Sparse EEG Source Localization using Bernoulli Laplacian Priors

Facundo Costa, *Student Member, IEEE*, Hadj Batatia, *Member, IEEE*, Lotfi Chaari, *Member, IEEE*, and Jean-Yves Tourneret, *Senior Member, IEEE*

***Abstract*—Source localization in electroencephalography has received an increasing amount of interest in the last decade. Solving the underlying ill-posed inverse problem usually requires choosing an appropriate regularization. The usual ℓ_2 norm has been considered and provides solutions with low computational complexity. However, in several situations, realistic brain activity is believed to be focused in a few focal areas. In these cases, the ℓ_2 norm is known to overestimate the activated spatial areas. One solution to this problem is to promote sparse solutions for instance based on the ℓ_1 norm that are easy to handle with optimization techniques. In this paper, we consider the use of an $\ell_0 + \ell_1$ norm to enforce sparse source activity (by ensuring the solution has few non-zero elements) while regularizing the non-zero amplitudes of the solution. More precisely, the ℓ_0 pseudo-norm handles the position of the non-zero elements while the ℓ_1 norm constrains the values of their amplitudes. We use a Bernoulli-Laplace prior to introduce this combined $\ell_0 + \ell_1$ norm in a Bayesian framework. The proposed Bayesian model is shown to favor sparsity while jointly estimating the model hyperparameters using a Markov chain Monte Carlo sampling technique. We apply the model to both simulated and real EEG data, showing that the proposed method provides better results than the ℓ_2 and ℓ_1 norms regularizations in the presence of pointwise sources. A comparison with a recent method based on multiple sparse priors is also conducted.**

***Index Terms*—EEG, inverse problem, source localization, sparse Bayesian restoration, Markov chain Monte Carlo, $\ell_0 + \ell_1$ norm regularization**

I. Introduction

EEG source localization is an ill-posed inverse problem that has been receiving an intensive attention in the literature and has been solved using several interesting methods. These methods can be classified into two main groups based on different models: (i) the dipole-fitting models that try to estimate the localization of a small number of dipoles; and (ii) the distributed-source models that assume that there is a large number of dipoles and try to estimate their amplitudes and orientations. The dipole-fitting models [1, 2] assume that the brain activity is concentrated in a small-number of point-like sources and estimate the localization, amplitude and orientation of a few dipoles (usually no more than ten) to explain the measured data. A particularity of these models is that they lead to solutions that can vary extremely with the initial guess about the number of dipoles, their locations and their orientations because of the existence of many local minima in the optimized cost function [3]. To solve this problem, the MUSIC algorithm [4] and its variants R-MUSIC [5], RAP-MUSIC [6] and FINES [7] were developed. Another recent dipole-fitting model whose parameters are estimated using sequential Monte Carlo [8] formulates the EEG source localization as a semi-linear problem due to the measurements having a linear dependency with respect to the dipole amplitudes and a non-linear one with respect to the positions. If few and clustered sources are present in the underlying brain activity, the dipole-fitting algorithms generally yield good results [9, 10]. However, the performance of these algorithms can be altered in the case of multiple spatially extended sources [3]. On the other hand, the distributed source models represent the brain activity as a big number of dipoles (usually between 1.000 and 50.000) with fixed positions (that occupy either the entire volume or just the cortical surface of the brain), and estimate their amplitudes and orientations. Since the amount of dipoles is much larger than the number of electrodes, the problem to be solved is under-determined and has an infinite number of possible brain activity distributions associated with the measured data. The EEG source localization problem is therefore known to be ill-posed [3].

In order to choose from the possible solutions of the EEG source localization problem, a regularization trying to enforce desirable properties is usually applied. The minimum norm estimation algorithm [11] tries to minimize the ℓ_2 norm of the solution favoring low power activity estimations. However, the resulting model also prefers weak surface-power over strong depth-power [3]. Several algorithms have tried to correct this problem including the weighted minimum norm, Loreta [12] and sLoreta [13]. Among these solutions, sLoreta can estimate the maximum excitation point with zero localization error in noiseless conditions [3]. A drawback of these algorithms is that they require to adjust a regularization parameter, which is usually performed with empirical methods such as cross-validation, Bayesian estimation [14] or the L-curve method [15]. Another problem is that they overestimate the active spatial area in the cortex when the actual source activity is focal [3]. Motivated by the fact that realistic source activity is likely to be associated with few active brain regions at the same time, some algorithms that encourage sparse solutions have been developed. An ℓ_0 pseudo-norm can be used to provide sparse solutions [16]. However, since the minimization of this norm requires a very expensive combinatory search, it is usually approximated by an ℓ_1 norm [17, 18] that is easier to handle with classical optimization techniques due to its convexity. These two types of regularizations have been shown to be equivalent under certain conditions [16].

In addition to the optimization techniques mentioned above it is also possible to estimate the brain activity by modeling its time evolution and applying Kalman filtering [19, 20] or

particle filters [21–23]. Another way of looking for a solution with sparse properties is to consider Bayesian techniques with appropriate priors assigned to the model parameters. These Bayesian techniques have the advantage of estimating the model hyperparameters directly from the data using hierarchical algorithms. Note that Bayesian methods have been used both for dipole-fitting [24–27] and distributed source models [28–30]. In [29] Friston *et al.* developed the multiple sparse priors (MSP) framework in which they parcel the brain in different pre-selected regions and promote a sparse subset of these regions to be active. This is performed by using a Gaussian prior distribution for the brain activity and estimating its correlation matrix as a linear combination of pre-defined base matrices with finite support defined over each brain parcel. This method is designed to estimate brain activities that have multiple spatially extended sources and requires choosing a criterion for the prior parcellation of the brain in different regions and the selection of the base matrices. Even though the results of the MSP framework proved to be good in several scenarios, since our focus is to estimate point-like focal source activity we propose to consider separate dipoles instead of grouping them together in different pre-selected regions. This avoids the need to select a criterion for brain parcellation and constructing the base matrices required by the MSP framework.

For these reasons we propose to introduce a new algorithm relying on the ℓ_1 norm and ℓ_0 pseudo norm regularizations. Since the ℓ_0 pseudo-norm is too costly for direct implementation and the ℓ_1 norm does not always yield the same solution, we propose to combine them into a Bayesian framework to pursue sparse solutions. It can be easily shown that using a Laplace prior is the Bayesian equivalent of ℓ_1 norm regularization whereas a Bernoulli prior can be associated with the ℓ_0 pseudo-norm. As a consequence, we propose the use of a Bernoulli-Laplace prior in a distributed source model for the estimation of the activity of focal point-like sources. The combination of the two norms allows the non zero elements to be localized (via the Bernoulli part of the prior) and their amplitudes to be estimated (with the Laplace distribution). Note that the Laplace distribution prior is able to estimate both small and high amplitudes due to its large value around zero and its fat tails. In addition one could introduce a spatial-regularizing prior to yield a region-based sparse solution equivalent to the MSP model without the preprocessing and associated hypotheses.

In order to compute estimators associated with the proposed Bayesian model, we derive a Markov chain Monte Carlo (MCMC) method allowing samples to be generated according to the posterior of interest. These samples are then used to estimate the unknown model parameters. The resulting algorithm has several attractive properties, including the fact that it is able to estimate the model hyperparameters in an unsupervised framework and provides better estimations of the source activity than the traditional ℓ_2 or ℓ_1 norms.

The paper is organized as follows: The observation model is introduced in Section II. The Bayesian model proposed to estimate the unknown parameters is defined in Section III. Section IV studies a hybrid Gibbs sampler that will be used to generate samples asymptotically distributed according to the posterior of interest. Simulation results obtained with synthetic and real EEG data are reported in Section V. Section VI concludes the paper.

II. Problem statement

The main objective of the EEG source localization problem is to estimate the brain activity of the patient from EEG measurements. It is well known that the activity can be represented as a continuous current distribution [31] which can be approximated by a discrete number of dipoles located in the brain [3, 31]. Therefore we propose to model the brain activity by using a finite number of active dipoles located in the cortex. More precisely, we consider a distributed-source model with a fixed amount of dipoles with given locations. Since the electric potential at any point of the scalp can be calculated as a linear combination of the dipole amplitudes, the relationship between the potential at the scalp and the dipole amplitudes can be represented as follows [3, 31]

$$\boldsymbol{y} = \mathcal{H}\boldsymbol{x} + \boldsymbol{e} \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^{3M}$ contains the amplitudes of the M dipoles along the three spatial dimensions, $\boldsymbol{y} \in \mathbb{R}^N$ is the measurement vector acquired by the N electrodes, the $N \times 3M$ lead field matrix $\mathcal{H}$ models the propagation of the electromagnetic field from the sources to the sensors [24, 32] and $\boldsymbol{e}$ is an additive white Gaussian noise. Since the activity measured by an EEG is mainly generated by groups of neurons that are oriented orthogonal to the brain surface [31, 33], we assume the source orientations to be normal to the cortex. Moreover, as in many works [3, 28, 34], we limit the position of the dipoles to the cortex surface. As a consequence, the moment of each dipole can be determined using only one value, such that $\boldsymbol{x} \in \mathbb{R}^M$ and $\mathcal{H} \in \mathbb{R}^{N\times M}$. Note that even though the orientations of the dipoles are constrained, the directions of their currents are not known and have to be estimated by the algorithm (via the signs of the elements of $\boldsymbol{x}$) [3, 34].

The EEG source localization problem based on the observation model (1) mainly consists of estimating the vector $\boldsymbol{x}$ from the observed data $\boldsymbol{y}$. The next section introduces a hierarchical Bayesian model appropriate to solve this inverse problem. The likelihood of this model and the priors assigned to its unknown parameters and hyper-parameters are defined in the next sections.

III. Bayesian model

A. Likelihood

It is very classical in EEG analysis to consider an additive white Gaussian noise with a variance σ_n^2 [3]. When this assumption does not hold, it is classical to estimate the noise covariance matrix from the data and to whiten the data before applying the algorithm [35]. This strategy will be considered in our experiments related to real data presented in Section V-C. The Gaussian noise assumption for the noise samples leads to the following probability density function (pdf)

$$f(\boldsymbol{y}|\boldsymbol{x}, \sigma_n^2) = \left(\frac{1}{2\pi\sigma_n^2}\right)^{\frac{N}{2}} \exp\Big(-\frac{||\boldsymbol{y} - \mathcal{H}\boldsymbol{x}||^2}{2\sigma_n^2}\Big) \tag{2}$$

where $||.||$ denotes the Euclidean norm.

B. Prior Distributions

The unknown parameter vector associated with the proposed model (1) is $\boldsymbol{\theta} = \{\boldsymbol{x}, \sigma_n^2\}$. In order to perform Bayesian inference, we assign priors to these parameters as follows.

Prior for the dipole amplitudes

In order to encourage sparse solutions whose non zero elements have small amplitudes, we introduce the Bayesian analogous of an $\ell_0 + \ell_1$ regularization by using a Bernoulli-Laplace prior distribution for each element of the vector $\boldsymbol{x}$. The corresponding pdf for the ith element of $\boldsymbol{x}$ is

$$f(x_i|\omega, \lambda) = (1-\omega)\delta(x_i) + \frac{\omega}{2\lambda}\exp\left(-\frac{|x_i|}{\lambda}\right) \tag{3}$$

where $\delta(.)$ is the Dirac delta function, λ is the parameter of the Laplace distribution, and ω the weight balancing the effects of the Dirac delta function and the Laplace distribution. Assuming the random variables x_i are *a priori* independent, the prior distribution of $\boldsymbol{x}$ can be written

$$f(\boldsymbol{x}|\omega, \lambda) = \prod_{i=1}^{M} f(x_i|\omega, \lambda). \tag{4}$$

Prior for the noise variance

A non-informative Jeffrey's prior is assigned to the noise variance

$$f(\sigma_n^2) \propto \frac{1}{\sigma_n^2} 1_{\mathbb{R}^+}(\sigma_n^2) \tag{5}$$

where $1_{\mathbb{R}^+}(\xi) = 1$ if $\xi \in \mathbb{R}^+$ and 0 otherwise. This prior is a very classical choice for a non-informative prior (see, e.g., [36] for motivations). Note that a more informative *prior* distribution related to the signal to noise ratio could also be investigated. However, we have chosen here to design an algorithm requiring as little prior information as possible.

C. Hyperparameter priors

The hyperparameter vector associated with the previous priors is $\boldsymbol{\Phi} = \{\omega, \lambda\}$ as displayed in the direct acyclic graph of Fig. 1. This paper considers a hierarchical Bayesian model which allows the hyperparameters to be estimated from the data. This strategy requires to assign priors to the hyperparameters (referred to as hyperpriors) that are defined in this section.

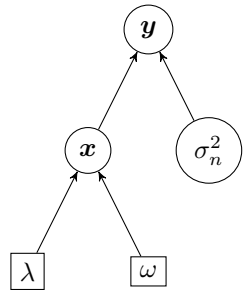


Fig. 1: Directed acyclic graph of the hierarchy used for the Bayesian model.

Hyperprior for ω

An independent uniform distribution on $[0, \omega_{\max}]$ is assigned to the weight ω

$$\omega \sim \mathcal{U}_{[0,\omega_{\max}]} \tag{6}$$

where $\omega_{\max} \in [0, 1]$ is an upper bound on ω that is fixed in order to ensure a minimum level of sparsity.[1].

Hyperprior for λ

Using similar arguments as for the noise variance σ_n^2, a Jeffrey's prior is assigned to λ in order to define the following non-informative prior

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

D. Posterior distribution

Taking into account the likelihood and priors introduced above, the joint posterior distribution of the model used for EEG source localization can be expressed using the following hierarchical structure

$$f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{\Phi}) f(\boldsymbol{\Phi}) \tag{8}$$

where $\{\boldsymbol{\theta}, \boldsymbol{\Phi}\}$ is the vector containing the model parameters and hyperparameters. Because of the complexity of this posterior distribution, the Bayesian estimators of $\{\boldsymbol{\theta}, \boldsymbol{\Phi}\}$ cannot be computed with simple closed-form expressions. Section IV studies an MCMC method that can be used to sample the joint posterior distribution (8) and build Bayesian estimators of the unknown model parameters using the generated samples.

IV. Markov chain Monte Carlo method

This section considers a Gibbs sampler [36] which generates samples iteratively from the conditional distributions of (8), *i.e.*, from $f(\sigma_n^2|\boldsymbol{y}, \boldsymbol{x})$, $f(\lambda|\boldsymbol{x})$, $f(\omega|\boldsymbol{x})$ and $f(x_i|\boldsymbol{y}, \boldsymbol{x}_{-i}, \omega, \lambda, \sigma_n^2)$. Note that this strategy showed interesting results for many image processing problems including image denoising [37], sparse image reconstruction [38], hyperspectral image unmixing [39] and fusion of hyperspectral and panchromatic images [40]. The next sections explain how to sample from the conditional distributions of the unknown parameters and hyperparameters associated with the posterior of interest (8). The resulting algorithm is also summarized in Algorithm 1.

A. Sampling according to $f(\sigma_n^2|\boldsymbol{y}, \boldsymbol{x})$

Using (8), it is straightforward to derive the conditional posterior distribution of σ_n^2 which is the following inverse gamma distribution

$$\sigma_n^2|\boldsymbol{x}, \boldsymbol{y} \sim \mathcal{IG}\left(\sigma_n^2 \middle| \frac{N}{2}, \frac{||\boldsymbol{y} - \mathcal{H}\boldsymbol{x}||^2}{2}\right) \tag{9}$$

where $||.||$ is the the Euclidean norm.

[1]We have observed that setting $\omega_{\max} = 0.5$ (instead of $\omega_{\max} = 1$) yields faster convergence of the sampler studied in Section IV.

Algorithm 1 Gibbs sampler.

Initialize $\boldsymbol{x}$ with the sLoreta solution
repeat
  Sample σ_n^2 according to (9).
  Sample λ according to (10).
  Sample ω according to (11).
  for $i = 1$ to M **do**
    Sample x_i according to (12).
  end for
until convergence

B. Sampling according to $f(\lambda|\boldsymbol{x})$

By using $f(\boldsymbol{x}|\omega, \lambda)$ and the prior distribution of λ, it is easy to derive the conditional distribution of λ which is also an inverse gamma distribution

$$\lambda|\boldsymbol{x} \sim \mathcal{IG}\Big(\lambda \mid ||\boldsymbol{x}||_0, ||\boldsymbol{x}||_1\Big) \tag{10}$$

where $||.||_0$ and $||.||_1$ are the ℓ_0 and ℓ_1 norms, respectively.

C. Sampling according to $f(\omega|\boldsymbol{x})$

Using $f(\boldsymbol{x}|\omega, \lambda)$ and the prior of ω it can be shown that the conditional distribution of ω is a truncated Beta distribution defined on the interval $[0, \omega_{\text{max}}]$

$$\omega|\boldsymbol{x} \sim \mathcal{B}_{[0,\omega_{\text{max}}]}(1 + ||\boldsymbol{x}||_0, 1 + M - ||\boldsymbol{x}||_0). \tag{11}$$

D. Sampling according to $f(x_i|\boldsymbol{y}, \boldsymbol{x}_{-i}, \omega, \lambda, \sigma_n^2)$

Using the likelihood and the prior distribution of $\boldsymbol{x}$, the conditional distribution of each signal element x_i can be expressed as follows

$$\begin{aligned} f(x_i|\boldsymbol{y}, \boldsymbol{x}_{-i}, \omega, \lambda, \sigma_n^2) = \omega_{1,i}\delta(x_i) + \omega_{2,i}\mathcal{N}_+(\mu_{i,+}, \sigma_i^2) \\ + \omega_{3,i}\mathcal{N}_-(\mu_{i,-}, \sigma_i^2) \end{aligned} \tag{12}$$

where $\mathcal{N}_+$ and $\mathcal{N}_-$ denote the truncated Gaussian distributions on $\mathbb{R}^+$ and $\mathbb{R}^-$, respectively. The vector $\boldsymbol{x}$ can be decomposed on the orthonormal basis $B = \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_M\}$ such that $\boldsymbol{x} = \widetilde{\boldsymbol{x}}_{-i} + x_i\boldsymbol{e}_i$ where $\widetilde{\boldsymbol{x}}_{-i}$ is obtained by setting the i^{th} element of $\boldsymbol{x}$ to 0. Defining $\boldsymbol{v}_i = \boldsymbol{y} - \mathcal{H}\widetilde{\boldsymbol{x}}_{-i}$ and $\boldsymbol{h}_i = \mathcal{H}\boldsymbol{e}_i$, the weights $(\omega_{l,i})_{1\leq l\leq 3}$ are defined as

$$\omega_{l,i} = \frac{u_{l,i}}{\sum_{l=1}^{3} u_{l,i}} \tag{13}$$

where

$$\begin{aligned} u_{1,i} &= 1 - \omega \\ u_{2,i} &= \frac{\omega}{2\lambda} \exp\left(\frac{(\mu_i^+)^2}{2\sigma_i^2}\right)\sqrt{2\pi\sigma_i^2}C(\mu_i^+, \sigma_i^2) \\ u_{3,i} &= \frac{\omega}{2\lambda} \exp\left(\frac{(\mu_i^-)^2}{2\sigma_i^2}\right)\sqrt{2\pi\sigma_i^2}C(-\mu_i^-, \sigma_i^2) \end{aligned} \tag{14}$$

and

$$\begin{aligned} \sigma_i^2 &= \frac{\sigma_n^2}{||\boldsymbol{h}_i||^2} \\ \mu_i^+ &= \sigma_i^2\left(\frac{\boldsymbol{h}_i^\intercal\boldsymbol{v}_i}{\sigma_n^2} - \frac{1}{\lambda}\right) \\ \mu_i^- &= \sigma_i^2\left(\frac{\boldsymbol{h}_i^\intercal\boldsymbol{v}_i}{\sigma_n^2} + \frac{1}{\lambda}\right) \\ C(\mu, \sigma^2) &= \frac{1}{2}\left[1 + \text{erf}\left(\frac{\mu}{\sqrt{2\sigma^2}}\right)\right]. \end{aligned} \tag{15}$$

Finally, it is interesting to mention how to sample efficiently from (12), which will also be useful to define appropriate estimators (see Section IV-E). Since (12) is a mixture of three distributions, we introduce a discrete random variable b_i taking the value 0 with probability $\omega_{1,i}$ (corresponding to $x_i = 0$), the value 1 with probability $\omega_{2,i}$ (corresponding to $x_i > 0$), and the value -1 with probability $\omega_{3,i}$ (corresponding to $x_i < 0$). To sample from (12), we can draw the discrete random variable b_i, and generate x_i conditionally upon b_i. The role of the discrete random variable b_i is to detect whether x_i is zero, positive or negative.

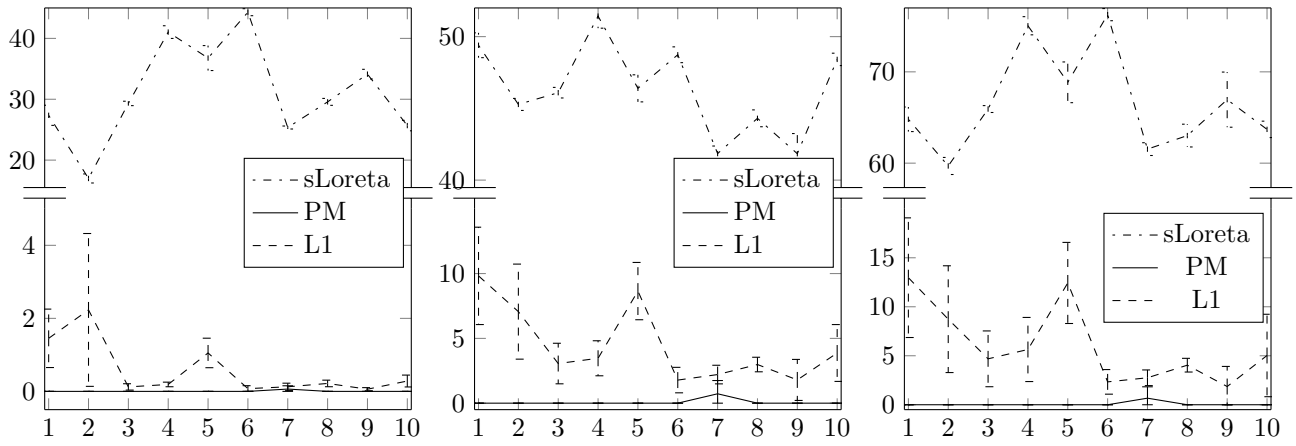
E. Parameter estimation

In order to estimate the different parameters associated with the posterior (8) using the samples generated by the MCMC method described above, we first detect the non-zero values of x_i corresponding to the non-zero dipole amplitudes. This detection can be easily achieved by using the generated discrete random variables b_i used to sample according to x_i. More precisely, we estimate b_i following the maximum a posteriori principle as the most likely value generated by the sampler[2]. Once we have detected whether x_i is zero, positive or negative (*i.e.*, once we have estimated b_i), we keep the vectors generated by the MCMC method satisfying these conditions for the different variables b_i and we estimate the dipole amplitudes, the noise variance and the hyperparameters using the mean of these vectors. In other words, we estimate the indicators b_i by the maximum a posteriori principle and the amplitudes x_i by the minimum mean square error principle (MMSE) which corresponds to the mean of the posterior.

V. Experimental results

This section reports different experiments conducted to evaluate the performance of the proposed EEG source localization algorithm for synthetic and real data. In these experiments, our algorithm was initialized with the sLoreta solution obtained after estimating the regularization parameter by minimizing the cross-validation error as recommended by Pascual-Marqui *et al.* in [13]. The upper bound of the sparsity level was set to $\omega_{\text{max}} = 0.5$, which is much larger than the expected value of ω. The results obtained with synthetic and real data are reported in two separate sections.

[2]We have noted that when the number of zeros is close to the number of non-zeros, it can be interesting to promote the non zero value which will result in estimating a small dipole amplitude.



(a) Center-mass localization error (mm) (b) Excitation extension (mm) (c) Transportation cost (mm)

Fig. 2: Simulation results for single dipole experiments. The horizontal axis indicates the activation number. The error bars show the standard deviation over 12 Monte Carlo runs.

A. Synthetic data

Simulation scenario

Synthetic data with few pointwise source activations were generated using a realistic BEM head model with 19 electrodes placed according to the $10-20$ international system of electrode placement. Three different kinds of activations were investigated: (i) single dipole activations, (ii) multiple distant dipole activations and (iii) multiple close dipole activations. The default subject anatomy of the Brainstorm software [41] was considered. This model corresponds to MRI measurements of a 27 year old male using the boundary element model implemented by the OpenMEEG package [42]. The default brain cortex of this subject was downsampled to generate a 1003-dipole head model. These dipoles are distributed along the cortex surface and have an orientation normal to it, as discussed in the previous sections. The resulting head model is such that $\mathcal{H} \in \mathbb{R}^{19\times 1003}$.

For each of the activations that are described in what follows, three independent white Gaussian noise realizations were added to the observed signal $\boldsymbol{Hx}$, resulting in three sets of measurements $\boldsymbol{y}$ with an SNR of 20dB. For each of these three noisy signals $\boldsymbol{y}$, four MCMC independent chains were run, resulting in 12 total simulations for each of the considered activations $\boldsymbol{x}$. The ℓ_1 and sLoreta methods were applied to the same three different sets of measurements $\boldsymbol{y}$ resulting in 3 simulations for each activation.

Performance evaluation

To assess the quality of the localization results, the following indicators were used

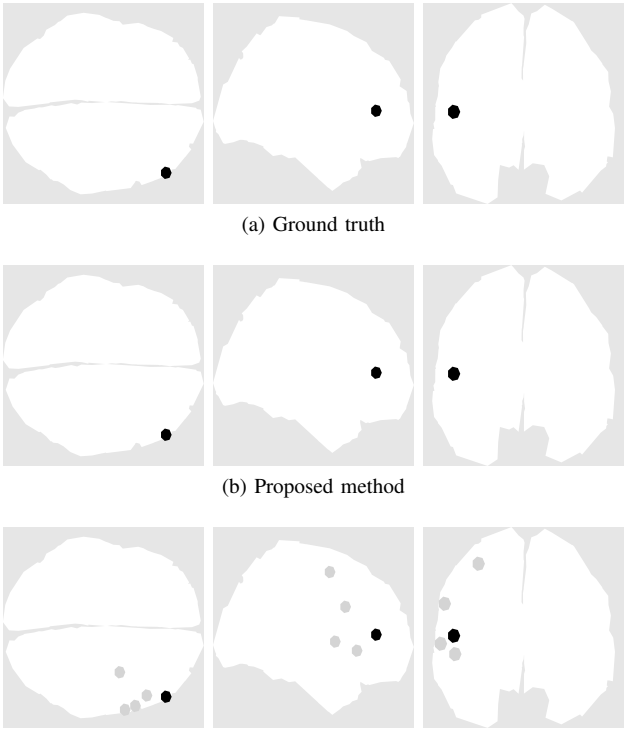
- Localization error [3]: The euclidean distance between the maximum of the estimated activity and the real source location is used to determine whether the algorithm is able to find the point of highest activity correctly or not.
- Center-mass localization error [43]: The euclidean distance between the barycenter of the estimated activity and the real source location allows us to appreciate if the activity estimated by the algorithm is centered around the correct point or if it is biased.
- Excitation extension: The spatial extension of the spatial area of the brain cortex that is estimated to be active was considered in [43]. Since the synthetic data only contains pointwise sources, this criterion should ideally be equal to zero.
- Transportation cost: This indicator evaluates the performance in a multiple dipole situation where the activity estimates from different dipoles may overlap. It is computed as the solution of an optimal mass transport optimization problem [44] considering the known ground truth to be the initial mass distribution and the activity estimated by the algorithm to be the target mass distribution. The activities associated with the ground truth and the estimated data are first normalized. The total cost of moving the activity from the non zero elements of the ground truth to the non zero elements of the estimated activity is then computed. It is obtained by finding the weights $w_{j\to k}$ that minimize (16)

$$\sum_j \sum_k w_{j\to k} |\boldsymbol{r}_{x_j^{\text{nz}}} - \boldsymbol{r}_{\widehat{x}_k^{\text{nz}}}| \tag{16}$$

where x_j^{nz} denotes a non-zero element of the ground truth, $\widehat{x}_k^{\text{nz}}$ is a non-zero element of the estimated activity and $\boldsymbol{r}_d$ represents the spatial position of dipole d in euclidean coordinates, with the following constraints for the weights (in order to avoid the trivial zero solution)

$$\sum_j w_{j\to k} = \widehat{x}_k^{\text{nz}}, \quad \sum_k w_{j\to k} = x_j^{\text{nz}}. \tag{17}$$

Note that (16) defines a similarity measure between the amount of activity in the non-zero elements of the ground truth and the estimated solution. The minimum cost of

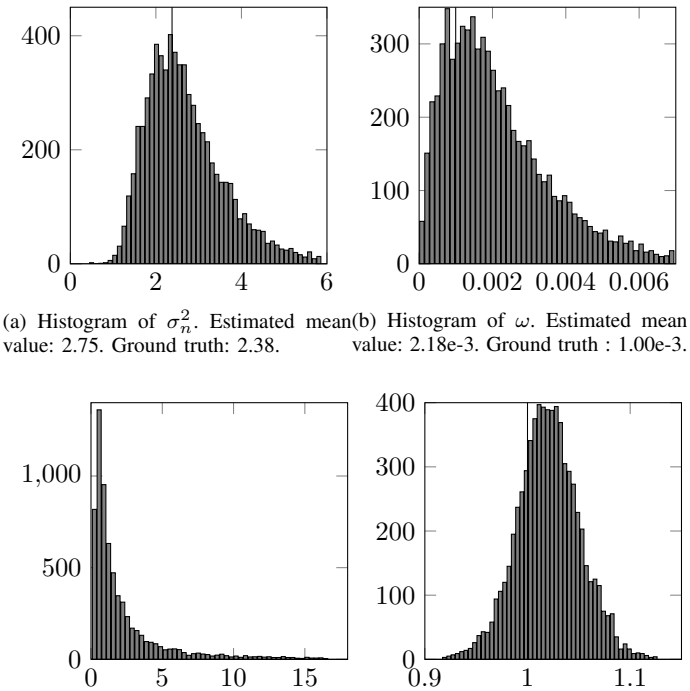


(a) Ground truth

(b) Proposed method

(c) Weighted ℓ_1 norm regularization

Fig. 3: Brain activity for one single dipole experiment (activation #5).



(a) Histogram of σ_n^2. Estimated mean value: 2.75. Ground truth: 2.38.

(b) Histogram of ω. Estimated mean value: 2.18e-3. Ground truth : 1.00e-3.

(c) Histogram of λ. Estimated mean value: 8.36.

(d) Histogram of x_i. Estimated mean value: 1.02. Ground truth: 1.00.

Fig. 4: Histogram of samples generated by the MCMC method for one of the single dipole simulations. The estimated mean values and ground truth are indicated in the figures.

(16) can be obtained using the simplex method of linear programming. The transportation cost of an estimated solution is finally defined as the minimum transportation cost calculated between the estimated solution and the ground truth. Since the activity has been normalized in the first step, this parameter is measured in millimeters.

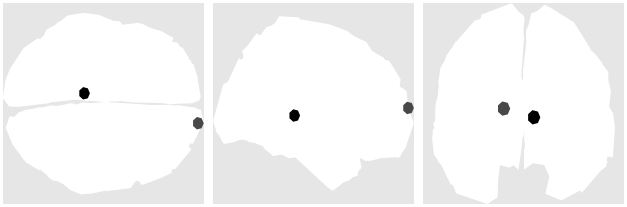
The proposed method is compared to the more traditional weighted ℓ_1 norm [45] and sLoreta. The regularization parameter of sLoreta was computed by cross-validation using the method recommended by Pascual-Marqui *et al.* in [13]. The weighted ℓ_1 norm was implemented using the alternating direction method of multipliers (ADMM) with the technique used by Boyd *et al.* in [46]. The regularization parameter was chosen so that $||\boldsymbol{y} - \mathcal{H}\widehat{\boldsymbol{x}}|| \approx ||\boldsymbol{y} - \mathcal{H}\boldsymbol{x}||$ according to the discrepancy principle [47].

Single dipole

The first kind of experiment consisted of activating only one dipole. Ten dipoles were randomly chosen from the 1003 dipoles of the head model. For each of these ten positions a measurement vector $\boldsymbol{x}$ was formed that had only the corresponding dipole activated (referred as single dipole activations #1 through #10). After generating the different sets of measurements $\boldsymbol{y}$, the localization error was found to be 0.00mm for all the dipoles with the three methods. The other performance parameters are displayed in Fig. 2 showing the good performance of the proposed method (indicated by PM).

The brain activities detected by the proposed method and the weighted ℓ_1 norm solution are illustrated in Fig. 3 for a representative simulation. Our algorithm managed to perfectly recover the activity for 9 out of the 10 activations (with center-mass localization error, extension and transportation cost equal to 0.00mm) for all simulations. The dipole corresponding to activation #7 was located precisely for some simulations and with very reduced error in the others. The mean transportation cost of the proposed method for activation #7 is 0.7mm. In comparison, sLoreta has an excitation extension that is significantly larger (between 41mm and 51mm for the different dipole positions) (as expected for an ℓ_2 norm regularization) and a larger transportation cost (bigger than 55mm in all cases). The weighted ℓ_1 norm regularization provides better estimations than sLoreta with a mean extension of up to 10mm and transportation cost of up to 13mm but is still outperformed by the proposed method.

In addition, Fig. 4 shows the histograms of the samples of σ_n^2, ω, λ and the amplitude of the active dipole x_i for one of the simulations corresponding to activation #3 as a representative case. It is shown that the ground truth values of ω (the proportion of non-zeros $1/1003$), σ_n^2 and $\mathbf{x}_i$ are inside the support of their histograms and are close to their estimated mean values, *i.e.*, close to their minimum mean square error estimates.

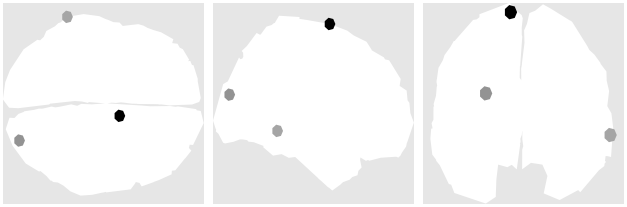


(a) Ground truth

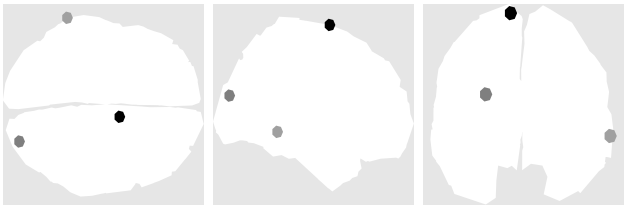
(b) Proposed method

(c) Weighted ℓ_1 norm regularization

Fig. 5: Brain activity for a multiple distant dipole experiment (activation # 2 that has two active dipoles).



(a) Ground truth



(b) Proposed method

(c) Weighted ℓ_1 norm regularization

Fig. 6: Brain activity for a multiple distant dipole experiment (activation # 3 that has three active dipoles).

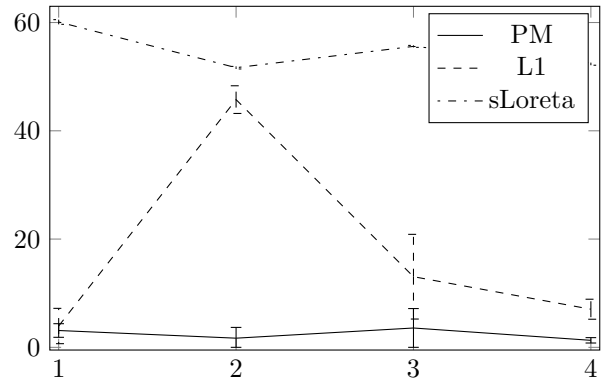


Fig. 7: Transportation cost for multiple distant dipoles experiments, the horizontal axis indicates the activation number (1 and 2: two active dipoles, 3 and 4: three active dipoles). The error bars show the standard deviation over 12 Monte Carlo runs.

Multiple distant dipoles

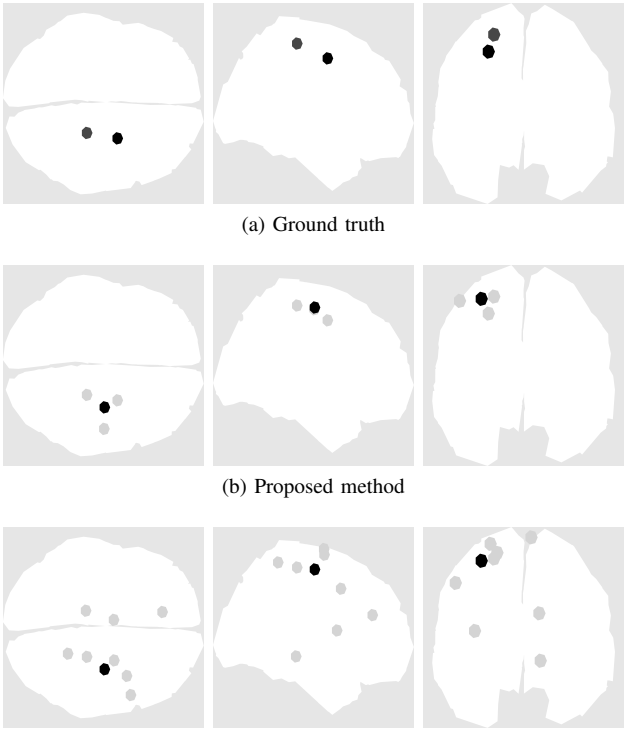
The second kind of experiments evaluates the performance of the proposed algorithm when several dipoles are activated at the same time and when these activated dipoles have distant space positions. More precisely, we chose randomly the following sets of dipoles from the 1003 dipoles present in the head model

- (i) Two pairs of $N = 2$ simultaneous dipoles spaced more than 100mm (activations #1 and #2).
- (ii) Two sets of $N = 3$ simultaneous dipoles spaced more than 100mm (activations #3 and #4).

For each of these four sets of dipoles, a vector $\boldsymbol{x}$ having only the corresponding active dipoles was formed and the measurements $\boldsymbol{y}$ were calculated as described previously.
The brain activities associated with two representative simulations corresponding to two and three dipoles are illustrated in Figs. 5 and 6. The activation #2 associated with two distant dipoles displayed in Fig. 5 is an interesting case for which the weighted ℓ_1 norm regularization fails completely to recover one of the dipoles. The activation #4 displayed in Fig. 6 shows that the proposed method detects an activity more concentrated in the activated dipoles while the ℓ_1 norm regularization provides less-sparse solutions.
Quantitative results in terms of transportation costs are displayed in Fig. 7 for the different experiments. The transportation costs obtained with the proposed method are below 3.6mm in all cases and are clearly smaller than those obtained with the other methods. Indeed, the sLoreta transportation costs are between 50 and 62mm, and the transportation costs associated with the weighted ℓ_1 norm regularization are between 3.9mm and 13mm, except for the activation #2 where it fails to recover one of the dipoles as previously stated.

Multiple close dipoles

The third kind of experiments evaluates the performance of the proposed algorithm for active dipoles that have close

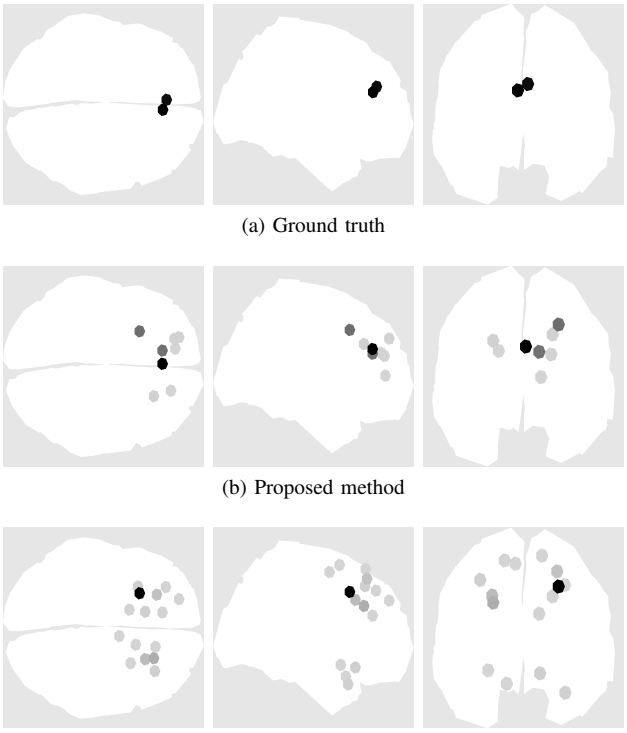


(a) Ground truth

(b) Proposed method

(c) Weighted ℓ_1 norm regularization

Fig. 8: Brain activity for multiple close dipoles (activation # 4 that has a 30mm separation between dipoles).



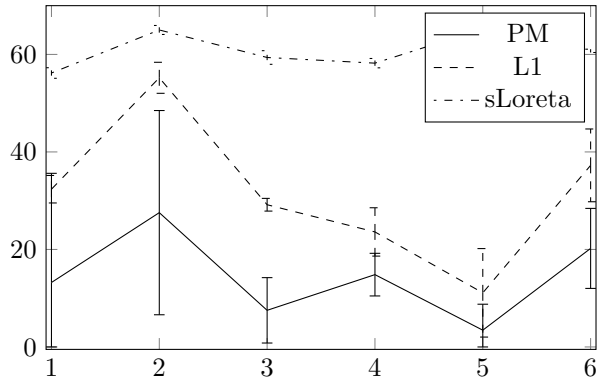
(a) Ground truth

(b) Proposed method

(c) Weighted ℓ_1 norm regularization

Fig. 9: Brain activity for multiple close dipoles (activation # 6 that has a 10mm separation between dipoles).

spatial positions. More precisely, we randomly chose the following sets of dipoles



(a)

Fig. 10: Transportation cost for multiple close dipoles experiments. The horizontal axis indicates the activation number (1 and 2: 50mm separation, 3 and 4: 30mm separation, 5 and 6: 10mm separation). The error bars show the standard deviation over 12 Monte Carlo runs.

- (i) Two pairs of dipoles spaced between 49 and 51mm (activations #1 and #2).
- (ii) Two pairs of dipoles spaced between 29 and 31mm (activations #3 and #4).
- (iii) Two pairs of dipoles spaced between 9 and 11mm (activations #5 and #6).

For each of these six sets of dipoles, a vector $\boldsymbol{x}$ having only the corresponding active dipoles was formed and the measurements $\boldsymbol{y}$ were simulated as described previously.
Figure 10 compares the transportation costs obtained with the different methods. Since it is much harder to distinguish the activity produced by two close dipoles, the transportation costs associated with the proposed method and the weighted ℓ_1 norm regularization are considerably higher than those obtained previously. However, the transportation costs obtained with the proposed algorithm are still below those obtained with the two other estimation strategies. Some interesting cases can be observed in Figs. 8 and 9. Figure 8 corresponds to one case where both algorithms fail to identify two dipoles and fuse them into a single dipole located in the middle of the two actual locations. In this particular activation, our algorithm adds considerably less extra activity than the weighted ℓ_1 norm regularization. In the case illustrated in Fig. 9, the proposed method correctly identifies two dipoles (but moves one of them from its original positions) while the weighted ℓ_1 norm regularization estimates a single dipole located very far from the two excited dipoles.

B. Computational cost

This section compares the computational cost of the different algorithms (weighted ℓ_1 norm, sLoreta and the proposed method) that were run on a modern Xeon CPU E3-1240 @ 3.4GHz processor using a Matlab implementation. In the case of the weighted ℓ_1 norm, we used the stopping criterion recommended by Boyd *et al.* in [46]. The proposed method was run with four parallel chains (as stated in section V-A) until the potential scale reduction factor (PSRF) of all the

Experiment type	sLoreta	ℓ_1 norm	Proposed method
Single dipole	3.65×10^{-3}	1.09	651.9
Multiple distant dipole	3.64×10^{-3}	0.84	886.6
Multiple close dipole	3.61×10^{-3}	0.91	1128.0

TABLE I: Computation costs of the different algorithms (in seconds).

generated samples was below 1.2 as recommended in [48]. The average running times of each of the methods are presented in Tab. I. As we can see, the weighted ℓ_1 norm is three orders of magnitude slower than sLoreta (taking seconds instead of milliseconds) while the proposed method is three orders of magnitude slower than the weighted ℓ_1 norm. Note that the first two algorithms seem to have a running time that does not depend on the kind of simulations while the proposed method is faster for simpler cases and slower for more complex dipole distributions. Having a higher computational cost is a typical disadvantage of MCMC sampling techniques when compared to optimization approaches. However, it is important to note that our algorithm is able to estimate its hyperparameters in one run while the other two require to be run several times in order to set their regularization parameters by cross-validation for instance.

C. Real data

Two different sets of real data were considered. The first one consists of an auditory evoked response while the second one is the evoked response to facial stimulus. In addition to the weighted ℓ_1 norm regularization, we also compared our results with the MSP algorithm [29] using the default parameters in the SPM software. [3]

Auditory evoked responses

The used data set was extracted from the MNE software [49, 50]. It corresponds to an evoked response to left-ear auditory pure-tone stimulus using a realistic BEM head model sampled with 60 EEG electrodes. The data was acquired using 60 EEG electrodes sampled at 600 samples/s. The samples were low-pass filtered at 40Hz and downsampled to 150 samples/s. The noise covariance matrix was estimated from 200ms of data preceding each stimulus and was used to whiten the data. Fifty one epochs were averaged to generate the measurements $\boldsymbol{y}$.

The sources associated with this example are composed of 1844 dipoles located on the cortex with orientations that are normal to the brain surface. One channel that had technical artifacts was ignored resulting in an operator $\mathcal{H} \in \mathbb{R}^{59\times 1884}$. We processed the eight time samples corresponding to 80ms $\leq$ t $\leq$ 126ms, *i.e.*, associated with the highest activity period in the EEG measurements as displayed in Fig. 11.

The sum of the estimated brain activities over the 8 time samples obtained by the proposed method, the weighted ℓ_1 norm regularization and the MSP algorithm are presented in Fig. 12. The proposed method consistently detects most of the activity concentrated in both the ipsilateral and contralateral auditory cortices. The weighted ℓ_1 norm regularization detects the activity in the right cortex in a similar position, but moves the activity detected in the left cortex to a lower point of the brain that is further away from the auditory cortex. The MSP algorithm finds the activity correctly in the auditory cortices but spreads it around a region instead of focusing it on a small number of dipoles. This is due to the fact

[3]The SPM software is freely avaiable at http://www.fil.ion.ucl.ac.uk/spm.

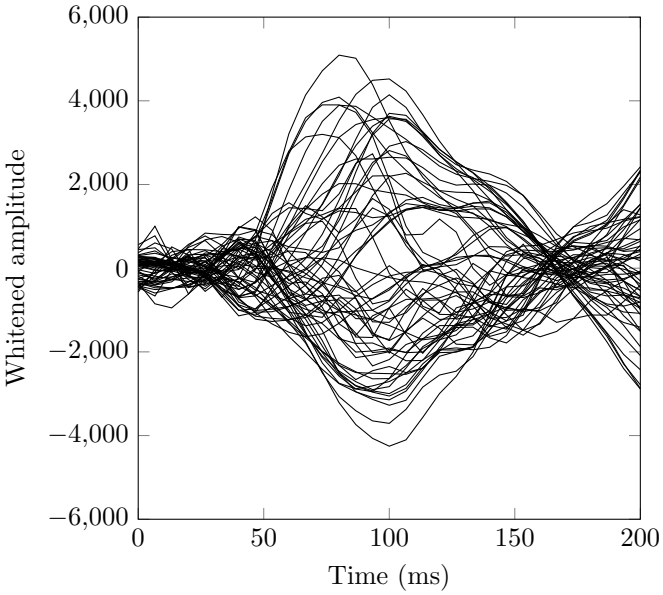
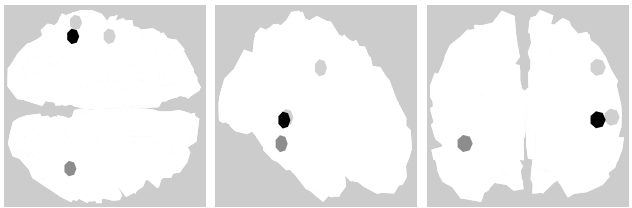
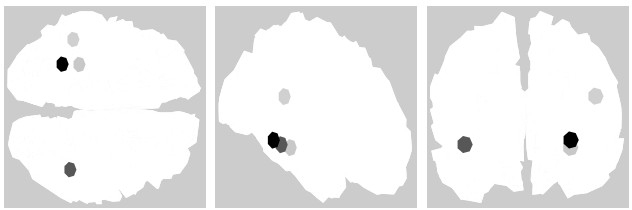


Fig. 11: EEG measurements for the real data application.



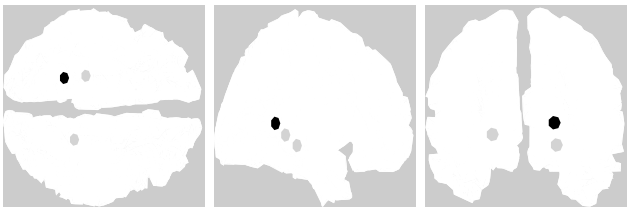
(a) Proposed method



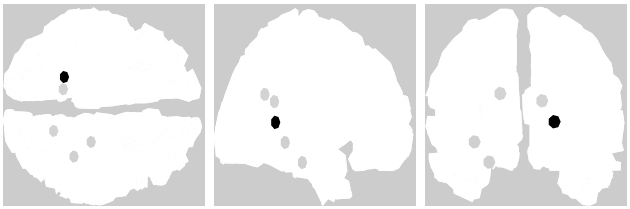
(b) Weighted ℓ_1 norm regularization

(c) MSP algorithm

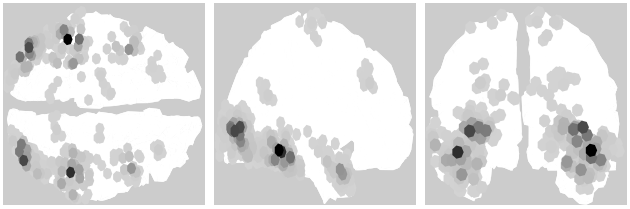
Fig. 12: Brain activity for the auditory evoked responses from 80ms to 126ms.



(a) Proposed method



(b) Weighted ℓ_1 norm regularization



(c) MSP algorithm

Fig. 13: Brain activity for the faced evoked responses for 160ms.

that both the proposed method and the weighted ℓ_1 norm promote sparsity over the dipoles while MSP promotes sparsity over pre-selected brain regions that depend on the parcellation scheme used.

Face-evoked responses

The data used in this section is one of the sample data sets available in the SPM software. It was acquired from a face perception study in which the subject had to judge the symmetry of a mixed set of faces and scrambled faces. Faces were presented during 600ms with an interval of 3600ms (for more details on the paradigm see [51]). The acquisition system was a 128-channel ActiveTwo system with a sampling frequency equal to 2048 Hz. The data was downsampled to 200 Hz and, after artifact rejection, the 299 epochs corresponding to non-scrambled faces were averaged and lowpass filtered at 40Hz. The head model is based on a T1 MRI scan of the patient downsampled to have 3004 dipoles, resulting in an operator $\mathcal{H} \in \mathbb{R}^{128\times 3004}$.

The brain activities detected by the three algorithms for $t = 160$ms (time sample with highest brain activity) are presented in Fig. 13. The MSP method locates the activity spread over several brain regions due to its pre-parcellation of the brain. In particular, it locates activity in the lateral and posterior regions. In comparison both the weighted ℓ_1 norm and the proposed method focus the activity closer to the fusiform regions of the temporal lobes, areas of the brain that are speculated to be specialized in facial recognition [52]. Note that the proposed method provides the most focal solution of the three. It is interesting to note that the MSP algorithm divides the brain in symmetric parcels in both hemispheres which very often causes the solution to have a high degree of symmetry while the other two methods only rely on the measurements to estimate the brain activity.

VI. Conclusion

This paper proposed a new Bayesian model for EEG source localization promoting sparsity for the dipole activities via a Bernoulli-Laplace prior. To compute the Bayesian estimators of this model, we introduced a Markov chain Monte Carlo method sampling the posterior distribution of interest and estimating the model parameters using the generated samples. The resulting EEG source localization strategy was compared to ℓ_2 norm (sLoreta) and weighted ℓ_1 norm regularizations for synthetic data and with the multiple sparse priors (MSP) algorithm for real data, showing promising results in both cases. More precisely, several experiments with synthetic data were constructed using single and multiple dipoles, both for close and distant locations. For the single dipole scenario, the proposed algorithm showed better performance than the more traditional ℓ_2 and ℓ_1 norm regularizations in terms of several evaluation criteria used in the literature. In multiple dipole scenarios, the estimated activities from different dipoles can overlap, making the classical evaluation criteria difficult to apply. In order to asses the performance in these scenarios, we proposed a new evaluation criterion denoted as transportation cost defined as the solution of an optimal mass transportation problem. This criterion showed that the proposed localization method performed better than the standard ℓ_2 norm and weighted ℓ_1 norm regularizations. We also considered two sets of real data consisting of the evoked responses to a left-ear auditory stimulation and to facial stimulus respectively. In both cases the algorithm showed better performance than the weighted ℓ_1 norm regularization and the MSP method to estimate brain activity generated by point-like sources.
Future work includes adapting the proposed model for multi-temporal EEG signals and generalizing the Bayesian model to cases where the head model is not known perfectly.

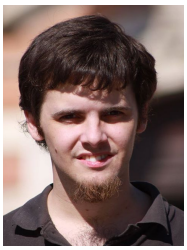
Acknowledgement

The authors would like to thank K. Friston and V. Litvak for their feedback about the multiple sparse priors algorithm.

References

[1] H. Buchner, G. Knoll, M. Fuchs, A. Rienäcker, R. Beckmann, M. Wagner, J. Silny, and J. Pesch, "Inverse localization of electric dipole current sources in finite element models of the human head," *Electroencephalogr. Clin. Neurophysiol.*, vol. 102, no. 4, pp. 267–278, 1997.
[2] B. N. Cuffin, "A method for localizing EEG sources in realistic head models," *IEEE Trans. Biomed. Eng.*, vol. 42, no. 1, pp. 68–71, 1995.
[3] 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.
[4] J. C. Mosher, P. S. Lewis, and R. M. Leahy, "Multiple dipole modeling and localization from spatio-temporal MEG data," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 6, pp. 541–557, 1992.
[5] J. C. Mosher and R. M. Leahy, "Recursive MUSIC: a framework for EEG and MEG source localization," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 11, pp. 1342–1354, 1998.

[6] J. Mosher and R. Leahy, "Source localization using recursively applied and projected (RAP) MUSIC," *IEEE Trans. Signal Process.*, vol. 47, no. 2, pp. 332–340, 1999.
[7] X.-L. Xu, B. Xu, and B. He, "An alternative subspace approach to EEG dipole source localization," *Phys. Med. Biol.*, vol. 49, no. 2, pp. 327–343, 2004.
[8] S. Sommariva and A. Sorrentino, "Sequential Monte Carlo samplers for semi-linear inverse problems and application to magnetoencephalography," *Inverse Probl.*, vol. 30, no. 11, pp. 114 020–114 043, 2014.
[9] F. L. da Silva and A. Van Rotterdam, "Biophysical aspects of EEG and magnetoencephalogram generation," in *Electroencephalography: Basic Principles, Clinical Applications and Related Fields*. Baltimore: Williams & Wilkins, 1998.
[10] H. Liu, P. H. Schimpf, G. Dong, X. Gao, F. Yang, and S. Gao, "Standardized shrinking LORETA-FOCUSS (SSLOFO): a new algorithm for spatio-temporal EEG source reconstruction," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 10, pp. 1681–1691, 2005.
[11] R. D. Pascual-Marqui, "Review of methods for solving the EEG inverse problem," *Int. J. Bioelectromagn.*, vol. 1, no. 1, pp. 75–86, 1999.
[12] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann, "Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain," *Int. J. Psychophysiol.*, vol. 18, no. 1, pp. 49–65, 1994.
[13] R. Pascual-Marqui *et al.*, "Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details," *Methods Findings Exp. Clin. Pharmacol.*, vol. 24D, pp. 5–12, 2002.
[14] L. Chaâri, J. Pesquet, J.-Y. Tourneret, P. Ciuciu, and A. Benazza-Benyahia, "A hierarchical Bayesian model for frame representation," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5560–5571, 2010.
[15] P. C. Hansen, "The L-Curve and Its Use in the Numerical Treatment of Inverse Problems," in *Computational Inverse Problems in Electrocardiology*. Southampton, UK: WIT Press, 2000.
[16] 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.
[17] K. Matsuura and Y. Okabe, "Selective minimum-norm solution of the biomagnetic inverse problem," *IEEE Trans. Biomed. Eng.*, vol. 42, no. 6, pp. 608–615, 1995.
[18] K. Uutela, M. Hämäläinen, and E. Somersalo, "Visualization of magnetoencephalographic data using minimum current estimates," *NeuroImage*, vol. 10, no. 2, pp. 173–180, 1999.
[19] A. Galka, O. Yamashita, T. Ozaki, R. Biscay, and P. Valdés-Sosa, "A solution to the dynamical inverse problem of EEG generation using spatiotemporal Kalman filtering," *NeuroImage*, vol. 23, no. 2, pp. 435–453, 2004.
[20] C. J. Long, P. L. Purdon, S. Temereanca, N. U. Desai, M. S. Hämäläinen, and E. N. Brown, "State-space solutions to the dynamic magnetoencephalography inverse problem using high performance computing," *The annals of applied statistics*, vol. 5, no. 2B, pp. 1207–1228, 2011.
[21] E. Somersalo, A. Voutilainen, and J. Kaipio, "Non-stationary magnetoencephalography by Bayesian filtering of dipole models," *Inverse Probl.*, vol. 19, no. 5, pp. 1047–1063, 2003.
[22] A. Sorrentino, A. M. Johansen, J. A. Aston, T. E. Nichols, W. S. Kendall *et al.*, "Dynamic filtering of static dipoles in magnetoencephalography," *The annals of applied statistics*, vol. 7, no. 2, pp. 955–988, 2013.
[23] X. Chen and S. Godsill, "Multiple dipolar sources localization for MEG using Bayesian particle filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Vancouver, Canada, May 2013.
[24] S. J. Kiebel, J. Daunizeau, C. Phillips, and K. J. Friston, "Variational Bayesian inversion of the equivalent current dipole model in EEG/MEG," *NeuroImage*, vol. 39, no. 2, pp. 728–741, 2008.
[25] T. Auranen, A. Nummenmaa, M. S. Hämäläinen, I. P. Jääskeläinen, J. Lampinen, A. Vehtari, and M. Sams, "Bayesian inverse analysis of neuromagnetic data using cortically constrained multiple dipoles," *Hum. Brain Mapp.*, vol. 28, no. 10, pp. 979–994, 2007.
[26] S. C. Jun, J. S. George, J. Paré-Blagoev, S. M. Plis, D. M. Ranken, D. M. Schmidt, and C. Wood, "Spatiotemporal Bayesian inference dipole analysis for MEG neuroimaging data," *NeuroImage*, vol. 28, no. 1, pp. 84–98, 2005.
[27] S. C. Jun, J. S. George, S. M. Plis, D. M. Ranken, D. M. Schmidt, and C. Wood, "Improving source detection and separation in a spatiotemporal Bayesian inference dipole analysis," *Phys. Med. Biol.*, vol. 51, no. 10, pp. 2395–2414, 2006.
[28] S. Baillet and L. Garnero, "A Bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 5, pp. 374–385, 1997.
[29] K. Friston, L. Harrison, J. Daunizeau, S. Kiebel, C. Phillips, N. Trujillo-Barreto, R. Henson, G. Flandin, and J. Mattout, "Multiple sparse priors for the M/EEG inverse problem," *NeuroImage*, vol. 39, no. 3, pp. 1104–1120, 2008.
[30] C. Stahlhut, H. T. Attias, K. Sekihara, D. Wipf, L. K. Hansen, and S. S. Nagarajan, "A hierarchical Bayesian M/EEG imaging method correcting for incomplete spatio-temporal priors," in *Proc. IEEE 10th Int. Symp. Biomed. Imagi. (ISBI)*, San Fransisco, USA, April 2013.
[31] H. Hallez, B. Vanrumste, R. Grech, J. Muscat, W. De Clercq, A. Vergult, Y. D'Asseler, K. P. Camilleri, S. G. Fabri, S. Van Huffel *et al.*, "Review on solving the forward problem in EEG source analysis," *J. Neuroeng. Rehabil.*, vol. 4, pp. 46–75, 2007.
[32] J. C. Mosher, R. M. Leahy, and P. S. Lewis, "EEG and MEG: forward solutions for inverse methods," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 3, pp. 245–259, 1999.
[33] L. Garnero, S. Baillet, J. Lachaux, and B. Renault, "Data operating in a PET/EEG/MRI experiment," *Hum. Brain Mapp.*, vol. 1, no. 8, pp. 1–11, 1995.
[34] 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, p. 1937, 2012.
[35] E. Maris, "A resampling method for estimating the signal subspace of spatio-temporal EEG/MEG data," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 8, pp. 935–949, 2003.
[36] G. Casella and C. P. Robert, *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 1999.
[37] L. Chaari, J.-Y. Tourneret, and H. Batatia, "Sparse Bayesian regularization using Bernoulli-Laplacian priors," in *Proc. European Signal Processing Conference (EUSIPCO)*, Marrakech, Morocco, Sept. 2013.
[38] N. Dobigeon, A. O. Hero, and J.-Y. Tourneret, "Hierarchical Bayesian sparse image reconstruction with application to MRFM," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 2059–2070, 2009.
[39] N. Dobigeon, J.-Y. Tourneret, and C.-I. Chang, "Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2684–2695, Jul. 2008.
[40] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of hyperspectral and multispectral images," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Florence, Italy, May 2014.
[41] F. Tadel, S. Baillet, J. C. Mosher, D. Pantazis, and R. M. Leahy, "Brainstorm: a user-friendly application for MEG/EEG analysis," *Comput. Intell. Neurosci.*, vol. 2011, no. 8, pp. 1–13, 2011.
[42] A. Gramfort, T. Papadopoulo, E. Olivi, M. Clerc *et al.*, "OpenMEEG: opensource software for quasistatic bioelectromagnetics," *Biomed. Eng. Online*, vol. 9, no. 45, pp. 1–20, 2010.
[43] C. Silva, J. Maltez, E. Trindade, A. Arriaga, and E. Ducla-Soares, "Evaluation of L1 and L2 minimum norm performances on EEG localizations," *Clin. Neurophys.*, vol. 115, no. 7, pp. 1657–1668, 2004.
[44] S. T. Rachev, "The Monge-Kantorovich Mass Transference Problem and Its Stochastic Applications," *Theory of Probability and Its Applications*, vol. 29, no. 4, pp. 647–676, 1984.
[45] P. Xu, Y. Tian, H. Chen, and D. Yao, "Lp norm iterative sparse solution for EEG source localization," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 3, pp. 400–409, 2007.
[46] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learning*, vol. 3, no. 1, pp. 1–122, 2011.
[47] V. A. Morozov, "On the solution of functional equations by the method of regularization," *Soviet Math. Dokl.*, vol. 7, pp. 414–417, 1966.
[48] 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.
[49] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, and M. S. Hämäläinen, "MNE software for processing MEG and EEG data," *NeuroImage*, vol. 86, pp. 446–460, 2014.
[50] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen *et al.*, "MEG and EEG data analysis with MNE-Python," *Front. Neurosci.*, vol. 7, no. 267, pp. 1–13, 2013.
[51] R. Henson, J. Mattout, K. Singh, G. Barnes, A. Hillebrand, and K. Friston, "Population-level inferences for distributed meg source localization under multiple constraints: application to face-evoked fields," *NeuroImage*, vol. 38, no. 3, pp. 422–438, 2007.
[52] N. Kanwisher, J. McDermott, and M. M. Chun, "The fusiform face area: a module in human extrastriate cortex specialized for face perception," *J. Neurosci.*, vol. 17, no. 11, pp. 4302–4311, 1997.



Facundo Costa was born in Buenos Aires, Argentina in 1988. He received the 5 years Eng. degree in Electronics Engineering as the top of his class from the Buenos Aires Institute of Technology (ITBA) in 2013. He worked as a development engineer specialized in radiofrequency for the Field Technology group of DirecTV Latin America until 2014. Currently he is a Ph.D. candidate in Signal Processing with the National Polytechnic Institute of Toulouse (University of Toulouse, INP-ENSEEIHT). He is also part of the Traitement et Compréhension d'Image (TCI) Group of the Institut de Recherche en Informatique de Toulouse (IRIT). His main research interest includes image processing, bayesian estimation and statistics.

Hadj Batatia is currently associate professor at the university of Toulouse (FR). He received his MSc and PhD in 1987 and 1992 from the university of Toulouse. He worked as lecturer at university of Malaysia Sarawak (MY) and then senior lecturer at university of Teesside (UK) till 1999. He then joined the university of Toulouse in 1999 as assistant professor. His research focus on medical image processing. His work include proposing statistical models to explain variability in various medical imaging modalities. Based on this models, he develops variational and Bayesian methods to solve reconstruction, segmentation and motion analysis problems. His interests include also efficient algorithms to sample complex distributions. In addition, he teaches complex systems engineering and heads the CNAM school of industrial sciences and information technology in Toulouse.

Lotfi Chaari was born in Sfax in Tunisia. He received the ingineering and master degrees from the High School of Telecommunication in Tunis (SUP'COM) in 2007. In 2010 he received the PhD degree from the university of Paris-Est, where he worked on medical imaging with the LIGM lab. He then moved to Grenoble as a postdoctoral fellow at Inria in the Mistis team where he worked on functional MRI data analysis and activation detection-estimation. Since september 2012, he is assistant professor at INP-Toulouse. He is doing his research with the TCI team of the IRIT lab, still on medical image processing.

Jean-Yves Tourneret received the ingénieur degree in electrical engineering from the Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique, d'Hydraulique et des Télécommunications (ENSEEIHT) de Toulouse in 1989 and the Ph.D. degree from the National Polytechnic Institute from Toulouse in 1992. He is currently a professor in the university of Toulouse (ENSEEIHT) and a member of the IRIT laboratory (UMR 5505 of the CNRS). His research activities are centered around statistical signal and image processing with a particular interest to Bayesian and Markov chain Monte Carlo (MCMC) methods. He has been involved in the organization of several conferences including the European conference on signal processing EUSIPCO'02 (program chair), the international conference ICASSP'06 (plenaries), the statistical signal processing workshop SSP12 (international liaisons), the International Workshop on Computational Advances in Multi-Sensor Adaptive Processing CAMSAP 2013 (local arrangements), the statistical signal processing workshop SSP'2014 (special sessions), the workshop on machine learning for signal processing MLSP'2014 (special sessions). He has been the general chair of the CIMI workshop on optimization and statistics in image processing hold in Toulouse in 2013 (with F. Malgouyres and D. Kouam) and of the International Workshop on Computational Advances in Multi-Sensor Adaptive Processing CAMSAP 2015 (with P. Djuric). He has been a member of different technical committees including the Signal Processing Theory and Methods (SPTM) committee of the IEEE Signal Processing Society (2001-2007, 2010-present). He has been serving as an associate editor for the IEEE Transactions on Signal Processing (2008-2011,2015-present) and for the EURASIP journal on Signal Processing (since july 2013).