Diffeomorphisms, Optimal Transport and Applications to Imaging

François-Xavier Vialard

UNIVERSITÉ PARIS-DAUPHINE
CEREMADE

MÉMOIRE D'HABILITATION À DIRIGER LES RECHERCHES

Diffeomorphisms, Optimal Transport and Applications to Imaging

---

présenté par

François-Xavier VIALARD

spécialité: Mathématiques

coordinateur des travaux: Jean-David Benamou

soutenu le 8 décembre 2017 après avis de

Michor Peter	Professeur à l'université de Vienne
Mielke Alexander	Professeur au Weirstrass Institute, Berlin
Younes Laurent	Professeur à l'université Johns Hopkins, Baltimore

devant le jury composé de :

Benamou Jean-David	Directeur de recherche à INRIA, Paris
Brenier Yann	Directeur de recherche au CNRS, École Polytechnique
Glass Olivier	Professeur à l'université Paris Dauphine
Masnou Simon	Professeur à l'université de Lyon 1
Michor Peter	Professeur à l'université de Vienne
Trouvé Alain	Professeur à l'ENS Cachan
Younes Laurent	Professeur à l'université Johns Hopkins, Baltimore.

Université Paris-Dauphine

Abstract

MIDO
Ceremade

Habilitation à diriger les recherches

Diffeomorphisms, Optimal Transport and Applications to Imaging

by François-Xavier VIALARD

In this document, we present our past five years research activities on diffeomorphic image matching, optimal transport and related mathematical questions. Most of our work is motivated by applications to shape analysis and in particular to medical imaging data. We aim at developing well-posed mathematical models and methods which are computationally feasible on real data in order to address concrete problems of interests. Thus, we are also involved in the software implementation of these methods, which are freely available.

In the introduction, we present a unified point of view between two areas in shape analysis, namely shape matching using a diffeomorphism group endowed with a right-invariant metric and optimal transport.

The first part of the document concerns the mathematical foundations of the so-called field of Computational Anatomy in medical imaging, and more generally in shape analysis. The questions we are interested in are often of analytical and variational nature. Our most important achievement in this direction is the geodesic completeness result of the group of diffeomorphisms endowed with a right-invariant Sobolev metric of high enough order. An other variational problem of interest is the minimization of the acceleration, for which we compute the tight relaxation in a particular case.

The second part of the document presents the natural extension of the L^2 Wasserstein metric to the case of unbalanced (nonnegative measures), that we called Wasserstein Fisher-Rao. We introduce it via its so-called Benamou and Brenier formulation, then we present the particularly simple equivalent static formulation and efficient numerical algorithms based on entropic regularization. We also draw the link between the Camassa-Holm equation and the Wasserstein-Fisher-Rao metric which is similar to the link Brenier has pioneered between optimal transport and the incompressible Euler equation. A surprising consequence is the new formulation of the Camassa-Holm equation as an incompressible Euler equation for a singular density.

The last part addresses the actual implementation of the framework of diffeomorphic registration and in particular the crucial choice of the right-invariant metric. Interestingly, we give a mathematical framework for each of the methods we proposed. In the same context, we show the efficiency of a new similarity measure between surfaces or curves based on unbalanced optimal transport in comparison with kernels methods. On the extension of diffeomorphic registration, we introduced a shooting algorithm on the space of images which allows to perform geodesic regression on time sequences of images. Last, we used this shooting algorithm to perform longitudinal shape data classification based on the shooting algorithm applied to the Alzheimer disease.

Remerciements

Tout d'abord, je suis très reconnaissant envers Peter Michor, Alexander Mielke et Laurent Younes d'avoir accepté de donner leur avis d'expert sur mes travaux de recherche résumés dans ce manuscrit. Je remercie en particulier Peter Michor et Laurent Younes pour leur présence en tant que membre du jury. Je remercie beaucoup Jean-David Benamou, non seulement pour avoir pris le soin de coordonner cette habilitation, mais aussi pour des interactions scientifiques fécondes: en effet, une partie des travaux présentés dans ce manuscrit a été développée à la suite d'une discussion autour du transport optimal. La présence d'Alain Trouvé dans le jury me fait très plaisir car sa démarche scientifique profonde, que j'ai pu connaître pendant ma thèse, reste une ligne directrice forte dans mes propres travaux. Pour avoir accepté de faire partie du jury, je remercie très chaleureusement Yann Brenier, Olivier Glass et Simon Masnou. Certains développements dans ce manuscrit sont reliés à leurs travaux antérieurs, parfois très étroitement, pour le transport optimal par exemple, et en ont bénéficié.

En dehors de l'activité de recherche elle-même, j'apprécie beaucoup le travail de recherche en collaboration. En particulier, mes travaux doivent beaucoup à mes nombreux collaborateurs, desquels j'ai parfois beaucoup appris. Ces nombreuses collaborations s'expliquent par un environnement de recherche dynamique, et ma participation à de nombreuses conférences. Je tiens à remercier sans ordre particulier, Darryl D. Holm, Martin Bauer, Martins Bruveris, Klas Modin, Stefan Sommer, Xavier Pennec, Gabriel Peyré, Stanley Durrleman pour leur implication dans l'organisation de ces conférences, qui étaient indispensables pour le développement de mes travaux.

J'ai aussi beaucoup profité au Ceremade d'un environnement propice, notamment sur des thématiques proches, du à la présence de Gabriel Peyré et Laurent Cohen, et par la suite de Jean-David Benamou et Guillaume Carlier. Je remercie beaucoup Gabriel qui a toujours été une force active pour la vie scientifique du laboratoire par l'organisation du groupe de travail stat-images par exemple. Je le remercie aussi très chaleureusement pour son ouverture à la collaboration, j'ai beaucoup appris à son contact, par l'intérmédiaire de l'encadrement de post-doc, comme Giacomo Nardi ou Bernhard Schmitzer et l'encadrement de la thèse de Lenaïc Chizat. Ce mémoire lui doit beaucoup. La création de l'équipe projet Inria Mokaplan, initiée par Jean-David Benamou, Guillaume Carlier et Gabriel Peyré a été un élément important pour conserver une dynamique forte autour des thématiques du transport optimal et d'imagerie. Je les remercie de m'y avoir associé.

Enfin, je remercie mes collègues du Ceremade et du département d'enseignement Mido pour un environnement de travail toujours agréable: plus particulièrement, Guillaume Legendre, mon voisin de bureau, qui supporte gentiment d'écouter mes nouveaux sujets d'excitations mathématiques.

Sans aucun doute, ma reconnaissance et mes remerciements les plus profonds vont à ma famille, mes parents et ma femme, qui m'ont toujours soutenu et aidé; ils ont motivé l'écriture de ce manuscrit.

Contents

Abstract **iii**

1 Introduction **1**
1.1 Motivation 1
1.2 A geometric framework for diffeomorphic image matching and optimal transport 3
1.3 Geodesic equations 10
1.4 Contributions 14

2 Variational problems on diffeomorphisms and shape spaces **17**
2.1 Introduction 17
2.2 Completeness results on the group of diffeomorphisms 19
2.3 Minimization of the acceleration 21
2.4 Boundary value problem on the space of curves 26

3 Unbalanced optimal transport and the Camassa-Holm equation **27**
3.1 Introduction 27
3.2 The Wasserstein-Fisher-Rao metric 29
3.3 Entropic regularization and scaling algorithms 35
3.4 Perspectives 38
3.5 From the Wasserstein-Fisher-Rao metric to the Camassa-Holm equation 39

4 Applications to diffeomorphic image matching **45**
4.1 Introduction 45
4.2 Sum of kernels and semidirect product of groups 45
4.3 Left-invariant metrics 48
4.4 Learning the metric 50
4.5 Optimal transport for diffeomorphic matching 51
4.6 Geodesic regression 52
4.7 Longitudinal deformation model for Alzheimer disease 53

Chapter 1

Introduction

1.1 Motivation

This document is concerned with mathematical questions in shape analysis and their applications in image registration and Computational Anatomy. More precisely, our work deals with the development of shape matching with medical imaging as target applications. We study in particular two complementary models of shape matching; the first one is concerned with the matching of any shapes or objects of the same class that can be embedded in the Euclidean space and the second one deals with the particular case when shapes are measures, actually nonnegative measures. Although some of the results are theoretical, for instance on infinite dimensional Riemannian geometry or optimal transport, these developments were essentially motivated by medical image registration and the so-called emerging domain of Computational Anatomy, a rather informative webpage is available at https://en.wikipedia.org/wiki/Computational_anatomy.

One goal of Computational Anatomy consists in developing quantitative tools to analyse and quantify the statistical variability of anatomical shapes and help the practitioner with the diagnosis. Underlying this approach, there is a constitutive hypothesis which is that pathologies can be detected out of images of organs. Let us mention, as a convincing example, the case of the Alzheimer disease which entails a decay of the hippocampi (some regions of interest in the brain). Going beyond the change of global indicators (such as volume) by using the whole geometrical information might improve the results of statistical studies, as shown for instance in [THD+14, LAFP11].

The mathematical modeling of shapes is far from being new and d'Arcy Thompson was probably the first to introduce the idea of studying their variability through an underlying *deformation* of a template (an average shape), see [Boo76]. Due to the increase of medical imaging data in the last twenty years, there was a need to develop quantitative methods applying such kind of ideas. To this end, Grenander laid down the foundations of pattern theory [Gre93, GM98]. It was further theoretically developed and also numerically implemented by Miller, Trouvé and Younes and collaborators [Tro95, BMTY05, JM00, MTY06, YAM09]. In this introduction, we give a short introduction to this area from a mathematical point of view.

Even before being interested by any statistical description of shape variability, a problem of interest in medical imaging consists in registering two biomedical images. The main application is to establish correspondences between different image acquisitions. In order to make the problem concrete, we instantiate it in more mathematical terms below, presented in the case of images but it extends readily to the case of any embedded object in the Euclidean space: An image can be modeled as a scalar valued function defined on a domain (smooth domain) $\Omega \subset \mathbb{R}^d$ with $d = 2, 3$. Finding correspondences between two images may or may not be symmetric with

respect to the given images (for example, one image is noisy and the other is not so that their role in the registration is not symmetric).Thus, standard approaches for registration distinguish the given images: there is the source image I that will be deformed by an invertible map $\varphi : \mathbb{R}^d \to \mathbb{R}^d$, the deformed image is $I \circ \varphi^{-1}$. The other image J will be called the target image. For concrete applications in biomedical imaging, constraints have to be imposed on the transformation φ. A usual constraint is to require that φ is invertible and preserves the topology: it is a priori difficult to make sense of folding or collapsing tissues. In particular, diffeomorphisms of the domain satisfy this condition. Due to noise or differences in acquisition modalities, there is in general no diffeomorphic transformation φ such that the deformed source image denoted by $\varphi \cdot I$ is equal to J. And if, by chance, there exists such a transformation, it need not be unique. From a computational viewpoint, solving for φ the equation $I \circ \varphi^{-1} = J$ is usually done via an optimization problem on the set of transformations denoted by $\mathcal{S}$:

$$\operatorname{argmin}_{\varphi \in \mathcal{S}} d(I \circ \varphi^{-1}, J)$$

where d is a function satisfying $d(I, J) = 0 \implies I = J$, such as a distance. In general, the first remark on the existence of φ implies that this minimization does not have a solution in $\mathcal{S}$ but this solution might exist in a bigger space. A practical solution for ill-posed inverse problems consists in adding a regularization term on the deformation, that penalizes "weird" deformations. The new minimization problem is then

$$\mathcal{J}(\varphi) = R(\varphi) + d(I \circ \varphi^{-1}, J)\,, \tag{1.1.1}$$

where $R(\varphi)$ is the "cost" associated with the deformation φ and controls its smoothness. This variational problem is most of the time non-convex. With respect to the previous issues (existence and uniqueness), existence is obtained and also, most of the time, a generic uniqueness result holds. Most of the available registration methods can be formulated as the minimization of a functional of type (1.1.1). Remark that, on purpose, we did not specify the domain for the optimization of the functional $\mathcal{J}$. Indeed, at this level of generality, one can incorporate the domain constraint in the function R, e.g. $R(\varphi)$ if $\varphi \notin \mathcal{S}$. In particular, the function R may include a prior knowledge on the space of admissible deformations, as well as statistical informations on the deformations: An idealized and informal situation would be that the distribution of deformations among a population is known a priori so that it makes sense to use $R(\varphi) = -\log(p(\varphi))$, which would correspond to log-likelihood maximization when $p(\varphi)$ is the probability of occurrence of φ.

In the situation where no prior information is available and the only constraint is the diffeomorphic one, different methods have been proposed to satisfy it: generate diffeomorphisms (1) using the Lie-exponential, i.e. flows of vector fields constant in time and (2) using time-dependent vector fields.
For (1), the Lie-exponential introduces a regularization on the Lie algebra on the velocity field

$$R(\varphi) := \frac{1}{2}\|v\|_V^2$$

where V is a Hilbert space of smooth vector fields and φ is determined by v by the following autonomous ode: $\partial_t \varphi(t) = v \circ \varphi(t)$. It is well-known that the Lie exponential is not surjective in infinite dimension: an example is given in [KW09]

on $\mathrm{Diff}(S_1)$.
For (2), the regularization is given by

$$R(\varphi) := \inf_{v_t} \frac{1}{2} \int_0^1 \|v_t\|_V^2 \, \mathrm{d}t$$

where $\partial_t \varphi(t) = v_t \circ \varphi(t)$ is the flow equation that determines φ for a given time dependent vector field v_t (the subscript t stands for the time variable).

While the latter approach is more computationnally demanding than the former, yet it enjoys more mathematical properties that are of interest for applications, in particular it provides a Riemannian framework. This framework is often called *Large Deformation Diffeomorphic Metric Mapping* (LDDMM) developed by the work of Dupuis, Grenander, Miller, Trouvé and Younes and many others. The Riemannian framework enables the generalization of simple tools from geometry, probability and statistics such as distance, geodesic, linear regression, cubic splines, principal component analysis, etc... which have been developed by many authors , including us, in the last ten years.

In the functional (1.1.1), the data fidelity denoted by d actually has an important impact on the minimizers as well as on the actual computed solutions, whatever the minimization scheme used in practice. These data fidelities are often simple to compute such as the square of the L^2 difference between the target image and the deformed initial image, called sum of squared differences (SSD) or such as fidelities built on cross-correlation quantities. Ideally, this data fidelity should be fast to compute as well as its gradient, convex and should vanish at its minimum when the two images are equal. When the data of interest is sparse, like a surface or a curve in the Euclidean space, data fidelities using kernel metrics have been proposed. These methods consists in representing the data as a measure, current or varifold and using some "weak" dual norm defined by a smoothing kernel. Actually, these data fidelity are used in many inverse problems in imaging and new developments in this area can have a substantial impact outside diffeomorphic image matching.

This document presents a selection of our contributions in the area of shape analysis and in particular in Computational Anatomy. These contributions range from the theoretical analysis of the existing framework, its development and its applications to real data. Before presenting them, we give a brief introduction to our area of research.

1.2 A geometric framework for diffeomorphic image matching and optimal transport

In this section, we introduce the LDDMM framework, optimal transport and we give a geometric picture in which the two can be understood.

1.2.1 A brief introduction to the LDDMM framework

The basic construction of the LDDMM framework consists in the diffeomorphic registration problem which aims at finding an optimal diffeomorphic map of the ambient space that maps a source shape or image onto a target one. It thus takes the form of an infinite dimensional optimal control problem: Minimize the cost functional

$$\mathcal{L}(\xi) = \int_0^1 \|\xi(t)\|_V^2 \, \mathrm{d}t + S(\varphi \cdot q) \tag{1.2.1}$$

under the constraints

$$\partial_t \varphi(t,x) = \xi(t,\varphi(t,x)) \tag{1.2.2}$$
$$\varphi(0,x) = x \ \forall x \in D\,, \tag{1.2.3}$$

and V is a Hilbert space of vector fields on a Euclidean domain $D \subset \mathbb{R}^d$, left invariant by their flows, such that the inclusion map $V \hookrightarrow W^{1,\infty}(D,\mathbb{R}^d)$ is continuous. That is, the norm on V controls the $W^{1,\infty}$ norm, we call such a space V an *admissible* space of vector fields. In particular, these spaces are included in the family of reproducing kernel Hilbert spaces (RKHS), [Aro50] since the pointwise evaluation is a continuous linear map, which implies that such spaces are completely defined by their kernel. In the rest of this document, we denote this kernel k and the isomorphism from V to its dual V^* by K.

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, see [You08, Appendix C]. Then, the set of flow at time 1 defines a group of diffeomorphisms denoted by $\mathcal{G}_V$. Denoting

$$\mathrm{Fl}_1(\xi) = \varphi(1) \text{ where } \varphi \text{ solves (1.2.2)}\,, \tag{1.2.4}$$

define

$$\mathcal{G}_V \overset{\text{def.}}{=} \{\varphi(1) \,:\, \exists \xi \in L^2([0,1],V) \text{ s.t. } \mathrm{Fl}_1(\xi)\}\,, \tag{1.2.5}$$

which has been introduced by Trouvé in [Tro95]. On this group, Trouvé defines a metric

$$\mathrm{dist}(\psi_1,\psi_0)^2 = \inf\left\{\int_0^1 \|\xi\|_V^2 \,\mathrm{d}t \,:\, \xi \in L^2([0,1],V) \text{ s.t. } \psi_1 = \mathrm{Fl}_1(\xi)\circ\psi_0\right\} \tag{1.2.6}$$

under which he proves that $\mathcal{G}_V$ is complete. In full generality, that is for a general space of vector fields V, very few properties are known on this group. For instance, it is a priori not a topological group, or more precisely, there is no known topological structure making it a topological group. Moreover, there does not need to be a differentiable structure on this group. The availability of these additional properties crucially depends on the space V. A remarkable property of this distance is that it is right-invariant which means for every $\psi_1,\psi_2,\psi_3 \in \mathcal{G}_V$, one has

$$\mathrm{dist}(\psi_1\circ\psi_3,\psi_0\circ\psi_3) = \mathrm{dist}(\psi_1,\psi_0)\,. \tag{1.2.7}$$

Instead of formulating the variational problem on the group of diffeomorphisms $\mathcal{G}_V$, it is often possible to rewrite (under mild conditions) the optimization problem on the manifold Q. In the case when Q is "smaller" than the group of diffeomorphisms, for instance the space of landmarks, it is of interest for computational purposes. For a shape q that is embedded in the domain D, this optimal control problem can be rewritten in the following form that can be of interest to reduce the dimension of the optimization set.

$$\mathcal{L}(\xi) = \int_0^1 \|\xi(t)\|_V^2 \,\mathrm{d}t + S(q(1)) \tag{1.2.8}$$

under the constraints

$$\partial_t q(t,x) = \xi(t,q(t,x))$$
$$q(0,x) = q_0(x) \ \forall x \in D\,.$$

In the case of an image $I : D \mapsto \mathbb{R}$, the constraint is replaced by the advection equation $\partial_t I + \langle \nabla I, \xi \rangle = 0$. We will see later in the introduction how the optimization set can be reduced to a space which has the dimensionality of Q.

Importantly, under mild conditions, it is possible to prove that this functional induces a Riemannian metric on the orbit of an object q when Q is a finite dimensional manifold. Thus, a right-invariant metric on a group G which has a left action on a manifold Q has an induced metric for which the action is a Riemannian submersion. The Riemannian structure on such a manifold Q, thought of as the space of shapes, is of interest in order to extend natural tools from Euclidean geometry such as distance, principal component analysis (PCA), linear regression, cubic splines, parallel transport, etc... We contributed to such extensions in [NHV11] for extending geodesic regression on the orbit of images and the introduction of cubic splines in [VT12, SVN15]. This last extension has motivated some theoretical developments presented in this document.

1.2.2 A brief introduction to optimal transport

Optimal transport is a way to lift a metric on a space M, a Riemannian manifold for instance endowed with its induced metric, to a metric on the space of probability densities. In this section, we only present the formulations used in this document, for a rather complete review on optimal transport we refer to the gigantic monograph of Villani [Vil08].

We present two equivalent formulations of optimal transport. We call the first formulation "static" since time does not intervene, in contrast to the second formulation which we call "dynamic".

Static formulation of optimal mass transport: The optimal mass transport problem as introduced by Monge in 1781 consists in finding, between two given probability measures ν_1 and ν_2, a map $\varphi : M \mapsto M$ such that $\varphi_* \nu_1 = \nu_2$, i.e. the image measure of ν_1 by φ is equal to ν_2 and which minimizes a cost given by

$$\int_M c(x, \varphi(x)) \, \mathrm{d}\nu_1(x) \,, \tag{1.2.9}$$

where c is a positive function that represents the cost of moving a particule of unit mass from location x to location y. This problem is ill-posed in the sense that solutions may not exist and the Kantorovich formulation of the problem is the correct relaxation of the Monge formulation, which can be presented as follows: On the space of probability measures on the product space $M \times M$, denoted by $\mathcal{P}(M \times M)$, find a minimizer to

$$\mathcal{I}(m) = \int_{M^2} c(x, y) \, \mathrm{d}m(x, y) \text{ such that } p^1_*(m) = \nu_1 \text{ and } p^2_*(m) = \nu_2 \,, \tag{1.2.10}$$

where $p^1_*(m), p^2_*(m)$ denote respectively the image measure of $m \in \mathcal{P}(M \times M)$ under the projections on the first and second factors on $M \times M$. Most often in the litterature, the cost c is chosen as a power of a distance. From now on, we will only discuss the case $c(x, y) = d(x, y)^2$ where d is the distance associated with a Riemannian metric on M. In this case, the Kantorovich minimization problem defines W_2, the so-called L^2-Wasserstein distance on the space of probability measures which metrizes the weak convergence of measures, as well as other L^p-Wasserstein distances. The Monge formulation can be expressed as a minimization problem as follows, for μ, ν

being absolutely continuous w.r.t. the Riemannian volume measure,

$$W_2(\mu,\nu)^2 \stackrel{\text{def.}}{=} \inf_{\varphi\in \mathrm{Diff}(M)} \left\{ \int_M d(\varphi(x),x)^2 \, \mathrm{d}\nu_1(x) \, : \, \varphi_*\nu_1 = \nu_2 \right\} , \tag{1.2.11}$$

where $\mathrm{Diff}(M)$ denotes the group of smooth diffeomorphisms of M. Usually, the Monge formulation is rather introduced with maps instead of diffeomorphisms, which is equivalent in this case.

Dynamic formulation: In [BB00], Benamou and Brenier introduced a dynamical version of optimal transport which was inspired and motivated by the study of the incompressible Euler equation. Let $\rho \in C^\infty(M,\mathbb{R}_+)$ be a positive function, note that all the quantities will be implicitly time dependent. The dynamic formulation of the Wasserstein distance consists in minimizing

$$\mathcal{E}(\xi) = \int_0^1 \int_M \|\xi(t,x)\|^2 \rho(t,x) \, \mathrm{dvol}(x) \, \mathrm{d}t \ , \tag{1.2.12}$$

subject to the constraints $\dot\rho + \mathrm{div}(\rho\,\xi) = 0$ and initial and final condition $\rho(0) = \rho_0$, $\rho(1) = \rho_1$. The notation $\|\cdot\|$ stands for the Euclidean norm.

Equivalently, following [BB00], a convex reformulation using the momentum $\mathsf{m} = \rho\,\xi$ reads

$$\mathcal{E}(\mathsf{m}) = \int_0^1 \int_M \frac{\|\mathsf{m}(t,x)\|^2}{\rho(t,x)} \, \mathrm{dvol}(x) \, \mathrm{d}t \, , \tag{1.2.13}$$

subject to the constraints $\dot\rho + \mathrm{div}(\mathsf{m}) = 0$ and the corresponding initial and final conditions on the density ρ. Let us underline that the functional $\mathcal{E}$ is convex in ρ, m and the continuity equation is linear in (ρ,m), therefore convex optimization methods can be applied for numerical simulations. Due to the continuity equation, the problem is feasible (or controllable) if and only if the initial and final densities have the same total mass using Moser's lemma [Mos65].

1.2.3 A common geometric framework

In this section, we present a geometric construction which is common to LDDMM and optimal transport. This is based on a result written in [Mic08, Claim of Section 29.21] infinite dimensions and we do not discuss here the complications arising from the infinite dimensional setting of shape spaces. We first introduce the notion of a group action.

Group actions. Let G_V be a group which is a smooth manifold with a tangent space at the identity Id denoted by V, acting from the *left* on a smooth manifold Q. We denote the action by

$$\Phi : G \times Q \to Q \, , \quad (g,q) \mapsto g\cdot q := \Phi_g(q). \tag{1.2.14}$$

Being a left action means that Φ satisfies the composition law $g_1\cdot(g_2\cdot q) = (g_1 g_2)\cdot q$ and $\mathrm{Id}\cdot q = q$ for any $q\in Q$ and $g_1, g_2 \in G$.

Definition 1. The *infinitesimal generator* of the action corresponding to $\xi \in V$ is the vector field on Q given by

$$\xi_Q(q) := \left.\frac{d}{dt}\right|_{t=0} \exp(t\xi)\cdot q \, . \tag{1.2.15}$$

We will often use the notation $\xi \cdot q$ instead of $\xi_Q(q)$.

We also define the right-trivialization map.

Definition 2 (Right-trivialization)**.** Let H be a group and a smooth manifold at the same time, possibly of infinite dimensions, the right-trivialization of TH is the bundle isomorphism $\tau : TH \mapsto H \times T_{\mathrm{Id}}H$ defined by $\tau(h, X_h) \overset{\text{def.}}{=} (h, dR_{h^{-1}}X_h)$, where X_h is a tangent vector at point h and $\mathcal{R}_{h^{-1}} : H \to H$ is the right multiplication by h^{-1}, namely, $R_{h^{-1}}(f) = fh^{-1}$ for all $f \in H$.

In fluid dynamics, the right-trivialized tangent vector $dR_{h^{-1}}X_h$ corresponds to the spatial or Eulerian velocity and X_h is the Lagrangian velocity. Importantly, on the group of Sobolev diffeomorphisms (defined in Section 2), this right-trivialization map is continuous but not differentiable with respect to the variable h. Indeed, right-multiplication R_h is smooth, yet left multiplication is continuous and usually not differentiable, due to a loss of smoothness.

This result presents a standard construction to obtain Riemannian submersions from a transitive group action (in finite dimensions).

Proposition 3. *Let G be a Lie group, whose tangent space at identity is denoted by V. Consider a smooth left action of G onto a smooth manifold Q, which is transitive and such that, for every $q \in Q$, the infinitesimal action $\xi \mapsto \xi \cdot q$ is a surjective map. Let $g_V : Q \to V^* \otimes V^*$ be a map such that for every $q \in Q$, $g_V(q)$ is a positive definite quadratic form. Let $q_0 \in Q$ be a "reference" point. One defines the Riemannian metric g_G on G by*

$$g_G(h)(X_h, X_h) = g_V(h \cdot q_0)(dR_{h^{-1}}X_h, dR_{h^{-1}}X_h)\,. \tag{1.2.16}$$

Let $X_q \in T_qQ$ be a tangent vector at point $h \cdot q_0 = q \in Q$, one defines the Riemannian metric g_Q on Q by

$$g_M(q)(X_q, X_q) \overset{\text{def.}}{=} \min_{\xi \in T_{\mathrm{Id}}G} g(q)(\xi, \xi) \text{ under the constraint } X_q = \xi \cdot q\,, \tag{1.2.17}$$

where $\xi = X_h \cdot h^{-1}$.

Then, the map $\pi_0 : G \to Q$ defined by $\pi_0(h) = h \cdot q_0$ is a Riemannian submersion of the metric g_G on G to the metric g_Q on Q.

We now formally apply the previous result to the infinite dimensional cases of optimal transport and LDDMM.

Example 4 (LDDMM)**.** *The LDDMM framework uses a single scalar product on the tangent space at identity to define the metric $g(\varphi \cdot q_0)(\xi, \xi) = \|\xi\|_V^2$. Thus, the metric on the group of diffeomorphisms is $g(\varphi)(X_\varphi, X_\varphi) = \|X_\varphi \circ \varphi^{-1}\|_V^2$, which is a right-invariant metric.*

Example 5 (Optimal transport)**.** *The Benamou-Brenier formula [BB00] reformulates optimal transport as an optimal control problem on the space of probability densities which takes the form written in Proposition 3.*

The group G is the group of diffeomorphisms $\mathrm{Diff}(M)$ of a given closed manifold M and the manifold Q which is acted upon is the affine space of probability densities on M. The action is the pushforward of a density by an element of the group. We fix a reference probability measure μ_0 and we consider the space of probability densities with respect to this reference measure, that we denote by $\mathrm{Dens}_p(M)$. That is, the left group action

$$\pi : \mathrm{Diff}(M) \times \mathrm{Dens}_p(M) \to \mathrm{Dens}_p(M) \tag{1.2.18}$$

is defined by

$$\pi(\varphi, \rho) = \varphi_*(\rho) = \mathrm{Jac}(\varphi^{-1})\rho(\varphi^{-1})\,. \tag{1.2.19}$$

The infinitesimal action defined in (1.2.15) *reads, by a straightforward computation,*

$$\xi \cdot \rho = -\operatorname{div}(\rho\,\xi)\,, \tag{1.2.20}$$

and the metric on the group of diffeomorphisms is given by

$$g(\varphi)(\xi, \xi) = \int_M |\xi(x)|^2 \rho(x)\,\mathrm{d}\mu_0(x)\,, \tag{1.2.21}$$

where $\rho(x) = \frac{\mathrm{d}\varphi_*\mu_0}{\mathrm{d}\mu_0}$.

Since $\xi = X_\varphi \circ \varphi^{-1}$, by the change of variable $y = \varphi^{-1}(x)$, the metric on the group of diffeomorphisms reads

$$g(\varphi)(X_\varphi, X_\varphi) = \int_M h(X_\varphi(y), X_\varphi(y))\,\mathrm{d}\mu_0(y)\,, \tag{1.2.22}$$

which is an L^2 metric on the group of diffeomorphisms. Note that for this L^2 metric to be well defined, one needs a Riemannian metric, denoted by h, on the codomain and a volume measure on the domain, μ_0, in this case.

Thus, Proposition 3 applies to this case and we retrieve the submersion introduced by Otto [Ott01], which shows that the map

$$\begin{aligned}&\pi : \mathrm{Diff}(M) \to \mathrm{Dens}_p(M)\\ &\pi(\varphi) = \varphi_*(\rho_0)\end{aligned}$$

is a formal Riemannian submersion of the metric $L^2(\rho_0)$ on $\mathrm{Diff}(M)$ to the L^2-Wasserstein metric on $\mathrm{Dens}_p(M)$, which is illustrated in Figure 1.1.

Example 6 (LDDMM). *The LDDMM framework is based on the action of a group of diffeomorphisms of a Euclidean space (or domain denoted by* $\Omega \subset \mathbb{R}^d$*) and a left action of this group of diffeomorphisms on a space of objects that can be images (or* 0*-forms in mathematical terms), surfaces or lines in this domain. For instance, in the case of an image, denoting by* $I : \Omega \mapsto \mathbb{R}$ *an image, the action of a diffeomorphism and the infinitesimal action are*

$$\begin{aligned}\varphi \cdot I &= I \circ \varphi^{-1}\\ \xi \cdot I &= -\langle \nabla I, \xi \rangle\end{aligned}$$

The boundary value problem in LDDMM is thus the following

$$\mathcal{L}(\xi) = \int_0^1 \|\xi(t)\|_V^2\,\mathrm{d}t \tag{1.2.23}$$

under the constraints

$$\begin{aligned}&\xi \cdot I = -\langle \nabla I, \xi \rangle\\ &I(0, x) = I_0(x) \text{ and } I(1, x) = I_1(x)\,.\end{aligned}$$

In fact, in order to clearly look at the similarities and differences with optimal transport, instead of acting on images, one can act on densities. Let us consider the action on densities defined above and its infinitesimal action (1.2.20)*. Therefore, the corresponding optimization*

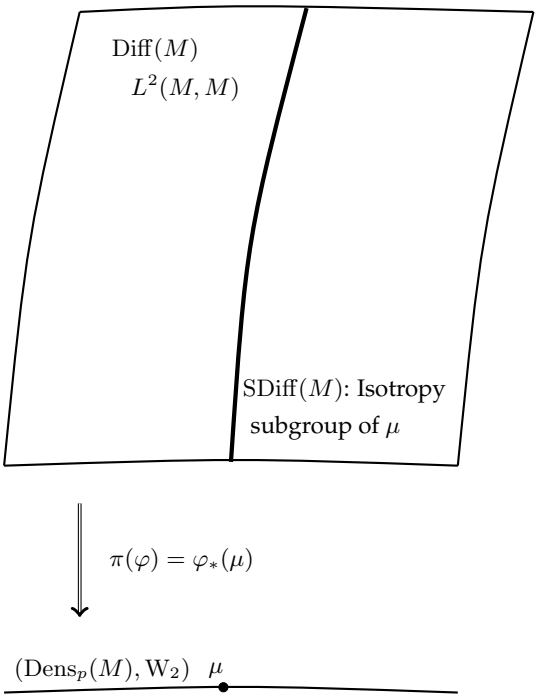


FIGURE 1.1: A Riemannian submersion and SDiff(M) as a Riemannian submanifold of $L^2(M,M)$: Incompressible Euler equation on SDiff(M)

problem leaves the action (1.2.23) *unchanged and the constraint becomes:*

$$\begin{aligned} &\partial_t \rho + \mathrm{div}(\rho\, \xi) = 0 \\ &\rho(0,x) = \rho_0(x) \text{ and } \rho(1,x) = \rho_1(x)\,. \end{aligned}$$

This problem is closer to the (1.2.12) *formulation, however, it is* ***not a convex optimization problem****. Note in particular that the action* (1.2.23) *does not involve the current density ρ, contrary to optimal transport. In contrast to optimal transport which applies to probability densities, an important feature of the LDDMM framework is its straightforward applications to natural objects: embeddings, differential forms, metric tensors,..., which makes it versatile for applications in imaging.*

In the situation of Proposition 3, one may be interested by the Riemannian manifold (Q, g_Q) and also by the fibers and their induced metric. In particular, $H_0 \overset{\text{def.}}{=} \pi_0^{-1}(\{q_0\})$ is a subgroup of G called the isotropy subgroup at point q_0. Note that H_0 is a submanifold of G and it is endowed with the Riemannian metric induced by g_G, which will be denoted by g_{H_0}. Immediately from the definition, since $h \cdot q_0 = q_0$ for every $h \in H_0$, the metric g_{H_0} is *right-invariant* under the action of H_0. It is thus a standard way to build right-invariant Riemannian metrics. Therefore, we have

Proposition 7. *Consider a Riemannian submersion as constructed in Proposition 3. Let H_0 be the isotropy subgroup of $q_0 \in Q$, then, H_0, seen as a Riemannian submanifold of G, has an induced metric g_{H_0} which is right-invariant (on H_0).*

This proposition shows that H_0 can be seen as a Riemannian submanifold in the group G endowed with its metric, or as a group with a right-invariant metric.

Note that this remark is useless in the LDDMM framework since the metric is already right-invariant on the full group, so the induced metric on the isotropy group

is a fortiori invariant under itself. However, in the optimal transport framework, it leads to the L^2 right-invariant metric on the group of volume preserving diffeomorphisms, denoted by SDiff, for which the geodesic equation is the incompressible Euler equation, as shown by Arnold in [Arn66]. This point of view has been investigated by Ebin and Marsden in [EM70] where the authors have taken an intrinsic point of view on the group of diffeomorphisms as an infinite dimensional weak Riemannian manifold. Formulating the geodesic equation as an ordinary differential equation in a Hilbert manifold of Sobolev diffeomorphisms, they proved, among others, local well-posedness of the geodesic equation for smooth enough initial conditions.

The Riemannian submanifold point of view was used by Brenier, motivated by the variational interpretation of geodesics as minimizers of the action functional. In particular, his polar factorization theorem [Bre91] was motivated by a numerical scheme approximating geodesics on the group of volume preserving diffeomorphisms. Optimal transport then appeared as a key tool to project a map onto this group by minimizing the L^2 distance and it can be interpreted as a non-linear extension of the pressure in the incompressible Euler equation. Back to the variational problem on SDiff, Brenier also used optimal transport in order to define the notion of generalized geodesics for the incompressible Euler equation in [Bre99].

Let us rewrite the polar factorization theorem of Brenier [Bre91] in a slightly more general formulation, which is shown in Figure 1.2.

Proposition 8 (A pre-polar factorization)**.** *Under the hypothesis of Proposition 3, consider $g \in G$, then there exists $h_0 \in H_0$ such that the geodesic between h_0 and g is the horizontal lift at point g of a geodesic between $g \cdot q_0$ and q_0. Moreover, if the geodesic between $g \cdot q_0$ and q_0 is unique, then h_0 is unique.*

This theorem does not explain the full polar factorization result of Brenier since more structure is available due, in that case, to the convexity associated with the Wasserstein metric on $\mathrm{Dens}_p(M)$. In this case, the element gh_0^{-1} is the gradient of a convex function, thus writing $g = \nabla\psi \circ h_0$ with $h_0 \in \overline{\mathrm{SDiff}}$, the set of volume preserving maps (see [Bre91]).

To conclude this section, let us underline that this geometric point of view obviously does not alleviate the analytical issues which have to be treated in each particular case. However, this explains the extension of the polar factorization theorem in Chapter 3.

1.3 Geodesic equations

In this section, we recall the computation of the geodesic equations in the LDDMM case. More precisely, the geodesic on the space of shapes and its horizontal lift. We first start with the geodesic equations on a Lie group with a right-invariant metric and then detail it for the induced metric.

1.3.1 Geodesics for right-invariant metric and fluid dynamic equations

We quickly describe the derivation of the geodesic equation for a right-invariant metric on a Lie group. A short proof of the derivation of this equation is given in [MP10, Theorem 3.2] in the case of a kinetic energy which holds true for general Lagrangians that are right-invariant. We need the definition of the adjoint and coadjoint operators:

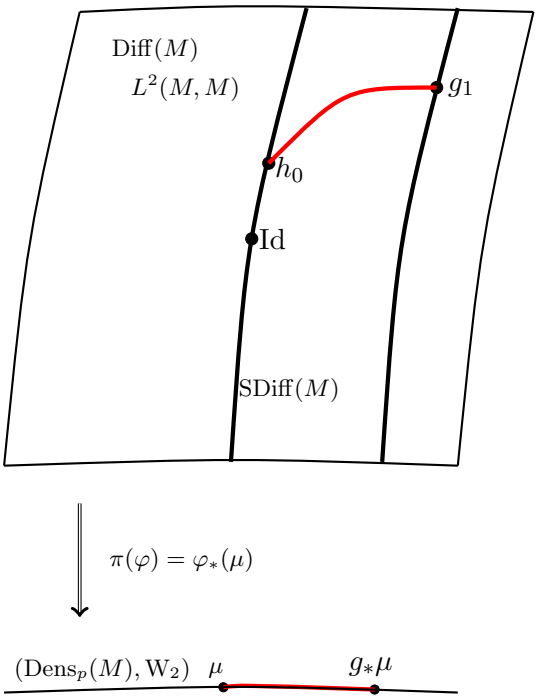


FIGURE 1.2: Polar factorization: $h_0 = \arg\min_{h\in \mathrm{SDiff}} \|g-h\|_{L^2}$

Definition 9. Let G be a Lie group and $h \in G$, the adjoint operator $\mathrm{Ad}_h : \mathfrak{g} \mapsto \mathfrak{g}$ is defined by

$$\mathrm{Ad}_h(\xi) \overset{\text{def.}}{=} dL_h \cdot dR_{h^{-1}}(\xi)\,. \tag{1.3.1}$$

Then, Ad_h^* is the adjoint of Ad_h defined by duality on $\mathfrak{g}$. Their corresponding differential map at Id are respectively denoted by ad and ad^*.

Let G be a Lie group, and $\mathcal{L} : TG \mapsto \mathbb{R}$ be a Lagrangian which satisfies the following property,

$$\mathcal{L}(g,\dot g) = \mathcal{L}(\mathrm{Id}, dR_{g^{-1}}(\dot g))\,. \tag{1.3.2}$$

The reduced Lagrangian is $\ell : \mathfrak{g} \mapsto \mathbb{R}$ defined by $\ell(\xi) = \mathcal{L}(\mathrm{Id},\xi)$ for $\xi \in \mathfrak{g}$. Thus, the variational problem for a reduced Lagrangian reads

$$\inf \int_0^1 \ell(\xi)\,\mathrm{d}t \quad \text{subject to} \quad \begin{cases} \dot g = dR_g(\xi) \\ g(0) = g_0 \in G \text{ and } g(1) = g_1 \in G\,. \end{cases} \tag{1.3.3}$$

In order to compute the Euler-Lagrange equation for (1.3.3), one needs to compute the variation of ξ in terms of the variation of g. It is given by $\dot w - \mathrm{ad}_\xi w$ for any path $w(t) \in T_{\mathrm{Id}}G$, therefore, the Euler-Lagrange equation reads

$$\left(\partial_t + \mathrm{ad}^*_\xi\right) \frac{\partial \ell}{\partial \xi} = 0\,, \tag{1.3.4}$$

and is called Euler-Poincaré equation. When the Lagrangian is a kinetic energy, $\ell(\xi) = \frac{1}{2}\langle \xi, L\xi\rangle$, which will be also denoted by $\frac{1}{2}\|\xi\|^2_{\mathfrak{g}}$, where $L : \mathfrak{g} \mapsto \mathfrak{g}$ is a quadratic form and $\langle\cdot,\cdot\rangle$ denotes the dual pairing, one has $\frac{\partial \ell}{\partial \xi} = L\xi$ and $L\xi$ is the so-called momentum, which will be denoted by m in what follows. Then, the critical curves are determined by their initial conditions $(g(0),\dot g(0))$ and the Euler-Poincaré equation (1.3.4). In the context of infinite dimensional Riemannian manifolds enjoying a group structure, this equation is called the Euler-Arnold equation [Arn66].

Let us compute more explicitely the Euler-Arnold equation and detail the expression of the adjoint Ad_h^* which acts on 1-forms. Let m be a 1-form density, then $\mathrm{Ad}_\varphi^*(\mathsf{m}) = D\varphi^T(\mathsf{m}\circ\varphi)\,\mathrm{Jac}(\varphi)$ and therefore the differentiation w.r.t. φ, the flow of the vector field u, gives

$$\mathrm{ad}_u^*(\mathsf{m}) = \mathrm{div}(u)\mathsf{m} + Du^T\cdot\mathsf{m} + D\mathsf{m}\cdot u\,. \tag{1.3.5}$$

Thus, the Euler-Arnold equation reads

$$\begin{cases} \partial_t \mathsf{m}_t + \mathrm{div}(u_t)\mathsf{m}_t + Du_t^T\cdot\mathsf{m}_t + D\mathsf{m}_t\cdot u_t = 0 \\ Lu_t = m_t\,, \end{cases} \tag{1.3.6}$$

where L is the differential operator defining the metric. A more geometrical way of writing this equation is the following,

$$\partial_t\mathsf{m}_t + \mathcal{L}_{u_t}\mathsf{m}_t + \mathrm{div}(u_t)\mathsf{m}_t = 0\,, \tag{1.3.7}$$

or alternatively

$$(\partial_t + \mathcal{L}_{u_t})\,(\mathsf{m}_t\otimes\mathrm{vol}) = 0\,, \tag{1.3.8}$$

together with the relation $Lu_t = \mathsf{m}_t$.

Some important examples in fluid dynamics of the Euler-Arnold equation are given hereafter. For the L^2 metric in one dimension, $Lu = u$, one has

$$\partial_t u + 3\partial_x u\, u = 0\,, \tag{1.3.9}$$

which is the inviscid Burgers equation. For the H^{div} metric in one dimension, $Lu = u - \partial_{xx}u$, one has the Camassa-Holm equation (actually when $a = b = 1$)

$$a^2\partial_t u - b^2\partial_{txx}u + 3a^2\partial_x u\, u - 2b^2\partial_{xx}u\,\partial_x u - b^2\partial_{xxx}u\, u = 0\,. \tag{1.3.10}$$

The Korteweg-de Vries equation can also be understood in this setting on a central extension of the group $\mathrm{Diff}(S_1)$. In the case where $G = \mathrm{SDiff}(M)$ is the group of volume preserving diffeomorphisms, the Euler-Arnold equation is the incompressible Euler equation

$$\partial_t u + \nabla_u u = -\nabla p\,,\ \mathrm{div}(u) = 0\,. \tag{1.3.11}$$

Let us detail the case of the $H^{\mathrm{div}}(\mathbb{T}_d)$ where $\mathbb{T}^d \overset{\text{def.}}{=} \mathbb{R}^d/\mathbb{Z}^d$ metric which is given in [KLMP13, Theorem A.1]. The differential operator takes the form $Lu = a^2u + b^2\nabla\,\mathrm{div}(u)$ which gives

$$\begin{aligned} \partial_t Lu + a^2\left(\mathrm{div}(u)u + \frac{1}{2}\nabla\langle u,u\rangle + Du\cdot u\right) + \\ b^2\Big(\mathrm{div}(u)\nabla\,\mathrm{div}(u) + Du^T\cdot\nabla\,\mathrm{div}(u) + D[\nabla\,\mathrm{div}(u)]\cdot u\Big) = 0\,. \end{aligned} \tag{1.3.12}$$

In fact, this equation can be extended to a general Riemannian manifold.

1.3.2 Geodesic for LDDMM

The geodesic equations on the orbits for the induced metric (by the Riemannian submersion) of a right-invariant metric are in correspondence with their horizontal lifts on the group.

We start with a few useful definitions. The tangent lift of Φ is defined as the action of G on TQ,

$$G \times TQ \to TQ, \quad (g, v_q) \mapsto g \cdot v_q := T\Phi_g(v_q), \tag{1.3.13}$$

with infinitesimal generator ξ_{TQ} corresponding to $\xi \in V$. Similarly, one defines the cotangent lifted action as

$$G \times T^*Q \to T^*Q, \quad (g, p) \mapsto g \cdot p := (T\Phi_{g^{-1}})^*(p)\,, \tag{1.3.14}$$

and the associated infinitesimal action will be denoted by $-\xi^* \cdot p$. The momentum map $\mathbf{J} : T^*Q \to V^*$ associated with the cotangent lift of Φ is given by

$$\langle \mathbf{J}(p, q), \xi \rangle_{V^* \times V} = \langle p, \xi \cdot q \rangle_{T^*Q \times TQ}\,, \tag{1.3.15}$$

for arbitrary $p \in T^*_qQ$ and $\xi \in V$.

Using the Lagrange multiplier method, the minimizers of the LDDMM functional are critical points of the augmented functional

$$\mathcal{J}(\xi, p, q) = \int_0^1 \frac{1}{2}\|\xi(t)\|_V^2 \,\mathrm{d}t + S(q(1)) + \int_0^1 (p(t), \dot{q}(t) - \xi(t) \cdot q(t))_{T^*_qQ, T_qQ} \,\mathrm{d}t\,, \tag{1.3.16}$$

which leads to the Euler-Lagrange equations:

$$\begin{cases} \dot{q}(t) = \xi(t) \cdot q(t) \\ \dot{p}(t) = -\xi^*(t) \cdot p(t) \\ \xi(t) = K\mathbf{J}(p(t), q(t))\,, \end{cases} \tag{1.3.17}$$

where the notation $-\xi(t)^* \cdot p(t)$ is the infinitesimal coadjoint action. For instance, in the case of functions $I : M \mapsto \mathbb{R}$, the system is

$$\begin{cases} \dot{I}(t) + \langle \nabla I(t), v(t) \rangle = 0 \\ \dot{P}(t) + \operatorname{div}(P(t)v(t)) = 0 \\ v(t) + K(P(t)\nabla I(t)) = 0\,, \end{cases} \tag{1.3.18}$$

which is used in Chapter 4. Note, that written on densities instead of functions, the corresponding system simply swaps the two first equations of system (1.3.22). The gradient of Functional (1.3.16) can be written as:

$$\nabla \mathcal{J}(t) = \xi(t) - K\mathbf{J}(p(t), q(t))\,, \tag{1.3.19}$$

where $q(t)$ solves the forward flow equation $\dot{q} = \xi(t) \cdot q(t)$ with $q(0) = q_0$ and $p(t)$ solves the adjoint equation

$$\dot{p}(t) = -\xi^*(t) \cdot p(t)\,, \tag{1.3.20}$$

with the boundary condition $p(1) = [\partial_q]_{q=q(1)} S(q(1))$. In particular, one has to solve the adjoint equation backward in time, to compute the gradient.

The horizontal lift of the geodesic $q(t)$ is simply given by the vector field $\xi(t) = K\mathbf{J}(p(t), q(t))$, so that $\xi(t)$ satisfies the Euler-Poincaré equation (1.3.4), as is shown

by the following computation,

$$\begin{aligned}\frac{d}{dt}\left(\mathbf{J}(p(t),q(t)),w\right)_{V^*,V} &= (\dot{p}(t), w\cdot q(t)) + (p(t), w\cdot \dot{q}(t))\\ &= (-d\xi(t)^*(q(t))(p(t)), w\cdot q(t)) + (p(t), dw(\xi(t))\cdot q(t))\\ &= (p(t), -d\xi(t)(w)\cdot q(t)) + (p(t), dw(\xi(t))\cdot q(t))\\ &= \big(p(t), -\operatorname{ad}_{\xi(t)}(w)\cdot q(t)\big)\\ &= \Big(-\operatorname{ad}^*_{\xi(t)}\mathbf{J}(p(t),q(t)), w\Big)_{V^*,V}\end{aligned}$$

which implies

$$\frac{d}{dt}\mathbf{J}(p(t),q(t)) + \operatorname{ad}^*_{\xi(t)}\mathbf{J}(p(t),q(t)) = 0\,, \tag{1.3.21}$$

and the conservation of the momentum: $\frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{Ad}_{\varphi(t)^*}(\mathbf{J}(p(t),q(t)))] = 0$.
In addition, the system (1.3.17) are the Hamiltonian equations associated to the Hamiltonian $H(p,q) = \frac{1}{2}\|\mathbf{J}(p(t),q(t))\|^2_V$.

Remark 10. *These geodesic equations* (1.3.17) *can be compared with optimal transport, for which the geodesic equations can be formulated in a smooth context,*

$$\begin{cases}\dot{\rho} + \operatorname{div}(\rho\nabla P) = 0\\ \dot{P} + \frac{1}{2}|\nabla P|^2 = 0\,.\end{cases} \tag{1.3.22}$$

In particular, the equation on P does not depend on the current density ρ.

From a numerical point of view, the formulation (1.3.17) is preferably used when the dimensionality of Q is lower than that of the (discretized) diffeomorphisms group, for instance often in the case of landmarks (group of points). Note that in this case, a lot is known on the geometry of the group [MMM13] and in particular curvature.

1.4 Contributions

In this section, we list the contributions we made in our research articles and we detail the cases where the work has been developed in collaboration with a student or a post-doc. Note that this document is based on existing material from our research articles, although sometimes presented with a slightly different point of view. Moreover, the introduction presents a common geometric framework between LDDMM and optimal transport, which, to the best of our knowledge, has not been written elsewhere.

Contributions to the theory of LDDMM and shape spaces. (Chapter 2)

1. In [BV17], we prove that completeness results for the group of diffeomorphisms equipped with a strong right-invariant Sobolev metric hold in all the meanings of the Hopf-Rinow theorem (which is valid in infinite dimensions). The strategy developed in the paper consists in merging the approach of Ebin and Marsden in [EM70] and the direct method in calculus of variation to obtain the desired results.

2. We introduced in [VT12] the use of second-order variational problems on the group of diffeomorphisms and shape spaces and, together with our coauthors,

we extended to higher-order in [GHM+12a]. In [TV16], we study the corresponding variational problem on the group of diffeomorphisms in a simple case and compute the relaxation of it.

3. On the space of curves in the plane, we studied a Finsler type of metric which favors piecewise rigid motion, motivated by [CNPV16]. The main result of [NPV16] is a proof of existence of shortest paths. This result is also generalized to the case of H^s inner metric on closed curves, for $s > 3/2$.

Contributions to unbalanced optimal transport and its link to fluid dynamic equations (Chapter 3).

1. We solved one issue related with the application of optimal transport to real data, the problem of total volume normalization:
We introduced the equivalent to the L^2 Wasserstein metric in the case of non-negative Radon measures. The study of this new metric, which we called the Wasserstein-Fisher-Rao metric was the starting project of Lenaïc Chizat's PhD thesis [CSPV16]. Generalizing Otto's Riemannian submersion, we proposed to Lenaïc Chizat and Bernhard Schmitzer to study the corresponding static formulation [CPSV15], with in mind the goal of developing fast numerical algorithms based on entropic regularization, later developed in [CPSV16].

2. In [GV16], we make the link between the Wasserstein-Fisher-Rao metric and the Camassa-Holm equation in fluid dynamics, which is the same than the link between the L^2 Wasserstein metric and the incompressible Euler equation. Among others, we prove an extension of the polar factorization theorem to this unbalanced case and we show a rather surprising result: The Camassa-Holm equation, which is a 1D PDE can be lifted in 2D as a sort of incompressible Euler equation for a radial density which has a singularity at 0.

Contributions on the practical implementation and applications of LDDMM (Chapter 4).

1. We made practical and substantial contributions to the choice of the right-invariant metric in the LDDMM method, sometimes by contributing to the theory or extending the framework. The paper [RVW+11] emphasizes the choice of the kernel and introduce the use of sum of gaussian kernels to produce more plausible deformations, while its corresponding geometrical structure is uncovered in [BRV12]. In [SRV13], we depart from the LDDMM framework to make sense of spatially varying kernels. Finally, in [VR14], we propose a simple variational approach to learn the parameters of the metric.
These methods are implemented in the software *Utilzreg* freely available at `http://sourceforge.net/projects/utilzreg/`. We were involved in the development of this C++ code from 2010 to 2014.

2. In [FCVP17], we introduced the use of entropic optimal transport as a similarity measure in diffeomorphic image matching. This new similarity measure is smooth, convex and represents a global similarity measure which helps avoiding bad local minima.

3. In [NHV11], we extended geodesic regression to the case of images in the LDDMM framework, based on an algorithm we developed in [VRRC12]. In [FRR+14], together with J.B. Fiot, a PhD student of L. Cohen, we develop a

method based on the tangent space representation and the algorithm developed in [VRRC12] to perform classification on longitudinal data in medical imaging. His results showed that statistically meaningful informations were present in the longitudinal evolutions.

Chapter 2

Variational problems on diffeomorphisms and shape spaces

Our contributions are the following:

1. We prove in [BV17] metric completeness on the group of diffeomorphisms endowed with a strong right-invariant Sobolev metric, provided its smoothness and that between any two diffeomorphisms of a given connected component, there exists a minimizing geodesic.

2. Minimizing the acceleration on the group of diffeomorphisms was proposed in [GHM$^+$12a, GHM$^+$12b, VT12] but the theoretical study was left open. We study this variational problem in the particular case of the group of diffeomorphisms of the interval $[0, 1]$, endowed with a right-invariant Sobolev metric of order 2. We compute the relaxation of this variational problem, in which appears the so-called the Fisher-Rao functional, a convex one-homogeneous functional on the space of measures.

3. On the space of closed curves in the plane, we studied a Finsler type of metric which favors piecewise rigid motion. The main result of [NPV16] is a proof of existence of shortest paths, which is also generalized to strong Sobolev metrics.

2.1 Introduction

In this chapter, we give a summary of our theoretical contributions on theoretical questions on shape spaces [BBM14]. Our contributions are concerned with variational calculus on the space of shapes and in particular completeness results once a Riemannian or Finsler structure is given. That is global well posedness of the geodesic equation, metric completeness (which is stronger) and the existence of shortest paths. Such questions are also relevant for higher-order models, which we also address in a simple situation.

We gave a particular interest to the framework of LDDMM in which we studied the case of strong right-invariant metric on the group of diffeomorphisms. Let M be either $\mathbb{R}^d$ or a compact manifold without boundary of dimension d. The group $\mathcal{G}_V$, defined in formula (1.2.5), is a priori not a differentiable manifold nor a topological group (the inversion need not be continuous). However, for certain choices of spaces V, such structures are available and therefore more properties can be derived in this situation. Indeed, consider the group $\mathcal{D}^s(M)$, with $s > d/2 + 1$, which consists of all C^1-diffeomorphisms of Sobolev regularity H^s. It is known since the work of Ebin

and Marsden [EM70] that $\mathcal{D}^s(M)$ is a smooth Hilbert manifold and a topological group. It only remains to prove that $\mathcal{G}_{H^s} = \mathcal{D}^s_0(M)$ (the connected component of identity) which is done in [BV17, Section 8]. When such structures are available, it is possible to study higher-order problems on the group of diffeomorphisms such as acceleration minimizing curves.

Strong right-invariant Sobolev metrics. The first contribution is concerned with the problem of geodesic completeness of the group of diffeomorphisms endowed with a strong right-invariant Sobolev metric. Historically most papers dealt with right-invariant Sobolev metrics on diffeomorphism groups in the weak setting, that is one considered H^r-metrics on $\mathcal{D}^s(M)$ with $s > r$; a typical assumption is $s > 2r + d/2 + 1$, in order to ensure that Lu is still C^1-regular. The disconnect between order of the metric and regularity of the group arose, because one was mostly interested in L^2 or H^1-metrics, but $\mathcal{D}^s(M)$ is a Hilbert manifold only when $s > d/2 + 1$. It was however noted already in [EM70] and again in [MP10], that the H^s-metric is well-defined and more importantly smooth on $\mathcal{D}^s(M)$, for integer s when the inner product is defined in terms of a differential operator. The smoothness of the metric is not obvious, since it is defined via

$$G_\varphi(X_\varphi, Y_\varphi) = \langle X_\varphi \circ \varphi^{-1}, Y_\varphi \circ \varphi^{-1} \rangle_{H^s}$$

and the definition uses the inversion, which is only a continuous, but not a smooth operation on $\mathcal{D}^s(M)$. In order to understand the smoothness of the metric, one needs to make the change of variable $y = \varphi^{-1}(x)$ and write

$$G_\varphi(X_\varphi, Y_\varphi) = \int_M \langle X_\varphi(y), (L(Y_\varphi \circ \varphi^{-1}) \circ \varphi \rangle \operatorname{Jac}(\varphi) \, \mathrm{d}\mu \,, \tag{2.1.1}$$

which makes appear the operator $L_\varphi \overset{\text{def.}}{=} \operatorname{Jac}(\varphi)(L(Y_\varphi \circ \varphi^{-1}) \circ \varphi$ which can be alternatively written as $L_\varphi = R^*_{\varphi^{-1}} \circ L \circ R_{\varphi^{-1}} = \operatorname{Jac}(\varphi) R_\varphi \circ L \circ R_{\varphi^{-1}}$ where R_ψ denotes the right composition with ψ and R^*_ψ its adjoint. There are at least two different ways to understand why the operator has a chance to be smooth with respect to φ. The first one consists in writing $L_\varphi(X)$ using the derivatives of Y and φ and realizing that the expression is rational in $[D\varphi]^{-1}$ and $\operatorname{Jac}(\varphi)$ and the derivatives of Y. Therefore, using the fact that H^s is a Hilbert algebra for $s > d/2$, one can conclude that it is smooth. It is actually a proof of its smoothness. However, this proof does not extend to the case when L is not a differential operator but a general elliptic Fourier multiplier, although the result still holds, see [Kol16]. The proof of it uses the similarity of L_φ with the adjoint action of Lie groups. Indeed, differentiation w.r.t. φ leads to a commutator between two operators. If the operator L were a differential operator, then the order of the commutator would be the same than L and this would give another proof of the smoothness. Indeed, one has

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} R_\varphi \circ L \circ R_{\varphi^{-1}} = R_\varphi \circ (\nabla_\xi \circ L_\varphi - L_\varphi \circ \nabla_\xi) \circ R_{\varphi^{-1}} \tag{2.1.2}$$

where φ is the flow of the vector field ξ. In the general case, one is left with the estimation of this commutator in order to prove the smoothness which is done in [Kol16]. Higher order Sobolev metrics have been studied recently on diffeomorphism groups of the circle [CK03], of the torus [KLT08] and of general compact manifolds [MP10].

One interesting property of strong Riemannian metrics on the diffeomorphism group is that the exponential map is locally well defined by the Gauss lemma valid in this infinite dimensional setting.

Higher-order variational problems: In [VT12], we introduced the use of cubic splines *in the space of shapes* to interpolate a sequence of shapes that are time dependent. Riemannian cubics (also called Riemannian splines) and probably more famous, its constrained alternative called Elastica belong to a class of problems that have been studied since the work of Euler (see the discussion in [Mum94]). Let us present the variational problem in a Riemannian setting. Riemannian splines are minimizers of

$$\mathcal{J}(x) = \int_0^1 g\left(\frac{D}{Dt}\dot{x}, \frac{D}{Dt}\dot{x}\right) \mathrm{d}t\,, \tag{2.1.3}$$

where (M, g) is a Riemannian manifold, $\frac{D}{Dt}$ is its associated covariant derivative and x is a sufficiently smooth curve from $[0, 1]$ in M satisfying first order boundary conditions, i.e. $x(0), \dot{x}(0)$ and $x(1), \dot{x}(1)$ are fixed. The case of Elastica consists in restricting the previous optimization problem to the set of curves that are parametrized by unit speed (when the problem is feasible), namely $g(\dot{x}, \dot{x}) = 1$ for all time.

This type of variational problems has been several times introduced and studied in applied mathematics [BBT65, LF73, NHP89, CLP95, CL95, GG02, Koi92] as well as in pure mathematics [Bis11, LS84, BG86] and it was then extensively used and numerically developed in image processing and computer vision [Mum94, CGMP11, USK15, SASK12, BK94, CKKS02]. In the past few years, higher-order models have been introduced in biomedical imaging for interpolation of a time sequence of shapes. They have been introduced in [VT12] for a diffeomorphic group action on a finite dimensional manifold and further developed for general invariant higher-order lagrangians in [GHM$^+$12a, GHM$^+$12b] on a group. A numerical implementation together with a generalized model have been proposed in [SVN15] in the context of medical imaging applications. What is still unsolved in the case of a group of diffeomorphisms, is the question of existence and regularity where the main obstacle is caused by the infinite dimensional setting.

In infinite dimensions, to the best of our knowledge, only the linear case has been addressed [Mic02]. Actually, in the non-linear case, the case of Riemannian metrics in infinite dimensions, existence of minimizing geodesics is already non-trivial as shown by Atkin in [Atk97], where an example of a geodesically complete Riemannian manifold is given such that the exponential map is not surjective. Therefore, Elastica or Riemannian splines will preferably be studied on Riemannian manifolds where all the properties of the Hopf-Rinow theorem fully hold.

As presented below, only recently it has been proven in [BV17] that the group of diffeomorphisms endowed with a right-invariant Sobolev metric of high enough order is complete in the sense of the Hopf-Rinow theorem. Motivated by this positive result, we explore the minimization of the acceleration in the one dimensional case, which is the first step towards its generalizations to higher dimensions.

2.2 Completeness results on the group of diffeomorphisms

This section is based on joint work with M. Bruveris in [BV17].

In this section, we present our contributions on this topic. First we want to show that strong, smooth Sobolev metrics on $\mathcal{D}^s(M)$ are complete both geodesically, metrically and that there exist minimizing geodesics between any two diffeomorphisms.

We recall here that the Hopf-Rinow theorem is not valid in infinite dimensions, namely Atkin gives in [Atk75] an example of a geodesically complete Riemannian manifold where the exponential map is not surjective. For the Sobolev diffeomorphism group with $s > d/2 + 1$, the best known result can be found in [MP10, Thm. 9.1] which is an improvement of the positive result of Ekeland [Eke78].

Geodesic completeness was shown for diffeomorphism group of the circle in [EK13] and in weaker form on $\mathbb{R}^d$ in [TY05a] and [MM13]. Metric completeness and existence of minimizing geodesics in the context of groups of Sobolev diffeomorphisms and its subgroups is — as far as we know – new. We prove the following theorem:

Theorem 11. *Let M be $\mathbb{R}^d$ or a closed manifold and $s > d/2 + 1$. If G^s is a smooth, right-invariant Sobolev-metric of order s on $\mathcal{D}^s(M)$, then*

1. *$(\mathcal{D}^s(M), G^s)$ is geodesically complete;*
2. *$(\mathcal{D}^s(M)_0, \operatorname{dist}^s)$ is a complete metric space;*
3. *Any two elements of $\mathcal{D}^s(M)_0$ can be joined by a minimizing geodesic.*

The statements also hold for the subgroups $\mathcal{D}^s_\mu(M)$ and $\mathcal{D}^s_\omega(M)$ of diffeomorphisms preserving a volume form μ or a symplectic structure ω.

The crucial ingredient in the proof is showing that the flow map

$$\operatorname{Fl}_t : L^1(I, \mathfrak{X}^s(M)) \to \mathcal{D}^s(M) \tag{2.2.1}$$

exists and is continuous. The existence was known for vector fields in $C(I, \mathfrak{X}^s(M))$ and the continuity as a map into $\mathcal{D}^{s'}$ for $s' < s$ was shown in [Inc12]. We extend the existence result to vector fields that are L^1 in time and show continuity with respect to the manifold topology. The flow map allows us to identify the space of H^1-paths with the space of right-trivialized velocities,

$$\mathcal{D}^s(M) \times L^2(I, \mathfrak{X}^s(M)) \xrightarrow{\cong} H^1(I, \mathcal{D}^s(M)), \quad (\varphi, u) \mapsto \operatorname{Fl}(u) \circ \varphi\,.$$

Since $L^2(I, \mathfrak{X}^s(M))$ is a Hilbert space, we can use variational methods to show the existence of minimizing geodesics.

In order to show metric completeness, we derive the following estimate on the geodesic distance,

$$\|\varphi - \psi\|_{H^s} \leq C \operatorname{dist}^s(\varphi, \psi)\,,$$

which is valid on a bounded metric dist^s-ball. In other words, the identity map between the two metric spaces

$$\operatorname{Id} : \left(\mathcal{D}^s(\mathbb{R}^d), \|\cdot\|_{H^s}\right) \to \left(\mathcal{D}^s(\mathbb{R}^d), \operatorname{dist}^s\right)$$

is locally Lipschitz continuous. For compact manifolds we show a similar inequality in coordinate charts. The Lipschitz continuity implies that a Cauchy sequence for dist^s is a Cauchy sequence for $\|\cdot\|_{H^s}$, thus giving us a candidate for a limit point. One then needs proceeds to show that the limit point lies in the diffeomorphism group and that the sequence converges to it with respect to the geodesic distance.
Finally, we show the existence of minimizing geodesics between any two diffeomorphisms in the same connected component. This extends Thm. 9.1 in [MP10], where existence of minimizing geodesics was shown only for an open and dense subset.

This existence result is shown using the direct method of the calculus of variations. Namely, the variational problem we consider consists in the minimization of an energy which is, under a change of variables, a weakly lower semi-continuous functional on a weakly closed constraint set. The change of variables is simply given by the vector field associated with the path and in the next lemma, we also prove that the constrained set is weakly closed.

On the analysis of the LDDMM framework, the paths solving the image registration problem are smooth. We also obtain, in [BV17, Section 8], using the proximal calculus on Riemannian manifolds [AF05] that Karcher means (points that minimize the sum of the squared distances to a given finite number of points) of k diffeomorphisms – and more generally shapes – are unique on a dense subset of the k-fold product $\mathcal{D}^s \times \ldots \times \mathcal{D}^s$.

Open question 12. *Consider the group $\mathcal{D}^s(M)$ and two diffeomorphisms in $D^k(M)$ with $k > s$. Does any geodesic joining the two lies in $D^k(M)$? Locally, due to the Gauss lemma and the "no loss no gain" result of [Kol16], the answer is positive but the answer is not known globally, in particular when the exponential map is not invertible. In the article [TV16], we show that the answer is positive in a particular case $M = [0, 1]$ on the group of H^2 diffeomorphisms which is identity on the points 0 and 1 (at first order).*

2.3 Minimization of the acceleration

We present hereafter some of the results obtained in [GHM+12a] and the relaxation result of [TV16], wich is a joint work with R. Tahraoui. One of the motivation of this work can be found in [VT12] where the finite dimensional case of landmarks was treated. Figure 2.1 shows three different types of interpolations: the first column is geodesic interpolation where the time sequence of curves (in white) is interpolated in time (time is the z axis); the second column is piecewise geodesic interpolation; the last one is cubic spline interpolation, which is our interest in this section. Note that for application, only the first and third columns present an interesting dimension reduction compared to the full sequence of data. A question of mathematical interest

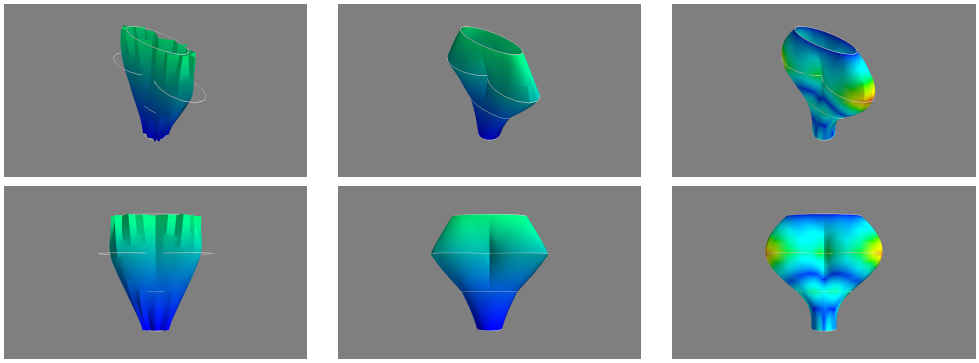


FIGURE 2.1: Different types of shape interpolation: geodesic interpolation, piecewise geodesic, cubic spline interpolation. The interpolated shapes are represented in white and are group of points in 2D. On the first two columns, the color change represents the time evolution.

is the extension of cubic splines to the infinite dimensional setting. We give some insights on the question of existence hereafter.

2.3.1 Invariant higher-order variational problems

In [GHM+12a], higher-order models are proposed on groups of diffeomorphisms but for the standard Riemannian cubics functional, no analytical study was provided. However, the formulation of the Euler-Lagrange equation in reduced coordinates is simple enough to summarize some of the results in [GHM+12a]. Namely, let $\mathcal{L} : T^kG \mapsto \mathbb{R}$ be a Lagrangian defined on the k^{th} order tangent bundle, then a curve $g : [t_0, t_1] \to G$ is a critical curve of the action

$$\mathcal{J}[q] = \int_{t_0}^{t_1} \mathcal{L}\left(g(t), \dot{g}(t), ..., g^{(k)}(t)\right) dt \tag{2.3.1}$$

among all curves $g(t) \in G$ whose first $(k-1)$ derivatives are fixed at the endpoints: $g^{(j)}(t_i)$, $i = 0, 1$, $j = 0, ..., k-1$, if and only if $g(t)$ is a solution of the k^{th}-order Euler-Lagrange equations

$$\sum_{j=0}^{k} (-1)^j \frac{d^j}{dt^j} \frac{\partial \mathcal{L}}{\partial g^{(j)}} = 0\,. \tag{2.3.2}$$

Now, an invariant higher-order Lagrangian is completely defined by its restriction on the higher-order tangent space at identity. As a consequence, the Lagrangian (2.3.1) can be rewritten as

$$\mathcal{L}\left([g]^{(k)}_{g(t_0)}\right) = \ell\left(v(t_0), \dot{v}(t_0), \ldots, v^{(k-1)}(t_0)\right),$$

where $v \overset{\text{def.}}{=} dR_{g^{-1}}(\dot{g})$ as detailed in [GHM+12a]. The corresponding higher-order Euler-Poincaré equation is

$$(\partial_t + \operatorname{ad}^*_v) \sum_{j=0}^{k-1} (-1)^j \partial_t^j \frac{\delta \ell}{\delta v^{(j)}} = 0\,. \tag{2.3.3}$$

Let us instantiate it in the case of the Lagrangian (2.1.3) for which the previous setting applies. Indeed, in the case of a Lie group G with a right-invariant metric, the covariant derivative can be written as follows: Let $V(t) \in T_{g(t)}G$ be a vector field along a curve $g(t) \in G$, $\nu(t) = dR_{g^{-1}}(V(t))$ and $\xi(t) = dR_{g^{-1}}(\dot{g})$,

$$\frac{D}{Dt}V = \left(\dot{\nu} + \frac{1}{2}\operatorname{ad}^\dagger_\xi \nu + \frac{1}{2}\operatorname{ad}^\dagger_\nu \xi - \frac{1}{2}[\xi, \nu]\right)_G(g)\,, \tag{2.3.4}$$

where $\operatorname{ad}^\dagger$ is the metric adjoint defined by

$$\operatorname{ad}^\dagger_\nu \kappa := (\operatorname{ad}^*_\nu(\kappa^\flat))^\sharp$$

for any $\nu, \kappa \in \mathfrak{g}$ where $\flat$ is the isomorphism associated with the metric from$\mathfrak{g}$ to $\mathfrak{g}^*$ and $\sharp$ is its inverse. They correspond to raising and lowering indices in tensor notation.

Therefore, the reduced lagrangian for (2.1.3) is

$$\mathcal{J}(g) = \int_0^1 \|\dot{\xi} + \operatorname{ad}^\dagger_\xi \xi\|^2_{\mathfrak{g}}\, \mathrm{d}t\,. \tag{2.3.5}$$

For this Lagrangian, the Euler-Lagrange equation (2.3.3) reads

$$\left(\partial_t + \mathrm{ad}^\dagger_\xi\right)\left(\partial_t \eta + \mathrm{ad}^\dagger_\eta \xi + \mathrm{ad}_\eta \xi\right) = 0 \quad \text{where} \quad \eta := \dot{\xi} + \mathrm{ad}^\dagger_\xi \xi. \tag{2.3.6}$$

While this formula is compact, it is a formal calculation in the case of the group of diffeomorphisms since there is a loss of smoothness which is already present in the acceleration formula (2.3.6). This formula also resembles closely to the Jacobi field equations for such metrics which was used in [MP10] and for which an integral formulation is available [MP10, Proposition 5.5]. This is of course expected due to Formula (2.3.7) below.

There is however a clear obstacle to use reduction for variational analysis since the operator $\mathrm{ad}^\dagger$ is unbounded on the tangent space at identity due to the loss of derivative. Following [EM70], one can use the smooth Riemannian structure on $\mathcal{D}^s$ to check that the functional (2.1.3) is well defined. The following proposition of [GG02] is valid in infinite dimensions:

Proposition 13. *Let (M, g) be an infinite dimensional strong Riemannian manifold and*

$$\Omega_{0,1}(M) := \{x \in H^2([0,1], M) \,|\, x(i) = x_i\,,\ \dot{x}(i) = v_i \text{ for } i = 0, 1\}$$

be the space of paths with first order boundary constraints for given $(x_0, v_0) \in TM$ and $(x_1, v_1) \in TM$. The functional (2.1.3) is smooth on $\Omega_{0,1}(M)$ and

$$\mathcal{J}'(x)(v) = \int_0^1 g\left(\frac{D^2}{Dt^2}v, \frac{D}{Dt}\dot{x}\right) - g\left(R\left(\dot{x}, \frac{D}{Dt}\dot{x}\right), v\right)\,\mathrm{d}t\,.$$

A smooth curve x is a critical point of $\mathcal{J}$ if and only if it satisfies the Riemannian cubic equation

$$\frac{D^3}{Dt^3}\dot{x} - R\left(\dot{x}, \frac{D}{Dt}\dot{x}\right)\dot{x} = 0\,. \tag{2.3.7}$$

The critical points of $\mathcal{J}$ are called Riemannian cubics or cubic polynomials. In Euclidean space, the curvature tensor vanishes and one recovers standard cubic polynomials. In this paper, we will be interested in existence of minimizers for the functional (2.1.3) in the case of the group of diffeomorphisms endowed with a strong right-invariant metric. The existence of minimizers (and the fact that $\mathcal{J}$ satisfies the Palais-Smale condition) does not follow from the corresponding proof in [GG02] since it strongly relies on the finite dimension hypothesis and compactness of balls. Moreover, as shown above, it is not possible to follow the proof of [BV17] since the reduced functional (2.3.5) is not well defined on the tangent space at identity. Therefore, it is useful to write the variational problem in Lagrangian coordinates, taking advantage of the smoothness of the metric.

2.3.2 Main result and strategy of proof

The smoothness of the metric is not enough to deal with the problem of existence of minimizing geodesics and the well-known example is the work of Brenier on generalized solutions of Euler equation [Bre89]. As explained above, a technical important difference is that the L^2 metric on the space of diffeomorphisms is a weak metric in the sense of Ebin and Marsden [EM70], whereas we work with a strong metric. The group of diffeomorphisms (actually H^s diffeomorphisms) endowed with a right-invariant Sobolev metric of order $s > d/2 + 1$ is complete for all the meanings of the Hopf-Rinow theorem as proven in [BV17]. Passing to second-order derivatives has

been less treated from a variational point of view, and to the best of our knowledge it has never been addressed in the case of right-invariant metrics on the group of diffeomorphisms.

We now present the three main steps developed in [TV16]. The first step is to choose a convenient formulation of the acceleration which is done . The first technical choice follows Ebin and Marsden [EM70] and it consists in writing the acceleration in Lagrangian coordinates instead of Eulerian coordinates. The point is to avoid the loss of smoothness of the Eulerian formulation. The second choice which appears the most important from an analytical point of view consists in using the second derivative of the diffeomorphism as the main variable to compute the geodesic equation. We therefore work on $H_0^2([0,1])$ in order to avoid boundary terms. At this step, we strongly use the one dimensional setting. This simple change of variable leads to geodesic equation that have a Hamiltonian formulation enjoying important analytical properties. Let us give an overview of the new set of equations. Now the variable q represents a function in $L^2([0,1])$ and thus the dual space can be identified with $L^2([0,1])$, we have the following formulation, with $\mathcal{H}$ being the Hamiltonian

$$\begin{cases} \dot{q} = \partial_p \mathcal{H}(p,q) = K(q)(p) \\ \dot{p} = -\partial_q \mathcal{H}(p,q) = -B(q)(p,p)\,, \end{cases} \tag{2.3.8}$$

where $K(q)$ is a bounded linear operator on $L^2([0,1])$ which is continuous w.r.t. q for the weak topology. As is well-known, $K(q)$ is the inverse of the metric tensor. The operator $B(q)$ is bounded as a bilinear operator on L^2 and is continuous w.r.t. q for the weak topology. However, it is not continuous w.r.t. to p for the weak topology due to the bilinear structure.

Importantly, the operator B is non-local and therefore, the acceleration functional (2.3.11) written below is not the integral of a Caratheodory type integrand. This non-local term is more precisely $\mathcal{U}(p^2,q)$ where $\mathcal{U}$ is defined by

$$\mathcal{U}(f,q)(x) \overset{\text{def.}}{=} \frac{1}{2}\int_x^1 \eta(q) f \,\mathrm{d}y\,, \tag{2.3.9}$$

where $\eta(q)$ is defined by

$$\eta(q)(x) = \exp\left(\int_0^x q(u)\,\mathrm{d}u\right)\,. \tag{2.3.10}$$

In such cases, there does not exist a general theory of relaxation and the relaxation formulation has to be studied directly.

Thus, the second step consists in studying the relaxation of the acceleration functional, which is the lower semi continuous enveloppe of it. The functional can be written as follows

$$\mathcal{J}(p,q) = \int_0^1 \|K(q)^{1/2}(\dot{p} + B(q)(p,p))\|_{L^2}^2 \,\mathrm{d}t + P(p(1),q(1)) \tag{2.3.11}$$

where P is a relaxation of the endpoint constraint at time 1 which is lower continous for the weak topology. Expanding the quadratic term, we have to deal with the weak limit of $\langle \dot{p}, K(q)\dot{p}\rangle$ denoted by ν and the weak limit of $B(q)(p,p)$ which only involves the weak limit of p^2 denoted by μ. These two weak limits are related to each other.

The relation is given by the following inequality

$$(\partial_t \sqrt{\mu})^2 \leq \nu \,. \tag{2.3.12}$$

In fact, Equation (2.3.12) can be made rigorous using the Fisher-Rao functional which is a convex functional on measures on the time space domain $D = [0,1]^2$ defined by

$$\mathrm{FR}_f(\mu,\nu) = \int_D \frac{1}{4}\frac{\rho_\nu^2}{\rho_\mu} f \,\mathrm{d}\lambda$$

where f is a positive and continuous weight function on D and ρ_ν and ρ_μ are the densities of μ and ν with respect to a dominating measure λ. Now, formula (2.3.12) can be rewritten as

$$\mathrm{FR}(\mu, \partial_t \mu) \leq \nu \,. \tag{2.3.13}$$

as linear operators on continuous positive functions f on the domain D. Therefore, the relaxation of $\mathcal{J}$ will make appear the Fisher-Rao term $\mathrm{FR}(\mu, \partial_t \mu)$. Let us underline that informally, this is the cost associated with the oscillations that are generated on p. However, in order to prove that the relaxation of the functional exactly involves this quantity we have to construct explicitly the oscillations that generate the measure μ at the given cost $\mathrm{FR}(\mu, \partial_t \mu)$. This is a technical step that relies on the construction of solutions to the first equation of the Hamiltonian system and also on an explicit construction of the oscillations. In summary, the main result is thus the following,

Main result. *The relaxation of the acceleration functional,*

$$\mathcal{J}(p,q) = \int_0^1 \left\| \left(\frac{D}{Dt}\dot{q}\right)^\flat \right\|_q^2 \mathrm{d}t + P(p(1), q(1)) \,, \tag{2.3.14}$$

with boundary conditions $p(0), q(0)$ and P a continuous (w.r.t. weak convergence) penalization term can be written as

$$\mathcal{F}(\Delta, p, q) = \mathrm{FR}_\eta(\Delta, \partial_t \Delta) + \int_0^1 \left\| \left(\frac{D}{Dt}\dot{q}\right)^\flat + \pi_q^*(\mathcal{U}(\Delta, q)) \right\|_q^2 \mathrm{d}t + P(p(1), q(1)) \,, \tag{2.3.15}$$

where Δ is a time dependent nonnegative Radon measure and $\partial_t \Delta$ is a Radon measure. In addition, $\|\cdot\|_q$ is the norm of the scalar product on the cotangent space, $\left(\frac{D}{Dt}\dot{q}\right)^\flat$ is the acceleration on the cotangent space, π_q^* is a projection which depends on q through integral terms. The map $\mathcal{U}$ defined in (2.3.9) is linear with respect to Δ.

In particular, at the expense of the Fisher-Rao energy, one could possibly decrease the relaxed energy $\mathcal{F}$. The last step consists in standard analysis of optimality conditions of the above functional by means of convex analysis. Note that this new variational problem is convex in Δ but non convex in (p, q, Δ).

2.3.3 Open questions and applications

We have studied the relaxation of the acceleration in the case of a right-invariant Sobolev metric on the group of diffeomorphisms in one dimension. We have shown, that for the relaxed endpoint constraint, the relaxation of the functional involves the Fisher-Rao functional. Being a convex and one-homogeneous functional on measures, we have derived several optimality conditions by means of convex analysis:

optimality is linked to the existence of a solution to a Riccati equation that involves the acceleration of the curve.

Several interesting questions remain open:
(1) We have proved the existence of an example where the minimizer of the relaxed acceleration functional has a non zero defect measure. In other words, we ask whether all the minimizers of the relaxed acceleration functional satisfy $\Delta = 0$. At least, prove that smooth solutions of the Euler-Lagrange equation are solutions of the relaxed problem at least for short times.
(2) Another important point is to add a boundary hard constraint at time 1 on the couple (p, q). It is probable that the relaxation will involve a different form of the Fisher-Rao metric.
(3) Our proof relies at different steps on the one dimensional case. For instance, the formulation of the geodesic equations was a key point for obtaining the properties on the derivative of the metric. This step can probably be extended to higher dimensional case, i.e. on $\mathrm{Diff}^s([0,1]^d)$ for $s > d/2 + 1$. However, the relaxation of the functional will certainly differ from the Fisher-Rao functional.
(4) In a different direction, we could also have written the Euler-Lagrange equations for the relaxed functional on the space $\mathcal{C}_R$. An interesting question would be to use this Euler-Lagrange equation to derive regularity properties of the solutions from the endpoints such as in the geodesic case.
(5) We also have left behind a technical question about the characterization of the subdifferential of the Fisher-Rao functional similarly to what has been done for the TV norm on functions. This would lead to a finer characterization than the one presented in the paper.

> For practical applications, and in particular interpolation of a time sequence of shapes, one has often the freedom to design the cost functional as well. Therefore, as long as the interpolation is feasible, it is possible to propose a variational model that will be well posed, by measuring the acceleration with a stronger norm, for instance Sobolev of order $s' > s + 1$. This type of approach has been implemented in [SVN15].

2.4 Boundary value problem on the space of curves

This section is based on joint work with G. Nardi (Post-doc) together with G. Peyré in [CNPV16].

Motivated by our work on piecewise rigid curve evolution [CNPV16], we study in [NPV16] on a Finsler metric on the space of curves embedded in $\mathbb{R}^2$. This Finsler metric is designed to reproduce piecewise regular deformations. We actually proposed and studied a BV metric on the first derivative of the curve, therefore denoted by BV^2. For this metric, piecewise regular motion is possible and in fact, favored due to the BV penalization. The main contribution of the paper consists in proving existence to the boundary value problem, i.e. there exists a shortest path between any two curves in the space. The proof of this result is based on two ingredients. The first one consists in using the constant speed reparametrization and the second one puts forward a somewhat unusual approach in this field: it is based on the construction of a martingale with values in $BV^2(S_1, \mathbb{R})$, where S_1 is the unit circle, which appears to be a bounded martingale in $H^1(S_1, \mathbb{R})$ and therefore, since

this space has the Radon-Nicodym property, the martingale converges p.s., its limit defining a length minimizing path.

We also applied the same method to the case of Sobolev metrics to prove existence of length minimizing paths to give a complementary result to [BMM19].

Chapter 3

Unbalanced optimal transport and the Camassa-Holm equation

When this work started, no equivalent to the L^2 Wasserstein distance in the case of nonnegative Radon measures was available. We introduced this new distance, using a dynamic formulation. We proved the equivalence of this dynamic model with a static formulation, in view of applying entropic regularization for numerical computation. This chapter is a summary of some of two related different contributions: (1) Unbalanced optimal transport developed in [CSPV16, CPSV15, GV16]. Note that this extension of optimal transport is the model that I proposed as a starting project for the PhD of Lenaïc Chizat. (2) The link of this unbalanced optimal transport model with the Camassa-Holm equation explained in [GV16], which is similar to the one between standard optimal transport and the incompressible Euler equation. In particular, this gives a new geometric interpretation of the Camassa-Holm and new results on the properties of geodesics.

3.1 Introduction

For practical applications in imaging or machine learning, a major bottleneck of standard optimal transport is that it is restricted to measures that have the same total mass. A usual approach to get around this problem consists in dividing all the densities by their total mass. This renormalization is global and it would be desirable to take change of mass into account, locally in space and to quantify the amount of transportation, growth and decay of mass that a shape can occur. Such models have been proposed in the literature and we now review some of them.

3.1.1 Previous works

Extending standard optimal transport to more general measures can be addressed in, at least, two different ways, characterized by their formulation: "dynamic" (with a virtual time) or "static" (without time) as described in the introduction Chapter 1.

Static formulations of unbalanced optimal transport. The simplest way to define a metric that extends the Wasserstein metric to the space of nonnegative Radon measure is probably the bounded Lipschitz distance, which is an extension of the L^1 Wasserstein metric by its formulation as a dual norm on the space of Lipschitz functions. This class of distances was proposed in the early work of Kantorovich and Rubinstein in [KR58], later developed by [Han99]. These directions have been

followed in practical applications by [Gui02] and by [GPC15] for applications in neuroimaging.
Slightly different, partial optimal transport, first introduced in computer vision works [RGT97, PW08] and theoretically studied in [CM10], consists in transferring a given fixed amount of mass between two measures ρ_0 and ρ_1 as cheaply as possible. In [CM10] and [Fig10], regularity results on the boundary between the active and inactive sets are proved.

Dynamic formulations of unbalanced optimal transport. Since the introduction of the Benamou-Brenier formulation [BB00], several dynamical formulations [Ben03, MRSS15, LM13, PR14, PR13] have been proposed. In [Ben03], the marginal constraint in the original BB formulation is relaxed using the L^2 norm. In the rest these works, a source term is introduced in the continuity equation. They differ in the way this source is penalized or chosen. As noted in [CSPV16], optimal partial transport is tightly related to the generalized transport proposed in [PR14, PR13] which allows a dynamic formulation of the optimal partial transport problem. In [MRSS15], the penalization follows the work of [TY05b] where an L^2 penalization is used, and consequently, cannot be considered as a proper generalization of optimal transport.

3.1.2 Simultaneous works

Perhaps a bit surprisingly, no equivalent of the L^2 Wasserstein distance was proposed until 2015, when many simultaneous works introduced the same model, sometimes for different purposes. This works include ours [CSPV16, CPSV15] motivated by imaging applications, the work of Liero, Mielke and Savaré [LMS15, LMS16] motivated by gradient flows as well as [KMV16]. Rezakhanlou, in [Rez15], introduced the model to generalize optimal transport of densities to contact structures. All these developments were based on the dynamical formulation of optimal transport and only one developed directly the static formulation [FZM+15], motivated by machine learning problems. The authors relaxed the marginal constraints of standard optimal transport using the Kullback-Leibler Let us give an attempt to compare the contributions at a rather high-level point of views: The first[1] article to appear on the web was [KMV16]. The authors proposed the dynamic model and showed that this new metric metrizes weak convergence on the space of positive Radon measures and looked at some gradient flow applications. The articles [LMS15, LMS16] proposed the dynamical model and show the equivalence with the static formulation and study it in very general spaces, and in particular the dual formulation is studied in details. Among many other results such as the metrizability of weak convergence or the generalization of the static formulation, they show a perhaps surprising link between optimal transport on the space of position and mass, called the cone, and the new metric.[2] Another important object introduced in [LMS15] is the so-called Gaussian-Hellinger distance which can be seen as a generalization of the Wasserstein-Fisher-Rao metric. Interestingly, the induced length space by the Gaussian-Hellinger distance is the Wasserstein-Fisher-Rao metric. In comparison to these works, the contributions of our articles [CPSV15, CPSV15] which are not contained elsewhere are (1) a geometrical point of view by generalizing Otto's

[1] According to the historical section at the end of [LMS15], it appears that Liero, Mielke and Savaré are certainly the first to have presented this new model.

[2] It is precisely this link which guided our new formulation of the Camassa-Holm equation as an incompressible Euler equation.

Riemannian submersion, (2) a general dynamical framework in which the other L^p equivalent distances can be formulated and in particular the link to partial optimal transport, (3) a numerical algorithm for the dynamic formulation which implements a Douglas-Rachford's method.

3.2 The Wasserstein-Fisher-Rao metric

This section is based on joint works with L. Chizat (PhD student co-advised with G. Peyré), B. Schmitzer (Post-doc) and G. Peyré in [CSPV16, CPSV15] and with T. Gallouët in [GV16].

Since our goal was to build a metric that has a Riemannian-like behaviour, our starting point is the dynamic formulation of Benamou and Brenier, who highlighted the use of the metric tensor associated with optimal transport. We did not want to design any metric on nonnegative densities but a metric that shares as many properties as possible with standard optimal transport. Our proposal consisted in using the infimal convolution of the metric tensor of optimal transport given by the Benamou-Brenier formula and the so-called Fisher-Rao metric tensor, defined below.

3.2.1 Definition and first properties

The continuity equation enforces the mass conservation property in the Benamou-Brenier formulation (1.2.12). This constraint can be relaxed by introducing a source term $\mu \in C^\infty(M, \mathbb{R})$,

$$\dot\rho = -\operatorname{div}(\rho\,\xi) + \mu\,. \tag{3.2.1}$$

For a given variation of the density $\dot\rho$, there exist a priori many couples (v, μ) that reproduce this variation. Following [TY05b], it can be determined via the minimization of the norm of (v, μ), for a given choice of norm. The penalization of μ was chosen in [MRSS15] as the L^2 norm but a natural choice is rather the Fisher-Rao metric

$$\mathrm{FR}^2(\mu) = \int_M \frac{\mu(t,x)^2}{\rho(t,x)}\,\operatorname{dvol}(x)\,,$$

because it is (one) homogeneous w.r.t. the couple (ρ, μ). It is actually similar to the Benamou-Brenier tensor. Indeed, let us rewrite Equation (3.2.2), in the (non-equivalent) formulation

$$\dot\rho = -\operatorname{div}(\rho\,\xi) + \alpha\rho\,. \tag{3.2.2}$$

Then, the functional FR is the L^2 norm of the growth rate w.r.t. the density ρ since it can be written as $\int_M \alpha(t,x)^2\rho(t,x)\,\operatorname{dvol}(x)$ where α is the growth rate $\alpha(t,x) \overset{\text{def.}}{=} \frac{\mu(t,x)}{\rho(t,x)}$. Note in particular that this action is 1-homogeneous with respect to the couple (μ, ρ). This point is important for convex analysis properties and especially, in order to define the action functional on singular measures via the same formula.

Thus, the Wasserstein-Fisher-Rao functional also known as Hellinger-Kantorovich [LMS15] is simply given by the infimal convolution between the Wasserstein and the Fisher-Rao metric tensor, that is

$$\mathcal{J}(\xi, \alpha) = \int_0^1 \int_M (\|\xi(x)\|^2 + \alpha(x)^2)\rho(x)\,\operatorname{dvol}(x)\,\mathrm{d}t\,, \tag{3.2.3}$$

under the generalized continuity equation constraint and the time boundary constraints

$$\dot{\rho} = -\operatorname{div}(\rho\,\xi) + \alpha\rho$$
$$\rho(0,x) = \rho_0(x) \text{ and } \rho(1,x) = \rho_1(x)\,.$$

Definition 14 (WF metric)**.** Let (M,g) be a smooth Riemannian manifold compact and without boundary, $a,b \in \mathbb{R}_+^*$ be two positive real numbers and $\rho_0,\rho_1 \in \mathcal{M}_+(M)$ be two nonnegative Radon measures. The Wasserstein-Fisher-Rao metric is defined by

$$\mathrm{WF}^2(\rho_0,\rho_1) = \inf_{\rho,\mathsf{m},\mu} \mathcal{J}(\rho,\mathsf{m},\mu)\,, \tag{3.2.4}$$

where

$$\mathcal{J}(\rho,\mathsf{m},\mu) = a^2\int_0^1\int_M \frac{g^{-1}(x)(\tilde{\mathsf{m}}(t,x),\tilde{\mathsf{m}}(t,x))}{\tilde{\rho}(t,x)}\,\mathrm{d}\nu(t,x) + b^2\int_0^1\int_M \frac{\tilde{\mu}(t,x)^2}{\tilde{\rho}(t,x)}\,\mathrm{d}\nu(t,x) \tag{3.2.5}$$

over the set (ρ,m,μ) satisfying $\rho \in \mathcal{M}([0,1]\times M)$, $\mathsf{m} \in (\Gamma^0_M([0,1]\times M, TM))^*$ which denotes the dual of time dependent continuous vector fields on M (time dependent sections of the tangent bundle), $\mu \in \mathcal{M}([0,1]\times M)$ subject to the constraint

$$\int_0^1\int_M \partial_t f\,\mathrm{d}\rho + \int_0^1\int_M \mathsf{m}(\nabla_x f) - f\mu\,\mathrm{d}\nu = \int_M f(1,\cdot)\,\mathrm{d}\rho_1 - \int_M f(0,\cdot)\,\mathrm{d}\rho_0 \tag{3.2.6}$$

satisfied for every test function $f \in C^1([0,1]\times M,\mathbb{R})$. Moreover, ν is chosen such that ρ,m,μ are absolutely continuous with respect to ν and $\tilde{\rho},\tilde{\mathsf{m}},\tilde{\mu}$ denote their Radon-Nikodym derivative with respect to ν.

Remark 15. *Note that, in the previous definition, the divergence operator* $\operatorname{div}(\cdot)$ *is defined by duality on the space of* C^1 *functions. In addition, since the functions in the integrand of formula* (3.2.4) *are one homogeneous with respect to the triple of arguments* $(\tilde{\rho},\tilde{\mathsf{m}},\tilde{\mu})$*, the functional does not depend on the choice of* ν *which dominates the measures. Last, the Radon-Nikodym theorem applied to the measure* m *gives* $\mathsf{m} = \tilde{\mathsf{m}}\nu$ *where* $\tilde{\mathsf{m}}$ *is a measurable section of* T^*M*.*

The following property can be proven:

Proposition 16. *The* WF^2 *functional is convex and positively one homogeneous on the space of Radon measures. Moreover,* WF *defines a distance on the space of nonnegative Radon measures which is continuous w.r.t. to the weak-* topology.*

This dynamical formulation enjoys most of the analytical properties of the initial Benamou-Brenier formulation (1.2.12) and especially convexity. An important consequence is the existence of optimal paths in the space of time-dependent measures [CSPV16] by application of the Fenchel-Rockafellar duality theorem as stated in the next proposition. We omit the proof here since it is similar to the one given in [CPSV15] in the Euclidean case and it is also proven in more general spaces in [LMS15].

Proposition 17 (Hamilton-Jacobi)**.** *There exists a minimum to the minimization problem* (3.2.4) *and it holds*

$$\mathrm{WF}^2(\rho_0,\rho_1) = \sup_{q\in C}\int_M q(1,\cdot)\,\mathrm{d}\rho_1 - \int_M q(0,\cdot)\,\mathrm{d}\rho_0 \tag{3.2.7}$$

where C is the set of functions $q \in C^1([0,1]\times M,\mathbb{R})$ such that

$$\partial_t q(t,x) + \frac{1}{2a^2}\|\nabla q(t,x)\|^2 + \frac{1}{2b^2} q(t,x)^2 \leq 0\,. \tag{3.2.8}$$

At this point, it is interesting to note that from a numerical point of view, this formulation can be used to implement Douglas-Rachford algorithms, as done in [CSPV16]. However, recent numerical advances in optimal transport have introduced the use of relative entropy on the coupling plan in the static formulation as a regularization of the linear programming problem. Therefore, motivated by computational purposes, we asked for the existence of an equivalent static formulation for this new unbalanced optimal transport model. As a key step for its justification, we generalized the Riemannian submersion highlighted by Otto, as explained in the introduction.

3.2.2 The geometric interpretation.

Not only analytical properties of standard optimal transport are conserved but also some interesting geometrical properties such as the Riemannian submersion highlighted by Otto, as explained in the introduction. More precisely, the group of diffeomorphisms can be replaced by a semi-direct product of group between $\mathrm{Diff}(M)$ and the space $\Lambda(M) = C^\infty(M,\mathbb{R}_+^*)$ which is a group under pointwise multiplication. *Actually, the continuity equation with a source term can be interpreted as the infinitesimal action of this semi-direct product of group.* We define the semi-direct product of group between $\mathrm{Diff}(M)$ and $\Lambda(M)$ in order to turn the map π defined by

$$\begin{aligned}
&\pi : (\mathrm{Diff}(M) \ltimes_\Psi \Lambda(M)) \times \mathrm{Dens}(M) \mapsto \mathrm{Dens}(M)\\
&\pi\left((\varphi,\lambda),\rho\right) \overset{\text{def.}}{=} \varphi\cdot\lambda\,\varphi_*\rho = \varphi_*(\lambda\rho)
\end{aligned}$$

into a left-action of the group $\mathrm{Diff}(M)\ltimes_\Psi\Lambda(M)$ on the space of (generalized) densities. The group composition law is defined by:

$$(\varphi_1,\lambda_1)\cdot(\varphi_2,\lambda_2) = (\varphi_1\circ\varphi_2, (\lambda_1\circ\varphi_2)\lambda_2) \tag{3.2.9}$$

It is obvious that the metric written in Formula (3.2.3) is of the form defined in Proposition 3. Therefore, it explains why Otto's Riemannian submersion can be extended. What remains to be computed is an explicit formula for the metric on the group. This formula is obtained passing from Eulerian to Lagrangian coordinates by

$$\begin{aligned}
G(\varphi,\lambda)((X_\varphi,X_\lambda),&(X_\varphi,X_\lambda)) = \int_M g(v,v)\rho\,\mathrm{d}x + \int_M \alpha^2\rho\,\mathrm{d}x\\
&= \int_M g(X_\varphi\circ\varphi^{-1}, X_\varphi\circ\varphi^{-1})\varphi_*(\lambda\rho_0)\mathrm{d}x + \int_M (X_\lambda\lambda^{-1})^2\circ\varphi^{-1}\varphi_*(\lambda\rho_0)\mathrm{d}x\\
&= \int_M g(X_\varphi,X_\varphi)\lambda\rho_0\,\mathrm{d}x + \int_M \frac{1}{\lambda}X_\lambda^2\,\rho_0\,\mathrm{d}x\,.
\end{aligned} \tag{3.2.10}$$

Therefore, we see appearing a Riemannian metric on the space $M\times\mathbb{R}_+^*$ which is $mg + \frac{\mathrm{d}m^2}{m}$. This metric, up to the change of variable $r^2 = m$ is a cone metric, as usually presented in Riemannian geometry. Let us motivate the introduction of the cone metric from a different perspective by first discussing informally what happens for a particle of mass $m(t)$ at a spatial position $x(t)$ in a Riemannian manifold (M,g)

under the generalized continuity constraint (3.2.2); If the control variables $v(t,x)$ and $\alpha(t,x)$ are Lipschitz, then the solution of the continuity equation with initial data $m(0)\delta_{x(0)}$ has the form $m(t)\delta_{x(t)}$ where $m(t) \in \mathbb{R}_+^*$ is the mass of the Dirac measure and $x(t) \in M$ its location; The system reads

$$\begin{cases} \dot{x}(t) = v(t,x(t)) \\ \dot{m}(t) = \alpha(t,x(t))m(t)\,, \end{cases} \tag{3.2.11}$$

which is directly obtained by duality since the flow map associated with (v,α) is well defined. The action now reads $\int_0^1 a^2|v(x(t))|^2 m(t) + b^2\frac{\dot{m}(t)^2}{m(t)}\,\mathrm{d}t$. Thus, considering the particle as a point in $M \times \mathbb{R}_+^*$, the Riemannian metric seen by the particle is $a^2 mg + b^2\frac{\mathrm{d}m^2}{m}$. As said in the previous section, this space is isometric to the standard Riemannian cone defined below.

Definition 18 (Cone)**.** Let (M,g) be a Riemannian manifold. The cone over M denoted by $\mathcal{C}(M)$ is the quotient space $(M\times\mathbb{R}_+)\ /\ (M\times\{0\})$. The cone point $M\times\{0\}$ is denoted by $\mathcal{S}$. The cone will be endowed with the metric $g_{\mathcal{C}(M)} \overset{\text{def.}}{=} r^2 g + \mathrm{d}r^2$ defined on $M\times\mathbb{R}_+^*$ and r is the variable in $\mathbb{R}_+^*$.

The explicit formula for the distance on the Riemannian cone can be found in [BBI01] and the isometry is given by the square root change of variable on the mass, as stated in the following proposition.

Proposition 19. *Let a,b be two positive real numbers and (M,g) be a Riemannian manifold. The distance on $(M\times\mathbb{R}_+^*, a^2 mg + \frac{b^2}{m}\mathrm{d}m^2)$ is given by*

$$d((x_1,m_1),(x_2,m_2))^2 = 4b^2\left(m_2+m_1-2\sqrt{m_1m_2}\cos\left(\frac{a}{2b}\min(d_M(x_1,x_2),\pi)\right)\right)\,. \tag{3.2.12}$$

The space $(M\times\mathbb{R}_+^*, mg + \frac{1}{4m}\mathrm{d}m^2)$ is isometric to $(\mathcal{C}(M), g_{\mathcal{C}(M)})$ by the change of variable $r=\sqrt{m}$. If c is a geodesic for the metric $\frac{a^2}{4b^2}g$, an isometry $S:\mathbb{C}\setminus\mathbb{R}_- \to M\times\mathbb{R}_+^*$ is defined by $S(\sqrt{m}e^{i\theta}) = (c(\theta), 2bm)$.

In physical terms, it implies that mass can "appear" and "disappear" at finite cost. In other words, the Riemannian cone is not complete but adding the cone point, which represents $M\times\{0\}$, to $M\times\mathbb{R}_+^*$ turns it into a complete metric space when M is complete. Importantly, the distance associated with the cone metric (3.2.12) is 1-homogeneous in (m_1,m_2).

To be coherent with this literature, we decided to change the group action in order to obtain this cone metric. We thus introduce the group $\mathrm{Diff}(M)\ltimes_\Psi \Lambda_{1/2}(M)$ where $\Lambda_{1/2}(M)$ is also the space of positive functions on M but the action on $\mathrm{Dens}(M)$ now differs and is given by

$$\begin{aligned} &\pi : \big(\mathrm{Diff}(M)\ltimes_\Psi \Lambda_{1/2}(M)\big)\times\mathrm{Dens}(M) \mapsto \mathrm{Dens}(M) \\ &\pi\left((\varphi,\lambda),\rho\right) \overset{\text{def.}}{=} \varphi_*(\lambda^2\rho)\,. \end{aligned} \tag{3.2.13}$$

Note that the infinitesimal action is slightly changed to

$$(v,\alpha)\cdot\rho = -\operatorname{div}(v\rho) + 2\alpha\rho\,. \tag{3.2.14}$$

Indeed, one has

$$(\varphi(t),\lambda(t))\cdot\rho = \mathrm{Jac}(\varphi(t)^{-1})(\lambda^2(t)\rho)\circ\varphi^{-1}(t)\,.$$

First recall that $\partial_t \varphi(t) = v \circ \varphi(t)$ and $\partial_t \lambda = \lambda(t)\alpha \circ \varphi(t)$. Once evaluated at time $t = 0$ where $\varphi(0) = \mathrm{Id}$ and $\lambda(0) = 1$, the differentiation with respect to φ gives $-\operatorname{div}(v\rho)$ and the second term $2\alpha\rho$ is given by the differentiation with respect to λ.

The semidirect product of groups is endowed with the L^2 metric on the space of maps between M endowed with the reference measure ρ_0 and the space $M \times \mathbb{R}_+$ endowed with the metric defined above. Let us recall the definition of this L^2 metric.

Definition 20 (L^2 metric)**.** Let M be a manifold endowed with a measure μ and (N, g) be a Riemannian manifold. Consider a measurable map $\varphi : M \to N$ and two measurable maps, $X, Y : M \mapsto TN$ such that $p_N(X) = p_N(Y) = \varphi$ where $p_N : TN \to N$ is the natural projection. Then, the L^2 Riemannian metric w.r.t. to the volume form μ and the metric g at point φ is defined by

$$\langle X, Y \rangle_\varphi = \int_M g(\varphi(x))(X(\varphi(x)), Y(\varphi(x)))\, \mathrm{d}\mu(x)\,. \tag{3.2.15}$$

This is probably the simplest type of (weak) Riemannian metrics on spaces of mappings and it has been studied in details in [EM70] in the case $L^2(M, M)$ and also in [FG89] where, in particular, the curvature is computed for $L^2(M, N)$ for N an other Riemannian manifold. Note in particular that this metric is *not* the right-invariant metric L^2 on the semidirect product of groups as in [HMR98] which would lead to an EPDiff equation on a principal fibre bundle as developed in [GTV13].

3.2.3 The Monge-Kantorovich formulation

A Monge-type formulation By the geometric point of view developed above, it is possible to derive a Monge formulation directly and also to derive a pre-formulation of the Monge-Ampère equation. We first derive formally the equations for which a precise meaning will be given in the next sections. The first important consequence of the L^2 metric on the group and the Riemannian submersion is that one can define a Monge formulation of the Wasserstein-Fisher-Rao metric as follows:

$$\mathrm{WF}(\rho_0, \rho_1) = \inf_{(\varphi,\lambda)} \left\{ \|(\varphi, \lambda) - (\mathrm{Id}, 1)\|_{L^2(\rho_0)} \,:\, \varphi_*(\lambda^2 \rho_0) = \rho_1 \right\}. \tag{3.2.16}$$

As in the case of standard optimal transport, it is therefore compulsory to find the relaxation of this Monge formulation, which we also call a Kantorovich formulation.

The Kantorovich-type formulation The generalization to any positive Radon measures of the Kantorovich relaxation requires the definition of a convex functional which is one-homogeneous on the space of Radon measures described below. The next theorem is proven in [CPSV15] and in another form in [LMS15], both only in the Euclidean case. We extend it to a Riemannian setting in [GV16].

Theorem 21. *For two given positive Radon measures ρ_1, ρ_2, we define, for $\mathcal{M}_+(M^2)$ the space of positive Radon measures on M^2,*

$$\Gamma(\rho_1, \rho_2) \overset{\text{def.}}{=} \left\{ (\gamma_1, \gamma_2) \in \left(\mathcal{M}_+(M^2)\right)^2 : p^1_* \gamma_1 = \rho_1,\, p^2_* \gamma_2 = \rho_2 \right\}, \tag{3.2.17}$$

where p^1 and p^2 denote the projection on the first and second factors of the product M^2. The variational problem associated with the Wasserstein-Fisher-Rao distance is

$$\mathrm{WF}^2(\rho_1, \rho_2) = \min_{(\gamma_1,\gamma_2)\in\Gamma(\rho_1,\rho_2)} \int_{M^2} d^2_{C(M)} \left((x, \frac{\mathrm{d}\gamma_1}{\mathrm{d}\gamma}), (y, \frac{\mathrm{d}\gamma_2}{\mathrm{d}\gamma}) \right) \mathrm{d}\gamma(x, y)\,, \tag{3.2.18}$$

where $d^2_{C(M)}$ is the square of the cone distance given in definition 18 and γ is any measure that dominates ρ_1 and ρ_2.

Remark 22. *The fact that $S(\gamma_1,\gamma_2) \overset{\text{def.}}{=} \int_{M^2} d^2_{C(M)}\left((x,\frac{\mathrm{d}\gamma_1}{\mathrm{d}\gamma}),(y,\frac{\mathrm{d}\gamma_2}{\mathrm{d}\gamma})\right)\mathrm{d}\gamma(x,y)$ is well defined follows from the application of [Roc71, Theorem 5]. It does not depend on the choice of the measure γ since the function d^2 is one-homogeneous w.r.t. the mass variables. As a consequence of Rockafellar's theorem [Roc71, Theorem 5], S is convex and lower-semicontinuous on the space of Radon measures as the Legendre-Fenchel transform of a convex functional on the space of continuous functions.*

We also state without proof the dual formulation which is given by the application of Fenchel-Rockafellar duality theorem, see [CSPV16].

Proposition 23. *It holds*

$$\mathrm{WF}^2(\rho_0,\rho_1) = \sup_{(\phi,\psi)\in C(M)^2} \int_M \phi(x)\,\mathrm{d}\rho_0(x) + \int_M \psi(y)\,\mathrm{d}\rho_1(y) \tag{3.2.19}$$

subject to $\forall (x,y)\in M^2$,

$$\begin{cases} \phi(x)\le 1\,, \quad \psi(y)\le 1\,, \\ (1-\phi(x))(1-\psi(y)) \ge \cos^2\left(d(x,y)\wedge(\pi/2)\right)\,. \end{cases} \tag{3.2.20}$$

A reformulation of this linear optimization problem is

$$\mathrm{WF}^2(\rho_0,\rho_1) = \sup_{(z_0,z_1)\in C(M)^2} \int_M 1-e^{-z_0(x)}\,\mathrm{d}\rho_0(x) + \int_M 1-e^{-z_1(y)}\,\mathrm{d}\rho_1(y) \tag{3.2.21}$$

subject to $\forall (x,y)\in M^2$,

$$z_0(x)+z_1(y) \le -\log\left(\cos^2\left(\min(d(x,y),\pi/2)\right)\right)\,. \tag{3.2.22}$$

Interestingly, the last formulation can be further reduced since the exponential $r\mapsto e^r$ is the Fenchel-Legendre conjugate associated with the Kullback-Leibler divergence defined below. Therefore, using duality again, it is proven in [LMS15] that the static problem in Proposition 23 can be rewritten as

$$\begin{aligned}\mathrm{WF}^2(\rho_0,\rho_1) = \inf_{\gamma\in\mathcal{M}_+(M^2)} &\mathrm{KL}(\mathrm{Proj}^1_*\,\gamma,\rho_0) + \mathrm{KL}(\mathrm{Proj}^2_*\,\gamma,\rho_1) \\ &- \int_{M^2} \log\left(\cos^2\left(\min(d(x,y),\pi/2)\right)\right)\mathrm{d}\gamma(x,y)\end{aligned} \tag{3.2.23}$$

with

$$\mathrm{KL}(\mu,\nu) = \int \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right)\mathrm{d}\nu + |\nu| - |\mu|\,,$$

the Kullback-Leibler divergence. Formulation (3.3.1) of unbalanced optimal transport and its extensions have been intensively developed in [LMS15], where generalizations of this metric are studied in spaces such as Hausdorff topological spaces endowed with a (pseudo) distance satisfying mild conditions.

We only focused in this chapter on the equivalent to the L^2 Wasserstein distance, however, the dynamical approach readily extends to any L^p Wasserstein - L^q Fisher-Rao type of Lagrangian, which gives the framework for unbalanced optimal transport in general, see [CPSV15]. For instance, it encompasses the partial optimal transport model which is an L^2 Wasserstein - L^1 Fisher-Rao (which means TV) type of

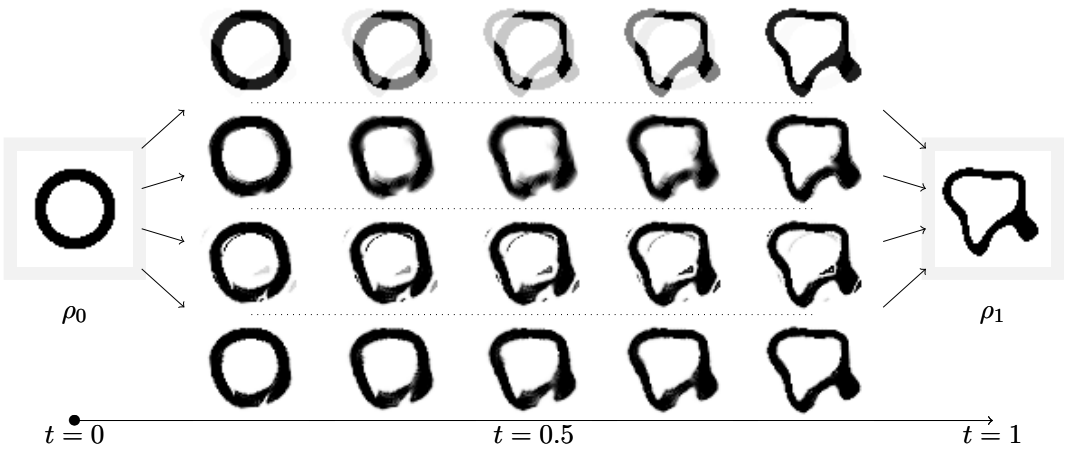


FIGURE 3.1: Geodesics between ρ_0 and ρ_1 having same total mass (number of black pixels). The first row is pure Fisher-Rao, the second is standard L^2 Wasserstein distance, the third one is partial optimal transport and the last one is Wasserstein-Fisher-Rao.

metric. Note that the corresponding static formulation requires the computation of the cost on the cone $M \times \mathbb{R}_+^*$ which is not always available in closed form.

In Figure 3.1, we compare the different geodesics between two binary densities of the same total mass (simulated by the use of Douglas-Rachford algorithm on the dynamical formulation). Thus, one can apply standard optimal transport and compare it to the Wasserstein-Fisher-Rao metric and other unbalanced metrics. In particular, we clearly see that optimal transport has to move masses on rather long distances to explain the deformation. Therefore, this generalization of optimal transport can be seen as a regularization with respect to noise in the data.

3.3 Entropic regularization and scaling algorithms

This section is based on joint works with L. Chizat, B. Schmitzer (Post-doc) and G. Peyré in [CPSV16].

Entropic regularization simply consists in adding the negative entropy of the coupling plan to the optimal transport functional. Entropic regularization goes back to Schrödinger [Léo14, Léo12] but it was introduced for numerical purposes in [Cut13]. This approach was then further developed and applied to multimarginal problems such as Wasserstein barycenters (Karcher means) in [SdGG+15, BCC+15, CD14]. The main reason for the use of relative entropy as a regularizer is that it leads to a particularly simple algorithm. This algorithm enables fast computations of an approximation of the optimal coupling plan using an alternate projection algorithm, known as Sinkhorn algorithm [Sin64].

Developing efficient numerical algorithm was our motivation to find a corresponding static formulation. The two static formulations (3.2.18) and (3.3.1), especially the last one, enable the introduction of the entropic regularization on the coupling plan γ. This direction is developed in [CPSV16] and we briefly describe it hereafter. Let ε a positive real and two reference measures γ_1, γ_2 on M which dominate respectively ρ_1, ρ_2, consider the previous functional regularized using the negative entropy of $\tilde{\gamma}$, the density of γ w.r.t. $\gamma_0 = \gamma_1 \otimes \gamma_2$. The negative entropy is

defined as $H(\gamma) = \int_{M^2} \tilde{\gamma}(x,y) \log(\tilde{\gamma}(x,y)) \, \mathrm{d}\gamma_0(x,y)$. We introduce

$$\mathrm{WF}_\varepsilon^2(\rho_1,\rho_2) = \inf_{\gamma \in \mathcal{M}_+(M^2)} \mathrm{KL}(\mathrm{Proj}_*^1 \gamma, \rho_1) + \mathrm{KL}(\mathrm{Proj}_*^2 \gamma, \rho_2) \\ - \int_{M^2} \log(\cos^2(d(x,y) \wedge (\pi/2))) \, \mathrm{d}\gamma(x,y) + \varepsilon H(\gamma) \,. \quad (3.3.1)$$

The scaling algorithm which is a generalization of the Sinkhorn algorithm for standard optimal transport is an alternate minimization on the dual problem which reads, with $c(x,y) = -\log(\cos^2(d(x,y) \wedge (\pi/2)))$,

$$\mathrm{WF}_\varepsilon^2(\rho_0,\rho_1) = \sup_{p_1,p_2} - \int_M (e^{-p_1(x)} - 1) \, \mathrm{d}\rho_1(x) \; - \int_M (e^{-p_2(y)} - 1) \, \mathrm{d}\rho_2(y) \\ - \varepsilon \int_{M^2} e^{(p_1(x)+p_2(y)-c(x,y))/\varepsilon)} \, \mathrm{d}\gamma_0(x,y) \,, \quad (3.3.2)$$

where the supremum is taken on functions in $L^\infty(M,\gamma_1) \times L^\infty(M,\gamma_2)$ as explained in [CPSV16]. The primal solution can be written as

$$\tilde{\gamma}(x,y) = a(x) e^{-c(x,y)/\varepsilon)} b(y) \,, \quad (3.3.3)$$

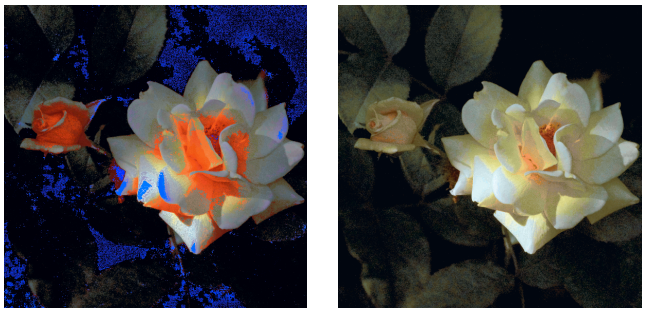
where $a(x) = e^{p_1(x)/\varepsilon}$ and $b(y) = e^{p_2(y)/\varepsilon}$. The alternate minimization has a closed formulation and the generalized Sinkhorn algorithm consists in iterating, after the initialization of $a(x), b(y)$:

$$a(x) \leftarrow (\rho_1(x)/\Sigma_1(x))^{1/(\varepsilon+1)} \\ b(y) \leftarrow (\rho_2(y)/\Sigma_2(y))^{1/(\varepsilon+1)} \,,$$

where $\Sigma_1(x) = \int_M b(y) e^{-c(x,y)/\varepsilon} \, \mathrm{d}\gamma_2(y)$ and $\Sigma_2(y) = \int_M a(x) e^{-c(x,y)/\varepsilon} \, \mathrm{d}\gamma_1(x)$. Similarly to the Sinkhorn algorithm for standard optimal transport [Cut13], this algorithm has linear convergence rate w.r.t. a Hilbert type metric, see [CPSV16, Section 3.5]. Note that this alternate algorithm can be generalized to other divergences than KL, with preferably a closed formula for the alternate minimization for computational efficiency, but the linear convergence proof need not apply.
Finally, let us remark that, in this algorithm, the cost c can differ from the possibly surprising $-\log(\cos^2(\min(d(x,y),\pi/2)))$. The distance on the Euclidean space has been proposed in [FZM+15] and [LMS15] showed that its induced length space is the Wasserstein-Fisher-Rao metric. Last, introducing positive weights on the Kullback-Leibler divergences in (3.3.1) is also possible but, depending on this parameters, the resulting object might not be a distance.

In [CPSV16], applications to the transport of histograms is shown on color transfer. Figure 3.2 shows the initial and final images on which the color histograms are computed and the results of the color transfer. In comparison with standard optimal transport, the Kullback-Leibler penalization gives a visually more satisfying result, see Figure 3.3. We refer to [CPSV16] for more details and where other applications are shown on gradient flows (Hele-Shaw type model) and on barycenters (Karcher means) for this new distance.

FIGURE 3.2: Transporting the color histograms: initial and final images.



Optimal transport Range constraint

Kullback-Leibler Total variation

FIGURE 3.3: Different metrics for the color transfer application. The Kullback-Leibler divergence shows a visually satisfying result in comparison with standard optimal transport.

3.4 Perspectives

This section was devoted to the extension of the L^2 Wasserstein metric to the space of nonnegative Radon measures. We discuss hereafter other extensions of the L^2 Wasserstein metric which seem natural or of interest.

First, such an extension to signed Radon measure seems difficult [Mai12], although some ad-hoc constructions can be proposed.

In a similar direction, an interesting perspective of research consists in extending these optimal transport metrics to cone-valued measures, and especially L^2 type of metrics. One cone of particular interest for applications is the cone of positive semidefinite matrices, for which several works have appeared in the past few years [NGT15, NG14, CGT16]. In particular, since these metrics should metrize the weak convergence of measures, significantly different behavior than standard optimal transport can be expected. Let us explain what we expect when the cone is the space of PSD matrices $S_2^+(\mathbb{R})$ and the underlying space is $\mathbb{R}$. Let ε be a positive real number, and using complex plane notations, define $\mu_1 = i \otimes i\delta_0 + 1 \otimes 1\delta_\varepsilon$ and $\mu_2 = e^{i\pi/4} \otimes e^{i\pi/4}\delta_0 + e^{i3\pi/4} \otimes e^{i3\pi/4}\delta_\varepsilon$. The total mass of μ_1, equal to Id is also equal to that of μ_2, thus if ε is small, the optimal transport distance should be close to zero, since the optimal transport distance should metrize the weak convergence of measures. Indeed, the path shown in Figure 3.4 is of order ε and it exhibits a "shock"

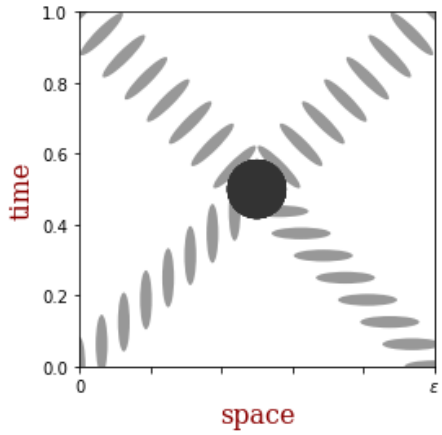


FIGURE 3.4: A possible trajectory from μ_1 at time 0 to μ_2 at time 1 whose cost is of order ε.

between the two Dirac masses at midpoint in time. Actually, at time $1/2$, where the two tensors add up to identity, they can be decomposed an infinite number of ways. If there were no transport and pure transformation of the matrices, the cost would be of order $\sin(\pi/4)$. Unfortunately, this might rule out the existence of a "static" model and possibly the efficient algorithm developed in [PCVS16]. To conclude, let us insist on the striking difference of this extension with standard optimal transport which excludes "shocks" during the transport but at the initial and final timepoints. In this type of model, shocks are needed in order to fulfill the metrizability of weak convergence. These models are also part of L. Chizat's PhD thesis.

The last direction consists in "smoothing" optimal transport. Since optimal transport is a distance between probability measures which can be Dirac measures, the optimal maps cannot be expected to be smooth mappings. However, the deep theory of the Ma-Trudinger-Wang tensor gives conditions under which smoothness of the optimal maps for the L^2 Wasserstein metric is guaranteed [DPF14], when the

source and target measures are have positive smooth density w.r.t. the volume measure. These conditions do not hold when M has negative sectional curvature, in this case the maps can be discontinuous. Therefore, it is natural to ask for an "optimal transport like" metric on the space of densities which has, built-in, the smoothness property. Attempts in this direction include J. Louet's PhD thesis [Lou14], and in imaging applications in [FPR+13]. These two attempts aim at smoothing the static formulation of optimal transport. In general, these static formulations do not provide a metric on the space of densities. To propose a metric on the space of densities which includes smoothness constraints, we currently work on an extension of the Benamou-Brenier formulation.

3.5 From the Wasserstein-Fisher-Rao metric to the Camassa-Holm equation

This section is based on joint work with T. (Thomas) Gallouët [GV16].

Since Otto's Riemannian submersion has been extended to the case of the Wasserstein-Fisher-Rao metric, it is natural to look for a fluid dynamic equation which is the counterpart of the incompressible Euler equation for the standard L^2 Wasserstein metric. In other words, what is the corresponding geometry on the isotropy subgroup and if there is an associated polar factorization theorem. It turns out that what replaces incompressible Euler equation is the geodesic equation for the right-invariant H^{div} metric on the group of diffeomorphisms. In the one dimensional case, this geodesic equation is known as the Camassa-Holm equation introduced in [CH93]. Since its introduction, the Camassa-Holm equation has attracted a lot of attention since it is a bi-Hamiltonian system as well as an integrable system, it exhibits peakon solutions and it is a model for waves in shallow water [CL08, Con01, Len05, CE98, BC07, Dan01, GHR11]. In particular, this equation is known for its well understood blow-up in finite time and it is a model for wave breaking [McK04].

3.5.1 The link to the Camassa-Holm equation

The Riemannian submersion $\pi_0 : \mathrm{Diff}(M) \ltimes_\Psi \Lambda_{1/2}(M) \mapsto \mathrm{Dens}(M)$ defined in Formula (3.2.2) enables to study the equivalent fluid dynamic equation. The fiber of the Riemannian submersion at vol is $\pi_0^{-1}(\{\mathrm{vol}\})$ and it will be denoted temporarily by H_0. More explicitely, we have

$$\pi_0^{-1}(\{\mathrm{vol}\}) = \{(\varphi, \lambda) \in \mathrm{Diff}(M) \ltimes_\Psi \Lambda_{1/2}(M) \ : \ \varphi_*(\lambda^2 \,\mathrm{vol}) = \mathrm{vol}\}\,. \tag{3.5.1}$$

The constraint $\varphi_*(\lambda^2 \,\mathrm{vol}) = \mathrm{vol}$ can be made explicit as follows

$$H_0 = \{(\varphi, \sqrt{\mathrm{Jac}(\varphi)}) \in \mathrm{Diff}(M) \ltimes_\Psi \Lambda_{1/2}(M) \ : \ \varphi \in \mathrm{Diff}(M)\}\,. \tag{3.5.2}$$

Note that this isotropy subgroup can be identified with the group of diffeomorphims of M by the map $\varphi \mapsto (\varphi, \sqrt{\mathrm{Jac}(\varphi)})$. Now, the vertical space at point $(\varphi, \sqrt{\mathrm{Jac}(\varphi)})$ is

$$\mathrm{Ker}\left(d\pi_0(\varphi, \sqrt{\mathrm{Jac}(\varphi)})\right) = \{(v, \alpha) \circ (\varphi, \sqrt{\mathrm{Jac}(\varphi)}) \ : \ \mathrm{div}\, v = 2\alpha\}\,, \tag{3.5.3}$$

and equivalently

$$\mathrm{Ker}\left(d\pi_0(\varphi, \sqrt{\mathrm{Jac}(\varphi)})\right) = \left\{\left(v, \frac{1}{2}\,\mathrm{div}\, v\right) \circ (\varphi, \sqrt{\mathrm{Jac}(\varphi)}) \ : \ v \in \mathrm{Vect}(M)\right\}\,. \tag{3.5.4}$$

The induced (weak) Riemannian metric on H_0 reads

$$G_\varphi(X_\varphi, X_\varphi) = \int_M |v|^2 \,\mathrm{dvol} + \frac{1}{4}\int_M |\operatorname{div} v|^2 \,\mathrm{dvol}\,, \tag{3.5.5}$$

where $v = X_\varphi \circ \varphi^{-1}$. In conclusion, we have $H_0 \simeq \mathrm{Diff}(M)$ and the induced metric on $\mathrm{Diff}(M)$ is the right-invariant H^{div} metric, isometrically embedded in $\mathrm{Diff}(M) \ltimes_\Psi \Lambda_{1/2}(M)$ endowed with its $L^2(M, \mathcal{C}(M))$ metric.

As a straightforward application, we retrieve a theorem by Michor and Mumford [MM05].

Corollary 24. *The distance on* $\mathrm{Diff}(M)$ *with the right-invariant metric* H^{div} *is non degenerate.*

Proof. Let $\varphi_0, \varphi_1 \in \mathrm{Diff}(M)$ be two diffeomorphisms and c be a path joining them. The length of the path c for the right-invariant metric H^{div} is equal to the length of the lifted path $\tilde{c}$ in $\mathrm{Diff}(M) \ltimes_\Psi \Lambda_{1/2}(M)$. Since $L^2(M, \mathcal{C}(M))$ is a Hilbert manifold, the length of the path $\tilde{c}$ is bounded below by the length of the geodesic joining the natural lifts of φ_0 and φ_1 in $L^2(M, \mathcal{C}(M))$. Therefore, it leads to

$$d_{H^{\mathrm{div}}}(\varphi_0, \varphi_1) \geq d_{L^2(M,\mathcal{C}(M))}\left((\varphi_0, \sqrt{\mathrm{Jac}(\varphi_0)}), (\varphi_1, \sqrt{\mathrm{Jac}(\varphi_1)})\right). \tag{3.5.6}$$

If $d_{H^{\mathrm{div}}}(\varphi_0, \varphi_1) = 0$ then $d_{L^2(M,\mathcal{C}(M))}\left((\varphi_0, \sqrt{\mathrm{Jac}(\varphi_0)}), (\varphi_1, \sqrt{\mathrm{Jac}(\varphi_1)})\right) = 0$ which implies $\varphi_0 = \varphi_1$. □

On the well-posedness of the initial value problem. Global well-posedness does not hold in one dimension since there exists smooth initial conditions for the Camassa-Holm equation such that the solutions blow up in finite time.

In higher dimensions, the initial value problem has been studied by Michor and Mumford [MM13, Theorem 3]. This is not a direct result of [EM70] since the differential operator associated to the metric is not elliptic. They prove that the initial value problem on the space of vector fields is well posed for initial data in a Sobolev space of high enough order. Although the proof could probably be adapted to the case of a Riemannian manifold, in that case, the result of local well posedness is not known.

Open question 25. *Is it possible to adapt the proof of local well-posedness in [EM70] to treat the case of the non-elliptic operator associated with the right-invariant* H^{div} *metric ?*

3.5.2 The Camassa-Holm equation as an incompressible Euler equation

This is a possibly surprising result since there is a priori no link between the semi-direct product of group and the group of volume preserving diffeomorphisms. The hint comes from the formulation of the Wasserstein-Fisher-Rao distance, as a the Wasserstein L^2 distance on the cone $\mathcal{C}(M)$, which was put forward in [LMS15, Section 7.2]. One has

$$\mathrm{WF}(\mu_0, \mu_1) = \min_{(\nu_0,\nu_1)\in \mathbb{E}_m^{-1}(\nu_0)\times\mathbb{E}_m^{-1}(\nu_1)} W_2(\nu_0, \nu_1)\,, \tag{3.5.7}$$

where $\mathbb{E}_m : \mathcal{P}^2_{\mathcal{C}(M)} \mapsto \mathcal{M}_+(M)$ defined by $\mathbb{E}(\nu) = \int_{\mathbb{R}_+} r^2 \,\mathrm{d}\nu(r)$, which means integration w.r.t. the (square root of) mass variable for probability measures ν for which $\int_{\mathcal{C}(M)} r^2 \,\mathrm{d}\nu < +\infty$. In other words, $\mathcal{P}^2_{\mathcal{C}(M)}$ is the set of probability measures

which have a finite second moment w.r.t. to the cone distance. Since this formulation draws the link between standard optimal transport and the Wasserstein-Fisher-Rao distance, it is then natural to look for the relation between Camassa-Holm and the incompressible Euler equation.

The first step to make this link clear consists in interpreting the semidirect product of group as a subset of the group of diffeomorphisms of the cone. The cone has a particular structure and can be viewed as the trivial principle bundle, since $\mathbb{R}_+^*$ is a multiplicative group. Actually the cone can be viewed as the fiber bundle of half-densities once a reference measure is chosen, which trivializes it. Diffeomorphisms that respect this group structure are called automorphisms, thus the semidirect product of groups is the automorphism group of the cone, denoted by $\mathrm{Aut}(\mathcal{C}(M))$. In other words, it is the group of diffeomorphisms of $\mathcal{C}(M)$ that are linear in the radial component r. Let us denote $\mathrm{Aut}_{\mathrm{vol}}(\mathcal{C}(M))$ the isotropy subgroup of vol, the Riemannian volume form on M, for the group action defined in (3.2.2).

Writing the geodesic equation of the H^{div} right-invariant metric as a Riemannian submanifold of $\mathrm{Aut}_{\mathrm{vol}}(\mathcal{C}(M)) \subset \mathrm{Aut}(\mathcal{C}(M)) \subset \mathrm{Diff}(\mathcal{C}(M))$ leads to

$$\frac{D}{Dt}(\dot{\varphi}, \dot{\lambda r}) = -\nabla \Psi_p \circ (\varphi, \lambda r)\,, \tag{3.5.8}$$

where $\Psi_p(x,r) \overset{\text{def.}}{=} \frac{1}{2} r^2 p(x)$. At this point, this equation looks like the incompressible Euler equation, however, the Riemannian volume measure is not left invariant by pushforward of an element in $\mathrm{Aut}_{\mathrm{vol}}(\mathcal{C}(M))$. However, a straightforward computation shows that the density $r^{-3}\,\mathrm{d}r\,\mathrm{dvol}$ is left invariant by pushforward. Therefore, we have the reformulation of the geodesic equation of the H^{div} right-invariant metric as an incompressible Euler equation *but for a density on the cone which has a radial form and not integrable at the cone point.*

Theorem 26. *A solution (φ, λ) of (3.5.8) is a solution of the incompressible Euler equation for the density $\tilde{\nu} = r^{-3-d}\,\mathrm{dvol}_{\mathcal{C}(M)}$ where $\mathrm{dvol}_{\mathcal{C}(M)}$ is the volume form on the cone $\mathcal{C}(M)$ and d is the dimension of M.*

This result is also possibly surprising since it means that $\mathrm{Aut}_{\mathrm{vol}}(\mathcal{C}(M))$ is left invariant by the (modified) Euler equation. In other words, this theorem underlines that $\mathrm{Aut}_{\mathrm{vol}}(\mathcal{C}(M)) = \mathrm{Aut}(\mathcal{C}(M)) \cap \mathrm{SDiff}_{\tilde{\nu}}(\mathcal{C}(M))$. Importantly, it can be shown that $\mathrm{Aut}(\mathcal{C}(M))$ is a totally geodesic subspace of $\mathrm{Diff}(\mathcal{C}(M))$, which explains the fact that the geodesic equation on $\mathrm{Aut}_{\mathrm{vol}}(\mathcal{C}(M))$ is actually a geodesic equation on $\mathrm{SDiff}_{\tilde{\nu}}(\mathcal{C}(M))$. We illustrate this situation in Figure 3.5.

Theorem 26, applied to the 1D case, leads to

Corollary 27. *Solutions to the Camassa-Holm equation*

$$\partial_t u - \frac{1}{4}\partial_{txx} u + 3\partial_x u\, u - \frac{1}{2}\partial_{xx} u\, \partial_x u - \frac{1}{4}\partial_{xxx} u\, u = 0 \tag{3.5.9}$$

are mapped to solutions of the incompressible Euler equation on $\mathbb{R}^2 \setminus \{0\}$ for the density $\rho = \frac{1}{r^4}\,\mathrm{Leb}$, that is

$$\begin{cases} \dot{v} + \nabla_v v = -\nabla P\,, \\ \nabla \cdot (\rho v) = 0\,, \end{cases} \tag{3.5.10}$$

by the map $u \mapsto \big(u(\theta), \frac{r}{2}\partial_x u(\theta)\big)$.

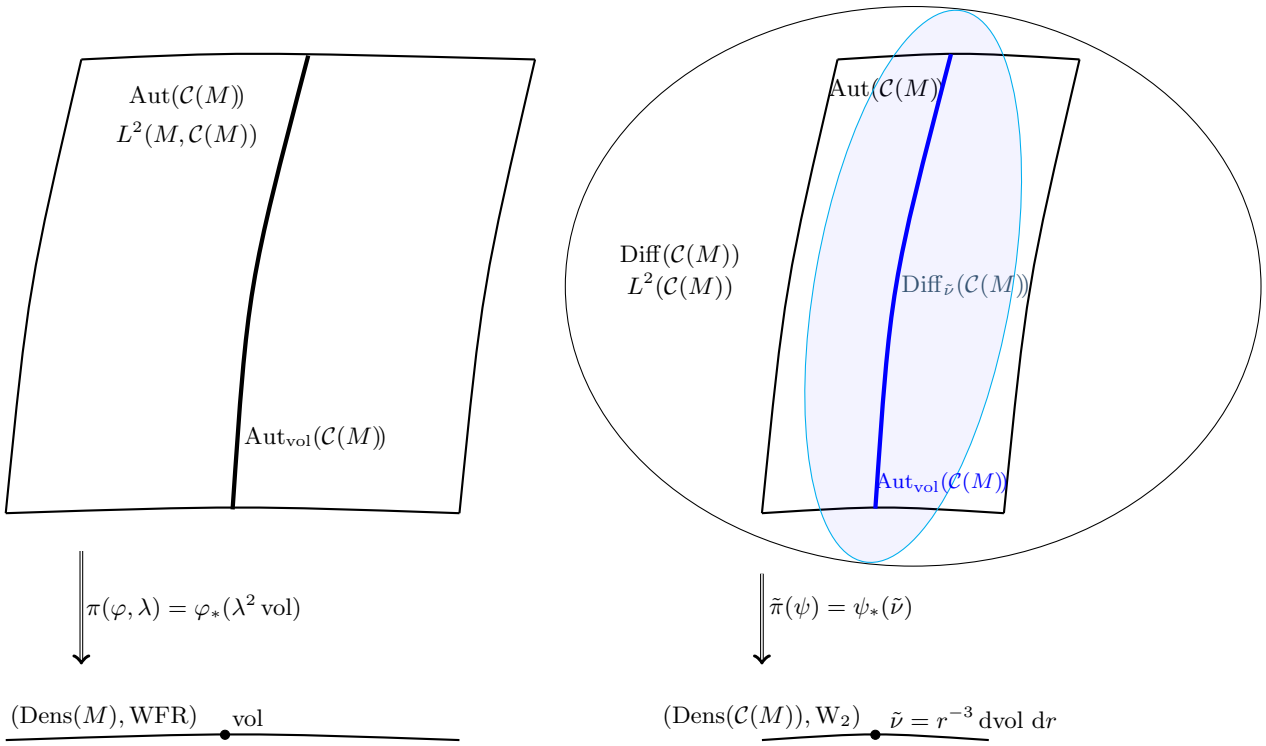


FIGURE 3.5: On the left, the picture represents the Riemannian submersion between $\mathrm{Aut}(\mathcal{C}(M))$ and the space of positive densities on M and the fiber above the volume form is $\mathrm{Aut}_{\mathrm{vol}}(\mathcal{C}(M))$. On the right, the picture represents the automorphism group $\mathrm{Aut}(\mathcal{C}(M))$ isometrically embedded in $\mathrm{Diff}(\mathcal{C}(M))$ and the intersection of $\mathrm{Diff}_{\tilde{\nu}}(\mathcal{C}(M))$ and $\mathrm{Aut}(\mathcal{C}(M))$ is equal to $\mathrm{Aut}_{\mathrm{vol}}(\mathcal{C}(M))$.

A natural application of this result concerns minimizing properties of solutions of the geodesic equation. It appears possible to adapt the proof of Y. Brenier [Bre03] in this case. The following result cannot be seen as a direct consequence of his result since the cone is not a compact manifold, but it is a simple adaptation of it.

Theorem 28. *Smooth solutions to the Camassa-Holm equation (3.5.9) are length minimizing for short times.*

This result can be also extended in the Riemannian case in any dimension, for a compact manifold without boundaries, see [GV16, Section 6].

Open question 29. *In this section, we put forward the link between the Camassa-Holm equation and the incompressible Euler equation. It is possible to define an equivalent to the generalized flows introduced by Y. Brenier in [Bre89]. However, we conjecture that this convex relaxation is not tight in 1D and tight in higher dimensions, similarly to what happens for the incompressible Euler equation (not tight in 2D but tight in 3D due to Schnirelman's results). Therefore, the relaxation of the Camassa-Holm equation is open.*

Open question 30. *Recently, the article [ABLZ17] shows the correspondence between the entropic regularization of incompressible Euler and a certain form of the Navier-Stokes equation. What is the corresponding regularization of the Camassa-Holm equation ?*

3.5.3 The corresponding polar factorization theorem

Following the geometric picture in Proposition 8, we explain the corresponding polar factorization theorem. It turns out that this factorization can be extended to a larger class containing the automorphism group of the cone $\mathrm{Aut}(\mathcal{C}(M))$. In the following, we state a polar factorization theorem for a class of maps from M to $C(M)$. We start with definitions.

Definition 31. We define the generalized automorphism semigroup of $C(M)$ as the set of mesurable maps (φ, λ) from M to $C(M)$

$$\overline{\mathrm{Aut}}(\mathcal{C}(M)) = \left\{(\varphi,\lambda) \in \mathcal{M}es(M,M) \ltimes \mathcal{M}es(M,\mathbb{R}_+^*)\right\}, \tag{3.5.11}$$

endowed with the semigroup law

$$(\varphi_1,\lambda_1)\cdot(\varphi_2,\lambda_2) = (\varphi_1\circ\varphi_2, (\lambda_1\circ\varphi_2)\lambda_2)\,.$$

We also consider the stabilizer of the volume measure in the automorphisms of $C(M)$. It is a subsemigroup and is defined by

$$\overline{\mathrm{Aut}}_{\mathrm{vol}}(\mathcal{C}(M)) = \left\{(s,\lambda) \in \overline{\mathrm{Aut}}(\mathcal{C}(M)) \,:\, \pi\left((s,\lambda),\mathrm{vol}\right) = \mathrm{vol}\right\}. \tag{3.5.12}$$

By abuse of notation, any $(s,\lambda) \in \overline{\mathrm{Aut}}_{\mathrm{vol}}(\mathcal{C}(M))$ will be denoted $\left(s, \sqrt{\mathrm{Jac}(s)}\right)$ meaning that for every continuous function $f \in C(M,\mathbb{R})$

$$\int_M f(s(x))\sqrt{\mathrm{Jac}(s)}^2\,\mathrm{dvol}(x) = \int_M f(x)\,\mathrm{dvol}(x)\,. \tag{3.5.13}$$

Using the notation $c(x,y) = -\log(\cos^2(\min(d(x,y),\pi/2))$ and the usual definition of c-convex functions, we have:

Theorem 32 (Polar factorization)**.** *Let $(\phi,\lambda) \in \overline{\mathrm{Aut}}(\mathcal{C}(M))$ be an element of the generalized automorphism group of the half-densities bundle such that $\rho_1 = \pi_0\left[(\phi,\lambda),\mathrm{vol}\right]$ is*

an absolute continuous admissible measure. Then, there exists a unique minimizer, characterized by a c-convex function z_0, to the Monge formulation (3.2.16) between vol *and* ρ_1 *and there exists a unique measure preserving generalized automorphism* $(s, \sqrt{\mathrm{Jac}(s)}) \in \overline{\mathrm{Aut}}_{\mathrm{vol}}(\mathcal{C}(M))$ *such that* vol *a.e.*

$$(\phi, \lambda) = \exp^{\mathcal{C}(M)}\left(-\frac{1}{2}\nabla p_{z_0}, -p_{z_0}\right) \circ (s, \sqrt{\mathrm{Jac}(s)}) \tag{3.5.14}$$

or equivalently

$$(\phi, \lambda) = \left(\varphi, e^{-z_0}\sqrt{1 + \|\nabla z_0\|^2}\right) \cdot (s, \sqrt{\mathrm{Jac}(s)}), \tag{3.5.15}$$

where $p_{z_0} = e^{z_0} - 1$ *and*

$$\varphi(x) = \exp_x^M\left(-\arctan\left(\frac{1}{2}\|\nabla z_0(x)\|\right)\frac{\nabla z_0(x)}{\|\nabla z_0(x)\|}\right). \tag{3.5.16}$$

Moreover $(s, \sqrt{\mathrm{Jac}(s)})$ *is the unique* $L^2(M, \mathcal{C}(M))$ *projection of* (ϕ, λ) *onto* $\overline{\mathrm{Aut}}_{\mathrm{vol}}(\mathcal{C}(M))$.

Note that underlying the Wasserstein-Fisher-Rao distance, there corresponds a Monge-Ampère equation which we state informally as follows, for a c-convex function z:

$$\det\left[-\nabla^2 z(x) + (\nabla^2_{xx} c)(x, \varphi(x))\right] = \left|\det\left[(\nabla_{x,y} c)(x, \varphi(x))\right]\right| e^{-2z(x)}\left(1 + \frac{1}{4}\|\nabla z(x)\|^2\right)\frac{f(x)}{g \circ \varphi(x)}, \tag{3.5.17}$$

where φ is given by

$$\varphi(x) = \exp_x^M\left(-\arctan\left(\frac{1}{2}\|\tilde{\nabla} z(x)\|\right)\frac{\tilde{\nabla} z(x)}{\|\tilde{\nabla} z(x)\|}\right) \tag{3.5.18}$$

and satisfies the second boundary value problem: φ maps the support of ρ_0 to the support of ρ_1. A current direction of our research is the smoothness properties of the optimal maps following the Ma-Trudinger-Wang approach.

Open question 33. *We conjecture that the positivity of the MTW tensor for the cost c on M implies the positivity of the MTW tensor of the Riemannian distance on $\mathcal{C}(M)$.*

Chapter 4

Applications to diffeomorphic image matching

This section summarizes the contributions we had on the practical implementation of the LDDMM framework, and some of its possible extensions. We proposed the "sum of kernels" [RVW+11] in order to produce more plausible deformations and give a mathematical interpretation in [BRV12]. In the same direction, we proposed the introduction of spatially varying kernels in the framework of left-invariant metrics instead of right-invariance [SRV13], and a simple optimization method to learn the metric parameters in [VR14].

We also introduced geodesic regression on the space of images [NHV11] and a corresponding shooting algorithm [VRRC12] and present an application of this algorithm as a longitudinal deformation model for the development of Alzheimer disease [FRR+14].

4.1 Introduction

As presented in the introduction, the elementary method of LDDMM consists in the minimization of the functional

$$\mathcal{L}(\xi) = \int_0^1 \|\xi(t)\|_V^2 \, \mathrm{d}t + S(q(1)) \tag{4.1.1}$$

under the constraints

$$\begin{aligned} &\partial_t q(t,x) = \xi(t, q(t,x)) \\ &q(0,x) = q_0(x) \ \forall x \in D\,. \end{aligned} \tag{4.1.2}$$

For practical applications of the method, two crucial points are (1) the choice of the reproducing kernel space (RKHS) of vector fields V or equivalently the choice of the kernel defining it and (2) the choice of the similarity measure S. In this chapter, we will mostly discuss the first issue and present a similarity measure based on optimal transport for curves or surfaces.

4.2 Sum of kernels and semidirect product of groups

This section is based on the work with Laurent Risser and Martins Bruveris in [RVW+11, BRV12].

In most applications, a single Gaussian kernel was used or a kernel corresponding to the differential operator $(\mathrm{Id} + \sigma\Delta)^k$ for a well chosen k with a single parameter

σ. The Gaussian width, also denoted by σ, was chosen so that the matching "quality" was sufficiently good, which means a small σ. What is the effect of this parameter on the shape of the deformation ? Let us choose for instance a small translation of a shape as a simple example. If one chooses a small kernel size, the deformation will be explained by the motion of the boundaries and the center point of the shape will almost not move at all. On the contrary, this translation will be well captured by a Gaussian kernel with large σ, however, this kernel will not give a good matching on real data, where small scale features matter. This practical question is reflected by

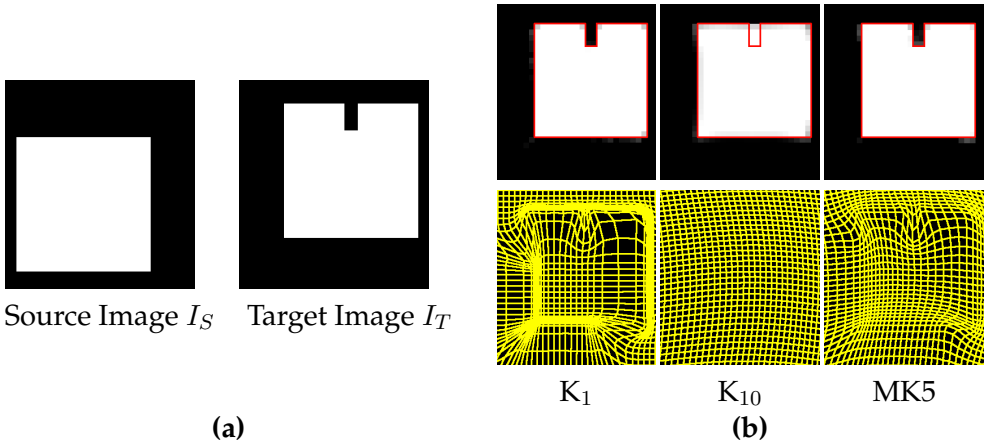


FIGURE 4.1: Influence of the smoothing kernel when registering two images containing differences at several scales simultaneously in the LDDMM framework. **(a)** Region of interest of the source and target images I_S and I_T containing the registered shapes. **(b)** Registration of the images I_S and I_T using different kernels: (k_1 and k_{10}) Gaussian kernels of width $\sigma = 1$ and $\sigma = 10$ pixels; (MK5) sum of 5 Gaussian kernels linearly sampled between 10 and 1 pixels. Diffeomorphic transformations of I_S at $t = 1$ (final deformation) are shown on the top and corresponding homogeneous grids (step = 1 pixel) are on the bottom.

the following property of the LDDMM method: Where there is no gradient information, for instance in a flat region of the image, the kernel spatially interpolates the rest of the information (i.e. the momentum) to drive the motion of the points. Therefore, it is natural to introduce a sum of kernels to fill in the missing information while preserving the quality of the matching. Therefore, more plausible deformations are obtained since the correlation of the motions of the points is higher. This is illustrated in Figure 4.1.

In practice, this method works really well and the mathematical insight for this efficiency is probably the variational interpretation of the sum of kernel. For the sake of simplicity, we only treat the case of a finite set of RKHS Hilbert spaces H_i with kernels k_i and Riesz isomorphisms K_i between H_i^* and H_i for $i = 1, \ldots, n$. Denoting $H = H_1 + \ldots + H_n$, the space of all functions of the form $v_1 + \ldots + v_n$ with $v_i \in H_i$, the norm is defined by

$$\|v\|_H^2 = \inf \left\{ \sum_{i=1}^{n} \|v_i\|_{H_i}^2 \,\Big|\, \sum_{i=1}^{n} v_i = v \right\} . \tag{4.2.1}$$

The minimum is achieved for a unique n-tuple of vector fields and the space H,

endowed with the norm defined by (4.3.5), is complete. The result is the following, there exists a unique element $p \in \bigcap_{i=1}^n H_i^*$ for which one has $v_i = K_i p$ and

$$v = \sum_{i=1}^{n} K_i p \,, \tag{4.2.2}$$

the family $(v_i)_{i=1,\dots,n}$ realizing the (unique) infimum of the variational problem. The formula (4.3.5) induces a scalar product on H which makes H a RKHS, and its associated kernel is $\mathsf{k} := \sum_{i=1}^n \mathsf{k}_i$, where k_i denotes the kernel of the space H_i. This property was written in [Aro50]. Note also that this property is the particular case of an elementary result in convex analysis, at least in finite dimensions: the convex conjugate of an infimal convolution is equal to the sum of the convex conjugates.

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. This is probably explained because of the shape of the kernel, in relation with the similarity measure is responsible for the local minimum of the functional. We suspect that a small kernel size increases the number of local minimums, although we have no quantitative argument in this direction.

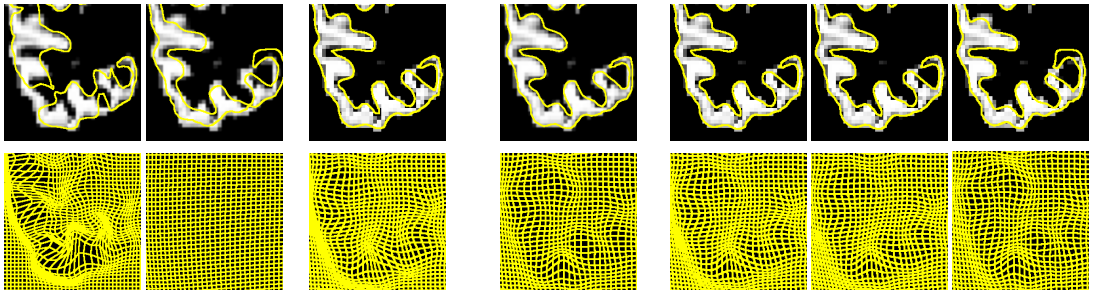


FIGURE 4.2: Different regularizations that show the effect of the sum of kernels and in particular that it avoids bad local minima. MK stands for the sum of kernels. In [RVW+11], we propose a rule of thumb on how to choose the weights semi-automatically and the scale parameters.

This infimal convolution at the level of the Lie algebras has its counterpart (actually not uniquely defined from the linear situation) at the level of the groups. Let us assume that we build the groups $\mathcal{G}_{H_i}$ as in Formula (1.2.5), we can now use these groups to act on an image by, for instance,

$$(\varphi_1, \dots, \varphi_n) \cdot q = \varphi_1 \cdot (\dots \varphi_{n-1} \cdot (\varphi_n \cdot q)) \,, \tag{4.2.3}$$

deciding in particular on the order of the scales. Now, there are at least two questions: (1) Does there exist a group structure on $\mathcal{G}_1 \times \dots \times \mathcal{G}_n$ such that the above map is a left action ? (2) What would be the Lagrangian to minimize in order to "choose" one particular deformation ?

The natural answer to the latter consists in minimizing

$$\frac{1}{2} \sum_{i=1}^{n} \int_0^1 \|v_i(t)\|_{H_i}^2 \, \mathrm{d}t + d^2(\varphi(1).I_0, I_{\text{target}}) \;\; , \tag{4.2.4}$$

where $\varphi(1) = \varphi_1(1) \circ \ldots \circ \varphi_n(1)$. It turns out that the answer to the first is positive by defining the semidirect product of groups $\mathcal{G} = \mathcal{G}_1 \rtimes \cdots \rtimes \mathcal{G}_n$ and this group is equal to $\mathcal{G}_H$, and the minimization problem (4.2.4) is equivalent to the standard LDDMM formulation with a RKHS defined by (4.3.5). However, the caveat in this formulation is that the diffeomorphisms $\varphi_i(t)$ are not the flows of $v_i(t)$ unless $i = 1$, but they are defined by

$$\partial_t \varphi_k(t) = \left(v_k(t) + (\mathrm{Id} - \mathrm{Ad}_{\varphi_k(t)}) \sum_{i=k+1}^{n} v_i(t) \right) \circ \varphi_k(t) \,. \tag{4.2.5}$$

We can also sum over all scales to form $v(t) = \sum_{k=1}^{n} v_k(t)$ and compute the flow $\varphi(t)$ of $v(t)$. Then a simple calculation shows that

$$\varphi(t) = \varphi_1(t) \circ \ldots \circ \varphi_n(t) \quad . \tag{4.2.6}$$

4.3 Left-invariant metrics

This section is based on the articles with T. Schmah and L. Risser [SRV13, SRV15].

A natural extension of the sum of kernels consists in having a kernel which may depend on the location. However, the right-invariant point of view is meant for a homogeneous material whose properties are translation invariant although this is not required by the theory. In practice the kernel used in diffeomorphic methods (LDDMM and the other methods cited above) has always been chosen to be translationally-invariant and isotropic. In LDDMM, spatially varying or non isotropic ("direction-dependent") kernels have no obvious interpretation, because the norm is defined in Eulerian coordinates, so that as t varies during the deformation, a fixed point in the source image moves through space, and conversely, a fixed point in space will correspond to different points on the source image. Similarly, the directions in a direction-dependent kernel are defined with respect to Eulerian coordinates, not the coordinates of the moving source image. Nonetheless, spatially-varying kernels are potentially of great interest in medical applications, if they can be made to represent spatially-variable (or non-isotropic) deformability of tissue. This is indeed already done in [RVBS13] to model sliding conditions between the lungs and the ribs.

In this section, we present a slightly different registration framework than LDDMM which naturally supports the use of spatially varying kernels. It is based on a *left*-invariant metric, i.e. based on a norm in the body (Lagrangian) coordinates of the source image, which explains its naming LIDM (Left Invariant Diffeomorphic Metrics). This means that instead of the V norm in (4.1.1) being applied to the spatial (Eulerian) velocity defined by (4.1.2), it is applied to the *convective velocity* defined by

$$\partial_t \phi(t) = d\phi(t) \cdot v(t) \,, \tag{4.3.1}$$

where $d\phi(t)$ is the spatial derivative of $\phi(t)$.

It is well-known that left trivialization and right trivialization are "isomorphic" by the inverse map. The setting we propose consists in using a left action with a left-invariant metric, which differs from the standard LDDMM setting, which is a left action together with a right-invariant metric. In particular, this new framework looses the induced Riemannian metric on the orbit.

Thus, we are left with proving the existence of the flow of (4.3.1) for an element $v \in L^2([0,1], V)$ for instance, which is done in [SRV15]. Now, we define the set of

associated transformations.

$$\mathcal{G}_V^L := \left\{\phi(1) \in B \,\middle|\, \partial_t\phi(t) = d\phi(t)\cdot v(t) \text{ and } v \in L^2([0,1],V)\right\},$$
$$\mathcal{G}_V = \left\{\phi(1) \in B \,\middle|\, \partial_t\phi(t) = u(t)\circ\phi^{-1}(t) \text{ and } u \in L^2([0,1],V)\right\}.$$

It can be shown that $\mathcal{G}_V^L = \mathcal{G}_V$, however, the sets of paths $\phi(t)$ in the definitions of $\mathcal{G}_V^L$ and $\mathcal{G}_V$ do not coincide in general. Once $\mathcal{G}_V^L$ is equipped with the left-invariant metric (that is the norm of $v \in L^2([0,1],V)$) denoted by d_L, and denoting the right-invariant metric by d_R we get

Proposition 34. *1. The inverse mapping is an isometry:*

$$(\mathcal{G}_V^L, d_L) \to (\mathcal{G}_V, d_R)$$
$$\phi \to \phi^{-1}$$

2. *ϕ is a left-geodesic if and only if ϕ^{-1} is a right-geodesic.*
3. *Left translation is an isometry of $(\mathcal{G}_V^L, d_L)$, and right translation is an isometry of $(\mathcal{G}_V, d_R)$.*
4. *The left translation of a left-geodesic is a left-geodesic (and similarly for right-geodesics).*

Therefore, as expected, we obtain the following result:

Corollary 35. *[Equivalence of Optimal Matches in Left- and Right- LDM] Consider the problem of minimising*

$$\mathcal{J}(\phi) = \frac{1}{2}\int_0^1 \|v(t)\|_V^2 \,\mathrm{d}t + E(\phi_1\cdot I, J), \tag{4.3.2}$$

for $\phi_0 = Id_\Omega$, and with either constraint

$$\partial_t\phi_t = d\phi_t\cdot v_t \qquad \textit{(Left-LDM constraint)} \tag{4.3.3}$$

or

$$\partial_t\phi_t = v_t\circ\phi_t \qquad \textit{(Right-LDM constraint)}. \tag{4.3.4}$$

Then

1. *The optimal endpoint ϕ_1 is the same with either constraint.*
2. *If ϕ_t minimises $\mathcal{J}$ in Left-LDM, then $\psi_t := \phi_{1-t}^{-1}\circ\phi_1$ minimises $\mathcal{J}$ in Right-LDM.*
3. *If ψ_t minimises $\mathcal{J}$ in Right-LDM, then $\phi_t := \psi_1\circ\psi_{1-t}^{-1}$ minimises $\mathcal{J}$ in Left-LDM.*

Optimal paths in Left-LDM are left-geodesics, while optimal paths in Right-LDM are right-geodesics.

Although not surprising, this result enables the use of spatially varying kernels that can defined using a variational approach. Let us consider, as in the previous section, a family of RKHS $(H_i)_{i=1,\dots,n}$ and an operator $A : H_1\oplus\ldots\oplus H_n \mapsto H = H_1+\ldots+H_n$. On the space H, we introduce

$$\|v\|_H^2 = \inf\left\{\sum_{i=1}^n \|v_i\|_{H_i}^2 \,\middle|\, A(v_1,\dots,v_n) = v\right\}. \tag{4.3.5}$$

Using again duality and under mild assumptions, the kernel associated with H is $H^* \ni p \mapsto \sum_{i=1}^n K_i(A^*p)_i \in H$.

Let us give an instance of it in the context of biomedical images. Suppose we have a partition of unity $((\chi_i)_{i=1,\dots,n})$ of the domain of interest (a manual segmentation of the biological shape) where we have some knowledge of the deformability properties of the shape, modeled by the kernel K_i. The map A can be chosen as $\sum_{i=1}^n \chi_i v_i$ and the corresponding kernel is

$$K = \sum_{i=1}^{n} \chi_i K_i \chi_i \,. \tag{4.3.6}$$

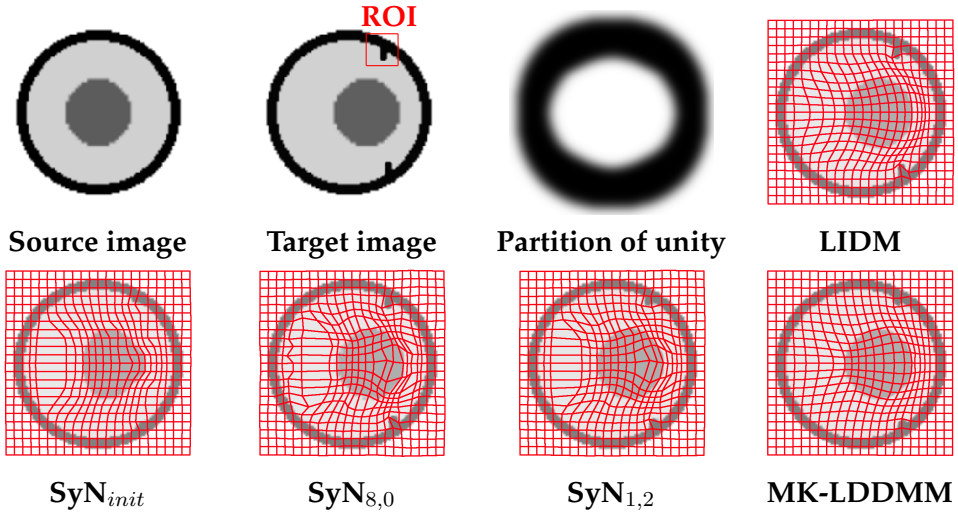


FIGURE 4.3: Results of image registration tests on a synthetic example.

The experiment of Figure 4.3 is taken from [SRV13], and it shows registration results for a synthetic example which includes features at different scales. LIDM shows the results of the registration using a kernel which is defined accordingly to the partition of unity shown in the figure (two gaussian kernels with a large σ on the white and a small σ on the black). As expected, it performs better than the sum of kernel because it captures the small scale deformations.

The use of spatially varying kernels provably improves the registration results on real data. However, the shortcoming of this approach is that the kernel does not evolve with the deformed shape. For small/medium deformations, it may not be a problem but it cannot be applied in the case of large deformations. In such a case, the kernel has to depend on the shape itself and it can be formulated in the framework of Proposition 3. Such approaches have actually been developed in [You12, ATTY15] in which the operator A depends on the shape itself, but developing models for images associated with an efficient implementation remains open.

4.4 Learning the metric

This section is based on [VR14], joint work with L. Risser.

Instead of having a partition of unity drawn by the user, it is natural to ask whether the smoothing kernel can be learnt from data. We summarize here an approach proposed in [VR14] to learn the parameters from a given population and a given template.

Building on the LIDM model, we aim at designing a set of kernels expressing spatially-varying metrics. We use (symmetric) positive definite matrices M as a parametrization of this set of kernels. In order to ensure 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 propose the following parametrization,

$$\mathcal{K} = \{\hat{K} M \hat{K} \mid M \text{ SDP operator on } L^2(\mathbb{R}^d, \mathbb{R}^d)\}\,, \tag{4.4.1}$$

where $\hat{K}$ is a spatially-homogeneous smoothing kernel (typically Gaussian). Now, the variational model consists in minimizing the functional, with β a positive real:

$$\mathcal{F}(M) = \frac{\beta}{2} d^2_{S^{++}}(M, Id) + \frac{1}{N}\sum_{n=1}^{N} \min_v \mathcal{J}_{I_n}(v, M)\,, \tag{4.4.2}$$

where M is a symmetric The first term is a regularizer of the kernel parameters so that 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 $d^2_{S^{++}}(Id, M)$ 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 [VR14] where we used simple method of dimension reduction (since the matrix M is of size n^2 where n is the number of voxels) and it gave promising results. We refer to the article for the experiments. This direction of research should probably be revisited and should incorporate some of the technics developed in deep learning in order to learn the shape dependent metric while still constraining the evolution to remain diffeomorphic.

4.5 Optimal transport for diffeomorphic matching

This section is based on joint work with J. Feydy, B. Charlier and G. Peyré in [FCVP17].

Using entropic optimal transport as a similarity measure has been advocated in [FCVP17]. Entropic regularization of (unbalanced) optimal transport not only permitted the use of fast algorithms to solve it but it also defines a smooth convex data similarity between positive Radon measures. Therefore, due to its global nature, we expect such similarities to be better behaved than the ones currently used in shape matching. Let us discuss a bit state of the art data similarity measures for embedded shapes such as surfaces or curves: Such a similarity should be parametrization invariant and weak dual norms on the space of currents, or varifolds, see [VG05, CT13] for instance, have been put forward. For instance, in the case of measures, such norms takes the form, for H a usually Gaussian RKHS

$$D_H(\mu, \nu) = \sup_{f \in B_1(H)} \langle f, \nu - \mu \rangle\,, \tag{4.5.1}$$

whereas the L^1 Wasserstein distance, denoted by W_1 is defined by

$$W_1(\mu,\nu) = \sup_{f\in \mathrm{Lip}_1(H)} \langle f, \nu-\mu\rangle\,, \tag{4.5.2}$$

where the supremum is taken over the space of Lipschitz function of Lipschitz constant less or equal than 1. Although it may have a similar formulation, the two norms are very different, for instance in terms of the rate of convergence of an empirical measure to its limit measure (see for instance [FG15]). In our experiments [FCVP17], the behavior seems in favor of the regularized optimal transport distance, since the kernel metrics behave as a sort of semi-local L^2 norm.

Let us note that for the use of the entropic regularized optimal transport on an embedding, one first need a map from the space of embeddings Emb to the space of nonnegative Radon measures on a space X, $\Psi : \mathrm{Emb} \mapsto \mathcal{M}_+(X)$. There is a choice of space X, that is for instance representing a surface given by a triangulation as just a measure supported by the vertices of the triangulation, or as a measure on the space of (centers of triangles, normals) where the weight is given by the surface of the triangle. Then, different costs can be used such as the ones described in [VG05, CT13] depending on the context.

Open question 36. *A question of practical interest is how to extend such a similarity measure to the space of images. A natural attempt would be to lift the image as a measure supported on its graph, as done in [TPK+16]. However, a different point of view can be taken by considering a binary image as a singular measure on its level line and use standard optimal transport on this embedded object, to fit in the approach described above. Therefore, a probably more promising extension consists in lifting the image as a collection of level lines weighted by the gradient. This is future work.*

4.6 Geodesic regression

This section is based on joint work with M. Niethammer and his student [NHV11].

As advocated in the introduction of this document, the LDDMM framework enables the generalizations of Euclidean tools since it is Riemannian in the strong sense, when one chooses, as presented in Chapter 2, a high order Sobolev metric. The simplest one is probably geodesic regression in which the goal consists in fitting some given shape time sequence with a single geodesic where the unknown parametrizing the geodesic are the shape and the slope. Therefore, the data is represented as a low dimensional subspace in the space of shapes.

In [NHV11], we developed this model on the space of images. Recall that it is possible to reformulate the registration problem as an optimization problem on the initial momentum as explained in the introduction 1.3.

$$\mathcal{J}(P_0) = \frac{1}{2}\langle P_0, KP_0\rangle_{L^2} + \sum_{k=1}^{n} S(I(t), J_k) \tag{4.6.1}$$

under the PDE constraint

$$\begin{cases} \partial_t I + \langle \nabla I, v\rangle = 0 \\ \partial_t P + \nabla\cdot(Pv) = 0 \\ v + K(P\nabla I) = 0\,, \end{cases} \tag{4.6.2}$$

and the initial conditions $P(0) = P_0$ and $I(0) = I_0$. This requires in particular an algorithm for solving the variational problem and it has been proposed in [VRRC12],

which uses an efficient implementation of the adjoint equations associated with the system (4.6.2). The adjoint equations in terms of $\hat{I}, \hat{P}, \hat{V}$, the adjoint variables to I, P, v read

$$\begin{cases} \dot{\hat{P}} + \nabla \hat{P} \cdot v - \nabla I \cdot K \star \hat{V} = 0 \\ \dot{\hat{I}} + \nabla \cdot (Iv) + \nabla \cdot PK \star \hat{V} = 0 \\ \hat{V} + \hat{I} \nabla I - P \nabla \hat{P} = 0 \,, \end{cases} \tag{4.6.3}$$

subject to boundary conditions

$$\begin{cases} \hat{I}(1) + I(1) - J = 0 \,, \\ \hat{P}(1) = 0 \,, \end{cases} \tag{4.6.4}$$

in the case where $n = 1$ and d is the squared L^2 distance. Note in particular that the solution to this system can be computed explicitly if one has computed the inverse of the flow of the vector field v. This vector field being smooth, numerical algorithms perform better on the flow, rather than solving directly the advection and continuity equations. More precisely, we have that the solution satisfies the integral relation

$$\begin{cases} \hat{P}(t) = \hat{P}(1) \circ \phi_{t,1} - \int_t^1 [\nabla I(s) \cdot \hat{v}(s)] \circ \phi_{t,s} \, ds \,, \\ \hat{I}(t) = \mathrm{Jac}(\phi_{t,1}) \hat{I}(1) \circ \phi_{t,1} \\ \qquad + \int_t^1 \mathrm{Jac}(\phi_{t,s}) [\nabla \cdot (P(s) \hat{v}(s))] \circ \phi_{t,s} \, ds \,, \end{cases} \tag{4.6.5}$$

with:

$$\begin{cases} \hat{v}(t) = K \star [P(t) \nabla \hat{P}(t) - \hat{I}(t) \nabla I(t)] \,, \\ P(t) = \mathrm{Jac}(\phi_{t,0}) P(0) \circ \phi_{t,0} \,, \\ I(t) = I(0) \circ \phi_{t,0} \,, \end{cases} \tag{4.6.6}$$

where $\phi_{s,t} \doteq \phi_{0,t} \circ \phi_{0,s}^{-1}$ defined by the flow of the time-dependent velocity field $v(t) = -K \star P(t) \nabla I(t)$. Note that we used a second-order finite volume method using the Total Variation Diminishing (TVD) MinMod limiter to compute the inverse of the flow [LeV07]. In the same direction, we extended in [SVN15] the shooting system of cubic splines in order to represent data with few parameters.

4.7 Longitudinal deformation model for Alzheimer disease

This section is based on joint work with J.B. Fiot and L. Risser and other collaborators [FRR+14].

In this section, we briefly present the work done in [FRR+14] which aims at classifying, based on imaging data, patients who present mild cognitive impairment (MCI), between those who are going to develop the Alzheimer disease and those who will remain MCI. The available data are from the ADNI database and concerns a hundred of patients: for each patient, we have the shape of the segmented hippocampus at two different timepoints on a one year time interval. We proposed to use the LDDMM framework and the shooting algorithm [VRRC12], to represent the longitudinal deformation of the hippocampus by their initial shape at the first timepoint and the initial momentum, or equivalently, the initial vector field given by the shooting algorithm that minimizes Formulation (4.6.1), after a rigid alignment of the shapes. Then, a common template is defined via a Karcher mean algorithm, as in [VRHR11] and all the momentums, that are defined on the patient hippocampus shape, are mapped onto this template. Different mappings have been tried, such as

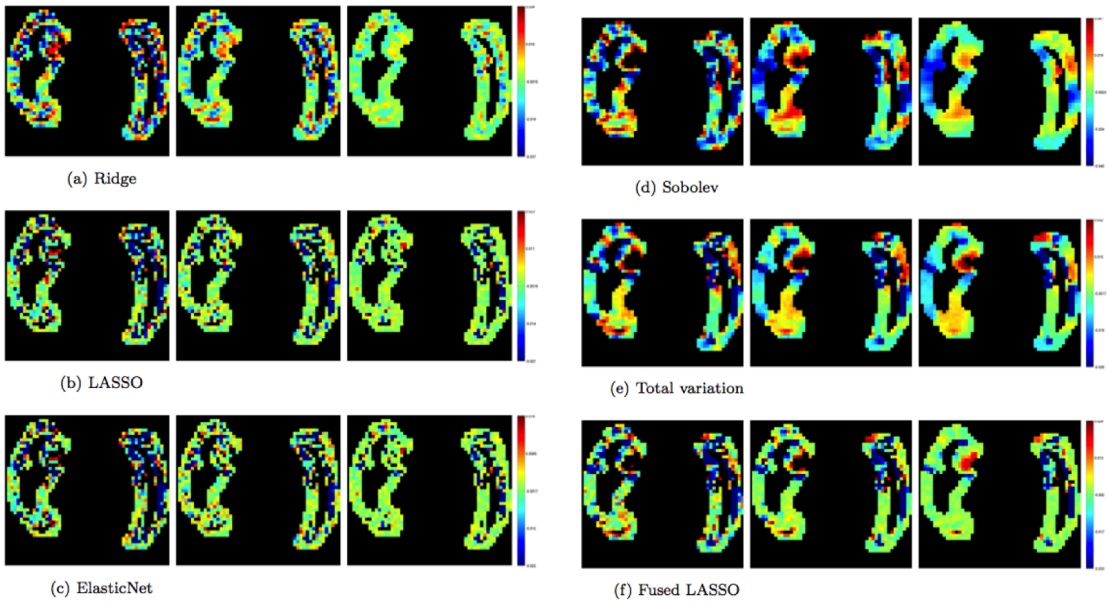


FIGURE 4.4: Different regularizations for the classification task.

coadjoint transport (i.e. in this case pushforward), adjoint transport of the vector field, and the former has finally been used due to better performance with respect to the classification results. The last step consists in using machine learning methods of supervised learning to separate the two populations: each patient evolution is represented by a density in the common domain of the template. Thus, the dimensionality of data is huge in comparison with the number of samples. Therefore, we used a regularized SVM method, with different regularizations such as Sobolev, total variation and fused lasso. Ideally, our aim was to identify some regions of particular interest using regularizations such as total variation. However, the best performances were obtained by the fused lasso regularization, see Figure 4.4.

The results were promising and in particular we have shown that initial momenta of the hippocampus deformations do capture information relevant to the disease progression. There is a wide range of perspectives opened by this paper and in particular the question of how to transport the momentum associated with the patient hippocampus evolution to the common template in which machine learning method can be used. Such a transport has to be a nonlinear map and we unfortunately did not include parallel transport because of the implementation difficulties. This motivated the conference paper [NV13] which studies the construction of Riemannian metrics on the space of shapes in order to perform statistics on small longitudinal evolutions and to preserve the statistical power of global indicators, such as volume variation. In particular, we discussed the absence of scale invariance properties of parallel transport in the LDDMM framework and we used the De Rham decomposition theorem (Riemannian geometry) to propose Riemannian metrics that leave invariant by parallel transport a collection of global evolution indicators.

Bibliography

[ABLZ17] M. Arnaudon, A. Bela Cruzeiro, C. Léonard, and J.-C. Zambrini. An entropic interpolation problem for incompressible viscid fluids. *ArXiv e-prints*, April 2017.

[AF05] Daniel Azagra and Juan Ferrera. Proximal calculus on Riemannian manifolds. *Mediterr. J. Math.*, 2(4):437–450, 2005.

[Arn66] Vladimir Arnold. Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits. *Ann. Inst. Fourier (Grenoble)*, 16(fasc. 1):319–361, 1966.

[Aro50] N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337–404, 1950.

[Atk75] C. J. Atkin. The Hopf-Rinow theorem is false in infinite dimensions. *Bull. London Math. Soc.*, 7(3):261–266, 1975.

[Atk97] Christopher J. Atkin. Geodesic and metric completeness in infinite dimensions. *Hokkaido Math. J.*, 26(1):1–61, 1997.

[ATTY15] S. Arguillere, E. Trélat, A. Trouvé, and L. Younes. Shape deformation analysis from the optimal control viewpoint. *J. Math. Pures Appl.*, 104, 2015.

[BB00] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.

[BBI01] D. Burago, Y. Burago, and S. Ivanov. A course in metric geometry. *American Mathematical Soc.*, 2001.

[BBM14] M. Bauer, M. Bruveris, and P. W. Michor. Overview of the geometries of shape spaces and diffeomorphism groups. *Journal of Mathematical Imaging and Vision*, pages 60–97, 2014.

[BBT65] G. Birkhoff, H. Burchard, and D. Thomas. Non-linear interpolation by splines, pseudosplines and elastica. *General Motors Research Lab report*, 468, 1965.

[BC07] Alberto Bressan and Adrian Constantin. Global conservative solutions of the Camassa-Holm equation. *Arch. Ration. Mech. Anal.*, 183(2):215–239, 2007.

[BCC$^+$15] J-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.

[Ben03] J-D. Benamou. Numerical resolution of an "unbalanced" mass transport problem. *ESAIM: Mathematical Modelling and Numerical Analysis*, 37(05):851–868, 2003.

[BG86] Robert Bryant and Phillip Griffiths. Reduction for constrained variational problems and $\int(\kappa^2/2)ds$. *Am. J. Math.*, 108:525–570, 1986.

[Bis11] Jean-Michel. Bismut. *Hypoelliptic Laplacian and orbital integrals*, volume 177. Princeton University Press, 2011.

[BK94] Guido Brunnett and Johannes Kiefer. Interpolation with minimal-energy splines. *Computer-Aided Design*, 26(2):137 – 144, 1994.

[BMM19] M. Bruveris, P. W. Michor, and D. Mumford. Geodesic completeness for Sobolev metrics on the space of immersed plane curves. *Forum Math. Sigma, 2*, 2014, e19.

[BMTY05] M. Faisal Beg, Michael I. Miller, Alain Trouvé, and Laurent Younes. Computing large deformation metric mappings via geodesic flow of diffeomorphisms. *International Journal of Computer Vision*, 61:139–157, 2005.

[Boo76] F. L. Bookstein. The study of shape transformations after d'arcy thompson. *Mathematical Biosciences*, 34:177–219, 1976.

[Bre89] Yann Brenier. The least action principle and the related concept of generalized flows for incompressible perfect fluids. *J. Am. Math. Soc.*, 2(2):225–255, 1989.

[Bre91] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.

[Bre99] Yann Brenier. Minimal geodesics on groups of volume-preserving maps and generalized solutions of the Euler equations. *Comm. Pure Appl. Math.*, 52(4):411–452, 1999.

[Bre03] Yann Brenier. Topics on hydrodynamics and volume preserving maps. *Handbook of Mathematical Fluid Dynamics*, 2:55 – 86, 2003.

[BRV12] Martins Bruveris, Laurent Risser, and François-Xavier Vialard. Mixture of kernels and iterated semidirect product of diffeomorphisms groups. *Multiscale Model. Simul.*, 10(4):1344–1368, 2012.

[BV17] Martins Bruveris and François-Xavier Vialard. On completeness of groups of diffeomorphisms. *J. Eur. Math. Soc. (JEMS)*, 19(5):1507–1544, 2017.

[CD14] M. Cuturi and A. Doucet. Fast computation of Wasserstein barycenters. In *Proceedings of ICML*, volume 32, 2014.

[CE98] Adrian Constantin and Joachim Escher. Wave breaking for nonlinear nonlocal shallow water equations. *Acta Math.*, 181(2):229–243, 1998.

[CGMP11] Frédéric Cao, Yann Gousseau, Simon Masnou, and Patrick Pérez. Geometrically guided exemplar-based inpainting. *SIAM Journal on Imaging Sciences*, 4(4):1143–1179, 2011.

[CGT16] Yongxin Chen, Tryphon T. Georgiou, , and Allen Tannenbaum. Distances and riemannian metrics for multivariate spectral densities. *Preprint arXiv:1610.03041*, 2016.

[CH93] Roberto Camassa and Darryl D. Holm. An integrable shallow water equation with peaked solitons. *Phys. Rev. Lett.*, 71(11):1661–1664, 1993.

[CK03] A. Constantin and B. Kolev. Geodesic flow on the diffeomorphism group of the circle. *Comment. Math. Helv.*, 78(4):787–804, 2003.

[CKKS02] Tony F. Chan, Sung Ha Kang, Kang, and Jianhong Shen. Euler's elastica and curvature based inpaintings. *SIAM J. Appl. Math*, 63:564–592, 2002.

[CL95] P. Crouch and F. Silva Leite. The dynamic interpolation problem: On Riemannian manifold, Lie groups and symmetric spaces. *Journal of dynamical & Control Systems*, 1:177–202, 1995.

[CL08] Adrian Constantin and David Lannes. The hydrodynamical relevance of the camassa–holm and degasperis–procesi equations. *Archive for Rational Mechanics and Analysis*, 192(1):165–186, 2008.

[CLP95] M. Camarinha, F. Silva Leite, and P.Crouch. Splines of class C^k on non-euclidean spaces. *IMA Journal of Mathematical Control & Information*, 12:399–410, 1995.

[CM10] L. Caffarelli and R. J. McCann. Free boundaries in optimal transport and Monge-Ampere obstacle problems. *Annals of mathematics*, 171(2):673–730, 2010.

[CNPV16] Guillaume Charpiat, Giacomo Nardi, Gabriel Peyré, and François-Xavier Vialard. Piecewise rigid curve deformation via a Finsler steepest descent. *Interfaces Free Bound.*, 18(1):1–44, 2016.

[Con01] Adrian Constantin. On the scattering problem for the Camassa-Holm equation. *R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci.*, 457(2008):953–970, 2001.

[CPSV15] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. Unbalanced Optimal Transport: Geometry and Kantorovich Formulation. *ArXiv e-prints*, August 2015.

[CPSV16] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. Scaling Algorithms for Unbalanced Transport Problems. *ArXiv e-prints*, July 2016.

[CSPV16] L. Chizat, B. Schmitzer, G. Peyré, and F.-X. Vialard. An Interpolating Distance between Optimal Transport and Fisher-Rao. *Found. Comp. Math.*, 2016.

[CT13] Nicolas Charon and Alain Trouvé. The varifold representation of nonoriented shapes for diffeomorphic registration. *SIAM Journal on Imaging Sciences*, 6(4):2547–2580, 2013.

[Cut13] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 2292–2300. Curran Associates, Inc., 2013.

[Dan01] Raphaël Danchin. A few remarks on the Camassa-Holm equation. *Differential Integral Equations*, 14(8):953–988, 2001.

[DPF14] Guido De Philippis and Alessio Figalli. The Monge-Ampère equation and its link to optimal transportation. *Bull. Amer. Math. Soc. (N.S.)*, 51(4):527–580, 2014.

[EK13] Joachim Escher and Boris Kolev. Geodesic completeness for Sobolev H^s-metrics on the diffeomorphisms group of the circle, 2013.

[Eke78] Ivar Ekeland. The Hopf-Rinow theorem in infinite dimension. *J. Differential Geom.*, 13(2):287–301, 1978.

[EM70] David G. Ebin and Jerrold Marsden. Groups of diffeomorphisms and the motion of an incompressible fluid. *Ann. of Math. (2)*, 92:102–163, 1970.

[FCVP17] Jean Feydy, Benjamin CHARLIER, François-Xavier Vialard, and Gabriel Peyré. Optimal Transport for Diffeomorphic Registration. In *MICCAI 2017*, Proc. MICCAI 2017, Quebec, Canada, September 2017.

[FG89] D. S. Freed and D. Groisser. The basic geometry of the manifold of riemannian metrics and of its quotient by the diffeomorphism group. *Michigan Math. J.*, 36(3):323–344, 1989.

[FG15] Nicolas Fournier and Arnaud Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3):707–738, Aug 2015.

[Fig10] A. Figalli. The optimal partial transport problem. *Archive for rational mechanics and analysis*, 195(2):533–560, 2010.

[FPR+13] Sira Ferradans, Nicolas Papadakis, Julien Rabin, Gabriel Peyré, and Jean-François Aujol. *Regularized Discrete Optimal Transport*, pages 428–439. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[FRR+14] Jean-Baptiste Fiot, Hugo Raguet, Laurent Risser, Laurent D. Cohen, Jurgen Fripp, and François-Xavier Vialard. Longitudinal deformation models, spatial regularizations and learning strategies to quantify alzheimer's disease progression. *NeuroImage: Clinical*, 4(Supplement C):718 – 729, 2014.

[FZM+15] C. Frogner, C. Zhang, H. Mobahi, M. Araya, and T. Poggio. Learning with a Wasserstein loss. In *Advances in Neural Information Processing Systems*, volume 28, pages 2044–2052. 2015.

[GG02] Roberto Giambo and Fabio Giannoni. An analytical theory for riemannian cubic polynomials. *IMA Journal of Mathematical Control & Information*, 19:445–460, 2002.

[GHM+12a] F. Gay-Balmaz, D. D. Holm, D. M. Meier, T. S. Ratiu, and F.-X. Vialard. Invariant Higher-Order Variational Problems. *Communications in Mathematical Physics*, 309:413–458, January 2012.

[GHM+12b] F. Gay-Balmaz, D. D. Holm, D. M. Meier, T. S. Ratiu, and F.-X. Vialard. Invariant Higher-Order Variational Problems II. *Journal of NonLinear Science*, 22:553–597, August 2012.

[GHR11] Katrin Grunert, Helge Holden, and Xavier Raynaud. Lipschitz metric for the periodic camassa–holm equation. *Journal of Differential Equations*, 250(3):1460 – 1492, 2011.

[GM98] U. Grenander and M. I. Miller. Computational anatomy: An emerging discipline. *Quarterly of Applied Mathematics*, LVI(4):617–694, 1998.

[GPC15] A. Gramfort, G. Peyré, and M. Cuturi. Fast optimal transport averaging of neuroimaging data. In *Proc. IPMI'15*, 2015.

[Gre93] Ulf Grenander. *General Pattern Theory.* Oxford Science Publications, 1993.

[GTV13] F. Gay-Balmaz, C. Tronci, and C. Vizman. Geometric dynamics on the automorphism group of principal bundles: geodesic flows, dual pairs and chromomorphism groups. *Journal of Geometric Mechanics*, 5:39–84, 2013.

[Gui02] K. Guittet. Extended Kantorovich norms: a tool for optimization. Technical report, Tech. Rep. 4402, INRIA, 2002.

[GV16] T. Gallouët and F.-X. Vialard. From unbalanced optimal transport to the Camassa-Holm equation. *ArXiv e-prints*, September 2016.

[Han99] L.G. Hanin. An extension of the Kantorovich norm. *Contemporary Mathematics*, 226:113–130, 1999.

[HMR98] D. D. Holm, J. E. Marsden, and T. S. Ratiu. The Euler-Poincaré equations and semidirect products with applications to continuum theories. *Adv. Math.*, 137:1–81, 1998.

[Inc12] Hasan Inci. *On the Well-posedness of the Incompressible Euler Equation*. PhD thesis, Universität Zürich, 2012.

[JM00] S.C. Joshi and M.I. Miller. Landmark matching via large deformation diffeomorphisms. *International Journal of Computer Vision*, 2000.

[KLMP13] B. Khesin, J. Lenells, G. Misiołek, and S. C. Preston. Geometry of Diffeomorphism Groups, Complete integrability and Geometric statistics. *Geom. Funct. Anal.*, 23(1):334–366, 2013.

[KLT08] Thomas Kappeler, Enrique Loubet, and Peter Topalov. Riemannian exponential maps of the diffeomorphism groups of $\mathbb{T}^2$. *Asian J. Math.*, 12(3):391–420, 2008.

[KMV16] S. Kondratyev, L. Monsaingeon, and D. Vorotnikov. A new optimal trasnport distance on the space of finite Radon measures. *Adv. Differential Equations*, 21(11):1117–1164, 2016.

[Koi92] Norihito Koiso. Elasticae in a Riemannian submanifold. *Osaka J. Math.*, 29(3):539–543, 1992.

[Kol16] B. Kolev. Local well-posedness of the EPDiff equation: a survey. *ArXiv e-prints*, November 2016.

[KR58] L.V. Kantorovich and G.S. Rubinshtein. On a space of totally additive functions. *Vestn Lening. Univ.*, 13:52–59, 1958.

[KW09] Boris A. Khesin and Robert Wendt. *The Geometry of Infinite-Dimensional Groups*, volume 51. Springer-Verlag, 2009.

[LAFP11] Marco Lorenzi, Nicholas Ayache, Giovanni B. Frisoni, and Xavier Pennec. Mapping the effects of Aβ_{1-42} levels on the longitudinal changes in healthy aging: hierarchical modeling based on stationary velocity fields. In *Proceedings of Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 6892 of *LNCS*, pages 663–670. Springer, 2011.

[Len05] Jonatan Lenells. Conservation laws of the Camassa-Holm equation. *J. Phys. A, Math. Gen.*, 38(4):869–880, 2005.

[Léo12] C. Léonard. From the schrödinger problem to the monge–kantorovich problem. *Journal of Functional Analysis*, 262(4):1879–1920, 2012.

[Léo14] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst. A*, 34(4):1533–1574, 2014.

[LeV07] Randall J LeVeque. *Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems*, volume 98. Siam, 2007.

[LF73] E. H. Lee and G. E. Forsythe. Variational study of nonlinear spline curves. *SIAM Review*, 15(1):120–133, 1973.

[LM13] D. Lombardi and E. Maitre. Eulerian models and algorithms for unbalanced optimal transport. <hal-00976501v3>, 2013.

[LMS15] M. Liero, A. Mielke, and G. Savaré. Optimal Entropy-Transport problems and a new Hellinger-Kantorovich distance between positive measures. *ArXiv e-prints*, August 2015.

[LMS16] M. Liero, A. Mielke, and G. Savaré. Optimal transport in competition with reaction: the Hellinger-Kantorovich distance and geodesic curves. *SIAM J. Image Analysis*, 48(4):2869–2911, 2016.

[Lou14] Jean Louet. *Optimal transport problems with gradient penalization*. Theses, Université Paris Sud - Paris XI, July 2014.

[LS84] Joel Langer and David A. Singer. The total squared curvature of closed curves. *J. Differential Geom.*, 20(1):1–22, 1984.

[Mai12] E. Mainini. A description of transport cost for signed measures. *Journal of Mathematical Sciences*, 181(6):837–855, Mar 2012.

[McK04] Henry P. McKean. Breakdown of the Camassa-Holm equation. *Comm. Pure Appl. Math.*, 57(3):416–418, 2004.

[Mic02] Gheorghe Micula. A variational approach to spline functions theory. *General Mathematics*, 10:21–51, 2002.

[Mic08] P. W. Michor. *Topics in Differential Geometry*, volume 93 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2008.

[MM05] Peter W. Michor and David Mumford. Vanishing geodesic distance on spaces of submanifolds and diffeomorphisms. *Doc. Math.*, 10:217–245, 2005.

[MM13] Peter W. Michor and David Mumford. On Euler's equation and 'EPDiff'. *J. Geom. Mech.*, 5(3):319–344, 2013.

[MMM13] Mario Micheli, Peter W. Michor, and David Mumford. Sobolev metrics on diffeomorphism groups and the derived geometry of spaces of submanifolds. *Izvestiya: Mathematics*, 77(3):541–570, 2013.

[Mos65] J. Moser. On the volume elements on a manifold. *Trans. Amer. Math. Soc.*, 120:286–294, 1965.

[MP10] Gerard Misiołek and Stephen C. Preston. Fredholm properties of Riemannian exponential maps on diffeomorphism groups. *Invent. Math.*, 179(1):191–227, 2010.

[MRSS15] J. Maas, M. Rumpf, C. Schönlieb, and S. Simon. A generalized model for optimal transport of images including dissipation and density modulation. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6), Apr 2015. arXiv:1504.01988.

[MTY06] Michael I. Miller, Alain Trouvé, and Laurent Younes. Geodesic shooting for computational anatomy. *J. Math. Imaging Vis.*, 24(2):209–228, 2006.

[Mum94] David Mumford. Elastica and computer vision. In ChandrajitL. Bajaj, editor, *Algebraic Geometry and its Applications*, pages 491–506. Springer New York, 1994.

[NG14] L. Ning and T. T. Georgiou. Metrics for matrix-valued measures via test functions. In *53rd IEEE Conference on Decision and Control*, pages 2642–2647. IEEE, 2014.

[NGT15] L. Ning, T. T. Georgiou, and A. Tannenbaum. On matrix-valued monge–kantorovich optimal mass transport. *IEEE transactions on automatic control*, 60(2):373–382, 2015.

[NHP89] L. Noakes, G. Heinzinger, and B. Paden. Cubic splines on curved spaces. *IMA Journal of Mathematical Control & Information*, 6:465–473, 1989.

[NHV11] Marc Niethammer, Yang Huang, and François-Xavier Vialard. *Geodesic Regression for Image Time-Series*, pages 655–662. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[NPV16] Giacomo Nardi, Gabriel Peyré, and François-Xavier Vialard. Geodesics on shape spaces with bounded variation and Sobolev metrics. *SIAM J. Imaging Sci.*, 9(1):238–274, 2016.

[NV13] Marc Niethammer and François-Xavier Vialard. Riemannian metrics for statistics on shapes: Parallel transport and scale invariance. In *Proceedings of Miccai workshop, MFCA*, 2013.

[Ott01] Felix Otto. The geometry of dissipative evolution equations: The porous medium equation. *Communications in Partial Differential Equations*, 26(1-2):101–174, 2001.

[PCVS16] G. Peyré, L. Chizat, F.-X. Vialard, and J. Solomon. Quantum Optimal Transport for Tensor Field Processing. *ArXiv e-prints*, December 2016.

[PR13] B. Piccoli and F. Rossi. On properties of the Generalized Wasserstein distance. arXiv:1304.7014, 2013.

[PR14] B. Piccoli and F. Rossi. Generalized Wasserstein distance and its application to transport equations with source. *Archive for Rational Mechanics and Analysis*, 211(1):335–358, 2014.

[PW08] O. Pele and M. Werman. A linear time histogram metric for improved sift matching. In *European Conference on Computer Vision, ECCV 2008*, pages 495–508. 2008.

[Rez15] F Rezakhanlou. Optimal transport problems for contact structures, 2015.

[RGT97] Y. Rubner, L.J. Guibas, and C. Tomasi. The earth mover's distance, multi-dimensional scaling, and color-based image retrieval. In *Proceedings of the ARPA Image Understanding Workshop*, pages 661–668, 1997.

[Roc71] R.T. Rockafellar. Integrals which are convex functionals. II. *Pacific Journal of Mathematics*, 39(2):439–469, 1971.

[RVBS13] Laurent Risser, François-Xavier Vialard, Habib Y. Baluwala, and Julia A. Schnabel. Piecewise-diffeomorphic image registration: Application to the motion estimation between 3D CT lung images with sliding conditions. *Medical Image Analysis*, 17:182–193, 2013.

[RVW+11] L. Risser, F.-X. Vialard, R. Wolz, M. Murgasova, D. D. Holm, and D. Rueckert. Simultaneous multi-scale registration using large deformation diffeomorphic metric mapping. *IEEE Trans. Med. Imaging*, 30(10):1746–1759, 2011.

[SASK12] Chafik Samir, P.-A. Absil, Anuj Srivastava, and Eric Klassen. A gradient-descent method for curve fitting on riemannian manifolds. *Foundations of Computational Mathematics*, 12(1):49–73, 2012.

[SdGG+15] J. Solomon, F. de Goes, G. Peyré, M. Cuturi, A. Butscher, A. Nguyen, T. Du, and L. Guibas. Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics (Proc. SIGGRAPH 2015)*, 34(4):66:1–66:11, 2015.

[Sin64] R. Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *Ann. Math. Statist.*, 35:876–879, 1964.

[SRV13] Tanya Schmah, Laurent Risser, and François-Xavier Vialard. Left-invariant metrics for diffeomorphic image registration with spatially-varying regularisation. In Kensaku Mori, Ichiro Sakuma, Yoshinobu Sato, Christian Barillot, and Nassir Navab, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013: 16th International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part I*, pages 203–210, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[SRV15] Tanya Schmah, Laurent Risser, and François-Xavier Vialard. *Diffeomorphic Image Matching with Left-Invariant Metrics*, pages 373–392. Springer New York, New York, NY, 2015.

[SVN15] Nikhil Singh, François-Xavier Vialard, and Marc Niethammer. Splines for diffeomorphisms. *Medical Image Analysis*, 25(1):56–71, 2015.

[THD+14] Xiaoying Tang, Dominic Holland, Anders M. Dale, Laurent Younes, Michael I. Miller, and for the Alzheimer's Disease Neuroimaging Initiative. Shape abnormalities of subcortical and ventricular structures in mild cognitive impairment and alzheimer's disease: Detecting, quantifying, and predicting. *Human Brain Mapping*, 35(8):3701–3725, 2014.

[TPK+16] M. Thorpe, S. Park, S. Kolouri, G. K. Rohde, and D. Slepčev. A Transportation L^p Distance for Signal Analysis. *ArXiv e-prints*, September 2016.

[Tro95] Alain Trouvé. Action de groupe de dimension infinie et reconnaissance de formes. (infinite dimensional group action and pattern recognition). 1995.

[TV16] R. Tahraoui and F.-X. Vialard. Riemannian cubics on the group of diffeomorphisms and the Fisher-Rao metric. *ArXiv e-prints*, June 2016.

[TY05a] A. Trouvé and L. Younes. Local geometry of deformable templates. *SIAM J. Math. Anal.*, 37(1):17–59, 2005.

[TY05b] A. Trouvé and L. Younes. Metamorphoses through lie group action. *Foundations of Computational Mathematics*, 5(2):173–198, 2005.

[USK15] J. Ulen, P. Strandmark, and F. Kahl. Shortest paths with higher-order regularization. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 37(12):2588–2600, Dec 2015.

[VG05] M. Vaillant and J. Glaunès. Surface matching via currents. In *Information Processing in Medical Imaging*, volume 3565 of *Lecture Notes in Computer Science*, pages 381–392. Springer Berlin / Heidelberg, 2005.

[Vil08] C. Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.

[VR14] François-Xavier Vialard and Laurent Risser. Spatially-varying metric learning for diffeomorphic image registration: A variational framework. In Polina Golland, Nobuhiko Hata, Christian Barillot, Joachim Hornegger, and Robert Howe, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14-18, 2014, Proceedings, Part I*, pages 227–234, Cham, 2014. Springer International Publishing.

[VRHR11] 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'11)*, 2011.

[VRRC12] François-Xavier Vialard, Laurent Risser, Daniel Rueckert, and Colin J. Cotter. Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation. *Int. J. Comput. Vision*, 97:229–241, 2012.

[VT12] F.-X. Vialard and A. Trouvé. Shape Splines and Stochastic Shape Evolutions: A Second Order Point of View. *Quart. Appl. Math.*, 2012.

[YAM09] Laurent Younes, Felipe Arrate, and Michael I. Miller. Evolutions equations in computational anatomy. *NeuroImage*, 45(1, Supplement 1):S40 – S50, 2009. Mathematics in Brain Imaging.

[You08] Laurent Younes. *Shapes and Diffeomorphisms*. Springer, 2008.

[You12] Laurent Younes. Constrained diffeomorphic shape evolution. *Foundations of Computational Mathematics*, 12(3):295–325, Jun 2012.