) = \prod_{t=1}^{T} \mathcal{N}\Big(\boldsymbol{y}^t \Big| \boldsymbol{H}\boldsymbol{x}^t, \sigma_n^2 \mathbb{I}_M\Big) \tag{2}$$

where $\mathbb{I}_M$ is the $M \times M$ identity matrix, σ_n^2 is the noise variance and $\boldsymbol{\theta} = \{\boldsymbol{X}, \sigma_n^2\}$ is the unknown parameter vector.

C. Priors

Dipole amplitudes $\boldsymbol{X}$

Considering the brain activity as sparse and structured spatio-temporally, we adopt an $\ell_{2,0}$ pseudo norm regularization for $\boldsymbol{X}$. In the Bayesian framework, we propose to approximate $\ell_{2,0}$ using a multivariate Bernoulli Laplace prior for each row $\boldsymbol{x}_i$ of $\boldsymbol{X}$ (for $i = 1, ..., N$) by considering the prior

$$f(\boldsymbol{x}_i|z_i, a, \sigma_n^2) \propto \begin{cases} \delta(\boldsymbol{x}_i) & \text{if } z_i = 0 \\ \exp\Big(-\sqrt{\frac{v_i a}{\sigma_n^2}} ||\boldsymbol{x}_i||_2 \Big) & \text{if } z_i = 1 \end{cases} \tag{3}$$

where $\boldsymbol{z}$ and a are hyperparameters, $v_i = ||\boldsymbol{h}^i||_2$ is a weight to compensate the depth-weighting effect (a known problem in the literature [9] that is due to the fact that the different dipoles produce measurements of different amplitude) and $||\boldsymbol{v}||_2$ is the ℓ_2 norm. Parameter a regulates the amplitude of the non-zero elements whereas the elements of $\boldsymbol{z}$ indicate which vectors $\boldsymbol{x}_i$ are zeros. They are assigned a Bernoulli prior (such that $P(z_i = 1) = \omega = 1 - P(z_i = 0)$) defined by

$$z_i|\omega \sim \mathcal{B}\left(z_i|\omega\right), \quad \omega \in [0, 1]. \tag{4}$$

Inspired by [10], we introduced in [8] a latent variable τ_i^2 for each row $\boldsymbol{x}_i$ allowing the indicators z_i to be sampled more efficiently. The resulting prior distribution of $(\tau_i^2, \boldsymbol{x}_i)$ is

$$f(\tau_i^2|a) = \mathcal{G}\Big(\tau_i^2 \Big| \frac{T+1}{2}, \frac{v_i a}{2}\Big) \tag{5}$$

$$f(\boldsymbol{x}_i|z_i, \tau_i^2, \sigma_n^2) = \begin{cases} \delta(\boldsymbol{x}_i) & \text{if } z_i = 0 \\ \mathcal{N}\Big(\boldsymbol{x}_i \Big| 0, \sigma_n^2 \tau_i^2 I_T\Big) & \text{if } z_i = 1 \end{cases} \tag{6}$$

where $\mathcal{G}$ and $\mathcal{N}$ denote the gamma and normal distributions. It is straightforward to show that the marginal distribution of $\boldsymbol{x}_i$ computed from (5) and (6) agrees with (3).

Noise variance σ_n^2

The noise variance σ_n^2 is assigned a Jeffrey's prior

$$f(\sigma_n^2) \propto \frac{1}{\sigma_n^2} 1_{\mathbb{R}^+}(\sigma_n^2) \tag{7}$$

where $1_{\mathbb{R}^+}(\xi) = 1$ if $\xi \in \mathbb{R}^+$ and 0 otherwise.

D. Hyperparameter priors

The hyperparameters ω and a are assigned uniform and conjugate gamma priors

$$f(\omega) = \mathcal{U}(\omega|0, 1), \quad f(a|\alpha, \beta) = \mathcal{G}\left(a \middle| \alpha, \beta\right)$$

with $\alpha = \beta = 1$ (corresponding to a vague prior for a).

E. Posterior distribution

Using the priors defined above and denoting the hyperparameter vector $\boldsymbol{\phi} = \{\omega, a\}$, the posterior distribution of the proposed Bayesian model can be derived as follows

$$f(\theta, \boldsymbol{z}, \boldsymbol{\tau}^2, \boldsymbol{\phi}|\boldsymbol{Y}) \propto f(\boldsymbol{Y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{z}, \boldsymbol{\tau}^2) f(\boldsymbol{z}, \boldsymbol{\tau}^2|\boldsymbol{\phi}) f(\boldsymbol{\phi}) \tag{8}$$

where $f(\boldsymbol{Y}|\boldsymbol{\theta})$ has been defined in (2) and

$$f(\boldsymbol{\theta}|\boldsymbol{z}, \boldsymbol{\tau}^2) \propto f(\sigma_n^2) \prod_{i=1}^{N} f(\boldsymbol{x}_i|z_i, \tau_i^2, \sigma_n^2)$$

$$f(\boldsymbol{z}, \boldsymbol{\tau}^2|\boldsymbol{\phi}) = \prod_{i=1}^{N} f(z_i|\omega) f(\tau_i^2|a)$$

$$f(\boldsymbol{\phi}) = f(a|\alpha, \beta) f(\omega).$$

F. Partially collapsed Gibbs sampler

To estimate the model parameters, we proposed in [8] to draw samples from (8) using a PCGS, sampling z_i and $\boldsymbol{x}_i$ jointly. The corresponding conditional distributions are summarized in Table I, where $\mathcal{GIG}$, $\mathcal{IG}$ and $\mathcal{B}e$ are the generalized inverse Gaussian, inverse gamma and beta distributions (see also [8]). Note that $\boldsymbol{X}_{-i}$ denotes the matrix $\boldsymbol{X}$ whose i-th row has been set to zero and that the following notations have been used

$$\boldsymbol{\mu}_i = \frac{\sigma_i^2 \boldsymbol{h}^{i^T}(\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{X}_{-i})}{\sigma_n^2}, \sigma_i^2 = \frac{\sigma_n^2 \tau_i^2}{1 + \tau_i^2 \boldsymbol{h}^{i^T} \boldsymbol{h}^i}$$

$$k_0 = 1 - \omega, k_1 = \omega \left(\frac{\sigma_n^2 \tau_i^2}{\sigma_i^2}\right)^{-\frac{T}{2}} \exp\left(\frac{||\boldsymbol{\mu}_i||^2}{2\sigma_i^2}\right).$$

τ_i^2	$\mathcal{G}\Big(\frac{T+1}{2}, \frac{v_i a}{2}\Big)$ if $z_i = 0$ $\mathcal{GIG}\Big(\frac{1}{2}, v_i a, \frac{\|\|\boldsymbol{x}_i\|\|^2}{\sigma_n^2}\Big)$ if $z_i = 1$
z_i	$\mathcal{B}\Big(1, \frac{k_1}{k_0 + k_1}\Big)$
$\boldsymbol{x}_i$	$\delta(\boldsymbol{x}_i)$ if $z_i = 0$ $\mathcal{N}\Big(\boldsymbol{\mu}_i, \sigma_i^2\Big)$ if $z_i = 1$
a	$\mathcal{G}\Big(\frac{N(T+1)}{2} + \alpha, \frac{\sum_i [v_i \tau_i^2]}{2} + \beta\Big)$
σ_n^2	$\mathcal{IG}\Big(\frac{(M + \|\|\boldsymbol{z}\|\|_0)T}{2}, \frac{1}{2}\Big[\|\|\boldsymbol{H}\boldsymbol{X} - \boldsymbol{Y}\|\|^2 + \sum_i \frac{\|\|\boldsymbol{x}_i\|\|^2}{\tau_i^2}\Big]\Big)$
ω	$\mathcal{B}e\Big(1 + \|\|\boldsymbol{z}\|\|_0, 1 + N - \|\|\boldsymbol{z}\|\|_0\Big)$

TABLE I: Conditional distributions $f(\tau_i^2|\boldsymbol{x}_i, \sigma_n^2, a, z_i)$, $f(z_i|\boldsymbol{Y}, \boldsymbol{X}_{-i}, \sigma_n^2, \tau_i^2, \omega)$, $f(\boldsymbol{x}_i|z_i, \boldsymbol{Y}, \boldsymbol{X}_{-i}, \sigma_n^2, \tau_i^2)$, $f(a|\boldsymbol{\tau}^2)$, $f(\sigma_n^2|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\tau}^2, \boldsymbol{z})$ and $f(\omega, \boldsymbol{z})$ used in [8].

III. Metropolis-Hastings moves

A. Multiple dipole shift proposals

We have observed that the standard PCGS developed in [8] gets sometimes stuck around local maxima of the posterior 8. This section studies two kinds of moves which allow the sampler to escape from these local maxima, and thus ensure a faster convergence of the PCGS. The multiple dipole shift (MDS) is a move changing several elements of $\boldsymbol{z}$ simultaneously after each sampling iteration. The proposed move is accepted or rejected using a Metropolis-Hastings criterion to preserve the target distribution. This move is inspired by an idea developed by Bourguignon *et al* [11]. The authors of [11] proposed to move a single non-zero element of a binary sequence to a random neighboring position after each iteration of the MCMC sampler. We generalize here their scheme by proposing to move a random subset of K non-zeros simultaneously to random neighboring positions. Our experiments have shown empirically that $K = 1$ is not enough to escape from local maxima but that $K = 2$ provides an improved convergence. Since there is a high correlation between the variables $\boldsymbol{\tau}^2$ and $\boldsymbol{z}$, we propose to update their values jointly. The proposal is detailed in Algorithm 1.

Algorithm 1 Multiple dipole shift proposal.

$\bar{\boldsymbol{z}} = \boldsymbol{z}$
repeat K times
  Set ind_{old} to be the index of a random non-zero of $\boldsymbol{z}$
  Set $\boldsymbol{p} = [\text{ind}_{\text{old}}, \text{neigh}_\gamma(\text{ind}_{\text{old}})]$
  Set ind_{new} to be a random element of $\boldsymbol{p}$
  Set $\bar{z}_{\text{ind}_{\text{old}}} = 0$ and $\bar{z}_{\text{ind}_{\text{new}}} = 1$
end
Sample $\bar{\boldsymbol{X}}$ from $f(\bar{\boldsymbol{X}}|\bar{\boldsymbol{z}}, \boldsymbol{Y}, \sigma_n^2, \boldsymbol{\tau}^2)$.
Sample $\bar{\boldsymbol{\tau}}^2$ from $f(\bar{\boldsymbol{\tau}}^2|\bar{\boldsymbol{X}}, \sigma_n^2, a, \bar{\boldsymbol{z}})$.
Set $\{\boldsymbol{z}, \boldsymbol{\tau}^2\} = \{\bar{\boldsymbol{z}}, \bar{\boldsymbol{\tau}}^2\}$ with probability $\min\left(\frac{f(\bar{\boldsymbol{z}}, \bar{\boldsymbol{\tau}}^2|.)}{f(\boldsymbol{z}, \boldsymbol{\tau}^2|.)}, 1\right)$
Resample $\boldsymbol{X}$ if the proposal was accepted

Neighborhood

The difficulty of sampling the posterior distribution (8) is mainly due to the fact that some columns of $\boldsymbol{H}$ are highly correlated. Consequently, the neighborhood should not be strictly topological but should depend on the structure of $\boldsymbol{H}$. Thus we propose to define the neighborhood of $\boldsymbol{x}_i$ as follows

$$\text{neigh}_\gamma(i) \triangleq \left\{ j \neq i \;\middle|\; |\text{corr}(\boldsymbol{h}^i, \boldsymbol{h}^j)| \geq \gamma \right\} \tag{9}$$

where $\text{corr}(\boldsymbol{v}_1, \boldsymbol{v}_2)$ is the correlation between the two vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ and where the neighborhood size can be adjusted by the value of $\gamma \in [0, 1]$ ($\text{neigh}_\gamma(i)$ contains all the dipoles for $\gamma = 0$, whereas $\text{neigh}_\gamma(i)$ is empty for $\gamma = 1$).

In order to maximize the move efficiency, the value of γ has to be selected carefully. Indeed, a too large value of γ will prevent the algorithm to escape from local maxima. Conversely, a too low value of γ yields many useless proposals requiring a large number of iterations to obtain useful moves. Our results obtained by cross validation have shown that a good compromise is obtained for $\gamma = 0.8$.

Algorithm 2 Inter-chain proposals.

Define $\boldsymbol{c} = \{1, .., L\}$ where L is the amount of chains
for $i = \{1, .., L\}$
  Choose (and remove) a random element k from $\boldsymbol{c}$
  Denote as $\{\bar{\boldsymbol{z}}_k, \bar{\boldsymbol{\tau}}_k^2\}$ the sampled values of $\{\boldsymbol{z}, \boldsymbol{\tau}^2\}$ of MCMC chain number $\#k$
  For the chain $\#i$ set $\{\boldsymbol{z}_i, \boldsymbol{\tau}_i^2\} = \{\bar{\boldsymbol{z}}_k, \bar{\boldsymbol{\tau}}_k^2\}$ with probability $\frac{f(\bar{\boldsymbol{z}}_k, \bar{\boldsymbol{\tau}}_k^2|.)}{f(\boldsymbol{z}, \boldsymbol{\tau}^2|.)}$
  Resample $\boldsymbol{X}$ if the proposal has been accepted
end

B. Inter-chain proposals

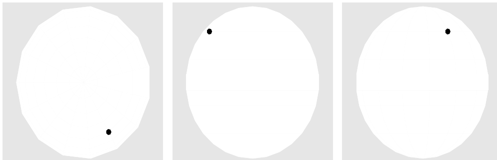
The MDS proposal described in the previous section allows the algorithm to find the active dipoles correctly provided the number of active dipoles is small. However, when a higher amount of non-zeros is present in the ground truth, it is possible for the chains to get stuck around different values of $\boldsymbol{z}$. In order to help these chains to converge to the same value, we propose to exchange information between parallel chains as suggested in [12–14]. We do this by introducing inter-chain (IC) moves exchanging the values of $\boldsymbol{z}$ and $\boldsymbol{\tau}^2$ between different chains. These moves are accepted with a Metropolis-Hastings acceptance probability to ensure the target posterior distribution is preserved. More precisely, an IC proposal is made after each iteration with probability p (adjusted to $\frac{1}{100}$ by cross validation) according to Algorithm 2.

IV. Experimental validation

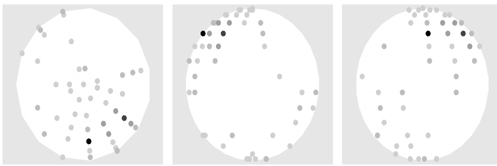
Realistic synthetic data was used to illustrate the effectiveness of the proposals. A three-shell head model with 41 electrodes and 212 dipoles was built. Two different values of the dipole activity $\boldsymbol{X}$ were considered, one with a single active dipole and one with five active dipoles. Active dipoles were assigned damped sinusoidal excitations with frequencies between 5 and 20Hz, a time duration of 500ms and a sampling frequency 200Hz. The single dipole activation was used to show the effectiveness of MDS proposals whereas the five dipole activation was used to test the IC moves. In each case, eight MCMC chains were run in parallel with 10.000 iterations. The probability of dipole activity was estimated as the proportion of chains that found an activity for this dipole.

Single dipole

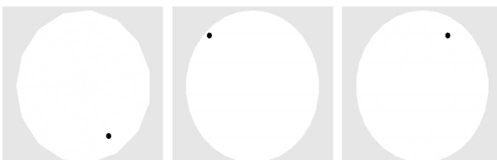
The first way of comparing the convergence of the samplers (with and without using MDS proposals) is to consider the potential scale reduction factor (PSRF) defined in [15],[16, p. 332]. After 10.000 iterations, the highest value of PSRF calculated using the MDS moves was 1.15 whereas



(a) Ground truth



(b) 10.000 iterations without MDS moves



(c) 80 iterations with MDS moves

Fig. 1: Actual/estimated activity probabilities (single dipole). Axial, coronal and sagittal views respectively.

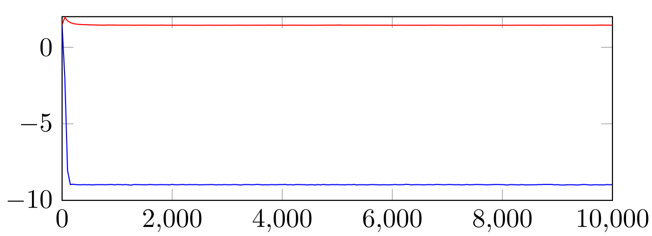
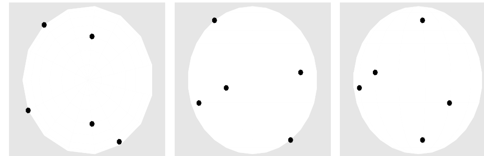
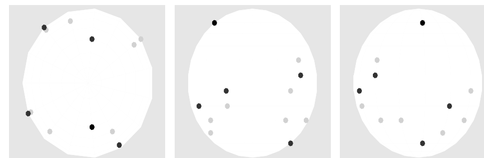


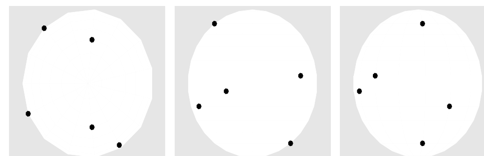
Fig. 2: Logarithmic MSE of $\boldsymbol{X}$ with (blue) and without (red) MSD moves for the single dipole simulation.



(a) Ground truth



(b) 10.000 iterations without IC moves



(c) 1.000 iterations with IC moves

Fig. 3: Actual/estimated activity probability (five dipoles). Axial, coronal and sagittal views respectively.

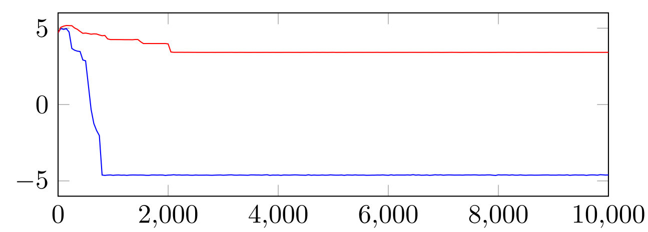


Fig. 4: Logarithmic MSE of $\boldsymbol{X}$ with (blue) and without (red) IC moves for the five-dipole simulation.

it was 32 without using these moves (we recall here that a PRSF less than 1.2 is recommended for deciding that the sampler has converged). The probability of finding activity in each dipole with and without proposals is illustrated in Fig. 1 and is compared to the ground truth. Without using our proposals, the different chains are unable to converge to the correct solution in 10.000 iterations. Conversely, the MDS proposal allows the correct dipole activity to be estimated in less than 100 iterations. A last way of analyzing the MDS proposal is to study the evolution of the mean square error (MSE) of $\boldsymbol{X}$ versus the number of iterations as displayed in Fig. 2 that shows that MDS moves accelerate convergence.

Multiple dipoles

This section illustrates the use of IC moves in addition to MDS moves in presence of multiple active dipoles. For this, we considered 5 active dipoles and used both kinds of proposals. The highest value of PSRF (after 10.000 iterations) obtained when using the MDS proposal only is 171 whereas it equals 1.01 when using the MDS and IC proposals, which shows the benefit of introducing IC proposals. Each estimated activity probability is displayed in Fig. 3 with its corresponding ground truth. The advantage of introducing IC moves is finally confirmed in Fig. 4, which shows the MSEs of $\boldsymbol{X}$ obtained with and without them.

V. Conclusion

This paper studied the efficiency of two Metropolis-Hastings moves to accelerate the convergence of a partially collapsed Gibbs sampler used for EEG source localization. These moves were based on multiple dipole shifts for the elements of a given chain and on inter-chain exchanges. The advantages of these moves were clearly shown on synthetic data with controlled ground truth. Even if the results were obtained for a specific hierarchical Bayesian model based on a multivariate Bernoulli Laplacian prior, we think that these moves are also of interest in other contexts. Our future work will be devoted to extend our results to situations where the head operator is only partially known and has to be estimated jointly with the other model parameters.

REFERENCES

[1] C. Robert, *The Bayesian Choice: a decision theoric motivation*. New York: Springer-Verlag, 1994.

[2] V. Šmídl and A. Quinn, *The Variational Bayes Method in Signal Processing*. New York: Springer-Verlag, 2005.

[3] G. Casella and C. P. Robert, *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 1999.

[4] C. Robert, *Discretization and MCMC convergence assessment*. New York: Springer-Verlag, 2012.

[5] D. A. Van Dyk and T. Park, "Partially collapsed Gibbs samplers: Theory and methods," *J. Amer. Stat. Soc.*, vol. 103, no. 482, pp. 790–796, 2008.

[6] T. Park and D. A. van Dyk, "Partially collapsed Gibbs samplers: Illustrations and applications," *J. Comput. and Graph. Stat.*, vol. 18, no. 2, pp. 283–305, 2009.

[7] M. Pereyra, P. Schniter, E. Chouzenoux, J.-C. Pesquet, J.-Y. Tourneret, A. Hero, and S. McLaughlin, "Tutorial on Stochastic Simulation and Optimization Methods in Signal Processing," *arXiv preprint arXiv:1505.00273*, 2015.

[8] F. Costa, H. Batatia, T. Oberlin, and J.-Y. Tourneret, "EEG source localization based on a structured sparsity prior and a partially collapsed gibbs sampler," in *Proc. of International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP15)*, Cancun, Mexico, 2015.

[9] R. Grech, T. Cassar, J. Muscat, K. P. Camilleri, S. G. Fabri, M. Zervakis, P. Xanthopoulos, V. Sakkalis, and B. Vanrumste, "Review on solving the inverse problem in EEG source analysis," *J. Neuroeng. Rehabil.*, vol. 4, pp. 5–25, 2008.

[10] S. Raman, T. J. Fuchs, P. J. Wild, E. Dahl, and V. Roth, "The Bayesian group-lasso for analyzing contingency tables," in *Proc. 26th ACM Annu. Int. Conf. Mach. Learn. (ICML)*, Montreal, Quebec, Jun. 2009.

[11] S. Bourguignon and H. Carfantan, "Bernoulli-Gaussian spectral analysis of unevenly spaced astrophysical data," in *Proc. IEEE Workshop on Stat. Signal Processing (SSP)*, Bordeaux, France, Jul. 2005.

[12] C. J. Geyer, "Markov chain Monte Carlo maximum likelihood," in *Proc. 23rd Symp. Interface Comput. Sci. Statist.*, Seattle, USA, Oct. 1991.

[13] K. B. Laskey and J. W. Myers, "Population markov chain monte carlo," *Mach. Learn.*, vol. 50, pp. 175–196, 2003.

[14] C. J. Geyer and E. A. Thompson, "Annealing Markov chain Monte Carlo with applications to ancestral inference," *J. Amer. Stat. Soc.*, vol. 90, no. 431, pp. 909–920, 1995.

[15] 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.

[16] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian data analysis*. London, U.K.: Chapman & Hall, 1995.