Bayesian M/EEG source localization with possible joint skull conductivity estimation

Facundo Hernan Costa

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National Polytechnique de Toulouse (INP Toulouse)

Discipline ou spécialité :

Image, Information et Hypermédia

Présentée et soutenue par :

M. FACUNDO HERNAN COSTA

le jeudi 2 mars 2017

Titre :

Bayesian M/EEG Source Localization with possible Joint Skull Conductivity Estimation

---

Ecole doctorale :

Mathématiques, Informatique, Télécommunications de Toulouse (MITT)

Unité de recherche :

Institut de Recherche en Informatique de Toulouse (I.R.I.T.)

Directeurs de Thèse :

M. JEAN YVES TOURNERET

M. HADJ BATATIA

Rapporteurs :

M. DAVID BRIE, UNIVERSITÉ LORRAINE

M. LAURENT ALBERA, UNIVERSITE RENNES 1

Membres du jury :

M. ANDRE FERRARI, UNIVERSITE DE NICE SOPHIA ANTIPOLIS, Président

M. HADJ BATATIA, INP TOULOUSE, Membre

M. JEAN YVES TOURNERET, INP TOULOUSE, Membre

M. MATTHIEU KOWALSKI, UNIVERSITE PARIS 11, Membre

ABSTRACT

M/EEG mechanisms allow determining changes in the brain activity, which is useful in diagnosing brain disorders such as epilepsy. They consist of measuring the electric potential at the scalp and the magnetic field around the head. The measurements are related to the underlying brain activity by a linear model that depends on the lead-field matrix. Localizing the sources, or dipoles, of M/EEG measurements consists of inverting this linear model. However, the non-uniqueness of the solution (due to the fundamental law of physics) and the low number of dipoles make the inverse problem ill-posed. Solving such problem requires some sort of regularization to reduce the search space. The literature abounds of methods and techniques to solve this problem, especially with variational approaches.

This thesis develops Bayesian methods to solve ill-posed inverse problems, with application to M/EEG. The main idea underlying this work is to constrain sources to be sparse. This hypothesis is valid in many applications such as certain types of epilepsy. We develop different hierarchical models to account for the sparsity of the sources.

Theoretically, enforcing sparsity is equivalent to minimizing a cost function penalized by an ℓ_0 pseudo norm of the solution. However, since the ℓ_0 regularization leads to NP-hard problems, the ℓ_1 approximation is usually preferred. Our first contribution consists of combining the two norms in a Bayesian framework, using a Bernoulli-Laplace prior. A Markov chain Monte Carlo (MCMC) algorithm is used to estimate the parameters of the model jointly with the source location and intensity. Comparing the results, in several scenarios, with those obtained with sLoreta and the weighted ℓ_1 norm regularization shows interesting performance, at the price of a higher computational complexity.

Our Bernoulli-Laplace model solves the source localization problem at one instant of time. However, it is biophysically well-known that the brain activity follows spatiotemporal patterns. Exploiting the temporal dimension is there-

fore interesting to further constrain the problem. Our second contribution consists of formulating a structured sparsity model to exploit this biophysical phenomenon. Precisely, a multivariate Bernoulli-Laplacian distribution is proposed as an a priori distribution for the dipole locations. A latent variable is introduced to handle the resulting complex posterior and an original Metropolis-Hastings sampling algorithm is developed. The results show that the proposed sampling technique improves significantly the convergence. A comparative analysis of the results is performed between the proposed model, an ℓ_{21} mixed norm regularization and the Multiple Sparse Priors (MSP) algorithm. Various experiments are conducted with synthetic and real data. Results show that our model has several advantages including a better recovery of the dipole locations.

The previous two algorithms consider a fully known lead-field matrix. However, this is seldom the case in practical applications. Instead, this matrix is the result of approximation methods that lead to significant uncertainties. Our third contribution consists of handling the uncertainty of the lead-field matrix. The proposed method consists in expressing this matrix as a function of the skull conductivity using a polynomial matrix interpolation technique. The conductivity is considered as the main source of uncertainty of the lead-field matrix. Our multivariate Bernoulli-Laplacian model is then extended to estimate the skull conductivity jointly with the brain activity. The resulting model is compared to other methods including the techniques of Vallaghé et al and Guttierez et al. Our method provides results of better quality without requiring knowledge of the active dipole positions and is not limited to a single dipole activation.

RÉSUMÉ

Les techniques M/EEG permettent de déterminer les changements de l'activité du cerveau, utiles au diagnostic de pathologies cérébrales, telle que l'épilepsie. Ces techniques consistent à mesurer les potentiels électriques sur le scalp et le champ magnétique autour de la tête. Ces mesures sont reliées à l'activité électrique du cerveau par un modèle linéaire dépendant d'une matrice de mélange liée à un modèle physique.

La localisation des sources, ou dipôles, des mesures M/EEG consiste à inverser le modèle physique. Cependant, la non-unicité de la solution (due à la loi fondamentale de physique) et le faible nombre de dipôles rendent le problème inverse mal-posé. Sa résolution requiert une forme de régularisation pour restreindre l'espace de recherche. La littérature compte un nombre important de travaux traitant de ce problème, notamment avec des approches variationnelles.

Cette thèse développe des méthodes Bayésiennes pour résoudre des problèmes inverses, avec application au traitement des signaux M/EEG. L'idée principale sous-jacente à ce travail est de contraindre les sources à être parcimonieuses. Cette hypothèse est valide dans plusieurs applications, en particulier pour certaines formes d'épilepsie. Nous développons différents modèles Bayésiens hiérarchiques pour considérer la parcimonie des sources.

En théorie, contraindre la parcimonie des sources équivaut à minimiser une fonction de coût pénalisée par la norme ℓ_0 de leurs positions. Cependant, la régularisation ℓ_0 générant des problèmes NP-complets, l'approximation de cette pseudo-norme par la norme ℓ_1 est souvent adoptée. Notre première contribution consiste à combiner les deux normes dans un cadre Bayésien, à l'aide d'une loi a priori Bernoulli-Laplace. Un algorithme Monte Carlo par chaîne de Markov est utilisé pour estimer conjointement les paramètres du modèle et les positions et intensités des sources. La comparaison des résultats, selon plusieurs scénarii, avec ceux obtenus par sLoreta et la régularisation par la

norme ℓ_1 montre des performances intéressantes, mais au détriment d'un coût de calcul relativement élevé.

Notre modèle Bernoulli-Laplace résout le problème de localisation des sources pour un instant donné. Cependant, il est admis que l'activité cérébrale a une certaine structure spatio-temporelle. L'exploitation de la dimension temporelle est par conséquent intéressante pour contraindre d'avantage le problème. Notre seconde contribution consiste à formuler un modèle de parcimonie structurée pour exploiter ce phénomène biophysique. Précisément, une distribution Bernoulli-Laplacienne multivariée est proposée comme loi a priori pour les dipôles. Une variable latente est introduite pour traiter la loi a posteriori complexe résultante et un algorithme d'échantillonnage original de type Metropolis-Hastings est développé. Les résultats montrent que la technique d'échantillonnage proposée améliore significativement la convergence de la méthode MCMC. Une analyse comparative des résultats a été réalisée entre la méthode proposée, une régularisation par la norme mixte ℓ_{21}, et l'algorithme MSP (Multiple Sparse Priors). De nombreuses expérimentations ont été faites avec des données synthétiques et des données réelles. Les résultats montrent que notre méthode a plusieurs avantages, notamment une meilleure localisation des dipôles.

Nos deux précédents algorithmes considèrent que le modèle physique est entièrement connu. Cependant, cela est rarement le cas dans les applications pratiques. Au contraire, la matrice du modèle physique est le résultat de méthodes d'approximation qui conduisent à des incertitudes significatives. Notre troisième contribution consiste à considérer l'incertitude du modèle physique dans le problème de localisation de sources. La méthode proposée consiste à exprimer la matrice de mélange du modèle comme une fonction de la conductivité du crâne, en utilisant une technique d'interpolation polynomiale. La conductivité est considérée comme la source principale de l'incertitude du modèle physique et elle est estimée à partir des données. La distribution Bernoulli-Laplacienne multivariée est étendue pour estimer la conductivité conjointement avec l'activité cérébrale. Le modèle résultant est comparé à d'autres méthodes en particulier les techniques de Vallaghé et al. et Guttierez et al. Notre méthode fournit des résultats de meilleure qualité sans connaissance préalable de la position des dipôles actifs, et n'est pas limitée à un dipôle unique.

CONTENTS

ACRONYMS

BEM	Boundary Element Method
ECG	Electrocardiography
EEG	Electroencephalography
EIT	Electrical Impedance Tomography
FDM	Finite Difference Method
FEM	Finite Element Method
FINES	First priNcipal vEctorS
FMRI	Functional MRI
LORETA	LOw-Resolution Electromagnetic Tomography Algorithm
MCMC	Markov chain Monte Carlo
MEG	Magnetoencephalography
M/EEG	Magnetoelectroencephalography
MH	Metropolis-Hastings
MMSE	Minimum MSE
MNE	Minimum Norm Estimates
MRI	Magnetic Resonance Imaging
MSE	Mean Square Error
MSP	Multiple Sparse Priors
MUSIC	MUltiple SIgnal Classification
PDF	Probability Density Function
PSRF	Potential Scale Reduction Factor
R-MUSIC	Recursive MUSIC
RAP-MUSIC	Recursively APplied MUSIC
SLORETA	Standardized LORETA
SNR	Signal to Noise Ratio
SPM	Statistical Parametric Mapping

PROBABILITY DISTRIBUTIONS

$\mathscr{B}(p)$	Bernoulli distribution with parameter p
$\mathscr{B}e(\alpha,\beta)$	Beta distribution with shape parameter α and rate parameter β
$\mathscr{G}(\alpha,\beta)$	Gamma distribution with shape parameter α and rate parameter β
$\mathscr{GIG}(a,b,p)$	Generalized Inverse Gaussian distribution with parameters a, b and p
$\mathscr{IG}(\alpha,\beta)$	Inverse gamma distribution with shape parameter α and rate parameter β
$\mathscr{N}(\boldsymbol{\mu},\Sigma)$	Normal distribution with mean $\boldsymbol{\mu}$ and covariance Σ
$\mathscr{N}_+(\boldsymbol{\mu},\Sigma)$	Normal distribution with mean $\boldsymbol{\mu}$ and covariance Σ truncated on $\mathbb{R}^+$
$\mathscr{N}_-(\boldsymbol{\mu},\Sigma)$	Normal distribution with mean $\boldsymbol{\mu}$ and covariance Σ truncated on $\mathbb{R}^-$
$\mathscr{U}(a,b)$	Uniform distribution with lower limit a and upper limit b

List of Figures

List of Tables

List of Algorithms

CHAPTER 1

INTRODUCTION

Contents

1.1 MOTIVATION

Recent World Health Organization data suggest that neurological disorders, including epilepsy, are one of the most important contributors to the global burden of human suffering [1]. M/EEG measurement is one of the main tools used by specialists to examine epilepsy patients. Other uses of M/EEG include evaluation of encephalopaties and focal brain lesions [1].

M/EEG is a powerful non-invasive technique that measures the electric potentials at the scalp and the magnetic fields around the head. These measurements depend upon 1) the underlying brain activity and 2) the geometric composition of the head. The recovery of the brain activity from the measurements is an ill-posed inverse problem. Thus, a regularization is needed to have a narrow the search space. This regularization should typically be chosen to constrain the solution to have some realistic properties.

In this thesis we propose new Bayesian approaches to solve the M/EEG source localization problem. Our work is specifically focused on cases where the brain activity is spatially concentrated, such as in certain forms of epilepsy.

The main idea underlying the thesis is to develop Baysian sparse models for the sources. More precisely, three contributions are presented:

(i) Developing a hierarchical Bayesian model that solves the source localization problem for one instant of time by promoting sparsity using Bernoulli-Laplacian priors [2].

(ii) Investigating a Bayesian structured sparsity model to exploit the temporal dimension of the M/EEG measurements [3, 4].

(iii) Developing Metropolis-Hasting sampling scheme that improves significantly the speed of the convergence of our MCMC algorithm [5].

(iv) Expanding the model to estimate the skull conductivity jointly with the brain activity [6].

1.2 Organization of the manuscript

Chapter 2: Medical Context
This chapter provides an introduction to M/EEG, the link between electric brain activity and the M/EEG measurements and the preprocessing techniques that are typically used to clean the signal. We then focus on how these measurements are used in solving the source localization problem and describe state-of-the-art algorithms that have been developed in the literature with their advantages and disadvantages.

Chapter 3: Bayesian Sparse M/EEG Source Localization
In this chapter, we present a hierarchical Bayesian model aimed to solving the source localization problem by promoting sparsity. Ideally the ℓ_0 pseudo norm regularization should be used to regularize this problem. However, due to its intractability, the ℓ_0 pseudo norm is usually replaced by the ℓ_1 norm. Our model proposes to combine both of them, resulting in ℓ_0 and ℓ_1 regularizations in a Bayesian framework to pursue sparse solutions.

Chapter 4: Structured Sparsity Bayesian M/EEG Source Localization
In this chapter, we improve the model of Chapter 3 to take advantage of the temporal structure of the M/EEG measurements. This is done by changing the prior associated with the brain activity and computing a

Bayesian approximation of the M/EEG source localization problem using an ℓ_{20} pseudo mixed norm regularization. This type of norm promotes sparsity among different dipoles (via the ℓ_0 portion of the norm) and groups all the time samples of the same dipole together, forcing them all to be either jointly active or inactive (with the ℓ_2 norm portion).

Chapter 5: Myopic Skull Conductivity Estimation

When working with real data, certain parameters needed to calculate the leadfield matrix are not always known. Out of these, the skull conductivity stands out since it varies significantly among subjects and can affect the results of the brain activity reconstruction significantly. In this chapter we generalize the model to estimate the skull conductivity jointly with the brain activity. This is done by approximating the dependency of the operator with respect to the skull conductivity with a polynomial matrix.

Chapter 6: Conclusion and Future Work

This chapter presents some conclusions based on our work and some ideas that should be pursued in the near future related to it.

Appendix A: Conditional probability distributions derivations

This appendix shows the algebraic derivations of the conditional probability distributions associated with the Bayesian models introduced in Chapters 4 and 5.

1.3 PUBLICATIONS

Journal papers

(i) F. Costa, H. Batatia, L. Chaari, and J.-Y. Tourneret, "Sparse EEG Source Localization using Bernoulli Laplacian Priors," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 12, pp. 2888–2898, 2015.

(ii) F. Costa, H. Batatia, T. Oberlin, C. D'Giano, and J.-Y. Tourneret, "Bayesian EEG source localization using a structured sparsity prior," *NeuroImage*, vol. 144, pp. 142–152, jan. 2017.

(iii) F. Costa, H. Batatia, T. Oberlin, and J.-Y. Tourneret, "Skull Conductivity Estimation for EEG Reconstruction," *IEEE Signal Process. Lett.*, to be published.

Conference papers

(i) 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 (CAMSAP'15)*, Cancun, Mexico, 2015.

(ii) F. Costa, H. Batatia, T. Oberlin, and J.-Y. Tourneret, "A partially collapsed gibbs sampler with accelerated convergence for EEG source localization," in *Proc. of IEEE Workshop on Stat. Sig. Proc. (SSP'16)*, Palma de Mallorca, Spain, 2016.

(iii) Y. Mejía, F. Costa, H. Argüello, J.-Y. Tourneret and H. Batatia "Bayesian reconstruction of hyperspectral images by using compressed sensing measurements and a local structured prior," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, New Orleans, USA, 2017.

CHAPTER 2

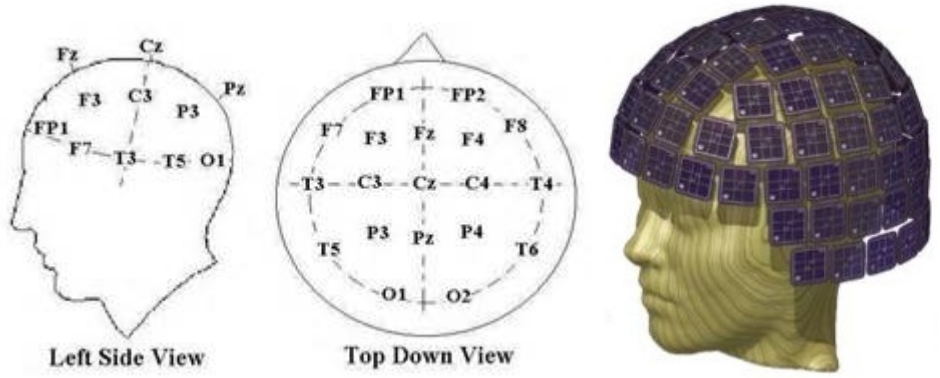
MEDICAL CONTEXT

Contents

2.1 M/EEG MEASUREMENTS

In the past decades, several non-invasive techniques for measuring brain activity have been developed. Among them, M/EEG has the noticeable advantage of being able to track brain activity with a temporal resolution in the order of the milliseconds. This makes it an invaluable tool in a variety of medical applications including the diagnosis of epilepsy, sleeping disorders, coma, encephalopathies, and brain death [7, 8].

M/EEG is a combination between EEG and MEG. EEG consists in recording the electrical voltage, typically in the order of μV, measured by electrodes placed over the scalp (shown in Fig. 2.1a). On the other side, MEG consists



(a) EEG electrodes. (b) MEG magnetometers.

Figure 2.1: Typical sensors setups for EEG and MEG measurements.

in recording the magnetic fields using very sensitive magnetometers placed around the head (shown in Fig. 2.1b). Both the electrical voltage measured by EEG and the magnetic fields measured by MEG are generated by the brain activity of the subject that can be modeled by currents [9].

2.2 PREPROCESSING

The measured M/EEG signal is typically highpass filtered at approximately 0.5 Hz to remove very low frequency interference (such as breathing). A lowpass filter is also applied to eliminate measurement noise higher than 50 – 70 Hz approximately. Unfortunately, the measured M/EEG signal can also be contaminated by artifacts that are not related with the brain activity of interest and cannot be filtered out as easily [10]. The most important artifact source is the power-line interference at 50 Hz (or 60 Hz). The easiest way to deal with this interference is to just discard measurements in which the artifact noise is higher than a certain threshold. There are also other techniques to deal with power-line noise such as physical solutions [10, 11] and adaptive noise cancelling techniques [12, 13].

In addition to power-line noise, M/EEG measurements can also be contaminated by other signals that originate from the patient but are not related to brain activity, such as eye movement, muscle noise and heart signal. To compensate for eye movement artifacts an additional EEG electrode is typi-

cally placed in the nose in order to obtain measurements that are used to identify eye movement artifacts [14]. ECG artifacts are produced by electrical heart activity. Due to its high amplitude, this activity can be measured by placing electrodes in any non-cephalic part of the body. Cardiac artifacts can be suppressed by recording an ECG of the heart activity. In addition to algorithms aimed to reduce measurement artifacts, other techniques can also be applied in the pre-processing stage to improve the source localization such as tensor-based preprocessing [15–17].

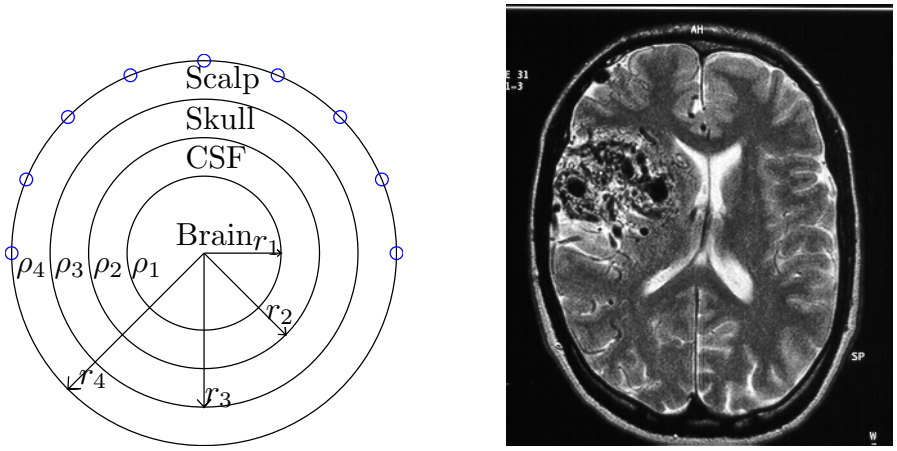
Once the M/EEG measurements have been preprocessed, they can be used to solve the source localization problem.

2.3 Source Localization

The source localization problem consists in using the cleaned M/EEG measurements to determine the electrical activity of the subject's brain. The relationship between electric brain activity and M/EEG measurements is represented by the leadfield matrix (also called head operator). The leadfield matrix depends on the shape and composition of the subject's head and will be explained in what follows.

2.3.1 Leadfield matrix

Several head models with varying precision and complexity have been used throughout the years, being mainly divided in two categories (1) the shell head models and (2) the realistic head models [18] that are displayed in Fig. 2.2. The former model represents the human head using a fixed number of concentric spheres (typically 3 or 4). Each sphere corresponds to the interface between two different tissues of the human head considered to be uniform with constant conductivities [19]. In the three-shell head model, the skull, cerebrospinal fluid and brain tissues are considered whereas the four-shell model adds an additional outer-most sphere to model the head tissue. In order to calculate the head operator with the shell models it is necessary to set different parameters: (1) the radius of the spheres, (2) the conductivity of each tissue, (3) the amount and locations of the dipoles inside the brain and (4) the amount and locations of the electrodes in the scalp. On the other hand, the realistic head models are typically computed from the MRI of the patient in order to better represent the distribution of the different tissues inside the human head. To perform this



(a) Shell model

(b) MRI Scan

Figure 2.2: Shell and realistic brain model representations.

calculation it is possible to use several methods [18] including the boundary element method (BEM) [20], the finite element method (FEM) [21] and the finite difference method (FDM) [22]. However, in order to be able to compute a head model operator from the MRI it is still necessary to set several parameters including (1) the conductivity of each tissue, (2) the amount and locations of the dipoles inside the brain and (3) the amount and locations of the electrodes in the scalp.

To ensure the quality of the activity estimation, the electrode positions and tissue conductivities should be set as close as possible to their real values. Most of the techniques developed for M/EEG source localization assume that these parameters are known in advance whereas a few consider that there can also be uncertainty in some of them. In the following we will first consider the case that assumes that the operator parameters are perfectly known in advance.

2.3.2 Methods with known leadfield matrix

The methods that have been developed to solve the M/EEG source localization problem can be classified in two groups: (i) the dipole-fitting models that represent the brain activity as a small number of dipoles with unknown positions; and (ii) the distributed-source models that represent it as a large number of dipoles at fixed positions.

The dipole-fitting models [23, 24] assume that the brain activity is concentrated in a small-number of point-like sources and estimate the location, amplitude and orientation of a few dipoles 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 [25]. To solve this problem, the MUSIC algorithm [26] and its variants R-MUSIC [27], RAP-MUSIC [28] and FINES [29] were developed. Another recent dipole-fitting model whose parameters are estimated using sequential Monte Carlo [30] formulates the M/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 [31, 32]. However, the performance of these algorithms can be altered in the case of multiple spatially extended sources [25]. An alternative use of dipole-fitting models is as a way to find an initial iteration point for distributed-source methods [33].

On the other hand, the distributed-source methods model the brain activity as the result of a large number of discrete dipoles with fixed positions and try to estimate their amplitudes and orientations [25]. Since the amount of dipoles used in the brain model is typically much larger than the amount of measurements, the inverse problem is ill-posed in the sense that there is an infinite amount of brain activities that can justify the measurements [25]. A regularization is thus needed in order to incorporate additional information to solve this inverse problem. The kind of regularization to use should be chosen to promote realistic properties of the solution. For instance, one of the most simple regularizations consists of penalizing the ℓ_2 norm of the solution using the minimum norm estimation algorithm [34] or its variants based on the weighted minimum norm: Loreta [35] and sLoreta [36]. However, these methods have been shown to overestimate the size of the active area if the brain activity is focused [25], which is believed to be the case in a number of medical applications. A better way to estimate focal brain activity is to promote sparsity, by applying an ℓ_0 pseudo norm regularization [37]. Unfortunately, this procedure is known to be intractable in an optimization framework. As a consequence, the ℓ_0 pseudo norm is usually approximated by the ℓ_1 norm via convex relaxation [38], in spite of the fact that these two approaches do not always provide the same solution [37].

All the distributed-source methods presented so far consider each time in-

dependently. However, to improve source localization, it is possible to make use of the temporal structure of the data by promoting structured sparsity, which is known to yield better results than standard sparsity when applied to strongly group sparse signals [39]. Structured sparsity has been shown to improve results in several applications including audio restoration [40], image analysis [41] and machine learning [42]. One way of applying structured sparsity in M/EEG source localization is to use mixed-norms regularization such as the ℓ_{21} mixed norm [43] (also referred to as group-lasso).

In addition to optimization techniques, several approaches have tried to model the time evolution of the dipole activity and estimate it using either Kalman filtering [44, 45] or particle filters [46–48]. Several Bayesian methods have been used as well, both for dipole-fitting models [49, 50] and distributed source models [51, 52]. In [51], the multiple sparse priors (MSP) approach was developed, which parcellates the brain in different pre-defined regions and promotes all the dipoles in each region to be active or inactive jointly. Doing this the brain activity is encouraged to extend over an area instead of being focused in point-like sources. Conversely, we are mainly interested in estimating point-like focal source activity which has been proved to be relevant in clinical applications [53]. In order to do this, we will consider each dipole separately instead of grouping them together. Note that this approach avoids the need of choosing a criterion for brain parcellization as required in the MSP method.

2.3.3 Methods with unknown leadfield matrix

In a more general case, we can consider that some of the parameters needed to calculate the operator are not perfectly known. Several authors have analyzed the influence of having errors in these parameters in the estimation of brain activity. Minor errors in the electrode positions have been shown not to affect significantly the results [54] whereas there is a much higher sensibility to variations in the tissue conductivities, making their values critical [55–58]. The conductivities of the human head tissue, cerebrospinal fluid and brain have well known values that are accepted in the literature [55]. However, there has been some controversy regarding the conductivity of the human skull [18]. The ratio between scalp and skull conductivities was initially reported to be 80 [59] but since then other authors have published values as low as 15 [60]. The value of the human skull conductivity is also known to vary significantly across different subjects [18, 61]. Because of this, it remains of interest to develop methods that estimate the skull conductivity to improve the quality of the brain activity re-

construction. This can be done using techniques such as electrical impedance tomography (EIT) [61], using intracranial electrodes [62] or measuring it directly during surgery [60]. However it is also possible to estimate it directly from the M/EEG measurements.

Several methods have been proposed to estimate the conductivity of the skull and the brain activity jointly, albeit requiring very restrictive conditions to yield good results. For instance, some methods require having a very good a-priori knowledge about the active dipole positions [63, 64], others assume there is only one dipole active [65] and others limit the estimation of the skull conductivity to a few discrete values [66, 67].

2.3.4 M/EEG Source Localization

In the most general case, M/EEG source localization can be formulated as an inverse problem that consists in estimating the brain activity of a patient from M/EEG measurements taken from M sensors during T time samples. In a distributed source model, the brain activity is represented by a finite number of dipoles located at fixed positions in the brain cortex. The M/EEG measurement matrix $\boldsymbol{Y} \in \mathbb{R}^{M \times T}$ can be written

$$\boldsymbol{Y} = \boldsymbol{H}(\rho)\boldsymbol{X} + \boldsymbol{E} \qquad [2.1]$$

where $\boldsymbol{X} \in \mathbb{R}^{N \times T}$ contains the dipole amplitudes, $\boldsymbol{H}(\rho) \in \mathbb{R}^{M \times N}$ represents the head operator (usually called "leadfield matrix") that depends on the skull conductivity ρ and $\boldsymbol{E}$ is the measurement noise.

If the value of ρ is known in advance then we have the common M/EEG source localization problem, which consists in estimating the matrix $\boldsymbol{X}$ from the known operator $\boldsymbol{H}(\rho)$ and the measurements $\boldsymbol{Y}$. This problem will be considered in Chapters 3 and 4. If the skull conductivity ρ is not known in advance then it can be estimated jointly with $\boldsymbol{X}$, resulting in a myope inverse problem considered in Chapter 5.

2.4 Conclusion

This chapter presented a brief illustration of the M/EEG principles and the different information that can be extracted from M/EEG measurements using signal processing techniques. Among them, the source localization problem was analyzed in more detail, summarizing some state-of-the-art algorithms that are

currently being used to tackle the problem. In Chapter 3 we will introduce a distributed-source method that combines ℓ_0 and ℓ_1 regularizations to promote sparsity. This results in a model that solves the inverse problem considering each time sample independently. The initial model is later improved in Chapter 4 to promote structured sparsity using an ℓ_{20} mixed norm regularization in Chapter 4. This model shows improved results since it takes advantage of the temporal structure of the M/EEG measurements. Finally, uncertainty in the leadfield matrix is considered. Precisely, this matrix is expressed as a function of the skull conductivity ρ, which is the most important source of uncertainty. The structured sparsity model is extended in Chapter 5 to estimate jointly the brain activity and the skull conductivity.

CHAPTER 3

BAYESIAN SPARSE M/EEG SOURCE LOCALIZATION

Contents

3.1 INTRODUCTION

When using a distributed-source model to solve the M/EEG source localization problem, it is necessary to apply a regularization in order to have a unique solution. In certain medical conditions, such as certain types of epilepsy or focal lesions, brain activity is believed to be focalized in a small brain region [8], so it makes sense to choose a regularization that promotes sparsity. This should be done by applying an ℓ_0 pseudo norm regularization [37]. However, due to its intractability, it is typically approximated to an ℓ_1 norm despite the fact that the results of both regularizations is not always the same [37]. We propose to combine both ℓ_0 and ℓ_1 regularizations in a Bayesian framework to pursue sparse solutions.

3.2 PROPOSED BAYESIAN MODEL

This chapter is devoted to the presentation of a new Bayesian model for M/EEG source localization. The model is described in Section 3.2. Section 3.3 studies a hybrid Gibbs sampler that will be used to generate samples asymptotically distributed according to the posterior of the model introduced in Section 3.2. Simulation results obtained with synthetic and real M/EEG data are presented in Section 3.4.

3.2.1 LIKELIHOOD

It is very classical in M/EEG analysis to consider an additive white Gaussian noise whose variance will be denoted as σ_n^2 [25]. When this assumption does not hold, it is always possible to estimate the noise covariance matrix from the data and to whiten the data before applying the algorithm [68]. The Gaussian noise assumption for the noise samples leads to the following probability density function (pdf)

$$f(\boldsymbol{Y}|\boldsymbol{X},\sigma_n^2) = \prod_{t=1}^{T} \mathcal{N}\left(\boldsymbol{y}^t \middle| \boldsymbol{H}\boldsymbol{x}^t, \sigma_n^2 I_M\right) \qquad [3.1]$$

where $\boldsymbol{Y} \in \mathbb{R}^{M\times T}$ contains the M/EEG measurements, $\boldsymbol{X} \in \mathbb{R}^{N\times T}$ the dipole amplitudes and $\boldsymbol{H} \in \mathbb{R}^{M\times N}$ represents the head operator (usually called "lead-field matrix"). $\boldsymbol{m}^t$ is the tth column of matrix $\boldsymbol{M}$ and I_M is the identity matrix of size M.

3.2.2 PRIOR DISTRIBUTIONS

The unknown parameter vector associated with the likelihood [3.1] is $\boldsymbol{\theta} = \{\boldsymbol{X}, \sigma_n^2\}$. In order to perform Bayesian inference, we assign priors to these parameters as follows.

3.2.2.1 PRIOR FOR THE DIPOLE AMPLITUDES

As stated in Section 3.1, we want to build a regularization combining an ℓ_0 pseudo norm with an ℓ_1 norm of the M/EEG source localization solution. It can be easily shown that using a Laplace prior is the Bayesian equivalent of an ℓ_1 norm regularization whereas a Bernoulli prior can be associated with the ℓ_0 pseudo-norm. As a consequence, we propose to use a Bernoulli-Laplace prior. 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 is able to estimate both small and high amplitudes due to its large value around zero and its fat tails. The corresponding prior for the ijth element of $\boldsymbol{X}$ is

$$f(x_{ij}|\omega,\lambda) = (1-\omega)\delta(x_{ij}) + \frac{\omega}{2\lambda}\exp\left(-\frac{|x_{ij}|}{\lambda}\right) \qquad [3.2]$$

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_{ij} are *a priori* independent, the prior distribution of $\boldsymbol{X}$ can be written

$$f(\boldsymbol{X}|\omega,\lambda) = \prod_{j=1}^{T}\prod_{i=1}^{N} f(x_{ij}|\omega,\lambda). \qquad [3.3]$$

3.2.2.2 PRIOR FOR THE NOISE VARIANCE

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

$$f(\sigma_n^2) \propto \frac{1}{\sigma_n^2} 1_{\mathbb{R}^+}(\sigma_n^2) \qquad [3.4]$$

where $1_{\mathbb{R}^+}(x)$ is the indicator function on $\mathbb{R}^+$. This prior is a very classical choice for a non-informative prior (see, e.g., [69] for motivations).

3.2.3 HYPERPARAMETER PRIORS

The hyperparameter vector associated with the previous priors is $\boldsymbol{\Phi} = \{\omega, \lambda\}$ as displayed in the direct acyclic graph of Fig. 3.1. We consider 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.

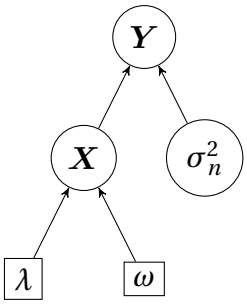


Figure 3.1: Directed acyclic graph of the hierarchy used for the Bayesian model.

3.2.3.1 HYPERPRIOR FOR ω

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

$$\omega \sim \mathcal{U}\Big(0, \omega_{\max}\Big) \tag{3.5}$$

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]

3.2.3.2 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

$$f(\lambda) \propto \frac{1}{\lambda} 1_{\mathbb{R}^+}(\lambda). \tag{3.6}$$

[1] We have observed that setting $\omega_{\max} < 1$ (instead of $\omega_{\max} = 1$) yields faster convergence of the sampler studied in Section 3.3. $\omega_{\max} = 0.5$ was set by cross-validation.

3.2.4 POSTERIOR DISTRIBUTION

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

$$f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\boldsymbol{Y}) \propto f(\boldsymbol{Y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{\Phi}) f(\boldsymbol{\Phi}) \quad [3.7]$$

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 3.3 studies an MCMC method that can be used to sample the joint posterior distribution [3.7] and build Bayesian estimators of the unknown model parameters using the generated samples.

3.3 GIBBS SAMPLER

This section considers a Gibbs sampler [69] which generates samples iteratively from the conditional distributions of [3.7], *i.e.*, from $f(\sigma_n^2|\boldsymbol{Y}, \boldsymbol{X})$, $f(\lambda|\boldsymbol{X})$, $f(\omega|\boldsymbol{X})$ and $f(x_{ij}|\boldsymbol{Y}, \boldsymbol{X}_{-ij}, \omega, \lambda, \sigma_n^2)$ where $\boldsymbol{M}_{-ij}$ denotes the matrix $\boldsymbol{M}$ whose ijth element has been replaced by zero. The next sections explain how to sample from the conditional distributions of the unknown parameters and hyperparameters associated with the posterior of interest [3.7]. The resulting algorithm is also summarized in Algorithm 3.1.

Algorithm 3.1 Gibbs sampler.

Initialize $\boldsymbol{X}$ with the sLoreta solution
repeat
  Sample σ_n^2 according to $f(\sigma_n^2|\boldsymbol{X}, \boldsymbol{Y})$.
  Sample λ according to $f(\lambda|\boldsymbol{X})$.
  Sample ω according to $f(\omega|\boldsymbol{X})$.
  for $j = 1$ to T **do**
    for $i = 1$ to M **do**
      Sample x_{ij} according to $f(x_{ij}|\boldsymbol{Y}, \boldsymbol{X}_{-ij}, \omega, \lambda, \sigma_n^2)$.
    end for
  end for
until convergence

3.3.1 CONDITIONAL DISTRIBUTIONS

3.3.1.1 CONDITIONAL DISTRIBUTION OF σ_n^2

Using [3.7], it is straightforward to derive the conditional posterior distribution of σ_n^2 which is the following inverse gamma distribution

$$\sigma_n^2|\boldsymbol{X},\boldsymbol{Y} \sim \mathscr{IG}\Big(\sigma_n^2\Big|\frac{M}{2},\frac{||\boldsymbol{Y}-\boldsymbol{H}\boldsymbol{X}||^2}{2}\Big) \tag{3.8}$$

where $||.||$ represents the euclidean norm.

3.3.1.2 CONDITIONAL DISTRIBUTION OF λ

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

$$\lambda|\boldsymbol{X} \sim \mathscr{IG}\Big(\lambda \mid ||\boldsymbol{X}||_0, ||\boldsymbol{X}||_1\Big) \tag{3.9}$$

where $||.||_1$ is the ℓ_1 norm and $||.||_0$ the ℓ_0 norm.

3.3.1.3 CONDITIONAL DISTRIBUTION OF ω

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 $[0,\omega_{\max}]$

$$\omega|\boldsymbol{X} \sim \mathscr{B}e_{[0,\omega_{\max}]}(1+||\boldsymbol{X}||_0, 1+M-||\boldsymbol{X}||_0). \tag{3.10}$$

3.3.1.4 CONDITIONAL DISTRIBUTION OF x_{ij}

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

$$f(x_{ij}|\boldsymbol{Y},\boldsymbol{X}_{-ij},\omega,\lambda,\sigma_n^2,z_{ij}) = \begin{cases} \delta(x_{ij}) & \text{if } z_{ij}=0 \\ \mathscr{N}_+(\mu_{ij,+},\sigma_i^2) & \text{if } z_{ij}=1 \\ \mathscr{N}_-(\mu_{ij,-},\sigma_i^2) & \text{if } z_{ij}=-1 \end{cases} \tag{3.11}$$

where $\mathscr{N}_+$ and $\mathscr{N}_-$ denote the truncated Gaussian distributions on $\mathbb{R}^+$ and $\mathbb{R}^-$, respectively. The variable z_{ij} is a discrete random variable that takes value 0 with probability $\omega_{1,ij}$, value 1 with probability $\omega_{2,ij}$ and -1 with probability

$\omega_{3,ij}$. More precisely, defining $\boldsymbol{v}_{ij} = \boldsymbol{y}^j - \boldsymbol{H}\boldsymbol{x}^j_{-i}$, the weights $(\omega_{l,ij})_{1\leq l\leq 3}$ are defined as

$$\omega_{l,ij} = \frac{u_{l,ij}}{\sum_{l=1}^{3} u_{l,ij}} \tag{3.12}$$

where

$$\begin{aligned}
u_{1,ij} &= 1-\omega \\
u_{2,ij} &= \frac{\omega}{2\lambda}\exp\left(\frac{(\mu^+_{ij})^2}{2\sigma_i^2}\right)\sqrt{2\pi\sigma_i^2}\,C(\mu^+_{ij},\sigma_i^2) \\
u_{3,ij} &= \frac{\omega}{2\lambda}\exp\left(\frac{(\mu^-_{ij})^2}{2\sigma_i^2}\right)\sqrt{2\pi\sigma_i^2}\,C(-\mu^-_{ij},\sigma_i^2)
\end{aligned} \tag{3.13}$$

and

$$\begin{aligned}
\sigma_i^2 &= \frac{\sigma_n^2}{||\boldsymbol{h}_i||^2} \\
\mu^+_{ij} &= \sigma_i^2\left(\frac{\boldsymbol{h}_i^\top \boldsymbol{v}_{ij}}{\sigma_n^2} - \frac{1}{\lambda}\right) \\
\mu^-_{ij} &= \sigma_i^2\left(\frac{\boldsymbol{h}_i^\top \boldsymbol{v}_{ij}}{\sigma_n^2} + \frac{1}{\lambda}\right) \\
C(\mu,\sigma^2) &= \frac{1}{2}\left[1+\operatorname{erf}\left(\frac{\mu}{\sqrt{2\sigma^2}}\right)\right].
\end{aligned} \tag{3.14}$$

The truncated Gaussian distributions are sampled according to the algorithm specified in [70].

3.3.2 Parameter estimators

Using all the generated samples, the potential scale reduction factors (PSRFs) of all the model parameters and hyperparameters were calculated for each iteration. The samples corresponding to iterations where the PSRFs were above 1.2 were discarded since they were considered to be part of the burn-in period of the sampler, as recommended by [71]. The rest were kept for calculating the estimators of the model parameters and hyperparameters following

$$\hat{\boldsymbol{Z}} \triangleq \underset{\bar{\boldsymbol{Z}} \in \{0,1,-1\}^N}{\arg\max} \left(\#\mathscr{M}(\bar{\boldsymbol{Z}}) \right) \quad [3.15]$$

$$\hat{p} \triangleq \frac{1}{\#\mathscr{M}(\hat{\boldsymbol{Z}})} \sum_{m \in \mathscr{M}(\hat{\boldsymbol{Z}})} p^{(m)} \quad [3.16]$$

where $\#\mathscr{S}$ denotes the cardinal of set $\mathscr{S}$, $\mathscr{M}(\bar{\boldsymbol{Z}})$ is the set of iteration numbers m for which $\boldsymbol{Z}^{(m)} = \bar{\boldsymbol{Z}}$ after the burn-in period of the Gibbs sampler and $p^{(m)}$ is the m-th sample of $p \in \{\boldsymbol{X}, \lambda, \sigma_n^2, \omega\}$. Thus the estimator $\hat{\boldsymbol{Z}}$ in [3.15] is the maximum a posteriori estimator of $\hat{\boldsymbol{Z}}$ whereas the estimator used for all the other sampled variables in [3.16] is the minimum mean square error (MMSE) estimator conditionally to $\hat{\boldsymbol{Z}}$.

3.4 Experimental results

This section reports different experiments conducted to evaluate the performance of the proposed M/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 in [36]. The upper bound of the sparsity level was set to $\omega_{\max} = 0.5$, which is much larger than the expected value of ω. The results obtained with synthetic and real data are reported in two separate sections.

3.4.1 Synthetic data

3.4.1.1 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 [72] was considered. This model corresponds to MRI measurements of a 27 year old male using the boundary element model implemented by the OpenMEEG package [73]. 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 $\boldsymbol{H} \in \mathbb{R}^{19\times 1003}$.
For each of the activations $\boldsymbol{X} \in \mathbb{R}^{1003\times 1}$ 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.

3.4.1.2 PERFORMANCE EVALUATION

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

- Localization error [25]: 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 [74]: 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 [74]. 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 [75] 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 [3.17]

$$\sum_j \sum_k w_{j\to k} |\boldsymbol{r}_{x_j^{\mathrm{nz}}} - \boldsymbol{r}_{\widehat{x}_k^{\mathrm{nz}}}| \qquad [3.17]$$

where x_j^{nz} denotes a non-zero element of the ground truth, $\widehat{x}_k^{\mathrm{nz}}$ 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)

$$\sum_j w_{j\to k} = \widehat{x}_k^{\mathrm{nz}}, \quad \sum_k w_{j\to k} = x_j^{\mathrm{nz}}. \qquad [3.18]$$

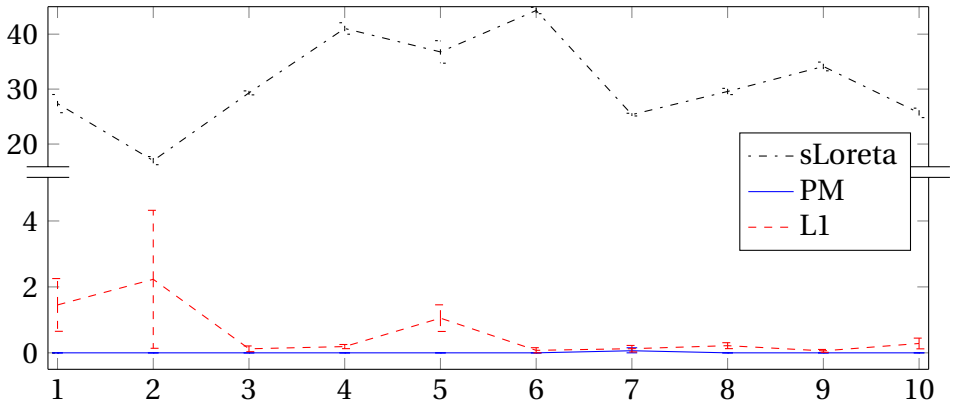
Note that [3.17] defines a similarity measure between the amount of activity in the non-zero elements of the ground truth and the estimated solution. The minimum cost of [3.17] can be obtained using the simplex method of linear programming. The transportation cost of an estimated solution is finally 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 [76] and sLoreta. The regularization parameter of sLoreta was computed by cross-validation using the method recommended in [36]. The weighted ℓ_1 norm was implemented using the alternating direction method of multipliers (ADMM) with the technique used in [77]. The regularization parameter was chosen so that $||\boldsymbol{Y} - \boldsymbol{H}\widehat{\boldsymbol{X}}|| \approx ||\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{X}||$ according to the discrepancy principle [78].

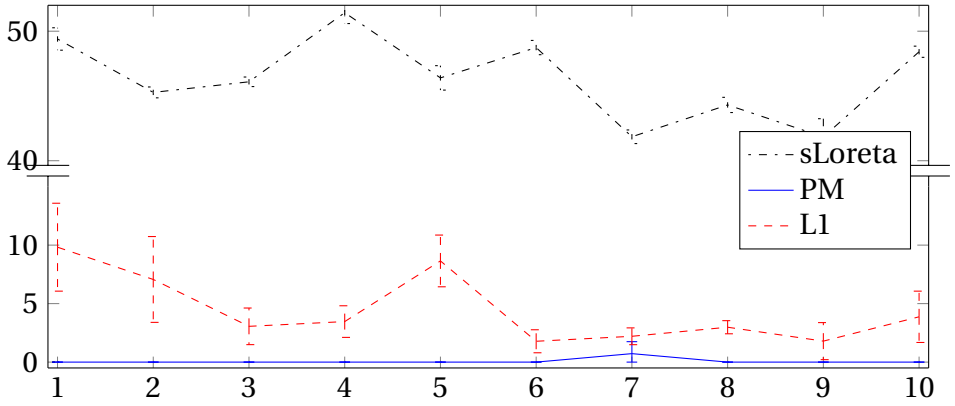
3.4.1.3 SINGLE DIPOLE

The first kind of experiment consisted of a set of ten simulations that have a single dipole active referred in the following as single dipole activations #1 through #10. With these activations, 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. 3.2 showing the good performance of the proposed method (indicated by PM).

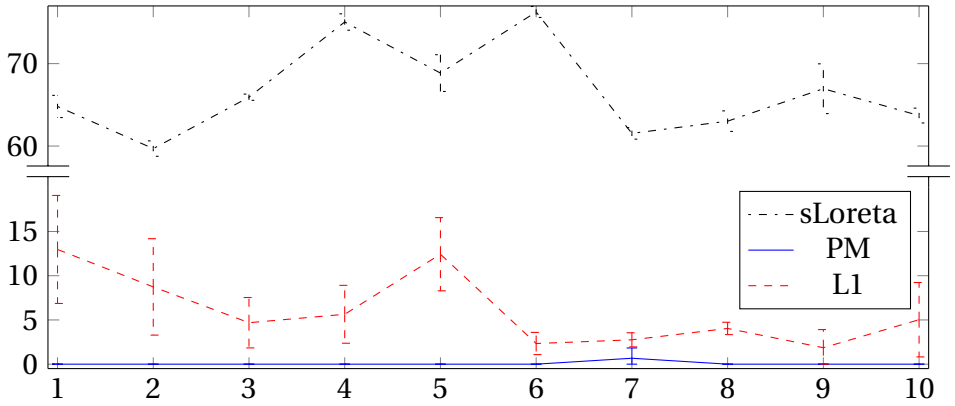
The brain activities detected by the proposed method and the weighted ℓ_1 norm solution are illustrated in Fig. 3.3 for a representative simulation. Our algorithm managed to perfectly recover the activity for 9 out of the 10 activations



(a) Center-mass localization error (mm)

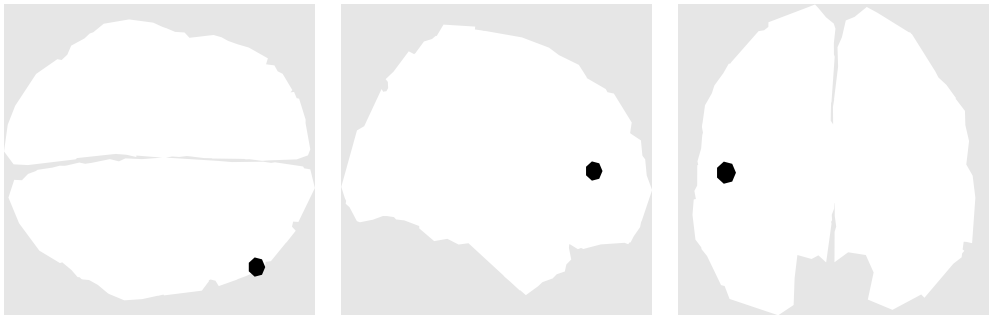


(b) Excitation extension (mm)

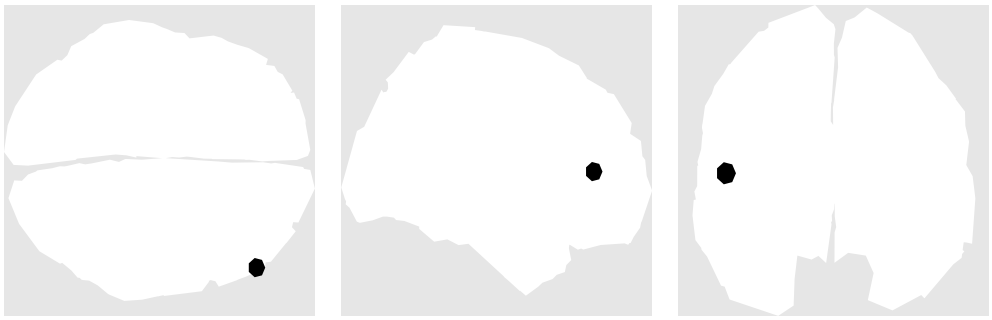


(c) Transportation cost (mm)

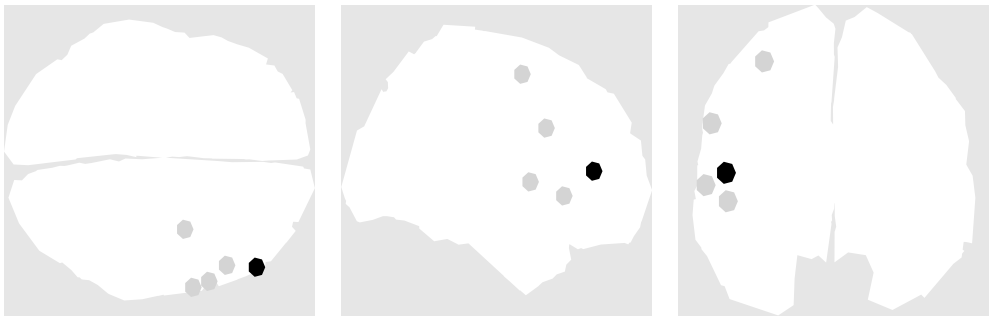
Figure 3.2: Simulation results for single dipole experiments. The horizontal axis indicates the activation number. The error bars show the standard deviation over 12 Monte Carlo runs.



(a) Ground truth

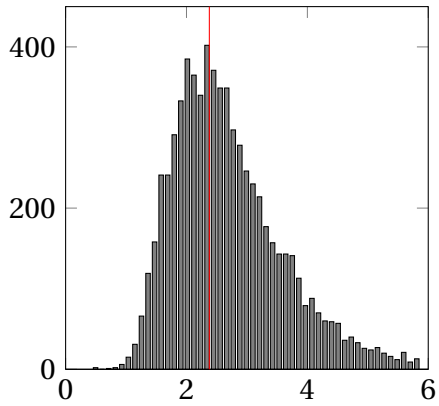


(b) Proposed method

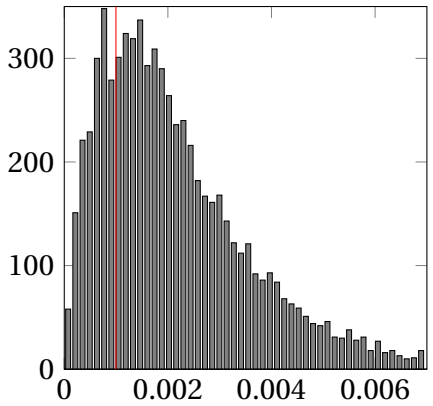


(c) Weighted ℓ_1 norm regularization

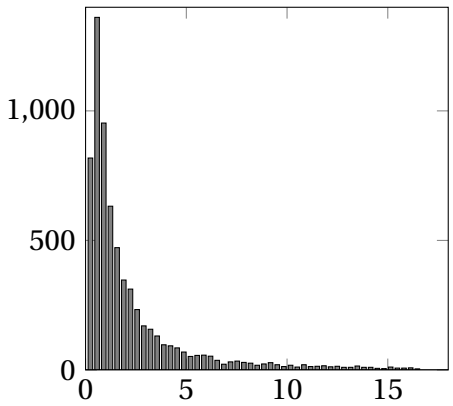
Figure 3.3: Brain activity for one single dipole experiment (activation #5).



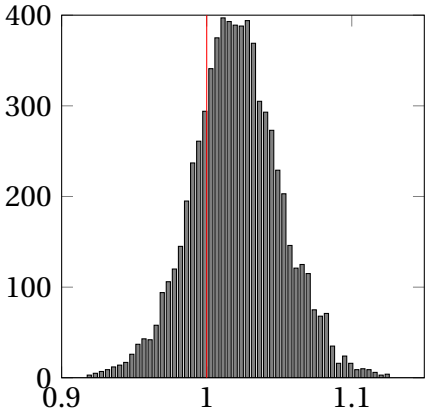
(a) Histogram of σ_n^2. Estimated mean value: 2.75. Ground truth: 2.38.



(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.

Figure 3.4: 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.

(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. 3.4 shows the histograms of the samples of σ_n^2, ω, λ and the amplitude of the active dipole x_{ij} for one of the simulations corresponding to simulation #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 x_{ij} 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.

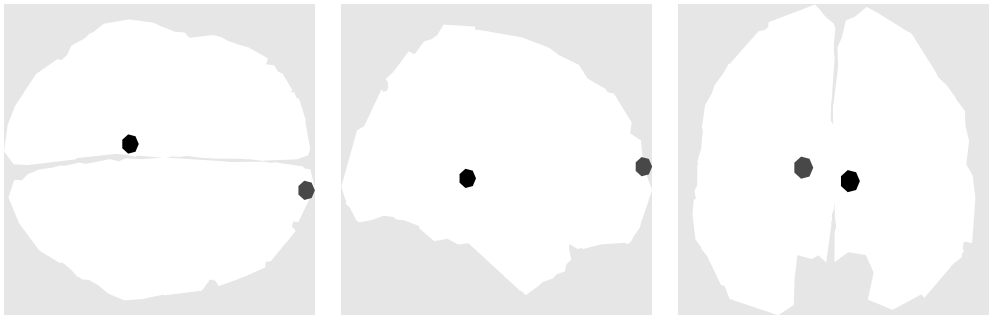
3.4.1.4 MULTIPLE DISTANT DIPOLES

The second kind of experiments evaluates the performance of the proposed algorithm when several dipoles are activated at the same time in 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).

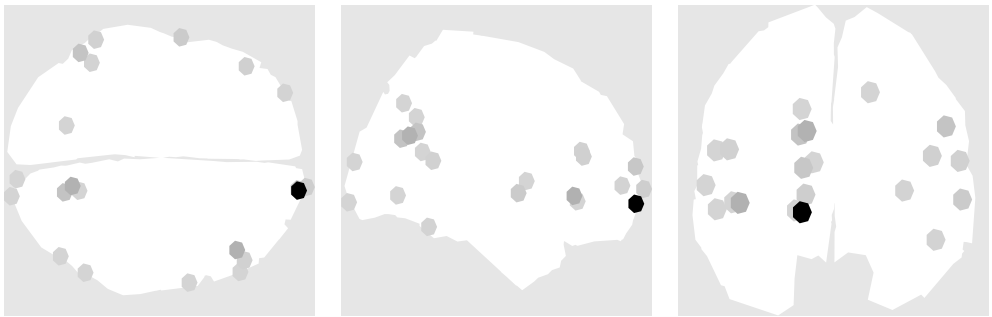
The brain activities associated with two representative simulations corresponding to two and three dipoles are illustrated in Figs. 3.5 and 3.6. The activation #2 associated with two distant dipoles displayed in Fig. 3.5 is an interesting case for which the weighted ℓ_1 norm regularization fails completely to recover one of the dipoles. The activation #4 displayed in Fig. 3.6 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. 3.7 for the different experiments. The transportation costs obtained with the



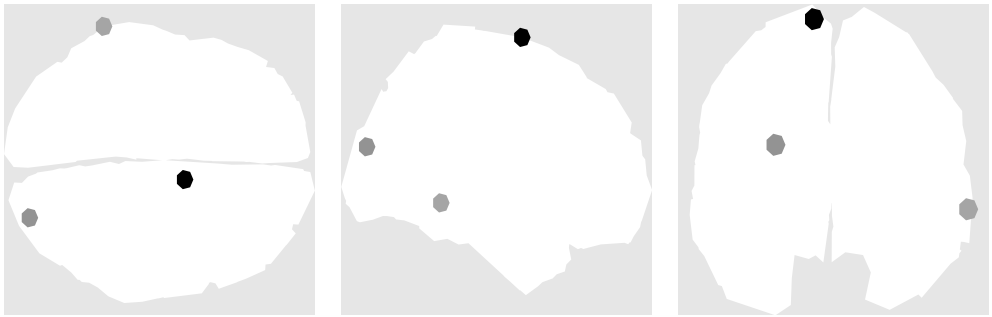
(a) Ground truth

(b) Proposed method

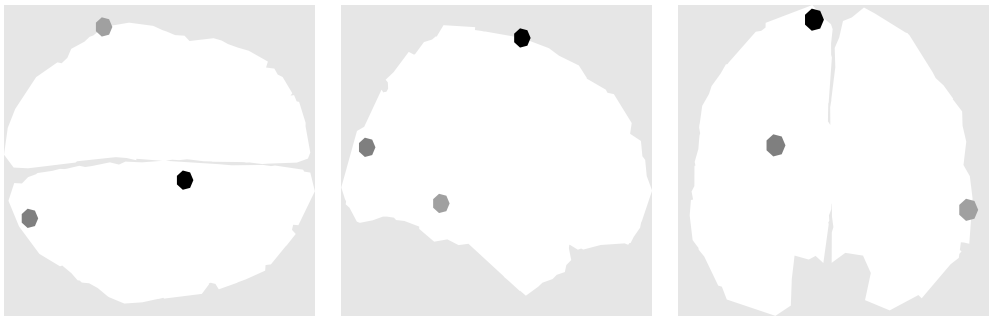


(c) Weighted ℓ_1 norm regularization

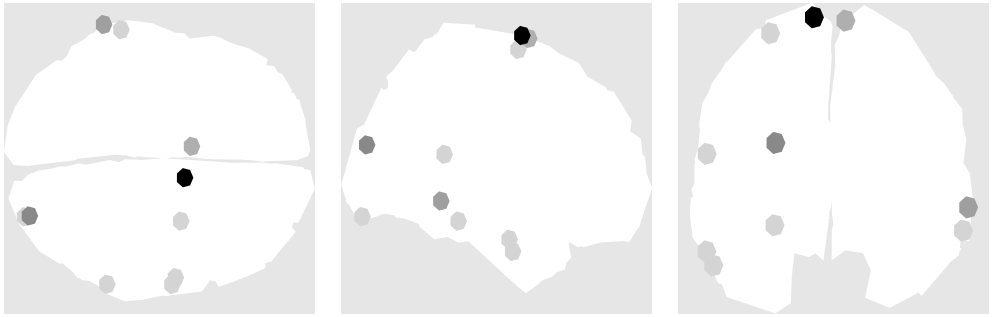
Figure 3.5: Brain activity for a multiple distant dipole experiment (activation # 2 that has two active dipoles).



(a) Ground truth



(b) Proposed method



(c) Weighted ℓ_1 norm regularization

Figure 3.6: Brain activity for a multiple distant dipole experiment (activation # 3 that has three active dipoles).

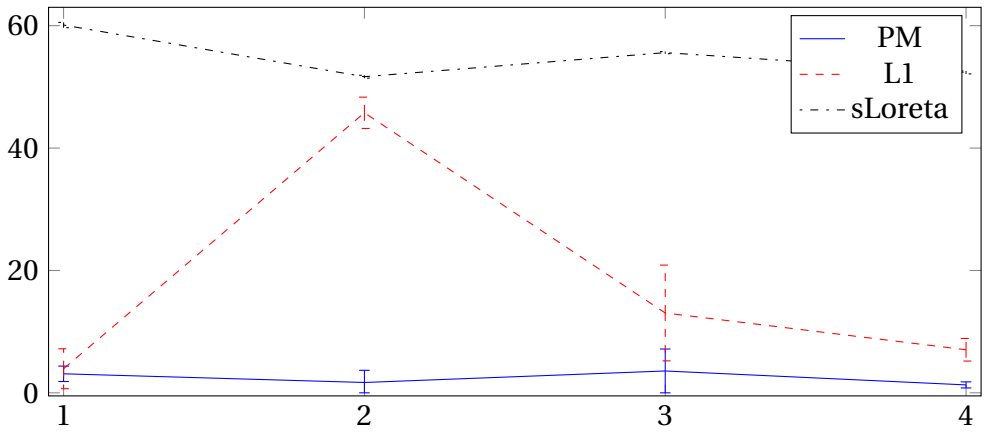


Figure 3.7: Transportation cost for multiple distant dipoles experiments, the horizontal axis indicates the activation number (1 and 2: two active dipoles, 3 and 4: three active dipoles). The error bars show the standard deviation over 12 Monte Carlo runs.

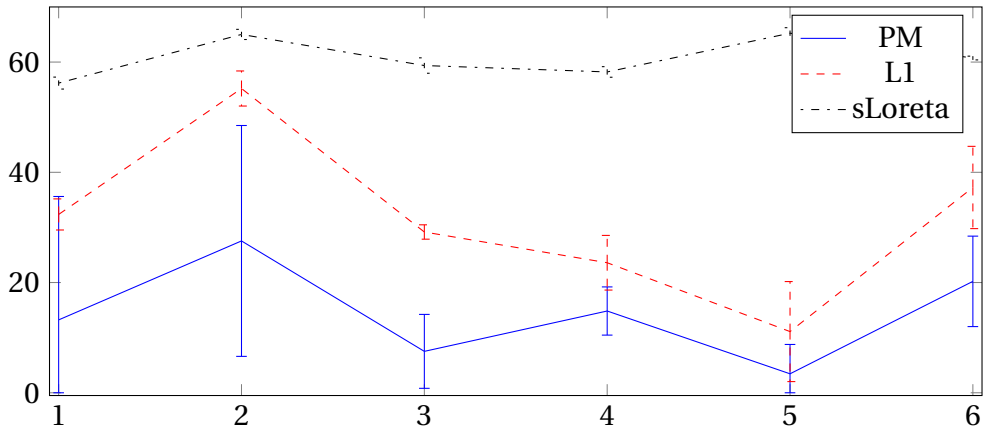
proposed method are below 3.6mm in all cases and are clearly smaller than those obtained with the other methods. Indeed, the sLoreta transportation costs are between 50 and 62mm, and the transportation costs associated with the weighted ℓ_1 norm regularization are between 3.9mm and 13mm, except for the activation #2 where it fails to recover one of the dipoles as previously stated.

3.4.1.5 MULTIPLE CLOSE DIPOLES

The third kind of experiments evaluates the performance of the proposed algorithm for active dipoles that have close spatial positions. More precisely, we randomly chose the following sets of dipoles

- (i) Two pairs of dipoles spaced approximately 50 (activations #1 and #2).
- (ii) Two pairs of dipoles spaced approximately 30mm (activations #3 and #4).
- (iii) Two pairs of dipoles spaced approximately 10mm (activations #5 and #6).

Figure 3.8 compares the transportation costs obtained with the different methods. Since it is much harder to distinguish the activity produced by two



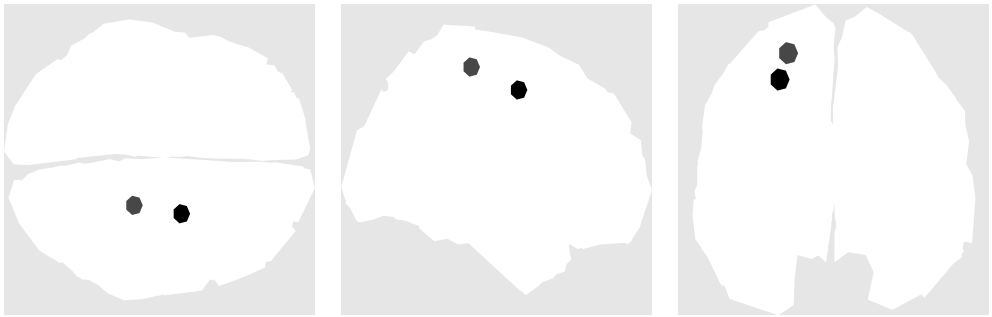
(a)

Figure 3.8: 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.

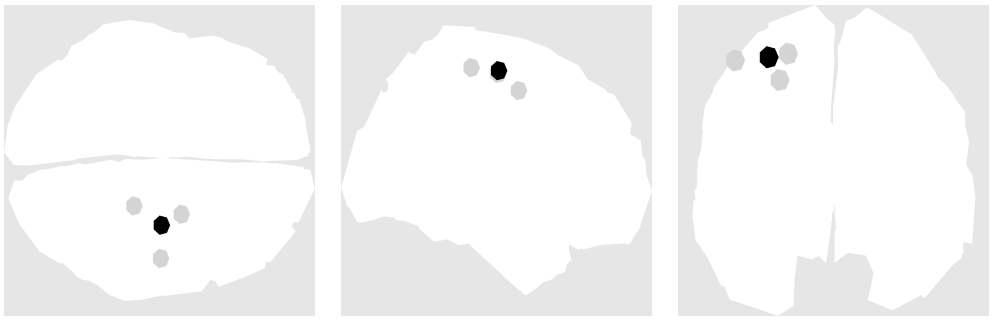
close dipoles, the transportation costs associated with the proposed method and the weighted ℓ_1 norm regularization are considerably higher than those obtained previously. However, the transportation costs obtained with the proposed algorithm are still below those obtained with the two other estimation strategies. Some interesting cases can be observed in Figs. 3.9 and 3.10. Figure 3.9 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. 3.10, 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.

3.4.2 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 [51] using the default parameters in the SPM



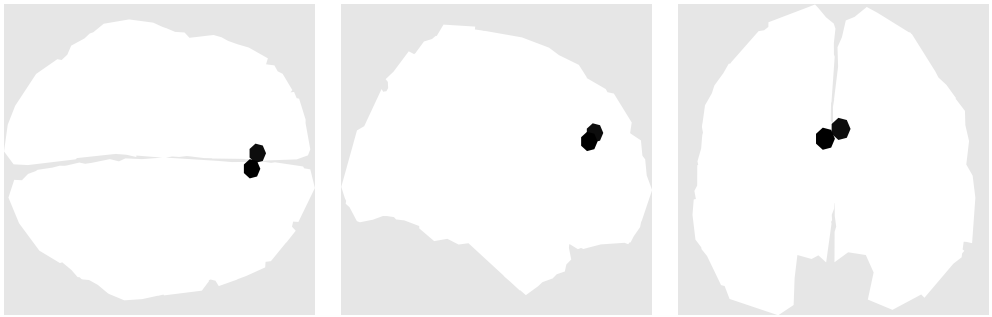
(a) Ground truth



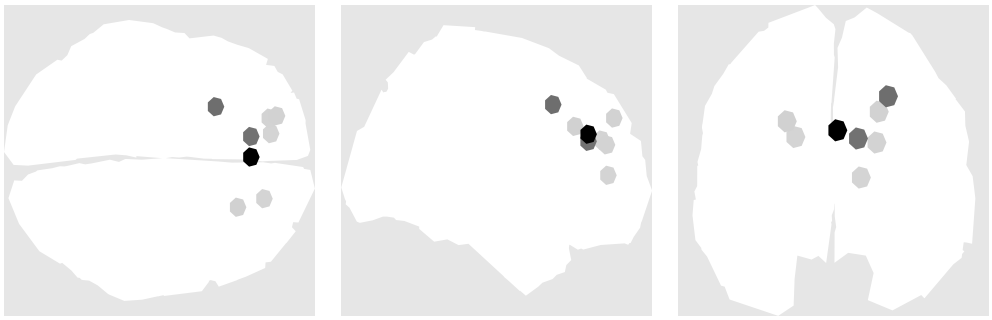
(b) Proposed method

(c) Weighted ℓ_1 norm regularization

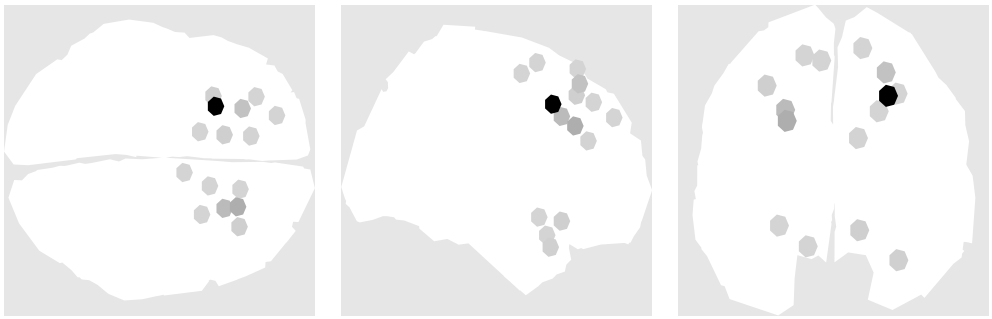
Figure 3.9: Brain activity for multiple close dipoles (activation # 4 that has a 30mm separation between dipoles).



(a) Ground truth



(b) Proposed method



(c) Weighted ℓ_1 norm regularization

Figure 3.10: Brain activity for multiple close dipoles (activation # 6 that has a 10mm separation between dipoles).

software. [2]

3.4.2.1 AUDITORY EVOKED RESPONSES

The used data set was extracted from the MNE software [79, 80]. It corresponds to an evoked response to left-ear auditory pure-tone stimulus using a realistic BEM head model. 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 $\boldsymbol{H} \in \mathbb{R}^{59\times 1884}$. We processed the eight time samples corresponding to 80ms ≤ t ≤ 126ms, *i.e.*, associated with the highest activity period in the M/EEG measurements as displayed in Fig. 3.11.

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

3.4.2.2 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

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

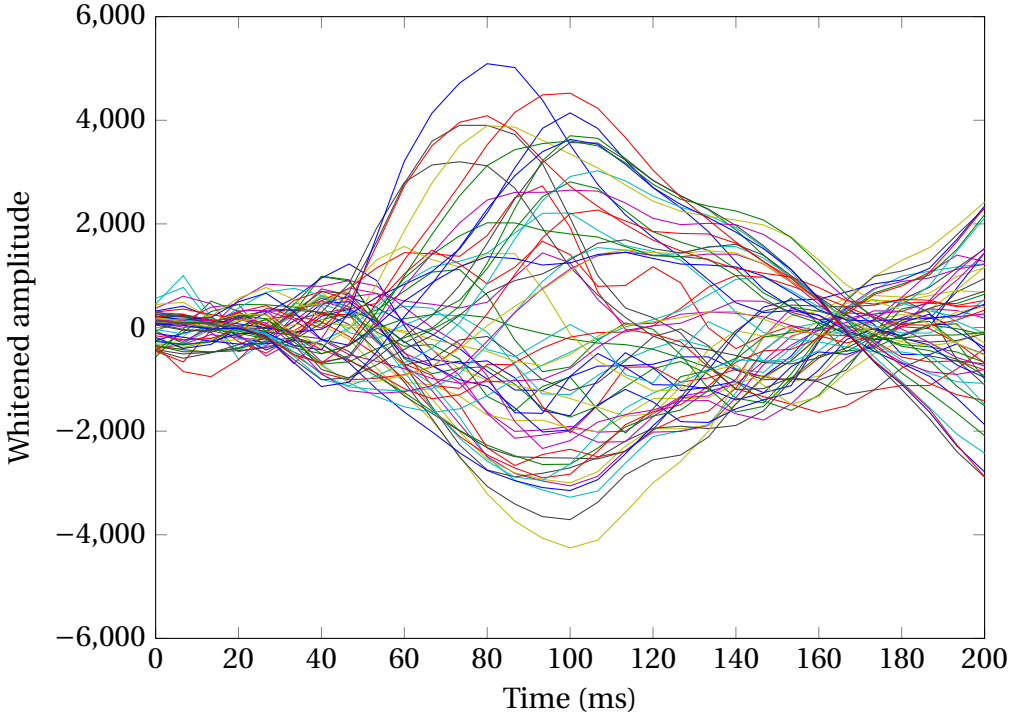
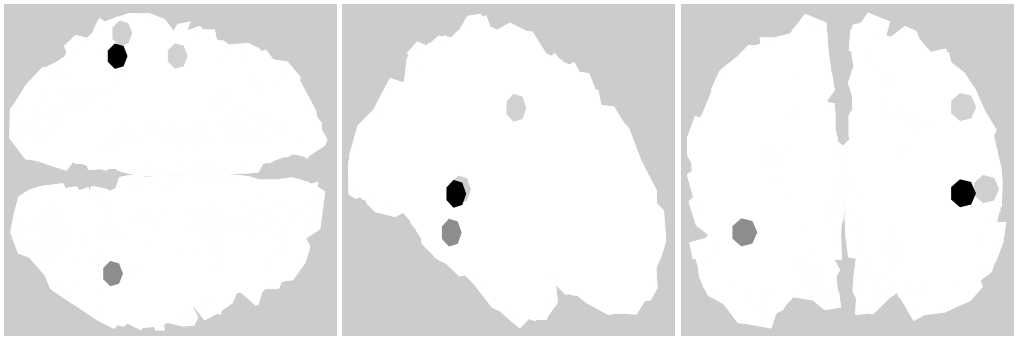
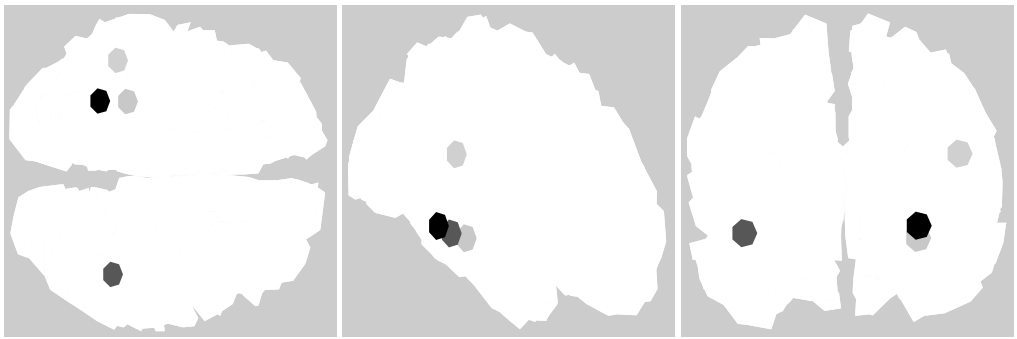


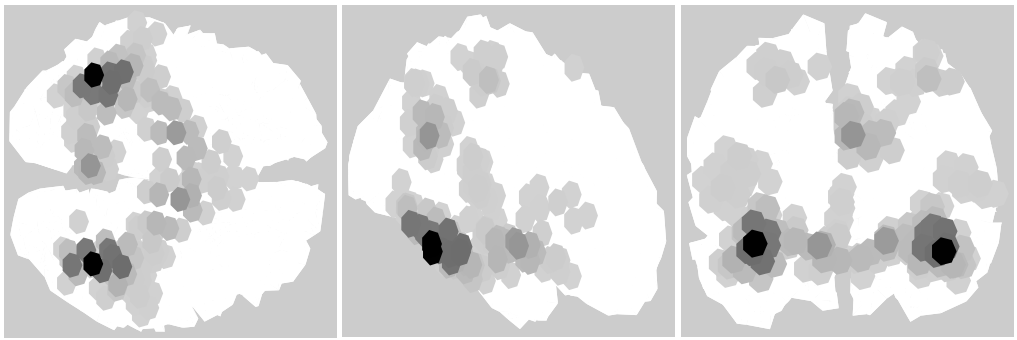
Figure 3.11: M/EEG measurements for the real data application.



(a) Proposed method

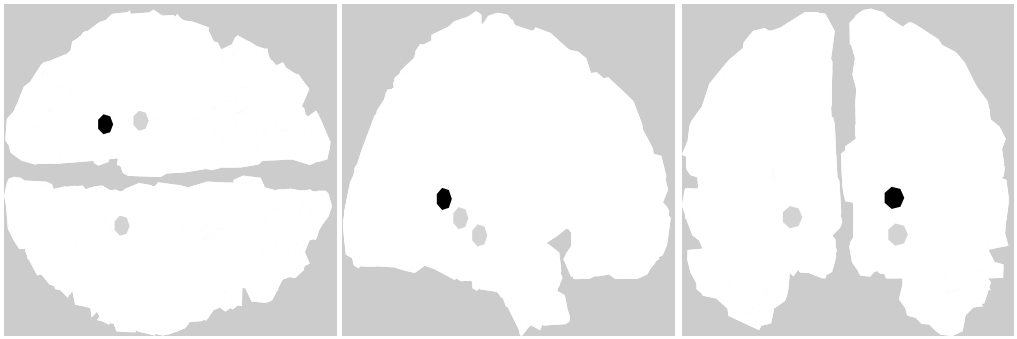


(b) Weighted ℓ_1 norm regularization



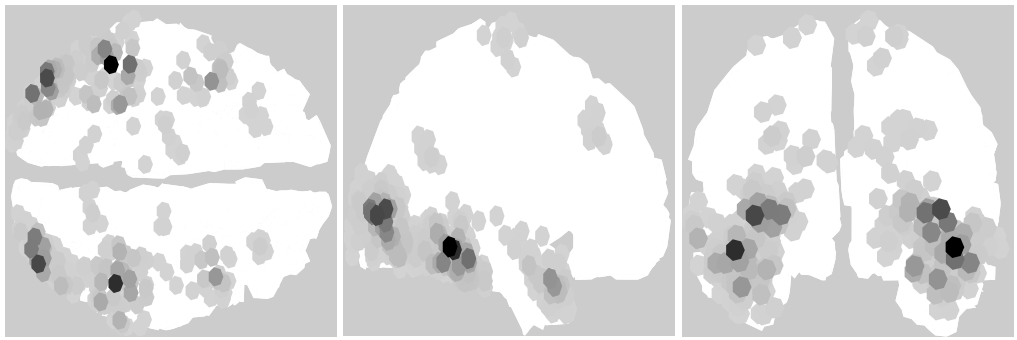
(c) MSP algorithm

Figure 3.12: Brain activity for the auditory evoked responses from 80ms to 126ms.



(a) Proposed method

(b) Weighted ℓ_1 norm regularization



(c) MSP algorithm

Figure 3.13: Brain activity for the faced evoked responses for 160ms.

paradigm see [81]). 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 $\boldsymbol{H} \in \mathbb{R}^{128\times 3004}$.

The brain activities detected by the three algorithms for $t = 160$ms (time sample with highest brain activity) are presented in Fig. 3.13. 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 [82]. 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.

3.4.3 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 in [77]. The proposed method was run with four parallel chains (as stated in Section 3.4.1.1) until the potential scale reduction factor (PSRF) of all the generated samples was below 1.2 as recommended in [71]. The average running times of each of the methods are presented in Tab. 3.1. 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

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

times in order to set their regularization parameters by cross-validation.

3.5 CONCLUSION

In this chapter we presented a Bayesian model for M/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 M/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.

The following chapter extends the proposed Bayesian model in order to reduce its computational cost significantly and to take advantage of the temporal

structure of the M/EEG measurements by promoting structured sparsity.

CHAPTER 4

BAYESIAN M/EEG SOURCE LOCALIZATION USING STRUCTURED SPARSITY

Contents

4.1 INTRODUCTION

The model introduced in the previous chapter is able to reconstruct the brain activity with better results than the algorithm based on a re-weighted ℓ_1 regularization. However, it does not take advantage of the temporal structure of the M/EEG measurements. This can be done by promoting structured sparsity with mixed-norms regularizations such as the ℓ_{21} mixed norm [43]. This kind of regularization promotes sparsity among different dipoles (via the ℓ_1 portion of the norm) but groups all the time samples of the same dipole together, forcing them all to be either jointly active or inactive (with the ℓ_2 norm portion). This chapter studies a new model for M/EEG source localization that uses a Bayesian approximation of the ℓ_{20} mixed norm regularization. It is expected to provide better sparsity than ℓ_{21} (since ℓ_0 promotes sparsity better than ℓ_1) since it exploits the temporal structure of the brain activation unlike the model presented in the previous chapter.

4.2 PROPOSED BAYESIAN MODEL

This chapter presents a hierarchical Bayesian model that estimates the brain activity with a structured sparsity constraint by using a multivariate Bernoulli Laplace prior (approximating the weighted ℓ_{20} mixed norm regularization) [4]. The posterior distribution of this model is too complex to derive closed form expressions for the conventional Bayesian estimators of its unknown parameters. Consequently, a Markov chain Monte Carlo method is investigated to generate samples asymptotically distributed according to the posterior distribution of interest. In order to avoid the sampler to get stuck around local maxima, specific Metropolis-Hastings moves are introduced, allowing new modes of the posterior to be explored. These moves are based on multiple dipole shifts (moving active dipoles to neighboring positions) and inter-chain proposals (exchanging samples between parallel MCMC chains) that significantly accelerate the convergence speed of the proposed sampler. These proposals generate candidates that are accepted or rejected using a Metropolis-Hastings criterion. The

method is applied to both synthetic and real data showing promising results compared to the more traditional ℓ_{21} mixed norm and the MSP methods.

The Bayesian model proposed to estimate the unknown parameters is defined in Section 4.2. Section 4.3 studies the partially collapsed Gibbs sampler that will be used to generate samples asymptotically distributed according to the posterior of interest. The Metropolis-Hastings moves introduced to accelerate the convergence of the sampler are presented in Section 4.4. Simulation results obtained with synthetic and real M/EEG data are reported in Section 4.5.

4.2.1 Likelihood

As in the previous chapter, we consider an additive white Gaussian noise with a constant variance σ_n^2 over the considered time samples [25] which leads to the same Gaussian likelihood

$$f(\boldsymbol{Y}|\boldsymbol{X},\sigma_n^2) = \prod_{t=1}^{T} \mathcal{N}\left(\boldsymbol{y}^t \middle| \boldsymbol{H}\boldsymbol{x}^t, \sigma_n^2 I_M\right) \tag{4.1}$$

where $\boldsymbol{Y} \in \mathbb{R}^{M\times T}$ contains the M/EEG measurements, $\boldsymbol{X} \in \mathbb{R}^{N\times T}$ the dipole amplitudes and $\boldsymbol{H} \in \mathbb{R}^{M\times N}$ represents the head operator (usually called "lead-field matrix"). $\boldsymbol{m}^t$ is the tth column of matrix $\boldsymbol{M}$ and I_M is the identity matrix of size M.

4.2.2 Prior distributions

4.2.2.1 Prior for the brain activity $\boldsymbol{X}$

To promote structured sparsity of the source activity, we first consider the weighted ℓ_{20} mixed pseudo-norm

$$||\boldsymbol{X}||_{20} = \#\{i : \sqrt{v_i}||\boldsymbol{x}_i|| \neq 0\} \tag{4.2}$$

where $\boldsymbol{m}_i$ is the ith row of $\boldsymbol{M}$, $||.||$ represents the euclidean norm, $v_i = ||\boldsymbol{h}^i||$ is a weight introduced to compensate the depth-weighting effect [25, 38] and $\#\mathcal{S}$ denotes the cardinal of the set $\mathcal{S}$. Since this prior leads to intractable computations, we propose to approximate it by a multivariate Laplace Bernoulli prior for each row of $\boldsymbol{X}$

$$f(\boldsymbol{x}_i|z_i,\lambda) \propto \begin{cases} \delta(\boldsymbol{x}_i) & \text{if } z_i = 0 \\ \exp\left(-\frac{1}{\lambda}\sqrt{v_i}||\boldsymbol{x}_i||\right) & \text{if } z_i = 1 \end{cases} \tag{4.3}$$

where $\delta(.)$ is the Dirac delta function, λ is the exponential distribution parameter and $\boldsymbol{z} \in \{0,1\}^N$ is a vector indicating which rows of $\boldsymbol{X}$ are zero ($z_i = 0$ if the ith row of $\boldsymbol{X}$ is zero) and different from zero ($z_i = 1$ if the ith row of $\boldsymbol{X}$ is different from zero) . To make the analysis easier it is convenient to define the hyperparameter $a = \frac{\sigma_n^2}{\lambda^2}$ which transforms the prior to

$$f(\boldsymbol{x}_i|z_i, a, \sigma_n^2) \propto \begin{cases} \delta(\boldsymbol{x}_i) & \text{if } z_i = 0 \\ \exp\left(-\sqrt{\frac{v_i a}{\sigma_n^2}}||\boldsymbol{x}_i||\right) & \text{if } z_i = 1 \end{cases} \quad [4.4]$$

The elements z_i are assigned a Bernoulli prior distribution with parameter $\omega \in [0,1]$:

$$f(z_i|\omega) = \mathscr{B}\Big(z_i|\omega\Big) \quad [4.5]$$

Note that the Dirac delta function $\delta(.)$ in the prior of $\boldsymbol{x}_i$ promotes sparsity while the multivariate Laplace distribution regulates the amplitudes of the non-zero rows. The parameter ω tunes the balance between them. Indeed, $\omega = 0$ yields $\boldsymbol{X} = 0$ whereas $\omega = 1$ reduces the prior to the Bayesian formulation of the group-lasso [83]. Unfortunately the prior [4.4] leads to an intractable posterior. It is possible to fix this problem by introducing a latent variable vector $\boldsymbol{\tau}^2 \in (\mathbb{R}^+)^N$ as proposed in [84]. More precisely, we use the following gamma and Bernoulli-Gaussian priors for τ_i^2 and $\boldsymbol{x}_i$

$$f(\tau_i^2|a) = \mathscr{G}\Big(\tau_i^2\Big|\frac{T+1}{2}, \frac{v_i a}{2}\Big) \quad [4.6]$$

$$f(\boldsymbol{x}_i|z_i, \tau_i^2, \sigma_n^2) = \begin{cases} \delta(\boldsymbol{x}_i) & \text{if } z_i = 0 \\ \mathscr{N}\Big(\boldsymbol{x}_i\Big|0, \sigma_n^2\tau_i^2 I_T\Big) & \text{if } z_i = 1 \end{cases} \quad [4.7]$$

which can be shown to lead to the desired marginal distribution of $\boldsymbol{x}_i$ [4.4] [84].

In addition assuming the rows of $\boldsymbol{\tau}^2$, $\boldsymbol{z}$, and $\boldsymbol{X}$ are a priori independent leads to the following

$$f(\boldsymbol{\tau}^2|a) = \prod_i^N f(\tau_i^2|a)$$
$$f(\boldsymbol{z}|\omega) = \prod_i^N f(z_i|\omega)$$
$$f(\boldsymbol{X}|\boldsymbol{z},\boldsymbol{\tau}^2,\sigma_n^2) = \prod_i^N f(\boldsymbol{x}_i|z_i,\tau_i^2,\sigma_n^2)$$

4.2.2.2 PRIOR FOR THE NOISE VARIANCE σ_n^2

As in our previous chapter, the noise variance is assigned a Jeffrey's prior

$$f(\sigma_n^2) \propto \frac{1}{\sigma_n^2} 1_{\mathbb{R}^+}(\sigma_n^2) \qquad [4.8]$$

where $1_{\mathbb{R}^+}(x)$ is the indicator function on $\mathbb{R}^+$.

4.2.3 HYPERPARAMETER PRIORS

The proposed method allows one to balance the importance between sparsity of the solution and fidelity to the measurements using two hyperparameters: 1) ω that adjusts the proportion of non-zero rows, and 2) a that controls the amplitudes of the non-zeros. The corresponding hierarchy of parameters and hyperparameters is shown in Fig. 4.1. In contrast to the ℓ_{21} mixed norm our algorithm does not require to adjust these hyperparameters but is able to estimate their values from the data by assigning hyperpriors to them following a so-called hierarchical Bayesian analysis. This section defines the priors assigned to the model hyperparameters.

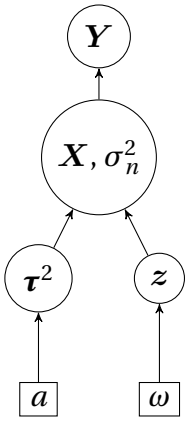


Figure 4.1: Directed acyclic graph for the proposed Bayesian model.

4.2.3.1 HYPERPRIOR OF a

A conjugate gamma prior is assigned to a

$$f(a|\alpha,\beta) = \mathcal{G}\left(a\middle|\alpha,\beta\right) \tag{4.9}$$

with $\alpha = \beta = 1$. These values of α and β yield a vague hyperprior for a. The conjugacy of this hyperprior will make the analysis easier in the sense that the conditional distribution of a required in the Gibbs sampler will also be a gamma distribution.

4.2.3.2 HYPERPRIOR OF ω

A uniform prior on [0, 1] is used for ω

$$f(\omega) = \mathcal{U}\left(\omega\middle|0,1\right) \tag{4.10}$$

reflecting the absence of knowledge for this hyperparameter.[1]

[1] In the model presented in Chapter 3, ω had to be limited to $\omega_{max} = 0.5$ to improve the convergence of the sampler. In this chapter, due to the introduction of latent variable $\boldsymbol{\tau}^2$ and the Metropolis-Hastings moves (that will be explained in the following) the use of the limit $\omega_{max} < 1$ is not necessary (ie: we set $\omega_{max} = 1$).

4.2.4 Posterior distribution

Using the previously described priors and hyperpriors, the posterior distribution of the proposed Bayesian model is

$$f(\boldsymbol{Y},\sigma_n^2,\boldsymbol{X},\boldsymbol{z},a,\boldsymbol{\tau}^2,\omega) \propto f(\boldsymbol{Y}|\boldsymbol{X},\sigma_n^2) f(\boldsymbol{X}|\boldsymbol{\tau}^2,\boldsymbol{z},\sigma_n^2) f(\boldsymbol{z}|\omega) f(\boldsymbol{\tau}^2|a) f(\sigma_n^2) f(a) f(\omega) \quad [4.11]$$

The following section investigates a partially collapsed Gibbs sampler that is used to sample according to the posterior distribution [4.11] and to build estimators of the unknown model parameters and hyperparameters using these generated samples.

4.3 Partially collapsed Gibbs sampler

The posterior distribution [4.11] is intractable and does not allow us to derive closed-form expressions for the estimators of its parameters and hyperparameters. Thus we propose to draw samples from [4.11] using an MCMC method and to use them to estimate the brain activity jointly with the model hyperparameters. More precisely, we investigate a partially collapsed Gibbs sampler that samples the variables z_i and $\boldsymbol{x}_i$ jointly in order to exploit the strong correlation between these two variables. The resulting sampling strategy is summarized in Algorithm 4.1 where $\boldsymbol{M}_{-i}$ denotes the matrix $\boldsymbol{M}$ whose ith row has been replaced by zeros. The corresponding conditional distributions are shown hereafter and their exact derivation can be found in Appendix A.

4.3.1 Conditional distributions

4.3.1.1 Conditional distribution of τ_i^2

The conditional distribution of τ_i^2 is a gamma distribution or a generalized inverse Gaussian distribution depending on the value of z_i. More precisely

$$f(\tau_i^2|\boldsymbol{x}_i,\sigma_n^2,a,z_i) = \begin{cases} \mathcal{G}\left(\tau_i^2 \middle| \frac{T+1}{2}, \frac{v_i a}{2}\right) & \text{if } z_i = 0 \\ \mathcal{GIG}(\tau_i^2 \middle| \frac{1}{2}, v_i a, \frac{\|\boldsymbol{x}_i\|^2}{\sigma_n^2}) & \text{if } z_i = 1. \end{cases} \quad [4.12]$$

Algorithm 4.1 Partially Collapsed Gibbs sampler.

Initialize $\boldsymbol{X} = 0$ and $\boldsymbol{z} = 0$
Sample a and $\boldsymbol{\tau}^2$ from their prior distributions
repeat
  Sample σ_n^2 from $f(\sigma_n^2|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\tau}^2, \boldsymbol{z})$
  Sample ω from $f(\omega|\boldsymbol{z})$
  for $i = 1$ to N **do**
    Sample τ_i^2 from $f(\tau_i^2|\boldsymbol{x}_i, \sigma_n^2, a, z_i)$
    Sample z_i from $f(z_i|\boldsymbol{Y}, \boldsymbol{X}_{-i}, \sigma_n^2, \tau_i^2, \omega)$
    Sample $\boldsymbol{x}_i$ from $f(\boldsymbol{x}_i|z_i, \boldsymbol{Y}, \boldsymbol{X}_{-i}, \sigma_n^2, \tau_i^2)$
  end for
  Sample a from $f(a|\boldsymbol{\tau}^2)$
until convergence

4.3.1.2 Conditional distribution of $\boldsymbol{x}_i$

The conditional distribution of the ith row of $\boldsymbol{X}$ is

$$f(\boldsymbol{x}_i|z_i, \boldsymbol{Y}, \boldsymbol{X}_{-i}, \sigma_n^2, \tau_i^2) = \begin{cases} \delta(\boldsymbol{x}_i) & \text{if } z_i = 0 \\ \mathcal{N}\left(\boldsymbol{x}_i \middle| \boldsymbol{\mu}_i, \sigma_i^2 I_N\right) & \text{if } z_i = 1 \end{cases} \quad [4.13]$$

with

$$\boldsymbol{\mu}_i = \frac{\sigma_i^2 (\boldsymbol{h}^i)^T (\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{X}_{-i})}{\sigma_n^2}, \sigma_i^2 = \frac{\sigma_n^2 \tau_i^2}{1 + \tau_i^2 (\boldsymbol{h}^i)^T \boldsymbol{h}^i}. \quad [4.14]$$

4.3.1.3 Conditional distribution of z_i

The conditional distribution of z_i is a Bernoulli distribution

$$f(z_i|\boldsymbol{Y}, \boldsymbol{X}_{-i}, \sigma_n^2, \tau_i^2, \omega) = \mathcal{B}\left(z_i \middle| 1, \frac{k_1}{k_0 + k_1}\right) \quad [4.15]$$

with

$$k_0 = 1 - \omega, k_1 = \omega \left(\frac{\sigma_n^2 \tau_i^2}{\sigma_i^2}\right)^{-\frac{T}{2}} \exp\left(\frac{||\boldsymbol{\mu}_i||^2}{2\sigma_i^2}\right). \quad [4.16]$$

4.3.1.4 CONDITIONAL DISTRIBUTION OF a

The conditional distribution of $a|\boldsymbol{\tau}^2$ is the following gamma distribution

$$f(a|\boldsymbol{\tau}^2) = \mathscr{G}\Big(a\Big|\frac{N(T+1)}{2}+\alpha, \frac{\sum_i[\nu_i\tau_i^2]}{2}+\beta\Big). \quad [4.17]$$

4.3.1.5 CONDITIONAL DISTRIBUTION OF σ_n^2

The distribution of $\sigma_n^2|\boldsymbol{Y},\boldsymbol{X},\boldsymbol{\tau}^2,\boldsymbol{z}$ is the following inverse gamma distribution

$$f(\sigma_n^2|\boldsymbol{Y},\boldsymbol{X},\boldsymbol{\tau}^2,\boldsymbol{z}) = \mathscr{IG}\Big(\sigma_n^2\Big|\frac{(M+||\boldsymbol{z}||_0)T}{2}, \frac{1}{2}\Big[||\boldsymbol{HX}-\boldsymbol{Y}||^2+\sum_{i\in I_1}\frac{||\boldsymbol{x}_i||^2}{\tau_i^2}\Big]\Big). \quad [4.18]$$

where $||.||_0$ the ℓ_0 norm.

4.3.1.6 CONDITIONAL DISTRIBUTION OF ω

Finally, $\omega|\boldsymbol{z}$ has the following beta distribution

$$f(\omega|\boldsymbol{z}) = \mathscr{B}e\Big(\omega\Big|1+||\boldsymbol{z}||_0, 1+N-||\boldsymbol{z}||_0\Big). \quad [4.19]$$

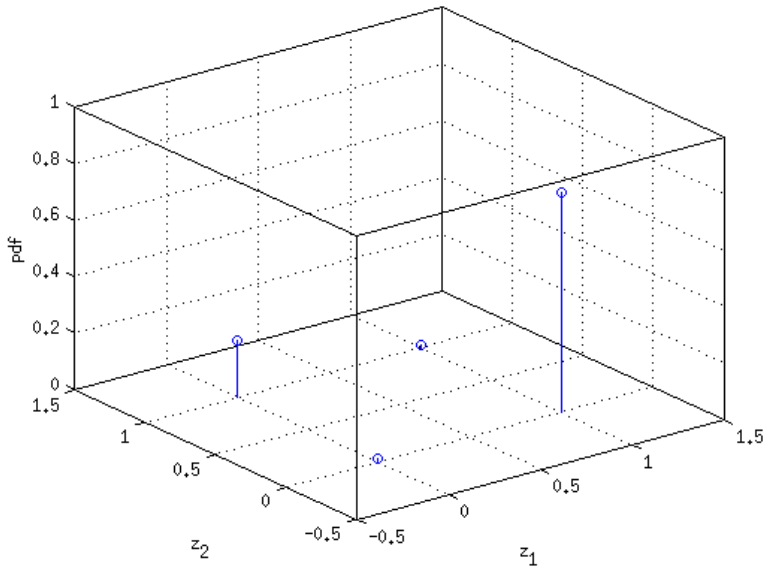


Figure 4.2: Example of posterior distribution of z with local maxima

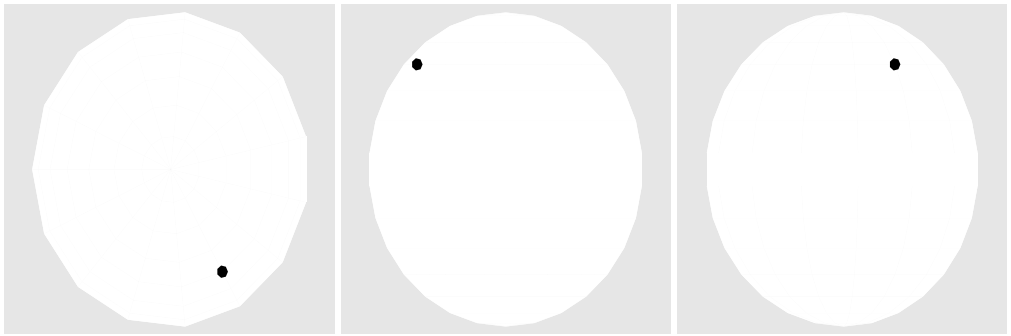
4.4 Improving convergence

4.4.1 Local maxima

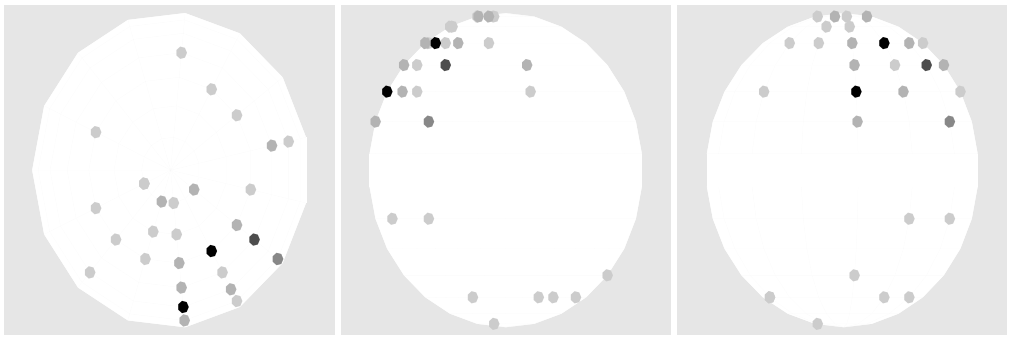
We have observed that the proposed partially collapsed Gibbs sampler may be stuck around local maxima of the variable z from which it is very difficult to escape in a reasonable amount of iterations. Fig. 4.2 illustrates via a simple example that if the sampler gets to $z = \{0,1\}$ it requires to go through intermediary states with very low probability to move to the correct value $z = \{1,0\}$. This kind of situation can occur if the dipoles corresponding to z_1 and z_2 produce similar measurements $\boldsymbol{Y}$ when active so that the probability of having either of them active is much higher than having them both on (or off) at the same time.

4.4.2 Multiple dipole shift proposals

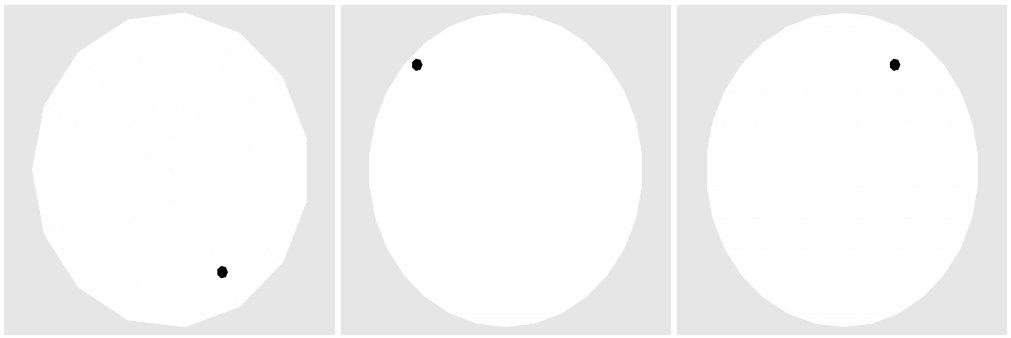
In order to solve this problem, we introduce a Metropolis-Hastings move, that consists in changing several elements of z simultaneously (which would allows us to go from $\{0,1\}$ to $\{1,0\}$ in one step in the previous example) after each sampling iteration. The proposal is accepted or rejected within the Gibbs-Sampler



(a) Ground truth - Axial, coronal and sagittal views respectively



(b) Estimation without proposals - Axial, coronal and sagittal views respectively



(c) Estimation with proposals - Axial, coronal and sagittal views respectively

Figure 4.3: Illustration of the effectiveness of the multiple dipole shift proposals.

using a Metropolis-Hastings criteria which guarantees that the target distribution is preserved [69].

We have observed that one of the most common ways for the algorithm to get stuck in local maxima is by failing to estimate the position of the non-zero elements of $\boldsymbol{z}$. In other words, the sampling scheme detects a non-zero in a position that is not correct but is close (in the sense that it produces similar measurements) to the correct one, which causes the problem described above.

4.4.2.1 SHIFT BASED PROPOSALS

Before describing the proposed move, it is interesting to mention that it was inspired by an idea developed in [85] to perform a spectral analysis of astrophysical data obtained after irregular sampling. The authors of [85] proposed to move a single non-zero element of a binary sequence to a random neighboring position after each iteration of the MCMC sampler. In the presence of a single non-zero element, this move is sufficient to escape from a local maximum of the posterior associated with our M/EEG source localization model. However, when there are several non-zero elements located at wrong locations, proposing to move each one of them separately may not be sufficient to escape from a local maximum. For this reason, we have generalized the scheme presented in [85] by proposing to move a random subset of D estimated non-zeros simultaneously to random neighboring positions. According to our experiments (some of them described in Section 4.5), the simple choice $D = 2$ provides good results in most practical cases. Since there is a high correlation between the variables $\boldsymbol{\tau}^2$ and $\boldsymbol{z}$, it is convenient to update their values jointly. The resulting multiple dipole shift proposal is detailed in Algorithm 4.2. Note that the algorithm of [85] works with 1-dimensional data so they define the neighborhood of the element z_k as $\{z_{k-1}, z_{k+1}\}$. In contrast, we are working with dipoles located in a 3-dimensional brain so the neighborhood definition is non-trivial and will be described in the following.

In Fig. 4.3 we can see the effect of introducing multiple dipole shift proposals (with $D = 2$) in a practical case. The first row of images is the ground truth (a single dipole activation) while the second row shows the probability of finding each dipole active with eight MCMC parallel chains without using proposals after 10.000 iterations. As we can see, the activity is in the correct area but the algorithm is not able to converge to the correct value of $\boldsymbol{z}$. After introducing the multiple dipole shift proposals the sampler converges in less than 1.000 iterations as shown in the third row of the figure.

Algorithm 4.2 Multiple dipole shift proposal.

$\bar{\boldsymbol{z}} = \boldsymbol{z}$
repeat D 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

4.4.2.2 Acceptance probability

In order to guarantee that the generated samples are asymptotically distributed according to the posterior [4.11], all moves resulting from the multiple dipole shift proposal are accepted or rejected with the acceptance probability described in Algorithm 4.2. This acceptance probability requires to compute the following probability distribution

$$f(\boldsymbol{z}_r, \boldsymbol{\tau}_r^2|\boldsymbol{Y}, a, \sigma_n^2, \omega) \propto (1-\omega)^{C_0}\omega^{C_1}(\sigma_n^2)^{-\frac{TC_1}{2}}|\Sigma|^{\frac{T}{2}} \prod_{i\in\boldsymbol{I}_1}(\tau_i^2)^{-\frac{T}{2}} \exp\left(-\frac{\sum_{t=1}^T D^t}{2}\right)\prod_{i=1}^N \mathcal{G}\left(\tau_i^2\middle|\frac{T+1}{2}, \frac{v_i a}{2}\right) \quad [4.20]$$

where $\boldsymbol{r} = \{i : z_i \neq \bar{z}_i\}$, $\boldsymbol{I}_k = \{i : z_{r_i} = k\}$, $C_k = \#\boldsymbol{I}_k$ for $k = \{0, 1\}$ and

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_n^2}\left[(\boldsymbol{H}^{\boldsymbol{I}_1})^T\boldsymbol{H}^{\boldsymbol{I}_1} + \text{diag}\left(\frac{1}{\boldsymbol{\tau}_{\boldsymbol{r}}^2}\right)\right]$$
$$\boldsymbol{\mu}^t = -\frac{\boldsymbol{\Sigma}(\boldsymbol{H}^{\boldsymbol{I}_1})^T(\boldsymbol{H}^{-\boldsymbol{r}}\boldsymbol{x}_{-\boldsymbol{r}}^t - \boldsymbol{y}^t)}{\sigma_n^2}$$
$$\boldsymbol{D}^t = \frac{(\boldsymbol{H}^{-\boldsymbol{r}}\boldsymbol{x}_{-\boldsymbol{r}}^t - \boldsymbol{y}^t)^T(\boldsymbol{H}^{-\boldsymbol{r}}\boldsymbol{x}_{-\boldsymbol{r}}^t - \boldsymbol{y}^t)}{\sigma_n^2} - (\boldsymbol{\mu}^t)^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^t.$$

4.4.2.3 NEIGHBORHOOD DEFINITION

It is obvious that the definition of the neighborhood used to exchange non-zero elements is crucial. Initially, we used a geometrical neighborhood, defined in terms of vertex connexity in the triangular tessellation modeling the brain cortex. However, this definition usually yields very small neighborhoods. This may cause the proposals not to be flexible enough to help the algorithm escape from local maxima.

For this reason we propose a neighborhood definition that considers two dipoles to be neighbors if the correlation between their respective columns is higher than a certain threshold

$$\text{neigh}_\gamma(i) \triangleq \left\{ j \neq i \;\middle|\; |\text{corr}(\boldsymbol{h}^i, \boldsymbol{h}^j)| \geq \gamma \right\} \qquad [4.21]$$

where $\text{corr}(\boldsymbol{v}_1, \boldsymbol{v}_2)$ is the correlation between vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ and where the neighborhood size can be adjusted by setting $\gamma \in [0,1]$ ($\gamma = 0$ corresponds to a neighborhood containing all the dipoles and $\gamma = 1$ corresponds to an empty neighborhood). Note that an additional advantage of this definition is the fact that it allows the approach to be extended to other kinds of inverse problems (different from M/EEG source localization) where no geometrical disposition of the elements of $\boldsymbol{z}$ may be available.

In order to maximize the moves efficiency, the value of γ has to be selected carefully. A very large value of γ will result in proposals not being flexible enough to help the algorithm in escaping local maxima. A very low value of γ will result in a very large amount of possible proposals with many of them being useless leading to a large number of iterations to reach useful moves. Our experiments based on cross-validation have shown that a good compromise is obtained with $\gamma = 0.8$ (see Section 4.5 for illustrations).

4.4.3 INTER-CHAIN PROPOSALS

The multiple dipole shift proposal scheme previously described allows the algorithm to better sample the value of $\boldsymbol{z}$ present in the posterior distribution and is able to find the active dipoles correctly provided the number of active dipoles is small. However, when running multiple MCMC chains in parallel with a higher amount of non-zeros present in the ground truth, it is possible for the different chains to get stuck in different values of $\boldsymbol{z}$. In order to help them converge to the same (most probable) value, it is possible to exchange information between parallel chains to avoid local maxima during their runs

as other approaches do, including Metropolis-coupled MCMC [86], Population MCMC [87] and simulated tempering [88, 89].

In this report, we introduce inter-chain moves by proposing to exchange the values of $\boldsymbol{z}$ and $\boldsymbol{\tau}^2$ between different chains. These moves are accepted with the Metropolis-Hastings probability shown in Algorithm 4.3.

Algorithm 4.3 Inter-chain proposals.

Define a vector $\boldsymbol{c} = \{1, 2, ..., L\}$ where L is the number of chains
for $i = \{1, 2, ..., L\}$
  Choose (and remove) a random element from $\boldsymbol{c}$ and denote it by k
  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

One potential problem introduced by inter-chain proposals is the fact that they require synchronizing the parallel MCMC chains, which decreases the iteration speed of the algorithm. In order to minimize this effect, an inter-chain proposal will be made after each iteration with probability p (adjusted to $\frac{1}{1000}$ by cross validation) according to Algorithm 4.3.

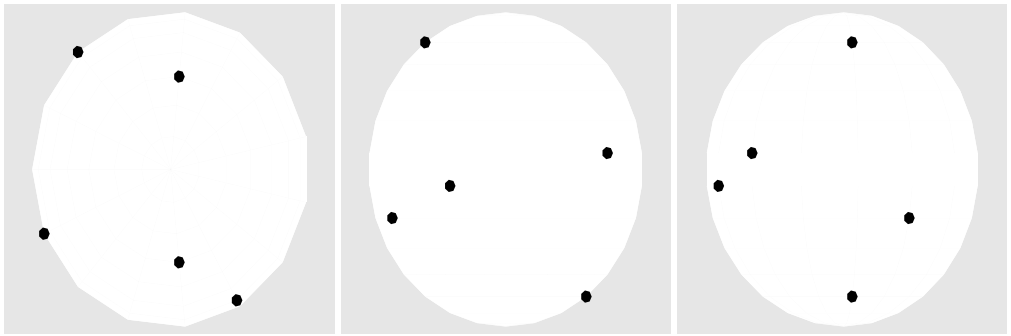
The benefit of introducing inter-chain proposals is illustrated in Fig. 4.4. The first row of images of this figure displays the five non-zeros present in the ground truth. Without using inter-chain proposals the different chains can converge to different modes after 100.000 iterations as illustrated in the second row (which displays the probability of finding each dipole active with 8 parallel MCMC chains). The introduction of inter-chain proposals causes all the different chains to converge to the same (correct) value of $\boldsymbol{z}$ in less than 5.000 iterations as illustrated in the third row of Fig. 4.4.

4.4.4 Parameter estimators

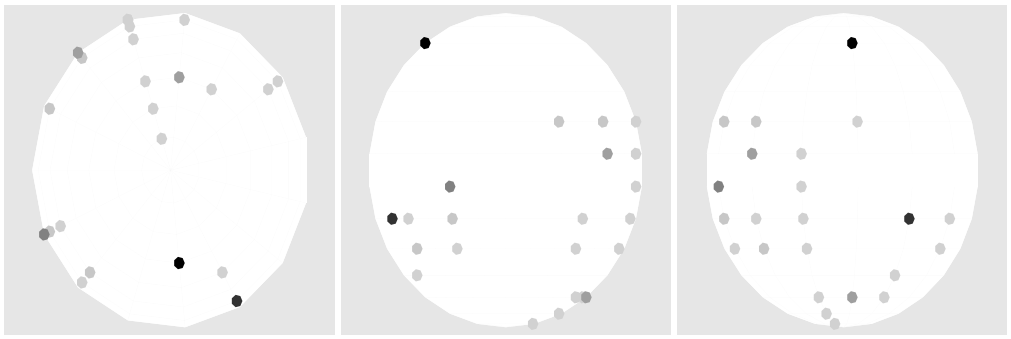
As in the previous chapter, we will use the point estimators defined as follows

$$\hat{\boldsymbol{z}} \triangleq \underset{\bar{\boldsymbol{z}} \in \{0,1\}^N}{\arg\max} \left(\#\mathcal{M}(\bar{\boldsymbol{z}}) \right) \tag{4.22}$$

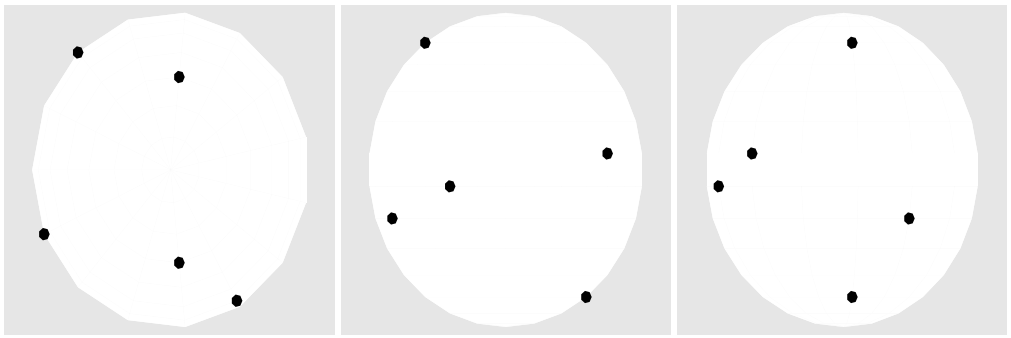
$$\hat{p} \triangleq \frac{1}{\#\mathcal{M}(\hat{\boldsymbol{z}})} \sum_{m \in \mathcal{M}(\hat{\boldsymbol{z}})} p^{(m)} \tag{4.23}$$



(a) Ground truth - Axial, coronal and sagittal views respectively



(b) Estimation without proposals - Axial, coronal and sagittal views respectively



(c) Estimation with proposals - Axial, coronal and sagittal views respectively

Figure 4.4: Efficiency of the inter-chain proposal.

where $\mathscr{M}(\bar{\boldsymbol{z}})$ is the set of iteration numbers m for which the sampled variable $\boldsymbol{z}^{(m)} = \bar{\boldsymbol{z}}$ after the burn-in period and p stands for any of the variables $\boldsymbol{X}$, a, σ_n^2, ω and $\boldsymbol{\tau}^2$. Thus the estimator $\hat{\boldsymbol{z}}$ [4.22] corresponds to a maximum a posteriori estimator whereas the estimator used for all the other sampled variables [4.23] is a minimum mean square error (MMSE) estimators conditionally upon the value of $\hat{\boldsymbol{z}}$.

It is interesting to note that the proposed method provides the full distribution of the unknown parameters and is not limited to point-estimate as the methods based on the ℓ_{21} mixed norm.

It will be shown in the next section that in certain conditions, such as low SNR, the a-posteriori distribution of $\boldsymbol{z}$ has components with comparable probabilities. These similar values of $\boldsymbol{z}$ are usually minor variations of each other (changing one of the dipoles to a neighboring position for instance). In this case, after convergence the algorithm oscillates between several values of $\boldsymbol{z}$ which allows the proposed method to identify several possible solutions (each of them corresponding to a different value of $\boldsymbol{z}$) with their corresponding probabilities. This is an advantage over the mixed ℓ_{21} mixed norm method that is only able to provide a point-estimate of the solution.

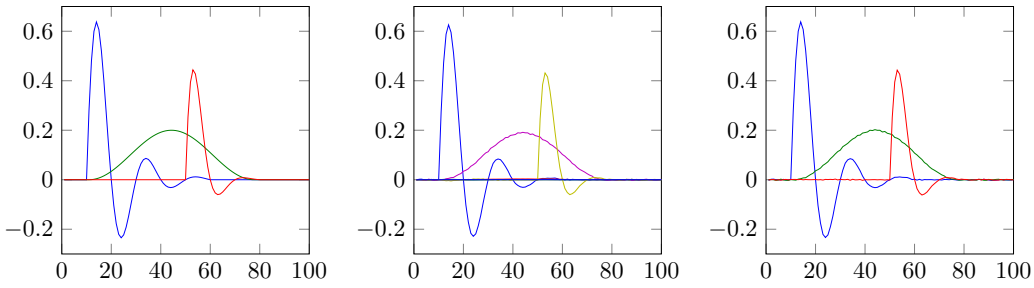
4.5 Experimental results

4.5.1 Synthetic data

Synthetic data are first considered to compare the ℓ_{21} mixed norm approach with the proposed method using a 212-dipole Stok three-sphere head model [19] with 41 electrodes. Three kinds of activations are performed: (1) three dipoles with low SNR, (2) five dipoles with high SNR and (3) multiple dipoles with high SNR.

4.5.1.1 Three-dipoles with low SNR

Three dipoles were assigned excitations defined as synthetic damped sinusoidal waves with frequencies between 5 and 20Hz. These excitations were 500ms long (a period corresponding to a stationary dipole activity) and sampled at 200Hz. Different levels of noise were used to compare the performance of the proposed method with the weighted ℓ_{21} mixed norm. The parameters of our multiple dipole shift proposal were set to $D = 2$ and $\gamma = 0.8$ and eight MCMC



(a) Ground truth (b) ℓ_{21}-mixed norm estimation (c) Proposed method

Figure 4.5: Estimated waveforms for three dipoles with SNR = 30dB.

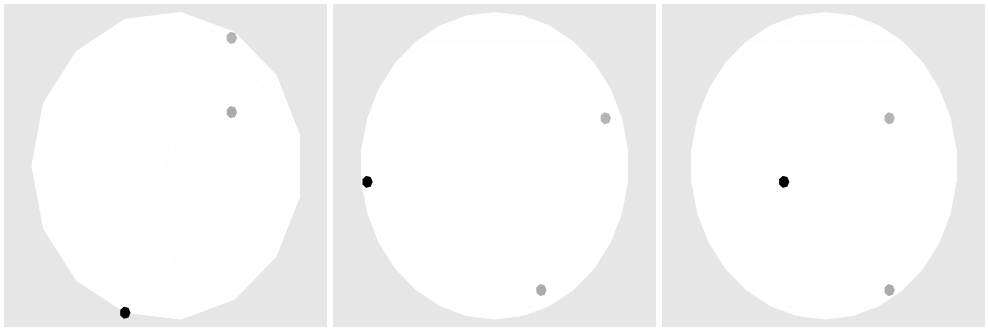
Active non-zeros	Percentage of samples
1, 2, 3	43%
1, 2, 4	22%
1, 2, 5	11%
1, 2, 6	7%
1, 2, 7	6%
Others	11%

Table 4.1: Three dipoles with SNR = -3dB: modes explored after convergence. Positions 1, 2 and 3 correspond to the non-zero elements of the ground truth.

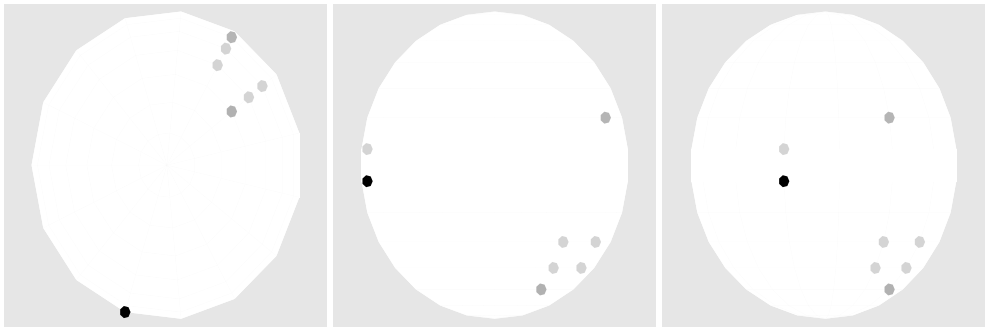
chains were run in parallel. For the ℓ_{21} mixed norm approach, the value of the regularization parameter λ was chosen using cross-validation to get the best possible result.

The results for SNR = 30dB are shown in Fig. 4.5 and Fig. 4.6. Both algorithms seem to provide the same solution that is equal to the ground truth. The only minor difference is that the ℓ_{21}-mixed norm regularization presents some dipoles (around ten) with very low but non-zero values, while our algorithm only detects the three non-zeros as real activity, as shown in Fig. 4.6.

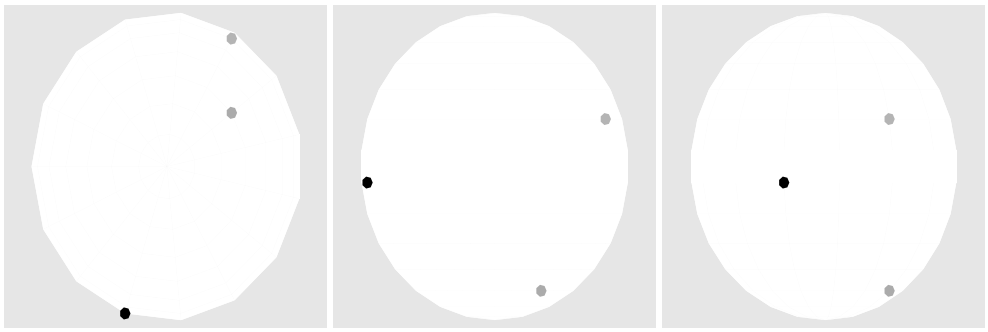
The estimated dipole locations with SNR = $-$3dB are shown in Fig. 4.8 whereas the corresponding estimated waveforms are shown in Fig. 4.7. Note that only the dipoles with highest activity are displayed for the ℓ_{21} approach. The approach based on the ℓ_{21} norm manages to recover only two of the three non-zero activities at the correct positions and seems to underestimate considerably the amplitude of the activity. This is a known problem caused by approximating



(a) Ground truth - Axial, coronal and sagittal views respectively

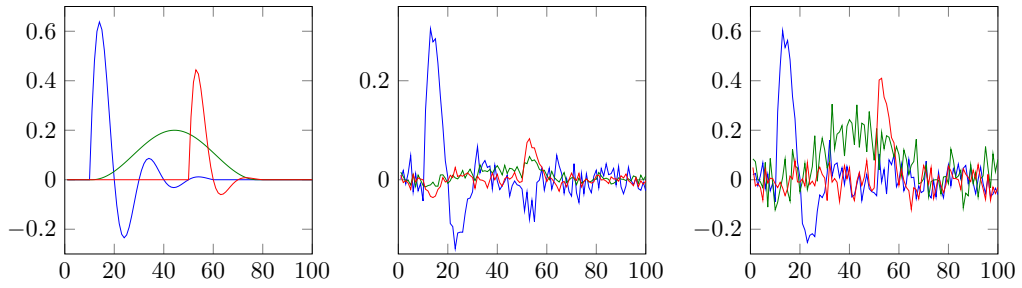


(b) ℓ_{21} - Axial, coronal and sagittal views respectively



(c) Proposed method - Axial, coronal and sagittal views respectively

Figure 4.6: Estimated activity for three dipoles and SNR = 30dB.



(a) Ground truth (b) ℓ_{21}-mixed norm estimation (c) Proposed method

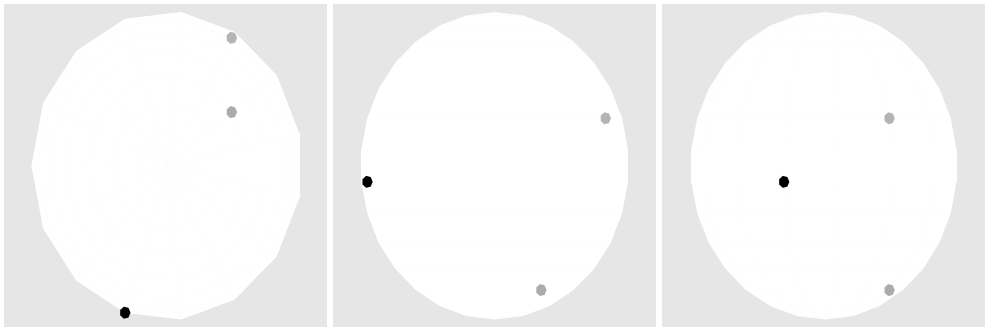
Figure 4.7: Estimated waveforms for three dipoles with SNR = -3dB.

the ℓ_0 pseudo-norm by the ℓ_1 norm, since the later penalizes high amplitudes while the former penalizes all non-zero values equally. Our algorithm oscillates between several values of $\boldsymbol{z}$ (specified in Table 4.1). However, the most probable value of $\boldsymbol{z}$ found by the algorithm is the correct one whereas the other ones have one of the active non-zeros moved to a close neighborhood.

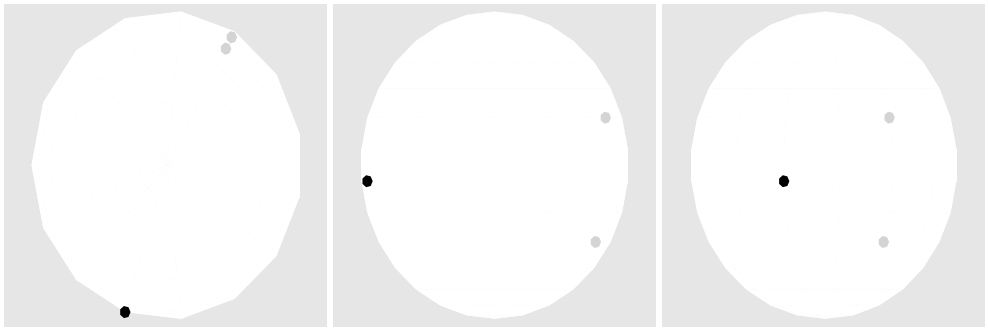
The proposed method does not only allow a point-estimation of the activity but can also be used to estimate uncertainties associated with the activity. For instance, Fig. 4.9 shows the confidence intervals of the activity estimation (mean $\pm$ 2 $\cdot$ standard deviations). The actual ground truth activation is clearly located within two standard deviations of the estimated mean value obtained with the proposed algorithm. The histogram of the generated hyperparameters ω, a and σ_n^2 are shown in Fig. 4.10. They are clearly in good agreement with the actual values of the corresponding parameters. The PSRF's are displayed in Fig. 4.11. It is possible to see that the PSRF's tend to 1 as the iterations increase, showing that the sampler convergence is satisfactory.

4.5.1.2 FIVE DIPOLES

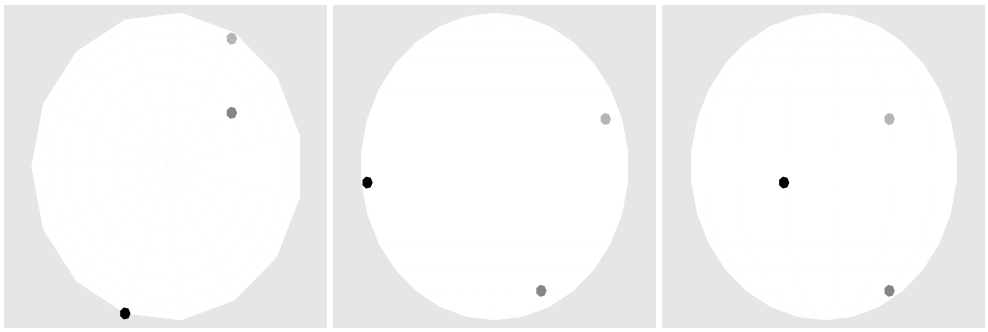
For the second kind of experiments, five dipoles were activated with the same damped sinusoidal wave of 5Hz. The activations were sampled at 200Hz and scaled in amplitude so that each of them produced the same energy in the measurements. Noise was added to the measurements to obtain SNR = 30dB. For the ℓ_{21} mixed norm regularization, the regularization parameter was set according to the uncertainty principle which consists in finding a solution $\hat{\boldsymbol{X}}$ such that $||\boldsymbol{H}\hat{\boldsymbol{X}} - \boldsymbol{Y}|| \approx ||\boldsymbol{H}\boldsymbol{X} - \boldsymbol{Y}||$ [78]. Eight MCMC chains were run in par-



(a) Ground truth - Axial, coronal and sagittal views respectively

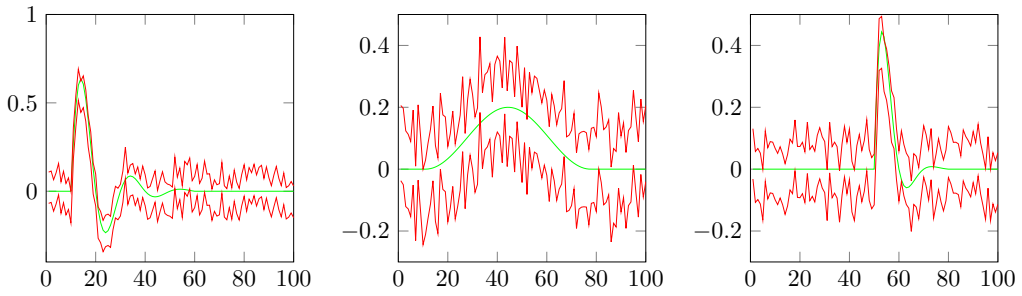


(b) ℓ_{21} - Axial, coronal and sagittal views respectively



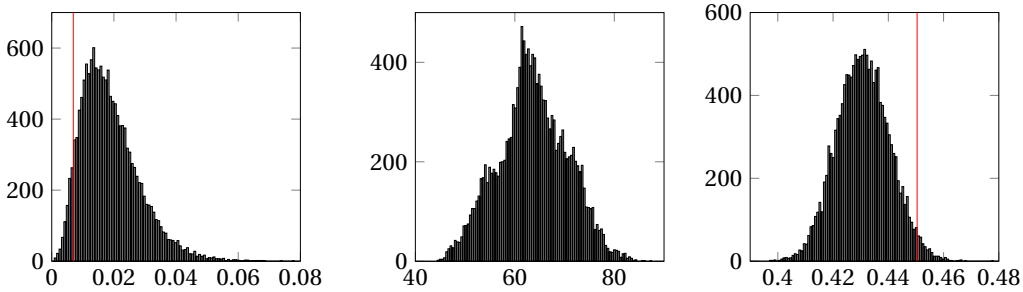
(c) Proposed method - Axial, coronal and sagittal views respectively

Figure 4.8: Estimated activity for three dipoles and SNR = -3dB.



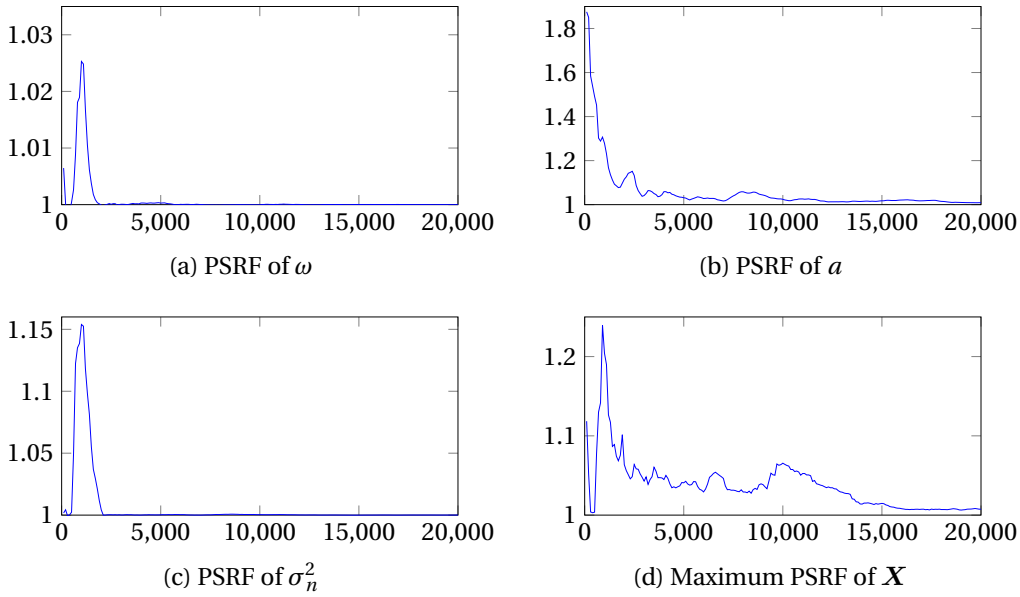
(a) Dipole waveform 1 (b) Dipole waveform 2 (c) Dipole waveform 3

Figure 4.9: Estimated boundaries $\mu \pm 2\sigma$ for the three dipole simulation with SNR = -3dB.



(a) Histogram of ω (b) Histogram of a (c) Histogram of σ_n^2

Figure 4.10: Three dipoles with SNR = -3dB: histograms of the hyperparameters. The actual values of ω and σ_n^2 are marked with a red vertical line.



(a) PSRF of ω

(b) PSRF of a

(c) PSRF of σ_n^2

(d) Maximum PSRF of $\boldsymbol{X}$

Figure 4.11: Three dipoles with SNR = -3dB: PSRFs of sampled variables.

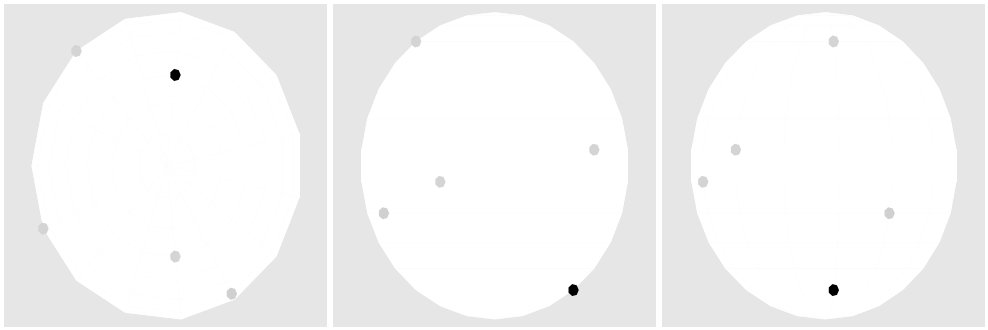
allel for the proposed method. Only the five non-zeros of the estimated activity with highest energy in the measurements were considered.

The results are displayed in Fig. 4.12 and 4.13. In the first figure we are able to see that the proposed method is able to recover the five locations perfectly while the ℓ_{21} norm only detects four activations, two of them being not at the correct locations. In the waveforms displayed in Fig. 4.13 we can see that, while both methods are able to recover the general activity pattern, the proposed method closely matches the waveform amplitude while the ℓ_{21} mixed norm does not. This is partially due to the fact that it detects some of the activations at the wrong positions (waveforms 4 and 5) and partially due to the tendency to underestimate the activation amplitudes inherent to the ℓ_0 to ℓ_1 convex relaxation.

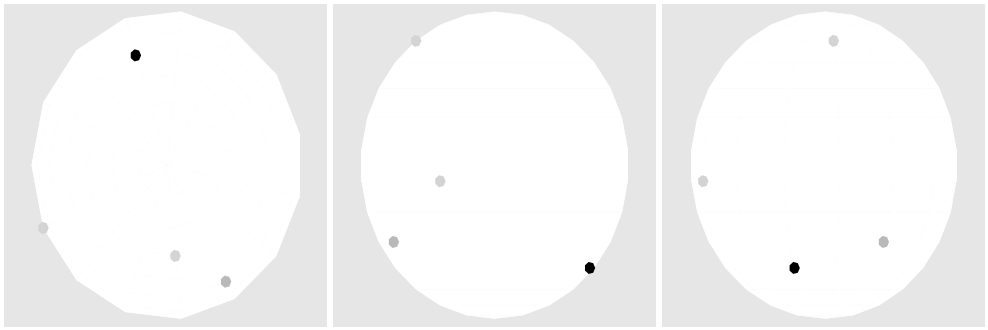
4.5.1.3 MULTIPLE DIPOLES

In this section, we compare the detection capabilities of the algorithm with respect to the ℓ_{21}-mixed norm approach by varying the amount of non-zeros present in the ground truth.

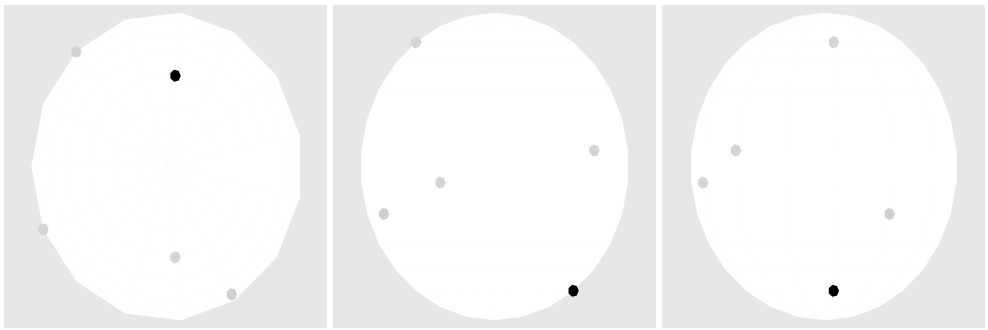
In each simulation of this section, P dipoles were activated with damped



(a) Ground truth - Axial, coronal and sagittal views respectively

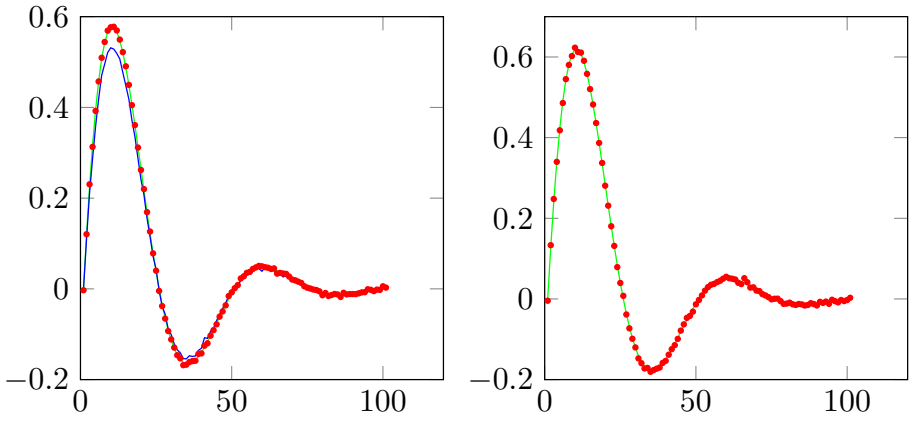


(b) ℓ_{21} dipole locations - Axial, coronal and sagittal views respectively



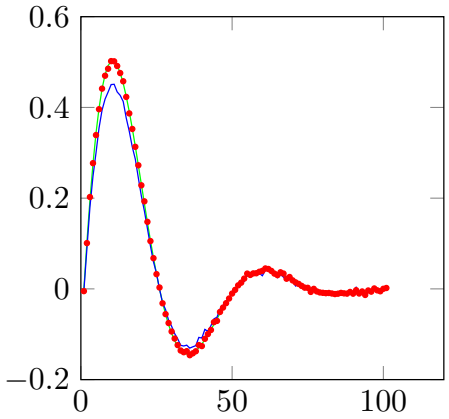
(c) Proposed Method - Axial, coronal and sagittal views respectively

Figure 4.12: Estimated activity for five dipoles and SNR = 30dB.

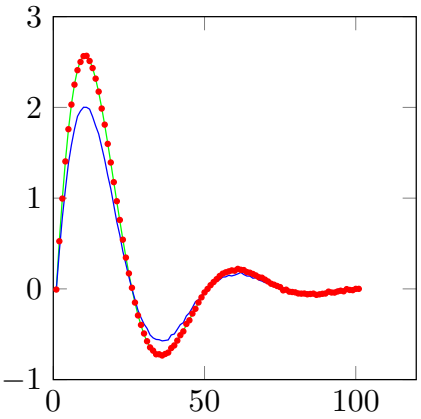


(a) Dipole waveform 1

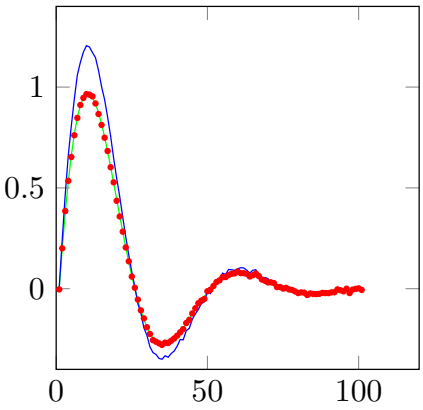
(b) Dipole waveform 2



(c) Dipole waveform 3

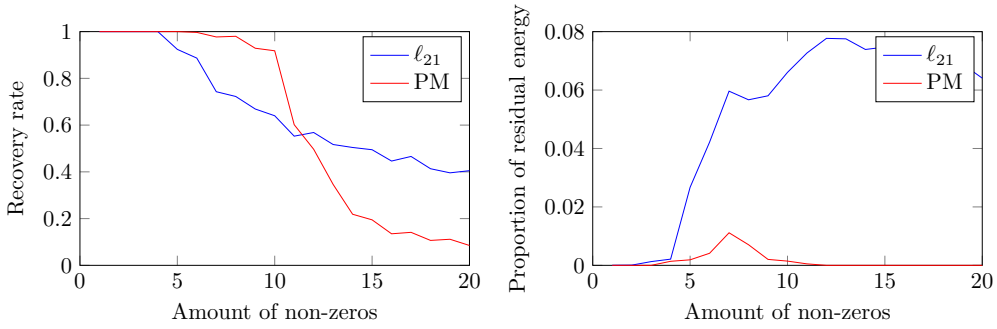


(d) Dipole waveform 4



(e) Dipole waveform 5

Figure 4.13: Estimated waveforms for five dipoles with SNR = 30dB. Green represents the ground truth, blue the ℓ_{21} mixed norm estimation and red the proposed method.



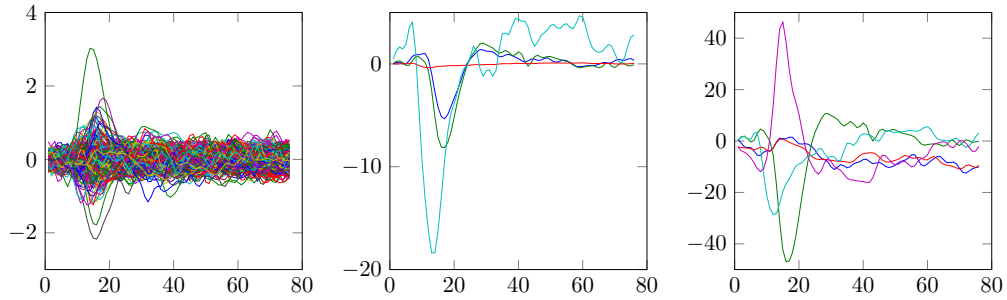
(a) Recovery rate as a function of P

(b) Proportion of residual energy as a function of P

Figure 4.14: Performance measures for multiple dipoles.

sinusoidal waves with frequencies varying between 5 and 20Hz. The activations were sampled at 200Hz and scaled in amplitude so that each of them produced the same energy in the measurements. Fifty different sets of localizations were used for the non-zero positions for each value of $P = 1, ..., 20$, resulting in a total of 1000 experiments. Noise was added to the measurements to obtain SNR = 30dB. For the ℓ_{21} mixed norm regularization, the regularization parameter was set according to the uncertainty principle.

For each simulation, the P non-zeros of the estimated activity associated with the highest energy in the measurements were considered as the estimated activity whereas the other elements were considered as residuals. We define the recovery rate as the proportion of non-zeros in the ground truth that are also present in the estimated activity. The average recovery rates of the proposed method and the ℓ_{21} mixed norm approach are presented in the first plot of Fig. 4.14 as a function of P. For $P \leq 10$ our algorithm detects the non-zeros with an accuracy higher than 90% which drops to 60.2% for $P = 11$ and 49.7% for $P = 12$. This drop of the recovery rate when a large number of non-zeros is present in the ground truth is well known, since the possible amount of non-zeros to recover correctly is limited by the span of the operator [37]. For comparison, the ℓ_{21} mixed norm regularization recovers up to $P = 5$ non-zeros with an accuracy higher than 90% and its recovery rate decreases slowly to reach 64% for $P = 10$. Note that our method performs better than the ℓ_{21} approach for $P \leq 11$. However, beyond this point, the performance of both methods is very poor preventing them from being used in practical applications.



(a) M/EEG whitened measurements (b) Estimated waveforms (ℓ_{21}-mixed norm) (c) Estimated waveforms (proposed method)

Figure 4.15: Measurements and estimated waveforms for the auditory evoked responses.

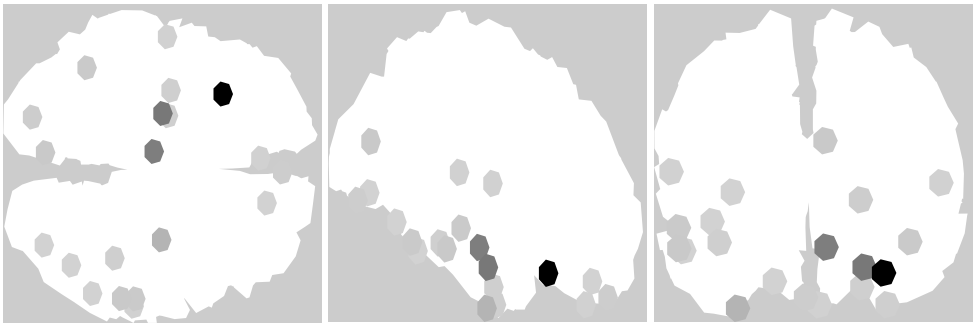
The recovery rate is calculated from the P main non-zero elements of the activity. However, it is also interesting to analyze how much activity is present in the residual non-zero elements. Thus, we define the proportion of residual energy as the amount of energy contained in the measurements generated by the residual non-zeros with respect to the total energy in the measurements. This residual energy serves as a measure of the sparsity of the solution. The second plot of Fig. 4.14 shows the value of the residual energy obtained for both algorithms as a function of P. The ℓ_{21} approach has up to 7.7% of the activity detected in residual non-zeros whereas our algorithm never exceeds 1.1% and always has lower residual activity than ℓ_{21}, confirming its good sparsity properties.

4.5.2 Real data

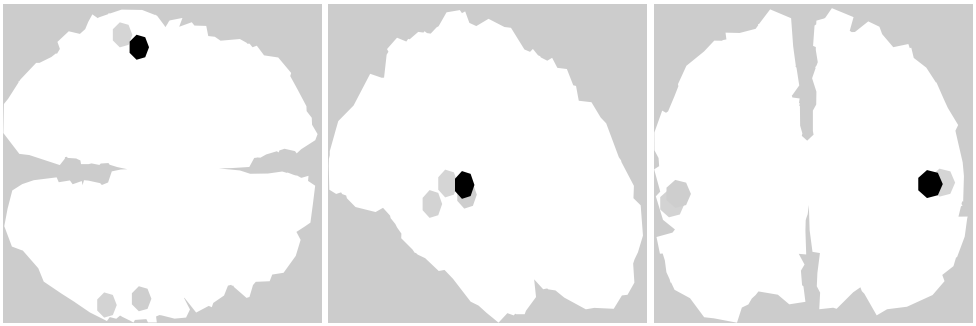
The same two real data sets used in Chapter 3 were considered. The first one corresponds to the auditory evoked responses to left ear pure tone stimulus while the second one consists of the evoked responses to facial stimulus. The results of the proposed method are compared with the weighted ℓ_{21} mixed norm [43] and the multiple sparse priors (MSP) method [51].

4.5.2.1 Auditory evoked responses

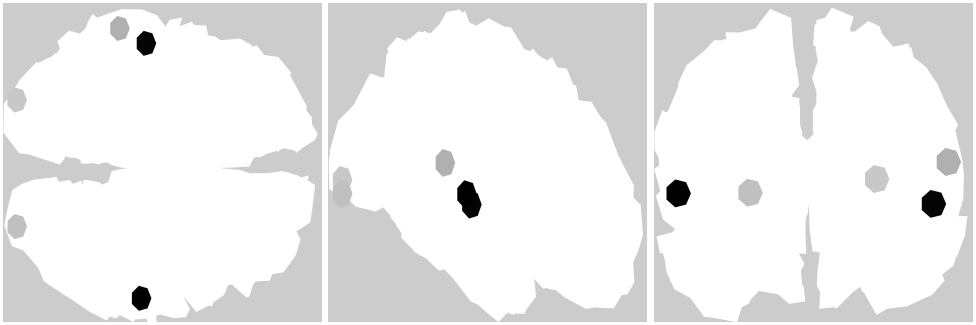
The default data set of the MNE software [79, 80] is used in this section. It consists of the evoked response to left-ear auditory pure-tone stimulus using a real-



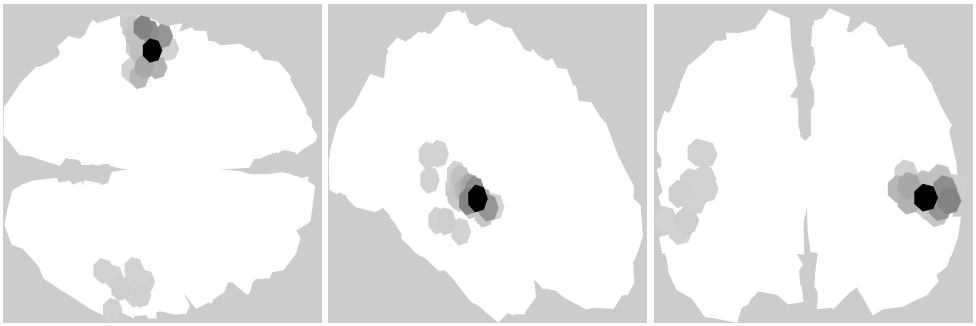
(a) Weighted ℓ_{21} mixed norm - Uncertainty principle of parameter a



(b) Weighted ℓ_{21} mixed norm - Manual adjustment of parameter a

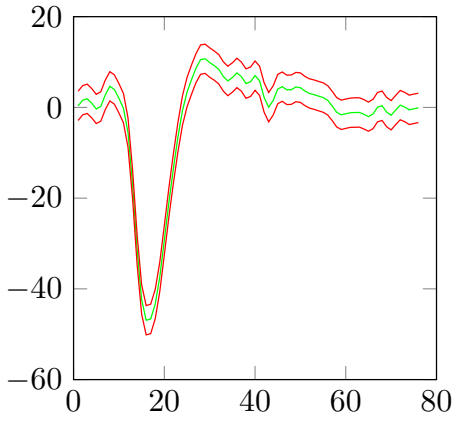


(c) Proposed method

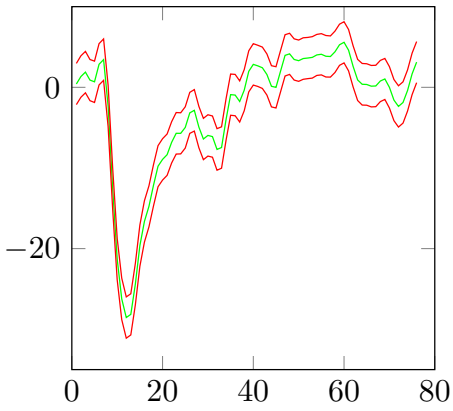


(d) MSP algorithm

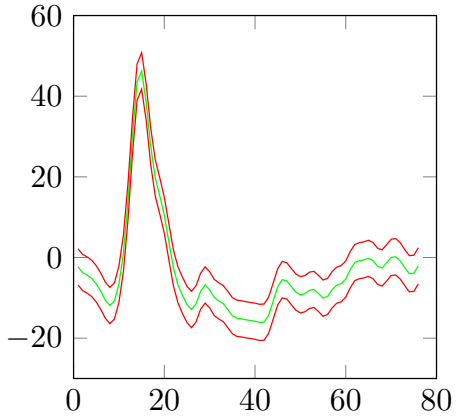
Figure 4.16: Estimated activity for the auditory evoked responses.



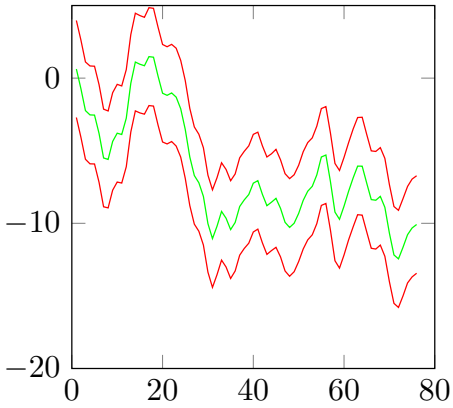
(a) Waveform 1 (Center right dipole)



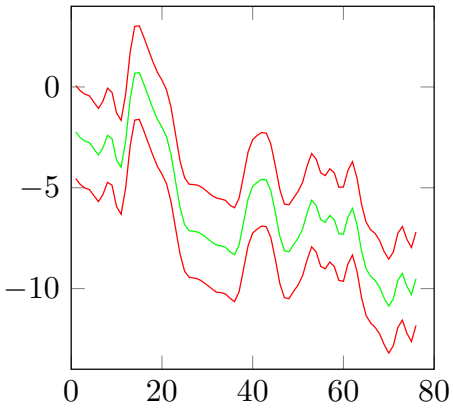
(b) Waveform 2 (Center left dipole 1, closer to the edge)



(c) Waveform 3 (Center left dipole 2)

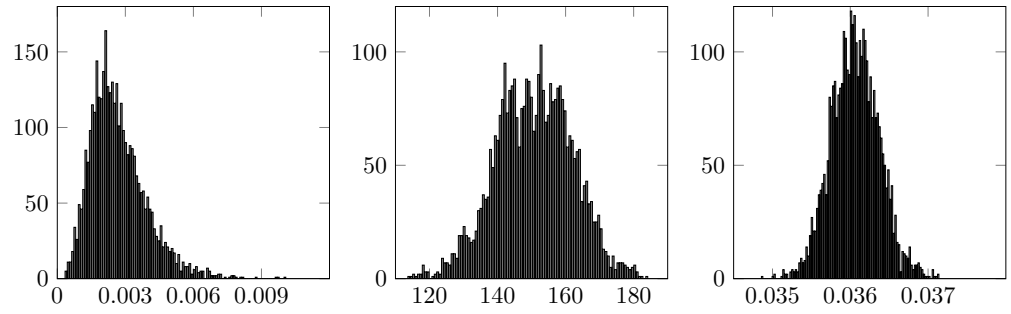


(d) Waveform 4 (Posterior right dipole)



(e) Waveform 5 (Posterior left dipole)

Figure 4.17: Estimated waveforms mean and boundaries $\mu \pm 2\sigma$ for the auditory evoked responses.



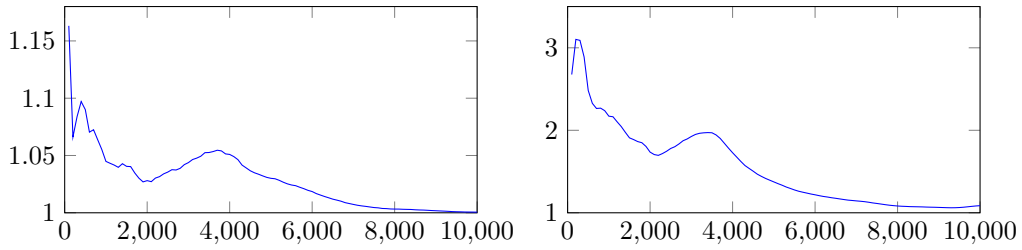
(a) Histogram of ω (b) Histogram of a (c) Histogram of σ_n^2

Figure 4.18: Hyperparameter histograms for the auditory evoked responses.

istic BEM (Boundary element method) head model sampled with 60 EEG electrodes and 306 MEG sensors. The head model contains 1.844 dipoles located on the cortex with orientations that are normal to the brain surface. Two channels that had technical artifacts were ignored. The data was sampled at 600Hz. The samples were low-pass filtered at 40Hz and downsampled to 150Hz. The noise covariance matrix was estimated from 200ms of the data preceding each stimulus and was used to whiten the measurements. Fifty-one epochs were averaged to calculate the measurements $\boldsymbol{Y}$. The activity of the source dipoles was estimated jointly for the period from 0ms to 500ms after the stimulus. From a clinical perspective it is expected to find the brain activity primarily focused on the auditory cortices that are located close to the ears in both hemispheres of the brain.

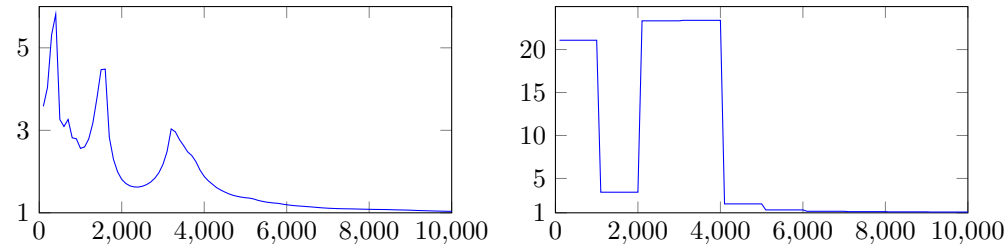
Since the measurements were whitened, it is possible to use the uncertainty principle to adjust the hyperparameter of the ℓ_{21} mixed norm. However, this provides an activity distributed all over the brain as shown in the first row of Fig. 4.16. By manually adjusting the hyperparameter to produce a sparser result, the ℓ_{21} mixed norm can obtain a solution that has activity in the auditory cortices as expected, shown in the second row of images. In contrast, our algorithm estimates its hyperparameters automatically and finds most of the activity in the auditory cortices without requiring any manual adjustment as displayed in the third row. On the other hand, the MSP method spreads the activity around the auditory cortices area since it groups the dipoles together in pre-defined regions.

The whitened measurements are displayed in Fig. 4.15 along with the activity estimation for both the ℓ_{21} approach (after manually adjusting the hyperpa-



(a) PSRF of ω

(b) PSRF of a



(c) PSRF of σ_n^2

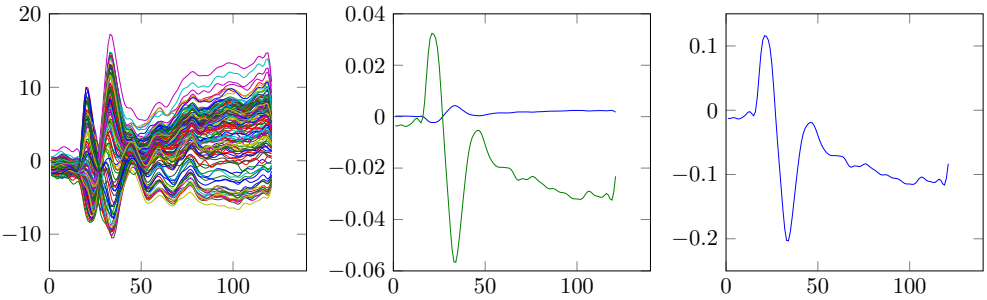
(d) Maximum PSRF of $\boldsymbol{X}$

Figure 4.19: PSRFs of sampled variables for the auditory evoked responses .

rameter) and our algorithm.

The five waveforms estimated by the proposed method with their confidence intervals of 2σ are shown in Fig. 4.17. Both results present sharp peaks in the activations of the auditory cortex dipoles between 80 and 100 milliseconds after the application of the stimulus. Note that the amplitudes estimated by the proposed method are much higher than the ones obtained with the ℓ_{21} approach due to the aforementioned amplitude underestimation of the latter.

The histograms of the hyperparameters of our algorithm are presented in Fig. 4.18 while their PSRF's are shown in Fig. 4.19. In the PSRF's we can see very abrupt changes in values for all the variables (most noticiably for $\boldsymbol{X}$) in the same iterations. These correspond to the iterations in which proposals were accepted by different chains and reflect the fact that the PSRF's can have very high values while the chains are in different modes. However, at the end of the simulation all chains converge to the same value of $\boldsymbol{z}$ which causes all the PSRF's to tend to 1, showing the correct convergence of all the chains to the same posterior distribution.



(a) M/EEG whitened measurements (b) ℓ_{21}-mixed norm estimated waveforms (c) Proposed method estimated waveforms

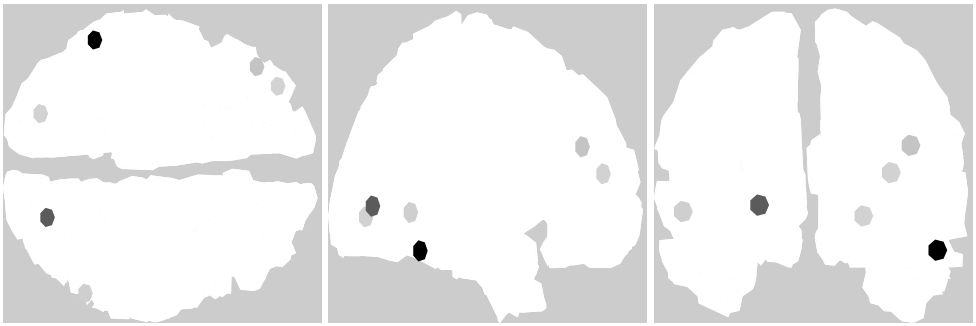
Figure 4.20: Measurements and estimated waveforms for the facial evoked responses.

4.5.2.2 FACIAL EVOKED RESPONSES

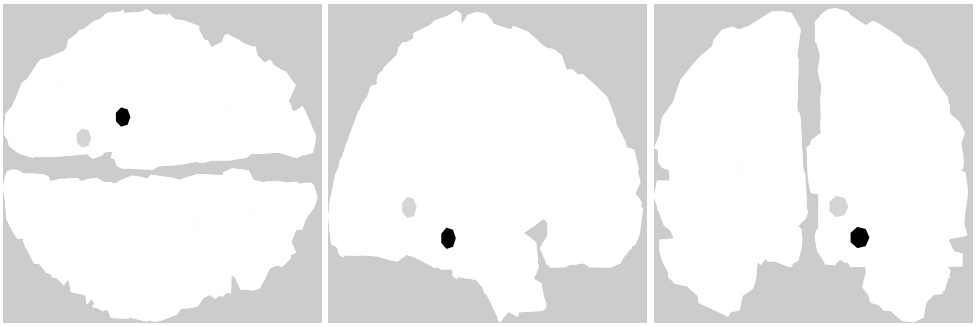
In a second experiment, we used data acquired from a face perception study where the subject was required to evaluate the symmetry of a mixed set of faces and scrambled faces, one of the default datasets of the SPM software[2]. Faces were presented during 600ms every 3600ms. The measurements were taken by the electrodes of a 128-channel ActiveTwo system that sampled at 2048 Hz. The measurements were downsampled to 200Hz and, after artifact rejection, 299 epochs corresponding to the non-scrambled faces were averaged and low-pass filtered to 40Hz. A T1 MRI scan was then downsampled to generate a 3004 dipole head model. The estimated activities are shown in Fig. 4.21. As in the previous case, we can see that the ℓ_{21} mixed norm response (with the regularization parameter adjusted according to the uncertainty principle) estimates the activity spread around the brain. In contrast, adjusting its regularization parameter manually results in a focal response concentrated in one of the fusiform regions in the temporal lobe associated with the facial recognition process [82] similar to the one obtained by our algorithm. The MSP algorithm spreads the activity over brain regions located more to the lateral and posterior parts of the brain, further away from the expected area.

Fig. 4.20 shows the M/EEG measurements and the estimated waveforms by the ℓ_{21} approach and the proposed method. As with the auditory evoked responses data, they differ in the scale due to the underestimation of the activity

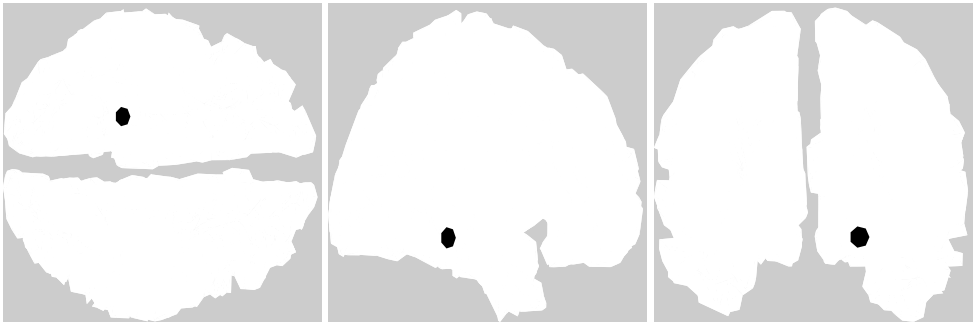
[2]The SPM software is freely avaiable at http://www.fil.ion.ucl.ac.uk/spm.



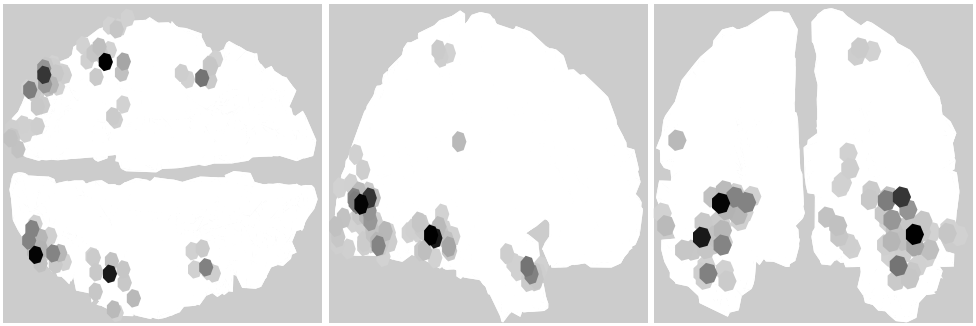
(a) Weighted ℓ_{21} mixed norm - Uncertainty principle for parameter a



(b) Weighted ℓ_{21} mixed norm - Manual adjustment of parameter a



(c) Proposed method



(d) MSP algorithm

Figure 4.21: Estimated activity for the facial evoked responses.

amplitude by the ℓ_{21} approach.

4.5.3 COMPUTATIONAL COST

The computational cost of the proposed method is much lower than the ℓ_0 + ℓ_1 method proposed in Chapter 3, mainly due to the introduction of the $\boldsymbol{\tau}$ latent variable that makes the sampling of $\boldsymbol{X}$ significantly faster. However, it is still higher than the ℓ_{21} mixed norm approach. The low SNR three-dipole experiment was processed in 6 seconds running in a modern Xeon CPU E3-1240 @ 3.4GHz processor (using a Matlab implementation with MEX files written in C) against 104ms for the ℓ_{21} mixed norm approach. However, it is interesting to mention that the ℓ_{21} norm approach requires running the algorithm multiple times to adjust the regularization parameter by cross-validation, contrary to the proposed algorithm that is able to estimate its own hyperparameters from one Markov chain produced by the partially collapsed Gibbs sampler.

4.6 CONCLUSION

In this chapter we generalized the Bayesian model of Chapter 3 to consider multiple time samples simultaneously. This generalized model promotes structured sparsity approximating the ℓ_{20} mixed norm in a Bayesian framework by using a multivariate Bernoulli Laplacian prior. A partially collapsed Gibbs sampler was investigated to sample from the target posterior distribution. We introduced multiple dipole shift proposals within each MCMC chain and exchange moves between different chains to improve the convergence speed of this sampler. Using the generated samples, the source activity was estimated jointly with the model hyperparameters in a fully unsupervised framework. The proposed method was compared with the ℓ_{21} mixed norm and the multiple sparse prior methods for a wide variety of situations including several multiple dipole synthetic activations and two different sets of real data. Our algorithm presented several advantages including better recovery of dipole locations and waveforms in low SNR conditions, the capacity of correctly detecting a higher amount of non-zeros, providing sparser solutions and avoiding underestimation of the activation amplitude. Finally, the possibility of providing several solutions with their corresponding probabilities is interesting.

The next chapter extends the proposed Bayesian model to consider the possibility of having uncertainty in the operator hyperparameters and making it

able to estimate those parameters jointly with the brain activity.

CHAPTER 5

JOINT M/EEG SOURCE LOCALIZATION AND SKULL CONDUCTIVITY ESTIMATION

Contents

5.1 Introduction

In the previous chapter we investigated a Bayesian model that is able to estimate focal brain activity promoting structured sparsity imitating an ℓ_{20} mixed norm regularization. However, in real M/EEG experiments the matrix $\boldsymbol{H}$ is not always completely known due to its dependency on several physical parameters. Of these, the skull conductivity has been shown to significantly affect the restoration process and depend significantly on the subject as mentioned in Chapter 2. In this chapter we generalize the model we proposed in Chapter 4 to estimate the skull conductivity jointly with the brain activity. This is done by approximating the dependency of the leadfield matrix with respect to the skull conductivity by a polynomial matrix.

5.2 Proposed Bayesian model

This chapter studies a new Bayesian model that solves the myope inverse problem presented in Section 2.3.4 for estimating the skull conductivity ρ jointly with the brain activity $\boldsymbol{X}$. The Bayesian model proposed to solve this problem is introduced in Section 5.2. Section 5.3 presents the partially collapsed Gibbs sampler that generates samples from the posterior distribution of the this model. The method used to model the leadfield matrix dependency with respect to the skull conductivity is presented in Section 5.4. Section 5.5 explains how to adapt the moves introduced in the previous chapter in order to make them relevant for the joint estimation of ρ and $\boldsymbol{X}$. Experimental results are finally presented in Section 5.6 in order to validate the proposed model and its estimation algorithm.

5.2.1 MATRIX NORMALIZATION

The depth biasing effect is a well known problem when solving the M/EEG source localization by minimizing a certain norm of the solution [25]. This effect is caused by the fact that each active dipole generates a signal of different amplitude. It is possible to compensate for this effect by either: (1) adding a weight vector $\boldsymbol{v}$ with elements $v_i = |\boldsymbol{h}^i|_2$ (being $||.||$ the euclidean norm and $\boldsymbol{h}^i$ the i-th column of matrix $\boldsymbol{H}$) that penalize the dipoles associated with large M/EEG measurements (as done in Chapters 3 and 4 or (2) normalizing the columns of the matrix so that they all have a unitary norm. When using the second option, the M/EEG source localization problem [2.1] can be written as

$$\boldsymbol{Y} = \bar{\boldsymbol{H}}(\rho)\bar{\boldsymbol{X}} + \boldsymbol{E} \tag{5.1}$$

where $\boldsymbol{Y} \in \mathbb{R}^{M\times T}$ contains the M/EEG measurements, $\boldsymbol{E} \in \mathbb{R}^{M\times T}$ is measurement noise, the columns of the normalized matrix are $\bar{\boldsymbol{h}}^i = \frac{\boldsymbol{h}^i}{|\boldsymbol{h}^i|_2}$ and the rows of the estimated activity are $\bar{\boldsymbol{x}}_i = |\boldsymbol{h}^i|_2 \boldsymbol{x}_i$ (being $\boldsymbol{x}_i$ the ith row of $\boldsymbol{X}$). Since our model has a variable leadfield matrix, using the first approach would cause the weight vector $\boldsymbol{v}$ to depend on the skull conductivity ρ which introduces unnecessary complexities. To avoid these, we use the second compensation model.

5.2.2 LIKELIHOOD

As in our previous models, we assume an independent white Gaussian noise with a constant variance σ_n^2, which leads to the following Gaussian likelihood

$$f(\boldsymbol{Y}|\bar{\boldsymbol{X}}, \sigma_n^2, \rho) = \prod_{t=1}^{T} \mathcal{N}\left(\boldsymbol{y}^t \middle| \bar{\boldsymbol{H}}(\rho)\bar{\boldsymbol{x}}^t, \sigma_n^2 I_M\right) \tag{5.2}$$

where I_M is the identity matrix of size M.

5.2.3 PRIOR AND HYPERPRIOR DISTRIBUTIONS

The dependencies between the different model parameters and hyperparameters is shown in Fig. 5.1. The priors used for each of them (apart from the skull conductivity ρ) are the same that were used in the model presented in the previous chapter with $v_i = 1$ (due to the leadfield matrix normalization described in Section 5.2.1) and are summarized in Table 5.1. In this table we can see that each row $\boldsymbol{x}_i$ is either jointly zero or non-zero, which is indicated by the discrete

variable z_i. ω is a variable between 0 and 1 that indicates the probability of having $z_i = 1$ whereas a controls the amplitude of the non-zero rows of $\boldsymbol{X}$. Note that $\boldsymbol{\tau}^2$ is a latent variable that is only introduced to accelerate the convergence of the algorithm. Finally, the noise variance σ_n^2 is assigned a Jeffrey's prior reflecting the absence of knowledge about this parameter. The prior used for the skull conductivity is detailed in the rest of this section.

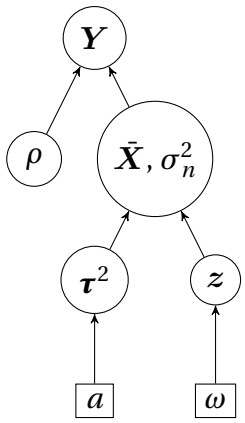


Figure 5.1: Directed acyclic graph for the proposed Bayesian model.

z_i	$\mathscr{B}\left(z_i \middle\vert \omega\right)$
τ_i^2	$\mathscr{G}\left(\tau_i^2 \middle\vert \frac{T+1}{2}, \frac{a}{2}\right)$
$\bar{\boldsymbol{x}}_i$	$\delta(\bar{\boldsymbol{x}}_i)$ if $z_i = 0$ $\mathcal{N}\left(0, \sigma_n^2 \tau_i^2 \mathbb{I}_T\right)$ if $z_i = 1$
σ_n^2	$\frac{1}{\sigma_n^2} 1_{\mathbb{R}^+}(\sigma_n^2)$
a	$\mathscr{G}\left(a \middle\vert \alpha, \beta\right)$ with $\alpha = \beta = 1$
ω	$\mathscr{U}\left(\omega \middle\vert 0, 1\right)$

Table 5.1: Prior distributions $f(z_i|\omega)$, $f(\tau_i^2|a)$, $f(\bar{\boldsymbol{x}}_i|z_i, \tau_i^2, \sigma_n^2)$, $f(\sigma_n^2)$, $f(a)$ and $f(\omega)$.

In order to keep the model simple we propose to assign a non-informative uniform prior for the skull conductivity

$$f(\rho) = \mathcal{U}\left(\rho \middle| \rho_{\min}, \rho_{\max}\right). \tag{5.3}$$

To select the range of this uniform distribution we consider an interval containing the scalp-to-skull conductivity ratios reported in the literature [59, 90], i.e, defined by $r_{\min} = 10$ and $r_{\max} = 100$. Considering the scalp conductivity to be $330\frac{mS}{m}$, this leads to $\rho_{\min} = 3.3\frac{mS}{m}$ and $\rho_{\max} = 33\frac{mS}{m}$.

5.2.4 Posterior distribution

The corresponding posterior distribution is defined as follows

$$\begin{aligned} f(\boldsymbol{Y}, \sigma_n^2, \bar{\boldsymbol{X}}, \boldsymbol{z}, a, \boldsymbol{\tau}^2, \omega, \rho) \propto & f(\boldsymbol{Y}|\bar{\boldsymbol{X}}, \sigma_n^2, \rho) f(\bar{\boldsymbol{X}}|\boldsymbol{\tau}^2, \boldsymbol{z}, \sigma_n^2) f(\boldsymbol{z}|\omega) \\ & f(\boldsymbol{\tau}^2|a) f(\sigma_n^2) f(a) f(\omega) f(\rho) \end{aligned} \tag{5.4}$$

5.3 Partially collapsed Gibbs sampler

Unfortunately, the posterior distribution [A.40] is intractable and does not allow Bayesian estimators of the different parameters and hyperparameters to be expressed in closed-form. As a consequence, we propose to draw samples from [A.40] and use them to estimate the brain activity jointly with the model hyperparameters. More precisely, we investigate a partially collapsed Gibbs sampler that samples the variables z_i and $\boldsymbol{x}_i$ jointly in order to exploit the strong correlation between these two variables. The proposed method samples the different variables according to their conditional distributions as shown in Algorithm 5.1. Note that the only difference between this sampler and the sampler developed in Section 4.3 is the last line corresponding to the sampling of the skull conductivity.

The corresponding conditional distributions of all the model parameters and hyperparameters apart from ρ are the same as in the ones corresponding to the model presented in Chapter 4, as explained in Appendix A, and can be found in Table 5.2.

Algorithm 5.1 Partially Collapsed Gibbs sampler.

Initialize all the parameters.
repeat
  for $i = 1$ to N **do**
    Sample τ_i^2 from $f(\tau_i^2|\bar{\boldsymbol{x}}_i, \sigma_n^2, a, z_i)$
    Sample z_i from $f(z_i|\boldsymbol{Y}, \bar{\boldsymbol{X}}_{-i}, \sigma_n^2, \tau_i^2, \omega, \rho)$
    Sample $\bar{\boldsymbol{x}}_i$ from $f(\bar{\boldsymbol{x}}_i|z_i, \boldsymbol{Y}, \bar{\boldsymbol{X}}_{-i}, \sigma_n^2, \tau_i^2, \rho)$
  end for
  Sample a from $f(a|\boldsymbol{\tau}^2)$
  Sample σ_n^2 from $f(\sigma_n^2|\boldsymbol{Y}, \bar{\boldsymbol{X}}, \boldsymbol{\tau}^2, \boldsymbol{z}, \rho)$
  Sample ω from $f(\omega|\boldsymbol{z})$
  Sample ρ from $f(\rho|\bar{\boldsymbol{X}}, \boldsymbol{Y}, \sigma_n^2)$
until convergence

5.3.1 Conditional distributions

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

Table 5.2: Conditional distributions $f(\tau_i^2|\boldsymbol{x}_i, \sigma_n^2, a, z_i)$, $f(z_i|\boldsymbol{Y}, \boldsymbol{X}_{-i}, \sigma_n^2, \tau_i^2, \omega, \rho)$, $f(\boldsymbol{x}_i|z_i, \boldsymbol{Y}, \boldsymbol{X}_{-i}, \sigma_n^2, \tau_i^2, \rho)$, $f(a|\boldsymbol{\tau}^2)$, $f(\sigma_n^2|\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\tau}^2, \boldsymbol{z}, \rho)$ and $f(\omega|\boldsymbol{z})$.

with

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

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

The conditional distribution of the skull conductivity can be written

$$f(\rho|\bar{\boldsymbol{X}}, \boldsymbol{Y}, \sigma_n^2) \propto \exp\left(- \frac{||\bar{\boldsymbol{H}}(\rho)\bar{\boldsymbol{X}} - \boldsymbol{Y}||^2}{2\sigma_n^2} \right) 1_{[\rho_{\min}, \rho_{\max}]}(\rho). \quad [5.5]$$

For arbitrary functions $\bar{\boldsymbol{H}}(\rho)$, [5.5] does not belong to a common family of distributions. The following section explains how to efficiently model the dependency of the matrix $\bar{\boldsymbol{H}}(\rho)$ in order to be able to sample from [5.5].

5.4 Leadfield matrix Approximation

5.4.1 Dependency analysis

Shell models can be used to derive closed form expressions for the leadfield matrix as a function of the conductivity that will denoted $\bar{\boldsymbol{H}}(\rho)$. However these models are quite complex [18] and would make the sampling from $f(\rho|\bar{\boldsymbol{X}}, \boldsymbol{Y}, \sigma_n^2)$ considerably difficult. In contrast, realistic head models can be calculated numerically for particular values of ρ but do not provide a closed-form expression for $\bar{\boldsymbol{H}}(\rho)$. In order to illustrate how the value of ρ affects the matrix, a four-shell 200-dipole head model with 41 electrodes was calculated for different values of ρ. Eight elements of the matrix $h_{i,j}(\rho)$ (chosen randomly) are displayed in Fig. 5.2 as a function of ρ. In order to have a simple expression of this dependency, in Sengül and Baysal [64] proposed to replace $\bar{\boldsymbol{H}}(\rho_k)$ by its linearization around the current value ρ_k. This method was shown to provide good results but requires evaluating the exact value of $\bar{\boldsymbol{H}}(\rho)$ at every iteration, which slows down the algorithm considerably. Since the variation of the matrix elements with respect to ρ is smooth and monotonic, we propose to approximate $\bar{\boldsymbol{H}}(\rho)$ using a simple mathematical expression as in [64]. However, the approximation is computed on the whole range $\rho_{\min} < \rho < \rho_{\max}$ with a polynomial matrix of small degree denoted by $\hat{\boldsymbol{H}}_L(\rho)$. This allows us to have a simple closed form

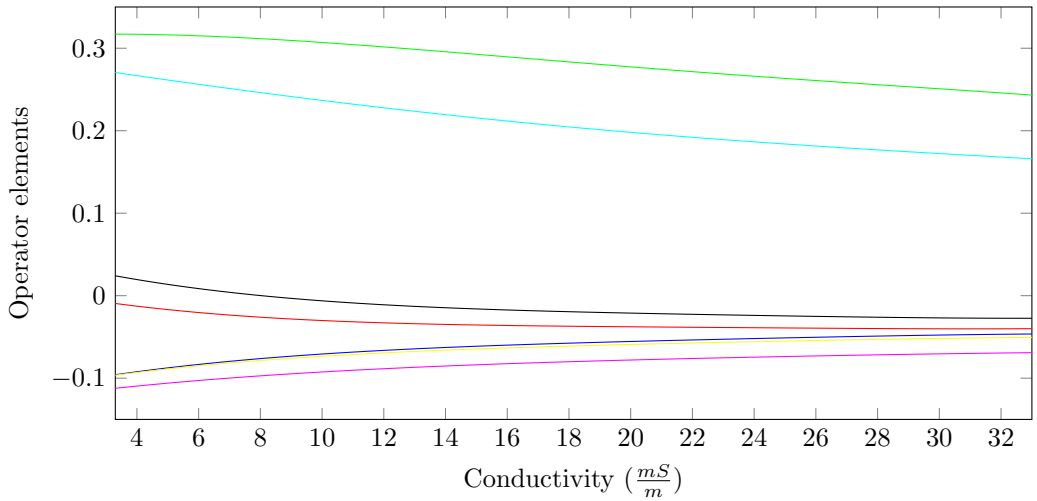


Figure 5.2: Variations of the matrix elements with respect to ρ.

expression of $\hat{\boldsymbol{H}}_L(\rho)$ for both shell and realistic head models. In addition, our method only requires calculating the exact value of $\bar{\boldsymbol{H}}(\rho)$ to construct the polynomial matrix $\hat{\boldsymbol{H}}_L(\rho)$ offline. Afterwards, the polynomial matrix is evaluated for the new value of ρ at every iteration, which accelerates the iteration speed considerably (since evaluating the matrix is much faster than recalculating the exact value of $\bar{\boldsymbol{H}}(\rho)$).

5.4.2 Polynomial approximation

Each element of $\hat{\boldsymbol{H}}_L(\rho)$ is constructed as a polynomial function of ρ

$$\hat{h}_{i,j}(\rho) = \sum_{l=0}^{L} c_{i,j,l}\rho^l. \qquad [5.6]$$

We propose to calculate the coefficients $c_{i,j,l}$ using least-squares fitting. This method consists in minimizing the following mean square error between $\bar{\boldsymbol{H}}(\rho)$ and $\hat{\boldsymbol{H}}_L(\rho)$ using K values of ρ

$$MSE(\rho) = \sum_{k=0}^{K-1} \left(\bar{h}_{i,j}(\rho_k) - \sum_{l=0}^{L} c_{i,j,l}\rho_k^l\right)^2 \qquad [5.7]$$

Note that the exact value of $\bar{\boldsymbol{H}}(\rho)$ is calculated for $\rho_k = \rho_{\min} + \frac{k(\rho_{\max}-\rho_{\min})}{K-1}$ with $0 \leq k \leq K-1$ by either using the shell model expression [19] or by eval-

uating numerically the head model from the patient's MRI [20]. After this we calculate the polynomial coefficients $c_{i,j,l}$ that minimize [5.7] using the classical least squares estimator

$$\boldsymbol{c}_{i,j} = \Big[\sum_{k=0}^{K-1} \boldsymbol{\psi}_k \boldsymbol{\psi}_k{}^T\Big]^{-1} \sum_{k=0}^{K-1} \bar{h}_{i,j}(\rho_k)\boldsymbol{\psi}_k \qquad [5.8]$$

with $\boldsymbol{\psi}_k = [\rho_k^0, ..., \rho_k^L]^T$ and $\boldsymbol{c}_{i,j} = [c_{i,j,0}, ..., c_{i,j,L}]^T$.

The next section will discuss how to sample the skull conductivity after the coeficients $c_{i,j,l}$ have been calculated.

5.4.3 SKULL CONDUCTIVITY SAMPLING

Approximating the relationship between the leadfield matrix and the skull conductivity with a polynomial matrix allows us to have the following simple closed form expression for the conditional distribution of the skull conductivity $f(\rho|\bar{\boldsymbol{X}}, \boldsymbol{Y}, \sigma_n^2)$

$$f(\rho|\bar{\boldsymbol{X}}, \boldsymbol{Y}, \sigma_n^2) \propto \exp\Big(-g(\rho)\Big) 1_{[\rho_{\min}, \rho_{\max}]}(\rho) \qquad [5.9]$$

where $g(\rho) = \frac{||\hat{\boldsymbol{H}}_{L,K}(\rho)\bar{\boldsymbol{X}} - \boldsymbol{Y}||^2}{2\sigma_n^2}$ is a polynomial of order $2L$.

Since it is not easy to sample from [5.9] directly, we propose to adopt a random-walk Metropolis-Hastings (MH) move. More precisely, this move consists in proposing at each iteration of the Gibbs sampler a new sample $\rho_{\text{prop}} = \rho_{\text{old}} + \epsilon$ and accepting it with probability

$$P_a = \begin{cases} \min\left(\frac{\exp(-g(\rho_{prop}))}{\exp(-g(\rho_{old}))}, 1\right) & \text{if } \rho_{\min} < \rho_{\text{prop}} < \rho_{\max} \\ 0 & \text{otherwise.} \end{cases} \qquad [5.10]$$

where ϵ is a variable defining the random walk whose distribution needs to be specified.

We propose to use a zero-mean Gaussian distribution for ϵ, i.e., $f(\epsilon) = \mathcal{N}(0, \sigma_\epsilon^2)$. Moreover, we propose to adjust the variance of the random walk σ_ϵ^2 empirically in order to obtain an appropriate acceptance rate, as recommended in [69]. Based on the experiments performed in Section 5.6, we have shown that $\sigma_\epsilon = \frac{\rho_{\max} - \rho_{\min}}{100}$ is a reasonable choice for the random walk variance.

5.5 IMPROVING THE CONVERGENCE OF THE GIBBS SAMPLER

5.5.1 MULTIPLE DIPOLE SHIFT PROPOSALS

Algorithm 5.2 Multiple dipole shift proposal.

$\hat{\boldsymbol{z}} = \boldsymbol{z}$
repeat D times
  Set ind_{old} to be the index of a random non-zero of $\boldsymbol{z}$
  Set $\boldsymbol{p} = [\text{ind}_{\text{old}}, \text{neigh}_{\gamma,\rho}(\text{ind}_{\text{old}})]$
  Set ind_{new} to be a random element of $\boldsymbol{p}$
  Set $\hat{z}_{\text{ind}_{\text{old}}} = 0$ and $\hat{z}_{\text{ind}_{\text{new}}} = 1$
end
Sample $\hat{\rho}$ from $f(\hat{\rho}|\boldsymbol{Y}, \sigma_n^2, \boldsymbol{\tau}^2, \hat{\boldsymbol{z}})$.
Sample $\hat{\boldsymbol{X}}$ from $f(\hat{\boldsymbol{X}}|\hat{\boldsymbol{z}}, \boldsymbol{Y}, \sigma_n^2, \boldsymbol{\tau}^2, \hat{\rho})$.
Sample $\hat{\boldsymbol{\tau}}^2$ from $f(\hat{\boldsymbol{\tau}}^2|\hat{\boldsymbol{X}}, \sigma_n^2, a, \hat{\boldsymbol{z}})$.
Set $\{\boldsymbol{z}, \boldsymbol{\tau}^2, \rho\} = \{\hat{\boldsymbol{z}}, \hat{\boldsymbol{\tau}}^2, \hat{\rho}\}$ with probability $\min\left(\frac{f(\hat{\boldsymbol{z}}, \hat{\boldsymbol{\tau}}^2, \hat{\rho}|\boldsymbol{Y}, a, \sigma_n^2, \omega)}{f(\boldsymbol{z}, \boldsymbol{\tau}^2, \rho|\boldsymbol{Y}, a, \sigma_n^2, \omega)}, 1\right)$
Else, do not change the values of $\{\boldsymbol{z}, \tau^2, \rho\}$
Resample $\boldsymbol{X}$ if the proposed move has been accepted

As explained in Chapter 4, the Gibbs sampler presented above tends to gets stuck around local maxima of the posterior distribution [A.40]. In particular, the MAP estimator of $\boldsymbol{z}$ does not usually indicate the correct positions of the active dipoles in a reasonable amount of time. To solve this problem we introduced in Chapter 4 multiple dipole shift proposals allowing D random non-zeros to be moved to neighboring positions. These moves were accepted or rejected using the classical MH acceptance rate. In order to build more efficient proposals adapted to the case of an unknown skull conductivity, we expand the multiple dipole shift moves to include the skull conductivity ρ. We do this by adding a step to sample the skull conductivity ρ marginally to $\boldsymbol{X}$. The final algorithm is illustrated in Algorithm 5.2 where we keep the neighborhood definition of Chapter 4

$$\text{neigh}_{\gamma,\rho}(i) \triangleq \{j \neq i \;\mid\; |\text{corr}(\hat{\boldsymbol{h}}_{L,K}(\rho)^{i}, \hat{\boldsymbol{h}}(\rho)^{j})| \geq \gamma\} \qquad [5.11]$$

where $\gamma \in [0,1]$ tunes the neighborhood size ($\gamma = 0$ corresponds to a neighborhood containing all the dipoles and $\gamma = 1$ corresponds to an empty neighbor-

hood). The values of D and γ were adjusted by cross validation to $D = 2$ and $\gamma = 0.8$.

To accept or reject the moves, we need to evaluate the conditional distribution $f(\boldsymbol{z}, \boldsymbol{\tau}^2, \rho | \boldsymbol{Y}, a, \sigma_n^2, \omega)$ up to a multiplicative constant. Considering only the values of the subset of $\boldsymbol{z}$ that change between the current value and the proposal (denoted by $\boldsymbol{z_r}$), the following result can be obtained

$$f(\boldsymbol{z}_r, \boldsymbol{\tau}_r^2, \rho | \boldsymbol{Y}, a, \sigma_n^2, \omega) \propto (1-\omega)^{C_0} \omega^{C_1} (\sigma_n^2)^{-\frac{TC_1}{2}} |\Sigma|^{\frac{T}{2}} \prod_{i \in \boldsymbol{I}_1} (\tau_i^2)^{-\frac{T}{2}} \exp\left(-\frac{\sum_{t=1}^{T} Q^t}{2}\right) \prod_{i=1}^{N} \mathcal{G}\left(\tau_i^2 \middle| \frac{T+1}{2}, \frac{a}{2}\right) \quad [5.12]$$

where $\boldsymbol{I}_j = \{i | z_{r_i} = j\}$, C_j is the cardinal of the set $\boldsymbol{I}_j$ (for $j = \{0, 1\}$) and

$$\begin{aligned}
\boldsymbol{\Sigma}^{-1} &= \frac{1}{\sigma_n^2} \left[\boldsymbol{H}^{\boldsymbol{I}_1}(\rho)^T \boldsymbol{H}^{\boldsymbol{I}_1}(\rho) + \text{diag}\left(\frac{1}{\tau_{\boldsymbol{r}}^2}\right) \right] \\
\boldsymbol{\mu}^t &= -\frac{\boldsymbol{\Sigma} \boldsymbol{H}^{\boldsymbol{I}_1}(\rho)^T [\boldsymbol{H}^{-\boldsymbol{r}}(\rho) \boldsymbol{x}_{-\boldsymbol{r}}^t - \boldsymbol{y}^t]}{\sigma_n^2} \\
Qb^t &= \frac{[\boldsymbol{H}^{-\boldsymbol{r}}(\rho) \boldsymbol{x}_{-\boldsymbol{r}}^t - \boldsymbol{y}^t]^T [\boldsymbol{H}^{-\boldsymbol{r}}(\rho) \boldsymbol{x}_{-\boldsymbol{r}}^t - \boldsymbol{y}^t]}{\sigma_n^2} - \boldsymbol{\mu}^{t\,T} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}^t
\end{aligned}$$

5.5.2 INTER-CHAIN PROPOSALS

As explained in Chapter 4 when running multiple MCMC chains in parallel it is possible for the different chains to get stuck around different values of $\boldsymbol{z}$. As a consequence, inter-chain proposals were considered to improve the sampler convergence. These proposals were exchanging active dipoles of different chains, with a probability defined according to the MH acceptance rate.

We now expand the inter-chain proposals to include the conductivity ρ to improve the sampling efficiency. More precisely, at each iteration, an exchange between the active dipoles of random pairs of chains is proposed with probability π (the value of π was fixed to 10^{-3} by cross validation) and this exchange is accepted with the MH probability as shown in Algorithm 5.3.

Algorithm 5.3 Inter-chain proposals.

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

5.6 Experimental Results

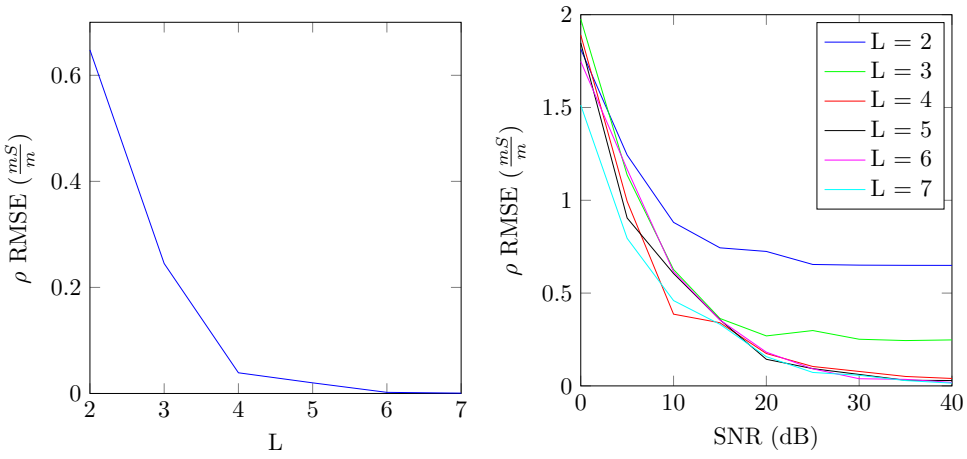
5.6.1 Synthetic data

Synthetic data is first considered using a 200-dipole Stok four-sphere head model [19] with 41 electrodes. The dipoles were uniformly distributed in the brain cortex oriented orthogonally to the brain surface. Two different kinds of activations were investigated: (1) single dipole activations with low SNR and (2) multiple dipole activations with high SNR. Before applying our method to the data we needed to construct the polynomial matrix $\hat{\boldsymbol{H}}_{L,K}(\rho)$, which is the objective of the next section.

5.6.1.1 Polynomial matrix construction

In order to construct $\hat{\boldsymbol{H}}_{L,K}(\rho)$ we have to select appropriate values of L and K (as defined in Section 5.4.2). Increasing the values of L and K clearly improves the quality of the approximation of $\hat{\boldsymbol{H}}_{L,K}$ at the price of a higher computational complexity. To choose appropriate values of these parameters, it is interesting to analyze their effect on the estimation of ρ when all the other parameters ($\boldsymbol{X}$ and σ_n^2) are known. To do this we used several ground truth values of $\boldsymbol{X}$ and σ_n^2 and assumed that the algorithm would ideally converge to the value $\hat{\rho}$ that minimizes the polynomial of order $2L$ defined as $g(\rho) = \frac{||\hat{\boldsymbol{H}}_{L,K}(\rho)\bar{\boldsymbol{X}} - \boldsymbol{Y}||^2}{2\sigma_n^2}$.

Using the head model described in the previous subsection, we used $K = 100$ and generated values of L in the range $2 \leq L \leq 7$. We also considered ten different ground truth values of ρ_{gt}, 200 different values of $\bar{\boldsymbol{X}} \in \mathbb{R}^{200\times100}$ (each of them having one of the 200 dipoles with a constant activity during the 100 time samples) and 9 values of SNR (from 0dB to 40dB in steps of 5dB) resulting



(a) Noiseless simulations

(b) Noisy simulations

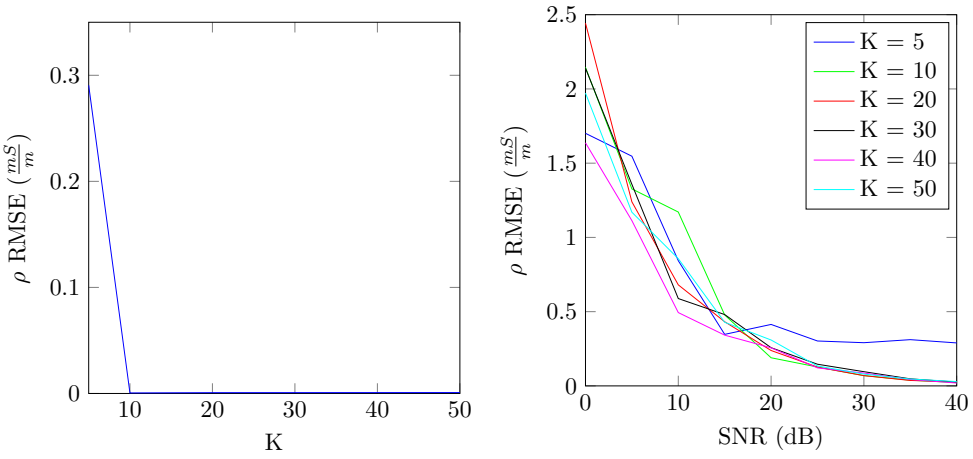
Figure 5.3: RMSE of ρ VS L (K = 100).

in a total of 108.000 experiments. Since the ground truth $\bar{\boldsymbol{X}}$ is used for the estimation of $\hat{\rho}$, the only factors that explain the difference between $\hat{\rho}$ and ρ_{gt} are the presence of noise and the approximation of $\bar{\boldsymbol{H}}(\rho)$ by $\hat{\boldsymbol{H}}_{L,K}(\rho)$, which allows us to illustrate the effect of L for different values of SNR. We define the root mean square error of the estimation of ρ for a particular value of L and SNR as

$$RMSE(L, SNR) = \sqrt{\frac{1}{200 \times 10} \sum_{i=1}^{200\times 10} (\widehat{\rho}_i - \rho_i)^2} \qquad [5.13]$$

ρ_i being the ground truth of ρ for the ith realization obtained for the specified values of L and SNR and $\hat{\rho}_i$ the corresponding estimated value of ρ.

Fig. 5.3 illustrates the RMSE of the estimation of ρ as a function of SNR for each value of L and for noiseless simulations. We can see that in noiseless situations, the error seems to tend asymptotically to 0 as the value of L increases, as expected. However, when the measurements are noisy the minimum value of RMSE is limited by the amount of noise. For instance, if the measurements correspond to SNR = 10dB , the estimation error only decreases until $L = 3$. This shows that for high SNR it makes sense to choose a high value of L. However, for common values of SNR (lower than 20dB) choosing high values of L does not improve the quality of the estimation of ρ. Be-



(a) Noiseless simulations.

(b) Noisy simulations.

Figure 5.4: RMSE of ρ VS K (L = 7).

cause of this, for a given SNR, we choose the smallest value of L such that $|RMSE(SNR, L+1) - RMSE(SNR, L)| < \epsilon$.

To analyze the effect of K, the same experiment was performed by varying $K \in \{5, 10, 20, 30, 40\}$ and fixing $L = 5$ for $K = 5$ and $L = 7$ for the higher values of K. Fig. 5.4 shows the RMSE of ρ for noisy and noiseless simulations. Both figures show that there is a significant improved approximation when K is increased from 5 to 10 if the SNR is high enough. However, choosing values of K higher than 10 does not improve the estimation of the skull conductivity ρ. Since we will be working with values of SNR between 10 and 30dB we have decided to use $L = 4$ and $K = 10$ in the remainder of the chapter.

5.6.1.2 SINGLE DIPOLE ACTIVATIONS

In the first kind of simulations, a random single dipole was assigned a damped sine activation of 5Hz. The activation was sampled at 200Hz and multiplied with the leadfield matrix with a chosen ground truth value of ρ denoted as ρ_{gt}. Gaussian white noise was added to the measurements to have a signal to noise ratio SNR = 10dB. The proposed estimation method is compared with two other methods: (1) a variation of the proposed method that uses a fixed value of the conductivity, more precisely $\rho_{fix} = \frac{\rho_{max}+\rho_{min}}{2} = 18.15\frac{mS}{m}$ in order to illustrate the advantages of estimating ρ (called default-ρ model) and (2) the optimization

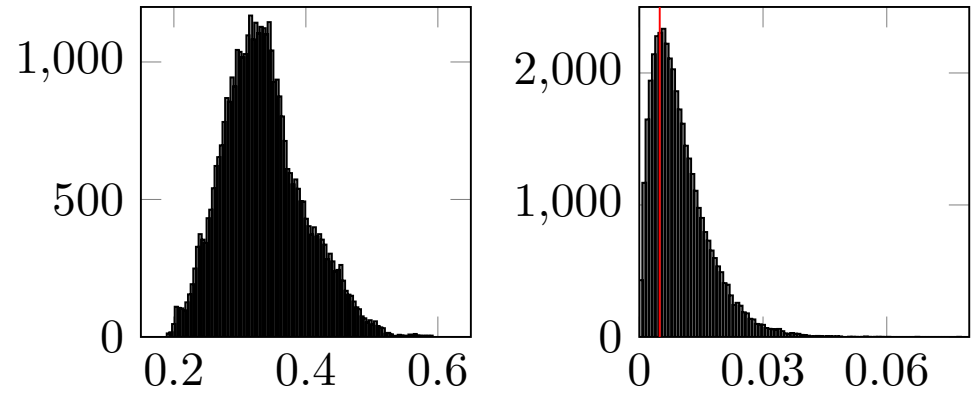
Method	Pos. error	$\frac{\lvert\hat{X}-X_{\text{gt}}\rvert_F^2}{\lvert X_{\text{gt}}\rvert_F^2}$	$\frac{\hat{\sigma_n}^2-\sigma_n^2}{\sigma_n^2}$	$\hat{\rho}-\rho_{\text{gt}}$	$\frac{\lvert\boldsymbol{H}(\rho_{\text{gt}})-\boldsymbol{H}(\hat{\rho})\rvert_F^2}{\lvert\boldsymbol{H}(\rho_{\text{gt}})\rvert_F^2}$
Default-ρ	0	2.53×10^{-3}	26.99×10^{-3}	4.46×10^{-3}	2.51×10^{-3}
PM	0	2.47×10^{-3}	6.21×10^{-3}	-43.89×10^{-6}	1.21×10^{-6}
Vallaghé	0	128×10^{-6}	N/A	-52.83×10^{-6}	1.15×10^{-6}

Table 5.3: Estimation errors for the different parameters (Simulation #1)

method studied in [65] that is able to estimate ρ and the brain activity jointly if there is only one active dipole. Finally, 8 parallel MCMC chains were used for both the proposed model and the default-ρ model. We illustrate two different cases. The first case corresponds to a conductivity ρ_{gt} close to ρ_{fix} (simulation #1) and the second case is characterized by a higher difference between ρ_{gt} and ρ_{fix} (simulation #2). More precisely, $\rho_{\text{gt}} = 13.68\frac{mS}{m}$ for simulation #1 and $\rho_{\text{gt}} = 3.59\frac{mS}{m}$ for simulation #2.

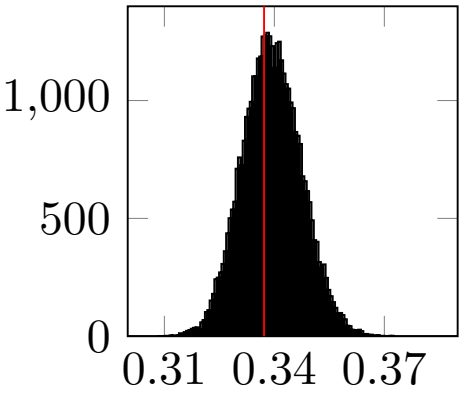
For simulation #1, all methods manage to find the active dipole in its exact location. The histograms of the simulated parameters for the two Bayesian models corresponding to a fixed value and an unknown value of ρ are shown in Figs. 5.5 and 5.6. First we comment the results obtained with the proposed model and its default-ρ variation. In order to understand why the model with a fixed value of ρ is able to correctly recover the activation despite the use of a wrong value of ρ, it is important to note on the different histograms that the proposed model is able to estimate all variables correctly. Note also that the default-ρ model seems to overestimate the noise variance σ_n^2. Thus, when the error in ρ (and thus, in the matrix) is small, the default-ρ model is able to mitigate the effect of using a wrong leadfield matrix by considering that there is additional noise in the system.

Vallaghé's optimization method also estimates the dipole location correctly for simulation #1. After running simulation #1 for 20 Monte Carlo runs with different noise realizations, the averaged MMSE estimate of the skull conductivity was $\hat{\rho} = 13.63\frac{mS}{m}$ (which is very close to the actual ground truth value $13.69\frac{mS}{m}$). An interesting property of the proposed estimation method is that it is able to estimate the posterior distribution of the skull conductivity (as shown in Fig. 5.5). This distribution can be used to determine the MMSE estimator of ρ (the mean of the posterior distribution, that is $13.64\frac{mS}{m}$) as well as uncertainties regarding this estimator. For instance the standard deviation of the MMSE estimator is $0.46\frac{mS}{m}$. Since the proposed method is not restricted to a point-estimate as Vallaghé"s optimization method, it is not necessary to run different

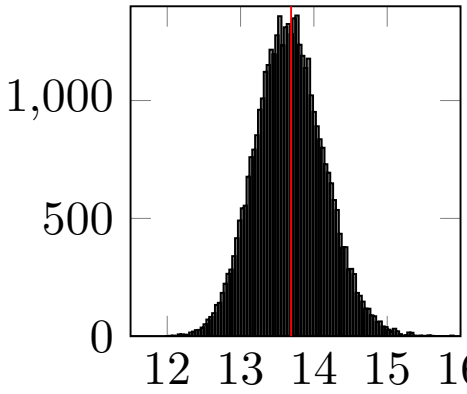


(a) Histogram of a.

(b) Histogram of ω.

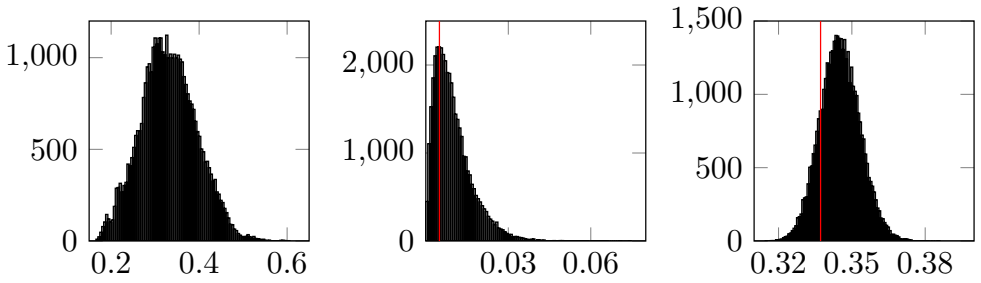


(c) Histogram of σ_n^2.



(d) Histogram of ρ.

Figure 5.5: Estimated marginal posterior distributions of the different model parameters for the proposed model (ρ estimated) for simulation #1 (single dipole)



(a) Histogram of a. (b) Histogram of ω. (c) Histogram of σ_n^2.

Figure 5.6: Estimated marginal posterior distributions of the default-ρ model ($\rho = \rho_{\text{fix}}$ is not estimated) for simulation #1 (single dipole).

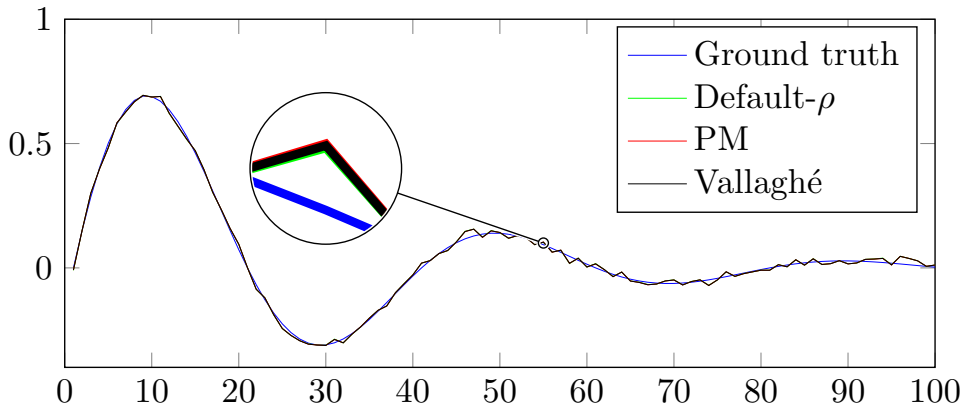


Figure 5.7: Recovered waveforms for the single dipole simulation #1.

Monte Carlo runs for different noise realizations in order to obtain measures of uncertainties for the different estimates.

Finally, Fig. 5.7 compares the estimated waveforms obtained with the proposed method (PM) and Vallaghé's one, where we can observe that both estimated waveforms are close to the ground truth. Quantitative results associated with simulation #1 are summarized in Table 5.3 and confirm that both methods perform similarly for this scenario.

The estimated locations for simulation #2 are shown in Fig. 5.8. In this case the default-ρ model fails to recover the correct dipole location and spreads the activity over a significant area of the brain, due to the fact that the difference between the leadfield matrices is significantly higher. The optimization method is still able to recover the dipole position correctly and gives an averaged MMSE estimate of the skull conductivity $\hat{\rho} = 3.85\frac{mS}{m}$ over 20 Monte Carlo runs while the proposed method estimates a mean value of $3.49\frac{mS}{m}$ (closer to the ground truth value of $3.59\frac{mS}{m}$) with a standard deviation of $0.12\frac{mS}{m}$. Fig. 5.9 shows that in this case the proposed method estimates a waveform that is considerably closer to the ground truth. Table 5.4 summarizes quantitative results associated with simulation #2, which confirm the remarks resulting from Fig. 5.9.

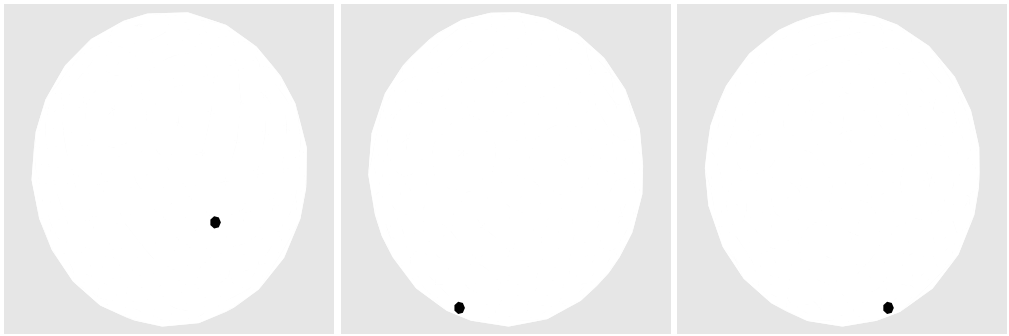
In order to asses the convergence of the experiment, the potential scale reduction factors (PSRFs) [71] for simulation #1 are shown in Fig. 5.10, showing they tend to 1 displaying good numerical convergence.

In summary, both the proposed and Vallaghé's method are able to find the correct dipole position for the two first simulations scenarios. However, the proposed method is able to better estimate the skull conductivity and the activation waveforms for simulation #2 while not being restricted to single dipole activations as Vallaghé's method.

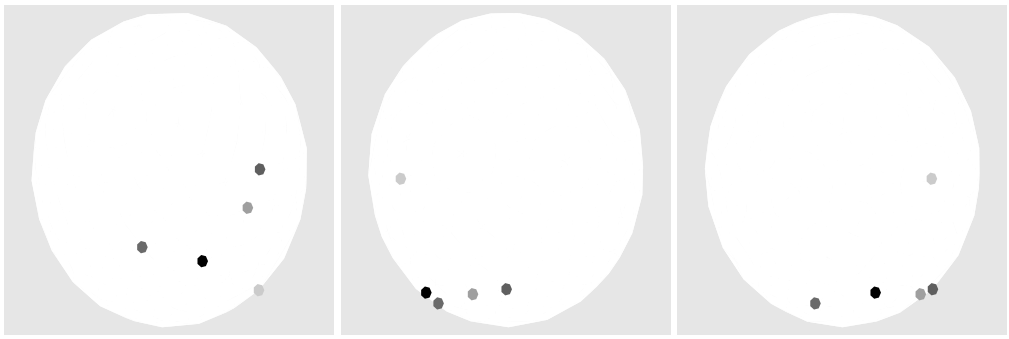
5.6.1.3 Deep dipole activations

In order to investigate how the algorithm performs for active dipoles at different depths, the following set of experiments was performed:

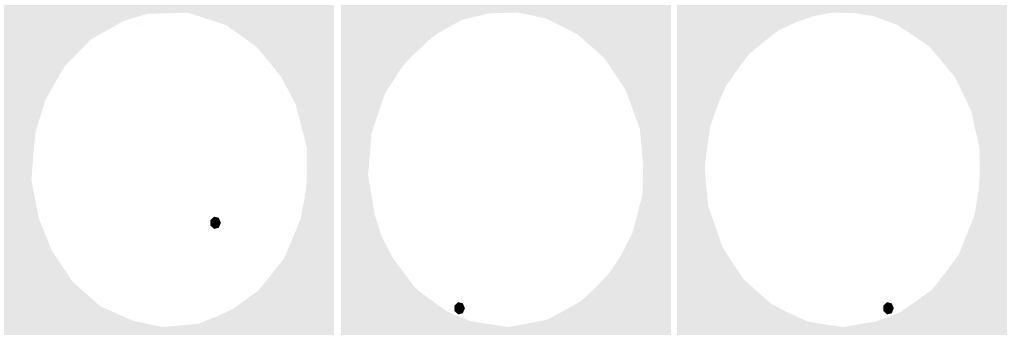
The original 200-dipole brain model was expanded to have 600 dipoles. The first 200 were localized in the same positions as in the original experiment (henceforth called "superficial dipoles"), the second set of 200 dipoles were localized 10% closer to the center of the sphere model compared to the first 200 dipoles (henceforth called "medium dipoles"). The final 200 dipoles were located 20% closer to the center of the sphere model than the superficial dipoles, and will be referred to as "deep dipoles".



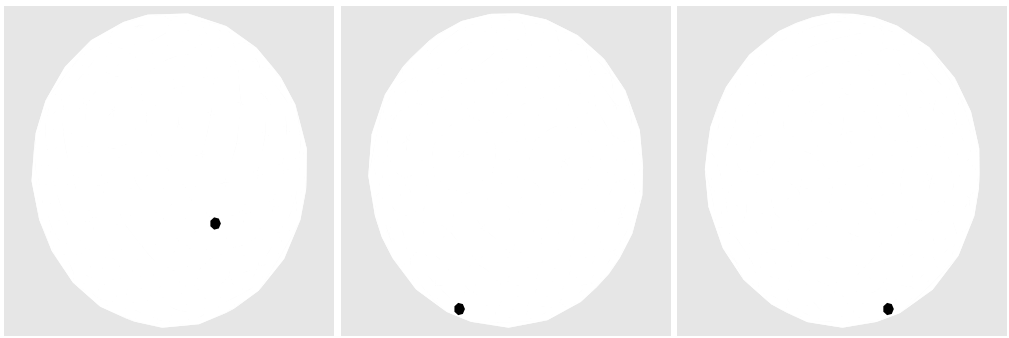
(a) Ground truth - Axial, coronal and sagittal views respectively



(b) Default-ρ model - Axial, coronal and sagittal views respectively



(c) Proposed method - Axial, coronal and sagittal views respectively



(d) Vallaghé's method - Axial, coronal and sagittal views respectively

Figure 5.8: Estimated activity for single dipole simulation #2.

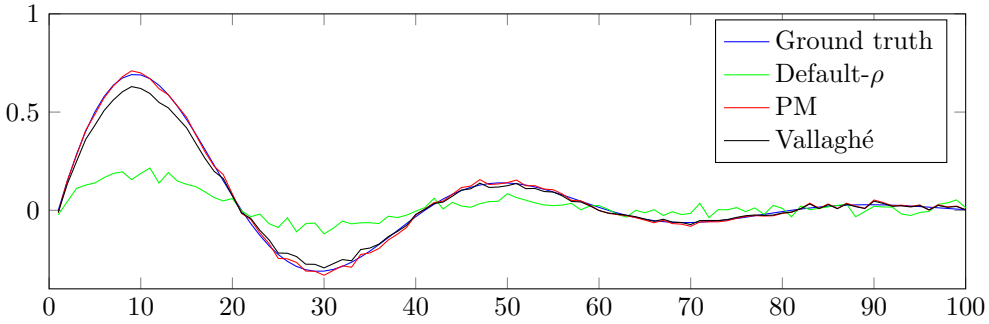
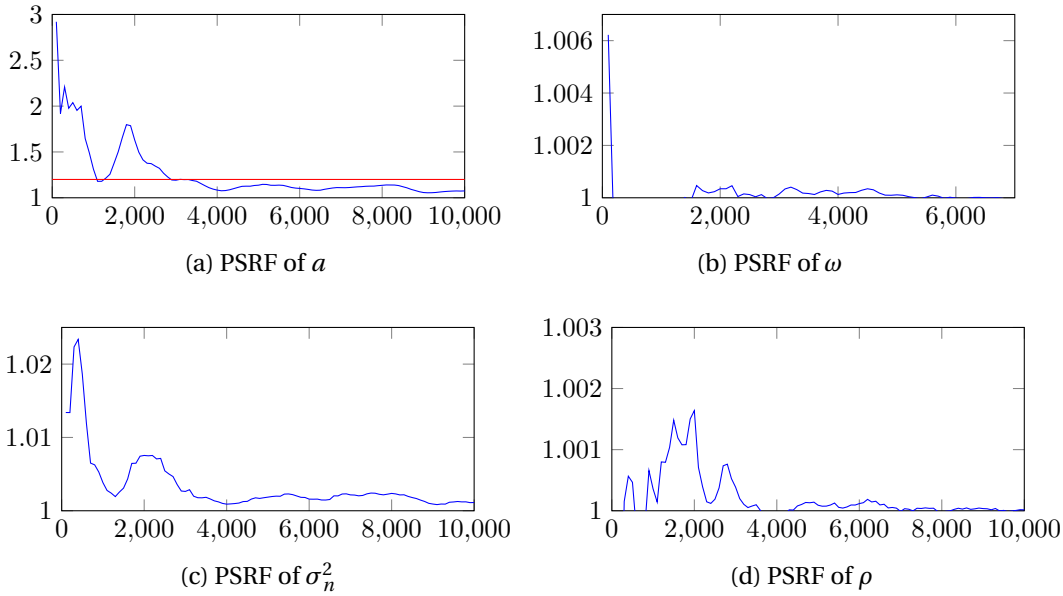


Figure 5.9: Estimated waveforms for simulation #2 (single dipole)



(a) PSRF of a

(b) PSRF of ω

(c) PSRF of σ_n^2

(d) PSRF of ρ

Figure 5.10: PSRFs for the proposed method versus the number of iterations for simulation #1 (single dipole).

Method	Pos. error	$\frac{\lvert\hat{X}-X_{\text{gt}}\rvert_F^2}{\lvert X_{\text{gt}}\rvert_F^2}$	$\frac{\hat{\sigma_n}^2-\sigma_n^2}{\sigma_n^2}$	$\hat{\rho}-\rho_{\text{gt}}$	$\frac{\lvert\boldsymbol{H}(\rho_{\text{gt}})-\boldsymbol{H}(\hat{\rho})\rvert_F^2}{\lvert\boldsymbol{H}(\rho_{\text{gt}})\rvert_F^2}$
Default-ρ	224×10^{-3}	1.36	108.4×10^{-3}	14.56×10^{-3}	75.85×10^{-3}
PM	0	2.47×10^{-3}	5.89×10^{-3}	10.33×10^{-6}	77.14×10^{-9}
Vallaghé	0	6.17×10^{-3}	N/A	265×10^{-6}	92.3×10^{-6}

Table 5.4: Estimation errors for the different parameters (Simulation #2)

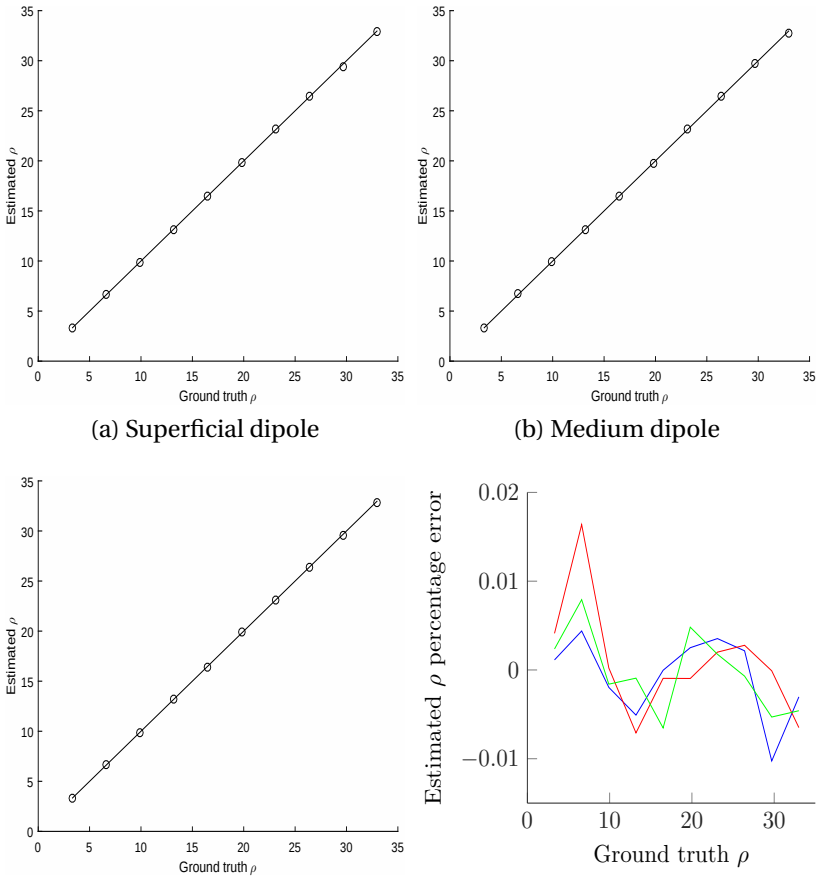
With this new brain model, a random superficial dipole was chosen. This superficial dipole was assigned a damped sinusoidal brain activity wave. In two separate scenarios, this activation wave was moved to the closest medium dipole and the closest deep dipole respectively, resulting in three values of **X**. For each of these three activations 10 linearly spaced values of ground truth ρ were chosen to calculate the leadfield matrix **H**, resulting in 30 noiseless measurements **HX**. Finally, noise was added to each of these noiseless measurements in order to obtain different values of the SNR, i.e., SNR = 10, 15, 20 and 30dB, resulting in a total of 120 noisy measurements **Y**. The proposed method was applied to each of the values of **Y** to recover the original value of **X** jointly with ρ.

In all 120 experiments, the proposed method was able to correctly estimate that there was a single active dipole and which one it was (without any confusion with ones that were slightly deeper or slightly more superficial than the active one). Figs. 5.11 and 5.12 show the estimation errors of ρ for SNR values of 30 and 20dB respectively. As we can see, the estimated values of ρ are all very close to the expected ideal diagonal for all dipole depths. This can also be seen in Figs.5.11d and 5.12d where the normalized error (expressed in percentage) is always below 2%. In comparison, Figs. 5.13 and 5.14 show the results when the SNR is 15 and 10dB respectively. As expected, increasing the noise in the system makes the estimation worse, causing the normalized error to be upto 6% for 15dB and upto 10% for 10dB of SNR.

These results indicate that the method is capable of estimating the depth of the active dipole and the skull conductivity jointly as long as the amount of noise present in the system is in a reasonable range.

5.6.1.4 MULTIPLE DIPOLE ACTIVATIONS

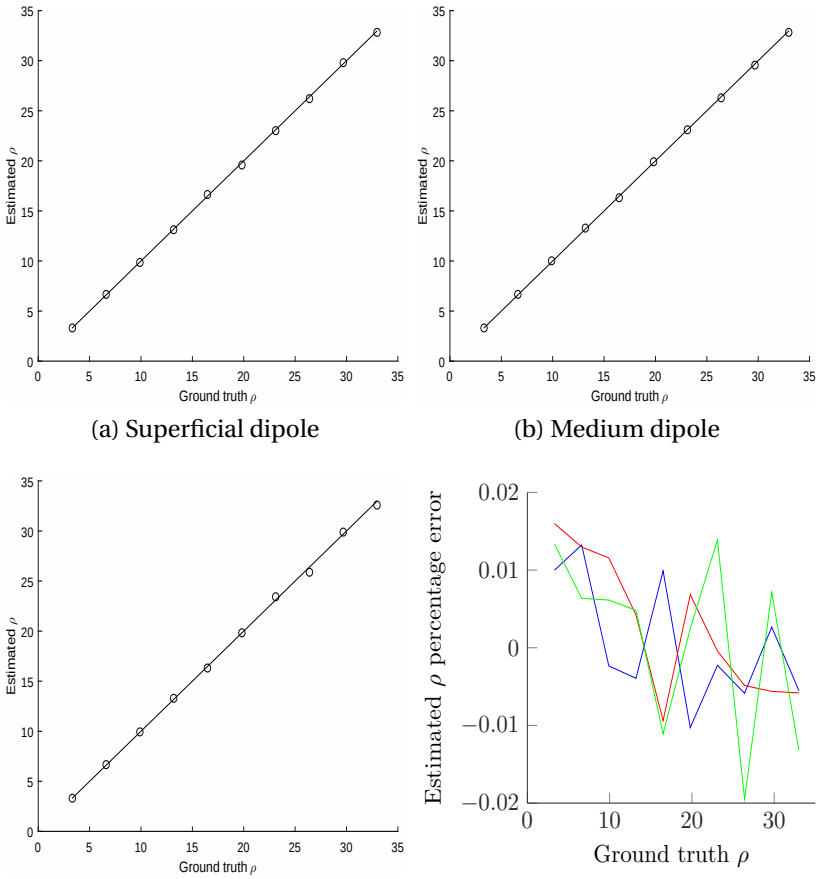
The second kind of simulations considered a variable amount of active dipoles to analyze the detection capabilities of the proposed algorithm. In each simu-



(a) Superficial dipole (b) Medium dipole

(c) Deep dipole (d) Percentage error as a function of ρ

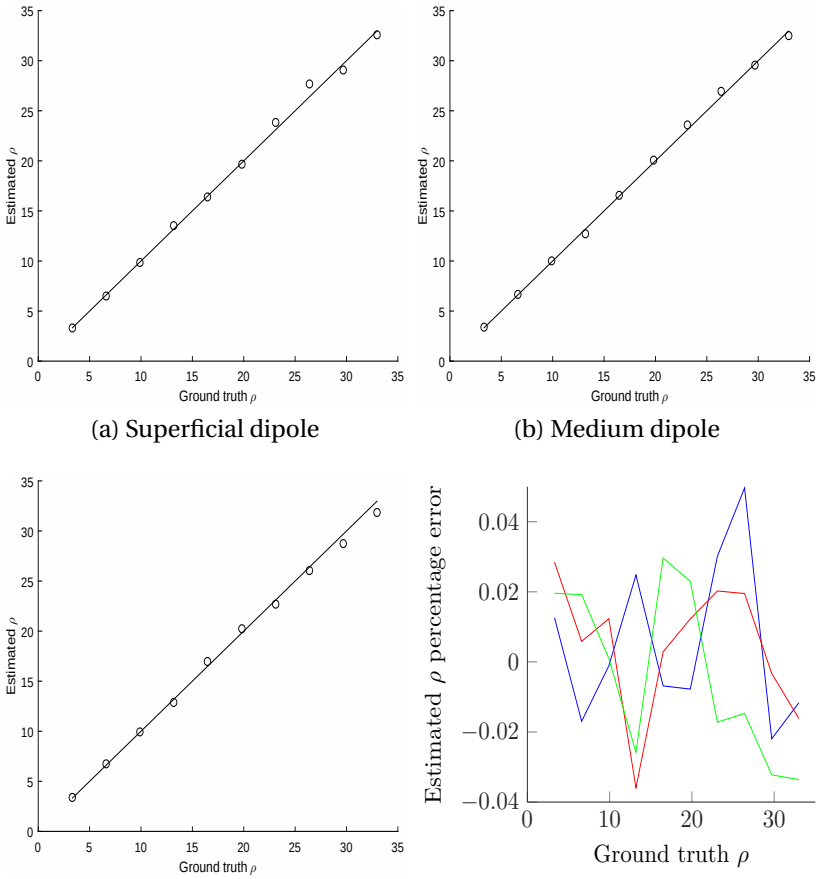
Figure 5.11: Estimated conductivity for the multiple dipole depths experiments (SNR = 30dB).



(a) Superficial dipole (b) Medium dipole

(c) Deep dipole (d) Percentage error as a function of ρ

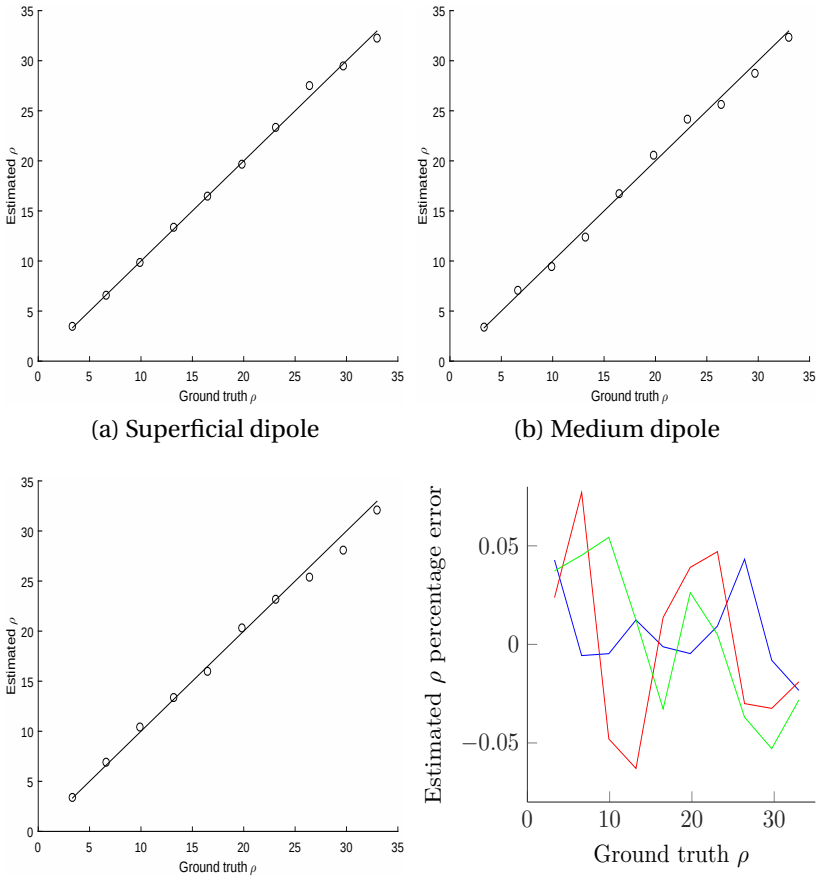
Figure 5.12: Estimated conductivity for the multiple dipole depths experiments (SNR = 20dB).



(a) Superficial dipole (b) Medium dipole

(c) Deep dipole (d) Percentage error as a function of ρ

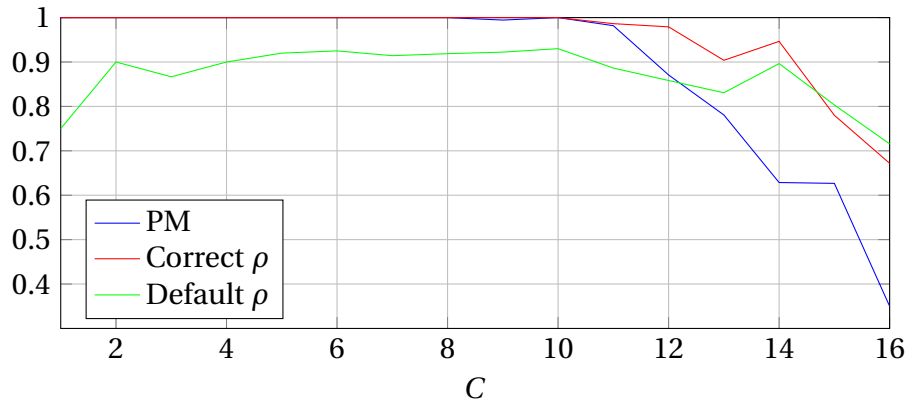
Figure 5.13: Estimated conductivity for the multiple dipole depths experiments (SNR = 15dB).



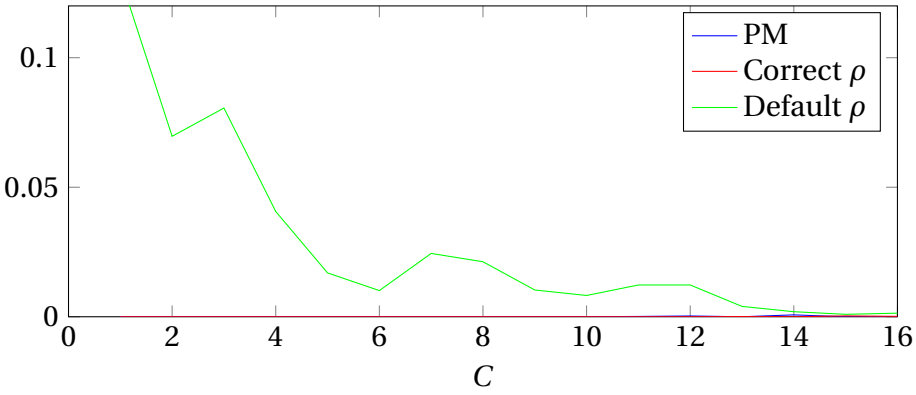
(a) Superficial dipole (b) Medium dipole

(c) Deep dipole (d) Percentage error as a function of ρ

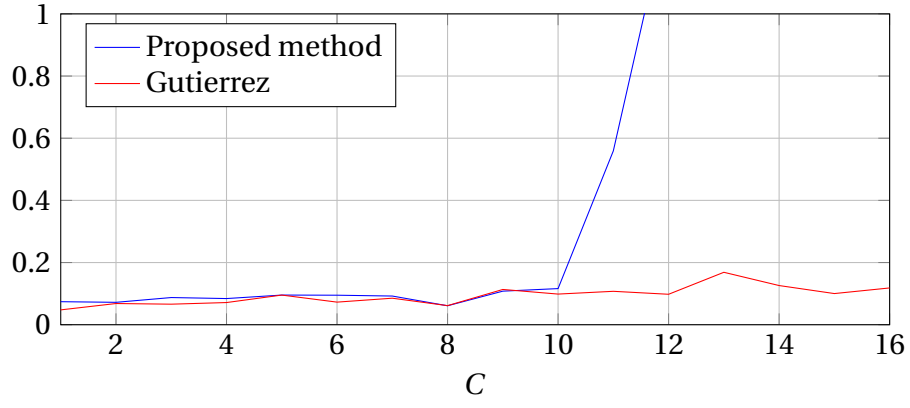
Figure 5.14: Estimated conductivity for the multiple dipole depths experiments (SNR = 10dB).



(a) Recovery rate versus C.



(b) Proportion of residual energy versus C.



(c) Estimated RMSE of $\hat{\rho}$

Figure 5.15: Performance measures for the estimation of multiple dipoles as a function of C.

lation, C dipoles were activated with damped sinusoidal waves with frequencies varying between 5 and 20Hz. The activations were sampled at 200Hz and scaled in amplitude in order to produce the same energy in the different measurements. Twenty different simulations were conducted for each value of $C = 1, ..., 16$, each one having a different set of active dipoles and a different uniform random value of conductivity in the range $\rho_{\min} < \rho_{gt} < \rho_{\max}$, resulting in a total of 320 experiments. Noise was added to the measurements to obtain SNR = 30dB.

Since Vallaghé's method cannot be applied to activations that have more than a single active dipole the performance of the proposed model was compared with two other recovery methods: (1) the correct-ρ model with $\rho_{\text{fix}} = \rho_{gt}$ (to evaluate the loss of performance when ρ is estimated) and (2) a default-ρ model with a $\rho_{\text{fix}} = \frac{\rho_{\min}+\rho_{\max}}{2} = 18.15\frac{mS}{m}$ (to illustrate the improvement due to the estimation of ρ). All models were run using 8 MCMC parallel chains. For each simulation result, the C estimated active dipoles that generated the strongest measurements in $\boldsymbol{Y}$ were considered to be the main dipoles recovered by the algorithm, while the other dipoles were used to compute the residual activity.

We define the recovery rate as the proportion of active dipoles that were recovered by a given method. Fig. 5.15a displays the average recovery rate as a function of C for the three models. The correct-ρ model is able to correctly recover the dipoles perfectly up to 10 dipoles, whereas its performance declines significantly when $C > 10$. The fact that the recovery performance decreases with more active non-zeros is well known since the leadfield matrix span limits the maximum possible amount of non-zeros that can be recovered correctly [37]. In comparison, the proposed model is able to estimate ρ jointly with the brain activity up to $C = 10$ practically without any recovery loss. For $C > 10$ its performance decreases faster than the other method because of the increasing error in the estimation of ρ (as shown in Fig. 5.15c). The recovery rate of the default-ρ model is significantly worse than the proposed method even for low values of C for the same reasons mentioned for single dipoles. Since the recovery rate only considers the main detected dipoles, it is also interesting to analyze the energy contained in the residual dipoles. Thus, we define the proportion of residual energy as the amount of energy contained in the measurements generated by the residual activity over the total energy in the measurements. This quantity is displayed in Fig. 5.15b where we can see that both the correct-ρ model and the proposed method have almost-zero energy (lower than 1%) in the residual dipoles for low values of C contrary to the default-ρ

model that has a larger proportion of residual energy.

Our algorithm was also compared with Gutierrez' method [63] which estimates the activity amplitude jointly with the skull conductivity (but requires knowing which are the active dipoles in advance). Because of this limitation, it makes no sense to analyze the recovery rate of Gutierrez' method but we can use it to compare the estimation of ρ as shown in Fig. 5.15c. We can see that the estimation performance of both methods is comparable up to $C = 10$. However, note again that the proposed algorithm is able to estimate the active dipoles positions contrary to Gutierrez' method. Conversely, for $C > 10$ the performance of our algorithm drops because it fails to recover the active dipole positions correctly.

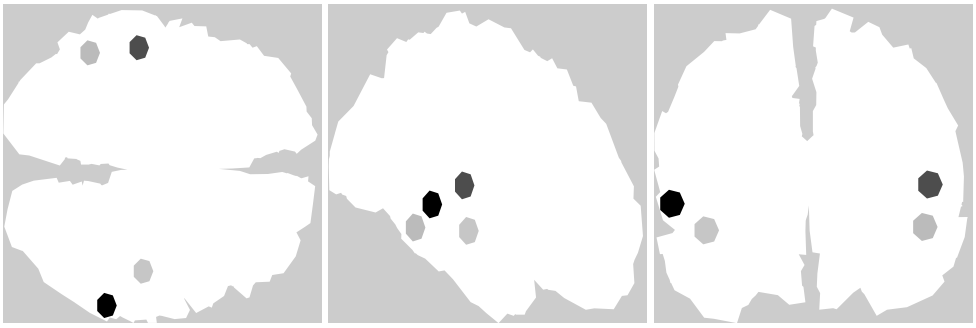
In summary, for $C \leq 10$, the quality of the reconstruction of the proposed method is very close to the method that knows the correct value of ρ in advance and significantly better than the default-ρ model and the error in the estimation of ρ is very close to the one obtained with Gutierrez' method (that knows the active dipole positions in advance). For $C > 10$ the performance of the proposed method drops faster than the method that knows the ground truth value of ρ since the error in the estimation of ρ increases considerably.

5.6.2 Real data

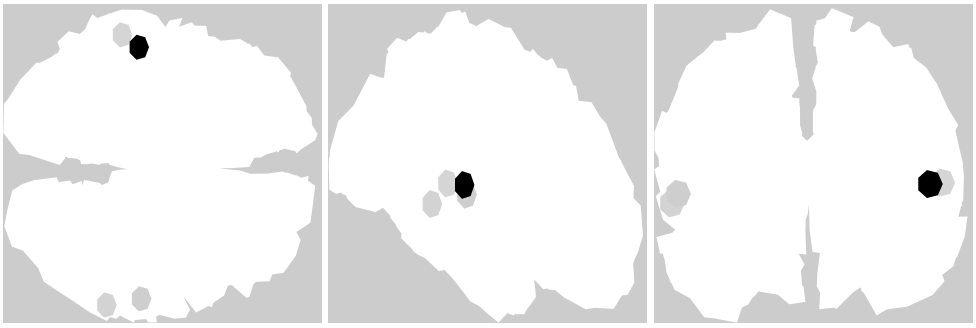
5.6.2.1 Auditory evoked responses

The left-ear auditory pure-tone stimulus data set from the MNE software [79, 80] that was used in Chapters 3 and 4 is considered in this section. It uses a realistic BEM (Boundary element method) head model containing 1.844 dipoles located on the cortex with orientations that are normal to the brain surface. The data was sampled with 306 MEG sensors at 600Hz, low-pass filtered at 40Hz and downsampled to 150Hz. One channel that had technical artifacts was ignored. The measurements corresponding to 200ms of data preceding each stimulus were considered to estimate the noise covariance matrix that was used to whiten the measurements. Fifty-one epochs were averaged to calculate $\boldsymbol{Y}$. The activity of the source dipoles was estimated jointly with the skull conductivity for the period lasting 500ms after the stimulus. From a clinical perspective it is expected to find the brain activity primarily focused on the auditory cortices that are located close to the ears in both brain hemispheres.

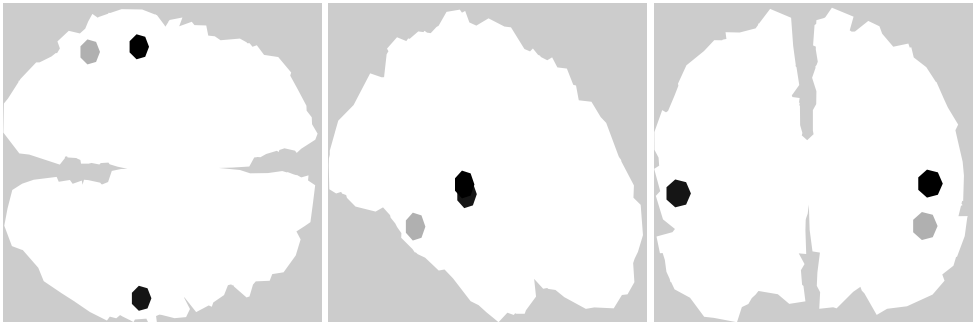
Our method was compared with (1) a default-ρ model that uses $\rho = 6\frac{mS}{S}$, corresponding to a ratio of 50 between the scalp and skull conductivities (the



(a) Default-ρ model

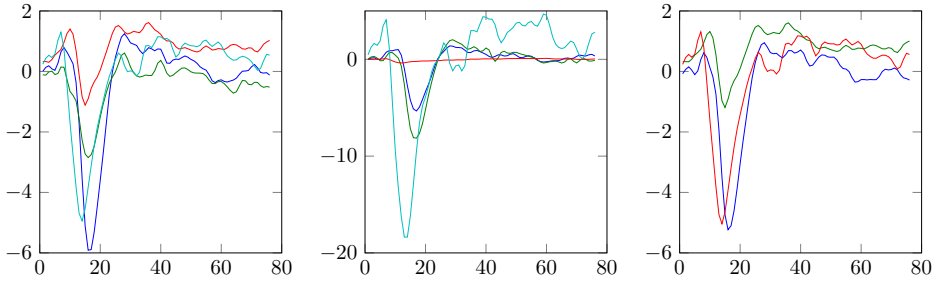


(b) ℓ_{21} mixed norm



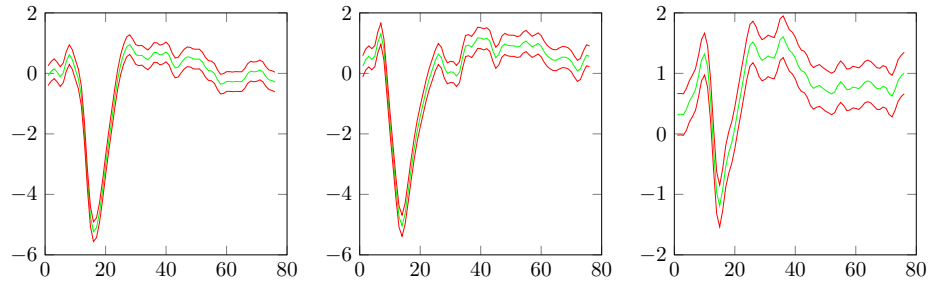
(c) Proposed method

Figure 5.16: Estimated activity for the auditory evoked responses.



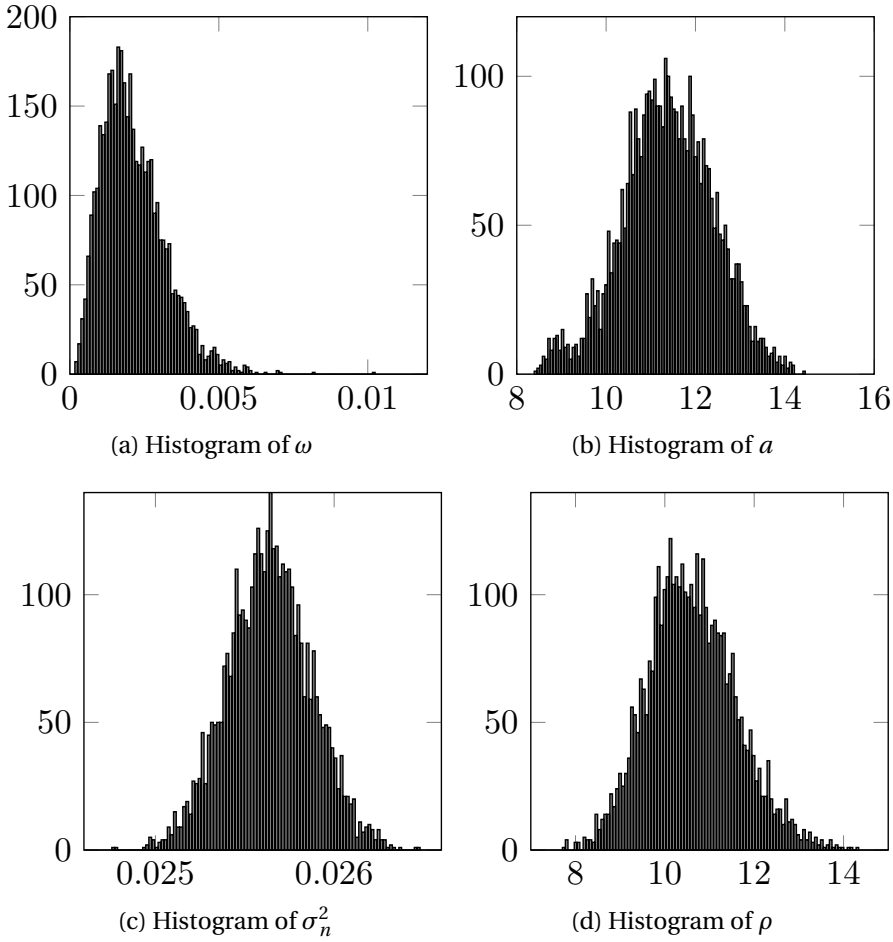
(a) Default-ρ model (b) ℓ_{21} mixed norm (c) Proposed method

Figure 5.17: Estimated waveforms for the auditory evoked responses.



(a) Waveform 1 (Center right dipole) (b) Waveform 2 (Center left dipole) (c) Waveform 3 (Posterior left dipole)

Figure 5.18: Estimated waveforms mean and boundaries $\mu \pm 2\sigma$ for the auditory evoked responses using the proposed model.



(a) Histogram of ω

(b) Histogram of a

(c) Histogram of σ_n^2

(d) Histogram of ρ

Figure 5.19: Estimated parameter distributions for the auditory evoked responses.

default value used by the MNE software), and (2) the ℓ_{21} mixed norm regularization introduced in [43] also using a leadfield matrix with $\rho = 6\frac{mS}{S}$.

We can see in Fig. 5.16 that out of the three models, the solution found by our method is the one that best agrees with the clinically expected solution of finding the activity focused on the auditory cortices whereas the other two spread the activity over several dipoles around the area. In addition, the MMSE estimator of the skull conductivity is $\hat{\rho} = 10.6\frac{mS}{S}$, corresponding to a ratio of 31 between the scalp and skull conductivities. This shows that this ratio is considerably lower than the value of 80 typically used in earlier research [59] and in agreement with recent studies [60, 90].

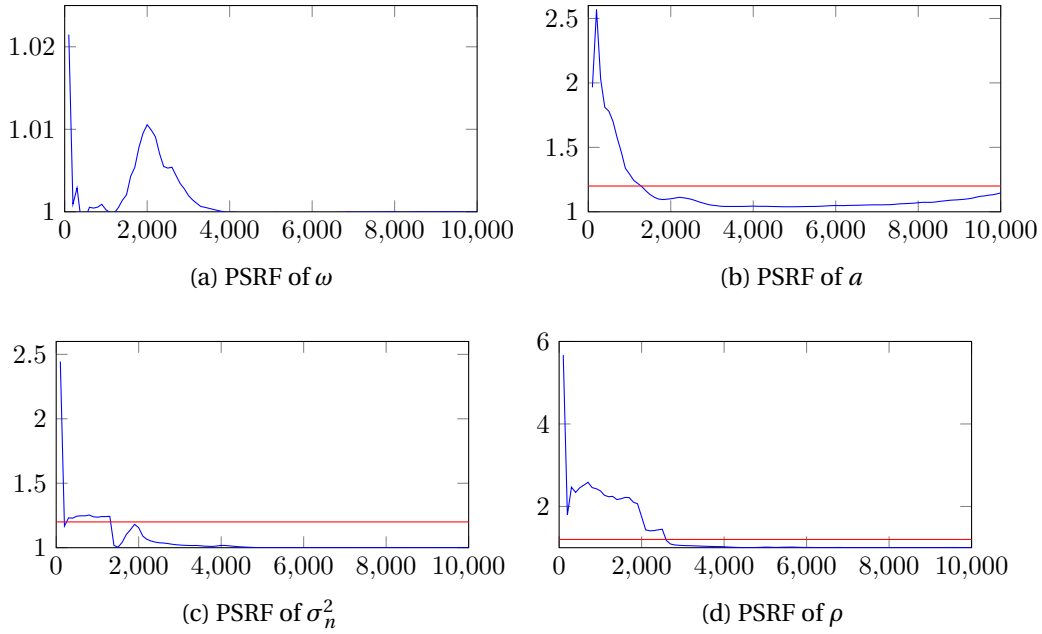
The estimated waveforms are presented in Fig. 5.17, where we can see the similarities of the estimated waveforms. However, the proposed method concentrates the activity of several dipoles in the same one. The different waveforms detected by our algorithm are presented separately in Fig. 5.18 where the mean value for each activation and the confidence intervals (mean $\pm$ two standard deviations) are displayed. We can see that the two dipoles located in the auditory cortices have most of their activity concentrated in strong peaks located around 90 ms after the stimulus, as it is clinically expected. The histograms of the sampled variables are shown in Fig. 5.19.

The potential scale reduction factors (PSRFs) [71] for the experiments are shown in Fig. 5.20, showing the good numerical convergence of the proposed partially collapsed Gibbs sampler.

In summary, the proposed method is able to concentrate the brain activity more strongly in the auditory cortices (where it is expected to be) than the other two methods and estimates a value of ρ that is more compatible with the findings of recent studies than the default value used by the MNE software.

5.6.3 Computational cost

The price to pay with the good performance of the proposed method is its computational complexity. Using Matlab implementations in a modern Xeon CPU E3-1240 processor, each simulation for the single dipole synthetic data experiments was processed on average in 96.1 seconds by the proposed model, 51.29 seconds by the default-ρ model and 23.8 seconds by Vallaghé's method. This is a common disadvantage of Bayesian methods when compared with optimization techniques.



(a) PSRF of ω

(b) PSRF of a

(c) PSRF of σ_n^2

(d) PSRF of ρ

Figure 5.20: PSRFs of the sampled variables for the auditory evoked responses along iterations.

5.7 CONCLUSION

This chapter investigated a new hierarchical Bayesian model for estimating the skull conductivity jointly with the brain activity directly from M/EEG measurements. A polynomial approximation of the leadfield matrix was used. The results of the proposed method were compared with the ones obtained with a fixed-conductivity model, showing that the proposed model is able to better estimate the underlying brain activity due its better flexibility. This improved performance was particularly observed when the value of the skull conductivity used in the fixed-conductivity model was far away from its ground truth value or in the presence of multiple dipoles. In addition, the proposed method was compared to two different optimization techniques introduced in [65] and [63]. Our method was shown to provide results of similar or better quality without requiring a single active dipole or knowledge of the active dipole positions in advance. Our algorithm was also applied to an auditory evoked response real dataset, showing that estimating the skull conductivity improves the quality of the reconstruction when compared with the fixed-conductivity model and with the ℓ_{21} mixed norm regularization.

CHAPTER 6

CONCLUSION AND FUTURE WORK

Contents

6.1 CONCLUSIONS

M/EEG is a powerful non-invasive technique that serves as a crucial tool in the analysis of several medical conditions including epilepsy. It has a much higher temporal resolution than other techniques such as fMRI but it has a lower spatial resolution. Due to this, a number of algorithms have arisen aiming to improve the spatial resolution of M/EEG in order to make it a more valuable tool for doctors. This is done by solving the source localization problem, that aims to estimate brain activity from the measurements. The main issue is that source localization is an ill posed inverse problem and, as such, admits an infinite number of solutions. Thus, in order to alleviate the issue, it is necessary to apply a regularization that promotes realistic properties of the solution of the inverse problem. An additional problem in source localization is that the head operator, although typically assumed to be completely known by most of the algorithms, can present uncertainties that significantly affect the quality of the brain activity restoration.

In this PhD, we have investigated several hierarchical Bayesian models that aim to solve the source localization problem. In Chapter 3 we developed a first model that promoted sparsity of the brain activity to solve the M/EEG source

localization problem. More precisely, a Bernoulli-Laplace prior was assigned to the brain activity corresponding to an $\ell_0 + \ell_1$ regularization. This model was shown to provide competitive results when compared to the more traditional sLoreta and weighted ℓ_1 algorithms. In addition, it was also able to estimate its hyperparameters automatically, which is a very interesting property. However, this initial model had some drawbacks. From a theoretical point of view, it failed to exploit the temporal structure of the data and, from an implementation point of view, the direct sampling of the dipole amplitudes resulted in an algorithm that was several orders of magnitude slower than optimization algorithms.

In order to improve our first model, we proposed to take advantage of the fact that the dipoles that are active or inactive in the brain are not likely to change in a millisecond scale. Based on this observation, we assigned a multivariate Laplace-Bernoulli prior to the brain activity (corresponding to an ℓ_{20} mixed norm regularization) that forced all the time samples of the same dipoles to be either jointly active or inactive. In addition, two other modifications were implemented to improve the sampling algorithm. The first one was the introduction of a latent variable that made the conditional distributions of the posterior distribution of interest much easier to sample. The second one was the introduction of Metropolis-Hastings moves allowing active dipoles to be moved in an appropriate neighborhood. Both modifications accelerated the algorithm speed in several orders of magnitude. The resulting algorithm was compared with the weighted ℓ_{21} mixed norm regularization, showing much better results in detecting non-zeros with synthetic data. Using real data, its results were compared with several state-of-the-art algorithms, showing consistency with expected results.

Finally, uncertainties in the operator were considered. Even though most state-of-the-art techniques assume the head operator to be perfectly known, in reality it depends on several physical parameters. Some of these parameters can be known with a good precision (such as the position of the electrodes or the conductivity of the scalp) whereas others cannot. In particular, the conductivity of the skull has been shown to vary strongly among subjects and to affect the estimated brain activity significantly. Therefore, we have investigated an extended Bayesian model to accounts for uncertainties in the skull conductivity, aiming to estimate it jointly with the brain activity. This resulted in the algorithm presented in Chapter 5. The proposed approach models the head operator as a polynomial matrix depending upon the skull conductivity. Random walk Metropolis-Hastings moves were then used to sample the skull con-

ductivity. The algorithm was compared with other state-of-the-art algorithms that allow us to estimate the skull conductivity jointly with the brain activity, showing that it is able to obtain results of similar or better quality while being applicable to more general situations.

6.2 FUTURE WORK

The work conducted during this PhD could be continued in the frame of several prospects. On the short term we have two interesting ideas to pursue:

Accelerate the sampling efficiency

The model presented in Chapter 4 provides good results but is still significantly slower than optimization algorithms (with the noticeable advantage of being able to estimate its hyperparameters directly from the data) due to the iterative MCMC sampling algorithm. It would be very interesting to consider alternative techniques trying to maximize the posterior distribution of our Bayesian model. For instance, the authors have tried replacing algorithm 4.1 by a message passing method. To develop our message passing algorithm, the Generalized Approximate Message Passing (GAMP) model presented in [91] was adapted to use a multivariate Bernoulli-Laplace prior for $\boldsymbol{X}$. The algorithm was tested with synthetic data with some interesting results. If the operator used $\boldsymbol{H}$ has columns that are not very correlated (for instance, if the operator columns have been generated independently), the algorithm is able to obtain the same or even better results than the algorithm in Chapter 4 with a significant reduction of the execution time, finding the correct non-zero positions in a very small amount of iterations. However, if there is a high correlation between the columns of the operator, the algorithm tends to get stuck and overestimate several correlated columns to be “active" at the same time. Unfortunately, due to the fact that spatially close dipoles generate very similar M/EEG measurements, head operators have columns that are generally highly correlated. As a consequence, the message passing method developed does not work for M/EEG source localization. We think that trying to improve the GAMP algorithm to avoid this problem would be an interesting research topic.

Application of Bernoulli-Laplace models to other inverse problems

Even though the main focus of the PhD was to solve the M/EEG source

localization problem, the models and algorithms that have been developed are not restricted to this application. The model presented in Chapter 4 could be used as an ℓ_{20} regularization-like technique for any inverse problem characterized by a sparse solution. For instance, we have started some work in compressive spectral imaging (CSI). The CSI problem consists in recovering the full spatial and spectral information of a scene from a significantly undersampled set of random projections acquired by a compressive spectral imager, such as the Coded Aperture Snapshot Spectral Imaging (CASSI) [92, 93]. For significant undersampling rates, the resulting inverse problem is ill-posed and requires a regularization to have a unique solution. One of the most common regularizations is promoting the hyperspectral image to be sparse in a given basis [94]. The quality of the reconstruction is further improved if other properties of natural images are included in the regularization such as their high degree of local structure similarity [95]. In [96] we introduced a hierarchical Bayesian model for solving the CASSI problem. Our model promoted the solution to be sparse in a selected wavelet basis and smooth in the image domain. Experiments showed that the results obtained with our method were competitive with state-of-the-art algorithms. The priors used in the model were very similar to the ones used in the model of Chapter 4 of the present thesis. However, sparsity was promoted using a multivariate Laplace distribution (instead of the multivariate Bernoulli-Laplace distribution used in the thesis). This was due to the fact that in the CASSI problem the high dimensionality of the problem did not allow us to implement the multivariate Bernoulli-Laplace prior distribution as easily as for the M/EEG source localization problem. It would be interesting to follow this track to make it possible to use the multivariate Bernoulli-Laplace prior for high dimensionality problems. In addition, the polynomial interpolation used for the operator in Chapter 5 could be used in any inverse problem where the operator is not perfectly known and depends on a scalar variable. To summarize, we think that it should be interesting to try to apply the Bayesian models investigated in this PhD thesis to different types of inverse problems to see how they perform.

whereas in the long term we would like to

Find ways of representing the operator using more than one scalar variable
In Chapter 5 we used a polynomial matrix to represent the head operator.

This allowed us to sample the skull conductivity easily and provided interesting results. However, the proposed method is only applicable if the head operator depends on a single scalar parameter. If we want to model an operator that depends on two or more scalar parameters, interpolating it with multidimensional polynomials is a more challenging problem. It would be interesting to pursue other ways of modeling the operator as a function of several scalar parameters in a more tractable way. We are for instance thinking to using splines interpolations that have been used sucessfully in many applications [97].

APPENDIX A

CONDITIONAL PROBABILITY DISTRIBUTIONS DERIVATIONS

A.1 INTRODUCTION

In this appendix we will show the algebraic derivation of the conditional probability distributions of the associated models presented in Chapter 4 (the structured sparsity model) and Chapter 5 (the skull conductivity joint estimation model).

A.2 STRUCTURED SPARSITY MODEL

A.2.1 POSTERIOR DISTRIBUTION

As specified in Chapter 4, the posterior distribution associated with the proposed model is:

$$f(\boldsymbol{Y},\sigma_n^2,\boldsymbol{X},\boldsymbol{z},a,\boldsymbol{\tau}^2,\omega) = f(\boldsymbol{Y}|\boldsymbol{X},\sigma_n^2)f(\boldsymbol{X}|\boldsymbol{\tau}^2,\boldsymbol{z},\sigma_n^2)f(\boldsymbol{z}|\omega)f(\boldsymbol{\tau}^2|a)f(\sigma_n^2)f(a)f(\omega) \quad [A.1]$$

From it we can derive the conditional distributions of all the associated pa-

rameters and hyperparameters using Bayes' theorem:

$$f(z_i, \boldsymbol{x}_i | \boldsymbol{Y}, \boldsymbol{X}_{-i}, \sigma_n^2, \tau_i^2, \omega) \propto f(\boldsymbol{Y} | \boldsymbol{X}, \sigma_n^2) f(\boldsymbol{x}_i | \tau_i^2, z_i, \sigma_n^2) f(z_i | \omega) \quad [A.2]$$

$$f(\tau_i^2 | \boldsymbol{x}_i, \sigma_n^2, a, z_i) \propto f(\boldsymbol{x}_i | \tau_i^2, z_i, \sigma_n^2) f(\tau_i^2 | a) \quad [A.3]$$

$$f(a | \boldsymbol{\tau}^2) \propto f(\boldsymbol{\tau}^2 | a) f(a) \quad [A.4]$$

$$f(\sigma_n^2 | \boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{\tau}^2, \boldsymbol{z}) \propto f(\boldsymbol{Y} | \boldsymbol{X}, \sigma_n^2) f(\boldsymbol{X} | \boldsymbol{\tau}^2, \boldsymbol{z}, \sigma_n^2) f(\sigma_n^2) \quad [A.5]$$

$$f(\omega | \boldsymbol{z}) \propto f(\boldsymbol{z} | \omega) f(\omega) \quad [A.6]$$

A.2.2 CONDITIONAL DISTRIBUTIONS

Conditional distribution of τ_i^2

The conditional distribution of τ_i^2 is

$$f(\tau_i^2 | \boldsymbol{x}_i, \sigma_n^2, a, z_i) \propto f(\boldsymbol{x}_i | \tau_i^2, z_i, \sigma_n^2) f(\tau_i^2 | a) \quad [A.7]$$

that is equal to

$$f(\tau_i^2 | \boldsymbol{x}_i, \sigma_n^2, a, z_i) = \begin{cases} \delta(\boldsymbol{x}_i) \mathcal{G}\left(\tau_i^2 \middle| \frac{T+1}{2}, \frac{v_i a}{2}\right) & \text{if } z_i = 0 \\ \mathcal{N}\left(\boldsymbol{x}_i \middle| 0, \sigma_n^2 \tau_i^2 I_T\right) \mathcal{G}\left(\tau_i^2 \middle| \frac{T+1}{2}, \frac{v_i a}{2}\right) & \text{if } z_i = 1. \end{cases} \quad [A.8]$$

Based on the development of [84] it can be seen that the conditional distribution of τ_i^2 is a generalized inverse gaussian when $z_i = 1$ and a gamma distribution when $z_i = 0$

$$f(\tau_i^2 | \boldsymbol{x}_i, \sigma_n^2, a, z_i) = \begin{cases} \mathcal{G}\left(\tau_i^2 \middle| \frac{T+1}{2}, \frac{v_i a}{2}\right) & \text{if } z_i = 0 \\ \mathcal{G}\mathcal{I}\mathcal{G}\left(\tau_i^2 \middle| \frac{1}{2}, v_i a, \frac{||\boldsymbol{x}_i||^2}{\sigma_n^2}\right) & \text{if } z_i = 1. \end{cases} \quad [A.9]$$

Conditional distribution of z_i and $\boldsymbol{x}_i$

In our model z_i and $\boldsymbol{x}_i$ are jointly sampled from

$$f(z_i, \boldsymbol{x}_i | \boldsymbol{Y}, \boldsymbol{X}_{-i}, \sigma_n^2, \tau_i^2, \omega) \propto f(\boldsymbol{Y} | \boldsymbol{X}, \sigma_n^2) f(\boldsymbol{x}_i | \tau_i^2, z_i, \sigma_n^2) f(z_i | \omega) \quad [A.10]$$

that is equal to

$$f(z_i, \boldsymbol{x}_i | \boldsymbol{Y}, \boldsymbol{X}_{-i}, \sigma_n^2, \tau_i^2, \omega) \propto$$
$$\exp\Big(-\frac{||\boldsymbol{HX}-\boldsymbol{Y}||^2}{2\sigma_n^2}\Big)\Big[(1-z_i)\delta(\boldsymbol{x}_i) + z_i \mathcal{N}\Big(\boldsymbol{x}_i\Big|0, \sigma_n^2\tau_i^2 I_T\Big)\Big]\Big[(1-\omega)\delta(z_i) + \omega\delta(z_i-1)\Big] =$$
$$\exp\Big(-\frac{||\boldsymbol{HX}-\boldsymbol{Y}||^2}{2\sigma_n^2}\Big)\Big[(1-\omega)\delta(z_i)\delta(\boldsymbol{x}_i) + \omega\delta(z_i-1)\mathcal{N}\Big(\boldsymbol{x}_i\Big|0, \sigma_n^2\tau_i^2 I_T\Big)\Big] =$$
$$(1-\omega)\delta(z_i)\exp\Big(-\frac{||\boldsymbol{HX}_{-i}-\boldsymbol{Y}||^2}{2\sigma_n^2}\Big)\delta(\boldsymbol{x}_i) +$$
$$\omega(2\pi\tau_i^2\sigma_n^2)^{-\frac{T}{2}}\delta(z_i-1)\exp\Big(-\frac{||\boldsymbol{HX}-\boldsymbol{Y}||^2}{2\sigma_n^2}\Big)\exp\Big(-\frac{||\boldsymbol{x}_i||^2}{2\sigma_n^2\tau_i^2}\Big). \quad [A.11]$$

The marginal distribution of z_i is of the form

$$f(z_i|\boldsymbol{Y}, \boldsymbol{X}_{-i}, \sigma_n^2, \tau_i^2, \omega) = \int f(z_i, \boldsymbol{x}_i|\boldsymbol{Y}, \boldsymbol{X}_{-i}, \sigma_n^2, \tau_i^2, \omega) d\boldsymbol{x}_i \propto k_0\delta(z_i) + k_1\delta(z_i-1) \quad [A.12]$$

with

$$k_0 = \int (1-\omega)\exp\Big(-\frac{||\boldsymbol{HX}_{-i}-\boldsymbol{Y}||^2}{2\sigma_n^2}\Big)\delta(\boldsymbol{x}_i)d\boldsymbol{x}_i = (1-\omega)\exp\Big(-\frac{||\boldsymbol{HX}_{-i}-\boldsymbol{Y}||^2}{2\sigma_n^2}\Big) \quad [A.13]$$

$$k_1 = \omega(2\pi\tau_i^2\sigma_n^2)^{-\frac{T}{2}}\int \exp\Big[-\frac{1}{2\sigma_n^2}\Big(||\boldsymbol{HX}-\boldsymbol{Y}||^2 + \frac{||\boldsymbol{x}_i||^2}{\tau_i^2}\Big)\Big]d\boldsymbol{x}_i. \quad [A.14]$$

This implies that z_i has the following Bernoulli distribution

$$f(z_i|\boldsymbol{Y}, \boldsymbol{X}_{-i}, \sigma_n^2, \tau_i^2, \omega) = \mathcal{B}\Big(z_i\Big|\frac{k_1}{k_0+k_1}\Big). \quad [A.15]$$

To find the value of k_1 we calculate the minus logarithm of the integrand

$$-\log\Big(\exp\Big[-\frac{1}{2\sigma_n^2}\Big(||\boldsymbol{HX}-\boldsymbol{Y}||^2+\frac{||\boldsymbol{x}_i||^2}{\tau_i^2}\Big)\Big]\Big) = \frac{1}{2\sigma_n^2}\Big(||\boldsymbol{HX}-\boldsymbol{Y}||^2+\frac{||\boldsymbol{x}_i||^2}{\tau_i^2}\Big) \quad [A.16]$$

and express it as a sum for the different values of t

$$\frac{1}{2\sigma_n^2}\Big(||\boldsymbol{HX}-\boldsymbol{Y}||^2 + \frac{||\boldsymbol{x}_i||^2}{\tau_i^2}\Big) = \frac{1}{2\sigma_n^2}\sum_{t=1}^{T}\Big(||\boldsymbol{H}\boldsymbol{x}^t-\boldsymbol{y}^t||^2 + \frac{(\boldsymbol{x}_i^t)^2}{\tau_i^2}\Big). \quad [A.17]$$

If we denote $\boldsymbol{h}^i$ each column of the operator $\boldsymbol{H}$ we can express term number t of the sum as

$$\frac{1}{2\sigma_n^2}\left(||\boldsymbol{H}\boldsymbol{x}^t-\boldsymbol{y}^t||^2+\frac{(x_i^t)^2}{\tau_i^2}\right)=\frac{1}{2\sigma_n^2}\left[\left(\sum_{j\neq i}\boldsymbol{h}^j x_j^t+\boldsymbol{h}^i x_i^t-\boldsymbol{y}^t\right)^T\left(\sum_{j\neq i}\boldsymbol{h}^j x_j^t+\boldsymbol{h}^i x_i^t-\boldsymbol{y}^t\right)+\frac{(x_i^t)^2}{\tau_i^2}\right]. \tag{A.18}$$

By denoting $\boldsymbol{D}_i^t=\boldsymbol{Y}^t-\sum_{j\neq i}\boldsymbol{h}^j x_j^t$ and expanding this expression we have

$$[\text{A.18}]=\frac{1}{2\sigma_n^2}\left((\boldsymbol{h}^i)^T\boldsymbol{h}^i(x_i^t)^2-2x_i^t(\boldsymbol{h}^i)^T\boldsymbol{D}_i^t+(\boldsymbol{D}_i^t)^T\boldsymbol{D}_i^t+\frac{(x_i^t)^2}{\tau_i^2}\right). \tag{A.19}$$

Matching the terms of the previous expression with $\frac{(x_i^t-\mu_i^t)^2}{2\sigma_i^2}+K_i^t$ we have

$$\frac{1}{2\sigma_n^2}\left((\boldsymbol{h}^i)^T\boldsymbol{h}^i(x_i^t)^2-2x_i^t(\boldsymbol{h}^i)^T\boldsymbol{D}_i^t+(\boldsymbol{D}_i^t)^T\boldsymbol{D}_i^t+\frac{(x_i^t)^2}{\tau_i^2}\right)=\frac{(x_i^t)^2}{2\sigma_i^2}-\frac{\mu_i^t x_i^t}{\sigma_i^2}+\frac{(\mu_i^t)^2}{2\sigma_i^2}+K_i^t \tag{A.20}$$

$$\frac{1}{2\sigma_i^2}=\frac{(\boldsymbol{h}^i)^T\boldsymbol{h}^i+\frac{1}{\tau_i^2}}{2\sigma_n^2}=>\sigma_i^2=\frac{\sigma_n^2\tau_i^2}{1+\tau_i^2(\boldsymbol{h}^i)^T\boldsymbol{h}^i} \tag{A.21}$$

$$\frac{\mu_i^t}{\sigma_i^2}=\frac{(\boldsymbol{h}^i)^T\boldsymbol{D}_i^t}{\sigma_n^2}=>\mu_i^t=\frac{\sigma_i^2(\boldsymbol{h}^i)^T\boldsymbol{D}_i^t}{\sigma_n^2} \tag{A.22}$$

$$\frac{(\mu_i^t)^2}{2\sigma_i^2}+K_i^t=\frac{(\boldsymbol{D}_i^t)^T\boldsymbol{D}_i^t}{2\sigma_n^2}=>K_i^t=\frac{(\boldsymbol{D}_i^t)^T\boldsymbol{D}_i^t}{2\sigma_n^2}-\frac{(\mu_i^t)^2}{2\sigma_i^2}. \tag{A.23}$$

Summing over all time samples and applying the function $\exp(-x)$ (to compensate the steps done in A.16 and A.18) results in

$$\exp\left(-\frac{||\boldsymbol{H}\boldsymbol{X}-\boldsymbol{Y}||^2}{2\sigma_n^2}\right)\exp\left(-\frac{||\boldsymbol{x}_i||^2}{2\sigma_n^2\tau_i^2}\right)=(2\pi\sigma_i)^{\frac{T}{2}}\prod_{t=1}^{T}\exp(-K_i^t).\mathcal{N}\left(x_i^t\middle|\mu_i^t,\sigma_i^2\right). \tag{A.24}$$

We can calculate the final value of k_1 by combining [A.14] and [A.24]

$$k_1 = \omega(2\pi\tau_i^2\sigma_n^2)^{-\frac{T}{2}} \int \exp\Big(-\frac{||\boldsymbol{HX}-\boldsymbol{Y}||^2}{2\sigma_n^2}\Big)\exp\Big(-\frac{||\boldsymbol{x}_i||^2}{2\sigma_n^2\tau_i^2}\Big)d\boldsymbol{x}_i =$$

$$\omega(2\pi\tau_i^2\sigma_n^2)^{-\frac{T}{2}}(2\pi\sigma_i)^{\frac{T}{2}}\prod_{t=1}^{T}\exp(-K_i^t) =$$

$$\omega\Big(\frac{\sigma_n^2\tau_i^2}{\sigma_i^2}\Big)^{-\frac{T}{2}}\exp\Big(-\frac{||\boldsymbol{HX}_{-i}-\boldsymbol{Y}||^2}{2\sigma_n^2}\Big)\exp\Big(\frac{||\boldsymbol{\mu}_i||^2}{2\sigma_i^2}\Big). \quad [A.25]$$

Using [A.11] and [A.24] we obtain the conditional distribution of $\boldsymbol{x}_i$

$$f(\boldsymbol{x}_i|z_i,\boldsymbol{Y},\boldsymbol{X}_{-i},\sigma_n^2,\tau_i^2) = \begin{cases} \delta(\boldsymbol{x}_i) & \text{if } z_i = 0 \\ \mathcal{N}\Big(\boldsymbol{x}_i\Big|\boldsymbol{\mu}_i,\sigma_i^2 I_T\Big) & \text{if } z_i = 1. \end{cases} \quad [A.26]$$

Conditional distribution of a

The conditional distribution of a is

$$f(a|\boldsymbol{\tau}^2) \propto f(a)\prod_{i=1}^{N} f(\tau_i^2|\boldsymbol{x}_i,\sigma_n^2,a) \quad [A.27]$$

$$f(a|\boldsymbol{\tau}^2) \propto a^{\alpha-1}\exp\Big(-\beta a\Big)\prod_{i=1}^{N}\Big[\Big(\frac{av_i}{2}\Big)^{\frac{T+1}{2}}(\tau_i^2)^{\frac{T-1}{2}}\exp\Big(-\frac{av_i\tau_i^2}{2}\Big)\Big] \propto$$

$$a^{\alpha+\frac{N(T+1)}{2}-1}\exp\Big(-\Big(\beta+\frac{1}{2}\sum_{i=1}^{N}[v_i\tau_i^2]\Big)a\Big) \quad [A.28]$$

which corresponds to the following gamma distribution

$$f(a|\boldsymbol{\tau}^2) = \mathcal{G}\Big(a\Big|\frac{N(T+1)}{2}+\alpha,\frac{\sum_i[v_i\tau_i^2]}{2}+\beta\Big). \quad [A.29]$$

Conditional distribution of σ_n^2

The conditional distribution of σ_n^2 is

$$f(\sigma_n^2|\boldsymbol{Y},\boldsymbol{X},\boldsymbol{\tau}^2,\boldsymbol{z}) \propto f(\boldsymbol{Y}|\boldsymbol{X},\sigma_n^2)f(\boldsymbol{X}|\boldsymbol{\tau}^2,\boldsymbol{z},\sigma_n^2)f(\sigma_n^2) \quad [A.30]$$

$$f(\sigma_n^2|\boldsymbol{Y},\boldsymbol{X},\boldsymbol{\tau}^2,\boldsymbol{z}) \propto (\sigma_n^2)^{-1}(2\pi\sigma_n^2)^{-\frac{MT}{2}} \exp\Big(-\frac{||\boldsymbol{HX}-\boldsymbol{Y}||^2}{2\sigma_n^2}\Big) \prod_{i=1}^{N}\Big[(z_i-1)\delta(\boldsymbol{x}_i)+z_i\mathcal{N}\Big(\boldsymbol{x}_i\Big|0,\sigma_n^2\tau_i^2 I_T\Big)\Big]. \quad [A.31]$$

Denoting $I_k = \{i : z_i = k\}$ for $k = \{0,1\}$ and using the identity

$$\prod_{i=1}^{N}\Big[(z_i-1)f(x)+z_i g(x)\Big] = \prod_{i\in I_0} f(x)\prod_{i\in I_1} g(x) \quad [A.32]$$

we can express this by

$$f(\sigma_n^2|\boldsymbol{Y},\boldsymbol{X},\boldsymbol{\tau}^2,\boldsymbol{z}) \propto (\sigma_n^2)^{-1}(2\pi\sigma_n^2)^{-\frac{MT}{2}} \exp\Big(-\frac{||\boldsymbol{HX}-\boldsymbol{Y}||^2}{2\sigma_n^2}\Big)\prod_{i\in I_0}\delta(\boldsymbol{x}_i) \prod_{i\in I_1}\Big[(2\pi\tau_i^2\sigma_n^2)^{-\frac{T}{2}}\exp\Big(-\frac{||\boldsymbol{x}_i||^2}{2\tau_i^2\sigma_n^2}\Big)\Big]. \quad [A.33]$$

The previous expression leads to

$$f(\sigma_n^2|\boldsymbol{Y},\boldsymbol{X},\boldsymbol{\tau}^2,\boldsymbol{z}) \propto (\sigma_n^2)^{-\left(1+\frac{(M+||\boldsymbol{z}||_0)T}{2}\right)}\exp\Big(-\frac{1}{2\sigma_n^2}\Big[||\boldsymbol{HX}-\boldsymbol{Y}||^2+\sum_{i\in I_1}\frac{||\boldsymbol{x}_i||^2}{\tau_i^2}\Big]\Big) \quad [A.34]$$

which corresponds to the following inverse gamma distribution

$$f(\sigma_n^2|\boldsymbol{Y},\boldsymbol{X},\boldsymbol{\tau}^2,\boldsymbol{z}) = \mathcal{IG}\Big(\sigma_n^2\Big|\frac{(M+||\boldsymbol{z}||_0)T}{2},\frac{1}{2}\Big[||\boldsymbol{HX}-\boldsymbol{Y}||^2+\sum_{i\in I_1}\frac{||\boldsymbol{x}_i||^2}{\tau_i^2}\Big]\Big). \quad [A.35]$$

Conditional distribution of ω

The conditional distribution for ω is

$$f(\omega|\boldsymbol{z}) \propto f(\boldsymbol{z}|\omega)f(\omega) \quad [A.36]$$

$$f(\omega|\boldsymbol{z}) \propto 1_{0,1}\prod_{i=1}^{N}\Big[\delta(z_i)(1-\omega)+\delta(z_i-1)\omega\Big] \quad [A.37]$$

where $1_{[0,1]}$ represents a function that is 1 for $0 < \omega < 1$ and 0 elsewhere. Using the identity [A.32] this can be shown to be equal to

$$f(\omega|\boldsymbol{z}) \propto 1_{0,1}(1-\omega)^{N-||\boldsymbol{z}||_0}\omega^{||\boldsymbol{z}||_0}\prod_{i\in I_0}\delta(z_i)\prod_{i\in I_1}\delta(z_i-1) \quad [A.38]$$

which corresponds to the following Beta distribution

$$f(\omega|\boldsymbol{z}) = \mathscr{B}e\Big(\omega\Big|1+||\boldsymbol{z}||_0, 1+N-||\boldsymbol{z}||_0\Big). \quad [A.39]$$

A.3 Skull conductivity joint estimation model

A.3.1 Posterior distribution

The associated posterior distribution for the model introduced in Chapter 5 is:

$$f(\boldsymbol{Y},\sigma_n^2,\bar{\boldsymbol{X}},\boldsymbol{z},a,\boldsymbol{\tau}^2,\omega,\rho) \propto f(\boldsymbol{Y}|\bar{\boldsymbol{X}},\sigma_n^2,\rho)f(\bar{\boldsymbol{X}}|\boldsymbol{\tau}^2,\boldsymbol{z},\sigma_n^2)f(\boldsymbol{z}|\omega) \quad [A.40]$$
$$f(\boldsymbol{\tau}^2|a)f(\sigma_n^2)f(a)f(\omega)f(\rho)$$

It is possible to derive the conditional distributions of the parameters and hyperparameters using Bayes' theorem:

$$f(z_i,\bar{\boldsymbol{x}}_i|\boldsymbol{Y},\bar{\boldsymbol{X}}_{-i},\sigma_n^2,\tau_i^2,\omega,\rho) \propto f(\boldsymbol{Y}|\bar{\boldsymbol{X}},\sigma_n^2,\rho)f(\bar{\boldsymbol{x}}_i|\tau_i^2,z_i,\sigma_n^2)f(z_i|\omega) \quad [A.41]$$
$$f(\tau_i^2|\bar{\boldsymbol{x}}_i,\sigma_n^2,a,z_i) \propto f(\bar{\boldsymbol{x}}_i|\tau_i^2,z_i,\sigma_n^2)f(\tau_i^2|a) \quad [A.42]$$
$$f(a|\boldsymbol{\tau}^2) \propto f(\boldsymbol{\tau}^2|a)f(a) \quad [A.43]$$
$$f(\sigma_n^2|\boldsymbol{Y},\bar{\boldsymbol{X}},\boldsymbol{\tau}^2,\boldsymbol{z},\rho) \propto f(\boldsymbol{Y}|\bar{\boldsymbol{X}},\sigma_n^2,\rho)f(\bar{\boldsymbol{X}}|\boldsymbol{\tau}^2,\boldsymbol{z},\sigma_n^2)f(\sigma_n^2) \quad [A.44]$$
$$f(\omega|\boldsymbol{z}) \propto f(\boldsymbol{z}|\omega)f(\omega) \quad [A.45]$$
$$f(\rho|\boldsymbol{Y},\bar{\boldsymbol{X}},\sigma_n^2) \propto f(\boldsymbol{Y}|\bar{\boldsymbol{X}},\sigma_n^2,\rho)f(\rho) \quad [A.46]$$

A.3.2 Conditional distributions

Due to the fact that the likelihood and the priors of this model are the same that were used in the previous model setting $v_i = 1$ and changing $\boldsymbol{H}$ for $\bar{\boldsymbol{H}}(\rho)$ and $\boldsymbol{X}$ for $\bar{\boldsymbol{X}}$, the derivations of all the conditional distributions of the model parameters and hyperparameters are identical to the ones of the previous model. The only exception to this being the newly introduced parameter ρ.

Conditional distribution of ρ

The conditional distribution of ρ is

$$f(\rho|\boldsymbol{Y},\bar{\boldsymbol{X}},\sigma_n^2) \propto f(\boldsymbol{Y}|\bar{\boldsymbol{X}},\sigma_n^2,\rho)f(\rho) \quad \propto \exp\Big(-\frac{||\bar{\boldsymbol{H}}(\rho)\bar{\boldsymbol{X}}-\boldsymbol{Y}||^2}{2\sigma_n^2}\Big)1_{[\rho_{min},\rho_{max}]} \tag{A.47}$$

BIBLIOGRAPHY

[1] J. C. Moreira Bermudez, M. Holsbach Costa, M. H. Maruo, and F. A. Marengo Rodriguez, "DynBrain – EEG dynamic image reconstruction and segmentation for brain tissue characterization," Tech. Rep., Universidad Federal de Santa Catarina, 2013.

[2] F. Costa, H. Batatia, L. Chaari, and J.-Y. Tourneret, "Sparse EEG Source Localization using Bernoulli Laplacian Priors," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 12, pp. 2888–2898, 2015.

[3] 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 (CAMSAP'15)*, Cancun, Mexico, 2015.

[4] F. Costa, H. Batatia, T. Oberlin, C. D'Giano, and J.-Y. Tourneret, "Bayesian EEG source localization using a structured sparsity prior," *NeuroImage*, vol. 144, pp. 142–152, jan. 2017.

[5] F. Costa, H. Batatia, T. Oberlin, and J.-Y. Tourneret, "A partially collapsed gibbs sampler with accelerated convergence for EEG source localization," in *Proc. of IEEE Workshop on Stat. Sig. Proc. (SSP'16)*, Palma de Mallorca, Spain, 2016.

[6] F. Costa, H. Batatia, T. Oberlin, and J.-Y. Tourneret, "Skull Conductivity Estimation for EEG Reconstruction," *IEEE Signal Process. Lett.*, to be published.

[7] A. S. Fauci et al., *Harrison's principles of internal medicine*, vol. 2, McGraw-Hill, Medical Publishing Division, 2008.

[8] R. Flink, B. Pedersen, A. Guekht, K. Malmgren, R. Michelucci, B. Neville, F. Pinto, U. Stephani, and C. Özkara, "Guidelines for the use of eeg methodology in the diagnosis of epilepsy," *Acta Neurologica Scandinavica*, vol. 106, no. 1, pp. 1–7, 2002.

[9] J. De Munck, B. Van Dijk, and H. Spekreijse, "Mathematical dipoles are adequate to describe realistic generators of human brain activity," *IEEE Transactions on Biomedical Engineering*, vol. 35, no. 11, pp. 960–966, 1988.

[10] A. B. Usakli, "Improvement of eeg signal acquisition: An electrical aspect for state of the art of front end," *Computational intelligence and neuroscience*, vol. 2010, pp. 12, 2010.

[11] H. W. Ott and H. W. Ott, *Noise reduction techniques in electronic systems*, vol. 442, Wiley New York, 1988.

[12] D. O. Olguın, F. Bouchereau, and S. Martınez, "Adaptive notch filter for eeg signals based on the lms algorithm with variable step-size parameter," in *Proceedings of the 39th International Conference on Information Sciences and Systems*, 2005.

[13] M. H. Costa and M. C. Tavares, "Removing harmonic power line interference from biopotential signals in low cost acquisition systems," *Computers in Biology and Medicine*, vol. 39, no. 6, pp. 519–526, 2009.

[14] D. M. White and C. A. Van Cott, "Eeg artifacts in the intensive care unit setting," *American journal of electroneurodiagnostic technology*, vol. 50, no. 1, pp. 8–25, 2010.

[15] H. Becker, L. Albera, P. Comon, M. Haardt, G. Birot, F. Wendling, M. Gavaret, C.-G. Bénar, and I. Merlet, "EEG extended source localization: tensor-based vs. conventional methods," *NeuroImage*, vol. 96, pp. 143–157, 2014.

[16] H. Becker, P. Comon, L. Albera, M. Haardt, and I. Merlet, "Multi-way space–time–wave-vector analysis for EEG source separation," *Signal Processing*, vol. 92, no. 4, pp. 1021–1031, 2012.

[17] H. Becker, *Denoising, separation and localization of EEG sources in the context of epilepsy*, Ph.D. thesis, Université Nice Sophia Antipolis, 2014.

[18] 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.

[19] C. J. Stok, *The inverse Problem in EEG and MEG with Application to Visual Evoked Responses*, Ph.D. thesis, Univ. Twente, Enschede, The Netherlands, 1986.

[20] B. He, T. Musha, Y. Okamoto, S. Homma, Y. Nakajima, and T. Sato, "Electric dipole tracing in the brain by means of the boundary element method and its accuracy," *IEEE Trans. Biomed. Eng.*, vol. 34, no. 6, pp. 406–414, 1987.

[21] C. R. Johnson, *Numerical methods for bioelectric field problems*, Boca Rato (FL): CRC Press, 1995.

[22] F. Marino, E. Halgren, J.-M. Badier, M. Gee, and V. Nenov, "A finite difference model of electric field propagation in the human head: Implementation and validation," in *Proc. of the 19th Annual Northeast Bioengineering Conference*, New Jersey, USA, 1993, IEEE, pp. 82–85.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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. 114020–114043, 2014.

[31] 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.

[32] 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.

[33] G. Birot, L. Albera, F. Wendling, and I. Merlet, "Localization of extended brain sources from EEG/MEG: the ExSo-MUSIC approach," *NeuroImage*, vol. 56, no. 1, pp. 102–113, 2011.

[34] R. D. Pascual-Marqui, "Review of methods for solving the EEG inverse problem," *Int. J. Bioelectromagn.*, vol. 1, no. 1, pp. 75–86, 1999.

[35] 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.

[36] 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.

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

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

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

[40] M. Kowalski, K. Siedenburg, and M. Dorfler, "Social sparsity! neighborhood systems enrich structured shrinkage operators," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2498–2511, 2013.

[41] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: from Gaussian mixture models to structured sparsity," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2481–2499, 2012.

[42] J. Huang, T. Zhang, and D. Metaxas, "Learning with structured sparsity," *J. Mach. Learn. Res.*, vol. 12, pp. 3371–3412, 2011.

[43] 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.

[44] 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.

[45] 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," *Ann. Appl. Stat.*, vol. 5, no. 2B, pp. 1207–1228, 2011.

[46] 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.

[47] A. Sorrentino, A. M. Johansen, J. A. Aston, T. E. Nichols, W. S. Kendall, et al., "Dynamic filtering of static dipoles in magnetoencephalography," *Ann. Appl. Stat.*, vol. 7, no. 2, pp. 955–988, 2013.

[48] 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.

[49] 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.

[50] 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.

[51] 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.

[52] 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.

[53] H.-J. Huppertz, S. Hoegg, C. Sick, C. Lücking, J. Zentner, A. Schulze-Bonhage, and R. Kristeva-Feige, "Cortical current density reconstruction of interictal epileptiform activity in temporal lobe epilepsy," *Clin. Neurophys.*, vol. 112, no. 9, pp. 1761–1772, 2001.

[54] Y. Wang and J. Gotman, "The influence of electrode location errors on EEG dipole source localization with a realistic head model," *Clin. Neurophys.*, vol. 112, no. 9, pp. 1777–1780, 2001.

[55] Z. A. Acar and S. Makeig, "Effects of forward model errors on EEG source localization," *Brain Topogr.*, vol. 26, no. 3, pp. 378–396, 2013.

[56] S. Vallaghé and M. Clerc, "A global sensitivity analysis of three-and four-layer EEG conductivity models," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 988–995, 2009.

[57] B. Vanrumste, G. Van Hoey, R. Van de Walle, M. D'havé, I. Lemahieu, and P. Boon, "Dipole location errors in electroencephalogram source analysis due to volume conductor model errors," *Med. Biol. Eng. Comput.*, vol. 38, no. 5, pp. 528–534, 2000.

[58] N. G. Gençer and C. E. Acar, "Sensitivity of EEG and MEG measurements to tissue conductivity," *Phys. Med. Biol.*, vol. 49, no. 5, pp. 701, 2004.

[59] L. Geddes and L. Baker, "The specific resistance of biological material - a compendium of data for the biomedical engineer and physiologist," *Med. Biol. Eng. Comput.*, vol. 5, no. 3, pp. 271–293, 1967.

[60] R. Hoekema, G. Wieneke, F. Leijten, C. Van Veelen, P. Van Rijen, G. Huiskamp, J. Ansems, and A. Van Huffelen, "Measurement of the conductivity of skull, temporarily removed during epilepsy surgery," *Brain Topogr.*, vol. 16, no. 1, pp. 29–38, 2003.

[61] S. Gonçalves, J. C. De Munck, J. Verbunt, F. Bijma, R. M. Heethaar, F. Lopes da Silva, et al., "In vivo measurement of the brain and skull resistivities using an EIT-based method and realistic models for the head," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 6, pp. 754–767, 2003.

[62] Y. Lai, W. Van Drongelen, L. Ding, K. Hecox, V. Towle, D. Frim, and B. He, "Estimation of in vivo human brain-to-skull conductivity ratio from simultaneous extra-and intra-cranial electrical potential recordings," *Clin. Neurophys.*, vol. 116, no. 2, pp. 456–465, 2005.

[63] D. Gutiérrez, A. Nehorai, and C. H. Muravchik, "Estimating brain conductivities and dipole source signals with EEG arrays," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 12, pp. 2113–2122, 2004.

[64] G. Şengül and U. Baysal, "An extended Kalman filtering approach for the estimation of human head tissue conductivities by using EEG data: a simulation study," *Physiol. Meas.*, vol. 33, no. 4, pp. 571, 2012.

[65] S. Vallaghé, M. Clerc, and J.-M. Badier, "In vivo conductivity estimation using somatosensory evoked potentials and cortical constraint on the source," in *Proc. IEEE 4th Int. Symp. Biomed. Imagi. (ISBI)*, Washington D.C., USA, 2007, pp. 1036–1039.

[66] S. Lew, C. H. Wolters, A. Anwander, S. Makeig, and R. S. MacLeod, "Improved EEG source analysis using low-resolution conductivity estimation in a four-compartment finite element head model," *Hum. Brain Mapp.*, vol. 30, no. 9, pp. 2862–2878, 2009.

[67] S. Lew, C. Wolters, A. Anwander, S. Makeig, and R. MacLeod, "Low resolution conductivity estimation to improve source localization," in *International Congress Series*. Elsevier, 2007, vol. 1300, pp. 149–152.

[68] 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.

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

[70] C. P. Robert, "Simulation of truncated normal variables," *Statistics and computing*, vol. 5, no. 2, pp. 121–125, 1995.

[71] 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.

[72] 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.

[73] 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.

[74] 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.

[75] 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.

[76] 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.

[77] 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.

[78] V. A. Morozov, "On the solution of functional equations by the method of regularization," *Soviet Math. Dokl.*, vol. 7, pp. 414–417, 1966.

[79] 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.

[80] 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.

[81] 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.

[82] 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.

[83] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Statist. Soc.*, vol. 68, no. 1, pp. 49–67, 2006.

[84] 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.

[85] 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.

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

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

[88] 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.

[89] E. Marinari and G. Parisi, "Simulated tempering: a new Monte Carlo scheme," *Europhy. Lett.*, vol. 19, no. 6, pp. 451–458, 1992.

[90] T. F. Oostendorp, J. Delbeke, and D. F. Stegeman, "The conductivity of the human skull: results of in vivo and in vitro measurements," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 11, pp. 1487–1492, 2000.

[91] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*. IEEE, 2011, pp. 2168–2172.

[92] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Appl. Opt.*, vol. 47, no. 10, pp. B44–B51, Apr 2008.

[93] R. M. Willett, M. F. Duarte, M. A. Davenport, and R. G. Baraniuk, "Sparsity and structure in hyperspectral imaging : Sensing, reconstruction, and target detection," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 116–126, Jan 2014.

[94] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Applied and Numerical Harmonic Analysis. Springer New York, 2013.

[95] J. Fessler, "Image reconstruction," in *Image Reconstruction*, chapter 1, pp. 1–18. 2016.

[96] Y. Mejía, F. Costa, H. Argüello, J.-Y. Tourneret, and H. Batatia, "Bayesian reconstruction of hyperspectral images by using compressed sensing measurements and a local structured prior," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, New Orleans, USA, 2017.

[97] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal processing magazine*, vol. 16, no. 6, pp. 22–38, 1999.