Universidade de São Paulo
Instituto de Física

Esparsidade Estruturada em Reconstrução de Fontes de EEG

André Biasin Segalla Francisco

Orientador: Prof. Dr. Nestor Felipe Caticha Alfonso

Dissertação de mestrado apresentada ao Instituto de Física da Universidade de São Paulo, como requisito parcial para a obtenção do título de Mestre em Ciências.

Banca Examinadora:

Prof. Dr. Nestor Felipe Caticha Alfonso (IFUSP)

Prof. Dr. Luiz Antonio Baccalá (POLI/USP)

Prof. Dr. Raphael Yokoingawa de Camargo (UFABC)

São Paulo
2018

FICHA CATALOGRÁFICA
Preparada pelo Serviço de Biblioteca e Informação
do Instituto de Física da Universidade de São Paulo

Francisco, André Biasin Segalla

Esparsidade estruturada em reconstrução de fontes de EEG. São Paulo, 2018.

Dissertação (Mestrado) – Universidade de São Paulo. Faculdade de Educação, Instituto de Física, Instituto de Química e Instituto de Biociências

Orientador: Prof. Dr. Nestor Felipe Caticha Alfonso
Área de Concentração: Física Geral

Unitermos: 1. Física – Estudo e Ensino; 2. Eletroencefalografia; 3. Magnetoencefalografia; 4. Física Médica.

USP/IF/SBI-003/2018

University of São Paulo
Physics Institute

Structured Sparsity in EEG Source Reconstruction

André Biasin Segalla Francisco

Supervisor: Prof. Dr. Nestor Felipe Caticha Alfonso

Dissertation submitted to the Physics Institute of the
University of São Paulo in partial fulfillment of the re-
quirements for the degree of Master of Science.

Examining Committee:

Prof. Dr. Nestor Felipe Caticha Alfonso (IFUSP)

Prof. Dr. Luiz Antonio Baccalá (POLI/USP)

Prof. Dr. Raphael Yokoingawa de Camargo (UFABC)

São Paulo
2018

Contents

1	Introduction	5
1.1	Outline	10
1.2	EEG generators	11
1.2.1	Action Potential and Postsynaptic Potential	11
1.3	Maxwell's equations	13
1.4	Source Model - The Current Dipole	15
2	Forward Problem	18
2.1	Formal Solution	19
2.2	Border Element Method	22
3	Inverse Problem	29
3.1	Introduction	29
3.2	A Brief Review of Inverse Methods and the Wiener Filter Estimate	30
3.3	A Bayesian Approach of the Inverse Problem	32
3.4	Sparse Bayesian Learning (SBL)	33
3.5	Inclusion of structures as degrees of freedom into the SBL framework: Structured Sparsity	36
3.5.1	Simple example of inclusion of a structure into the SBL framework	38
3.6	Multiscale Approach: How to include degrees of freedom of several scales simultaneously	41
4	Simulations and Results	44
4.1	Experiment 1 - Reconstruction of Active Cortical Areas of Multiple Scales	50
4.2	Experiment 2 - Reconstruction of Focal Activities through inclusion of ap- propriate structures	67
5	Conclusion and further work	83
	Appendices	87
A	Quasi-Static Approximation of the Maxwell's equations in head tissues	88

<i>CONTENTS</i>	2
B Champagne Algorithm	90
B.1 Update equation for z	91
B.2 Update equation for γ	92
C Origin of Sparsity in the Evidence Maximization Framework	93
D Bayesian Model Selection	96
E Expectation Maximization (EM) Algorithm	98
E.1 Derivation of the EM equations for a general case	98
E.1.1 Update Equation for Γ	100
E.1.2 Update Equation for Λ	101
E.1.3 EM Algorithm	102
E.2 EM algorithm for the MNE estimate	103
F K-Means Clustering	106
G Woodbury Identity	107

Resumo - Esparsidade Estruturada em Reconstrução de Fontes de EEG

Neuroimagem funcional é uma área da neurociência que visa o desenvolvimento de diversas técnicas para mapear a atividade do sistema nervoso e esteve sob constante desenvolvimento durante as últimas décadas devido à sua grande importância para aplicações clínicas e pesquisa. Técnicas usualmente utilizadas, como imagem por ressonância magnética funcional (fMRI) e tomografia por emissão de pósitrons (PET) têm ótima resolução espacial ($\sim mm$), mas uma resolução temporal limitada ($\sim s$), impondo um grande desafio para nossa compreensão a respeito da dinâmica de funções cognitivas mais elevadas, cujas oscilações podem ocorrer em escalas temporais muito mais finas ($\sim ms$). Tal limitação ocorre pelo fato destas técnicas medirem respostas biológicas lentas que são correlacionadas de maneira indireta com a atividade elétrica cerebral. As duas principais técnicas capazes de superar essa limitação são a Eletro- e Magnetoencefalografia (EEG/MEG), que são técnicas não invasivas para medir os campos elétricos e magnéticos no escalpo, respectivamente, gerados pelas fontes elétricas cerebrais. Ambas possuem resolução temporal na ordem de milissegundo, mas tipicamente uma baixa resolução espacial ($\sim cm$) devido à natureza mal posta do problema inverso eletromagnético. Um imenso esforço vem sendo feito durante as últimas décadas para melhorar suas resoluções espaciais através da incorporação de informação relevante ao problema de outras técnicas de imagens e/ou de vínculos biologicamente inspirados aliados ao desenvolvimento de métodos matemáticos e algoritmos sofisticados. Neste trabalho focaremos em EEG, embora todas técnicas aqui apresentadas possam ser igualmente aplicadas ao MEG devido às suas formas matemáticas idênticas. Em particular, nós exploramos **esparsidade** como uma importante restrição matemática dentro de uma abordagem Bayesiana chamada **Aprendizagem Bayesiana Esparsa** (SBL), que permite a obtenção de soluções únicas significativas no problema de reconstrução de fontes. Além disso, investigamos como incorporar diferentes estruturas como graus de liberdade nesta abordagem, que é uma aplicação de **esparsidade estruturada** e mostramos que é um caminho promissor para melhorar a precisão de reconstrução de fontes em métodos de imagens eletromagnéticos.

Palavras-chave: Eletroencefalografia; Problema Inverso; Esparsidade Estruturada; Aprendizagem Bayesiana Esparsa.

Abstract - Structured Sparsity in EEG Source Reconstruction

Functional Neuroimaging is an area of neuroscience which aims at developing several techniques to map the activity of the nervous system and has been under constant development in the last decades due to its high importance in clinical applications and research. Common applied techniques such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) have great spatial resolution ($\sim mm$), but a limited temporal resolution ($\sim s$), which poses a great challenge on our understanding of the dynamics of higher cognitive functions, whose oscillations can occur in much finer temporal scales ($\sim ms$). Such limitation occurs because these techniques rely on measurements of slow biological responses which are correlated in a complicated manner to the actual electric activity. The two major candidates that overcome this shortcoming are Electro- and Magnetoencephalography (EEG/MEG), which are non-invasive techniques that measure the electric and magnetic fields on the scalp, respectively, generated by the electrical brain sources. Both have millisecond temporal resolution, but typically low spatial resolution ($\sim cm$) due to the highly ill-posed nature of the electromagnetic inverse problem. There has been a huge effort in the last decades to improve their spatial resolution by means of incorporating relevant information to the problem from either other imaging modalities and/or biologically inspired constraints allied with the development of sophisticated mathematical methods and algorithms. In this work we focus on EEG, although all techniques here presented can be equally applied to MEG because of their identical mathematical form. In particular, we explore **sparsity** as a useful mathematical constraint in a Bayesian framework called **Sparse Bayesian Learning** (SBL), which enables the achievement of meaningful unique solutions in the source reconstruction problem. Moreover, we investigate how to incorporate different structures as degrees of freedom into this framework, which is an application of **structured sparsity** and show that it is a promising way to improve the source reconstruction accuracy of electromagnetic imaging methods.

Keywords: Eletroencephalography; Inverse Problem; Structured Sparsity; Sparse Bayesian Learning.

Chapter 1

Introduction

EEG/MEG (Electro/Magnetoencephalography) are non invasive techniques to infer brain activity from the external electromagnetic field on the scalp due to intercranial electric sources. In particular, EEG measures electric potential differences and MEG the magnetic field. In practice, EEG is much more common in clinical applications than MEG mainly because of its low cost and portability, whereas MEG requires superconducting sensors (SQUID - Superconducting Quantum Interference Device) to measure the extremely weak biomagnetic brain signals ($\sim fT$) and an advanced isolation setting [23]. EEG common clinical applications include diagnosis of epilepsy, sleep disorders, depth of anesthesia, coma and brain death. Now, both EEG and MEG can be used as neuroimaging techniques through estimation of the underlying generating sources from the measured bioelectromagnetic signals, which translates into the so called source reconstruction or electromagnetic inverse problem (EIP). The limitation of the information that we are able to extract from both of these techniques is such that the EIP is typically a severely ill posed mathematical problem in the sense that there are many radically different source configurations that explain the measured signals equally well, which in turn means that suitable constraints, either biologically and/or mathematically inspired, have to be taken into account to get meaningful unique solutions. The spatial resolution of these techniques tends to be rather poor ($\sim cm$) compared to the typical resolution achieved ($\sim mm$) by commonly used functional imaging techniques such as functional magnetic resonance imaging (fMRI) or positron emission tomography (PET). Nevertheless, there has been a huge effort on improving the spatial resolution of EEG/MEG in the last decades because they offer a unique window into the brain dynamics through their high temporal resolution ($\sim ms$), which can't be achieved by any other technique because they typically rely on measures of slow biological responses. Such improvement could have a great impact on our understanding of cognitive processes, consciousness and in applications such as brain computer interfaces (BCI).

Despite the fact that MEG requires much more resources than EEG, there has been

considerable interest in research in MEG mainly because its signals are much less affected by the head tissues properties (volume conduction effects) [15] than EEG, which is a central feature in source reconstruction and severely limits EEG spatial resolution [15]. For instance, EEG signals are strongly affected (smeared) by the very high brain to skull conductivity ratio (BSCR), which greatly varies between subjects and across age, and is generally very hard to access in vivo. On the contrary, MEG signals are almost unaffected by it and therefore access to this information is not required. This limitation lead to a lot of ongoing research focusing on either trying to indirectly estimate a subject's BSCR [1] or on directly extracting related information through other imaging techniques [16, 27, 28]. On the other hand, EEG can detect signals from either radial and tangential sources on the cortex, while MEG is almost blind to the former [15]. In general, EEG and MEG are therefore seen as complementary techniques [22] and thus the combination of them can effectively increase source reconstruction accuracy. In this work we will focus on EEG, but all formalism of the inverse problem explored here can be equally applied to MEG since both problems have the same form. As stated above, the EIP is an ill posed problem, which means that we have to add information in order to restrict the solution space. The first information we add are the head tissues geometric information acquired with high resolution structural MRI from a sample subject. More specifically, the geometric relevant information in this problem are the boundary surfaces between head tissues with different electrical properties among themselves, such as brain, cerebrospinal fluid (CSF), skull and scalp and also the surface where we will place the electrical sources, which is chosen to be the inner surface of the cortex, called white surface. A representation of the boundary surfaces and the white matter are depicted in Figures 1.1 and 1.2 respectively.

Another relevant medium information we need are the conductivities of each of these tissues, which we set to constant values inside each of them for simplicity, although in reality the tissues electric properties are anisotropic and not even truly homogeneous, which leads to important modelling errors in the EEG, as already mentioned.

The brain electrical sources are modelled as current dipoles placed over the whole white surface and perpendicularly oriented to it as means of approximating pyramidal neurons structures in the cortex, which are believed to be the greatest contributors to the EEG signals [22] and will be described below in more detail. Once we fixed the positions and directions of the current dipoles, the only remaining quantity to be determined are their amplitudes and through the use of Maxwell's equations we arrive at a linear equation known as the **forward equation**, which establishes the connection between the source amplitudes and the signals:

Figure 1.1: *Sample BEM surfaces corresponding to inner skull, outer skull and outer skin together with the pial surface (outer cortex surface) reconstructed from high resolution MRI and available in the MNE software [13, 14].*

$$V = GJ + \xi, \quad (1.1)$$

where $V \in \mathbb{R}^{M \times T}$ is the matrix of the measured potential differences on the scalp with each row corresponding to one of the M electrodes and each column corresponding to a set measurements at a specific time instant of a total of T instants, $J \in \mathbb{R}^{N \times T}$ are the source amplitudes over the time course T , $G \in \mathbb{R}^{M \times N}$ is the so called lead field matrix, which connects J with V and contains the relevant head and sensor information, and $\xi \in \mathbb{R}^{M \times T}$ is the noise, which includes sensor noise and also biological artifacts such as eye blinking, muscle movements, heart beating or even brain activity not related to the activity under investigation. Now, in a realistic case we have typically $M \ll N$, which means that the dimensionality of the source space ($\sim 10^4 - 10^5$) is much higher than the dimensionality of the signal space ($\sim 10^1 - 10^2$), revealing the undercompleteness of the EIP of estimating J given G and V . Several regularization techniques have been used over the last decades to constrain the solution space by enforcing desired solution properties and in this work we focus on **sparsity**. Sparse solutions and representations are widely studied in areas such as Machine Learning, Signal Processing and Image Reconstruction and in our case it relates to situations where typically only small regions of the cortex

Figure 1.2: *Fine representation of a sample white surface with 312273 vertices provided by the MNE software.*

relative to its whole extension are active during some specific task or experiment, such as in brain responses to auditori and visual stimuli or in investigations of epileptic foci, which in our problem translates into situations where J is row-sparse, meaning that only relatively few rows are different from zero. We use a Bayesian approach to enforce sparsity called **Sparse Bayesian Learning** (SBL) [43, 44, 49, 50, 51, 52], which also allows us to easily include arbitrary structures of the original source space as degrees of freedom. This is important because regions of different shapes and extensions of the cortex activate at a given moment rather than isolated tiny regions as represented by our current dipoles. Inclusion of structures in a sparse model is known as **structured sparsity** and it has not been used in the EEG/MEG source reconstruction problem until recently [6, 25, 38].

In this work we investigate the inclusion of structures through two numerical experiments to investigate its impact on the source reconstruction accuracy under the SBL framework. In the first one, we define several regions of different characteristic sizes as degrees of freedom corresponding to different scales of resolution of our space (Figure 1.3) and try to recover them from simulated corrupted signals under different noise levels. These regions were defined randomly using the K-Means Clustering Algorithm [31], although in real applications they should be defined according to anatomical or functional information, such as the Brodmann areas [5] as used in [38]. Moreover, we use the degrees of freedom of all scales **simultaneously** to reflect a situation where we do not know the typical extension of the activated areas under investigation. Another approach which deals with multiple scales is related to renormalization group ideas such as in [2, 7], where a step wise procedure is applied to go recursively from coarser to finer scales of resolution of the source space. More specifically, a solution from a coarser scale obtained through a Bayesian learning step is incorporated into the next finer scale through a **Backward Renormalization** procedure developed in [7] as prior information. These two steps are

repeated until the solution in the finest desired scale of resolution is achieved, and despite the great success of this method in reconstructing localized sources, it has not yet been extended to reconstruct more general regions of greater characteristic sizes. Such an approach was attempted in this work by, for example, simulating a signal from an arbitrary region from an arbitrary scale from Figure 1.3 and trying to recover it from the former corrupted with noise, but by choosing regions of coarser scales as the simulated activities, they could typically not be recovered. In other words, information was inevitably lost by going from coarser to finer scales. A solution to this problem was therefore to avoid the information transfer step between the scales of resolutions by considering all regions of all scales as degrees of freedom **simultaneously** and to use the SBL framework to find the sparsest combination of them to best describe the simulated noisy signals. The key idea here is that the solution at a fixed time instant is a sparse vector whose components reflect regions of activity with almost arbitrary characteristic sizes, so that all scales are treated on an equal footing, thus appropriately reflecting our lack of knowledge of the extents of the involved activities. Promising results were obtained through this approach as regions of all scales could be successfully recovered for different noise levels, as discussed in more detail in chapter 4. A major advance in this direction would be the development of an approach to automatically learn structures instead of predefining them as is done in other contexts such as in [47, 54].

In the second experiment, we tried to reconstruct focal activities from their simulated signals by including appropriate structures into our framework in a two step fashion. Under focal activity we mean an activity profile with a center dipole with highest amplitude and the surrounding dipoles with amplitudes falling with distance from it. In the first step we used large regions of activities around each dipole in the original source space as degrees of freedom to scan the whole space for regions with high probability of being active and in the second step we used structures of different sizes around each of the dipoles of the surviving regions of the first step to fine tune the simulated activity profile. The structures used in this experiment are represented in Figure 1.4 and a simulation example where we compared colormaps of the true simulated activity profile with the estimated profiles of our described approach (sSBL), without any inclusion of structures (SBL) and weighted Minimum Norm Estimate (wMNE) [15]. We notice obvious superiority of sSBL, indicating that inclusion of appropriate structures when prior information is available, which in this case is the focal nature of the sources, is a promising way to improve source reconstruction accuracy. As mentioned before, focal activities are widely studied in some contexts such as in focal epilepsies, where the identification of the epileptic focus is crucial for a presurgical evaluation of patients that do not respond to drug treatments.

Figure 1.3: *Scales of resolutions of the source space. Upper left: 20 regions. Upper right: 80 regions. Lower left: 320. Lower right: 1280 regions.*

1.1 Outline

The rest of this work is organized as follows. In the rest of this first chapter we give an introduction on the EEG/MEG generators and state the bioelectromagnetic problem through the Maxwell's equations. In chapter 2 we make a self contained review on the EEG forward problem solution in a realistic head model with the Boundary Element Method (BEM) [19]. The final goal of this chapter is to derive equation (1.1), which is the starting point of many EEG/MEG source reconstruction problems. In the third chapter we focus on the electromagnetic inverse problem and describe the Sparse Bayesian Learning framework (SBL) in some detail as a mean of achieving sparse solutions, its relation to Bayesian model selection and how to add structures into it. In chapter 4 we apply this formalism by conducting the two numerical experiments described above and show their results, where in the first one we investigate an approach to deal with multiple different scales of activity simultaneously and in the second one we try to reconstruct general focal activity profiles through inclusion of appropriate structures in a two step fashion of coarse scanning and fine tuning, and finally, in chapter 5, we summarize our

Figure 1.4: Structures defined around a center dipole (blue dot) used to reconstruct focal activities in the second experiment. Each vertex of the triangular lattice correspond to a current dipole and each filled blue region corresponds to a degree of freedom where all dipoles within are constrained to activate with the same amplitudes. Source reconstruction is conducted in two steps: a scanning step and a fine tuning step. In the scanning step large structures as shown above around each dipole of the original source space are used as degrees of freedom to scan the whole space for regions of probable activity and in the fine tuning step regions of different sizes around each dipole of the regions with high probability of being active obtained from the first step are used as degrees of freedom to fine tune the activity profile.

conclusions and indicate possible directions of further works.

1.2 EEG generators

Here we give a simple description to the EEG/MEG signal generators by highlighting the most important aspects of the brain's processes that are believed to be relevant to EEG/MEG signals. It should be kept in mind that it is by no means a detailed description of the brain's components and processes, which can be found in great detail in [18].

1.2.1 Action Potential and Postsynaptic Potential

The neuron is the responsible cell for processing and transmitting information in the nervous system through electric and chemical signals and for giving rise to the measured electromagnetic signals on the scalp. Its basics constituents are dendrites, cell body and axons, and there is about 86 billions neurons in the human nervous system. Communication between two neurons (Figure 1.6) starts at the so called presynaptic neuron (PreN) firing a so called **action potential**, which is a fast propagation of a perturbation of its membrane potential inducing a fast hyperpolarization followed by a repolarization and ultimately returning to its resting state. This perturbation propagates through the axon of the PreN reaching its synaptic terminals in the region called synapse (Figure 1.7). Neurotransmitters are then released from the terminals into the synaptic cleft and attach to the receptors of the postsynaptic neuron (PosN) changing the properties of specific ion channels, like K^+ , Na^+ and Cl^- channels, which induces a net ion current, called synaptic

Figure 1.5: Colormaps of an example simulation under noise level $SNR=10$ dB. (Upper Left) True activity profile. (Upper Right) *sSBL estimate*(SBL with focal structures as degrees of freedom as shown in Figure 1.4). (Lower Left) SBL with original fine source space as degrees of freedom. (Lower Right) Weighted Minimum Norm Estimate (*wMNE*).

current, through its membrane that changes its resting potential (~ -70 mV). This change is called **postsynaptic potential** (PSP) and it can be excitatory (EPSP), which means that the PosN depolarizes, increasing its probability of firing its own action potential when its membrane potential rises above a threshold of about -55 mV, allowing it to repeat this whole process and communicate with another neuron. On the other hand, the PSP can also be inhibitory (IPSP), which leads to a hyperpolarization of the PosN membrane potential, bringing it further from the firing threshold. A scheme of a typical profile of the action potential is depicted in Figure 1.8. It is believed that most contributors to the measurable EEG signals are due to many synchronously active PSPs on parallel aligned structures on the cortex, whose alignment allows for constructive superposition of the associated synaptic currents. On the other hand, it is believed that action potentials have minor contribution to the EEG signals due to its very short latency ($\sim ms$), which

decreases their probability of synchronization in comparison to PSPs [45], which have much larger latencies that can reach several hundreds of milliseconds. It should be emphasized that EEG signals can originate from areas of very different sizes of coherently active neurons over the cortex [22, 45], which can even overlap. Moreover, the complicated folded cortex structure allows the existence of silent sources, such as activated fissures on the cortex, also called sulci (Figure 1.9), where close dipoles of opposite orientations tend to cancel each others signals out, revealing thus another source of limitation of electromagnetic imaging techniques. Furthermore, deeper sources are well known to produce much weaker signals than superficial sources located in the gyri (Figure 1.9) because the electromagnetic fields fall rapidly with distance and therefore the latter are believed to produce the most significant part of the EEG signals.

Figure 1.6: *Scheme with presynaptic neuron at the left and postsynaptic neuron at the right, [56].*

Figure 1.7: *Scheme of the synapse [57].*

1.3 Maxwell's equations

Since the phenomena we want to describe are from electromagnetic nature, we have to start our modelling from the Maxwell equations:

Figure 1.8: *Scheme of action potential* [55]. Figure 1.9: *Representation of gyrus and sulcus of the human cortex along with the cortical dipoles. Activated dipoles on opposite walls of the sulcus tend to cancel each others signals out, being therefore a silent EEG source* [55].

$$\nabla \cdot \vec{D} = \rho, \quad (1.2)$$

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \quad (1.3)$$

$$\nabla \cdot \vec{B} = 0, \quad (1.4)$$

$$\nabla \times \vec{H} = \vec{J} + \frac{\partial \vec{D}}{\partial t}, \quad (1.5)$$

where ρ e $\vec{J}$ refer to the free charge density and free current density, respectively. Applying the divergence operator in (1.5) and using (1.2), we arrive at the continuity equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \vec{J} = 0. \quad (1.6)$$

When the medium is linear, we have $\vec{D} = \epsilon \vec{E}$ e $\vec{H} = \vec{B}/\mu$. Moreover, when the typical frequency is small enough, we can use the *quasi-static* approximation of Maxwell's equations, which consists of neglecting the terms with partial derivatives in time in (1.5) and (1.3), thereby decoupling the electric and magnetic fields, which will be described in the next section. Thus, we have $\nabla \times \vec{E} = 0$ and therefore we can write $\vec{E} = -\nabla \phi$, where $\phi(\vec{x})$ is a scalar potential. Furthermore, equation (1.6) becomes $\nabla \cdot \vec{J} = 0$, which means that there is no charge accumulation anywhere.

1.4 Source Model - The Current Dipole

As stated in the previous section, it is believed that most of the measured EEG signal are due to synchronous activity of synaptic potentials of big patches of neurons that are parallel aligned to each other and perpendicular oriented to the cortex surface, the so called pyramidal neurons [22]. The brain tissues, such as white matter, gray matter, cerebrospinal fluid (CSF), skull and scalp, can be considered as linear conductors and therefore the Ohm's law is valid [22]. It is common practice [15, 22, 23] to split the total current in the brain into two different terms $\vec{J} = \vec{J}_p + \vec{J}_v$, where $\vec{J}_p$ is the synaptic current, also called *primary current*, and $\vec{J}_v$ is the passive current in the rest of the brain tissues, called *secondary* or *volume* current, that results from the effect of the electric field on the charge carriers in the brain volume. Since we have a linear medium [22], we further use Ohm's Law $\vec{J}_v = \sigma \vec{E} = -\sigma \nabla \phi$, where σ is the medium conductivity, ϕ the electric potential and ∇ the nabla operator. The relation $\vec{E} = -\nabla \phi$ is valid in the quase-static approximation of the Maxwell's equations, which will be described later. Finally, from (1.6), we get

$$\nabla(\sigma \nabla \phi) = \nabla \cdot \vec{J}_p =: -s(\vec{x}), \quad (1.7)$$

where $s(\vec{x})$ is the so called current source. Note that (1.7) is the Poisson equation, having therefore the identical form of the better known equation $\nabla(\epsilon \nabla \phi) = -\rho(\vec{x})$, where $\rho(\vec{x})$ is the charge density and ϵ is the medium permissivity. If the medium conductivity is constant, we can write a particular solution of (1.7) as

$$\phi(\vec{x}) = \frac{1}{\sigma} \int G(\vec{x}, \vec{x}') s(\vec{x}') d^3 \vec{x}', \quad (1.8)$$

where G is the Green's function of the Poisson equation, given by $G(\vec{x}, \vec{x}') = 1/4\pi|\vec{x} - \vec{x}'|$ and the integral is over the whole space. Furthermore, if the characteristic dimension of an active patch of dipoles is small compared to its distance to an observation point, we can model its associated primary current as a dipole, called *current dipole*, which is a widely used source model in the EEG/MEG literature [15, 22, 23, 30, 37]:

$$\vec{J}_p(\vec{x}) = \vec{J} \delta(\vec{x} - \vec{x}_p), \quad (1.9)$$

where $\vec{J}$ dipole moment given by $\vec{J} = \int \vec{J}_p d^3 \vec{x}$, $\vec{x}_p$ is the dipole position and $\delta(x)$ is the Dirac delta function. Typical current magnitudes that generate signals are on the order of $10nAm$ and can be imagined to originate from several so called macro-columns [45], which are structures of about $1mm^3$ volume, where all neurons within are densely connected and tend to active coherently (Figure 1.11).

Figure 1.10: *Scheme of primary currents generated by pyramidal neurons in the cortex and their resulting electromagnetic fields. Adapted from [41].*

Figure 1.11: *Characteristic dimensions of a cortical macro-column [45].*

Chapter 2

Forward Problem

In this chapter we are going to derive the forward equation using the Border Element Method (BEM) following [10, 15, 19, 30]. The forward equation is the starting point for any source reconstruction attempt since it models the relationship between the electrical brain sources and scalp potentials.

The forward problem of EEG can be stated as follows: given a current source distribution inside the brain, calculate the electric potential distribution on the scalp. In other words, we have to solve (1.7) with appropriate boundary conditions. To do that, we first have to specify a volume conductor model of the head, and we are going to use one of the most common realistic head models used in the literature, as in [19, 23, 30]. We consider the head as a set of nested regions Ω_i , $i = 1, \dots, M + 1$, with respective constant conductivities: $\sigma(\vec{x}) = \sigma_i$ for $\vec{x} \in \Omega_i$. We denote by S_i the boundary surface between the volumes Ω_i e Ω_{i+1} and by $\vec{n}_i$ their respective outside normal oriented unit vectors. The head surface corresponds to S_M and the volume Ω_{M+1} therefore is a region of space around it, where the conductivity vanishes: $\sigma_{M+1} = 0$. Writing $\phi(\vec{x}) = \phi_i(\vec{x})$ for $\vec{x} \in \Omega_i$, we can write a Poisson equation for each region: $\nabla^2 \phi_i = -s(\vec{x})/\sigma_i$, $\vec{x} \in \Omega_i$. Furthermore we consider that both the potential and current density are continuous at each surface. This means that $\phi_i(\vec{x}) = \phi_{i+1}(\vec{x})$ and, by (1.7), $\sigma_i \partial_{n_i} \phi_i(\vec{x}) = \sigma_{i+1} \partial_{n_i} \phi_{i+1}(\vec{x})$ $\forall \vec{x} \in S_i$, $i = 1, \dots, M$, with $\partial_{n_i} := \vec{n}_i \cdot \nabla$. Note that since $\sigma_{M+1} = 0$, no current flows out of the head: $\partial_{n_M} \phi_M(\vec{x}) = 0$, $\forall \vec{x} \in S_M$, which corresponds to a Neumann boundary condition and thus defines the potential up to an arbitrary additive constant. Thus, we can impose $\langle \phi \rangle_M = 0$ [30], where $\langle \rangle_M$ corresponds to the average over the surface M. Finally we can state our forward problem as:

Figure 2.1: *Nested volumes Ω_i with their respective constant conductivities σ_i and outside unit normal vectors $\vec{n}_i$. Adapted from [30].*

$$\begin{aligned}
 \nabla^2 \phi_i(\vec{x}) &= -\frac{s(\vec{x})}{\sigma_i}, \quad \vec{x} \in \Omega_i, \quad i = 1, \dots, M. \\
 \phi_i(\vec{x}) &= \phi_{i+1}(\vec{x}), \\
 \sigma_i \partial_{n_i} \phi_i(\vec{x}) &= \sigma_{i+1} \partial_{n_i} \phi_{i+1}(\vec{x}), \quad \vec{x} \in S_i, \quad i = 1, \dots, M. \\
 \partial_{n_M} \phi(\vec{x}) &= 0, \quad \vec{x} \in S_{M+1}, \\
 \int_{S_M} \phi(\vec{x}) dS &= 0.
 \end{aligned} \tag{2.1}$$

In the next section will be presented a formal solution of (2.1) using the second Green identity [26], which results on an integral equation for the potential and then will be presented a method to solve it numerically.

2.1 Formal Solution

Let $\Omega \subset \mathbb{R}^3$ be a region in space with conductivity $\sigma(\vec{x})$, $\partial\Omega$ its border and $\vec{u} : \Omega \rightarrow \mathbb{R}^3$ a vector field.

From Gauss' law, we can write:

$$\int_{\Omega} (\nabla' \cdot \vec{u}) d^3 \vec{x}' = \int_{\partial\Omega} (\vec{u} \cdot \vec{n}') dS', \quad (2.2)$$

where ∇' is the nabla operator with respect to $\vec{x}'$ and $\vec{n}'$ is the unit outside normal oriented vector to the bounding surface $\partial\Omega$.

Let's partition the region Ω into $M+1$ nested subregions according to figure (2.1) with their respective unit normal vectors $\vec{n}_i$, $i = 1, \dots, M+1$. The M th surface represents the scalp and the $(M+1)$ th a surface around it. Using the following notation $\vec{u}_i := \vec{u}(\vec{x})$ for $\vec{x} \in \Omega_i$, we write (2.2) as

$$\sum_{i=1}^{M+1} \int_{\Omega_i} (\nabla' \cdot \vec{u}_i) d^3 \vec{x}' = \sum_{i=1}^{M+1} \int_{\partial\Omega_i} (\vec{u}_i \cdot \vec{n}'_i) dS'. \quad (2.3)$$

As $\partial\Omega_i = S_i \cup S_{i-1}$, $i > 0$ and $\partial\Omega_1 = S_1$, we have that the integral on the right-hand side of (2.3) can be written as

$$\int_{\partial\Omega_i} (\vec{u}_i \cdot \vec{n}'_i) dS' = \int_{S_i} (\vec{u}_i \cdot \vec{n}'_i) dS' - \int_{S_{i-1}} (\vec{u}_i \cdot \vec{n}'_{i-1}) dS', \quad (2.4)$$

where the second term on the right-hand side of (2.4) vanishes for $i = 1$. Thus, we rewrite (2.3) as:

$$\sum_{i=1}^{M+1} \int_{\Omega_i} (\nabla' \cdot \vec{u}_i) d^3 \vec{x}' = \sum_{i=1}^M \int_{S_i} (\vec{u}_i - \vec{u}_{i+1}) \cdot \vec{n}'_i dS' + \int_{S_{M+1}} \vec{u}_{M+1} \cdot \vec{n}'_{M+1} dS'. \quad (2.5)$$

Now we suppose that the conductivity on reach region is constant and denote $\sigma_i := \sigma(\vec{x})$, $\vec{x} \in \Omega_i$, $i = 1, \dots, M+1$.

Consider now $\vec{u}_i(\vec{x}') = G(\vec{x}, \vec{x}') \nabla'(\sigma_i \phi_i(\vec{x}')) - \sigma_i \phi_i(\vec{x}') \nabla' G(\vec{x}, \vec{x}')$, where $\vec{x}$ is an arbitrary fixed point inside Ω , $\phi_i(\vec{x}')$ is the electric potential inside Ω_i satisfying the Poisson's equation (1.7) and $G(\vec{x}, \vec{x}')$ its respective Green function:

$$G(\vec{x}, \vec{x}') = \frac{1}{4\pi|\vec{x} - \vec{x}'|}, \text{ with } \nabla'^2 G(\vec{x}, \vec{x}') = -\delta(\vec{x} - \vec{x}'), \quad (2.6)$$

where $\delta(\vec{x} - \vec{x}')$ is the Dirac delta function.

As boundary conditions we shall have continuity of both potential and current on the surfaces S_i :

$$\phi_i(\vec{x}) = \phi_{i-1}(\vec{x}), \quad (2.7a)$$

$$\sigma_i \nabla \phi_i(\vec{x}) \cdot \vec{n}_i = \sigma_{i-1} \nabla \phi_{i-1}(\vec{x}) \cdot \vec{n}_i, \quad \vec{x} \in S_i. \quad (2.7b)$$

Inserting (2.7), (2.6) and (1.7) in (2.5) and considering $\sigma_{M+1} = 0$, we obtain:

$$\sigma(\vec{x})\phi(\vec{x}) = \int_{\Omega} G(\vec{x}, \vec{x}') s(\vec{x}') d^3 \vec{x}' - \sum_{i=1}^M (\sigma_i - \sigma_{i+1}) \int_{S_i} \phi(\vec{x}') \partial_{n'_i} G dS', \quad (2.8)$$

with $\partial_{n'_i} G = \vec{n}'_i \cdot \nabla' G(\vec{x}, \vec{x}')$. We can rewrite the first term on the right-hand side of (2.8) as

$$\begin{aligned} \int_{\Omega} G(\vec{x}, \vec{x}') s(\vec{x}') d^3 \vec{x}' &= -\frac{1}{4\pi} \int_{\Omega} \frac{\nabla' \cdot \vec{J}_p(\vec{x}')}{|\vec{x} - \vec{x}'|} d^3 \vec{x}' = \frac{1}{4\pi} \int_{\Omega} \vec{J}_p(\vec{x}') \cdot \nabla' \left(\frac{1}{|\vec{x} - \vec{x}'|} \right) d^3 \vec{x}' - \\ &\quad \frac{1}{4\pi} \int_{\Omega} \nabla' \cdot \left(\frac{\vec{J}_p(\vec{x}')}{|\vec{x} - \vec{x}'|} \right) d^3 \vec{x}' \xrightarrow{0} = \frac{1}{4\pi} \int_{\Omega} \vec{J}_p(\vec{x}') \cdot \frac{(\vec{x} - \vec{x}')}{|\vec{x} - \vec{x}'|^3} d^3 \vec{x}', \end{aligned} \quad (2.9)$$

where we integrated by parts and used Gauss' theorem taking the integration volume to infinity where the integrand vanishes.

Now, note that term $\partial_{n'_i} G = \frac{(\vec{x} - \vec{x}')}{|\vec{x} - \vec{x}'|^3} \cdot \vec{n}'_i$ in (2.8) has a singularity when $\vec{x} \rightarrow \vec{x}'$, so when we choose $\vec{x} \in S_i$ in (2.8), we have to evaluate the integral on S_i in terms of its principal

value, whose result can be found in [46]:

$$\lim_{\vec{x}_0^\pm \rightarrow \vec{x} \in S} \int_S f(\vec{x}') \frac{(\vec{x}_0 - \vec{x}')}{|\vec{x}_0 - \vec{x}'|^3} \cdot \vec{n}' dS' = \pm 2\pi f(\vec{x}) + \int_S f(\vec{x}') \frac{(\vec{x} - \vec{x}')}{|\vec{x} - \vec{x}'|^3} \cdot \vec{n}' dS', \quad (2.10)$$

where the superscripts $\pm$ on $\vec{x}_0$ refers to points approximating from outside and from inside the surface, respectively.

Inserting (2.9) and (2.10) in (2.8) with $\vec{x}_0 \rightarrow \vec{x} \in S_k$, we finally obtain:

$$\frac{\sigma_k + \sigma_{k+1}}{2} \phi(\vec{x}) = \sigma_1 \phi_\infty(\vec{x}) - \frac{1}{4\pi} \sum_{i=1}^M (\sigma_i - \sigma_{i+1}) \int_{S_i} \phi(\vec{x}') \frac{(\vec{x} - \vec{x}')}{|\vec{x} - \vec{x}'|^3} \cdot \vec{n}'_i dS', \quad \vec{x} \in S_k, \quad (2.11)$$

where

$$\phi_\infty(\vec{x}) = \frac{1}{4\pi\sigma_1} \int_{\Omega} \vec{J}_p(\vec{x}') \cdot \frac{(\vec{x} - \vec{x}')}{|\vec{x} - \vec{x}'|^3} d^3\vec{x}'. \quad (2.12)$$

The term (2.12) can be interpreted as the electric potential due to charges in a volume of constant conductivity and the second term on the right-hand side of (2.11) is a correction term due to the conductivities discontinuities between the different regions.

2.2 Border Element Method

Equation (2.11) is a Fredholm integral equation of second kind and to solve it numerically for realistic head surfaces, we use the border element method (BEM) with the method of weighted residuals [19]. First, we note that equation (2.11) can be written as $L(\phi(\vec{x})) = \phi_\infty(\vec{x})$, where L is a linear operator given by

$$L(\phi(\vec{x})) = \frac{\sigma_k + \sigma_{k+1}}{2\sigma_1} \phi(\vec{x}) + \frac{1}{\sigma_1} \sum_{m=0}^M \Delta\sigma_m \mathcal{D}_{km} \phi, \quad (2.13)$$

where $\Delta\sigma_m = \sigma_m - \sigma_{m+1}$ and $\mathcal{D}$ is another linear operator defined by

$$\mathcal{D}_{km}\phi = \frac{1}{4\pi} \int_{S_m} \phi(\vec{x}') \frac{(\vec{x} - \vec{x}')}{|\vec{x} - \vec{x}'|^3} \cdot \vec{n}'_m dS', \quad \vec{x} \in S_k.$$

The method of the weighted residuals consists in choosing a weighting function $w(\vec{x})$ and solve the following problem:

$$\int (L(\phi(\vec{x}')) - \phi_\infty(\vec{x}')) w(\vec{x}') dS' = 0, \quad (2.14)$$

where the integration domain is the same as in (2.11). Defining the inner product of two functions f, g as $\langle f, g \rangle := \int f g dS'$, (2.14) can be written as

$$\langle w(\vec{x}), \phi_\infty(\vec{x}) \rangle = \langle w(\vec{x}), L(\phi(\vec{x})) \rangle. \quad (2.15)$$

Following the BEM method, the surfaces are discretized, usually in triangles, which we can represent as $S_k \rightarrow \bigcup_{l=1}^{N_k} \Delta_l^k$, for $k = 1, \dots, M$, where Δ_l^k is the l th triangle of the k th surface, whose total triangle numbers is N_k . We also write $w(\vec{x})$ e $\phi(\vec{x})$ as linear combinations of certain basis functions:

$$\phi(\vec{x}) = \sum_{j=1}^{N'_k} v_j^k \phi_j^k(\vec{x}), \quad (2.16a)$$

$$w(\vec{x}) = \sum_{j=1}^{N'_k} \beta_j^k w_j^k(\vec{x}), \quad \vec{x} \in S_k, \quad (2.16b)$$

where ϕ_j^k e w_j^k are the basis functions and v_j^k , β_j^k their respective coefficients, where k is an index which refers to the k th surface.

There are two broad classes as common choices of the basis functions w_i :

$$w_i^k(\vec{x}) = \delta(\vec{x} - \vec{x}_{ik}), \quad (\text{Collocation}) \quad (2.17a)$$

$$w_i^k(\vec{x}) = \phi_i^k(\vec{x}). \quad (\text{Garlekin}) \quad (2.17b)$$

In general, the Garlekin results in a smaller error in the forward problem, [19].

The basis functions ϕ_i^k are defined through their respective *nodal points* $\vec{x}$, where $\phi_i^k(\vec{x}) = 1$ and $\phi = 0$ in all others. The simplest choice is $\vec{x} = \vec{x}_{ik}$, where $\vec{x}_{ik}$ is the center of the triangle Δ_i^k and choosing $\phi(\vec{x}) = \phi(\vec{x}_{ik}) = v_i^k$, $\forall \vec{x} \in \Delta_i^k$, i.e., the potential is constant within each triangle of the discretized surface and therefore the number of basis functions and triangles coincide: $N_k = N'_k$. An increase in the methods precision is achieved by choosing the nodal points as the triangle vertices, such that for triangle Δ_i^k with vertices $\vec{x}_a, \vec{x}_b, \vec{x}_c$, where a, b and c are ordered according to the right hand rule, we have $\phi(x) = v_a^k \phi_a^k(\vec{x}) + v_b^k \phi_b^k(\vec{x}) + v_c^k \phi_c^k(\vec{x})$, for all $\vec{x} \in \Delta_i^k$, where

$$\phi_a^k(\vec{x}) = \frac{\vec{x} \cdot (\vec{x}_b \times \vec{x}_c)}{\vec{x}_a \cdot (\vec{x}_b \times \vec{x}_c)}, \quad \phi_b^k(\vec{x}) = \frac{\vec{x} \cdot (\vec{x}_c \times \vec{x}_a)}{\vec{x}_a \cdot (\vec{x}_b \times \vec{x}_c)}, \quad \phi_c^k(\vec{x}) = \frac{\vec{x} \cdot (\vec{x}_a \times \vec{x}_b)}{\vec{x}_a \cdot (\vec{x}_b \times \vec{x}_c)},$$

that is, the potential inside a triangle is given by a linear combination of the basis functions at its vertices. In this case a closed surface discretized in N_k triangles will have $N'_k = N_k/2 + 2$ basis functions, which corresponds to its total number of vertices.

Of course we could go on and improve our precision by choosing higher order approximations, but a linear one is sufficient for our purposes.

Inserting (2.16) in (2.15) and remembering that the basis elements are independent, we have

$$\langle w_j^k(\vec{x}), L(\phi(\vec{x})) \rangle = \langle w_j^k(\vec{x}), \phi_\infty(\vec{x}) \rangle, \quad j = 1, \dots, N'_k, \quad \vec{x} \in S_k. \quad (2.18)$$

Writing (2.18) explicitly, we obtain

$$\frac{\sigma_k + \sigma_{k+1}}{2\sigma_1} \sum_{i=1}^{N'_k} v_i^k \langle \phi_i^k, w_j^k \rangle + \frac{1}{\sigma_1} \sum_{m=1}^M \Delta\sigma_m \sum_{i=1}^{N'_m} v_i^m \langle \mathcal{D}_{km} \phi_i^m, w_j^k \rangle = \langle \phi_\infty, w_j^k \rangle, \quad (2.19)$$

$$j = 1, \dots, N_k.$$

Or, in matrix form:

$$\underbrace{\begin{pmatrix} b^1 \\ \vdots \\ b^M \end{pmatrix}}_{=:b} = \underbrace{\begin{pmatrix} \Theta_1 + D_{11} & D_{12} & \cdots & D_{1,M} \\ D_{21} & \Theta_2 + D_{12} & \cdots & D_{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ D_{M1} & D_{M2} & \cdots & \Theta_M + D_{M,M} \end{pmatrix}}_{=:H} \underbrace{\begin{pmatrix} v^1 \\ \vdots \\ v^M \end{pmatrix}}_v, \quad (2.20)$$

where

$$\begin{aligned} (\Theta_k)_{ij} &= \frac{\sigma_k + \sigma_{k+1}}{2\sigma_0} \langle \phi_i^k, w_j^k \rangle, \\ (D_{km})_{ij} &= \frac{\Delta\sigma_m}{\sigma_0} \langle \mathcal{D}_{km} \phi_i^m, w_j^k \rangle, \\ (b^k)_i &= \langle \phi_\infty, w_i^k \rangle, \quad (v^k)_i = v_i^k, \end{aligned} \quad (2.21)$$

or yet briefly:

$$Hv = b, \quad (2.22)$$

where H is a $(MN') \times (MN')$ matrix with $N' = \sum_k N'_k$. Analytical expressions for the integrals in (2.21) can be found in [34, 48].

It is a commonly known fact that in problems with Neumann's boundary conditions the solution is determined up to an additive constant, that is, if $\phi(\vec{x})$ is a solution, then $\phi(\vec{x}) + c$ is also a solution, with c constant. This means that matrix H of (2.22) has a zero eigenvalue, being therefore singular. To remove the singularity the matrix is deflated [19, 23, 37], which leads to an invertible matrix $\tilde{H} = H - \frac{1}{N} \mathbf{1}_N \mathbf{1}'_N$, where $\mathbf{1}_N$ is an N

dimensional vector with all components equal to 1. Thereby, we obtain:

$$v = H'b, \quad (2.23)$$

where $H' = \tilde{H}^{-1}$.

In practice we are interested in potential measurements on the head surface, which in our case corresponds to the M th surface, which means that we only have to deal with the coefficients $v^M = \sum_{k=1}^M H'_{Mk} b^k$.

Writing the primary current as a linear combination of N current dipoles with fixed directions $\{\hat{d}_i\}$ and positions $\{\vec{x}_i\}$: $\vec{J}_p(\vec{x}) = \sum_{i=1}^N J_i \hat{d}_i \delta(\vec{x} - \vec{x}_i)$ and inserting in (2.12), we obtain

$$\phi_\infty(\vec{x}) = \frac{1}{4\pi\sigma_1} \sum_{i=1}^N J_i \hat{d}_i \cdot \frac{(\vec{x} - \vec{x}_i)}{|\vec{x} - \vec{x}_i|^3} = \sum_{i=1}^N \kappa_i J_i, \quad (2.24)$$

where $\kappa_i = \frac{1}{4\pi\sigma_1} \frac{(\vec{x} - \vec{x}_i)}{|\vec{x} - \vec{x}_i|^3} \cdot \hat{d}_i$ and J_i is the amplitude of the i th current dipole. As $(b^k)_l = \langle \phi_\infty, w_l^k \rangle$ we write b^k as

$$b^k = \begin{pmatrix} b_1^k \\ \vdots \\ b_{N'_M}^k \end{pmatrix} = \sum_{i=1}^N \underbrace{\begin{pmatrix} \langle \kappa_i, w_1^k \rangle \\ \vdots \\ \langle \kappa_i, w_{N'_M}^k \rangle \end{pmatrix}}_{=: \eta_i^k} J_i = \sum_{i=1}^N \eta_i^k J_i. \quad (2.25)$$

Using the above result we write the coefficients v^M as

$$v^M = \sum_{i=1}^N \underbrace{\left(\sum_{k=1}^M H'_{Mk} \eta_i^k \right)}_{=: g_i^M} J_i = \sum_{i=1}^N g_i^M J_i. \quad (2.26)$$

Lastly, the surface potential is given by

$$\phi(\vec{x}) = \sum_{j=1}^{N'_m} v_j^M \phi_j^M = \varphi^M v^M = \sum_{i=1}^N \underbrace{\varphi^M g_i^M}_{=: \mathcal{G}_i} J_i = \sum_{i=1}^N \mathcal{G}_i J_i, \quad \vec{x} \in S_M, \quad (2.27)$$

where $\varphi^M = (\phi_1^M, \dots, \phi_{N'_m}^M)$. Notice that $\mathcal{G}_i$ depends on the potential measurement position $\vec{x}$ and on the i th dipole position $\vec{x}_i$ and direction $\hat{d}_i$.

In practice the electric potential is measured in discrete head positions corresponding to the electrode positions of the EEG system. Assuming we have S electrodes and denoting their positions as $\vec{x}_j$, $j = 1, \dots, S$ and defining $\phi(\vec{x}_j) =: V_j$, $\mathcal{G}_i(\vec{x}_j) =: G_{ij}$, we can write from (2.27), $V_j = \sum_{i=1}^N G_{ij} J_i$. In matrix notation we have

$$V = GJ, \quad (2.28)$$

where $V = (V_1, \dots, V_S)'$, $J = (J_1, \dots, J_N)'$ and G is the matrix with components $\{G\}_{ij} = G_{ij}$. G , which is the Green function of our problem, is called *lead field* or *gain matrix* in the EEG literature and contains all head and electrodes positions informations. Observe that equation (2.28) refers to a fixed time instant and we can immediately generalize it for a time period of T instants by considering matrices $V \in \mathbb{R}^{S \times T}$ and $J \in \mathbb{R}^{N \times T}$, where each column of them corresponds to a time instant. Furthermore, in real situations we have to consider noise to our measurements, which can be of instrumental, external or biological nature. Under biological noise is to be understood signals due to undesired biological sources, such as eye blinks, heart beat, head movements or even brain activities that are not under investigation in a given experiment. For instance, if an experiment is being conducted to observe the evoked response of auditory stimuli, any brain activity not related to it, such as due to visual stimuli, could be considered as noise or interference. To sum up, our forward equation can be written as

$$V = GJ + \xi, \quad (2.29)$$

where $V \in \mathbb{R}^{S \times T}$, $G \in \mathbb{R}^{S \times N}$, $J \in \mathbb{R}^{N \times T}$ and $\xi \in \mathbb{R}^{S \times T}$ is the noise matrix. Equation (2.29) is usually the starting point for most works on EEG/MEG source reconstruction

and will be addressed in next chapter, where **sparsity** will be explored as a key property to overcome its undercompleteness.

Chapter 3

Inverse Problem

3.1 Introduction

In this chapter we focus on the electromagnetic inverse problem, whose purpose is to estimate the brain activity from electromagnetic signals measured on the scalp. First we are going to give a very brief review of some basics about the inverse problem and derive the classical Wiener filter solution. Then we describe the inverse problem from a general Bayesian perspective, which includes common regularization methods as special cases. In the following section we introduce a Bayesian framework to enforce sparse solutions, called Sparse Bayesian Learning (SBL) and show its relation to Bayesian model selection. Sparsity is the key property we will add to our inverse problem to constrain our solution space, which reflects situations where relatively small regions of interest of the cortex are active during some task. Then we explain a way to include **structures** in our source space as degrees of freedom of our problem. This is important because despite the fact that we are considering situations where small regions of the cortex are active at a specific time, in real cases they usually have different forms and extensions. If we do not include any structures, our estimates tend to be composed of only few active dipoles of our original source space, which correspond to tiny active regions of the cortex, even in the case that the simulated signals are generated by active regions comprising of several hundreds of neighbouring active dipoles. This is because a signal generated from some arbitrary extended region can be usually well explained by radically different active source configurations due to the undercompleteness of the inverse problem and in particular from much sparser ones, which in turn will be favoured by our sparsity enforcing approach. Thus addition of arbitrary structures as degrees of freedom can overcome this limitation because an extended activated region, which corresponds to several degrees of freedom in our the original source space, can now be represented by just one new degree of freedom becoming therefore a sparse object. Addition of structures under a sparsity enforcing approach is known as **structured sparsity**, which is an important subject of

compressed sensing. We show how to add structures in the SBL framework through a very simple example and finally we describe an approach to deal with degrees of freedom corresponding to region of different sizes (or scales) **simultaneously**, which reflects situations where there is lack of information about characteristic sizes of cortical activated regions during some experiment.

The developed tools in this chapter will then be tested in two numerical experiments in the next chapter. In the first one, we will try to recover activated regions from a set of predefined regions of different sizes by considering them as degrees of freedom simultaneously in the SBL framework. In the second experiment we will try to reconstruct focal extended sources with addition of appropriate structures in a two step fashion and compare its performance with other common algorithms.

3.2 A Brief Review of Inverse Methods and the Wiener Filter Estimate

A large number of different inverse methods have been developed in the literature and they can be classified into three main types: parametric methods, scanning methods and imaging methods. Parametric approaches consist in assuming that the brain activity can be well described by a few current dipoles with unknown positions, orientations and amplitudes. This procedure is called dipole fitting and a major drawback is that it is usually hard to choose the number of dipoles to be fitted. Scanning or Beamforming methods [15, 29] perform spatial filtering on the sensor data to discriminate between signals arriving from a location of interest and those originating elsewhere, thus enabling source reconstruction by varying the location of interest over the whole source space. Imaging methods, as used in this work, consist of assuming current dipoles densely distributed over the whole cortex with fixed positions. As we have seen, this approach leads to a underdetermined linear system. Furthermore, the dipole orientations can also be fixed, but it is not necessary to ensure linearity. Nevertheless we fix our dipole orientations perpendicular to the cortical mantle as it is compatible with our knowledge about pyramidal neuron structures in the cortex and also conveniently reduces the number of degrees of freedom by a factor of 3.

Our main objective from now on is thus to find a suitable 'inversion' of the forward equation (2.29), which means to estimate the activity J from given signals V . The ill-posed nature of the problem automatically arises from the fact that G is a $M \times N$ matrix with typically $M \gg N$ in a realistic case, which means that the problem has an infi-

nite number of equivalent solutions. Inclusion of suitable constraints to the problem is therefore mandatory in order to get a meaningful and hopefully unique solution. Note, however, that we already used some constraints to derive (2.28), like fixed dipole positions and directions as mentioned before, but these are clearly not enough.

A common procedure to find the vector of the current amplitudes J from given V is to minimize a cost or energy function, whose specific choice gives rise to a big family of methods used in the literature, which commonly fall into the following form:

$$E(J) = \|V - GJ\|_p + \lambda \mathcal{R}(J), \quad (3.1)$$

where $\|\cdot\|_p$ is the L_p -norm, the first term is the measurement error, $\mathcal{R}(J)$ is called *regularizer*, whose purpose is to add information to the problem in order to restrict the possible solutions and $\lambda > 0$ is the *regularization parameter*, which measures the tradeoff between the two terms.

The most common norm used is the L_2 and $\mathcal{R}(J)$ is usually chosen as a quadratic form $\mathcal{R}(J) = J^t W J$, where W is a $N \times N$ symmetric positive definite matrix and the superscript $'$ stands for *transpose*, leading to

$$E(J) = \|V - GJ\|^2 + \lambda J^t W J. \quad (3.2)$$

The minimum of (3.2) is easily obtained by setting its derivative with respect to J equals to zero:

$$\begin{aligned} \frac{\partial E(J)}{\partial J} &= \frac{\partial}{\partial J} (V^2 - 2J^t G^t V + J^t G^t G J + \lambda J^t W J) \\ &= -2G^t V + 2(G^t G + \lambda W)J = 0. \end{aligned} \quad (3.3)$$

Rearranging terms we obtain:

$$\hat{J} = (G^t G + \lambda W)^{-1} G^t V = W^{-1} G^t (G W^{-1} G^t + \lambda I_M)^{-1} V = L V, \quad (3.4)$$

where $L := W^{-1}G'(GW^{-1}G' + \lambda I_M)^{-1}$ is the associated Wiener filter operator. A simple way to evaluate the performance of the Wiener filter L is to calculate its associated resolution matrix $R = LG$, which would correspond to the identity matrix in the case of perfect reconstruction because we have the relation $\hat{J} = LV = LGJ = RJ$. A big family of estimators is related to the specific choice of the matrix W . For instance, one of the first developed estimators, called the minimum norm estimate (MNE) [23], which is equivalent to the Tikhonov regularization [42], consists in setting $W = I_N$, leading to $\mathcal{R}(J) = \|J\|^2$, which in turn means that sources with small norms are favoured. A well known limitation of this method is its tendency to estimate deep sources as superficial wide spread sources, which happens because of the fact that the electric potential falls off with distance and therefore signals generated by deep sources can be usually well described by closer sources to the sensors (superficial sources) with smaller norms. Several methods have been developed to address this so called depth bias, such as the weighted MNE (wMNE), which sets $W = \text{diag}(w)$, with $w_j = \sum_i G_{ij}^2$ and statistical parametric mapping methods like dynamic statistical parametric mapping (dSPM) [9] and standardized low resolution electromagnetic tomography (sLoreta) [20]. Interestingly, sLoreta corresponds to the first step of the Champagne algorithm [49, 50], an algorithm to efficiently minimize the cost function of the Sparse Bayesian Learning framework (Appendix B), which means that the latter can be seen as a refinement of the former. Conveniently, all of the imaging methods of the form (3.1) can be described from a Bayesian perspective [51], which we will adopt from now on and describe next.

3.3 A Bayesian Approach of the Inverse Problem

As mentioned in the preceding section, regularization methods of the form (3.1) are special cases of Bayesian methods. From our forward equation (2.29) we can immediately use Bayes' theorem to write the posterior distribution of the currents J :

$$P(J|V, I) = \frac{P(V|J, I)P(J|I)}{P(V|I)}, \quad (3.5)$$

where I refers to all additional or prior information that we include in our model, $P(V|J, I)$ is the likelihood of the measurements given the currents, which essentially contains the forward model, $P(J|I)$ is the prior distribution of the currents and $P(V|I)$ is the evidence of the data V given by

$$P(V|I) = \int P(V|J)P(J|I)dJ, \quad (3.6)$$

where dJ is a measure on the current source space. If we assume that the errors are i.i.d from a Gaussian distribution with mean 0 and unit covariance matrix, and choose the prior of the currents from the exponential family $P(J|I) \propto \exp(\lambda\mathcal{R}(J))$, we have

$$-\log P(J|V, I) \propto \|V - GJ\|_2^2 + \lambda\mathcal{R}(J). \quad (3.7)$$

This means that maximizing the log-posterior (3.7), which is equivalent to finding the maximum a posteriori solution (MAP), is identical to minimizing (3.1). Thus, including anatomical or physiological information to the problem corresponds to appropriately choosing the prior distribution $P(J|I)$, and is therefore an essential step to restrict our space of solutions to overcome the ill posed nature of the inverse problem. A key property that we will require is sparsity, which will be included in our model through an approach called Sparse Bayesian Learning (SBL), to be described below.

3.4 Sparse Bayesian Learning (SBL)

Sparse descriptions of objects are of great interest to the Machine Learning (ML), Signal Processing and Image Reconstruction communities as means of achieving parsimonious representations of their studied objects. A sparse vector, for instance, consists of a vector whose only few components are different from zero. In our context of EEG source reconstruction we desire sparsity in the sense that only relatively small regions are assumed to be active during some specific task, which means that we are going to look for sparse J . There are different ways of achieving sparsity, with the most obvious ones being explicitly choosing sparsity enforcing regularization terms in the cost function using, for example, the L_0 and L_1 norms. Here we will use a Bayesian approach called Sparse Bayesian Learning (SBL), which relies on ideas developed in [32, 35, 40, 43, 44] and aims at estimating the model hyperparameters from the data (Empirical Bayes [3]). The idea is first to choose a suitable prior family $P(J|\theta)$ parametrized by θ and then to approximate the posterior on θ by a delta function around its mode $\hat{\theta}$, obtaining

$$P(J|V) = \int P(J|V, \theta) P(\theta|V) d\theta \approx \int P(J|V, \theta) \delta(\theta - \hat{\theta}) d\theta = P(J|V, \hat{\theta}), \quad (3.8)$$

with $\hat{\theta} = \arg \max_{\theta} P(\theta|V)$. Using Bayes' theorem, $P(\theta|V)$ can be written as

$$P(\theta|V) = \frac{P(V|\theta)P(\theta)}{P(V)}. \quad (3.9)$$

If we now consider a uniform hyperprior $P(\theta)$, then maximizing $P(\theta|V)$ is equivalent to maximizing $P(V|\theta)$. But $P(V|\theta)$ is the model evidence, which appears in the denominator of (3.5) and is given by (3.6) with $I = \theta$. Hence this procedure is called evidence maximization, corresponding to choosing the most probable or simplest model from a Bayesian perspective [32]. Other approaches, known as hierarchical models, explicitly choose a family for the hyperprior $P(\theta)$ and use some approximation procedure to make inference as the posterior $P(J|V)$ is usually not analytically tractable [39].

Returning now to our problem, let's first rewrite our main equation:

$$V = GJ + \xi, \quad (3.10)$$

with $V, \xi \in \mathbb{R}^{M \times T}$, $J \in \mathbb{R}^{N \times T}$ and $G \in \mathbb{R}^{M \times N}$, with $N > M$, and we want to estimate J given V . First we assume that the noise is uncorrelated over time, which means that each column ξ_i of ξ is an i.i.d gaussian noise $\xi_i \sim \mathcal{N}(\xi_i|0, \beta^{-1}I_M)$, where $\beta > 0$ is the precision and I_M is the $M \times M$ identity matrix, which leads to a gaussian likelihood $P(V|J) = \mathcal{N}(V|GJ, \beta^{-1}I_M)$. Now we have to choose a prior distribution that will lead to sparse solutions and perhaps surprisingly it is enough to choose a simple gaussian of the form $P(J|I) = P(J|\gamma) = \mathcal{N}(J|0, \Gamma)$, with $\Gamma = \text{diag}(\gamma)$ and $\gamma = (\gamma_1, \dots, \gamma_N)^t$, $\gamma_i > 0$, i.e. :

$$P(J|\gamma) = \prod_{t=1}^T P(J_t|0, \gamma) = \prod_{t=1}^T \mathcal{N}(J_t|0, \Gamma). \quad (3.11)$$

So the next step is to estimate the variance vector γ through maximization of the model evidence, which is tractable since it's a convolution of gaussians:

$$P(V|\gamma) = \int P(V|J)P(J|\gamma)dJ = \mathcal{N}(V|0, \Sigma_V), \quad (3.12)$$

$$P(V|\Gamma) = \int P(V|J)P(J|\Gamma)dJ = \mathcal{N}(V|0, \Sigma_V), \quad (3.13)$$

$$V = \tilde{G}\tilde{J} + \xi$$

$$\psi_i \in \mathbb{R}^N$$

where $\Sigma_V = \beta^{-1}I_M + GTG'$ is called the model data covariance. Maximization of (3.13) can also be written as a minimization of a cost function defined by $\mathcal{L}(\gamma) = -2\log P(V|\gamma)$, explicitly:

$$\hat{\gamma} = \arg \min_{\gamma} \mathcal{L}(\gamma)$$

$$\mathcal{L}(\gamma) = \log |\Sigma_V| + \frac{1}{T} \sum_{t=1}^T V_t' \Sigma_V^{-1} V_t, \quad (3.14)$$

where $|\Sigma_V|$ is the determinant of Σ_V and irrelevant additive constants were omitted. It can be shown that the first term in (3.14) enforces sparsity [29] (Appendix C), while the second minimizes the data empirical error. Moreover, the second term can be rewritten as $Tr(C_V \Sigma_V^{-1})$ where $C_V = VV'/T$ is the empirical covariance.

There are many possible approaches to minimize (3.14) and in this work we used the so called **Champagne algorithm** [52] (Appendix B), since it is orders of magnitude faster for highly ill-posed problems than more common approaches like the EM algorithm. Once an estimate $\hat{\gamma}$ of γ is found, inference proceeds by calculating the posterior of $J|V$ using Bayes' rule, which is normal distributed since it is a product of normal distributions:

$$P(J|V, \hat{\gamma}) = \prod_{t=1}^T P(J_t|V_t, \hat{\gamma}) = \prod_{t=1}^T \mathcal{N}(J_t|\hat{J}_t, \Sigma_J), \quad (3.15)$$

where the mean and covariances are given by

$$\hat{J}_t = \beta \Sigma_J G' V_t, \quad (3.16)$$

$$\Sigma_J^{-1} = \Gamma^{-1} + \beta G' G. \quad (3.17)$$

The mean value (3.16) coincides with the Maximum a Posteriori Probability (MAP) estimate, which is a widely used Bayesian estimate in many inference problems and corresponds to the Maximum Likelihood estimate in the case where no prior information is available (uniform prior distribution). Observe that computation of (3.16) requires the inversion of a $N \times N$ matrix, where N is the dimensionality of the source space and can be large in realistic cases. To alleviate the computational burden we can use the Woodbury identity [53] (Appendix G) to obtain a new expression for the mean posterior current vector:

$$\hat{J}_t = \Gamma G' \left(\beta^{-1} I_M + G \Gamma G' \right)^{-1} V_t = \Gamma G' \Sigma_V^{-1} V_t, \quad (3.18)$$

which now requires the inversion of a $M \times M$ matrix, where M is the dimensionality of the signal space and which is usually much smaller than N as we already discussed.

3.5 Inclusion of structures as degrees of freedom into the SBL framework: Structured Sparsity

A fundamental question that can be posed in general imaging reconstruction problems is about the characteristic **length scales** present in the generators of the objects to be reconstructed. It's common that one does not have prior information about these relevant scales and/or there can be several different scales simultaneously present in the signal

generation. The addition of structures in a sparse context is known as **structured sparsity** in the compressed sensing community [11], and it is an area of continuously growing interest. Spatial and temporal structures can be either imposed from prior information or even learned [47, 54], and only recently such ideas have been used in the EEG/MEG source reconstruction community [2, 6, 8, 25].

Let's briefly discuss how we can include structures into our previously described SBL formalism in a natural way. The key idea of the SBL is to introduce a normal prior distribution with a variance for each free parameter, which in our case means introducing a variance parameter γ_i for each J_i , where J_i is the amplitude of the i th current dipole. Let's set $T = 1$ for simplicity and thereby omit the time index t in our analysis. Our chosen prior distribution on the amplitudes is given by $P(J|\gamma) = \mathcal{N}(J|0, \Gamma)$, with $\gamma = (\gamma_1, \dots, \gamma_N)$ and $\Gamma = \text{diag}(\gamma)$. So, our prior covariance is diagonal with a free parameter for each dipole, which means that we are letting each dipole behave independently and therefore no structure is enforced. To include some structure, we just have to introduce more general covariances, which can be easily achieved by first observing that the diagonal prior covariance can be written as

$$\Gamma = \begin{pmatrix} \gamma_1 & & \\ & \ddots & \\ & & \gamma_N \end{pmatrix} = \sum_{i=1}^N \gamma_i e_i e_i', \quad (3.19)$$

where e_i is the N dimensional standard unit vector with 1 in its i th component and 0 elsewhere. The sum on the right-hand side of (3.19) is the eigendecomposition of Γ , where the e_i 's are its eigenvectors and γ_i 's their corresponding eigenvalues. To generalize (3.19) we can just set ([50, 51]) :

$$\Gamma = \sum_{i=1}^{N_S} \gamma_i \psi_i \psi_i', \quad (3.20)$$

where the ψ_i 's are general basis vectors that in general have multiple nonzero components and N_S is the total number of structures.

3.5.1 Simple example of inclusion of a structure into the SBL framework

Here we show how to add structures using (3.20) through a very simple example. Suppose we have four dipoles indexed from 1 to 4 and for some reason we want to add some a priori structure considering that the first 3 dipoles are perfectly correlated and the fourth is independent. This can be achieved by choosing $\psi_1 = (1, 1, 1, 0)'$ and $\psi_2 = (0, 0, 0, 1)'$, which leads to (Figure 3.20):

$$\begin{aligned}
 \Gamma &= \gamma_1 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} (1, 1, 1, 0) + \gamma_2 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} (0, 0, 0, 1) \\
 &= \gamma_1 \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \gamma_2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
 &= \begin{pmatrix} \gamma_1 & \gamma_1 & \gamma_1 & 0 \\ \gamma_1 & \gamma_1 & \gamma_1 & 0 \\ \gamma_1 & \gamma_1 & \gamma_1 & 0 \\ 0 & 0 & 0 & \gamma_2 \end{pmatrix}.
 \end{aligned} \tag{3.21}$$

We see that the structures we have introduced are controlled by the parameters γ_1 and γ_2 . Notice that the values of the components can be arbitrary, so we can for example introduce weaker correlations or even anti-correlation by changing signs between components. Furthermore, it is also worth mentioning that by adding these structures we have decreased our number of free parameters from 4 to 2 in this example. Reducing the number of free parameters is obviously extremely helpful to solve highly ill posed problems, so imposing structures in the EEG inverse problem from either functional or anatomical information is a promising way to improve reconstruction accuracy. As is showed in the example above, we can introduce structures of different sizes or even as overlapping regions, meaning that a dipole might be included in more than one structure simultaneously. For example, regions of different sizes around a center dipole can be considered as degrees of freedom simultaneously, as is done in the second experiment in next chapter. Another way to approach it, is by solving the problem stepwise, where in each step the problem is solved in some scale of resolution corresponding to specific degrees of freedom of the source space and then its solution can be used as prior informa-

Figure 3.1: Representation of the addition of a simple structure into a 4 spin model. (Left) Four independent spins. (Right) Addition of structures $\psi = (\psi_1, \psi_2)$, $\psi_1 = (1, 1, 1, 0)'$, $\psi_2 = (0, 0, 0, 1)'$, constrains the first 3 spins to change together and the fourth to behave independently. With these structures a reduction from four to two degrees of freedom was achieved.

tion for the same problem in a finer scale with new degrees of freedom, like it is done in [2].

In [2] renormalization group ideas have been used in the EEG source reconstruction problem where several scales of resolution of the source space are defined through downsampling of a fine resolution of the cortex and an efficient method to transfer information between these scales is developed. More precisely, the solution of the problem at a coarser scale is incorporated as prior information in the next finer scale through a *backward renormalization* procedure [7], which is repeated until some desired resolution is achieved. Addition of these multiple scales of resolution has been shown to clearly improve localization accuracy of focal activities. Despite the great success of this so called Backward Renormalization Group Priors (BRGP) method in finding focal sources compared to classic methods, it has not been extended to reconstruct more general activity profiles and inclusion of general structures as free parameters is not straightforward.

Through the SBL framework, once we have introduced some structure into our general prior covariance matrix (3.20), the procedure follows identical as before by finding γ through evidence maximization using the Champagne algorithm. It turns out that the only change that has to be made in the update rule (Appendix B.10) [50] is following replacement:

$$G_j \rightarrow \tilde{G}_j = G\psi_j, \quad (3.22)$$

which in turn can be compactly summarized as $G \rightarrow \tilde{G} = G\psi$, where $\psi = [\psi_1, \dots, \psi_{N_S}]$ is the matrix with columns ψ_j and N_S is the total number of structures or new degrees of freedom of our problem. There is a very intuitive way to see what this replacement means. Take again the four spins example from before where we grouped the first three spins into a single degree of freedom and let the last one behave independently. Then, (3.22) is given by

$$\begin{aligned}
 \tilde{G} = G\psi &= \begin{pmatrix} G_{11} & G_{12} & G_{13} & G_{14} \\ G_{21} & G_{22} & G_{23} & G_{24} \\ G_{31} & G_{32} & G_{33} & G_{34} \\ G_{41} & G_{42} & G_{43} & G_{44} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \\
 &= \begin{pmatrix} G_{11} + G_{12} + G_{13} & G_{14} \\ G_{21} + G_{22} + G_{23} & G_{24} \\ G_{31} + G_{32} + G_{33} & G_{34} \\ G_{41} + G_{42} + G_{43} & G_{44} \end{pmatrix} \\
 &= \begin{pmatrix} \tilde{G}_{11} & \tilde{G}_{12} \\ \tilde{G}_{21} & \tilde{G}_{22} \\ \tilde{G}_{31} & \tilde{G}_{32} \\ \tilde{G}_{41} & \tilde{G}_{42} \end{pmatrix}.
 \end{aligned} \tag{3.23}$$

So, from (3.23) we see that the associated lead field to our new degree of freedom of grouped dipoles is given by the sum of the lead fields at each sensor, which can be written more generally as

$$V_i = \sum_j^N G_{ij} J_j = \sum_{\alpha=1}^{\tilde{N}} J_{\alpha} \sum_{j \in \alpha} G_{ij} = \sum_{\alpha=1}^{\tilde{N}} \tilde{G}_{i\alpha} J_{\alpha}, \tag{3.24}$$

with

$$\tilde{G}_{i\alpha} = \sum_{j \in \alpha} G_{ij}, \quad J_{\alpha} = J_{j \in \alpha}, \tag{3.25}$$

where the α 's correspond to the indexes of the new $\tilde{N}$ degrees of freedom and $j \in \alpha$ means the j th dipole which is inside the structure indexed by α . Notice that we are forcing every dipole j within a structure α to have the same amplitude J_α .

3.6 Multiscale Approach: How to include degrees of freedom of several scales simultaneously

With the formalism described in the previous section to include structures in our source reconstruction problem, we can now imagine a situation where our active sources can spread through regions of different characteristic sizes and describe it as follows. First, we define D scales of resolution of the source space, where under a scale of resolution is to be understood that the source space is partitioned into regions (or structures) of a characteristic size with uniform activities. Specifically, if we assume that the potential is generated by the degrees of freedom of a certain scale d , we have the respective forward equation:

$$V = G_d J_d + \xi, \quad (3.26)$$

where $d \in \{1, \dots, D\}$, $V \in \mathbb{R}^{M \times T}$, $G_d \in \mathbb{R}^{M \times N_d}$, $J_d \in \mathbb{R}^{N_d \times T}$. Here, a component $(J_d)_i =: J_i^d$ represents a region where all dipoles within are simultaneously active with the same magnitude, $(G_d)_i$ is its corresponding lead field and N_d is the number of regions of activity of scale d , which corresponds to the degrees of freedom of our problem. Under above considerations, we can obtain G_d from the original leadfield G by simply considering the expression for the potential at a specific electrode i and write

$$V_i = \sum_{j=1}^N G_{ij} J_j = \sum_{\alpha=1}^{N_d} J_\alpha^d \sum_{j \in \alpha} G_{ij} = \sum_{\alpha=1}^{N_d} G_{i\alpha}^d J_\alpha^d, \quad (3.27)$$

where

$$G_{i\alpha}^d = \sum_{j \in \alpha} G_{ij}, \quad J_\alpha^d = J_{j \in \alpha}. \quad (3.28)$$

Under ' $j \in \alpha$ ' is meant a dipole indexed by j that is inside the region indexed by α of scale d and we are forcing all dipoles within region α to have the same magnitude J_α^d . Now, our activity could originate from any possible regions of any possible scales and instead of iteratively solving and transferring the solution between scales, we try to solve the problem at all scales **simultaneously** by constructing the following composed Green function:

$$\mathcal{G} := [G_1, G_2, \dots, G_D], \quad (3.29)$$

where (3.29) just means that we are stacking all columns of all G_d together. Then we solve the following problem for the new degrees of freedom $\mathcal{J} := \left\{ J_\alpha^d \right\}_{\alpha,d}^{N_d,D}$:

$$V = \mathcal{G}\mathcal{J} + \xi, \quad (3.30)$$

with $\mathcal{G} \in \mathbb{R}^{M \times \mathfrak{N}}$, $\mathcal{J} \in \mathbb{R}^{\mathfrak{N} \times T}$, where $\mathfrak{N} = \sum_{d=1}^D N_d$. To gain a better insight to what this means, observe a signal at the i th sensor at a fixed time instant is explicitly given from (3.30) as

$$V_i = \sum_{d=1}^D \sum_{\alpha=1}^{N_d} G_{i\alpha}^d J_\alpha^d + \xi_i. \quad (3.31)$$

From (3.31) we see that there are terms from all scales that can contribute to the signal V_i , and we let our SBL algorithm choose the appropriate sparse configuration of terms that most likely give rise to the measured signal. In this framework all scales are treated on an equal footing, which in turns appropriately reflects a situation where there is lack of knowledge of the typical sizes of activated regions during some task. Such approaches of simultaneously considering generators of multiple scales to reconstruct an object have been used for example in the context of image and video restoration [33]. Here, we want to simulate this scenario by pre-defining regions of different characteristic scales and to evaluate if the SBL can recover activated regions of any scale. Of course, in real scenarios such partitioning of the brain should be done according to some functional and/or anatomical information, possibly application dependent. For example, in [38] the source space is partitioned into the famous Brodmann Areas [5]. Another possibility is to use

information of other functional imaging modalities, like fMRI or PET to define regions that are more likely to be active during some task. Here, however, we define the activity regions artificially through the use of a simple clustering algorithm (KMeans - Appendix F.1) for simplicity.

To close this section, it is still worth mentioning that through the formalism described above it is even possible to consider several different partitions of the same scale simultaneously as degrees of freedom, which in turn would allow arbitrary linear combinations of them as possible solutions, what might improve the methods capacity of fitting more general activity profiles. Nonetheless, such approaches would require a careful analysis on how to efficiently and meaningfully choose such partitions, which are therefore left to further works.

In the next chapter we are going to apply the formalism described above in two experiments and analyse the results.

Chapter 4

Simulations and Results

In this chapter we are going to apply the formalism described in the previous chapter of adding different structures as degrees of freedom into the SBL framework to reconstruct active cortical regions in two experiments.

In the first one we define several possible regions of different sizes as degrees of freedom simultaneously and try to recover them from simulated noisy signals. The key here is that regions of different sizes are treated on an equal footing, reflecting situations where we do not have a priori knowledge about the typical extensions of active cortical regions. This is in contrast to hierarchical approaches based on renormalization group ideas such as used in [2] that aim at reconstructing small active regions in a fine representation of the cortex. The idea is to solve the inverse problem using different resolutions of the cortex and to use the solution of a coarser scale obtained through a Bayesian learning step as prior information at the next finer scale through a **Backward Renormalization** (BR) procedure [7] and iterate this procedure until the finest desired scale is achieved. Such an approach was attempted in this work in order to try to recover more extended regions of activity but without success because information was lost in the BR step between coarser and finer scales. In other words, we couldn't design a robust BR step to satisfactorily conserve information obtained from the coarser scales. As a solution to this limitation we decided to drop the BR step and apply instead just one learning step using all regions of all scales simultaneously. In this way we treated all scales on an equal footing and let our SBL framework chose the sparsest combination of them that best described the simulated signals.

In the second experiment we used focal structures of different sizes to reconstruct focal activities, i.e., activity profiles with a center dipole with highest amplitude and a decaying amplitude profile with distance. The reconstruction was done in two steps, a scanning step, where large structures around each dipole of the original source space were used to scan for regions with high activity probability, and a fine tuning step, where structures

of different sizes around each dipole of these probably active regions were used to try to fine tune the true activity profile.

For all simulations was used the MNE software [13, 14] with Python, which provide several EEG and MEG analysis/visualization tools and a sample subject with real data experiments and high resolution surfaces reconstructed from MRI through Freesurfer [12]. The lead field matrix was calculated using linear collocation method in a three compartment head volume consisting of inner skull, outer skull and outer skin (head) BEM surfaces, represented in Figure 4.1. Furthermore, high resolution images of the sample inner and outer cortex surfaces can be seen in Figures 4.3 and 4.4.

Moreover, in all simulations was used a 256 electrodes montage, depicted in Figure 4.2. This is a very high electrode density montage, which provide a great advantage in comparison to commonly used low density montages in clinical applications of about 10 – 20 electrodes as it allows for a much more accurate description of the scalp potential profile. However, as opposed to what one might think, it is widely believed that indefinite addition of electrodes leading to increasingly denser montages do not increase the amount of usable information to be extracted from the scalp potential because the skull has a smearing effect on the signals and also acts as a low pass filter on its spatial frequencies. In other words, it can be shown that components of higher spatial frequencies of the measured signals are stronger attenuated by the high skull conductivity than components of lower frequencies [22], imposing theoretically a natural limitation on the possible achievable spatial resolution through EEG. However, there are also claims in the opposite direction [36], which puts this issue as a subject of still ongoing debate.

Figure 4.1: *BEM surfaces corresponding to inner skull, outer skull and outer skin and pial surface.*

Figure 4.2: *Head Montage with 256 electrodes used in all simulations.*

Figure 4.3: *Fine representation of the sample white surface (inner cortex surface) with 312273 vertices provided by the MNE software.*

Figure 4.4: *Fine representation of the sample pial surface (outer cortex surface) with 312273 vertices provided by the MNE software.*

4.1 Experiment 1 - Reconstruction of Active Cortical Areas of Multiple Scales

In this first numerical experiment we applied SBL to find activated regions of the source space that can have in principle different characteristic sizes. Here we use a coarser representation of the sample fine white surface (Figure 4.3) with 20480 vertices (Figure 4.5). The current dipoles are placed on the vertices and are normal oriented to it, i.e., their direction correspond to the mean of the normals of their neighbouring triangles (Figure 4.6).

Figure 4.5: *Coarse representation of the white surface with 20480 vertices used in the first experiment.*

Figure 4.6: Vertices of the coarse representation of the white surface used in the first experiment. The blue arrows are the current dipoles with directions corresponding to the mean of normals of its surrounding triangles

In order to define the different activity regions we apply the K-Means Clustering algorithm (KM) (Appendix F) to our solution space multiple times in a hierarchical fashion. It should be kept in mind that this choice is by no means fundamental and is just used as a simple way to create regions of different sizes and shapes. The algorithm groups together the given points into n clusters of close points, where n is a free parameter. Given that our source space points are more or less uniformly distributed through the cortical mantle, the clusters, although random, will have similar sizes and therefore a scale of resolution in our problem can be identified by the number of clusters we ask the KM to find. We chose four scales of resolution indexed by 0 to 3 corresponding, from the coarsest to finest, to 20, 80, 320 and 1280 regions, respectively. First, we applied the KM to group together the 20480 points of the original source space into 20 regions. Then, to achieve the next resolution we applied the KM to each of these 20 regions separately to partition them into 4 new regions, leading to a total of $4 \times 20 = 80$ new smaller regions composing the next finer scale. This procedure is repeated until we reach our last scale of 1280 regions. A summary of the information of these scales are shown in table 4.1.

scale	$\langle d \rangle (mm)$	$\sigma_d (mm)$
0	26.70	7.26
1	13.87	4.54
2	8.45	2.87
3	5.28	1.73

Table 4.1: *Mean distance and standard deviation of nearest neighbours cluster centers defined through the K-Means Clustering algorithm for each scale of resolution in experiment 1.*

Figure 4.7: *First scale (scale 0) of resolution of the cortex with 20 regions defined with the K-Means Clustering algorithm. The mean distance between first neighbours cluster centers is 26.70(7.26)mm.*

Figure 4.8: *Second scale (scale 1) of resolution of the cortex with 80 regions defined with the K-Means Clustering algorithm. The mean distance between first neighbours cluster centers is $13.87(4.54)\text{mm}$.*

Figure 4.9: *Third scale (scale 2) of resolution of the cortex with 320 regions defined with the K-Means Clustering algorithm. The mean distance between first neighbours cluster centers is $8.45(2.87)mm$.*

Figure 4.10: *Fourth scale (scale 3) of resolution of the cortex with 320 regions defined with the K-Means Clustering algorithm. The mean distance between first neighbours cluster centers is 5.28(1.73)mm.*

Once we have our regions we apply the formalism described in the last chapter to determine the corresponding lead field matrix by summing over the lead field values of the dipoles within each region, such that we obtain a resulting lead field matrix $\mathcal{G}$ of size 256×1700 , where 1700 ($20 + 80 + 320 + 1280$) is the total amount of possible regions of activity.

The simulation proceeds as follows. We select a region at random, apply a current during some period of time, project it to the signal space through the lead field and add some noise. The corrupted signal is then used to estimate the active region through the Champagne algorithm. The current profile for the simulations was modulated through an amplitude of the form $A(t) = e^{-\alpha t^2}$, which resembles a bump (Figure 4.11). Moreover, the current was applied over a time range between -20 and 20 with step 1, yielding a total of $T = 40$ time instants and α was set to $1/50$. Typically, a realistic time unit for such evoked potentials is millisecond and the current amplitude reaches values on the order of 10 nAm , nevertheless we omit the units since they are not relevant for the simulations and could in principle be anything else. Furthermore, our inverse algorithm was evaluated at the signal over the range -10 to 10, as delimited by the black vertical lines in the right graph of Figure 4.12, since it's about the range where some activity can be identified beyond noise, leading therefore to a total of $T = 30$ time instants as input to the Champagne algorithm.

Formally, we have a current matrix $\mathcal{J} \in \mathbb{R}^{N_4 \times T}$, with $N_4 = 1700$ and $T = 40$, from which was then obtained the signals V through $V = \mathcal{G}\mathcal{J}$. The signals were then corrupted with independent gaussian noise $\xi \in \mathbb{R}^{M \times T}$, $V \rightarrow V + \xi$, whose level was defined through the signal-to-noise ratio (SNR):

$$SNR = 20 \log_{10} \left(\frac{\|V\|_{\mathcal{F}}}{TM\sigma_{\xi}} \right), \quad (4.1)$$

where $M = 256$ is the number of channels. So, for a given SNR level (in dB), we added a gaussian noise $\xi \sim \mathcal{N}(\xi|0, \sigma_{\xi})$ with σ_{ξ} defined through (4.1) to the true signal matrix V . This corrupted signal was then used as input to the Champagne algorithm, which gives us some estimate of the underlying current $\hat{\mathcal{J}}$. To measure the algorithms performance first both $\mathcal{J}$ and $\hat{\mathcal{J}}$ were represented in our original finest scale (f) of the source space with $N = 20480$ dipoles $\mathcal{J} \rightarrow J_f$, $\hat{\mathcal{J}} \rightarrow \hat{J}_f$ and their overlap was evaluated through the **Jaccard Index**, also known as **Intersection over Union** (IoU). Given two sets A and B , their IoU defined by:

Figure 4.11: *Amplitude profile over time given by $A(t) = e^{-t^2/50}$ applied to the simulated active regions with 40 time instants.*

$$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}, \quad (4.2)$$

where $|A|$ is the cardinality of set A . Notice that $0 \leq IoU \leq 1$, where 1 means perfect overlap and 0 means disjoint. In this first experiment we are not concerned with matching the amplitude profile of the true sources, but only to infer which are active or not. The components of the estimated current vector were therefore considered as active (1) if they were greater than a threshold and inactive (0) otherwise, which was accomplished through following transformation for each time instant:

$$(\hat{J}_f)_i \rightarrow H(|(\hat{J}_f)_i| - \eta \max_{1 \leq k \leq N} |(\hat{J}_f)_k|), \quad (4.3)$$

where $H(x)$ is the Heaviside step function given by $H(x) = 1$ if $x \geq 0$ and 0 otherwise,

Figure 4.12: *Example of simulated signal without noise (left) and with noise level SNR = 10 dB (right). Each curve on both graphs corresponds to one of the 256 sensors. The evaluation of the algorithms performance was done within the range delimited by the black vertical lines.*

and η is a threshold parameter, which was set to 0.9. Moreover, with (4.3) the overlap measure (4.2) for our transformed binary currents at a specific time becomes

$$IoU(J_f, \hat{J}_f) = \frac{J_f \cdot \hat{J}_f}{\sum_{i=1}^N [(J_f)_i + (\hat{J}_f)_i] - J_f \cdot \hat{J}_f}, \quad (4.4)$$

where $\cdot$ is the usual dot product. Since we have such an estimate for each time instant index t , which we omitted until now, we used the time average over the whole activity period T as our final metric:

$$IoU := \langle IoU(J, \hat{J}) \rangle_T = \frac{1}{T} \sum_{t=1}^T IoU(J_t, \hat{J}_t). \quad (4.5)$$

$$IoU := \langle IoU(J_f, \hat{J}_f) \rangle_T = \frac{1}{T} \sum_{t=1}^T IoU(J_{ft}, \hat{J}_{ft}). \quad (4.6)$$

We simulated 32 random active regions, 8 for each scale and half of them being superficial and the other half deep source configurations with noise levels of SNR ranging from 0 dB to 20 dB. Plots of the mean values of the IoU can be seen in figures 4.13 and 4.14 grouped by regions and scales, respectively.

In Figure 4.13 we see that the estimation performance between superficial and deep sources using SBL are very similar, indicating the ability of sparsity of better resolving deep source configurations than typical classical methods, which tend to have trouble in estimating deeper sources because they tend to generate much weaker signals than superficial ones. More evidence of this important feature can be found in [25]. Furthermore, we see that for high SNR all regions could be perfectly recovered and for low SNR we get a slight performance decrease but still reaching reasonable values above 0.9.

Next we looked into the performance for the different scales as represented in Figure 4.14. We see again that performance is maximum at low noise levels but drops with increasing scale for high noises, which can be explained by analysing the proximity between the signals in the signal space. For instance, consider two arbitrary regions α and β with corresponding signals V_α and V_β , respectively, obtained through their associated lead fields $\mathcal{G}_\alpha$ and $\mathcal{G}_\beta$, where $\mathcal{G}_\gamma$ is the γ th column of the generalized lead field matrix $\mathcal{G}$. As a proximity measure between these signals we can consider the cosine of the angle θ between them: $\cos \theta = V_\alpha \cdot V_\beta / |V_\alpha| |V_\beta| = \mathcal{G}_\alpha \cdot \mathcal{G}_\beta / |\mathcal{G}_\alpha| |\mathcal{G}_\beta|$. Since here we are not concerned with the amplitude values, we can consider only angles $\theta \in [0, \pi/2]$ where $\cos \theta \in [0, 1]$, which can be achieved by appropriately changing signs of the involved vectors. Now, we can calculate the cosines between all signals and summarize this information in a histogram like plot (Figure 4.16). The x axis is the cosine and the y axis is the probability density estimated from histograms for all scales. For instance, the black curve for scale 0 is obtained by calculating the cosines between the signals of scale 0 and all others and then calculating its normalized histogram with bars of width 0.01, leading to a total of 100 bins between 0 and 1. Finally, the x values of the plotted points correspond to the centers of the histogram bars. From Figure 4.16 we observe that there are relatively (higher cosines) signals to signals originating from coarser than from finer activity scales, leading to a higher probability of wrongly estimating sources of the former type, specially for higher noise levels where true signal vectors tend to change their direction by greater angles than for lower noise levels, which can be easily seen by noting that the corrupted signal V can

Figure 4.13: Mean IoU for activity estimations of 16 deep and 16 superficial random activated regions for different noise levels.

be written as $V = V_S + \xi$, where V_S is the true signal and $\xi \sim \mathcal{N}(\xi|0, \sigma_\xi)$ the noise vector with $\xi \propto 1/SNR$. We can make a simple estimate of the cosine between V and V_S as a function of the $SNR = |V_S|^2/\sigma_\xi^2$ by choosing $|\xi| = \sigma_\xi$ and considering $V_S \cdot \xi = 0$, yielding

$$\cos(V, V_S) = \frac{V \cdot V_S}{|V||V_S|} \simeq \sqrt{\frac{SNR}{1 + SNR}}. \quad (4.7)$$

A graph of (4.7) is plotted in Figure 4.15 with the SNR represented in logarithmic scale $10 \log_{10}(SNR)$. The curve gives a rough estimate of the cosine of the typical angle between the true and corrupted signals for different SNR levels, so there is a high probability that two signals of two different sources being considered as equivalent by the inverse algorithm at a given noise level if their angle is such that its cosine is considerably greater than the estimated value $\cos(V, V_S)$, leading therefore to a high chance of the inverse algorithm to commit a mistake.

Figure 4.14: *mean IoU over time for activity estimations of all scales.*

A more informative way to visualize this issue is by plotting the complementary cumulative distribution function (ccdf) from the estimated probability density in Figure 4.16. The ccdf of a random variable X is given by $P(X > x)$, which is the complementary function of the cumulative distribution $\Phi(x) = P(X < x)$, since $P(X > x) = 1 - \Phi(x)$. A graph of the ccdf for each scale is shown in Figure 4.17, which reads as follows: for a randomly chosen signal S of a given scale (corresponding to a specific curve), the y value of a point P corresponds to an estimate of the probability of randomly choosing a signal from the whole source space within an angle θ from S , whose cosine is greater than the corresponding x value of P . We see that the curve of the coarsest scale (black) is slightly higher than the others, indicating a higher probability of the inverse algorithm to commit a mistake when a signal of a coarse scale is simulated, which we can better represent by combining the ccdf (Figure 4.16) with the relation (4.7) (Figure 4.15), yielding the curve depicted in Figure 4.18, where the x values of Figure 4.16 were substituted by the corresponding estimated noise levels given by relation (4.7) and the associated ccdf values were renamed to P_{error} because it roughly reflects the probability of committing an error.

Figure 4.15: *Estimates of the cosine of the angle between true (V_S) and corrupted (V) signals for different noise levels of SNR (dB) in logarithmic scale. At a given noise level, sources that generate signals whose cosine is much greater than $\cos(V, V_S)$ can be considered as equivalent with high probability by the inverse algorithm.*

Figure 4.16: *Estimation of the probability density of the cosine of the angles between all signals grouped by the different activity scales. Each point represents a histogram bar with width 0.01 from a normalized histogram for each scale. We note that there are generally points (higher $\cos\theta$) to the signals generated by sources of the coarsest scale (scale 0 in black) than of finer scales, which increases the probability of the inverse algorithm to misestimate the former relatively to the latter.*

Figure 4.17: *Estimation of the complementary cumulative distribution function of the cosine of the angles between all signals grouped by the different activity scales.*

Figure 4.18: *Estimation of the probability of the inverse algorithm committing a mistake ($P_{error} = P(X > \cos \theta)$) for a simulated signal of a given scale at a given noise level.*

We see that the simultaneous multiscale approach is a promising way to tackle the EEG source reconstruction problem when the typical scale of the activity is unknown, although careful analysis has to be carried out in order to define the relevant regions of activity to avoid defining degrees of freedom that are too close in the signal space because they can be hard to discriminate, specially for high noise levels. Moreover, in practice such regions should be meaningfully defined through relevant information from anatomical or/and functional nature, possibly acquired through other imaging techniques.

4.2 Experiment 2 - Reconstruction of Focal Activities through inclusion of appropriate structures

In our second experiment we aim at reconstructing focal activities. Under focal we mean that there is a center source with highest amplitude and an amplitude profile that falls off with distance from it. Focal sources are widely studied because they are simple to deal with and have important applications, like in focal epilepsy, where a pre-surgical evaluation of the patients interictal spikes is of great importance. As our source space here we use a coarser representation of the white surface with 5120 vertices (Figure 4.19) and put a current dipole with fixed direction on each one of them as before (Figure 4.20). The amplitude profile was chosen to be $J \sim \frac{1}{(d+1)^\epsilon}$, where d is the distance from the center dipole and ϵ is a positive exponent. Since our dipoles are arranged as vertices of a triangular lattice, we can count distance in units of triangle edges and choose $\epsilon = \ln 6 / \ln 2 \approx 2.58$, which means that the first nearest neighbours ($d = 1$) get an amplitude of $1/6$ of the center dipole amplitude. Moreover, we define the k th nearest neighbours (KNN) of some dipole as all dipoles that are k edges apart from it, which correspond to 3×2^k dipoles in a triangular lattice, $k = 1, 2, \dots$. Apart from this static amplitude profile, we also applied a temporal modulation to all of them of the same form as in the previous experiment through the amplitude $A(t) = e^{-\alpha t^2}$, with $\alpha = 1/50$.

Figure 4.19: *White sample surface with 5120 vertices used in the second experiment.*

Figure 4.20: *Source space with 5120 current dipoles (blue arrows) placed on each vertex and normally oriented used in the second experiment.*

Figure 4.21: *Amplitudes of the focal activities simulated in the second experiment. The highest curve corresponds to the amplitude of the center dipole (J) and each of the others corresponds to the amplitudes $J/(1+d)^\epsilon$, $\epsilon \approx 2.58$, of all dipoles that are d edges apart from the center dipole. Furthermore, all amplitudes are modulated by the bump like profile $A(t) = e^{-t^2/50}$, as in the last experiment.*

To try to estimate a given focal activity we are going to add structures and use SBL in two steps. In the first one the degrees of freedom are chosen to be large regions around each dipole of our whole source space. Specifically, for each dipole i we define a degree of freedom consisting of all surrounding dipoles within some range, which we represent by B_i^N , meaning all dipoles up to the N th nearest neighbours of i , including i itself, or, in other words, all dipoles within a range of N edges apart from dipole i . Furthermore we define B_i^0 just as the dipole i . These structures are included using the formalism described in the last chapter by adding covariance basis eigenvectors ψ_i with all components corresponding to the dipoles within the specified range from the center dipole equal to 1 and 0 otherwise and are depicted in Figure 4.22.

The SBL methods estimate amplitudes for each one of these regions and we choose a threshold to decide which of them probably contains active sources and survives into the next step. So this first step is actually scanning the whole source space looking for

Figure 4.22: Structures defined around a center dipole i (blue dot). From left to right we have B_i^0 , B_i^1 , B_i^2 and B_i^3 . At each one of them all vertices from blue triangles correspond to perfect correlated current dipoles.

potentially active regions and once we have them, we go to the second step, which is a fine estimation of the amplitude profile. This is done as follows. Say B_k^N is a surviving structure from the first step, i.e., has achieved a amplitude above some given threshold, then we introduce a sequence of F structures around each of its belonging dipoles $i \in B_k^N$ as $\{B_i^0, B_i^1, B_i^2, \dots, B_i^F\}$. This means that our degrees of freedom in the second step consist of regions of different sizes around all dipoles of the surviving regions from the first step, allowing the algorithm to search for the sparsest combination of them and fine tune the simulated amplitude profile, which could, for instance, fall rapidly or slow with distance from the activity center. We can formalize these two steps as follows:

- **Step 1 (Scanning).** Choose as degrees of freedom the structures B_i^N around each dipole i from the original source space and obtain an estimate of the activity $\hat{J}$ through the Champagne algorithm. Each component $\hat{J}_i$ of $\hat{J}$ corresponds to the amplitudes of all dipoles $j \in B_i^N$. Take the set of structures $\mathcal{S} := \left\{ B_i^N : |\hat{J}_i| > \eta \max_j \{|\hat{J}_j|\}, \eta > 0 \right\}$ to the next step and drop the rest, where η is a threshold parameter.
- **Step 2 (Fine tuning).** Choose as the new degrees of freedom the set of structures $\mathcal{D} := \left\{ \left\{ B_j^k \right\}_{k=0}^F : j \in B_i^N \subset \mathcal{S} \right\}$ and obtain the final estimate $\hat{J}$ from a new run of the Champagne algorithm. The number of degrees of freedom in this step is $|\mathcal{D}| = |\mathcal{S}| \times (3 \times 2^N) \times F$.

For the simulations was chosen $N = F = 10$ and 50 superficial and 50 deep source configurations were simulated for noise levels of SNR ranging from 0 to 20 dB and the performance was evaluated through the time average of the cosine between true and estimated source amplitudes:

$$\langle \cos(J, \hat{J}) \rangle := \frac{1}{T} \sum_{t=1}^T \frac{J_t \cdot \hat{J}_t}{|J_t| |\hat{J}_t|}. \quad (4.8)$$

The performance of SBL with the above described structures (sSBL) was compared with SBL without any structures, i.e., with the degrees of freedom corresponding to the dipoles of the original source space, MNE and weighted MNE (wMNE). In Figure 4.23 we see the mean values of the metric (4.8) over the 100 trials for the three methods, where we clearly notice the superiority of SBL. Classical MNE performed very poorly for all noise levels, whereas sSBL had a stable performance around 0.8. Moreover, SBL also had a very poor performance indicating that inclusion of appropriate structures can indeed significantly improve the reconstruction accuracy. In Figure 4.24 we plotted the same data but separated by superficial (solid lines) and deep (dotted lines) source configuration and we see that the performance of all algorithms was consistently higher for superficial source configurations.

Figure 4.23: Average cosines between estimated and true activity amplitudes for sSBL (SBL with structures), SBL, wMNE and MNE for all simulated activities.

Figure 4.24: Average cosines between estimated and true activity amplitudes for sSBL (SBL with structures), SBL, wMNE and MNE for all simulated activities separated by superficial (solid lines) and deep (dotted lines) source configurations (right).

Another interesting feature that is worth investigating is the algorithms capacity of recovering the centers of the simulated activities. For a given true activity center $\vec{x}_i$, the estimated center $\hat{x}_i$ was defined as the closest dipole position to the average position $\hat{x}$ weighted by the power of the corresponding estimated activity amplitudes $\hat{J}_k$, yielding the desired distance d :

$$d = |\vec{x}_i - \hat{x}_i|, \quad (4.9)$$

where

$$\hat{x}_i = \arg \min_{\vec{x}_j} |\vec{x}_j - \hat{x}|, \quad \hat{x} = \frac{\sum_{k=1}^N \hat{J}_k^2 \vec{x}_k}{\sum_{k=1}^N \hat{J}_k^2}, \quad (4.10)$$

where $\vec{x}_i$ is the position and $\hat{J}_i$ is the estimated amplitude associated with the i th dipole and N is the total number of dipoles. Moreover, we have a distance estimate (4.9) for all time instants t with $1 \leq t \leq T$, so as the performance metric was used the time average:

$$\langle d \rangle = \frac{1}{T} \sum_{t=1}^T d_t, \quad (4.11)$$

where d_t is distance between true and estimated activity center at time t .

The results can be visualized in figures 4.25, 4.26, 4.27, 4.28. Again, in Figure 4.25 we see the the estimates of the sSBL, SBL, wMNE and MNE for all 100 simulated source configurations and in 4.26 we see the same metric for the activities separated by superficial (solid lines) and deep (dotted lines) source configurations. In both of them we observe the superiority of the sSBL with a localization error of about 10 mm for all noise levels. MNE performed poorly for all noise levels and in particular for deep sources, as shown in 4.26, indicating the well known depth bias associated with the method, whereas the performance of the wMNE was better and increased with lower noises, although there can still be observed a significant performance difference between deep and superficial sources. Both sSBL and SBL had comparatively a much better and stable performance and in figures 4.27 and 4.28 we plotted just their performance for better visualization. We notice that while SBL is not suited for reconstructing the overall activity profile as shown in figures 4.25 and 4.26, it still can reasonably well recover the activity center.

This happens because it finds solutions comprising of only few active dipoles near the true activity center instead of extended structures around it, which can be visualized in colormaps of an example in Figure 4.29. MNE is not shown because it did not appear to have recovered any pattern. Nonetheless, the performance of sSBL was found to be superior than of SBL for all noise levels (Figure 4.27). In particular, in Figure 4.28 we see that sSBL performed better in finding deep source centers than SBL performed in finding .

Figure 4.25: Average distances between true activity centers and estimated activity centers of sSBL (SBL with structures), SBL, MNE and wMNE for all simulated activities.

Figure 4.26: Average distances between true and estimated activity centers of *sSBL* (*SBL with structures*), *SBL*, *MNE* and *wMNE* for all simulated activities separated by superficial (solid lines) and deep (dotted lines) source configurations.

Figure 4.27: Average distances between estimated and true activity amplitudes of *sSBL* (*SBL with structures*) and *SBL* for all simulated activities.

Figure 4.28: Average distances between estimated and true activity amplitudes of sSBL (SBL with structures) and SBL for all simulated activities separated by superficial (solid lines) and deep (dotted lines) source configurations.

Figure 4.29: Colormaps of an example simulation under noise level $\text{SNR}=10$ dB. (Upper left) True simulated activity. (Upper right) *sSBL* (*SBL with structures*) estimate. (Lower left) *SBL* (*without structures*) estimate. (Lower right) *wMNE*.

It is still worth examining if the second fine tuning step of our sSBL method does indeed improve reconstruction accuracy after the first scanning step. If this improvement was just marginal, there would be no reason for including the second step into our procedure. To investigate this issue, we plotted the same metrics used before to evaluate the sSBL's performance using just the first step (without step 2) and both steps (with step 2) in our estimation. In Figure 4.30 we see an obvious improvement obtained through inclusion of step 2 in our estimation of the amplitude profile, with our metric increasing the for lower noise levels, which, however, does not happen if we do not include it. Moreover, its performance is again slight worse for deep sources than for superficial ones (Figure 4.31). A similar behaviour can be observed in the estimation of the activity centers in figures 4.32 and 4.33, where we notice that inclusion of the fine tuning step greatly improves localization accuracy. Therefore we conclude that the fine tuning step is by no means redundant and should be included into our procedure to improve reconstruction accuracy. We can visualize the reconstruction of the same simulated activity as in Figure 4.29 for both steps in 4.34. As expected, we see that step 1 activates a large region around the true activity focus and that step 2 is able to find a focus near the true one. An important question that should still be addressed is how the reconstruction accuracy changes with the sizes and number of the structures used in this method and even if there is an optimal configuration of these parameters that might be automatically learned from the data, which would require more extensive simulations and is therefore left for further works.

Figure 4.30: Average cosines between true and estimated activity amplitudes of the sSBL without (green) and with (blue) the second fine tuning step for all simulated activities.

Figure 4.31: Average cosines between true and estimated activity amplitudes of the sSBL without (green) and with (blue) the second fine tuning step for all simulated activities separated by superficial (solid lines) and deep (dotted lines) source configurations.

Figure 4.32: Average distances between true and estimated activity centers of the sSBL without (green) and with (blue) the second fine tuning step for all simulated activities.

Figure 4.33: Average distances between true activity centers and estimated activity centers of *sSBL* (*SBL with structures*) and *SBL* for all simulated activities separated by superficial (solid lines) and deep (dotted lines) source configurations.

Figure 4.34: Colormaps of an example simulation under noise level $\text{SNR}=10$ dB. (Upper) True simulated activity. (Lower Left) First step of the sSBL estimate. (Lower right) Second step of the sSBL estimate.

Chapter 5

Conclusion and further work

In this work we applied structured sparsity to the EEG source reconstruction problem to deal with variable spatial extents of activity through the Sparse Bayesian Learning formalism [43, 44, 49] with the Champagne algorithm [29, 52]. EEG source reconstruction refers to the problem of estimating the electric brain sources from measured EEG signals, which offers a millisecond temporal resolution that is not achievable by most widely used functional imaging techniques such as fMRI and PET. Such a temporal resolution offers unique window into the brain dynamics underlying general cognitive processes, which justifies the huge efforts in the last decades to improve its typically low spatial resolution due to the ill-posed nature of the electromagnetic inverse problem. The fact that multiple radically different source configurations can give rise to the same measured signals requires the inclusion of additional information, typically biologically and/or mathematically inspired, to constrain the source space. Sparsity is a convenient mathematical property to ensure unique solutions in severely ill-posed inverse problems and can be biologically translated into situations where small regions of interest are active during some task, such as in brain responses to auditory or visual stimuli. There are several methods to enforce sparse solutions and here was used Sparse Bayesian Learning (SBL), a Bayesian method that consists in choosing the model hyperparameters that maximize the model evidence, which is equivalent to selecting the simplest model from a Bayesian perspective (Appendix D) and is shown to enforce sparsity as it leads to a minimization of a cost function with a regularizer that approximates the L_1 norm in some sense (Appendix C). Moreover, this cost function is minimized with the Champagne Algorithm (CA) (Appendix B), which is an efficient algorithm that is shown to be orders of magnitude faster than classical ones such as the EM algorithm (Appendix E) or the Mackay update rules [32] in highly ill-posed problems. Our original source space consists of small current dipoles placed over the whole cortical mantle with directions fixed normal to it, which models structures of parallel oriented neurons in the cortex that tend to activate coherently, so called pyramidal neurons. However, in more realistic situations structures of different extensions activate at some instant instead of such single dipoles, which motivates us to

include structures into our model by correlating neighbouring neurons and thereby defining new degrees of freedom as regions of different sizes. This inclusion of such structures is straightforward in the SBL framework and we investigated its performance by conducting two simulations. In the first one random regions of different sizes were defined through the K-Means Clustering algorithm in a hierarchical fashion, where regions at a given scale were defined as subregions from the regions of the previous coarser scale. In practice such regions could be defined according to anatomical or functional information, but since at this point we are just interested in the algorithms performance, it's not crucial. Four different scales of resolutions were used, with 20, 80, 320 and 1280 regions (1.3), respectively. A simulation run consisted of randomly choosing a region of a scale and generating a signal using the forward equation in a time window modulated by a gaussian amplitude (Figure 4.11), which was then corrupted with noise of different SNR levels (0 - 20 dB). The corrupted signal was then used as input to the CA with all regions of all scales being **simultaneously** considered as degrees of freedom of our problem to reflect a situation where no a priori knowledge about the typical activity extension is available. 32 runs were conducted, 8 for each scale with half of them corresponding to superficial and the other half to deep source configurations. To measure the algorithms performance was used the time average of the Jaccard Index, also known as Intersection over Union, which measures the overlap between two sets (4.4), over the simulated activity time window between the estimate and the true activity vector. For low noise levels we found that all activities could be recovered perfectly and for high noise levels we noticed that performance decreased stronger for regions from coarser scales (Figures 4.14 and 4.13), which could be explained by the observation that the sources from the coarser scales had more closer signals originating from other sources than the sources from finer scales, leading to a higher probability of the inverse algorithm to commit mistakes for higher noise levels where the signals of different sources are too close from each other. This indicates that careful analysis is required to avoid defining degrees of freedom that are too close from each other in the signal space and some efficient procedure to do it is still to be developed.

In the second experiment we tried to recover focal activities with spatial extent by using focal structures (Figure 4.22) as degrees of freedom in a two step fashion, where in the first one large structures around each dipole of the original source space of 5120 dipoles (Figure 4.19) were used as degrees of freedom to scan regions of high activity probability over the whole cortex and in the second step we used focal structures of different sizes around each dipole of the surviving regions of the first step, which were chosen through an amplitude threshold of 0.9 of the maximum estimated amplitude, to allow a fine tuning of the amplitude profile of the true activity. The simulated activities were chosen to have a dipole with highest amplitude as the activity center and surrounding dipoles having amplitudes falling with distance (in triangle lattice edges) from it, which

were then all modulated by a common gaussian amplitude in a time window like in the first experiment to simulate an evoked response like dynamic. We made 100 simulation runs by randomly choosing activity centers from superficial and deep regions (50 for each) and simulated the signals through projection into the signal space with the corresponding lead field and adding gaussian independent noises with SNR ranging again from 0 to 20 dB according to definition (4.1). The resulting estimates were evaluated in two ways. In the first was used the time average over the simulated time window of the cosine between the angle of the true and the estimated activity vectors to analyse the proximity of the overall amplitude profiles. In the second one was used a metric to evaluate the distance between the estimated and true activity centers, where an estimated center was defined as the closest dipole of the original source space to the center of mass of the estimated activity vector weighted by power of the estimated amplitudes. We compared the performance of our algorithm with SBL without inclusion of structures, MNE and wMNE and observed a clear superiority of the former in both metrics, depicted in Figures 4.23 and 4.25, indicating the clear advantage of including structures when some prior knowledge is available. Focal activities are extensively investigated in focal epilepsies, where typically a small region of malfunctioning neurons are responsible for initiating a seizure and whose signature can be often identified as rapid signal spikes. Precise identification of such regions are crucial for a presurgical evaluation to eventually remove them in cases where patients do not respond to any drug treatment, which nowadays is commonly done through invasive methods by recording the brain activity through intracranial electrodes. Therefore a development of non invasive methods to precisely identify regions of epileptic foci would be a huge advance for both practical and ethical reasons. Another interesting application of finding activity foci is Brain Computer Interface (BCI), where a device should translate a person's electrical brain signals into meaningful associated actions. For evoked responses that are typically generated from small regions, an accurate source reconstruction method can greatly improve a BCI's performance.

Despite our positive preliminary simulation results, advances have to be made in future studies in at least to directions to evaluate and improve source reconstruction accuracy. In the first, more extensive simulations of more general activity dynamics and profiles should be carried out and the estimations results compared with other well known sparsity enforcing algorithms. Besides, an important improvement of the here presented methods would be the development of a procedure to automatically learn relevant structures such as in [47, 54]. The second direction refers to the development of techniques to extract more precise tissues information, such as the skull conductivity profile, which is crucial in EEG since it has a strong attenuating influence and smearing effect on its signals, thus severely limiting its spatial resolution. Such limitation can be overcome by MEG, which can be successfully integrated with EEG to improve reconstruction accuracy due

to their complementary nature, although it requires a much greater amount of resources. In addition to important applications such as in epilepsy investigation, electromagnetic brain imaging techniques allows us to gain a deeper understanding on brain dynamics, being thus a potential useful tool for investigating any conditions related with cognitive processes and even to open a path towards a deeper understanding of consciousness.

Appendices

Appendix A

Quasi-Static Approximation of the Maxwell's equations in head tissues

We show the validity of the quasi-static approximation of the Maxwell's Equations by comparing the conductive with capacitive effects in head tissues, as in [22].

First, we rewrite the continuity equation (1.6) using (1.2) and linearity $\vec{D} = \epsilon \vec{E}$:

$$\nabla \cdot \left(\vec{J} + \frac{\partial(\epsilon \vec{E})}{\partial t} \right) = 0. \quad (\text{A.1})$$

Now, consider a component of the Fourier Transform of the oscillating electric field:

$$\vec{E}(\vec{x}, t) = \vec{E}(\vec{x}) e^{i\omega t}, \quad (\text{A.2})$$

where $\omega = 2\pi f$ is the angular frequency, f is the frequency, i is the imaginary unit and $\vec{E}(\vec{x})$ the amplitude. Substituting (A.2) and Ohm's Law $\vec{J} = \sigma \vec{E}$ in (A.1), we get

$$\nabla \cdot \left(\sigma \vec{E} + i\omega \epsilon \vec{E} \right) = \nabla \cdot \left[\sigma \vec{E} \left(1 + \frac{i\omega \epsilon}{\sigma} \right) \right] = 0. \quad (\text{A.3})$$

Now, if we have that the ratio $R := \omega \epsilon / \sigma \ll 1$, then the second term of (A.3) can be neglected, thus equation (A.1) leads to $\nabla \cdot \vec{J} = 0$. But this is equivalent as approximating

$\frac{\partial \vec{D}}{\partial t} \approx 0$ in equation (1.5), thus yielding $\nabla \times \vec{H} = \vec{J}$. It turns out that the ratio R is indeed typically small in living tissues ($\sim 10^{-2} - 10^{-3}$) [22], such that $\frac{\partial \vec{D}}{\partial t} \approx 0$ in (1.5) is a reasonable approximation. Likewise, to neglect magnetic induction, take the curl of equation (1.3):

$$\nabla \times \nabla \times \vec{E} = -\frac{\partial}{\partial t}(\nabla \times \vec{B}). \quad (\text{A.4})$$

Substituting (1.5), (A.1) and $\vec{H} = \vec{B}/\mu$ in the right-hand side of (A.4) yields

$$\nabla \times \nabla \times \vec{E} = -\mu \frac{\partial}{\partial t} \left(\sigma \vec{E} + \epsilon \frac{\partial \vec{E}}{\partial t} \right). \quad (\text{A.5})$$

To approximate $\nabla \times \vec{E} \approx 0$ we require that:

$$\frac{\mu \left| \frac{\partial}{\partial t} \left(\sigma \vec{E} + \epsilon \frac{\partial \vec{E}}{\partial t} \right) \right|}{|\nabla \times \nabla \times \vec{E}|} \ll 1. \quad (\text{A.6})$$

Again, [22] shows that (A.6) is valid for frequencies up to roughly $1MHz$ and typical EEG frequencies are on the order of Hz up to kHz . Thus, we can safely approximate $\nabla \times \vec{E} \approx 0$ and together with $\nabla \times \vec{B} \approx \mu \vec{J}$ we have obtained the quasi-static approximation of the Maxwell's equations, where both fields $\vec{E}$ and $\vec{B}$ are decoupled and can therefore be treated separately.

Appendix B

Champagne Algorithm

The Champagne Algorithm (CA) [52] is an efficient algorithm to minimize following cost function:

$$\mathcal{L}(\gamma) = \log |\Sigma_V| + \frac{1}{T} \sum_{t=1}^T V_t' \Sigma_V^{-1} V_t, \quad (\text{B.1})$$

with $\Sigma_V = \beta I_M + G \Gamma G'$, $\Gamma = \text{diag}(\gamma)$, $\gamma = (\gamma_1, \dots, \gamma_N)$, $\gamma_i > 0$. We will describe the CA in some detail following [29]. First, we introduce an alternative cost function, which is an upper bound of (B.5). To do that, first notice that the term $\log |\Sigma_V|$ is a concave function of all γ_i and therefore we can introduce auxiliary variables $z \in \mathbb{R}^N$ and $z_0 \in \mathbb{R}$ such that:

$$z' \gamma - z_0 \geq \log |\Sigma_V|. \quad (\text{B.2})$$

For the second term of (3.14) we have following relation:

$$V_t' \Sigma_V^{-1} V_t = \min_{J_t} \left(\beta \|V_t - G J_t\|^2 + J_t' \Gamma^{-1} J_t \right). \quad (\text{B.3})$$

This can be shown by first observing that the minimum of the argument on the right-hand side of (B.3) is given at the mean value $\hat{J}_t$, since it is the negative of the argument of the posterior on $J_t|V_t$. This gives

$$\begin{aligned}
\frac{\beta}{2} \|V_t - G\hat{J}_t\|^2 + \hat{J}_t' \Gamma^{-1} \hat{J}_t &= \frac{\beta}{2} (V_t' V_t - 2\hat{J}_t' G' V_t + \hat{J}_t' G' G \hat{J}_t) + \frac{1}{2} \hat{J}_t' \Gamma^{-1} \hat{J}_t \\
&= \frac{1}{2} (\beta V_t' V_t - 2\hat{J}_t' \beta G' V_t + \hat{J}_t' \Sigma_J^{-1} \hat{J}_t) \\
&= \frac{1}{2} (\beta V_t' V_t - \hat{J}_t' \Sigma_J^{-1} \hat{J}_t) \\
&= \frac{1}{2} (\beta V_t' V_t - (\beta \Sigma_J G' V_t)' \Sigma_J^{-1} (\beta \Sigma_J G' V_t)) \\
&= \frac{1}{2} V_t' (\beta I_M - \beta G \Sigma_J G' \beta) V_t \\
&= \frac{1}{2} V_t' (\beta I_M - \beta G (\Gamma^{-1} + \beta G' G)^{-1} G' \beta) V_t \\
&= \frac{1}{2} V_t' (\beta I_M + G \Gamma G')^{-1} V_t \\
&= \frac{1}{2} V_t' \Sigma_V^{-1} V_t.
\end{aligned} \tag{B.4}$$

With (B.2) and (B.4) we can now define a new cost function $\tilde{\mathcal{L}}(\gamma, J, z)$ which is an upper bound on $\mathcal{L}(\gamma)$, given by

$$\tilde{\mathcal{L}}(\gamma, J, z) = \frac{1}{T} \sum_{t=1}^T (\beta \|V_t - G J_t\|^2 + J_t' \Gamma^{-1} J_t) + z' \gamma - z_0, \tag{B.5}$$

with $\tilde{\mathcal{L}}(\gamma, J, z) \geq \mathcal{L}(\gamma)$. Moreover, minimizing $\tilde{\mathcal{L}}(\gamma, J, z)$ with respect to γ, J and z is equivalent to minimizing $\mathcal{L}$ with respect to γ .

B.1 Update equation for z

To minimize $\tilde{\mathcal{L}}(\gamma, J, z)$ with respect to z , observe that it's the same as solving:

$$\hat{z} = \arg \min(z' \gamma - z_0), \quad s.t. \quad z' \gamma - z_0 \geq \log |\Sigma_V|. \tag{B.6}$$

Denoting E_j the diagonal $M \times M$ matrix with the j th element equal to 1 and the rest equal to 0, the above problem is solved by:

$$\begin{aligned}
\hat{z}_j &= \frac{\partial}{\partial \gamma_j} \log |\Sigma_V| = \text{Tr} \left(\Sigma_V^{-1} \frac{\partial}{\partial \gamma_j} \Sigma_V \right) = \text{Tr} \left(\Sigma_V^{-1} \frac{\partial}{\partial \gamma_j} (\beta^{-1} I_M + G \Gamma G') \right) = \\
&= \text{Tr} \left(\Sigma_V^{-1} G E_j G' \right) = \text{Tr} \left(\Sigma_V^{-1} G_j G'_j \right) = \text{Tr} \left(G'_j \Sigma_V^{-1} G_j \right) = \\
&= G'_j \Sigma_V^{-1} G_j.
\end{aligned} \tag{B.7}$$

where G_j is the j th column of G .

B.2 Update equation for γ

To minimize $\tilde{\mathcal{L}}$ with respect to γ , we have to solve

$$\hat{\gamma} = \arg \min_{\gamma} \left(\frac{1}{T} \sum_{t=1}^T J'_t \Gamma^{-1} J_t + z' \gamma \right) = \arg \min_{\gamma} \sum_{i=1}^N \left(\gamma_i z_i + \frac{\frac{1}{T} \sum_{t=1}^T J_{it}^2}{\gamma_i} \right). \tag{B.8}$$

Derivating the argument of the term on the right-hand side of (B.8) with respect to γ_j and equating to zero, we get

$$\hat{\gamma}_j = \sqrt{\frac{\frac{1}{T} \sum_{t=1}^T J_{it}^2}{z_j}}. \tag{B.9}$$

Now, the update equation for J_t is already given by (3.16) and putting all together, we can write an iterative update rule for the γ_j 's as

$$\hat{\gamma}_j^{(n+1)} = \frac{\hat{\gamma}_j^{(n)} \|G'_j (\Sigma_V^{(n)})^{-1} V\|_{\mathcal{F}}}{\sqrt{T} \sqrt{G'_j (\Sigma_V^{(n)})^{-1} G_j}}, \quad j = 1, \dots, N \tag{B.10}$$

$e_j = (0, \dots, 1, \dots, 0)^t$
 where $\Sigma_V^{(n)} = \beta^{-1} I_M + G \Gamma^{(n)} G'$, $\|\cdot\|_{\mathcal{F}}$ is the Frobenius norm and n is the iteration number.

Appendix C

Origin of Sparsity in the Evidence Maximization Framework

It is by no means obvious that minimization of the cost function (3.14) leads to sparse solutions. To gain an intuition about its sparsity enforcing property let's set $T = 1$ and consider both relations of our estimate $\hat{\gamma}$ and (B.3):

$$\hat{\gamma} = \arg \min \left(\log |\Sigma_V| + V' \Sigma_V^{-1} V \right), \quad (\text{C.1})$$

$$V' \Sigma_V^{-1} V = \min_J \left(\beta \|V - GJ\|^2 + J' \Gamma^{-1} J \right). \quad (\text{C.2})$$

Denoting the minimum value of our cost function as $\epsilon := \min_{\gamma} \left(\log |\Sigma_V| + V' \Sigma_V^{-1} V \right)$ and taking the minimum of both sides of (C.2) with respect to γ , we get

$$\min_{\gamma} (V' \Sigma_V^{-1} V) = \min_J \left(\beta \|V - GJ\|^2 + \min_{\gamma} J' \Gamma^{-1} J \right). \quad (\text{C.3})$$

Inserting (C.3) into the expression of ϵ and changing the order of the minimums leads to

$$\epsilon = \min_J \left(\beta \|V - GJ\|^2 + \min_{\gamma} \left(J' \Gamma^{-1} J + \log |\Sigma_V| \right) \right). \quad (\text{C.4})$$

From (C.4) we can immediately write our estimate of J as:

$$\hat{J} = \arg \min_J \left(\beta \|V - GJ\|^2 + \mathcal{R}(J) \right), \quad (\text{C.5})$$

where

$$\mathcal{R}(J) = \min_{\gamma} \left(J' \Gamma^{-1} J + \log |\Sigma_V| \right) = \min_{\gamma} \left(\sum_{j=1}^N \frac{J_j^2}{\gamma_j} + \log |\Sigma_V| \right) \quad (\text{C.6})$$

is the regularization term.

To be able to evaluate (C.6) analytically, we consider the simplification $G_i' G_j = \delta_{ij}$, where G_j is the j th column of G and δ_{ij} is the Kronecker function, which gives

$$\log |\Sigma_V| = \sum_{j=1}^N \log(\beta^{-1} + \gamma_j). \quad (\text{C.7})$$

So the regularization term (C.6) becomes:

$$\mathcal{R}(J) = \min_{\gamma} \left(\sum_{j=1}^N \frac{J_j^2}{\gamma_j} + \log(\beta^{-1} + \gamma_j) \right). \quad (\text{C.8})$$

The argument on the right-hand side of (C.8) is evaluated by setting its derivative with respect to γ_j to zero, yielding

$$\mathcal{R}(J) = \sum_{j=1}^N h(J_j), \quad (\text{C.9})$$

where

$$h(x) = \frac{2|x|}{\sqrt{x^2 + 4\beta^{-1} + |x|}} + \log \left(\beta^{-1} + \frac{x^2}{2} + \frac{1}{2}|x|\sqrt{x^2 + 4\beta^{-1}} \right). \quad (\text{C.10})$$

In graph (C.1) we see the function $h(x)$ (for $\beta = 1$) in some sense approximates the $L1$ norm $|x|$, which is commonly used as a sparsity enforcing regularizer due to its sharp peak around 0.

Figure C.1: *Graphs of $|x|$ and $h(x)$ for different precision values β with appropriate scaling and shifting to enable visualization of the similar shapes. Notice that $h(x)$ approximates $|x|$, which is well known for its sparsity enforcing property [29].*

Appendix D

Bayesian Model Selection

Suppose we have some data $\mathcal{D}$ available regarding a phenomenon obtained through an experiment and want to choose some underlying model $\mathcal{M}$ to describe it, which is the central problem of science. A model family can, for example, be specified by a set of functions and equations and a particular model of this family is given by a specific choice of the set of parameters and hyperparameters involved. Since there is almost always missing information, it is justified to apply probability theory and write the posterior probability $P(\mathcal{M}|\mathcal{D})$, meaning the probability of model $\mathcal{M}$ being "true" given the dataset $\mathcal{D}$, i.e., the probability of the model $\mathcal{M}$ describing the data $\mathcal{D}$. Suppose further, for simplicity, that we have two candidate models $\mathcal{M}_1$ and $\mathcal{M}_2$ available to describe $\mathcal{D}$ and want to choose one among them. This can be done by choosing the most probable one given the data, i.e. the one with higher posterior probability $P(\mathcal{M}|\mathcal{D})$, which can be calculated through application of the Bayes' theorem:

$$P(\mathcal{M}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{M})P(\mathcal{M})}{P(\mathcal{D})}, \quad (\text{D.1})$$

where $P(\mathcal{D}|\mathcal{M})$ is the likelihood, $P(\mathcal{M})$ is the model prior and $P(\mathcal{D})$ is the normalization factor. To compare our models $\mathcal{M}_1$ and $\mathcal{M}_2$ we can evaluate the ratio of their posterior probabilities, also called odds ($\mathcal{O}$):

$$\mathcal{O} = \frac{P(\mathcal{M}_1|\mathcal{D})}{P(\mathcal{M}_2|\mathcal{D})} = \frac{P(\mathcal{D}|\mathcal{M}_1)P(\mathcal{M}_1)}{P(\mathcal{D}|\mathcal{M}_2)P(\mathcal{M}_2)}. \quad (\text{D.2})$$

Observe that the term $P(\mathcal{D})$ vanished, which is desirable since it is typically infeasible to calculate. If we further do not have any reason to favour any of the models, their prior

probabilities coincide and the ratio simplifies to

$$\mathcal{O} = \frac{P(\mathcal{D}|\mathcal{M}_1)}{P(\mathcal{D}|\mathcal{M}_2)}. \quad (\text{D.3})$$

Model selection would then proceed by choosing $\mathcal{M}_1$ if $\mathcal{O} > 1$ and $\mathcal{M}_2$ otherwise. It can be shown that this procedure tends to favour simpler models, which can be interpreted as the essence of the famous Occam's Razor [17, 32]. Lastly, observe that the likelihood $P(\mathcal{D}|\mathcal{M})$ corresponds to the normalizing factor $P(V|I)$ of equation (3.5) with $V = \mathcal{D}$ and $\mathcal{M} = I$, which is the key of the Sparse Bayesian Learning framework that aims at evidence maximization in order to choose the models hyperparameters. By maximizing the evidence, one is choosing the model with lowest complexity from a Bayesian perspective.

Appendix E

Expectation Maximization (EM) Algorithm

E.1 Derivation of the EM equations for a general case

Here we derive the EM algorithm [4, 29], a commonly used algorithm used in optimization problems. To begin with we write the forward equation at a time instant t :

$$V_t = GJ_t + \xi_t, \quad (\text{E.1})$$

where $V_t \in \mathbb{R}^M$ is the signal vector, $J_t \in \mathbb{R}^N$ the source vector, $G \in \mathbb{R}^{M \times N}$ the lead field and $\xi \in \mathbb{R}^M$ the noise vector, all considered at time instant $t = 1, \dots, T$. We assume for now general noise and prior precision matrices Λ and Γ , giving $\xi_t \sim \mathcal{N}(\xi_t|0, \Lambda^{-1})$ and therefore $P(V_t|J_t, \Lambda) = \mathcal{N}(V_t|J_t, \Lambda^{-1})$ and $P(J_t|\Gamma) = \mathcal{N}(J_t|0, \Gamma^{-1})$. The posterior on J_t we write using Bayes' theorem:

$$P(J_t|V_t, \Lambda, \Gamma) = \frac{P(V_t|J_t, \Lambda)P(J_t|\Gamma)}{P(V_t|\Lambda, \Gamma)} = \mathcal{N}(J_t|\bar{J}_t, \Omega^{-1}), \quad (\text{E.2})$$

with posterior mean and empirical data precision given by, respectively:

$$\begin{aligned} \bar{J}_t &= \Gamma^{-1}G'\Psi V_t, \\ \Psi &= \left(\Lambda^{-1} + G\Gamma^{-1}G'\right)^{-1}. \end{aligned} \quad (\text{E.3})$$

Since we are considering no time correlation, the joint distributions on all t 's are just products of independent gaussians:

$$\begin{aligned}
P(J|\Gamma) &= P(J_1, \dots, J_T|\Gamma) = \prod_{t=1}^T P(J_t|\Gamma) = \prod_{t=1}^T \mathcal{N}(J_t|0, \Gamma^{-1}), \\
P(V|J, \Lambda) &= P(V_1, \dots, V_T|J_1, \dots, J_T, \Lambda) = \prod_{t=1}^T P(V_t|J_t, \Lambda) = \prod_{t=1}^T \mathcal{N}(V_t|GJ_t, \Lambda^{-1}), \\
P(J|V, \Lambda, \Gamma) &= P(J_1, \dots, J_T|V_1, \dots, V_T, \Lambda, \Gamma) = \prod_{t=1}^T P(J_t|V_t, \Lambda, \Gamma) = \prod_{t=1}^T \mathcal{N}(J_t|\bar{J}_t, \Omega^{-1}).
\end{aligned} \tag{E.4}$$

Now, to use the inferred parameters (E.3) of the posterior we have to estimate the precision matrices Λ and Γ , which are our hyperparameters. As discussed before, a natural way would be to select simplest model from a Bayesian perspective by choosing the set of hyperparameters that maximize the model evidence $P(V|\Gamma, \Lambda)$, given by:

$$P(V|\Gamma, \Lambda) = \int P(V|J, \Lambda)P(J|\Gamma)dJ = \mathcal{N}(V|0, \Gamma), \tag{E.5}$$

with $\Gamma^{-1} = \Lambda^{-1} + G\Gamma G'$, or equivalently that minimize the convenient cost function $-2\log P(V|\Gamma, \Lambda)$. This approach corresponds to the Sparse Bayesian Learning (SBL) framework, and although direct maximization of $P(V|\Gamma, \Lambda)$ is possible, it is however complicated enough to motivate us to pursue another path. We consider instead the logarithm of the joint distribution $P(J, V|\Gamma, \Lambda)$:

$$\log P(V, J|\Gamma, \Lambda) = \log P(V|J, \Lambda) + \log p(J|\Gamma), \tag{E.6}$$

which we would like to maximize. Since J is not observable, we can average it out using the posterior $P(J|V, \Gamma, \Lambda)$, which results in the average data likelihood Ξ :

$$\begin{aligned}
\Xi(\Gamma, \Lambda) &= \int P(J|V, \Gamma, \Lambda) \log P(V, J|\Gamma, \Lambda) dJ, \\
&= \langle \log P(V, J|\Gamma, \Lambda) \rangle, \\
&= \langle \log P(V|J, \Lambda) \rangle + \langle \log P(J|\Gamma) \rangle,
\end{aligned} \tag{E.7}$$

where $\langle \cdot \rangle$ refers to the expected value with respect to the posterior $P(J|V, \Gamma, \Lambda)$. It turns out that maximizing (E.7) is much easier than maximizing (E.5) [4]. Finally, our precision matrices are estimated through:

$$\begin{aligned}
\hat{\Lambda} &= \arg \max_{\Lambda} \Xi(\Gamma, \Lambda), \\
\hat{\Gamma} &= \arg \max_{\Gamma} \Xi(\Gamma, \Lambda).
\end{aligned} \tag{E.8}$$

Using our expressions in (E.4), we can write $\log P(V, J|\Gamma, \Lambda)$ explicitly:

$$\log P(V, J|\Gamma, \Lambda) = \frac{T}{2} \log |\Gamma| - \frac{1}{2} \sum_{t=1}^T J_t' \Gamma J_t + \frac{T}{2} \log |\Lambda| - \frac{1}{2} \sum_{t=1}^T (V_t - G J_t)' \Lambda (V_t - G J_t), \tag{E.9}$$

which leads to

$$\Xi(\Gamma, \Lambda) = \frac{T}{2} \log |\Gamma| - \frac{1}{2} \left\langle \sum_{t=1}^T J_t' \Gamma J_t \right\rangle + \frac{T}{2} \log |\Lambda| - \frac{1}{2} \left\langle \sum_{t=1}^T (V_t - G J_t)' \Lambda (V_t - G J_t) \right\rangle. \tag{E.10}$$

E.1.1 Update Equation for Γ

To calculate $\hat{\Gamma} = \arg \max_{\Gamma} \Xi(\Gamma, \Lambda)$, we set the derivative of Ξ with respect to Γ equal to 0:

$$\frac{\partial}{\partial \Gamma} \Xi(\Gamma, \Lambda) = \frac{T}{2} \Gamma^{-1} - \frac{1}{2} \left\langle \sum_{t=1}^T J_t J_t' \right\rangle = 0. \tag{E.11}$$

Using the fact that $\langle J_t J_t' \rangle = \Omega^{-1} + \bar{J}_t \bar{J}_t'$, we get from (E.11):

$$\Gamma^{-1} = \frac{1}{T} \sum_{t=1}^T \bar{J}_t \bar{J}_t' + \Omega. \quad (\text{E.12})$$

E.1.2 Update Equation for Λ

To calculate $\hat{\Lambda} = \arg \max_{\Lambda} \Xi(\Gamma, \Lambda)$, we set the derivative of Ξ with respect to Λ equal to 0:

$$\frac{\partial}{\partial \Lambda} \Xi(\Gamma, \Lambda) = \frac{T}{2} \Lambda^{-1} - \frac{1}{2} \left\langle \sum_{t=1}^T (V_t - G J_t)(V_t - G J_t)' \right\rangle = 0. \quad (\text{E.13})$$

Rearranging terms, we obtain

$$\Lambda^{-1} = \frac{1}{T} \left\langle \sum_{t=1}^T (V_t - G J_t)(V_t - G J_t)' \right\rangle. \quad (\text{E.14})$$

Again, using $\langle J_t J_t' \rangle = \Omega^{-1} + \bar{J}_t \bar{J}_t'$, the terms of the sum on the right-hand side of (E.14) can be rewritten as

$$\begin{aligned} \langle (V_t - G J_t)(V_t - G J_t)' \rangle &= \langle V_t V_t' - 2G J_t V_t' + G J_t J_t' G' \rangle \\ &= V_t V_t' - 2G \bar{J}_t V_t' + G(\bar{J}_t \bar{J}_t' + \Omega^{-1})G' \\ &= (V_t - G \bar{J}_t)(V_t - G \bar{J}_t)' + G \Omega^{-1} G'. \end{aligned} \quad (\text{E.15})$$

Finally, inserting (E.15) in (E.14) results in

$$\Lambda^{-1} = \frac{1}{T} \sum_{t=1}^T (V_t - G \bar{J}_t)(V_t - G \bar{J}_t)' + G \Omega^{-1} G'. \quad (\text{E.16})$$

E.1.3 EM Algorithm

The EM algorithm consists in recursively estimating the posterior parameters $\bar{J}_t$ and Ω (Expectation or E step) and the hyperparameters Γ and Λ (Maximization or M step). Starting from initial parameters $\Gamma_{(0)}$ and $\Lambda_{(0)}$ we have following iteration steps:

- **E step:**

$$\begin{aligned}\Psi_{(n)} &= \left(\Lambda_{(n)}^{-1} + G' \Gamma_{(n)}^{-1} G \right)^{-1}, \\ \bar{J}_{t(n)} &= \Gamma_{(n)}^{-1} G' \Psi_{(n)} V_t.\end{aligned}\tag{E.17}$$

- **M step:**

$$\begin{aligned}\Gamma_{(n+1)}^{-1} &= \frac{1}{T} \sum_{t=1}^T \bar{J}_{t(n)} \bar{J}_{t(n)}' + \Omega_{(n)}, \\ \Lambda_{(n+1)}^{-1} &= \frac{1}{T} \sum_{t=1}^T (V_t - G \bar{J}_{t(n)})(V_t - G \bar{J}_{t(n)})' + G \Omega_{(n)}^{-1} G',\end{aligned}\tag{E.18}$$

where n is the iteration number. It can be shown that this recursion does indeed increase the evidence $P(V|\Gamma, \Lambda)$ [4]. The EM algorithm is widely used in several applications where the equations can't be solved analytically and because of its relative simplicity compared to other methods. Moreover, it is not restricted to linear problems, so in general if we consider a problem with X , Z and θ corresponding to observed data, latent data and unknown parameters, respectively, we can summarize the E and M steps as

- **E step:**

$$\Xi(\theta|\theta^{(n)}) = \langle \log P(X, Z|\theta) \rangle_{Z|X, \theta^{(n)}},\tag{E.19}$$

- **M step:**

$$\theta^{(n+1)} = \arg \max_{\theta} \Xi(\theta|\theta^{(n)}),\tag{E.20}$$

where n is the iteration number and $\theta^{(0)}$ is chosen to some appropriate initial values. In our problem we have $X = V$, $Z = J$ and $\theta = (\Gamma, \Lambda)$. A possible stop criterium of this recursion can be chosen to be based on the change of the average data likelihood Ω or and the parameters θ between two iterations, i.e., if the change of either are below some threshold between two iterations, then the algorithm can be stopped.

E.2 EM algorithm for the MNE estimate

As a particular application of the EM algorithm, we derive the update equations for the special case $\Gamma = \gamma I_N$ and $\Lambda = \beta I_M$, where I_K is the $K \times K$ identity matrix, $\beta > 0$ is the precision or inverse of the variance and $\gamma > 0$ is the regularization parameter of the MNE estimator. Using equation (E.10), we get

$$\Xi(\gamma, \beta) = \frac{NT}{2} \log \gamma - \frac{\gamma}{2} \left\langle \sum_{t=1}^T J_t' J_t \right\rangle + \frac{MT}{2} \log \beta - \frac{\beta}{2} \left\langle \sum_{t=1}^T (V_t - GJ_t)' (V_t - GJ_t) \right\rangle. \quad (\text{E.21})$$

Further, using

$$\frac{\partial}{\partial \gamma} \Xi(\gamma, \beta) = \frac{NT}{2\gamma} - \frac{1}{2} \left\langle \sum_{t=1}^T J_t' J_t \right\rangle = 0 \quad (\text{E.22})$$

and

$$\begin{aligned} \left\langle \sum_{t=1}^T J_t' J_t \right\rangle &= Tr \left(\sum_{t=1}^T \langle J_t J_t' \rangle \right) \\ &= Tr \left(\sum_{t=1}^T \bar{J}_t \bar{J}_t' + K\Gamma^{-1} \right) \\ &= \sum_{t=1}^T \bar{J}_t' \bar{J}_t + KTr(\Gamma^{-1}), \end{aligned} \quad (\text{E.23})$$

we get the update equation for γ :

$$\hat{\gamma}^{-1} = \frac{1}{NT} \left\langle \sum_{t=1}^T J_t' J_t \right\rangle = \frac{1}{N} \left(\sum_{t=1}^T \|\bar{J}_t\|^2 + Tr(\Gamma^{-1}) \right) = \frac{1}{N} \left(\sum_{t=1}^T \|\bar{J}_t\|^2 + \gamma^{-1}N \right). \quad (\text{E.24})$$

Regarding the parameter β we have:

$$\hat{\beta}^{-1} = \frac{1}{MT} \left\langle \sum_{t=1}^T (V_t - GJ_t)' (V_t - GJ_t) \right\rangle. \quad (\text{E.25})$$

Now, we using the relations

$$\begin{aligned} \left\langle \sum_{t=1}^T (V_t - GJ_t)' (V_t - GJ_t) \right\rangle &= \left\langle \sum_{t=1}^T (V_t' V_t - J_t' G' - V_t' G J_t + J_t' G' G J_t) \right\rangle \\ &= \sum_{t=1}^T \left(V_t' V_t - \bar{J}_t' G' J_t - V_t' G \bar{J}_t + \langle J_t' G' G J_t \rangle \right) \end{aligned} \quad (\text{E.26})$$

and

$$\begin{aligned} \langle J_t' G' G J_t \rangle &= \langle \text{Tr} (G' G J_t J_t') \rangle \\ &= \text{Tr} (G' G \langle J_t J_t' \rangle) \\ &= \text{Tr} \left(G' G \left(\bar{J}_t \bar{J}_t' + \Gamma^{-1} \right) \right) \\ &= \bar{J}_t' G' G \bar{J}_t + \gamma^{-1} \text{Tr} (G' G), \end{aligned} \quad (\text{E.27})$$

we finally obtain

$$\hat{\beta}^{-1} = \frac{1}{M} \left(\frac{1}{T} \sum_{t=1}^T \|V_t - G\bar{J}_t\|^2 + \gamma^{-1} \text{Tr} (G' G) \right). \quad (\text{E.28})$$

To summarize, we have following update equations:

• **E step:**

$$\begin{aligned} \Psi_{(n)} &= \left(\beta_{(n)}^{-1} I_M + \gamma_{(n)}^{-1} G G' \right)^{-1}, \\ \bar{J}_{t(n)} &= \gamma_{(n)}^{-1} G' \Psi_{(n)} V_t. \end{aligned} \quad (\text{E.29})$$

• **M step:**

$$\begin{aligned}
\gamma_{(n+1)}^{-1} &= \frac{1}{N} \sum_{t=1}^T \|\bar{J}_{t(n)}\|^2 + \gamma_{(n)}^{-1}, \\
\beta_{(n+1)}^{-1} &= \frac{1}{M} \left(\frac{1}{T} \sum_{t=1}^T \|V_t - G\bar{J}_{t(n)}\|^2 + \gamma_{(n)}^{-1} \text{Tr}(G'G) \right).
\end{aligned}
\tag{E.30}$$

Appendix F

K-Means Clustering

In order to define our areas of activity of different scales, we use the K-Means Clustering algorithm [31]. The idea is to group points into a given number of clusters such that within cluster variance is as small as possible. More specifically, the problem can be stated as follows. Given a set of n d -dimensional points $P = \{\vec{x}_i\}_{i=1}^n$, $\vec{x}_i \in \mathbb{R}^d \forall i$, the k -means clustering aims to partition P into k sets $S = \{S_j\}_{j=1}^k$, $k \leq n$, such that the sum of within-cluster variances is a minimum. Formally, the problem is to solve:

$$\arg \min_S \sum_{i=1}^k \sum_{\vec{x} \in S_i} \|\vec{x} - \vec{\mu}_i\|^2, \quad (\text{F.1})$$

where $\vec{\mu}_i$ are the mean of points in S_i , i.e. :

$$\vec{\mu}_i = \frac{1}{|S_i|} \sum_{\vec{x}_j \in S_i} \vec{x}_j,$$

where $|A|$ is the cardinality of set A . Finding the optimal solution to (F.1) is a NP-hard problem and therefore there are several heuristic algorithms like the classic Lloyd's algorithm, or just k-means algorithm [31]. These limitations, however, do not concern us since our only purpose here is just to be able to define regions of activities on the cortex which have some typical desired sizes, so any other clustering methods could be equally applied.

Appendix G

Woodbury Identity

If A , C , V and U denote matrices with the correct (conformable) sizes and with A and C invertible, we have the the Woodbury identity [53]:

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U \left(C^{-1} + VA^{-1}U \right)^{-1} VA^{-1}. \quad (\text{G.1})$$

To show (G.1), first consider following identities:

$$\begin{aligned} U + UCV A^{-1}U &= UC \left(C^{-1} + VA^{-1}U \right) = (A + UCV) A^{-1}U \\ (A + UCV)^{-1} UC &= A^{-1}U \left(C^{-1} + VA^{-1}U \right)^{-1}. \end{aligned} \quad (\text{G.2})$$

Then,

$$\begin{aligned} A^{-1} &= (A + UCV)^{-1} (A + UCV) A^{-1} \\ &= (A + UCV)^{-1} (I + UCV A^{-1}) \\ &= (A + UCV)^{-1} + (A + UCV)^{-1} UCV A^{-1} \\ &= (A + UCV)^{-1} + A^{-1}U \left(C^{-1} + VA^{-1}U \right)^{-1} VA^{-1}. \end{aligned} \quad (\text{G.3})$$

In chapter 3 we wrote two expressions (3.4, 3.18) for the posterior mean of the source vector $\hat{J}_t$, which we summarize here:

$$\hat{J}_t = \beta \left(\Gamma^{-1} + \beta G'G \right)^{-1} G'V_t = \Gamma G' \left(\beta^{-1}I + GTG' \right)^{-1} V_t. \quad (\text{G.4})$$

The advantage of using the right-hand side of (G.4) is that it usually requires an inversion of a much smaller matrix ($M \times M$) than it is required in the middle expression ($N \times N$), where N is the dimensionality of the source space and M is the dimensionality of the signal space. Computationally, the inversion of a $n \times n$ matrix has typically a complexity of $O(n^3)$, which means that for typical ratios N/M around 1000, inverting a $N \times N$ matrix is about 10^9 times slower than inverting a $M \times M$ matrix!

We show (G.4) by first identifying $A = \Gamma^{-1}$, $C = \beta I$, $U = G'$, $V = G$ and using (G.1) to write

$$\left(\Gamma^{-1} + \beta G'G\right)^{-1} = \Gamma - \Gamma G' \left(\beta^{-1}I + G\Gamma G'\right)^{-1} G\Gamma. \quad (\text{G.5})$$

Now, we left multiply the factor middle term of (G.4) by the inverse of the factor on the right-hand side of (G.4) and use (G.5):

$$\begin{aligned} & \left[\Gamma G' \left(\beta^{-1}I + G\Gamma G'\right)^{-1}\right]^{-1} \beta \left(\Gamma^{-1} + \beta G'G\right)^{-1} G' = \\ &= \left(\beta^{-1}I + G\Gamma G'\right) (\Gamma G')^{-1} \beta \left[\Gamma - \Gamma G' \left(\beta^{-1}I + G\Gamma G'\right)^{-1} G\Gamma\right] G' = \\ &= \left(\beta^{-1}I + G\Gamma G'\right) (\Gamma G')^{-1} \Gamma G' \left[\beta I - \beta \left(\beta^{-1}I + G\Gamma G'\right)^{-1} G\Gamma G'\right] = \\ &= \left(\beta^{-1}I + G\Gamma G'\right) \left[\beta I - \beta \left(\beta^{-1}I + G\Gamma G'\right)^{-1} G\Gamma G'\right] = \\ &= I + \beta G\Gamma G' - \beta I G\Gamma G' = \\ &= I. \end{aligned} \quad (\text{G.6})$$

Bibliography

- [1] Z. A. Acar, C. E. Acar and S. Makeig. *Simultaneous head tissue conductivity and EEG source location estimation*. NeuroImage, Volume 124, Pages 168-180, 2016.
- [2] S. da Rocha Amaral, L. A. Baccalá, L. S. Barbosa and N. Caticha. *Backward renormalization-group inference of cortical dipole sources and neural connectivity efficacy*. Physical Review E 95, 062415, 2017.
- [3] J. Berger. *Statistical Decision Theory and Bayesian Analysis 2nd ed.* Springer-Verlag, New York, 1985.
- [4] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag New York, Inc. Secaucus, NJ, USA, 2006.
- [5] K. Brodmann. *Vergleichende Lokalisationslehre der Großhirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Barth, Leipzig, 1909
- [6] C. Cai, K. Sekihara and S. Nagarajan. *Hierarchical Multiscale Bayesian Algorithms for Biomagnetic Brain Imaging*. 8th International IEEE EMBS Conference on Neural Engineering Shanghai, China, 2017.
- [7] N. Caticha. *Source Localization by Entropic Inference and Backward Renormalization Group Priors*. Entropy, 17:2573-2589, 2015.
- [8] F. Costa, H. Batatia, T. Oberlin, C. D’Giano and J. Y. Tournier. *Bayesian EEG Source Localization Using a Structured Sparsity Prior*. NeuroImage, Volume 144, Part A., Pages 142-152, 2017.
- [9] A. M Dale , A. K. Liu, B. R. Fischl, R. L. Buckner, J. W. Belliveau , J. D. Lewine and E. Halgren *Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity*. Neuron 26(1):55-67, 2000.
- [10] D. Geselowitz. *On Bioelectric Potentials in an Inhomogeneous Volume Conductor*. Biophysical Journal 7(1):1-11, 1967.
- [11] G. Kutyniok. *Compressed Sensing: Theory and Applications*. <https://arxiv.org/abs/1203.3815>, 2012.

- [12] <http://surfer.nmr.mgh.harvard.edu/>
- [13] A. Gramfort, M. Luessi, E. Larson, D. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen and M. Hämäläinen. *MNE software for processing MEG and EEG data*. NeuroImage, Volume 86, Pages 446-460, ISSN 1053-8119, 2014.
- [14] A. Gramfort, M. Luessi, E. Larson, D. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen and M. Hämäläinen. *MEG and EEG data analysis with MNE-Python*. Frontiers in Neuroscience, Volume 7, ISSN 1662-453X, 2013.
- [15] S. Baillet, J. C. Mosher and R. M. Leahy. *Electromagnetic Brain Mapping*. IEEE Signal Processing Magazine, 18(6):14-30, 2001.
- [16] T. C. Ferree, K. J. Eriksen and D. M. Tucker. *Regional Head Tissue Conductivity Estimation for Improved EEG Analysis*. IEEE Trans. Biomed. Engineering 47(12), 2000.
- [17] E. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, doi:10.1017/CBO9780511790423, 2003.
- [18] Eric R Kandel, James H. Schwartz and Thomas M. Jessell. *Principles of Neural Science*. 4th ed. New York: McGraw-Hill, Health Professions Division, 2000.
- [19] J. C. Mosher, R. M. Leahy and P. S. Lewis. *EEG and MEG: Forward Solutions for Inverse Methods*. IEEE transactions on bio-medical engineering, 46(3):245-59, 1999.
- [20] R. D. Pascual-Marqui. *Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details*. *Methods Find Exp Clin Pharmacol* 24 Suppl D:5-12, 2002.
- [21] J. W. Phillips, R. M. Leahy and J. C. Mosher. *MEG-Based Imaging of Focal Neuronal Current Sources*. IEEE transactions on medical imaging, vol. 16, NO. 3, 1997.
- [22] P. L. Nunez, R. Srinivasan. *Electric Fields of the Brain: The Neurophysics of EEG, 2nd Edition*. Oxford University Press.
- [23] M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila and O. V. Lounasmaa. *Magnetoencephalography - theory, instrumentation, and applications to noninvasive studies of the working human brain*, 1993.
- [24] M. Hämäläinen and R. Ilmoniemi. *Interpreting magnetic fields of the brain: minimum norm estimates*. Medical and Biological Engineering and Computing, 32:35-42. 10.1007/BF02512476

- [25] P. Krishnaswamy, G. O. Henao, J. Ahveninen, S. Khan, B. Babadi, J. E. Iglesias, M. S. Hamalainen and P. L. Purdon. *Sparsity Enables Estimation of both Subcortical and Cortical Activity from MEG and EEG*. <https://arxiv.org/abs/1706.08041>, 2017.
- [26] J. D. Jackson. *Classical Electrodynamics, Third Edition*. John Wiley & Sons, Inc, 1998.
- [27] M. Joy, A. Nachmann, T. DeMonte, D. Wang and W. Ma. *Electrical Conductivity Imaging using MRI Measurement of the Magnetic Field Vector*. 31st Annual International Conference of the IEEE EMBS Minneapolis, Minnesota, USA, September 2-6, 2009.
- [28] U. Katscher, D.-H. Kim and J. K. Seo. *Recent Progress and Future Challenges in MR Electric Properties Tomography*. Computational and Mathematical Methods in Medicine, 2013. <http://dx.doi.org/10.1155/2013/546562>.
- [29] K. Sekihara and S. Nagarajan-. *Electromagnetic Brain Imaging, A Bayesian Perspective*. Springer, 2015.
- [30] J. Kybic, M. Clerc, T. Abboud, O. Faugeras, R. Keriven and T. Papadopoulos. *A common formalism for the integral formulations of the forward EEG problem*. IEEE Transactions on Medical Imaging, 24(1):12-18, 2005.
- [31] S. P. Lloyd. *Least Squares Quantization in PCM*. IEE Transactions on Information, 28(2):129-137, 1982.
- [32] D. J. C. MacKay. *Bayesian interpolation*. Neural Computation, 4(3):415-447, 1992.
- [33] J. Mairal, G. Sapiro and M. Elad. *Learning Multiscale Sparse Representations for Image and Video Restoration*. Multiscale Model. Simul. Vol. 7, No. 1, pp. 214-241. Society for Industrial Applied Mathematics, 2008.
- [34] J. C. de Munck. *A linear discretization of the volume conductor boundary integral equation using analytically integrated elements*. IEEE Trans. Biomed. Eng. 39(9), 986-990, 1992.
- [35] R. M. Neal. *Bayesian Learning for Neural Networks*. Springer-Verlag, New York, 1996.
- [36] Y. Petrov, J. Nador, C. Hughes, S. Tran, O. Yavuzcetin and S. Sridhar. *Ultra-dense EEG sampling results in two-fold increase of functional brain information*. NeuroImage, Volume 90, Pages 140-145, 2014.

- [37] J. Rahola and S. Tissari. *Iterative solution of dense linear systems arising from the electrostatic integral equation in MEG*. Physics in Medicine and Biology 47:961-975, 2002.
- [38] S. Saha, Y. a. I. Nesterets, R. Rana, M. Tahtali, F. d. Hoog and T. E. Gureyev. *EEG source localization using a sparsity prior based on Brodmann areas*. <https://arxiv.org/abs/1406.2434>, 2014.
- [39] M. Sato, T. Yoshioka, S. Kajihara, K. Toyama, N. Goda, K. Doya and M. Kawato. *Hierarchical Bayesian estimation for MEG inverse problem*. Science and Technology, 25:806-856, 2004.
- [40] J. Skilling. *On parameter estimation and quantified MaxEnt*. In Maximum Entropy and Bayesian Methods, Laramie, 1990, W. T. Grandy, Jr., and L. H. Schick, eds., pp.267-273. Kluwer, Dordrecht.
- [41] I. Tanzer. *Numerical Modelling in Electro- and Magnetoencephalography*. PhD thesis, Helsinki University of Technology, 2006.
- [42] Tikhonov, A. N.; V. Y. Arsenin. *Solution of Ill-posed Problems*. Washington: Winston & Sons. ISBN 0-470-99124-0, 1977.
- [43] M. E. Tipping. *The relevance vector machine*. Neural Inform. Process. Syst., vol. 12, pp. 652-658, 2000.
- [44] M. E. Tipping. *Sparse Bayesian Learning and the relevance vector machine*. J. Mach. Learn. Res. 1,211-244, 2001.
- [45] A. Varsavsky, I. Mareels and M. Cook. *Epileptic seizures and the EEG: measurement, models, detection and prediction*. CRC Press, 2016.
- [46] V.S. Vladimirov. *A Collection of Problems on the Equations of Mathematical Physics*. Mir Publishers Moscow, Springer-Verlag Berlin Heidelberg GmbH, 1986.
- [47] Y. Wang, D. Wipf, J. M. Yun, W. Chen and I. Wassel. *Clustered Sparse Bayesian Learning*. Uncertainty in Artificial Intelligence (UAI), 2015.
- [48] D. R. Wilton, S. M. Rao, A. W. Glisson, D. H. Schaubert, O. M. Al-Bundak and C. M. Butler. *Potential integrals for uniform and linear source distributions on polygonal and polyhedral domains*. IEEE Trans. Antenn. Propag. 32(3), 276-281, 1984.
- [49] D. Wipf, B. D. Rao. *Sparse Bayesian Learning for Basis Selection*. IEEE Trans. Signal Process. 52(8), 2153-2164.

- [50] D. Wipf, R. Ramírez, J. Palmer, S. Makeig and B. Rao. *Analysis of Empirical Bayesian Methods for Neuroelectromagnetic Source Localization*. Adv. Neural Inf. Process. Syst. 19, MIT Press, 2007.
- [51] D. Wipf and S. Nagarajan. *A unified Bayesian framework for MEG/EEG source imaging*. NeuroImage, Volume 44, Pages 947-966, 2009.
- [52] D. Wipf, J. P. Owen, H. T. Attias, K. Sekihara and S. Nagarajan. *Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG*. NeuroImage, Volume 49, Pages 641-655, 2010.
- [53] M. A. Woodbury. *Inverting modified matrices*. Memorandum Rept. 42, Statistical Research Group, Princeton University, Princeton, NJ, 1950.
- [54] A. Wu, M. Park, O. Koyejo and J. W. Pillow. *Sparse Bayesian structure learning with dependent relevance determination prior*. Advances in Neural Information Processing Systems 27, 1628-1636, 2014.
- [55] <https://www.moleculardevices.com/what-action-potential>
- [56] http://hplusmagazine.com/wp-content/uploads/neuron_structure2.jpg
- [57] <https://thesalience.files.wordpress.com/2013/05/synapse-labeled1-300x271.jpg>