



Mathematical models and numerical methods for the analysis of medical images and complex data.

Laurent Risser

► To cite this version:

Laurent Risser. Mathematical models and numerical methods for the analysis of medical images and complex data.. Machine Learning [stat.ML]. Université Toulouse 3 Paul Sabatier, 2018. tel-01945575

HAL Id: tel-01945575

<https://theses.hal.science/tel-01945575>

Submitted on 5 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Paul Sabatier - Toulouse III

Manuscrit

présenté pour l'obtention de

L'Habilitation à Diriger des Recherches

par

Laurent Risser

de l'Institut de Mathématiques de Toulouse

**Mathematical models and numerical
methods for the analysis of medical
images and complex data.**

Soutenance le 30 novembre 2018 devant le jury composé de :

Rapporteurs :	Jean-François Aujol	Professeur, Université de Bordeaux
	Gabriel Peyré	Directeur de Recherche, ENS Paris
	Laurent Younes	Professeur, Johns Hopkins University, Etats-Unis
Examineurs :	Agnès Desolneux	Directrice de Recherche, ENS Paris-Saclay
	Jean-Michel Loubes	Professeur, Université Toulouse 3
	Xavier Pennec	Directeur de Recherche, INRIA Sophia-Antipolis
Parrain :	Fabrice Gamboa	Professeur, Université Toulouse 3

Abstract

This manuscript synthesizes my scientific activity between September 2007 (end of my PhD thesis) and July 2018. This scientific activity was first carried out in the context of three postdoctoral positions at Neurospin - *CEA Saclay*, the departments of Applied Mathematics and of Biomedical Image Analysis at *Imperial College London*, and the department of Biomedical Image Analysis at *University of Oxford*. It was then pursued at the *Mathematics Institute of Toulouse*, where I obtained a CNRS Research Engineer position in 2012. This has lead me to work on different projects in collaboration with various researchers and PhD students. A selection of relevant scientific contributions related to *Mathematical models and numerical methods in medical image and complex data analysis* are synthesized in this manuscript and other contributions are only mentioned.

A general synthesis of my scientific activity is first developed Chapter 1. It first gives a global overview of my carrier and my PhD work. It then describes the different scientific projects in which I have been involved and who were my main collaborators. As almost all my contributions were carried out in the context of collaborative projects, so it also develops what was my personal contribution to selected communications. It finally gives my bibliographic record. A more detailed presentation of the selected research projects is given in the following chapters. I distinguish my contributions in *Mathematical models in medical image analysis*, *Numerical methods for stochastic modeling* and *Statistical learning on complex data*. Chapters 2, 3 and 4 then deal with each these fields.

Acknowledgements

Je voudrais d'abord remercier Fabrice Gamboa et Jean Michel Loubes de m'avoir particulièrement aidé dans ce joli projet qu'est la préparation d'une HDR. Je souhaite de même remercier chaleureusement mes rapporteurs Jean-François Aujol, Gabriel Peyré et Laurent Younes pour leurs rapports ainsi qu'Agnès Desolneux et Xavier Pennec qui ont accepté de participer à mon jury de soutenance. Un gros merci aussi à Laure Coutin pour son support dans ce projet ainsi que Sébastien Gadat et Gersende Fort pour m'avoir impliqué dans leurs projets et ainsi relancer mon activité scientifique sur des thématiques à la croisée entre programmation numérique et aléatoire. Un merci aussi à Jérôme Fehrenbach et François Bachoc pour leurs conseils dans la rédaction du manuscrit et à Michèle Antonin pour son support administratif.

La liste des personnes qui ont compté dans ma carrière ces 11 dernières années est bien longue et je leur suis infiniment reconnaissant. Mes encadrants de postdoc ont notamment eu un impact significatif sur ma carrière : Philippe Ciuciu m'a fait découvrir le monde de l'imagerie médicale, m'a beaucoup poussé et ainsi largement fait progresser scientifiquement parlant. Darryl D. Holm m'a permis de découvrir une vision plus Anglo-Saxonne de la recherche que celle que j'avais connu jusqu'alors. Travailler avec Daniel Rueckert et Julia A. Schnabel a été un réel plaisir de par leur vision pragmatique et efficace de la maturation d'une intuition vers une communication scientifique. Merci beaucoup ! Thank you very much! Danke sehr!

Une rencontre déterminante pour moi a aussi été celle de François-Xavier Vialard avec qui j'ai été particulièrement complémentaire sur plusieurs projets liés au recalage d'images. Ceci nous a permis de développer des idées que je considère parmi les plus belles de ma carrière. Un gros merci !

Je n'oublie aussi pas tous ceux qui m'ont permis de me lancer dans la recherche, avant et pendant ma thèse. Je pense en particulier à Céline Badufle, Benjamin Vidal et Renaud Marty sans qui je ne me serais sans doute jamais orienté vers une thèse à la fin de mon Master. Patrice Dalle et Jean-Denis Duroux ont ensuite été les premiers à me guider dans le monde de la recherche. Ensuite, Franck Plouraboué, Caroline Fonta et Xavier Descombes ont été les piliers pour me former au métier de chercheur. Merci à tous !

Enfin, je voulais avoir une pensée pour tout l'amour et le support que m'apporte ma famille. Merci !

Contents

Manuscript organization	1
1 General synthesis	3
1.1 Preamble	3
1.1.1 Career overview	3
1.1.2 PhD thesis work	4
1.2 Projects and collaborations	6
1.2.1 Projects in medical image analysis	6
1.2.2 Projects in numerical methods for stochastic modeling . .	9
1.2.3 Projects in statistical learning on complex data	10
1.3 Personal contributions	10
1.4 Teaching activity and students supervision	19
1.5 Bibliographic record	21
2 Mathematical models in medical image analysis	29
2.1 Introduction	29
2.1.1 Medical image analysis	29
2.1.2 Medical image registration	29
2.1.3 LDDMM image registration	32
2.1.4 LogDemons image registration	36
2.2 Medical image registration models	37
2.2.1 Summary of contributions	37
2.2.2 Diffeomorphic image matching using geodesic shooting . .	38
2.2.3 Karcher mean estimations for 3D images	39
2.2.4 Left-invariant metrics for diffeomorphic image matching .	42
2.2.5 Image matching based on a reaction-diffusion model. . . .	44
2.3 Regularization metrics in medical image registration	46
2.3.1 Summary of contributions	46
2.3.2 Multi-scale metrics	48
2.3.3 Diffeomorphic image registration with sliding conditions .	52
2.3.4 Learning optimal regularization metrics	54
2.4 Similarity metrics in medical image registration	55
2.4.1 Summary of contributions	55
2.4.2 Local estimation of mutual information gradients	56
2.5 Image segmentation models	59
2.5.1 Summary of contributions	59
2.5.2 Regularization model for the Fast Marching segmentation	60
2.6 Outlook	62

3	Numerical methods for stochastic modeling	63
3.1	Summary of contributions	63
3.2	Numerical methods for the analysis of the brain activity	64
3.2.1	A general model for the analysis of fMRI time series	64
3.2.2	Estimation of 3D Ising and Potts field partition functions	67
3.2.3	Results and discussion	69
3.3	A stochastic framework for the online graph barycenter estimation	71
3.3.1	Motivation	71
3.3.2	Methodology	72
3.3.3	Results and discussion	75
3.4	Outlook	77
4	Statistical learning for complex data	79
4.1	Summary of contributions	79
4.2	Regularization models on 3D image domains	80
4.2.1	Motivation	80
4.2.2	Methodology	80
4.2.3	Results	82
4.3	Distribution regression with a RKHS approach	82
4.3.1	Motivation	82
4.3.2	Methodology	82
4.3.3	Results and discussion	83
4.4	Representative variable detection for complex data	84
4.4.1	Motivation	84
4.4.2	Methodology	85
4.4.3	Results and discussion	86
4.5	Outlook	87
	Bibliography	88
	Main publications	98
A	Mathematical models in medical image analysis	99
A.1	Simultaneous Multiscale Registration using Large Deformation Diffeomorphic Metric Mapping [IJ-6]	99
A.2	Diffeomorphic 3D Image Registration via Geodesic Shooting us- ing an Efficient Adjoint Calculation [IJ-7]	114
A.3	Diffeomorphic Atlas Estimation using Geodesic Shooting on Vol- umetric Images [IJ-8]	128
A.4	Mixture of Kernels and Iterated semidirect Product of Diffeomor- phisms Groups [IJ-10]	141
A.5	Piecewise-Diffeomorphic Image Registration: Application to the Motion Estimation between 3D CT Lung Images with Sliding Conditions [IJ-11]	167
A.6	Construction of Diffeomorphic Spatio-temporal Atlases using Kärcher means and LDDMM [IC-22]	180
A.7	Piecewise-diffeomorphic registration of 3D CT/MR pulmonary images with sliding conditions [IC-27]	189

A.8	Hybrid Feature-based Diffeomorphic Registration for Tumour Tracking in 2-D Liver Ultrasound Images [IJ-13]	194
A.9	Diffeomorphic image matching with left-invariant metrics [B-2]	205
A.10	Spatially-varying metric learning for diffeomorphic image registration. A variational framework [IC-32]	226
A.11	Diffeomorphic registration with self-adaptive spatial regularization for the segmentation of non-human primate brains [IC-33]	235
A.12	Filling Large Discontinuities in 3D Vascular Networks using Skeleton- and Intensity-based Information [IC-34]	240
A.13	A DCE-MRI Driven 3-D Reaction-Diffusion Model of Solid Tumour Growth [IJ-20]	249
A.14	Regularized Multi-Label Fast Marching and Application to Whole-Body Image Segmentation [IC-37]	261
B	Numerical methods for stochastic modeling	266
B.1	Unsupervised spatial mixture modelling for within-subject analysis of fMRI data [IJ-4]	267
B.2	Min-max extrapolation scheme for fast estimation of 3D Potts field partition functions [IJ-5]	284
B.3	How to calculate the barycenter of a weighted graph [IJ-19]	299
B.4	Online Barycenter Estimation of Large Weighted Graphs [SJ-6]	333
C	Statistical learning on complex data	356
C.1	Longitudinal deformation models, spatial regularizations and learning strategies to quantify Alzheimer’s disease progression [IJ-15]	356
C.2	A representative variable detection framework for complex data based on CORE-clustering [SC-1]	369
C.3	Distribution regression model with a Reproducing Kernel Hilbert Space approach [SJ-5]	390

Manuscript organization

This manuscript contains a synthesis of my scientific activity between September 2007 (end of my PhD thesis work) and July 2018. A selection of representative communications is also given in appendix.

A general synthesis of my scientific activity is first developed Chapter 1. Its first section is a preamble giving a global overview of my carrier and my PhD work. Section 1.2 then describes the different scientific projects in which I have been involved after my PhD work and who were my main collaborators. Note that almost all my research projects were carried out in collaboration with other researchers or PhD students. My personal contribution to selected journal papers and conference proceedings is then developed in Section 1.3. Then, a presentation of the students I have formally supervised and of the different courses and practicals I have taught is given in Section 1.4. My bibliographic record is finally given in Section 1.5, where I distinguish different kinds of contributions (Refereed international journal papers, Refereed international conference proceedings, ...). In this manuscript, all these communications are cited with a different style as the other citations. Table 1 represents their style.

A more detailed presentation of selected research projects is then given in Chapters 2 to 4. For clarity purposes, I only present the communications in which I consider that my scientific contribution was significant. Among them, I distinguished scientific contributions related to three themes: *Mathematical models in medical image analysis*, *Numerical methods for stochastic modeling* and *Statistical learning on complex data*. Explanations related to these themes will be developed in Chapters 2, 3 and 4, respectively. Each of these chapters starts with an introductory section and its following sections develop the contributions of specific projects.

Refereed international journal papers	[IJ-.]
Refereed national journal papers	[NJ-.]
Papers submitted to journals	[SJ-.]
Books and book chapters	[B-.]
Submitted book chapters	[SB-.]
Refereed international conference proceedings	[IC-.]
Refereed national conference proceedings	[NC-.]
Submitted international conference proceedings	[SC-.]

Table 1: Citation types of the references in the bibliographic record of Section 1.5. The dots (.) correspond to the communication numbers in chronological order.

Chapter 1

General synthesis

1.1 Preamble

1.1.1 Career overview

After I graduated with my bachelor degree, I have for a long time hesitated between studying applied mathematics or computer science. I finally found myself between the two with a Master 1 degree in applied mathematics and a Master 2 degree in computer science. I discovered the fields of Scientific Computing (PDEs, Optimization) and Statistics during my Master 1 degree and realized this year that I wanted to make my career in numerical mathematics. Then, I decided to be even more specialized in programming for data analysis. This has lead me to go to a Master 2 degree in computer science applied to signal and image analysis.

This profile in-between applied mathematics and computer science gave me the opportunity to get a funding for a PhD thesis in engineering science with Pr F. Plouraboué at the Fluid Mechanics Institute of Toulouse. My PhD work consisted in analyzing large 3D images of the cerebral micro-vasculature, in order to quantify the blood flow properties at the brain scale. In terms of programming, I coded image analysis algorithms where the memory management and the algorithmic complexity were critical constraints. Different statistical issues related to small datasets in high dimension also became concrete for me. This finally gave me the taste for applications in life sciences. In parallel to my PhD thesis work, I was also junior lecturer (*moniteur*) in fluid mechanics at the Paul Sabatier University. I gave a Master 1 level course in computational models for fluid dynamics (non-linear 2D/3D PDEs) during three years, which strengthened my knowledge in numerical simulation. I also gave different courses in mechanics and realized how important are pertinent approximations when one mathematically solves a real-life problem. Selecting the most influential properties of a modeled phenomenon is indeed what often makes such problems solvable with a negligible approximation error.

My goal after my PhD thesis work was to focus on *numerical mathematics applied to medical image analysis*, in order to find either an academic or an industrial position in this field. I first obtained a postdoctoral position at Neu-

rospin/CEA Saclay with Dr P. Ciuciu where I worked for almost two years on Bayesian optimization models to estimate the brain activity in functional Magnetic Resonance Imaging. Then obtained two postdoctoral positions at Imperial College London (with Pr D. Rueckert and Pr D.D. Holm) and at University of Oxford (with Dr J.A. Schnabel), where I worked for three years on the development of image registration strategies in medical imaging. In addition to work experiences in data analysis, these almost five years of postdoctoral positions were probably those where I learned my most important lessons about how to communicate in science and how to build a research project. I then wished to continue my career at the interaction between numerical mathematics and real-life applications.

In 2012, I obtained a CNRS Research Engineer position at the Toulouse Institute of Mathematics (IMT). My two main missions there were (and are still) first to give a high level technical support to the scientific activity at IMT, and to additionally have my own scientific activity. I have then continued different collaborations in medical imaging with former colleagues. Medical imaging was however a very minor research theme at IMT at the time, so I involved myself in other research projects of the Probability and Statistics team (ESP) of IMT. In particular, I have initiated collaborations related to statistical learning and the analysis of complex data. These themes are indeed close to medical imaging from a methodological point of view and were also interesting to me. After working on several projects with technical contributions in statistical learning (C++, Python, OpenCL and R programming; results interpretation; trainees co-supervision; ...), I was gradually more and more involved in their scientific aspects, so I also developed a research activity in this field.

This path has lead me to have scientific contributions in relatively varied fields, although being all related to *Mathematical models and numerical methods in medical image and complex data analysis*. In this manuscript, I will synthesize the scientific contributions I have made after my PhD work (September 2007) and before July 2018. I will focus on my most significant scientific contributions, in my opinion, and mention the other ones.

1.1.2 PhD thesis work

My PhD thesis work [B-1] was carried out at the Fluid Institute of Toulouse (IMFT, UMR 5502) between 2003 and 2007, under the supervision of Franck Plouraboué. My goal was to quantify the anatomy of intracortical vascular networks in order to evaluate their blood flow properties using statistical mechanics techniques. Image resolution was 1.4^3 microns per voxel for acquired volumes of about 3 cubic millimeters, as shown Fig. 1.1. The whole vasculature was then captured in the acquired images which made it possible to estimate the blood flow properties at the millimeter scale. This scale is interesting as it is similar to the cortical areas size. The results of this work were then of importance to compare the blood flow properties of different brain regions, or of a single brain region in different groups of subjects. This also made it possible to quantify the local impact of a brain stroke in terms of nutriment supply for the brain tissues, and to quantify how a tumor strongly increases its nutriment supply by transforming the vascular network in its neighborhood. Remark that these

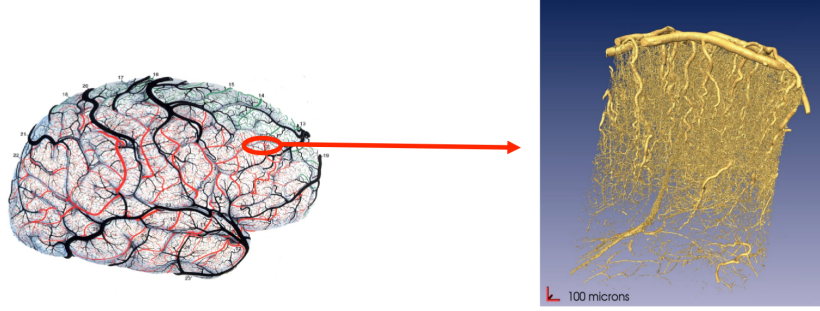


Figure 1.1: Data studied during my PhD work: Samples of the intracortical vasculature of about 3 mm^3 were acquired using synchrotron tomography at a resolution of $1.4 \mu\text{m}$ per voxel. All vessels are then distinguished in the volumes, which makes it possible to evaluate blood flow properties at the cortical region scale using statistical mechanics techniques. Illustration out of [B-1].

results were also of interest to understand the observed signal in BOLD fMRI as developed in Section 3.2. This work was rewarded in 2008 by the French national prize *La recherche* with the *human health* mention. Although I had a minor research activity in vascular image analysis after my PhD thesis [IJ-9, IC-34, SJ-3], this experience had an important impact on the research themes I developed later:

1. I used various tools of statistical mechanics and image analysis in [IJ-3, IJ-1, IC-9, IC-7, IC-6, IC-5, IC-3, IC-2, IC-1, NC-1]. This made me familiar with the statistical analysis of 3D medical images.
2. The project which had the highest impact on my future career was the one published in [IJ-2, IC-4]. I developed a novel image processing strategy to fill the discontinuities that can be obtained when segmenting vascular networks in 3D images. These developments were made in a scientific community different to the one of F. Plouraboué. He therefore put me in contact with X. Descombes (INRIA Sophia-Antipolis) who is specialist of bio-medical imaging and who advised me in this context. This allowed me to move from the fluid mechanics community to the medical imaging community after my PhD.
3. In the very end of my PhD thesis, I also worked on graph representations of the vascular networks in order to detect clusters of influent vessels using graph clustering techniques (see Fig. 1.2). This work is the last methodological part of my PhD manuscript [B-1] and was only published four years later in a journal [IJ-9], after further bio-mechanical developments. Although it had little impact in my PhD work, it was the basis for the development of a new research activity in graphs analysis eight years later, at the Mathematics Institute of Toulouse [IJ-19, SJ-6, SC-1].
4. I finally worked on large data volumes: each 3D image was about 8 times larger than what I could allocate on my workstation. This developed my taste for low-level algorithms suitable for real-data analysis, which I kept later in all my methodological developments.

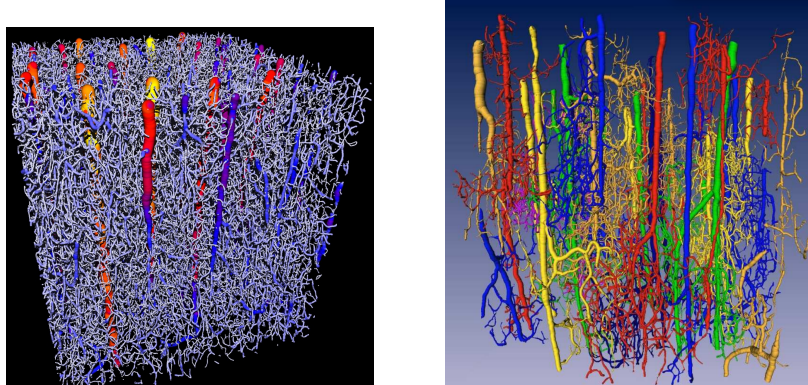


Figure 1.2: Automatic detection of the most influent structures in intracortical vascular networks, based on graph clustering. **(Left)** A segmented and skeletonized sample of vascular network. The vascular network is represented by a graph augmented with spatial coordinates and a local diameter at each node. The colors represent here local vessels diameters. **(Right)** Subgraphs extracted out of the graph on the left. Each color represent an influent cluster of vessels with respect to the blood flow properties in the sample. Illustration out of [B-1].

The rest of this manuscript focuses on the scientific contributions I developed during my postdoctoral works and at the Mathematics Institute of Toulouse between 2007 and 2018. I refer to [B-1] as well as the citations of this section for further details about my PhD work.

1.2 Projects and collaborations

This section presents the main research projects in which I have been involved after my PhD work. In particular, I explain their working environment and who were my main collaborators. The scientific contributions and overviews of the main publications out of these projects is developed in Chapters 2 to 4.

1.2.1 Projects in medical image analysis

My projects related to medical image analysis are developed in Chapter 2. Here is below a synthesis of their motivations and contributions.

Medical image registration in the LDDMM framework An important project for me has been the development of methodologies in the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework (see Section 2.1.2). I started working on this framework during my postdoctoral work at Imperial College London where my goal was to develop medical image registration strategies in interaction with D.D. Holm at the Department of Applied Mathematics and D. Rueckert in the Biomedical Image Analysis department. F.X. Vialard (former PhD student of A. Trounev at ENS Cachan) was hired as a postdoctoral researcher in the same project as me and we started a long term collaboration.

Our starting point was to deeply understand [BMTY05a] and how to make it work on practical cases given by D. Rueckert.

The first work in this collaboration with F.X. Vialard, was [IJ-6,IJ-10,IC-20,IC-18,IC-24] (Subsection 2.3.2). We defined multi-scale metrics in LDDMM, motivated by real applications in 3D medical imaging. From the application side, I collaborated with M. Murgasova (former PhD student with D. Rueckert) who motivated the problems related to the registration of pre-term babies MR images. I also collaborated with M. Bruveris (former PhD student with D.D. Holm) who extended [IJ-6] in [IJ-10] with a more mathematically rigorous approach. We then worked on an extension of [BMTY05a] where sliding constraints could be modeled in the context of my postdoctoral work at University of Oxford with J.A. Schnabel (former researcher in the Biomedical image analysis department of Univ. Oxford) and published this work in [IJ-11,IC-24] (Subsection 2.3.3). This work motivated the development of alternative formulation to LDDMM, where spatially-varying metrics would make sense. This was done in collaboration with T. Schmah (Univ. Toronto) and published in [B-2,IC-31], where left-invariant metrics mathematically justified the use of LDDMM with spatially-varying metrics (Subsection 2.2.4). After having justified the use of spatially-varying registration in LDDMM, we finally built a strategy that learns optimal spatially-varying regularization metrics with respect to a learning set of reference images [IC-32] (Subsection 2.3.4). Remark that these methods were recently summarized in [SB-1].

An alternative project on which I worked with F.X. Vialard, based on the same starting point, was the development of a formulation of LDDMM where geodesic shooting is used to register the images [IJ-7] (Subsection 2.2.2). This strategy was then used to define the Karcher means of shapes (average shapes) in 3D images [IJ-8,IC-22,IC-21] (Subsection 2.2.3). An original statistical learning pipeline based on the initial momenta computed using [IJ-7] and the averaged shapes of [IJ-8] was also presented in [IC-28,IJ-15] in collaboration with J.B. Fiot (former PhD student with L.D. Cohen at Univ. Paris Dauphine).

Medical image registration in other frameworks Personal contributions in medical image registration, outside of the LDDMM framework, were made mostly by collaborating with different PhD students. Tight collaborations with PhD students are much more frequent for postdoctoral researchers in the UK than in France. One of their role is indeed to give a scientific support to the PhD students of their advisor. I worked for three years as a postdoctoral researcher in the UK, which explains these collaborations. Note that I also continued having such collaborations after having been hired at IMT, in particular with J.A. Schnabel’s team. In [IJ-17,IC-30], I worked with B. Papiez (former PhD student of J.A. Schnabel) to extend the sliding motion strategy of [IJ-11]. A simpler image registration formulation was used but the location of the sliding constraints was automatically detected. I also worked with J. Ferhenbach in this project to justify the developments. In parallel, I also had also a strong scientific implication in the PhD work of A. Cifor (former PhD student of J.A. Schnabel) where we worked on the robust tracking of liver tumors in 2D Ultrasound image series [IJ-13,IC-29,IC-26]. In the same vein, I also had a strong implication in the PhD thesis work of T. Roque (former PhD student of J.A. Schnabel), in particular in [IJ-20] where image deformations are driven by a

physiologically motivated reaction-diffusion model (Subsection 2.2.5). By regularly talking with M.P. Heinrich (former PhD student of J.A. Schnabel) about the definition of new realistic frameworks for multi-modal image registration, I also defined the mutual information gradient estimation technique of [IC-27,IC-23]. More secondary collaborations with J.A. Schnabel’s students were first in [IJ-12] with H. Baluwala (former PhD student of J.A. Schnabel) where we mostly shared image pre-processing tasks with [IJ-11]. I also worked with M. Bhushan [IC-25] in order to make diffeomorphic his motion correction strategy.

In addition to these collaborations with J.A. Schnabel’s students, I also collaborated with the team of C. Fonta (DR CNRS) and M. Mescam (lecturer, Univ. Toulouse) at the Brain and Cognition (CerCo) laboratory of Toulouse. In this context, I developed the image registration with automatic selection of the deformations scale in [IC-33] for marmoset monkeys brains.

Image segmentation Several image segmentation projects on which I have contributed are related to image registration: A reference segmentation can be thoroughly performed once for all on a template (average) image containing the shape of interest. The segmented image is then registered to the template and the reference segmentation is transported from the template domain to the segmented image using the mapping. Using such techniques I had a strong implication in the PhD thesis work of D.P. Zhang (former PhD student of D. Rueckert) where we developed different strategies for the segmentation of the coronary artery in 3D+time cardiac CT sequences [IC-19,IC-17,IC-16]. More recently, I also developed a plugin for the 3Dslicer software in order to perform the template-based segmentation of marmoset brain images [SJ-2] with C. Fonta and M. Mescam from CerCo.

As an extension of my PhD thesis gap filling strategy of [IJ-2], I also worked with R. Bates (former PhD student of J.A. Schnabel) on a post-treatment strategy for tubular structures segmentation. We indeed extended [IJ-2] to micro-CT images of tumorous vascular networks [IC-34]. By collaborating with J.M. Mirebeau (CR CNRS, Univ. Paris Dauphine) and J. Fehrenbach in the context of an ANR project, we also presented in [IJ-16] the ITK implementation of an efficient anisotropic non-linear diffusion technique for 2D or 3D images.

I also work with F. Gamboa (Pr Univ. Toulouse, IMT), A. Gossé (CR CEA Saclay) and A. Quaini (CR CEA Saclay) since 2015 on the segmentation and the feature extraction of 2D image sequences representing rotating and levitating balls which are extremely warmed-up. The goal here is to understand the mechanical properties of the balls under extreme conditions. The main technical issues deal with artifacts, occlusions and low boundary contrasts observed during the experimental protocol. A communication explaining the results of this work should be submitted over the following months.

Finally, I currently collaborate with F. Malgouyres (Pr Univ. Toulouse, IMT) and colleagues from the *Toulouse Cancer University Institute*, in particular S. Ken (Research Engineer INSERM, IUCT), on a 4 years INSERM project where we work on segmentation strategies for nodules and other structures in multi-modal whole body images. For now, we have published a regularization strategy for the Fast Marching algorithm [IC-37] and I co-supervise with F. Malgouyres V.K. Ghorpade (postdoctorate IMT) who works on these developments since January 2018.

1.2.2 Projects in numerical methods for stochastic modeling

I mainly developed original numerical methods for stochastic modeling in the context of two projects and had besides minor contributions. Here is below a synthesis of their motivations and contributions. Further explanations are developed Chapter 3.

My first work experience with numerical methods for stochastic modeling was during my postdoctoral work at CEA Saclay with P. Ciuciu (former CR CEA Saclay). This project was related to the analysis of the brain activity in functional Magnetic Resonance Imaging (fMRI) time series, (Section 3.2). An original Bayesian model was first developed in collaboration with T. Vincent (former PhD student of P. Ciuciu) to analyze fMRI time series [IJ-4,IJ-14,NJ-1,IC-11,IC-15,IC-8]. I have been particularly involved in the development of a strategy to efficiently compute the partition function of Potts field with respect to their inverse temperature β [IJ-5,IC-13,IC-14,IC-12,NC-3], in order to make unsupervised the spatial regularization of this model. Insights about the proposed methodology were also developed with F. Forbes (DR INRIA Grenobles) and J. Idier (Pr Central Nantes), who temporarily hired me for two months in the end of my postdoctoral work at CEA Saclay and before my postdoctoral work at Imperial College London. Note that I also supervised the Master 2 project of A.L. Fouque (ENS Cachan) during this postdoctoral work and we developed a statistical clustering strategy for fMRI time series [IC-10,NC-2].

More recently, I collaborated with S. Gadat (Pr Toulouse School of Economics, IMT) and I. Gavra (former PhD student of S. Gadat) on the definition of a barycenter estimation strategy for graphs in which a probability measure reflects observation occurrences on the graph nodes (Section 3.3). We first published [IJ-19] and extended this work to the online and high dimensional context in [IJ-6]. I. Gavra has recently obtained a lecturer position at University of Rennes and we plan continuing our collaboration.

I also worked with S. Ribes who was PhD student under the supervision of O. Caselles at a laboratory of Univ. Toulouse 3 (SIMAD). She had the potential and the material to write a paper in a medical image analysis journal but her supervisors had little experience in this field. After talking together about her project, we agreed that I would informally advise her in this part of her PhD thesis work. I mainly helped her to develop the image segmentation pipeline based on Bayesian model as in [IJ-4], and led the paper redaction at the IEEE Trans. Medical Imaging format [IJ-18].

I have finally started a collaboration last year with G. Fort (DR CNRS, IMT) about Maximum Likelihood inference algorithms in statistical models. My goal in this collaboration is to develop original numerical methodologies to make scalable Gibbs sampling strategies. For now, my contributions were mostly technical and I worked on an efficient implementation of the algorithms in [IC-36,NC-4]. Note that, I also had a similar technical contribution with S. Gadat and M. Costa (Lecturer at Univ. Toulouse, IMT) [SJ-1] which deals with atomic deconvolution *i.e.* with deconvolution in density estimation.

1.2.3 Projects in statistical learning on complex data

The third field in which I had a scientific activity is statistical learning applied to complex data. My research activity in this field is developed Chapter 4 and is synthesized hereafter. Note that I give different links between complex data analysis and medical image registration in Section 4.1. In this manuscript, I distinguish my research activity in these fields by considering medical image registration as a specific subfield of complex data analysis.

As mentioned Subsections 1.2.1 and 1.2.2, I first had a minor contribution in this field while supervising the M2 project of A.L. Fouque with P. Ciuciu during my postdoctoral work at CEA Saclay [IC-10,NC-2]. This contribution was about statistical learning techniques to clusterize hemodynamic parameters out of fMRI time series. Several years later, I worked with J.B. Fiot and F.X. Vialard on the exploration of different spatial regularization models for logistic regression on 3D image domains [IC-28,IJ-15] (Section 4.2). I also developed an image registration strategy with automatic selection of the deformations scale in [IC-33] based on LASSO regularization. In these contributions, statistical learning techniques were mostly applied to specific applicative cases.

I started developing new models in statistical learning by co-supervising two PhD theses with J.M. Loubes (Pr Univ. Toulouse, IMT). I first work with T. Bui since September 2016 on the development of statistical models for the classification of 3D coiled shapes out of the inner ear as well as distributions of the response of the ear to otoacoustic emission (OAE) [SJ-5] (Section 4.3). I also work with C. Champion since September 2017 on the extraction of representative variables in complex systems [SC-1,NC-5] (Section 4.4).

Other collaborations in this theme have also started with F. Gamboa (Pr Univ Toulouse, IMT), F. Bachoc (Lecturer Univ Toulouse, IMT), and S. Déjean (Research Engineer, Univ. Toulouse, IMT) and should lead to new developments in the future.

1.3 Personal contributions

Almost all my scientific contributions were developed in the context of collaborative works. In this section, I therefore make clear what was my personal contribution to selected journal papers and conference proceedings. These selected communications are presented in chronological order and correspond to the communications given in appendix.

Unsupervised spatial mixture modelling for within-subject analysis of fMRI data [IJ-4]

(Motivation) Within-subject analysis of the brain activity in BOLD fMRI consists in detecting patterns of energy consumption in 3D+time image series. Each of these patterns is located at an image point and smoothly evolves in time during several seconds after an onset. It represent a brain activation and is related to local variations of oxygen consumption in the brain due to a cognitive task. In 2010, all existing approaches either detected the activations using a predefined energy consumption pattern, or estimated this pattern at specific locations and times. There is however a physiological evidence that these two tasks should be performed simultaneously as the energy consumption pattern strongly depends

of the local vasculature which varies across brain regions [IJ-1,IJ-3]. **(Main paper contributions)** This paper develops an original Bayesian model in which the activations detection and the energy consumption patterns are simultaneously estimated. The model makes physiologically realistic hypotheses to constrain the minimized energy, making it possible to detect local activations that are lost using more generic regularization models. **(Personal contributions)** T. Vincent (who was a PhD student of P. Ciuciu at CEA Saclay) was the main contributor to this work. I developed the unsupervised spatial regularization model and strategy to estimate the 3D Potts field partition functions. The partition function strategy was generalized later in [IJ-5]. This paper is shown in Appendix B.1.

Min-max extrapolation scheme for fast estimation of 3D Potts field partition functions [IJ-5]

(Motivation) Potts models are typically used as Hidden Markov Fields with K labels/colors when segmenting an image using a Bayesian formalism. They indeed allow to spatially regularize the optimal segmentation with a strength which is controlled by an inverse temperature β . When the regularization level is unsupervised, it is however mandatory to compute the partition function of the Potts field w.r.t. β , which can be extremely demanding in terms of computations. **(Main paper contributions)** In [IJ-5], we proposed a fast partition function estimation strategy for 2D and 3D Potts fields with irregular shapes. This technique was applied to the estimation of the brain activity in functional MRI time-series. In this application, about 100 Potts fields were used to spatially regularize the detection of brain activation/deactivation/inactivation, where the regularization level was automatically-tuned and region-wise. **(Personal contributions)** I was the main contributor to this work. I collaborated with T. Vincent (who was a PhD student of P. Ciuciu at CEA Saclay) to integrate the partition function estimation strategy to the brain activity analysis pipeline. I also worked with F. Forbes, J. Idier and P. Ciuciu to develop my insights about the proposed method. This paper is shown in Appendix B.2.

Simultaneous Multiscale Registration using Large Deformation Diffeomorphic Metric Mapping [IJ-6]

(Motivation) This paper was motivated by the lack of literature in 2011 on the choice of physiologically realistic regularizing metrics to register medical images with LDDMM [BMTY05b]. **(Main paper contributions)** The impact of the regularizing metric in medical image registration was first discussed. In particular, we have made clear that using unsuitable regularizing metrics with respect to the registered structures leads to physiologically implausible deformations, even if the shape boundaries are accurately matched. Motivated by real-life medical image registration cases, a strategy to define multi-scale metrics in LDDMM was presented, assessed and discussed. **(Personal contributions)** I was the main contributor to this work. The key ideas came by discussing with F.X. Vialard when developing our implementation of [BMTY05b] on 3D medical images. I realized that very little literature was dealing with the choice of the metric in LDDMM although this choice is fundamental in practice. This paper is shown in Appendix A.1.

Diffeomorphic 3D Image Registration via Geodesic Shooting using an Efficient Adjoint Calculation [IJ-7]

(Motivation) This work was motivated by the need for an accurate tool to compute the initial momenta that compare 3D images in the LDDMM framework. Initial momenta are indeed important in LDDMM, as they compactly encode local differences between the registered images and can therefore be used for further statistics on shape spaces or for the estimation of average shapes. Note that they are specific to LDDMM in the image registration community and are one of the main reasons that make this formalism appealing. **(Main paper contributions)** A new variational strategy for the diffeomorphic registration of 3D images is defined. It performs the optimization on the set of geodesic paths instead of on all the possible curves, and therefore directly estimates the initial momenta comparing two images. **(Personal contributions)** I tightly collaborated with F.X. Vialard on this paper. My main contributions have dealt with the resolution of implementation and numerical issues related to the use of the geodesic shooting strategy on 3D medical images. This paper is shown in Appendix A.2.

Diffeomorphic Atlas Estimation using Geodesic Shooting on Volumetric Images [IJ-8]

(Motivation) Computing the average shape of a given organ is fundamental for many medical image applications, in particular in brain imaging. It indeed makes it natural to propagate local information measured on a reference set of imaged organs into this average shape, denoted template. This information (typically a probabilistic segmentation) can then be propagated to other images. Another important application is to quantify the local variability of the reference images. The motivation of this paper is to define a computationally tractable strategy to compute average shapes out of 3D medical images. **(Main paper contributions)** A new algorithm to compute intrinsic means of organ shapes from 3D medical images was defined. This algorithm is based on the geodesic shooting algorithm of [IJ-7] and is fully diffeomorphic. Contrary to other template definition strategies, the intensities of the average shapes are then not the average intensities of several images registered to each other, leading to sharper region boundaries. This strategy also offers interesting properties for further statistical studies by using the information contained in initial momenta. **(Personal contributions)** F.X. Vialard and me contributed equally to this work. F.X. Vialard computed the gradients of the optimized energy and I developed the gradient descent based strategy. This paper is shown in Appendix A.3.

Mixture of Kernels and Iterated semidirect Product of Diffeomorphisms Groups [IJ-10]

(Motivation) This work directly follows [IJ-6], where we defined and discussed a practical method to use multi-scale metrics in LDDMM. A first attempt to distinguish scale-dependent deformations out of an optimal deformation between two shapes was given in [IJ-6], but this contribution was secondary compared with other ones. As it may be useful for further statistical studies, the work of [IJ-10] strongly develops this discussion. **(Main paper contributions)** The influence of different scales when comparing two shapes using LDDMM with

multi-scale kernels is studied with a more rigorous model than in [IJ-6]. A variational approach is developed for the multiscale analysis of diffeomorphisms and the semidirect product representation is generalized to several scales. **(Personal contributions)** F.X. Vialard and M. Bruveris (who was a PhD student of D.D. Holm at Imperial College London) were the main contributors to the developed model. My first contribution was to give the research directions of this work and to specifically make sure that the mathematical developments would have a practical impact from a image analysis point of view and would be algorithmically realistic on 3D images. I also implemented and assessed the strategy on 3D medical images. This paper is shown in Appendix A.4.

Piecewise-Diffeomorphic Image Registration: Application to the Motion Estimation between 3D CT Lung Images with Sliding Conditions [IJ-11]

(Motivation) Standard medical image registration models make the hypothesis that the deformations between registered images are smooth (and then continuous) everywhere. However, sliding conditions can be observed in medical images, for instance at the lung boundaries. This paper was then motivated by the need for diffeomorphic image registration models with sliding conditions. **(Main paper contributions)** We first defined a general strategy for modeling sliding conditions when registering 3D images in a piecewise-diffeomorphic framework. Compared with existing literature in 2012, this strategy ensured that the estimated deformations were invertible everywhere although they could be locally discontinuous. We also integrated the proposed strategy to the LD-DMM [BMTY05b] and the LogDemons [VPPA08] diffeomorphic registration frameworks. **(Personal contributions)** I was the main contributor to this work. I worked with F.X. Vialard to define the admissible Reproducing Kernel Hilbert Space in the LDDMM context. We also shared image pre-processing tasks with H.A. Baluwala (who was a PhD student of J.A. Schnabel at Univ. Oxford) that were also used in [IJ-12]. This paper is shown in Appendix A.5.

Construction of Diffeomorphic Spatio-temporal Atlases using Kärcher means and LDDMM [IC-22]

(Motivation) This work directly extends [IJ-8] where a fully diffeomorphic strategy was proposed to compute average shapes (atlases) out of 3D images. It was motivated by the need for the definition of fully diffeomorphic spatio-temporal atlases. In this context, each reference image is associated to an acquisition time and the average template evolves in time. The spatio-temporal atlas also spatially moves smoothly in time with no intensity change and any intensity blurring, which allows to preserve the sharpness of region boundaries, or to make move an atlas associated to a single segmentation. **(Main paper contributions)** Compared with [IJ-8], a straightforward contribution was to use a time-dependent kernel to weight the influence of each reference image at a given time. In order to make the temporal evolution of the template fully diffeomorphic and dense in time, our key contribution was to perform the spatio-temporal shape averaging on the tangent space of the evolution rather than on the space of images. **(Personal contributions)** I was the main contributor to this work. F.X. Vialard formalized the intuition I had about the spatio-temporal

averaging strategy on tangent spaces. This paper is shown in Appendix A.6.

Piecewise-diffeomorphic registration of 3D CT/MR pulmonary images with sliding conditions [IC-27]

(Motivation) The driving motivation of this paper was to make it possible to register multimodal 3D images with sliding conditions. The two technical issues addressed in this paper were (1) to make the estimation of local similarity gradients computationally tractable on large multimodal images, and (2) to strongly regularize the deformations, as the registered structures have a strongly different representations in the CT and MR images, while modeling local sliding conditions. **(Main paper contributions)** This paper directly applies the regularization strategy of [IJ-11] to locally constrain sliding deformations. Its main contribution is the use of approximated local gradient of mutual information to match the images which was first presented in [IC-23] and is developed here. **(Personal contributions)** I was the main contributor to this work. M.P. Heinrich (who was a PhD student of J.A. Schnabel at Univ. Oxford) helped me to pre-process the images and to assess the registration quality. This paper is shown in Appendix A.7.

Hybrid Feature-based Diffeomorphic Registration for Tumour Tracking in 2-D Liver Ultrasound Images [IJ-13]

(Motivation) Ultrasound (US) imaging is a widely accessible and low-cost image acquisition modality but also opens various questions in image analysis. This is due to the fact that it generates different artifacts and that it acquires 2D image sequences in a 3D domain. The specific driving motivation of [IJ-13] is to define a robust and accurate method to compensate for the breathing motion when tracking liver tumors in US imaging. **(Main paper contributions)** A whole diffeomorphic image registration pipeline was defined to follow the tumors. The PDE-based deformation model was inspired from the LogDemons framework of [VPPA08]. The main contribution of [IJ-13] was the definition of new matching forces that allow to robustly follow the tumor in 2D US image sequences. **(Personal contributions)** My main contribution to this paper was to scientifically lead the work of A. Cifor (who was a PhD student of J.A. Schnabel at Univ. Oxford) to integrate the image features she defined in an image registration framework and to assess the results. I also found a mathematical justification to the algorithm and its parameters through a PDE-based formulation of the registration algorithm. This paper is shown in Appendix A.8.

Longitudinal deformation models, spatial regularizations and learning strategies to quantify Alzheimer’s disease progression [IJ-15]

(Motivation) The early detection of Alzheimer’s disease (AD) is an important challenge for its efficient treatment through adapted drug delivery. In this paper, we worked on its detection based on local hippocampal shape changes in time. The hippocampus is indeed a subcortical structure which is known by the clinicians to be anatomically impacted by AD. **(Main paper contributions)** In this paper, we explored the use of different spatial regularization models in logistic regression to learn which local shape deformations optimally discriminate AD subjects from subject with Mild Cognitive Impairment. **(Personal**

contributions) J.B. Fiot (who was a PhD student of L.D. Cohen at Univ. Paris Dauphine) and F.X. Vialard were the main contributors to this work. I helped J.B. Fiot defining an average shape and aligning the images. We have also made the link between the regularization strategies and their physiological interpretation. This project is the one in which I started developing a scientific activity in machine learning. This paper is shown in Appendix C.1.

Diffeomorphic image matching with left-invariant metrics [B-2]

(Motivation) The Large Deformation by Diffeomorphic Metric Mapping (LDDMM) framework of [BMTY05b] was designed to regularize the deformations with the same smoothing properties in the whole image domain. This contribution presented an alternative problem formulation, denoted Left-LDM or LIDM, in which spatially-varying metrics make sense. **(Main paper contributions)** We first explored the use of left-invariant metrics on diffeomorphism groups based on reproducing kernels defined in the body coordinates of the source image. This approach differs from LDDMM, where right-invariant metric on a diffeomorphism group are used. A link with LDDMM was also established and a practical algorithm to register 3D images with LIDM was given. **(Personal contributions)** My main contribution was to guide the mathematical developments in strong collaboration with F.X. Vialard and T. Schmah so that the registration strategy would be realistically applied on 3D medical images. This has lead to a computationally tractable 3D image registration algorithm (very close to [BMTY05b]) where the final deformation is analytically the same as using LIDM although the path is different. This paper is shown in Appendix A.9.

Spatially-varying metric learning for diffeomorphic image registration. A variational framework [IC-32]

(Motivation) In medical image registration, it makes obvious sense that the deformations of different organs should be ideally regularized with different smoothing properties. Standard medical image registration algorithms however use spatially homogeneous smoothing properties for two main reasons: (1) This is mathematically and algorithmically much simpler, and (2) tuning spatially varying smoothing properties requires prior information on the registered structures that is generally not available. **(Main paper contributions)** In this paper, we build on the diffeomorphic registration model of [B-2] to define a strategy that learns optimal spatially-varying regularization metrics with respect to a learning set of reference images. The learning strategy is defined in a variational framework. **(Personal contributions)** I tightly collaborated with F.X. Vialard in this project. I gave research directions, so that the strategy would be computationally tractable on 3D medical images and would lead to meaningful results. I also defined a numerical solution to keep the learning dimension reasonable, implemented the strategy and tested it on real 3D medical images. This paper is shown in Appendix A.10.

Diffeomorphic registration with self-adaptive spatial regularization for the segmentation of non-human primate brains [IC-33]

(Motivation) The motivation of this paper is close to the one of [IC-32] and deals with the semi-automatic tuning of the smoothing properties in the reg-

istration of template images. Contrary to [IC-32], the goal of this paper is to alleviate the need for scale definition in the regularizing metric of a medical image registration algorithm. **(Main paper contributions)** The main methodological contribution of this paper is to explore a new strategy to automatically tune the spatial regularization of the deformations in medical image registration. To do so, the image registration model is an optimization strategy in which the deformations of the template are the weighted sum of reference deformations at different scales, and the weights are penalized with a L_1 norm (LASSO). Sparse non-null weights are then computed, leading to optimal scale selection. **(Personal contributions)** I was the main contributor to this work. L. Dolius, C. Fonta and M. Mescam acquired the images and helped me to interpret the results. This paper is shown in Appendix A.11.

Filling Large Discontinuities in 3D Vascular Networks using Skeleton- and Intensity-based Information [IC-34]

(Motivation) The segmentation of vascular networks often leads to discontinuities in the segmented vessels. For tumorous networks, no hypotheses can additionally be made on the network structures, due to the chaotic arrangement of their vessels. This makes it impossible to use standard gap filling algorithms for such vascular networks. **(Main paper contributions)** This paper extended [IJ-2], that I wrote during my PhD thesis work, with a gap filling strategy that combines both skeleton- and intensity-based information to fill large discontinuities. **(Personal contributions)** My main contribution to this paper was to scientifically lead the work of R. Bates (who was a PhD student of J.A. Schnabel at Univ. Oxford) in order to extend [IJ-2] and make it efficient with the data he had. This paper is shown in Appendix A.12.

How to calculate the barycenter of a weighted graph [IJ-19]

(Motivation) Undirected graphs with weighted edges and probability measures on their nodes are of particular interest to model complex phenomena. For instance, they may represent a social network with individuals talking about a given topic. In this case, the individuals are the nodes, the strength of the relation between two individuals is an edge weight, and the probability measures reflect the occurrences of a tag (e.g. an hashtag in twitter). There was no algorithm to compute the barycenter such structures in 2017 although this may be statistically informative. **(Main paper contributions)** In this paper, we introduced an original stochastic algorithm to find the Fréchet mean of such graphs. It relies on a noisy simulated annealing algorithm. **(Personal contributions)** I. Gavra (who was a PhD student of S. Gadat at Univ. Paul Sabatier/IMT) was the main contributor to this work. S. Gadat worked with here on the algorithm definition and its convergence. My main contribution was to lead I. Gavra's work to make her strategy usable on real data. This allowed us to make it algorithmically efficient on reasonably large graphs, to develop insights about its parametrization, and to identify practical issues which make this algorithm not scalable to large graphs. These issues were treated in the follow-up paper [SJ-6] where my scientific involvement was stronger. This paper is shown in Appendix B.3.

A DCE-MRI Driven 3-D Reaction-Diffusion Model of Solid Tumour Growth [IJ-20]

(Motivation) This work was motivated by the need for tumor growth prediction models to estimate the response to therapies. **(Main paper contributions)** This paper introduced an image-driven 3D reaction-diffusion model of avascular tumor growth in order to predict spatio-temporal tumor evolution. The model is calibrated using information derived from follow-up DCE-MRI images. It indeed consists in registering follow-up multi-layer images with constraints encoded in a non-linear reaction-diffusion model. The registration then consists in estimating the model parameters. Note that it can also be seen as a PDE-constrained optimization problem. **(Personal contributions)** I had two major contributions in this paper. The first one was to scientifically lead the work of T. Roque (who was a PhD student of J.A. Schnabel at Univ. Oxford) to transform the tumor growth equations she collected in the literature into a reaction-diffusion model which can be used based on DCE-MRI image information. I also discretized the equations to make the resolution scheme stable and sufficiently fast on 3D image domains. In addition, I advised T. Roque on a simple and pragmatic optimization strategy to automatically tune tumor specific model parameters. This paper is shown in Appendix A.13.

Regularized Multi-Label Fast Marching and Application to Whole-Body Image Segmentation [IC-37]

(Motivation) The segmentation of multiple structures such as lymph nodes in whole-body MR images of patients with tumors is a task which can be hardly automatized for two main reasons: (1) Structures boundaries are not visible everywhere, and (2) the patients and the structures to segment may have a large anatomical variability. User interventions are then necessary but should be as limited as possible, and related to particularly responsive algorithms. **(Main paper contributions)** We proposed a computationally efficient regularization strategy for the Fast Marching (FM) segmentation of multiple organs. The regularization stabilizes the segmentation of complex structures and has a low computational impact. We also integrated this regularized segmentation strategy to the 3DSlicer software so that clinicians could validate the methodology on real cases. **(Personal contributions)** I supervised this project in collaboration with F. Malgouyres with whom we deepened the first intuitions about the regularization strategy, S. Ken who gave us its driving motivation and participated to the results assessment, and S. Lebreton who integrated the algorithms to 3DSlicer. This paper is shown in Appendix A.14.

A representative variable detection framework for complex data based on CORE-clustering [SC-1]

(Motivation) Discovering representative information in high dimensional spaces with a limited number of observations is a recurrent problem in data analysis. Heterogeneity between the variables behavior and multiple similarities between variable subsets make the analysis of complex systems an ambiguous task. **(Main paper contributions)** This paper presents a formalism to robustly estimate the representative variables in such complex systems. The formalism is based on a novel graph clustering strategy, denoted CORE-clustering, adapted

to the addressed problem. The graphs encode the relations between different observed variables and the clusters are selected based on the number of variables they contain. The representative variables are finally the cluster centers, so the number of variables in each cluster can be seen as a regularization parameter. The method is additionally designed to be scalable to large datasets. **(Personal contributions)** The original idea of the CORE-clustering algorithm came from the PhD thesis work of A.C. Brunet with J.M. Loubes as a supervisor, but was only published in Arxiv [BAL⁺16]. I supervised C. Champion (PhD student IMT, co-supervised by J.M. Loubes and me) to totally re-design the methodology. We developed mathematical and algorithmic insights to make it efficient in the general complex data case. I also advised her in the experimental validation. We finally wrote together [SC-1], with advice from J.M Loubes. This paper is shown in Appendix C.2.

Distribution regression model with a Reproducing Kernel Hilbert Space approach [SJ-5]

(Motivation) Regression analysis is a predictive modeling technique that has been widely studied over the last decades with the goal to investigate relationships between predictors and responses. Extensions of the Reproducing Kernel Hilbert Space (RKHS) framework became popular to extend the results of the statistical learning theory in the context of regression of functional data as well as to develop estimation procedures of functional valued functions f . As far as the authors know, It has however not been extended so far to probability distribution spaces. **(Main paper contributions)** This paper introduces a strategy to solve the regression problem where the inputs belong to probability distribution spaces and the output predictors are real values. The regression function is composed of an unknown function f and an element of $\mathcal{H}(K)$, where $\mathcal{H}(K)$ is the RKHS induced by the kernel K defined on the set of mean embeddings of distributions to RKHS $\mathcal{H}(k)$. **(Personal contributions)** My main contribution in this paper was to guide the work of T. Bui (PhD student IMT, co-supervised by J.M. Loubes, P. Balaesque and me) in order to establish the link between the mathematical formalism she developed and the auto-acoustic response curves she studied. This was critical to understand the model and to obtain pertinent results. This paper is shown in Appendix C.3.

Online Barycenter Estimation of Large Weighted Graphs [SJ-6]

(Motivation) This paper follows [IJ-19], where an original strategy was proposed to compute the barycenter of undirected weighted graphs. The method of [IJ-19] has strong mathematical foundations but is not scalable to large graphs. In addition, although the formalism is general enough to address the online estimation of online graphs, this application is not clearly discussed in [IJ-19]. **(Main paper contributions)** In this paper, we extend [IJ-19] to efficiently estimate the barycenter of very large graphs. The online case, where empirical observations of the graph node probability measures are made in parallel to the barycenter estimation, is also discussed. Algorithmic aspects of the strategy are highlighted as they are directly related to the scalability of the method. **(Personal contributions)** I have tightly collaborated with I. Gavra to develop the methodology and we have similar contributions in [SJ-6]. This paper is shown

in Appendix B.4.

1.4 Teaching activity and students supervision

This section briefly develops the teaching activity I had in during career and gives an overview of the students I have supervised. Note that I have mixed in this section the activity I had during and after my PhD thesis work (defended in September 2007).

Teaching activity

- 2017-2018: Lectures and practical courses in Image Analysis (32 hours). Master 2 MAPI3 (Applied Mathematics). Paul Sabatier University, Toulouse.
- 2017-2018: Lectures and practical courses in Machine Learning (18 hours). Master students to University lecturers (context of a two weeks spring school). VNUHCM - University of Science, Ho-Chi-Minh city, Vietnam.
- 2017-2018: Practical courses in Statistics (16 hours). Master 2 of ISAE / Supaero, Toulouse.
- 2017-2018: Lecture and practical courses in GPU computing (4 hours). Master 2 of ISAE / Supaero, Toulouse.
- 2016-2017: Lectures and practical courses in Image Analysis (8 hours). Master 2 MAPI3 (Applied Mathematics). Paul Sabatier University, Toulouse.
- 2016-2017: Practical courses in Statistics (16 hours). Master 2 of ISAE / Supaero, Toulouse.
- 2006: Lectures and practical courses of Numerical Simulation in Fluid Mechanics (32 hours). Master 1 in Mechanics and Energetics, Paul Sabatier University, Toulouse.
- 2006: Practical courses of Stochastic Process applied to heterogeneous media (14 hours). Master 1 in Mechanics and Energetics, Paul Sabatier University, Toulouse.
- 2005: Lectures and practical courses of Numerical Simulation in Fluid Mechanics (44 hours). Master 1 in Mechanics and Energetics, Paul Sabatier University, Toulouse.
- 2005: Practical courses of Point Mechanics (20 hours). Licence 1 in Mathematics and Computer Science applied to Science (DEUG MIAS), Paul Sabatier University, Toulouse.
- 2004: Lectures and practical courses of Numerical Simulation in Fluid Mechanics (44 hours). Master 1 in Mechanics and Energetics, Paul Sabatier University, Toulouse.
- 2004: Practical courses of Point Mechanics (20 hours). Licence 1 in Mathematics and Computer Science applied to Science (DEUG MIAS), Paul Sabatier University, Toulouse.

Students supervision

Postdoc supervision

- 02/2018-: V. K. Ghorpade. Postdoctoral researcher in Applied Mathematics for medical image analysis: *Mutli-modal image registration of 3D whole-body medical images*. Co-supervision with F. Malgouyres (Pr IMT).

PhD student supervision

- 09/2017-: C. Champion (Applied Mathematics at l'INSERM/IMT): *Development of new strategies for the analysis of complex data*. Co-supervision with J.-M. Loubes (Pr IMT) and R. Burcelin (DR INSERM).
- 09/2016-: T. Trang Bui (Applied Mathematics at INSA Toulouse/IMT): *Regularization models for the analysis of auditory data*. Co-supervision with J.-M. Loubes (Pr IMT) and P. Balaesque (CR1 CNRS, UMR5288).

Master students supervision

- 2018: R. Vaysse (M1 Applied Mathematics/Data Analysis, Paul Sabatier University - 5 months): *Statistical analysis of data out of speech samples for Parkinson's disease detection*. Co-supervision with S. Déjean (IR UPS) and J. Farinas (Mcf UPS - UMR5505)
- 2017: V. Brès (M2 Applied Mathematics/Computer Science, ENSEEIHT - 6 months): *GPU computing with OpenCL to speed-up large graph clustering algorithms*.
- 2017: S. Lebreton (M2 Applied Mathematics/Computer Science, ENSEEIHT - 6 months): *Development of a C++ plugin in 3DSlicer for the semi-interactive segmentation of 3D medical images*. Co-supervision with F. Malgouyres (Pr IMT).
- 2017: N. Artigouha (M1 Computer Science, INSA Toulouse - 2 months): *Using the C++ Boost Graph Library to for the analysis of large graphs*.
- 2016: D. Grasselly (M1 Applied Mathematics, INSA Toulouse - 3 months): *Development of a Matlab code for the registration of lung images with sliding conditions*. Co-supervision with J. Fehrenbach (Mcf UPS/IMT).
- 2016: M. Verdier (M1 Applied Mathematics, INSA Toulouse - 3 months): *Induction of Bayesian networks from medical data*.
- 2016: M. Ralle (M2 Mathématiques, Paris Orsay University - 4 months): *Statistical analysis of the cochlear coil*. Co-supervision with J.M. Loubes (PR UPS/IMT).
- 2015: T. Berriat (M1 Applied Mathematics, INSA Toulouse - 3 months): *Statistical analysis of the cochlear coil*. Co-supervision with J.M. Loubes (PR UPS/IMT).
- 2014: A. Choury (M2 Applied Mathematics, INSA Toulouse - 6 months): *Statistical analysis of seismic wave propagation measures*. Co-supervision with J.M. Loubes (PR UPS/IMT) and P. Besse (PR INSA/IMT).

- 2013: L. Dolius (M2 Medical Imaging and Radiophysics, Paul Sabatier University - 6 months): *Quantitative Analysis of 3D brain images*. Co-supervision with C. Fonta (DR CerCo/CNRS) and M. Mescam (McF UPS/CerCo).
- 2010: A. Camphuis (M1 Supelec - 2 months): *Validation of a medical image registration algorithm*. Co-supervision with F.X. Vialard (Postdoctoral researcher at Imperial College London).
- 2008: A.L. Fouque (M2 ENS Cachan - 5 months): *Analysis of the BOLD signal in functional MRI*. Co-supervision with P. Ciuciu (CR CEA Saclay).
- 2005: V. Gratsac (M2 Computer Science, Nantes University - 6 months): *Segmentation of large vascular network images using Tensor Voting*.

I have finally supervised 8 trainees with License 2 and 3 levels from *prépa INPT*, *ENS Lyon* and *IUT Toulouse* on the implementation of different algorithms.

1.5 Bibliographic record

Refereed international journal papers

- [IJ-20] T. Roque, L. Risser, V. Kersemans, S. Smart, D. Allen, P. Kinchesh, S. Gilchrist, A. Gomes, J. A. Schnabel, and M. Chappell. A DCE-MRI driven 3-d reaction-diffusion model of solid tumour growth. *IEEE Transactions on Medical Imaging*, 37(3):712–23, 2018.
- [IJ-19] S. Gadat, I. Gavra, and L. Risser. How to calculate the barycenter of a weighted graph. *Informatics: Mathematics of Operations Research*, 2018.
- [IJ-18] S. Ribes, D. Didierlaurent, N. Decoster, E. Gonneau, L. Risser, V. Feillel, and O. Caselles. Automatic segmentation of breast MR images through a markov random field statistical model. *IEEE Transactions on Medical Imaging*, 2014.
- [IJ-17] B. W. Papiez, M. P. Heinrich, J. Fehrenbach, L. Risser, and J. A. Schnabel. An implicit sliding-motion preserving regularisation via bilateral filtering for deformable image registration. *Medical Image Analysis*, 2014.
- [IJ-16] J. Mirebeau, J. Fehrenbach, L. Risser, and S. Tobji. Anisotropic Diffusion in ITK. *The Insight Journal*, 2014.
- [IJ-15] J. B. Fiot, H. Raguet, L. Risser, L. D. Cohen, J. Fripp, F. X. Vialard, and ADNI. Longitudinal deformation models, spatial regularizations and learning strategies to quantify alzheimer’s disease progression. *NeuroImage: Clinical*, 2014.
- [IJ-14] T. Vincent, S. Badillo, L. Risser, L. Chaari, C. Bakhous, F. Forbes, and P. Ciuciu. Flexible multivariate hemodynamics fMRI data analyses and simulations with pyhrf. *Frontiers in Neuroscience*, 2014.

- [IJ-13] A. Cifor, L. Risser, D. Chung, E. M. Anderson, and J. A. Schnabel. Hybrid feature-based diffeomorphic registration for tumour tracking in 2-d liver ultrasound images. *IEEE Transactions on Medical Imaging*, 2013.
- [IJ-12] H. Y. Baluwala, L. Risser, J. A. Schnabel, and K. A. Saddi. Towards a physiologically motivated registration of diagnostic CT and PET/CT of lung volumes. *Medical Physics*, 40(2), 2013.
- [IJ-11] L. Risser, F. X. Vialard, H. Y. Baluwala, and J. A. Schnabel. Piecewise-diffeomorphic image registration: Application to the motion estimation between 3d CT lung images with sliding conditions. *Medical Image Analysis*, 2012.
- [IJ-10] M. Bruveris, L. Risser, and F. X. Vialard. Mixture of kernels and iterated semidirect product of diffeomorphisms groups. *SIAM Multiscale Modeling and Simulation*, 10(4):1344–68, 2012.
- [IJ-9] R. Guibert, C. Fonta, L. Risser, and Plouraboué F. Coupling and robustness of intra-cortical vascular territories. *NeuroImage*, 2012.
- [IJ-8] F. X. Vialard, L. Risser, D. Rueckert, and D. Holm. Diffeomorphic atlas estimation using geodesic shooting on volumetric images. *Annals of the British Machine Vision Association*, 2012.
- [IJ-7] F. X. Vialard, L. Risser, D. Rueckert, and C. J. Cotter. Diffeomorphic 3d image registration via geodesic shooting using an efficient adjoint calculation. *International Journal of Computer Vision*, 2011.
- [IJ-6] L. Risser, F. X. Vialard, R. Wolz, M. Murgasova, D. Holm, D. Rueckert, and ADNI. Simultaneous multiscale registration using large deformation diffeomorphic metric mapping. *IEEE Transactions on Medical Imaging*, 2011.
- [IJ-5] L. Risser, T. Vincent, F. Forbes, J. Idier, and P. Ciuciu. Min-max extrapolation scheme for fast estimation of 3d potts field partition functions. application to the joint detection-estimation of brain activity in fMRI. *Journal of Signal Processing Systems*, 60(1), 2010.
- [IJ-4] T. Vincent, L. Risser, P. Ciuciu, and J. Idier. Unsupervised spatial mixture modelling for within-subject analysis of fMRI data. *IEEE Transactions on Medical Imaging*, 29(4):1059–75, 2010.
- [IJ-3] L. Risser, F. Plouraboué, P. Cloetens, and C. Fonta. A 3d-investigation shows that angiogenesis in primate cerebral cortex mainly occurs at capillary level. *International Journal of Developmental Neuroscience*, 27(2):185–96, 2008.
- [IJ-2] L. Risser, F. Plouraboué, and X. Descombes. Gap filling in vessel networks by skeletonization and tensor voting. *IEEE Transactions on Medical Imaging*, 27(5):674–87, 2008.
- [IJ-1] L. Risser, F. Plouraboué, A. Steyer, P. Cloetens, G. Le Duc, and C. Fonta. From homogeneous to fractal normal and tumorous micro-vascular networks in the brain. *Journal of Cerebral Blood Flow and Metabolism*, 27:293–303, 2006.

Refereed national journal papers

- [NJ-2] J. Braga, P. Bouvier, J. R. Dherbey, P. Balaesque, L. Risser, J.-M. Loubes, J. Dumoncel, B. Duployer, and C. Tenailleau. Echoes from the past: new insights into the early hominin cochlea from a phylogenetic approach. *Comptes Rendus Palevol*, 2017.
- [NJ-1] P. Ciuciu, T. Vincent, L. Risser, and S. Donnet. A joint detection-estimation framework for analysing within-subject fMRI data. *Journal de la Société Française de Statistique*, 151(1):58–89, 2010.

Papers submitted to journals

- [SJ-6] I. Gavra and L. Risser. Online barycenter estimation of large weighted graphs. *Mathematical Programming Computation (submitted)*.
- [SJ-5] T. Bui, J.-M. Loubes, L. Risser, and P. Balaesque Distribution regression model with a Reproducing Kernel Hilbert Space approach. *The Canadian Journal of Statistics (submitted)*.
- [SJ-4] E. J. De Jager, A. N. Van Schoor, J. W. Hoffman, A. C. Oetl  , C. Fonta, M. Mescam, L. Risser, A. Beaudet Sulcal pattern variation in extant human endocasts. *Journal of Anatomy (submitted)*
- [SJ-3] L. Keller, Q. Wagner, D. Offner, M. Pugliano, Y. Arntz, N. Messadeq, A. Priya, W. Wolfgang, P. Schwint  , L. Risser, and N. Benkirane-Jessel Synergic therapeutic effect of mesenchymal stem cells together with angiogenic nanocontainers as a new strategy to vascularize bone substitutes. – *(submitted)*
- [SJ-2] L. Risser, A. Sadoun, M. Mescam, K. Strelnikov, S. Lebreton, S. Boucher, P. Girard, N. Vayssi  re, M. G. P. Rosa, and C. Fonta. In vivo probabilistic location of cortical areas in a 3D atlas of the marmoset brain. *Brain Structure and Function (major revision)*.
- [SJ-1] M. Costa, S. Gadat, P. Gonnord, and L. Risser. Cytometry inference through adaptive atomic deconvolution. *Journal of Nonparametric Statistics (minor revision)*.

Books and book chapters

- [B-2] T. Schmah, L. Risser, and F. X. Vialard. Diffeomorphic image matching with left-invariant metrics. In *Fields Institute Communications 73 - Geometry, Mechanics, and Dynamics: The Legacy of Jerry Marsden*. 2015.
- [B-1] L. Risser. Analyse quantitative de r  seaux micro-vasculaires intra-corticaux. *PhD thesis*, Universit   de Toulouse, 2007.

Submitted book chapters

- [SB-1] F. X. Vialard, and L. Risser. Spatially varying metrics for LDDMM. *Riemannian Geometric Statistics in Medical Image Analysis (submitted)*.

Refereed international conference proceedings

- [IC-37] L. Risser, S. Ken, S. Lebreton, E. Grossiord, S. Kanoun, and F. Malgouyres. Regularized multi-label fast marching and application to whole-body image segmentation. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*, 2018.
- [IC-36] G. Fort, L. Risser, Y. Atchadé, and E. Moulines. Stochastic fista algorithms: so fast? In *Proceedings of IEEE Statistical Signal Processing Workshop (SSP)*, 2018.
- [IC-35] E. A. Schmidt, O. Maarek, J. Despres, M. Verdier, and L. Risser. Icp: From correlation to causation. In *Intracranial Pressure & Neuromonitoring XVI*, pages 167–171. Springer International Publishing, 2018.
- [IC-34] R. Bates, L. Risser, B. Irving, B. Papiez, P. Kannan, V. Kersemans, and J. A. Schnabel. Filling large discontinuities in 3d vascular networks using skeleton- and intensity-based information. In *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2015.
- [IC-33] L. Risser, L. Dolius, C. Fonta, and M. Mescam. Diffeomorphic registration with self-adaptive spatial regularization for the segmentation of non-human primate brains. In *Proceedings of International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2014.
- [IC-32] F. X. Vialard and L. Risser. Spatially-varying metric learning for diffeomorphic image registration. a variational framework. In *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2014.
- [IC-31] T. Schmah, L. Risser, and F. X. Vialard. Left-invariant metrics for diffeomorphic image registration with spatially-varying regularisation. In *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2013.
- [IC-30] B. W. Papiez, M. P. Heinrich, L. Risser, and J. A. Schnabel. Complex lung motion estimation via adaptive bilateral filtering of the deformation field. In *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2013.
- [IC-29] A. Cifor, L. Risser, M. P. Heinrich, D. Chung, and J. A. Schnabel. Rigid registration of untracked freehand 2d ultrasound sweeps to 3d CT of liver tumours. In *Proceedings of MICCAI Workshop on Computational and Clinical Applications in Abdominal Imaging (MICCAI-ABDI)*, 2013.
- [IC-28] J. B. Fiot, L. Risser, L. Cohen, J. Fripp, and F. X. Vialard. Local vs global descriptors of hippocampus shape evolution for alzheimer’s longitudinal population analysis. In *Proceedings of MICCAI Workshop on Spatiotemporal Image Analysis for Longitudinal and Time-Series Image Data (MICCAI-STIA)*, 2012.

- [IC-27] L. Risser, M. P. Heinrich, T. Matin, and J. A. Schnabel. Piecewise-diffeomorphic registration of 3d CT/MR pulmonary images with sliding conditions. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*, 2012.
- [IC-26] A. Cifor, L. Risser, D. Chung, E. M. Anderson, and J. A. Schnabel. Hybrid feature-based log-demons registration for tumour tracking in 2-d liver ultrasound images. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*, 2012.
- [IC-25] M. Bhushan, J. A. Schnabel, M. P. Heinrich, L. Risser, M. Brady, and M. Jenkinson. Motion correction and parameter estimation in DCEMRI sequences: Application to colorectal cancer. In *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2011.
- [IC-24] L. Risser, H. Baluwala, and J. A. Schnabel. Diffeomorphic registration with sliding conditions: Application to the registration of lungs CT images. In *Proceedings of MICCAI Workshop on Pulmonary Image Analysis (MICCAI-PIA)*, 2011.
- [IC-23] L. Risser, M.P. Heinrich, D. Rueckert, and J. A. Schnabel. Multi-modal diffeomorphic registration using mutual information: Application to the registration of CT and MR pulmonary images. In *Proceedings of MICCAI Workshop on Pulmonary Image Analysis (MICCAI-PIA)*, 2011.
- [IC-22] L. Risser, F. X. Vialard, A. Serag, P. Aljabar, and D. Rueckert. Construction of diffeomorphic spatio-temporal atlases using k rcher means and lddmm: Application to early cortical development. In *Proceedings of MICCAI Workshop on Image Analysis of Human Brain Development (MICCAI-IAHBD)*, 2011.
- [IC-21] F. X. Vialard, L. Risser, D. D. Holm, and D. Rueckert. Diffeomorphic atlas estimation using karcher mean and geodesic shooting on volumetric images. In *Proceedings of Medical Image Understanding and Analysis (MIUA)*, 2011.
- [IC-20] L. Risser, F. X. Vialard, R. Wolz, D. Holm, and D. Rueckert. Simultaneous fine and coarse diffeomorphic registration: Application to the atrophy measurement in alzheimer’s disease. In *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2010.
- [IC-19] D. P. Zhang, L. Risser, F. X. Vialard, O. Friman, L. Neefjes, N. Mollet, W. Niessen, and D. Rueckert. Coronary motion estimation using probability atlas and diffeomorphic registration from CTA. In *Proceedings of International Workshop on Medical Imaging and Augmented Reality (MIAR)*, 2010.
- [IC-18] L. Risser, F. X. Vialard, M. Murgasova, D. Holm, and D. Rueckert. Large diffeomorphic registration using fine and coarse strategies. application to the brain growth characterization. In *Proceedings of International Workshop on Biomedical Image Registration (WBIR)*, 2010.

- [IC-17] D. P. Zhang, L. Risser, O. Friman, L. Neefjes, N. Mollet, W. Niessen, and D. Rueckert. Nonrigid registration and template matching for coronary motion modeling from 4d CTA. In *Proceedings of International Workshop on Biomedical Image Registration (WBIR)*, 2010.
- [IC-16] D. P. Zhang, L. Risser, C. Metz, N. R. Mollet, W. Niessen, and D. Rueckert. Coronary artery motion modeling from 3d cardiac CT sequences. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*, 2010.
- [IC-15] L. Risser, T. Vincent, F. Forbes, J. Idier, and P. Ciuciu. How to deal with brain deactivation in the joint detection-estimation framework? In *Proceedings of Human Brain Mapping (HBM)*, 2010.
- [IC-14] L. Risser, J. Idier, and P. Ciuciu. Extrapolation schemes for fast ising field partition functions estimation. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, 2009.
- [IC-13] L. Risser, T. Vincent, P. Ciuciu, and J. Idier. Robust extrapolation scheme for fast estimation of 3d ising field partition functions: Application to within-subject fMRI data analysis. In *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2009.
- [IC-12] L. Risser, T. Vincent, and P. Ciuciu. Extrapolation scheme for fast 3d ising field partition function estimation. application to bayesian analysis of within subject fMRI data. In *Proceedings of Human Brain Mapping (HBM)*, 2009.
- [IC-11] T. Vincent, L. Risser, and P. Ciuciu. Spatially adaptive mixture modeling for analysis of fMRI time series. In *Proceedings of Human Brain Mapping (HBM)*, 2009.
- [IC-10] A. L. Fouque, P. Ciuciu, and L. Risser. Multivariate spatial gaussian mixture modeling for statistical clustering of hemodynamic parameters in functional MRI. In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2009.
- [IC-9] L. Risser, F. Plouraboué, and X. Descombes. Vascular network topology extraction from 3d high resolution x-ray images using tensor voting. In *Proceedings of International Conference on Approximation Methods and numerical Modeling in Environment and Natural Resources (MAMERN)*, 2009.
- [IC-8] L. Risser, P. Ciuciu, T. Aso, and D. Le Bihan. Brain activation detection using diffusion-weighted and bold fMRI: a comparative study. In *Proceedings of MICCAI workshop on Computational Diffusion MRI (MICCAI-CDMRI)*, 2008.
- [IC-7] L. Risser, F. Plouraboué, and C. Fonta. Angiogenesis in brain : a comparative study of local and organizational vascular structures. In *Proceedings of Neuroscience*, 2008.

- [IC-6] L. Risser, F. Plouraboué, and C. Fonta. Micro-vascular maturation in cerebral cortex: a quantitative high resolution study. In *Proceedings of Forum of European Neuroscience*, 2008.
- [IC-5] L. Risser, F. Plouraboué, C. Fonta, and A. Steyer. Cortical micro-vascular inhomogeneities: a high resolution 3d investigation. In *Proceedings of Annual Meeting of the Organization for Human Brain Mapping (HBM)*, 2007.
- [IC-4] L. Risser, F. Plouraboué, and X. Descombes. Gap filling in 3d vessel like patterns with tensor voting. In *Proceedings of International Conference On Computer Vision Theory and Applications (VISAPP)*, 2007.
- [IC-3] L. Risser, F. Plouraboué, P. Cloetens, A. Steyer, and C. Fonta. Maturation of marmoset cortical cerebral vascular networks: high-resolution 3d investigations. In *Proceedings of annual meeting Society for Neuroscience*, 2005.
- [IC-2] L. Risser, F. Plouraboué, P. Cloetens, A. Steyer, G. Leduc, L. Renaud, and C. Fonta. Normal and tumoral cerebral vascular networks: high resolution 3d investigations. In *Proceedings of Novel targeting drugs and radiotherapy*, 2005.
- [IC-1] C. Fonta, L. Risser, F. Plouraboué, L. Renaud, P. Cloetens, and A. Steyer. X-ray high resolution vascular network analysis in the cerebral cortex. In *Proceedings of Forum of European Neuroscience*, 2004.

Refereed French conference proceedings

- [NC-5] C. Champion, R. Burcelin, J.-M. Loubes, L. Risser. A new package CORE-Clust for robust and scalable analysis of complex data. In *Proceedings of Journées de Statistique de la SFdS*, 2018.
- [NC-4] G. Fort, L. Risser, E. Moulines, E. Ollier, and A. Samson. Algorithmes gradient-proximaux stochastiques. In *Proceedings of Groupe de Recherche et d'Etudes du Traitement du Signal (GRETSI)*, 2017.
- [NC-3] L. Risser, T. Vincent, P. Ciuciu, and J. Idier. Schémas d'extrapolation de fonctions de partition de champs de potts. application à l'analyse d'image en IRM fonctionnelle. In *Proceedings of Groupe de Recherche et d'Etudes du Traitement du Signal (GRETSI)*, 2009.
- [NC-2] A. L. Fouque, P. Ciuciu, L. Risser, and T. Vincent. Mélanges spatiaux gaussiens multivariés pour la classification de paramètres hémodynamiques en IRM fonctionnelle. In *Proceedings of Groupe de Recherche et d'Etudes du Traitement du Signal (GRETSI)*, 2009.
- [NC-1] L. Risser, F. Plouraboué, C. Fonta, P. Cloetens, and A. Steyer. Volumes élémentaires représentatifs dans les réseaux microvasculaires. In *Proceedings of Journées d'Etude sur les Milieux Poreux (JEMP)*, 2005.

Submitted international conference proceedings

- [SC-1] C. Champion, A.-C. Brunet, J.-M. Loubes, and L. Risser A representative variable detection framework for complex data based on CORE-clustering *Proceedings of International Conference on Algorithmic Learning Theory (ALT 2019) (submitted)*.

Chapter 2

Mathematical models in medical image analysis

2.1 Introduction

2.1.1 Medical image analysis

The medical image analysis community is a wide multidisciplinary community composed of applied mathematicians, computer scientists, specialists of signal processing, and data scientists who develop data analysis models related to medical images. This community is particularly active due to the constant progress of medical imaging acquisition devices. Image acquisition is indeed faster and faster, making it possible to acquire images with an increasingly accurate spatial resolution and image time series having shorter repetition times. Different image acquisition modalities and parameterizations, as well as different contrast agents also allow to quantify different information in a same diseased patient. Back in 2000, [DA00] already reported the impressive progress made in acquisition and analysis of medical images since the early works in the 70's. These progress are likely to continue in the future as they have an important impact in the society.

2.1.2 Medical image registration

Most of my contributions in medical image analysis deal with image registration. I then give a brief introduction to this field in this subsection.

Problem definition Consider a fixed image I_F and a moving image I_M . These images are defined on 2D or 3D discrete domains Ω_F and Ω_M with regularly sampled points denoted pixels in 2D and voxels in 3D. Although medical images may contain vectors or tensor structures at each of their points, we only treat the case where the pixels/voxels contain scalar values here. Medical images are also augmented with image to world matrices W_F and W_M which transform their pixel/voxel coordinates into world coordinates which are common to all registered images and expressed in millimeters. For instance, if $\mathbf{p}$ is a voxel of image I_F , its millimeter coordinates are $W_F \mathbf{p}$. Images I_F and I_M may indeed

be scanned with different resolutions, orientations and origins. In real applications, they additionally often cover different spatial domains although they contain at least a common organ of interest. Medical image registration then consists in mapping the common structures of interest I_F in I_M .

Similarity metrics We denote I_D the deformation of the moving image I_M by the sought mapping. A similarity metric $S(I_F, I_D)$ has naturally to be used to establish an optimal mapping between I_F and I_M . This metric must have low values when the common structures of I_F and I_D are accurately mapped and high values when they are distant to each other. For images acquired using the same modality, the hypothesis that a given structure has a given gray level often holds, at least after gray level pre-alignments. The most classic similarity metric is then the so-called Sum of Squared Difference (SSD) and is equal to $S(I_F, I_D) = \|I_F - I_D\|_2^2$. The images I_F and I_M may however be acquired using different imaging modalities (*e.g.* using CT and MR imaging) typically to quantify different physiological properties of a studied organ. For such multimodal images, a given structure has generally different gray levels in I_F and I_M . As a consequence, more complex similarity measures such as the mutual information (MI) [MCV⁺97] or edge-based techniques such as [SADDdS15] or Modality Independent Neighborhood Descriptors (MIND) [HJB⁺12a] must be used. It finally worth mentioning that only the intensities of specific structures may be different in the registered images, typically due to a pathology. In this case, techniques mixing image registration with the detection of these structures [HDJR12] as well as metamorphosis models [TY05, GY05, NHP⁺11] may be used.

Classic deformation models When I_F and I_M are acquired using different imaging modalities, it is common practice to rigidly align these images. Medical image registration then consists in modifying the properties of W_M with a translation and a volume preserving rotation that optimizes the similarity between I_F and the deformed moving image $I_D = I_M \circ W_M^{-1} \circ \phi_r \circ W_F$, where I_D is sampled in the fixed image domain and ϕ_r is the rigid deformation. Note that a simple generalization of rigid registration is affine registration, where the volume preserving constraints are released and shearing constraints can be captured. It is often used to propagate the segmentation of an image to another one which contains the same structures, with little degrees of freedom to optimize.

In order to estimate local deformations, more generic deformation models are however required. This can be for instance essential to quantify the local growth or shrinkage of an organ in follow-up images. Such deformations are encoded in a displacement field ϕ , which is a vector field that maps the coordinates of the fixed image to corresponding coordinates in the moving image. To simplify the notations, we suppose here that I_F and I_M are in the same image domain. For a given point $\mathbf{p}$ in I_F , $\mathbf{p} + \phi(\mathbf{p})$ is then the corresponding point in I_M , and we denote $I_D = I_M \circ \phi$ the deformed moving image. The registration of I_F and I_M now consists in optimizing ϕ so that $S(I_F, I_D)$ is low and ϕ is also physiologically plausible. Such deformation problems are often denoted non-rigid registration problems. The vectors of ϕ should for instance have reasonably large amplitudes and the smoothness of this vector field should be coherent with the biomechanical properties of the compared images. Non-rigid registration is

then often formulated as the following optimization problem:

$$\hat{\phi} = \arg \min_{\phi} S(I_F, I_M \circ \phi) + R(\phi), \quad (2.1)$$

where S is the similarity energy, R is the regularization energy, and the degrees of freedom of ϕ are optimized. This problem opened a whole field of research with questions related to the model for ϕ , the deformations regularization, the similarity metric, and obviously the optimization method. Surveys and classifications of such methods can be found for instance in [VMK⁺16, SDP13, Bro92].

A straightforward idea in non-rigid registration would be to use PDE-based biomechanical models of the registered shapes to constrain the deformations, as for instance in [HRS⁺99]. These methods however require to segment the registered structures and to parametrize the biomechanical models with unknown patient-specific model parameters. They are also not necessarily simple to code, non-linear and multi-dimensional PDEs being known as relatively complex to solve in mathematical engineering. Finally, no biomechanically sound deformation model is very well established for some organs, for instance the brain. Classic non-rigid registration models then use smoothing models derived from the heat equation as in optical flow models [HS81, LK81a]. PDE-based regularization models with spatially homogeneous smoothing properties are also reasonable as in elastic based registration [Bro92]. Naturally smoothing the deformations by estimating the mapping on a grid of control point and interpolating the deformations on the rest of the domain as in [RSH⁺99] is also standard in non-rigid and multimodal image registration. Convolution-based registration was finally made extremely popular in medical image registration by the Demons algorithm of [Thi96]. It is finally interesting to remark that most classic non-rigid image registration algorithms are relatively generic with spatially-homogeneous regularization properties.

Diffeomorphic deformation models An important category of medical image registration algorithms, which has been very popular these 10 last years and which I used in various papers is diffeomorphic image registration. Instead of encoding the deformations in displacement fields directly or in a grid of control points, they are encoded in a velocity field integrated in time:

$$\partial_t \phi(t, x) = v(t, \phi(t, x)), \quad (2.2)$$

where $x \in \Omega$ and $t \in [0, 1]$. Initial conditions for ϕ are typically null deformations $\phi(0, x) = \mathbf{0}$ for all $x \in \Omega$, and the estimated displacement field is $\phi(1, x)$. The key advantage of diffeomorphic image registration is that if the integration of v is performed on a sufficiently fine temporal grid, displacement field $\phi(1, x)$ is ensured to be a one-to-one (invertible) mapping. This property is highly desirable in most medical image registration applications, where a point in I_F corresponds to a single point in I_M and vice-versa. This property cannot be easily ensured using non-diffeomorphic techniques too. Note that two important classes of diffeomorphic image registration algorithms exist, those using stationary velocity fields as in the LogDemons algorithm [VPPA08] for instance, and those based on time-varying velocity fields as in LDDMM [BMTY05a] for instance. There were long debates about what was the best choice: stationary velocity fields lead to clearly faster and less memory demanding algorithms, have far less degrees

of freedom to estimate φ , and can be efficiently integrated using the method of [ACPA06]. Time-varying velocity fields are however more versatile, more natural to optimize using gradient descent based approaches and, in particular in the LDDMM context, can offer unique properties for further statistical studies as *e.g.* in [SFP⁺10]. Both strategies then make sense depending on the application.

Discussion From a broad data analysis point of view, I would conclude this subsection by claiming that medical image registration is a class of data mapping problems in which (1) reliable data information is strongly related to image gradients and is therefore sparsely distributed in space, (2) the degrees of freedom are typically much larger than the amount of available and reliable information, and (3) the mapping regularization is what makes the problem well posed and should be physiologically pertinent. Remark finally that there has been these last years an extremely strong interest given to new image registration frameworks based on neural networks [LKB⁺17]. These methods will not be discussed in this manuscript but could be promising in the future.

2.1.3 LDDMM image registration

The Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework is one of the main medical image registration formalisms. More generally, it is related to the field of *Computational Anatomy* [GM98, Pen09, YAM09] where the analytical and statistical study of biological shapes variability has been actively developed in the past fifteen years. In this context, the Riemannian formalism on shape spaces has provided efficient tools [PARP06, MTY02, MTY06, YMSM08, TMT02] allowing the use of powerful statistical methods developed for Riemannian manifolds [FLPJ04, FVJ08]. This formalism is the one in which I have developed several of my main scientific contributions. I therefore present it in this subsection.

Problem definition LDDMM image registration is based on a variational setting and the choice of a Riemannian metric. Its goal is to solve the diffeomorphic image registration problem, *i.e.* to estimate optimal smooth and invertible maps (diffeomorphisms) of the ambient space that represent a mapping between the points of a source image I_S and those of a target image I_T [DGM98, JDJG04, BMTY05b]. Note that the notations I_S and I_T are common in the LDDMM formalism, and correspond to the moving and fixed images (I_M and I_F , resp.) using more general image registration notations. This formalism is particularly adapted to the registration of most 3D medical images, where the hypothesis that the organ deformations are smooth is reasonable, and the topology of the represented organs is preserved.

Image registration is then formulated as an energy minimization problem, where the energy $\mathcal{E}(v)$ is:

$$\mathcal{E}(v) = \frac{1}{2} \int_0^1 \|v(t)\|_V^2 dt + \|I_S \circ \varphi^{-1} - I_T\|_{L^2}^2, \quad (2.3)$$

with:

$$\begin{aligned}\partial_t \varphi(t, x) &= v(t, \varphi(t, x)) \\ \varphi(0, x) &= x \quad \forall x \in \Omega.\end{aligned}\tag{2.4}$$

As explained Section 2.1.2, a time-dependent velocity field v is the optimized structure with respect to $\mathcal{E}$. It is defined on the source image domain and for times $t \in [0, 1]$. The similarity measure between the deformed source image $I_S \circ \varphi^{-1}$ and the target image I_T is the sum of squared differences. This similarity measure is standard for gray level images (see Section 2.1.2). The norm $\|v(t)\|_V$ also controls the smoothness of the optimal deformations and will be further discussed in the next paragraph. At time t , it can be computed using $\|v(t)\|_V = \langle \mathcal{F}(v(t)) \mathcal{F}(K)^{-1}, \mathcal{F}(v(t)) \rangle_{L^2}$, where $\mathcal{F}(\cdot)$ represents the Fourier transform and $\langle \cdot, \cdot \rangle_{L^2}$ is the L^2 inner product. This makes appear the smoothing kernel K , which is directly related to the Reproducing Kernel Hilbert Space (RKHS) V , and is used to convolve the deformations when registering two images (see Alg. 1). As summarized Fig. 2.1, the flow constraints encode the trajectory of the points $x \in \Omega$: At time $t = 0$, a point x of the source image I_S is naturally at location $\varphi(0, x) = x$. Then, its motion at times $t \in [0, 1]$ is defined by the integration of the time-dependent velocity field $v(t, x)$. The transformed location of x at time $t = 1$ is finally $\varphi(1, x)$ and corresponds to the mapping of x in the target image I_T .

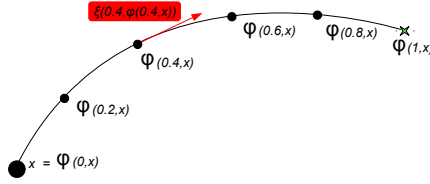


Figure 2.1: Transportation of the point $x \in \Omega$ through the diffeomorphism $\varphi(t, x)$, where Ω is the domain of the source image I_S . The point $\varphi(1, x)$ is the mapping of x in the target image I_T . Illustration out of [IJ-6].

Importantly, the gradient of $\mathcal{E}$ with respect to v can be analytically computed at each time t following [BMTY05b] as:

$$\nabla_v \mathcal{E}(t) = v(t) - K \star [Det J_{\phi_{t,1}} \nabla(\phi_{t,0} \circ I_S)((\phi_{t,0} \circ I_S) - (\phi_{t,1} \circ I_T))], \tag{2.5}$$

where ϕ_{t_j, t_i} transports an image from time t_i to time t_j through the diffeomorphism φ , and $Det J$ contains the determinant of the Jacobians of a deformation. Note that $Det J$ at a point x represents local volume variations. For instance, it is equal to 1 if the local volume is preserved, lower than 1 if it shrinks, and higher than 1 if it expands. Values lower to 0 also mean that the deformation is locally not invertible. The gradients Eq. (2.5) make it possible to minimize Eq. (2.3) with reasonable computational resources using gradient descent based optimization. LDDMM image registration of I_S and I_T therefore consists in minimizing Eq. (2.3) using the gradients Eq. (2.5) following the constraints Eq. (2.4).

Properties I now give an overview of the mathematical properties of LDDMM based on the developments of [SB-1]. In Eq. (2.3), V is a Hilbert space of vector fields on a Euclidean domain $\Omega \subset \mathbb{R}^d$. Importantly, the inclusion map $V \hookrightarrow W^{1,\infty}(\Omega, \mathbb{R}^d)$ (*i.e.* the space of vector fields which are Lipschitz continuous) is continuous. This assumption ensures that the computed maps are diffeomorphisms. The norm on V then controls the $W^{1,\infty}$ norm. These spaces are included in the family of the reproducing kernel Hilbert spaces (RKHS) [Aro50], which implies that they are completely defined by their kernel. The kernel is a function from the product space $\Omega \times \Omega$ into $\mathbb{R}^d$ which satisfies the above-mentioned assumption if it is sufficiently smooth. The direct consequence of this hypothesis on V is that the flow of a time dependent vector field in $L^2([0, 1], V)$ is well defined as explained [You08, Appendix C]. Then, the set of flows at time 1 defines a group of diffeomorphisms denoted by $\mathcal{G}_V$:

$$\mathcal{G}_V \stackrel{\text{def}}{=} \{\varphi(1) : \exists v \in L^2([0, 1], V) \text{ s.t. } \phi(v)\}, \quad (2.6)$$

where $\phi(v) = \varphi(1)$ and φ solves Eq. (2.4) [Tro95]. Trounev then defined a metric on this group:

$$\text{dist}(\psi_1, \psi_0)^2 = \inf \left\{ \int_0^1 \|v\|_V^2 dt : v \in L^2([0, 1], V) \text{ s.t. } \psi_1 = \phi(v) \circ \psi_0 \right\} \quad (2.7)$$

under which he proved that $\mathcal{G}_V$ is complete. It is finally important to emphasize that the deformations between two images in LDDMM are optimal paths, or geodesics, between the images. As discussed in [You07], these 3D+time deformations have therefore shooting properties and can be entirely encoded in a 3D scalar field: the initial momentum P_0 .

Implementation Different ideas related to the implementation of the LDDMM framework are now discussed. This discussion specifically builds on [BMTY05b] where a practical algorithm of LDDMM for image matching was given. We then give hereafter an overview of this algorithm, plus different numerical strategies we used to make it work efficiently.

When registering two images, one has first to define a discrete domain on which the time-dependent vector fields $\varphi(t, x)$ and $v(t, x)$ are computed, where $\varphi(t, x)$ is the mapping of x at time t through φ and $v(t, x)$ is velocity field integrated in time to compute φ . A natural choice is to use a spatial grid defined by the pixel/voxel coordinates of I_S . We denote $\hat{D}$ this discrete domain and recall that D is the dense image domain. As discussed Section 2.1.2, we also make the hypothesis that I_S and I_T are in the same image domain to simplify our notations. In our implementation, we also used an uniformly sampled grid to discretize t . The grid time step should also be sufficiently small to avoid generating non-invertible deformations when temporally integrating v . About 10 time steps are enough in most applications but more time steps may be necessary when sharp deformation are computed (see *e.g.* [IJ-11]).

We use the following notations to describe the registration algorithm: The t_θ , $\theta \in \{1, \dots, \Theta\}$ are the discrete time points. For each t_θ , several vector fields are required to encode useful deformations based on the diffeomorphism φ : $\phi_{t_j, t_i}(x)$ first transports $x \in \hat{D}$ from time t_i to time t_j through φ . The images I_{S, t_θ} and I_{T, t_θ} also correspond to I_S and I_T transported at time t_θ using ϕ_{0, t_θ}

and ϕ_{1,t_θ} respectively. Image registration is then a gradient descent algorithm where v is optimized with respect to I_S , I_T and the smoothing kernel K as shown Alg. 1.

Alg. 1 Interpreted LDDMM algorithm of [BMTY05b] to register the images I_S and I_T .

```

1: {Initialization}
2:  $\forall \theta \in \{1, \dots, \Theta\}$  and  $\forall x \in \hat{D}$ : the velocity  $\mathbf{v}(t_\theta, x) = \mathbf{0}$ .
3: repeat
4:   {Compute the mappings between  $t = 1$  and  $t_\theta$ }
5:   for  $\theta = \Theta - 1 \rightarrow 0$  do
6:      $\forall x \in \hat{D}$ : Compute  $\phi_{1,t_\theta}(x)$  and  $\phi_{t_\theta,1}(x)$ .
7:   end for
8:   {Compute the smooth energy gradients}
9:   for  $\theta = 1 \rightarrow \Theta$  do
10:     $\forall x \in \hat{D}$ : Compute  $\phi_{0,t_\theta}(x)$ .
11:     $\forall x \in \hat{D}$ : Compute  $I_{S,t_\theta}(x)$  and  $I_{T,t_\theta}(x)$ .
12:     $\forall x \in \hat{D}$ :  $\mathbf{u}(t_\theta, x) = \epsilon_1 (\text{DetJ}(\phi_{t_\theta,1}(x)) \nabla I_{S,t_\theta}(x) (I_{S,t_\theta}(x) - I_{T,t_\theta}(x)))$ .
13:     $\mathbf{u}(t_\theta, \cdot) = K \star \mathbf{u}(t_\theta, \cdot)$ .
14:     $\forall x \in \hat{D}$ :  $\nabla_{\mathbf{v}} \mathbf{E}(t_\theta, x) = \mathbf{v}(t_\theta, x) - \mathbf{u}(t_\theta, x)$ 
15:   end for
16:   {Update  $\mathbf{v}$ }
17:    $\forall \theta \in \{1, \dots, \Theta\}$  and  $\forall x \in \hat{D}$ :  $\mathbf{v}(t_\theta, x) = \mathbf{v}(t_\theta, x) - \epsilon_2 \nabla_{\mathbf{v}} \mathbf{E}(t_\theta, x)$ 
18: until Convergence

```

In Alg. 1 the mappings $\phi_{t_1,t_2}(x)$ are computed using an Euler method from time t_2 to time t_1 . Another remark is that a simple and very efficient technique can be used to speed-up the convergence of this registration algorithm. So-called momentum methods [RHW86] are widely known in machine learning to speed-up the convergence of gradient descent algorithms in high dimension. At each iteration, it simply consists in updating the optimized variables with a linear combination of the current gradients and the previous update. Personal (unpublished) experience has shown that this technique is particularly efficient in image registration where, at a given iteration, the mapping has only converged in some regions but not in all of them.

Another interesting point to discuss to make the practical use of the LDDMM algorithm efficient, is that it depends on two parameters ϵ_1 and ϵ_2 . In practice ϵ_1 should be sufficiently large so that $\mathbf{u}(t_\theta, x)$ has much more influence than $\mathbf{v}(t_\theta, x)$ in row 14 of Alg. 1. The vector field $\mathbf{u}(t_\theta, x)$ indeed pushes one image to the other and can be interpreted as a force field. The influence of $\mathbf{v}(t_\theta, x)$ should then be small but not negligible. This term is specific to LDDMM in the medical image registration community and indeed ensures the temporal consistency of the time-dependent deformations. The choice of ϵ_2 is more conventional in a gradient descent algorithm and controls the convergence speed. An empirical technique to tune it was given in [IJ-11]: At the first algorithm iteration, we compute $v_{max} = \max_{t_\theta, x} \|\nabla_{\mathbf{v}} \mathbf{E}(t_\theta, x)\|_2$. We then set ϵ_2 as equal to $0.5/v_{max}$, where 0.5 is in pixels/voxels, so that the maximum update at the first iteration is half a pixel/voxel. The updates have then a reasonable and automatically controlled amplitude.

Distributed codes Note finally that the implementation of [BMTY05b] on which I have worked is freely available on sourceforge in the *uTilzReg* package¹. It works on 2D and 3D nifti images, was implemented in *C++* and supports parallel computation using *openMP*. A GPU implementation of [IJ-7], which is methodologically close to [BMTY05b], was also developed by A. Martin (Master student at ENS Lyon) during his internship at the Institut de Mathématiques de Toulouse and is freely distributed on github². Its implementation is based on the codes of *uTilzReg* and uses the *openCL* programming language in the computationally intensive loops. On 3D medical images containing 192^3 voxels, it was shown to be about 60 times faster than the original *C++/openMP* code³. This makes obvious the interest of GPU computing in medical image registration, where the data are very well structured.

2.1.4 LogDemons image registration

Although most of my contributions in image registration were developed in the LDDMM formalism, I also worked in the LogDemons image registration formalism [VPPA08] in various papers (*e.g.* [IJ-11,IJ-17,IC-33]). This formalism has not the mathematical properties of LDDMM, given Subsection 2.1.3, but it has the important advantage to require far less memory and computational resources than LDDMM. It also ensures that the estimated deformations are diffeomorphic, and generally leads to similar deformations as by using LDDMM. It is therefore extremely popular in medical image registration, when the application is only to map two images. Note that I re-implemented in the LogDemons formalism different ideas developed (and mathematically justified) in the LDDMM framework, in order to apply them on real medical images, as in [IJ-11] for instance. Here is then a brief overview of this formalism.

Let I_S be a source image defined on the spatial domain $\Omega \subset \mathbb{R}^n$ and registered on a target image $I_T \in \Omega$. Here, I_S is transformed through the time-dependent diffeomorphic transformation $\phi_t^{\mathbf{v}}$, $t \in [0, 1]$ which is defined by a stationary velocity field $\mathbf{v}$ using: $\frac{\partial}{\partial t} \phi_t^{\mathbf{v}} = \mathbf{v}(\phi_t^{\mathbf{v}})$, where $\phi_0^{\mathbf{v}} = \text{Id}$. The final deformation is the exponential map of $\mathbf{v}$, $\exp(\mathbf{v}) \doteq \phi_1^{\mathbf{v}}$, and the deformed source image is then computed as $I_S \circ \exp(\mathbf{v})$. The optimal velocity field $\tilde{\mathbf{v}}$ is obtained by minimizing $\tilde{\mathbf{v}} = \arg \min_{\mathbf{v}} E(\mathbf{v}, \mathbf{v}_c)$, where the energy E is defined as:

$$E(\mathbf{v}, \mathbf{v}_c) = \frac{1}{\lambda_i^2} \|I_T - I_S \circ \phi_1^{\mathbf{v}_c}\|_{L^2}^2 + \frac{1}{\lambda_x^2} \|\log((\phi_1^{\mathbf{v}})^{-1} \circ \phi_1^{\mathbf{v}_c})\|_{L^2}^2 + \frac{1}{\lambda_d^2} \|\nabla \mathbf{v}\|_{L^2}^2 \quad (2.8)$$

where the logarithm is the inverse operation of the exponential. In this equation, the first term measures the sum of squared differences between the registered image intensities, the second term measures the correspondence between the smooth deformation $\phi_1^{\mathbf{v}}$ and the deformation $\phi_1^{\mathbf{v}_c}$, and the third term measures the spatial regularity of $\mathbf{v}$. Insights about the influence of the parameters λ_i , λ_x , λ_d are thoroughly developed in [MPS⁺11]. A key aspect of the demons algorithms is that it decouples the estimation of the optimal image matching (terms 1 and 2 of Eq. (2.8)) with the spatial regularization of the deformations

¹<https://sourceforge.net/projects/utlzreg/>

²<https://github.com/scalexm/GeoShoot/>

³2 minutes of computations on a Nvidia GTX 780 GPU using *openCL*, instead of 2 hours on a 32 cores Intel Xeon E5-2650 using *C++/openMP*

(terms 2 and 3 of Eq. (2.8)). As a result, the velocity field $\mathbf{v}_c$ encodes an intermediate transformation $\phi_1^{\mathbf{v}_c}$, called correspondence, which matches the two images without considering the regularity of the transformation. Minimization of $E(\mathbf{v}, \mathbf{v}_c)$ is as follows: The velocity field $\mathbf{v}$ is initialized as null. Then $E(\mathbf{v}, \mathbf{v}_c)$ is iteratively minimized using a two sub-step strategy: Firstly, $\mathbf{v}_c$ is computed using $\mathbf{v}$ and an update field $\delta\mathbf{v}$, defined as:

$$\delta\mathbf{v}(\mathbf{x}) = -\frac{I_T - I_S \circ \phi_1^{\mathbf{v}}}{\|\mathbf{J}(\mathbf{x})\|^2 + \lambda_i^2/\lambda_x^2} \mathbf{J}(\mathbf{x}) \quad (2.9)$$

where $\mathbf{J}(\mathbf{x})$ is the gradient of the image intensities, $\mathbf{J}(\mathbf{x}) = \nabla(I_S \circ \phi_1^{\mathbf{v}})$. Note that $\delta\mathbf{v}$ may be smoothed by a Gaussian kernel (fluid-like regularization). Ideally, $\mathbf{v}_c$ should be updated using $\mathbf{v}_c = \log(\phi_1^{\mathbf{v}} \circ \phi_1^{\delta\mathbf{v}})$. Since the logarithm of a deformation is computationally intractable in general, $\mathbf{v}_c$ is approximated using the Baker-Campbell-Hausdorff (BCH) formula: $\mathbf{v}_c \simeq \mathbf{v}_c + \delta\mathbf{v} + [\mathbf{v}, \delta\mathbf{v}]/2 + [\mathbf{v}, [\mathbf{v}, \delta\mathbf{v}]]$, where the Lie bracket $[\cdot, \cdot]$ is defined by $[\mathbf{v}_1, \mathbf{v}_2] = (\nabla\mathbf{v}_1)\mathbf{v}_2 - (\nabla\mathbf{v}_2)\mathbf{v}_1$. In the second sub-step, $\mathbf{v}$ is updated by smoothing $\mathbf{v}_c$ using a Gaussian kernel (diffusion-like regularization).

2.2 Medical image registration models

2.2.1 Summary of contributions

This section contains the main manuscripts related to my research activity on the definition of generic or application-specific image registration models.

Summarized papers Subsection 2.2.2 deals with a geodesic shooting model to perform diffeomorphic image registration in the LDDMM formalism and was published in [IJ-7] (see Appendix A.2). A project building directly on [IJ-7], was the one of [IJ-8, IC-22, IC-21] (see Appendices A.3 and A.6). In these papers, we developed new algorithms to compute intrinsic means of organ shapes from 3D medical images. They are presented Subsection 2.2.3. Subsection 2.2.4 then summarizes [B-2] (see Appendix A.9) which gives a mathematical justification to spatially-varying metrics in the LDDMM framework. Finally, Subsection 2.2.5 explains [IJ-20] (see Appendix A.13) which was not in developed the LDDMM framework. In this work, image deformations are driven by a physiologically motivated reaction diffusion model and only a few model parameters are optimized.

Other papers Other personal contributions in this topic are first in [IC-31], which is a preliminary proceeding to [B-2]. [IC-25] is also a strategy for the motion correction and parameter estimation in dcmMRI sequences. My minor contribution in this paper was to make the registration algorithm diffeomorphic. Finally, direct applications of the average shape estimation of [IJ-8] are found in [IC-28, IJ-15] for the brain hippocampus and in [SJ-2] for marmoset brains.

2.2.2 Diffeomorphic image matching using geodesic shooting

Introduction

In [IJ-7] (see Appendix A.2), a geodesic shooting model was developed to perform diffeomorphic image registration in the LDDMM formalism. This work was motivated by the need for an accurate tool to compute initial momenta that compare 3D images in the LDDMM framework. Initial momenta are indeed important in LDDMM, as they are a key for further statistics on shape spaces and they are specific to LDDMM in the image registration community. A new variational strategy for the diffeomorphic registration of 3D images was defined. It performs the optimization directly on the set of geodesic paths instead of on all the possible curves.

Methodology

The key of [IJ-7] is a reformulation of Eq. (2.3) with a slightly different functional

$$\mathcal{E}(P_0) = \frac{1}{2} \int_D K(P_0 \nabla I_0)(x) P_0(x) \nabla I_0(x) \, dx + S(I(1)), \quad (2.10)$$

under the constraints

$$\begin{cases} \partial_t I + \langle \nabla I, v \rangle = 0, \\ \partial_t P + \operatorname{div}(Pv) = 0, \\ v + K(P_0 \nabla I_0)(x) = 0, \end{cases} \quad (2.11)$$

and with initial conditions $P(t=0) = P_0$ and $I(t=0) = I_0$. In this context, the function $P_0 : D \mapsto \mathbb{R}$ is called the initial momentum. We then denote

$$K(P_0 \nabla I_0)(x) = \int_D k(x, y) P_0(y) \nabla I_0(y) \, dy. \quad (2.12)$$

This quantity can be reformulated as an L^2 norm of the quantity $P_0 \nabla I_0$ for the square root of the kernel k . Moreover, the system (2.11) encodes the fact that the evolution of $I(t)$ is geodesic in the LDDMM setting. Therefore, this formulation transforms the problem of optimizing on the time dependent d dimensional vector field v into optimizing on a function P_0 defined on the domain D . As in most image registration algorithms, [IJ-7] optimizes P_0 using a gradient descent. Estimation of these gradients is at the heart of this paper. It is important to remark that the dimension of the optimized scalar field P_0 is much lower than the dimension of the time-dependent vector field v although they are directly related by the system (2.11). The optimization procedure is then more constrained, which makes it more efficient for the estimation of initial momenta.

Results

The most interesting result of [IJ-7] is the one which shows that the initial momenta computed using the proposed method are more accurate than those computed of the standard strategy of [BMTY05b] on synthetic data. This is

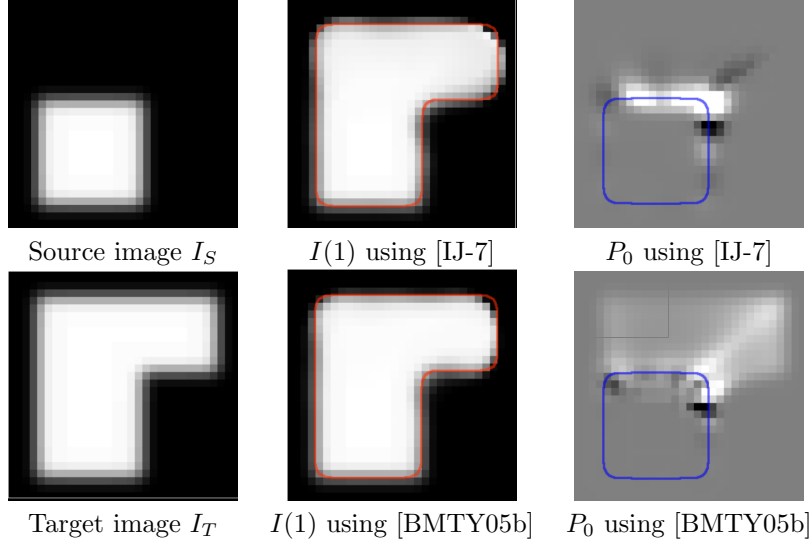


Figure 2.2: Registration of synthetic images using the methods of [BMTY05b] and [IJ-7]. The estimated deformed source image $I(1)$ and initial momenta P_0 are represented. Illustration out of [IJ-7].

illustrated Fig. 2.2. This tool to efficiently compute the initial momenta has then be the basis for further developments in which F.X. Vialard and me have developed new algorithms to compute intrinsic means of organ shapes from 3D medical images [IJ-8, IC-22, IC-21], as developed Subsection 2.2.3.

2.2.3 Karcher mean estimations for 3D images

Introduction

Computing the average shape of a given organ is fundamental for many medical image applications, in particular in brain imaging. It indeed makes it natural to propagate local information measured on a reference set of imaged organs into this average shape, denoted template. This information (typically a probabilistic segmentation) can then be propagated to other images. Another important application is to quantify the local variability of the reference images. The motivation of [IC-21, IJ-8] was then to define a computationally tractable strategy to compute average shapes out of 3D medical images. The proposed algorithm was based on the geodesic shooting algorithm of [IJ-7] presented Subsection 2.2.2 and is fully diffeomorphic. Contrary to standard template definition strategies, the intensities of the average shapes are then not the average intensities of several images registered to each other, leading to sharper region boundaries. This strategy also offers interesting properties for further statistical studies by using the information contained in initial momenta.

In [IC-22], the work of [IC-21, IJ-8] was extended as a fully diffeomorphic strategy to compute spatio-temporal atlases (time-dependent average shapes) out of 3D images on which a notion of time is associated (*e.g.* the age, time after a pathology onset). Compared with [IC-21, IJ-8], a straightforward contribution

was to use a time-dependent kernel to weight the influence of each reference image at a given time. The key contribution was however to perform the spatio-temporal shape averaging on the tangent space of the evolution rather than on the space of images. As a result, the temporal evolution of the template was fully diffeomorphic and dense in time. These contributions are summarized hereafter.

Volumetric atlas estimation using Karcher means

We build on the geodesic shooting formalism of [IJ-7] presented Subsection 2.2.2 and denote $P_0(A, B)$ the initial momentum representing the deformation to match a source image I_S to a target image I_T . We also consider a set of scans I^s weighted by h^s , $s \in [1, \dots, S]$ and representing shapes with the same topology. We use the methodology of [Pen06, FLPJ04] to estimate the weighted average of these shapes, denoted by A . Such an average shape is often denoted a template or atlas in medical imaging. The main advantage of this method is that it ensures that the structures topology of A is the same as in the I^s , even when their spatial variability is large. Accordingly to the Karcher mean [Kar77, FLPJ04], an average shape is the minimizer of $\mathcal{M}(A) \doteq \alpha^{-1} \sum_{s=1}^S h_s d(A, I^s)^2$, where $\alpha := \sum_{s=1}^S h_s$ and $d(A, I^s)$ is the distance between A and I^s . In our context, $d(A, I^s)^2 = \langle P_0(A, I^s), K \star P_0(A, I^s) \rangle_{L^2}$, where K is the smoothing kernel associated to the metric of the problem [BMTY05a]. Note that the uniqueness of A is generally not guaranteed. However, in finite dimensions, it can be proven that a unique minimizer to $\mathcal{M}(A)$ exists if the group of data lies in a sufficiently small neighborhood. The gradient of $\mathcal{M}$ with respect to the momentum variable is then given by:

$$\nabla^* \mathcal{M}(A) = \alpha^{-1} \sum_{s=1}^S h_s P_0(A, I^s). \quad (2.13)$$

After defining an initial guess, the average shape A is then estimated using a gradient descent: (1) For each image I^s , $s \in [1, \dots, S]$, compute the initial momentum $P_0(A, I^s)$ by registering A on I^s using LDDMM. (2) Compute the weighted mean of the momenta $P_0^{av} := P_0(A, I^s)$, $s \in [1, \dots, S]$ using Eq. (2.13). (3) Compute the deformed template image A_1 using the shooting equation Eq. (2.11) with initial conditions $P_0 = P_0^{av}$ and $I_0 = A$. (4) Update the template A with $A := A_1$. This iterative procedure is stopped when the norm of the gradient (Eq. 2.13) is below a given threshold.

Spatio-temporal atlas estimation using Karcher means

In the 3D+time context, we denote by I^s the s^{th} scan acquired at time τ^s for $s \in [1, \dots, S]$. We denote φ_τ^s , the unknown diffeomorphism which encodes the temporal evolution of the shape in I^s at times $\tau \in [\tau_{init}, \tau_{end}]$. We also denote by A_τ the temporal evolution of the average shape and by Φ_τ the diffeomorphism that deforms $A_{\tau_{init}}$ to A_τ , so $\Phi_{\tau_{init}}$ is the identity transformation. We measure the anatomical variability V_τ around A_τ using the PGA method of [FLPJ04]. In practice, the quantities A_τ and V_τ are estimated for values of τ regularly sampled between τ_{init} and τ_{end} and denoted τ_n , $n \in \{1, \dots, N\}$. Fig. 2.3 illustrates these notations. As discussed in [IC-22], a straightforward strategy to compute

A_τ is to replace h_s of Eq. (2.13) with $g(\tau_n - \tau^s)$, where g is a Gaussian kernel. This model however requires the estimation of $N \times S$ Karcher means at each iteration of the gradient descent algorithm, which is particularly time consuming. In addition, it does not control any relation between the diffeomorphisms φ_τ^s (between the images I^s and A_τ) and the diffeomorphism Φ_τ (between the averages $A_{\tau_{init}}$ and A_τ) when computing the initial momenta. We then developed another model which gives this control and requires the estimation of only S initial momenta P_0 at each gradient descent iteration, and summarize it below.

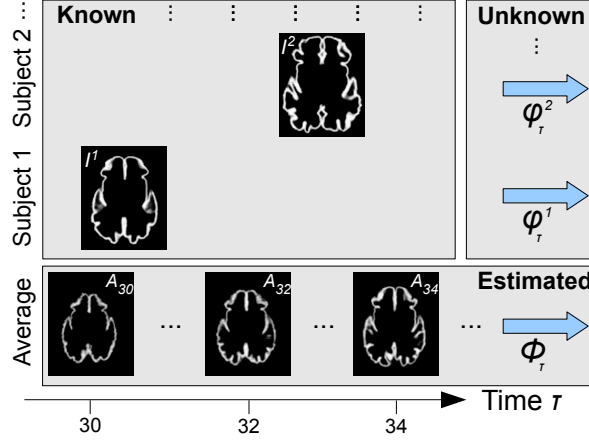


Figure 2.3: Estimation of the average spatio-temporal development of the cortex A_τ , $\tau \in [\tau_{init}, \tau_{end}]$ from segmented 3D images I^s , $s \in [1, \dots, S]$ of different subjects acquired at different time-points τ^s . Illustration out of [IC-22].

Subject-dependent spatio-temporal deformations of the cortical volumes φ_τ^s are supposed unknown. The initial momenta $P_0(A_{\tau_n}, I^s \circ \varphi_{\tau^s, \tau_n}^s)$ are then approximated with $T_{\tau^s \rightarrow \tau_n}(P_0(A_{\tau^s}, I^s))$, where $A_{\tau^s} = A_{\tau_n} \circ \Phi_{\tau^s, \tau_n}$ is known and $T_{\tau^s \rightarrow \tau_n}(\cdot)$ is the transportation of the momentum from time τ^s to time τ_n . We then estimate the unknown temporal evolution of I^s using this technique in a temporal window defined by the kernel g . The transportation of the momentum using $T_{\cdot \rightarrow \cdot}(\cdot)$ is similar to the problem of changing coordinate systems for spatio-temporal studies of shapes. We use a simple transportation defined as: $T_{\tau^s \rightarrow \tau_n}(P_0(A_{\tau^s}, I^s)) := P_0(A_{\tau^s}, I^s) \circ \Phi_{\tau_n, \tau^s}$, which is the first order approximation of the standard push-forward operation on momentum. Note that there is no agreement in the literature on the adequate methodology to address this problem. This transport of the momentum can be done using parallel transport [You07] or classical transformations such as in [RCSO⁺04].

The corresponding algorithm uses spatio-temporal averaging on the tangent space of the evolution and is summarized as follows: Define initial guesses of the A_{τ_n} , $n \in \{1, \dots, N\}$. Then compute a first estimation of Φ_τ , $\tau \in [\tau_{init}, \tau_{end}]$ using pairwise registration of successive images A_{τ_n} as done in the first algorithm. Using Φ_τ , A_τ can then be densely represented in time. Then repeat until convergence: (1) $\forall s$, estimate $P_0(A_{\tau^s}, I^s)$ (2) $\forall n$, compute $\nabla^* \mathcal{M}_{\tau_n}(A_{\tau_n}) = \alpha^{-1} \sum_{s=1}^S g(\tau_n - \tau^s) T_{\tau^s \rightarrow \tau_n}(P_0(A_{\tau^s}, I^s))$. (3) $\forall n$, update A_{τ_n} using the shooting system Eq. 2.11 and then update Φ_τ .

Results

We now give a brief overview of the results of [IC-22,IJ-8]. In [IJ-8], the convergence of the gradient descent strategy was assessed on 3D images representing shapes with three levels of complexity: simple hippocampi (subcortical brain structure), relatively flat brains, and well folded brains. In all cases, the algorithm converged after little iterations (from two iterations for the hippocampi to 5 iterations for the folded brains). In [IC-22], the spatio-temporal algorithm was also applied to compute the average brain growth in pre-term babies and their spatio-temporal anatomical variability, as shown Fig. 2.4. An amount of $S = 50$ reference T2 weighted MR images with about a millimetric resolution were used, and the gestational age of the scanned babies was from 29 to 37 weeks. The method was shown as being an interesting tool in this context, with a nice potential for further statistical studies. Its only limitation is that it can only be reasonably used on images in which the shape structures are clearly visible. In brain imaging this is the case for babies, subcortical structures, fossiles and all mammals except modern adult humans. For a reasonable use of this method on modern adult humans other landmarks should additionally be used.

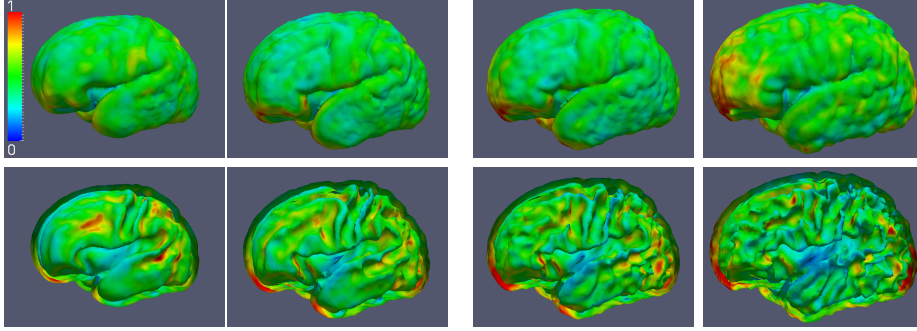


Figure 2.4: Average cortical surfaces of A_{τ_n} estimated at different times τ_n using spatio-temporal averaging on tangent spaces (second algorithm). From left to right, the age τ_n is 30, 32, 35 and 37 weeks of gestational age. Illustrations at the top represent the outer cortical surface and those at the bottom the inner cortical surface. Colors represent the normalized initial momentum variability V_{τ_n} at the cortical surface and are sampled identically in all images. Illustration out of [IC-22].

2.2.4 Left-invariant metrics for diffeomorphic image matching

Introduction

A natural extension of the LDDMM formalism with spatially homogeneous kernels sum of kernels [BMTY05b] consists in having a kernel which depend on their spatial location. Defining an extension of LDDMM justifying such kernels was then developed in [B-2,IC-31] (see Appendix A.9).

It is first important to mention that $\mathcal{G}_V$, defined Eq. (2.6), is right-invariant. This means that for every $\psi_1, \psi_2, \psi_3 \in \mathcal{G}_V$ and the distance of Eq. (2.7) the

following property holds:

$$\text{dist}(\psi_1 \circ \psi_3, \psi_0 \circ \psi_3) = \text{dist}(\psi_1, \psi_0). \quad (2.14)$$

However, the right-invariant point of view was designed for homogeneous materials with translation invariant properties. A consequence in LDDMM is that spatially varying or direction-dependent kernels have no obvious interpretation. More practically, the norm is defined in Eulerian coordinates but when t varies and the source image I_S is deformed by $\varphi(t, \cdot)$, a given point p of I_S moves through space (see Fig. 2.1). Conversely, a given point in space corresponds to different points in the deformed source image for different times t . Similarly, the directions in a direction-dependent kernel are defined with respect to Eulerian coordinates and not the coordinates of the moving source image.

Nonetheless, spatially-varying kernels have a high interest in medical image registration where the different registered structures have naturally different (and potentially non-isotropic) deformation properties. Before working on [IC-31] we already worked on such problematics in [IJ-11] (Section 2.3.3) to model sliding conditions between the lungs and the ribs and realized how important was this question for medical applications. Below is then an overview of this extension of LDDMM which naturally supports the use of spatially varying kernels.

Methodology

The framework of [B-2, IC-31] is based on a *left*-invariant metric, i.e. a norm in the body (Lagrangian) coordinates of the source image. We then denoted it LIDM for *Left Invariant Diffeomorphic Metrics*. Instead of applying the norm V to the spatial velocity defined by (2.4), it is then applied to the *convective velocity* $v(t)$ implicitly defined by

$$\partial_t \varphi(t) = d\varphi(t) \cdot v(t), \quad (2.15)$$

where $d\varphi(t)$ is the spatial derivative of $\varphi(t)$. The optimized energy Eq. (2.3) in LDDMM image registration is then constrained with $\partial_t \varphi_t = d\varphi_t \cdot v_t$ in LIDM, instead of the standard LDDMM constraint $\partial_t \varphi_t = v_t \circ \varphi_t$. An important result of [IC-31] to design a computationally tractable implementation of LIDM on 3D images is the following: both LIDM and LDDMM approaches lead to the same final deformation at time 1 if the same smoothing kernel is used. More generally:

- If ϕ_t minimizes $\mathcal{E}$ in LIDM, then $\varphi_t := \phi_{1-t}^{-1} \circ \phi_1$ minimizes $\mathcal{E}$ in LDDMM.
- If φ_t minimizes $\mathcal{E}$ in LDDMM, then $\phi_t := \varphi_1 \circ \varphi_{1-t}^{-1}$ minimizes $\mathcal{E}$ in LIDM.

Optimal paths in Left-LDM are left-geodesics and optimal paths in Right-LDM are right-geodesics, but all of them lead to the same deformation for a given metric. This result had an important consequence for the practical development of LDDMM-derived algorithms with spatially-varying metrics. If the final mapping is what matters for the application the LDDMM algorithm Alg. 1 can still be used with regularization properties depending on space. The only difference is the interpretation of the deformation.

Results

Fig. 2.5 illustrates the main result of [IC-31] on a synthetic example. In this example, LIDM registered the images using a kernel which is defined accordingly to a partition of unity which smoothly splits the spatial domain into two sub-regions. In the first subregion (white), where the disc is homogeneously translated, a Gaussian kernel with a large σ is used. A small σ is however used in the second region (black), where fine deformations are observed. In this example LIDM, performed better than [BMTY05b] with the sum of kernel [IJ-6,IJ-10] and the diffeomorphic and multi-scale strategy of [AEGG08]. Using a partition of unity and adapted regularization levels gave here an intuitive and efficient control to get the desired deformations.

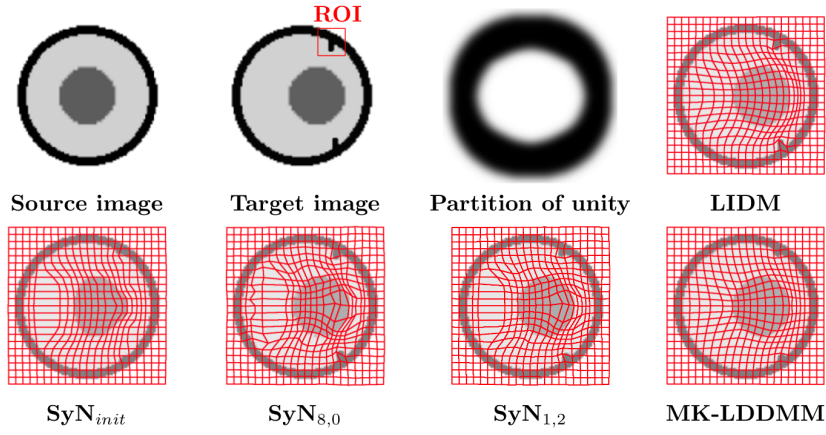


Figure 2.5: Comparison of the LIDM strategy of [IC-31] with spatially-varying regularization with two reference diffeomorphic registration algorithms with stationary regularization (LDDMM: [BMTY05b], [IJ-6] and SyN: [AEGG08]). Illustration out of [IC-31].

In general, the use of spatially varying kernels then appears as interesting for real medical images. However, the shortcoming of this approach is that the kernel still does not evolve with the deformed shape. For moderate deformations as in Fig. 2.5, this is not a problem. In particular, this was the starting point of the metric learning strategy in [IC-32] presented Subsection 2.3.4. This method however cannot reasonably be applied in the large deformations case where the kernel should depend on the shape itself. Such approaches have actually been developed in [You12, ATTY15, AMY16] in which the operator A depends on the shape itself, but developing models for images associated with an efficient implementation remains unsolved.

2.2.5 Image matching based on a reaction-diffusion model.

The last medical image registration model described in this section strongly differs from the other ones as it is not related to LDDMM and instead constraint image deformations with a physiologically motivated reaction-diffusion PDE model.

Motivation

This work was motivated by the need for avascular tumor growth prediction models to estimate the response to therapies. In this context, models of the tumor evolution with respect to physiological parameters are well established, so RKHS-based or other standard generic non-rigid registration algorithms are not adapted. We then introduced and assessed in [IJ-20] a novel image-driven 3D reaction-diffusion model of avascular tumor growth in order to predicts spatio-temporal tumor evolution. The model was calibrated using information derived from follow-up DCE-MRI images. It therefore corresponds to a PDE-constrained optimization problem where the registration consists in estimating the model parameters. Follow-up multi-layer images are indeed registered with deformations that are constrained by the PDE model. Tumor growth prediction finally consists in simulating the PDE model after the last acquisition time of the follow-up images that were used to calibrate the model.

Methodology

Deformation model The first main contribution of the paper was to gather different equations, out of the tumor evolution modeling literature, in a system of equations in a 3D image domain, where most physiological information can be derived from DCE-MRI images. This system is numerically solved in a 3D image domain and models the spatio-temporal evolution of the amount of proliferating p , hypoxic q and necrotic n cells in a tumor. Here is its formulation in a continuous domain at location $\mathbf{x}$ and time t :

$$\frac{\partial p}{\partial t} = \nabla(d \nabla p) + g(\eta)p \left(1 - \frac{r}{\Theta}\right) - f(\eta)p \quad (2.16)$$

$$\frac{\partial q}{\partial t} = \nabla(d \nabla q) + f(\eta)p - h(\eta)q \quad (2.17)$$

$$\frac{\partial n}{\partial t} = h(\eta)q, \quad (2.18)$$

where the diffusion coefficient d was modeled by a scalar field $d = \exp(-r/\kappa)$ in Eq. (21) of [IJ-20] and r is the total number of cells ($r = p + q + n$). Local proliferation, hypoxia and necrosis rate g , f and h are given by Equations (4)-(6) of [IJ-20] and are directly related to the nutrient distribution η (Equation (19) of [IJ-20]). Finally Θ is the carrying capacity of the tissue represented by the volume of a voxel as defined section 2.D of [IJ-20]. On a 3D domain, this system is therefore:

$$\frac{\partial p}{\partial t} = \frac{\partial d}{\partial x} \frac{\partial p}{\partial x} + \frac{\partial d}{\partial y} \frac{\partial p}{\partial y} + \frac{\partial d}{\partial z} \frac{\partial p}{\partial z} + d \frac{\partial^2 p}{\partial x^2} + d \frac{\partial^2 p}{\partial y^2} + d \frac{\partial^2 p}{\partial z^2} + A \quad (2.19)$$

$$\frac{\partial q}{\partial t} = \frac{\partial d}{\partial x} \frac{\partial q}{\partial x} + \frac{\partial d}{\partial y} \frac{\partial q}{\partial y} + \frac{\partial d}{\partial z} \frac{\partial q}{\partial z} + d \frac{\partial^2 q}{\partial x^2} + d \frac{\partial^2 q}{\partial y^2} + d \frac{\partial^2 q}{\partial z^2} + B \quad (2.20)$$

$$\frac{\partial n}{\partial t} = C, \quad (2.21)$$

where, at each point $(\mathbf{x}, t)$, $A = (1 - r/\Theta)gp - fp$, $B = fp - hq$, and $C = hq$. Discretization and numerical resolution of this system are developed in the additional document of [IJ-20], given in Appendix A.13.

Image registration The second main paper contribution was to show how to use this 3D non-linear model to predict the evolution of an avascular tumor. Consider two DCE-MRI time series acquired at time points TP1 and TP2. Different maps are extracted at TP1 in order to initiate the PDE-based model. These maps specifically represent at TP1 the number of proliferating cells (p), of hypoxic cells (q), of necrotic cells (n), as well as the nutrient distribution (η) and the diffusion coefficient of the hypoxic and proliferative cell densities (d).

Before solving the spatio-temporal PDE model, additional global parameters have also to be solved. In [IJ-20], three parameters related to the models of d , g and f were identified as pertinent to control tumor-specific behaviors. For given parameter values the model can be solved until TP2 and the propagated maps can be compared with the corresponding maps extracted from the DCE-MRI time series at TP2 to evaluate the parameters pertinence. The image registration model then consists in optimizing these parameters with the PDE system as a spatio-temporal constraint. In practice this optimization was performed using a simulated annealing strategy and the relative volume estimation error between the observed tumor volume and the true one was optimized.

Once the three parameters estimated, the model PDE system can then be solved after TP2 to predict the tumor evolution. An interesting application is then to compare the predicted evolution with the true one to evaluate the impact of a treatment after TP2.

Results and discussion

After validating the tumor registration strategy of [IJ-20] on synthetic data, it was assessed on nine preclinical cases of breast carcinoma. For each case, DCE-MRI derived physiological maps between TP1 and TP2 were registered. Each registration required about 15 hours and 1500 optimization step on a 3.4GHz Intel Xeon computer with 32GB of RAM, and image domains of size $128 \times 64 \times 64$. The optimal parameters were then used to predict the tumor evolution at follow-up time points TP3, TP4 and TP5. Propagated maps were then compared with the true ones. When excluding two cases in which the model hypotheses were not respected, the average tumor volume errors were 0.22 ± 0.19 , 0.40 ± 0.34 and 0.56 ± 0.37 at times TP3, TP4 and TP5, respectively. Mean relative errors in term of total cell number were a bit higher but still reasonable with 1.02 ± 0.95 , 1.65 ± 1.71 and 2.67 ± 2.33 at times TP3, TP4 and TP5, respectively. These higher values were expected as only average tumor volume errors was optimized by the algorithm. They however open interesting directions for refinements of the model in particular on the estimation of necrotic cells. Another extension of this method for vascularized tumors is also natural. A finer optimization model would finally make the parameters estimation much faster and robust.

2.3 Regularization metrics in medical image registration

2.3.1 Summary of contributions

This section presents different projects related to the regularization of the deformations in diffeomorphic registration, more specifically in the LDDMM formal-

ism. These projects were carried out in the context of a long term collaboration with F.X. Vialard and have lead to several of my main scientific contributions which I summarize here.

Summarized papers The paper [IJ-6], Subsection 2.3.2, was first motivated by the lack of literature in 2011 on the choice of physiologically realistic regularizing metrics to register medical images with LDDMM [BMTY05b]. It first discussed the impact of the regularizing metric in medical image registration. In particular, it made clear that using unsuitable regularizing metric, with respect to the registered structures, leads to physiologically implausible deformations even when the shape boundaries are accurately matched. Motivated by real medical image registration cases, a strategy to define multi-scale metrics in LDDMM was presented, assessed and discussed. A first attempt to measure the impact of each scale was given in [IJ-6] as this may be useful for further statistical studies. It was then strongly developed in [IJ-10] with a more mathematically-grounded model. This extension is also presented Subsection 2.3.2.

An extension of this work, which later motivated the work of [B-2] on spatially varying metrics, is [IJ-11] where we defined a general strategy for modeling sliding conditions when registering 3D images in a piecewise-diffeomorphic framework. This work is presented Subsection 2.3.3.

After having justified the use of spatially-varying registration in LDDMM [B-2] (Subsection 2.2.4), we also built on this paper to define a strategy that learns optimal spatially-varying regularization metrics with respect to a learning set of reference images. The learning strategy is defined in the variational framework of [IC-32] presented Subsection 2.3.4.

Other papers Other contributions which are not detailed here are first [IJ-17, IC-30] which deals with sliding constraints in diffeomorphic 3D image registration as in [IJ-11]. An important aspect treated in this paper was that sliding conditions location were automatically detected by the proposed algorithm, contrary to existing approaches in 2014. A spatially-dependent regularization scheme was introduced. It contains a non-linear term that depends on image intensities and the estimated deformations themselves, which makes it possible to detect discontinuous deformations. B. W. Papiez (who was a PhD student of J.A. Schnabel at Univ. Oxford) had the intuitions that have lead to the proposed algorithm and was the main contributor to this paper. I worked with J. Fehrenbach (lecturer at Univ. Toulouse, IMT) to cast the algorithm of B. W. Papiez into a PDE model. This allowed me to formally understand the algorithm and how to properly tune its parameters in a relatively unstable context. Subsection 2.5 and the recommendations section 3 are directly related to my collaboration with J. Fehrenbach. Another publication related to [IJ-11] is [IJ-12] and also deals with lungs registration with sliding motion. This approach was more application-grounded and dedicated to the registration of CT and PET images.

An alternative strategy to [IC-32], where optimal metrics are selected from the data, was also presented in [IC-33]. Contrary to [IC-32], where the spatially-varying regularization metric is learned before the registration, the strategy of [IC-33] pre-defined a reference basis of deformations at different scales and then registered the images with deformations encoded as a weighted sum of

the basis elements. The level of smoothing was then controlled by the most influential basis elements. To select optimal deformation scales, the weights were also regularized by using a LASSO term. This work is not presented in the main manuscript as I consider it is not mature enough. It is however given in Appendix A.11.

Finally, preliminary versions of [IJ-6] and [IJ-11] were published in the proceedings [IC-20, IC-18, IC-24].

2.3.2 Multi-scale metrics

Introduction

We now discuss the project presented in [IJ-6, IJ-10] (see Appendices A.1 and A.4) in which I worked on the development of multi-scale kernels in the LDDMM framework [BMTY05b]. In most applications, a Gaussian kernel is used to smooth the deformations. The Gaussian width σ is commonly chosen to obtain a good matching accuracy. This means that small values, close to the image resolution, are used for σ . One can then wonder what is the effect of this parameter on the structure of the deformation. This question is the starting point of this project.

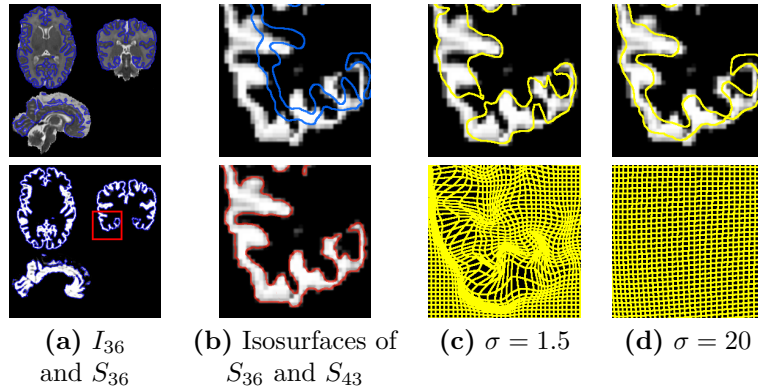


Figure 2.6: (a) gray matter extraction of the 3D MR image I_{36} (top) and resulting segmentation S_{36} (bottom). The red square indicates the 2D region of interest shown in (b-c). (b) The blue and red isolines represent the cortical surface of S_{36} and S_{43} , respectively. The gray levels are the segmented cortex of S_{43} . (c-d) The yellow isolines represent deformed cortical surfaces of S_{36} after LDDMM registration on S_{43} with $\sigma = 1.5$ and $\sigma = 20$, respectively. The grids represent the estimated dense deformations. Illustration out of [IJ-6].

In [IJ-6], we have first illustrated the influence of σ on the mapping obtained between two images of the gray matter acquired on a pre-term baby at about 36 and 43 weeks of gestational age, as summarized Fig. 2.6. Let us focus on the (b-top) subfigure of Fig. 2.6. The blue isoline represents the cortex boundary in a 2D region of interest (ROI) out of a 3D segmented image S_{36} and the ROI is located in the red square of the (a-bottom) subfigure. The gray levels of the same (b-top) subfigure also represent the segmented cortex in the same pre-term baby but 7 weeks later. It is obvious that the brain became globally larger as

the brain and the skull strongly grow at this age. The shapes should be almost shifted at the scale of this ROI to capture the amplitude of the deformation. Importantly, existing cortex folds also became deeper and new folds appeared, which is normal during brain maturation because the cortex growth is faster than the skull growth. Capturing the folding process requires registering the images at a scale close to the image resolution here. To conclude, the registration of these images requires at a same time a large σ and a small σ . If only a small or a large σ is used, the optimal path (and the optimization process) will either lead to physiologically implausible deformations as shown Fig. 2.6-(c) or does not capture fine deformations as shown Fig. 2.6-(d). This justifies the use of multi-scale kernels to establish geodesics between such follow-up medical images.

Sum of kernels

In LDDMM, the kernel spatially interpolates the rest of the information (i.e. the momentum) to drive the motion of the points where there is no gradient information, *e.g.* in flat image regions. Therefore, it is natural to introduce a sum of kernels to fill in the missing information while preserving the physiologically realistic matchings. Based on the practical implementation of LDDMM for images of [BMTY05b] and summarized Alg. 1, we have proposed to use smoothing kernels constructed as the sum of several Gaussian kernels [IJ-6]. These kernels denoted by MK, are the weighted sum of N Gaussian kernels K_{σ_n} , each of them being parametrized by its standard deviation σ_n :

$$\begin{aligned} MK(x) &= \sum_{n=1}^N a_n K_{\sigma_n}(x) \\ &= \sum_{n=1}^N a_n (2\pi)^{-3/2} |\Sigma_n|^{-1/2} \exp\left(-\frac{1}{2} x^T \Sigma_n^{-1} x\right), \end{aligned} \quad (2.22)$$

where Σ_n and a_n are respectively the covariance matrix and the weight of the n^{th} Gaussian function. Each Σ_n is only defined by a characteristic scale σ_n : $\Sigma_n = \sigma_n Id_{\mathbb{R}^d}$. Once this kernel is defined, the registration algorithm is the same as in Alg. 1.

A tricky aspect of these kernel construction for practical applications is however the tuning of their weights a_n . Although the choice of the σ_n has a rather intuitive influence on the optimal deformations, the tuning of the a_n strongly depends on the representation and the spatial organization of the registered shapes at the scales σ_n , $n \in [1, N]$. As described in [IJ-6] it depends on: (1) *Representation and spatial organization of the structures*: A same shape can be encoded in various ways. For instance, it can be a binary or a gray levels image. This representation has first a non-linear influence on the similarity metric (the sum of squared difference in LDDMM) *forces* (unsmoothed gradients) as shown row 12 of Alg. 1. The choice of optimal parameters a_n is even more complicated to do as the spatial relation between the shape structures should also be taken into account when smoothing the forces (row 13 of Alg. 1). (2) *Prior knowledge*: Prior knowledge about the amplitude of the structures displacement at each scale σ_n may be incorporated in a_n .

In [IC-18] we have then proposed to semi-automatically tune the a_n as follows:

$$a_n = a'_n / g(K_{\sigma_n}, I_S, I_T),$$

where $g(K_{\sigma_n}, I_S, I_T)$ represents the typical amplitude of the forces when registering I_S to I_T at a scale σ_n . This amplitude is related to (1) and cannot therefore be computed analytically. An empirical technique to tune it is the following: for each K_{σ_n} , the value of $g(K_{\sigma_n}, I_S, I_T)$ can be estimated by observing the maximum update of the velocity field v in a pre-iteration of registration of I_S on I_T using only the kernel K_{σ_n} with $a_n = 1$. The apparent weights a'_n , $n \in [1, N]$ provide an intuitive control of the amplitude of the displacements and are related to (2). To deform the largest features of I_S and I_T with a similar amplitude at each scale σ_n , the user should tune all the apparent weights a'_n with the same value. Typical results we obtained in [IJ-6] on the example of Fig. 2.6 are shown Fig. 2.7.

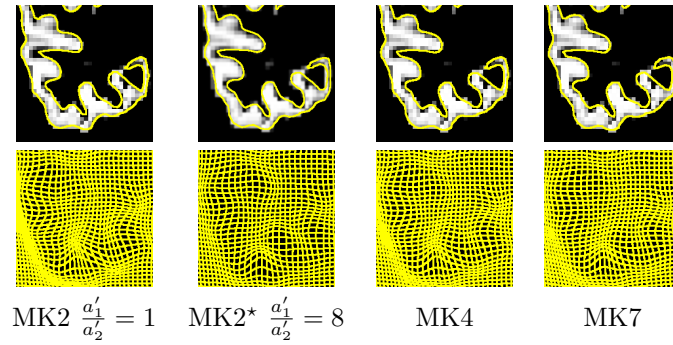


Figure 2.7: Registration results obtained on the example of Fig. 2.6 using multi-scale kernels. MKN stands for the sum of N kernels. Here $MK4$ and $MK7$ were automatically designed with apparent weights a'_i having the same value. Illustration out of [IJ-6].

The results of Fig. 2.7 make clear that multi-scale kernels with automatically tuned a_n following our method perform well in this example. More plausible deformations are indeed obtained since the correlation of the motions of the points is higher. Another phenomenon observed in practice is that a better quality of matching is obtained with a sum of kernels than with a single kernel of small width. Although we have no quantitative argument in this direction, we strongly believe that this is due to the convergence of the gradient descent algorithm to local minima. In standard image registration, coarse to fine techniques [LK81b] are ubiquitous. They consist in first registering two images with a strong regularization level and then iteratively decreasing the regularization level when the algorithm has converged at the current scale. At each considered scale, gradient descent based registration is then likely to be performed in a stable orbit *w.r.t.* the compared shapes scale. In LDDMM, using the sum of kernels at different scales instead of small scales only may then have a similar effect from an optimization point of view.

Influence of each scale

It is interesting to remark that the influence of each sub-kernel of the multi-scale kernels we defined can be measured. This property is particularly interesting for further statistical analyses. A first attempt to characterize this influence has

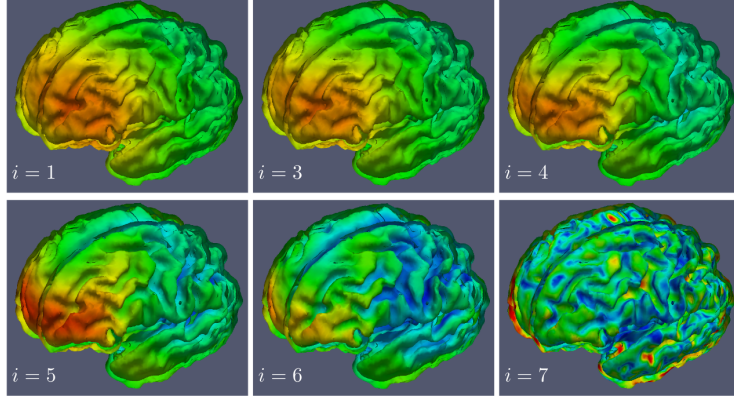


Figure 2.8: Representation of scale-dependent deformations φ_i out of a deformation φ obtained between two brain images using the method of [IJ-10]. The colors represent the amplitude of the scale-dependent deformations at the brain surface. Illustration out of [IJ-10].

been presented in [IJ-6] and was strongly developed in [IJ-10]. In [IJ-10], the registration of I_S on I_T is performed by minimizing an energy $\mathcal{E}_n$ with respect to the n -tuple $(v_1, \dots, v_n)$ where each time-dependent velocity field v_i is associated to scale-dependent deformations. The energy $\mathcal{E}(v)$ of Eq. (2.3) then becomes:

$$\mathcal{E}_n(v_1, \dots, v_n) = \frac{1}{2} \sum_{i=1}^n \int_0^1 \|v_i(t)\|_{H_i}^2 dt + \|I_S \circ \varphi^{-1} - I_T\|_{L^2}^2, \quad (2.23)$$

where the space H_i corresponds to kernel K_{σ_i} , the whole diffeomorphism $\varphi(t)$ is equal to $\varphi_1(t) \circ \dots \circ \varphi_n(t)$, and $\varphi_i(t)$ is defined by

$$\partial_t \varphi_k(t) = \left(v_k(t) + (\text{Id} - \text{Ad}_{\varphi_k(t)}) \sum_{i=k+1}^n v_i(t) \right) \circ \varphi_k(t). \quad (2.24)$$

Here $\text{Ad}_{\varphi} v$ also denotes the adjoint action of the group of diffeomorphisms on the Lie algebra of vector fields:

$$\text{Ad}_{\varphi} v(x) = (D\varphi.v) \circ \varphi^{-1}(x) = D_{\varphi^{-1}(x)} \varphi.v(\varphi^{-1}(x)). \quad (2.25)$$

These equations then allow to quantify scale-dependent deformations φ_i in the whole deformation φ . Results and algorithmic description of the solution for 3D images were given [IJ-10]. An illustration of this paper, where the deformations between two brain images were split into 7 scales is given Fig. 2.8. Note that [SLNP12] built on these ideas to incorporate sparsity priors on the scales. The space of kernels was also extended in [TQ18], with wavelet-based multi-scale kernels based. The project of [IJ-6, IJ-10] was a first step in such developments.

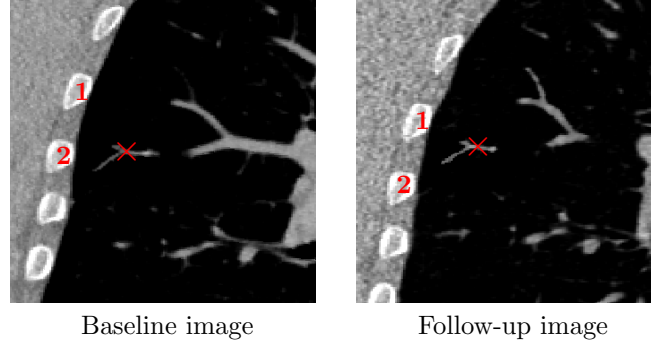


Figure 2.9: Illustration of the sliding motion at the lung boundary in the coronal view of two CT volumes acquired on the same subject. The motion of the vessel designated by the red cross, and the ribs (1) and (2) clearly demonstrate the sliding motion at the lung boundary. Images out of the EMPIRE10 challenge [MVGR⁺11] and illustration out of [IJ-11].

2.3.3 Diffeomorphic image registration with sliding conditions

Introduction

We now focus on how to model sliding constraints in the LDDMM formalism. Such constraints are observed *e.g.* at the lung boundaries as emphasized in Fig. 2.9. In [IJ-11], we have developed a smoothing strategy to solve this problem by using Alg. 1 (of [BMTY05b]), with specific smoothing properties. The central idea was to predefine different regions of interest Ω^k in the domain Ω of the registered images at the boundary of which discontinuous deformations are potentially estimated. Note first that these region of interest are fixed, so the source image I_S and the target image I_T must be aligned at the boundaries of the regions Ω^k . This is done by pre-registering the images with a very large amount of smoothing. This domain decomposition is illustrated Fig. 2.10.

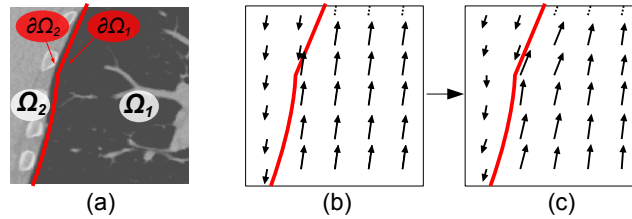


Figure 2.10: (a) Subdivision of the registration domain Ω into Ω^1 (inside the lung) and Ω^2 . Subdomain boundaries are represented by $\partial\Omega^1$ and $\partial\Omega^2$. (b) Velocity fields $\mathbf{v}$ which can be obtained in Ω after independent smoothing in Ω^1 and in Ω^2 , and (c) after enforcing sliding conditions in the neighborhood of $\partial\Omega^1$ and $\partial\Omega^2$. Illustration out of [IJ-11].

Methodology

Instead of considering a Reproducing Kernel Hilbert Space (RKHS) V embedded in $C^1(\Omega, \mathbb{R}^n)$ or $W^{1,\infty}$ as in the previous section, we used here N RKHS of vector fields $V^k \in C^1(\Omega^k, [0, 1])$ which can capture sliding motion, *i.e.* with an orthogonal component to the boundary vanishes at any point of $\partial\Omega^k$. The set of admissible vector fields is therefore defined by $V := \bigoplus_{k=1}^N V^k$, the direct sum of the Hilbert spaces $(V^k)_{k \in [1, N]}$. In particular, the norm on V of a vector field $\mathbf{v}_t$ is given by

$$\|\mathbf{v}_t\|_V^2 = \sum_{k=1}^N \|\mathbf{v}_t^k\|_{V^k}^2, \quad (2.26)$$

where $\mathbf{v}_t^k$ is the restriction of $\mathbf{v}_t$ to Ω^k . The flow of any $\mathbf{v} \in L^2([0, 1], V)$ is then well defined although the resulting deformations are piecewise-diffeomorphic and not diffeomorphic. As a consequence, the deformation is a diffeomorphism on each subdomain and allows for sliding motion along the boundaries.

Now that an admissible RKHS is defined, let us focus on the strategy we used to mimic the Gaussian smoothing of row 13 in Alg. 1 with the desired properties. In order to prevent from information exchange between the region Ω^k , the updates were diffused with Neumann boundary conditions at the boundaries of Ω^k . Independent Gaussian based convolution in each region Ω^k , would have been a quicker alternative in terms of computations but would not take into account the intrinsic region geometry. Then, in order to make sure that the orthogonal component to the boundary vanishes at any point of $\partial\Omega^k$, we use a projection strategy of the updates before and after smoothing so that they respect this constraint.

To do so, we considered the vector field $\mathbf{T}$ so that for each point $x \in \Omega$, $x + \mathbf{T}(x)$ is the nearest boundary between two subdomains in a limited neighborhood around the boundaries $\partial\Omega^k$. For the registration of pulmonary images, we empirically used a neighborhood of about $\gamma = 20$ millimeters. Consider a velocity field $\mathbf{w}$ defined on Ω . We used $\mathbf{T}$ to enforce the sliding conditions around $\partial\Omega^k$ by reducing the contributions of $\mathbf{w}(x)$ in the direction of $\mathbf{T}(x)$, when $\|\mathbf{T}(x)\|_{L^2} < \gamma$:

$$\mathbf{w}(x) = \mathbf{w}(x) - \alpha(x)\mathbf{T}(x) \frac{\langle \mathbf{w}(x), \mathbf{T}(x) \rangle_{L^2}}{\|\mathbf{T}(x)\|_{L^2}^2}, \quad (2.27)$$

where the weight $\alpha(x)$ equals $(\gamma - \|\mathbf{T}(x)\|)^2/\gamma$. For numerical stability, $\mathbf{w}(x)$ was set to 0 if $\|\mathbf{T}(x)\|_{L^2}^2 = 0$. The registration algorithm is then the same as Alg. 1 except row 13, where $\mathbf{u}$ is first projected using Eq. (2.27), then smoothed using the heat (diffusion) equation, and then projected again using Eq. (2.27).

Results

Results shown in [IJ-11] made clear the impact of this strategy compared with standard smoothing kernels. Fig. 2.11 shows the impact of such a piecewise diffeomorphic kernel when registering lung image where a sliding motion is clearly required at the lung boundaries. Note that to make this strategy tractable on large medical images (as in Fig. 2.11), we also coded it in the LogDemons formalism of [VPPA08]. The computational burden would have been too high in the LDDMM framework for such large 3D images. Both methods however led to similar results on smaller images.

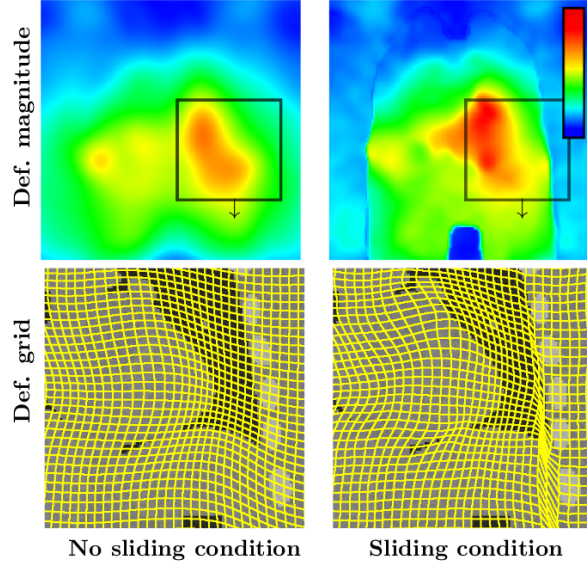


Figure 2.11: Deformation magnitude and deformed grids obtained when registering I_1 to I_5 using LogDemons using sliding motion modeling (S-LogD) or not (LogD MR). Colour bar is from 0 to 5 cm. Illustration out of [IJ-11].

2.3.4 Learning optimal regularization metrics

As mentioned Subsection 2.2.4, our work on spatially varying metrics opened the question of how to learn optimal metrics based on a set of reference images representing the same organ (shape) in different patients and a template (average) image. Incorporating mechanical or biological constraints to reproduce realistic results is indeed what we have described so far in this chapter. It is however also natural to learn the metric parameters using data driven methods if no mechanical model is well-established for the data of interest. This subsection then briefly presents the work of [IC-32], where an answer was given.

Building on [IC-31,B-2], we designed a set of kernels expressing spatially-varying metrics. We used symmetric positive definite matrices M as a parametrization of this set of kernels. In order to ensure the smoothness of the deformations, any kernel of this set has to satisfy the constraint that the Hilbert space of vector fields is embedded in the Banach space of C^1 vector fields. To enforce this constraint, we proposed the following parametrization,

$$\mathcal{K} = \{\hat{K}M\hat{K} \mid M \text{ SDP operator on } L^2(\mathbb{R}^d, \mathbb{R}^d)\}, \quad (2.28)$$

where $\hat{K}$ is a spatially-homogeneous smoothing kernel (typically Gaussian). The variational model then consisted in minimizing the functional:

$$\mathcal{F}(M) = \frac{\beta}{2} d_{S^{++}}^2(M, Id) + \frac{1}{N} \sum_{n=1}^N \min_v \mathcal{E}_{I_n}(v, M), \quad (2.29)$$

where β is a positive weight. The first term regularizes the kernel parameters, so that the minimization problem is well posed. Here, it favors parametrizations

of M close to the identity matrix but other a priori correlation matrix could be used. The term $d_{S++}^2(Id, M)$ can be chosen as the squared distance on the space of positive definite matrices given by $\|\log(M)\|^2$. Here again, other choices of regularizations could have been used such as the log-determinant divergence. This model has been implemented in [IC-32] where a simple method of dimension reduction was used since the matrix M is of size n^2 where n is the number of voxels and it gave promising results on the 40 subjects of the LONI Probabilistic Brain Atlas (LPBA40). An illustration of the matrix M obtained this paper is given Fig. 2.12. The pertinence of the learned metrics was further evaluated by comparing the mappings obtained using this metric with other mappings obtained using reference algorithms on 3D brain images. The algorithm was shown to lead to particularly competitive mappings in our tests, both in terms of accuracy and deformation smoothness. We finally mentioned that an exciting perspective of this work would also be to statistically analyze the obtained spatially-varying metric parameters.

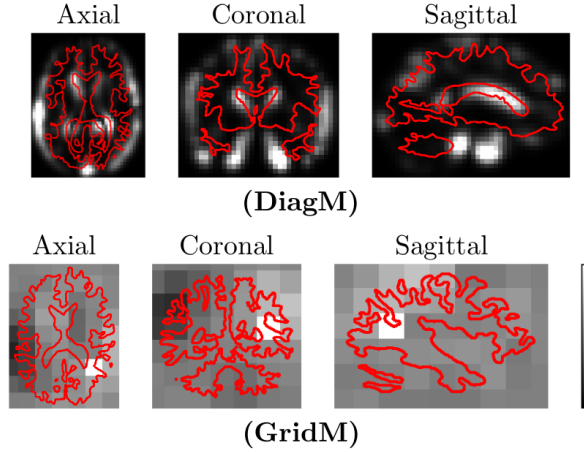


Figure 2.12: Values out of M learned on 40 subjects of the LONI Probabilistic Brain Atlas (LPBA40). The values are represented at their corresponding location in the template image T . **(DiagM)**: Values $M(j, j)$ for $j \in [1, \dots, N]$. Color bar ranges from 1 (black) to 1.04 (white). **(GridM)**: $M(i, j)$ for a fixed i and $j \in [1, \dots, L]$. White point corresponds to $i = j$ and has an intensity of 1.03. Color bar ranges from -0.05 (black) to 0.05 (white) for other points. Red curves represent the boundary between white and gray matter in T . Illustration out of [IC-32].

2.4 Similarity metrics in medical image registration

2.4.1 Summary of contributions

In all LDDMM or LogDemons methods presented Sections 2.2 and 2.3, the similarity metric is the sum of squared differences. It is then adapted to images

in which a given structure has the same intensity in all registered images (see Section 2.1.2). This section then deals with other similarity metric adapted to multimodal images or noisy images with potentially strong artifacts.

Summarized papers My main project in the use of similarity metrics adapted to multimodal images is the one of [IC-27,IC-23], where I defined a methodology to approximate local gradients of mutual information. This methodology is summarized Subsection 2.4.2.

Other paper Another project related to multi-modal image registration in which was involved, was motivated by the robust tracking of liver tumors in 2D Ultrasound image series [IJ-13]. These 2D images were acquired in a 3D domain and the tracked liver tumors moved because of the breathing motion. A whole diffeomorphic image registration pipeline was defined to follow the tumors. This PDE-based deformation model is inspired from the LogDemons framework of [VPPA08]. The main paper contribution is the integration of new matching forces, based on regional image properties, that allow to robustly follow the tumor in the 2D US images. My main personal contribution in [IJ-13] was mostly to drive the developments and to help writing the paper. My methodological contribution was the definition of the PDE-based deformation model but is minor compared with other works. This paper is then only given in Appendix A.8.

2.4.2 Local estimation of mutual information gradients

Motivation

The project of [IC-27,IC-23] was motivated by the need for multimodal images registration strategies with flexible deformation regularization properties. Different modalities capture different information about the imaged organs, which motivates their use from a diagnostic perspective. In this context, establishing automatically an accurate mapping between two multi-modal images gives access to clinicians the corresponding points between multimodal images representing the same organ. As explained, Subsection 2.1.2, mutual information is the most popular similarity metric to register multimodal images. Deformations regularization is also critical there as there not necessarily exists a mapping between all observed structures. This is illustrated Fig. 2.13 on a thoracic cage acquired using CT and MR imaging. Standard multimodal image registration algorithms therefore use rigid deformations or a very large amount of smoothing to tackle this issue. As an extension of [IJ-11], presented Subsection 2.3.3, we believed that spatially varying properties may be used in this context as a prior to solve more advanced registration models, as shown again Fig. 2.13 where a discontinuity can be seen at the thoracic cage boundary. Existing registration methods with adapted regularization properties are based on local gradients of the similarity metric. This motivated the development of a strategy for the fast and local estimation of mutual information gradients in 3D image registration.

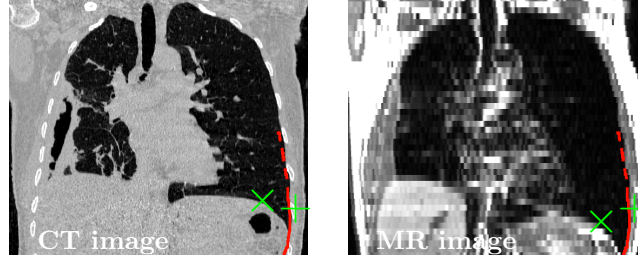


Figure 2.13: Two registered 3D CT/MR images. Sliding conditions may occur at the thoracic cage boundary (*e.g.* red curve). The crosses + and $\times$ respectively show a rib and a point at the boundary between the lungs and the diaphragm. Although + nearly remains at the same location, $\times$ clearly moves down with the rest of the boundary. The sliding motion is obvious here. Illustration out of [IC-27].

Methodology

Mutual information Let ω_S and ω_T be two Parzen windows that are related to the intensities of $I_S \circ \phi_1^Y$ and I_T . As in [TU00], we build the Parzen windows using cubic B-splines so that they have unit integral and respect the partition unity. The discrete sets of intensities associated with $I_S \circ \phi_1^Y$ and I_T are L_S and L_T , respectively. The joint Parzen discrete probability between $I_S \circ \phi_1^Y$ and I_T is then:

$$p(i, j; \mathbf{v}) = \frac{\alpha(\mathbf{v})}{\epsilon_S \epsilon_T} \sum_{\mathbf{x} \in \Omega} \omega_S \left(\frac{i - I_S \circ \phi_1^Y(\mathbf{x})}{\epsilon_S} \right) \omega_T \left(\frac{j - I_T(\mathbf{x})}{\epsilon_T} \right), \quad (2.30)$$

where $i \in L_S$, $j \in L_T$, ϵ_T and ϵ_S are scaling factors controlling the size of the Parzen windows, and $\alpha(\mathbf{v})$ is the normalizing constant of the probabilities. The marginal discrete probabilities are then $p_S(i; \mathbf{v}) = \sum_{j \in L_T} p(i, j; \mathbf{v})$ and $p_T(j; \mathbf{v}) = \sum_{i \in L_S} p(i, j; \mathbf{v})$. Since I_T is the fixed image, the probabilities p_T do not depend on $\mathbf{v}$. We then denote $p_T(j; \mathbf{v}) = p_T(j)$. The mutual information between $I_S \circ \phi_1^Y$ and I_T is then:

$$S(\mathbf{v}) = - \sum_{i \in L_S} \sum_{j \in L_T} p(i, j; \mathbf{v}) \log_2 \left(\frac{p(i, j; \mathbf{v})}{p_S(i; \mathbf{v}) p_T(j)} \right) \quad (2.31)$$

Mutual information gradients As shown in [TU00], when only I_S is deformed to match I_T , the derivative of the mutual information S with respect to a deformation parameter μ can be written as:

$$\frac{\partial S}{\partial \mu} = - \sum_{i \in L_S} \sum_{j \in L_T} \frac{\partial p(i, j; \mathbf{v})}{\partial \mu} \log_2 \left(\frac{p(i, j; \mathbf{v})}{p_S(i; \mathbf{v})} \right). \quad (2.32)$$

The authors of [TU00] then developed this equation in a parametric registration context. It is however interesting to note that this equation is sufficiently general

to model μ as any deformation of ϕ_1^Y . We then computed the derivative of $p(i, j; \mathbf{v})$ according to μ :

$$\begin{aligned} \frac{\partial p(i, j; \mathbf{v})}{\partial \mu} &= a_1 \sum_{\mathbf{x} \in \Omega} \frac{\partial \omega_S \left(\frac{i - I_S \circ \phi_1^Y(\mathbf{x})}{\epsilon_S} \right)}{\partial \mu} \omega_T \left(\frac{j - I_T(\mathbf{x})}{\epsilon_T} \right) \\ &= \frac{-a_1}{\epsilon_S} \sum_{\mathbf{x} \in \Omega} \frac{\partial I_S \circ \phi_1^Y(\mathbf{x})}{\partial \mu} \frac{\partial \omega_S(\xi)}{\partial \xi} \Big|_{\xi = \frac{i - I_S \circ \phi_1^Y(\mathbf{x})}{\epsilon_S}} \omega_T \left(\frac{j - I_T(\mathbf{x})}{\epsilon_T} \right) \end{aligned}$$

where a_1 is constant for all intensity levels i and j . Eq. (2.32) can then be written as:

$$\begin{aligned} \frac{\partial S}{\partial \mu} &\sim \\ \sum_{\mathbf{x} \in \Omega} \sum_{i \in L_S} \sum_{j \in L_T} \frac{\partial I_S \circ \phi_1^Y(\mathbf{x})}{\partial \mu} \frac{\partial \omega_S(\xi)}{\partial \xi} \Big|_{\xi = \frac{i - I_S \circ \phi_1^Y(\mathbf{x})}{\epsilon_S}} \omega_T \left(\frac{j - I_T(\mathbf{x})}{\epsilon_T} \right) \log_2 \left(\frac{p(i, j; \mathbf{v})}{p_S(i; \mathbf{v})} \right) \end{aligned} \quad (2.33)$$

The third and fourth terms of Eq. (2.33) can be straightforwardly computed from the current deformation. The second one can also be analytically computed since the ω_S is constructed using B-splines. The first term however depends on the type of deformation that μ represents. Importantly, the derivative $\partial S / \partial \mu$ is computed using a triple-sum on Ω , L_S and L_T , which is critical in terms of computational burden in the general case. We then describe hereafter how we locally estimated the derivative of Eq. (2.33) in the LogDemons framework [VPPA08] at a low algorithmic cost.

Approximated mutual information gradients As introduced in Subsection 2.1.4, the update field $\delta \mathbf{v}$ is computed in the LogDemons framework without considering the regularity of the transformation. When estimating $\delta \mathbf{v}$ in the point $\mathbf{x} \in \Omega$ using Eq. (2.33), we consider the parameter $\mu_{\mathbf{x}}$ as a local translation of $\mathbf{x}$ and do not perform any image or histogram smoothing. In addition, the local updates are constructed in the direction of the intensity gradients with an amplitude depending almost exclusively on the mutual information. To do so, we denote $\mathbf{b}(\mathbf{x}) = \frac{\nabla I_S \circ \phi_1^Y(\mathbf{x})}{|\nabla I_S \circ \phi_1^Y(\mathbf{x})|}$ as the normalized intensity gradient and consider the points $\mathbf{x}_n = \mathbf{x} + n\delta \mathbf{b}$, $n \in \{-1, 0, 1\}$, where δ is the spatial distance between the points $\mathbf{x}_n$. We also reduce the number of bins i and j considered in Eq. (2.33) by exploiting the fact that the cubic B-splines in ω_S and ω_T are non-null on a compact domain only. To do so, we first denote γ this domain's extent. We then denote $L_{S, \mathbf{x}}$ and $L_{T, \mathbf{x}}$, the subsets of L_S and L_T representing bins with less intensity difference than γ with $I_S \circ \phi_1^Y(\mathbf{x}_n)$ and $I_T(\mathbf{x})$ respectively. We then estimate the contribution of the points $\mathbf{x}_n$ to the mutual information S using:

$$\begin{aligned} S(\mathbf{x}) &= \sum_{\mathbf{x}_n} \sum_{i \in L_{S, \mathbf{x}_n}} \sum_{j \in L_{T, \mathbf{x}_n}} I_S \circ \phi_1^Y(\mathbf{x}_n) \frac{\partial \omega_S(\xi)}{\partial \xi} \Big|_{\xi = \frac{i - I_S \circ \phi_1^Y(\mathbf{x}_n)}{\epsilon_S}} \\ &\quad \omega_T \left(\frac{j - I_T(\mathbf{x}_n)}{\epsilon_T} \right) \log_2 \left(\frac{p(i, j; \mu)}{p_S(i; \mu)} \right) \end{aligned} \quad (2.34)$$

and the contribution $S^+(\mathbf{x})$ and $S^-(\mathbf{x})$, if the points $\mathbf{x}_n$ are translated by $\delta \mathbf{b}$, and by $-\delta \mathbf{b}$. The corresponding contributions are obtained by replacing $\mathbf{x}_n$

by $\mathbf{x}_n - \delta \mathbf{b}$ and $\mathbf{x}_n + \delta \mathbf{b}$ in the first term of Eq. (2.34). In the present context, the derivative Eq. (2.33) is then estimated in the direction where the variation of mutual information is the highest:

$$\frac{\partial S}{\partial \mu_{\mathbf{x}}}(\mathbf{x}) \sim \min \left(\frac{S^+(\mathbf{x}) - S(\mathbf{x})}{\delta}, \frac{S^-(\mathbf{x}) - S(\mathbf{x})}{\delta} \right) \quad (2.35)$$

Note that the minimum value is considered because the values of S , S^+ and S^- are negative. Using this strategy, only five trilinear interpolations of the gray levels in $I_S \circ \phi_1^{\mathbf{v}}$ are required to estimate $\partial S(\mathbf{x})/\partial \mu_{\mathbf{x}}$.

Results

In [IC-27], these approximated mutual information gradients were incorporated in the LogDemons framework of [VPPA08] in addition to the sliding motion strategy of [IJ-11]. Results of [IC-27] were in the same vein as those of [IJ-11] but for multimodal images. Compared with the standard Free-Form Deformation (FFD) algorithm of [RSH⁺99] for non-rigid multimodal image registration, a better trade-off between organs matching accuracy and deformation smoothness (outside of the discontinuity) was also obtained. Computational times were also reduced compared with FFD for 3D images of $300 \times 300 \times 70$ voxels (about 40 minutes for [IC-27] and several hours for FFD).

2.5 Image segmentation models

2.5.1 Summary of contributions

In this section, I present my research activity in medical image segmentation. It was secondary compared with the one in image registration but one paper worth being presented here in my opinion.

Summarized paper The segmentation strategy of [IC-37] is first presented subsection 2.5.2. The manuscript is also given in Appendix A.14. This work was motivated by the segmentation of multiple structures such as lymph nodes in whole-body MR images of patients with tumors. This task can be hardly automatized for two main reasons: (1) Structures boundaries are not visible everywhere due to very similar gradients in neighbor structures, and (2) the patients and the segmented structures have large anatomical variability. User interventions are then necessary but should be as limited as possible and with particularly responsive algorithms. In [IC-37], we then proposed a computationally efficient regularization strategy for the Fast Marching (FM) segmentation of multiple organs. This regularization stabilizes the semi-automatic registration of complex structures and has a low computational impact.

Other papers Other contributions which are not described here are first those made in collaboration with D.P. Zhang (former PhD student at Imperial College London) about the coronary artery motion modeling from 3D cardiac CT sequences using template matching [IC-19, IC-17, IC-16]. In addition to lead the methodological developments of D.P. Zhang, my methodological contribution

in these papers was mainly to incorporate existing diffeomorphic registration algorithms in a robust template matching context.

The paper [SJ-2] also presents a Python plugin for 3Dslicer⁴, which I developed to segment MR images representing the marmoset monkey brain. The segmentation was again based on a template propagation strategy and the *Insight Segmentation and Registration Toolkit* (ITK)⁵. This contribution is however mainly technical and therefore not developed here.

The work of [IC-34] extended [IJ-2] with a gap filling strategy that combines both skeleton- and intensity-based information to fill large discontinuities. It was motivated by the segmentation of vascular networks, where discontinuities are obtained in the segmented vessels using standard segmentation techniques, due to complex morphology and weak signals. This work is only given in Appendix A.12 as it extends a methodology developed during my PhD work.

Finally, [IJ-16] presents the ITK implementation of an efficient anisotropic non-linear diffusion technique for 2D or 3D images. This technique is based on the adaptive scheme of [FM14] making the diffusion stable and requiring limited numerical resources. Anisotropic Non-Linear Diffusion is a powerful image processing technique, which allows to simultaneously remove the noise and enhance sharp features in two or three dimensional images. Note that the anisotropy was not considered in the Perona and Malik sense [PM90] which is common in image processing but rather Weickert sense [Wei96], where the orientation is truly taken into account. In a sense, this method can be considered as a segmentation technique. The contributions of [IJ-16] are however too technical to be presented here.

2.5.2 Regularization model for the Fast Marching segmentation

Motivation

The work of [IC-37] was motivated by the quantitative analysis of Chronic Lymphocytic Leukemia (CLL), which is the most common B-cell malignancy and mostly affects elderly people. This requires the segmentation of more than ten organs in whole-body MR images. Due to high inter-patient variability and lack of intensity gradients at some organ boundaries, the only way to obtain a good segmentation is to ask a trained clinician to draw the organ shape according to his experience. This task is particularly time consuming (about 100 minutes). Our goal was then to develop a pertinent semi-interactive strategy [1-5] to reduce the time dedicated to this task. We then proposed a computationally efficient regularization strategy for the Fast Marching (FM) segmentation [Set96] of multiple organs.

Methodology

We consider L regions of interests (organs) in image I . The clinician iteratively places the seeds and then runs the Multi-label Fast-Marching strategy [Set96]. The label of each segmented region is related to one or several pre-defined seeds. The iterative propagation of the labels is also then performed using a standard

⁴<https://www.slicer.org/>

⁵<https://itk.org/>

was compared with their manual segmentation by experienced clinicians, which is the clinical routine, on MR images of $250 \times 250 \times 200$ voxels. The segmentation accuracy was similar using both methods and the time dedicated to the segmentation was reduced by a about a factor 3 using regularized Fast-Marching. A typical results is illustrated Fig. 2.14.

2.6 Outlook

There are several perspectives to the works presented in this section. Immediate research directions will be carried out in the context of the supervision of V. K. Ghorpade’s postdoctoral work on the registration of 3D and mutli-modal whole-body images. I am in particular interested in developing the polyaffine registration framework of [ACAP09, TMK11] to constrain the registration of CT-MR whole body images, in which the bones are locally rigid and other structures can be modeled with elastic constraints. In this context, we also work on an extension of the MIND model of [HJB⁺12b] that transforms the representation of multi-modal images, so that they can be registered using the sum-of-squared-difference similarity metric.

I also have a long term collaboration with A. Gossé (CR CEA Saclay), A. Quaini (CR CEA Saclay) and F. Gamboa (Pr Univ. Toulouse, IMT) in which I work on features extraction of 2D image sequences representing rotating and levitating balls which are extremely warmed-up. We started writing a paper explaining the results of this work. I also plan to write an applied communication explaining the pipeline I developed for these specific images.

I established first contacts with S. Chafik (Pr Univ. Clermont-Auvergne) to work on deformation models and shape analysis. We could work on an extension of [SJ-4], where we would estimate the variability of the sulcal pattern variability in human endocasts. I have indeed started a collaboration last year with E. De Jager (University of Pretoria, South Africa) and C. Fonta (CNRS, CerCo Toulouse) to study such structures. After applying existing techniques to analyse the data of E. De Jager, our work has lead to open questions dealing with how to address the specificity of human endocasts data to understand their anatomical variability.

I have also developed last year a collaboration with L. Keller (Regenerative Nanomedecine team, Strasbourg hospital) to study the vasculature of broken bones in micro-CT images. Until now, I used the techniques I developed during my PhD thesis [SJ-3]. We have written a proposal to extend the tools I developed, as the vessels segmentation could be clearly improved. In particular I would like to make it possible to segment these very large images by reducing the algorithmic cost of computationally inefficient, but robust, algorithms using stochastic techniques.

Finally, I would be interested to work on developing self-adaptive regularization models in medical image registration inspired by neural network models. There has indeed been a large interests on such models recently in the medical image registration community, *e.g.* [MWZJL16]. Constraining properly such models and then the underlying optimization algorithms, could lead to more accurate and physiologically realistic image mappings than using current methods.

Chapter 3

Numerical methods for stochastic modeling

3.1 Summary of contributions

This chapter presents my main scientific contributions related to numerical methods for stochastic modeling. My main contributions in this field were made in the context of two projects summarized in Sections 3.2 and 3.3.

Summarized papers The first project in I developed numerical methods for stochastic modeling was motivated by the analysis of the brain activity in 3D+time fMRI time series and is developed in Section 3.2. It was carried out in the context of my postdoctoral work at CEA Saclay with P. Ciuciu (DR CEA Saclay). Related papers in the appendices are [IJ-5] in Appendix B.2 and [IJ-4] in Appendix B.1. An original Bayesian model for the simultaneous detection of brain activations and the estimation of the their response in the image sequences was first developed. This model constraints the problem using physiologically realistic hypotheses, making it possible to detect local brain activations that are lost using more generic models. In this context, I was particularly involved in the development of a strategy to efficiently compute the partition function of irregular 3D Potts fields with respect to their inverse temperature β , which can be extremely demanding in terms of computations using standard methods. This contribution allowed to spatially regularize the detection of brain activation/deactivation/inactivation in fMRI time series using automatically-tuned and region-wise regularization levels.

The second project was about the barycenter estimation of graphs in which a probability measure on its nodes reflects the observation occurrences. For instance, the graph may represent a social network and the observations are topics discussed by network members (*e.g.* a hashtag in twitter). In this example, the notion of graph barycenter may allow to define a typical user interested by this topic. It is also a first step to define a notion of distance between two observation sets observed on the same graph. Specifically, the developed strategy estimates the Fréchet mean of such graphs and relies on a noisy simulated annealing algorithm. This project was carried out in collaboration with S. Ga-

dat (Pr Toulouse School Economics, IMT) and I. Gavra (former PhD student Univ. Toulouse 3, IMT) and is presented in Section 3.3. Related papers in the appendices are [IJ-19] in Appendix B.3 and [IJ-6] in Appendix B.4.

Other papers I also had secondary contributions in other projects related to numerical methods for stochastic modeling. In [IJ-18], I worked on the development of an algorithm dedicated to the automatic segmentation of the breast in magnetic resonance images. This algorithm was based on a statistical segmentation strategy regularized by using a hidden Markov model. My main contributions there were to advise S. Ribes (former PhD student, Univ. Toulouse 3, SIMAD) in the development of her pipeline and in presenting her methodology at the *IEEE Trans. Medical Imaging* format. Her PhD work was indeed not supervised by researchers in medical image analysis and I helped her to communicate in this community.

I have also recently started a collaboration with G. Fort (DR CNRS, IMT) about Maximum Likelihood inference algorithms in statistical models [IC-36,NC-4]. Although my goal in this collaboration is to develop original numerical methodologies for such strategies, my contributions were technical so far (efficient *C++* algorithm implementation). I also had a similar technical contribution in [SJ-1] which deals with atomic deconvolution *i.e.* with deconvolution in density estimation. Note finally that I also supervised the Master 2 project of A.L. Fouque (ENS Cachan), dealing with statistical clustering of fMRI image series, during my postdoctoral work at CEA Saclay. This work was published in [IC-10,NC-2] but was too preliminary to be developed here.

3.2 Numerical methods for the analysis of the brain activity

3.2.1 A general model for the analysis of fMRI time series BOLD fMRI

In medical imaging, the Blood-Oxygen-Level-Dependent (BOLD) signal locally captures temporal variations of oxyhemoglobin and deoxyhemoglobin relative levels. This signal is of particular importance in functional Magnetic Resonance Imaging (fMRI) applied to the detection of the brain activity, where it is established that each brain region is specialized to specific cognitive tasks. It can indeed be used to detect active brain regions, which have a higher oxyhemoglobin and deoxyhemoglobin (or more simply energy) consumption than at rest. For a given activation, it is fundamental to note that the energy consumption starts briefly after the neuronal activity and then follows a pattern that smoothly evolves during several seconds. This is due to local vascular properties which somehow induce an inertia between the energy requirement and its supply by the blood.

As quantified during my PhD thesis work [IJ-1,IJ-3,IJ-9], vascular network properties strongly vary across brain regions. These local patterns of energy consumption, called the hemodynamic response functions (HRF), therefore vary

in space in BOLD fMRI time-series. The methodological developments of this project were related to the analysis of the brain activity in BOLD fMRI by making such physiologically motivated hypotheses on the observed 3D+t signal to robustly capture fine activations.

The Joint Detection Estimation (JDE) framework

The framework in which I worked to analyze BOLD fMRI time-series was the Joint Detection-Estimation (JDE) framework introduced in [VCI07, MIV⁺08]. This approach relies on a prior parcellation of the brain into $\mathcal{P} = (\mathcal{P}_\gamma)_{\gamma=1:r}$ functionally homogeneous and connected parcels [TFP⁺06]. The shape of such parcels can be seen at the top-right of Fig. 3.4.

Every parcel $\mathcal{P}_\gamma$ comprising voxels $(V_j)_{j=1:J}$ is characterized by a single hemodynamic response function (HRF) $\mathbf{h}$. This HRF should be ideally estimated at each voxel but that would induce too much degrees of freedom to the model. Existing methods in 2009 also used a single HRF for the whole brain. Using a different HRF in each brain region then appears as a reasonable trade-off. Within a given $\mathcal{P}_\gamma$, stimulus-related fluctuations of the BOLD signal magnitude are however voxel-dependent and encoded by $\mathbf{a} = (a_j^m)_{j=1:J, m=1:M}$, the response levels where m stands for the stimulus type index.

The fMRI time course measured in voxel V_j then reads: $\mathbf{y}_j = \sum_{m=1}^M a_j^m \mathbf{x}^m \star \mathbf{h} + \mathbf{b}_j$, where $\mathbf{x}^m$ stands for the binary stimuli vector, whose non-zero entries encode the arrival times of the m^{th} stimulus type, and $\mathbf{b}_j$ stands for the noise component [VCI07, MIV⁺08]. Note that a drift term $\mathbf{Pl}$ is also considered in the model to capture low frequency variations due to breathing. We will however not further discuss this term here as its use is straightforward and refer to [IJ-4] for more details. These notations are illustrated in Fig. 3.1.

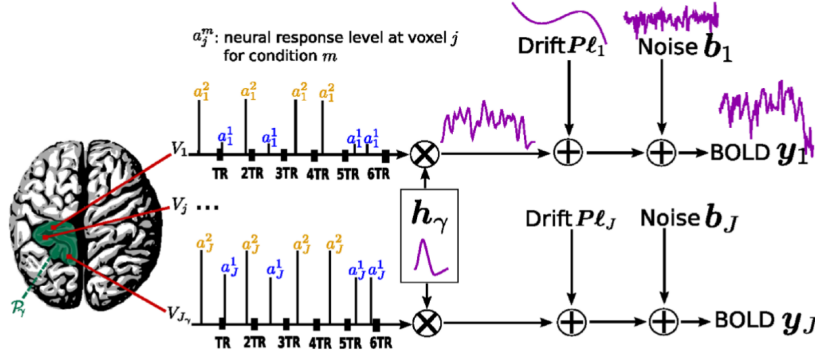


Figure 3.1: Parcel-based regional BOLD model of the Joint Detection-Estimation framework of [VCI07, MIV⁺08] and [IJ-4, IJ-5]. Illustration out of [IJ-4].

Prior probability density functions are introduced within a Bayesian framework on every $(\mathbf{a}, \mathbf{h})$. A Gaussian prior on $\mathbf{h}$ with a smooth constraint is also

used in addition to non-informative priors on each hyper-parameter [VCI07]. Spatial Gaussian mixture models are expressed on $\mathbf{a}$ through the introduction of hidden variables $\mathbf{q} = (\mathbf{q}_j^m)_{j=1:J}^{m=1:M}$ that encode whether voxel V_j is activated in response to stimulus m or not. In this case, q_j^m is in $\{1, 0\}$ but it may have more than two colors/labels if other states are modeled, as for instance in [IJ-5] where deactivations are also modeled.

The main contribution of [IJ-4], compared with previous work was to spatially smooth $\mathbf{q}$ with a regularization level β_m that is automatically tuned in each parcel and for each stimulus type m . The stochastic optimization problem is then independent from a parcel to the other. Stimulus-dependent hidden Ising fields were introduced on the states such that the global prior pdf reads:

$$p(\mathbf{a} | \Theta_{\mathbf{a}}) = \prod_m \sum_{\mathbf{q}^m} [(\prod_j f(a_j^m | q_j^m, \boldsymbol{\theta}_m)) \Pr(\mathbf{q}^m | \beta_m)]$$

and $f(a_j^m | q_j^m = i) \sim \mathcal{N}(\mu_{i,m}, v_{i,m})$. Parameters $\mu_{i,m}$ and $v_{i,m}$ define the prior mean and variance of class $i = 0, 1$, respectively for the stimulus type m . The set $\boldsymbol{\theta}_m$ comprises four prior mixture parameters $\boldsymbol{\theta}_m = \{\mu_{0,m}, \mu_{1,m}, v_{0,m}, v_{1,m}, \beta_m\}$ since non-activating voxels are modeled using a zero-mean Gaussian density ($\mu_{0,m} = 0$).

Samples of the full posterior pdf $p(\mathbf{h}, \mathbf{a}, \mathbf{q}, \boldsymbol{\Theta} | \mathbf{y})$ are simulated using a Gibbs sampler algorithm and posterior mean estimates are then computed from these samples. Inner Metropolis-Hastings (MH) steps are also use in case of inability to draw samples from the full conditional posterior probability. Here, we focus on the sampling of parameter β_m , which is achieved using a *symmetric* random walk Metropolis-Hastings (MH) step: At iteration k , a candidate $\beta_m^{(k+1/2)} \sim \mathcal{N}(\beta_m^{(k)}, \sigma_\epsilon^2)$ is generated. It is accepted (*i.e.*, $\beta_m^{(k+1)} = \beta_m^{(k+1/2)}$) with probability: $\alpha(\beta_m^{(k)} \rightarrow \beta_m^{(k+1/2)}) = \min(1, A_{k,k+1/2})$, where the acceptance ratio $A_{k,k+1/2}$ follows Eq. (3.1):

$$\begin{aligned} A_{k,k+1/2} &= \frac{p(\beta_m^{(k+1/2)} | \mathbf{q}_m^{(k)})}{p(\beta_m^{(k)} | \mathbf{q}_m^{(k)})} = \frac{p(\mathbf{q}_m^{(k)} | \beta_m^{(k+1/2)}) p(\beta_m^{(k+1/2)})}{p(\mathbf{q}_m^{(k)} | \beta_m^{(k)}) p(\beta_m^{(k)})} \\ &= \frac{Z(\beta_m^{(k)})}{Z(\beta_m^{(k+1/2)})} \exp\left((\beta_m^{(k+1/2)} - \beta_m^{(k)}) U(\mathbf{q}_m^{(k)})\right), \end{aligned}$$

using Bayes' rule and considering a uniform prior for β_m .

It is important to remark that this approach requires to estimate ratios of partition functions $Z(\cdot)$ for all $\mathcal{P}_\gamma$ parcels prior to exploring the full posterior pdf. This aspect of the algorithm is the one on which I had a major contribution. I therefore develop next section the solution I developed to quickly and robustly estimate parcel-related partition functions for Ising models in [IJ-4] and 3-class (colors) Potts models in [IJ-5]. Results obtained using the JDE model with and without automatically-tuned β are also shown in Subsection 3.2.3.

3.2.2 Estimation of 3D Ising and Potts field partition functions

Introduction

As developed above, the JDE framework was proposed in [MIV⁺08, TFP⁺06] and [IJ-4] as a generalization of regression methods for the analysis of fMRI data. It was initially developed to discriminate activating voxels from nonactivating ones. To this end, spatially adaptive 2-class mixture models and irregular 3D Ising fields embodied the spatial correlation over these hidden states in each predefined brain region. This method was extended in [IJ-5] to also account for putative deactivations that may appear for instance in pathologies (epilepsy). To this end, the 2-class mixture models were then extended to 3 classes and the Ising models became Potts models.

Both for 2-class Ising models and for 3-class Potts models, the regularization was made spatially adaptive in different brain regions. This required the estimation of numerous partition functions on irregular 3D domains. This section then explains the original partition function estimation strategy we developed.

Problem statement

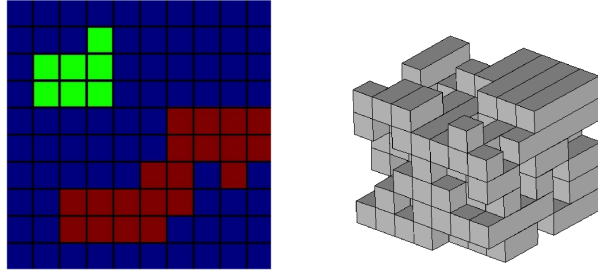


Figure 3.2: Instantiations of typical Potts fields which partition functions are estimated. **(left)** 3-colors Potts field in a 2D regular grid $\mathcal{T}$. In [IJ-5], the colors model the voxels states (**activated**, **non-activated**, and **de-activated**) **(right)** Irregular 3D grid $\mathcal{T}$ in which a partition function may be estimated. In [IJ-5] or [IJ-4], this grid represents a cortical region in which the brain activity is estimated.

Let us consider a grid $\mathcal{T}$ characterized by a set of sites (voxels) $\mathbf{s} = (s_i)_{i=1:n}$. Importantly here, the grid $\mathcal{T}$ is in a 3D connected domain and is not necessarily regular. In the driving motivation of this work it indeed represents a segmented region in a 3D image. A label $q_i \in \{0, \dots, L\}$ is associated to each site s_i , where L is the possible set of labels (or colors). Fig. 3.2 illustrates these notations. A pair of adjacent sites s_i and s_j ($i \neq j$) is denoted $i \sim j$ and is called a clique c , so the set of all cliques is an undirected graph denoted $\mathcal{G}$. Let $\mathbf{q} = (q_1, q_2, \dots, q_n) \in \{0, \dots, L\}^n$ be the set of labels associated to $\mathbf{s}$. In what follows, we assume $\mathbf{q}$ to be distributed according to a symmetric Potts model:

$$\Pr(\mathbf{q}|\beta) = Z(\beta)^{-1} \exp(\beta U(\mathbf{q})) , \quad (3.1)$$

where the global *negative energy* is:

$$U(\mathbf{q}) = \sum_{(i,j) \in \mathcal{G}} I(q_i == q_j), \quad (3.2)$$

and $I(A) = 1$ whenever A is true and 0 otherwise. The inverse temperature $\beta \geq 0$ controls the amount of spatial correlation between the components of $\mathbf{q}$ according to $\mathcal{G}$. The *partition function* $Z(\beta)$ reads

$$Z(\beta) = \sum_{\mathbf{q} \in \{0, \dots, L\}^n} \exp(\beta U(\mathbf{q})) \quad (3.3)$$

and depends on the geometry of $\mathcal{G}$ in addition to β . Its exact evaluation in a reasonable amount of time is impossible except on tiny grids as the number of combinations $\mathbf{q} \in \{0, \dots, L\}^n$ can be extremely large even for medium sized domains, and L equals 2 or 3. Most image segmentation models based on Potts or Ising models therefore work with a fixed β so the partition function is considered as an unknown constant variable. Robust and fast estimation of $Z(\beta)$ is however a key issue for numerous 3D medical imaging problems involving Potts models and more generally discrete MRFs when β is not fixed.

Partition function estimation

Path sampling Different strategies exist to estimate the partition function of Potts models. The most common one is the *path sampling* strategy [TIP08] in which $Z(\beta)$ is first straightforwardly computed for $\beta = 0$ and then iteratively computed for larger values of β using:

$$\begin{cases} Z(0) &= L^n \\ Z(\beta) &\approx Z(\beta_0) \frac{1}{M} \sum_{m=1}^M \frac{\exp(\beta U(\mathbf{q}_m))}{\exp(\beta_0 U(\mathbf{q}_m))} \end{cases} \quad (3.4)$$

Once a $Z(\beta_i)$ computed, $Z(\beta_{i+1})$ can then be estimated for β_{i+1} slightly higher than β_i . Remark that at each iteration, this algorithm also requires to instantiate M label sets $\mathbf{q}_m$ in the considered graph $\mathcal{G}$ at the current inverse temperature β_0 . This is made using the Swendsen-Wang algorithm [HBJ⁺97]. We have shown in [IJ-5] that this strategy performed particularly well for the domains on which we worked. It is however particularly time consuming when used in the joint detection estimation framework of [IJ-4, IJ-5] as it still requires the instantiation of numerous $\mathbf{q}_m$. This is why we developed our extrapolation strategy.

Alternative strategies Remark that alternative methods to the path sampling strategy exist to estimate partition functions. An important one is based on the mean field theory as in [FP03], which iteratively computes a fixed point equation. We have however shown in [IJ-5] that it was not accurate in our application. A far faster one is the Onsager's formula [Ons44] which gives the analytic expression of the partition functions for 2D square grids with toroidal boundary assumptions. This formula was not generic enough to be used in our project but allowed us to compare our approximations to ground truth results in relatively simple domains.

Fast and robust partition function estimation The technique proposed in [IJ-5] consists in (1) precomputing a set of partition functions on various reference grids $(\mathcal{G}_p)_{p=\{1,\dots,P\}}$ and then (2) quickly estimating the partition function of a test grid $\mathcal{T}$ by extrapolating a partition function $(\mathcal{G}_{ref})$ which is selected out of $(\mathcal{G}_p)_{p=\{1,\dots,P\}}$. As justified in [IJ-5], the extrapolation function is:

$$\log \tilde{Z}_{\mathcal{T}}(\beta, \mathcal{G}_{ref}) = \left(\frac{c_{\mathcal{T}}}{c_{\mathcal{G}_{ref}}} (\log \hat{Z}_{\mathcal{G}_{ref}}(\beta) - \log L) + \log L \right), \quad (3.5)$$

where $c_{\mathcal{T}}$ and $c_{\mathcal{G}_{ref}}$ are the number of cliques in $\mathcal{T}$ and $\mathcal{G}_{ref}$, respectively. The strategies to estimate the $(\mathcal{G}_p)_{p=\{1,\dots,P\}}$ and select $\mathcal{G}_{ref}$ are described hereafter.

The partition function of the P reference grids $(\mathcal{G}_p)_{p=\{1,\dots,P\}}$ are computed once for all using path sampling with a fine step on different β values and a large amount of simulated $\mathbf{q}_m$. This strategy is particularly time consuming but leads to accurate estimates of the partition functions. Note that the configurations of the reference grids $(\mathcal{G}_p)_{p=\{1,\dots,P\}}$ should be inhomogeneous and cover diverse situations that may occur in the applications.

The grid $(\mathcal{G}_{ref})$ selected out of $(\mathcal{G}_p)_{p=\{1,\dots,P\}}$ to approximate the partition function of $\mathcal{T}$ is chosen based on geometric properties. Let n_i be the number of neighbors for site s_i of $\mathcal{T}$. We defined $r_{\mathcal{T}} = \sigma_{\mathcal{T}}/\mu_{\mathcal{T}}$ as a measure of grid homogeneity where $\mu_{\mathcal{T}}$ and $\sigma_{\mathcal{T}}$ are the mean and standard deviation of n_i over $\mathcal{T}$, respectively. We then define a similarity index between two grids as:

$$\mathcal{L}_{\mathcal{T}}(\mathcal{G}_p) = (r_{\mathcal{T}} - r_{\mathcal{G}_p})^2 \quad (3.6)$$

A second similarity index we use is the approximation error criterion between the partition function of $\mathcal{T}$ and its approximation using the one of $\mathcal{G}_p$. This approximation error criterion depends on the number of cliques c , and sites n , in $\mathcal{T}$ and $\mathcal{G}_p$.

$$\mathcal{A}_{\mathcal{T}}(\mathcal{G}_p) = \left(\frac{(n_{\mathcal{T}} - 1) - c_{\mathcal{G}_p}(n_{\mathcal{G}_p} - 1)/c_{\mathcal{G}_p}}{n_{\mathcal{T}}} \right)^2 \quad (3.7)$$

Justification of $\mathcal{A}_{\mathcal{T}}(\mathcal{G}_p)$ is at the heart of [IJ-5]. Importantly, it holds for $\mathcal{G}_p$ sufficiently similar to $\mathcal{T}$, which we quantify using $\mathcal{L}_{\mathcal{T}}$. The reference grid is then finally selected as:

$$\mathcal{G}_{ref} = \arg \min_{(\mathcal{G}_p)_{p=\{1,\dots,P\}}} \mathcal{A}_{\mathcal{T}}(\mathcal{G}_p) \quad \text{subject to } \mathcal{L}_{\mathcal{T}}(\mathcal{G}_p) \leq \epsilon \quad (3.8)$$

3.2.3 Results and discussion

A brief overview of the main results obtained in [IJ-4,IJ-5] is now given. Fig. 3.3 first compares the above-mentioned partition function estimation techniques in a case where the ground truth is known, using the Osanger's formula. One can see that the path sampling and the proposed extrapolation technique are those that perform best. The extrapolation technique is however much faster. We then present in Fig. 3.4 representative results obtained by using the proposed strategy in the proceedings of [IC-15] which corresponds to [IJ-5]. Here the three classes represent activations, non-activations and de-activations. In the example of Fig. 3.4(top), it can be shows that finer activations are estimated

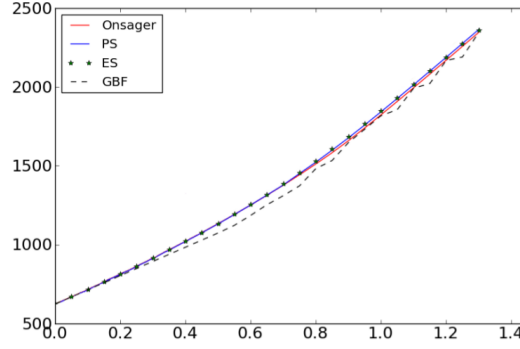


Figure 3.3: Partition function of a 2D 2-class Potts field defined over a 30×30 cyclic regular grid estimated using (Red curve) Onsager’s formula, (Blue curve) path sampling estimate, (Green stars) extrapolation technique from a reference set made up by 250 grids of various size and shape and (Black dashed curve) Mean field theory (denoted GBF) based approximation. Illustration out of [IJ-5].

than by using fixed β or no spatial regularization. In Fig. 3.4(top), the cortical areas that are known to be related to vision are those in which activations are found after a visual stimulation. Remark that in both cases, the areas in which no activations are found have large β values, *i.e.* a strong level of regularization while those with activations have optimal β values between 0.5 and 0.8. It worth finally mentioning that computational times with or without the estimation of β are almost the same. They required about 1 hour for each image series in 2010 and using Python.

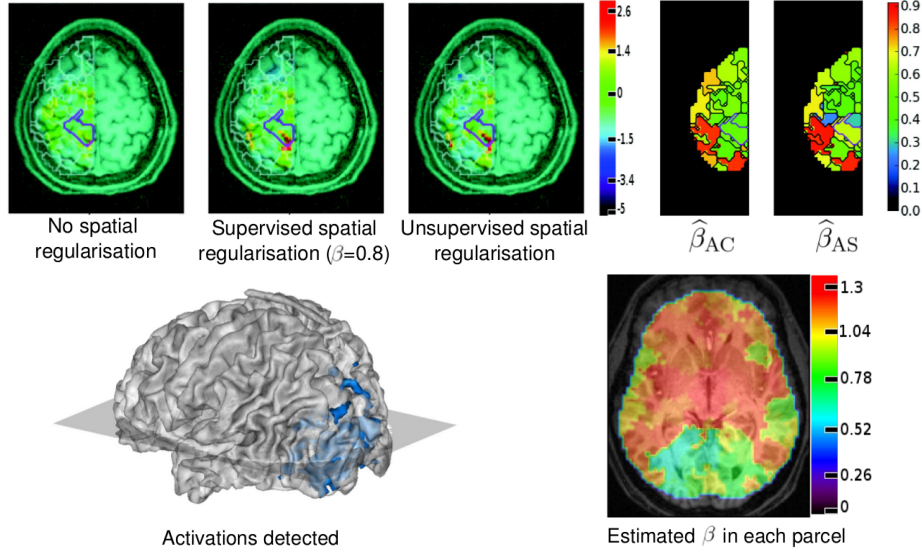


Figure 3.4: Application of the partition function estimation technique to the joint detection-estimation framework of [IJ-4]. **(top)** Estimated normalized contrast maps between an auditory computation task (AC) and sentence (AS) in a 2D slice out of a 3D domain. **(bottom)** Brain activation detected after a visual stimulus. Illustration out of [IC-15].

3.3 A stochastic framework for the online graph barycenter estimation

3.3.1 Motivation

Graph structures can model complex phenomena of high interest in a wide variety of domains and play an important role in various fields of data analysis. Although graphs have been used for quite a while in some fields, *e.g.* in sociology [Mor34], the recent explosion of available data and computational resources boosted the importance of studying such structures. Among the main application fields, one can count computer science (web understanding [PSV07]), biology (neural, protein, gene networks), social sciences (analysis of citations, social networks [HK10]), machine learning ([GA10]), statistical or quantum physics ([Est15]), marketing (consumers preference graphs) and computational linguistics ([NS06]).

Singling out the central node in a graph can be seen as a first step to understand a network structure. Different notions of *node centrality* have been introduced to measure the influence or the importance of nodes of interest in a network. Centrality notions are sometimes related to the mean distance from each node to all others [Bav50], to the degree of each node [Fre78] or even to the eigenvalues of the graph's adjacency matrix [Bon72]. A rather complete survey can be found in [Bor05]. These notions of centrality however can only take into

account predefined weights on the edges but not on the nodes, although there are numerous applications where this would be rather natural. For example, consider a graph representing a subway network, where its nodes are the subway stations. It seems quite reasonable to use the number of passengers getting in and getting out the subway network at the different station to establish the *central* station. In the case of a traffic network, the node-weight can model how many cars pass by a given intersection; in the case of a social network it can model the number of followers (or likes, or posts, *etc.*) of each individual.

To take this kind of information into account, we defined in [IJ-19] a method to estimate the barycenter of a graph with respect to a probability measure on the nodes set. In [SJ-6], this methodology was then made scalable to large graphs. It was also explicitly developed in an online context where the exact probability measure is unknown and only observations of this random distribution are required (*e.g.* observations of the subway stations at which passengers getting in or leaving the network). It may additionally be updated at the arrival of a new observation. Besides determining a central node, the knowledge of such a barycenter on a graph may be of multiple uses in future work. For example, the computation of the barycenter using two observation data sets on a single graph could be used to determine whether these sets are sampled using the same probability measure. The barycenter may also be useful in graph representation, since setting the graph barycenter in a central position provides an intuitive visualization.

3.3.2 Methodology

Main notations

Back in 1948, M. Fréchet presented a possible answer to define the mean of a probability measure on an Euclidean space [Fré48]. He introduced a notion of typical position of order p for a random variable Z defined on a general metric space $(\mathcal{E}, d)$ and distributed according to any probability measure ν . This is now known as the p -Fréchet mean, or simply the p -mean, and is defined as:

$$M_\nu^{(p)} := \arg \min_{x \in \mathcal{E}} \mathbb{E}_{Z \sim \nu} [d^p(x, Z)]. \quad (3.9)$$

For example, if Z is a random variable distributed according to a distribution ν on $\mathbb{R}^d$, its expected value, given by $m_\nu = \int_{\mathbb{R}^d} x d\nu(x)$ is also the point that minimizes:

$$x \mapsto \mathbb{E}_{Z \sim \nu} [|x - Z|^2]. \quad (3.10)$$

Now, let $G = (N, E)$ denote a finite weighted graph, where E is its edges set and ν is a probability measure on its nodes set N . The barycenter of a graph $G = (N, E)$ is then defined here as the following 2-Fréchet mean, that we simply denote Fréchet mean:

$$M_\nu = \operatorname{argmin}_{x \in N} \sum_{y \in N} d^2(x, y) \nu(y), \quad (3.11)$$

where $d(x, y)$ is the sum of the edge weights in the shortest path between nodes x and y , and $\nu(y)$ is the probability measure on node y . Importantly here, the lower an edge weight the closer the linked nodes. We then denote edge length

the weight given to an edge. One should finally remark that the Fréchet mean of a weighed graph is not necessarily unique. This is indeed due to the potentially complex structure of graph $G = (N, E)$.

As mentioned in Subsection 3.3.1, we place ourselves in an online estimation framework, in the sense that we suppose that the probability measure ν is unknown. A sequence $(Y_n)_{n \geq 0}$ of i.i.d. random variables distributed according to ν is instead available. For instance, in the subway networks example, an observation Y_n can be interpreted as the access of a passenger to a given station. Remark that if only a number of observations at each node is known, the Y_n can be randomly generated following a uniform law with more or less weight on the nodes depending on the number of observation. This makes it possible to use the strategies of this section by mimicking the observations.

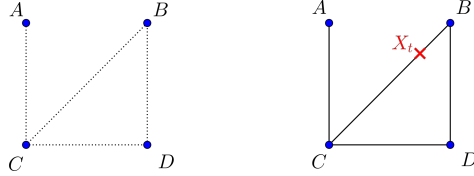


Figure 3.5: **(left)** Example of discrete graph G , and **(right)** corresponding continuous version Γ_G . X_t represents the current position of the algorithm in Γ_G . In this example its closest node in G is the node B . Illustration out of [SJ-6].

Our strategy to estimate the barycenter of $G = (N, E)$ finally runs on Γ_G , a continuous version of the discrete graph G . Given an edge $e = (u, v)$ of length L_e in $G = (N, E)$, the corresponding edge in Γ_G is an interval $[0, L_e]$ where its extremities are the linked nodes, as shown in Fig. 3.5. As a result, the estimated graph barycenter X_t can be on an edge and not only on a node.

Graph barycenter estimation

In [IJ-19], we proposed a method to estimate the barycenter of weighted graphs and established its convergence from a theoretical point of view. This algorithm is based on a simulated annealing algorithm with a random perturbation, as summarized in Alg. 2. Importantly, the random perturbation is modeled in order to escape potential local traps. Its impact is then decreased progressively in order to cool down the system and let the algorithm converge. This effect is parametrized by a continuous function $(\beta_t)_{t \geq 0}$, that represents the inverse of the so-called temperature schedule: when β_t is small, the system is *hot* and the random noise is quite important with respect to the gradient descent term. Then, when β_t goes to infinity, the random perturbation is negligible.

The convergence of the simulated annealing to the set of global minima, is guaranteed from a theoretical point of view for logarithmic evolutions of the temperature, *i.e.* $\beta_t = \beta \log t$. Large values of β increase the convergence rate of the algorithm. However, if its value is too large, the algorithm might converge to a local minimum instead of a global one (see for example [Haj88]).

Alg. 2 Graph barycenter estimation algorithm of [IJ-19]

Require: Continuous version of $G = (N, E)$, *i.e.* Γ_G .

Require: Observations sequence $Y = (Y_k)_{k \geq 1}$ on the nodes set N .

Require: Increasing inverse temperature $(\beta_t)_{t \geq 0}$ and intensity $(\alpha_t)_{t \geq 0}$.

1: Pick $X_0 \in \Gamma_G$ and set $K = \text{len}(Y) - 1$.¹

2: $T_0 = 0$.

3: **for** $k = 0 : K$ **do**

4: Generate T_k according to α_k .

5: Generate $\varepsilon_k \sim \mathcal{N}(0, \sqrt{T_k - T_{k-1}})$.

6: Randomly move X_k (Brownian motion): $X_k = X_k + \mathbf{h}_k \varepsilon_k$, where $\mathbf{h}_k$ is a direction uniformly chosen among the directions departing from X_k , and ε_k is a step size.

7: Deterministically move X_k towards Y_{k+1} : $X_{k+1} = X_k + \beta_{T_k} \alpha_{T_k}^{-1} \mathbf{X}_k \mathbf{Y}_{k+1}$, where $\mathbf{X}_k \mathbf{Y}_{k+1}$ represents the shortest (geodesic) path from X_k to Y_{k+1} in Γ_G .

8: **end for**

9: **return** Graph location X_K estimated as the barycenter of Γ_G . We consider the nearest node to X_K in G as its barycenter.

¹ Here $\text{len}(Y)$ represents the length of the sequence $(Y_k)_{k \geq 1}$.

Another important parameter comes from the online aspect of the algorithm. In our model, we simulate the arrival times T_n of the observations Y_n by an inhomogeneous Poisson process $(N_t^\alpha)_{t \geq 0}$ ², where $(\alpha_t)_{t \geq 0}$ is a continuous and increasing function that describes the rate at which we use the sequence of observations $(Y_n)_{n \geq 0}$. We refer to $(\alpha_t)_{t \geq 0}$ as the *intensity* of the process. On one hand, and from a theoretical point of view, using more observations improves the algorithm's accuracy and convergence rate, so it may seem natural to use large values for α_t . On the other hand, in practice, observations can be costly and limited, so one would like to limit their use as much as possible. More discussions about the algorithm and its convergence are given in [IJ-19].

Multi-scale graph barycenter estimation

The key practical issue with the strategy of [IJ-19] on large graphs is that the deterministic move (row 7 of Alg. 2) requires to compute the shortest path from X_k to Y_{k+1} . To achieve this, we used a standard Dijkstra's algorithm which is particularly demanding in terms of computational times, especially when computed $K + 1$ times. The solution of [IJ-19] was to pre-compute once for all the shortest distances between all node pairs and then to use this information for a quick algorithm execution. Computing these distances is $\#N$ times slower than computing the shortest path between two nodes, where $\#N$ is the number of nodes in $G = (N, E)$. This solution then makes sense when $K + 1$ is larger than $\#N$, or when multiple runs of the algorithm will be performed on the same graph, *e.g.* in order to evaluate the barycenters related to different observation sets Y . Its major drawback is however that it requires to store a $\#N \times \#N$ matrix in memory, which is unrealistic when $\#N$ is large. Moreover, the algorithmic cost of a Dijkstra's algorithm on weighted graphs is anyway $\mathcal{O}(\#N^2)$

² T_n is the n -th jumping time of the Poisson process N_t^α , $T_n := \inf\{t : N_t^\alpha = n\}$

and therefore does not scale at all to large graphs. We then proposed in [SJ-6] a multiscale extension of [IJ-19], where reasonable heuristics were used to make the problem scalable. The broad lines of this strategy are given in Alg. 3.

Alg. 3 Multiscale barycenter estimation

Require: Graph $G = (N, E)$.

- 1: Partition $G = (N, E)$ is partitioned into I sub-graphs $G_i = (C_i, E_i)$.
 - 2: Undersample $G = (N, E)$ in $\tilde{G} = (\tilde{N}, \tilde{E})$, where each node of $\tilde{G}$ represents a compact description of sub-graph G_i .
 - 3: Estimate the barycenter $\tilde{b}$ of $\tilde{G}$ using Alg. 2.
 - 4: Compute a multiscale graph $\hat{G} = (\hat{C}, \hat{E})$ with the nodes of G in the subgraph of $\tilde{b}$ and the nodes of $\tilde{G}$ elsewhere.
 - 5: Estimate the barycenter $\hat{b}$ of $\hat{G}$ using Alg. 2.
 - 6: **return** Node $\hat{b}$ estimated as the barycenter of G
-

In short, the graph $G = (N, E)$ is partitioned into I clusters using any clustering strategy that scales to large data, *e.g.* [pyt]. A connected subgraph $G_i = (C_i, E_i)$ is then defined for each cluster i and its center will be a node of the undersampled graph $\tilde{G}$. After estimating the barycenter $\tilde{b}$ of this undersampled graph, a multiscale graph $\hat{G}$ is generated. This graph contains all nodes and edges of $G = (N, E)$ the central cluster of $\tilde{b}$ and only the subsampled information in other graph clusters. As a result, a fine barycenter of G can be estimated in $\hat{G}$ with a strongly limited number of information. In addition to this multiscale model, [SJ-6] develops algorithmic strategies to make the estimation of $\tilde{G}$ and $\hat{G}$ scalable on large graphs. In particular, this requires to define a reasonable approximation strategy of the subsampled edge lengths in $\tilde{G}$ and $\hat{G}$, which is discussed and assessed in the paper.

3.3.3 Results and discussion

The strategies of [IJ-19] and [SJ-6] were tested on graphs of different sizes and structures (subway and road networks as well as social networks). We give here representative results of [SJ-6] obtained on the subway network of Paris and the road network of the New-York City urban area.

In Fig. 3.6, we first represent subsampled and multiscale versions of the Parisian metro graph, based on Alg. 3. The whole graph has 296 nodes and 353 edges and it was partitioned into 19 subsets of subway stations. Node related observations were also sampled based on the number of passengers getting in the subway network at the corresponding stations during one year. Note that when running 100 times Alg. 2 on the whole graph, the subway station *Chatelet* was always estimated as the graph barycenter, which is not a surprise for most Parisians. This also satisfactorily assessed the methodology in this simple example. Interestingly, one can see in Fig. 3.6(top-right) that Chatelet is not included in $\tilde{G}$ and thus cannot be estimated as the subsampled graph center, which justifies the use of the multiscale graph $\hat{G}$ here. By running 100 times the multi-scale extension Alg. 2, we then estimated 97 times *Chatelet*. This extension was then shown efficient and stable in this test.

More exhaustive assessment of the proposed framework was made on a graph representing the road network of the New York city urban area. In particular,

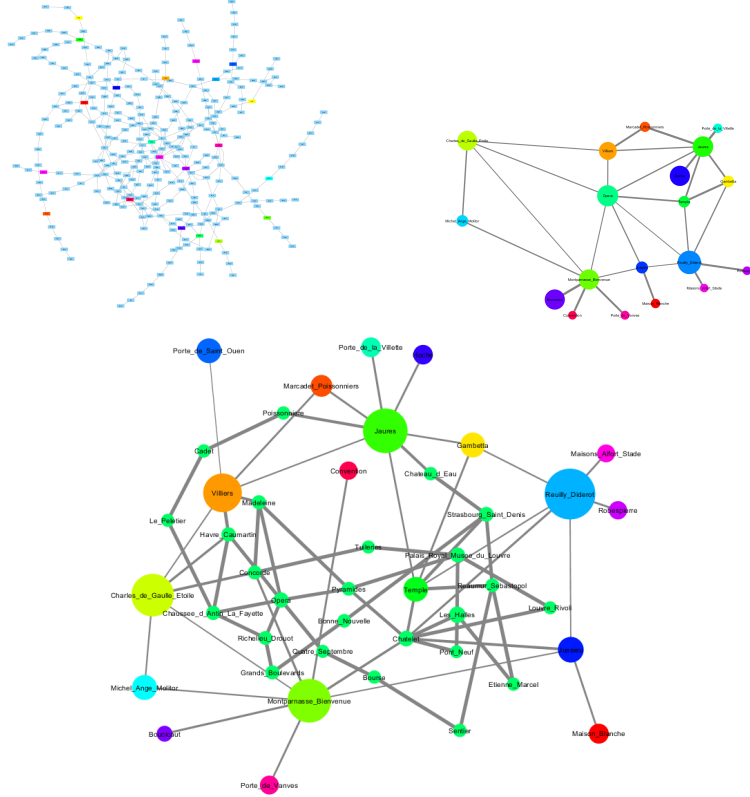


Figure 3.6: Graph representations of the Parisian Metro. Nodes locations do not reflect the actual GPS coordinates and the edges width is inversely proportional to the time needed to go from a station to another one. **(Top-left)** Complete graph G . The colored nodes are the estimated barycenters of the precomputed sub-graphs G_i and are be the nodes $\tilde{N}$ of $\tilde{G}$. **(Top-right)** Subsampled graph $\tilde{G}$. **(Bottom)** Multiscale graph $\hat{G}$. Illustration out of [SJ-6].

we tested the algorithm sensitivity to the parameter β , the stopping time T , the number of observations O and the graph parcellation. The initial graph had 264.346 nodes and 733.846 edges and its multiscale version had slightly less than 1% of this volume. This made it possible to estimate this network barycenter on a laptop with 32GB memory in less than 1 minute after 3h and 30minutes for the pre-computation of the subsampled graph. Our key result was that the variability of the barycenter estimate on different runs and with recommended parameters was less than 3% of the graph diameter, which is relatively accurate. This result is illustrated in Fig. 3.7.

Remark finally that the strategies were also tested on social networks graphs out of Facebook, zbMATH and Youtube. Results were also stable in these graph although slightly less than in the subway and road networks presented here, as discussed in [IJ-19,SJ-6]. A Python package implementing these strategies will finally be distributed subject to [SJ-6] acceptance.

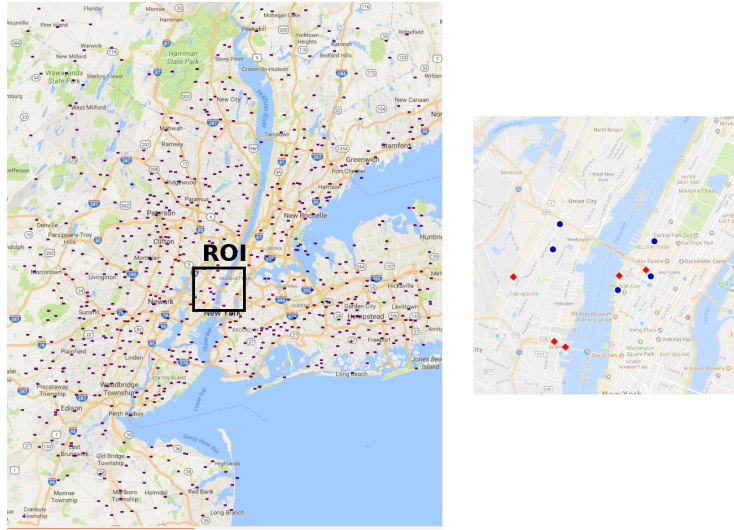


Figure 3.7: Results obtained on the road network of the New-York city urban area. **(Left)** General view of the complete New York graph. The red diamond-shaped points are the barycenters obtained on a pre-defined graph partition. **(Right)** Barycenters obtained using two different graph partitions. The barycenter estimations are represented in the region of interest (ROI) defined in the general view. The red and blue dots were obtained using each of the two partitions. Illustration out of [SJ-6].

3.4 Outlook

Immediate extensions of the works presented in this chapter will be carried-out by continuing my collaboration with I. Gavra. We work on an extension of [SJ-6] where we mathematically study the impact of the graph parcelation, on which our multi-scale model is based, on the barycenter estimates. We also believe that estimating average paths would also be a nice extension of this work.

I also continue working with G. Fort and S. Gadat on extensions of [IC-36] and [SJ-1], respectively. For now my contributions are mostly technical but I expect developing a stronger scientific activity in numerical methods for stochastic modeling based on these collaborations.

Finally, I have developed these two last years a good experience in GPGPU computing (General-purpose computing on graphics processing units) which allows to massively parallelize the computations in scientific computing. Graphics processing units are indeed specific architectures that were designed to solve very efficiently common linear algebra computations such as matrix/vector multiplications. These architectures are however not particularly suited to process irregular data such as graphs, in the sense that the data are not regularly organized in the memory. GPUs are however cheap and popular. I was also able to speed-up by a factor 10 a large graph clustering algorithm using simple GPGPU techniques (internship of V. Brès at IMT). I therefore expect to develop stochastic algorithms for the analysis of irregular data on GPU architectures. Specifically, I believe that by developing stochastic algorithms adapted to the

material constraints of GPU architectures, the time required to clusterize very large graphs could be strongly reduced.

Chapter 4

Statistical learning for complex data

4.1 Summary of contributions

This chapter presents my scientific contributions in statistical learning for complex data. Complex data are understood here in the sense that they have a high dimension with a potentially limited number of observations. Heterogeneity between the variables behavior and multiple similarities between variable subsets can also make ambiguous their analysis. The definition of pertinent regularization techniques and adapted analysis models are then central issues here. It is interesting to make a link between these issues and those addressed by the image registration models and regularization techniques of Chapter 2. The methods of Chapter 2 may indeed be seen as a category of complex data analysis methods in which the geometry of the information is strongly related to physiological tissue properties and the reliable information is sparsely distributed at the location of pertinent intensity gradients. To make clear the distinction between the methods presented here and those of Chapter 2, no strategy of the present chapter mimics physiologically inspired models for image matching.

Summarized papers The project of [IC-28,IJ-15] (Appendix C.1) is the first one in which I participated to scientific contributions in statistical learning for complex data. It consisted in empirically exploring the use of different spatial regularization models for logistic regression on 3D image domains. The classified observations were indeed initial momenta computed using [IJ-7], which is presented in Subsection 2.2.2. The goal of the logistic regression model was to learn local shape changes which optimally discriminate subjects with Alzheimer's disease from subject with Mild Cognitive Impairment. This work consists in the empirical evaluation of regularization methods in the specific high dimensional context of 3D medical images, it is briefly presented in Subsection 4.2.

I also work on the development of new statistical learning techniques for complex data by co-supervizing two PhD theses with J.M. Loubes (Pr Univ. Toulouse, IMT) . I first work with T. Bui on the development of statistical

models adapted to the analysis of 3D coiled shapes out of the inner ear and response distributions of the ear to otoacoustic emissions. In this context, we developed the distribution regression strategy of [SJ-5] for probability distributions which is based on a Reproducing Kernel Hilbert Space (RKHS) regression framework. An overview of this strategy is given in Subsection 4.3 and the submitted manuscript can be found in Appendix C.3. I also work with C. Champion on the extraction of representative variables in complex systems. In particular, we developed an original and scalable graph clustering method which is directly applied to regularize the detection of representative variables in high dimensional systems with a little amount of observations. The submitted paper [SJ-1] is given in Appendix C.2 and an overview of the method is developed in Subsection 4.4.

Other papers Among the three themes distinguished in this manuscript, this is the one in which my research activity is the newest although I applied such techniques in various projects. In [IC-10,NC-2], I first applied statistical learning techniques to clusterize hemodynamic parameters out of functional MRI time series in the framework of [NJ-1]. In [IC-33] (Appendix A.11), I also used LASSO regularization to select the optimal scale of the deformations between two registered images, as discussed in Section 2.3.1. In [NJ-2], permutation tests and the Akaike information criterion (AIC) were also used to clusterize different evolution models in anatomical data out of the inner ear of early hominins fossils. In these three projects, different learning strategies were however only applied, so they will not be further discussed in this manuscript.

4.2 Regularization models on 3D image domains

4.2.1 Motivation

The early detection of Alzheimer’s disease (AD) is an important challenge for its efficient treatment through adapted drug delivery. In [IC-28,IJ-15], we worked on its detection based on local hippocampal shape changes in time. The hippocampus is indeed a subcortical structure which is known by the clinicians to be anatomically impacted by AD. The main paper contribution was to explore the use of different spatial regularization models in logistic regression to learn local shape changes that optimally distinguish AD subjects from subject with Mild Cognitive Impairment. The novelty in this context was the development and the assessment of a mathematically rigorous image analysis pipeline for such applications. Here is the below a brief overview of the method and its results.

4.2.2 Methodology

The row material of [IJ-15] was a set of MR images of the brain acquired on $n = 103$ patients out of the ADNI¹ database [SGMWLJ⁺05]. An example of such images is illustrated in Fig. 4.1. All considered patients had Mild Cognitive Impairment (MCI) at a first image acquisition time, denoted baseline.

¹<http://www.loni.ucla.edu/ADNI>

Follow-up acquisitions of the same patients were made one year after. A diagnostic made several years later also distinguished 19 patients who converted to Alzheimer’s disease (AD) and 84 who are still MCI. The main question addressed in this project is then whether the early anatomical transformations of the hippocampus can predict this diagnostic.

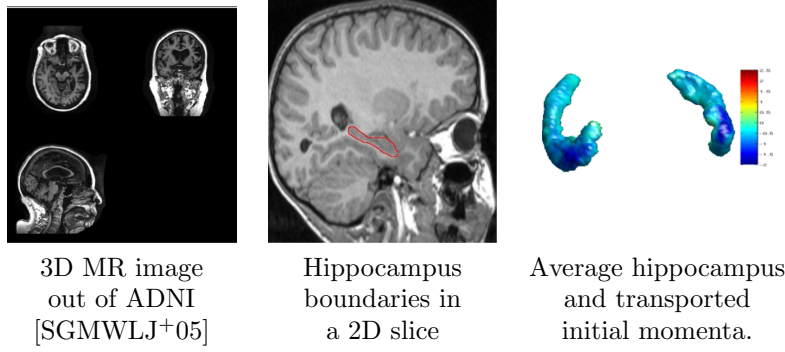


Figure 4.1: Data on which the supervised classification of [IJ-15] is performed. **(Left)** Input MR image acquired on a patient at baseline. **(Center)** Representation of the hippocampus surface by the red curve, in a 2D frame out of the MR image. **(Right)** Volumetric representation of the average hippocampus in all data, *i.e.* the template. The color represent the amount of temporal deformations observed in one patient and transported on the template. Illustration out of [IJ-15].

To solve this problem, the follow-up images related to each patient were registered using [IJ-7]. The local deformations of each of the 103 hippocampi were then encoded in a 3D field of initial momenta (see Subsections 2.2.2) containing about $p = 20000$ scalars. As all hippocampi at baseline do not have the same shape, all the initial momenta were transported to a common template. Template definition was made on a subset of 20 segmented hippocampal volumes using the Karcher mean strategy of [IJ-8] with a metric of [IJ-6]. The predictive model of [IJ-15] was then

$$\mathbf{y} = F(\mathbf{X}\mathbf{w} + b), \quad (4.1)$$

where $\mathbf{y} \in \{\pm 1\}^n$ is the behavioral variable (conversion or not to AD), $\mathbf{X} \in \mathbb{R}^{n \times p}$ contains the n observations of patient specific deformations, *i.e.* the n transported initial momenta. The vector $\mathbf{w} \in \mathbb{R}^p$ and the scalar b are then the parameters to estimate. This is made using a logistic regression model where the probability for a patient i to have label y_i given the observation $\mathbf{x}_i$ is

$$p(y_i | \mathbf{x}_i, \mathbf{w}, b) \stackrel{\text{def.}}{=} \frac{1}{1 + \exp(-y_i (\mathbf{x}_i^T \mathbf{w} + b))}. \quad (4.2)$$

By considering a log likelihood model and considering the observations as independent, the problem then consists in finding $\arg \min_{\mathbf{w}, b} \mathcal{L}(\mathbf{w}, b) + \lambda J(\mathbf{w})$, where the loss term $\mathcal{L}$ is

$$\mathcal{L}(\mathbf{w}, b) \stackrel{\text{def.}}{=} \frac{1}{n} \sum_{i=1}^n \log(1 + \exp(-y_i (\mathbf{x}_i^T \mathbf{w} + b))). \quad (4.3)$$

We recall that $n \ll p$ in this context as $n = 103$ subjects are studied and each observation is about in dimension $p = 20000$. One of the main contribution of the paper was then to use different regularization strategies in this context by testing different models for $J(\mathbf{w})$. Ridge [HK70], LASSO [Tib94] and elastic net [ZH05] regularizations which do not take into account any spatial structure of the coefficients were first tested. They were compared with the total variation, Sobolev and fused LASSO regularizations [TSR⁺05] in 3D volumes which take into account the spatial distribution of the initial momenta. Total variation favors sharp spatial transitions between the values of $\mathbf{w}$ while Sobolev enforces smooth transition. Fused LASSO adds to Total variation a sparsity constraint on the values with an additional LASSO constraint. The energy was finally optimized using the HANSO algorithm [LO12].

4.2.3 Results

In the results section, different regularization strategies and weights λ were tested. The predictive power of each strategy was measured using a leave-10%-out scheme. Results were particularly interesting for a main reason: The regularization strategies leading to the highest predictive power were clearly Total Variation and Fused LASSO. These methods are those that model sharp spatial transitions for the values of $\mathbf{w}$, which is consistent with physiological studies showing that specific regions of the hippocampus are likely to be related to AD conversion.

4.3 Distribution regression with a RKHS approach

4.3.1 Motivation

Regression analysis is a predictive modeling technique which has been widely studied over the last decades. Its goal is to investigate relationships between predictors and responses. In the context functional data regression, extensions the Reproducing Kernel Hilbert Space (RKHS) framework has become popular to extend the results of the statistical learning theory. As far as the authors know, It has however not been extended so far to probability distribution spaces although this may open various applications. This is the driving motivation of this work.

4.3.2 Methodology

We consider the model:

$$y_i = f(\mu_i) + \epsilon_i \quad (4.4)$$

where $\{\mu_i\}_{i=1}^n$ are n observed probability distributions on $\mathbb{R}$ and the $\{y_i\}_{i=1}^n$ are real numbers (scores) associated to each μ_i . In addition, ϵ_i represents an i.i.d. noise and f is a functional valued function [Pre07]. As in classical regression models, we estimate the unknown function f based on the observations $\{(\mu_i, y_i)\}_{i=1}^n$. The regression problem will be regularized using the Wasserstein distance on probability distributions.

Consider two probability distributions μ and ν in $\mathcal{W}_2(\mathbb{R})$, F_μ^{-1} the quantile function associated to μ , and U a uniform random variable on $[0, 1]$. In this case, the simplest expression for the Wasserstein distance W_2 is given in [Whi76]:

$$W_2(\mu, \nu) = \mathbb{E}(F_\mu^{-1}(U) - F_\nu^{-1}(U))^2. \quad (4.5)$$

Following [BGLV17], a positive definite kernel $k_\Theta : \mathcal{W}_2(\Omega) \times \mathcal{W}_2(\Omega) \rightarrow \mathbb{R}$ can be defined as follows:

$$k_\Theta(\mu, \nu) := \gamma^2 \exp\left(-\frac{W_2^{2H}(\mu, \nu)}{l}\right), \quad (4.6)$$

where the parameter $\Theta := (\gamma, H, l)$ is such that $\gamma \neq 0$ and $l > 0$. As shown in [BGLV17], k_Θ is a positive definite kernel if $0 < H \leq 1$.

A Reproducing Kernel Hilbert Space (RKHS) space $\mathcal{F}_0$ is then defined by considering the following inner product

$$\langle f_n, g_m \rangle_{\mathcal{F}_0} = \sum_{i=1}^n \sum_{j=1}^m \alpha_i \beta_j k_\Theta(\mu_i, \nu_j),$$

where $f_n(\bullet) = \sum_{i=1}^n \alpha_i k_\Theta(\bullet, \mu_i)$ and $g_m(\bullet) = \sum_{j=1}^m \beta_j k_\Theta(\bullet, \nu_j)$. Now, let $\mathcal{F}$ be the space of all continuous real-valued functions from $\mathcal{W}_2(\Omega)$ to $\mathbb{R}$. We prove in [SJ-5] that $\mathcal{F}_0 \subset \mathcal{F}$, so our estimation strategy can be written as:

$$\hat{f} = \underset{f \in \mathcal{F}}{\operatorname{argmin}} \left(\sum_{i=1}^n |y_i - f(\mu_i)|^2 + \lambda \|f\|_{\mathcal{F}}^2 \right), \quad (4.7)$$

where the weight λ is strictly positive. Using the Representer theorem [KW70, SHS01], this leads to the following expression for $\hat{f}$,

$$\hat{f} : \mu \mapsto \hat{f}(\mu) := \sum_{j=1}^n \alpha_j k_\Theta(\mu, \mu_j), \quad (4.8)$$

where $\{\alpha_j\}_{j=1}^n$ are obtained from the data by solving the system of linear equations given Eq. (16) of [SJ-5].

4.3.3 Results and discussion

After studying the influence of the parameters on synthetic data, interesting results were obtained on 48 oto-emission curves, representing the response of the ear to different frequencies (here between 0Hz and 10kHz). The curves are illustrated in Fig. 4.2. Each of them is associated to the age of the patient on which it was acquired. A leave-one-out procedure was then used to validate the effectiveness of the regression approach to predict the patient age based on its oto-emission curve. Results are given in Fig. 4.3, where it can be observed that the regression strategy performed well in the range of ages at which most acquisitions were made. Note that the curves were considered as probability distributions after a suitable normalization. More details are given in Appendix C.3.

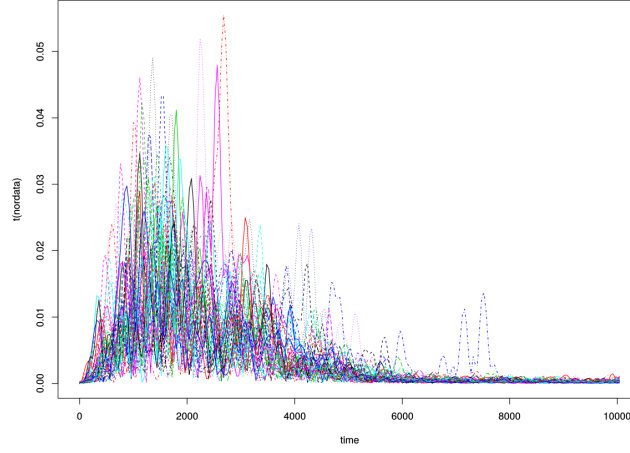


Figure 4.2: Oto-emission curves obtained on 48 subjects with frequencies ranging from 0Hz to 10kHz. Illustration out of [SJ-5].

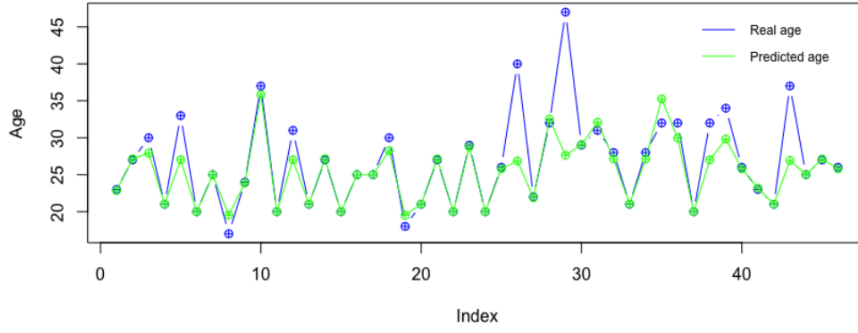


Figure 4.3: Prediction of the age using a leave-one-out procedure, based on the functions of Fig. 4.2. Illustration out of [SJ-5].

4.4 Representative variable detection for complex data

4.4.1 Motivation

Discovering representative information in high dimensional spaces with a limited number of observations is a recurrent problem in data analysis. Heterogeneity between the variables behavior and multiple similarities between variable subsets make the analysis of complex systems an ambiguous task. In [SC-1], we then presented a formalism to regularize the selection of representative variables in such complex systems. The formalism is based on a specific graph clustering strategy, denoted CORE-clustering, adapted to the addressed problem. In particular, each cluster must contain a minimum number of variables and all its variables must have a coherent observed behavior. The representative variables

are then the cluster centers. The key feature of this contribution compared with the literature is that the regularization level is then controlled by the minimum number of variables in each cluster, which is a particularly intuitive parameter to tune.

4.4.2 Methodology

Let us consider a complex system of p quantitative variables $X = (X^1, \dots, X^p)$ and n observations of these variables $(X_1^i, \dots, X_n^i)$, where $n \ll p$. The similarity between the observed variables behavior can be modeled in a graph $\mathcal{G}$, where each node represents a variable and each edge weight encodes the similarity between two variable behaviors. The norm of the correlation between variable pairs or other similarity measures may be used for this weight.

In order to define a notion of coherence in a subset of variables, we first define the coherence $c(X^i, X^j)$ between X^i and X^j

$$c(X^i, X^j) = \max_{P \in \mathbf{P}_{i,j}} \min_{e_{m,n} \in P} w_{m,n}, \quad (4.9)$$

where $\mathbf{P}_{i,j}$ is the set of all possible paths in $\mathcal{G}$ between the nodes representing X^i and X^j , and $w_{m,n}$ is the weight between the nodes representing X^m and X^n . Note that $\min_{e_{m,n} \in P} w_{m,n}$ is known as the capacity of path P in $\mathcal{G}$ [Pol60, Hu61]. We now denote S a subset of the variables X . The coherence $\mathbf{c}(S)$ of this variable subset is the minimal coherence between the variables it contains:

$$\mathbf{c}(S) = \min_{(X^i, X^j) \in S^2} c(X^i, X^j) \quad (4.10)$$

If all the variables of S have a coherent observed behavior, then $\mathbf{c}(S)$ is high, and $\mathbf{c}(S)$ drops if at least one variable of S does not behave like another variable of this subset. The justification of this measure is further discussed and illustrated in [SC-1].

In order to regularize the detection of the representative variables in X , we consider these variables as the center of particularly coherent clusters in $\mathcal{G}$ containing between τ to 2τ variables. A cluster S is considered as particularly coherent if $\mathbf{c}(S) > \xi$. For instance, if the In the norm of the correlation between variable pairs is used $\xi = 0.75$ can be a reasonable choice. We denote CORE-clusters a clusters which respects these two constraints. The variable selected in each CORE-cluster is then considered as an influential variable as its behavior is coherent with at least τ other variables. We proposed in [SC-1] to find optimal CORE-clusters $\hat{S} = \{\hat{S}_\lambda\}_{\lambda \in \{1, \dots, \Lambda\}}$ by optimizing:

$$\hat{S} = \arg \max_S \sum_{\lambda=1}^{\Lambda} \mathbf{c}(S_\lambda), \quad (4.11)$$

where Λ is not fixed, $S_{\lambda_1} \cap S_{\lambda_2} = \emptyset$ for all possible (λ_1, λ_2) , and each S_λ contains τ to $2\tau - 1$ variables and has a coherence $\mathbf{c}(S_\lambda) > \xi$. Note that the amount of combinations to explore is huge and computing Eq. (4.10) is particularly demanding. In order to make the strategy computationally scalable, the problem is first solved on the maximum spanning tree [Kru56] of the whole

graph. Two aggregative and divisive optimization algorithms which do not explicitly compute Eq. (4.11) are then presented in [SC-1]. We show in this paper that their algorithmic cost is lower than $K \log(K)$, where K is the number of graph edges. This is much lower than existing clustering strategies with a relatively similar control on the sought clusters [Sei83, GMT⁺16] as discussed in [SC-1].

4.4.3 Results and discussion

The two proposed CORE-clustering algorithms were first tested on synthetic data, which allowed to evaluate their performance depending on the noise in the data and the number of observations n for a given variables number p . The influence of the parameter τ was also evaluated. Results on real data were obtained on the classic Yeast dataset [SSZ⁺98] which can be found on the UCI Machine Learning Repository² and includes $n = 77$ yeast samples under various time during the cell cycle and a total of $p = 1660$ genes, so $n \ll p$. The goal of this analysis was to detect CORE-clusters among the correlation patterns in the time series of gene expressions of yeast measured along the cell cycle. The similarity between all gene pairs was measured using the absolute value of Pearson's correlation. A total of about 1.3×10^6 weighted edges was then used when representing the variable correlations in the resulting graph.

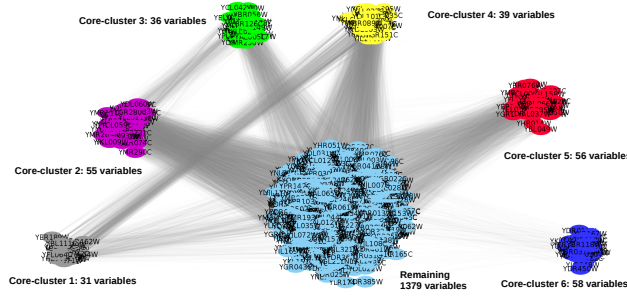


Figure 4.4: CORE-clusters obtained using the strategy of [SC-1] on the yeast dataset of [SSZ⁺98] and the granularity coefficient $\tau = 30$. CORE-clusters containing 30 to 59 variables were estimated.

The two proposed algorithms required about 160 and 3 seconds on an Intel Core i7 computer, respectively. An illustration of the CORE-clusters selected using $\tau = 30$ is given in Fig. 4.4. We have shown in [SC-1] that the results obtained using the two algorithms were coherent to each other. We have also shown that by using only 30 observations instead of 77 (specifically the 30 first observations of the dataset), we selected two representative variables out of the four that were detected using 77 observations. The observed influence of the two other representative variables was probably lost in the 37 observations we removed. This suggests that our strategy detects stable representative variables, even when the number of observations is very low compared with the dimension of the observations. The submitted paper can be found in Appendix C.2.

²<https://archive.ics.uci.edu/ml/datasets/Yeast>

4.5 Outlook

Immediate research directions in complex data analysis will be carried-out in the context T. Trang Bui and C. Champion PhD theses. In the context of T. Trang Bui's thesis, we establish a model to learn a link between the inner-ear shape properties and the frequencies that can be heard. We also work with C. Champion on regularization models to make more robust the representative variable detection model of [SC-1].

An important project to me is the development of new methods to detect and correct selection biases in machine learning. We indeed started with J.M. Loubes, F. Gamboa and F. Bachoc a tight collaboration to propose new Fair learning solutions. We also plan to submit an ANR proposal on this topic this year.

I also have an ANR project with S. Déjean (IR, IMT), J. Farinas (Mcf IRIT), K. Daoudi (INRIA Bordeaux) and members of Toulouse and Bordeaux hospitals to discriminate two types of pathologies (Alzheimer's Disease and MSA) based on the voice. We then hope to find characteristics which would make an early diagnosis efficient based on machine learning techniques.

I currently write a proposal with F. Malgoures (Pr IMT) in order to study satellite images of fire propagations. Image segmentation should be reasonably performed using standard image analysis techniques. We are however interested in developing statistical-learning based techniques to predict the fire propagation.

I finally established links with P. Payoux and the TONIC team (INSERM, Toulouse hospital) to distinguish healthy and pathological aging based on longitudinal multi-modal data. We currently start this collaboration by co-supervising a student project and we write a proposal for a post-doctoral position on this topic.

Bibliography

- [ACAP09] Vincent Arsigny, Olivier Commowick, Nicholas Ayache, and Xavier Pennec. A fast and log-euclidean polyaffine framework for locally linear registration. *Journal of Mathematical Imaging and Vision*, 33(2):222–238, 2009.
- [ACPA06] V. Arsigny, O. Commowick, X. Pennec, and N. Ayache. A log-Euclidean framework for statistics on diffeomorphisms. In *Proc. MICCAI*, pages 924–931, 2006.
- [AEGG08] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12:26–41, 2008.
- [AMY16] S. Arguillère, M. Miller, and L. Younes. Diffeomorphic surface registration with atrophy constraints. *SIAM Journal on Imaging Sciences*, 9(3):975–1003, 2016.
- [Aro50] N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337–404, 1950.
- [ATTY15] S. Arguillere, E. Trélat, A. Trouvé, and L. Younes. Shape deformation analysis from the optimal control viewpoint. *J. Math. Pures Appl.*, 104, 2015.
- [BAL⁺16] A.-C. Brunet, J.-M. Azais, J.-M. Loubes, J. Amar, and R. Burcelin. A new gene co-expression network analysis based on core structure detection (csd). *Arriv*, 2016.
- [Bav50] Alex Bavelas. Communication patterns in task-oriented groups. *Journal of the Acoustical Society of America*, 1950.
- [BGLV17] F. Bachoc, F. Gamboa, J.-M. Loubes, and N. Venet. A gaussian process regression model for distribution inputs. *IEEE Transactions on Information Theory*, 2017.
- [BMTY05a] Faisal M. Beg, Michael I. Miller, Alain Trouvé, and Laurent Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *IJCV*, 61(2):139–157, 2005.
- [BMTY05b] M F Beg, M I Miller, A Trouvé, and L Younes. Computing large deformation metric mappings via geodesic flow of diffeomorphisms. *IJCV*, 61:139–157, 2005.

- [Bon72] Phillip Bonacich. Factoring and weighting approaches to status scores and clique identification. *The Journal of Mathematical Sociology*, 2(1):113–120, 1972.
- [Bor05] Stephen P. Borgatti. Centrality and network flow. *Social Networks*, 1:55–71, 2005.
- [Bro92] L. G. Brown. A survey of image registration techniques. *ACM Comput. Surv.*, 24(4):325–376, 1992.
- [DA00] James S. Duncan and Nicholas Ayache. Medical image analysis: Progress over two decades and the challenges ahead. *IEEE trans. Pattern Analysis and Machine Intelligence*, 22(1):1030–1069, 2000.
- [DGM98] P. Dupuis, U. Grenander, and M. I. Miller. Variational problems on flows of diffeomorphisms for image matching. *Quart. Appl. Math.*, 56:587–600, 1998.
- [Est15] E. Estrada. *Introduction to Complex Networks. Structure and Dynamics, chapter of Evolutionary Equations with Applications to Natural Sciences*. Lecture Notes in Mathematics. Springer, 2015.
- [FLPJ04] P. T. Fletcher, C. Lu, M. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions Medical Imaging*, pages 995–1005, 2004.
- [FM14] J. Fehrenbach and J.-M. Mirebeau. Sparse non-negative stencils for anisotropic diffusion. *Journal of Mathematical Imaging and Vision*, 49(1):123 – 147, 2014.
- [FP03] F. Forbes and N. Peyrard. Hidden Markov Random Field selection criteria based on Mean Field-like approximations. *IEEE trans. Pattern Analysis and Machine Intelligence*, 8(25):1055–67, 2003.
- [Fré48] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l’Institut Henri Poincaré (B)*, 10:215–310, 1948.
- [Fre78] Linton C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215 – 239, 1978.
- [FVJ08] P. T. Fletcher, S. Venkatasubramanian, and S. Joshi. Robust statistics on riemannian manifolds via the geometric median. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, 2008.
- [GA10] Fienberg S.E. Airoldi E. M. Goldenberg A., Zheng A. X. A survey of statistical network models. *Found. Trends Mach. Learn.*, 2(2):129–233, Feb. 2010.

- [GM98] U. Grenander and M. I. Miller. Computational anatomy: An emerging discipline. *Quarterly of Applied Mathematics*, LVI(4):617–694, 1998.
- [GMT⁺16] Christos Giatsidis, Fragkiskos D. Malliaros, Nikolaos Tziortziotis, Charanpal Dhanjal, Emmanouil Kiagias, Dimitrios M. Thilikos, and Michalis Vazirgiannis. A k-core decomposition framework for graph clustering. *CoRR*, 2016.
- [GY05] L. Garcin and L. Younes. Geodesic image matching: A wavelet based energy minimization scheme. In *Proc. EMMCVPR*, pages 349–364, 2005.
- [Haj88] B. Hajek. Cooling schedules for optimal annealing. *Mathematics of operations research*, 13(2):311–329, 1988.
- [HBJ⁺97] David M. Higdon, James E. Bowsher, Valen E. Johnson, Timothy G. Turkington, David R. Gilland, and Ronald J. Jaszczak. Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data. *IEEE Transactions on Medical Imaging*, 16(5):516–526, October 1997.
- [HDJR12] Mohamed Hachama, Agnès Desolneux, and Frédéric J.P. Richard. A bayesian technique for image classifying registration. *IEEE transactions on image processing*, 21(9):4080–91, 2012.
- [HJB⁺12a] M. P. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F. V. Gleeson, S. M. Brady, and J. A. Schnabel. MIND: modality independent neighbourhood descriptor for multi-modal deformable registration. *Med. Im. Analysis*, 16(7):1423–1435, 2012.
- [HJB⁺12b] M. P. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F. V. Gleeson, Brady S. M., and J.A. Schnabel. Mind: modality independent neighbourhood descriptor for multi-modal deformable registration. *Med Image Analysis*, 16(7):1423–35, 2012.
- [HK70] A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12:55–67, 1970.
- [HK10] M. Handcock and K. Gile. Modeling social networks from sampled data. *The Annals of Applied Statistics*, 2010.
- [HRS⁺99] A. Hagemann, K. Rohr, H. S. Stiehl, U. Spetzger, and J. M. Gilsbach. Biomechanical modeling of the human head for physically based, non-rigid image registration. *IEEE Transactions on Medical Imaging*, 18(10):875–884, 1999.
- [HS81] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17, 1981.

- [Hu61] T. C. Hu. The maximum capacity route problem. *Operations research*, 9:898–900, 1961.
- [JDJG04] S. Joshi, Brad Davis, Matthieu Jomier, and Guido Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23:151–160, 2004.
- [Kar77] H. Karcher. Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.*, 30(5):509–541, 1977.
- [Kru56] J. B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7:48–50, 1956.
- [KW70] G. S. Kimeldorf and G. Wahba. A correspondence between bayesian estimation on stochastic processes and smoothing by splines. *The Annals of Mathematical Statistics*, 41(2):495–502, 1970.
- [LK81a] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. Seventh International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.
- [LK81b] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2, IJCAI’81*, pages 674–679, 1981.
- [LKB⁺17] G Litjens, T Kooi, B. E. Bejnordi, Setio A. A. A., F Ciompi, M Ghafoorian, Van der Laak n J. A. W. M., B. Van Ginneken, and C. I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60 – 88, 2017.
- [LO12] A. S. Lewis and M. L. Overton. Nonsmooth optimization via quasi-newton methods. *Mathematical Programming*, pages 1–29, 2012.
- [MCV⁺97] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Trans. on Med. Imaging*, 16(2):187–198, 1997.
- [MIV⁺08] S. Makni, J. Idier, T. Vincent, B. Thirion, G. Dehaene-Lambertz, and P. Ciuciu. A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI. *NeuroImage*, 41(3):941–969, July 2008.
- [Mor34] J. Moreno. *Who Shall Survive?* Nervous and Mental Disease Publishing Company, 1934.
- [MPS⁺11] T Mansi, X. Pennec, M. Sermesant, H. Delingette, and N. Ayache. iLogDemons: A demons-based registration algorithm for tracking incompressible elastic biological tissues. *IJCV*, 92(1):92–111, 2011.

- [MTY02] M. I. Miller, A. Trouvé, and L. Younes. On the metrics and Euler-Lagrange equations of computational anatomy. *Annual Review of biomedical Engineering*, 4:375–405, 2002.
- [MTY06] M. I. Miller, A. Trouvé, and L. Younes. Geodesic shooting for computational anatomy. *Journal of Mathematical Imaging and Vision*, 24(2):209–228, 2006.
- [MVGR⁺11] K. Murphy, B. Van Ginneken, J. M. Reinhardt, S. Kabus, K. Ding, X. Deng, J. P. W. Pluim, and et. al. Evaluation of registration methods on thoracic CT: The EMPIRE10 challenge. *IEEE Trans. Med. Imaging*, 2011.
- [MWZJL16] S. Miao and R. Wang Z. J. and Liao. A cnn regression approach for real-time 2d/3d registration. *IEEE Trans. Medical Imaging*, 35(5):1352–63, 2016.
- [NHP⁺11] M. Niethammer, G. L. Hart, D. F. Pace, P. M. Vespa, A. Irimia, J. D. Horn, and S. R. Aylward. Geometric metamorphosis. In *Proc. MICCAI*, pages 639–646, 2011.
- [NS06] Vivi Nastase and Stan Szpakowicz. A study of two graph algorithms in topic-driven summarization. pages 29–32, 07 2006.
- [Ons44] L. Onsager. A two-dimensional model with an order-disorder transition. *Physical Review*, 3&4(65):117–49, 1944.
- [PARP06] X. Pennec, N. Ayache, A. Roche, and P. Cachier. Non-rigid MR/US registration for tracking brain deformations. In R. S. Blum and Z. Liu, editors, *Multi-Sensor Image Fusion and Its Applications*. CRC Press, Boca Raton, FL, first edition, 2006.
- [Pen06] Xavier Pennec. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25:127–154, 2006.
- [Pen09] Xavier Pennec. Statistical computing on manifolds: From riemannian geometry to computational anatomy. *Emerging Trends in Visual Computing*, 5416:347–386, 2009.
- [PM90] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.
- [Pol60] M Pollack. The maximum capacity through a network. *Operations research*, 8:733–36, 1960.
- [Pre07] C. Preda. Regression models for functional data by reproducing kernel hilbert spaces methods. *Journal of statistical planning and inference*, 137(3):829–40, 2007.
- [PSV07] R. Pastor-Satorras and A. Vespignani. Evolution and structure of the internet : A statistical physics approach. *Cambridge University Press*, 2007.

- [pyt]
- [RCSO⁺04] Anil Rao, Raghavendra Chandrashekara, Gerardo I. Sanchez-Ortiz, Raad Mohiaddin, Paul Aljabar, Joseph V. Hajnal, Basant K. Puri, and Daniel Rueckert. Spatial transformation of motion and deformation fields using nonrigid registration. *IEEE TMI*, 23(9):1065–1076, 2004.
- [RHW86] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–36, 1986.
- [RSH⁺99] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Trans. Med. Imaging*, 18(8):712–721, 1999.
- [SADDdS15] Camille Sutour, Jean-François Aujol, Charles-Alban Deledalle, and Baudouin Denis de Senneville. Edge-based multi-modal registration and application for night vision devices. *Journal of Mathematical Imaging and Vision*, 53(2):131–150, 2015.
- [SDP13] Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging*, 32(7):1153 – 1190, 2013.
- [Sei83] S Seidman. Network structure and minimum degree. *Social Networks*, 5:269–87, 1983.
- [Set96] J. A Sethian. A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences*, 93(4):1591–1595, 1996.
- [SFP⁺10] Nikhil Singh, P. Thomas Fletcher, J. Samuel Preston, Linh Ha, Richard King, J. Stephen Marron, Michael Wiener, and Sarang Joshi. Multivariate statistical analysis of deformation momenta relating anatomical shape to neuropsychological measures. In *Proc. MICCAI*, volume 6363, pages 529–537, 2010.
- [SGMWLJ⁺05] Mueller S. G., Weiner M. W., Thal L. J., Petersen R. C., Jack C. R., Jagust W., Trojanowski J. Q., Toga A. W., and Beckett L. Ways toward an early diagnosis in alzheimer’s disease: the alzheimer’s disease neuroimaging initiative (adni). *Alzheimers Dement.*, 1:55–66, 2005.
- [SHS01] Bernhard Schölkopf, Ralf Herbrich, and Alex J. Smola. A generalized representer theorem. *Computational Learning Theory. Lecture Notes in Computer Science.*, 2111(2):416–426, 2001.
- [SLNP12] Stefan Sommer, Francois Lauze, Mads Nielsen, and Xavier Pennec. Sparse multi-scale diffeomorphic registration: The kernel bundle framework. *J. of Mathematical Imaging and Vision*, 46, 07 2012.

- [SSZ⁺98] P Spellman, G Sherlock, M Zhang, I Vishwanath, K Anders, M Eisen, P Brown, D Botstein, and Futcher B. Comprehensive identification of cell-cycle-regulated genes of yeast *saccharomyces cerevisiae* by microarray hybridization. *Molecular biology of the cell*, 9:3273–3297, 1998.
- [TFP⁺06] B. Thirion, G. Flandin, P. Pinel, A. Roche, P. Ciuciu, and J.B. Poline. Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fmri datasets. In *Hum. Brain Mapp.*, volume 27, page 678–693, 2006.
- [Thi96] Jean-Philippe Thirion. Non-rigid matching using demons. In *Proc. CVPR*, page 245, 1996.
- [Tib94] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–88, 1994.
- [TIP08] Adrien Trillon, Jérôme Idier, and Pierre Peureux. Unsupervised Bayesian 3D reconstruction for non-destructive evaluation using gammagraphy. In *EUSIPCO*, Lausanne, Suisse, August 2008.
- [TMK11] M. Taquet, B Macq, and Warfield S. K. Spatially adaptive log-euclidean polyaffine registration based on sparse matches. In *Proceedings of MICCAI*, volume 14, page 590–597, 2011.
- [TMT02] C. Twinings, S. Marsland, and C. Taylor. Measuring geodesic distances on the space of bounded diffeomorphisms. In *British Machine Vision Conference*, 2002.
- [TQ18] Mingzhen Tan and Anqi Qiu. Multiscale frame-based kernels for large deformation diffeomorphic metric mapping. *IEEE Transactions on Medical Imaging*, PP:1–1, 05 2018.
- [Tro95] Alain Trouvé. Action de groupe de dimension infinie et reconnaissance de formes. *C. R. Acad. Sci. Paris Sér. I Math.*, 321(8):1031–1034, 1995.
- [TSR⁺05] Robert Tibshirani, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society Series B*, pages 91–108, 2005.
- [TU00] P. Thévenaz and M. Unser. Optimization of mutual information for multiresolution image registration. *IEEE Trans. Imag. Proc.*, 9(12):2083–2099, 2000.
- [TY05] A. Trouvé and L. Younes. Metamorphoses through Lie group action. *Foundations of Computational Mathematics*, 5(2):173–198, 2005.

- [VCI07] T. Vincent, P. Ciuciu, and J. Idier. Spatial mixture modelling for the joint detection-estimation of brain activity in fMRI. In *proc. ICASSP*, volume I, pages 325–328, Honolulu, Hawaii, Apr. 2007.
- [VMK⁺16] Max A. Viergever, J.B. Antoine Maintz, Stefan Klein, Keelin Murphy, Marius Staring, and Josien P.W. Pluim. A survey of medical image registration – under review. *Medical Image Analysis*, 33:140 – 144, 2016.
- [VPPA08] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Symmetric log-domain diffeomorphic registration: A demons-based approach. In *Proc. MICCAI*, pages 754–761, 2008.
- [Wei96] Joachim Weickert. *Anisotropic Diffusion In Image Processing*. 1996.
- [Whi76] W. Whitt. Bivariate distributions with given marginals. *The Annals of statistics*, 4(6):1280–89, 1976.
- [YAM09] L. Younes, F. Arrate, and M. I. Miller. Evolutions equations in computational anatomy. *NeuroImage*, 45(1, Supplement 1):S40 – S50, 2009. Mathematics in Brain Imaging.
- [YMSM08] L. Younes, P. Michor, J. Shah, and D. Mumford. A metric on shape spaces with explicit geodesics. *Rend. Lincei Mat. Appl.*, 9:25–57, 2008.
- [You07] L. Younes. Jacobi fields in groups of diffeomorphisms and applications. *Quart. Appl. Math.*, 65:113–134, 2007.
- [You08] L. Younes. *Shapes and Diffeomorphisms*. Springer, 2008.
- [You12] Laurent Younes. Constrained diffeomorphic shape evolution. *Foundations of Computational Mathematics*, 12(3):295–325, Jun 2012.
- [ZH05] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67:301–320, 2005.