EEG SOURCE LOCALIZATION BASED ON A STRUCTURED SPARSITY PRIOR AND A PARTIALLY COLLAPSED GIBBS SAMPLER

Facundo Costa, Hadj Batatia, Thomas Oberlin and Jean-Yves Tourneret

University of Toulouse, IRIT - INP-ENSEEIHT, France
firstname.lastname@enseeiht.fr

ABSTRACT

In this paper, we propose a hierarchical Bayesian model approximating the ℓ_{20} mixed-norm regularization by a multivariate Bernoulli Laplace prior to solve the EEG inverse problem by promoting spatial structured sparsity. The posterior distribution of this model is too complex to derive closed-form expressions of the standard Bayesian estimators. However, this posterior can be sampled using an MCMC method and the generated samples can be used to compute Bayesian estimators of the unknown model parameters. The proposed MCMC algorithm is based on a partially collapsed Gibbs sampler and a dual dipole random shift proposal for the non-zero positions. Note that the proposed method estimates the brain activity and all other model parameters jointly in a completely unsupervised framework. The results obtained on synthetic data with controlled ground truth show the good performance of the proposed method when compared to the ℓ_{21} approach in different scenarios, and its capacity to estimate point-like source activity.

Index Terms**—** EEG, MCMC, inverse problem, source localization, structured-sparsity, hierarchical Bayesian model, ℓ_{20}-norm regularization

1. INTRODUCTION

EEG source localization is an ill-posed inverse problem [1] that continues to attract a significant amount of interest in the signal and image processing literature. The problem is classically addressed using some regularization that enforces realistic properties on the solution. Among the proposed regularizations, the ℓ_0 pseudo-norm is known to estimate correctly sparse focal brain activity [2]. Unfortunately, the minimization of the ℓ_0 pseudo-norm is intractable. Thus it is usually approximated by the convex ℓ_1 norm that can be handled more easily using classical optimization techniques [3] but does not provide the same solution [2]. In addition, both methods consider each time sample independently which can lead to unrealistic solutions [4]. It has been shown that structured sparsity can provide better results by exploiting the temporal dimension of the data [5]. Structured sparsity can be enforced for EEG source localization using mixed-norms such as the ℓ_{21} norm [4] (also known as group-lasso), which constrains all the time samples of a dipole to be either completely active or inactive during the time period. As an alternative to the ℓ_{21} norm, we introduce a new hierarchical Bayesian model based on a multivariate Bernoulli Laplacian prior on the dipole activity. This paper will show that this prior allows sparser solutions to be obtained. Since the posterior associated with this prior is intractable, a Markov chain Monte Carlo sampling technique is used to draw samples of the unknown parameters asymptotically distributed according to this posterior. A dual dipole random shift proposal is also added in order to improve convergence. The generated samples are then used to estimate both the brain activity and the model parameters and hyperparameters in a completely unsupervised framework.

The paper is organized as follows: Section 2 introduces the source localization problem. The proposed Bayesian model is defined in Section 3. Section 4 presents the partially collapsed Gibbs sampler that can generate samples asymptotically distributed according to the posterior of this model. Results obtained with synthetic data are presented in Section 5. Section 6 concludes the paper.

2. PROBLEM STATEMENT

We consider a distributed-source model [1] that has a fixed number of dipoles on the cortical surface whose orientations are supposed orthogonal to the cortex.

$$\boldsymbol{Y} = \mathcal{H}\boldsymbol{X} + \mathbf{E} \tag{1}$$

where $\boldsymbol{X} \in \mathbb{R}^{N\times T}$ contains the amplitudes of the N dipoles for the corresponding T time samples, $\boldsymbol{Y} \in \mathbb{R}^{M\times T}$ contains the measurements of the M electrodes for these T time samples, $\mathcal{H} \in \mathbb{R}^{M\times N}$ models the propagation of the electromagnetic field from the sources to the sensors and $\mathbf{E} \in \mathbb{R}^{M\times T}$ is a noise term. The EEG source localization problem consists of estimating the matrix $\boldsymbol{X}$ from the measurements $\boldsymbol{Y}$.

3. BAYESIAN MODEL

3.1. Likelihood

It is very classical in the literature to consider an additive white Gaussian noise with a constant variance σ_n^2 for the T considered time instants [1]. Note that when this assumption does not hold, it is possible to estimate the noise covariance matrix from the data and to whiten the measurements in a preprocessing stage [4]. This assumption leads to the likelihood

$$f(\boldsymbol{y}^t|\boldsymbol{x}^t,\sigma_n^2) = \left(\frac{1}{2\pi\sigma_n^2}\right)^{\frac{M}{2}} \exp\left(-\frac{||\boldsymbol{y}^t - \boldsymbol{H}\boldsymbol{x}^t||^2}{2\sigma_n^2}\right) \quad (2)$$

where $\boldsymbol{y}^t$ is the tth column of $\boldsymbol{Y}$ and $||.||$ denotes the Euclidean norm. Different time samples are assumed to be associated with independent noises.

3.2. Priors

3.2.1. Dipole amplitudes $\boldsymbol{X}$

The weighted ℓ_{20} pseudo norm of a matrix $\boldsymbol{X}$ with rows $\boldsymbol{x}_1, ..., \boldsymbol{x}_N$ is defined by

$$||\boldsymbol{X}||_{20} = \#\{i : \sqrt{v_i}||\boldsymbol{x}_i||_2 \neq 0\} \quad (3)$$

where $\#\mathcal{S}$ is the cardinal of the set $\mathcal{S}$ and $v_i = ||\boldsymbol{h}^i||_2$ ($\boldsymbol{h}^i$ being the i-th column of the operator $\mathcal{H}$) is a weight used to compensate for the depth-weighting effect as explained in [1, 3]. We propose to approximate the ℓ_0 norm using a Bernoulli distribution via a multivariate Bernoulli Laplace prior for each row $\boldsymbol{x}_i$ of $\boldsymbol{X}$, i.e., by considering the following prior

$$f(\boldsymbol{x}_i|z_i, a, \sigma_n^2) \propto \begin{cases} \delta(\boldsymbol{x}_i) & \text{if } z_i = 0 \\ \exp\left(-\sqrt{\frac{v_i a}{\sigma_n^2}}||\boldsymbol{x}_i||_2\right) & \text{if } z_i = 1 \end{cases} \quad (4)$$

where a is a hyperparameter that controls the amplitudes of the non-zero rows of $\boldsymbol{X}$ and $\boldsymbol{z} \in \{0,1\}^N$ is a vector indicating which rows of $\boldsymbol{X}$ are non-zero. The elements of $\boldsymbol{z}$ are assigned a Bernoulli prior with parameter $\omega \in [0,1]$

$$z_i|\omega \sim \mathcal{B}(z_i|\omega). \quad (5)$$

Note that the prior of $\boldsymbol{x}_i$ defined in (4) contains two different parts: the Dirac delta function $\delta(.)$ that promotes sparsity by ensuring absence of activity and the multivariate Laplace distribution that adjusts the amplitudes of the non-zero rows. Setting $\omega = 0$ reduces to $\boldsymbol{X}$ whereas $\omega = 1$ corresponds to the ℓ_{21}-mixed norm regularization introduced in the Bayesian formulation of the group-lasso. To be able to sample efficiently from the posterior distribution of the model parameters, it is interesting to introduce a latent variable τ_i^2 for each row $\boldsymbol{x}_i$ as in [6]. More precisely, the joint prior distribution of $(\tau_i^2, \boldsymbol{x}_i)$ can be defined as

$$f(\tau_i^2|a) = \mathcal{G}\left(\tau_i^2 \middle| \frac{T+1}{2}, \frac{v_i a}{2}\right) \quad (6)$$

$$f(\boldsymbol{x}_i|z_i,\tau_i^2,\sigma_n^2) = \begin{cases} \delta(\boldsymbol{x}_i) & \text{if } z_i = 0 \\ \mathcal{N}\left(\boldsymbol{x}_i \middle| 0, \sigma_n^2\tau_i^2 I_T\right) & \text{if } z_i = 1 \end{cases} \quad (7)$$

where $\mathcal{G}$ and $\mathcal{N}$ denote the gamma and normal distributions. Indeed, the prior distribution specified above is such that the marginal distribution of $\boldsymbol{x}_i$ is (4) [6].

3.2.2. Noise variance σ_n^2

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

$$f(\sigma_n^2) \propto \frac{1}{\sigma_n^2} 1_{\mathbb{R}^+}(\sigma_n^2) \quad (8)$$

where $1_{\mathbb{R}^+}(\xi) = 1$ if $\xi \in \mathbb{R}^+$ and 0 otherwise. Motivations for using this prior can be found in [7].

3.3. Hyperparameter priors

In the ℓ_{21} norm based approach, the regularization parameter makes a compromise between the sparsity of the solution and the fidelity to the measurements. In the proposed Bayesian model, this compromise is adjusted by two hyperparameters: (1) ω that determines the proportion of the rows of $\boldsymbol{X}$ that are non-zero and (2) a that controls the amplitudes of the non-zero rows of $\boldsymbol{X}$. We will denote the hyperparameter vector by $\boldsymbol{\phi} = \{\omega, a\}$. To make our algorithm capable of estimating the values of ω and a from the data, we need to assign priors to these hyperparameters (usually called hyperpriors).

A conjugate gamma prior is chosen for a for simplicity

$$f(a|\alpha,\beta) = \mathcal{G}\left(a\middle|\alpha,\beta\right) \quad (9)$$

with $\alpha = \beta = 1$. This choice of (α, β) corresponds to a vague hyperprior for a.

A non-informative uniform prior on $[0,1]$ is used for ω

$$f(\omega) = \mathcal{U}(\omega|0,1). \quad (10)$$

also reflecting the absence of knowledge for this parameter.

3.4. Posterior distribution

Using the priors and hyperpriors defined in Section 3.2 and 3.3, the posterior distribution of the proposed Bayesian model can be derived as follows

$$f(\theta, \boldsymbol{z}, \boldsymbol{\tau}^2, \boldsymbol{\phi}|\boldsymbol{Y}) \propto f(\boldsymbol{Y}|\boldsymbol{\theta})f(\boldsymbol{\theta}|\boldsymbol{z},\boldsymbol{\tau}^2)f(\boldsymbol{z},\boldsymbol{\tau}^2|\boldsymbol{\phi})f(\boldsymbol{\phi}) \quad (11)$$

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

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

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

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

4. A PARTIALLY COLLAPSED GIBBS SAMPLER

The Bayesian estimators of the unknown model parameters $\sigma_n^2, \boldsymbol{X}, \boldsymbol{z}, a, \boldsymbol{\tau}^2, \omega$ are clearly difficult to express in closed form using (11). Thus, we propose to draw samples from the posterior distribution (11) and use these samples to estimate the model parameters and hyperparameters using a partially collapsed Gibbs sampler which samples the variables z_i and $\boldsymbol{x}_i$ jointly. The corresponding conditional distributions are detailed in the following sections.

4.1. Conditional distributions

The conditional distributions of the different parameters and hyperparameters are provided in Table 1, where $\mathcal{G}$, $\mathcal{GIG}$, $\mathcal{N}$, $\mathcal{B}$, $\mathcal{IG}$ and $\mathcal{B}e$ stand for the gamma, generalized inverse Gaussian, normal, Bernoulli, inverse gamma and beta distributions respectively (for the definition of the $\mathcal{GIG}$ distribution, see [6]).

τ_i^2	$\mathcal{G}\left(\frac{T+1}{2}, \frac{v_i a}{2}\right)$ if $z_i = 0$ $\mathcal{GIG}\left(\frac{1}{2}, v_i a, \frac{\lVert \boldsymbol{x}_i \rVert^2}{\sigma_n^2}\right)$ if $z_i = 1$
z_i	$\mathcal{B}\left(1, \frac{k_1}{k_0+k_1}\right)$
$\boldsymbol{x}_i$	$\delta(\boldsymbol{x}_i)$ if $z_i = 0$ $\mathcal{N}\left(\boldsymbol{\mu}_i, \sigma_i^2\right)$ if $z_i = 1$
a	$\mathcal{G}\left(\frac{N(T+1)}{2} + \alpha, \frac{\sum_i [v_i \tau_i^2]}{2} + \beta\right)$
σ_n^2	$\mathcal{IG}\left(\frac{(M+\lVert \boldsymbol{z} \rVert_0)T}{2}, \frac{1}{2}\left[\lVert \boldsymbol{HX} - \boldsymbol{Y} \rVert^2 + \sum_i \frac{\lVert \boldsymbol{x}_i \rVert^2}{\tau_i^2}\right]\right)$
ω	$\mathcal{B}e\left(1 + \lVert \boldsymbol{z} \rVert_0, 1 + N - \lVert \boldsymbol{z} \rVert_0\right)$

Table 1: 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 $f(\omega, \boldsymbol{z})$.

We denote by $\boldsymbol{X}_{-i}$ the matrix $\boldsymbol{X}$ with its i-th row set to zero and

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

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

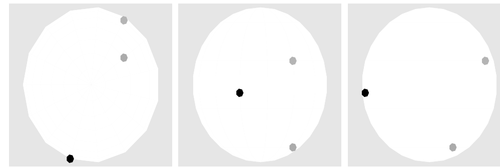
4.2. Dual dipole random shift proposal

In practice, the Gibbs sampler can get trapped in local maxima of the target distribution, especially when the indicator variables z_i have to be sampled. This problem has been reported in several works such as [8] and has been observed for the proposed partially collapsed Gibbs sampler. To solve this problem, after each sampling iteration, a new value of $\boldsymbol{z}$ can be proposed in order to escape from a possible local maximum. This value is accepted or rejected using the Metropolis-Hastings acceptance ratio to keep the same target distribution.

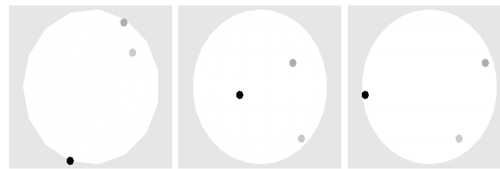
In this work, we have implemented dual dipole random shift proposals which consist of moving up to two indicators within their neighborhoods (defined as the operator columns that have a correlation with the dipole column higher than γ). The parameter $\gamma \in [0, 1]$ tunes the neighborhood size ($\gamma = 0$ corresponds to a neighborhood containing all the dipoles and $\gamma = 1$ corresponds to an empty neighborhood). In our experiments, we have used the value $\gamma = 0.8$ that has been adjusted by cross validation.

5. EXPERIMENTAL VALIDATION

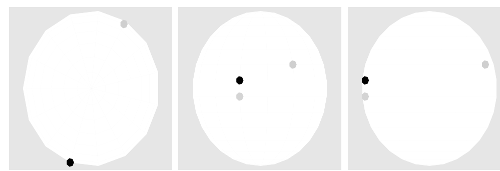
This section aims at comparing the proposed method with the weighted ℓ_{21} approach. The EEG source localization problem considered in our experiments uses the Stok three-shell head model with 41 electrodes and a source space of 212 dipoles uniformly distributed in the cortex surface. Synthetic damped sinusoidal excitations with frequencies between 5 and 20Hz were assigned to the active dipoles dipoles. These excitations are 500ms long (a period in which the dipole activity is known to be stationary) and sampled at 200Hz, resulting in $T = 100$. The regularization parameter of the weighted ℓ_{21} norm was set according to the uncertainty principle.



(a) Ground truth



(b) Proposed method

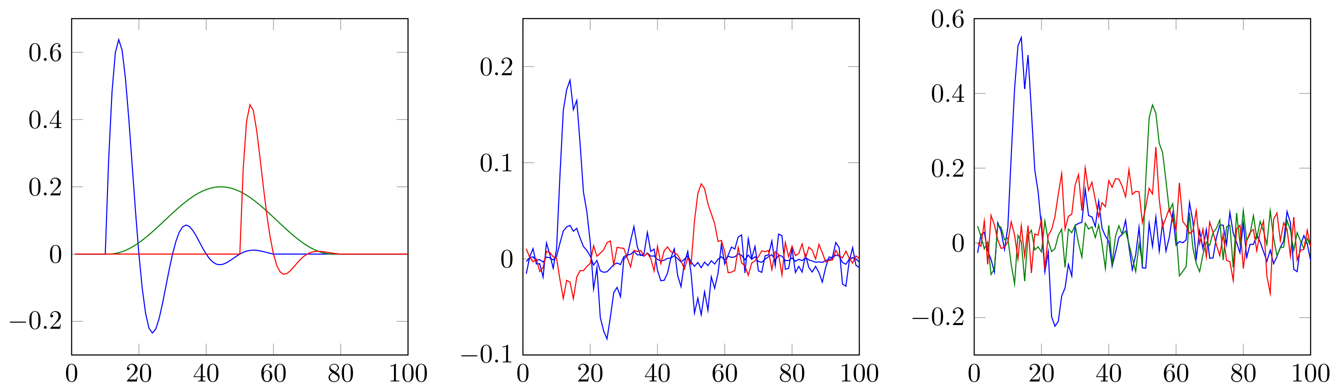


(c) Weighted ℓ_{21}-norm

Fig. 1: Typical brain activity localization (SNR = -3dB).

Two different kind of simulations were run, the first one has a fixed amount of active dipoles in the ground truth and a variable level of SNR whereas the second one presents a fixed level of SNR with a variable amount of active dipoles.

For the first kind of simulations three dipoles were active in the ground truth. For high SNR values (20dB or more), both methods are able to correctly detect the dipole locations and estimate their activation waveforms. However, as the SNR decreases, the proposed method outperforms the approach based on the ℓ_{21} norm. A representative example is illustrated in Figs. 1 and 2 (obtained for SNR = -3dB). As we can see in this particular case, the proposed algorithm based on a structured sparsity prior manages to recover correctly the three activations while concentrating each of the activations in only one dipole. In comparison, the ℓ_{21} norm only recovers two activations and spreads some of the activity between neighboring dipoles. One can also see that the waveforms recovered by the proposed method are much closer to the original excitations than those obtained with the ℓ_{21} norm (note the presence of a bias with the latter). This result can be explained by the fact that the ℓ_1 norm tends to overpenalize large amplitudes whereas the selected prior penalizes all non-zero coefficients equally.



(a) Ground Truth

(b) Weighted ℓ_{21} mixed norm

(c) Proposed method

Fig. 2: Ground truth and typical estimated time waveforms with SNR = -3dB.

For the second kind of simulations, the SNR was set to 30dB while the amount of active dipoles in the ground truth (denoted by P) was varied from 1 to 7. Fifty different active dipole localizations were used for each value of P. After each simulation run, the P dipoles with highest estimated activity were considered to be active. The recovery rate (defined as the probability of detecting an active dipole in its correct location) for both methods is shown in Table 2. The proposed method is able to detect up to 5 active dipoles with a near perfect recovery rate while the performance of the ℓ_{21} norm method starts decreasing at $P = 3$.

It is important to note that the price to pay with the proposed method is its computational complexity. One simulation of the previous examples was processed in 6 seconds with a modern Xeon CPU E3-1240 @ 3.4GHz processor (using a Matlab implementation with MEX files written in C) against 104 milliseconds for the ℓ_{21} mixed norm. However, also note that the ℓ_{21} norm approach requires running the algorithm multiple times to adjust the regularization parameter by cross-validation.

	1 - 2	3	4	5	6	7
PM	100%	100%	100%	98.8%	84.0%	65.1%
ℓ_{21}	100%	97.3%	93.5%	78.8%	61.7%	49.1%

Table 2: Recovery rate as a function of P for the proposed method and the weighted ℓ_{21} norm (computed with 50 Monte Carlo runs).

6. CONCLUSION

This paper introduced a new hierarchical Bayesian model for EEG source localization promoting structured sparsity using a multivariate Bernoulli Laplacian prior. A partially collapsed Gibbs sampler was developed to draw samples from its posterior distribution. A specific Metropolis-Hastings move (called dual dipole random shift) was also introduced in order to speed up the algorithm convergence. The generated samples were used to estimate the source activity and the model hyperparameters jointly in an unsupervised framework. The resulting algorithm was compared to the ℓ_{21} mixed norm regularization showing promising results for synthetic data composed by point-like source activations. More precisely, the proposed method showed better detection results and a better recovery of the activation waveforms for small SNRs, while avoiding the amplitude underestimation observed with the ℓ_{21} approach. In addition, the proposed method presented a better recovery rate for different amounts of active dipoles. The method is currently being applied to real data and is already showing promising results which will be published in the near future. Future work will try to generalize the method to practical cases where $\mathcal{H}$ is only partially known.

7. REFERENCES

[1] 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. 5, no. 1, pp. 25, 2008.

[2] E. J. Candes, “The restricted isometry property and its implications for compressed sensing,” *C. R. l'Académie des Sciences*, vol. 346, no. 9, pp. 589–592, 2008.

[3] K. Uutela, M. Hämäläinen, and E. Somersalo, “Visualization of magnetoencephalographic data using minimum current estimates,” *NeuroImage*, vol. 10, no. 2, pp. 173–180, 1999.

[4] A. Gramfort, M. Kowalski, and M. Hämäläinen, “Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods,” *Phys. Med. Biol.*, vol. 57, no. 7, pp. 1937, 2012.

[5] J. Huang and T. Zhang, “The benefit of group sparsity,” *Ann. Statist.*, vol. 38, no. 4, pp. 1978–2004, Aug. 2010.

[6] 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.

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

[8] 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.