

Linköping Studies in Science and Technology
Dissertation No. 2307

Modern multimodal methods in brain MRI

David Abramian

Linköping Studies in Science and Technology
Dissertations, No. 2307

Modern multimodal methods in brain MRI

David Abramian



Linköping University
Department of Biomedical Engineering
SE-581 83 Linköping, Sweden

Linköping 2023



This work is licensed under a Creative Commons Attribution 4.0 International License.

<https://creativecommons.org/licenses/by/4.0/>

Edition 1.0

© David Abramian, 2023

ISBN ISBN 978-91-8075-135-3 (print)

ISBN ISBN 978-91-8075-136-0 (PDF)

ISSN 0345-7524

URL <https://doi.org/10.3384/9789180751360>

Published articles have been reprinted with permission from the respective copyright holder.

Typeset using X_YT_EX

Printed by LiU-Tryck, Linköping 2023

POPULÄRVETENSKAPLIG SAMMANFATTNING

Magnetresonansavbildning (MR) är en av de viktigaste metoderna inom medicinsk avbildning. Genom att mäta ändringar i de magnetiska egenskaperna hos olika vävnadstyper kan tre-dimensionella bilder av människokroppen skapas. MR är en flexibel teknik, och olika metoder kan användas för att mäta olika fysiologiska egenskaper. För studier av hjärnan är de viktigaste metoderna strukturell MR, som ger bilder med hög upplösning och bra kontrast mellan olika vävnadstyper, funktionell MR, som kan användas för att studera vilka delar av hjärnan som är aktiva under olika experiment, och diffusions MR, som främst används för att studera hur olika delar av hjärnan är sammankopplade via axoner i nervfibrer.

Den här avhandlingen presenterar flera avancerade metoder för att analysera MR bilder, och dessa metoder följer två spår. Det första spåret handlar om hur bilder från olika MR-metoder kan komplettera varandra, och hur kombinationen kan leda till bättre resultat. Det andra spåret handlar om att använda moderna verktyg inom matematik och beräkningsvetenskap, såsom artificiell intelligens, Bayesiansk statistik och signalbehandling på grafer, för att analysera MR-bilder.

Publikationerna i den här avhandlingen täcker flera olika ämnen inom hjärnavbildning. De två första artiklarna presenterar metoder för att förbättra analysen av funktionell MR, genom att använda information från strukturell MR och diffusions MR. Den tredje artikeln presenterar en metod för att lokalisera aktiva områden i hjärnan med funktionell MR, och hur denna information kan användas när en hjärntumör tas bort med hjälp av strålning (för att inte skada dessa viktiga områden). I den fjärde artikeln används artificiell intelligens för att öka den spatiala upplösningen i diffusions MR, vilket kan förbättra analysen av nervfibrernas orientering i vit hjärnsubstans. Slutligen så visar den femte artikeln att strukturella MR-bilder som har anonymiserats, genom att ta bort ansiktet, delvis kan återskapas genom att lägga tillbaka ansiktet. Detta illustrerar att artificiell intelligens har en stor potential att förbättra medicinsk avbildning och diagnostik, men att den också medför vissa faror.

ABSTRACT

Magnetic resonance imaging (MRI) is one of the pillars of modern medical imaging, providing a non-invasive means to generate 3D images of the body with high soft-tissue contrast. Furthermore, the possibilities afforded by the design of MRI sequences enable the signal to be sensitized to a multitude of physiological tissue properties, resulting in a wide variety of distinct MRI modalities for clinical and research use.

This thesis presents a number of advanced brain MRI applications, which fulfill, to differing extents, two complementary aims. On the one hand, they explore the benefits of a multimodal approach to MRI, combining structural, functional and diffusion MRI, in a variety of contexts. On the other, they emphasize the use of advanced mathematical and computational tools in the analysis of MRI data, such as deep learning, Bayesian statistics, and graph signal processing.

Paper I introduces an anatomically-adapted extension to previous work in Bayesian spatial priors for functional MRI data, where anatomical information is introduced from a T1-weighted image to compensate for the low anatomical contrast of functional MRI data.

It has been observed that the spatial correlation structure of the BOLD signal in brain white matter follows the orientation of the underlying axonal fibers. Paper II argues about the implications of this fact on the ideal shape of spatial filters for the analysis of white matter functional MRI data. By using axonal orientation information extracted from diffusion MRI, and leveraging the possibilities afforded by graph signal processing, a graph-based description of the white matter structure is introduced, which, in turn, enables the definition of spatial filters whose shape is adapted to the underlying axonal structure, and demonstrates the increased detection power resulting from their use.

One of the main clinical applications of functional MRI is functional localization of the eloquent areas of the brain prior to brain surgery. This practice is widespread for various invasive surgeries, but is less common for stereotactic radiosurgery (SRS), a non-invasive surgical procedure wherein tissue is ablated by concentrating several beams of high-energy radiation. Paper III describes an analysis and processing pipeline for functional MRI data that enables its use for functional localization and delineation of organs-at-risk for Elekta GammaKnife SRS procedures.

Paper IV presents a deep learning model for super-resolution of diffusion MRI fiber ODFs, which outperforms standard interpolation methods in estimating local axonal fiber orientations in white matter. Finally, Paper V demonstrates that some popular methods for anonymizing facial data in structural MRI volumes can be partially reversed by applying generative deep learning models, highlighting one way in which the enormous power of deep learning models can potentially be put to use for harmful purposes.

Acknowledgments

Now that this project is nearing its end I can look back on it with the benefit of perspective. It has been an eventful time, often challenging, but always engaging. Most of all, the last five years have been a true education, for which I am deeply grateful. I would like to extend my thanks to all who have made it possible, by your contributions, by your encouragement, or by being there.

First and foremost, I want to express my gratitude to my supervisor Anders Eklund, whose constant support, endless drive, encouragement, and understanding have been at the root of making this thesis possible. I also want to give big thanks to my fantastic co-supervisors Evren Özarslan, Ida Blystad, and Hamid Behjat, for your encouragement and expertise. Big thanks to all my co-authors, for your valuable contributions to the works presented here, and to my knowledge and understanding.

Special thanks to Hamid Behjat, for nurturing my interest in research and biomedical imaging, and starting me on this path. Thanks to Martin Larsson, with whom I greatly enjoyed collaborating.

I want to thank Alfredo Ordinola, Cem Yolcu, Deneb Boito, Hans Knutsson, Iulian Emil Tampu, Johan Jönemo, Magnus Herberthson, Muhammad Usman Akbar, Marco Domenico Cirillo, Neda Haj Hosseini, and Xuan Gu, for sharing the struggles and making work always pleasant and interesting. Thanks also to my teaching colleagues Martin Hultman and Anette Karlsson, from whom I have learned a lot, and to all my co-workers at IMT for your welcoming friendliness.

I gratefully acknowledge the funding received towards my PhD from CENIIT (Center for industrial information technology) and LiU Cancer, as well as the ITEA/VINNOVA-funded projects IMPACT and ASSIST. Likewise, I wish to acknowledge the research support provided by the Center for Medical Image Science and Visualization (CMIV) at Linköping University.

Thanks to Felicia, Daphne, and all the staff at Yogi. Your role in making this thesis possible cannot be overlooked!

Finally, I want to thank and dedicate this thesis to the friends and family I left abroad to pursue this work. Getting to see you again is the highlight of every year. Thanks to my mother, to whom I owe everything; to my father, who instilled my interest in all things; to my brother, for your love and affection, and for all the things we share; and to Yasmin, for the brightness you bring to my life.

Contents

Abstract	iii
Acknowledgments	v
Contents	vii
List of Figures	ix
1 Introduction	1
1.1 Thesis outline	2
1.2 List of publications	2
1.3 Additional publications	3
1.4 Abbreviations	4
1.5 Data, code, and ethics	5
2 Magnetic Resonance Imaging	7
2.1 The MRI scanner	7
2.2 Nuclear magnetism	8
2.3 MR signals	10
2.4 Image formation	11
3 Functional Magnetic Resonance Imaging	15
3.1 Origins of fMRI: the BOLD signal	15
3.2 Design of fMRI experiments	17
3.3 Preprocessing of fMRI data	18
3.4 Statistical analysis of fMRI data	20
4 Diffusion Magnetic Resonance Imaging	25
4.1 The diffusion process	25
4.2 Acquisition of dMRI	26
4.3 Reconstruction models	27
4.4 Tractography	31
5 Graph Signal Processing	33

5.1	Introduction to graphs	33
5.2	Graph signals	34
5.3	Graph spectral domain	35
6	Deep Learning	41
6.1	Machine learning	41
6.2	Neural networks	43
6.3	Convolutional neural networks	48
6.4	Generative adversarial networks	50
7	Summary of papers	53
7.1	Paper I - Anatomically informed Bayesian spatial priors for fMRI analysis	53
7.2	Paper II - Diffusion-informed spatial smoothing of fMRI data in white matter using spectral graph filters	53
7.3	Paper III - Evaluation of inverse treatment planning for Gamma Knife radiosurgery using fMRI brain activation maps as organs at risk	54
7.4	Paper IV - Super-resolution mapping of anisotropic tissue structure with diffusion MRI and deep learning	55
7.5	Paper V - Refacing: reconstructing anonymized facial features using GANs	55
	Bibliography	57
	Paper I	65
	Paper II	73
	Paper III	93
	Paper IV	127
	Paper V	139

List of Figures

2.1	Magnetized spin precession	9
2.2	Sampling of k -space and image reconstruction	13
3.1	fMRI experiment	17
3.2	fMRI activation mapping with voxel-level inference	23
4.1	Comparison of dMRI reconstruction methods	29
4.2	Deterministic tractography streamlines	32
5.1	Weighted undirected graph	34
5.2	Graph Laplacian eigenvectors	36
6.1	Two layer neural network	45
6.2	Convolutional neural network for classification	50
6.3	Synthetic T_1 -weighted brain volumes generated with a 3D GAN	51
6.4	T_1 - T_2 image translation with CycleGAN	52

1

Introduction

Medical imaging has become an ubiquitous component of medicine, and magnetic resonance imaging (MRI) is one of its principal modalities. Its capacity for generating high-resolution 3-dimensional images of the body without relying on ionizing radiation makes it a mainstay for both research environments and clinical practice. Its high soft-tissue contrast makes MRI especially suited for the study of the brain.

One of the main strengths of MRI is its flexibility. Although complex physical mechanisms underlie the generation and measurement of the MRI signal, these mechanisms enable the design of pulse sequences that can sensitize the measured signal to a variety of tissue properties. Thus, rather than being a tool solely for the imaging of brain structure, the various MRI modalities provide a multifaceted picture of the brain, with functional MRI providing insight into brain activity by measuring changes in blood oxygenation, and diffusion MRI revealing aspects of tissue microstructure by studying the random motion of water molecules.

Another aspect of MRI flexibility manifests after the data have been acquired. The process of gleaning insights from MRI data is based on substantial theoretical and computational components. The data are limited and noisy, requiring a combination of preprocessing stages, model fitting, and statistics in order to reveal their meaning.

The multidisciplinary nature of MRI presents multiple avenues for development, with advances in scanner technology, acquisition sequence design, physical models, preprocessing tools, and statistical analysis methods, all having the potential to increase the usefulness of MRI.

The aim of this thesis is to explore two such avenues for improvement. On the one hand, it takes a multimodal approach to MRI, presenting several ways in which information from multiple MRI modalities can be productively combined. On the other, it focuses on ways in which modern mathematical and computational tools, such as Bayesian statistics, graph signal processing, and deep learning, can be usefully applied in MRI. It is the hope of the author that the works presented in this thesis will lay another brick on the road to uncovering the possibilities afforded by MRI.

1.1 Thesis outline

Chapter 2 is a general introduction to the physics of MRI. Chapters 3 and 4 delve deeper into specific physical and processing aspects of the MRI modalities this thesis focuses on, namely, functional MRI and diffusion MRI. Following this, Chapters 5 and 6 provide background on graph signal processing and deep learning, respectively, two of the main computational tools employed in this thesis. Finally, the main publications resulting from this work are reviewed in Chapter 7.

1.2 List of publications

This thesis is based on the following publications, referred to in the text with their roman numeral:

- I. **David Abramian**, Per Sidén, Hans Knutsson, Mattias Villani, and Anders Eklund. “Anatomically informed Bayesian spatial priors for fMRI analysis.” In: *IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. 2020, pp. 1026–1030. DOI: 10.1109/ISBI45749.2020.9098342.
- II. **David Abramian**, Martin Larsson, Anders Eklund, Iman Aganj, Carl-Fredrik Westin, and Hamid Behjat. “Diffusion-informed spatial smoothing of fMRI data in white matter using spectral graph filters.” In: *Neuroimage* 237 (2021), p. 118095. DOI: 10.1016/j.neuroimage.2021.118095.
- III. **David Abramian**, Ida Blystad, and Anders Eklund. “Evaluation of inverse treatment planning for Gamma Knife radiosurgery using fMRI brain activation maps as organs at risk.” In: *medRxiv preprint*. 2023. DOI: 10.1101/2022.12.12.22283334. Submitted to *Medical Physics*.
- IV. **David Abramian**, Anders Eklund, and Evren Özarslan. “Super-resolution mapping of anisotropic tissue structure with diffusion MRI and deep learning.” In manuscript.
- V. **David Abramian** and Anders Eklund. “Refacing: reconstructing anonymized facial features using GANs.” In: *IEEE 16th International Symposium on Biomedical Imaging (ISBI)*. 2019, pp. 1104–1108. DOI: 10.1109/ISBI.2019.8759515.

1.3 Additional publications

This thesis work also resulted in the following publications, which are not the main focus of this work:

- I. Anjali Tarun, **David Abramian**, Hamid Behjat, and Dimitri Van De Ville. “Graph spectral analysis of voxel-wise brain graphs from diffusion-weighted MRI.” in: *IEEE 16th International Symposium on Biomedical Imaging (ISBI)*. 2019, pp. 159–163. DOI: 10.1109/ISBI.2019.8759496.
- II. **David Abramian**, Martin Larsson, Anders Eklund, and Hamid Behjat. “Improved functional MRI activation mapping in white matter through diffusion-adapted spatial filtering.” In: *IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. 2020, pp. 539–543. DOI: 10.1109/ISBI45749.2020.9098582.
- III. Anjali Tarun, Hamid Behjat, Thomas Bolton, **David Abramian**, and Dimitri Van De Ville. “Structural mediation of human brain activity revealed by white-matter interpolation of fMRI.” in: *Neuroimage* 213 (2020), p. 116718. DOI: 10.1016/j.neuroimage.2020.116718.
- IV. Hamid Behjat, Iman Aganj, **David Abramian**, Anders Eklund, and Carl-Fredrik Westin. “Characterization of spatial dynamics of fMRI data in white matter using diffusion-informed white matter harmonics.” In: *IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. 2021, pp. 1586–1590. DOI: 10.1109/ISBI48211.2021.9433958.
- V. Marco Domenico Cirillo, **David Abramian**, and Anders Eklund. “Vox2Vox: 3D-GAN for brain tumour segmentation.” In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Springer. 2021, pp. 274–284. DOI: 10.1007/978-3-030-72084-1_25.
- VI. Marco Domenico Cirillo, **David Abramian**, and Anders Eklund. “What is the best data augmentation for 3D brain tumor segmentation?” In: *IEEE International Conference on Image Processing (ICIP)*. 2021, pp. 36–40. DOI: 10.1109/ICIP42928.2021.9506328.
- VII. Hamid Behjat, Anjali Tarun, **David Abramian**, Martin Larsson and Dimitri Van De Ville. “Voxel-wise brain graphs from diffusion MRI: Intrinsic eigenspace dimensionality and application to functional MRI.” In: *medRxiv preprint*. 2023. DOI: 10.1101/2022.09.29.510097. Accepted for publication in *IEEE Open Journal of Engineering in Medicine and Biology*.

1.4 Abbreviations

The following is a list of abbreviations appearing in this thesis:

ADC	apparent diffusion coefficient
AR	autoregressive
BOLD	blood oxygenation level-dependent
CSD	constrained spherical deconvolution
CNN	convolutional neural network
dHb	deoxyhemoglobin
dmRI	diffusion magnetic resonance imaging
DNN	dense neural network
dODF	diffusion orientation distribution function
DSI	diffusion spectrum imaging
DTI	diffusion tensor imaging
EPI	echo-planar imaging
FDR	false discovery rate
FID	free induction decay
FA	fractional anisotropy
fMRI	functional magnetic resonance imaging
fODF	fiber orientation distribution function
FWE	family-wise error
FWHM	full width at half maximum
GAN	generative adversarial network
GLM	general linear model
GQI	generalized q-sampling imaging
GRE	gradient echo
GSP	graph signal processing
HCP	Human Connectome Project
HRF	hemodynamic response function
LTI	linear and time-invariant
MD	mean diffusivity
MR	magnetic resonance
MRI	magnetic resonance imaging
ODF	orientation distribution function
PD	proton density
PGSE	pulsed gradient spin echo
OAR	organ at risk
QBI	Q-ball imaging
ReLU	rectified linear unit
RF	radiofrequency
rs-fMRI	resting state functional magnetic resonance imaging
SD	spherical deconvolution
SDF	spin density function
SE	spin echo
SGD	stochastic gradient descent
SH	spherical harmonic
SNR	signal-to-noise ratio

1.5 Data, code, and ethics

All the papers included in this thesis are based on open data [1, 2, 3], and all processing scripts are made available on GitHub¹. These two aspects facilitate the reproducibility of the reported results [4, 5]. The ethics board of Linköping decided that no additional ethics approvals are required to analyze these open datasets.

¹<https://github.com/DavidAbramian>

2

Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is a medical imaging modality which relies on the magnetic properties of chemical elements to generate images of the body. It is non-invasive, and largely safe for the subject, as it does not utilize ionizing radiation. This chapter will review the basic physical mechanisms that underlie the generation of images through MRI. More in-depth treatments of the principles of MRI can be found in [6] and [7].

2.1 The MRI scanner

At its simplest, an MRI scanner consists of a large tube, the scanner *bore*, into which a sample is placed to be scanned. In our case, this sample is a healthy human subject or a patient, but it can also be a non-human animal, a phantom featuring some desirable property, or any other substance.

The walls of the scanner bore are lined with several types of magnets. Typically, these are electromagnets, and are therefore often referred to as *coils*. These magnets are the main elements responsible for generating a measurable signal from the sample and acquiring said signal. Much of the remaining bulk of a clinical MRI scanner consists of the electronics needed to control the magnets.

There are three primary systems responsible for generating the various magnetic fields required to generate and acquire signal in an MRI scanner:

- The main magnet generates the static B_0 field, a very strong and uniform field oriented along the bore of the scanner, which by convention is assigned the z axis. Typical field strengths for current medical scanners are between 0.5 and 3 T, while 7 T units have become available in recent years. Higher static field strengths allow for larger signal-to-noise ratios in the acquired signal.
- The gradient system consists of three orthogonal gradient coils capable of generating a linear magnetic field gradient along arbitrary spatial orientations. They are fundamental for signal localization, as well as for traversing k -space.

- A set of radiofrequency (RF) transmitter coils produce a rotating magnetic field B_1 which excites the spin systems of the sampled nuclei. The same or a separate set of receiver coils are then used to read out the signal from the excited nuclear spins.

2.2 Nuclear magnetism

Atomic nuclei possess an intrinsic angular momentum \mathbf{J} , called *spin angular momentum*. Although a quantum-mechanical property, spin can be interpreted in classical mechanics as a rotation of the nucleus about its axis, yielding a microscopic magnetic field around it which causes it to behave as a tiny bar magnet. The magnetic moment of a nucleus $\boldsymbol{\mu}$ is related to its spin angular momentum \mathbf{J} by

$$\boldsymbol{\mu} = \gamma \mathbf{J}, \quad (2.1)$$

where γ is the gyromagnetic ratio of the element. For hydrogen, the main nucleus investigated in MRI, this constant takes the value $2.675 \cdot 10^8$ rad/s/T.

In the absence of an external magnetic field, the magnetic moments $\boldsymbol{\mu}$ within a sample have random spatial orientations, resulting in no measurable bulk magnetization. However, when placed within the strong B_0 field of an MRI scanner, nuclear spins begin precessing around the orientation of B_0 at the *Larmor frequency* of

$$\omega_0 = \gamma B_0. \quad (2.2)$$

Furthermore, each individual spin takes one of a limited set of possible angles with respect to the orientation of B_0 . For hydrogen nuclei, only two spin orientations are possible: parallel and antiparallel with respect to B_0 (see Figure 2.1). In a phenomenon known as *Zeeman splitting*, spins in the parallel state have a lower energy level than those in the antiparallel state, which results in the former being very slightly more numerous in a sample. This small difference in the spin populations is responsible for generating a bulk magnetization vector \mathbf{M} oriented along z . As the precession phase of individual spins is still random, there is no transversal component to \mathbf{M} in the xy plane. Therefore, the bulk magnetization \mathbf{M} at equilibrium is given by

$$\begin{cases} M_z^0 = \frac{\gamma^2 \hbar^2 B_0 N_s}{4KT_s}, \\ M_{xy}^0 = 0, \end{cases} \quad (2.3)$$

where \hbar is Planck's constant divided by 2π , K is Boltzmann's constant, N_s is the total number of spins in the sample, and T_s is the absolute temperature of the spin system.

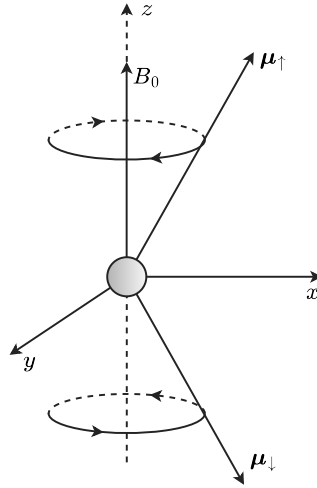


Figure 2.1: Magnetized spins precess about the B_0 field, taking either a parallel (μ_{\uparrow}) or an antiparallel (μ_{\downarrow}) orientation with respect to it.

RF excitation

In order to produce measurable signal from the spin system, its bulk magnetization \mathbf{M} needs to be tilted onto the xy plane. This can be done by subjecting the spin system to a transient magnetic field B_1 , itself rotating in the xy plane at the Larmor frequency of the spin system. This is referred to as an *RF pulse*, due to its transient nature and the fact that the Larmor frequency of hydrogen is in the RF band.

Subjecting the spin system to a magnetic field perpendicular to B_0 and rotating at its Larmor frequency fulfills the *resonance condition* of the spin system, prompting the bulk magnetization \mathbf{M} of the system to experience a *forced precession* about the B_1 field, with

$$\omega_1 = \gamma B_1. \quad (2.4)$$

The spin system is shifted away from the z axis by the *flip angle*

$$\alpha = \int_0^{\tau_p} \omega_1(t) dt = \int_0^{\tau_p} \gamma B_1^e(t) dt, \quad (2.5)$$

where $B_1^e(t)$ is the temporal envelope of the RF pulse, and τ_p is the duration of the pulse. The shape of RF pulses can be controlled in order to influence various aspects of the excitation process, such as its duration and the bandwidth of excited spins.

Relaxation

After a spin system has been excited by a transient RF pulse, the tilted magnetization vector \mathbf{M} continues to precess around the B_0 field. At the

same time, the spins are subjected to two exponential relaxation processes that cause the system to return to its earlier equilibrium state:

- Longitudinal relaxation, causing the longitudinal component of magnetization M_z to progressively return to the equilibrium state of M_z^0 . This process is characterized by time constant T_1 .
- Transverse relaxation, by which the rotation of the individual spins about B_0 continuously become dephased with each other, resulting in the progressive disappearance of the transversal magnetization component M_{xy} . This process is characterized by time constant T_2 , and is faster than longitudinal relaxation for all ordinary spin systems [8]. Furthermore, in practice, field inhomogeneities cause the transversal magnetization components to decay at an increased rate characterized by time constant T_2^* (“ T_2 star”).

The temporal evolution of the longitudinal and transversal components of the magnetization vector \mathbf{M} are given by

$$\begin{cases} M_z(t) = M_z^0 (1 - e^{-t/T_1}) + M_z(0_+)e^{-t/T_1} \\ M_{xy}(t) = M_{xy}(0_+)e^{-t/T_2} \end{cases} \quad (2.6)$$

Due to their different chemical environment, various bodily tissues differ in their T_1 and T_2 relaxation constants. This difference in relaxation properties is one of the fundamental ways in which MR images achieve contrast between different tissues. The magnitude of the magnetization is also proportional to the proton density (PD) of the sample, making this a third source of contrast.

2.3 MR signals

As was discussed previously, when a spin system is excited by an RF pulse, its bulk magnetization is tilted away from the z axis and begins precessing around B_0 field at the Larmor frequency. By Faraday’s law, this induces an oscillating voltage in the receiver RF coils of the scanner. After being subjected to demodulation and low-pass filtering, the resulting signal can be described in simplified form as

$$S(t) = \int_{\text{object}} M_{xy}(\mathbf{r}, 0)e^{-t/T_2^*(\mathbf{r})}e^{-i\Delta\omega(\mathbf{r})t}d\mathbf{r} \quad (2.7)$$

For a spin system of a spin spectral density of $\rho(\omega)$ this equation can be rewritten as

$$S(t) = \int_{-\infty}^{\infty} \rho(\omega)e^{-t/T_2(\omega)}e^{-i\Delta\omega t}d\omega \quad (2.8)$$

In the following we describe the main forms that these signals take.

Free induction decays

Free induction decays (FIDs) represent the signals obtained when the spins are permitted to precess freely after a single excitation. In this scenario, the transversal magnetization component M_{xy} , and thus the measured signal, decays exponentially with time constant T_2^* .

Spin echoes

After the transversal magnetization is allowed to dephase for a period of time, the FID signal can decay completely. However, it is possible to apply a 180° RF pulse to flip the magnetization of the spins around an axis determined by the phase of the pulse in the xy plane. As their magnetization has been flipped, yet their precession continues in the same direction, their continued motion results in their progressive rephasing. If the 180° pulse is applied at time $T_E/2$, then the spins will become fully rephased at the *echo time* T_E , producing what is known as a *spin echo* (SE) signal. The time between iterations of excitation-dephasing-refocusing is referred to as *repetition time* T_R . The SE signal is only affected by T_2 relaxation, as the field inhomogeneity effects causing faster T_2^* relaxation are reverted by the rephasing. SE is one of the fundamental MR sequence designs, capable of producing T_1 -, T_2 -, and PD-weighted images [9], and used in standard diffusion-weighted sequences [10] (further discussed in Chapter 4).

Gradient echoes

A different way to generate MR echoes is by using a single RF excitation pulse, and achieving dephasing and rephasing by using magnetic field gradients. Following the excitation pulse, a magnetic field gradient is used to induce a spatial dependence on the rotation frequency of the nuclear spins, causing quick dephasing. After a time $T_E/2$, this gradient is inverted, resulting in a progressive rephasing which culminates in the generation of a *gradient echo* (GRE) at echo time T_E . Unlike for SE, GRE echoes are affected by T_2^* relaxation, as field inhomogeneity effects are not compensated. In addition to allowing for very fast T_E and T_R , resulting in faster acquisition, the T_2^* weighting in GRE sequences makes them sensitive to susceptibility changes brought about by blood oxygenation effects, and are thus widely used for functional MRI (further discussed in Chapter 3).

2.4 Image formation

As implied by equation 2.7 the signal measured by a receiver coil after an RF excitation pulse will include contributions from the whole scanned object, which is not conducive to producing an image. The gradient system of the

scanner is fundamental for enabling the differentiation of signal components by their source location. This is done through the mechanisms of slice selection and spatial encoding.

Slice selection

Slice selection is used to limit the region of nuclear spins that will be excited to a single slice. This is done by applying a magnetic field gradient at the time of RF excitation. The gradient causes a spatial dependence in the Larmor frequency of protons across the imaged object, with only those matching the frequency of the RF pulse being excited. This limits the excited volume to a single slice of the object, whose thickness can be selected by controlling the bandwidth of the RF pulse.

Spatial encoding

Beyond selective excitation through slice selection, the fundamental way in which magnetic field gradients are used to disentangle the spatial source of MR signals is by imparting a spatial dependence on the frequency and phase of the spin precession across the scanned object. These two mechanisms are known as *frequency encoding* and *phase encoding*, respectively.

Frequency encoding is achieved by applying a continuous gradient \mathbf{G}_f during acquisition. Thus, the Larmor frequency across the object becomes

$$\omega(\mathbf{r}) = \omega_0 + \gamma(\mathbf{G}_f \cdot \mathbf{r}). \quad (2.9)$$

Likewise, the signal can be phase encoded by applying a transient gradient \mathbf{G}_p , which imparts a spatially-dependent phase

$$\phi(\mathbf{r}) = -\gamma(\mathbf{G}_p \cdot \mathbf{r})T_p \quad (2.10)$$

to the system. These two effects can act in unison to influence the frequency and phase of precession of spins in different locations.

By rewriting the received MR signal as

$$S(\mathbf{k}) = \int \rho(\mathbf{r})e^{-i\mathbf{k} \cdot \mathbf{r}} d\mathbf{r}, \quad (2.11)$$

it can be shown to be the Fourier transform of the magnetization, where

$$\mathbf{k}(t) = \gamma \int_0^t \mathbf{G}(\tau) d\tau. \quad (2.12)$$

Therefore, the position in k -space is given by the history of applied gradients after RF excitation. By performing repeated RF excitations and readouts while applying various frequency and phase encoding gradients it is possible to obtain a sampling of k -space. In the simplest case of uniform Cartesian sampling, the image data can be recovered as the inverse Fourier transform of the sampled data (see Figure 2.2).

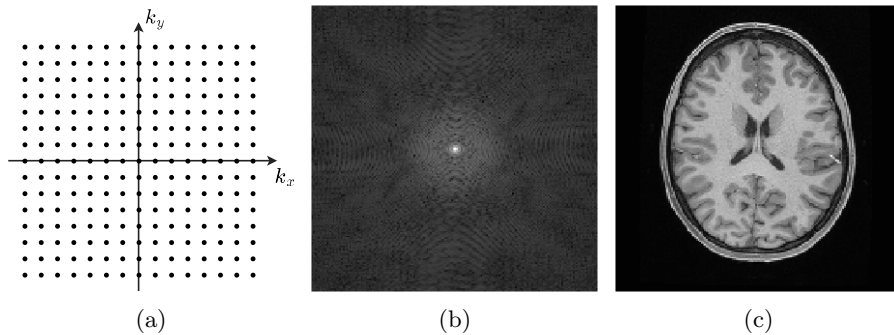


Figure 2.2: Sampling of k -space and image reconstruction. (a) 2-dimensional Cartesian sampling of k -space along the x and y axes. (b) Example k -space data. (c) Reconstructed T_1 -weighted image obtained by taking the 2-dimensional Fourier transform of (b).

3

Functional Magnetic Resonance Imaging

Functional magnetic resonance imaging (fMRI) is an MRI modality used to study the ongoing activity of the brain in time. Due to its noninvasive nature, high spatial resolution, and good temporal resolution, it has become an indispensable tool for studying the brain in both research and clinical contexts. Despite this, it does have some limitations [11]. It does not measure neuronal activity directly, but only through the proxy of changes in blood oxygenation. Furthermore, the signal it measures is weak, requiring a substantial processing and analysis apparatus in order to extract its meaning.

There are two fundamental modalities of fMRI, each spanning their own separate, although overlapping, field of study. *Task-based fMRI* is concerned with studying the brain's response to a given set of stimuli or the performance of specific tasks. Among its applications are functional localization for clinical interventions, pharmacological studies on the effects of drugs, as well as research in neuroscience and psychology. Conversely, in *resting state fMRI* (rs-fMRI), subjects are scanned in a state of rest, i.e., without any specific stimuli or task prompts. The similarities and differences in the spontaneous brain activity across brain regions can be studied to extract information about the inherent organization of the brain, what is known as *functional connectivity*. Although rs-fMRI is a fertile field of study, the fMRI work carried out for this thesis concerns only task-based fMRI. This chapter will describe the fundamental aspects of fMRI and the acquisition and analysis of task-based fMRI data.

3.1 Origins of fMRI: the BOLD signal

The possibility of studying brain activity through MRI results from two observations: that brain activity is followed by local changes in blood oxygenation, and that blood oxygenation can be readily imaged with MRI.

When part of the brain is recruited for a task, its oxygen requirements are increased. In order to supply it with oxygen, its local bloodflow is increased after a delay of several seconds. Importantly, bloodflow is increased

beyond the immediate needs of the region, leading to a temporary surplus of oxygenated blood, which eventually recedes to pre-stimulus levels.

Deoxyhemoglobin (dHb), whose concentration follows the changes in blood oxygenation, acts as a natural paramagnetic contrast agent, causing changes in magnetic susceptibility detectable through MRI. This is known as the *blood oxygenation level-dependent* (BOLD) contrast, the signal measured in fMRI, first described in rodent brains by Ogawa et al. in the 1990s [12, 13], and used for imaging human brain activity shortly afterwards [14, 15, 16].

The temporal relationship between momentary brain activity and the changes in local blood flow that follow it can be described by a *hemodynamic response function* (HRF), which is a fundamental component in the analysis of fMRI data. It is a relatively slow response, taking anywhere around 10-20 seconds to return to baseline levels, and is characterized by a short initial dip followed by a peak and culminating in a long post-stimulus undershoot. Importantly, the relationship between neural activity and hemodynamic response is largely linear and time-invariant (LTI). This makes it possible to predict the expected hemodynamic response for an arbitrary set of stimuli just from the timing of the stimuli and the HRF.

White matter fMRI

The BOLD signal has been traditionally thought to originate from local field potentials, which reflect intracortical activity and are thus constrained to the gray matter of the brain [17]. At the same time, although activations in white matter are not unusual, they are commonly dismissed, with the BOLD signal in white matter being sometimes used as a nuisance regressor for activation mapping.

However, recent publications reflect the beginning of a shift in perspective, where the possibility and functional relevance of white matter fMRI is acknowledged [18, 19, 20, 21]. Part of the reason for the lack of attention that white matter fMRI has received can be attributed to the anatomical and physiological differences between white matter and gray matter. An overall weaker BOLD signal, a different HRF, and acquisition sequences optimally tuned for gray matter, among others, together make it difficult to detect relevant activity in white matter [18], and point to the need for developing methods tailored to white matter fMRI.

Paper II is focused on developing one such method. In particular, it relies on the observation that the spatiotemporal correlation structure of the BOLD signal in white matter is anisotropic and generally matches the orientation of the underlying axonal fiber bundles [22, 23, 24]. This observation suggests that the isotropic Gaussian filters generally used in fMRI analysis (discussed further) are not well suited to detecting activations in white matter based on the matched-filter argument. We propose and evaluate an alternative design based on diffusion-weighted imaging and graph signal processing (see Chap-

ters 4 and 5, respectively) which allows filters to adapt to the shape of white matter fiber bundles, and demonstrate its increased detection power.

3.2 Design of fMRI experiments

In a standard fMRI experiment, a subject is placed in an MRI scanner and subjected to a timed sequence of stimuli, or asked to perform a certain task following a specific time scheme. All the while, the subject's brain is being repeatedly scanned, resulting in a 4-dimensional dataset describing both the spatial and temporal aspects of the subject's brain activity. After the experiment this data is preprocessed to account for various imperfections and artifacts, and subjected to statistical analysis to identify the brain regions that were involved in the experiment.

It is common to acquire data for multiple stimuli or tasks (known as *conditions*) within a single fMRI experiment, often with the purpose of comparing the subject's responses to each. At the same time, each condition needs to be repeated multiple times in a single experiment in order to produce data with enough statistical power to properly identify the brain activity. The set of individual conditions and their specific timing parameters constitute a *task paradigm*. These can generally be either *block-based* designs, where conditions are presented in blocks of constant duration intercut with periods of rest, or *event-related* designs, where stimuli are presented in short blocks with uneven timing. Figure 3.1 illustrates an example fMRI experiment with a block-based design.

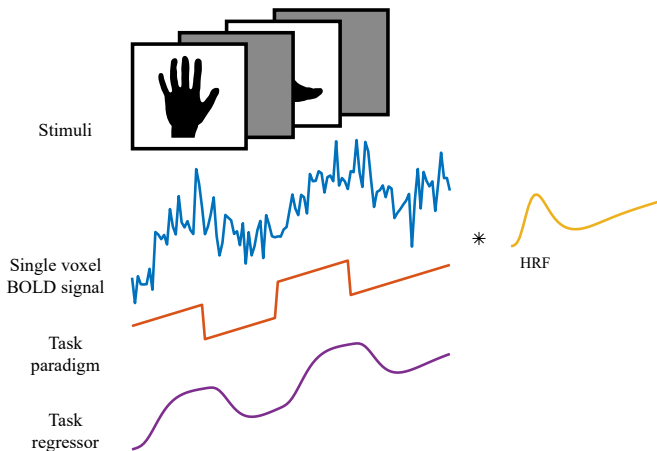


Figure 3.1: Example fMRI experiment with a block-based design and a motor task. The subject alternates block of activity and rest. The BOLD signal is noisy, but the signal attributed to the performance of the task can be modeled as the convolution of the task paradigm with the HRF.

Since fMRI is concerned with the temporal dynamics of blood oxygenation, it requires fast scanning sequences, with common T_R values in the 0.5 – 5 second range. The most common acquisition sequence used for fMRI is GRE echo-planar imaging (EPI), first described by Peter Mansfield in 1977 [25, 26]. This is a 2D sequence where k -space is traversed in a zig-zagging pattern using frequency and phase encoding along perpendicular axes. The use of GRE is optimal to obtain T_2^* -weighted signal, which is sensitive to the magnetic susceptibility effects of the BOLD signal, in addition to being relatively fast.

3.3 Preprocessing of fMRI data

The analysis of fMRI data begins with a preprocessing stage, with the purpose of correcting various data artifacts, increasing the detection power of the subsequent statistical analysis, and optionally taking the data of multiple subjects into a common space where they can be compared. This section will briefly describe some of the most common preprocessing steps taken and their motivation:

Distortion correction The acquisition of fMRI data with an EPI sequence is prone to suffer from *susceptibility artifacts*, substantial distortions that happen along the phase encoding direction of acquisition due to inhomogeneities in the B_0 field, which generally appear at the boundaries between air and tissue. These artifacts are quite pernicious, resulting in a compression or elongation of the brain in the phase encoding direction which complicates the precise localization of activity in the brain as well any registration efforts within the fMRI series or with other, unaffected modalities. Methods to compensate for these artifacts generally involve nonlinearly unwarping the images after estimating the B_0 field, either directly by the calculation of a *fieldmap* from the phase of the acquired signal [27, 28], or using methods based on scanning twice with opposite phase encoding directions, which also inverts the direction of distortion [29].

Slice timing correction Given that fMRI volumes are acquired using 2D sequences, different slices of the data are acquired at slightly different times, with delays between the first and last slice of a volume being potentially as high as nearly one full T_R . This difference in timing can have detrimental effects for the subsequent statistical analysis of the data, as it relies on defining a model with precise timing information. Slice timing is usually corrected by interpolating all the slices to the timing of a reference slice [30, 31].

Motion correction During the performance of an fMRI experiment, it is inevitable that the subject in the scanner will move their head, even if by a small amount. Unfortunately, this can have relatively severe

negative consequences for the remainder of the analysis. For one, head motion causes the individual time-series at each voxel in the brain to correspond to different spatial locations at different points in time. The largest effects from this are seen at the edges of the brain, where voxels can have large fluctuations in their signal, often resulting in spurious activations. This effect can be largely controlled by using linear or rigid registration to spatially align all the fMRI volumes with a reference volume.

Head motion causes a further, more pernicious effect by influencing the magnetization of excited spins, which, as discussed in Chapter 2, is spatially encoded. Because of this, head motion can result in spurious activations even after applying motion correction [32]. In order to account for these effects, it is common to extract a set of motion parameters (translations and rotations) from the motion correction procedure and use them in the statistical analysis to model any remaining signal components correlated with the head motion.

Spatial smoothing There are a number of reasons given for the need of applying spatial smoothing to fMRI data. First, the BOLD signal is relatively weak, and given that activations are generally expected to extend over a region, spatial smoothing can increase the *signal-to-noise ratio* (SNR) of the data. Furthermore, due to individual subject variations, there can be some misalignment when trying to compare activations across subjects. In such cases, spatial smoothing can blur the regions of activity and increase the potential for overlap across subjects. Finally, some statistical analyses, such as those based on Gaussian random field theory (discussed further), require that the data present a certain degree of smoothness, which can be achieved by applying spatial smoothing.

The most common approach to spatial smoothing is using isotropic Gaussian filters, which are generally described (in mm) by their full width at half maximum (FWHM), given by

$$\text{FWHM} = 2\sigma\sqrt{2\ln(2)}, \quad (3.1)$$

where σ represents the standard deviation of the Gaussian. However, there have been a number of methods incorporating additional assumptions or information from other MRI modalities to better match the smoothing filter to the expected shape of activations. Paper I presents an analogous method incorporating information from a T_1 -weighted image to adapt the smoothing to the anatomical features of the brain, albeit within a Bayesian analysis framework which does not rely on filtering to achieve smoothness (discussed further). Moreover, as mentioned previously, Paper II presents a spatial smoothing method tailored specifically to increasing the SNR in white matter.

Detrending It is common for the fMRI voxel time-series to present with low frequency drifting and trends, which can influence the estimation of activity if left unaccounted. There are several ways to account for these, the most common being their removal by high-pass filtering or by including detrending regressors in the statistical model.

Spatial normalization As mentioned previously, variability between individual brains precludes any simple comparison of activations between subjects. In order to enable such comparisons, it is necessary to spatially warp the fMRI data of all subjects to a common template space in which they can be compared. Several such templates have been proposed, with the most popular being the Talairach atlas [33, 34], the MNI305 atlas [35], and the derived ICBM-152 template.

Spatial normalization generally involves a nonlinear registration step. As fMRI data has very low anatomical contrast, the nonlinear warping is usually estimated from a high resolution anatomical image (e.g. T_1 -weighted). The fMRI data itself needs to be registered to the anatomical image using methods that can account for the difference in contrast between both modalities. In order to avoid loss of spatial resolution through repeated interpolation, it is common to combine all steps involving registration (e.g., T_1 to template space, fMRI to T_1 , motion correction of individual fMRI volumes) into a single transformation.

It should be noted that in practice there is considerable variation in the preprocessing steps performed for individual fMRI experiments, with no clear consensus on any single set of procedures. The issue is further complicated by the widespread reliance in the field on several fMRI software packages [36, 37, 38] developed separately by different groups, which do not share the same preprocessing and statistical methods, and present some idiosyncrasies regarding the correct way to analyze fMRI data.

There have been some attempts to increase the interoperability of existing tools, two of which should be highlighted. Nipype [39] allows the creation of processing pipelines in Python that can combine individual tools from a number of software packages. Based on it, fMRIPrep [40] is an attempt to provide a standardized, complete and reproducible preprocessing pipeline for fMRI data based on best practices. This was used for preprocessing the fMRI data used in Paper III.

3.4 Statistical analysis of fMRI data

There exists a multitude of procedures for determining the location and extent of brain activations in an fMRI dataset. This section will give a standard presentation of single-subject analysis based on the general linear model (GLM).

General linear model

A standard way to locate the active voxels in the brain is to formulate a general linear model problem, in which the signal observed at each voxel is modeled as a linear combination of a set of known explanatory variables, the *regressors*, plus an unaccounted remaining component, the *error terms*.

Considering an fMRI dataset of T volumes (i.e., time points), with N voxels and K explanatory variables, the GLM can be formulated in matrix form as

$$\underset{[T \times N]}{\mathbf{Y}} = \underset{[T \times K]}{\mathbf{X}} \underset{[K \times N]}{\mathbf{W}} + \underset{[T \times N]}{\mathbf{E}}, \quad (3.2)$$

where the columns of \mathbf{Y} contain the observed time-series at each voxel, the columns of \mathbf{X} contain the time-series of the regressors, \mathbf{W} is the parameter matrix to be estimated, describing the degree to which each regressor is present in each observed time-series, and \mathbf{E} is the error matrix, also to be estimated, containing the portion of the signals in \mathbf{Y} that can not be explained by the regressors in \mathbf{X} .

The *design matrix* \mathbf{X} is composed of several regressors of interest, those associated with the various task conditions and whose presence we want to identify in the signals, and a set of nuisance regressors, which are not specifically of interest but must be included in order to remove spurious components from the signal. The task regressors are obtained by convolving the time-series of each condition with the HRF, giving the expected hemodynamic response for the given time-series of stimuli. This is made possible by the LTI behavior of the hemodynamic response of the brain. The nuisance regressors typically include a set of motion regressors, and, optionally, a set of detrending regressors, but it is possible to include a large number of additional regressors that account for various imperfections in the BOLD signal [41, 42]. Furthermore, it is possible to include the temporal derivative of regressors, which increases the flexibility of the model by accounting for small imprecisions in the timing of events.

The error terms in \mathbf{E} are commonly assumed to be normally distributed with zero mean and a variance of σ_n^2 for the n -th voxel. If, in addition, the error terms are temporally uncorrelated, an unbiased and least-squares optimal solution for the regression parameters \mathbf{W} can be found by

$$\hat{\mathbf{W}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}, \quad (3.3)$$

while the error variance for each voxel σ_n^2 can be estimated as

$$\hat{\sigma}_n^2 = \frac{(\mathbf{Y}_{:,n} - \mathbf{X} \hat{\mathbf{W}}_{:,n})^T (\mathbf{Y}_{:,n} - \mathbf{X} \hat{\mathbf{W}}_{:,n})}{T - (K + 1)} \quad (3.4)$$

where $T - (K + 1)$ is the number of *degrees of freedom* of the model. In practice, however, the errors for a voxel at different time points do tend to be correlated. A common way to address this issue is to model the serially-correlated errors as an autoregressive (AR) process of order p .

Hypothesis testing

To determine whether any one voxel was engaged by a task, we attempt to reject the hypothesis that one of its regression coefficients is 0. More generally, it is possible to test for the involvement of a linear combination of regressors by specifying a *contrast vector* \mathbf{c} and evaluating the null hypothesis of $\mathbf{c}^T \hat{\mathbf{W}}_{:,n} = 0$. To test this hypothesis, a t-statistic is first obtained by

$$t_n = \frac{\mathbf{c}^T \hat{\mathbf{W}}_{:,n}}{\sqrt{\hat{\sigma}_n^2 \mathbf{c}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c}}}, \quad (3.5)$$

which, under the null, follows a t-distribution with $T - (K + 1)$ degrees of freedom. The t-statistics for all voxels can then be arranged into a t-map (see Figure 3.2a), which can be used to infer the activate voxels in several ways. The most common inference methods are:

Voxel-level inference The simplest way to determine which voxels are active is to threshold the t-map at a given significance level and consider the voxels that survive the thresholding as active (see Figure 3.2b). This type of inference provides high spatial specificity in the activations, but requires correction for the large number of statistical tests performed (discussed further). All the fMRI analysis presented in this thesis used voxel-level inference.

Cluster-level inference In an alternative approach, the t-maps are first thresholded at an arbitrary *cluster-forming threshold* (e.g., $p = 0.001$) to define clusters of activity, i.e., contiguous regions that survive the threshold. Clusters are then deemed to be significant if they survive a *cluster size* or *cluster mass threshold* [43]. Although this approach has increased sensitivity, it also results in decreases spatial specificity. For one, small clusters cannot be deemed active by design. Furthermore, activity cannot be ascribed to any specific set of voxels within an active cluster.

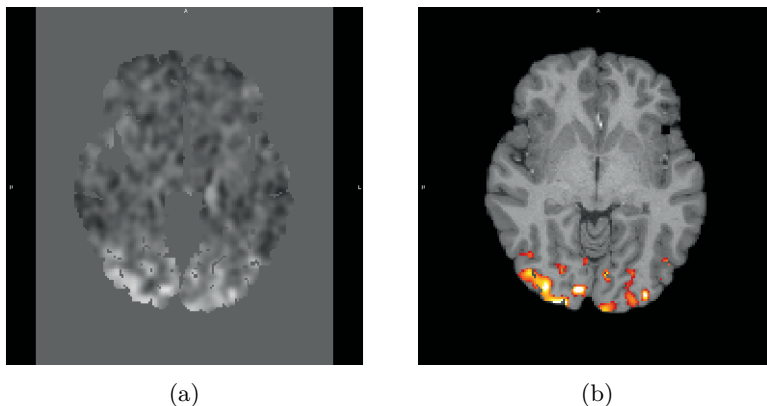


Figure 3.2: Activation mapping. (a) t-map from emotion processing task. (b) Activation map obtained by voxel-level inference on the t-map from (a), overlaid on subject's T_1 -weighted image.

Family-wise error correction

The analysis of fMRI data requires performing a large number of statistical tests. For example, voxel-level inference requires one test per voxel, resulting in as many as several hundred thousand tests for a single subject. Individual tests are performed at a significance level of α , which represents the possibility of obtaining a false positive. However, α vastly underestimates the probability of obtaining false positives when multiple tests are performed. Instead, it is reasonable to consider the probability of obtaining any false positives across all the tests performed, a measure known as the *family-wise error rate* (FWE rate).

The simplest approach to control the FWE rate is applying *Bonferroni correction*, where individual tests are performed at the adjusted level α/N , where N is the total number of tests performed. Under such correction, there is a probability α of having one or more false positives across all of the tests performed. However, this correction can be excessively strict, with few voxels surviving thresholding. A less stringent approach is provided by *false discovery rate* (FDR) correction [44, 45], which, when controlled at level α , ensures that of the voxels deemed active, a fraction α is expected to be false positives. In the case that no correction is applied, a fraction α of all the tests performed is expected to result in false positives.

Another set of approaches for controlling the FWE rate is based on Gaussian random field theory [46]. These methods are based on the observation that, for sufficiently smooth data, the number of independent observations is substantially smaller than the number of voxels in the dataset, and thus require less stringent correction. These methods can be applied to both voxel-level and cluster-level inference. However, they rely on assumptions about the

smoothness of the data, e.g., that it is constant across the brain. Furthermore, these methods require the application of large smoothing filters in order to obtain data of sufficient smoothness, which negatively impacts the spatial specificity of detected activations.

Bayesian analysis of fMRI data

What has been discussed so far covers a large portion of the common practice for fMRI analyses. However, Paper I involves a substantially different formulation of the statistical aspects of fMRI analysis based on Bayesian statistics, for which a brief introduction is in order.

The statistical perspective presented so far is *frequentist*, where probabilities are regarded as long-term rates of occurrence of events. Bayesian statistics is a different formalization of statistics, based on the concept of subjective probabilities and the use of Bayes' theorem to update degrees of belief in the face of data. This updating procedure can be written as

$$\underbrace{P(\boldsymbol{\theta}|\mathbf{Y})}_{\text{Posterior probability of } \boldsymbol{\theta}} = \frac{\overbrace{P(\mathbf{Y}|\boldsymbol{\theta})P(\boldsymbol{\theta})}^{\text{Likelihood function}}}{\underbrace{P(\mathbf{Y})}_{\text{Prior probability of } \boldsymbol{\theta}}} \propto \underbrace{P(\mathbf{Y}|\boldsymbol{\theta})}_{\text{Likelihood function}} \underbrace{P(\boldsymbol{\theta})}_{\text{Prior probability of } \boldsymbol{\theta}}, \quad (3.6)$$

which expressed how our *prior* knowledge of parameters $\boldsymbol{\theta}$ is modified by the observation of data \mathbf{Y} , producing an updated *posterior* probability distribution. The *likelihood function* expresses the way in which the specific set of observed data \mathbf{Y} would arise from a model with hypothetical parameters $\boldsymbol{\theta}$. One disadvantage of Bayesian inference is that, with the exception of very simple models, it is not possible to obtain a closed-form solution for the posterior distribution of a model. Instead it is possible to use approximation methods, such as variational Bayes, to obtain a closed-form approximation of the posterior, or sampling methods, such as Markov chain Monte Carlo (MCMC) to obtain samples from the exact posterior, which can be used to calculate any point statistics.

The Bayesian fMRI analysis framework employed in Paper I was introduced by Penny et al. [47, 48, 49, 50] and further developed by Sidén et al. [51, 52]. One fundamental way in which it differs from the frequentist approach described previously is in its handling of the smoothness of data. While the frequentist approach requires smoothing the data with a spatial filter of arbitrary size, the existing smoothness of the data can instead be included in the Bayesian model in the form of spatial priors. Hyperparameters for these spatial priors are estimated for each regressor and AR coefficient separately, allowing them to model different degrees of smoothness. Paper I illustrates another advantage of this formulation, namely that the priors can be easily modified to make the smoothing imposed by the model anatomically-adaptive.

4

Diffusion Magnetic Resonance Imaging

Diffusion magnetic resonance imaging (dMRI) is an MRI modality used to probe and characterize the diffusive movements of spin-bearing particles in the specimen. This, in turn, can reveal important aspects of their local environment, e.g., the microstructure of tissues.

This chapter will introduce the main aspects of dMRI related to the publications presented in this thesis.

4.1 The diffusion process

Diffusion refers to a stochastic process by which particles move from regions of high concentration to regions of low concentration without bulk motion. A typical illustration of diffusion is provided by a drop of dye slowly spreading in a glass of water. From this perspective, the diffusion process is described by Fick's first law, which states

$$\mathbf{J} = -D\nabla C, \quad (4.1)$$

where \mathbf{J} is the net particle flux vector, C is the particle concentration, and D is the diffusion coefficient. The latter characterizes the rate at which diffusion takes place, and its value depends on various aspects of the situation, such as the size of the diffusing molecules, the temperature, and the microstructural features of the environment. It is this dependence of D on environmental aspects which makes it possible to infer microstructural features from its measurement.

On a molecular level, the mechanism responsible for the particle flux is *self-diffusion*, the random motion of particles, which takes place even in the absence of a concentration gradient for all particles at temperatures over zero Kelvin. Einstein showed that, under unconstrained conditions (i.e. *free diffusion*), the displacements of a large number of particles follow a Gaussian distribution, with the mean squared displacement, in 3-dimensional space, being given by

$$\langle r^2 \rangle = 6Dt, \quad (4.2)$$

where t is the diffusion time.

4.2 Acquisition of dMRI

The diffusion of water molecules in a scanned sample has measurable effects on the MRI signal. Due to the existence of inhomogeneities in the B_0 field, the random motion of particles causes them to accrue different precession phases. This dephasing results in a slightly weaker spin echo signal than could be expected from magnetic relaxation effects alone. As the loss of signal is proportional to the diffusion distance, it can be used to measure D .

The *diffusion-weighting* of the signal can be controlled and emphasized by the use of magnetic field gradients. This is the basis of the pulsed gradient spin echo (PGSE) sequence introduced by Stejskal and Tanner [10], which is the basis of most modern dMRI sequences in common usage.

In this sequence, the initial RF excitation pulse is followed by two brief gradient pulses of duration δ (*encoding time*) separated by a time Δ (*diffusion time*), and with a 180° RF pulse applied between them. The first of these gradient pulses induces a position-dependent phase change on the particles, which is proportional to $q = \gamma\delta G$, where G is the magnitude of the magnetic field gradient. Afterwards, the 180° RF pulse flips the sign of the first phase change, after which the second pulse applies an equal phase change to that of the first gradient. If the particles remained stationary, the phase changes from both gradients would cancel out. However, given their random motion and the location-dependence of the phase changes from the gradients, the degree to which the phase changes are canceled out becomes a function of the distance traveled by the particles in the direction in which the gradients were applied.

When considering the signal produced by this sequence, it is useful to focus on the effects of diffusion-weighting and disregard the effects of relaxation. To that end, we can consider the *MR signal attenuation* $E(q) = S(q)/S(0)$, where $S(q)$ is the measured signal and $S(0)$ refers to the signal acquired in the absence of diffusion-encoding gradients. For the case of free diffusion, this signal becomes

$$E(q) = e^{-q^2 D(\Delta - \delta/3)} = e^{-bD}, \quad (4.3)$$

where $b = q^2(\Delta - \delta/3) = (\gamma\delta G)^2(\Delta - \delta/3)$ is the b -value, a metric quantifying the degree of diffusion sensitization imposed by the sequence.

In practice, most water molecules are found in a complex environment, where they are hindered by cells and other molecules. Because of this, the assumption of free diffusion rarely applies, with the true diffusion distances of particles being smaller than what would be expected. Consequently, it is common to specify that the quantity generally measured by dMRI is the *apparent diffusion coefficient* (ADC) [53], which is lower than the free diffusion

coefficient D , and is given by

$$ADC = -\frac{\log(E(q))}{b}. \quad (4.4)$$

4.3 Reconstruction models

As has been discussed so far, MRI can be used to measure the diffusivity of water molecules in tissue. The orientation of the measurement is the same as that of the diffusion-encoding gradients, while the degree of diffusion-weighting can be controlled by setting the b -value. Using these tools, there are multiple models for characterizing the diffusion properties of tissue, which differ in their complexity, the precision of the description they provide, and the amount and type of data they require.

For regions of isotropic diffusion, such as the brain gray matter, a single measurement of ADC in any direction may provide sufficient characterization. However, for the study of anisotropic regions, where the measured diffusivity depends on the orientation along which it is being measured, more complex models are necessary to provide a full description. This is the case for brain white matter, composed of long axonal fiber bundles which hinder the diffusion of water molecules across their orientation more than they do along their orientation. This makes dMRI useful for measuring the orientation of white matter fibers.

Diffusion tensor imaging

One simple way to model anisotropic diffusion is by retaining the Gaussian diffusion shape, but introducing a matrix for representing diffusivity, known as the *diffusion tensor*, that allows the distribution to extend unevenly in 3 dimensions. Adaptation of this to dMRI has led to *diffusion tensor imaging* (DTI) [54, 55]. The diffusion tensor \mathbf{D} is a real, symmetric 3×3 matrix written as

$$\mathbf{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix}, \quad (4.5)$$

whose diagonal elements represent the measured diffusivities along the x -, y - and z -axes, while the off-diagonal elements are related to the covariance of displacements along different axes.

In order to derive the most information from the diffusion tensor, it is useful to take its eigendecomposition. Its eigenvectors $\mathbf{Q} = [\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3]$ and eigenvalues $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \lambda_3]$ can be used to represent the diffusion tensor as an ellipsoid oriented in 3-dimensional space, which provides a visual description of the diffusion process. Furthermore, the eigenvalues can be used to produce some of the most common scalar metrics used to characterize the diffusion

process. The most common of these are the mean diffusivity (MD) and the fractional anisotropy (FA) [56], given by

$$MD = \langle \lambda \rangle = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} = \frac{D_{xx} + D_{yy} + D_{zz}}{3}, \quad (4.6)$$

$$FA = \sqrt{\frac{3}{2} \frac{\sqrt{(\lambda_1 - \langle \lambda \rangle)^2 + (\lambda_2 - \langle \lambda \rangle)^2 + (\lambda_3 - \langle \lambda \rangle)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}}. \quad (4.7)$$

The MD provides an orientation-independent estimate of the overall diffusivity from a tissue, and has been shown to be clinically relevant, for example, for the detection of ischemic stroke lesions [57]. The FA index measures the degree to which the diffusion tensor differs from an isotropic sphere. Among its clinical uses are the quantification of white matter integrity.

Although DTI provides valuable insights into tissue microstructure, its underlying assumptions limit it to recovering a single diffusion peak for each voxel, causing it to fail in voxels containing more complex fiber configurations, such as crossing or kissing fibers. The next sections will introduce more complex models capable of representing such fiber configurations.

Nonparametric reconstruction methods

In order to move into more complex representations of the diffusion process taking place in a voxel, it is useful to consider the *ensemble average propagator* $p(\mathbf{r}, \Delta)$, a probability distribution describing the spatial spreading pattern of water molecules during the diffusion time Δ . Reconstruction models attempt to estimate the shape of this distribution, relying on modeling assumptions to different extents. DTI, for example, assumes that the propagator is a 3-dimensional Gaussian distribution, which prevents it from giving useful results in regions of multiple fiber populations. DTI can be generalized by assuming that p at each voxel can be represented as a combination of a fixed number of Gaussian components [58].

In contrast with these approaches, a number of nonparametric reconstruction methods have been proposed, which attempt to estimate the distribution p directly with no or minimal modeling assumptions. These approaches generally represent the diffusion in each voxel as a spherical function known as an *orientation distribution function* (ODF), which they try to estimate. The various methods differ in the type of ODF that they reconstruct, of which there are two types. Diffusion ODFs (dODFs) are an estimate of the propagator p , representing the degree to which water molecules diffuse in various directions. In contrast, fiber ODFs (fODFs) provide an estimate of the underlying fiber populations in each voxel. Although these represent different formalism, the most immediately apparent difference between both types of ODFs is that, although they are broadly similar, dODFs are comparatively smoother than

fODFs. Figure 4.1 shows a comparison between the diffusion tensor, dODFs, and fODFs.

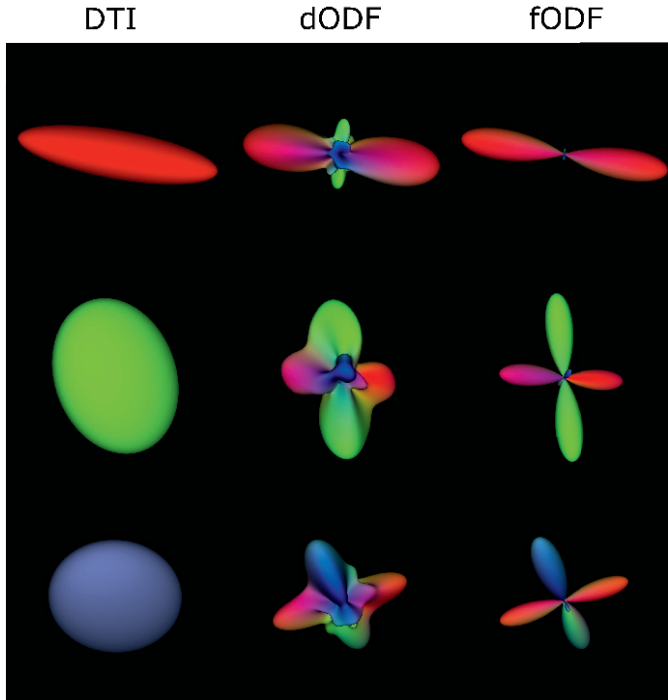


Figure 4.1: Comparison of various reconstruction methods. Left to right: diffusion tensor [54], dODF from constant solid angle QBI [59], and fODF from CSD [60]. Each row represents the same voxel. DTI is sufficient to represent single fiber populations, but more advanced methods are needed for voxels with multiple fiber populations.

What follows is a brief high-level description of some popular nonparametric reconstruction methods.

Diffusion spectrum imaging In diffusion spectrum imaging (DSI) [61, 62], the propagator p is estimated directly by performing an extensive Cartesian 3-dimensional sampling of $\mathbf{q} = \gamma\delta G\mathbf{g}$, where \mathbf{g} represents the orientation along which the diffusion-encoding gradients are applied. Taking the Fourier transform of this sampling yields a discrete 3-dimensional estimate of the propagator p . The value of the dODF along direction \mathbf{r} is then obtained by radially integrating the estimated p from the origin along \mathbf{r} .

The major disadvantage of DSI is that its Cartesian sampling scheme requires a very large number of measurements of \mathbf{q} , frequently exceeding 500.

Q-ball imaging Q-ball imaging (QBI) [63, 64] provides an alternative approach for measuring the dODF using more conventional spherical shell sampling, albeit with high density. Instead of estimating p using the Fourier transform, QBI produces samples of the dODF directly by applying the Funk-Radon transform on the spherical measurements of \mathbf{q} .

QBI has reduced acquisition and computation requirements in comparison to DSI, but results in some loss of accuracy in the estimated dODF.

Generalized q-sampling imaging Unlike the previous methods, generalized q-sampling imaging (GQI) [65] estimates a spin density function (SDF) directly from the q-space samples, and can be applied on data using either a Cartesian or a spherical shell sampling. The SDF is the product of the dODF and the spin density, from which the dODF can be obtained by normalization.

Spherical deconvolution Spherical deconvolution (SD) methods [66, 67] attempt to estimate the fODF directly by modeling the measured signal as the expected response for a single fiber population (the *fiber response function*) convolved with an ideal fODF representing the fiber populations present in a voxel. An estimate of the fODF is then obtained by deconvolution of measured signal and the fiber response function.

Although these methods excel at resolving peaks, they are quite susceptible to noise, resulting in the detection of spurious peaks. Constrained spherical deconvolution (CSD) [60] is a popular enhancement which imposes constraints to improve the conditioning of the deconvolution problem, with improvements to the angular resolution and noise susceptibility.

Spherical harmonics

Spherical harmonics (SH) are a set of orthonormal functions defined on the surface of a sphere. They provide a convenient and compact way to represent arbitrary spherical functions, and as such are commonly used to express ODFs.

The SH of order l and index m is given by

$$Y_{l,m}(\theta, \phi) = \sqrt{\frac{(2l+1)}{4\pi} \frac{(l-m)!}{(l+m)!}} P_{l,m}(\cos \theta) e^{im\phi}, \quad (4.8)$$

where $P_{l,m}$ is an l -th order, m -th degree associated Legendre polynomial. A complex function defined over the surface of the sphere can be written as a weighted sum of SH functions as

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l c_{l,m} Y_{l,m}, \quad (4.9)$$

where $c_{l,m}$ is a complex number. Truncation of the series at a finite $l_{\max} = L$ gives a low-frequency approximation of the function, akin to a truncated Fourier transform. $L = 8$ is commonly used for representing ODFs, resulting in $2(L+1)^2$ unique coefficients. For real-valued functions, this representation can be simplified, as $c_{l,m}^* = (-1)^m c_{l,-m}$, requiring the calculation of only $(L+1)^2$ unique coefficients. For antipodally-symmetric functions, such as ODFs, this can be further simplified, as $c_{l,m} = 0$ for odd values of l , requiring only $\frac{1}{2}(L+1)(L+2)$ unique coefficients (45 for $L = 8$).

Descoteaux et al. [68] employed a modified spherical harmonic basis that incorporates the antipodal symmetry, and uses only a single index $j(l, m) = (l^2 + l + 2)/2 + m$. This symmetric, real and orthonormal basis was the one used to represent fODFs in Paper IV.

4.4 Tractography

One of the main clinical and research uses of dMRI is the mapping of axonal connections in the brain, through a process known as *tractography* [69, 70, 71]. The fundamental idea behind tractography is that, to the extent that the local diffusion orientation information is a reflection of the orientation of underlying axonal fiber bundles, it can be iteratively “stepped through” in order to map global brain pathways.

Although a large variety of algorithms have been proposed, tractography approaches can be broadly placed into two categories:

Deterministic tractography Deterministic tractography is concerned with estimating individual connections in the brain, called *streamlines*. Starting from a specified seed point somewhere in the brain, the streamline is grown by iteratively sampling the peak diffusion direction and taking a small step in that direction, until a set of stopping criteria are met (see Figure 4.2).

Importantly, as the grid sampling pattern for which diffusion data is available is not anatomically representative of the axonal connections that are being mapped, the streamline tracking process is done in continuous space. This requires that interpolation be used in order to sample the peak diffusion direction in each iteration.

Probabilistic tractography The process of fiber tracking is subject to multiple sources of uncertainty due to the effects of noise and limited resolution. Furthermore, these errors propagate and accumulate over iterations, which can lead to increasing uncertainty as the tracking progresses. Probabilistic tractography approaches account for this uncertainty by considering a distribution of directions which streamlines can follow from any given position. Such methods allow the estimation of the uncertainty associated with alternative streamline paths.

Although tractography faces challenges with regards to its reliability and accuracy [72], its capacity to identify axonal fiber bundles in vivo has no counterpart, which has lead to its broad clinical use for, among other things, preoperative planning for brain tumors and other conditions [73].

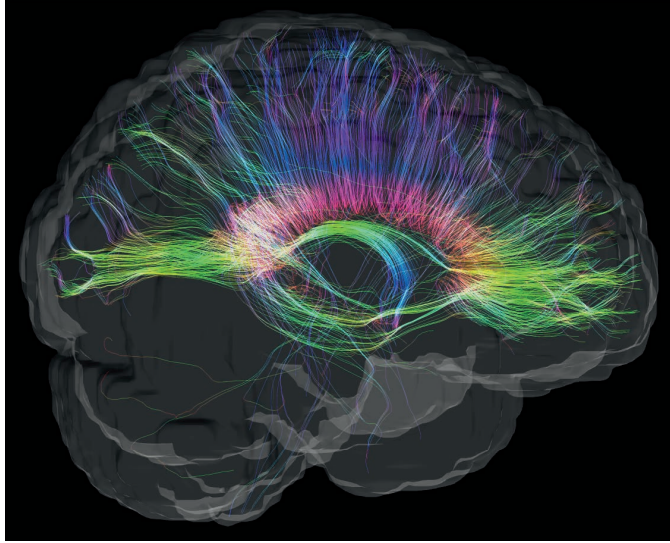


Figure 4.2: Streamlines obtained through deterministic tractography, with seeding of the corpus callosum.

5

Graph Signal Processing

Graphs are mathematical constructs that allow the definition of domains with arbitrary geometry and connectedness. As a tool for investigating the brain, graphs constitute a natural setting for representing and studying the structural and functional connectivity of brain regions. Furthermore, the rich graph-theoretical literature has provided a multitude of perspectives and metrics for characterizing and gaining insight from brain graphs [74].

The recently-developed field of graph signal processing (GSP) provides an alternative approach for the study of the brain. Rather than the graph being the main object of study, in GSP graphs provide the substrate on which signals can be defined and studied. Thanks to its continued development, GSP is largely analogous to traditional signal processing in regular domains, with operations such as the Fourier transform, spectral filtering, and wavelet decomposition of signals being readily available.

Within this thesis, GSP is the fundamental mathematical tool underlying the methods developed in Paper II. This chapter will provide a brief introduction to the GSP principles relevant to that work. The presentation is based on Paper II and our previous works [75, 76]. For the interested reader, a number of introductory and advanced treatments of this topic are available [77, 78].

5.1 Introduction to graphs

A graph is a discrete mathematical object consisting of a set of *vertices* and a set of *edges* connecting pairs of vertices. Together, these elements can be used to describe a set of objects and their relationships to each other.

Graphs can be classified on the basis of the types of connections that they allow between vertices. *Simple graphs* are those which allow only a single edge to exist between any pair of vertices, as well as disallowing loops, that is, edges connecting a vertex to itself. In *weighted graphs*, each edge has an associated weight specifying the degree of relatedness or connectedness it represents, whereas *unweighted graphs* do not allow specifying the strength of connections. Regarding the orientation of connections, graphs can be *undirected* when connections between vertices are symmetric, or *directed* in case connections can be asymmetric.

The ensuing discussion of GSP is focused specifically on simple, weighted, undirected graphs (see Figure 5.1). Formally, we define a simple, weighted, undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$ as a set \mathcal{V} of N_v vertices, a set \mathcal{E} of edges connecting pairs (i, j) of vertices, and an *adjacency matrix* \mathbf{A} , whose nonzero elements $a_{i,j}$ represent the weights of edges $(i, j) \in \mathcal{E}$. Given that the graph is undirected, $a_{i,j} = a_{j,i}$, i.e., \mathbf{A} is symmetric.

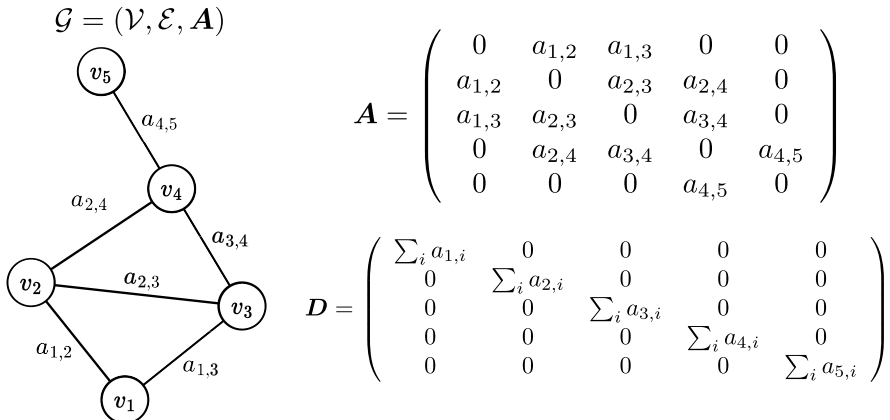


Figure 5.1: Example of weighted, undirected graph, with corresponding adjacency matrix \mathbf{A} and degree matrix \mathbf{D} .

Graphs provide a versatile domain on which to define signals. Common regular domains such as time-series or images can be represented as simple cyclic or lattice graphs, respectively. However, the same graph formalism is equally capable of representing irregular and abstract domains of arbitrary complexity.

5.2 Graph signals

In order to perform any kind of signal processing on graphs it is first necessary to bring the concept of *signal* to the graph domain. Informally, graph signals can be seen as vectors of data defined on a graph vertex set \mathcal{V} . Formally, for a given graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$, real graph signals are defined as elements of $\ell^2(\mathcal{G})$, denoting the Hilbert space of all square-summable vectors $\mathbf{f} \in \mathbb{R}^{N_v}$ with inner product

$$\langle \mathbf{f}_1, \mathbf{f}_2 \rangle = \sum_{n=1}^{N_v} \mathbf{f}_1[n] \mathbf{f}_2[n], \quad \forall \mathbf{f}_1, \mathbf{f}_2 \in \ell^2(\mathcal{G}) \quad (5.1)$$

and norm

$$\|\mathbf{f}\|_2^2 = \langle \mathbf{f}, \mathbf{f} \rangle = \sum_{n=1}^{N_v} |\mathbf{f}[n]|^2 < \infty, \quad \forall \mathbf{f} \in \ell^2(\mathcal{G}). \quad (5.2)$$

A graph signal $\mathbf{f} : \mathcal{V} \rightarrow \mathbb{R}$ is a vector whose n -th component is the value of the signal at the n -th vertex of the graph. Although in this work we focus on real signals, complex graph signals can be similarly defined in this way.

5.3 Graph spectral domain

We can define a diagonal *degree matrix* \mathbf{D} , with elements $d_{i,i} = \sum_j a_{i,j}$. The i -th diagonal element of \mathbf{D} is the degree of the i -th vertex of \mathcal{G} , that is, the sum of all edge weights associated to it.

The adjacency and degree matrices of \mathcal{G} can be used to define a desired *Laplacian matrix* of \mathcal{G} , which can be defined in either combinatorial form \mathbf{L} or normalized form \mathcal{L} as [79]:

$$\mathbf{L} = \mathbf{D} - \mathbf{A}, \quad (5.3)$$

$$\mathcal{L} = \mathbf{D}^{-1/2} \mathbf{L} \mathbf{D}^{-1/2} = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}. \quad (5.4)$$

Given that both definitions of the Laplacian matrix are symmetric and positive semi-definite, their eigendecomposition leads to a set of real non-negative eigenvalues:

$$\Lambda(\mathcal{G}) = \{0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{N_v} \stackrel{\text{def}}{=} \lambda_{\max}\}. \quad (5.5)$$

The associated eigenvectors $\{\mathbf{u}_l\}_{l=1}^{N_v}$ can be complex, but a real set can always be found, given that the Laplacian matrices are real and symmetric. The eigenvectors form an orthonormal basis, i.e., $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \delta_{i,j}$, and span the $\ell^2(\mathcal{G})$ space.

The set $\Lambda(\mathcal{G})$ represents the spectrum of the graph \mathcal{G} . The smallest eigenvalue is always 0, and the multiplicity of zeros matches the number of connected components of the graph. The maximum eigenvalue is unbounded for \mathbf{L} , whereas for \mathcal{L} it is always ≤ 2 . The spectrum of a graph shows similarities to frequency in the classical domain, with the eigenvectors associated with higher eigenvalues being less smooth (i.e., more variable) than those of lower eigenvalues [80]. However, as opposed to the complex exponentials in the classical domain, which constitute the basis of the Fourier domain, the Laplacian eigenvectors are not necessarily delocalized, especially those associated to higher eigenvalues. Figure 5.2 provides an illustration of the relationship between graph Laplacian eigenvalues and frequency.

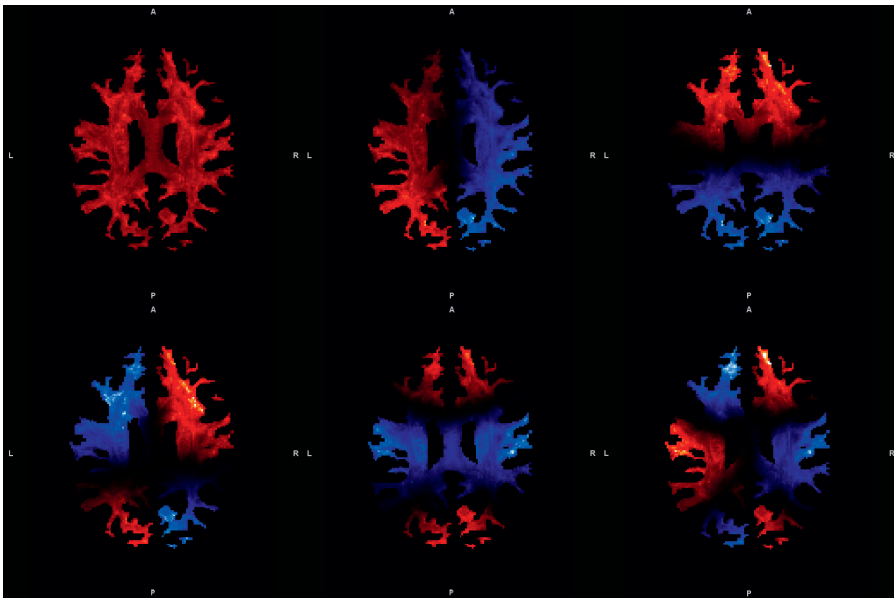


Figure 5.2: Axial slices of eigenvectors corresponding to first six Laplacian eigenvalues of a white matter graph (left to right and top to bottom). Red and blue correspond to positive and negative values, respectively. As can be seen, higher eigenvalues are related to higher spatial frequencies. Image reproduced from Abramian and Larsson [75].

Graph Fourier transform

The classical continuous Fourier transform is defined in terms of complex exponentials $e^{i\omega x}$, which, in turn, represent the eigenfunctions of the one-dimensional Laplacian operator:

$$\frac{d^2}{dx^2} e^{i\omega x} = -\omega^2 e^{i\omega x}. \quad (5.6)$$

The inverse Fourier transform can therefore be seen as the expansion of a continuous signal f in terms of the eigenfunctions of the Laplacian operator:

$$f(x) = \frac{1}{2\pi} \int \hat{f}(\omega) e^{i\omega x} d\omega. \quad (5.7)$$

In analogy to this, Hammond et al. [81] defined the graph Fourier transform as the expansion of a graph signal in terms of the eigenvectors of the graph Laplacian. For any graph signal $\mathbf{f} \in \ell^2(\mathcal{G})$, its graph Fourier transformed representation, denoted $\hat{\mathbf{f}} \in \ell^2(\mathcal{G})$, is defined as

$$\hat{\mathbf{f}}[l] = \mathcal{F}\{\mathbf{f}\} = \langle \mathbf{u}_l, \mathbf{f} \rangle = \sum_{n=1}^{N_v} \mathbf{f}[n] \mathbf{u}_l[n], \quad (5.8)$$

with the corresponding inverse transform given by

$$\mathbf{f}[n] = \mathcal{F}^{-1}\{\hat{\mathbf{f}}\} = \sum_{l=1}^{N_v} \hat{\mathbf{f}}[l] \mathbf{u}_l[n]. \quad (5.9)$$

Using this definition for the graph Fourier transform, it can be shown that the Parseval relation holds [81]:

$$\langle \mathbf{f}_1, \mathbf{f}_2 \rangle = \langle \hat{\mathbf{f}}_1, \hat{\mathbf{f}}_2 \rangle, \quad \forall \mathbf{f}_1, \mathbf{f}_2 \in \ell^2(\mathcal{G}). \quad (5.10)$$

Continuous spectral kernels

Having established the graph spectral domain and the graph Fourier transform, it is possible to extend some general signal processing procedures from the classical Euclidean setting to the graph setting. In particular, we are interested in filtering graph signals with a filter defined by its profile in the graph spectrum. A convenient way of defining such a filter is by sampling a continuous function $\mathcal{K} : [0, \lambda_{\max}] \rightarrow \mathbb{R}^+$, which we denote as a *spectral graph kernel*:

$$\hat{\mathbf{k}}[l] = \mathcal{K}(\lambda_l), \quad l = 1, \dots, N_v. \quad (5.11)$$

Signal processing on graphs

For any two graph signals $\mathbf{f}_1, \mathbf{f}_2 \in \mathbb{R}^{N_v}$, their convolution product is defined as

$$\begin{aligned} (\mathbf{f}_1 * \mathbf{f}_2)[n] &\stackrel{\text{def}}{=} \sum_{l=1}^{N_v} \hat{\mathbf{f}}_1[l] \hat{\mathbf{f}}_2[l] \mathbf{u}_l[n] \\ &= \mathcal{F}^{-1}\{\hat{\mathbf{f}}_1[l] \hat{\mathbf{f}}_2[l]\}. \end{aligned} \quad (5.12)$$

As can be seen, the convolution operation in the vertex domain is equivalent to multiplication in the spectral domain, just like in conventional signal processing. The filtering of a graph signal is then defined based on this operation. Given a graph signal $\mathbf{f} \in \ell^2(\mathcal{G})$ and a spectral graph kernel $\hat{\mathbf{k}}$, the filtered signal $(F_{\mathbf{k}}\mathbf{f})$ is obtained as

$$\begin{aligned} (F_{\mathbf{k}}\mathbf{f})[n] &\stackrel{\text{def}}{=} (\mathbf{k} * \mathbf{f})[n] \\ &= \sum_{l=1}^{N_v} \hat{\mathbf{k}}[l] \hat{\mathbf{f}}[l] \mathbf{u}_l[n]. \end{aligned} \quad (5.13)$$

The impulse response associated with a given spectral kernel $\hat{\mathbf{k}}$ can be obtained, as in conventional signal processing, by filtering an impulse signal δ . However, this impulse response is not shift-invariant, and varies depending on the vertex m on which the impulse is localized:

$$\hat{\delta}_m[l] = \langle \mathbf{u}_l, \delta_m \rangle = \mathbf{u}_l[m], \quad (5.14)$$

$$\begin{aligned}
 \psi_{\mathcal{K},m}[n] &\stackrel{\text{def}}{=} (F_{\mathbf{k}}\boldsymbol{\delta}_m)[n] \\
 &= \sum_{l=1}^{N_v} \hat{\mathbf{k}}[l] \hat{\boldsymbol{\delta}}_m[l] \mathbf{u}_l[n] \\
 &= \sum_{l=1}^{N_v} \underbrace{\hat{\mathbf{k}}[l] \mathbf{u}_l[m]}_{=\hat{\psi}_{\mathcal{K},m}} \mathbf{u}_l[n].
 \end{aligned} \tag{5.15}$$

where $\psi_{\mathcal{K},m}$ denotes the impulse response, also commonly referred to as an *atom*, associated to spectral kernel $\hat{\mathbf{k}}$ and localized at vertex m . Therefore, for a given spectral kernel there are N_v possible atoms, produced by filtering an impulse localized on each vertex of the graph.

Finally, it can be shown that filtering a signal \mathbf{f} with spectral kernel $\hat{\mathbf{k}}$ is equivalent to calculating the inner product between the signal and the atoms of the kernel:

$$\begin{aligned}
 (F_{\mathbf{k}}\mathbf{f})[m] &= \sum_{l=1}^{N_v} \hat{\mathbf{k}}[l] \hat{\mathbf{f}}[l] \mathbf{u}_l[m] \\
 &= \sum_{l=1}^{N_v} \hat{\psi}_{\mathcal{K},m}[l] \hat{\mathbf{f}}[l] \\
 &= \langle \hat{\psi}_{\mathcal{K},m}, \hat{\mathbf{f}} \rangle \\
 &\stackrel{(5.10)}{=} \langle \psi_{\mathcal{K},m}, \mathbf{f} \rangle.
 \end{aligned} \tag{5.16}$$

Polynomial kernel approximation

Both the filtering of graph signals and their decomposition onto sets of spectral kernels are operations realized through (5.16). However, this equation depends on the availability of the full set of atoms for every kernel, which in turn requires the calculation of all the eigenvectors of the Laplacian matrix. Such an approach becomes infeasible for larger graphs. Instead, a fast approximation algorithm can be used.

Let \mathcal{P} be a polynomial approximation of kernel \mathcal{K} . For a graph signal \mathbf{f} , its filtering with kernel \mathcal{K} , or equivalently, its decomposition coefficients when projected on the atoms of \mathcal{K} , can be approximated using \mathcal{P} as

$$\tilde{\mathbf{c}}_{\mathcal{K}} = \sum_{l=1}^{N_v} \mathcal{P}(\lambda_l) \hat{\mathbf{f}}[l] \mathbf{u}_l \tag{5.17}$$

$$= \mathcal{P}(\mathcal{L}) \sum_{l=1}^{N_v} \hat{\mathbf{f}}[l] \mathbf{u}_l \tag{5.18}$$

$$= \mathcal{P}(\mathcal{L}) \mathbf{f}, \tag{5.19}$$

where $\tilde{\mathbf{c}}_{\mathcal{K}} \in \ell^2(\mathcal{G})$ with $\tilde{\mathbf{c}}_{\mathcal{K}}[m] = c_{\mathcal{K},m}$. In (5.18) we use the fact that for any polynomial \mathcal{P}

$$\mathcal{L} \mathbf{u}_l = \lambda_l \mathbf{u}_l \Rightarrow \mathcal{P}(\mathcal{L}) \mathbf{u}_l = \mathcal{P}(\lambda_l) \mathbf{u}_l. \tag{5.20}$$

This approximation has the benefit that it does not require the calculation of the Laplacian eigenvectors. Instead, a polynomial of the Laplacian matrix is applied to the signal, which can be efficiently implemented with matrix-vector multiplication. In Hammond et al. [81], a truncated Chebyshev expansion was used as it has the benefit of approximating a minimax polynomial. This minimizes an upper bound on the approximation error in the coefficients and is the method we used in Paper II.

6

Deep Learning

In the last few decades, machine learning methods, and in particular the subset commonly referred to as *deep learning*, have exploded in popularity in all fields for their combination of relative ease of use with remarkable versatility and performance. Medical imaging is no exception, with deep learning gaining traction for many different tasks ranging from disease classification and organ segmentation to view synthesis for surgical recording [82].

It is not surprising that deep learning will constitute a substantial element in a work concerned with modern approaches to medical imaging. This chapter will introduce the deep learning aspects relevant to Papers IV and V. For an in-depth treatment of deep learning, several reference texts are available [83, 84].

6.1 Machine learning

Before discussing neural networks, it is useful to introduce some general machine learning concepts. *Machine learning* refers to a set of model-based optimization methods that *learn* to perform tasks in a data-driven way, that is, the model parameters are iteratively optimized in a *training* process to maximize the task performance of the model on a given set of data.

A machine learning model can generally be seen as a function that transforms a set of inputs \mathbf{x} into a set of outputs \mathbf{y} on the basis of some parameters $\boldsymbol{\theta}$:

$$\mathbf{y} = f(\mathbf{x}, \boldsymbol{\theta}). \quad (6.1)$$

The learning aspect consists in finding the set of parameters $\boldsymbol{\theta}$ that best fulfill the task of the model. While machine learning can be applied to a wealth of different tasks, we will narrowly focus on the subset of *supervised learning* tasks. This describes task where for every training input \mathbf{x} there is a known target output \mathbf{t} that the network is trained to generate. This encompasses common tasks such as classification and regression, among many others.

Loss functions

Machine learning relies on various optimization strategies to arrive at the optimal parameters for a model. For this to be possible, it is first necessary to formalize the goals of the machine learning model and express them in mathematical terms. This is done with a *loss function* L , which provides a quantitative expression for the “badness” of fit of a given model. In the case of supervised learning, the loss function usually expresses the degree to which the model outputs \mathbf{y} differ from the target outputs \mathbf{t} . Optimal model parameters are then obtained by minimizing the loss function on a set of training data:

$$\tilde{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} L(\mathbf{y}, \mathbf{t}) = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} L(f(\mathbf{x}, \boldsymbol{\theta}), \mathbf{t}). \quad (6.2)$$

A trained model can then be applied to make predictions on data outside of its training set.

Optimizers

Non-trivial machine learning models are generally complex and nonlinear, requiring numerical methods to arrive at optimal parameters. In the context of neural networks, the algorithm responsible for minimizing the loss function is called the *optimizer*. The fundamental optimizer relevant to neural networks is *gradient descent*, a greedy algorithm which consists on iteratively generating model outputs for the training data, calculating their loss, and applying a small update on the model parameters in the direction of the negative gradient of the loss with respect to the parameters:

$$\mathbf{y}^{(e)} = f(\mathbf{x}, \boldsymbol{\theta}^{(e)}), \quad (6.3)$$

$$\boldsymbol{\theta}^{(e+1)} = \boldsymbol{\theta}^{(e)} - \alpha \frac{dL(\mathbf{y}^{(e)}, \mathbf{t})}{d\boldsymbol{\theta}^{(e)}}, \quad (6.4)$$

where e is the iteration index, known as the current *epoch*, and α represents the *learning rate*, specifying the size of the parameter update.

In gradient descent, the loss is evaluated on all the training data at once. Stochastic gradient descent (SGD) [85] is an essential development of this method, where gradient updates are calculated on portions of the data, known as *batches*, with a full epoch being completed when all the training data has been used once. This modification presents several distinct advantages over ordinary gradient descent. First, as parameters are updated multiple times per epoch, training is faster and more computationally-efficient. Furthermore, although SGD parameter updates are not globally-optimal for the whole dataset, a certain degree of noise in the updates can be beneficial for avoiding suboptimal local minima and exploring a wider region of the parameter space. Lastly, computational limitations (e.g., limited CPU or GPU memory) may preclude calculating gradient updates from an entire dataset at once, making the use of batches a necessity in many cases.

The SGD algorithm forms the basis of a number of modern optimizers with better stability and convergence properties. Among the most significant of these are RMSProp [86] and Adam [87].

Model generalization

Machine learning models are trained by minimizing the loss on a training dataset. However, training data performance is not necessarily a good estimate of *generalization performance*, that is, the performance of the model on new, yet unseen data. One of the main reasons for this is *overfitting*, which describes a situation when models learn to reproduce minute aspects of the training data that are not representative of all similar data. Overfitting is generally caused by an excess of model parameters with respect to the number of training datapoints available. *Regularization* describes the set of approaches aimed at minimizing overfitting.

The simplest strategy for measuring the generalization performance of machine learning models is based on dividing the available data into training, validation and test sets. Training data is used to calculate parameter updates, but the model performance on validation data is also periodically evaluated during training, giving an unbiased estimate of generalization performance. Validation data is then used as the primary measure of model performance, and to determine the optimal time to stop training.

It is commonly desirable to compare the performance of multiple models. One common case is the search for optimal model *hyperparameters*, non-learnable model parameters that need to be set before training and can have a substantial performance effect (e.g., the learning rate of the optimizer). When used for model comparison, validation performance itself becomes biased, as it favors models that happen to perform well on the specific set of validation datapoints. Because of this, after selecting a model on the basis of validation performance, the final estimate of generalization performance is obtained by evaluating it on test data reserved for this purpose.

6.2 Neural networks

Neural networks are a multipurpose machine learning framework in which data is transformed by a sequence of operations, called *layers*. The specific choice and arrangement of layers define the *architecture* of the network. Furthermore, they serve as one of the primary means to categorize neural networks, either by their reliance on certain types of layers (e.g., dense neural networks, convolutional neural networks), or by features of their arrangement (e.g., recurrent neural networks are characterized by the presence of feedback connections).

When unqualified, the term *neural networks* generally refers to dense neural networks (DNNs), also known as artificial neural networks or fully-

connected neural networks. This is the original style of neural networks, inspired in the workings of biological neurons, and with origins in the *perceptron* algorithm [88]. Each network layer consists of a number of artificial neurons, each of which produces as output a linear combination of its inputs, further subjected to a nonlinear *activation function*. The input features of the model are the inputs of the first layer of neurons, with every subsequent layer taking as inputs the outputs of the previous layer.

The output y of the operation performed by a single neuron can be expressed as

$$y = \sigma(\mathbf{w}^T \mathbf{x} + b), \quad (6.5)$$

where \mathbf{x} is a $[N_i \times 1]$ vector of input features, \mathbf{w} is a $[N_i \times 1]$ vector of linear *weights*, b is a scalar bias, and σ is a nonlinear activation function. For a layer of N_o neurons, the same can be written as

$$\mathbf{y} = \sigma(\mathbf{W}\mathbf{x} + \mathbf{b}), \quad (6.6)$$

where \mathbf{X} is a $[N_o \times N_i]$ weight matrix and \mathbf{b} is a $[N_o \times 1]$ vector of biases. Neurons in subsequent layers perform a similar operation, taking as input the neuron outputs from the previous layer. For example, a two layer network with output \mathbf{y} is described by

$$\mathbf{z} = \sigma(\mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)}), \quad (6.7)$$

$$\mathbf{y} = \sigma(\mathbf{W}^{(2)}\mathbf{z} + \mathbf{b}^{(2)}) \quad (6.8)$$

$$= \sigma(\mathbf{W}^{(2)}\sigma(\mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)}) + \mathbf{b}^{(2)}), \quad (6.9)$$

where superindices represent the layer number. This network is schematically presented in Figure 6.1. The number of neurons in a layer is referred to as its *width*, while the number of layers in a network is referred to as its *depth*. The weights and biases of each layer constitute the set of model parameters $\boldsymbol{\theta}$ to be optimized by training the network.

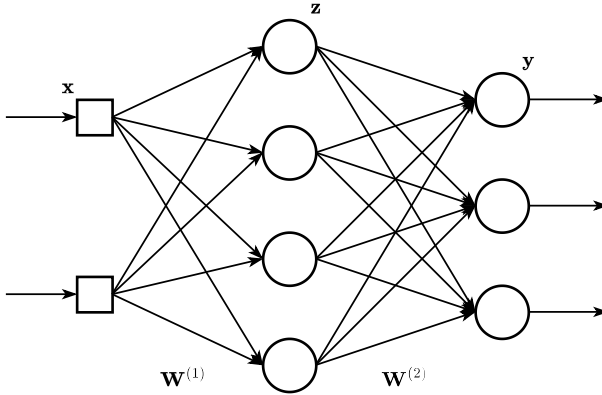


Figure 6.1: Example of a two layer neural network. Square nodes represent inputs, while circular nodes are neurons. Layer biases are omitted for simplicity.

Activation functions

Neural networks are remarkable in that, given sufficient width [89, 90] or depth [91], they are capable of approximating any well-behaved function. However, these results crucially depend on the use of nonlinear activation functions after each layer, which enable the network to learn nonlinear mappings. In the absence of activation functions, neural networks of any depth will be constrained to learning linear mappings of their inputs.

There are multiple aspects involved in the choice of an activation function. Some of the most common activation functions will be briefly described.

Sigmoid One of the early activation functions, the sigmoid has since been phased out in favor of variations of the rectified linear unit, due to their improved convergence properties. The sigmoid function constrains its output to the $(0, 1)$ range, which makes it useful as an output activation for binary classification networks.

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \in (0, 1). \quad (6.10)$$

Hyperbolic tangent The hyperbolic tangent function has a similar shape and properties to the sigmoid function, but gives output in the $(-1, 1)$ range instead.

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \in (-1, 1). \quad (6.11)$$

Softmax The softmax takes a vector of inputs and returns a vector of positive elements which sums to 1. As such, it is used in the output of multi-class

classification networks.

$$\text{softmax}(\mathbf{x}) = \frac{e^{x_i}}{\sum_j e^{x_j}} \in (0, 1). \quad (6.12)$$

ReLU The rectified linear unit (ReLU) is a general purpose activation function which zeros out negative values from its input. It is arguably the most popular activation function currently in use, as its large positive gradients are favorable for training.

$$\text{ReLU}(x) = \max(0, x) \in (0, \infty). \quad (6.13)$$

Leaky ReLU The use of ReLU activations can leave some network neurons constantly operating in the negative range, where they produce no output. These are known as *dead neurons*, as the lack of a gradient for the negative input range of ReLU makes these neurons irrecoverable. Leaky ReLU activations introduce a small negative gradient α for the negative input range to prevent this problem.

$$\text{Leaky ReLU}(x) = \max(-\alpha x, x) \in (-\infty, \infty). \quad (6.14)$$

Layer types

The main learning units of a DNN are *dense layers*, implementing the mapping between successive layers of neurons described in Eq. (6.6). However, there exist additional layer types that fulfill other purposes in the network:

Dropout layers Dropout layers [92] generally follow dense layers, and randomly set a fraction of their inputs to zero during training, effectively disabling certain neurons in each training iteration. This has a regularizing effect on the network, forcing it not to rely too strongly on any one set of connections.

Normalization layers Normalization layers [93, 94, 95] are generally placed after a linear transformation but before the activation function in a dense layer. These layers have the effect of replacing the mean and standard deviation of a set of data with learned values β and γ , respectively. Thus, they update the data range to a more useful one learned during training, which increases the speed of training convergence and improves gradient flow (discussed further).

Although many of these layers were introduced in the context of convolutional neural networks (discussed further), they can equally be applied to DNNs.

Practical training aspects

Neural networks are trained using iterative optimizers in accordance to what was previously discussed. However, several of their unique aspects have important consequences for training performance, driving many practical decisions of practitioners as well as new developments in the field.

Each training iteration begins by sending training data through the network to generate output, in what is known as the *forward pass*. It is then necessary to calculate the gradient of the loss with respect to the network weights, which is done using the *backpropagation* algorithm, so called because the gradient flows backwards from the network output to the input. The calculation of these gradients constitutes the *backwards pass* of training.

The successful backwards *flow* of the gradient is essential for the network to converge to a useful solution, and generally implies achieving a usable middle ground between two undesirable situations: *exploding gradients*, where excessively large gradients cause the training to diverge, and *vanishing gradients*, where zero or negligible gradients prevent any training from happening. Multiple factors influence the proper flow of gradients:

- The gradient at each layer is generally proportional to the layer output. Therefore, excessively large or small layer outputs can result in exploding or vanishing gradients, respectively.
- Gradients are also affected by the activation function at each layer. Specifically, incoming gradients are multiplied by the derivative of the activation function of a given layer. This can cause the gradients to become nullified, as in the case of sigmoidal functions with large positive or negative inputs.
- As the gradient flows backwards, exploding or vanishing gradient effects are compounded for each successive layer. This constitutes a limiting factor on the practical depth of a network, beyond which training becomes unstable or stops happening.

In practice, these issues are addressed by a combination of several interacting factors:

Activation function As was discussed previously, ReLU and its variants have largely replaced sigmoidal functions as the standard due to their improved gradient flow properties.

Data normalization The specific range of values that input data takes determines the size of layer activation, impacting gradient flow. It is common to normalize the input data to have zero mean and unit standard deviation, or conversely to shift it to a fixed small range, such as $[-1, 1]$.

Weight initialization Prior to training, network weights are initialized to random values. As the output of a layer is a function of the inner product of input and weights, the optimal initialization scheme depends on the range of the input data, the number of neurons, and the chosen activation function [96, 97].

Network architecture Certain layers (e.g., normalization layers) and network architectures (e.g., residual networks) are more conducive to the flow of gradients.

It can be said that *deep learning*, the development of ever deeper neural networks capable of fulfilling increasingly complex tasks, has largely been enabled by the progressive development of methods for managing the flow of gradients throughout the full length of a network.

6.3 Convolutional neural networks

Although theoretically capable of learning any well-behaved function, practical considerations make DNNs unsuited to many tasks. For a wide variety of image- and signal-based tasks, convolutional neural networks (CNNs) have been the dominant paradigm for more than two decades. As their name suggests, rather than linearly combining all inputs, CNNs are fundamentally based on the convolution operation, which is given by

$$y[n] = (x * h)[n] = \sum_m x[m]h[n - m] \quad (6.15)$$

for 1-dimensional signal x and convolutional kernel h , and can be extended to any number of dimensions. Convolutional kernels are also referred to as *filters*.

CNNs can be seen as a subset of DNNs, where neurons are replaced by learned filters. This offers distinct advantages for image and other signal data:

- Convolutional kernels have limited spatial extents, and as such, their output is based on a limited set of points from their input. Unlike the full weight matrix \mathbf{W} of DNNs, CNNs have very sparse weight matrices, which makes them more parameter-efficient. Furthermore, the number of parameters is specified by the number and size of filters, and is therefore decoupled from the size of the input image.
- Filters are applied by sliding them over the extent of the input image, giving as output a *feature map*. Thus, unlike the distinct rows of the \mathbf{W} matrix of DNNs, in CNNs the rows are identical but shifted, as outputs at different places are based on different pixels but the same filter weights. This is known as *weight sharing*, and is a further way in which CNNs are more parameter-efficient than DNNs.

- Learned filter kernels can perform ordinary image-processing operations, customized to the particular needs of the task.

Spatial context in CNNs

As the output of a convolution is based on a limited set of input data, it is only based on a limited spatial context, which can prevent the network from making determinations based on distant regions of an image. For a given point of a layer output, *receptive field* describes the region of the input data on which it is based. The receptive field increases as a function of layer depth, as every layer output is based on a region of the previous layers' output.

The receptive field of a CNN can also be increased by forcibly decreasing the size of a feature map, causing subsequent convolutions to cover a larger area of the input. This can be done by the use of pooling layers, which replace each $[N \times N]$ group of pixels with a summary metric of it (e.g., maximum, average), or by the use of strided convolutions.

It is often also necessary to increase the size of feature maps, such as for encoder-decoder architectures or in the generation of images from small latent vectors (discussed further). This is done by the use of upsampling convolution layers, such as transposed convolution and resize-convolution.

CNN architectures

Like DNNs, CNNs are primarily composed of a set of convolutional layers arranged in sequence, each of which is made of a number of learned filters. Each input image is convolved with each filter in a layer, producing feature maps as output. The set of feature maps produced by filters from one layer constitutes the input for the next layers' filters.

CNN filters are generally very small, with sizes of $[3 \times 3]$ being most common. As such, they act as local feature detectors, with early layers detecting basic features such as lines, edges and corners in various orientations. Subsequent layers combine these simple features into increasingly complex and high level features. In this way, features are learned in a hierarchical pattern.

For certain tasks like classification, a common solution uses a stack of convolutional layers as a *feature extractor*, with several dense layers performing the classification on the basis of the obtained features (see Figure 6.2). On the other hand, tasks where the output takes the form of image data, as the input, can often be performed with *fully-convolutional networks*, i.e., without the use of dense layers.

Encoders and *decoders* are common elements in CNN architectures, with the former progressively reducing the size of feature maps and the latter increasing them. Both can be combined in the encoder-decoder architecture, which features a spatial bottleneck in the middle that leads to a very compressed representation of the input data. The U-Net architecture [98] is a

very successful example of this architecture, originally developed for medical image segmentation and since then applied to multiple image tasks.

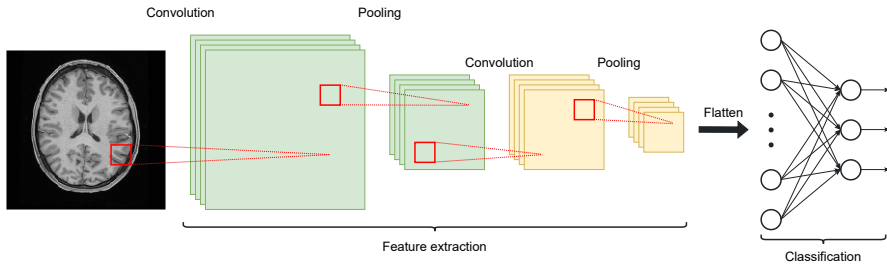


Figure 6.2: Example CNN architecture for image classification. The network consists of a convolutional portion for feature extraction, followed by a set of dense layers for classification.

6.4 Generative adversarial networks

Generative adversarial networks (GANs) are a type of deep learning model used to generate realistic images in the style of a provided dataset. Specifically, during training the model learns the distribution of the training data. The trained model can then generate new samples from this distribution using random seed vectors as input. In contrast with earlier generative models, such as variational autoencoders, GANs are capable of generating highly realistic images. Figure 6.3 presents synthetic brain images generated with a 3D GAN.

The fundamental idea of GANs resides in their novel loss function, which requires training two separate networks: a generator, G , responsible for creating synthetic images from noise seeds, and a discriminator, D , which judges the realism of images. The two networks are trained simultaneously in an adversarial manner, where the discriminator judges the images generated by the generator, and the generator updates its weight based on this judgment. For image data, both the generator and discriminator are CNNs.

GANs were originally introduced in 2014 by Goodfellow et al. [99]. Since then the fundamental idea has been developed further and extended to more tasks. Conditional GANs [100] were proposed in the same year, where models learn conditional distributions of data, and can be sampled to generate data of specific classes. These, in turn, led to the development image-to-image GANs.

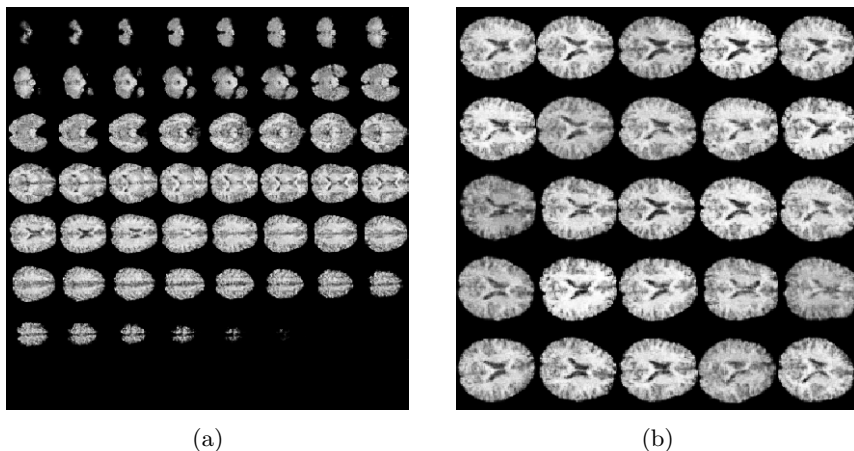


Figure 6.3: Synthetic T_1 -weighted brain volumes generated with a 3D GAN. (a) Axial slices from a single synthetic volume. (b) Single axial slice from multiple synthetic volumes.

Image-to-image GANs

One of the most exciting developments in the fields of GANs is the development of image-to-image models, which enable the “translation” of images between different domains or representations. Although this is a somewhat abstract idea, we can consider cases where a piece of data can be represented in two different domains $\{A, B\}$ between which there exists some mapping. For example, the same scene \mathbf{x} can be represented as a color image, \mathbf{x}_A , or a grayscale image, \mathbf{x}_B . An image-to-image GAN can be trained to convert data between these domains, i.e., $G_{A \rightarrow B}(\mathbf{x}_A) = \hat{\mathbf{x}}_B$. In practice, image-to-image GANs can be trained to perform difficult and underdetermined tasks, such as colorization and inpainting. Two particularly influential models will be highlighted:

Pix2Pix Pix2Pix [101] is a *paired* image-to-image translation GAN. As it uses supervised loss terms, it requires the training data to be paired, that is, the same data needs to be expressed in both domains. Once trained, it can convert data from one domain to another.

CycleGAN In contrast to Pix2Pix, CycleGAN [102] is an *unpaired* image-to-image translation network, that is, it requires training data from both domains, but it does not require it to be the same data. This widely expands its range of applicability, as there are numerous applications where paired data is effectively impossible to obtain. For example, it can be used to translate pictures between the styles of various painters, a scenario which, for a paired model, would have required multiple painters to have painted identical scenes in their own style.

In order to achieve unpaired translation between domains, CycleGAN relies on two sets of image-translation GANs: one from domain A to domain B , and one in the opposite direction. Having translation networks in both directions, CycleGAN incorporates a cycle-consistency loss that ensures that the transformations given by both networks are invertible:

$$G_{A \rightarrow B}(G_{B \rightarrow A}(\mathbf{x}_A)) \approx \mathbf{x}_A, \quad (6.16)$$

$$G_{B \rightarrow A}(G_{A \rightarrow B}(\mathbf{x}_B)) \approx \mathbf{x}_B. \quad (6.17)$$

Image-to-image GANs have found multiple applications in medical imaging, such as translating between image modalities (e.g., CT to MRI or vice versa, conversion between MRI modalities, see Figure 6.4), generating high-resolution images from low-resolution images, correcting image distortions, and performing segmentation [103].

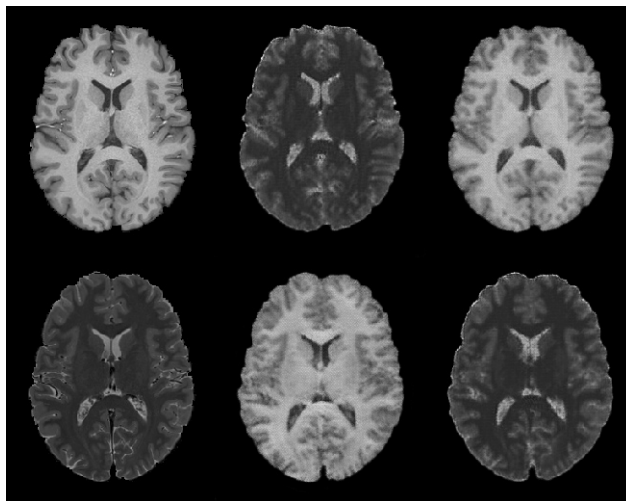


Figure 6.4: Image-to-image translation between T_1 - and T_2 -weighted MRI images using CycleGAN, which simultaneously learns to translate both domains into each other. Top: T_1 to T_2 translation. Bottom: T_2 to T_1 translation. Left: original image. Center: translated image. Right: cycle consistency image, obtained by translating the central image back to its original domain.

7

Summary of papers

This chapter provides a brief summary of the papers included in this thesis.

7.1 Paper I - Anatomically informed Bayesian spatial priors for fMRI analysis

In this paper we propose a modification to the Bayesian fMRI data analysis framework developed by Penny et al. [36, 47, 48] and refined by Sidén et al. [51]. The existing model uses isotropic spatial priors, where every voxel is considered to be equally related to all its neighbors, causing fMRI signal to be averaged across anatomical boundaries such as sulci. We propose to make these priors anatomically adaptive by incorporating anatomical information from T_1 -weighted images. We estimate the local orientation of anatomical structures using structure tensor methods, and incorporate it into two proposed 2-dimensional adaptive priors which prevent the averaging of signal across anatomical boundaries. The priors are evaluated on data from a single subject from the Human Connectome Project (HCP), with the resulting regressor coefficients and posterior probability maps more closely following the anatomical features of the brain.

7.2 Paper II - Diffusion-informed spatial smoothing of fMRI data in white matter using spectral graph filters

As was discussed in Chapter 3, activation mapping in white matter is a controversial topic, and in order to be successful may require the development of tailored methods [18]. It has been observed that the correlation structure of the BOLD signal in white matter is anisotropic, and its orientation follows that of the underlying axonal fiber bundles [22, 23, 24]. This fact suggests that isotropic Gaussian smoothing is suboptimal for activation mapping in white matter. In this paper, we propose a diffusion-informed, GSP-based spatial smoothing approach that accounts for this correlation structure. We specify a dense, voxel-level graph representation of brain white matter. The connec-

tions between adjacent voxels are weighted by the coherence in the orientation of their fiber populations, estimated from dODFs calculated from dMRI data. fMRI signals are defined on this graph, and low-pass filtered with graph heat kernel filters. Due to the shift-variant properties of filters defined on the graph spectral domain, and in combination with the specific graph weighting scheme used, filters instantiated at any graph vertex (i.e., voxel) adapt to the shape of the underlying fiber bundles. The proposed method is exhaustively tested on data from 95 HCP subjects, showing improved activation mapping performance for both single-subject and group analyses.

7.3 Paper III - Evaluation of inverse treatment planning for Gamma Knife radiosurgery using fMRI brain activation maps as organs at risk

Stereotactic radiosurgery (SRS) is a noninvasive surgical modality where tissue is ablated by the applications of focused beams of ionizing radiation, and is a common treatment option for small intracranial tumors. However, the use of ionizing radiation can cause damage to important brain regions. SRS treatment planning requires the careful specification of placement, intensity, and collimator settings for a potentially large set of radiation shots. Fortunately, the introduction of inverse planning systems automates the treatment planning process, producing treatment plans by formulating and solving an optimization problem. This facilitates the introduction of organs at risk (OARs), critical brain regions where radiation doses should be limited, to the treatment planning process. These typically include anatomically-identifiable regions, such as the brainstem and optic nerve. However, the use of functional OAR to spare functionally-relevant areas of the brain is more rare.

The Gamma Knife, developed by the company Elekta, is one of the principal radiation delivery systems for SRS. In this paper, done in collaboration with Elekta, we make use of the recently-introduced Lightning inverse planning system for Gamma Knife SRS to examine the possibility of incorporating fMRI-derived functional OARs into the treatment planning process. We implement a pipeline for analyzing fMRI data, transferring the resulting activation maps to the treatment planning software, and using them as OARs for treatment planning. We evaluate the effects of incorporating such functional OARs on brain tumor data from 5 patients, with results showing that the Lightning optimizer successfully follows the dose constraints imposed by functional OARs, reducing the radiation dose incident on eloquent brain regions.

7.4 Paper IV - Super-resolution mapping of anisotropic tissue structure with diffusion MRI and deep learning

Diffusion MRI is limited by the spatial resolution of the available data. In this paper, we propose a deep-learning based approach to upsample fODFs obtained through CSD. We use dMRI data from 95 HCP subjects to generate CSD fODFs, and convert them to a SH representation. In parallel, we downsample the dMRI data by averaging $2 \times 2 \times 2$ voxel regions to produce a low-resolution version of the data, and apply the same procedure to generate SHs of CSD fODFs. We use both sets of data to train DNN networks that take a small cube of low-resolution data as input and predict the high-resolution version of the central voxel. By applying this single-voxel model throughout the brain, it is possible to double the resolution along each spatial axis of a brain fODF dataset. Our results show that the proposed model, while relatively simple, shows upsampling performance superior to spline interpolation of any order. In addition, the model has a substantial denoising effect, removing spurious fODF peaks present even on the original high-resolution data.

7.5 Paper V - Refacing: reconstructing anonymized facial features using GANs

The removal of facial features is a standard anonymization procedure for MRI data, especially important for open datasets, as the 3-dimensional volumes can be rendered, revealing the face of subjects. In the final paper we investigate the potential of image-to-image GANs for reverting the facial anonymization of MRI datasets. We apply two different face-removal methods on T_1 -weighted volumes from 581 subjects from the IXI dataset, which is available without any facial anonymization. We then train CycleGAN models to translate 2D sagittal slices of data between the original and anonymized domains, which learn to deanonymize the data. Our results show that the tested anonymization algorithms can be partially reverted, illustrating that deep learning models can pose some danger to data anonymization procedures.

Bibliography

- [1] David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy EJ Behrens, Essa Yacoub, Kamil Ugurbil, Wu-Minn HCP Consortium, et al. “The WU-Minn human connectome project: an overview.” In: *Neuroimage* 80 (2013), pp. 62–79.
- [2] Cyril R Pernet, Krzysztof J Gorgolewski, Dominic Job, David Rodriguez, Ian Whittle, and Joanna Wardlaw. “A structural and functional magnetic resonance imaging dataset of brain tumour patients.” In: *Scientific data* 3.1 (2016), pp. 1–6.
- [3] *IXI Dataset*. URL: <https://brain-development.org/ixi-dataset/>.
- [4] Russell A Poldrack and Krzysztof J Gorgolewski. “Making big data open: data sharing in neuroimaging.” In: *Nature neuroscience* 17.11 (2014), pp. 1510–1517.
- [5] Stephen J Egle, Ben Marwick, Yaroslav O Halchenko, Michael Hanke, Shoaib Sufi, Padraig Gleeson, R Angus Silver, Andrew P Davison, Linda Lanyon, Mathew Abrams, et al. “Toward standard practices for sharing computer code and programs in neuroscience.” In: *Nature neuroscience* 20.6 (2017), pp. 770–773.
- [6] Zhi-Pei Liang and Paul C Lauterbur. *Principles of magnetic resonance imaging*. SPIE Optical Engineering Press Bellingham, 2000.
- [7] Robert W Brown, Y-C Norman Cheng, E Mark Haacke, Michael R Thompson, and Ramesh Venkatesan. *Magnetic resonance imaging: physical principles and sequence design*. John Wiley & Sons, 2014.
- [8] Daniel D Traficante. “Relaxation. Can T2, be longer than T1?” In: *Concepts in Magnetic Resonance* 3.3 (1991), pp. 171–177.
- [9] Bernd André Jung and Matthias Weigel. “Spin echo magnetic resonance imaging.” In: *Journal of Magnetic Resonance Imaging* 37.4 (2013), pp. 805–817.
- [10] Edward O Stejskal and John E Tanner. “Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient.” In: *The journal of chemical physics* 42.1 (1965), pp. 288–292.
- [11] Nikos K Logothetis. “What we can do and what we cannot do with fMRI.” In: *Nature* 453.7197 (2008), pp. 869–878.
- [12] Seiji Ogawa, Tso-Ming Lee, Asha S Nayak, and Paul Glynn. “Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields.” In: *Magnetic resonance in medicine* 14.1 (1990), pp. 68–78.
- [13] Seiji Ogawa, Tso-Ming Lee, Alan R Kay, and David W Tank. “Brain magnetic resonance imaging with contrast dependent on blood oxygenation.” In: *proceedings of the National Academy of Sciences* 87.24 (1990), pp. 9868–9872.
- [14] Seiji Ogawa, David W Tank, Ravi Menon, Jutta M Ellermann, Seong G Kim, Helmut Merkle, and Kamil Ugurbil. “Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging.” In: *Proceedings of the National Academy of Sciences* 89.13 (1992), pp. 5951–5955.
- [15] Kenneth K Kwong, John W Belliveau, David A Chesler, Inna E Goldberg, Robert M Weisskoff, Brigitte P Poncelet, David N Kennedy, Bernice E Hoppel, Mark S Cohen, and Robert Turner. “Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation.” In: *Proceedings of the National Academy of Sciences* 89.12 (1992), pp. 5675–5679.

- [16] Peter A Bandettini, Eric C Wong, R Scott Hinks, Ronald S Tikofsky, and James S Hyde. "Time course EPI of human brain function during task activation." In: *Magnetic resonance in medicine* 25.2 (1992), pp. 390–397.
- [17] Nikos K Logothetis, Jon Pauls, Mark Augath, Torsten Trinath, and Axel Oeltermann. "Neurophysiological investigation of the basis of the fMRI signal." In: *nature* 412.6843 (2001), pp. 150–157.
- [18] Jodie R Gawryluk, Erin L Mazerolle, and Ryan CN D'Arcy. "Does functional MRI detect activation in white matter? A review of emerging evidence, issues, and future directions." In: *Frontiers in neuroscience* 8 (2014), p. 239.
- [19] Lukas A Grajauskas, Tory Frizzell, Xiaowei Song, and Ryan CN D'Arcy. "White matter fMRI activation cannot be treated as a nuisance regressor: Overcoming a historical blind spot." In: *Frontiers in neuroscience* 13 (2019), p. 1024.
- [20] John C Gore, Muwei Li, Yurui Gao, Tung-Lin Wu, Kurt G Schilling, Yali Huang, Arabinda Mishra, Allen T Newton, Baxter P Rogers, Li Min Chen, et al. "Functional MRI and resting state connectivity in white matter-a mini-review." In: *Magnetic resonance imaging* 63 (2019), pp. 1–11.
- [21] Erin L Mazerolle, Lisa Ohlhauser, Chantel D Mayo, Abu Sheriff, and Jodie R Gawryluk. "Evidence of underreporting of white matter fMRI activation." In: *Journal of Magnetic Resonance Imaging* 51.5 (2020), pp. 1596–1597.
- [22] Zhaohua Ding, Allen T Newton, Ran Xu, Adam W Anderson, Victoria L Morgan, and John C Gore. "Spatio-temporal correlation tensors reveal functional structure in human brain." In: *PloS one* 8.12 (2013), e82107.
- [23] Zhaohua Ding, Ran Xu, Stephen K Bailey, Tung-Lin Wu, Victoria L Morgan, Laurie E Cutting, Adam W Anderson, and John C Gore. "Visualizing functional pathways in the human brain using correlation tensors and magnetic resonance imaging." In: *Magnetic resonance imaging* 34.1 (2016), pp. 8–17.
- [24] Zhaohua Ding, Yali Huang, Stephen K Bailey, Yurui Gao, Laurie E Cutting, Baxter P Rogers, Allen T Newton, and John C Gore. "Detection of synchronous brain activity in white matter tracts at rest and under functional loading." In: *Proceedings of the National Academy of Sciences* 115.3 (2018), pp. 595–600.
- [25] Peter Mansfield. "Multi-planar image formation using NMR spin echoes." In: *Journal of Physics C: Solid State Physics* 10.3 (1977), p. L55.
- [26] Michael K Stehling, Robert Turner, and Peter Mansfield. "Echo-planar imaging: magnetic resonance imaging in a fraction of a second." In: *Science* 254.5028 (1991), pp. 43–50.
- [27] Chloe Hutton, Andreas Bork, Oliver Josephs, Ralf Deichmann, John Ashburner, and Robert Turner. "Image distortion correction in fMRI: a quantitative evaluation." In: *Neuroimage* 16.1 (2002), pp. 217–240.
- [28] Mark Jenkinson. "Fast, automated, N-dimensional phase-unwrapping algorithm." In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 49.1 (2003), pp. 193–197.
- [29] Jesper LR Andersson, Stefan Skare, and John Ashburner. "How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging." In: *Neuroimage* 20.2 (2003), pp. 870–888.
- [30] RNA Henson, C Buechel, O Josephs, and KJ Friston. "The slice-timing problem in event-related fMRI." In: *NeuroImage* 9 (1999), p. 125.
- [31] Ronald Sladky, Karl J Friston, Jasmin Tröstl, Ross Cunnington, Ewald Moser, and Christian Windischberger. "Slice-timing effects and their correction in functional MRI." In: *Neuroimage* 58.2 (2011), pp. 588–594.

- [32] Karl J Friston, Steven Williams, Robert Howard, Richard SJ Frackowiak, and Robert Turner. "Movement-related effects in fMRI time-series." In: *Magnetic resonance in medicine* 35.3 (1996), pp. 346–355.
- [33] Jean Talairach. *Atlas d'anatomie stereotaxique du telencephale: etudes anatomo-radiologiques*. Masson, 1967.
- [34] Jean Talairach. "Co-planar stereotaxic atlas of the human brain." In: *3-D proportional system: An approach to cerebral imaging* (1988).
- [35] Alan C Evans, D Louis Collins, SR Mills, Edward D Brown, Ryan L Kelly, and Terry M Peters. "3D statistical neuroanatomical models from 305 MRI volumes." In: *1993 IEEE conference record nuclear science symposium and medical imaging conference*. IEEE. 1993, pp. 1813–1817.
- [36] William D Penny, Karl J Friston, John T Ashburner, Stefan J Kiebel, and Thomas E Nichols. *Statistical parametric mapping: the analysis of functional brain images*. Elsevier, 2011.
- [37] Mark Jenkinson, Christian F Beckmann, Timothy EJ Behrens, Mark W Woolrich, and Stephen M Smith. "Fsl." In: *Neuroimage* 62.2 (2012), pp. 782–790.
- [38] Robert W Cox. "AFNI: software for analysis and visualization of functional magnetic resonance neuroimages." In: *Computers and Biomedical research* 29.3 (1996), pp. 162–173.
- [39] Krzysztof Gorgolewski, Christopher D Burns, Cindee Madison, Dav Clark, Yaroslav O Halchenko, Michael L Waskom, and Satrajit S Ghosh. "Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python." In: *Frontiers in neuroinformatics* (2011), p. 13.
- [40] Oscar Esteban, Christopher J Markiewicz, Ross W Blair, Craig A Moodie, A Ilkay Isik, Asier Erramuzpe, James D Kent, Mathias Goncalves, Elizabeth DuPre, Madeleine Snyder, et al. "fMRIPrep: a robust preprocessing pipeline for functional MRI." In: *Nature methods* 16.1 (2019), pp. 111–116.
- [41] Yashar Behzadi, Khaled Restom, Joy Liau, and Thomas T Liu. "A component based noise correction method (CompCor) for BOLD and perfusion based fMRI." In: *Neuroimage* 37.1 (2007), pp. 90–101.
- [42] Theodore D Satterthwaite, Mark A Elliott, Raphael T Gerraty, Kosha Ruparel, James Loughhead, Monica E Calkins, Simon B Eickhoff, Hakon Hakonarson, Ruben C Gur, Raquel E Gur, et al. "An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data." In: *Neuroimage* 64 (2013), pp. 240–256.
- [43] Karl J Friston, Keith J Worsley, Richard SJ Frackowiak, John C Mazziotta, and Alan C Evans. "Assessing the significance of focal activations using their spatial extent." In: *Human brain mapping* 1.3 (1994), pp. 210–220.
- [44] Yoav Benjamini and Yosef Hochberg. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." In: *Journal of the Royal statistical society: series B (Methodological)* 57.1 (1995), pp. 289–300.
- [45] Christopher R Genovese, Nicole A Lazar, and Thomas Nichols. "Thresholding of statistical maps in functional neuroimaging using the false discovery rate." In: *Neuroimage* 15.4 (2002), pp. 870–878.
- [46] Keith J Worsley, Jonathan E Taylor, Francesco Tomaiuolo, and Jason Lerch. "Unified univariate and multivariate random field theory." In: *Neuroimage* 23 (2004), S189–S195.
- [47] Will Penny and Guillaume Flandin. "Bayesian analysis of fMRI data with spatial priors." In: *Proceedings of the Joint Statistical Meeting (JSM)*. American Statistical Association. Citeseer. 2005.

- [48] Will Penny, Guillaume Flandin, and Nelson Trujillo-Barreto. “Bayesian comparison of spatially regularised general linear models.” In: *Human brain mapping* 28.4 (2007), pp. 275–293.
- [49] Lee M Harrison, W Penny, Jean Daunizeau, and Karl J Friston. “Diffusion-based spatial priors for functional magnetic resonance images.” In: *Neuroimage* 41.2 (2008), pp. 408–423.
- [50] Lee M Harrison, Will Penny, Guillaume Flandin, Christian C Ruff, Nikolaus Weiskopf, and Karl J Friston. “Graph-partitioned spatial priors for functional magnetic resonance images.” In: *NeuroImage* 43.4 (2008), pp. 694–707.
- [51] Per Sidén, Anders Eklund, David Bolin, and Mattias Villani. “Fast Bayesian whole-brain fMRI analysis with spatial 3D priors.” In: *NeuroImage* 146 (2017), pp. 211–225.
- [52] Per Sidén, Finn Lindgren, David Bolin, Anders Eklund, and Mattias Villani. “Spatial 3D Matérn priors for fast whole-brain fMRI analysis.” In: *Bayesian Analysis* 16.4 (2021), pp. 1251–1278.
- [53] Denis Le Bihan, Eric Breton, Denis Lallemand, Philippe Grenier, Emmanuel Cabanis, and Maurice Laval-Jeantet. “MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders.” In: *Radiology* 161.2 (1986), pp. 401–407.
- [54] Peter J Basser, James Mattiello, and Denis LeBihan. “Estimation of the effective self-diffusion tensor from the NMR spin echo.” In: *Journal of Magnetic Resonance, Series B* 103.3 (1994), pp. 247–254.
- [55] Peter J Basser, James Mattiello, and Denis LeBihan. “MR diffusion tensor spectroscopy and imaging.” In: *Biophysical journal* 66.1 (1994), pp. 259–267.
- [56] P J Basser and C Pierpaoli. “Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI.” In: *Journal of Magnetic resonance. Series B* 111.3 (1996), pp. 209–219.
- [57] Mark F Lythgoe, Albert L Busza, Fernando Calamante, Christopher H Sotak, Martin D King, Anna C Bingham, Stephen R Williams, and David G Gadian. “Effects of diffusion anisotropy on lesion delineation in a rat model of cerebral ischemia.” In: *Magnetic resonance in medicine* 38.4 (1997), pp. 662–668.
- [58] BA Inglis, EL Bossart, DL Buckley, ED Wirth III, and TH Mareci. “Visualization of neural tissue water compartments using biexponential diffusion tensor MRI.” In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 45.4 (2001), pp. 580–587.
- [59] Iman Aganj, Christophe Lenglet, Guillermo Sapiro, Essa Yacoub, Kamil Ugurbil, and Noam Harel. “Reconstruction of the orientation distribution function in single- and multiple-shell q-ball imaging within constant solid angle.” In: *Magnetic resonance in medicine* 64.2 (2010), pp. 554–566.
- [60] J-Donald Tournier, Fernando Calamante, and Alan Connolly. “Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution.” In: *Neuroimage* 35.4 (2007), pp. 1459–1472.
- [61] David Solomon Tuch et al. “Diffusion MRI of complex tissue structure.” PhD thesis. Massachusetts Institute of Technology, 2002.
- [62] Van J Wedeen, Patric Hagmann, Wen-Yih Isaac Tseng, Timothy G Reese, and Robert M Weisskoff. “Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging.” In: *Magnetic resonance in medicine* 54.6 (2005), pp. 1377–1386.

- [63] David S Tuch, Timothy G Reese, Mette R Wiegell, and Van J Wedeen. "Diffusion MRI of complex neural architecture." In: *Neuron* 40.5 (2003), pp. 885–895.
- [64] David S Tuch. "Q-ball imaging." In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 52.6 (2004), pp. 1358–1372.
- [65] Fang-Cheng Yeh, Van Jay Wedeen, and Wen-Yih Isaac Tseng. "Generalized q-sampling imaging." In: *IEEE transactions on medical imaging* 29.9 (2010), pp. 1626–1635.
- [66] A Anderson and Z Ding. "Sub-voxel measurement of fiber orientation using high angular resolution diffusion tensor imaging." In: *Book of abstracts: Tenth Annual Meeting of the International Society for Magnetic Resonance in Medicine. Berkeley, CA: ISMRM*. Vol. 10. 2002, p. 440.
- [67] J-Donald Tournier, Fernando Calamante, David G Gadian, and Alan Connelly. "Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution." In: *Neuroimage* 23.3 (2004), pp. 1176–1185.
- [68] Maxime Descoteaux, Elaine Angelino, Shaun Fitzgibbons, and Rachid Deriche. "Apparent diffusion coefficients from high angular resolution diffusion imaging: Estimation and applications." In: *Magnetic resonance in medicine: an official journal of the international society for magnetic resonance in medicine* 56.2 (2006), pp. 395–410.
- [69] Thomas E Conturo, Nicolas F Lori, Thomas S Cull, Erbil Akbudak, Abraham Z Snyder, Joshua S Shimony, Robert C McKinstry, Harold Burton, and Marcus E Raichle. "Tracking neuronal fiber pathways in the living human brain." In: *Proceedings of the National Academy of Sciences* 96.18 (1999), pp. 10422–10427.
- [70] Susumu Mori, Barbara J Crain, Vadappuram P Chacko, and Peter CM Van Zijl. "Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging." In: *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 45.2 (1999), pp. 265–269.
- [71] Peter J Basser, Sinisa Pajevic, Carlo Pierpaoli, Jeffrey Duda, and Akram Aldroubi. "In vivo fiber tractography using DT-MRI data." In: *Magnetic resonance in medicine* 44.4 (2000), pp. 625–632.
- [72] Klaus H Maier-Hein, Peter F Neher, Jean-Christophe Houde, Marc-Alexandre Côté, Eleftherios Garyfallidis, Jidan Zhong, Maxime Chamberland, Fang-Cheng Yeh, Ying-Chia Lin, Qing Ji, et al. "The challenge of mapping the human connectome based on diffusion tractography." In: *Nature communications* 8.1 (2017), p. 1349.
- [73] Kei Yamada, Koji Sakai, Kentaro Akazawa, Sachiko Yuen, and Tsunehiko Nishimura. "MR tractography: a review of its clinical applications." In: *Magnetic resonance in medical sciences* 8.4 (2009), pp. 165–174.
- [74] Ed Bullmore and Olaf Sporns. "Complex brain networks: graph theoretical analysis of structural and functional systems." In: *Nature reviews neuroscience* 10.3 (2009), pp. 186–198.
- [75] David Abramian and Martin Larsson. "Diffusion-adapted spatial filtering of fMRI data for improved activation mapping in white matter." In: ().
- [76] David Abramian, Martin Larsson, Anders Eklund, and Hamid Behjat. "Improved functional MRI activation mapping in white matter through diffusion-adapted spatial filtering." In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2020, pp. 539–543.

- [77] David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. “The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains.” In: *IEEE signal processing magazine* 30.3 (2013), pp. 83–98.
- [78] Antonio Ortega, Pascal Frossard, Jelena Kovačević, José MF Moura, and Pierre Vandergheynst. “Graph signal processing: Overview, challenges, and applications.” In: *Proceedings of the IEEE* 106.5 (2018), pp. 808–828. DOI: 10.1109/JPROC.2018.2820126.
- [79] Fan RK Chung. *Spectral graph theory*. Vol. 92. American Mathematical Soc., 1997.
- [80] David I Shuman, Benjamin Ricaud, and Pierre Vandergheynst. “Vertex-frequency analysis on graphs.” In: *Applied and Computational Harmonic Analysis* 40.2 (2016), pp. 260–291.
- [81] David K Hammond, Pierre Vandergheynst, and Rémi Gribonval. “Wavelets on graphs via spectral graph theory.” In: *Applied and Computational Harmonic Analysis* 30.2 (2011), pp. 129–150.
- [82] Hayit Greenspan, Bram Van Ginneken, and Ronald M Summers. “Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique.” In: *IEEE transactions on medical imaging* 35.5 (2016), pp. 1153–1159.
- [83] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
- [84] Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola. “Dive into Deep Learning.” In: *arXiv preprint arXiv:2106.11342* (2021).
- [85] Herbert Robbins and Sutton Monro. “A stochastic approximation method.” In: *The annals of mathematical statistics* (1951), pp. 400–407.
- [86] Geoffrey Hinton. “Lecture 6a: Overview of mini-batch gradient descent.” In: *Courses for lectures: Neural Networks for Machine Learning* (2012).
- [87] Diederik P Kingma and Jimmy Ba. “Adam: A method for stochastic optimization.” In: *arXiv preprint arXiv:1412.6980* (2014).
- [88] Frank Rosenblatt. “The perceptron: a probabilistic model for information storage and organization in the brain.” In: *Psychological review* 65.6 (1958), p. 386.
- [89] George Cybenko. “Approximation by superpositions of a sigmoidal function.” In: *Mathematics of control, signals and systems* 2.4 (1989), pp. 303–314.
- [90] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. “Multilayer feedforward networks are universal approximators.” In: *Neural networks* 2.5 (1989), pp. 359–366.
- [91] Zhou Lu, Hongming Pu, Feicheng Wang, Zhiqiang Hu, and Liwei Wang. “The expressive power of neural networks: A view from the width.” In: *Advances in neural information processing systems* 30 (2017).
- [92] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. “Dropout: a simple way to prevent neural networks from overfitting.” In: *The journal of machine learning research* 15.1 (2014), pp. 1929–1958.
- [93] Sergey Ioffe and Christian Szegedy. “Batch normalization: Accelerating deep network training by reducing internal covariate shift.” In: *International conference on machine learning*. pmlr. 2015, pp. 448–456.
- [94] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. “Layer normalization.” In: *arXiv preprint arXiv:1607.06450* (2016).
- [95] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. “Instance normalization: The missing ingredient for fast stylization.” In: *arXiv preprint arXiv:1607.08022* (2016).

-
- [96] Xavier Glorot and Yoshua Bengio. “Understanding the difficulty of training deep feedforward neural networks.” In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings. 2010, pp. 249–256.
 - [97] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification.” In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 1026–1034.
 - [98] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation.” In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18*. Springer. 2015, pp. 234–241.
 - [99] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. “Generative Adversarial Nets.” In: *Advances in Neural Information Processing Systems*. Vol. 27. Curran Associates, Inc., 2014.
 - [100] Mehdi Mirza and Simon Osindero. “Conditional generative adversarial nets.” In: *arXiv preprint arXiv:1411.1784* (2014).
 - [101] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. “Image-to-image translation with conditional adversarial networks.” In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1125–1134.
 - [102] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. “Unpaired image-to-image translation using cycle-consistent adversarial networks.” In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2223–2232.
 - [103] Xin Yi, Ekta Walia, and Paul Babyn. “Generative adversarial network in medical imaging: A review.” In: *Medical image analysis* 58 (2019), p. 101552.

Papers

The papers associated with this thesis have been removed for copyright reasons. For more details about these see:

<https://doi.org/10.3384/9789180751360>

FACULTY OF SCIENCE AND ENGINEERING

Linköping Studies in Science and Technology, Dissertation No. 2307, 2023
Department of Biomedical Engineering

Linköping University
SE-581 83 Linköping, Sweden

www.liu.se