

Inverse problems in functional brain imaging

Reference material for the MVA and MSV Master 2
course

Maureen Clerc
Théo Papadopulo
Bertrand Thirion

October 3, 2022

Contents

1	Introduction	7
I	Physical and Physiological modeling	17
2	Origin of functional measurements	21
2.1	Origin of the fMRI signal: the BOLD effect	21
2.2	Origin of magnetic and electric measurements	22
2.2.1	Neural current sources	22
2.2.2	Action potentials and postsynaptic potentials	24
2.2.3	Estimates of dipole strengths	24
2.2.4	Electromagnetic propagation	26
3	Forward problem computation	31
3.1	Introduction	31
3.2	Simple geometries	31
3.2.1	Current dipole in infinite homogeneous medium	31
3.2.2	Silent sources	32
3.3	Semi-realistic model	33
3.3.1	Magnetic field computation	33
3.3.2	Electric potential computation	36
3.4	Realistic model	38
3.4.1	A variational formulation of the forward problem . . .	40
3.4.2	Discretization of the FEM forward problem	43
3.4.3	Solving the FEM forward problem	47
4	Leadfield computation	49
4.1	The leadfield	49
4.2	The adjoint method	50
4.3	Application to M/EEG leadfields	51
4.3.1	EEG leadfield	52
4.3.2	MEG leadfield	53

4.4	Incorporating sensor geometry	54
4.4.1	MEG: Squids geometry	54
4.4.2	EEG: surface electrodes	54
4.5	Results	55
5	Geometric modeling of the head	57
5.1	Magnetic Resonance Imaging	57
5.1.1	Basic principle of NMR	57
5.1.2	MRI scanning	59
5.2	Segmentation of Magnetic Resonance Images (MRI)	61
II	Analysis of functional imaging data	63
6	Functional Magnetic Resonance Imaging	65
6.1	Background on fMRI data analysis	65
6.1.1	Principle of an activation study	65
6.1.2	A short perspective on functional neuroimaging.	66
6.1.3	What we need to know about the data	66
6.2	fMRI data analysis: State of the art	67
6.2.1	Data pre-processing: temporal and spatial realignments.	67
6.2.2	The general linear model (GLM)	67
6.2.3	Thresholding	71
6.2.4	Detecting modules versus establishing functional connections.	73
6.3	Spatial models for fMRI data	74
6.3.1	Gaussian Random Fields	75
6.3.2	Markov Random Fields	76
6.3.3	Spatial wavelets	78
6.4	Dealing with the absence of a brain template	78
6.4.1	Spatial normalization	79
6.4.2	Shortcomings of this procedure	80
6.5	A quick overview of anatomical data processing	81
7	Statistical inference on multi-subject fMRI data	83
7.1	Classical statistical models	83
7.1.1	Fixed effects and mixed effects models	84
7.1.2	Classical hypothesis testing	85
7.1.3	The Gaussian mixed effect model	86
7.1.4	Non-parametric assessment of the statistics	87
7.1.5	Making inference on images	88
7.1.6	A Bayesian perspective	89
7.2	Analysing the reproducibility of group studies	90

7.2.1	Specificity, sensitivity, reproducibility	90
7.2.2	Quantifying the reproducibility in group studies	91
7.2.3	Observations from a large cohort	93
7.2.4	Limitations of the standard Gaussian mixed-effects model	100
7.3	Spatially relaxed inference: parcel-based RFX	106
7.3.1	Parcellation for group inference: definition and constraints	106
7.3.2	Solution: parcellation algorithm	109
7.3.3	The Parcel-based random effects (PRFX) procedure .	113
7.3.4	Results and discussion	113
8	Localizing cortical activity from MEEG	117
8.1	Pseudoinverse solution	119
8.2	Dipole fitting	120
8.3	Scanning methods	121
8.3.1	MULTiple SIgnal Classification (MUSIC)	121
8.3.2	Beamforming methods	124
8.4	Estimating distributed activity: imaging approach	126
8.4.1	Tikhonov regularization	127
8.4.2	Selecting the regularization parameter: the L-curve .	127
III	Appendices	129
A	Formulas and mathematical lemmas	131
A.1	Differential operators in \mathbb{R}^3	131
A.1.1	Multiplicative properties	132
A.1.2	Second derivative properties	132
A.1.3	Conversion from volume to surface integrals	132
A.1.4	The Green function for the Laplacian in \mathbb{R}^3	133
A.1.5	The Green formula	134
A.2	Maxwell equations	135
A.2.1	Current density	135
A.2.2	Maxwell-Gauss equation	135
A.2.3	Maxwell-Ampere equation	136
A.2.4	Maxwell-Faraday equation	137
A.2.5	Maxwell-Gauss equation for magnetism	138
A.3	The Poincaré and Hölder inequalities	139
A.4	Integral equalities	141

B Computational methods	143
B.1 Minimization under constraints: the Lagrange multiplier approach	143
B.2 Singular Value Decomposition	144
B.2.1 Moore-Penrose pseudoinverse	145
B.2.2 SVD and least-squares problems	146
B.2.3 Least-squares problems: a useful formula	147
C Statistical Results	149
C.1 On the Likelihood ratio and the F statistic	149
C.2 Bayesian derivation of the t test	151
C.3 D'Agostino-Pearson normality test	152
C.4 The localizer dataset	154
Bibliography	157
Index	171

Chapter 1

Introduction

This course on functional imaging of the brain principally deals with non-invasive functional imaging modalities: functional Magnetic Resonance Imaging, and Electro- and Magneto-encephalography. In this introduction we recall general facts on *functional brain imaging* considered as a domain of research. We start by describing the main challenge, which is merely brain complexity. Then we recall a few general facts on neuroscience and functional neuroimaging.

Brain complexity: Let us start with a review of some aspects of the brain complexity. Brain complexity is first *combinatorial*, since the brain comprises about 10^{11} neurons and 10^{15} synapses. Producing a realistic model or a simulation of such a huge amount of entities is thus a challenge that will need decades to be addressed (see e.g. the blue brain project <http://bluebrain.epfl.ch/>). For this reason, neuroscience can only proceed by revealing some snapshots of brain structure and function, depending on the modality of observation. This yields a second aspect of brain complexity: Several levels of organization at *different scales* must be considered, e.g. (in the case of the cortex) neurons, cortical columns, cortical maps, brain systems and lobes

Brain modelling is thus addressed at these different levels, but it is yet not clear how to integrate all these levels in a common framework.

A third aspect of brain complexity will be discussed further in this document: *inter-subject variability* (see Chapter 7). It means that there is no one-to-one mapping between any two different brains, so that it is difficult to identify and describe brain structures at the population level.

Observing the brain: The modalities for exploring brain structure and function are numerous, and they evolve quite rapidly according to technological breakthroughs that allow the observation of new phenomena. In this

section, we give only a superficial overview of these modalities, with an emphasis on human studies.

Historically, many insights on the functional architecture of the brain have been gained through electro-physiology studies in monkeys: Electrical recordings provided local information on brain function (e.g. firing neurons) which is the basis for the characterization of functionally specific brain areas. On the other hand, this technique does not provide a global view on brain activity, is expensive (due to the acquisition setup and monkey training), and generalization to the human brain raises cross-species issues [150]. Direct electrophysiological recordings within the brain can sometimes be performed on humans, either during neurosurgery, or with intracerebral electrodes implanted for surgical planning.

A second source of information, historically the first one, comes from lesion studies: it consists in describing the impact of local lesions on brain function. This description is of course very macroscopic. Brain architecture has been explored through post-mortem analyses (histology), so that the cell-level organization (cyto-architecture) has become one of the major ways to describe different brain regions (Brodmann areas). This description is limited, however, by the large inter-subject variability in brain structure.

Today, the main source of information is neuroimaging: Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET, SPECT), Optical Imaging (OI) and Magneto- and Electro-encephalography (MEEG).

- Before functional MRI became the standard tool for producing brain map, PET was used to map metabolic activity of the brain. However this modality is invasive, slow, and does not yield a very high spatial resolution. PET is now used for pharmacological studies, due to its ability to trace specific chemical components.
- MRI has provided, for a long time now, anatomical images (description of the brain in terms of tissues and macroscopic model of brain shape). More recently, it has also been used to acquire functional images (metabolic activity) [114]. It is now the main tool for structural and functional brain investigation. Finally, diffusion MRI provides a unique picture on brain anatomical connectivity [125].
- OI is probably the most recent modality for brain investigation. It can provide images of brain metabolism or intrinsic activity. To date, it does not allow a full coverage of the brain volume.
- MEEG, unlike the other modalities, performs some measurement of the intrinsic electro-magnetic activity of the brain. But it is recorded on the scalp, with a rather coarse resolution. The localization of the activity is then quite challenging.

These modalities are represented, with their resolution and invasivity in Fig. 1.1.

We conclude this introductory chapter with an overview on fMRI and MEG/EEG.

fMRI

As the first functional Magnetic Resonance Images were acquired in the early nineties, the domain is still young and the current techniques are evolving quickly.

Modern human brain mapping: The main aim of modern human brain mapping is to build a cartography of human brain functions (see e.g. [51]). Functionally, this is based on a subtractive reasoning principle: given activity maps related to two experimental conditions A and B, the difference map is expected to represent the spatial substrate of the cognitive processes involved in condition A and not in condition B. Localizing the corresponding regions in a population of subjects is a fundamental issue, which is not reduced to the construction of brain maps, but also involves the understanding of the underlying brain connectivity (see e.g. [79, 11]). More recently, there has also been a trend toward getting more insight on the mechanisms of brain activity, i.e. to understand and model what effects regions exert on each other in a certain experimental context [122, 12].

Human brain mapping is certainly an important project as such, but it also has some impact on the study of brain disease. Currently, one of the main hopes is that imaging might be used as an objective diagnosis tool and therapy follow-up for several diseases. In particular, one of the aims of anatomical and functional imaging is to finding some *biomarkers* for psychiatric diseases by comparing different population of patients: this is the case for autism [14], schizophrenia [20], Alzheimer's disease [146], or various cognitive and sensori-motor impairments [35]. The main anatomical biomarkers can be the cortical thickness [42, 91], the cortical shape [38], differences of diffusion anisotropy in the white matter [35], structure of fiber tracts [37]. Functional biomarkers are simply the amount of activity in certain regions which may characterize certain disabilities, or the amount of correlation of the brain signals across different regions (functional connectivity patterns, see e.g. [81]).

However, diagnosis is not the only question of interest, and in many cases, brain imaging may help to understand or provide some mechanisms for some cognitive diseases, e.g. dyslexia [75] or dyscalculia [104]. This is a major challenge for the next years.

MEG and EEG

The study of human bioelectricity was initiated with the discovery of electrocardiography at the turn of the twentieth century, followed by electroencephalography (EEG) in the 1920's, magnetocardiography in the 1960's, and magnetoencephalography (MEG) in the 1970's. Biomagnetic and bioelectric fields have the same biological origin: the displacement of charges within active cells called neurons [95].

Nowadays, EEG is relatively inexpensive and used commonly to detect and qualify neural activities (epilepsy detection and characterisation, neural disorder qualification, Brain Computer Interfaces, ...). MEG is, comparatively, much more expensive as SQUIDS work in very challenging conditions (at liquid helium temperature) and as a specially shielded room must be used to separate the signal of interest from the ambient noise. However, as it reveals a complementary vision to that of EEG and as it is less sensitive to the head structure, it also bears great hopes and more and more MEG machines are installed throughout the world. There are several scales at which bioelectricity can be described and measured: the microscopic scale, with microelectrodes placed inside or in a very close vicinity to neurons, and the mesoscopic scale, with intracortical recording of local field potentials (i.e. the electric potential within the cortex), below a square millimeter. Non-invasive measurements of the electric potential via EEG or the magnetic field via MEG are taken on the scalp, and the spatial extent of brain activity to which these measurements can be related has not yet been elucidated, but lies between a square millimeter and a square centimeter. Given the size of the head, and the time scale of interest - the millisecond - the quasistatic approximation can be applied to the Maxwell equations [68]. The electromagnetic field is thus related to the electric sources by two linear equations: the Poisson equation for the electric potential, and the Biot-Savart equation for the magnetic field.

MEG and EEG can be measured simultaneously and reveal complementary properties of the electrical fields. The two techniques have temporal resolutions of about the millisecond, which is the typical granularity of the measurable electrical phenomena that arise in the brain. This high temporal resolution is what makes MEG and EEG attractive for the functional study of the brain. The spatial resolution, on the contrary, is somewhat poor as only a few hundreds of simultaneous data points can be acquired simultaneously (about 300-400 for MEG and up to 256 for EEG). MEG and EEG are somewhat complementary with fMRI and SPECT in that those provide a very good spatial resolution but a rather poor temporal one (about the second for fMRI and the minute for SPECT). Contrarily to fMRI, which "only" measures an haemodynamic response linked to the metabolic demand, MEG

and EEG also measure a direct consequence of the electrical activity of the brain: it is admitted that the MEG and EEG measured signals correspond to the variations of the post-synaptic potentials of the pyramidal cells in the cortex. Pyramidal neurons compose approximately 80% of the neurons of the cortex, and it requires at least about 50,000 active such neurons to generate some measurable signal.

Reconstructing the cortical sources from the electromagnetic field measured on EEG and MEG sensors requires that the inverse problem of MEG and EEG (denoted collectively as MEEG for short) be solved. It is a difficult problem because it is ill-posed, and this has led to a large body of literature, concerning both theoretical [9, 21] and computational aspects [8, 29, 41].

There are two main domains of application for EEG and MEG: clinical applications and the study of cognition.

Clinical research in neurophysiology aims at understanding the mechanisms leading to disorders of the brain and the central nervous system, in order to improve diagnosis and eventually propose new therapies. The clinical domains in which EEG and MEG are most routinely used include epilepsy, schizophrenia, depression, attention deficit disorders. Clinicians are especially interested in the time courses of the measured signals: their experience in the visual analysis of superimposed sensor time courses allows them to detect abnormal patterns. The source localization performed by clinicians is generally limited to simple dipole scanning methods.

EEG and MEG rely on passive measurements, with no applied electromagnetic field. In contrast, active techniques using bioelectric stimulation, are currently being developed and tested on patients, to treat disorders such as Parkinson's disease, chronic pain, dystonia and depression. Implanted intracortical or extradural electrodes can deliver a electrical stimulation to specific brain areas (Deep Brain Stimulation or DBS). Less invasive, Transcranial Magnetic Stimulation (TMS), uses a time-varying magnetic field to induce a current within the cortex. All of these techniques can be studied with the same equations, models, and numerical tools as the ones used for MEEG. Moreover, in order to understand the physiological mechanisms triggered by these stimulations, simultaneous TMS/EEG and DBS/MEEG can be performed and analyzed [148].

Outline

This class material is divided in two parts: the first on modeling, which goes into details in explaining geometrical, physiological and numerical models used in the field of human brain mapping. The second part deals with the analysis of functional imaging data, coming from fMRI or from MEG/EEG: in the second part, an emphasis is put on the solution of ill-posed inverse

problems, and on statistical tests. An Appendix presents all the required technical tools, making the material self-contained.

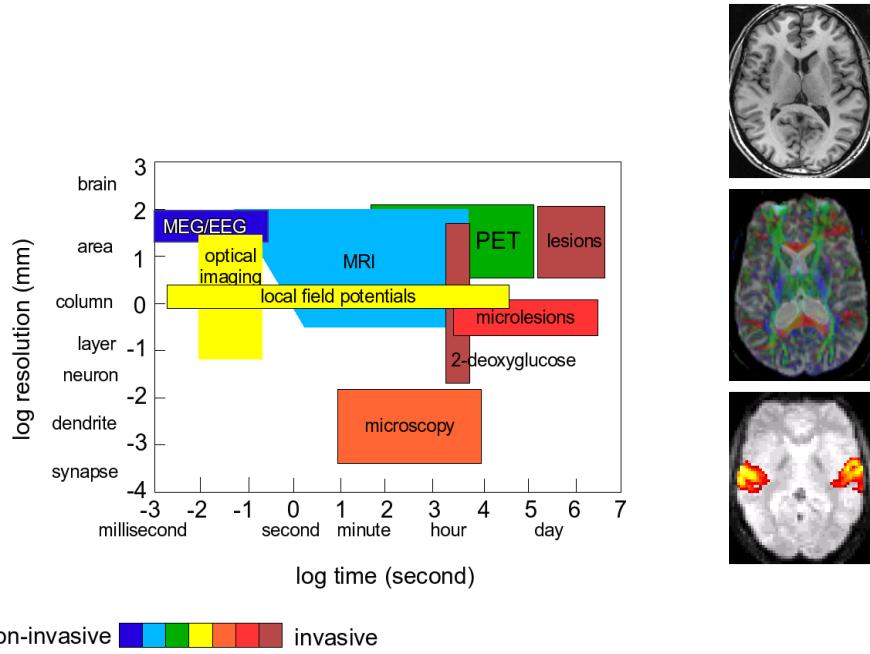


Figure 1.1: (left) The different imaging modalities for brain mapping. MRI and functional MRI have the unique property to yield high-resolution information while being minimally invasive. Unlike other modalities, MRI allows whole brain imaging. (right) Typical example of T1 anatomical MRI (top), pre-processed Diffusion-Weighted (DW) MRI (middle) and fMRI (bottom) images, presented in axial views. These images are from the Neurospin 3T scanner. For the DW-MRI image (courtesy from P.Fillard), the main direction of water diffusion is color-coded: green for antero-posterior diffusion, red for lateral diffusion, blue for vertical diffusion. The functional image has been analyzed to yield the regions activated in an auditory task.

Acronyms

BCI	Brain Computer Interface
BEM	Boundary Element Method
DBS	Deep Brain Stimulation
EEG	electro-encephalography
FEM	Finite Element Method
fMRI	functional Magnetic Resonance Imaging
HRF	haemodynamic response function
MEG	magneto-encephalography
MRF	Markov random field
MRI	Magnetic Resonance Imaging
PET	Positron Emitted Tomography
SQUID	Supraconductive Quantum Imaging Device
SVD	Singular Value Decomposition

Part I

Physical and Physiological modeling

Understanding the functioning of the brain through measurements such as fMRI or M/EEG requires to understand the links between the brain activity and the measurements. The first part of this course presents the physical and physiological principles that guide the development of models for linking brain activity to measurements, i.e. the development of “forward models”. The estimation of brain activity from the analysis of the measurements will be covered in the second part of the course, on “inverse problems” and statistical analysis.

This part is divided in four chapters:

- **Origin of functional measurements** in which we explain the current knowledge on the origin of fMRI and electrophysiological (M/EEG) signals. Notably, the origins of both types of measurements are quite distinct, and few models exist that combine the two.
- **Forward problem computation** deals with the M/EEG forward problem, explaining the different levels of detail that can be used and the resulting differences in computational complexity.
- **Leadfield computation** gives insight into adjoint, or reciprocity principles, that simplify the forward problem computation.
- **Geometric modeling of the head** explains the extraction of geometrical information from Magnetic Resonance Images. This geometrical information is the substrate on which the functional information resides, and its precise knowledge is important for deciphering brain activity.

Chapter 2

Origin of functional measurements

2.1 Origin of the fMRI signal: the BOLD effect

The origin of the fMRI signal is still not very well understood. MRI machines can be tuned to acquire images that measure different phenomena. We focus here on the Blood Oxygen Level Dependent (BOLD) effect, which is the principal acquisition mode used nowadays.

As sketched in section 2.2.1, brain activity translates into action potentials, which correspond to the depolarization of the neuron membrane, which occurs through the exchange of ions – essentially potassium K^+ and sodium Na^+ (see figure 2.1).

At the level of synapses, this information is exchanged through the release of neurotransmitters that bind to receptor sites on post-synaptic terminals, yielding the next neuron to depolarization. Recovery from neuronal signaling requires uptake and repackaging of neurotransmitter and restoration of ionic gradients, all processes that consume Adenosin Triphosphate (ATP). ATP

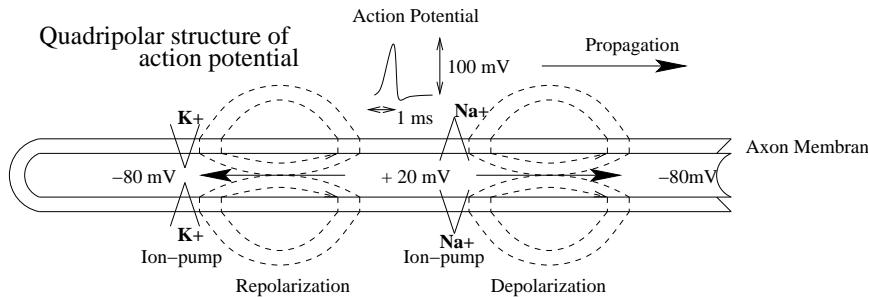


Figure 2.1: The ion exchanges corresponding to action potentials.

consumption, in turn, requires a continuous supply of glucose and oxygen which is provided by Cerebral Blood Flow (CBF)¹. It was noticed quite early [77] that brain activity can be assessed by the measurement of regional CBF. This is the effect measured eg with Positron Emission Tomography (PET). The BOLD effect measured by fMRI is more subtle.

It is important to note that while the oxygenated blood has basically the same magnetic susceptibility as the surrounding brain tissues, deoxyhemoglobin (the de-oxygenated hemoglobin) is paramagnetic and thus incurs a signal drop when its rate increases. Brain activity increases the CBF, but actually this is more governed by the glucose consumption than by the oxygen one. This means that the rate of deoxyhemoglobin/oxyhemoglobin actually decreases, which induces an increase of the MRI signal. In practice, the phenomenon is even more complicated as the CBF increases, blood volume and blood velocity also increase, but the blood oxygenation effect dominates. The signal increase with current machines is of a few percent at 1.5T and 5-15% for 4T machines.

The BOLD signal is composed of CBF contributions from larger arteries and veins, smaller arterioles and venules, and capillaries. Experimental results indicate that the BOLD signal increases roughly as the square of the magnetic field strength. Furthermore, it can be weighted to the smaller vessels, and hence closer to the active neurons, by using larger magnetic fields. For example, whereas about 70% of the BOLD signal arises from larger vessels in a 1.5 Tesla scanner, about 70% arises from smaller vessels in a 4 Tesla scanner. This explains the race for larger field scanners: a few 7 Tesla commercial scanners have become operational, and experimental 8 and 9 Tesla scanners are under development.

2.2 Origin of magnetic and electric measurements

Neuronal currents generate magnetic and electric fields according to principles stated in Maxwell's equations. The neural current distribution can be described as the primary current, and viewed as the "battery" in a resistive circuit. The postsynaptic currents in the cortical pyramidal cells are the main primary currents which give rise to measurable MEG and EEG signals.

2.2.1 Neural current sources

¹As an illustration, the brain receives 15% of the total cardiac output of blood, and yet accounts for 2% of the body weight; in particular, the flow per gram tissue to gray matter is comparable to that in the heart muscle, the most energetic organ in the body [18]. Yet the brain has no reserve store of oxygen, and depends on continuous delivery by CBF.

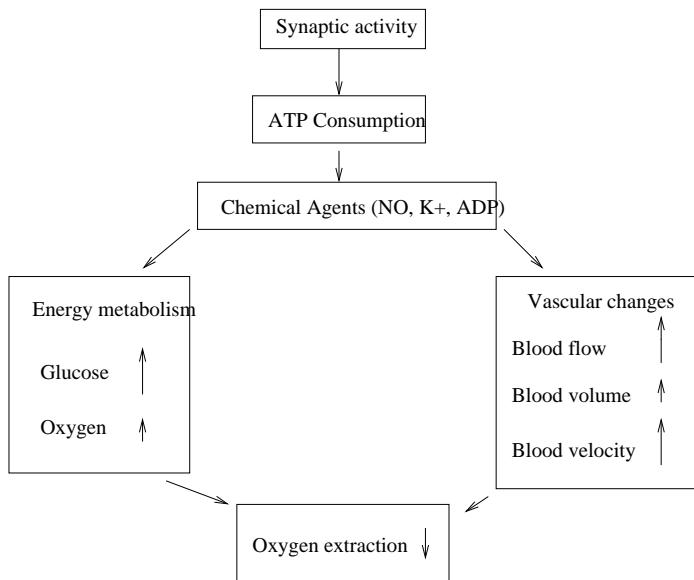


Figure 2.2: Physiological changes accompanying brain activation. Arrows indicate the increase or decrease of the corresponding item and its magnitude. Functional neuroimaging is largely based on the metabolism and flow changes in the lower three blocks: the drop in oxygen extraction is the basis for the BOLD signal changes measured with fMRI, but the MR signal is potentially sensitive to blood flow, volume, and velocity as well. This figure is taken from [18].

2.2.2 Action potentials and postsynaptic potentials

Electric signals propagate within the brain along nerve fibers (axons) as a series of action potentials (APs). The corresponding primary current can be approximated by a pair of opposite current dipoles, one at the depolarization and one at the repolarization front (figure), and this quadrupolar source moves along the axon as the activation propagates. The distance between the two dipoles depends on the duration of the AP and on the conduction velocity of the fiber. For a cortical axon with a conduction speed of 5 m/s, the opposite dipoles would be about 5mm apart.

In synapses, the chemical transmitter molecules change the ion permeabilities of the postsynaptic membrane and a postsynaptic potential (PSP) and current are generated. In contrast to the currents associated with an action potential, the postsynaptic current can be adequately described by a single current dipole oriented along the dendrite. The magnetic field of a current dipole falls off with distance less strongly (in order $1/r^2$) than the field associated with the quadrupolar AP currents (in order $1/r^3$).

Furthermore, temporal summation of currents flowing in neighboring fibers is more effective for synaptic currents, which last up to tens of milliseconds, than for the about 1 ms-long action potentials. Thus the electromagnetic signals observed outside and on the surface of the head seem to be largely due to the synaptic current flow. In special cases, currents related to action potentials might also significantly contribute to cortical MEG and EEG signals, such as, e.g., high-frequency somatosensory responses.

The pyramidal cells are the principal types of neurons in the cortex, with their apical dendrites oriented parallel to each other and perpendicular to the cortical surface. Since neurons guide the current flow, the resultant direction of the electrical current flowing in the dendrites is also perpendicular to the cortical sheet of gray matter.

2.2.3 Estimates of dipole strengths

Each PSP may contribute as little as a 20 fAm current dipole, probably too small to measure in MEG or EEG. The current dipole moments required to explain the measured MEG and EEG fields outside the head are on the order of 10 nAm [68, 67]. This would correspond to about a million of synapses simultaneously active during a typical evoked response. Although such a synchronous activity only forms a part of the total activity, it can have high functional relevance. For example, invasive recordings from monkeys have shown surprisingly large temporal overlap of neuronal firing in many visual cortical areas (Schmolesky et al, 1998). Epileptic discharges also are typically associated with strong current densities due to highly synchronous activity.

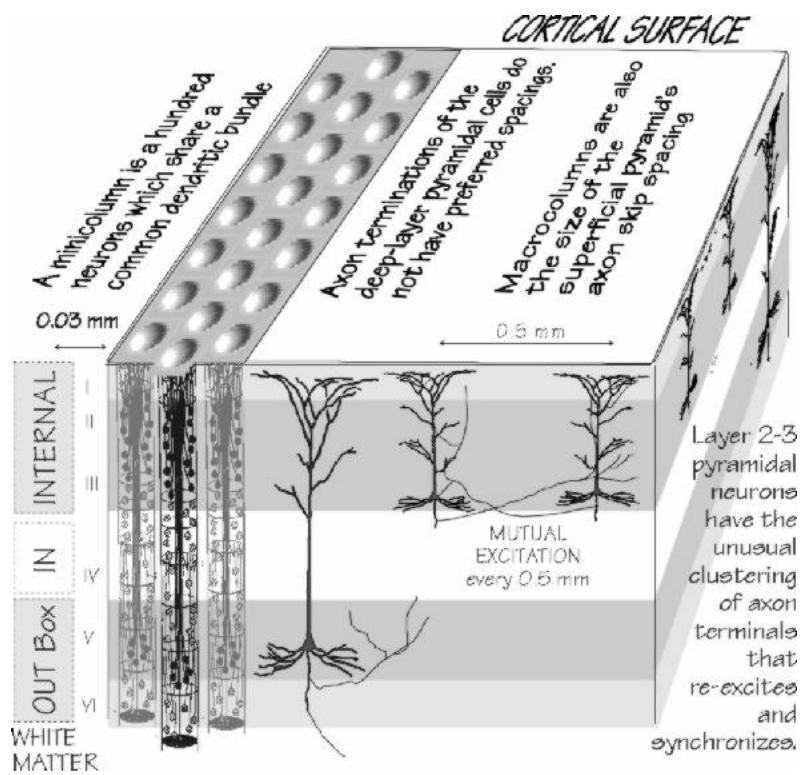


Figure 2.3: The organisation of the cortex.

2.2.4 Electromagnetic propagation

The electric current, The electric potential V , the magnetic field \mathbf{B} and the electric current \mathbf{J} are related through the Maxwell Equations, whose local and integral forms are summarized here:

Name	Differential form	Integral form
Gauss's law	$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}$	$\int_{\partial\Omega} \mathbf{E} \cdot \mathbf{n} ds = \int_{\Omega} \frac{\rho}{\epsilon_0} dr$
Gauss's law for magnetism	$\nabla \cdot \mathbf{B} = 0$	$\int_{\partial\Omega} \mathbf{B} \cdot \mathbf{n} ds = 0$
Faraday's law	$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$	$\int_{\partial S} \mathbf{E} \cdot \vec{dl} = - \int_S \frac{\partial \mathbf{B}}{\partial t} ds$
Ampère's circuital law	$\nabla \times \mathbf{B} = \mu_0 (\mathbf{J} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t})$	$\int_{\partial S} \mathbf{B} \cdot \vec{dl} = \mu_0 \int_S (\mathbf{J} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}) \cdot \mathbf{n} ds$

$$\epsilon_0 = 8.85 \cdot 10^{-12} kg^{-1} m^{-3} A^2 s^4, \mu_0 = 4\pi \cdot 10^{-7} kg m A^{-2} s^{-2}, \epsilon_0 \mu_0 c^2 = 1.$$

\mathbf{E} is expressed in $V m^{-1}$, \mathbf{B} in T (tesla), \mathbf{J} in $A m^{-2}$ and ρ in $C m^{-3}$.

More details on these relations are provided in Appendix A.2.

Quasistatic approximation

The Maxwell equations presented above are written in a general, time-varying, setting. For EEG and MEG modeling, the spatial scale, the frequencies, and the medium properties make it possible to neglect the inductive, capacitive and displacement effects, and to effectively omit the time-derivatives in Ampere's and Faraday's laws. Omitting the time-derivatives in Ampere's law is called magneto-quasi-stationarity, while omitting the time-derivatives in Faraday's laws is called the electro-quasistatic regime. This considerably simplifies the resulting system, because the magnetic and electric fields become uncoupled, and can be solved separately.

Let us briefly justify the quasistatic assumption. We note that, in a magnetic medium free of charges or current generators, the volumic current density \mathbf{J} is the sum of an *volumic ohmic current* and a *polarization current*:

$$\mathbf{J} = \sigma \mathbf{E} + \frac{\partial \mathbf{P}}{\partial t} \quad (2.1)$$

where $\mathbf{P} = (\epsilon - \epsilon_0) \mathbf{E}$ is the polarization vector, ϵ the *permittivity* of the medium, and σ the *conductivity*.

Let us examine the Maxwell-Ampere equation (in above table and A.19 in the appendix). Using (2.1), we are able to express the right-hand side as a function of \mathbf{E} only:

$$\nabla \times \mathbf{B} = \mu_0 \left(\sigma \mathbf{E} + \epsilon \frac{\partial \mathbf{E}}{\partial t} \right)$$

Let us further assume that the electrical field can be modelled using a planar waveform at frequency f , $\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0 e^{i(\omega t - \mathbf{k} \cdot \mathbf{r})}$, with $\omega = 2\pi f$ and the added condition $\mathbf{E}_0 \cdot \mathbf{k} = 0$.

Using this model, we have:

$$\frac{\partial \mathbf{E}}{\partial t} = i\omega \mathbf{E}$$

and

$$\begin{aligned}\left\| \varepsilon \frac{\partial \mathbf{E}}{\partial t} \right\| &= \omega \varepsilon \|\mathbf{E}\| \\ \|\sigma \mathbf{E}\| &= \sigma \|\mathbf{E}\|\end{aligned}$$

The term $\varepsilon \frac{\partial \mathbf{E}}{\partial t}$ is negligible compared to $\sigma \mathbf{E}$, at a frequency f , if

$$\frac{2\pi f \varepsilon}{\sigma} \ll 1.$$

For the brain, $\sigma = 0.3 \Omega^{-1} m^{-1}$, the permittivity ε is of the order of $10^5 \varepsilon_0 = 8.8510^{-7}$, and the frequencies of interest are typically lower than $f = 100$ Hz. With these values, $\frac{2\pi f \varepsilon}{\sigma}$ is of the order of $2 \cdot 10^{-3}$. Therefore, the time derivative in Ampere's law can be neglected. The Maxwell-Ampere equation becomes time-invariant:

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J}, \quad (2.2)$$

and the current density is consequently divergence-free

$$\nabla \cdot \mathbf{J} = 0. \quad (2.3)$$

To show that the time-derivative can be neglected in Faraday's law, we compute the rotational of this equation. The left hand side becomes:

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - \Delta \mathbf{E}.$$

But with our choice for \mathbf{E} , we have:

$$\nabla \cdot \mathbf{E} = \mathbf{E}_0 \nabla e^{i(\omega t - \mathbf{k} \cdot \mathbf{r})} = -i \mathbf{E}_0 \cdot \mathbf{k} e^{i(\omega t - \mathbf{k} \cdot \mathbf{r})} = 0.$$

Thus:

$$\begin{aligned}\nabla \times (\nabla \times \mathbf{E}) &= -\Delta \mathbf{E} = \|\mathbf{k}\|^2 \mathbf{E} = -\frac{\partial}{\partial t} \nabla \times \mathbf{B} \\ &= -\mu_0 \frac{\partial}{\partial t} \left(\sigma \mathbf{E} + \varepsilon \frac{\partial \mathbf{E}}{\partial t} \right) \\ &= -\mu_0 i \omega (\sigma + \varepsilon i \omega) \mathbf{E}\end{aligned}$$

Consequently, we get: $\|\mathbf{k}\|^2 = |\mu_0 i \omega (\sigma + \varepsilon i \omega)|$ and the corresponding wavelength is $\lambda = \frac{1}{\|\mathbf{k}\|} = \frac{1}{\sqrt{|\mu_0 i \omega (\sigma + \varepsilon i \omega)|}}$. For the head, this quantity is equal to $65m \gg \emptyset_{\text{head}}$. This means that the time-derivative can be neglected in Faraday's law, leading to the time-invariant Maxwell-Faraday equation

$$\nabla \times \mathbf{E} = 0. \quad (2.4)$$

Poisson equation

A consequence of the time-invariant Maxwell-Faraday equation (2.4) is that the electric field \mathbf{E} derives from a potential, which we call the electric potential and denote V :

$$\mathbf{E} = -\nabla V .$$

It is useful to divide the current density \mathbf{J} , into two components: the passive *Ohmic* or *return* current $\sigma\mathbf{E}$ and the remaining *primary current* \mathbf{J}^P

$$\mathbf{J} = -\sigma\nabla V + \mathbf{J}^P . \quad (2.5)$$

Although this equation holds at different scales, it is not possible to include all the microscopic conductivity details in models of MEEG activity and thus σ refers to macroscopic conductivity with at least 1mm scale. The division of neuronal currents to primary and volume currents is physiologically meaningful. For instance, chemical transmitters in a synapse give rise to primary current mainly inside the postsynaptic cell, whereas the volume current flows passively in the medium, with a distribution depending on the conductivity profile. By finding the primary current, we can locate the active brain regions.

The current density is divergence-free, and using the decomposition (2.5) shows that the electric potential and the primary current are related by a simple equation, called a Poisson equation:

$$\nabla \cdot (\sigma\nabla V) = \nabla \cdot \mathbf{J}^P . \quad (2.6)$$

Biot and Savart law

We derive in this section the Biot and Savart law, relating the magnetic field to the current density. Recall the time-invariant Maxwell-Ampere equation

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J}$$

and take its curl:

$$\nabla \times \nabla \times \mathbf{B} = \mu_0 \nabla \times \mathbf{J} . \quad (2.7)$$

The left-hand side can be rewritten

$$\nabla \times \nabla \times \mathbf{B} = -\Delta \mathbf{B} + \nabla(\nabla \cdot \mathbf{B})$$

where the Laplacian acts coordinatewise on the vector field \mathbf{B} . Since $\nabla \cdot \mathbf{B} = 0$ (Maxwell equation expressing absence of magnetic charges), (2.7) rewrites:

$$-\Delta \mathbf{B} = \mu_0 \nabla \times \mathbf{J} .$$

Recalling that a fundamental solution to the Laplacian in \mathbb{R}^3 is $-\frac{1}{4\pi\|r\|}$ (see appendix A.1.4), in the sense that

$$\Delta \left(-\frac{1}{4\pi\|r\|} \right) = \delta_0$$

this implies that

$$\mathbf{B} = \frac{\mu_0}{4\pi} \int \nabla' \times \mathbf{J}(\mathbf{r}') \frac{1}{\|\mathbf{r} - \mathbf{r}'\|} d\mathbf{r}' + \mathbf{B}_H$$

where \mathbf{B}_H is a *harmonic* function, i.e. such that

$$\Delta \mathbf{B}_H = 0 .$$

With the condition that \mathbf{B} vanishes at infinity, the harmonic term can be discarded, an integration by parts leads to the Biot and Savart law:

$$\mathbf{B} = \frac{\mu_0}{4\pi} \int \mathbf{J}(\mathbf{r}') \times \frac{\mathbf{r} - \mathbf{r}'}{\|\mathbf{r} - \mathbf{r}'\|^3} d\mathbf{r}' . \quad (2.8)$$

Replacing the total current \mathbf{J} by its decomposition (2.5), the Biot-Savart law becomes

$$\mathbf{B} = \mathbf{B}_0 - \frac{\mu_0}{4\pi} \int \sigma \nabla' V(\mathbf{r}') \times \frac{\mathbf{r} - \mathbf{r}'}{\|\mathbf{r} - \mathbf{r}'\|^3} d\mathbf{r}' , \quad (2.9)$$

where \mathbf{B}_0 is the contribution to the magnetic field coming from the primary current:

$$\mathbf{B}_0 = \frac{\mu_0}{4\pi} \int \mathbf{J}^P(\mathbf{r}') \times \frac{\mathbf{r} - \mathbf{r}'}{\|\mathbf{r} - \mathbf{r}'\|^3} d\mathbf{r}' .$$

Note: if the medium is infinite, with homogeneous conductivity σ , a simplification occurs in the Biot-Savart equation, since

$$\nabla \times \mathbf{J}(\mathbf{r}) = \nabla \times (\mathbf{J}^P(\mathbf{r}) - \sigma \nabla V(\mathbf{r})),$$

and in a homogeneous medium, $\nabla \times (\sigma \nabla V) = \sigma \nabla \times \nabla V = 0$. From Eq. (2.8), the magnetic field hence becomes independent of the ohmic contribution:

$$\mathbf{B} = \mathbf{B}_0 = \frac{\mu_0}{4\pi} \int \mathbf{J}^P \times \frac{\mathbf{r} - \mathbf{r}'}{\|\mathbf{r} - \mathbf{r}'\|^3} d\mathbf{r}' . \quad (2.10)$$

Chapter 3

Forward problem computation

3.1 Introduction

The forward problem of magneto-electro-encephalography aims at computing the electromagnetic field produced by a known primary current, in a known geometry.

This chapter is organized in increasing model complexity. Section 3.2 presents the forward problem in simple geometrical settings for which the calculations can be done by hand. Section 3.3 considers more general nested surface models, which we call “semi-realistic”, in which subject-dependent surfaces are designed to match the main tissue interfaces in the head. Computations in the semi-realistic setting are performed using Boundary Element Methods (BEM). Finally, Section 3.4 presents the most sophisticated model, which we call “realistic”, which models the tissue conductivity voxel-wise, does not necessitate to define interfaces between tissues of homogeneous conductivity, and allows for tensor-valued conductivity.

3.2 Simple geometries

3.2.1 Current dipole in infinite homogeneous medium

Electrical potential

A current dipole with moment \mathbf{q} and position \mathbf{p} is represented by $\mathbf{J}^P(\mathbf{r}) = \mathbf{q} \delta_{\mathbf{p}}(\mathbf{r})$. The potential created by such a dipole follows the Poisson equation

$$\nabla \cdot (\sigma \nabla V) = \mathbf{q} \cdot \nabla \delta(\mathbf{r} - \mathbf{p}) .$$

As the medium is infinite with constant conductivity σ ,

$$\sigma \Delta V = \mathbf{q} \cdot \nabla \delta(\mathbf{r} - \mathbf{p}) .$$

As detailed in Appendix A.1.4, in three dimensions, the fundamental solution to the Laplacian is $\frac{-1}{4\pi\|\mathbf{r}\|}$ in the sense that

$$\Delta \left(\frac{-1}{4\pi\|\mathbf{r}\|} \right) = \delta_0 .$$

Hence,

$$V(\mathbf{r}) = -\frac{1}{4\pi\sigma} \int \mathbf{q} \cdot \nabla' \delta(\mathbf{r}' - \mathbf{p}) \frac{1}{\|\mathbf{r} - \mathbf{r}'\|} d\mathbf{r}' = \frac{1}{4\pi\sigma} \mathbf{q} \cdot \frac{\mathbf{r} - \mathbf{p}}{\|\mathbf{r} - \mathbf{p}\|^3} .$$

Magnetic field

In an infinite, homogeneous domain, only the primary current contributes to the magnetic field as was shown in the derivation of 2.10, therefore, for a dipolar source,

$$\mathbf{B} = \frac{\mu_0}{4\pi} \mathbf{q} \times \frac{\mathbf{r} - \mathbf{p}}{\|\mathbf{r} - \mathbf{p}\|^3} .$$

3.2.2 Silent sources

Helmholtz, in 1853, was the first to point out the existence of sources which are electromagnetically silent, i.e. produce a null electro-magnetic field. First note that a solenoidal source, such that $\nabla \cdot \mathbf{J}^P = 0$, is electrically silent, since the source term of the Poisson equation vanishes in this case.

Next, we exhibit an electromagnetically silent source, in the form of a primary current \mathbf{J}^P , supported on a surface S , and such that $\mathbf{J}^P = q \mathbf{n}$ where \mathbf{n} is the normal vector to S and q is a constant. We prove that \mathbf{J}^P is electromagnetically silent if the medium is infinite and homogeneous. We will extend this result to more general domains in the course of this chapter.

In an infinite homogeneous medium the potential V can be written as an integral over the support of the primary current:

$$V(\mathbf{r}) = -\frac{1}{4\pi\sigma} \int \nabla' \cdot \mathbf{J}^P(\mathbf{r}') \frac{1}{\|\mathbf{r} - \mathbf{r}'\|} d\mathbf{r}' = \frac{1}{4\pi\sigma} \int \mathbf{J}^P(\mathbf{r}') \cdot \frac{\mathbf{r} - \mathbf{r}'}{\|\mathbf{r} - \mathbf{r}'\|^3} d\mathbf{r}' .$$

For the particular case of $\mathbf{J}^P = q \mathbf{n} \delta_S$,

$$V(\mathbf{r}) = \frac{1}{4\pi\sigma} \int_S q \mathbf{n} \cdot \frac{\mathbf{r} - \mathbf{r}'}{\|\mathbf{r} - \mathbf{r}'\|^3} ds = \frac{q}{4\pi\sigma} \int_S \frac{\mathbf{r} - \mathbf{r}'}{\|\mathbf{r} - \mathbf{r}'\|^3} \cdot \mathbf{n} ds$$

The integral on the right represents the solid angle of S viewed from \mathbf{r} , and vanishes if \mathbf{r} is exterior to S .

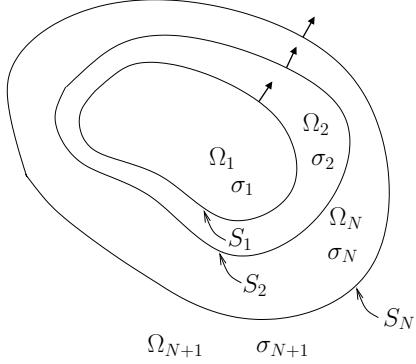


Figure 3.1: The head is modeled as a set of nested regions $\Omega_1, \dots, \Omega_{N+1}$ with constant isotropic conductivities $\sigma_1, \dots, \sigma_{N+1}$, separated by interfaces S_1, \dots, S_N . Arrows indicate the normal directions (outward).

The magnetic field can be represented by the Biot-Savart equation, yielding, if Ω_S denotes the volume contained inside S :

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int_S q \mathbf{n} \times \nabla' \left(\frac{1}{\|\mathbf{r} - \mathbf{r}'\|} \right) ds = \frac{\mu_0 q}{4\pi} \int_{\Omega_S} \nabla' \times \left(\nabla' \frac{1}{\|\mathbf{r} - \mathbf{r}'\|} \right) d\mathbf{r}' = 0.$$

3.3 Semi-realistic model

In an infinite, homogeneous domain, the electric potential and the magnetic field decay at the same rate. However, when measured on the scalp, the two fields have very different spatial properties: the magnetic field appears more ‘‘focal’’, and the electric potential more ‘‘diffuse’’. The main reason for this qualitative difference is that the magnetic field is less sensitive than the electric potential to conductivity differences in the tissues of the head. The electric potential, in particular, is subject to diffusion because of the low conductivity of the skull.

In this section, we will consider a piecewise-constant conductivity, organised in layers, as depicted in Figure 3.1.

3.3.1 Magnetic field computation

Section 2.2.4 has established the Biot and Savart law, decomposing the magnetic field into a primary current contribution and an ohmic contribution

$$\mathbf{B} = \mathbf{B}_0 - \frac{\mu_0}{4\pi} \int \sigma \nabla' V(\mathbf{r}') \times \frac{\mathbf{r} - \mathbf{r}'}{\|\mathbf{r} - \mathbf{r}'\|^3} d\mathbf{r}',$$

where

$$\mathbf{B}_0 = \frac{\mu_0}{4\pi} \int \mathbf{J}^P(\mathbf{r}') \times \frac{\mathbf{r} - \mathbf{r}'}{\|\mathbf{r} - \mathbf{r}'\|^3} d\mathbf{r}' .$$

With the piecewise-constant conductivity model, the ohmic term can be decomposed as a sum over volumes of constant conductivity:

$$\int \sigma \nabla' V(\mathbf{r}') \times \frac{\mathbf{r} - \mathbf{r}'}{\|\mathbf{r} - \mathbf{r}'\|^3} d\mathbf{r}' = \sum_i \sigma_i \int_{\Omega_i} \nabla V \times \frac{\mathbf{r} - \mathbf{r}'}{\|\mathbf{r} - \mathbf{r}'\|^3} d\mathbf{r}' = \sum_i \sigma_i I_i \quad (3.1)$$

In the above identity, note that the conductivities must be assumed constant in each domain Ω_i , in order to take σ_i out of the integral over Ω_i . The volume integral I_i can be expressed as a surface integral on $\partial\Omega_i = S_{i-1} \cup S_i$ with $S_0 = \emptyset$. With this in view, we use the Stokes formula, and the identity

$$\nabla \times (V \nabla g) = \nabla V \times \nabla g .$$

Thus

$$\begin{aligned} I_i &= \int_{\Omega_i} \nabla \times \left(V(\mathbf{r}') \frac{\mathbf{r} - \mathbf{r}'}{\|\mathbf{r} - \mathbf{r}'\|^3} \right) d\mathbf{r}' = \int_{\partial\Omega_i} \mathbf{n} \times \left(V(\mathbf{r}') \frac{\mathbf{r} - \mathbf{r}'}{\|\mathbf{r} - \mathbf{r}'\|^3} \right) ds \\ &= \int_{S_i} \mathbf{n}' \times \left(V(\mathbf{r}') \frac{\mathbf{r} - \mathbf{r}'}{\|\mathbf{r} - \mathbf{r}'\|^3} \right) ds - \int_{S_{i-1}} \mathbf{n}' \times \left(V(\mathbf{r}') \frac{\mathbf{r} - \mathbf{r}'}{\|\mathbf{r} - \mathbf{r}'\|^3} \right) ds , \end{aligned}$$

with $\mathbf{n} = \frac{\mathbf{r}'}{\|\mathbf{r}'\|}$. This expression is then inserted in (3.1) and, recalling that $\sigma_{N+1} = 0$,

$$\mathbf{B}(\mathbf{r}) = \mathbf{B}_0(\mathbf{r}) + \frac{\mu_0}{4\pi} \sum_{i=1}^N (\sigma_i - \sigma_{i+1}) \int_{S_i} V(\mathbf{r}') \frac{\mathbf{r} - \mathbf{r}'}{\|\mathbf{r} - \mathbf{r}'\|^3} \times \mathbf{n}' ds(\mathbf{r}') \quad (3.2)$$

The case of a concentric spherical head

In the case where the surfaces S_i are spherical, and concentric, the above expression shows that the radial component of the magnetic field is independent of the conductivity profile:

$$\mathbf{B}(\mathbf{r}) \cdot \mathbf{r} = \mathbf{B}_0(\mathbf{r}) \cdot \mathbf{r} .$$

This results from the identity

$$((\mathbf{r} - \mathbf{r}') \times \mathbf{r}') \cdot \mathbf{r} = 0 .$$

The independence to the conductivity profile can be extended to the three components of the magnetic field for a spherical geometry. Indeed,

outside Ω , since $\sigma = 0$ and $\mathbf{J}^p = 0$, $\mathbf{J} = 0$ and hence $\nabla \times \mathbf{B} = 0$. The magnetic field thus derives from a “potential” which we denote U :

$$\mathbf{B} = -\nabla U \quad (3.3)$$

The potential U is only defined up to a constant, but since $\mathbf{B} \rightarrow 0$ at infinity, we adjust the constant so that $U \rightarrow 0$ at infinity.

Denote $e_r = \mathbf{r}/\|\mathbf{r}\|$ the unit radial vector. For \mathbf{r} outside Ω ,

$$U(\mathbf{r}) = - \int_0^\infty \nabla U(\mathbf{r} + t e_r) \cdot e_r dt$$

The above expression shows that \mathbf{B} outside Ω is independent of σ . Furthermore, if the source is a single dipole located at \mathbf{r}_0 , with dipolar moment \mathbf{q} , then $U(\mathbf{r})$ can be written as:

$$\begin{aligned} U(\mathbf{r}) &= \int_0^\infty \mathbf{B}(\mathbf{r} + t e_r) \cdot e_r dt = \int_0^\infty \mathbf{B}_0(\mathbf{r} + t e_r) \cdot e_r dt \\ &= \frac{\mu_0}{4\pi} \mathbf{q} \times (\mathbf{r} - \mathbf{r}_0) \cdot e_r \int_0^\infty \frac{dt}{\|\mathbf{r} + t e_r - \mathbf{r}_0\|^3} \end{aligned} \quad =$$

Thus, for a radial dipole, $\mathbf{q} \times (\mathbf{r} - \mathbf{r}_0) \cdot e_r = 0$, hence $U(\mathbf{r}) = 0$ and $\mathbf{B}(\mathbf{r})$ vanishes outside Ω . This is still true for a set of radial dipoles by additivity of the magnetic field. The integral in the above formula can be computed analytically, leading to:

$$U(\mathbf{r}) = \frac{\mu_0}{4\pi} \frac{(\mathbf{q}, \mathbf{a}, e_r)}{a(a + \mathbf{a} \cdot e_r)} = \frac{\mu_0}{4\pi} \frac{(\mathbf{q}, \mathbf{a}, \mathbf{r})}{a(ar + \mathbf{a} \cdot \mathbf{r})} = -\frac{\mu_0}{4\pi F} (\mathbf{q}, \mathbf{r}_0, \mathbf{r}) ,$$

with $\mathbf{a} = \mathbf{r} - \mathbf{r}_0$, $a = \|\mathbf{a}\|$, $r = \|\mathbf{r}\|$, $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ being the determinant of the 3 vectors \mathbf{a} , \mathbf{b} and \mathbf{c} and $F = a(ar + \mathbf{a} \cdot \mathbf{r})$.

Applying Eq. 3.3 to the previous formula gives the so-called Sarvas formula for the magnetic field:

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi F^2} (F \mathbf{q} \times \mathbf{r}_0 - (\mathbf{q}, \mathbf{r}_0, \mathbf{r}) \nabla F) ,$$

with

$$\begin{aligned} \nabla F &= \left(a + 2r + \frac{\mathbf{a} \cdot \mathbf{r}}{a} \right) \mathbf{a} + \left(\frac{a^2}{r} + a \right) \mathbf{r} , \\ \nabla F &= \left(\frac{a^2}{r} + \frac{\mathbf{a} \cdot \mathbf{r}}{a} + 2(a + r) \right) \mathbf{r} - \left(a + 2r + \frac{\mathbf{a} \cdot \mathbf{r}}{a} \right) \mathbf{r}_0 . \end{aligned}$$

3.3.2 Electric potential computation

The geometrical setting is again the one of Figure 3.1. In each domain Ω_i , the potential follows a Poisson equation

$$\sigma_i \Delta V = f_i$$

where f_i is the restriction of $\nabla \cdot \mathbf{J}^P$ to Ω_i . At the interface S_i between Ω_i and Ω_{i+1} , the following jump conditions hold:

$$[V]_{S_i} = 0 \quad (3.4)$$

$$[\sigma \partial_{\mathbf{n}} V]_{S_i} = 0 . \quad (3.5)$$

We define the jump of a function $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ at interface S_j as

$$[f]_j = f_{S_j}^- - f_{S_j}^+,$$

the functions f^- and f^+ on S_j being respectively the interior and exterior limits of f :

$$\text{for } \mathbf{r} \in S_j, \quad f_{S_j}^\pm(\mathbf{r}) = \lim_{\alpha \rightarrow 0^\pm} f(\mathbf{r} + \alpha \mathbf{n}).$$

Note that these quantities depend on the orientation of \mathbf{n} .

Using the same type of technique as for the magnetic field, one can show that the values of the potential (and of the normal current flow) on surfaces S_i are related by integral operators.

Green formula

We recall the Green formula (see proof in appendix A.1.5)

$$\int_{\Omega} (u \Delta v - v \Delta u) d\mathbf{r} = \int_{\partial\Omega} (u \partial_{\mathbf{n}} v - v \partial_{\mathbf{n}} u) ds(\mathbf{r}) .$$

Consider $v = -\frac{1}{4\pi \|\mathbf{r} - \mathbf{r}'\|} = G_L(\mathbf{r} - \mathbf{r}')$ and a harmonic function u .

The left-hand side integral $I(\mathbf{r}) = \int_{\Omega} (u \Delta' v - v \Delta' u) d\mathbf{r}'$ takes different values according to the position of \mathbf{r} with respect to Ω , as summarized below:

$\mathbf{r} \in \Omega$	$I(\mathbf{r}) = u(\mathbf{r})$
$\mathbf{r} \in \mathbb{R}^3 \setminus \bar{\Omega}$	$I(\mathbf{r}) = 0$
$\mathbf{r} \in \partial\Omega$	$I(\mathbf{r}) = \frac{u^-(\mathbf{r})}{2}$

The first two lines of the above table are trivial to prove, and the third relies on solid angle computations (refer to [110] for the proof). Thus seen from inside,

$$I(\mathbf{r}) = \int_{\partial\Omega} u^-(\mathbf{r}) \partial_{\mathbf{n}'} G_L(\mathbf{r} - \mathbf{r}') - G_L(\mathbf{r} - \mathbf{r}') \partial_{\mathbf{n}'} u^-(\mathbf{r}') ds(\mathbf{r}') \quad (3.6)$$

The same treatment can be applied to the volume $\Omega' = \mathbb{R}^3 \setminus \overline{\Omega}$, and seen from Ω , this yields

$$J(\mathbf{r}) = \int_{\partial\Omega} -u^+(\mathbf{r}) \partial_{\mathbf{n}'} G_L(\mathbf{r} - \mathbf{r}') + G_L(\mathbf{r} - \mathbf{r}') \partial_{\mathbf{n}'} u^+(\mathbf{r}') ds(\mathbf{r}') \quad (3.7)$$

with the integral term $J(\mathbf{r})$ equal to

$\mathbf{r} \in \mathbb{R}^3 \setminus \overline{\Omega}$	$J(\mathbf{r}) = u(\mathbf{r})$
$\mathbf{r} \in \Omega$	$J(\mathbf{r}) = 0$
$\mathbf{r} \in \partial\Omega$	$J(\mathbf{r}) = \frac{u^+(\mathbf{r})}{2}$

Summing (3.6) and (3.7), for $\mathbf{r} \in \Omega$,

$$u(\mathbf{r}) = \int_{\partial\Omega} [u] \partial_{\mathbf{n}'} G_L(\mathbf{r} - \mathbf{r}') ds(\mathbf{r}') - \int_{\partial\Omega} [\partial_{\mathbf{n}'} u] G_L(\mathbf{r} - \mathbf{r}') ds(\mathbf{r}')$$

and, for $\mathbf{r} \in \partial\Omega$,

$$\frac{u^+(\mathbf{r}) + u^-(\mathbf{r}')}{2} = \int_{\partial\Omega} [u] \partial_{\mathbf{n}'} G_L(\mathbf{r} - \mathbf{r}') ds(\mathbf{r}') - \int_{\partial\Omega} [\partial_{\mathbf{n}'} u] G_L(\mathbf{r} - \mathbf{r}') ds(\mathbf{r}')$$

To simplify notation, we introduce two integral operators, called the “double-layer” and “single-layer” operators, which map a scalar function f on $\partial\Omega$ to another scalar function on $\partial\Omega$:

$$\begin{aligned} (\mathcal{D}f)(\mathbf{r}) &= \int_{\partial\Omega} \partial_{\mathbf{n}'} G_L(\mathbf{r} - \mathbf{r}') f(\mathbf{r}') ds(\mathbf{r}') \\ (\mathcal{S}f)(\mathbf{r}) &= \int_{\partial\Omega} G_L(\mathbf{r} - \mathbf{r}') f(\mathbf{r}') ds(\mathbf{r}') . \end{aligned}$$

The two above relations become, for $\mathbf{r} \in \Omega$,

$$u(\mathbf{r}) = \mathcal{D}[u] - \mathcal{S}[\partial_{\mathbf{n}} u]$$

and for $\mathbf{r} \in \partial\Omega$,

$$u^\mp(\mathbf{r}) = (\pm \frac{I}{2} + \mathcal{D}) [u] - \mathcal{S}[\partial_{\mathbf{n}} u]$$

This also holds when $\Omega = \Omega_1 \cup \Omega_2 \cup \dots \cup \Omega_N$, with $\partial\Omega = S_1 \cup S_2 \cup \dots \cup S_N$. In this case, for $\mathbf{r} \in S_j$,

$$\frac{u^-(\mathbf{r}) + u^+(\mathbf{r})}{2} = \sum_{k=1}^N \mathcal{D}_{jk} [u]_{S_k} - \mathcal{S}_{jk} [\partial_{\mathbf{n}} u]_{S_k} \quad (3.8)$$

Geselowitz formula

Supposing the primary current \mathbf{J}^P to be restricted to one volume Ω_i , consider V^∞ such that $\sigma_i \Delta V^\infty = \nabla \cdot \mathbf{J}^P$ holds in all \mathbb{R}^3 . Across each surface S_j , the potential V^∞ and its normal derivative $\partial_{\mathbf{n}} V$ are continuous. Consider the function $u = \sigma V - \sigma_i V^\infty$; it is harmonic in each Ω_j , therefore satisfies (3.8). Since $[u]_{S_j} = (\sigma_j - \sigma_{j+1})V_j$ and $[\partial_{\mathbf{n}} u] = 0$ (see proof below), we obtain, on each surface S_j ,

$$\frac{\sigma_j + \sigma_{j+1}}{2} V_j - \sum_{k=1}^N (\sigma_k - \sigma_{k+1}) \mathcal{D}_{jk} V_k = \sigma_i V_j^\infty , \quad (3.9)$$

a formula which was established in 1967 by Geselowitz.

Proof.

$$\begin{aligned} [u]_{S_j} &= u_{S_j}^- - u_{S_j}^+ \\ &= (\sigma_j V_j^- - \sigma_i V_j^\infty) - (\sigma_{j+1} V_j^+ - \sigma_i V_j^\infty) \\ &= (\sigma_j - \sigma_{j+1}) V_j , \end{aligned}$$

because of the continuity of the potentials V and V^∞ across surface S_j .

$$\begin{aligned} [\partial_{\mathbf{n}} u] &= \partial_{\mathbf{n}} u^- - \partial_{\mathbf{n}} u^+ \\ &= (\sigma_j \partial_{\mathbf{n}} V_j^- - \sigma_i \partial_{\mathbf{n}} V_j^\infty) - (\sigma_{j+1} \partial_{\mathbf{n}} V_j^+ - \sigma_i \partial_{\mathbf{n}} V_j^\infty) \\ &= [\sigma \partial_{\mathbf{n}} V]_j = 0 , \end{aligned}$$

because of the continuity of $\sigma \partial_{\mathbf{n}} V$ and $\sigma_i \partial_{\mathbf{n}} V^\infty$ across surface S_j . \square

3.4 Realistic model

For even more realistic models, the piecewise constancy of conductivity that has been made in the previous section needs to be relaxed. Indeed, the brain is known to have strong anisotropies in the conductivities at least in two domains:

- the skull is a non-homogeneous material. It is a porous material with marrow insertions and all kinds of holes filled with air or various liquids (eg. sinuses). Also, its shape is extremely complex and difficult to extract from MRI images, so that it is often “guessed” by using its relative position with respect to the other interfaces. In practice, researchers have found that its conductivity plays a fundamental role in EEG and that it would best modelled (in absence of more direct measurements) with distinct radial and tangential conductivities.

- the white matter is even less homogeneous: it is made of an entanglement of fibers connecting different pieces of the cortex. It has thus a strong anisotropic behavior. While the importance of taking into account this anisotropy for MEG/EEG reconstruction has less been investigated, it is certainly interesting to evaluate its effects and fortunately, there exists a way of measuring it (contrarily to the case of the skull). Diffusion MRI is able to measure the diffusion of water molecules in various directions. Intuitively, the water flow more easily along the direction of the fibers in the white matter than across them. This anisotropy of diffusion of water can be used to model an anisotropic conductivity as currents are certainly better conducted along the fibers than across them.

Dealing with such anisotropies with a BEM like method is impossible most of the time (it would be possible to deal with radial and tangential anisotropies for a spherical head but not much more). Thus, this problem needs to be tackled using directly the Maxwell equation in the quasistatic case. So we start again with the Poisson equation

$$\nabla \cdot (\sigma \nabla V) = f = \nabla \cdot \mathbf{J}^P .$$

To obtain a unique solution, this equation needs to be supplemented with a boundary condition. To do so, we hypothesize that no current flows outside of the head (which is mostly true except at the spinal column which is “far” from most EEG/MEG measurements). We thus have to solve the following problem:

$$\left\{ \begin{array}{l} \nabla \cdot (\sigma \nabla V) = \nabla \cdot \mathbf{J}^P \quad \text{in } \Omega \\ \sigma \frac{\partial V}{\partial \mathbf{n}} = \sigma \nabla V \cdot \mathbf{n} = 0 \quad \text{on } S = \partial \Omega. \end{array} \right. \quad (3.10)$$

This problem will be solved using a Finite Element Approach. We will first show that the PDE 3.10 can be formulated as a variational problem (section 3.4.1), which is then discretized to obtain a linear system (section 3.4.2), and then solved (section 3.4.3).

Anisotropic model Note that in the above formulation, σ can either be taken as a simple scalar function of \mathbf{r} but it can as well be a function that associates a 3D symmetric definite positive matrix at each point of the space.

This matrix is a tensorial description of the anisotropic conductivity (the eigenvalues represent the conductivity along the corresponding eigenvector). Denoting by Σ this matrix, the anisotropic system becomes:

$$\left\{ \begin{array}{l} \nabla \cdot (\Sigma \nabla V) = \nabla \cdot \mathbf{J}^p \quad \text{in } \Omega \\ \Sigma \frac{\partial V}{\partial \mathbf{n}} = \Sigma \nabla V \cdot \mathbf{n} = 0 \quad \text{on } S = \partial \Omega. \end{array} \right. \quad (3.11)$$

For simplicity, we will develop the scalar model hereafter but most of the results can be trivially adapted to the anisotropic case. Notationally, almost nothing changes except that $\sigma(\mathbf{r}) \nabla V(\mathbf{r}) \cdot \nabla w(\mathbf{r})$ is replaced by $\nabla V(\mathbf{r})^T \Sigma \nabla w(\mathbf{r})$ and $\sigma(\mathbf{r}) \|\nabla \phi(\mathbf{r})\|^2$ is replaced by $\nabla \phi(\mathbf{r})^T \Sigma \nabla \phi(\mathbf{r})$.

3.4.1 A variational formulation of the forward problem

Let us first define some functional spaces that will be needed hereafter.

$$H^1(\Omega) = \{w \in L^2(\Omega), \nabla w \in L^2(\Omega)^3\} .$$

$$H^2(\Omega) = \{w \in L^2(\Omega), \nabla w \in H^1(\Omega)^3\} .$$

These spaces simply provide functions that can be simply plugged within the equations that will be used (with all integrals and differentiations well defined).

We first show that the following three problems are equivalent:

① $V \in H^2(\Omega)$ is solution of:

$$\left\{ \begin{array}{l} \nabla \cdot (\sigma \nabla V) = f \quad \text{in } \Omega \\ \sigma \frac{\partial V}{\partial \mathbf{n}} = \sigma \nabla V \cdot \mathbf{n} = g \quad \text{on } S = \partial \Omega. \end{array} \right.$$

② $V \in H^1(\Omega)$ is such that

$$\forall w \in H^1(\Omega) \quad \int_{\Omega} \sigma(\mathbf{r}) \nabla V(\mathbf{r}) \cdot \nabla w(\mathbf{r}) d\mathbf{r} + \int_{\Omega} f(\mathbf{r}) w(\mathbf{r}) d\mathbf{r} - \int_S g(\mathbf{r}) w(\mathbf{r}) ds = 0 .$$

③ $V = \arg \min_{\phi \in H^1(\Omega)} E(\phi)$ with:

$$E(\phi) = \frac{1}{2} \int_{\Omega} \sigma(\mathbf{r}) \|\nabla \phi(\mathbf{r})\|^2 d\mathbf{r} + \int_{\Omega} f(\mathbf{r}) \phi(\mathbf{r}) d\mathbf{r} - \int_S g(\mathbf{r}) \phi(\mathbf{r}) ds .$$

Notice that the PDE in ① is exactly the same as the one in 3.10: we just have renamed $f = \nabla \cdot \mathbf{J}^P$ and allowed for a more general Neumann boundary condition g . This makes the presentation slightly more general and shows that the basic method will remain the same even if we were able to model the currents in the neck. The functions f and g are supposed to be square integrable, that is $f \in L^2(\Omega)$ and $g \in L^2(\Omega)$.

Theorem 3.1. *Problems ①, ② and ③ are equivalent.*

Proof. Problem ② will be used as a pivot. The proof is thus in two parts: equivalence of ① and ②, and equivalence of ② and ③.

① \implies ②

Using the formula: $\nabla \cdot (\sigma w \nabla V) = \sigma \nabla V \cdot \nabla w + w \nabla \cdot \sigma \nabla V$ and integrating it over the domain Ω , we have:

$$\int_{\Omega} w(\mathbf{r}) \nabla \cdot \sigma(\mathbf{r}) \nabla V(\mathbf{r}) d\mathbf{r} = \int_{\Omega} \nabla \cdot (\sigma(\mathbf{r}) w(\mathbf{r}) \nabla V(\mathbf{r})) d\mathbf{r} - \int_{\Omega} \sigma(\mathbf{r}) \nabla V(\mathbf{r}) \cdot \nabla w(\mathbf{r}) d\mathbf{r}$$

In the left hand side of this equation, $\nabla \cdot \sigma \nabla V$ can be replaced by f because of ①. The Green theorem can be used to transform the first term of the right hand side giving:

$$\int_{\Omega} \sigma(\mathbf{r}) \nabla V(\mathbf{r}) \cdot \nabla w(\mathbf{r}) d\mathbf{r} + \int_{\Omega} f(\mathbf{r}) w(\mathbf{r}) d\mathbf{r} - \int_S w(\mathbf{r}) \sigma(\mathbf{r}) \nabla V(\mathbf{r}) \cdot \mathbf{n} ds = 0 .$$

Replacing $\sigma(\mathbf{r}) \nabla V(\mathbf{r}) \cdot \mathbf{n}$ by its value on S as given by the boundary condition of ① yields the result.

$$\int_{\Omega} \sigma(\mathbf{r}) \nabla V(\mathbf{r}) \cdot \nabla w(\mathbf{r}) d\mathbf{r} + \int_{\Omega} f(\mathbf{r}) w(\mathbf{r}) d\mathbf{r} - \int_S g(\mathbf{r}) w(\mathbf{r}) ds = 0 .$$

② \implies ①

If ② is true for any $w \in H^1(\Omega)$, it is also true for $w \in D(\Omega)$ the space of C^∞ functions with compact support in Ω . The dual of $D(\Omega)$ is the space of distributions over Ω , $D'(\Omega)$. If $\nabla \cdot (\sigma \nabla V) \in L^2(\Omega)$, then $\nabla \cdot (\sigma \nabla V) - f \in L^2(\Omega)$ since $f \in L^2(\Omega)$ by hypothesis. Denoting by $\langle \dots \rangle$ the duality bracket between the spaces $L^2(\Omega)$ and $D'(\Omega)$. Thus Eq. ② can be written as $\langle \nabla \cdot (\sigma \nabla V) - f, w \rangle = 0$. From a standard result in functionnal analysis [Brezis 88], $\nabla \cdot (\sigma \nabla V) - f$ is zero almost everywhere.

$\textcircled{3} \implies \textcircled{2}$

If $\textcircled{3}$ is true, then for all $w \in H^1(\Omega)$ and for any real number λ , we have:

$$E(V) \leq E(V + \lambda w) . \quad (3.12)$$

This is true because:

$$\forall V \in H^2(\Omega) \quad \forall w \in H^1(\Omega) \quad \forall \lambda \in \mathbb{R} \quad V + \lambda w \in H^1(\Omega)$$

By definition of E :

$$\begin{aligned} E(V + \lambda w) &= \frac{1}{2} \int_{\Omega} \sigma(\mathbf{r}) \|\nabla(V + \lambda w)(\mathbf{r})\|^2 d\mathbf{r} + \int_{\Omega} f(\mathbf{r})(V + \lambda w)(\mathbf{r}) d\mathbf{r} - \\ &\quad \int_S g(\mathbf{r})(V + \lambda w)(\mathbf{r}) ds \\ &= \frac{1}{2} \int_{\Omega} \sigma(\mathbf{r}) \|\nabla V(\mathbf{r})\|^2 d\mathbf{r} + \int_{\Omega} f(\mathbf{r})V(\mathbf{r}) d\mathbf{r} - \int_S g(\mathbf{r})V(\mathbf{r}) ds + \\ &\quad \lambda \int_{\Omega} \sigma(\mathbf{r}) \nabla V(\mathbf{r}) \cdot \nabla w(\mathbf{r}) d\mathbf{r} + \lambda \int_{\Omega} f(\mathbf{r})w(\mathbf{r}) d\mathbf{r} - \lambda \int_S g(\mathbf{r})w(\mathbf{r}) ds + \\ &\quad \frac{\lambda^2}{2} \int_{\Omega} \sigma(\mathbf{r}) \|\nabla w(\mathbf{r})\|^2 d\mathbf{r} . \end{aligned} \quad (3.13)$$

$$\begin{aligned} E(V + \lambda w) &= E(V) + \frac{\lambda^2}{2} \int_{\Omega} \sigma(\mathbf{r}) \|\nabla w(\mathbf{r})\|^2 d\mathbf{r} \\ &\quad \lambda \left(\int_{\Omega} \sigma(\mathbf{r}) \nabla V(\mathbf{r}) \cdot \nabla w(\mathbf{r}) d\mathbf{r} + \int_{\Omega} f(\mathbf{r})w(\mathbf{r}) d\mathbf{r} - \int_S g(\mathbf{r})w(\mathbf{r}) ds \right) \end{aligned} \quad (3.14)$$

For λ sufficiently small and positive, Eq. 3.12 implies:

$$\int_{\Omega} \sigma(\mathbf{r}) \nabla V(\mathbf{r}) \cdot \nabla w(\mathbf{r}) d\mathbf{r} + \int_{\Omega} f(\mathbf{r})w(\mathbf{r}) d\mathbf{r} - \int_S g(\mathbf{r})w(\mathbf{r}) ds \geq 0 .$$

For λ sufficiently small and negative, Eq. 3.12 implies:

$$\int_{\Omega} \sigma(\mathbf{r}) \nabla V(\mathbf{r}) \cdot \nabla w(\mathbf{r}) d\mathbf{r} + \int_{\Omega} f(\mathbf{r})w(\mathbf{r}) d\mathbf{r} - \int_S g(\mathbf{r})w(\mathbf{r}) ds \leq 0 .$$

Thus:

$$\int_{\Omega} \sigma(\mathbf{r}) \nabla V(\mathbf{r}) \cdot \nabla w(\mathbf{r}) d\mathbf{r} + \int_{\Omega} f(\mathbf{r})w(\mathbf{r}) d\mathbf{r} - \int_S g(\mathbf{r})w(\mathbf{r}) ds = 0 .$$

$\textcircled{2} \implies \textcircled{3}$

From Eq. 3.14 with $\textcircled{2}$, denoting by ϕ the value $V + \lambda w$, it is clear that $E(V)$ is the minimum value of $E(\phi)$. This is true since when w spans $H^1(\Omega)$ and for all real λ , $\phi = V + \lambda w$ spans $H^1(\Omega)$.

□

3.4.2 Discretization of the FEM forward problem

General discrete framework

The FEM forward problem is implemented using the variational formulation ③. The continuous functional spaces are approximated using the Galerkin method yielding a discrete problem.

The 3D space Ω is tessellated with bounded cells (eg tetrahedra or hexahedra) (C_i), $i = 1 \dots N_C$. This tessellation Ω_h also introduces a set of points (V_i), $i = 1 \dots N_V$ (the vertices of the cells) and the space of the continuous functions over Ω is approximated by a vectorial space using some basis functions (w^i), $i = 1 \dots N_V$ defined at each vertex.

$$H_h^1(\Omega_h) = \left\{ \phi_h, \exists (\phi_1, \dots, \phi_{N_V}) \in \mathbb{R}^{N_V}, \quad \phi_h(\mathbf{r}) = \sum_{i=1}^{N_V} \phi_i w^i(\mathbf{r}) \right\},$$

The boundary of the tessellation S_h also defines a tessellation of S the boundary of Ω . Without loss of generality, we assume that the vertices of the tessellation that are on the boundary of the tessellation are (V_i), $i = 1 \dots N_S$ with $N_S < N_V$.

$$H_h^1(S_h) = \left\{ \phi_h, \exists (\phi_1, \dots, \phi_{N_S}) \in \mathbb{R}^{N_S}, \quad \phi_h(\mathbf{r}) = \sum_{i=1}^{N_S} \phi_i w_{S_h}^i(\mathbf{r}) \right\},$$

where $w_{S_h}^i$ is the restriction to S_h of the function w^i .

The discretization of the criterion E in ③ is obtained by using the discretized versions of all the involved functions ϕ , f and g . For σ , we will use a different discretization scheme where σ is given by a constant σ_i over the cell C_i .

$$\begin{aligned} E(\phi_h) &= \frac{1}{2} \int_{\Omega} \sigma(\mathbf{r}) \left\| \nabla \left(\sum_{i=1}^{N_V} \phi_i w^i(\mathbf{r}) \right) \right\|^2 d\mathbf{r} + \int_{\Omega} f(\mathbf{r}) \left(\sum_{i=1}^{N_V} \phi_i w^i(\mathbf{r}) \right) d\mathbf{r} - \\ &\quad \int_S \left(\sum_{i=1}^{N_S} \phi_i w_{S_h}^i(\mathbf{r}) \right) g(\mathbf{r}) ds \\ &= E_h(\Phi), \end{aligned}$$

where $\Phi = (\phi_1, \dots, \phi_{N_V}) \in \mathbb{R}^{N_V}$.

$$E_h(\Phi) = \frac{1}{2} \sum_{i,j=1}^{N_V} \phi_i \phi_j \int_{\Omega} \sigma(\mathbf{r}) \nabla w^i(\mathbf{r}) \cdot \nabla w^j(\mathbf{r}) d\mathbf{r} + \sum_{i=1}^{N_V} \phi_i \int_{\Omega} f(\mathbf{r}) w^i(\mathbf{r}) d\mathbf{r} - \sum_{i=1}^{N_S} \phi_i \int_S w_{S_h}^i(\mathbf{r}) g(\mathbf{r}) ds$$

The minimization of $E(\phi)$ then becomes a simple problem of minimizing in finite dimension the quadratic criterion $E_h(\Phi)$. Denoting by A_{ij} the second order derivative of this criterion $\frac{\partial^2 E_h}{\partial \phi_i \partial \phi_j}$, we have:

$$A_{ij} = \int_{\Omega} \sigma(\mathbf{r}) \nabla w^i(\mathbf{r}) \cdot \nabla w^j(\mathbf{r}) d\mathbf{r} .$$

Note that the matrix \mathbf{A} is naturally symmetric. Usually the basis functions will have a very local support, so that \mathbf{A} will also be very sparse. We also introduce the vector \mathbf{B} :

$$B_i = \begin{cases} \int_{\Omega} f(\mathbf{r}) w^i(\mathbf{r}) d\mathbf{r} - \int_S w_{S_h}^i(\mathbf{r}) g(\mathbf{r}) ds & \text{for } i \leq N_S \\ \int_{\Omega} f(\mathbf{r}) w^i(\mathbf{r}) d\mathbf{r} & \text{otherwise} . \end{cases}$$

The criterion $E_h(\Phi)$ can be written as $\frac{1}{2} \Phi^T \mathbf{A} \Phi + \mathbf{B} \cdot \Phi$ and optimality is obtained when $\mathbf{A} \Phi + \mathbf{B} = 0$.

$$\Phi^T \mathbf{A} \Phi = \sum_{i,j=1}^{N_V} \phi_i \phi_j A_{ij} = \sum_{i,j=1}^{N_V} \int_{\Omega} \phi_i \phi_j \sigma(\mathbf{r}) \nabla w^i(\mathbf{r}) \cdot \nabla w^j(\mathbf{r}) d\mathbf{r} = \int_{\Omega} \sigma(\mathbf{r}) \|\nabla \phi_h(\mathbf{r})\|^2 d\mathbf{r} \quad (3.15)$$

This proves that the matrix \mathbf{A} is positive because $\sigma > 0$ over Ω . Note, however that the matrix is not definite. Indeed, $\Phi^T \mathbf{A} \Phi$ is zero iff $\nabla \phi_h(\mathbf{r}) = 0$ on Ω almost everywhere. This is natural as the original equation is insensitive to the addition to V of a constant function over Ω . On our discretized spaces, this happens for $\Phi = Cst \mathbf{1}$ (this is the case whenever the constant function over Ω_h belongs the space $H_h^1(\Omega_h)$ which is the case for the standard basis functions P^1 or Q^1 used for tetrahedric or hexahedric cells respectively). Similarly to Eq. 3.15, we can prove that $\Phi^T \mathbf{A} \Psi = \int_{\Omega} \sigma(\mathbf{r}) \nabla \phi_h(\mathbf{r}) \cdot \nabla \psi_h(\mathbf{r}) d\mathbf{r}$. Applying this result to $\Psi = \mathbf{1}$, proves that the kernel of the matrix \mathbf{A} is the constant vector $\mathbf{1}$. Rewriting this result for each line of the matrix gives:

$$\forall i = 1 \dots N_V \quad \sum_{j=1}^{N_V} A_{ij} = 0 . \quad (3.16)$$

This result can be used to reduce the amount of memory used to store the matrix \mathbf{A} . Indeed, Eq 3.16 can be rewritten as:

$$A_{ii} = - \sum_{j \neq i} A_{ij} . \quad (3.17)$$

This can be used to rewrite conveniently the part of the criterion $E_h(\phi_h)$ containing \mathbf{A} .

$$C(\Phi) = \frac{1}{2} \Phi^T \mathbf{A} \Phi = \frac{1}{2} \sum_{i,j=1}^{N_V} \phi_i \phi_j A_{ij} = \frac{1}{2} \sum_{i=1}^{N_V} \left(A_{ii} \phi_i^2 + \sum_{i \neq j} A_{ij} \phi_i \phi_j \right) .$$

Replacing A_{ii} by its value given by Eq. 3.17 yields:

$$\begin{aligned}
C(\Phi) &= \frac{1}{2} \sum_{i=1}^{N_V} \sum_{i \neq j} (-A_{ij}\phi_i^2 + A_{ij}\phi_i\phi_j) \\
&= \frac{1}{2} \sum_{i=1}^{N_V} \sum_{i \neq j} A_{ij} (\phi_j - \phi_i) \phi_i \\
&= \frac{1}{2} \sum_{i=1}^{N_V} \sum_{i < j} A_{ij} [(\phi_j - \phi_i) \phi_i + (\phi_i - \phi_j) \phi_j] \quad \text{since } A_{ji} = A_{ij} \\
&= -\frac{1}{2} \sum_{i=1}^{N_V} \sum_{i < j} A_{ij} (\phi_j - \phi_i)^2
\end{aligned}$$

This formulation has a double advantage:

- it totally removes the need for the terms A_{ii} which do not need to be computed nor stored.
- if we recall that ϕ_h represents a potential, the criterion explicitly involves the differences of these potential values. This is physically very natural.

A P^1 implementation

In practice, we will use P^1 basis function over a tetrahedral mesh. In dimension d , a tetrahedron T_j is defined by $d+1$ vertices $\mathbf{V}_{i_k}, k = 1 \dots d+1$. The restriction w_j^i of w^i to T_j is defined by (for notational simplicity, we assume without loss of generality that $\mathbf{V}_{i_k} = \mathbf{V}_k$):

$$w_j^i(\mathbf{r}) = \begin{cases} 0 & \text{if the point } \mathbf{r} \text{ does not belong to } T_j \\ \frac{|\mathbf{r} \cdot \mathbf{V}_2 \dots \mathbf{V}_{d+1}|}{|\mathbf{V}_1 \dots \mathbf{V}_{d+1}|} & \text{otherwise,} \end{cases}$$

where the vectors in the determinants are written with $d+1$ coordinates: the usual d coordinates are augmented with a final 1 (this is called projective or homogeneous coordinates). Because w_j^i is a linear function of \mathbf{r} , its gradient is a constant vector which can be computed as the first row of the inverse matrix $[\mathbf{V}_1 - \mathbf{V}_{d+1} \dots \mathbf{V}_d - \mathbf{V}_{d+1}]^{-1}$. Given this value of the gradient ∇w_j^i , the matrix \mathbf{A} is computed as:

$$A_{ij} = \sum_{k: \mathbf{V}_i \in T_k, \mathbf{V}_j \in T_k} \int_{T_k} \sigma_k \nabla w_k^i \cdot \nabla w_k^j d\mathbf{r} = \sum_{k: \mathbf{V}_i \in T_k, \mathbf{V}_j \in T_k} \sigma_k \nabla w_k^i \cdot \nabla w_k^j \text{vol}(T_k),$$

This shows that the only non-diagonal (since diagonal elements do not need to be computed) non-zero coefficients of the matrix \mathbf{A} correspond to the edges of the mesh. Consequently, the matrix \mathbf{A} is very sparse.

Since we usually assume that $g = 0$ for the forward problem, the computation of \mathbf{B} includes only the term containing $f = \nabla \cdot \mathbf{J}^P$.

$B_i = \int_{\Omega} \nabla \cdot \mathbf{J}^P(\mathbf{r}) w^i(\mathbf{r}) d\mathbf{r}$ becomes:

$$\begin{aligned} B_i &= \int_{\Omega} \nabla \cdot (w^i(\mathbf{r}) \mathbf{J}^P(\mathbf{r})) d\mathbf{r} - \int_{\Omega} \nabla w^i(\mathbf{r}) \cdot \mathbf{J}^P(\mathbf{r}) d\mathbf{r} \\ &= \int_S w^i(\mathbf{r}) \mathbf{J}^P(\mathbf{r}) \cdot ds(\mathbf{r}) - \int_{\Omega} \nabla w^i(\mathbf{r}) \cdot \mathbf{J}^P(\mathbf{r}) d\mathbf{r} \\ &= \int_S w^i(\mathbf{r}) \mathbf{J}^P(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r}) ds(\mathbf{r}) - \int_{\Omega} \nabla w^i(\mathbf{r}) \cdot \mathbf{J}^P(\mathbf{r}) d\mathbf{r} \end{aligned}$$

To go further, one needs to choose a proper model for \mathbf{J}^P .

The continuous case In this first case, \mathbf{J}^P is represented as $\mathbf{J}^P(\mathbf{r}) = \sum_{j=1}^{N_V} w^j(\mathbf{r}) \mathbf{J}_j^P$. Consequently:

$$B_i = \sum_{j=1}^{N_V} \mathbf{J}_j^P \cdot \int_S w^i(\mathbf{r}) w^j(\mathbf{r}) \mathbf{n}(\mathbf{r}) ds(\mathbf{r}) - \sum_{j=1}^{N_V} \mathbf{J}_j^P \cdot \int_{\Omega} w^j(\mathbf{r}) \nabla w^i(\mathbf{r}) d\mathbf{r}$$

Assuming that there are no sources on S gives:

$$B_i = - \sum_{j=1}^{N_V} \mathbf{J}_j^P \int_{\Omega} w^j(\mathbf{r}) \nabla w^i(\mathbf{r}) d\mathbf{r}$$

The Dirac case In this case, the current distribution is localized at a single point in space \mathbf{r}_0 (or at a few places by linear combination). At this point in space the current orientation and strength are described by a vector μ . Thus, $\mathbf{J}^P(\mathbf{r})$ can be written as: $\mathbf{J}^P(\mathbf{r}) = \mu \delta(\mathbf{r} - \mathbf{r}_0)$. Consequently,

$$B_i = \mu \int_S w^i(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}_0) \cdot \mathbf{n}(\mathbf{r}) d\mathbf{r} - \mu \cdot \nabla w^i(\mathbf{r}_0)$$

Assuming again that there are no sources on S gives:

$$B_i = -\mu \cdot \nabla w^i(\mathbf{r}_0)$$

3.4.3 Solving the FEM forward problem

The discretized problem solution can be found by solving $\mathbf{AV} + \mathbf{B} = \mathbf{0}$. This is a fairly simple matrix problem, which could in theory be solved using a pseudo-inverse (the inverse of \mathbf{A} does not exist since its kernel is the vector $\mathbf{1}$). However, \mathbf{A} is a very big matrix which can be stored only because it is sparse. Since the pseudo-inverse (or the inverse) of a sparse matrix is usually not a sparse matrix, it is unwise to try to compute it, as the amount of memory needed to store it will be huge. It is thus a much better idea to solve $\mathbf{AV} + \mathbf{B} = \mathbf{0}$ using an iterative method for each \mathbf{B} . Since the matrix \mathbf{A} is symmetric and positive, the conjugate gradient method can be used. Strictly speaking, this method can only be used with definite matrices, but it actually works in this case provided that \mathbf{B} is in the range of matrix \mathbf{A} . Because \mathbf{A} is symmetric, the property (3.16) also holds for column vectors of \mathbf{A} , which means that these columns are all orthogonal to the constant vector $\mathbf{1}$. These vectors thus all have the property that their mean value is zero. It can be verified that this property is true for the various versions of \mathbf{B} detailed above.

The conjugate gradient method only needs to evaluate the quantity $\mathbf{g} = \mathbf{A}\Phi + \mathbf{B}$ for any given parameter vector Φ and the quantity $\mathbf{g}^T \mathbf{Ag}$ to compute the optimal step at each iteration. These two quantities can be easily evaluated using simple traversal of all the edges of the mesh.

In general, the conjugate gradient method only requires a few steps to converge (typically one or two magnitude order less iterations than N_V the size of the matrix). However, due to the very different values of the conductivities for the various parts of the head, this leads to badly conditionned matrices \mathbf{A} . To improve the speed of converge, it is thus preferable to use a preconditioned conjugate gradient method. A simple diagonal preconditioner is already very effective. Such a preconditioner is obtained by the inverting a diagonal matrix whose entries are the diagonal elements of the matrix \mathbf{A} .

Chapter 4

Leadfield computation

The inverse source problem in electroencephalography (EEG) and magnetoencephalography (MEG) which will be the topic of chapter 8 traditionnally “compares” actual measurements to quantities simulated using a forward model (as depicted in chapter 3) that predicts the electric potential or magnetic field at the sensor level.

Consequently, the inverse problem needs only a subset of the values computed by forward problems (the values at sensors positions). This subset of values for every possible source is called the leadfield, which basically directly relates brain sources to measurements at the sensors. For complex geometries, there is no analytical formula of the leadfield. The common approach is to numerically compute the value of the leadfield for a finite number of point sources (dipoles). There are several drawbacks: the model of the source space is fixed (a set of dipoles) and the computation can be expensive for as much as 10 000 dipoles.

There are more efficient methods to compute the leadfield using what is called the adjoint operator. This is the main topic of this chapter.

4.1 The leadfield

From the previous chapter, it is obvious that both the potential V and the magnetic field \mathbf{B} are linear functions of the sources \mathbf{J}^p . Because of this property, the predictions of the forward problem at the sensor locations (the only information used by inverse problems) can be conveniently summarized by a leadfield, which is a linear operator \mathbf{L} that maps source activations to simulated potentials of magnetic fields at sensor locations. Since it is linear, this operator can be written using the Riesz representation theorem:

$$\mathbf{L}(\mathbf{J}^p) = \left\langle \frac{\partial \mathbf{L}}{\partial \mathbf{J}^p}, \mathbf{J}^p \right\rangle, \quad (4.1)$$

where $\langle \cdot, \cdot \rangle$ denotes for the inner product¹ of functions in \mathbb{R}^3 :

$$\langle u, v \rangle = \int_{\Omega} u(\mathbf{r}) \cdot v(\mathbf{r}) dr .$$

Traditionally, the forward problem equations are discretized and the leadfield is represented by the matrix $\frac{\partial \mathbf{L}}{\partial \mathbf{J}^P}$ whose number of lines equals the number of sensors and whose number of columns equals the number of parameters of the dipoles considered as possible sources (around 10 000 in distributed source models). Each column of this matrix gives the values of the electric or magnetic measurements at the sensor locations for a given dipole parameter. A simple way to compute the leadfield matrix is thus to solve as many forward problems as there are source parameters, which can be quite time consuming with detailed realistic geometries for the head. One solution to bypass this problem relies on some matrix manipulations on the discretized problem so as to estimate rows of the leadfield matrix (instead of columns) [160, 161]. This is quite interesting as there are generally much less sensors than source positions and as the resulting computation is very similar to the original forward problem. Another solution is to work in the continuous domain, and to use the Helmholtz reciprocity principle to express the leadfields as the solutions of different forward problems [160, 113, 133].

Both these approaches can be explained using the adjoint method [87], which is applied here in the continuous domain to both the EEG and MEG leadfields. As the approach is very general, it is quite easy to include within this framework new information such as sensor geometry.

4.2 The adjoint method

Let p represent a parameter which can be a real scalar or vector function. For any p , we consider the (hopefully unique) function v which satisfies the state equation:

$$\mathbf{A}v = b(p) , \quad (4.2)$$

where \mathbf{A} is a linear operator of \mathbb{R}^3 and b is a differentiable function of \mathbb{R}^3 . To give some intuition, in our case the state function v is the electric potential $V(\mathbf{r})$, the parameter p is the primary source current vector $\mathbf{J}^P(\mathbf{r})$, the operator \mathbf{A} is the generalized Laplacian $\nabla \cdot (\sigma \nabla \cdot)$ and b is the divergence operator.

Let us consider a “measurement function” g which is a linear functional of p and v . From the Riesz representation theorem, there exists m and $n \in E$ such that $g(p, v) = \langle m, v \rangle + \langle n, p \rangle$. To compute the differential of g with respect to p (knowing that v also depends linearly on p), we first build

¹For scalar functions the inner product inside the integral becomes a simple product.

a Lagrangian \mathcal{L} by adding to the measurement equation the inner product of the state equation with a Lagrange multiplier w :

$$\mathcal{L}(p, v, w) = \langle m, v \rangle + \langle n, p \rangle + \langle w, \mathbf{A}v - b(p) \rangle .$$

Notice that $g(p, v(p)) = \mathcal{L}(p, v(p), w)$ for all w as $\mathbf{A}v(p) = b(p)$. Assuming that \mathcal{L} is differentiable with respect to both v and p , the Lagrange multiplier w is chosen such that $\frac{\partial \mathcal{L}}{\partial v}(p, v(p), w)\delta v = \langle m, \delta v \rangle + \langle w, \mathbf{A}\delta v \rangle = 0$ for all δv . Introducing the adjoint operator \mathbf{H}^* of a linear operator \mathbf{H} , such that $\langle x, \mathbf{H}y \rangle = \langle \mathbf{H}^*x, y \rangle$, the previous equation can be rewritten as $\langle m + \mathbf{A}^*w, \delta v \rangle = 0$ for all δv , which is equivalent to:

$$\mathbf{A}^*w = -m . \quad (4.3)$$

This equation is called the adjoint equation. With this choice for w , the differential of g can be written as:

$$\delta g = \frac{\partial \mathcal{L}}{\partial p}(p, v(p), w)\delta p = \langle n - \frac{\partial b^*}{\partial p}w, \delta p \rangle . \quad (4.4)$$

Since g is linear in p , $g(p) = \langle \frac{\partial g}{\partial p}, p \rangle$. $g(p)$ being a measurement, $\frac{\partial g}{\partial p}$ is exactly the leadfield \mathbf{L} and Eq. (4.4) exactly expresses this operator as a function of the adjoint variable w :

$$\mathbf{L} = n - \frac{\partial b^*}{\partial p}w . \quad (4.5)$$

4.3 Application to M/EEG leadfields

This section briefly explains the application of the previous result to M/EEG leadfields. For more detailed calculations, the reader is referred to [153]. Using the previous result first requires the computation of the adjoint operators associated to our state equation (3.10) (i.e. \mathbf{A}^* and $\frac{\partial b^*}{\partial p}$). In our case, v is the potential $V(\mathbf{r})$, p is the primary source $\mathbf{J}^P(\mathbf{r})$ and b is the divergence operator, which is linear so that $\frac{\partial \nabla \cdot \mathbf{J}^P}{\partial \mathbf{J}^P} \mathbf{J}^P = \nabla \cdot \mathbf{J}^P$. Using the divergence theorem:

$$\langle w, \nabla \cdot \mathbf{J}^P \rangle = \int_{\Omega} w \nabla \cdot \mathbf{J}^P dr = - \int_{\Omega} \mathbf{J}^P \cdot \nabla w dr = \langle -\nabla w, \mathbf{J}^P \rangle ,$$

because sources are restricted to be strictly inside the head, so the boundary integrals vanish. Thus:

$$\frac{\partial b^*}{\partial p}w = \frac{\partial \nabla \cdot \mathbf{J}^P}{\partial \mathbf{J}^P}w = -\nabla w . \quad (4.6)$$

For \mathbf{A}^* , we start with $\langle w, \mathbf{A}v \rangle$:

$$\langle w, \mathbf{A}v \rangle = \int_{\Omega} w \nabla \cdot (\sigma \nabla V) dr .$$

Assuming that w is continuous on Ω , but that $\sigma \nabla w \cdot \mathbf{n}$ can be discontinuous at tissue interfaces, the divergence theorem applied twice leads to (using the facts that $\sigma \nabla V \cdot \mathbf{n} = 0$ on $\partial\Omega$ and that V and $\sigma \nabla V \cdot \mathbf{n}$ are continuous functions over Ω):

$$\begin{aligned} \langle A^*w, v \rangle &= \int_{\Omega} w \nabla \cdot (\sigma \nabla V) dr \\ &= \int_{\Omega} V \nabla \cdot (\sigma \nabla w) dr - \sum_k \int_{S_k} V [\sigma \nabla w \cdot \mathbf{n}]_{S_k} ds , \end{aligned} \quad (4.7)$$

where S_k are the tissue interfaces and $[.]_{S_k}$ denotes the jump of a function on a given surface S_k .

4.3.1 EEG leadfield

One line of the EEG leadfield is the potential of an electrode located at point \mathbf{r}_i with respect to some reference electrode located at \mathbf{r}_0 (mean reference can also easily be handled with slightly more complicated formulae). Computing a leadfield for dipoles with fixed orientations can also be done easily following the same principle.

$$\begin{aligned} g_{EEG}(p, v) &= g_{EEG}(\mathbf{J}^P, V) = V(\mathbf{r}_i) - V(\mathbf{r}_0) \\ &= \int_{\partial\Omega} V(\mathbf{r}) (\delta(\mathbf{r}_i) - \delta(\mathbf{r}_0)) dr , \end{aligned}$$

where the notation $\delta(\mathbf{r})$ refers to the Dirac distribution at $\mathbf{r} \in \partial\Omega$. The integral is restricted to $\partial\Omega$ since all the EEG sensors are located on this surface. As $g(\mathbf{J}^P, V)$ does not depend on \mathbf{J}^P , this leads to $m_{EEG} = \delta(\mathbf{r}_i) - \delta(\mathbf{r}_0)$ and $n_{EEG} = 0$. Equating all the surfacic and volumic terms in the linear forms shows that $[\sigma_k \nabla w \cdot \mathbf{n}]_{S_k} = 0$ for all inner interfaces (as there are no other surfacic terms for those), hence only the last surface ($\partial\Omega$) is to be kept in the sum in Eq. (4.7):

$$\langle \mathbf{A}^*w, V \rangle = \int_{\Omega} V \nabla \cdot (\sigma \nabla w) dr - \int_{\partial\Omega} V [\sigma \nabla w \cdot \mathbf{n}]_{\partial\Omega} ds .$$

From Eq. (4.3), w satisfies:

$$\boxed{\begin{cases} \nabla \cdot (\sigma \nabla w) = 0 & \text{in } \Omega \\ \sigma \nabla w \cdot \mathbf{n} = \delta_{\mathbf{r}_i} - \delta_{\mathbf{r}_0} & \text{on } \partial\Omega \end{cases}} \quad (4.8)$$

Combining Eq. (4.5) and Eq. (4.6), the EEG leadfield for punctual electrodes is:

$$\boxed{\mathbf{L}_{EEG} = \nabla w} \quad (4.9)$$

4.3.2 MEG leadfield

For MEG, the only difference is in the type of measurements: $g_{MEG}(p, v) = g_{MEG}(\mathbf{J}^P, V) = \mathbf{B}(\mathbf{r}_i) \cdot \mathbf{d}_i$, with \mathbf{B} given by the Biot-Savart law (2.8). This law has two terms, \mathbf{B}_0 depending on \mathbf{J}^P and the other term depending on V . Applying the adjoint state approach requires the computation of the corresponding m_{MEG} and n_{MEG} .

n_{MEG} corresponds to the \mathbf{B}_0 term:

$$\begin{aligned}\mathbf{B}_0(\mathbf{r}_i) \cdot \mathbf{d}_i &= \frac{\mu_0}{4\pi} \int_{\Omega} \left(\mathbf{J}^P \times \nabla \left(\frac{1}{R} \right) \right) \cdot \mathbf{d}_i \, dr \\ &= \frac{\mu_0}{4\pi} \int_{\Omega} \mathbf{J}^P \cdot \left(\nabla \left(\frac{1}{R} \right) \times \mathbf{d}_i \right) \, dr.\end{aligned}$$

Thus

$$n_{MEG} = \frac{\mu_0}{4\pi} \left(\nabla \left(\frac{1}{R} \right) \times \mathbf{d}_i \right). \quad (4.10)$$

m_{MEG} corresponds to the Ohmic current term:

$$\begin{aligned}< m_{MEG}, \mathbf{J}^P > &= -\frac{\mu_0}{4\pi} \int_{\Omega} \sigma \nabla V \cdot \left(\nabla \left(\frac{1}{R} \right) \times \mathbf{d}_i \right) \, dr \\ &= \frac{\mu_0}{4\pi} \int_{\Omega} V \nabla \cdot \sigma \left(\left(\frac{1}{R} \right) \times \mathbf{d}_i \right) \, dr - \\ &\quad \frac{\mu_0}{4\pi} \int_{\Omega} \nabla \cdot V \sigma \left(\left(\frac{1}{R} \right) \times \mathbf{d}_i \right) \, dr \\ &= \frac{\mu_0}{4\pi} \int_{\Omega} V \nabla \cdot \sigma \left(\left(\frac{1}{R} \right) \times \mathbf{d}_i \right) \, dr - \\ &\quad \frac{\mu_0}{4\pi} \sum_k \int_{S_k} V[\sigma]_{S_k} \left(\left(\frac{1}{R} \right) \times \mathbf{d}_i \right) \cdot \mathbf{n} \, ds.\end{aligned}$$

Thus m_{MEG} has both a volumic term and several surfacic terms. Equating all the surfacic and volumic terms in the linear forms yields the following equation for the adjoint:

$\begin{cases} \nabla \cdot (\sigma \nabla w) &= -\frac{\mu_0}{4\pi} \nabla \cdot (\sigma \nabla \left(\frac{1}{R} \right) \times \mathbf{d}_i) & \text{in } \Omega_k, \\ [\sigma \nabla w \cdot \mathbf{n}]_{S_k} &= -\frac{\mu_0}{4\pi} [\sigma]_{S_k} \nabla \left(\frac{1}{R} \right) \times \mathbf{d}_i \cdot \mathbf{n} & \text{on } S_k \end{cases}$
--

Note that if the conductivity σ is constant and isotropic in a domain Ω_k , then $\nabla \cdot (\sigma \nabla \left(\frac{1}{R} \right) \times \mathbf{d}_i) = 0$ and w is harmonic in each domain Ω_k . This result has been used previously for the numerical computation of the MEG leadfield [113, 133].

Combining Eq. (4.5), Eq.(4.10) and Eq. (4.6), the MEG leadfield for a punctual magnetometer is:

$\mathbf{L}_{MEG} = \nabla w + \frac{\mu_0}{4\pi} \left(\nabla \left(\frac{1}{R} \right) \times \mathbf{d}_i \right).$
--

4.4 Incorporating sensor geometry

4.4.1 MEG: Squids geometry

A magnetometer measures the flux of the magnetic field through a small loop. For a magnetometer i , let M_i be the surface enclosed by the loop, and \mathbf{d}_i the unitary vector normal to M_i . The function g can thus be written as $g(\mathbf{J}^P) = \int_{M_i} \mathbf{d}_i \cdot \mathbf{B}(\mathbf{r}') d\mathbf{r}'$. Switching the integrals over M_i and Ω and since the only dependence of $\mathbf{B}(\mathbf{r}')$ on \mathbf{r}' is in R , it is quite easy to show that following the lines of section III.B yields:

$$\begin{cases} \nabla \cdot (\sigma \nabla w) &= -\frac{\mu_0}{4\pi} \int_{M_i} \nabla \cdot (\sigma \nabla (\frac{1}{R}) \times \mathbf{d}_i) d\mathbf{r}' \quad \text{in } \Omega_k, \\ [\sigma \nabla w \cdot \mathbf{n}]_{S_k} &= -\frac{\mu_0}{4\pi} [\sigma]_{S_k} \nabla (\frac{1}{R}) \times \mathbf{d}_i \cdot \mathbf{n} \quad \text{on } S_k \end{cases}$$

As in the point wise case, the term $\int_{M_i} \nabla \cdot (\sigma \nabla (\frac{1}{R}) \times \mathbf{d}_i) d\mathbf{r}'$ vanishes if σ is constant and isotropic in a domain Ω_k . The resulting expression for the leadfield is:

$$\mathbf{L}_{MEG} = \nabla w + \frac{\mu_0}{4\pi} \int_{M_i} \left(\nabla \left(\frac{1}{R} \right) \times \mathbf{d}_i \right) d\mathbf{r}' .$$

This formulation is the same as the one given in [161]. Gradiometers can be treated similarly. Generally, MEG manufacturers give a set of positions and weights for each sensor, and the linear combination of the magnetic field at these positions using these weights is meant to recreate the measurement. But this requires as many adjoint problems as there are positions. The main advantage of the proposed approach is to require a single adjoint problem per sensor, reducing notably the needed for the leadfield computation.

4.4.2 EEG: surface electrodes

Incorporating the surface of EEG electrodes in the model is more complicated as the electrode perturbs the electrical field due to the *shunt effect*: potential is constant at the surface of the electrodes due to the high conducting electrode (gel and metal). This effect has been modeled for EEG [115] and requires a modification of the boundary condition of the state equation (3.10), which becomes:

$$\begin{cases} V + z_k \sigma \nabla V \cdot \mathbf{n} &= v_k \quad \text{on } e_k \\ \int_{e_k} \sigma \nabla V \cdot \mathbf{n} dr &= 0 \\ \sigma \nabla V \cdot \mathbf{n} &= 0 \quad \text{on } S' \cup e_k \end{cases} \quad (4.11)$$

where e_k is the k -th electrode, v_k is the constant value of the potential on the k -th electrode and z_k is the effective contact impedance which models the

electrochemical effect at the skin-electrode interface. Following the adjoint approach yields the same leadfield equation as in the pointwise EEG case (Eq. (4.9)). However, the boundary condition of the adjoint state equation for w is transformed into:

$$\begin{cases} w + z_k \sigma \nabla w \cdot \mathbf{n} = W_k & \text{on } e_k, \\ \int_{e_k} \sigma \nabla w \cdot \mathbf{n} dr = \begin{cases} 1 & k = i \\ -1 & k = 0 \\ 0 & \text{otherwise} \end{cases}, \\ \sigma \nabla w \cdot \mathbf{n} = 0 & \text{on } S \setminus \cup e_k, \end{cases} \quad (4.12)$$

where W_k is a constant.

4.5 Results

To illustrate the adjoint approach, leadfields for magnetometers are computed for a standard three nested spheres model with radii of 0.87, 0.92, 1, meant to represent brain, skull and scalp tissues with respective conductivities of 1, 0.02, 1. While it is known that in a spherical geometry the magnetic field outside the conductor does not depend on the conductivities, we intentionally put different conductivities in our model to test that the numerical solution is actually similar to the case of a homogeneous sphere. Computations were made using a finite element method presented in [117] and results were compared with the ground truth given by the analytical formulation. The magnetometer is oriented in the x direction, positioned at a distance of 0.03 of the outermost sphere, and has a radius of 0.015. The differential equation of section IV A was solved, using our finite element method. The integral term over the magnetometer loop was computed after applying Stokes theorem with a Gauss-Kronod method using 61 points. Figure 4.1 shows the absolute error is similar and small for three dipole orientations. More results can be found in [153].

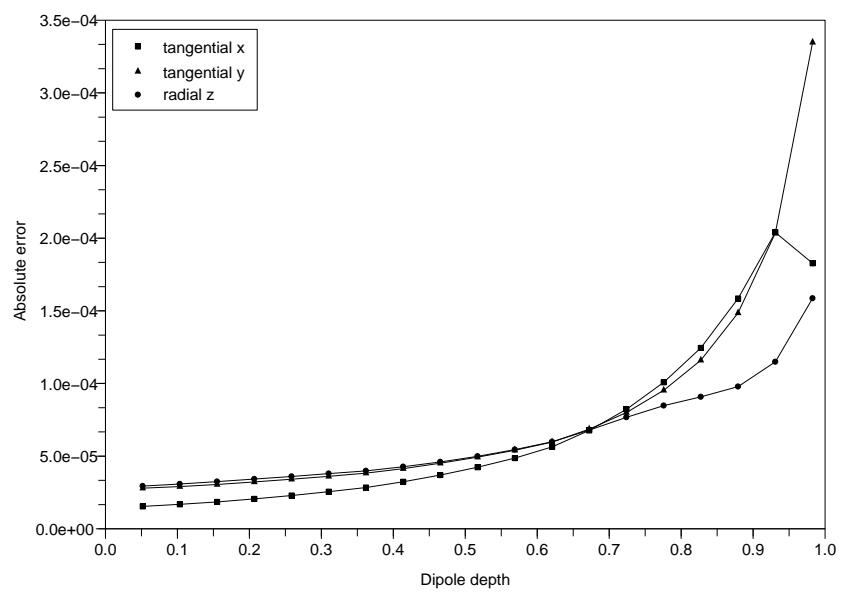


Figure 4.1: Absolute error between numerical and analytical solution with respect to dipole depth.

Chapter 5

Geometric modeling of the head

5.1 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) was developed from knowledge gained in the study of nuclear magnetic resonance. The acronym NMR (Nuclear Magnetic Resonance) is still often used to describe the technique.

5.1.1 Basic principle of NMR

MRI relies (most often) on the relaxation properties of excited hydrogen nuclei in water. Since hydrogen nuclei have a single proton, they have a spin, which is an intrinsic angular moment. It can be associated with a magnetic dipole moment: each hydrogen nucleus behaves as a tiny magnet, with the north/south axis parallel to the spin axis. The sum of the moments of a sample of molecules is zero in the absence of a magnetic field. When an object to be imaged is placed in the powerful, uniform magnetic field \mathbf{B}_0 , the spins of the nuclei within the tissue precess around the direction of \mathbf{B}_0 . The resulting magnetic moment of a sample is oriented in the direction of \mathbf{B}_0 . The frequency ν_0 (Larmor frequency) of the precession is linearly related to the field by the gyromagnetic ratio γ , whose value depends on the nature of the nuclei.

$$\nu_0 = \gamma |\mathbf{B}_0| \quad (5.1)$$

Besides the precession of the nuclei, a second phenomenon is important to us: the relaxation of the nuclei. In the presence of a constant field \mathbf{B}_0 , the spin axes of the nuclei slowly tend to align with \mathbf{B}_0 . The Radio Frequency pulse (RF pulse) technique consists in applying in addition to \mathbf{B}_0 a transient field pulse \mathbf{B}_1 , orthogonal to \mathbf{B}_0 , rotating at the resonance frequency of the

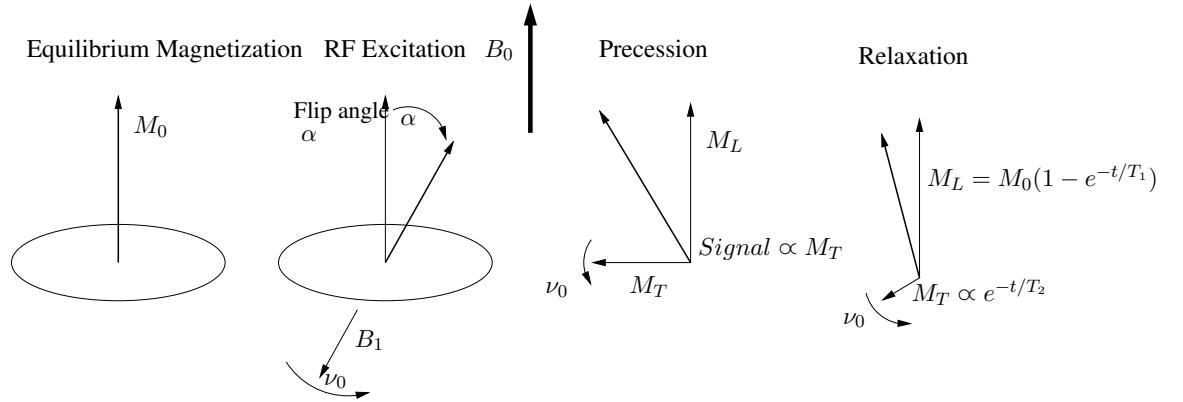


Figure 5.1: The basic physics of the NMR experiment: in a magnetic field \mathbf{B}_0 , an equilibrium magnetization M_0 forms due to the alignment of nuclear dipoles (left). An RF pulse tips over M_0 creating a longitudinal component M_L and a transverse component M_T (middle). M_T precesses around the direction of \mathbf{B}_0 , generating a detectable MR signal. Over time M_T decays to zero with a relaxation time T_2 and M_L returns to M_0 with a relaxation time T_1 (right). This picture is taken from [18].

nuclei ν_0 , and several orders of magnitude smaller. When such an RF pulse is applied the resulting moments M of the nuclei are flipped (usually by 30 or 90 degrees, according to the duration of the pulse). After the pulse, M precesses around \mathbf{B}_0 and finally aligns with \mathbf{B}_0 : the transient transversal moment iM_T also called *Free Induction Decay* (FID) cancels with a time constant T_2 while the longitudinal moment M_L reaches its equilibrium with a time constant T_1 . The values of M_L and M_T are measured using coils. T_1 and T_2 depend on the environment, so that their local value can be used to discriminate between the tissues, proton density being a third signature to discriminate between tissues. An illustration of the phenomenon is given in figure 5.1.

Measuring M_L (resp. M_T) leads to T_1 - (resp. T_2 -) weighted images. A subtle but important variant of the T_2 technique is called T_{2*} imaging. In order to obtain an image, RF pulses are applied repeatedly: such a repetition is called a pulse sequence. Many kinds of pulse sequences are possible, and the sensitivity of the MR images to the different parameters can be adjusted by tuning the repetition time (TR) between consecutive pulses and the time between the RF pulse and the measurement TE (time to echo). Typical sequences may be of several kinds:

- The Gradient Echo Pulse Sequence simply consists of the repetition of the FID described previously. It is simply described by the value of

the flip angle α and the repetition time TR.

- The Spin Echo Pulse Sequence consists in applying a first 90 degrees pulse, then after a time $TE/2$ a 180 degrees pulse in the transverse plane. The effect of this pulse is to refocus the signal whose phase has been quickly dispersed by local field inhomogeneities. Thus, an echo of signal appears at time TE and is measured. This echo can be repeated many times to sample the T_2 decay.
- The Inversion Recovery Pulse Sequence begins with a 180 degrees pulse and after a delay TI a 90 degrees pulse. It enhances the T_1 weighting of the image.

5.1.2 MRI scanning

To selectively image different voxels (picture elements) of a subject, magnetic gradients are applied. Because of the relation (5.1), the spatial variation of the magnetic field magnitude induces a Larmor frequency variation which can be used to localize the piece of material that generated it. Gradients are typically applied during the RF pulses, during the recording of the generated signal and between these two time instants to encode a slice selection, and a position in the slice with a frequency and a phase (see figure 5.2). The same coils are used for the transmission of gradients and the reception of the signal. Since a coil typically encompasses the body (the head of the subject), it measures a sum of the signals from each tissue in the head.

More precisely, the sequence of events that occurs is:

- The magnetic field \mathbf{B}_0 is added with a gradient in the z direction. The selection of a particular frequency at the receiver part is then equivalent to the selection of a slice -a plane with a thickness of typically 1 to 10 mm- along the z direction. This procedure is called the *slice selection*.
- Then, within each slice or plane spanned by the resulting directions (x and y), two gradients are applied during the relaxation.
 - In the x direction, a negative gradient is applied after the RF pulse, and a positive one during acquisition, which creates a gradient echo during data acquisition, halfway through the second gradient pulse; The effect is that the precession frequency varies along the x axis, so that a Fourier transform of the signal gives its amplitude along the axis. This procedure is called *frequency encoding*.
 - In the remaining y direction, a gradient field is applied for a short interval between the RF pulse and data acquisition. After cancellation of this field, the precession is at the uniform frequency, but

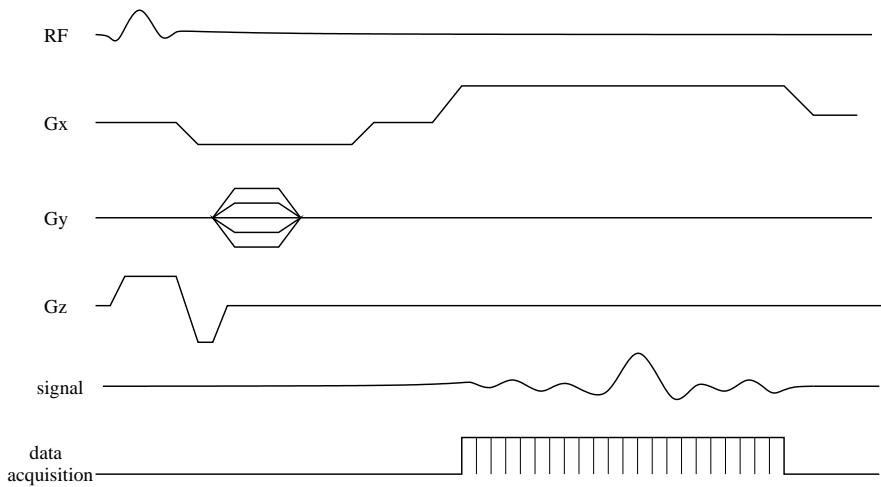


Figure 5.2: A basic imaging pulse sequence.

During the RF excitation pulse a gradient in z is applied (slice selection), and during read-out of the signal a gradient in x is applied (frequency encoding). Between these gradient pulses, a gradient pulse in y is applied, and the amplitude of this pulse is stepped through a different value each time the pulse is repeated (phase encoding). Typically 128 or 256 phase-encoding steps (repeats of the pulse sequence) are required to collect sufficient information to reconstruct an image. This figure is taken from [18].

with a phase shift determined by the position on y . Repeating the frequency encoding many times with different phase shifts creates information on the y position. This third procedure is known as *phase encoding*.

For a Repetition Time TR of about 1s, acquiring a single 256×256 slice (which needs 256 phase encoding steps so 256 RF pulses) would require 256s (4min and 16s). Acquiring a full volume like in such a way is impractical. Fortunately, there is a lot of “dead time” in a TR time slice. This “dead time” can be used to acquire several slices in parallel leading to much more reasonable whole head acquisition times. The amount of slices that can be acquired in parallel as well as the exact gradient patterns that have to be applied can vary. These parameters can be tuned to optimize various aspects (acquisition time, contrast, resolution, . . .). Designing pulse sequences for a given effect is a complicated task, which is achieved by specialists.

After acquisition of the data in the frequency space, also known as k -space, the data is mapped into the 3D space by Fourier transform.

5.2 Segmentation of Magnetic Resonance Images (MRI)

Among several alternatives for segmenting data [43, 40], we take as a reference the Brainvisa pipeline (among others, see [128, 98]). Anatomical data processing typically includes the following steps:

- An approximate pre-registration of the data to the Talairach template based on the anterior commissure/posterior commissure definition.
- A correction of the bias in the anatomical image, which is important to improve the detection of grey/white matter [96].
- The detection of the grey/white matter and cerebro-spinal fluid (csf) based on the analysis of the histogram of the data. This may or may not include spatial constraints.
- Segmentation of the grey matter under the constraint of making it have a spherical topology. This provides two interfaces, the grey/white matter interface and the grey matter/csf interface.
- Segmentation of the brain sulci, i.e. the inward folds of the cortical surface. The difficulty here is to decide when sulci end and start, and what the meaningful portions are

- Sulci recognition follows yet, using a basis of 20 manually annotated brain database, based on pattern recognition technique [128, 121].
- Projection of the sulci fundus onto the cortical surface. The difficulty here is to defined reliably the ends of the sulcal lines.
- Derivation of a surface-based coordinate system. This may take into account identified sulci [27] or simply the surface geometry [43]. In general a spherical coordinate is used; since the spherical coordinate system can certainly not be isometric, it has to decide which property has to be preserved. For instance, conformal mapping preserve the angles, possibly at the expense of large areal deformation [158].
- Segmentation of the cortical surface into gyri; this operation can be merely coordinate-based: the definition of cortical gyri in the reference sphere is simply pulled back onto the individual brain meshes using the previously defined coordinate system [26].

The whole procedure can be considered as performing a registration of cortical surfaces; as such it may be relatively efficient [44]. However, it is also rooted in deeper considerations on the structure and the evolution of brain folding patterns [151, 27].

Part II

Analysis of functional imaging data

Chapter 6

Functional Magnetic Resonance Imaging

This chapter is an introduction to functional Magnetic Resonance Imaging (fMRI): it presents a synthetic view of the principles, goals and techniques of intra-subject or first-level fMRI data analysis, with a short glimpse of anatomical MRI data processing. It does not aim at being exhaustive but proposes a specific point of view on this domain of research. In particular, we prepare the discussion for multi-subject or second-level fMRI data analysis, which will be the main focus of this course.

6.1 Background on fMRI data analysis

6.1.1 Principle of an activation study

The principle of an activation study with fMRI is simple. It consists in placing a subject or a patient in the scanner and acquiring data that reflect the haemodynamic activity of the brain. While the acquisition occurs, subjects are submitted to various stimuli or asked to perform certain tasks. These will influence the regional neural activity of the brain which in turn will modulate locally the haemodynamic (blood flow, blood volume and oxygenation) of some brain regions. The data acquired is a series of slices, each of which can be recorded typically in less than 80ms, with an in plane resolution and thickness of the order of the mm (1-3 mm). Therefore, the whole brain volume can be covered in a few seconds with current scanners but equipment is improving fast and these numbers will get down. Data acquisition is continuous for 5 to 20 minutes, and is called a *run*, an MR session usually consists of several runs for a subject, together with the acquisition of other kind of images reflecting the anatomy of the brain (e.g. T1 images with grey white matter contrast, diffusion images from which fibre bundles connecting

brain regions can be reconstructed). The data acquired for one subject is of the order of the Giga byte, and an actual experiment usually involves 15 to 40 subjects (for instance using different groups of subjects such as normal versus patients). During a run, an experimental paradigm designed to elicit a certain component of perception or cognition is proposed to the subject. It consists in several tasks or stimuli (the *conditions*) and fMRI data is commonly used to detect brain regions whose haemodynamic signal varies between conditions. Other data such as reaction time, eye movement, cardiac or breathing rhythms can be conjointly acquired. A comparison of two or more experimental conditions is called a *contrast*.

6.1.2 A short perspective on functional neuroimaging.

As previously described, fMRI data are often used for detecting brain regions whose haemodynamic signal varies across experimental conditions. Applications in human studies can generally be considered as addressing one of the two aspects: group versus single subject studies and normal versus patient studies. On the latter axis, the understanding of human brain functions is opposed to dysfunctions in psychiatric or neurological diseases. On the former axis, the specific information obtained from a particular subject is contrasted to the description of the information obtained at the population level. Using groups of normal subjects, cognitive neuroscientists that use fMRI to probe the brain functions are seriously challenged by philosophers or psychologists such as J. Fodor [50] who claim that localizing brain regions that respond to certain stimuli does not help to understand how the brain functions. Fodor uses the mechanistic analogy of a car engine, and asks how the knowledge of the localisation of pieces such as the piston or carburettor helps to understand the function of these pieces in the engine. It does not, unless one is interested in mending some parts, as the neurosurgeon might be for pathologies involving brain surgery. While the argument is potent, it does not account for the numerous occasions where the spatial organisation is a reflection of the functioning, such as the retinotopic organisation of the early visual cortex, and that brain region characterisation and localisation might be a necessary first step in the process of defining models of brain functioning.

6.1.3 What we need to know about the data

As described previously, the origin of the data is haemodynamic. Following an increase of synaptic and spiking activity, neurons require more energy conveyed through the blood in the form of oxygen and glucose [103]. Through a mechanism not fully known, the result is first an increase of oxygen extraction followed by an increase of blood flow that over compensates the loss of

oxy-haemoglobin. The Blood Oxygen Level Dependent (BOLD) contrast can be measured with MRI because the oxy-haemoglobin is diamagnetic while the deoxy-haemoglobin is paramagnetic, such that the MR signal increases with the ratio oxy over deoxy-haemoglobin [114]. The timing of this effect is relatively slow: it peaks around four to six seconds after a short stimulation and returns to baseline in about 25s. This behaviour is usually modeled by a haemodynamic filter h . It is interesting to note that even if the absolute timing may be considered as poor, the differential timing between two conditions can be precise to the order of 100 ms. The BOLD contrast does not show many anatomical details, and is sensitive to artifacts due to magnetic field variation around air/tissue interfaces inducing loss of signal or volume deformation in these areas.

6.2 fMRI data analysis: State of the art

The most common approach that has dominated the past decade can be decomposed in the following steps. Its success is linked to freely available tools such as SPM (see www.fil.ucl.ac.uk/SPM).

6.2.1 Data pre-processing: temporal and spatial realignments.

First, some distortions are specific to echo-planar images (EPI) commonly used in fMRI; in order to unbias the localization of brain regions, it is thus important to correct these distortions [78]. However, part of the signal cannot be recovered. Fortunately, the use of modern multi-channel antennas currently limits these distortions.

Second, subjects are never completely still in the scanner and movement needs to be corrected with a strong impact on the signal obtained. Movements correlated to the experimental paradigm are particularly difficult to correct if not impossible. Most current techniques assume a rigid body movement between two brain scans. Temporally, the slices of one brain volume are not acquired at the same time, and all voxel-based time series are usually interpolated to impose a unique timing for all voxel of one brain volume.

Spatial coregistration to a template is usually considered as a pre-processing, although it might also be done after the first level analysis. We defer the description of that processing step to section 6.4.1 since we consider it as crucial for group analyses.

6.2.2 The general linear model (GLM)

The BOLD response This step consists in modelling the BOLD signal and constructing statistical maps. As a first approximation the BOLD

haemodynamic response function (HRF) can be considered as a linear phenomenon with respect to stimulation. In practice, the following model [63] is used:

$$h(t) = \mathbb{I}_{t>0} \left[\left(\frac{t}{a_1 b_1} \right)^{a_1} \exp \left(b_1 - \frac{t}{a_1} \right) - c \left(\frac{t}{a_2 b_2} \right)^{a_2} \exp \left(b_2 - \frac{t}{a_2} \right) \right], \quad (6.1)$$

where $a_1 = 6$, $a_2 = 12$, $b_1 = 0.9s$, $b_2 = 0.9s$ and $c = 0.35$. However, the hypothesis that a certain filter h can be used to fit the temporal responses in a linear way, whatever the experimental paradigm, in any region of the brain and in any individual, is not fulfilled in general [1, 2, 36], and important work has been performed to better model these response [24, 156, 25, 94].

More fundamentally, there is a need to model the set of phenomena that occur during the BOLD response. To date, the most satisfactory generative model of the BOLD is the so-called balloon model [17], which has been further discussed and assessed in a series of works [60, 127, 32, 138, 76]. However, it is important to recall that this model is over-parameterized, hence can hardly be reliably tested based on the noisy and short times series available in fMRI datasets. To give an idea of the state of the art, we simply provide in Fig. 6.1 a description of the historical model of [17] completed by [60].

Let u be the neuronal activity at a certain cortical location, (f, v) be the blood flow and the blood venous volume at that location, q the vein deoxyhemoglobin content (all these variables are normalized to be 1 at rest), and y be the bold signal. The balloon model is then

$$\begin{aligned} \ddot{f} &= \varepsilon u - \kappa_s \dot{f} - \kappa_f (f - 1) \\ \dot{v} &= \frac{1}{\tau} (f - v^{1/\alpha}) \\ \dot{q} &= \frac{1}{\tau} \left(f \frac{1 - (1 - E_0)^{1/f}}{E_0} - v^{1/\alpha-1} q \right) \\ y &= V_0 (a_1 (1 - q) - a_2 (1 - v)) \end{aligned}$$

Here ε is the variable of interest, i.e. the neural efficacy. The constants $(\kappa_s, \kappa_f, \tau, \alpha, E_0, V_0)$ have to be estimated, since they may vary from locations to locations or from subjects to subjects. (a_1, a_2) are usually assumed to be known.

Figure 6.1: Formulation of the Balloon model according to [60].

To conclude this overview on BOLD modelling, the standard model based on the haemodynamic filter in Eq. 6.1 works astonishingly well in many

contexts, and has never been seriously challenged by concurrent approaches for group studies, which is our perspective here.

The design matrix. A linear model is constructed that includes all experimental factors which are believed to have an impact on the BOLD signal. Each regressor is constructed as the convolution of a standard HRF as in Eq. (6.1) with a time series representing the occurrence and duration of the experimental condition (stimuli or tasks). Given a set of time courses describing the occurrence of experimental condition $(p_1(t), \dots, p_C(t))_{t=1..N_t}$, where N_t is the number of scans, the fMRI design matrix is thus obtained by

$$X = \begin{bmatrix} (h * p_1)(1) & \dots & (h * p_C)(1) \\ \vdots & & \vdots \\ (h * p_1)(N_t) & \dots & (h * p_C)(N_t) \end{bmatrix} \quad (6.2)$$

Note that in practice, the design matrix has to be computed at higher temporal resolution and then down-sampled at the fMRI acquisition resolution. The intensity, or more generally parameters (loudness, speed, angle etc.) of the stimuli may vary in some experimental designs; in such cases, the time courses $p_c(t)$ take scalar values corresponding to the encoded quantities. Moreover, the temporal derivative of the C regressors $\left(\frac{d(h * p_c)(t)}{dt} = \frac{dh}{dt} * p_c\right)_{c=1..C}$ may be included in order to avoid that temporal misfit of the provided model inflates the residual variance.

X is further completed by adding a constant regressor, low frequency drifts models and possibly other known confounds (motion estimates, cardiac and respiratory measurements). It is thus an $N_t \times R$ matrix, N_t being the number of scans and $R > C$ being the number of model regressors. The remaining noise is modelled as a Gaussian autoregressive process. For simplicity, we assume that X is a full rank matrix, so that $\text{rank}(X) = R$. The generative model of the data $y(t)$ at a given site is:

$$y(t) = \sum_{r=1}^R X_{t,r} \beta_r + \epsilon(t), \epsilon \sim \mathcal{N}(0, \sigma^2 V) \quad (6.3)$$

where β is a set of parameters that relates the observed signal to the model, σ^2 the noise variance and V is the normalized noise covariance matrix; $\mathbb{E}(\epsilon^T \epsilon) = \sigma^2 V$.

Let Y be an fMRI dataset, written as a $(N_t \times N_v)$ matrix (scans \times voxel), and let X be the design matrix that describes effects of interest and confounds; the GLM proceeds by estimating the effect vectors β such that

$$Y = X\beta + \epsilon, \epsilon \sim \bigodot_{v=1}^{N_v} \mathcal{N}(0, \sigma^2(v)V(v)) \quad (6.4)$$

where ϵ represents the residual matrix. Importantly, the noise magnitude and covariance structure are voxel-dependent.

Estimation of the GLM parameters. Assuming that $V(v)$ is known in each voxel, the model parameters are estimated using the best linear unbiased estimate:

$$\begin{aligned}\hat{\beta}(v) &= (X^T V(v)^{-1} X)^{-1} X^T V(v)^{-1} Y(v) \\ \hat{\sigma}(v)^2 &= \frac{1}{N_t - R} (Y^T(v) V(v)^{-1} Y(v) - Y^T(v) V(v)^{-1} X (X^T V(v)^{-1} X)^{-1} X^T V(v)^{-1} Y(v))\end{aligned}\quad (6.5)$$

However, the voxel-based noise covariance matrix $V(v)$ is unknown. In general, the noise is assumed to be an AR(1) process [165, 167, 5]: assuming such a noise structure, the best linear unbiased estimator of β is obtained as follows: let $\epsilon = Y - X(X^T X)^{-1} X^T Y$ be the ordinary least squares residual, the order 1 auto-correlation ρ is estimated in each voxel, possibly with some regularization [165], yielding an estimated covariance matrix \hat{V} and the associated whitening operator $\hat{V}^{-\frac{1}{2}}$ is applied to both the data Y and the design matrix X , thus yielding their filtered counterpart \tilde{Y}, \tilde{X} . The effects, variance and degrees of freedom are then estimated using ordinary least squares on the whitened data

$$\hat{\beta}(v) = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \tilde{Y}(v) \quad (6.7)$$

$$\hat{\sigma}(v)^2 = \frac{1}{N_t - R} (\tilde{Y}(v)^T \tilde{Y}(v) - \tilde{Y}(v)^T \tilde{X} (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \tilde{Y}(v)) \quad (6.8)$$

$$\hat{\nu} = N_t - R \quad (6.9)$$

Contrasts and tests. Let c be the linear combination of the experimental conditions that is of particular interest; c is also called a functional contrast. The next part consists in making a statistical decision on the presence of an activation at any voxel v of the dataset (the dependence on v is implicit in the following equations). In terms of hypothesis testing, the alternatives are typically:

$$(H_0) \quad c^T \beta = 0 \quad (6.10)$$

$$(H_1) \quad c^T \beta > 0 \quad (6.11)$$

According to Neyman-Pearson lemma, the uniformly most powerful test to decide which hypothesis is true is the Likelihood ratio test: Let $\mathcal{L}(Y)$ be the likelihood of the data,

$$\Lambda = \frac{\sup_{c^T \beta = 0} \mathcal{L}(Y | \beta)}{\sup_{c^T \beta > 0} \mathcal{L}(Y | \beta)} \quad (6.12)$$

Moreover, assuming that the noise covariance is known, Λ is a monotonous function of the exhaustive statistic t (Student, for a signed test) or F (Fisher, which is unsigned and can be defined for multidimensional contrasts); see Appendix C.1 for a sketch of the proof

$$t(v) = \frac{c^T \hat{\beta}(v)}{\hat{\sigma}(v) \sqrt{c^T (X^T V X)^{-1} c}} \quad (6.13)$$

$$F(v) = \frac{\text{Tr}(c^T \hat{\beta}(v) \hat{\beta}(v)^T c)}{\hat{\sigma}(v)^2 \text{Tr}(c^T (X^T V X)^{-1} c)} = \frac{Y(v)^T (P_X - P_R) Y(v)}{Y(v)^T (I - P_X) Y(v)} \frac{N_t - R}{R - \text{rank}(c)} \quad (6.14)$$

where $P_X = X(X^T X)^{-1} X^T$, and $P_R = Xc(c^T X^T Xc)^{-1} c^T X^T$. Note that we write t and F as functions.

Note that in general, the equivalence between the F test and the likelihood ratio test does not hold when the noise covariance is unknown [31]. However, the t or F tests are preferred in general in the neuroimaging community due to their simplicity, and because their significance can be readily obtained by comparing their values to Student/Fisher distributions, while the likelihood ratio test cannot be assessed without (unfulfilled) asymptotic assumptions. In the remainder of this document, we will refer to the first level decision statistic as ϕ .

Although we did not write it explicitly, the model can easily deal with multiple session in this formalism: the data are concatenated along the temporal dimension, while the design matrix has to be written as a Kronecker product of the mono-session design matrices; the only minor difference is that the residual parameters have to be estimated for each session.

6.2.3 Thresholding

Once a statistical map is constructed at the individual or group level, it remains to be decided which brain regions are worth reporting, hence to derive a threshold defining brain regions with significant activity. The principle on which this threshold is chosen varies across studies. Three types of statistical threshold are used. Uncorrected thresholds ϕ_c are defined to control for the risk of false positive at the voxel level.

$$p(\phi > \phi_c | H_0) < \alpha \quad (6.15)$$

If only one region or voxel is tested, this is a valid approach. Note that when the GLM is written in a Bayesian framework as in appendix C.2, the test is performed on posterior probability maps

$$p(c^T \beta > \delta | Y) > 1 - \alpha, \quad (6.16)$$

where δ is some reference value (0 typically). When uninformative priors are used, and when $\delta = 0$ the tests corresponding to 6.15 and 6.16 are equivalent. We do not further consider Bayesian tests in this section.

Often however, a priori localisation is not known, and because a brain volume contains several tens of thousands of voxels, a large number of regions are likely to be reported by chance with this kind of threshold. To prevent false positive results, several techniques have been derived for controlling the risk of error family wise (across voxels or regions). The most straightforward way consists in correcting the threshold ϕ_c for the number of tests performed -i.e. the number N_v of voxels- assuming independence. This is known as the Bonferroni correction. ϕ_c has to adjusted so that

$$p(\phi > \phi_c | H_0) < 1 - (1 - \alpha)^{1/N_v} \simeq \frac{\alpha}{N_v}. \quad (6.17)$$

This approach is correct in general cases [166], but, since *i*) the independence hypothesis is trivially violated and *ii*) the spatial structure of the data is not accounted for, this approach is not very sensitive.

More sophisticated approaches have based this thresholding on the notion of the expected number of above threshold regions approximated by the Euler characteristic of random fields. These approximations have been extended to a number of statistical fields (t, F, χ^2) and geometry (any volume, surface, 4D data) [167]. For instance, the expected Euler characteristic χ_c of a Gaussian Random Field ϕ thresholded of dimension D (2 or 3) in a volume V , at level ϕ_c is

$$\mathbb{E}(\chi_c) = \Omega |\Lambda|^{\frac{1}{2}} (2\pi)^{-\frac{D+1}{2}} H_D(\phi_c) e^{-\frac{\phi_c^2}{2}} \quad (6.18)$$

where Λ is the covariance matrix of the field (basically, its smoothness or inverse point spread function) and H_D the Hermite polynomial of degree D . Now, for a high value of z , one has

$$P(\phi_{max} \geq \phi_c) \approx P(\chi_c > 1) \approx \mathbb{E}(\chi_c) \quad (6.19)$$

Equations (6.18) and (6.19) together give a new way of setting a threshold on a smooth Gaussian random field. A concept of interest is the -dimensionless- notion of the number of resolution elements (*RESELS*): This number is given by

$$RESELS = \frac{\Omega}{\prod_{i=1}^D FWHM_i}, \quad (6.20)$$

where $FWHM_i$ represents the full width at half maximum of the spatial filter associated with the map, in direction i , and related to the smoothness of the field by

$$\Omega |\Lambda|^{\frac{1}{2}} = RESELS (4 \log(2))^{\frac{D}{2}} \quad (6.21)$$

This method has been further improved with the introduction of the spatial extent of the activated areas [57], joint test on the height and size of supra-threshold clusters [124], and the robust estimation of smoothness in presence of activations [85]. Note that this approach is compatible with spatial smoothing of the data (more precisely, this model encourages smoothing, to fulfill the hypotheses of the model [71]).

These tests are computationally efficient but also rely on several assumptions, including a heavy smoothing of the volume to be thresholded which contradicts the MRI physicist efforts to improve image resolution. A few approaches have been proposed to use permutation-based non-parametric framework, but they face several difficulties due to the non-whiteness of the residuals of the GLM [53].

The third approach uses the false detection rate (FDR) threshold that controls for the false positive rate as a function of the number of reported voxels. The threshold ϕ_c corresponds to a FDR value of α is

$$p(H_0|\phi > \phi_c) = \frac{p(\phi > \phi_c|H_0)p(H_0)}{p(\phi > \phi_c)} < \alpha \quad (6.22)$$

Following [62], the estimation of the best threshold given α is usually carried out based on the Benjamini-Hochberg procedure; some more robust alternatives have been proposed recently [134]. Due to its definition, the FDR yields typically a less severe threshold than the family-wise error control procedures (6.17-6.18).

In general, thresholding is a trade-off between sensitivity and specificity, which may depend on the application purpose although this is rarely acknowledged in the literature. The defined regions are then reported using their local maxima in the standard space.

Note that cluster-level thresholding can be more sensitive, given that the truly active voxels are generally spatially clustered. Among the previously discussed techniques, only the Gaussian Random Field Model can be used in this perspective. Several alternatives have been proposed, such as cluster mass thresholding [137], but the calibration of this statistic in the intra-subject framework is still problematic.

6.2.4 Detecting modules versus establishing functional connections.

While the techniques described previously aim at localizing the activity in the brain, and therefore defining spatially defined functional modules, an increasingly large part of the literature is now devoted to establishing the functional connections between brain regions. The original observation in [13] shows that even during no motor activity (resting state or other such

as visual stimulation), BOLD signal of a series or region that respond to motor tasks are correlated. Since then, two main approaches are concurrently explored. The first one tries to extract networks of correlated activity (or sharing some information) and techniques such as principal component analysis, independent component analysis (ICA) [101], probabilistic ICA [10], partial least square (PLS)[99, 84], various clustering techniques, self organizing maps [118]. Escaping from the linear decomposition framework, kernel- and manifold learning-based techniques have also been proposed and can certainly be more sensitive [144, 142]. Such methods aim at defining the various functional networks that underlie brain activity and their relation to external tasks or stimulation. These exploratory approaches often suffer from a lack of interpretability.

The alternative approach consists in defining a specific network, a graphical model, choosing a priori the nodes and the structure of the graph, and to estimate the functional links given the experimental paradigm. This led to the development of structural equation models [100] and dynamical causal models [59, 119, 23]. By definition, the results of such procedures depend heavily on the regions introduced as nodes of the graphical model, and introducing many regions may yield ambiguous results [16].

Furthermore, the steady states used in many functional connectivity studies are not always well-defined states, and hence it would be of interest to extend the notion of functional connectivity to states that are controlled to a larger extent by the experimenter and that follow a predefined dynamic. Finally, signal similarity does not only come about by functional interaction, but can be influenced by confounding physiological effects of no interest like heart beat or artifacts like subject motion [34].

6.3 Spatial models for fMRI data

Several alternatives spatial models have been considered for modelling the spatial correlations fMRI data modelling.

- The first one consists in using the theory of Gaussian Random Field (GRF) to obtain estimates of the significance of activity peaks in fMRI maps
- The second one consists in using a Markov Random Field (MRF) model to model the spatial smoothness of the *true* activity maps and/or of the background noise and detect the active regions more sensitively.
- The third one consists in using spatial wavelets to decompose the fMRI signal.

Given our own interest on spatial models, we give a quick overview of these approaches. It should be reminded that all these developments have been performed in the framework of intra-subject analyses.

6.3.1 Gaussian Random Fields

Modelling the noise as a GRF: In many softwares used for activation detection, the thresholding procedure is based on a Gaussian Random Field (GRF) model, where the reference GRF has the smoothness of the data and is assumed to be centered. As detailed in section 6.2.3, the Euler characteristic of the excursion sets of such GRFs is estimated analytically, and this readily provides a significance of activation values for high threshold. The problem is that these thresholds are correct only for very smooth data, and for high threshold values.

An interesting alternative has been proposed in [166], where the probability p_{dlm} of detecting local maxima above some threshold ϕ_c is controlled, instead of the probability of detecting any voxel above such a threshold:

$$p_{dlm}(\phi_c) = \sum_{v=1}^{N_v} p((\phi(v) > \phi_c) \text{ and } \phi(v) > \phi(w), \forall w \sim v), \quad (6.23)$$

where $v \sim w$ denotes neighboring relationship in the spatial domain. This model further requires that the spatial covariance function between two sites x_1 and x_2 is of the form $\exp(-(x_1 - x_2)^T \Sigma^{-1} (x_1 - x_2))$.

The derivation has been performed for Gaussian fields, hence it requires a conversion from standard Student or Fisher fields to Gaussian fields. Let \mathcal{N} be the standard Gaussian density, and $\mathcal{M}(z) = \int_z^\infty \mathcal{N}(u)du$; let ρ be the spatial autocorrelation, $\alpha = \arcsin\left(\sqrt{\frac{1-\rho^2}{2}}\right)$, $h = \sqrt{\frac{1-\rho}{1+\rho}}$ and

$$q(\rho, z) = \frac{1}{\pi} \int_0^\alpha \exp\left(-\frac{h^2 z^2}{2(\sin \theta)^2}\right) d\theta$$

Then, if we define $\mathcal{Q}(\rho, z) = 1 - 2\mathcal{M}(hz^+) + q(\rho, z)$, where $z^+ = z.(z > 0)$, p_{dlm} is defined as:

$$p_{dlm}(\phi_c) = \sum_v \int_{\phi_c}^\infty \left(\prod_{d=1}^D \mathcal{Q}(\rho_d, z) \right) \mathcal{N}(z) dz, \quad (6.24)$$

where the central product is over the volume dimensions. This procedure has been shown in [166] to provide more sensitive tests than the standard Gaussian random field theory to assess the significance of local maxima.

Modelling the data as a GRF: Another perspective has been proposed in [70], in which the noise covariance process is assumed to be (spatially) white, while the model parameters β are given a smooth spatial prior

$$p(\beta) = \mathcal{N}(0, \lambda \exp(-L(\beta)\tau)), \quad (6.25)$$

where $(\lambda > 0, \tau > 0)$ are the scalar parameters of the model (to be estimated), and L is some operator that represents the spatial structure of the image, and possibly depends on β itself (to model the image anisotropy). The estimation of the parameters follows through an EM procedure; the posterior of the parameters $p(\beta|Y)$ can finally be derived: the marginals at each voxel yield the probability that the given voxel is activated. With respect to traditional univariate estimates, this Gaussian process approach allows some (possibly anisotropic) spatial regularization; in [70], this is at the expense of the unrealistic assumption of spatially white noise. The merit of this approach has not been established yet in practical situations.

6.3.2 Markov Random Fields

Markov Random Fields (MRFs) have been introduced in fMRI thresholding procedures to introduce the prior knowledge that realistic brain activations are spatially extended, hence smooth fields, while part of the noise (measurement error) is spatially white; under such hypotheses the compromise between specificity and sensitivity can be improved by imposing a Markov Random Field prior on the voxel labels: let Z be the label map, i.e. the set of values $(Z(v))_{v=1..N_v}$; it is associated with the prior $p(Z)$; let ζ be a particular configuration for Z :

$$p(Z = \zeta) \propto f(\xi) \exp \left(-\frac{\xi}{4} \sum_{v=1}^{N_v} \sum_{w \sim v} \mathbb{I}(\zeta(v) \neq \zeta_w) \right), \quad (6.26)$$

where $f(\xi)$ is the normalization constant of the MRF, which cannot be computed explicitly, and the relation $w \sim v$ denotes the fact that v and w are neighbours. It is then standard to assume conditional independence of the likelihood:

$$p(Y|Z, \theta) = \prod_{v=1}^{N_v} p(Y(v)|Z(v), \theta), \quad (6.27)$$

where θ is a set of parameters of the likelihood. The MRF definition requires a parameter ξ that controls the smoothness of the MRF, i.e. the spatial regularization of the labels implied by the model. The posterior distribution of the unknown parameters given the observed data Y is then

$$p(Z = \zeta, \theta, \xi | Y) \propto p(Z = \zeta | \xi) p(\xi) p(\theta) \prod_{v=1}^{N_v} p(Y(v) | Z(v), \theta) \quad (6.28)$$

The likelihood model is generally fairly simple, each components being modeled by a Gaussian. Note that this model corresponds to a kind of spatially constrained mixture model. The estimation of θ is thus no more difficult than estimating the parameters of a Gaussian Mixture Model (GMM). A two-classes or three classes model is usually chosen, in order to model the three competing hypotheses:

$$\begin{aligned} H_0 &: c^T \beta = 0, \\ H_1 &: c^T \beta > 0, \\ H_{-1} &: c^T \beta < 0, \end{aligned}$$

where the relation of the hypotheses with the observed data conforms to the model detailed in Section 6.2.2. Assuming that ξ is known, inference of θ and Z can be carried out either through MCMC sampling or Variational Bayes inference (which is only approximate, but in general relatively accurate [162, 116]), or by using the efficient minimum cut approach [80]; the latter provides the optimal solution for a two-classes problem (only), but is still much more expensive than Variational Bayes [116].

The main problem with this model is that ξ is unknown a priori. One solution consists in estimating the model with different values of ξ , and to choose one a posteriori, but it is even not clear how to choose the best solution (moreover, this requires a partition of the data into training, validation and test set). A second possibility is to use a full Bayesian approach, where all the quantities are obtained through MCMC sampling, but this solution is particularly inefficient [163]. A third solution consists in replacing the discrete label model with a continuous model, which is reminiscent of the Gaussian Random Field approach described in Section 6.3.1; inference on this continuous model can be much more efficient than on the discrete model. An elegant solution based on Variational Bayes approximation can be found in [162].

An extension of the basic fMRI data-based MRF model has been proposed to account for anatomical information, such as the probability of one voxel being in grey matter, white matter or cerebro-spinal fluid. The generative model consists in introducing a second observation w , namely the anatomical voxel type in the model. The set of states Z becomes richer with a number of (activation status \times tissue type) combinations (9 in the present case). Eq (6.28) becomes

$$p(Z = \zeta, \theta, \xi | Y, W) \propto p(Z = \zeta | \xi) p(\xi) p(\theta) \prod_{i=1}^{N_v} p(Y(v) | Z(v), \theta) p(W(v) | Z(v)) \quad (6.29)$$

This model has been shown in [116] to provide sensible constraints in the detection of active regions.

In spite of this, Markov Random Fields have received little interest in the neuroimaging community. The shortcoming of this model is that the topography of the activation maps is relatively poorly modeled with the generative model $p(Y|Z, \theta)$; while this model is optimal for segmenting *flat* activation patterns corrupted by white noise, they do not carry striking improvements in real cases [162, 116]. Finally, the evaluation of these approaches in the context of group analyses still has to be done.

6.3.3 Spatial wavelets

Finally, another generic alternative to model the spatial distribution fMRI activity consists in using spatial wavelet bases. The approach is simply to decompose the signal onto the basis, and then to use a wavelet-based shrinkage to filter the activation and noise parameters i.e. the wavelets coefficients of the parameters maps associated with low ϕ values are discarded [154]. The reconstruction of the data putatively provides denoised activation estimates. However, this model is not generative, and should be regarded as a post-processing trick, which is not free of artifacts. More complex strategies may be necessary may then be necessary to reduce the bias in the resulting maps [155].

A more sophisticated framework has been presented in [49], in which the parameters are given a prior

$$\beta = \gamma W + E, \quad (6.30)$$

where W represents a set of spatial basis function (typically wavelet-based). At each resolution level, the wavelet coefficient are then modeled as a mixture of two Gaussians to separate the noise coefficients from the signal coefficients. The parameters and hyperparameters of the model are then estimated through a Variational Bayes approach, thus providing posterior estimates for $p(\beta|Y)$. The resulting posterior probability maps for the contrasts $p(c^T \beta|Y)$ can then be used for assessing the presence of effects. The whole procedure is finally shown in [49] to better preserve the shape of activations than Gaussian Markov Fields or smoothing.

Still, such a procedure is blind to the underlying anatomy, and the ultimate benefit in the case of group studies has not been proved.

6.4 Dealing with the absence of a brain template

In this section, we describe a problem that underpins neuroimaging group analysis: namely that statistical inference methods require that the brain locations or structures have to be matched across subjects. We first describe the standard framework, which is based on a so-called spatial normalization procedure, then discuss its shortcomings.

6.4.1 Spatial normalization

There is a consensus in the neuroimaging community on the way to perform multi-subject analyses: data are first coregistered to a reference template and re-sampled (stereotactic normalization, see e.g. [58]). Activation is detected in a mass univariate framework using a General Linear Model at the individual level, then at the group level through the use of mixed-effects models.

While there are many different techniques to perform the spatial coregistration, the most usual procedure is to first realign the functional volumes to the anatomical volume acquired in the same scanning session, and use this more detailed image to derive a deformation field that warps the subject brain anatomy to a standard template (generally the so called ICBM152 volume image which represents the average of 152 healthy T1 brain images). This template corresponds (but only approximately) to a brain neurosurgical atlas, the Talairach and Tournoux atlas [140].

The coregistration of the functional data onto the anatomy is relatively simple, because an affine deformation model is usually sufficient (if we ignore the local effect of residual EPI distortions, which is hard to correct for anyway [78]). The standard matching criterion is normalized mutual information, given that there no simple statistical relationship between the T1 and the T2 contrast.

The coregistration of anatomical image to the template is more involved [4]. The matching criterion is usually the sum of squared differences after histogram equalization of the smoothed input and target images. More recent approaches work on pre-processed anatomical images, where the tissue type probability maps have been extracted [6].

An affine coregistration is first performed to roughly align the images. Then, in [4] a set of deformation parameters are computed in a basis of smooth spatial deformations, typically the low frequency part of a discrete cosine transform of the ambient (3D) space. The number of degrees of freedom in the deformation can be controlled through the number of basis deformation used. The bending energy is also included in the criterion to minimize the risk of unrealistic deformation patterns. There is no guarantee that the deformation is invertible.

More recent approaches use diffeomorphic registration algorithms. Although he conceptual framework goes back to [22], the application to neuroimaging has become important only since [6]. In this framework, the deformation is defined within some pre-defined spatial basis, and the smoothness is also enforced by placing shrinking priors on the deformation coefficients. The diffeomorphic framework with constant vector field can then be handled as traditional small deformation problems using the scaling and squaring procedure. The efficient implementation described in [6] makes it usable in

large scale datasets. With respect to previous alternatives, this algorithm improves the coregistration of many brain structures (see [86], for a comparison of many coregistration approaches). Finally, it should be pointed that brain registration can be surface-based instead of merely anatomy-based. We discuss this in Section 6.5.

6.4.2 Shortcomings of this procedure

The SPM procedure within the stereotactic normalization framework is challenged by:

- The residual inter-subject variability that is known to remain after normalization. Indeed, the algorithms that warp a subject anatomy to a template do not and cannot perform a perfect match. The information used for the warping is the main anatomical contrasts (deep sulci, ventricles) but the variation of the anatomy between subjects is such that there is no obvious point to point correspondences between subjects. Based on anatomical [73] or functional landmarks [139] this variability has been evaluated as about 1cm in many cortical regions. This residual mismatch should not only be seen as a shortcoming of normalization procedures that can be improved, but rather as an effect of the intrinsic variability of cortical shape, folding patterns [128], cytoarchitecture [44] and functional organization [159].
- Indeed, even if the anatomy could be matched perfectly, this would not necessarily be the case for functional regions, given that the function/structure matching implicit in brain mapping is only approximately true. For instance, the Fusiform Face Area (responding more to face than to objects) may be localised more anteriorly in the fusiform gyrus in one subject compared to another.
- The difficulty of assessing the functional specialization of brain regions, given that regions of interest are not entirely correctly characterized by their coordinates in the template space or with respect to sulco-gyral anatomy, but also by their relative position with respect to anatomical landmarks, their extent or their connections to other regions [15].

Several approaches are currently used to limit the impact of inter-subject variability on functional MRI (fMRI) group analyses. First, spatial smoothing is frequently applied to the data, thus sacrificing resolution for sensitivity in voxel- and cluster-level detection. Another approach consists in defining functional Regions of Interests (ROIs) using a *localizer* experiment, and then to restrict the group analysis to these particular regions [132]. This approach is particularly well suited to study some fine-grained functionally specialized

regions [132]. While it is often more sensitive, the investigation is restricted to the functional ROI defined and is also questionable from the perspective of the reliability of region definition [56]. Anatomical atlases can be used to define ROIs in the common space (see e.g. [152, 39, 135], but this approach precludes any analysis of inter-subject anatomical variability.

6.5 A quick overview of anatomical data processing

In this section, we give a quick overview on how to process anatomical data in the perspective of functional data analysis and of assessing anatomo-functional structure correspondences. It is important to notice that this approach complements the iconic normalization described in the previous section, in the sense that it aims at recognizing different brain structures instead of standardizing brain shapes. This approach specifically aims at preserving the idiosyncratic structure of each individual dataset. We focus on the cortical parts, given that subcortical structures are usually handled separately, based on dedicated methods (see e.g. [126]). Among several alternatives [43, 40], we take as a reference the Brainvisa pipeline (among others, see [128, 98]). Anatomical data typically includes the following steps:

- An approximate pre-registration of the data to the Talairach template based on the anterior commissure/posterior commissure definition.
- A correction of the bias in the anatomical image, which is important to improve the detection of grey/white matter [96].
- The detection of the grey/white matter and cerebro-spinal fluid (csf) based on the analysis of the histogram of the data. This may or may not include spatial constraints.
- Segmentation of the grey matter under the constraint of making it have a spherical topology. This provides two interfaces, the grey/white matter interface and the grey matter/csf interface.
- Segmentation of the brain sulci, i.e. the inward folds of the cortical surface. The difficulty here is to decide when sulci end and start, and what the meaningful portions are
- Sulci recognition follows yet, using a basis of 20 manually annotated brain database, based on pattern recognition technique [128, 121].
- Projection of the sulci fundus onto the cortical surface. The difficulty here is to defined reliably the ends of the sulcal lines.

- Derivation of a surface-based coordinate system. This may take into account identified sulci [27] or simply the surface geometry [43]. In general a spherical coordinate is used; since the spherical coordinate system can certainly not be isometric, it has to decide which property has to be preserved. For instance, conformal mapping preserve the angles, possibly at the expense of large areal deformation [158].
- Segmentation of the cortical surface into gyri; this operation can be merely coordinate-based: the definition of cortical gyri in the reference sphere is simply pulled back onto the individual brain meshes using the previously defined coordinate system [26].

The whole procedure can be considered as performing a registration of cortical surfaces; as such it may be relatively efficient [44]. However, it is also rooted in deeper considerations on the structure and the evolution of brain folding patterns [151, 27]. Alternative procedures can also take place to combine 2D and 3D information in normalization, but they are not a standard yet.

Recently, it has been proposed to directly align cortical meshes by using an iterative search for optimal vertex correspondences based on feature fields (that basically represents the distance of the nodes to the outer envelope of the cortex) and a regularization to preserve local topology [92]. The procedure is applied in a multi-scale fashion with alternate matching/template estimation steps. The meshes are then re-parametrized, and cortical averages can be constructed, which preserve some salient features of cortical geometry (although some regions are more problematic). It is shown in [92] that about 50 subjects are necessary to yield a bias free template. Still folding variants exist and it is suggested to have several population averages to represent all the folding variants observed in the different subgroups of subjects. The problem of unbiased anatomy representation is and will probably remain open for several years.

Chapter 7

Statistical inference on multi-subject fMRI data

In this chapter, we describe and evaluate the methodology group analyses for fMRI data. This is organized in three contributions:

- In section 7.1, we introduce the main univariate models used for group analysis, and introduce the main definitions that are necessary for the discussion. We give an overview of some contributions published in [108, 129, 82].
- In section 7.2, we detail the question of the reproducibility of group-level brain activation maps, which has been published in [145].
- We end up with an original work on multi-subject parcellation for group inference; this is an updated version fo the work published in [143].

Here we concentrate on the problem of one-sample inference, in which it has to be decided where in a certain common brain space there is a positive effect across subjects. It is assumed that a contrast c of experimental conditions has been defined in a coherent manner across subjects; for simplicity, this contrast-related effect $c^T \beta$ will be denoted β .

7.1 Classical statistical models

Classical statistical models refer to mass-univariate models, in which the functional images are assumed to be coregistered. According to the first-level analysis described in Sec. 6.2.2, the data consists in vectors $\hat{\beta}_s = (\hat{\beta}_s(v))_{v \in \{1..N_v\}}$ where N_v denotes the number of spatial sites (voxels), and $s \in \{1, \dots, S\}$ denotes the subject. These vectors represent the estimated

effect of a certain contrast. Moreover, the variance of the estimator of β_s is also estimated, and provided as a corresponding set of scalars $\hat{\sigma}_s^2$ ¹. This provides variance values for each site and subject. The degrees of freedom $(\nu_s)_{s \in 1, \dots, S}$ are assumed to be known, and constant across voxels, but not necessarily across subjects.

7.1.1 Fixed effects and mixed effects models

The simplest problem consists in deciding, for a given position v , whether there is a positive effect at the population level, given the observations $(\hat{\beta}_s(v))_{s \in \{1, \dots, S\}}$ and the associated variances. This simple question can receive multiple answers, depending on whether the null hypothesis is that no subject activated (fixed-effect model, FFX) or that the average in the population is not different from zero (mixed-effect model, MFX). Usually, only the second question is considered, since the result of the fixed effect inference is valid only on the cohort at hand, and not on the whole population from which the cohort has been sampled [54]. Note that other question could be asked, such as whether at least half (or any other proportion) of the subjects showed a positive effect (see [72]).

Within the mixed-effects model, it is often possible to neglect the first-level variance $\hat{\sigma}_s^2(v)$, given that this variance source is in general largely dominated by the second level sample variance of $(\hat{\beta}_s(v))_{s \in 1, \dots, S}$. In that case, the model is traditionally referred to as *random effects* (RFX), under the convention that the subject is the random factor in the model. However, it may be important to take the first-level variance into account [129]. In fact, the so-called RFX model can be seen as an approximation of the correct MFX model.

The MFX model can be written formally as (here and in the sequel the dependence on v will be omitted, although all the quantities are in fact voxel-dependent):

$$\hat{\beta}_s = \beta_s + \varepsilon_s, \varepsilon_s \sim \mathcal{N}(0, \sigma_s^2) \quad (7.1)$$

$$\beta_s = \bar{\beta} + \eta_s, \quad (7.2)$$

where $\bar{\beta}$ denotes the parameter of interest, i.e. the group average response. The first equation is simply a restatement of the intra-subject General Linear Model, and the second one is the group model.

¹One should take care that this is not the same term as in Eq. 6.6, where σ represents the variance of the residuals of the model. Here this term represents in fact the contrast related variance $\sigma_c^2 = c^T(X^T V X)^{-1} c$. But, Given that c is implicit in this chapter, no confusion is possible, and we retain the notation σ .

7.1.2 Classical hypothesis testing

In order to perform hypothesis testing, the usual approach consists in assuming a certain distribution for the effects $(\beta_s)_{s \in 1, \dots, S}$, which will be parametrized by $\bar{\beta}$; this distribution will thus be denoted $p(\beta|\bar{\beta})$. This distribution is usually chosen as normal (we discuss the validity of this hypothesis in section 7.2)

The following likelihood ratio (LR) test can thus be defined:

$$\Lambda = \frac{\sup_{\bar{\beta}=0} \mathcal{L}(\beta|\bar{\beta})}{\sup_{\bar{\beta} \in \mathbb{R}} \mathcal{L}(\beta|\bar{\beta})}, \quad (7.3)$$

assuming that the empirical means is positive (otherwise, the test is obviously not significant). In that case, we denote the set of possible values for $\bar{\beta}$ as Θ (unconstrained case, here \mathbb{R}), and Θ_0 (constrained case, here $\{0\}$). The likelihood ratio test rejects the null hypothesis H_0 when its inverse exceeds a critical value λ_α . That is, the decision rule has the form:

- if $\Lambda^{-1} \geq \lambda_\alpha$ reject H_0 .
- if $\Lambda^{-1} < \lambda_\alpha$ accept (or do not reject) H_0 .

The critical value λ_α is usually chosen to obtain a specified significance level α , through the relation: $P_0(\Lambda^{-1} \geq \lambda_\alpha) = \alpha$. The Neyman-Pearson lemma states that this likelihood ratio test is the most powerful among all level- α tests for this problem. Under the normal hypothesis, and assuming that $\forall s \in \{1, \dots, S\}, \sigma_s^2 = 0$, the likelihood test simply boils down to the standard t-test,

$$t = \frac{\text{mean}_s(\hat{\beta}(s))\sqrt{S}}{\sqrt{\text{var}_s \hat{\beta}(s)}} \quad (7.4)$$

as shown in Section C.1.

In more general cases, however, the exact distribution of the likelihood ratio corresponding to specific hypotheses is very difficult to determine. A convenient result, though, says that as the sample size n approaches ∞ , the test statistic $-2 \log(\Lambda)$ will be asymptotically χ^2 distributed with degrees of freedom equal to the difference in dimensionality of Θ and Θ_0 , which is 1 here. This readily provides a statistical procedure, but this procedure is only asymptotically correct. This is an issue here, given that the number S of observations is low (typically 10 to 20). Non-parametric procedures can be used as an alternative (see below).

Before analyzing the Gaussian MFX model in more detail, it should be noted that different alternatives can be used as statistics: for instance, the sign statistic, the Wilcoxon signed rank statistic or the empirical likelihood

ratio. Mixed-effects model extensions (i.e. models that incorporate the information of first-level variance) have been introduced and discussed in [129]. Let us just recall the definition of Wilcoxon's signed rank statistic (WKX) [74], which sorts the absolute effects in ascending order, then sums up the ranks modulated by the corresponding effect's sign:

$$W = \sum_{s=1}^S \text{sign}(\hat{\beta}_s) \text{rank}(\hat{\beta}_s) \quad (7.5)$$

Interestingly, the calibration of this statistic is straightforward and voxel-independent; it depends only on S .

7.1.3 The Gaussian mixed effect model

Given its popularity in the literature [167, 51, 164, 129, 145], it is important to emphasize the detail of the Gaussian mixed effects model. This model writes:

$$\hat{\beta}_s = \beta_s + \varepsilon_s, \varepsilon_s \sim \mathcal{N}(0, \sigma_s^2) \quad (7.6)$$

$$\beta_s = \bar{\beta} + \eta_s, \eta_s \sim \mathcal{N}(0, v_g) \quad (7.7)$$

Let us recall that $(\hat{\beta}_s)$ and σ_s^2 are assumed to be known. This relies on the approximation $\sigma_s^2 = \hat{\sigma}_s^2$, which is usually accepted, assuming that the first-level model has been correctly specified, and given the number of degrees of freedom ν_s is large enough (typically ≥ 100) to provide an accurate estimate of σ_s^2 . Given that both v_g and $\bar{\beta}$ are unknown, there is no closed-form formula for estimating the model. An Expectation Maximization algorithm has been proposed in [129], restricted maximum likelihood has been used in [167], Markov Chain Monte Carlo methods in [164]. In fact, in this simple model, the maximum likelihood solution can be reached very efficiently (see Fig. 7.1.3)

Finally, the following decision statistic can be formed:

$$\phi = \sum_{s=1}^S \frac{\hat{\beta}_s}{\hat{\sigma}_s^2 + v_g} \left(\sum_{s=1}^S \frac{1}{\hat{\sigma}_s^2 + v_g} \right)^{-\frac{1}{2}} \quad (7.11)$$

This decision statistic can be simply interpreted as a Bayesian test where the average size of the group effect is compared to the uncertainty of this effect. Note that it is not in general equivalent to the LR test - which remains a plausible alternative [129].

The next problem is to assess the significance of this statistic (LR or (7.11)): although μ is formally constructed as a Student statistic, it cannot

The joint estimation of the group effect and the group variance proceeds from Eqs. (7.6-7.7). At a voxel v , S values of estimated effects $\hat{\beta}$ are available, together with S estimates of the associated variances $\hat{\sigma}^2$ (we drop the voxel index v for simplicity). We also assume that the estimated variance is correct, so that $\sigma^2 = \hat{\sigma}^2$ (note that the estimator generally relies on $\nu > 100$ degrees of freedom).

For this model, the log-likelihood of the data is written as:

$$\mathcal{L}((\hat{\beta})|\bar{\beta}, v_g) = cst - \frac{1}{2} \left(\sum_{s=1}^S \log(\sigma_s^2 + v_g) + \sum_{s=1}^S \frac{(\bar{\beta} - \hat{\beta}_s)^2}{\sigma_s^2 + v_g} \right) \quad (7.8)$$

maximizing \mathcal{L} with respect to $\bar{\beta}$ while keeping v_g fixed yields:

$$\bar{\beta} = \sum_{s=1}^S \frac{\hat{\beta}_s}{\sigma_s^2 + v_g} \left(\sum_{s=1}^S \frac{1}{\sigma_s^2 + v_g} \right)^{-1} \quad (7.9)$$

while the minimization of \mathcal{L} with respect to v_g , while $\bar{\beta}$ is fixed yields

$$\sum_{s=1}^S \frac{(\bar{\beta} - \hat{\beta}_s)^2}{(\sigma_s^2 + v_g)^2} = \sum_{s=1}^S \frac{1}{\sigma_s^2 + v_g} \quad (7.10)$$

Let $L(v_g)$ and $R(v_g)$ be the left and right hand side terms in Eq. (7.10). We solve it by iterating the solution of $L(v_g) = R$ in under the constraint $v_g > 0$ using Newton's method, then updating the right hand side term. This procedure always converges in a few iterations.

be identified as such because of the heterogeneous variance terms. One conservative solution consists in using a Student distribution with $S - 1$ degrees of freedom, see e.g. [164], but this solution is not very powerful. In that case, the significance of the statistic should rather be assessed in a non-parametric framework.

7.1.4 Non-parametric assessment of the statistics

Although the use of analytical bounds for statistical tests is both efficient and elegant, this can be performed in only very restricted cases; in the case of the Gaussian mixed-effect model (see above), there is no such solution. Thus we need to resort to non-parametric techniques to estimate the reference pdf $p_0 = p(\phi|H_0)$ for any statistic ϕ .

Resampling techniques are usually used for that purpose [129]. In

particular, in the case of group studies in neuroimaging, the sign permutation approach consists in tabulating p_0 by resampling the estimated effects across all possible flips of sign, the number of which is 2^S . Such a procedure is intended to test a simple null hypothesis $H_0 : \bar{\beta} = 0$ about the parameter of interest under the restrictive hypothesis that p_0 is symmetric.

In the present case, two additional conditions must be fulfilled [108]: *i*) the subjects are drawn independently *ii*) first-level estimators are location equivariant and scale-invariant. These conditions imply that the distribution of standard test statistics are stochastically increasing w.r.t. the population mean effect, which validates the use of sign permutations to perform one-sided tests, i.e. testing the composite hypothesis $H_0 : \bar{\beta} \leq 0$. Note that these hypotheses are weaker than the standard normality assumption.

The test is exact conditional on the effects values and first level variances up to the discretization induced by the finite number of permutations. The accuracy of the test is traditionally controlled with the statistical on binomial variates. The main challenge with this kind of procedure is of course computational, given the number of voxels and permutation to be considered.

7.1.5 Making inference on images

So far, we have considered only voxel-based inference. However, neuroimaging data consists of images, which implies that many tests are performed simultaneously or that inference is performed on more complex objects. A known issue is that the expected number of false detections for a test performed on N_v independent sites at significance α is $N_v\alpha$. Controlling this later number amounts to applying a Bonferroni correction (see Sec. 6.2.3), which is correct for independent voxels, and typically conservative for positively correlated data as in neuroimaging (see e.g. [166] for a discussion on that point). In the case of non-parametric assessment procedure, it is however much more useful to control the family-wise error rate by tabulating the maximum statistic at the map level under the null hypothesis [129].

However, this kind of procedure rests only on signal amplitude, and is not optimal for spatially extended but weaker activations. For this kind of case -which is especially relevant in group studies, given the shortcomings of normalization discussed in section 6.4.1- a dual procedure might be preferred: *i*) in a first step, the map is thresholded at a relatively arbitrary threshold ($p < 0.01$, uncorrected, typically); *ii*) in a second step the maximal cluster size is tabulated under using the non-parametric framework, i.e. sign flips of the individual data; under the null hypothesis, the resulting distribution represents the distribution of the maximal supre-threshold cluster size *iii*) finally the clusters whose size is greater than a proportion $(1 - \alpha)$ of the tabulated distribution are reported as active. Here, a *cluster* refers to a set

of spatially connected voxels assuming a certain model of spatial contiguity (6, 18 or 26 nearest neighbors typically; we use 18 in our implementation).

There exists an interpretational difficulty with this procedure: the fact that a cluster is significantly greater than expected by chance does not mean that all the voxels within this cluster is active, but that there is at least one voxel within the cluster for which the null hypothesis can be rejected; in statistical terms, this approach provides a weak control of false detections. This problem is not fundamental, given that voxel-level information has little relevance in group studies.

7.1.6 A Bayesian perspective

It is perfectly possible to rephrase the previous analysis based using Bayesian concepts [164]: Assuming a weakly informative prior distribution on β and v_g , posterior distribution can be derived for these parameters. Note that in that particular case, non-informative priors as in [164] yield an improper posterior, hence a non-identifiable model [109]. An interesting feature of Bayesian assessment is that the H_1 hypothesis can also be rejected, showing that a region is indeed inactive, but this possibility has rarely been used in the neuroimaging literature.

A recent contribution [82] has tried to include the cross-subject spatial variability in data modelling, as if the unknown displacement variable could be marginalized out for the estimation of the template map. But this turns out to be impossible, and the proposed model is not a global generative model of the data (see [3] for the definition of such a model). Still a statistical test can be defined, and the procedure proposed in [82] is shown to provide less biased group-level activation maps than standard univariate detection procedures (the final test is a classical test, not a Bayesian test).

In fact, the Bayesian point of view on inference has not been successful so far, for several reasons:

- Weakly informative priors have, by definition, little or no effect on the posterior, which thus essentially represents the likelihood term.
- Alternatively, non-trivial priors, e.g. spatial priors under a Markov Random Field can be used (see Section 6.3), but this boils down to a post-processing of the map, which is furthermore costly. There is still little evidence in favor of its usefulness in group studies ([64, 163]).
- More generally, Bayesian approach do not solve multiple comparison issues and do not fit well with cluster-level assessment of the activated regions. For all these aspects, the classical non-parametric procedures provide reliable p-values which are still preferred for interpretation purpose (and for publication in neuroimaging literature).

7.2 Analysing the reproducibility of group studies

7.2.1 Specificity, sensitivity, reproducibility

Let t be a statistic that measures the positivity of a group-level parameter as defined in model (7.1-7.2); usually t is defined at the voxel level in the template (normalized) space. The quality of the ensuing inference can be assessed by different criteria: specificity, sensitivity, and reproducibility.

Specificity is the number or rate of detections that are allowed by a certain detection procedure, and corresponds to assessing how frequently ϕ is above some threshold t_α under the null hypothesis:

$$p(\phi > \phi_\alpha | H_0) < \alpha \quad (7.12)$$

This is the most important criterion, since it governs the probability of making false inference from a dataset, which is the most prominent pitfall for neuroscientists. Thus statistical approaches have been designed to control specificity at the map level, in particular with the use of Gaussian Random Field Theory [167], or non-parametric procedures (see Sec. 7.1.4).

Sensitivity or power is the capacity of the analysis to detect truly activated regions.

$$p(\phi > \phi_\beta | H_1) \geq \beta \quad (7.13)$$

Although this is of primary interest to neuroscientists, a lack of sensitivity is not a fundamental scientific issue because the failure to detect an effect does not allow one to conclude that the effect is indeed absent: the (H_1) hypothesis is not rejected if $\phi \leq \phi_\alpha$, see Eq. (7.12). Sensitivity is clearly limited by the small number $S \sim 10 - 20$ of subjects in the experiments and by the lack of spatial correspondence between activated areas across subjects: in univariate procedure, the lack of spatial correspondence results in an inflated variance in the activated regions, hence a lower sensitivity of the test statistic [145]. Moreover, the data may be non-normal, which limits the sensitivity of the standard statistical tests.

Reproducibility has rarely been considered so far, but it is also a fundamental aspect for neuroimaging inference: the results should be stable e.g. with respect to the cohort under study.

$$p(\phi > \phi_\gamma | \phi' > \phi_\gamma) \geq \gamma, \quad (7.14)$$

where t and t' represent the same statistic obtained from two different samples of subjects. In realistic conditions ($S \leq 20$ subjects), reproducibility is clearly limited by the number of false negatives rather than the presence of false positives. This means that the requirement of increasing the reproducibility of the results in neuroimaging studies is largely consistent with the need to increase the sensitivity.

We now describe some of our contributions to quantify reproducibility in a useful manner.

7.2.2 Quantifying the reproducibility in group studies

We have proposed two measures to assess the reproducibility of the activation maps derived from group analysis. The first, based on a mixture of binomial distributions, characterizes the stability of the status (active/inactive) of each voxel of the dataset. The second measures how frequently clusters of voxels are found at similar locations in the normalized MNI/Talairach space across subjects. We use these measures in a bootstrap framework that enable us to characterize the reproducibility of activation maps obtained at the group level.

Reproducibility measure at the voxel level. In order to estimate the reproducibility of a statistical model, we need a method to compare statistical maps issuing from the same technique, but sampled from different groups of subjects. We use the reliability indexes elaborated in [61, 88, 89]. Assume that a statistical procedure (e.g., thresholding) yields binary maps g_1, \dots, g_R for different groups of subjects. At each voxel v , an R -dimensional binary vector $[g_1(v), \dots, g_R(v)]$ is thus obtained. At the image level, the distribution of $G(v) = \sum_{r=1}^R g_r(v)$ is modelled by a mixture of two binomial distributions, one for the null hypothesis, one for the converse hypothesis: Let π_A^1 be the probability that a truly active voxel is declared active, $\pi_A^0 = 1 - \pi_A^1$ the probability that a truly active voxel is declared inactive, π_I^1 , the probability that a truly inactive voxel is declared active, $\pi_I^0 = 1 - \pi_I^1$ the probability that an truly inactive voxel is declared inactive, and λ the proportion of truly activated voxels. Then, using a spatial independence assumption, the log-likelihood of the data is written as

$$\log(p(G)|\lambda, \pi_A^0, \pi_I^0) = cst + \sum_{v=1}^{N_v} \log \left(\lambda(\pi_A^0)^{R-G(v)}(\pi_A^1)^{G(v)} + (1-\lambda)(\pi_I^0)^{R-G(v)}(\pi_I^1)^{G(v)} \right) \quad (7.15)$$

Assuming $R \geq 3$ the three free parameters, $\pi_A^0, \pi_I^0, \lambda$ can be estimated using an EM algorithm. Note that optimizing the model over its different parameters sequentially, and using an adequate initialization, we could run the model for $R = 2$, though with higher variability in the estimation. An example of mixture of binomial distributions is given in Fig 7.1.

Given these estimates, the coherence index κ , known as Cohen's kappa is computed to measure the concordance of the different observations with the mixture model. Let $p_0 = \lambda\pi_A^1 + (1-\lambda)\pi_I^0$ be the fraction of voxels that are correctly classified by the mixture model. p_0 should be compared to

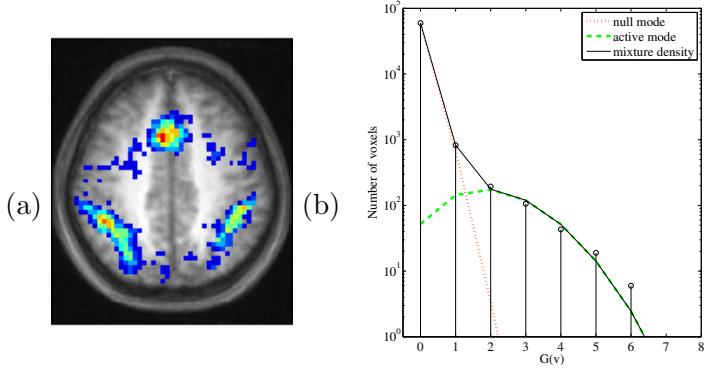


Figure 7.1: (a) Example of a reproducibility map G : the color codes for the number of times a given voxel has been declared active in a R -split of the population: from 0 (no color) to $R = 8$ (red). (b) Fitting the histogram of G with a mixture of binomial distribution. The empirical histogram of $G(v)$ is modelled by the model in Eq. (7.15), with $R = 8$. The Y axis is in log-coordinates for the sake of clarity.

the fraction of correct classifications that occur by chance $p_C = \lambda\pi^0 + (1 - \lambda)(1 - \pi^0)$, where $\pi^0 = \lambda\pi_A^0 + (1 - \lambda)\pi_I^0$ is the proportion of voxels declared inactive. The fraction of correct classifications corrected for chance is thus

$$\kappa = \frac{p_0 - p_C}{1 - p_C} \quad (7.16)$$

In this setting, $0 \leq \kappa \leq 1$ measures the fit of the mixture model to the data, which in turns reflects the concordance of the binary maps given as input to the model (7.15). If there is very little agreement on which voxels are active, the two components of the mixtures overlap, and κ is close to 0, whereas the separation between the components of the mixture increases and κ is close to 1 if there is a good agreement between binary maps. λ can also be retained as an index of the test sensitivity.

Note that more complex -and realistic- models have been proposed in the literature [93], in which the parameter λ is allowed to vary spatially. However, our main purpose is not activation detection, but obtaining a global reproducibility measurement; for this reason, we keep the basic setting.

Reproducibility measure at the cluster level Another way to assess the reproducibility of the results is to compare the positions of the clusters of supra-threshold voxels that arise through any group analysis. Assuming that the binary maps g_1, \dots, g_R are obtained from different groups of subjects

through a thresholding procedure, one can post-process them in order to yield connected components. The connected components with a size greater than a given threshold η are then retained, and their centre of mass (cm) is computed: let $x_i^r, i = 1..I(r)$ be the spatial coordinates of the *cms* derived from map g_r , we propose the following average distance between any two maps:

$$\Phi = \frac{1}{R(R-1)} \sum_{r=1}^R \sum_{s \in \{1, \dots, R\} - \{r\}} \frac{1}{I(r)} \sum_{i=1}^{I(r)} \min_{j \in \{1, \dots, I(s)\}} \phi(\|x_i^r - x_j^s\|), \quad (7.17)$$

where $\phi(x) = 1 - \exp\left(-\frac{x^2}{2\delta^2}\right)$ is a penalty function that is close to zero when the cluster centroids are properly matched and close to 1 otherwise. Φ represents the average mismatch between any *cm* of a supra-threshold component in a given map and the closest *cm* of any supra-threshold cluster obtained from another map. Appropriate penalty terms are used to handle the case $I(r) = 0$. We have performed some experiments using $\eta = 10$ voxels or $\eta = 30$ voxels, and use $\delta = 6mm$.

Procedure for the assessment of reproducibility The procedure consists in dividing the population of 81 subjects in $R = 2, 3, 4, 5, 6$ or 8 disjoint groups of $S = 40, 27, 20, 16, 13$ and 10 subjects respectively. The computation of different statistics, the derivation of an adequate threshold and the thresholding are performed in the different subgroups, and global reproducibility measures are derived from the ensuing binary maps. This procedure is repeated 100 times for each instance, yielding a distribution of the indexes κ , λ and Φ for each possible technique/parameter.

7.2.3 Observations from a large cohort

The dataset used in this experiment is the subset of the first $S = 81$ subjects from the localizer database (see Sec. C.4).

First, we choose the traditional RFX analysis procedure [see Eq. (7.4)], thresholded at $p < 0.001$ uncorrected, using an analytical threshold and evaluate the distribution of the different indexes for three contrasts of interest. This is important to understand how well the indexes are characteristic of the amount, the spread and the variability of supra-threshold activity. In particular, it is important that the estimated reproducibility indexes are less variable for a given contrast than across contrasts.

Second, we evaluate the choice of the threshold on the different indexes, in the case of the voxel-based t-test. While the sensitivity index certainly decreases while the threshold increases, the behaviour of the reproducibility may be more complex, due to the trade-off between false positive and

false negative rates (non-standard behaviours due to extremely low or high thresholds are not considered here).

Third, we study the behavior of the different measurements when the number of subjects in the group varies; while it is obvious that reproducibility increases with the group size, it is not clear whether there exists a plateau and at which level. Previous studies [33, 106] suggest also a steady increase of sensitivity with the group size.

Finally, we choose the following statistics: RFX, RFX on smoothed (12mm FWHM instead of 5mm) effect maps (SRFX), MFX, Wilcoxon(WKX), Cluster-level RFX (CRFX), Parcel-based RFX (PRFX) and Ψ FX. Ψ FX is formally defined as mixed effect statistic 7.11, in which the group variance v_g has been set to 0. This is also the statistic defined in [111]. However, we calibrate it with the non-parametric framework, making it a correct group inference statistic. RFX, SRFX, MFX, Ψ FX and PRFX maps are thresholded at the $p < 0.001$ level, uncorrected for multiple comparisons. CRFX is thresholded at $p < 0.01$, uncorrected level at the voxel level, then at $p < 0.01$, at the cluster level. Note that these choices are made in order to roughly balance the specificity of the methods, while using them in a standard way.

PRFX maps (see Sec. 7.3) are computed for $K = 500$ parcels. Since the parcel centres are defined at the group level in Talairach space, the voxels in the group result map are assigned to the parcel with the closest center in Talairach space. This results in a piecewise constant map, the pieces resulting from a Voronoi parcellation of the group mask into parcels. Note that in our bootstrap procedure, such boundaries are defined independently in each subgroup of subject. For parcellation, we use the hierarchical procedure presented in [143].

Reproducibility measurements for different cognitive contrasts. We computed the random effects z-variate for different cognitive contrasts, using $R = 5$ groups of $S = 16$ and a threshold $\theta = 3.1$ corresponding to $p < 0.001$ uncorrected for the contrasts *left-right button press*, *audio instructions-video instructions* and *computation-reading*. The reliability index κ , the proportion of putative true positives λ , and the inter-cluster distance penalty Φ are given in Fig. 7.2. It shows that κ and λ have different behaviours and are strongly dependent on the cognitive contrast under study. For instance, the left motor contrast activates relatively small regions with a relatively low reproducibility; the auditory-selective contrast activates larger regions with high reproducibility; the computation-selective contrast activates larger regions, but with low reproducibility. The inter-cluster distance penalty Φ does not discriminate between the different contrasts as strongly as κ . As could have been expected, it has the opposite behaviour (maximal for the computation contrast, minimal for the auditory contrast).

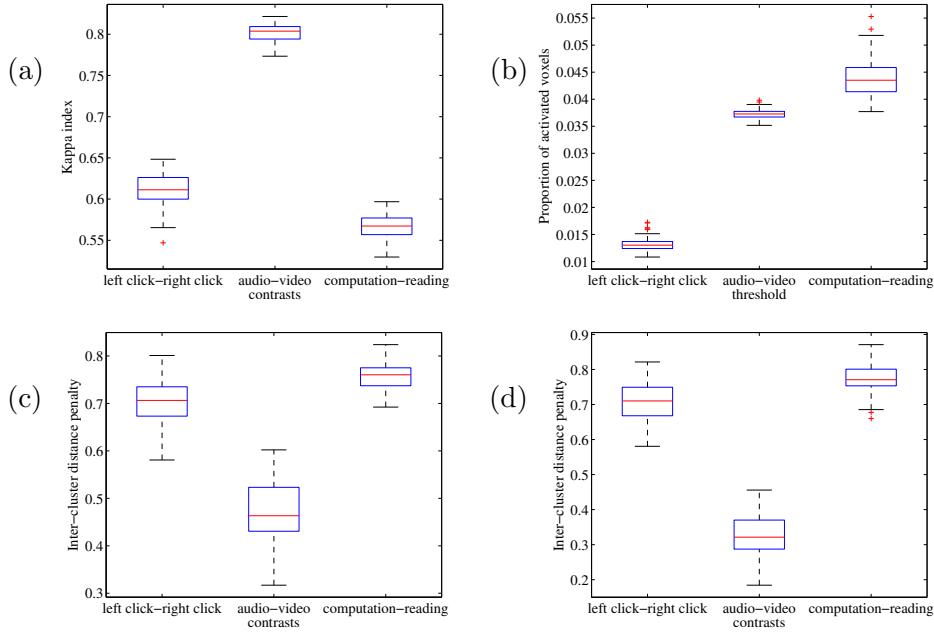


Figure 7.2: Dependence of the reproducibility and of the sensitivity of the random effects analysis on the functional contrast under consideration. These results are obtained by drawing 5 disjoint groups of $S = 16$ subjects in the population of 81 subjects, and applying the whole analysis procedure. The threshold is $\theta = 3.1$ ($p < 0.001$) (a) Over 100 replications, the reproducibility index is higher for the *audio instructions-video instructions* contrast than for the *left-right button press* and *computation-reading* contrast. (b) However, the size of the putatively activated areas is greater for the contrast that shows regions involved in computation, and smaller for the contrast that shows the regions involved in motor activity. (c-d) The cluster variability penalty Φ is presented for clusters of more than $\eta = 10$ (c) or $\eta = 30$ (d) voxels (the lower the better). The behaviour is as expected, with the smallest value for the auditory-specific contrast.

How the threshold affects the reproducibility of the analysis ?

Here we study the behaviour of our reproducibility measures when applied to a thresholded RFX map, when we let the threshold vary. The reproducibility measure is computed for 100 different splits of the population of subjects into $R = 5$ groups of $\mathcal{S} = 16$ subjects, in the case of the *left-right button press* contrast. The threshold (in z-variate scale) varies from $\theta = 2.2$ ($p < 0.015$, uncorrected) to $\theta = 4.0$ ($p < 3.2 \cdot 10^{-5}$, uncorrected) in steps of 0.2.

As expected, the sensitivity parameter λ decreases when θ increases (see Fig. 7.3(b)). More interestingly, κ reaches a maximum for $\theta^* \sim 2.7$, but the index remains close at least for $\theta < 3.5$ as can be seen in Fig. 7.3(a). Accordingly, the inter-cluster distance penalty Φ is minimized for a threshold $\theta^* \sim 3$. The correspondence of these results is interesting, given that these two similarity measures are obtained independently, and based on different considerations. Note that we have obtained similar results when studying the other contrasts with slightly higher (auditory contrast) or lower (computation contrast) threshold values. Thereafter, we retain the threshold $\theta = 3.1$ ($p < 0.001$, uncorrected for multiple comparisons) for random effects z-statistics.

How many subjects are necessary to obtain a reliable group map ?

We study the dependence of κ , λ and Φ when we let the size \mathcal{S} of the group vary. We base our investigation on the *left-right button press* contrast, with group maps thresholded at the $\theta = 3.1$ ($p < 0.001$, uncorrected) level. The results are presented in Fig. 7.4. It shows that the reproducibility increases with the group size, which was expected. The sensitivity also increases with the group size. Interestingly, the reproducibility reaches a plateau only for $\mathcal{S} \approx 25$. The inter-cluster distance penalty Φ has a similar behaviour, with a plateau for $\mathcal{S} = 27$ subjects when $\eta = 10$, while lower values are reached when using $\eta = 30$.

Comparison of different group analysis methods. Now we study how the reproducibility index behaves for different statistical methods: The t statistic [RFX, see Eq. (7.4)], the same test after 12 mm smoothing of the data -instead of 5mm- (SRFX), the mixed effects statistic, controlled by permutation [MFX, see Eq. (7.11)], the parcel-based RFX test (PRFX), the t-statistic thresholded at the cluster-level (CRFX), the Wilcoxon test (WKX), and the pseudo-MFX test Ψ FX. RFX, SRFX, MFX, WKX, PRFX and Ψ FX maps are thresholded at the $p < 0.001$ level, uncorrected for multiple comparisons. The CRFX map is first thresholded at the $p < 0.01$, uncorrected level, then at the $p < 0.01$ cluster-level. The results are obtained by bootstrapping in $R = 8$ groups of size $\mathcal{S} = 10$. The results are presented in Fig. 7.5.

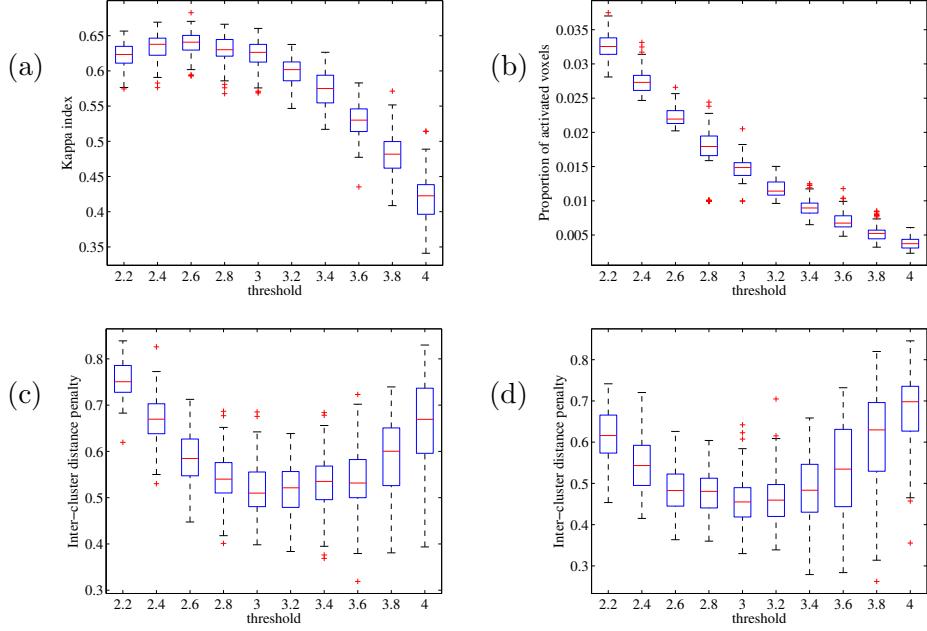


Figure 7.3: Dependence of the reproducibility, the sensitivity, and the distance between supra-threshold clusters of the group random effects analysis on the threshold chosen to binarize the statistic maps. These results are obtained by drawing 5 disjoint groups of $S = 16$ subjects in the population of 81 subjects, and applying the procedure described in section 7.2.2. This is performed on the images of the *left-right button press* contrast, with 100 resamplings. (a) The reproducibility index κ shows is maximized for $\theta \sim 2.7$. (b) The sensitivity decreases when θ increases. (c,d) The average distance between supra-threshold clusters of more than 10(c) or 30(d) voxels across groups has a minimum around $\theta \sim 3$.

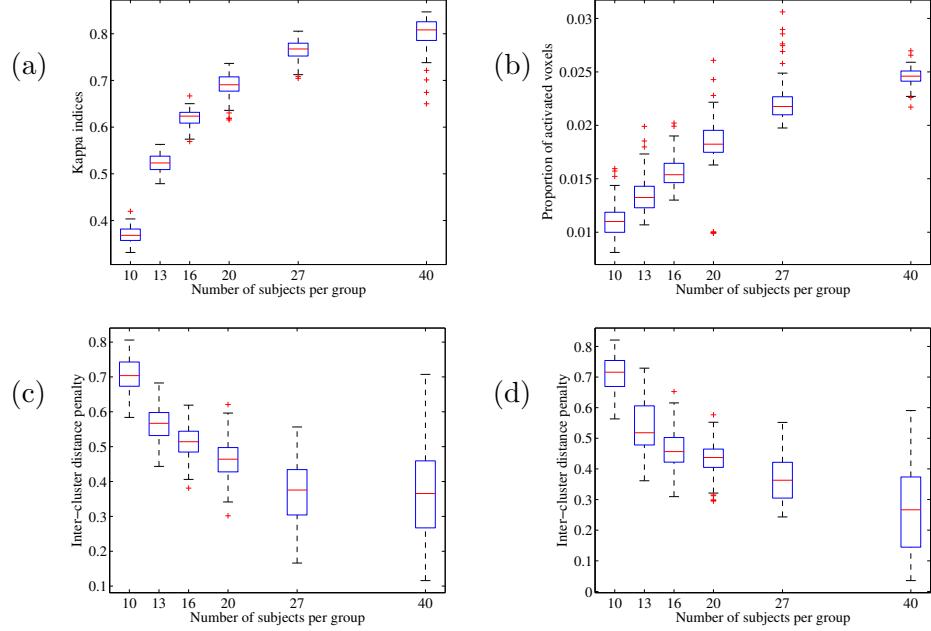


Figure 7.4: Effect of the RFX group size on reproducibility κ (a), sensitivity λ (b) and the average distance between supra-threshold cluster centroids Φ (c-d). The reproducibility is assessed considering disjoint groups of size $S = 10, 13, 16, 20, 27, 40$ within the population of 81 subjects. This is performed on the images of the *left-right button press* contrast, with 100 resamplings. (a) The reproducibility index increases with S and reaches a plateau for $S > 20$. (b) The size of putatively activated areas steadily increases with S . (c-d) The average intra-cluster distance decreases with S ; it reaches a plateau for $S > 20$ when $\eta = 10$ (c), whereas it further decreases when $\eta = 30$ (d).

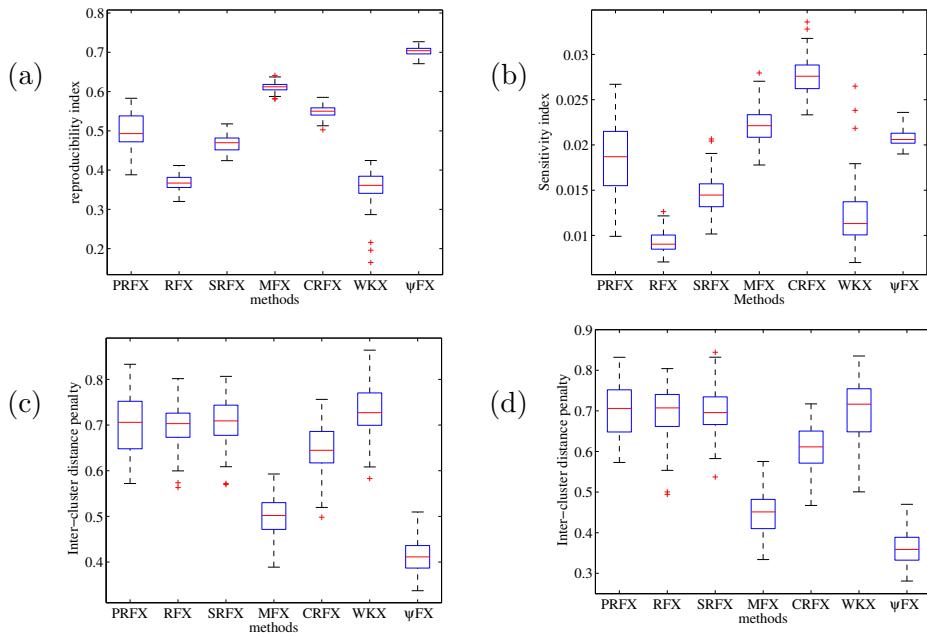


Figure 7.5: Dependence of the reliability κ (a), the sensitivity λ (b), inter-supra-threshold cluster distance penalty Φ (c-d) of the statistical analysis on the group statistic used. Φ is based clusters of size greater than $\eta = 10$ (c) or $\eta = 30$ (d). These quantities assessed considering $R = 8$ disjoint groups of size $S = 10$ within the population of 81 subjects, using the *left-right button press* contrast, and 100 resamplings.

From the point of view of reproducibility, the WKX and RFX tests have the worst performance overall, while the SRFX performs slightly better. CRFX, PRFX and MFX techniques yield higher reproducibility, but Ψ FX yield the highest values. The results are more variable with PRFX than with other techniques; this reflects the fact that PRFX is based on a smaller number of volume elements, so that statistical tests have a less stable behaviour purely due to fewer number of parcels compared to voxels.

CRFX, MFX, and to a lesser extent, PRFX tests are more sensitive, i.e. have a larger fraction of generally activated voxels, than voxel-based tests. Note however that the specificity control of CRFX matches the other approaches only approximately.

Finally, the average supra-threshold cluster distance Φ is minimal for Ψ FX, and relatively low for MFX. It is similar for the other techniques. Before discussing these results, let us first introduce some descriptive statistics to enlighten and complete the analysis of this dataset.

7.2.4 Limitations of the standard Gaussian mixed-effects model

Are the effects Gaussian distributed ? An important question that may arise is whether the effects $\hat{\beta}_s(v)$ are normally distributed or not, since this is a key assumption in standard (random effects) group analysis. We have used the D'Agostino-Pearson test (see Section C.3 and [168] for details), based on the computation of the skewness and the kurtosis (third and forth order cumulants) of the values $\{\hat{\beta}_s(v)\}, s = 1..S$ in each voxel v . This provides the p-value of the D'Agostino-Pearson statistic under the null (normal) hypothesis. For the sake of visualization, we convert the p-value into a z-value. We have then repeated the procedure based on the normalized effects $\{\bar{\beta}_s(v) = \frac{\hat{\beta}_s(v)}{\hat{\sigma}_s(v)}\}, s \in \{1,..,S\}, v \in \{1,..,N_v\}$ which removes a potential variability in signal scaling across the population. At the group level, the normalization through the residual magnitude has a much greater impact than the deviation from normality on the resulting tests due to the fact that $\hat{\sigma}_s(v)$ is estimated with a finite ($\nu = 100$) number of degrees of freedom.

Analysis of the second-level variance We performed the D'Agostino-Pearson test on the effects $\hat{\beta}(v)$ of all the voxels, as well as the normalized effects $\bar{\beta}(v)$, which yields two maps for each contrast. We present them for *left-right button press*, *audio instructions-video instructions* and *computation-reading*, thresholded at the $p < 0.001$ uncorrected level. We also present the inter-subject variance maps $v_g(v)$ computed in a mixed-effect model (see Sec. 7.1.3). We present these maps together with the RFX map (converted to a z-variate) based on 81 subjects in Figs. 7.6-7.8. Note that other contrasts, e.g. *horizontal-vertical checkerboards*, *sentence*

reading-low-level vision, cognitive trials-motor trials, and the opposite ones, not presented here due to space limitations, yield qualitatively similar results.

In each case, the regions with highest group variance are found in the regions with highest random effects statistics in absolute values; some of them are absent in the maps 7.6-7.8, where signed statistics are presented.

Inspection of these maps suggests that

- Areas of high variance tend to co-localize with the activated areas. This implies that the parameters $v_g(v)$ and $\bar{\beta}(v)$ are certainly not independent, and that statistics that are penalized by the group variance may not be very efficient in general.
- Non-normality is very significant in wide regions of the brain: deviation from normality of $\hat{\beta}$ across subjects concerns 22% of the brain voxels at ($p < 0.001$, uncorrected) for the *computation-reading* contrast, 27% for the *left-right button press* contrast and 30 % for the *audio instructions-video instructions* contrast.
- Deviation from the normality hypothesis is much lower for the normalized effects $\bar{\beta} = \frac{\hat{\beta}}{\hat{\sigma}}$ than for the raw effects $\hat{\beta}$. For instance, the rate of voxels with normality rejected at ($p < 0.001$, uncorrected) drops from 22% to 9.2 % for the *computation-reading* contrast, from 27% to 2.9% for the *left-right button press* contrast and from 30 % to 10% for the *audio instructions-video instructions* contrast. This means that dimensionless first-level statistics yield more homogeneous quantities across subjects than effects expressed in percents of baseline signal increase.
- Deviation from normality of the effects does not specifically co-localize with activated areas, but, in several cases it coincides with the boundaries of activated areas.

From Figures 7.6-7.8, one of the most striking effects is the co-localization of high second-level variance areas with large random effects areas. Such an effect is not expected since the RFX is defined as the quotient of the estimated mean effect by the standard deviation of this estimate.

The interpretation could be that 1) the contrast-to-noise ratio (CNR) of the BOLD effect is highly variable across subjects, and by definition this effect does not appear in non-activated areas and/or 2) spatial mis-registration² implies that at a given voxel, i.e. a given position in MNI space, some subjects have activity while other subjects do not, thus spatially widening the signal distribution. For simple contrasts such as those

²Spatial mis-registration may be artefactual (incorrect normalization) or not (intrinsically different functional anatomy.)

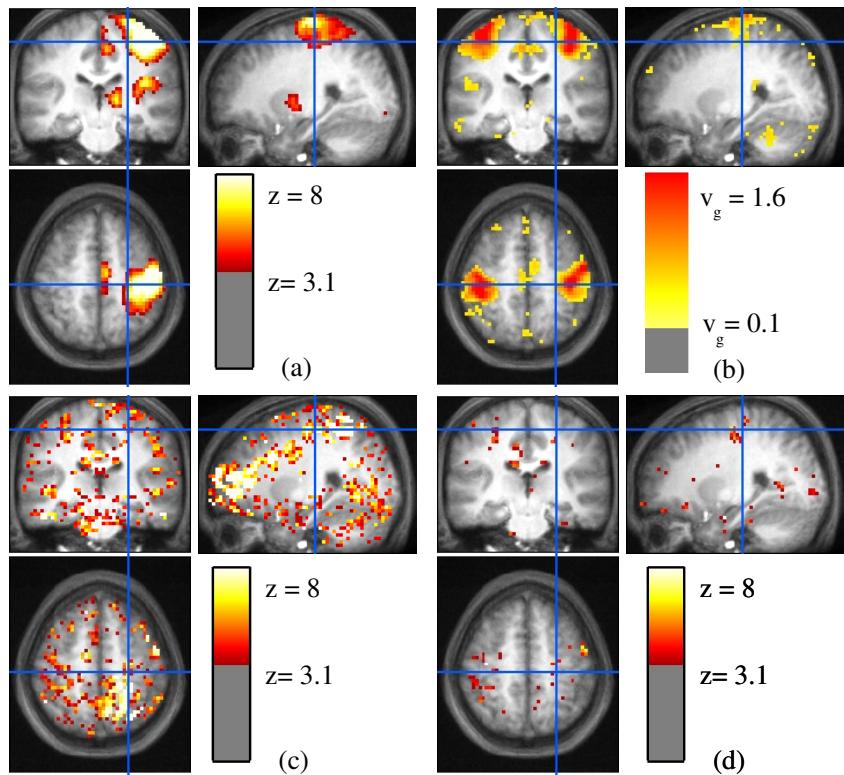


Figure 7.6: Statistical model of the effects for the *left-right button press* contrast, on $S = 81$ subjects. (a) z-value associated with the RFX test; (b) group variance estimate; (c) z-value of the D'Agostino-Pearson test for normality of the effects $\hat{\beta}$; (d) z-value of the D'Agostino-Pearson test applied to the normalized effects $\bar{\beta}$. Note that all the z values are limited to the $[-8, 8]$ range. The color scale of the variance image has been chosen arbitrarily in order to have supra-threshold areas that are comparable with the other maps. The variance is expressed in squared percentage of the BOLD mean signal. Cross position: $(-23, -28, 56)$ mm in the MNI space.

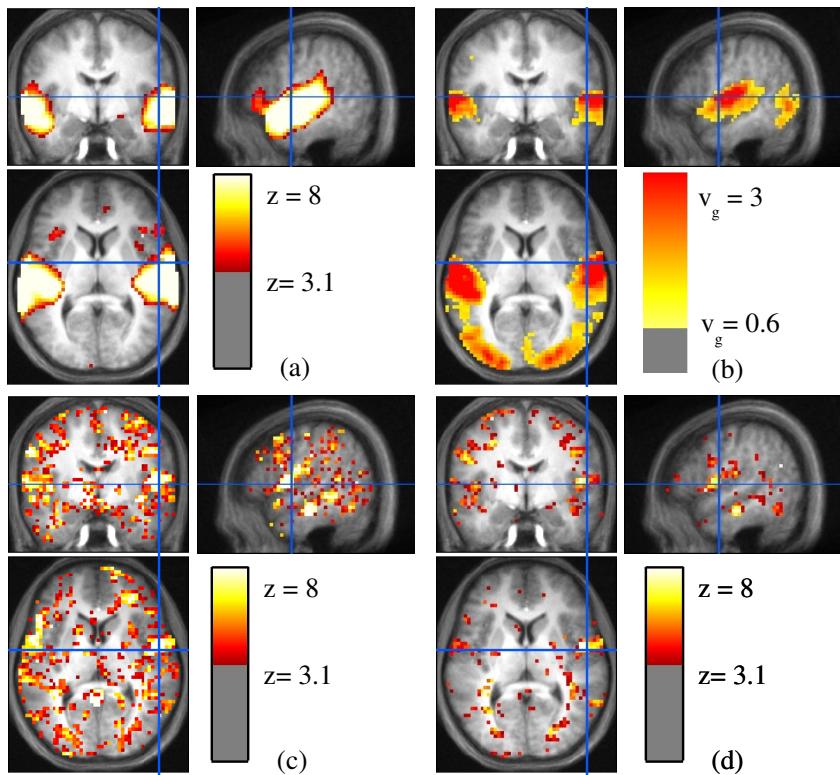


Figure 7.7: Statistical model of the effects for the *audio instructions-video instructions* contrast, on $S = 81$ subjects. (a) z-value associated with the RFX test; (b) group variance estimate; (c) z-value of the D'Agostino-Pearson test for normality of the effects $\hat{\beta}$; (d) z-value of the D'Agostino-Pearson test applied to the normalized effects $\bar{\beta}$. Note that all the z values are limited to the $[-8, 8]$ range. The color scale of the variance image has been chosen arbitrarily in order to have supra-threshold areas that are comparable with the other maps. The variance is expressed in squared percentage of the BOLD mean signal. Cross position: $(-54, -6, 8)$ mm in the MNI space.

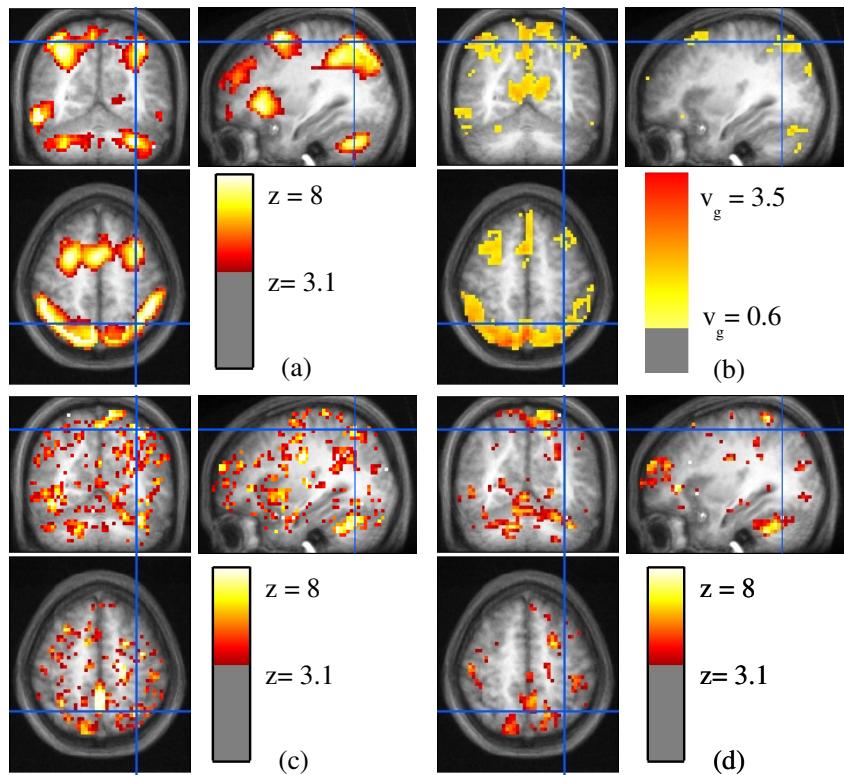


Figure 7.8: Statistical model of the effects for the *computation-reading* contrast, on $S = 81$ subjects. (a) z-value associated with the RFX test; (b) group variance estimate; (c) z-value of the D'Agostino-Pearson test for normality of the effects $\hat{\beta}$; (d) z-value of the D'Agostino-Pearson test applied to the normalized effects $\bar{\beta}$. Note that all the z values are limited to the $[-8, 8]$ range. The color scale of the variance image has been chosen arbitrarily in order to have supra-threshold areas that are comparable with the other maps. The variance is expressed in squared percentage of the BOLD mean signal. Cross position: (-33,-60,56)mm in the MNI space.

used (left or right button press, sentence listening), different cognitive strategies should be ruled out.

This inflated variance effect certainly deserves more investigation, given its prominent effect on statistics (sensitivity and reproducibility): for instance, the Ψ FX statistic -that does not take into account the group variance, hence is simply a weighted average of the subject-based effects- seems more reliable than the MFX statistic, which is itself much more reliable than the RFX statistic (see Fig. 7.5). The effect of group variance is also an argument in favour of Bayesian analysis of fMRI data, if the reference signal level is not 0 [55].

Non-normality is another important factor. To our knowledge, this has not been investigated before, since it requires a high number of subjects. Interestingly, the importance of non-normality is reduced when considering normalized effects $\bar{\beta}_s(v)$ instead of raw effects $\hat{\beta}_s(v)$. This shows that information from the first-level analysis that are used in group statistics have to be chosen carefully. In particular, the difference observed between the normality of $\hat{\beta}$ and $\bar{\beta}$ maps possibly indicates that the current way of normalizing signal magnitude with respect to the mean signal may not be optimal for inter-subject comparison (this is also an open question for inter-session variability). However, the normalization with respect to first-level variance might not be satisfactory, since it could in turn be highly dependent on acquisition artifacts, motion and physiology, whether these are modeled or not. We are not aware of any successful signal calibration strategy, but mixed-effects model may solve part of the problem. Interestingly, several areas with significant non-normality are found at the periphery of activation maxima, confirming the impact of spatial shifts on group statistics. Once again, further investigations on non-normality may be performed, e.g. searching different groups of subjects in the population or outlier subjects (see [83]). Robust statistics might also be used for inference [157], but at the risk of a weaker control on specificity. Moreover, such inference schemes raise the difficult question of the generalizability of group results to other groups of subjects (given that the concept of *outlier* is ill-defined when considering a small group). In general, it is advisable to use non-parametric assessment to obtain reliable thresholds [107]. However, the choice of robust statistics (statistics that adapt to non-normal data) is not necessarily advantageous: for instance, the Wilcoxon statistic did not perform better than other statistics in our experiments (see Fig. 7.5).

Practical conclusion : reliability of the different statistical tests

One of the most important practical questions is to describe or design the most efficient ways to perform group studies in neuroimaging. Based on this first study we can suggest some guidelines.

First of all, given the results on normality tests, non-parametric assessment of functional activity should be preferred to analytical tests, which may rely on incorrect hypotheses. This can be done using adapted tool-boxes e.g. SnPM [71] or Distance [107]). It is worthwhile to note that C implementation of the tests reduces computation time to a reasonable level (e.g. cluster-level p-values can be computed in less than one minute on a ten subjects dataset). Non-parametric estimation of the significance improves both the study sensitivity and reproducibility.

Second, mixed-effects models should systematically be preferred to mere random effects analyses: there is some information in the first level of the data that improves the estimation of the group effects/variance and statistic.

Third, cluster- and parcel-based inference should be preferred to voxel-based thresholding. Cluster-level inference is of frequent use, which benefits the sensitivity and the reliability of group analyses. However, it is based on the assumption that activated regions are large, which is not necessarily true. Parcel-based inference may thus be an interesting alternative, since it further allows some spatial relaxation in the subject-to-subject correspondence. The price to pay is a larger variability of the results due to a less stable decision function (activated vs non activated). We recommend the combination of one of these techniques together with MFX. By contrast, stronger smoothing (12mm FWHM instead of 5mm) did not increase significantly the results reliability.

Fourth, to our surprise, Ψ FX was found to be the most reliable technique. Although the statistic function does not take into account the group variance - as argued earlier, this is probably the reason for its higher performance - its distribution under the null hypothesis is tabulated by random flips of the effects signs, so that it is indeed a valid group inference technique. However, it should be used with care because first the thresholds have to be computed voxel per voxel, (i.e. are not spatially stationary), and second the statistic value itself has no obvious interpretation, in contrast to the RFX and MFX statistics.

7.3 Spatially relaxed inference: parcel-based RFX

7.3.1 Parcellation for group inference: definition and constraints

Definition of the problem As discussed in Sec. 7.2.1, the main limit of standard inference procedures is their weak sensitivity. This lack of sensitivity is, in turn largely related to the choice of mass-univariate analysis, which entails a severe multiple comparison problem. As noticed in Section 7.1.5, cluster-level inference is in general much more sensitive than voxel-

level inference. Moreover, given the shortcomings of spatial normalization discussed in section 6.4.1, there is little interest in inferring the presence of activation at the voxel level: a regional or super-voxel level of inference is clearly more adapted.

Recently, [48] proposed to group voxels into anatomically and functionally homogeneous parcels across subjects. This latter approach is appealing for several reasons: *i*) the voxels, which are the spatial units of functional images, define an arbitrary spatial resolution; small regions modelled as groups of voxels may better reflect true regions of activity, as they can be characterized in fMRI; *ii*) this makes the analysis less sensitive to -artefactual or intrinsic- mis-registrations in multi-subject studies; *iii*) technically, this under-sampling of the brain volume or surface reduces the well-known problem of multiple comparisons, allowing for less conservative Bonferroni corrections; *iv*) in general, it may also better take into account the true between subjects anatomical variability such as cyto-architectonic [130] or sulco-gyral [128] structure of the cortex. The model proposed here follows this parcellation idea.

Previous work related to parcellation Although the idea of constructing automatically anatomo-functional parcels is relatively recent in neuroimaging, some works are clearly related to this line of thought and we briefly review some of those attempts.

First, brain parcellations have been proposed in a purely anatomical context. Existing methods are based either on prior knowledge of anatomy and connectivity [102], on sulcal geometry [97, 147, 90, 141], on sulcal identification [19], or on probabilistic atlases [131, 45]. We note that these parcellations can be used in a functional context, after co-registration of anatomical and functional images. This may be done semi-automatically [112], or using atlases [152]. Note that this procedure might suffer from poor anatomo-functional correspondence related to EPI distortions; more generally the functional homogeneity of the resulting parcels should be checked for inference purposes.

By contrast, parcellation based on functional information is a relatively novel approach, with specific challenges. Clustering of similar time series yields homogeneous functional regions [136], but purely functional clusters are well-defined only locally [66], and specific methods have to be designed for the parcellation of the entire brain. For instance, [120] have proposed an EM algorithm to jointly model the spatial location of activation, together with the activation amplitude at each cluster. This joint modeling is reminiscent of [47]. However, this model is rather adapted to encode sparse activation patterns, while we aim at a parcellation of the entire brain volume or large brain regions, consistently across subjects.

Last, in [46, 47, 48] Flandin et al. launched fMRI-tailored parcellation approaches; they have proposed a clustering approach based on Gaussian Mixture Models (GMMs) that groups voxels from multiple datasets according to a spatio-functional criterion, blindly to the subject. This is a very effective approach, since it is based on quick algorithms. In this framework, the anatomy of the subject can be adequately introduced in the core algorithms and the technique naturally yields results such that inter subject analysis assumptions are naturally enforced. To our knowledge, this is a first approach to deal with anatomical and functional variability for multi-subject analysis. However, it has a few limitations. First, it does not guarantee the spatial connectivity of the parcels. Secondly, it does not necessarily produce multi-subject parcels (or cliques) in which each subject of the group is present, which should be a desirable feature³. This may or may not be a limitation depending on whether subjects are functionally homogeneous or not, but some application may require that corresponding parcels can be found in all or most of the subjects. Last, the definition of a spatial/functional criterion may be cumbersome, since it contains non-homogeneous terms and the relative weighting of the functional and the spatial information is still an open question.

Particular constraints when using parcellations for group-level inference Thereafter, it is understood that a parcellation can be performed on volumetric data (images), or surface-based data (meshes); it is only required that a coordinate system defines the position of any object in either case. Let us briefly describe the desired parcel characteristics:

Let $(\beta_s)_{s=1..S}$ be a set of fMRI data consisting in one or several contrast maps sampled from S different subjects (note that each β_s can be multi-dimensional). Basically, the information available on each site (voxel in 3D space or vertex of a mesh) consists of a position information τ and a feature information β . A parcellation of the dataset is a subdivision of each individual dataset into K regions which is consistent across subjects. Let $(\lambda_s)_{s=1..S}$ be the associated labelling functions -on in each subject⁴. Consistency refers to the following criteria

1. *Onto property*: Each subject should be present in each individual data
2. *Intra-subject connectivity*: Each parcel in each subject should be spatially connected.
3. *Intra subject homogeneity*: Each parcel in each subject should represent a functionally homogeneous region, which means that within parcel variance should be as small as possible in each subject.

³This may be called the *onto property* of the parcellation.

⁴Note that this is a hard labelling: each site belongs to a single parcel

Input : G : a graph describing spatial neighborhood relationships between the N_v sites F : feature matrix of shape (N_v, dim) K : number of desired clusters Initialize all the clusters for $k = 1 : N - K$ find (i, j) such that $i \neq j$ and $\text{var}_{q \in j \cup i}(F(q))$ is minimized assign i and j to cluster $(k + N)$ let $F[k + N] = F[i] \cup F[j]$ endow $(k + N)$ with the neighbors of i and j delete i and j end for OUTPUT: the K remaining components $k \in \{1, 2 * N - K\}$
--

Figure 7.9: Pseudo-code of an algorithm that performs a variance-minimizing and topologically correct parcellation of an individual dataset. This can be seen as a variant of Ward's algorithm.

4. *Inter-subject spatial coherence*: the relative position of parcels should be the same across individuals. This means that the spatial transformation implicit in the parcellation should be regular in some sense (e.g. diffeomorphism).
5. *Inter-subject homogeneity*: the parcels should match across subjects concerning some feature information (position, functional activity, possibly anatomical or functional connectivity)

Before proposing a solution, it is important to notice that the problem is relatively easily solved in the intra-subject case with an adapted clustering algorithm. In particular, conditions 4 and 5 disappear, while condition 1) is almost naturally enforced. Thus the problem boils down to clustering the 3D set in order to minimize the within-parcel variance, while enforcing the spatial connectivity of the parcels. We present in Fig. 7.9 a solution based on a modification of Ward's algorithm that includes spatial connectivity constraints.

7.3.2 Solution: parcellation algorithm

Unfortunately, solving the problem in the intra subject case does not help much for solving it in the inter-subject case: While it is a priori possible

to define parcels independently in each dataset and then to match them in a second step, the matching complex is at the same time complex and ill-posed, given that there may exist no natural one-to-one mapping between the individual structures.

A more adapted solution consists in building a population-level model of the parcellation, which reflects the average of individual information, and defines the topological constraints and the spatial and functional information related to the parcellation. This model is called *template*, and its counterpart in each subject will be called *prototype*. This template/prototype/data hierarchy can be loosely interpreted as a generative model, as shown in figure 7.10(left). The functional information/position of the template, prototypes, and voxels will be denoted $\beta(k)/\bar{\tau}(k)$ for $k \in 1,..,K$, $\tilde{\beta}_s(k)/\tilde{\tau}_s(k)$ for $k \in 1,..,K, s \in 1..S$ and $\beta_s(v)/\tau_s(v)$ for $v \in 1,..,N_v, s \in 1..S$. Recall that the assignment from voxels to parcels is denoted λ_s

We can now rephrase the problem Find $(\lambda_s)_{s \in 1,..,S}$ such that

1. $\forall s \in [1,..,S], \lambda_s ([1,..,N_v] \rightarrow [1,..,K])$ is onto,
2. $\forall k \in [1,..,K], (\lambda_s)^{-1}(k)$ is connected in 3D,
3. $\forall s \in [1,..,S], \sum_{k=1}^K var(\beta_{(\lambda_s)^{-1}(k)})$ is minimized, i.e. $J_{intra}(s) = \sum_{v=1}^{N_v} \|\tilde{\beta}_s(\lambda_s(v)) - \beta_s(v)\|^2$ is minimized, assuming that $\tilde{\beta}_s(k) = mean_{\lambda_s(v)=k} \beta_s(v)$.
4. The mapping $\forall(s, \sigma) \in [1,..,S]^2, (\tilde{\tau}_s) \rightarrow (\tilde{\tau}_\sigma)$ is smooth. Here we weaken this condition to: $\forall s \in [1,..,S], (\bar{\tau}) \rightarrow (\tilde{\tau}_s)$ is smooth, in a sense which is made precise below.
5. $\sum_{k=1}^K var_s(\tilde{\beta}_s(k))$ is minimized, i.e. $J_{inter} = \sum_{k=1}^K \|\tilde{\beta}_s(k) - \bar{\beta}(k)\|^2$ is minimized, assuming that $\bar{\beta}_s(k) = mean_{s \in 1,..,S} \beta_s(k)$

The estimation model essentially minimizes the criteria 3) and 5), while criteria 1) and 2) are taken as hard constraints. Note that, assuming that the number of voxels per parcel is constant across subjects, by Huygens formula, minimizing $J = J_{inter} + \sum_{s=1}^S J_{intra}(s)$ simply amounts to minimizing

$$J_T = J_{inter} + \sum_{s=1}^S J_{intra}(s) = \sum_{s=1}^S \sum_{v=1}^{N_v} \|\beta_s(v) - \bar{\beta}(\lambda_s(v))\|^2, \quad (7.18)$$

meaning that our procedure simply minimized the global inertia of the data clustering some hard constraints. Criterion 4) is handled in the following way: the group-level coordinates $(\bar{\tau}(k)_{k=1..K})$ are used to build a topological model of the parcellation which is simply the Delaunay triangulation of the ambient space. The resulting Delaunay graph defines the neighboring system on the template. $\forall s \in [1,..,S]$, the Jacobian of the mappings $(\bar{\tau}) \rightarrow (\tilde{\tau}_s)$

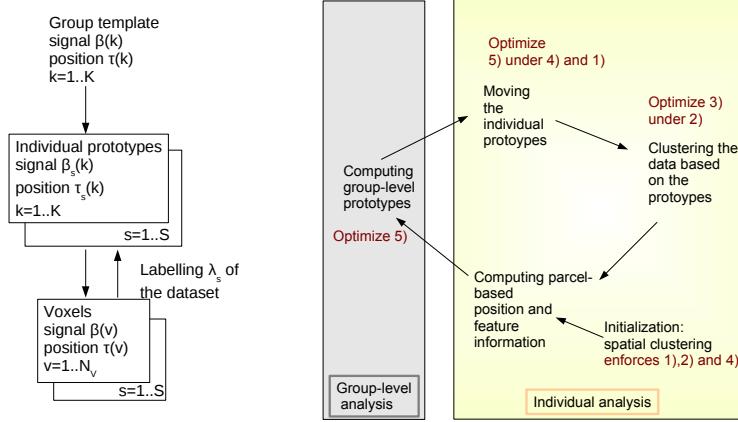


Figure 7.10: (left) Graphical interpretation of the parcellation model the descending arrows represent the *generative model*. (right) Scheme of the solution for the estimation of parcels for group inference.

w.r.t. the graph topology can then be computed and only assignments that have a positive Jacobian $\forall k \in [1, \dots, K]$ are accepted.

In practice, the model is estimated in a step-wise manner, as illustrated in Fig. 7.10(right). Assuming that an initial labellings λ_s^0 are available, and that these labellings enforce conditions 1),2) and 4), the following steps are alternated:

- Computation of parcel-based position $\tilde{\tau}$ in the anatomical space and feature-based information $\tilde{\beta}$, for each parcel $k \in 1, \dots, K$ and any subject $s \in 1, \dots, S$.

$$\tilde{\beta}_s(k) = \text{mean}_{\lambda_s(v)=k} \beta_s(v) \quad (7.19)$$

$$\tilde{\tau}_s(k) = \text{mean}_{\lambda_s(v)=k} \tau_s(v) \quad (7.20)$$

Note that these average values make sense only if the parcels are compact enough in both cases, which corresponds to the criterion 3) above, and requires that the number K of parcels is not too large.

- Computing the group-level prototypes or template, which means that population-average values are computed for both the position and functional information.

$$\bar{\beta}_s(k) = \text{mean}_{s \in 1, \dots, S} \tilde{\beta}_s(k) \quad (7.21)$$

$$\bar{\tau}_s(k) = \text{mean}_{s \in 1, \dots, S} \tilde{\tau}_s(k) \quad (7.22)$$

- In each subject $s \in 1, \dots, S$, recompute prototypes $(\tilde{\beta}, \tilde{\tau})$ that minimize criterion 5) given the constraints of 4). Here it is understood that $\forall k \in 1, \dots, K$, the searched prototype is a particular voxel v in subject s that minimizes $j(v) = \|\beta_s(v) - \bar{\beta}(k)\|^2$ under the constraint that the mappings $(\bar{\tau}) \rightarrow (\tilde{\tau}_s)$ have a positive Jacobian at each node of the template graph. In practice, the voxel v is thus the minimizer of j within a domain which is the intersection of the half spaces supported by its neighbours.
- The following step simply consists in computing the mapping λ^s in order to minimize $\|\tilde{\beta}_s(\lambda_s(v)) - \beta_s(v)\|^2$. In a spirit similar to the intra-subject case, a greedy procedure is used to group each voxel to the nearest prototype in functional space, but where assignments are allowed only when connectivity is ensured. This is done using a greedy procedure similar to Dijkstra's algorithm.

Convergence The convergence of the algorithm is checked by computing the criterion J . In practice, convergence proceeds in 5 steps.

Initialization The initialization is performed by drawing random positions $\bar{\tau}(k), k = 1, \dots, K$ in the common space and then by deriving $(\tilde{\tau}_s(k)), k = 1, \dots, K, s = 1, \dots, S$ and the initial assignments λ_s^0 without taking into account the functional information. This ensures that the basic constraints of the algorithm are met, while the criterion J is clearly not optimized. The final value of J , and a fortiori the assignments λ_s clearly depend on this initial choice, but the fluctuations are usually relatively low compared to the difference between the original and the final value of J (around 10%).

Spatial regularization. Running this algorithm (both in the intra- and inter-subject case) usually provides extremely non-regular shapes for the resulting clusters - although the connectivity constraints are met. In practice, it is extremely useful to include spatial regularization in the procedure. A simple way to do that is to replace the features $\beta_s(v)$ by

$$\underline{\beta}_s(v) = \left[\frac{1}{\beta_0} \beta_s(v); \frac{\mu}{\tau_0} \tau(v) \right], \quad (7.23)$$

where β_0 and τ_0 are arbitrary quantities -typically the standard deviation of the input data- which are used to make the coordinates dimensionless, and μ is a constant that weights the two relative information. This parameter can be set e.g. by qualitatively assessing the parcellation in an intra-subject case.

7.3.3 The Parcel-based random effects (PRFX) procedure

Once the parcels, i.e. the assignment functions (λ_s) are computed, it is about trivial to compute parcel-based average of informations of interest, e.g. activation values. The parcels are then simply handled as super-voxels in inference procedures. For simplicity, we concentrate on random-effects procedures only, given that a mixed-effects procedures entails several levels of variance: the voxel-level first-level variance $\sigma_s(v)$ in Eq. (7.1), the variance of the mean estimator (which can easily be taken into account by bootstrap), and the inter-subject variance. This point remains an interesting matter for future work.

We thus consider only the last variance term and compute the t-statistic associated with this model. The nice feature with parcellation is that the multiple comparison problem is alleviated by the replacement of N_v voxels with a much smaller number K of parcels. Assuming that the parcellation model is reasonable, this gain may not be lost by a reduction of spatial accuracy of the model.

Once again, the safest procedure consists in using a non-parametric assessment, i.e. recomputing the statistic for sign-flipped datasets, in order to get the distribution of the maximum value of the brain volume under the null hypothesis. Classical inference proceeds as usual. Still there remains the question of *what* is indeed active when the null hypothesis is rejected. According to the model, the *parcel* is active, but clearly, this corresponds to different voxels according to the subject under consideration. Our approach consists in clustering the common space with a Voronoi parcellation, based only on the spatial coordinates, i.e. assigning each site to the nearest template $\bar{\tau}_k$. This set is used for representation of the active region.

Caveat. Nothing prevents a user of this procedure to use the same data for the estimation of the parcels and the computation of random-effects. This is in fact a reasonable way to use the model. In this case, the parcellation procedure itself needs to be included in the non-parametric assessment. Note that choosing a corrected statistic, i.e. the map-level null, is necessary, to obtain a meaningful calibration of the statistic.

7.3.4 Results and discussion

We have used this method on a excerpt of the Localizer dataset that comprises 14 subjects. The contrast under study shows functional activity in a reading task (as opposed to viewing a checkerboard).

The voxel-based random effects map, thresholded at the cluster level $p < 0.5$ corrected, $z > 4.80$, yields few voxels around two foci, one in the

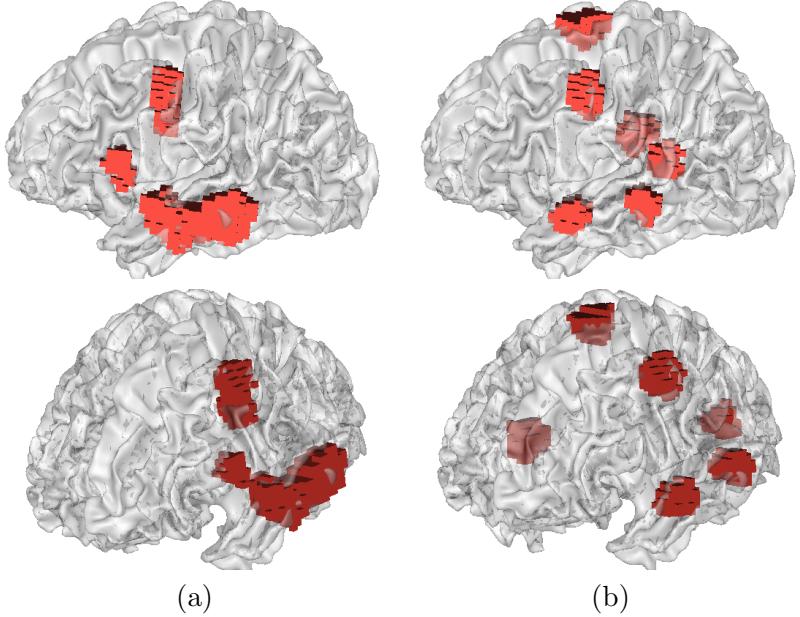


Figure 7.11: Comparison of a *naive* parcel-based RFX map (a) with the proposed one (b), both thresholded at ($p < 0.05$), corrected. The supra-threshold parcels have been meshed and are represented on a typical grey-white matter interface. The solution (a) finds activity along the left temporal lobe, in left Broca's are and in the left precentral gyrus. The solution (b) outlines a much dispersed network, with three distinct foci along the temporal lobe, the precentral gyrus and SMA. It also finds one significantly active region in the right temporal lobe, but misses left Broca's area.

prefrontal cortex $z = 5.18(-48, -3, 42)\text{mm}$ and one in the posterior part of the temporal cortex $z = 4.99, (-63, -36, -3)\text{mm}$.

We compare this with the parcel-based random effects map obtained from the procedure described above (where learning and test are performed on the same data) with 400 resamplings, and with a trivial parcellation which is simply a spatial partition unrelated to any functional information. The threshold is $p < 0.05$, corrected, which corresponds to z values of $z > 4.02$ and $z > 3.72$ respectively. However, the informed parcellation is more sensitive than the other one, as can be seen in figure 7.11.

Essentially, both parcellations provide significantly active parcels in the loci found by the voxel-level test. But both find a richer network. Part of this gain in sensitivity is clearly explained by the reduction of the multiple comparison problem associated with the resolution decrease. Using the functional data itself to drive the parcellation seems to segregate the active

set of regions into a finer grain pattern (see the temporal lobe in the left hemisphere), and to be effective in outlining isolated regions with significant activity (in the SMA and the right temporal region). On the other hand, this procedure does not detect Broca's area.

Chapter 8

Localizing cortical activity from MEEG

Source localization from external EEG or MEG measurements is an ill-posed inverse problem. Ill-posedness in the sense of Hadamard can have three causes: non-existence, non-uniqueness, and non-continuity with respect to a change in the data. Uniqueness of the source distribution is a well-known problem: there exist “silent sources” of EEG and MEG (see section 3.2.2). However, there are several uniqueness results for the inverse source problem, in the case ideal case of perfect measurements:

- If the source distribution \mathbf{J}^P is modelled as a linear combination of isolated dipoles,

$$\mathbf{J}^P = \sum_{k=1}^m \vec{q}_k \delta_{\mathbf{r}_k}(\mathbf{r})$$

then m , \vec{q}_k and \mathbf{r}_k are uniquely determined by the electric potential V on the whole outer surface (the scalp) [7].

- If \mathbf{J}^P is modelled as a surfacic distribution, and normal to a prescribed surface S : $\mathbf{J}^P = q\mathbf{n}$ then the amplitude of the source distribution, q , is uniquely determined, *up to a constant*, by perfect measurements of the electric potential V on the whole scalp.

Even if the uniqueness of the source distribution giving rise to external measurements can be proved in some cases, in practise, the number of sensors is limited, their response is not perfect, their positions are not perfectly known. The geometrical model is not exact either: the conductivity values are only approximately known, and the sources do not necessarily correspond to the model. Therefore, in practise, regularization is compulsory.

The principle behind all source reconstruction methods is to minimize a norm of the difference between measured and simulated data:

$$C(\mathbf{J}^P) = \left\| \begin{pmatrix} V_{\text{meas}} \\ \mathbf{B}_{\text{meas}} \end{pmatrix} - \begin{pmatrix} V(\mathbf{J}^P) \\ \mathbf{B}(\mathbf{J}^P) \end{pmatrix} \right\| .$$

The relationship between the simulated data (V, \mathbf{B}) and the sources \mathbf{J}^P may either be explicit or implicit. The explicit relationship consists in computing the so-called *leadfield matrix*, which relates the sources, discretized in the head volume, to the electromagnetic field at the sensor positions. With a Boundary Element or Finite Element Method, a leadfield matrix computation requires to invert the forward problem matrix. In the case of an implicit relationship, the Poisson equation and Biot-Savart relation are considered as constraints, and incorporated via the technique of Lagrange multipliers (see Appendix B.1, and section 4).

The sources can be modeled as the superposition of a small number of isolated dipoles, whose number, positions, orientations and amplitudes must be estimated, or as a spatial distribution of dipole parameters over a given region (distributed source) which is discretized over a grid or a mesh.

In the case of isolated dipoles, several models can be chosen to constrain the search. In particular, the spatio-temporal behavior of the dipoles fall into one of three categories:

- *moving dipole*: dipole positions and orientations are both free to vary over time,
- *rotating dipole*: positions are fixed but moment orientations are allowed to vary,
- *fixed dipole*: positions and orientations are both constrained to be fixed.

A difficulty of dipole-based methods is to find the right number of dipoles: an under-estimation will lead to a localization bias, whereas an over-estimation will lead to superfluous dipoles, indistinguishable from the true dipoles, and modify the localizations of the true dipoles.

After presenting the basic, pseudoinverse solution to mean-square minimization in Section 8.1, non-linear dipole fitting methods are presented in Section 8.2. Section 8.3 deals with MUSIC and Beamforming, two scanning methods, which do not rely on an assumption on the number of dipoles. In Section 8.4, we deal with reconstruction methods for distributed sources, where the number of unknowns is far greater than the number of measurements.

8.1 Pseudoinverse solution

The leadfield matrix \mathbf{G} relates the source amplitudes to the (ideal) sensor measurements. In presence of additive measurement noise,

$$\mathbf{M} = \mathbf{GJ} + \mathbf{N}. \quad (8.1)$$

Let us make dimensions more explicit: suppose that the number of sensors (both EEG and MEG) equals m , the number of time-points equals n and the source space dimension equals p . Then (8.1) becomes:

$$\begin{pmatrix} M_1(t_1) & \dots & M_1(t_n) \\ \vdots & \ddots & \vdots \\ M_m(t_1) & \dots & M_m(t_n) \end{pmatrix} = \begin{pmatrix} G_{11} & \dots & G_{1p} \\ \vdots & \ddots & \vdots \\ G_{m1} & \dots & G_{mp} \end{pmatrix} \begin{pmatrix} j_1(t_1) & \dots & j_1(t_n) \\ \vdots & \ddots & \vdots \\ j_p(t_1) & \dots & j_p(t_n) \end{pmatrix} + \mathbf{N},$$

and, denoting as $\mathbf{J}(t_i)$ the column vector $(j_1(t_i) \dots j_p(t_i))^T$ and $\mathbf{M}(t_i)$ the column vector $(M_1(t_i) \dots M_p(t_i))^T$,

$$(\mathbf{M}(t_1), \dots, \mathbf{M}(t_n)) = \mathbf{G} (\mathbf{J}(t_1), \dots, \mathbf{J}(t_n)) + \mathbf{N}$$

The minimum norm solution seeks a source matrix J which minimizes the Frobenius norm

$$\begin{aligned} \|\mathbf{M} - \mathbf{GJ}\|_F^2 &= \text{Tr} ((\mathbf{M} - \mathbf{GJ})^T (\mathbf{M} - \mathbf{GJ})) \\ &= \sum_i \|\mathbf{M}(t_i) - \mathbf{GJ}(t_i)\|^2 \end{aligned}$$

This problem is equivalent to finding the set $\{J(t_i), i = 1 \dots n\}$ which, for each i , minimize

$$\|\mathbf{M}(t_i) - \mathbf{GJ}(t_i)\|^2 = (\mathbf{M}(t_i) - \mathbf{GJ}(t_i))^T ((\mathbf{M}(t_i) - \mathbf{GJ}(t_i)))$$

A simple computation of the gradient with respect to \mathbf{J} shows that a minimizer $\hat{\mathbf{J}}$ of $\|\mathbf{M} - \mathbf{GJ}\|^2$ must satisfy

$$\mathbf{G}^T \mathbf{G} \hat{\mathbf{J}} = \mathbf{G}^T \mathbf{M}. \quad (8.2)$$

If \mathbf{G} has rank p , then $\mathbf{G}^T \mathbf{G}$ is invertible and

$$\hat{\mathbf{J}} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{M}.$$

In the case of distributed sources, the source space dimension is generally very large (typically 10^4), while the number of sensors is limited to a few hundred. Hence, \mathbf{G} is rank-deficient, and a minimizer must be selected among the solutions of (8.2). The ‘‘Moore-Penrose pseudoinverse’’ selects the minimizer whose norm is minimum, by imposing $\hat{\mathbf{J}}$ to be orthogonal to $\text{Ker}(\mathbf{G}^T \mathbf{G})$. It is denoted

$$\hat{\mathbf{J}} = (\mathbf{G}^T \mathbf{G})^\dagger \mathbf{G}^T \mathbf{M}. \quad (8.3)$$

and is computed by Singular Value Decomposition (SVD), as explained in Appendix B.2.

8.2 Dipole fitting

Let us assume a limited number of dipolar sources with fixed position over a time interval. For p dipoles, with position \mathbf{p}_i and moment $\mathbf{q}_i = q_i \mathbf{u}_i$, the relationship between sources and measurements is of the form

$$\mathbf{M}(t) = \sum_{i=1}^p g(\mathbf{p}_i, \mathbf{u}_i) q_i(t) + \mathbf{N}(t),$$

or equivalently

$$\mathbf{M}(t) = \mathbf{G} \mathbf{Q}(t) + \mathbf{N}(t),$$

where in this case the leadfield matrix \mathbf{G} is

$$\mathbf{G} = \begin{pmatrix} g_1(\mathbf{p}_1, \mathbf{u}_1) & \dots & g_1(\mathbf{p}_p, \mathbf{u}_p) \\ \vdots & \ddots & \vdots \\ g_m(\mathbf{p}_1, \mathbf{u}_1) & \dots & g_m(\mathbf{p}_p, \mathbf{u}_p) \end{pmatrix}$$

It is possible to separate the direction parameter from the position parameter, by decomposing the leadfield over the unit basis: if the coordinates of \mathbf{u}_i in the unit basis are (u_i^x, u_i^y, u_i^z) then $g_j(\mathbf{p}_i, \mathbf{u}_i) = g_j^x(\mathbf{p}_i) u_i^x + g_j^y(\mathbf{p}_i) u_i^y + g_j^z(\mathbf{p}_i) u_i^z$ and the product $\mathbf{G} Q$ can be rewritten:

$$\begin{pmatrix} g_1^x(\mathbf{p}_1) & g_1^y(\mathbf{p}_1) & g_1^z(\mathbf{p}_1) & \dots & g_1^x(\mathbf{p}_p) & g_1^y(\mathbf{p}_p) & g_1^z(\mathbf{p}_p) \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ g_m^x(\mathbf{p}_1) & g_m^y(\mathbf{p}_1) & g_m^z(\mathbf{p}_1) & \dots & g_m^x(\mathbf{p}_p) & g_m^y(\mathbf{p}_p) & g_m^z(\mathbf{p}_p) \end{pmatrix} \begin{pmatrix} u_1^x q_1 \\ u_1^y q_1 \\ u_1^z q_1 \\ \vdots \\ u_p^x q_p \\ u_p^y q_p \\ u_p^z q_p \end{pmatrix}$$

The resulting leadfield matrix \mathbf{G} thus only depends on the position parameter.

Note: for magnetoencephalography (MEG) in a spherical head model, the measurements do not depend on the radial component of the source, therefore only the two tangential directions need to be computed at each position. The resulting leadfield matrix \mathbf{G} has dimension $m \times 2p$ instead of $m \times 3p$ in the general case.

We suppose that

- the number of dipoles' parameters p is smaller than the number of sensors m ;
- the leadfield matrix \mathbf{G} has full rank, thus $\mathbf{G}^T \mathbf{G}$ is invertible.

In section 8.1, we have presented the pseudoinverse solution of $\|\mathbf{M} - \mathbf{GQ}\|$:

$$\hat{\mathbf{Q}} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{M}. \quad (8.4)$$

The amplitude term \mathbf{Q} is thus linearly related to the measurements, and easy to compute, once the matrix \mathbf{G} is known.

8.3 Scanning methods

The methods presented in this section are called scanning methods because they assume that the activated dipole(s) are located at a predefined set of positions. For example, the considered locations may belong to a regular sampling of the cortex region. The dipole orientations are either assumed to be known or can be left unconstrained. The scanning methods attempt to decide among the possible dipole locations where it is the most appropriate to place the sources so as to best describe some subset of the measurements (usually a small time window containing the events of interest). The two methods presented here rely on different assumptions on the model or on the measured signal and thus cannot be blindly used in all situations¹. They nonetheless give very stable results when the appropriate conditions are met.

Both methods presented here use a time-window of interest, which needs first to be identified. The signal is also usually prewhitened.

8.3.1 MULTiple SIgnal Classification (MUSIC)

The MUSIC method, proposed in 1992, recovers the positions by scanning potential locations, after having separated a “signal space” from a “noise space” by a Principal Component Analysis (PCA) decomposition [105].

Assuming the sensor space to be of dimensionality m and the time window of size n , an $m \times n$ measurement matrix \mathbf{M}_k is formed for each trial k . As seen above, and assuming, for the time being, that there is no noise, we have $\mathbf{M}_k = \mathbf{GQ}_k$, where \mathbf{G} and \mathbf{Q}_k are respectively matrices of sizes $m \times r$ and $r \times n$. The number of columns of \mathbf{G} , r , is the number of parameters needed to describe the dipoles at all the locations considered to form the matrix \mathbf{G} . For example, assuming we have p positions and that only the strengths of the dipoles are to be searched for (the dipole orientations being given), then $r = p$. If orientations as well as orientations are to be recovered, then $r = 3p$. Intermediate situations can also occur. MUSIC is able to treat simultaneously various types of dipoles, but for simplicity we assume here

¹This is actually true of every localisation method.

to have only one type of dipole search. Usually, it is assumed that p , and thus r are small with respect to m and n .

MUSIC fundamentally relies on rank assumptions over the matrices \mathbf{G} and \mathbf{Q} . More precisely, it is assumed that:

- **Gain matrix assumption:** The \mathbf{G} matrix for p dipoles is full rank (i.e. of rank r).
- **Asynchronous assumption:** The correlation matrix $R_{\mathbf{Q}} = E(\mathbf{Q}_k \mathbf{Q}_k^T)$ matrix for p dipoles is full rank (i.e. of rank r).
- **Noise whiteness assumption:** The noise is considered additive and temporally and spatially zero-mean white noise with variance σ^2 . When a good noise model can be established, a prewhitening phase ensures that this is the case. Additionally the signal and noise are assumed to be uncorrelated.

Denoting by \mathbf{N}_k the $m \times n$ noise matrix at trial k , the noisy data \mathbf{M}_k is given by $\mathbf{M}_k = \mathbf{G}\mathbf{Q}_k + \mathbf{N}_k$. The auto-correlation matrix of the measurements $E(\mathbf{M}_k \mathbf{M}_k^T)$, where E denotes the expectation across trials, can be written as:

$$\begin{aligned} E(\mathbf{M}_k \mathbf{M}_k^T) &= E((\mathbf{G}\mathbf{Q}_k + \mathbf{N}_k)(\mathbf{G}\mathbf{Q}_k + \mathbf{N}_k)^T), \\ E(\mathbf{M}_k \mathbf{M}_k^T) &= E(\mathbf{G}\mathbf{Q}_k \mathbf{Q}_k^T \mathbf{G}^T) + \mathbf{G}E(\mathbf{Q}_k \mathbf{N}_k^T) + E(\mathbf{N}_k \mathbf{Q}_k^T) \mathbf{G}^T + E(\mathbf{N}_k \mathbf{N}_k^T), \\ E(\mathbf{M}_k \mathbf{M}_k^T) &= \mathbf{G}R_{\mathbf{Q}}\mathbf{G}^T + \sigma^2\mathbf{I}, \end{aligned}$$

since $E(\mathbf{N}_k \mathbf{Q}_k^T) = E(\mathbf{Q}_k \mathbf{N}_k^T) = 0$ by the assumption that signal and noise are uncorrelated. Since $\mathbf{G}R_{\mathbf{Q}}\mathbf{G}^T$ is exactly of rank r with positive eigenvalue (by the Gain matrix and the asynchronous assumptions), the r highest eigenvalues (and therefore the first r higher singular values of \mathbf{M}) correspond to the the signal part of the matrix while the $m - r$ smallest once should (in theory) be identical (equal to σ^2) and correspond to noise.

Using the best rank r approximation of $E(\mathbf{M}_k \mathbf{M}_k^T)$ thus allows the recovery of the signal space. In practise, MUSIC approximates $E(\mathbf{M}_k \mathbf{M}_k^T)$ by $E(\mathbf{M}_k)E(\mathbf{M}_k)^T$, which is equivalent to working with the averaged signals. In the following, we denote by \mathbf{U} either the eigenvectors of $\mathbf{F} = E(\mathbf{M}_k \mathbf{M}_k^T)$ in the “exact” theory or the left singular vectors of $\mathbf{F} = E(\mathbf{M}_k)$ when $E(\mathbf{M}_k \mathbf{M}_k^T)$ is approximated by the above formula. In both cases, the column vectors of \mathbf{U} are ordered in such a way that the eigenvalues or the singular values are in decreasing order. The best rank r approximation of \mathbf{F} in terms of the Frobenius norm is simply given by the first r columns of \mathbf{U} . Indeed, this corresponds to setting to zero the $m - r$ smallest eigenvalues or singular values of \mathbf{F} .

By writing $\mathbf{U} = [\mathbf{U}_r \mathbf{U}_{m-r}]$, the signal and noise spaces are thus respectively spanned by \mathbf{U}_r and \mathbf{U}_{m-r} . By construction, these spaces are orthogonal and the projectors $P_{\mathbf{U}_r} = \mathbf{U}_r \mathbf{U}_r^T$ and $P_{\mathbf{U}_r}^\perp = \mathbf{U}_{m-r} \mathbf{U}_{m-r}^T = \mathbf{I} - P_{\mathbf{U}_r}$ can be introduced.

Having identified the signal and noise spaces, the best source locations need to be found. Considering the sub-matrix \mathbf{G}_i corresponding to the columns of \mathbf{G} that are associated with a dipole at position x_i , the most appropriate columns are those that best correspond to the subspace \mathbf{U}_r . It is equivalent to say that the columns of \mathbf{G}_i should be as orthogonal as possible to \mathbf{U}_{m-r} (it is best to think in terms of angles as otherwise deep sources corresponding to \mathbf{G}_i of small norm would be favored as their contribution to \mathbf{U}_{m-r} will automatically be smaller). Since the matrix $\mathbf{U}_{m-r}^T \frac{\mathbf{G}_i}{\|\mathbf{G}_i\|}$ basically contains all the cosines of all the angles between one vector of the space spanned by \mathbf{U}_{m-r} and one vector of the space spanned by \mathbf{G}_i , minimizing the criterion:

$$C(x_i) = \frac{\|\mathbf{U}_{m-r}^T \mathbf{G}_i\|}{\|\mathbf{G}_i\|},$$

is what we are looking for as it is equivalent to minimize the sum of the squares of all these cosines. Equivalently, this criterion can be written as:

$$C(x_i) = \frac{\|\mathbf{U} \mathbf{U}_{m-r}^T \mathbf{G}_i\|}{\|\mathbf{G}_i\|} = \frac{\|\mathbf{U}_{m-r} \mathbf{U}_{m-r}^T \mathbf{G}_i\|}{\|\mathbf{G}_i\|} = \frac{\|P_{\mathbf{U}_r}^\perp \mathbf{G}_i\|}{\|\mathbf{G}_i\|},$$

as \mathbf{U} preserves the norm and as $\mathbf{U}_r \mathbf{U}_{m-r}^T = \mathbf{0}$.

In practice, the image $\frac{1}{C(x_i)}$ is formed and the traditional MUSIC method just extracts p maxima from this image as the best locations for the dipoles. This selection scheme is however not very good as the p dipoles extracted may all explain the same part of the signal. Imagine a signal built from the superposition of a superficial source and a deeper one: the above selection scheme will have the tendency to extract sources that all explain the superficial source as it tends to give the strongest signal, hiding thus the deeper source. For this reason a recursive version of MUSIC called RAP-MUSIC (Recursively APplied MUSIC) has been proposed. In this method, only a single maximum is extracted, and then the signal contribution of this source is subtracted from the measurements. The RAP-MUSIC method is then re-applied to this new measurement set.

The advantage of MUSIC over traditional dipole-fitting methods is that it only requires to scan for a single dipole position at a time. Difficulties linked with the MUSIC method are

- its inability to handle spatially distinct but simultaneous sources,

- the distinction between fixed and rotating sources (not described in this text),
- the evaluation of the rank of the gain matrix.

8.3.2 Beamforming methods

Beamforming methods are easier to implement than MUSIC, but rely on an even more restrictive assumption: that sources which are spatially distinct are temporally uncorrelated. The relation between the source amplitudes $J(x_i) = [J^1(x_i) J^2(x_i) J^3(x_i)]$ and the measurements M is given by:

$$\mathbf{M} = \sum_{i=1}^p G(x_i) \cdot J(x_i) + \mathbf{N} . \quad (8.5)$$

The mean of the source amplitude is denoted $\overline{J(x_i)}$, and the covariance of the source at x_i is defined by

$$C_J(x_i) = E \left([J(x_i) - \overline{J(x_i)}] [J(x_i) - \overline{J(x_i)}]^T \right)$$

This covariance is a 3×3 matrix whose trace will be used to measure the strength of the source.

Some hypotheses are introduced:

- the noise \mathbf{N} is zero-mean, with covariance C_N ;
- the sources are decorrelated: if $i \neq k$, $E \left([J(x_i) - \overline{J(x_i)}] [J(x_k) - \overline{J(x_k)}]^T \right)$ is the 3×3 null matrix;
- the noise and the source amplitudes are decorrelated

Under these hypotheses, one can express the mean and covariance of the measurements as:

$$\begin{aligned} \overline{\mathbf{M}} &= E(\mathbf{M}) = \sum_{i=1}^p G(x_i) \cdot \overline{J(x_i)} \\ C_{\mathbf{M}} &= E \left([\mathbf{M} - \overline{\mathbf{M}}] [\mathbf{M} - \overline{\mathbf{M}}]^T \right) = \sum_{i=1}^p G(x_i) C_J(x_i) G(x_i)^T + C_N \end{aligned}$$

The concept behind beamforming is, for a given spatial position x_0 , to apply a spatial filtering to the measurements, which filters out sources which do not come from a small volume around x_0 . Let $W(x_0)$ be a $m \times 3$ matrix representing the spatial filter: the source amplitude in the vicinity of x_0 will be estimated by

$$S(x_0) = W(x_0)^T \mathbf{M} . \quad (8.6)$$

An ideal narrow-band filter should satisfy $W(x_0)^T G(x) = \mathbf{I}$ if $x = x_0$ and $W(x_0)^T G(x) = 0$ if not. With such constraints, in the absence of noise, the reconstructed amplitudes $S(x_0)$ would exactly match the source amplitudes $J(x_0)$. However, such an ideal filter response is impossible to achieve with a limited number of measurements m , and a different set of constraints must be sought.

Using (8.5) and (8.6), it is clear that

$$S(x_0) = \sum_{i=1}^p W(x_0)^T G(x_i) J(x_i) + W(x_0)^T \mathbf{N}$$

and taking the covariance of the above expression,

$$\begin{aligned} C_S(x_0) &= \sum_{i=1}^p W(x_0)^T G(x_i) C_J(x_i) G(x_i)^T W(x_0) + W(x_0)^T C_N W(x_0) \\ &= C_J(x_0) + \sum_{x_i \neq x_0} W(x_0)^T G(x_i) C_J(x_i) G(x_i)^T W(x_0) + W(x_0)^T C_N W(x_0) \end{aligned}$$

The estimated source strength $\text{Tr}(C_S(x_0))$ must match $\text{Tr}(C_J(x_0))$. Given the above expression, $\text{Tr}(C_S(x_0)) - \text{Tr}(C_J(x_0))$ is a positive quantity, and one must therefore seek the $W(x_0)$ which minimizes it.

Given that $\text{Tr}(C_J(x_0))$ is independent of $W(x_0)$, the beamforming method seeks $W(x_0)$ which minimizes $\text{Tr}(C_S(x_0)) = \text{Tr}(W(x_0)^T C_M W(x_0))$, under the constraint that $W(x_0)^T G(x_0) = \mathbf{I}$.

Using a Lagrange multiplier approach, let L denote a 3×3 matrix of multipliers and introduce the Lagrangian

$$\mathcal{L}(W, L) = \text{Tr}(W^T C_M W + (W^T G - \mathbf{I}) L).$$

At the saddle-point (W, L) of \mathcal{L} ,

$$W = -\frac{1}{2} C_M^{-1} G L.$$

Substituting this expression in $G^T W = \mathbf{I}$ yields

$$L = -2(G^T C_M^{-1} G)^{-1}$$

therefore

$$W = C_M^{-1} G (G^T C_M^{-1} G)^{-1}$$

and

$$W^T = (G^T C_M^{-1} G)^{-1} G^T C_M^{-1}$$

in which we can recognize an expression similar to a “pseudoinverse” of G .

Given the above expression for the spatial filter, computing the estimated source strength $\text{Tr}(C_S(x_0))$ leads to considerable simplifications: indeed,

$$\widehat{\text{Var}J(x_0)} = \text{Tr}(C_S(x_0)) = \text{Tr}\left((G(x_0)^T C_{\mathbf{M}}^{-1} G(x_0))^{-1}\right) \quad (8.7)$$

From this expression, it is clear that the spatial resolution of the reconstruction depends on the spatial resolution of the leadfield matrices $G(x_0)$, which in turn depends on the number of sensors and their positions, and on the source positions (superficial sources better resolved than deep sources).

Supposing the measurements to be only due to noise and $C_{\mathbf{M}} = C_{\mathbf{N}} = \mathbf{I}$, we have:

$$\widehat{\text{Var}J(x_0)} = \text{Tr}\left((G(x_0)^T G(x_0))^{-1}\right);$$

at positions x_0 far away from the sensors. $(G(x_0)^T G(x_0))$ only has small elements, and $\text{Tr}\left((G(x_0)^T G(x_0))^{-1}\right)$ can become unduly large, since there is no activity present in this model. To compensate for this, a renormalization is introduced:

$$\widehat{\text{Var}J(x_0)} = \frac{\text{Tr}\left((G(x_0)^T C_{\mathbf{M}}^{-1} G(x_0))^{-1}\right)}{\text{Tr}\left((G(x_0)^T C_{\mathbf{N}}^{-1} G(x_0))^{-1}\right)}.$$

In conclusion, beamformers offer a rapid exploration of the source space, under strict assumptions of decorrelation between different sources, and between the sources and the noise.

8.4 Estimating distributed activity: imaging approach

In this section, we consider the source to be distributed over a spatial region which describes the cortical mantle, its orientation fixed normal to the cortical sheet, and the source amplitude is to be estimated. This approach was first proposed by Dale and Sereno [29]. In this source model, tens of thousands degrees of freedom are typically needed to represent a source distribution on the surface of the cortex.

When the gain matrix \mathbf{G} can be computed for such source models, it is a rectangular matrix (p columns and m lines with $p \gg m$). We have seen in Section 8.1 a Moore-Penrose pseudoinverse solution to the source estimation problem (8.3). This solution unfortunately has the disadvantage of being unstable with respect to measurement noise and to model errors. This can be seen by considering the condition number of the Moore-Penrose pseudoinverse. Let σ_1 be the largest eigenvalue of $\mathbf{G}^T \mathbf{G}$, and σ_r its smallest non-null eigenvalue: the condition number of $(\mathbf{G}^T \mathbf{G})^\dagger$ is $\frac{\sigma_1}{\sigma_r}$. Inaccuracies due to noise in the measurements can potentially be amplified by this quantity.

The purpose of the regularization which is presented in the next section is to lower this condition number, in order to provide stability of the reconstructed source.

8.4.1 Tikhonov regularization

Instead of minimizing $\|\mathbf{M} - \mathbf{GJ}\|^2$, the Tikhonov-regularized problem adds a regularity term to the cost function to be minimized:

$$C_\lambda(\mathbf{J}) = \|\mathbf{M} - \mathbf{GJ}\|^2 + \lambda\|\mathbf{J}\|^2 .$$

The L^2 norm is the simplest regularization term which may be introduced, but many variants exist. To minimize $C_\lambda(\mathbf{J})$, a necessary condition is that the gradient of C_λ with respect to \mathbf{J} vanishes thus

$$(\mathbf{G}^T \mathbf{G} + \lambda \mathbf{I}) \mathbf{J}_\lambda = \mathbf{G}^T \mathbf{M} .$$

The addition of the regularization term has made the matrix invertible, and the solution to the above equation is therefore $\mathbf{J}_\lambda = (\mathbf{G}^T \mathbf{G} + \lambda \mathbf{I})^{-1} \mathbf{G}^T \mathbf{M}$. Here the condition number is $\frac{\sigma_1 + \lambda}{\sigma_r + \lambda}$ which is lower than the condition number of the pseudoinverse as soon as $\lambda > 0$.

When the number of sources is much greater than the number of sensors, it is computationally efficient to make use of the following trick (see proof in section B.2.3):

$$\begin{aligned} \mathbf{J}_\lambda &= (\mathbf{G}^T \mathbf{G} + \lambda \mathbf{I})^{-1} \mathbf{G}^T M \\ &= \mathbf{G}^T (\mathbf{G} \mathbf{G}^T + \lambda \mathbf{I})^{-1} M \end{aligned}$$

8.4.2 Selecting the regularization parameter: the L-curve

In the Tikhonov-type regularization, a parameter λ has been introduced, which must be chosen carefully: if λ is too small, the reconstruction will remain unstable, whereas if λ is too large, the measurements will no longer be properly explained. Selection of the regularization parameter can be performed with the help of the L-curve, which represents in a 2D plane, $\log \|\mathbf{M} - \mathbf{GJ}_\lambda\|$ versus $\log \|\mathbf{J}_\lambda\|$, for different values of the parameter λ . The optimal λ is the one for which the curve displays a corner [69].

Part III

Appendices

Appendix A

Formulas and mathematical lemmas

A.1 Differential operators in \mathbb{R}^3

Let (x, y, z) denote the canonical basis of \mathbb{R}^3 . The *nabla* operator is a notation representing $\nabla = \begin{pmatrix} \partial_x \\ \partial_y \\ \partial_z \end{pmatrix}$. The *gradient* of a scalar field $a(x, y, z)$ is a vector field defined by $\nabla a = \begin{pmatrix} \partial a / \partial x \\ \partial a / \partial y \\ \partial a / \partial z \end{pmatrix}$.

The *divergence* of a vector field $\mathbf{b} = \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix}$ is the scalar field denoted $\nabla \cdot \mathbf{b}$ defined by:

$$\nabla \cdot \mathbf{b} = \frac{\partial b_x}{\partial x} + \frac{\partial b_y}{\partial y} + \frac{\partial b_z}{\partial z}$$

The divergence of the gradient is called the *Laplacian* and denoted Δ :

$$\Delta a = \nabla \cdot \nabla a = \frac{\partial^2 a}{\partial x^2} + \frac{\partial^2 a}{\partial y^2} + \frac{\partial^2 a}{\partial z^2}$$

The *curl* of vector field \mathbf{b} is denoted $\nabla \times \mathbf{b}$ and defined by:

$$\nabla \times \mathbf{b} = \begin{pmatrix} \partial a_y / \partial z - \partial a_z / \partial y \\ \partial a_z / \partial x - \partial a_x / \partial z \\ \partial a_x / \partial y - \partial a_y / \partial x \end{pmatrix}.$$

A.1.1 Multiplicative properties

Product rule for the gradient

$$\nabla(a b) = a \nabla b + b \nabla a \quad (\text{A.1})$$

Product of a scalar and a vector

$$\nabla \cdot (a \mathbf{b}) = a \nabla \cdot \mathbf{b} + \mathbf{b} \cdot \nabla a \quad (\text{A.2})$$

$$\nabla \times (a \mathbf{b}) = a \nabla \times \mathbf{b} + \nabla a \times \mathbf{b} \quad (\text{A.3})$$

Vector dot product

$$\nabla(\mathbf{a} \cdot \mathbf{b}) = (\mathbf{a} \cdot \nabla)\mathbf{b} + (\mathbf{b} \cdot \nabla)\mathbf{a} + \mathbf{a} \times (\nabla \times \mathbf{b}) + \mathbf{b} \times (\nabla \times \mathbf{a}) \quad (\text{A.4})$$

Vector cross product

$$\nabla \cdot (\mathbf{a} \times \mathbf{b}) = \mathbf{b} \cdot (\nabla \times \mathbf{a}) - \mathbf{a} \cdot (\nabla \times \mathbf{b}) \quad (\text{A.5})$$

$$\nabla \times (\mathbf{a} \times \mathbf{b}) = \mathbf{a}(\nabla \cdot \mathbf{b}) - \mathbf{b}(\nabla \cdot \mathbf{a}) + (\mathbf{b} \cdot \nabla)\mathbf{a} - (\mathbf{a} \cdot \nabla)\mathbf{b} \quad (\text{A.6})$$

A.1.2 Second derivative properties

The curl of the gradient of any scalar field a is the zero vector:

$$\nabla \times \nabla a = 0 \quad (\text{A.7})$$

The divergence of the curl of any vector field \mathbf{b} is always zero:

$$\nabla \cdot (\nabla \times \mathbf{b}) = 0 \quad (\text{A.8})$$

The curl of the curl of a vector field \mathbf{b} satisfies:

$$\nabla \times \nabla \times \mathbf{b} = \nabla(\nabla \cdot \mathbf{b}) - \Delta \mathbf{b} \quad (\text{A.9})$$

where $\Delta \mathbf{b}$ is a vector field whose coordinates are the Laplacian of the coordinates of \mathbf{b} .

A.1.3 Conversion from volume to surface integrals

Let Ω be a volume in \mathbb{R}^3 and $\partial\Omega$ its boundary. Let \mathbf{n} be an outward-pointing normal field defined on the boundary. Integrating the divergence of a vector field \mathbf{b} on Ω is equivalent to summing the flux $\mathbf{b} \cdot \mathbf{n}$ over the boundary:

$$\int_{\Omega} \nabla \cdot \mathbf{b} d\mathbf{r} = \int_{\partial\Omega} \mathbf{b} \cdot \mathbf{n} ds \quad (\text{A.10})$$

The above identity is called the Stokes theorem. In particular, if $\mathbf{b} = \nabla a$, this implies that

$$\int_{\Omega} \Delta a \, d\mathbf{r} = \int_{\partial\Omega} \frac{\partial a}{\partial \mathbf{n}} \, ds . \quad (\text{A.11})$$

A similar relationship holds for the curl of a vector field \mathbf{b} , $\nabla \times \mathbf{b}$:

$$\int_{\Omega} \nabla \times \mathbf{b} \, d\mathbf{r} = \int_{\partial\Omega} \mathbf{n} \times \mathbf{b} \, ds . \quad (\text{A.12})$$

In the above surface integral, it is noteworthy that the order is $\mathbf{n} \times \mathbf{b}$ and *not* $\mathbf{b} \times \mathbf{n}$.

A.1.4 The Green function for the Laplacian in \mathbb{R}^3

The Green function G_L for the Laplacian in \mathbb{R}^3 is a solution in \mathbb{R}^3 of:

$$\Delta G_L(\mathbf{r}) = \delta_0(\mathbf{r}) ,$$

where δ_0 denotes the dirac mass positioned at the origin of the space.

Theorem A.1. *The Green function for the Laplacian in \mathbb{R}^3 with radial symmetry is:*

$$G_L(\mathbf{r}) = -\frac{1}{4\pi\|\mathbf{r}\|} .$$

Proof. To find a solution to this equation, we integrate it over a volume Ω that contains the origin of the space:

$$\int_{\Omega} \Delta G_L(\mathbf{r}) \, d\mathbf{r} = \int_{\Omega} \delta_0(\mathbf{r}) \, d\mathbf{r} = 1 .$$

Applying the Stokes theorem on the left hand side of this equation leads to:

$$\int_{\partial\Omega} \nabla G_L(\mathbf{r}) \cdot \mathbf{n} \, ds = 1$$

Let us call $r = \|\mathbf{r}\|$ and look for a solution with radial symmetry $G_L(\mathbf{r}) = u(r)$. In this case, the gradient $\nabla G_L(\mathbf{r}) \cdot \mathbf{n}$ is a constant $u_r(R)$ on each sphere of radius R . Applying the above formula to such a spherical domain Ω gives:

$$4\pi R^2 u_r(R) = 1 .$$

Thus:

$$u_r(r) = \frac{1}{4\pi r^2} ,$$

and:

$$G_L(\mathbf{r}) = u(r) = -\frac{1}{4\pi r} .$$

□

Theorem A.2. *A solution of the equation $\Delta u = f$ is given by:*

$$u(\mathbf{r}) = (G_L * f)(\mathbf{r}) = \int_{\mathbb{R}^3} G_L(\mathbf{r} - \mathbf{r}') f(\mathbf{r}') d\mathbf{r}' .$$

Proof. Verification of the above formula. By permuting the Laplacian and the integration, we get:

$$\begin{aligned} \Delta u(\mathbf{r}) &= \Delta \left(\int_{\mathbb{R}^3} G_L(\mathbf{r} - \mathbf{r}') f(\mathbf{r}') d\mathbf{r}' \right) \\ &= \int_{\mathbb{R}^3} \Delta(G_L(\mathbf{r} - \mathbf{r}') f(\mathbf{r}')) d\mathbf{r}' \\ &= \int_{\mathbb{R}^3} \delta(\mathbf{r} - \mathbf{r}') f(\mathbf{r}') d\mathbf{r}' \\ &= f(\mathbf{r}) . \end{aligned}$$

□

A.1.5 The Green formula

Theorem A.3. *The Green formula states that the following identity holds for any functions u and v*

$$\int_{\Omega} (u \Delta v - v \Delta u) d\mathbf{r} = \int_{\partial\Omega} (u \partial_{\mathbf{n}} v - v \partial_{\mathbf{n}} u) ds(\mathbf{r}) .$$

Proof. Starting from the formula:

$$\nabla \cdot \mathbf{A} \nabla \mathbf{B} = \mathbf{A} \nabla \cdot \nabla \mathbf{B} + \nabla \mathbf{A} \cdot \nabla \mathbf{B} = \mathbf{A} \Delta \mathbf{B} + \nabla \mathbf{A} \cdot \nabla \mathbf{B} ,$$

we have:

$$\begin{aligned} u \Delta v &= \nabla \cdot u \nabla v - \nabla u \cdot \nabla v , \\ v \Delta u &= \nabla \cdot v \nabla u - \nabla u \cdot \nabla v . \end{aligned}$$

Subtracting these two equations and integrating the result over the domain Ω yields:

$$\int_{\Omega} (u \Delta v - v \Delta u) d\mathbf{r} = \int_{\Omega} \nabla \cdot (u \nabla v - v \nabla u) d\mathbf{r} .$$

Finally, applying the Stockes theorem on the right hand side of the equations gives:

$$\int_{\Omega} (u \Delta v - v \Delta u) d\mathbf{r} = \int_{\partial\Omega} (u \nabla v - v \nabla u) \cdot \mathbf{n} ds(\mathbf{r}) = \int_{\partial\Omega} (u \partial_{\mathbf{n}} v - v \partial_{\mathbf{n}} u) ds(\mathbf{r}) .$$

□

A.2 Maxwell equations

A.2.1 Current density

The current density \mathbf{J} represents the current crossing a unit surface normal to \mathbf{J} . Its unit is $A \cdot m^{-2}$. The total intensity crossing an oriented surface S is

$$I = \int_S \mathbf{J} \cdot \mathbf{n} ds .$$

The electric charge conservation principle attributes the variation of charge inside a closed surface exclusively to exchanges with the outside medium. Let ρ denote the volumic charge density. For a closed surface, the orientation convention is for the normal vector to point outward. The charge conservation principle implies that, if Ω is a volume with boundary $\partial\Omega$,

$$\frac{d}{dt} \int_{\Omega} \rho d\mathbf{r} = - \int_{\partial\Omega} \mathbf{J} \cdot \mathbf{n} ds \quad (\text{A.13})$$

For a fixed volume Ω ,

$$\frac{d}{dt} \int_{\Omega} \rho d\mathbf{r} = \int_{\Omega} \frac{\partial \rho}{\partial t} d\mathbf{r} \quad (\text{A.14})$$

The Green identity implies that

$$\int_{\partial\Omega} \mathbf{J} \cdot \mathbf{n} ds = \int_{\Omega} \nabla \cdot \mathbf{J} d\mathbf{r} \quad (\text{A.15})$$

and replacing (A.14) and (A.15) in (A.13),

$$\int_{\Omega} \frac{\partial \rho}{\partial t} d\mathbf{r} = - \int_{\Omega} \nabla \cdot \mathbf{J} d\mathbf{r}$$

As this is true for any fixed volume Ω , we obtain the local charge conservation equation:

$$\nabla \cdot \mathbf{J} = - \frac{\partial \rho}{\partial t} \quad (\text{A.16})$$

A.2.2 Maxwell-Gauss equation

The electric field generated at a position M by a single charge q_i at position P_i is equal to

$$\frac{1}{4\pi\epsilon_0} q_i \frac{\overrightarrow{P_i M}}{\|P_i M\|^3}$$

where ε_0 is the electrical permittivity of the vacuum. The flow of the electric field across a surface S is defined by $\psi = \int_S \mathbf{E} \cdot \mathbf{n} ds$. The electric flow on S due to a charge q positioned at P (coordinate \mathbf{p}) is hence equal to

$$\psi = \frac{q}{4\pi\varepsilon_0} \int_S \frac{\mathbf{r} - \mathbf{p}}{\|\mathbf{r} - \mathbf{p}\|^3} \cdot \mathbf{n} ds = \frac{q}{4\pi\varepsilon_0} \Omega$$

where Ω is the solid angle spanning S from position P .

For a closed surface S , $\Omega = 0$ if P is outside S , and $\Omega = 4\pi$ if P is inside S . The electric flow generated on a surface S by a set of charges q_i is, by summation, equal to

$$\psi = \frac{1}{4\pi\varepsilon_0} \sum q_i \Omega_i$$

where Ω_i is equal to 0 (resp. 4π) if the corresponding charge is outside (resp. inside) S . This result leads to the Gauss theorem:

$$\int_{\partial\Omega} \mathbf{E} \cdot \mathbf{n} ds = \frac{Q_{int}}{\varepsilon_0} = \int_{\Omega} \frac{\rho}{\varepsilon_0} dr$$

where ρ is the (volumic) charge density. Using the Green identity, the above relation becomes

$$\int_{\Omega} \nabla \cdot \mathbf{E} dr = \int_{\Omega} \frac{\rho}{\varepsilon_0} dr ,$$

which provides, in its local version, the Maxwell-Gauss equation:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0} . \quad (\text{A.17})$$

A.2.3 Maxwell-Ampere equation

Ampere's law is first established for a time-invariant setting. Given a closed loop enclosing an open surface S , the magnetic field integrated along ∂S is proportional to the current I crossing S :

$$\int_{\partial S} \mathbf{B} \cdot \vec{dl} = \mu_0 I .$$

The coefficient μ_0 is the magnetic susceptibility of the vacuum (ε_0 and μ_0 satisfy the relation $\varepsilon_0\mu_0c^2 = 1$). Using Green's (Stokes) theorem and introducing the current density \mathbf{J} , the above relation becomes

$$\int_S \nabla \times \mathbf{B} \cdot \mathbf{n} ds = \mu_0 \int_S \mathbf{J} \cdot \mathbf{n} ds$$

As this must hold for any open surface S , this implies the local relationship:

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} . \quad (\text{A.18})$$

As a consequence, the current density must be divergence-free:

$$\nabla \cdot \mathbf{J} = 0 .$$

We have seen in A.2.1 that the charge conservation principle implies

$$\nabla \cdot \mathbf{J} = -\frac{\partial \rho}{\partial t} .$$

Using the Maxwell-Gauss equation (A.17),

$$\nabla \cdot \mathbf{J} = -\nabla \cdot \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}$$

In the time-variant case, the quantity which is divergence-free is no longer the current density, but

$$\mathbf{J} + \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} .$$

Ampere's law (A.18) must be adapted to account for the additional term $\varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}$, sometimes called "displacement current". This leads to the Maxwell-Ampere equation:

$$\nabla \times \mathbf{B} = \mu_0 \left(\mathbf{J} + \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right) \quad (\text{A.19})$$

A.2.4 Maxwell-Faraday equation

The Maxwell-Faraday equation is a structural relationship between the electric and magnetic fields. Consider the electric force

$$e(t) = \int_{\partial S} \mathbf{E} \cdot \vec{dl}$$

induced by a magnetic field \mathbf{B} on a lineic circuit enclosing an open surface S . The law of induction states that

$$e(t) = -\frac{d\phi}{dt}$$

where

$$\phi(t) = \int_S \mathbf{B} \cdot \mathbf{n} ds.$$

The Green (Stokes) theorem provides the local form of the induction theorem, called the Maxwell-Faraday equation:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (\text{A.20})$$

A.2.5 Maxwell-Gauss equation for magnetism

The last Maxwell equation is a conservation equation for the magnetic field, which basically states the absence of magnetic monopoles. In its local form, it is written as:

$$\nabla \cdot \mathbf{B} = 0 . \quad (\text{A.21})$$

The integral form of this equation being:

$$\int_{\partial\Omega} \mathbf{B} \cdot \mathbf{n} ds = 0 .$$

A.3 The Poincaré and Hölder inequalities

Lemma A.1. *Hölder inequality:* For $u \in L^2(\Omega)$ and $v \in L^2(\Omega)$, we have:

$$\left(\int_{\Omega} |uv| \, d\mathbf{r} \right)^2 \leq \int_{\Omega} u^2 \, d\mathbf{r} \int_{\Omega} v^2 \, d\mathbf{r}$$

Proof. Since $|uv| \leq \frac{1}{2} (u^2 + v^2)$, we have:

$$\int_{\Omega} |uv| \, d\mathbf{r} \leq \frac{1}{2} \left(\int_{\Omega} u^2 \, d\mathbf{r} + \int_{\Omega} v^2 \, d\mathbf{r} \right) \, d\mathbf{r}$$

Replacing u by λu in the previous formula, we have:

$$\begin{aligned} \forall \lambda > 0 \int_{\Omega} |uv| \, d\mathbf{r} &\leq \frac{1}{2} \left(\lambda \int_{\Omega} u^2 \, d\mathbf{r} + \frac{1}{\lambda} \int_{\Omega} v^2 \, d\mathbf{r} \right) \\ &\leq \min_{\lambda > 0} \frac{1}{2} \left(\lambda \int_{\Omega} u^2 \, d\mathbf{r} + \frac{1}{\lambda} \int_{\Omega} v^2 \, d\mathbf{r} \right) \end{aligned}$$

Since the minimum of $\lambda A + \frac{1}{\lambda} B$ for $\lambda > 0$ is obtained for $\lambda = \sqrt{\frac{B}{A}}$, we get:

$$\int_{\Omega} |uv| \, d\mathbf{r} \leq \sqrt{\int_{\Omega} u^2 \, d\mathbf{r}} \sqrt{\int_{\Omega} v^2 \, d\mathbf{r}}$$

□

Remark A.1. Applying the Hölder inequality for $v = 1$, we obtain:

$$\left(\int_{\Omega} |u| \, d\mathbf{r} \right)^2 \leq \text{vol}(\Omega) \int_{\Omega} u^2 \, d\mathbf{r}. \quad (\text{A.22})$$

Lemma A.2. *Poincaré inequality:* If Ω is bounded then there is a constant $C(\Omega) > 0$ such that

$$\forall w \in H_0^1(\Omega) \quad \int_{\Omega} w^2(\mathbf{r}) \, d\mathbf{r} \leq C(\Omega) \int_{\Omega} \|\nabla w(\mathbf{r})\|^2 \, d\mathbf{r}$$

Proof. The proof is established here only in the 1D case for $\Omega = [a, b]$. Since $w \in H_0^1(\Omega)$, $w(a) = w(b) = 0$. We have:

$$|w(x)| = |w(x) - w(a)| = \left| \int_a^x w'(\mathbf{r}) \, d\mathbf{r} \right| \leq \int_a^x |w'(\mathbf{r})| \, d\mathbf{r} \leq \int_a^b |w'(\mathbf{r})| \, d\mathbf{r}$$

Integrating the previous equation squared yields:

$$\int_a^b w^2(\mathbf{r}) d\mathbf{r} \leq (b-a) \left(\int_a^b |w'(\mathbf{r})| d\mathbf{r} \right)^2$$

Using Eq. A.22 for the right hand side of the previous equation, we get:

$$\int_a^b w^2(\mathbf{r}) d\mathbf{r} \leq (b-a)^2 \int_a^b w'(\mathbf{r})^2 d\mathbf{r}$$

□

A.4 Integral equalities

On a face f defined by the vertices $\mathbf{V}_k, k = 1..d + 1$, given a base function w^i for some index i , the integral $A_i = \int_f w^i(\mathbf{r}) d\mathbf{r}$ is zero (because $w^i = 0$ over f) if the index i does not correspond to one of the vertices defining f . Otherwise, without loss of generality, we assume that i corresponds to the vertex \mathbf{V}_1 .

Parameterizing the space by the affine basis defined by the vertices \mathbf{V}_k , we have $\mathbf{r} = \left(1 - \sum_{j=1}^d \lambda_j\right) \mathbf{V}_{d+1} + \sum_{j=1}^d \lambda_j \mathbf{V}_j$, where $\lambda = (\lambda_j, j = 1 \dots d)$ is the vector of affine parameters. \mathbf{r} is in the domain delimited by f iff all the coefficients in the previous formula are between 0 and 1. Furthermore, $d\mathbf{r} = |\mathbf{V}_1 \dots \mathbf{V}_{d+1}| d\lambda$ (the determinant is written with homogenous coordinates for the vectors \mathbf{V}_k).

$$\begin{aligned}
A_i &= \int_f \frac{|\mathbf{r} \cdot \mathbf{V}_2 \dots \mathbf{V}_{d+1}|}{|\mathbf{V}_1 \dots \mathbf{V}_{d+1}|} d\mathbf{r} \\
&= \int_0^1 \int_0^{1-\lambda_1} \dots \int_0^{1-\sum_{i=1}^{d-1} \lambda_i} \left| \left(1 - \sum_{j=1}^d \lambda_j\right) \mathbf{V}_{d+1} + \sum_{j=1}^d \lambda_j \mathbf{V}_j \cdot \mathbf{V}_2 \dots \mathbf{V}_{d+1} \right| d\lambda_d \dots d\lambda_1 \\
&= |\mathbf{V}_1 \dots \mathbf{V}_{d+1}| \int_0^1 \int_0^{1-\lambda_1} \dots \int_0^{1-\sum_{i=1}^{d-1} \lambda_i} \lambda_1 d\lambda_d \dots d\lambda_1 \\
&= \frac{1}{p!} |\mathbf{V}_1 \dots \mathbf{V}_{d+1}| \int_0^1 \lambda_1 \int_0^{1-\lambda_1} \dots \int_0^{1-\sum_{i=1}^{d-p-1} \lambda_i} \left(1 - \sum_{i=1}^{d-p} \lambda_i\right)^p d\lambda_{d-p} \dots d\lambda_1 \\
&= \frac{1}{(d-1)!} |\mathbf{V}_1 \dots \mathbf{V}_{d+1}| \int_0^1 \lambda_1 (1 - \lambda_1)^{d-1} d\lambda_1 \\
&= \frac{1}{(d+1)!} |\mathbf{V}_1 \dots \mathbf{V}_{d+1}| \\
&= \frac{1}{d+1} Volume(f)
\end{aligned}$$

Similarly :

$$\begin{aligned}
A_i^k &= \int_f \left(\frac{|\mathbf{r} \cdot \mathbf{V}_2 \dots \mathbf{V}_{d+1}|}{|\mathbf{V}_1 \dots \mathbf{V}_{d+1}|} \right)^k d\mathbf{r} \\
&= |\mathbf{V}_1 \dots \mathbf{V}_{d+1}| \int_0^1 \int_0^{1-\lambda_1} \dots \int_0^{1-\sum_{i=1}^{d-1} \lambda_i} \lambda_1^k d\lambda_d \dots d\lambda_1 \\
&= \frac{1}{(d-j)!} |\mathbf{V}_1 \dots \mathbf{V}_{d+1}| \int_0^1 \lambda_1^k \int \dots \int_0^{1-\sum_{i=1}^{j-1} \lambda_i} (1 - \sum_{i=1}^j \lambda_j)^{d-j} d\lambda_j \dots d\lambda_1 \\
&= \frac{1}{(d-1)!} |\mathbf{V}_1 \dots \mathbf{V}_{d+1}| \int_0^1 \lambda_1^k (1 - \lambda_1)^{d-1} d\lambda_1 \\
&= \frac{1}{(d-1)!} |\mathbf{V}_1 \dots \mathbf{V}_{d+1}| \frac{k!(d-1)!}{(d+k)!} \\
&= \frac{k!d!}{(d+k)!} Volume(f)
\end{aligned}$$

$$\begin{aligned}
B_i^{kl} &= \int_f \left(\frac{|\mathbf{r} \cdot \mathbf{V}_2 \dots \mathbf{V}_{d+1}|}{|\mathbf{V}_1 \dots \mathbf{V}_{d+1}|} \right)^k \left(\frac{|\mathbf{V}_1 \cdot \mathbf{r} \cdot \mathbf{V}_3 \dots \mathbf{V}_{d+1}|}{|\mathbf{V}_1 \dots \mathbf{V}_{d+1}|} \right)^l d\mathbf{r} \\
&= |\mathbf{V}_1 \dots \mathbf{V}_{d+1}| \int_0^1 \int_0^{1-\lambda_1} \dots \int_0^{1-\sum_{i=1}^{d-1} \lambda_i} \lambda_1^k \lambda_2^l d\lambda_d \dots d\lambda_1 \\
&= \frac{1}{(d-2)!} |\mathbf{V}_1 \dots \mathbf{V}_{d+1}| \int_0^1 \int_0^{1-\lambda_1} \lambda_1^k \lambda_2^l (1 - \lambda_1 - \lambda_2)^{d-2} d\lambda_2 d\lambda_1 \\
&= \frac{1}{(d+2)!} |\mathbf{V}_1 \dots \mathbf{V}_{d+1}| \text{ for } k = l = 1
\end{aligned}$$

Appendix B

Computational methods

B.1 Minimization under constraints: the Lagrange multiplier approach

Suppose one want to solve a constrained minimisation problem such as:

$$\mathbf{x} = \underset{x, f(x)=0}{\operatorname{argmin}} C(x), \quad (\text{B.1})$$

where x can represent a single or vectorial variable, $C(x)$ is the criterion to be minimized and $f(x) = 0$ represents a constraint on the solution x (again this constraint can be either scalar or vectorial). For simplicity, only the scalar version of the problem is developped hereafter. The Lagrange multiplier approach states that problem B.1 can be expressed equivalently as the un-constrained problem:

$$\mathbf{x} = \underset{x, \lambda}{\operatorname{argmin}} C(x) - \lambda f(x);, \quad (\text{B.2})$$

The normal equations associated to problem B.2 are:

$$\begin{cases} C'(x) - \lambda f'(x) &= 0, \\ f(x) &= 0. \end{cases}$$

which clearly shows that the constraint $f(x) = 0$ is taken into account for the solution of the minimization problem. λ is called the Lagrangian parameter. In the vectorial case, as many Lagrangian parameters as constraints must be introduced and the term $\lambda f(x)$ is replaced by a scalar product.

B.2 Singular Value Decomposition

A basic theorem of linear algebra states that any real $M \times N$ matrix \mathbf{A} with $M \geq N$ can be written as the product of an $M \times N$ column orthogonal matrix \mathbf{U} , an $N \times N$ diagonal matrix \mathbf{D} with non-negative diagonal elements (known as the *singular values*), and the transpose of an $N \times N$ orthogonal matrix \mathbf{V} [65]. In other words,

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T = \sum_{i=1}^N d_i \mathbf{U}_i \mathbf{V}_i^T, \quad (\text{B.3})$$

where d_i refers to the i -th non-negative element of a diagonal matrix \mathbf{D} and \mathbf{M}_i designates the i -th column of a matrix \mathbf{M} (applied here to the matrices \mathbf{U} and \mathbf{V}). The singular values are the square roots of the eigenvalues of the matrix $\mathbf{A}\mathbf{A}^T$ (or $\mathbf{A}^T\mathbf{A}$ since these matrices share the same non-zero eigenvalues) while the columns of \mathbf{U} and \mathbf{V} (the *singular vectors*) correspond to the eigenvectors of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ respectively. As defined in Eq.(B.3), the SVD is not unique since:

- It is invariant to arbitrary permutations of the singular values and their corresponding left and right singular vectors. Sorting the singular values (usually by decreasing magnitude order) solves this problem unless there exist equal singular values.
- Simultaneous changes in the signs of the vectors \mathbf{U}_i and \mathbf{V}_i do not have any impact on the leftmost part of Eq.(B.3). In practice, this has no impact on most numerical computations involving the SVD.

In the case where $M < N$, the above theorem can be applied to \mathbf{A}^T yielding basically the same result, with the matrices \mathbf{D} and \mathbf{U} being $M \times M$ and matrix \mathbf{V} being $M \times N$.

If \mathbf{A} is considered as a linear operator from a vector space \mathbf{E}_N of dimension N to a vector space \mathbf{E}_M of dimension M , then SVD can be interpreted as choosing specific orthogonal bases for \mathbf{E}_M (given by \mathbf{U} eventually completed if non square) and \mathbf{E}_N (given by \mathbf{V} eventually completed if non square), such that \mathbf{A} is diagonal (given by \mathbf{D}) when expressed in the coordinate frames associated with those bases.

If \mathbf{A} has null singular values (ie \mathbf{D} has null diagonal elements), then this means that \mathbf{A} is singular. Its rank R is exactly equal to the number of non-null singular values. From Eq. (B.3) is then possible to obtain a reduced form of the SVD, in which \mathbf{U}_R , \mathbf{D}_R and \mathbf{V}_R are respectively $M \times R$, $R \times R$ and $N \times R$ matrices yielding the general formula:

$$\mathbf{A} = \mathbf{U}_R \mathbf{D}_R \mathbf{V}_R^T = \sum_{i=1}^R d_{Ri} \mathbf{U}_{Ri} \mathbf{V}_{Ri}^T.$$

The matrices \mathbf{U}_R , \mathbf{D}_R , \mathbf{V}_R are obtained by taking the columns of matrices \mathbf{U} , \mathbf{D} , \mathbf{V} corresponding to the non-null elements d_i . Thus, \mathbf{U}_R and \mathbf{V}_R provide respectively orthogonal bases of the image of \mathbf{A} and of the orthogonal to the kernel of \mathbf{A} .

Standard libraries such as lapack provide efficient ways of computing the SVD of a matrix \mathbf{A} without having to rely on the matrices $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ (which is an advantage for both the numerical stability of the result and the computational burden). [65] describes the algorithms for computing such a decomposition. Usually singular values are ordered by decreasing order to remove the permutation ambiguity depicted above.

B.2.1 Moore-Penrose pseudoinverse

If the matrix \mathbf{A} is square and invertible, its inverse is very easily obtained as $\mathbf{A}^{-1} = \mathbf{V}\mathbf{D}^{-1}\mathbf{U}^T$. When some singular values are null, \mathbf{D}^{-1} does not exist, but it is still possible to define \mathbf{D}^\dagger as the diagonal matrix such that:

$$d_i^\dagger = \begin{cases} 1/d_i & d_i \neq 0 \\ 0 & d_i = 0 \end{cases} \quad (\text{B.4})$$

The $N \times M$ matrix \mathbf{A}^\dagger defined as:

$$\mathbf{A}^\dagger = \mathbf{V}\mathbf{D}^\dagger\mathbf{U}^T,$$

is defined whatever is the matrix \mathbf{A} (even for non-square and non-invertible matrices) and is called the Moore-Penrose pseudoinverse of matrix \mathbf{A} . From the consideration about the reduced SVD, it can be seen that basically the pseudo inverse behaves as a regular inverse between the sub-spaces defined by \mathbf{U}_R and \mathbf{V}_R and has the same kernel as the original matrix \mathbf{A} .

Actually, the Moore-Penrose pseudoinverse can be defined as the unique $N \times M$ matrix \mathbf{A}^\dagger that satisfies the following relations:

$$\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A} \quad (\text{B.5})$$

$$\mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger \quad (\text{B.6})$$

$$(\mathbf{A}\mathbf{A}^\dagger)^* = \mathbf{A}\mathbf{A}^\dagger \quad (\text{B.7})$$

$$(\mathbf{A}^\dagger\mathbf{A})^* = \mathbf{A}^\dagger\mathbf{A} \quad (\text{B.8})$$

where \mathbf{A}^* is the conjugate transpose of \mathbf{A} . Equation (B.5) simply states that even if $\mathbf{A}\mathbf{A}^\dagger$ is not the identity, its restriction to the image of \mathbf{A} (defined by its column vectors) is the identity. Equation (B.6) states that \mathbf{A}^\dagger is a weak inverse of \mathbf{A} for the multiplicative semigroup. Equations (B.7) and (B.8) state respectively that $\mathbf{A}\mathbf{A}^\dagger$ and $\mathbf{A}^\dagger\mathbf{A}$ are Hermitian matrices. Another property of interest for us is:

$$\mathbf{A}^\dagger = \lim_{\lambda \rightarrow 0} (\mathbf{A}^T \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^T = \lim_{\lambda \rightarrow 0} \mathbf{A}^T (\mathbf{A} \mathbf{A}^T + \lambda \mathbf{I})^{-1}.$$

Various other properties of the Moore-Penrose pseudoinverse and some proofs of the above claims can be found from
http://en.wikipedia.org/wiki/Moore-Penrose_pseudoinverse.

B.2.2 SVD and least-squares problems

Singular Value Decomposition is an important decomposition for least-squares methods because of the orthogonal properties of the matrices \mathbf{U} and \mathbf{V} .

Indeed, if one has to solve:

$$\mathbf{x} = \underset{\mathbf{x}, \|\mathbf{x}\|=1}{\operatorname{argmin}} \|\mathbf{Ax}\|^2 = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{\|\mathbf{Ax}\|^2}{\|\mathbf{x}\|^2}, \quad (\text{B.9})$$

Lagrange multiplier (see section B.1) can be used. The above minimisation is thus equivalent to solving the problem:

$$\mathbf{x} = \underset{\mathbf{x}, \lambda}{\operatorname{argmin}} \|\mathbf{Ax}\|^2 - \lambda (\|\mathbf{x}\|^2 - 1).$$

Writing the normal equations of this problem yields:

$$\begin{cases} (\mathbf{A}^T \mathbf{A} - \lambda \mathbf{I}) \mathbf{x} &= 0 \\ \|\mathbf{x}\|^2 &= 1 \end{cases}$$

Thus the solution \mathbf{x} is an eigenvector of $\mathbf{A}^T \mathbf{A}$. In such a case, the value of the criterion is precisely the corresponding eigenvalue. Since singular values and singular vectors of \mathbf{A} are precisely the eigenvalues and eigenvectors of $\mathbf{A}^T \mathbf{A}$, the above problem is thus minimized for the right singular vector corresponding to the smallest singular value of \mathbf{A} . Indeed, introducing the

SVD of \mathbf{A} yields¹:

$$\begin{aligned}
\mathbf{x} &= \underset{\mathbf{x}, \|\mathbf{x}\|=1}{\operatorname{argmin}} \|\mathbf{Ax}\|^2 \\
&= \underset{\mathbf{x}, \|\mathbf{x}\|=1}{\operatorname{argmin}} \|\mathbf{UDV}^T \mathbf{x}\|^2 \\
&= \underset{\mathbf{x}, \|\mathbf{x}\|=1}{\operatorname{argmin}} \|\mathbf{DV}^T \mathbf{x}\|^2 \text{ since } \mathbf{U} \text{ is an orthogonal matrix} \\
&= \underset{\mathbf{x}', \|\mathbf{x}'\|=1}{\operatorname{argmin}} \|\mathbf{Dx}'\|^2 \text{ with } \mathbf{x}' = \mathbf{V}^T \mathbf{x} \text{ and since } \mathbf{V}^T \text{ is an orthogonal matrix}
\end{aligned}$$

The last two transforms are true since orthogonal transforms (corresponding to orthogonal matrices) preserve the norm, which means that $\|\mathbf{Uz}\| = \|\mathbf{z}\|$ and that \mathbf{V}^T maps the unit sphere to itself. Assuming that the smallest singular value has index l , then the solution to this last problem is $\mathbf{x}' = \mathbf{e}_l$ (the vector with all zero components except at position l where the coordinate is 1). Consequently, the solution $\mathbf{x} = \mathbf{Vx}' = \mathbf{Ve}_l = \mathbf{V}_l$. The solution of the problem is thus given by \mathbf{V}_l the l -th column vector of \mathbf{V} , where the index l corresponds to the index of the smallest singular value d_l .

B.2.3 Least-squares problems: a useful formula

In this section, we will show the following property:

Theorem B.1. *For any matrix \mathbf{A} , we have:*

$$(\mathbf{A}^T \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^T = \mathbf{A}^T (\mathbf{A} \mathbf{A}^T + \lambda \mathbf{I})^{-1} \quad (\text{B.10})$$

Proof. Let us introduce the SVD decomposition of matrix \mathbf{A} :

$$\mathbf{A} = \mathbf{UDV}^T.$$

Replacing this expression in the left hand side of Eq. (B.10) yields:

$$\begin{aligned}
(\mathbf{A}^T \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^T &= (\mathbf{V} \mathbf{D}^T \mathbf{D} \mathbf{V}^T + \lambda \mathbf{I})^{-1} \mathbf{A}^T \\
&= (\mathbf{V} (\mathbf{D}^T \mathbf{D} + \lambda \mathbf{I}) \mathbf{V}^T)^{-1} \mathbf{V} \mathbf{D}^T \mathbf{U}^T \\
&= \mathbf{V} (\mathbf{D}^T \mathbf{D} + \lambda \mathbf{I})^{-1} \mathbf{D}^T \mathbf{U}^T.
\end{aligned}$$

Similarly, replacing this expression in the right hand side of Eq. (B.10) yields:

¹Here, the SVD of an $M \times N$ matrix \mathbf{A} is written in such a way that the matrices \mathbf{U} , \mathbf{V} and \mathbf{D} are respectively of sizes $M \times M$, $N \times N$ and $M \times N$. This can always be done by completing \mathbf{U} and \mathbf{V} with some additional orthogonal columns and \mathbf{D} with zero columns or lines (depending on whether $M < N$ or $M > N$)

$$\begin{aligned}
\mathbf{A}^T (\mathbf{AA}^T + \lambda \mathbf{I})^{-1} &= \mathbf{A}^T (\mathbf{UDD}^T \mathbf{U}^T + \lambda \mathbf{I})^{-1} \\
&= \mathbf{V} \mathbf{D}^T \mathbf{U}^T (\mathbf{V} (\mathbf{D}\mathbf{D}^T + \lambda \mathbf{I}) \mathbf{U}^T)^{-1} \\
&= \mathbf{V} \mathbf{D}^T (\mathbf{D}\mathbf{D}^T + \lambda \mathbf{I})^{-1} \mathbf{U}^T.
\end{aligned}$$

It is easy to check that the matrices $(\mathbf{D}^T \mathbf{D} + \lambda \mathbf{I})^{-1} \mathbf{D}^T$ and $\mathbf{D}^T (\mathbf{D}\mathbf{D}^T + \lambda \mathbf{I})^{-1}$ are equal for any diagonal (non necessarily square) matrix. This matrix is the diagonal matrix (same size as \mathbf{D}^T) which diagonal elements are $\frac{\sigma_i}{\lambda + \sigma_i^2}$, where σ_i are the diagonal elements of \mathbf{D} . \square

Appendix C

Statistical Results

C.1 On the Likelihood ratio and the F statistic

In this section, we simply show that the Likelihood Ratio is equivalent to an F test in a non-hierarchical model, assuming that the noise is Gaussian i.i.d.. This is meant to justify the use of the summary statistics t, F instead of the Likelihood Ratio (LR) in the inference problems encountered in fMRI. Let us recall that the generative model for a given voxel-based time course is

$$y = X\beta + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma^2) \quad (\text{C.1})$$

Given a contrast c , the two alternative hypotheses are

$$\begin{aligned} (H_0) \quad & c^T \beta = 0 \\ (H_1) \quad & c^T \beta \neq 0 \end{aligned}$$

Note that this corresponds to an unsigned or two-sided test. Without loss of generality, this contrast can be assumed to be of the form $c = [1, \dots, 1, \dots, 0]$, where the number of 1 corresponds to the dimension of c ; in this case the design matrix can be partitioned as $X = [X_i X_r]$ and similarly β is partitioned in $\begin{bmatrix} \beta_i \\ \beta_r \end{bmatrix}$. The model C.1 becomes

$$y = X_i \beta_i + X_r \beta_r + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma^2) \quad (\text{C.2})$$

The likelihood ratio simply writes:

$$\Lambda = \frac{\sup_{\{\beta_i, \beta_r, \sigma^2\}} \mathcal{N}(y; X_i \beta_i + X_r \beta_r, \sigma^2 I)}{\sup_{\{\beta_r, \sigma^2\}} \mathcal{N}(y; X_r \beta_r, \sigma^2 I)} \quad (\text{C.3})$$

Letting $P_X = X(X^T X)^{-1} X^T$ be the projection matrix to the column space of X and similarly $P_r = X_r(X_r^T X_r)^{-1} X_r^T$, it can simply be shown that

$$\log(\Lambda) = \frac{N_t}{2} \log \frac{y^T(I - P_r)y}{y^T(I - P_X)y} \quad (\text{C.4})$$

On this same problem, the F statistic used for testing this hypothesis is given by

$$F = \frac{y^T(P_X - P_R)y}{y^T(I - P_X)y} \frac{N_t - \text{rank}(X)}{\text{rank}(X) - \text{rank}(X_i)} \quad (\text{C.5})$$

It follows that

$$F = \frac{N_t - \text{rank}(X)}{\text{rank}(X) - \text{rank}(X_i)} (\Lambda^{\frac{2}{N_t}} - 1) \quad (\text{C.6})$$

and finally,

$$\Lambda = \left(\frac{\text{rank}(X) - \text{rank}(X_i)}{N_t - \text{rank}(X)} F + 1 \right)^{\frac{N_t}{2}} \quad (\text{C.7})$$

This shows that making a test on the likelihood ratio Λ or the F statistic is equivalent in that case. This can be easily be generalized in the case of known covariance V . However, the strict equivalence is broken in the case where V is unknown [31].

While Λ can be assessed only asymptotically by using the property that $\Lambda \propto \chi_1^2$ when the number of samples goes to infinity, the F test can be assessed whatever the number of degrees of freedom.

In the case of the one-sided test, the equivalence of the LR with the t test generalizes readily the equivalence between the LR and the F test.

C.2 Bayesian derivation of the t test

Let us consider the model defined in Eq. (6.4). In a Bayesian approach, this model can be shown to yield a student posterior distribution for the effect β [52]: The likelihood model writes simply:

$$p(Y|\beta, \sigma) = \mathcal{N}(Y; X\beta, \sigma^2 V) \quad (\text{C.8})$$

Uninformative (and improper) priors are then chosen on the (β, σ) parameters:

$$p(\beta, \sigma) \propto \frac{1}{\sigma} \quad (\text{C.9})$$

Then

$$p(\beta|Y) = \int p(\beta, \sigma|Y)d\sigma \propto \int \mathcal{N}(Y; X\beta, \sigma^2 V) \frac{d\sigma}{\sigma} = \frac{1}{Q(\beta)^{N_t/2}}, \quad (\text{C.10})$$

where $Q(\beta) = (Y - X\beta)V^{-1}(Y - X\beta)$ can be rewritten using the least squares estimates in (6.5-6.6):

$$Q(\beta) = \hat{\varepsilon}^T V^{-1} \hat{\varepsilon} + (\beta - \hat{\beta})^T X^T V^{-1} X (\beta - \hat{\beta}), \quad (\text{C.11})$$

where $\hat{\varepsilon} = Y - X\hat{\beta}$. Finally, we obtain:

$$p(\beta|Y) \propto \frac{1}{\left(1 + \frac{(\beta - \hat{\beta})^T X^T V^{-1} X (\beta - \hat{\beta})}{\hat{\varepsilon}^T V^{-1} \hat{\varepsilon}}\right)^{N_t/2}} \quad (\text{C.12})$$

It follows naturally that β follows a multivariate student distribution with mean $\hat{\beta}$, variance $\frac{\hat{\varepsilon}^T V^{-1} \hat{\varepsilon}}{N_t - \text{rank}(X)} (X^T V^{-1} X)^{-1}$ and $N_t - \text{rank}(X)$ degrees of freedom. The application of contrasts yields posteriori distributions for $c^T \beta$ that correspond to the test found in Eq. (6.13).

Obviously, slightly different results would arise if informative priors were used instead. The main advantage of the Bayesian approach is that it allows an interpretation of the parameters $\hat{\beta}, \hat{\sigma}$ as a parametrization of the distribution of the true effects. For instance, the probability $p(\beta > \theta|Y)$ for thresholds θ that are different from 0 [55].

C.3 D'Agostino-Pearson normality test

Let x be a sample of size n whose normality needs to be tested. The D'Agostino-Pearson test first analyzes data to determine skewness $\sqrt{b_1}$ (to quantify the asymmetry of the distribution) and kurtosis b_2 (to quantify the shape of the distribution).

$$\sqrt{b_1} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3\right)^{\frac{3}{2}}} \quad (\text{C.13})$$

$$b_2 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^2} \quad (\text{C.14})$$

It then calculates how far each of these values differs from the value expected with a normal distribution, and computes a single p-value from the sum of the squares of these discrepancies. This test is a combination of the D'Agostino skewness test and Anscombe-Glynn kurtosis test. The test statistic is

$$K^2 = \Xi^2(\sqrt{b_1}) + \psi^2(b_2), \text{ where} \quad (\text{C.15})$$

where $\Xi^2(\sqrt{b_1})$ and $\psi^2(b_2)$ are the standard normal deviates equivalent to observing $\sqrt{b_1}$ (skewness) and b_2 (kurtosis).

$$\begin{aligned} \Xi(\sqrt{b_1}) &= \delta \log \left(\frac{Y}{\alpha} + \sqrt{\frac{Y^2}{\alpha^2} + 1} \right), \\ \alpha &= \sqrt{\frac{2}{W^2 - 1}}, \\ \delta &= \frac{1}{\sqrt{\log(W)}}, \\ W^2 &= -1 + \sqrt{2\beta_2(\sqrt{b_1}) - 1}, \\ \beta_2(\sqrt{b_1}) &= \frac{3(n^2 + 27n - 70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)}, \\ Y &= \sqrt{b_1} \sqrt{\frac{(n+1)(n+3)}{6(n-2)}} \end{aligned}$$

and

$$\begin{aligned}
\psi^2(b_2) &= \left(\left(1 - \frac{2}{9A} \right) - \sqrt[3]{\frac{1 - 2/A}{1 + \chi \sqrt{2/(A-4)}}} \right) \sqrt{\frac{9A}{2}}, \\
A &= 6 + \frac{8}{\sqrt{\beta_1(b_2)}} \left[\frac{2}{\sqrt{\beta_1(b_2)}} + \sqrt{1 + \frac{4}{\sqrt{\beta_1(b_2)}}} \right], \\
\sqrt{\beta_1(b_2)} &= \frac{6(n^2 - 5n + 2)}{(n+7)(n+9)} \sqrt{\frac{6(n+3)(n+5)}{n(n-2)(n-3)}}, \\
\chi &= \frac{1}{\sigma_b} \left(b_2 - \frac{3(n-1)}{n+1} \right), \\
\sigma_b^2 &= \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}
\end{aligned}$$

The K^2 statistic in Eq. (C.15) has approximately a chi-squared distribution, with 2 degrees of freedom when the population is normally distributed. For sample sizes $n > 8$, a normal approximation that is easily computerized is available [28].

C.4 The localizer dataset

We use a Localizer dataset, which is a short (5 minutes) experiment used as a routine since 2003 at SHFJ, Orsay and Neurospin.

This experiment is based on an event-related experimental paradigm consisting of ten conditions. Subjects were presented with a series of stimuli and were engaged in tasks such as passive viewing of horizontal or vertical checkerboards, left or right button press after audio or video instruction, computation (subtraction) after video or audio instruction and sentence processing, from the audio or visual modality. Events were randomly occurring in time (mean inter stimulus interval: 3s), with ten occurrences per event type (except button presses for which there are only five trials per session). The experimental paradigm is described with more detail in [123]. Note that contrasts of experimental conditions rely in fact on the sum of number of trials of each condition. For instance, the *left-right button press* contrast combines four experimental conditions (left/ right button press after audio/video instruction) and relies on twenty trials. Similarly, the *audio-video* and *computation-sentences* contrasts rely on sixty and forty trials respectively.

Eighty-one right-handed subjects participated in the study. The subjects gave informed consent and the protocol was approved by the local ethics committee. Functional images were acquired either on a 3T Bruker (2003-2006) scanner using an EPI sequence ($TR = 2400ms$, $TE = 60ms$, matrix size= 64×64 , $FOV = 24cm \times 24cm$), or on the 1.5 GE Signa scanner (2006-2007), or on Neurospin Siemens Tro scanner (2007-2008). Each volume consisted of n_a 3mm- or 4mm-thick axial slices without gap, where n_a varied from 26 to 40 according to the session. A session comprised 130 scans. The first four functional scans were discarded to allow the MR signal to reach steady state. Anatomical T1 images were acquired on the same scanner, with a spatial resolution of $1 \times 1 \times 1.2 mm^3$.

fMRI data processing consisted in 1) temporal Fourier interpolation to correct for between-slice timing, 2) motion estimation. For all subjects, motion estimates were smaller than 1mm and 1 degree, 3) spatial normalization of the functional images, re-interpolation to $3 \times 3 \times 3 mm^3$, and 4) smoothing (5mm FWHM). This pre-processing was performed with the SPM5 software (see e.g. [5]). First-level statistical analysis was carried out on these images using the SPM5 software, using standard high-pass filtering and AR(1) whitening. For further analysis, the voxel-based estimated effects, as well as the corresponding statistical maps for several contrasts of interest were retained.

In group studies (whatever the approach), we need to determine a global brain mask for the group by considering all the voxels that belong to at least

half of the individual brain masks defined with SPM5. It comprises approximately 60000 voxels (this is the average size of individual brain masks).

Bibliography

- [1] G. K. Aguirre, E. Zarahn, and M. D'Esposito. Empirical analysis of BOLD fMRI statistics. II. Spatially smoothed data collected under null-hypothesis and experimental conditions. *Neuroimage*, 5(3):199–212, April 1997.
- [2] G. K. Aguirre, E. Zarahn, and M. D'Esposito. The variability of human BOLD hemodynamic responses. *Neuroimage*, 7:574, 1998.
- [3] S. Allassonnière, Y. Amit, and Trouvé A. Towards a coherent statistical framework for dense deformable template estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(1):3–29, January 2007.
- [4] J. Ashburner and K.J. Friston. Nonlinear spatial normalization using basis functions. *Human Brain Mapping*, 7(4):254–66, 1999.
- [5] J. Ashburner, K.J. Friston, and W. Penny, editors. *Human Brain Function, 2nd Edition*. Academic press, 2004.
- [6] John Ashburner. A fast diffeomorphic image registration algorithm. *Neuroimage*, 38(1):95–113, October 2007.
- [7] A. El Badia and T. Ha-Duong. An inverse source problem in potential analysis. *Inverse Problems*, 16:651–663, 2000.
- [8] Sylvain Baillet and Line Garnero. A bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem. *IEEE Transactions on Biomedical Engineering*, 44(5):374–385, May 1997.
- [9] L. Baratchart, J. Leblond, F. Mandrea, and E.B. Saff. How can the meromorphic approximation help to solve some 2D inverse problems for the Laplacian? *Inverse Problems*, 15:79–90, 1999.
- [10] C.F. Beckmann and S.M. Smith. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 23(2):137–152, 2004.
- [11] T. E J Behrens, M. Jenkinson, M. D. Robson, S. M. Smith, and H. Johansen-Berg. A consistent relationship between local white matter architecture and functional specialisation in medial frontal cortex. *Neuroimage*, 30(1):220–227, Mar 2006.

- [12] Timothy E J Behrens, Mark W Woolrich, Mark E Walton, and Matthew F S Rushworth. Learning the value of information in an uncertain world. *Nat Neurosci*, 10(9):1214–1221, Sep 2007.
- [13] B. Biswal, F. Z. Yetkin, V. M. Haughton, and J. S. Hyde. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, 34:537–541, 1995.
- [14] Nathalie Boddaert, Catherine Barthélémy, J.-B. Poline, Yves Samson, Francis Brunelle, and M. Zilbovicius. Autism: functional brain mapping of exceptional calendar capacity. *Br J Psychiatry*, 187:83–86, July 2005.
- [15] M Brett, IS Johnsrude, and AM. Owen. The problem of functional localization in the human brain. *Nature Reviews Neuroscience*, 3(3):243–249, March 2002.
- [16] E Bullmore, B Horwitz, G Honey, M Brammer, S Williams, and T Sharma. How good is good enough in path analysis of fMRI data? *Neuroimage*, 11(4):289–301, April 2000.
- [17] R. B. Buxton, E. C. Wong, and Frank. L. R. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magnetic Resonance in Medicine*, 39:855–864, June 1998.
- [18] Richard B. Buxton. *Introduction to Functional Magnetic Resonance Imaging*. Cambridge University Press, 2002.
- [19] A. Cachia, J.-F. Mangin, D. Rivière, D. Papadopoulos-Orfanos, F. Kherif, I. Bloch, and J. Régis. A generic framework for parcellation of the cortical surface into gyri using geodesic Voronoi diagrams. *Medical Image Analysis*, 7(4):403–416, 2003.
- [20] A. Cachia, M.-L. Paillère-Martinot, A. Galinowski, D. Januel, R. De Beaurepaire, F. Bellivier, E. Artiges, T. Gallarda, J. Andoh, D. Bartrés-Faz, E. Duchesnay, D. Rivière, Y. Cointepas, M. Plaze, J.-F. Mangin, and J.-L. Martinot. Cortical folding abnormalities in schizophrenia patients with resistant auditory hallucinations. *Neuroimage*, 39(3):927–935, 2008.
- [21] M. Chafik, A. El Badia, and T. Ha-Duong. On some inverse EEG problems. In M. Tanaka and G. S. Dulikravich, editors, *Inverse Problem in Engineering Mechanics II*, pages 537–544. Elsevier Science Ltd, 2000.
- [22] Gary A. Christensen, Richard D. Rabbitt, and Michael I. Miller. Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing*, 5(10):1435–1447, October 1996.
- [23] Justin R Chumbley, Karl J Friston, Tom Fearn, and Stefan J Kiebel. A metropolis-hastings algorithm for dynamic causal models. *Neuroimage*, 38(3):478–487, Nov 2007.
- [24] P. Ciuciu, J.-B. Poline, G. Marrelec, J. Idier, Ch. Pallier, and H. Benali. Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment. *IEEE Transactions on Medical Imaging*, 22(10):1235–1251, October 2003.

- [25] P. Ciuciu, T. Vincent, A.-L. Fouque, and A. Roche. Improved fMRI group studies based on spatially varying non-parametric BOLD signal modeling. In *5th Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1263–1266, Paris, France, May 2008.
- [26] C. Clouchoux, O. Coulon, J.-L Anton, , J.-F. Mangin, and J. Régis. A new cortical surface parcellation model and its automatic implementation. In *Proceedings of the 9th conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, LNCS 4191, pages 193–200, Copenhagen, Denmark, October 2006. Springer.
- [27] C. Clouchoux, O. Coulon, D. Rivière, A. Cachia, J.-F. Mangin, and J. Régis. Anatomically constrained surface parameterization for cortical localization. In *MICCAI'05*, pages 344–351, 2005.
- [28] Ralph B. D'Agostino and Albert Belanger. A suggestion for using powerful and informative tests of normality. *The American Statistician*, 44(4):316–321, November 1990.
- [29] A.M. Dale and M.I. Sereno. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *Journal of Cognitive Neuroscience*, 5(2):162–176, 1993.
- [30] V. Della-Maggiore, W. Chau, P.R. Peres-Neto, and A.R. McIntosh. An empirical comparison of SPM preprocessing parameters to the analysis of fMRI data. *NeuroImage*, 17(1):19–28, 2002.
- [31] A. J. den Dekker, D. H. J. Poot, R. Bos, and J. Sijbers. Likelihood based hypothesis tests for brain activation detection from mri data distributed by colored noise: a simulation study. *IEEE Transactions on*, pages 1–1, Volume PP,&a href='xpl/tocpreprint.jsp?isnumber=4359023&punumber=42';Forthcoming& 2003. Accepted for future publication Medical Imaging.
- [32] Thomas Deneux and Olivier Faugeras. Using nonlinear models in fmri data analysis: model selection and activation detection. *Neuroimage*, 32(4):1669–1689, Oct 2006.
- [33] John E Desmond and Gary H Glover. Estimating sample size in functional MRI (fMRI) neuroimaging studies: statistical power analyses. *J Neurosci Methods*, 118(2):115–128, August 2002.
- [34] S. Dodel, J.-B. Poline, J.-L. Anton, and M. Brett. The influence of heart beat and respiration on functional connectivity networks. In *Proceedings of the 2th Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 380–383, Arlington, VA, April 2004.
- [35] G. Douaud, C. Poupon, Y. Cointepas, J.-F. Mangin, V. Gaura, N. Golestani, P. Krystkowiak, C. Verny, P. Damier, A.-C. Bachoud-Lévi, P. Hantraye, and P. Remy. Diffusion tensor imaging (DTI) in Huntington's disease patients: analyses of fractional anisotropy (FA) maps and apparent diffusion coefficient (ADC) maps. In *ISMRM Workshop on Methods for Quantitative Diffusion MRI of Human Brain, Lake Louise, Canada*, 2005.

- [36] JR Duann, TP Jung, WJ Kuo, TC Yeh, S Makeig, Hsieh JC, and Sejnowski TJ. Single-trial variability in event-related BOLD signals. *Neuroimage*, 15(4):823–35, April 2002.
- [37] J. Dubois, L. Hertz-Pannier, G. Dehaene-Lambertz, Y. Cointepas, and D. Le Bihan. Potential of DTI and fiber tracking to evaluate the organization and maturation of white matter bundles in healthy infants. In *ISMRM Workshop on Methods for Quantitative Diffusion MRI of Human Brain, Lake Louise, Canada*, 2005.
- [38] E. Duchesnay, A. Cachia, A. Roche, D. Rivière, Y. Cointepas, D. Papadopoulos-Orfanos, M. Zilbovicius, J.-L. Martinot, and J.-F. Mangin. Classification from cortical folding patterns. *IEEE Transactions on Medical Imaging*, 26(4):553–565, 2007.
- [39] Simon B Eickhoff, Stefan Heim, Karl Zilles, and Katrin Amunts. Testing anatomically specified hypotheses in functional imaging using cytoarchitectonic maps. *Neuroimage*, 32(2):570–582, Aug 2006.
- [40] David C Van Essen. A population-average, landmark- and surface-based (pals) atlas of human cerebral cortex. *Neuroimage*, 28(3):635–662, Nov 2005.
- [41] O. Faugeras, F. Clément, R. Deriche, R. Keriven, T. Papadopoulo, J. Roberts, T. Viéville, F. Devernay, J. Gomes, G. Hermosillo, P. Kornprobst, and D. Lingrand. The inverse eeg and meg problems: The adjoint space approach i: The continuous case. Technical Report 3673, INRIA, May 1999.
- [42] B. Fischl and A. M. Dale. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci USA*, 97(20):11050–5, 2000.
- [43] B. Fischl, M. I. Sereno, and A. M. Dale. Cortical surface-based analysis II: inflation, flattening and a surface-based coordinate system. *Neuroimage*, 9:195–207, 1999.
- [44] Bruce Fischl, Niranjini Rajendran, Evelina Busa, Jean Augustinack, Oliver Hinds, B. T Thomas Yeo, Hartmut Mohlberg, Katrin Amunts, and Karl Zilles. Cortical folding patterns and predicting cytoarchitecture. *Cerebral Cortex*, Dec 2007.
- [45] Bruce Fischl, André van der Kouwe, Christophe Destrieux, Eric Halgren, Florent Ségonne, David H Salat, Evelina Busa, Larry J Seidman, Jill Goldstein, David Kennedy, Verne Caviness, Nikos Makris, Bruce Rosen, and Anders M Dale. Automatically parcellating the human cerebral cortex. *Cerebral Cortex*, 14(1):11–22, January 2004.
- [46] G. Flandin, F. Kherif, X. Pennec, G. Malandain, N. Ayache, and J.-B. Poline. Improved detection sensitivity of functional MRI data using a brain parcellation technique. In *Proceedings of the 5th conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, LNCS 2488 (Part I), pages 467–474, Tokyo, Japan, September 2002. Springer.

- [47] G. Flandin, F. Kherif, X. Pennec, D. Rivière, N. Ayache, and J.-B. Poline. Parcellation of brain images with anatomical and functional constraints for fMRI data analysis. In *Proceedings of the 1st Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 907–910, Washington, DC, July 2002.
- [48] G. Flandin, W. Penny, X. Pennec, N. Ayache, and J.-B. Poline. A multisubject anatomo-functional parcellation of the brain. In *Proceedings of the 9th Annual Meeting of the Organization for Human Brain Mapping*, New York,, June 19–22 2003.
- [49] G. Flandin and W. D. Penny. Bayesian fMRI data analysis with sparse spatial basis function priors. *Neuroimage*, 34(3):1108–1125, February 2007.
- [50] J. Fodor. Let your brain alone. *London review of books*, 1999.
- [51] R. S. J. Frackowiak, K.J. Friston, C. Frith, R. Dolan, C. Price, J. Ashburner, W. Penny, and S. Zeki. *Human Brain Function, Second Edition*. Academic Press, San Diego,, 2003.
- [52] L. R. Frank, R. B. Buxton, and E. C. Wong. Probabilistic analysis of functional magnetic resonance imaging data. *Magn Reson Med*, 39(1):132–148, Jan 1998.
- [53] Ola Friman and Carl-Fredrik Westin. Resampling fmri time series. *Neuroimage*, 25(3):859–867, Apr 2005.
- [54] K. J. Friston, A. P. Holmes, and K. J. Worsley. How many subjects constitute a study? *Neuroimage*, 10(1):1–5, Jul 1999.
- [55] K. J. Friston and W. Penny. Posterior probability maps and SPMs. *Neuroimage*, 19(3):1240–1249, July 2003.
- [56] K. J. Friston, P. Rotshtein, J. J. Geng, P. Sterzer, and R. N. Henson. A critique of functional localisers. *Neuroimage*, 30(4):1077–1087, May 2006.
- [57] K.J. Friston. Statistical parametric mapping. In R.W. Thatcher, M. Hallet, T. Zeffiro, E.R. John, and M. Huerta, editors, *Functional Neuroimaging : Technical Foundations*, pages 79–93, 1994.
- [58] K.J. Friston, J. Ashburner, C.D. Frith, J.-B. Poline, J.D. Heather, and R.S.J. Frackowiak. Spatial registration and normalization of images. *Human Brain Mapping*, 3(3):165–189, 1995.
- [59] KJ Friston, L Harrison, and W Penny. Dynamic causal modelling. *Neuroimage*, 19(4):1273–302, August 2003.
- [60] K.J. Friston, W. Penny, C. Phillips, S Kiebel, G. Hinton, and J. Ashburner. Classical and bayesian inference in neuroimaging: Theory. *Neuroimage*, 16(2):465–483, 2002.
- [61] C. R. Genovese, D. C. Noll, and W. F. Eddy. Estimating test-retest reliability in functional MR imaging. I: Statistical methodology. *Magn Reson Med*, 38(3):497–507, September 1997.

- [62] Christopher R Genovese, Nicole A Lazar, and Thomas Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15(4):870–878, Apr 2002.
- [63] G. H. Glover. Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage*, 9:416–429, 1999.
- [64] P. Golland, W. E. L. Grimson, M. E. Shenton, and R. Kikinis. Detection and analysis of statistical differences in anatomical shape. *Medical Image Analysis*, 9(1):69–86, February 2005.
- [65] G.H. Golub and C.F. van Loan. *Matrix Computations*. The John Hopkins University Press, Baltimore, Maryland, second edition, 1989.
- [66] Kalanit Grill-Spector, Nicholas Knouf, and Nancy Kanwisher. The fusiform face area subserves face perception, not generic within-category identification. *Nat Neurosci*, 7(5):555–62, May 2004.
- [67] Matti Hämäläinen and Riitta Hari. Magnetoencephalographic (meg) characterization of dynamic brain activation: Basic principles and methods of data collection and source analysis. In Toga and Mazziotta [149], chapter 10, pages 227–253.
- [68] Matti Hämäläinen, Riitta Hari, Risto J. Ilmoniemi, Jukka Knuutila, and Olli V. Lounasmaa. Magnetoencephalography— theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2):413–497, April 1993.
- [69] Per Christian Hansen. *Rank-deficient and discrete ill-posed problems: Numerical aspects of linear inversion*. SIAM Monographs on Mathematical Modeling and Computation. SIAM, Philadelphia, 1998.
- [70] L. M. Harrison, W. Penny, J. Daunizeau, and K. J. Friston. Diffusion-based spatial priors for functional magnetic resonance images. *Neuroimage*, 41(2):408–423, Jun 2008.
- [71] S. Hayasaka and T.E. Nichols. Validating Cluster Size Inference: Random Field and Permutation Methods. *Neuroimage*, 20(4):2343–2356, 2003.
- [72] Ruth Heller, Yulia Golland, Rafael Malach, and Yoav Benjamini. Conjunction group analysis: an alternative to mixed/random effect analysis. *Neuroimage*, 37(4):1178–1185, Oct 2007.
- [73] P. Hellier, C. Barillot, I. Corouge, B. Gibaud, G. Le Goualher., D. L. Collins, A. Evans, G. Malandain, N. Ayache, G. E. Christensen, and H. J. Johnson. Retrospective evaluation of intersubject brain registration. *IEEE Transactions on Medical Imaging*, 22(9):1120–1130, September 2003.
- [74] M. Hollander and D.A. Wolfe. *Nonparametric statistical inference*. John Wiley & Sons, New York, USA, second edition edition, 1999.
- [75] B. Horwitz, J.M. Rumsey, and BC Donohue. Functional connectivity of the angular gyrus in normal reading and dyslexia. *Proc Natl Acad Sci*, 95(15):8939–44, June 1998.

- [76] Daniel J Jacobsen, Lars Kai Hansen, and Kristoffer Hougaard Madsen. Bayesian model comparison in nonlinear bold fmri hemodynamics. *Neural Comput*, 20(3):738–755, Mar 2008.
- [77] William James. *The Principles of Psychology*. Harvard: Cambridge, MA, 1890.
- [78] P. Jezzard and R.S. Balaban. Correction for geometric distortion in echo planar images from b_0 field variations. *Magnetic Resonance in Medicine*, 34(1):65–73, July 1995.
- [79] Heidi Johansen-Berg, Timothy E.J. Behrens, Emma Sillery, Olga Ciccarelli, Alan J. Thompson, Stephen M. Smith, and Paul M. Matthews. Functional-Anatomical Validation and Individual Variation of Diffusion Tractography-based Segmentation of the Human Thalamus. *Cerebral Cortex*, July 2004.
- [80] Eric R. Cosman Jr., John W. Fisher III, and William M. Wells III. Exact map activity detection in f mri using a glm with an ising spatial prior. In *Medical Image Computing and Computer-Assisted Intervention– MICCAI 2004*, pages 703–710, 2004.
- [81] Marcel Adam Just, Vladimir L Cherkassky, Timothy A Keller, and Nancy J Minshew. Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity. *Brain*, 127(Pt 8):1811–21, August 2004.
- [82] M. Keller, A. Roche, A. Tucholka, and B.Thirion. Dealing with spatial normalization errors in fMRI group inference using hierarchical modeling. *Statistica Sinica*, 18(4):1357–1374, 2008.
- [83] F. Kherif, J.-B. Poline, S. Mériaux, H. Benali, G. Flandin, and M. Brett. Group analysis in functional neuroimaging: selecting subjects using similarity measures. *Neuroimage*, 20(4):2197–2208, January 2004.
- [84] Ferath Kherif, Jean-Baptiste Poline, Guillaume Flandin, Habib Benali, Olivier Simon, Stanislas Dehaene, and Keith J Worsley. Multivariate model specification for fmri data. *Neuroimage*, 16(4):1068–1083, Aug 2002.
- [85] S.J. Kiebel, J.-B. Poline, K.J. Friston, A. P. Holmes, and K. Worsley. Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model. *Neuroimage*, 10(6):756–766, 1999.
- [86] Arno Klein, Jesper Andersson, Babak A Ardekani, John Ashburner, Brian Avants, Ming-Chang Chiang, Gary E Christensen, Louis Collins, Pierre Helier, Joo Hyun Song, Mark Jenkinson, Claude Lepage, Daniel Rueckert, Paul Thompson, Tom Vercauteren, Roger P Woods, J. John Mann, and Ramin V Parsey. Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration. *Neuroimage*, Jan 2009.
- [87] J.L. Lions. *Optimal control of systems governed by partial differential equations*. Springer, 1971.
- [88] M. Liou, H.-R. Su, J.-D. Lee, P. E. Cheng, Huang C.-C., and C.-H. Tsai. Bridging functional MR images and scientific inference: Reproducibility maps. *Journal of Cognitive Neuroscience*, 15(7):935–945, 2003.

- [89] Michelle Liou, Hong-Ren Su, Juin-Der Lee, John A D Aston, Arthur C Tsai, and Philip E Cheng. A method for generating reproducible evidence in fMRI studies. *Neuroimage*, Oct 2005.
- [90] G. Lohmann and D. Y. von Cramon. Automatic labelling of the human cortical surface using sulcal basins. *Medical Image Analysis*, 4(3):179–188, 2000.
- [91] E. Luders, K.L. Narr, P.M. Thompson, D.E. Rex, R.P. Woods, H. DeLuca, L. Jancke, and A.W. Toga. Gender effects on cortical thickness and the influence of scaling. *Human Brain Mapping*, 27:314–324, 2005.
- [92] Oliver Lyttelton, Maxime Boucher, Steven Robbins, and Alan Evans. An unbiased iterative group registration template for cortical surface analysis. *Neuroimage*, 34(4):1535–1544, Feb 2007.
- [93] Ranjan Maitra, Steven R Roys, and Rao P Gullapalli. Test-retest reliability estimation of functional MRI data. *mrm*, 48(1):62–70, Jul 2002.
- [94] S. Makni, J. Idier, T. Vincent, B. Thirion, G. Dehaene-Lambertz, and P. Ciuciu. A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI. *Neuroimage*, 41(3):941–969, July 2008.
- [95] Jaakko Malmivuo and Robert Plonsey. *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford University Press, 1995.
- [96] J.-F. Mangin, O. Coulon, and V. Frouin. Robust brain segmentation using histogram scale-space analysis and mathematical morphology. In W. M. Wells, A. Colchester, and S. Delp, editors, *Proceedings of the 1st conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, LNCS-1496, pages 1230–1241, MIT, Boston, October 1998. Springer.
- [97] J.-F. Mangin, V. Frouin, I. Bloch, J. Régis, and J. Lopez-Krahe. From 3D magnetic resonance images to structural representations of the cortex topography using topology preserving deformations. *Journal of Mathematical Imaging and Vision*, 5(4):297–318, 1995.
- [98] J.-F. Mangin, D. Rivière, O. Coulon, C. Poupon, A. Cachia, Y. Cointepas, J.-B. Poline, D. Le Bihan, J. Régis, and D. Papadopoulos-Orfanos. Coordinate-based versus structural approaches to brain image analysis. *Artificial Intelligence in Medicine*, 30:177–197, 2004.
- [99] A.R. McIntosh, F.L. Bookstein, J.V. Haxby, and C.L. Grady. Spatial pattern analysis of functional brain images using partial least squares. *Neuroimage*, 3:143–157, 1996.
- [100] A.R. McIntosh, C.L. Grady, L.G. Ungerleider, J.V. Haxby, S.I. Rapoport, and B. Horwitz. Network analysis of cortical visual pathways mapped with pet. *J Neurosci.*, 2:655–666, 1994.
- [101] MJ McKeown, TP Jung, S Makeig, G Brown, SS Kindermann, TW Lee, and TJ Sejnowski. Spatially independent activity patterns in functional mri data during the stroop color-naming task. *Proc Natl Acad Sci*, 95(3):803–10, February 1998.

- [102] J. W. Meyer, N. Makris, J. F. Bates, V. S. Caviness, and D. N. Kennedy. MRI-based topographic parcellation of human brain cerebral white matter. 1. technical foundations. *Neuroimage*, 9:1–17, 1999.
- [103] M.A. Mintun, B.N. Lundstrom, A.Z. Snyder, A.G. Vlassenko, G.L. Shulman, and M.E. Raichle. Blood flow and oxygen delivery to human brain during functional activity: theoretical modeling and experimental data. *Proceedings of the National Academy of Sciences*, 98(12):6859–6864, June 2001.
- [104] N. Molko, A. Cachia, D. Rivière, J.-F. Mangin, M Bruandet, D. Le Bihan, L. Cohen, and S. Dehaene. Functional and structural alterations of the intraparietal sulcus in a developmental dyscalculia of genetic origin. *Neuron*, 40:847–858, Nov 2003.
- [105] John C. Mosher, Paul S. Lewis, and Richard M. Leahy. Multiple dipole modeling and localization from spatio-temporal meg data. *IEEE Transactions on Biomedical Engineering*, 39(6):541–553, 1992.
- [106] Kevin Murphy and Hugh Garavan. An empirical investigation into the number of subjects required for an event-related fMRI study. *Neuroimage*, 22(2):879–85, Jun 2004.
- [107] S. Mériaux, A. Roche, G. Dehaene-Lambertz, and J.-B. Poline. When do mixed-effect models fail to improve detection sensitivity in fMRI group activation maps? In *Neuroimage (HBM'06)*, Florence, Italy, 2006.
- [108] S. Mériaux, A. Roche, G. Dehaene-Lambertz, B. Thirion, and J.-B. Poline. Combined permutation test and mixed-effect model for group average analysis in fMRI. *Human Brain Mapping*, 27(5):402–410, May 2006.
- [109] R. Natarajan and R. E. Kass. Reference bayesian methods for generalized linear mixed models. *Journal of the American Statistical Association*, 95(449):227–237, 2000.
- [110] Jean-Claude Nédélec. *Acoustic and Electromagnetic Equations*. Springer Verlag, 2001.
- [111] J. Neumann and G. Lohmann. Bayesian second-level analysis of functional magnetic resonance images. *Neuroimage*, 20(2):1346–1355, 2003.
- [112] A. Nieto-Castanon, S.S. Ghosh, J.A. Tourville, and F.H. Guenther. Region of interest based analysis of functional imaging data. *Neuroimage*, 19(4):1303–1316, 2003.
- [113] G. Nolte. The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Physics in Medicine and Biology*, 48:3637–3652, October 2003.
- [114] S. Ogawa, T. Lee, A. Kay, and D. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87(24):9868–9872, 1990.
- [115] Jorma O. Ollikainen, Marko Vauhkonen, Pasi A. Karjalainen, and Jari P. Kaipio. Effects of electrode properties on eeg measurements and a related inverse problem. *Medical Engineering & Physics*, 22:535–545, October 2000.

- [116] Wanmei Ou and Polina Golland. From spatial regularization to anatomical priors in fMRI analysis. In *IPMI, Glenwood Springs, Colorado*, July 2005.
- [117] Théodore Papadopoulo and Sylvain Vallaghé. Implicit meshing for finite element methods using levelsets. In *Proceedings of MMBIA 07*, 2007.
- [118] Scott J Peltier, Thad A Polk, and Douglas C Noll. Detecting low-frequency functional connectivity in fMRI using a self-organizing map (SOM) algorithm. *Human Brain Mapping*, 20(4):220–6, Dec 2003.
- [119] W. D. Penny, K. E. Stephan, A. Mechelli, and K. J. Friston. Comparing dynamic causal models. *Neuroimage*, 22(3):1157–1172, July 2004.
- [120] W.D. Penny and K.J. Friston. Mixtures of general linear models for functional neuroimaging. *IEEE Transactions on Medical Imaging*, 22(4):504–514, 2003.
- [121] M. Perrot, D. Rivière, and J.-F. Mangin. Identifying cortical sulci from localizations, shape and local organization. In *5th Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 420–423, Paris, France, May 2008.
- [122] Mathias Pessiglione, Liane Schmidt, Bogdan Draganski, Raffael Kalisch, Hakwan Lau, Ray J Dolan, and Chris D Frith. How the brain translates money into force: a neuroimaging study of subliminal motivation. *Science*, 316(5826):904–906, May 2007.
- [123] P. Pinel, B. Thirion, S. Mériaux, A. Jobert, J. Serres, D. Le Bihan, J.-B. Poline, and S. Dehaene. Fast reproducible identification and large-scale databasing of individual functional cognitive networks. *BMC Neurosci*, 8(1):91, Oct 2007.
- [124] J.-B. Poline, A. P. Holmes, K. Worsley, and K.J. Friston. *Making Statistical inferences, in Human Brain Function*, chapter 5, pages 85–106. Academic Press, 1997.
- [125] C. Poupon, J.-F. Mangin, C. A. Clark, V. Frouin, J. Régis, D. Le Bihan, and I. Bloch. Towards inference of the human brain connectivity from MR diffusion tensor data. *Medical Image Analysis*, 5:1–15, 2001.
- [126] F. Poupon, J.-F. Mangin, D. Hasboun, I. Magnin, and V. Frouin. Multi-object Deformable Templates Dedicated to the Segmentation of Brain Deep Structures. In *Proceedings of the 1st conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, LNCS-1496, pages 1134–1143, MIT, Boston, October 1998. Springer.
- [127] J.J. Riera, J. Watanabe, I. Kazuki, M. Naoki, E. Aubert, T. Ozaki, and R. Kawashima. A state-space model of the hemodynamic approach: nonlinear filtering of BOLD signal. *Neuroimage*, 21:547–567, 2004.
- [128] D. Rivière, J.-F. Mangin, D. Papadopoulos-Orfanos, J.-M. Martinez, V. Frouin, and J. Régis. Automatic recognition of cortical sulci of the human brain using a congregation of neural networks. *Medical Image Analysis*, 6(2):77–92, 2002.

- [129] A. Roche, S. Mériaux, M. Keller, and B. Thirion. Mixed-effects statistics for group analysis in fMRI: A nonparametric maximum likelihood approach. *Neuroimage*, 38:501–510, 2007.
- [130] P. E. Roland, S. Geyer, K. Amunts, T. Schormann, A. Schleicher, A. Malikovic, and K. Zilles. Cytoarchitectural maps of the human brain in standard anatomical space. *Human Brain Mapping*, 5(4):222–227, 1997.
- [131] S. Sandor and R. Leahy. Surface-based labeling of cortical anatomy using a deformable atlas. *IEEE Transactions on Medical Imaging*, 16(1):41–54, February 1997.
- [132] Rebecca Saxe, Matthew Brett, and Nancy Kanwisher. Divide and conquer: A defense of functional localizers. *Neuroimage*, 30(4):1088–1096, May 2006.
- [133] P. H. Schimpf. Application of quasi-static magnetic reciprocity to finite element models of the meg lead-field. *IEEE Transactions on Biomedical Engineering*, 54(11):2082–2088, November 2007.
- [134] Armin Schwartzman, Robert F Dougherty, Jongho Lee, Dara Ghahremani, and Jonathan E Taylor. Empirical null and false discovery rate analysis in neuroimaging. *Neuroimage*, 44(1):71–82, Jan 2009.
- [135] David W Shattuck, Mubeena Mirza, Vitria Adisetiyo, Cornelius Hojatkashani, Georges Salamon, Katherine L Narr, Russell A Poldrack, Robert M Bilder, and Arthur W Toga. Construction of a 3d probabilistic atlas of human cortical structures. *Neuroimage*, 39(3):1064–1080, Feb 2008.
- [136] Olivier Simon, F. Kherif, G. Flandin, J.-B. Poline, D. Rivière, J.-F. Mangin, D. Le Bihan, and S. Dehaene. Automatized clustering and functional geometry of human parietofrontal networks for language, space, and number. *Neuroimage*, 23(3):1192–1202, 11 2004.
- [137] Stephen M Smith and Thomas E Nichols. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*, 44(1):83–98, Jan 2009.
- [138] Klaas Enno Stephan, Nikolaus Weiskopf, Peter M. Drysdale, Peter A. Robinson, and Karl J. Friston. Comparing hemodynamic models with DCM. *Neuroimage*, 38(3):387–401, November 2007.
- [139] Peter Stiers, Ronald Peeters, Lieven Lagae, Paul Van Hecke, and Stefan Sunaert. Mapping multiple visual areas in the human brain with a short fMRI sequence. *Neuroimage*, 29(1):74–89, Jan 2006.
- [140] J. Talairach and P. Tournoux. *Co-Planar Stereotaxic Atlas of the Human Brain. 3-Dimensional Proportional System : An Approach to Cerebral Imaging*. Thieme Medical Publishers, Inc., Georg Thieme Verlag, Stuttgart, New York, 1988.
- [141] Xiaodong Tao, Xiao Han, Maryam E. Rettmann, Jerry L. Prince, and Christos Davatzikos. Statistical study on cortical sulci of human brains. In *IPMI*, pages 475–487, 2001.

- [142] B. Thirion, S. Dodel, and J.-B. Poline. Detection of signal synchronizations in resting-state fMRI datasets. *Neuroimage*, 29(1):321–327, January 2006.
- [143] B. Thirion, G. Flandin, P. Pinel, A. Roche, P. Ciuciu, and J.-B. Poline. Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets. *Human Brain Mapping*, 27(8):678–693, August 2006.
- [144] Bertrand Thirion and Olivier Faugeras. Dynamical components analysis of fMRI data through kernel PCA. *Neuroimage*, 20(1):34–49, September 2003.
- [145] Bertrand Thirion, Philippe Pinel, Sébastien Mériaux, Alexis Roche, Stanislas Dehaene, and Jean-Baptiste Poline. Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *Neuroimage*, 35(1):105–120, Mar 2007.
- [146] P.M. Thompson, K.M. Hayashi, G. de Zubicaray, A.L. Janke, S.E. Rose, J. Semple, D. Herman, M.S. Hong, S.S. Dittmer, and A.W. Doddrell, D.M. Toga. Dynamics of gray matter loss in alzheimer’s disease. *Journal of neuroscience*, 23(3):994–1005, 2003.
- [147] PM Thompson, C Schwartz, and AW Toga. High-resolution random mesh algorithms for creating a probabilistic 3D surface atlas of the human brain. *Neuroimage*, 3(1):19–34, Feb 1996.
- [148] G. Thut, J.R. Ives, F. Kampmann, M. Pastor, and A. Pascual-Leone. A device and protocol for combining TMS and online recordings of EEG/evoked potentials. *J. Neurosci. Methods*, 2005.
- [149] Arthur W. Toga and John C. Mazziotta, editors. *Brain Mapping: The Methods*. Academic Press, 2nd edition edition, 2002.
- [150] R. B. Tootell, D. Tsao, and W. Vanduffel. Neuroimaging weighs in: Humans meet macaques in “primate” visual cortex. *J. Neurosci.*, 23:3981–3989, 2003.
- [151] R. Toro. *ontogenesis of the cortical organisation: modelling and analysis*. PhD thesis, Paris VI, 2003.
- [152] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, 15(1):273–89, Jan 2002.
- [153] Sylvain Vallaghé, Théodore Papadopulo, and Maureen Clerc. The adjoint method for general eeg and meg sensor-based lead field equations. *Physics in Medicine and Biology*, 54:135–147, 2009.
- [154] Dimitri Van De Ville, Thierry Blu, and Michael Unser. Integrated wavelet processing and spatial statistical testing of fMRI data. *Neuroimage*, 23(4):1472–1485, December 2004.
- [155] Dimitri Van De Ville, Mohamed Seghier, François Lazeyras, Thierry Blu, and Michael Unser. Integrated wavelet processing and spatial statistical testing of fMRI data. *Neuroimage*, 37(4):1205–1217, October 2007.

- [156] T. Vincent, P. Ciuciu, and J. Idier. Spatial mixture modelling for the joint detection-estimation of brain activity in fMRI. In *32th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume I, pages 325–328, Honolulu, Hawaii, April 2007.
- [157] T.D. Wager, M.C. Keller, S.C. Lacey, and J. Jonides. Increased sensitivity in neuroimaging analyses using robust regression. *Neuroimage*, 26(1):99–113, 2005.
- [158] Yalin Wang, Xianfeng Gu, Kiralee M Hayashi, Tony F Chan, Paul M Thompson, and Shing-Tung Yau. Brain surface parameterization using riemann surface structure. *Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv*, 8(Pt 2):657–665, 2005.
- [159] Xingchang Wei, Seung-Schik Yoo, Chandlee C Dickey, Kelly H Zou, Charles R G Guttmann, and Lawrence P Panych. Functional MRI of auditory verbal working memory: long-term reproducibility analysis. *Neuroimage*, 21(3):1000–8, Mar 2004.
- [160] D. Weinstein, L. Zhukov, and C. Johnson. Lead-field bases for electroencephalography source imaging. *Annals of Biomedical Engineering*, 28(9):1059–1164, September 2000.
- [161] C. H. Wolters, L. Grasedyck, and W. Hackbusch. Efficient computation of lead field bases and influence matrix for the fem-based eeg and meg inverse problem. *Inverse Problems*, 20:1099–1116, 2004.
- [162] M. Woolrich and T. Behrens. Variational Bayes inference of spatial mixture models for segmentation. *IEEE Transactions on Medical Imaging*, 25(10):1380–1391, October 2006.
- [163] M. Woolrich, T. Behrens, Ch. Beckmann, and S. Smith. Mixture models with adaptive spatial regularization for segmentation with an application to fMRI data. *IEEE Transactions on Medical Imaging*, 24(1):1–11, January 2005.
- [164] M. Woolrich, M. Jenkinson, J. Brady, and S. Smith. Fully Bayesian spatio-temporal modelling of fMRI data. *IEEE Transactions on Medical Imaging*, 23(2):213–231, February 2004.
- [165] M. Woolrich, B. Ripley, M. Brady, and S. Smith. Temporal autocorrelation in univariate linear modelling of fMRI data. *Neuroimage*, 14(6):1370–1386, December 2001.
- [166] K. J. Worsley. An improved theoretical P value for SPMs based on discrete local maxima. *Neuroimage*, 28(4):1056–1062, December 2005.
- [167] K.J. Worsley, C.H. Liao, J. Aston, V. Petre, G.H. Duncan, F. Morales, and A.C. Evans. A general statistical analysis for fMRI data. *Neuroimage*, 15(1):1–15, January 2002.
- [168] Jerrold H. Zar. *Biostatistical Analysis*. Prentice-Hall, Inc., Upper Saddle River, NJ., 1999.

Index

- action potentials, 19, 20
- adjoint problem, 41, 42
- balloon model, 63
- Biot and Savart law, 25
- BOLD effect, 62
- boundary elements, 29
- conductivity, 23
- cortical column, 7
- current
 - ohmic, 25
 - primary, 25
 - return, 25
- current density, 23, 120
- DBS, 11
- Deep Brain Stimulation, 11
- deoxy-haemoglobin, 62
- echo-planar image, 63
- EEG, 11
 - leadfield, 43
- Electroencephalography, 11
- EPI, 63
- False Discovery Rate (FDR), 68
- finite elements, 33
- fMRI, 11, 61
- forward model
 - realistic, 33
 - semi-realistic, 29
- forward problem, 27
- functional biomarkers, 9
- functional connectivity, 68
- functional Magnetic Resonance Imaging, 11
 - gain matrix, 113
 - Gaussian Random Field (GRF), 69
- General linear model (GLM), 63
- Generalized Linear Model, 60
- GLM, 63
 - parameter estimation, 65
- Green
 - function for Laplacian, 118
 - formula, 119
- haemodynamic activity, 62
- haemodynamic response function, 63
- harmonic function, 26
- Helmholtz
 - reciprocity principle, 42
- HRF, 63
- hypothesis testing, 65
- ill-posedness, 105
- inverse problem
 - beamforming, 111
 - dipole fitting, 107
 - imaging approach, 113
 - minimum norm solution, 106
 - MUSIC, 109
 - scanning methods, 108, 111
 - source localization, 105
 - Tikhonov regularization, 113
 - uniqueness, 105
- Lagrange multiplier, 127
- Lagrangian, 127
- Laplacian
 - fundamental solution, 118
 - Green function, 118
- leadfield, 41
 - EEG, 43
 - matrix, 106
 - MEG, 44
- Magnetic Resonance Imaging, 9, 49
 - functional, 11, 61

Magnetoencephalography, 11
Markov Random Fields, 70
Maxwell equations, 23, 120
 Maxwell-Ampere, 121
 Maxwell-Faraday, 122
 Maxwell-Gauss, 120
MEG, 11
 leadfield, 44
Moore-Penrose pseudoinverse, 107, 129
MRI, 9, 49
 functional, 57
MUSIC method, 109

neurons, 7, 10
Nuclear Magnetic Resonance, 49

OI, 9
Optical Imaging, 9
oxy-haemoglobin, 62

permittivity, 23
PET, 9, 11
Poisson equation, 25
Positron Emitted Tomography, 9, 11
postsynaptic potentials, 19, 20
Preprocessing fMRI data, 58

quasistatic approximation, 23

Radio Frequency pulse, 49
reciprocity principle, 42
regularization
 L-curve, 114
 Tikhonov, 113

silent sources, 28
Singular Value Decomposition (SVD), 107,
 128
source models, 20
SQUID, 10

TMS, 11
Transcranial Magnetic Stimulation, 11