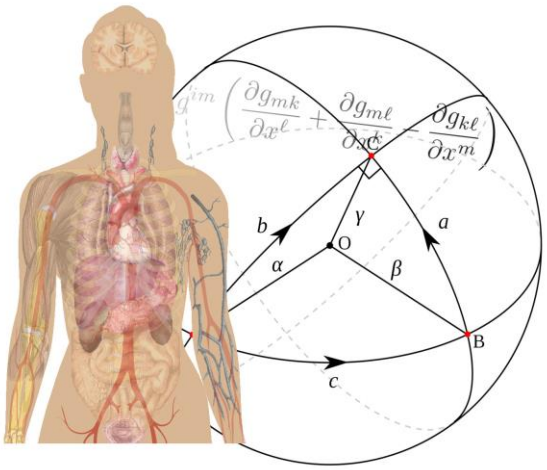
Proceedings of the fifth international workshop on Mathematical Foundations of Computational Anatomy (MFCA 2015)

Xavier Pennec, Sarang Joshi, Mads Nielsen, Thomas P. Fletcher, Stanley Durrleman, Stefan Sommer



MFCA 2015

5th MICCAI workshop on

Mathematical Foundations of of Computational Anatomy

Mathematical Foundations of Computational Anatomy

Geometrical and Statistical Methods for Biological Shape Variability Modeling

October 9th, 2015, Munich, Germany

http://www-sop.inria.fr/asclepios/events/MFCA15/
http://hal.inria.fr/MFCA/

Editors:

Xavier Pennec (INRIA Sophia-Antipolis, FR)
Sarang Joshi (SCI, University of Utah, USA)
Mads Nielsen (University of Copenhagen, DK)
Tom Fletcher (SCI, University of Utah, USA)
Stanley Durrleman (Inria / ICM, Paris, FR)
Stefan Sommer (University of Copenhagen, DK)

Inria informatics mathematics

MICCAI

Preface

Computational anatomy is an emerging discipline at the interface of geometry, statistics and image analysis which aims at modeling and analyzing the biological shape of tissues and organs. The goal is to estimate representative organ anatomies across diseases, populations, species or ages, to model the organ development across time (growth or aging), to establish their variability, and to correlate this variability information with other functional, genetic or structural information.

The **Mathematical Foundations of Computational Anatomy (MFCA)** workshop aims at fostering the interactions between the mathematical community around shapes and the MICCAI community in view of computational anatomy applications. It targets more particularly researchers investigating the combination of statistical and geometrical aspects in the modeling of the variability of biological shapes. The workshop is a forum for the exchange of the theoretical ideas and aims at being a source of inspiration for new methodological developments in computational anatomy. A special emphasis is put on theoretical developments, applications and results being welcomed as illustrations. Following the first edition of this workshop in 2006[1], the second edition in New-York in 2008[2], the third edition in Toronto in 2011[3], the forth edition in Nagoya Japan on September 22 2013[4], the fifth edition was held in Munich on October 9 2015[5].

Contributions were solicited in Riemannian, sub-Riemannian and group theoretical methods, advanced statistics on deformations and shapes, metrics for computational anatomy, statistics of surfaces, time-evolving geometric processes, stratified spaces, optimal transport, approximation methods in statistical learning and related subjects. Among the submitted papers, 14 were selected and organized in 4 oral sessions.

August 2015

Xavier Pennec
General Chair
MFCA'15

[1] http://www.inria.fr/sophia/asclepios/events/MFCA06/
[2] http://www.inria.fr/sophia/asclepios/events/MFCA08/
[3] http://www.inria.fr/sophia/asclepios/events/MFCA11/
[4] http://www.inria.fr/sophia/asclepios/events/MFCA13/
[5] http://www.inria.fr/sophia/asclepios/events/MFCA15/

Organization

Workshop Chairs

Xavier Pennec	(Asclepios team, INRIA Sophia-Antipolis, France)
Sarang Joshi	(SCI, University of Utah, USA)
Mads Nielsen	(University of Copenhagen, Denmark)
Tom Fletcher	(SCI, University of Utah, USA)
Stanley Durrleman	(ARAMIS team, Inria / ICM, Paris, FR)
Stefan Sommer	(University of Copenhagen, DK)

Program Committee

Stéphanie Allassonière	(Ecole Polytechnique, France)
Rachid Deriche	(INRIA, France)
Ian L. Dryden	(University of Nottingham, UK)
Aasa Feragen	(University of Copenhagen, DK)
Luc Florac	(Eindhoven U. of Technology, NL)
James Gee	(U. of Pennsylvania, USA)
Guido Gerig	(SCI, University of Utah, USA)
Polina Golland	(MIT, USA)
Darryl Holm	(Imperial College London, UK)
Susan Holmes	(Stanford U., USA)
Steve Marron	(UNC Chapel Hill, USA)
Stephen Marsland	(Massey University, NZ)
Yoshitaka Masutani	(U. of Tokyo Hosp, JP)
Michael Miller	(Johns Hopkins U., USA)
Marc Niethammer	(UNC Chapel Hill, USA)
Salvador Olmos	(U. of Saragossa, SP)
Jerry Prince	(Johns Hopkins U., USA)
Anand Rangarajan	(U. of Florida, USA)
Daniel Rueckert	(Imperial Coll. London, UK)
Kaleem Siddiqi	(McGill U., CA)
Martin Styner	(UNC Chapel Hill, USA)
Hemant D Tagare	(Yale University, USA)
Paul Thompson	(U. Southern California, USA)
Alain Trouvé	(ENS Cachan, FR)
Carole Twining	(U. of Manchester, UK)
Baba Vemuri	(U. of Florida, USA)
Francois Xavier Vialard	(Dauphine U., FR)
Rene Vidal	(Johns Hopkins U., USA)
Laurent Younes	(Johns Hopkins U., USA)
William M. Wells III	(MIT & Harvard, USA)

Table of Contents

Diffeomorphisms and matching of densities and shape

Longitudinal and multivariate analysis on manifolds

Matching the shape of curves and signals

Fréchet mean on manifolds and quotient spaces

Weighted Diffeomorphic Density Matching with Applications to Thoracic Image Registration

Caleb Rottman[1], Martin Bauer[2], Klas Modin[3], Sarang Joshi[1]

[1] Department of Bioengineering, Scientific Computing and Imaging Institute, University of Utah
[2] Fakultät für Mathematik, Universität Wien
[3] Department of Mathematical Sciences, Chalmers University of Technology and the University of Gothenburg

Abstract. In this article we study the problem of thoracic image registration, in particular the estimation of complex anatomical deformations associated with the breathing cycle. Using the intimate link between the Riemannian geometry of the space of diffeomorphisms and the space of densities, we develop an image registration framework that incorporates both the fundamental law of conservation of mass as well as spatially varying tissue compressibility properties. By exploiting the geometrical structure, the resulting algorithm is computationally efficient, yet widely general.

Keywords: density matching, image registration, Fisher-Rao metric, thoracic image registration

1 Introduction

In this paper we consider the problem of tracking organs undergoing deformations as a result of breathing in the thorax and imaged via computed tomography (CT). This problem has wide scale medical applications, in particular radiation therapy of the lung where accurate estimation of organ deformations during treatment impacts dose calculation and treatment decisions [8, 12, 18, 22]. The current state-of-the-art radiation treatment planning involves the acquisition of a series of respiratory correlated CT (RCCT) images to build 4D (3 spatial and 1 temporal) treatment planning data sets. Fundamental to the processing and clinical use of these 4D data sets is the accurate estimation of registration maps that characterize the motion of organs at risk as well as the target tumor volumes.

The 3D image produced from X-ray CT is an image of linear attenuation coefficients. The linear attenuation coefficient μ of a material is defined as $\mu = \alpha_m \rho_m$, where α_m is the mass attenuation coefficient of the material and ρ_m is the mass density. The linear attenuation coefficient is proportional to the true density and therefore exhibits conservation of mass.

Currently, the application of diffeomorphisms in medical image registration is mostly limited to the L^2 image action of the diffeomorphism group, which is not

a mass-preserving transformation. Furthermore, the diffeomorphisms estimated from typical image registrations algorithms (such as LDDMM [5] or ANTS [1]) do not accurately model the varying compressibility of different tissues types. In thoracic datasets, the lungs are highly compressible. Conversely, the bronchial tubes and the tissue surrounding the lungs are incompressible. During inhale, as air enters, the lung volume increases and the lung density decreases, while during exhale lung volume decreases and the lung density increases. But in both inhale and exhale, the lung mass is conserved.

In this paper we use a cone-beam CT dataset of a rat acquired at 11 time points of an inhale-exhale breathing cycle. Figure 1 shows the mass, volume, and density of the lungs of a rat at each time point of its breathing cycle, exemplifying these properties.

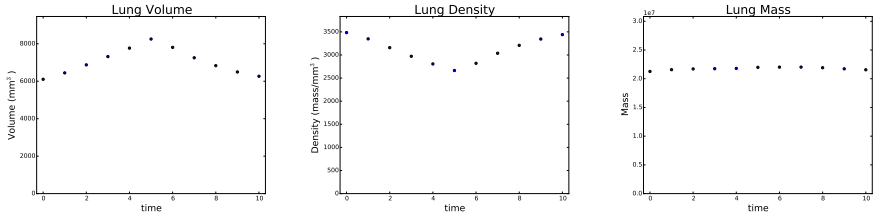


Fig. 1: Rat lung data: volume, density, and mass of the lungs during an inhale-exhale breathing cycle. As the volume increases, the density decreases, but mass is conserved.

Both of these effects can be clearly seen in the histograms of a full-inhale and a full-exhale image, as shown in Figure 2.

In 2010, the EMPIRE10 [15] challenge compared registration algorithms applied to intra-patient thoracic CT images. The winner of the competition used an LDDMM method using normalized cross correlation metric [21]. This method does not model conservation of mass or spatially varying tissue compressibility. While others in this competition used the density action on these images [6,9], none of these methods incorporate the spatially varying nature of tissue compressibility.

We present an image registration technique that incorporates conservation of mass and organ compressibility. Instead of the L^2 image action of diffeomorphisms, we use the physiologically appropriate density action. We also regularize the diffeomorphism by using a space-varying penalty which allows for high compressibility of the lung tissue while at the same time enforcing incompressibility of high density structures such as bone. The algorithm is based on the intimate link between the Riemannian geometry of the space of diffeomorphisms and the space of densities [4,13,14]. The resulting algorithm also has the added advantage that it is computationally efficient: orders of magnitude faster than existing diffeomorphic image registration algorithms.

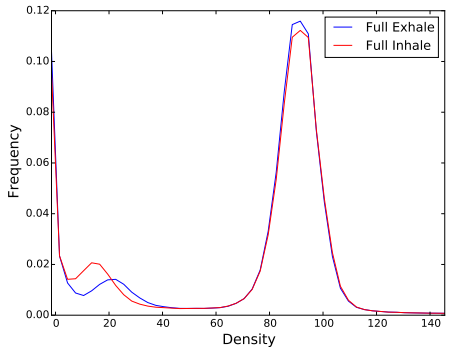


Fig. 2: Histograms of a full-inhale and full-exhale image. Each histogram has three peaks: the peak at 0 represents surrounding air, the middle peak represents lung tissue, and the peak at 90 represents soft tissue. For the lung tissue, the full inhale has higher volume but a lower image intensity than the full exhale, therefore showing conservation of mass. For the soft tissue, the average intensity does not change because it is incompressible. The slight drop in frequency of the full inhale is due to soft tissue leaving the image boundary.

2 Mathematical Formulation

Mathematically, the problem is to find a diffeomorphic (bijective and smooth) transformation between two densities on a subset $\Omega \subset \mathbb{R}^3$. With a 'density' we mean a volume form on Ω, i.e., an element of the form $I\,dx$ where $dx = dx^1 \wedge dx^2 \wedge dx^3$ is the standard volume element on $\mathbb{R}^3$ and $I = I(x)$ is a non-negative function on Ω. The space of all densities on Ω is denoted $\mathrm{Dens}(\Omega)$. One might, of course, identify $I\,dx$ with its function I, and thereby think of $\mathrm{Dens}(\Omega)$ as the set of non-negative functions on Ω. However, the invariance properties and geometry of the problem are remarkably more transparent when viewing $\mathrm{Dens}(M)$ as a space of volume forms.

The group of diffeomorphisms $\mathrm{Diff}(\Omega)$ acts from the right on $\mathrm{Dens}(\Omega)$ by *pullback:* the action of $\varphi \in \mathrm{Diff}(\Omega)$ on $I\,dx \in \mathrm{Dens}(\Omega)$ is given by

$$(\varphi, I\,dx) \mapsto \varphi^*(I\,dx) = (|D\varphi|\, I \circ \varphi)\,dx, \tag{1}$$

where $|D\varphi|$ denotes the Jacobian determinant of φ. The corresponding left action is given by *pushforward:*

$$(\varphi, I\,dx) \mapsto \varphi_*(I\,dx) = (\varphi^{-1})^*(I\,dx) = \left(|D\varphi^{-1}|\, I \circ \varphi^{-1}\right) dx. \tag{2}$$

The Riemannian geometry of the group of diffeomorphisms endowed with a suitable Sobolev H^1 metric is intimately linked to the Riemannian geometry of the space densities with the Fisher–Rao metric. This has been developed and extensively studied in [4, 13, 14]: the basic observation is that there are Sobolev

H^1-metrics on the space of diffeomorphisms that descend to the Fisher–Rao metric on the space of densities.

The distance associated with the Fisher–Rao metric is traditionally defined between *probability densities* (densities of total mass 1) and is given by

$$d_F(\mu_0, \mu_1) = \sqrt{\mathrm{vol}(\Omega)} \arccos\left(\frac{1}{\mathrm{vol}(\Omega)} \int_\Omega \sqrt{\frac{\mu_0}{dx}\frac{\mu_1}{dx}} dx \right) , \tag{3}$$

where μ_0 and μ_1 are probability densities. It naturally extends to the space of all densities and the case when $\mathrm{vol}(\Omega) = \infty$, for which it is given by

$$d_F^2(I_0\, dx, I_1\, dx) = \int_\Omega (\sqrt{I_0} - \sqrt{I_1})^2 dx\,. \tag{4}$$

Notice that $d_F^2(\cdot,\cdot)$ in this case is the *Hellinger distance*. For details, see [4].

The Fisher–Rao metric is the unique Riemannian metric on the space of probability densities that is invariant under the action of the diffeomorphism group [2,3]. This invariance property extends to the induced distance function, so

$$d_F^2(I_0\, dx, I_1\, dx) = d_F^2(\varphi_*(I_0\, dx), \varphi_*(I_1\, dx)) \qquad \forall \varphi \in \mathrm{Diff}(\Omega)\,. \tag{5}$$

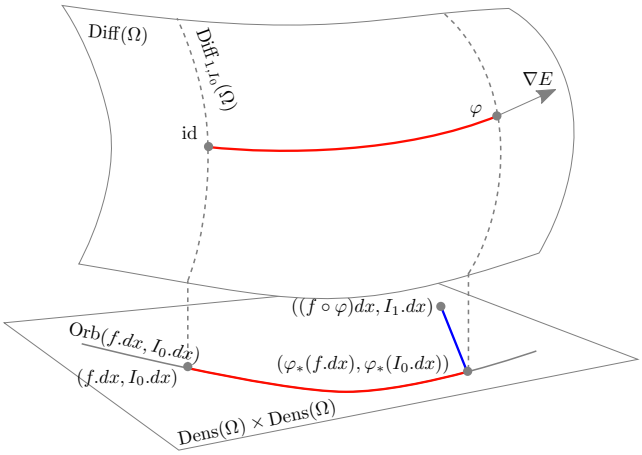


Fig. 3: Illustration of the geometry associated with the density matching problem. The gradient flow on $\mathrm{Diff}(\Omega)$ descends to a gradient flow on the orbit $\mathrm{Orb}(f\, dx, I_0\, dx)$. While constrained to $\mathrm{Orb}(f\, dx, I_0\, dx) \subset \mathrm{Dens}(\Omega) \times \mathrm{Dens}(\Omega)$, this flow strives to minimize the product Fisher-Rao distance to $((f \circ \varphi)\, dx, I_1\, dx)$.

Motivated by the aforementioned properties, we develop a weighted diffeomorphic matching algorithm for matching two density images. The algorithm

is based on the Sobolev H^1 gradient flow on the space of diffeomorphisms that minimizes the energy functional

$$E(\varphi) = d_F^2(\varphi_*(f\,dx), (f \circ \varphi^{-1})dx) + d_F^2(\varphi_*(I_0\,dx), I_1\,dx)). \tag{6}$$

This energy functional is only a slight modification of the energy functional studied in [4]. Indeed, if f in the above equation is a constant $\sigma > 0$, then (6) reduces to the energy functional of Bauer, Joshi, and Modin [4, § 5.1]. Moreover, the geometry described in [4, § 5.3] is valid also for the functional (6), and, consequently, the algorithm developed in [4, § 5.2] can be used also for minimizing (6). There the authors view the energy functional as a constrained minimization problem on the product space $\mathrm{Dens}(\Omega) \times \mathrm{Dens}(\Omega)$ equipped with the product distance, cf. Fig 3 and [4, § 5] for details on the resulting geometric picture. Related work on diffeomorphic density matching using the Fisher Rao metric can be found in [19, 20].

Using the invariance property of the Fisher-Rao metric and assuming infinite volume, the main optimization problem associated with the energy functional (6) is the following.

Given densities $I_0\,dx$, $I_1\,dx$, and $f\,dx$, find $\varphi \in \mathrm{Diff}(\Omega)$ minimizing

$$E(\varphi) = \underbrace{\int_\Omega (\sqrt{|D\varphi^{-1}|} - 1)^2\, f \circ \varphi^{-1}\, dx}_{E_1(\varphi)} + \underbrace{\int_\Omega \left(\sqrt{|D\varphi^{-1}| I_0 \circ \varphi^{-1}} - \sqrt{I_1}\right)^2 dx}_{E_2(\varphi)} . \tag{7}$$

The invariance of the Fisher-Rao distance can be seen with a simple change of variables $x \mapsto \varphi(y)$, $dx \mapsto |D\varphi| dy$, and $|D\varphi^{-1}| \mapsto \frac{1}{|D\varphi|}$. Then, Equation 7 becomes

$$E(\varphi) = \int_\Omega (1 - \sqrt{|D\varphi|})^2\, f\, dy + \int_\Omega \left(\sqrt{I_0} - \sqrt{|D\varphi| I_1 \circ \varphi}\right)^2 dy\,. \tag{8}$$

To better understand the energy functional $E(\varphi)$ we consider the two terms separately. The first term $E_1(\varphi)$ is a *regularity measure* for the transformation. It penalizes the deviation of the diffeomorphism φ from being volume preserving. The density $f\,dx$ acts as a weighting on the domain Ω. That is, change of volume (compression and expansion of the transformation φ) is penalized more in regions of Ω where f is large. The second term $E_2(\varphi)$ penalizes *dissimilarity* between $I_0\,dx$ and $\varphi^*(I_1\,dx)$. It is the Fisher–Rao distance between the initial density $I_0\,dx$ and the transformed target density $\varphi^*(I_1\,dx)$. Because of the invariance (5) of the Fisher–Rao metric, this is the same as the Fisher–Rao distance between $I_1\,dx$ and $\varphi_*(I_0\,dx)$.

Solutions to problem (7) are *not* unique. To see this, let $\mathrm{Diff}_I(\Omega)$ denote the space of all diffeomorphisms preserving the volume form $I\,dx$:

$$\mathrm{Diff}_I(\Omega) = \{\varphi \in \mathrm{Diff}(\Omega) \mid \; |D\varphi|\,(I \circ \varphi) = I\}. \tag{9}$$

If φ is a minimizer of $E(\cdot)$, then $\psi \circ \varphi$ for any

$$\psi \in \mathrm{Diff}_{1,I_0}(\Omega) \coloneqq \mathrm{Diff}_1(\Omega) \cap \mathrm{Diff}_{I_0}(\Omega) \tag{10}$$

is also a minimizer. Notice that this space is not trivial. For example, any diffeomorphism generated by a *Nambu–Poisson vector field* (see [16]), with I_0 as one of its Hamiltonians, will belong to it. A strategy to handle the degeneracy was developed in [4, § 5]: the fact that the metric is descending with respect to the H^1 metric on $\mathrm{Diff}(\Omega)$ can be used to ensure that the gradient flow is *infinitesimally optimal*, i.e., always orthogonal to the null-space. We employ the same strategy in this paper. The corresponding geometric picture can be seen in Fig. 3.

3 Gradient Flow Algorithm Development

We now derive in detail the algorithm used to optimize the functional defined in Equation 8. The H^1-metric on the space of diffeomorphisms is defined using the Hodge laplacian on vector fields and is given by:

$$G^I_\varphi(U,V) = \int_\Omega \langle -\Delta u, v\rangle dx \ . \tag{11}$$

Due to its connections to information geometry we also refer to this metric as *information metric*. Let $\nabla^{G^I} E$ denote the gradient with respect to the information metric defined above. Our approach to minimize the functional of (8) is to use a simple Euler integration of the discretization of the gradient flow:

$$\dot{\varphi} = -\nabla^{G^I} E(\varphi) \tag{12}$$

The resulting final algorithm (Algorithm 1) is order of magnitudes faster than LDDMM, since we are not required to time integrate the geodesic equations, as necessary in LDDMM [23].

In the following theorem we calculate the gradient of the energy functional:

Theorem 1. *The G^I–gradient of the matching functional* (8) *is given by*

$$\begin{aligned}\nabla^{G^I} E = -\Delta^{-1}\Big(&- \nabla\big(f \circ \varphi^{-1}(1 - \sqrt{|D\varphi^{-1}|})\big) - \\ &\sqrt{|D\varphi^{-1}|\, I_0 \circ \varphi^{-1}}\nabla\big(\sqrt{I_1}\big) + \nabla\big(\sqrt{|D\varphi^{-1}|\, I_0 \circ \varphi^{-1}}\big)\sqrt{I_1}\Big) \ .\end{aligned} \tag{13}$$

Remark 2. Notice that in the formula for $\nabla^{G^I} E$ we never need to compute φ, so in practice we only compute φ^{-1}. We update this directly via $\varphi^{-1}(y) \mapsto \varphi^{-1}(y + \epsilon \nabla^{G^I} E)$ for some step size ϵ.

Proof. We first calculate the variation of the energy functional. Therefore let φ_s be a family of diffeomorphisms parameterized by the real variable s, such that

$$\varphi_0 = \varphi \quad \text{and} \quad \frac{d}{ds}\Big|_{s=0} \varphi_s = v \circ \varphi. \tag{14}$$

We use the following identity, as derived in [10]:

$$\frac{d}{ds}\Big|_{s=0}\sqrt{|D\varphi_s|} = \frac{1}{2}\sqrt{|D\varphi|}\mathrm{div}(v)\circ\varphi. \tag{15}$$

The variation of the first term of the energy functional is

$$\frac{d}{ds}\Big|_{s=0}E_1(\varphi) = \int_\Omega f(x)(\sqrt{|D\varphi(x)|}-1)\sqrt{|D\varphi(x)|}\mathrm{div}(v)\circ\varphi(x)dx \tag{16}$$

We do a change of variable $x \mapsto \varphi^{-1}(y)$, $dx \mapsto |D\varphi^{-1}(y)|dy$, using the fact that $|D\varphi(x)| = \frac{1}{|D\varphi^{-1}(y)|}$;

$$= \int_\Omega f\circ\varphi^{-1}(y)(1-\sqrt{|D\varphi^{-1}(y)|})\mathrm{div}(v)(y)dy \tag{17}$$

$$= \left\langle f\circ\varphi^{-1}(1-\sqrt{|D\varphi^{-1}|}), \mathrm{div}(v)\right\rangle_{L^2(\mathbb{R}^3)} \tag{18}$$

$$= -\left\langle \nabla\left(f\circ\varphi^{-1}(1-\sqrt{|D\varphi^{-1}|})\right), v\right\rangle_{L^2(\mathbb{R}^3)} \tag{19}$$

using the fact that the adjoint of the divergence is the negative gradient. For the second term of the energy functional, we expand the square

$$E_2(\varphi) = \int_\Omega I_0(x) - 2\sqrt{I_0(x)I_1\circ\varphi(x)|D\varphi(x)|} + I_1\circ\varphi(x)|D\varphi(x)|dx \tag{20}$$

Now $\int_\Omega I_1\circ\varphi(x)|D\varphi(x)|dx$ is constant (conservation of mass), so we only need to minimize over the middle term. The derivative is then

$$\begin{aligned}\frac{d}{ds}\Big|_{s=0}E_2(\varphi) = -\int_\Omega & 2\sqrt{I_0(x)}\left(\nabla\sqrt{I_1}^T v\right)\circ\varphi(x)\sqrt{|D\varphi(x)|} \\ & - \sqrt{I_0(x)I_1\circ\varphi(x)|D\varphi(x)|}\mathrm{div}(v)\circ\varphi(x)dx.\end{aligned} \tag{21}$$

We do the same change of variables as before:

$$= -\int_\Omega \sqrt{I_0\circ\varphi^{-1}(y)}\frac{|D\varphi^{-1}(y)|}{\sqrt{|D\varphi^{-1}(y)|}}\left(2\nabla\sqrt{I_1(y)}^T v(y) + \sqrt{I_1(y)}\mathrm{div}(v)(y)\right) \tag{22}$$

$$\begin{aligned}= & -\left\langle 2\sqrt{|D\varphi^{-1}|\, I_0\circ\varphi^{-1}}\nabla\sqrt{I_1}, v\right\rangle_{L^2(\mathbb{R}^3)} \\ & - \left\langle \sqrt{|D\varphi^{-1}|\, I_0\circ\varphi^{-1}I_1}, \mathrm{div}(v)\right\rangle_{L^2(\mathbb{R}^3)}\end{aligned} \tag{23}$$

$$\begin{aligned}= & \left\langle -\sqrt{|D\varphi^{-1}|\, I_0\circ\varphi^{-1}}\nabla\sqrt{I_1}, v\right\rangle_{L^2(\mathbb{R}^3)} \\ & + \left\langle \nabla\left(\sqrt{|D\varphi^{-1}|\, I_0\circ\varphi^{-1}}\right)\sqrt{I_1}, v\right\rangle_{L^2(\mathbb{R}^3)}.\end{aligned} \tag{24}$$

From the above equations we conclude that:

$$\begin{aligned}-\Delta(\nabla^{G^I}E) = & -\nabla\left(f\circ\varphi^{-1}(1-\sqrt{|D\varphi^{-1}|})\right) \\ & - \sqrt{|D\varphi^{-1}|\, I_0\circ\varphi^{-1}}\nabla\sqrt{I_1} + \nabla\left(\sqrt{|D\varphi^{-1}|\, I_0\circ\varphi^{-1}}\right)\sqrt{I_1}\end{aligned} \tag{25}$$

Since we are taking the Sobolev gradient of E, we apply the inverse Laplacian to the right hand side of Equation 25 to solve for $\nabla^{G^I} E$.

Algorithm 1 Final Algorithm

Choose $\epsilon > 0$
Set $\varphi^{-1} = \text{id}$
Set $|D\varphi^{-1}| = 1$
for $iter = 1...\text{NumIters}$ **do**
  Compute $\varphi_* I_0 = I_0 \circ \varphi^{-1} |D\varphi^{-1}|$
  Compute $u = -\nabla\big(f \circ \varphi^{-1}(1 - \sqrt{|D\varphi^{-1}|})\big) - \sqrt{\varphi_* I_0}\nabla\sqrt{I_1} + \nabla(\sqrt{\varphi_* I_0})\sqrt{I_1}$
  Compute $v = -\Delta^{-1}(u)$
  Update $\varphi^{-1} \mapsto \varphi^{-1}(y + \epsilon v)$
  Update $|D\varphi^{-1}| \mapsto |D\varphi^{-1}| \circ \varphi^{-1} \mathrm{e}^{-\epsilon \mathrm{div}(v)}$
end for

Remark 3. Algorithm 1 constructs the mapping φ^{-1} by numerically integrating the vector field v. Thus, for small enough ϵ, the computed transformation φ^{-1} is a diffeomorphism (as is also the case in LDDMM).

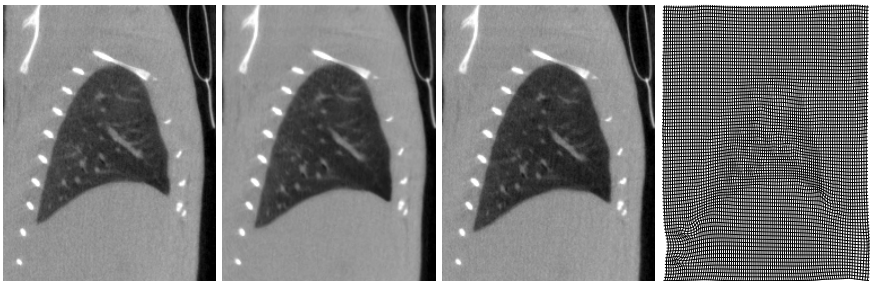
4 Results

We applied the proposed method to the previously mentioned rat dataset. In this dataset, an anesthetized rat was placed on a mechanical ventilator. This ventilator sent 11 gate signals to the cone-beam CT per breathing cycle, assuring that all projections would all be acquired at a consistent points of the breathing cycle [11]. Previous literature has shown that cone-beam CT is inadequate in estimating the true linear attenuating coefficient density [7], so we empirically estimated the density as the square of the the original data.

For these results we estimated the deformation from the full-exhale to the full-inhale image. The deformation was computed on the resolution of the original 3D volume ($245 \times 189 \times 217$); all the figures show the same 2D coronal slice of this volume. Shown in Fig. 4 are the coronal sections of full exhale, the full exhale deformed via the density action, and the corresponding image at full inhale and the estimated deformation.

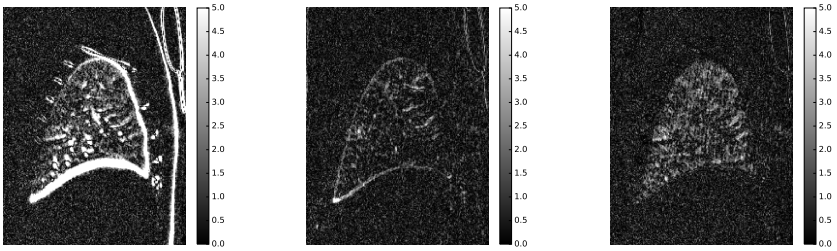
For the compressibility penalty f, we used a soft thresholding of the intensity values of the initial image using the logistic function. High intensity regions of the CT image (corresponding to bone and soft tissue) were given a high penalty ($f(x) = 10\sigma$) and low intensity regions of the CT image (corresponding to air and lungs) were given a low penalty ($f(x) = .1\sigma$) (see Figure 7)

We implemented the proposed algorithm and LDDMM on a single Titan-Z GPU (using the PyCA software package [17] bitbucket.org/scicompanat/pyca) for comparison. The difference images are pictured in Figure 5. The problem of



I_{ex} $\varphi_*(I_{ex}dx)$ I_{in} φ^{-1}

Fig. 4: Density action results. This figure shows the lung image at the full exhale, the full exhale deformed via the density action, and the corresponding image at full inhale. Shown in the right panel is the estimated deformation.

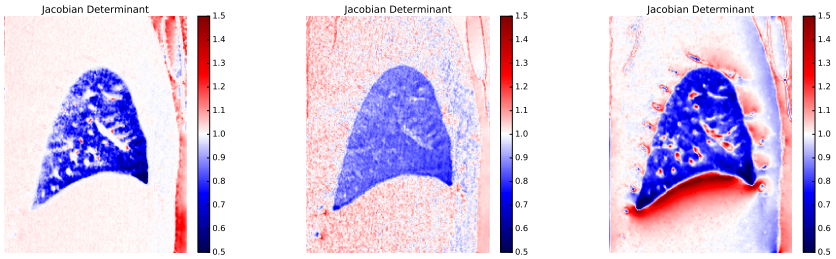


$|I_{in} - I_{ex}|$ $|I_{in} - \varphi_*(I_{ex}dx)|$ $|I_{in} - I_{ex} \circ \varphi^{-1}|$ (LDDMM)

Fig. 5: Absolute value of image differences: The left panel shows the difference between the original full exhale and the full inhale images. The center panel shows the result after registration using the proposed method. The right image shows the result using LDDMM with the L^2 image action. In LDDMM, there is significant error inside the lung due to the L^2 action not preserving mass.

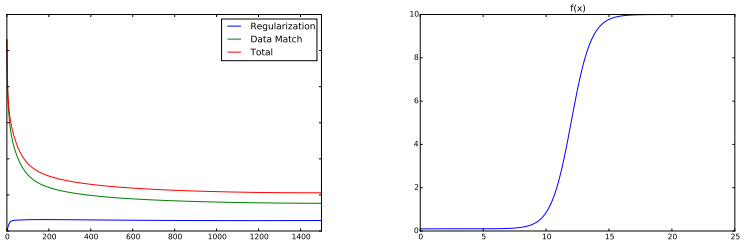
LDDMM using the L^2 action can be seen in this image. The Jacobian determinants are in Figure 6. The proposed method constrains the contraction and expansion to inside the lung and outside the body. In this figure we also show the results of using the density action with a constant penalty function ($f(x) = \sigma.$).

The proposed algorithm is significantly faster than LDDMM; it runs at 400 iterations per minute while LDDMM runs at 45 iterations per minute. We used 10 time steps to integrate the geodesic equations associated with the LDDMM formulation. Since we are not required to integrate the geodesic equations in the proposed algorithm, we have nearly a 10x speedup compared to LDDMM.



$f(x) = \text{sig}(I_0(x))$ $f(x) = 1$ LDDMM

Fig. 6: Jacobian determinants: On the left is the Jacobian determinant of the transformation estimated by the proposed method. Notice that the volume change is confined to inside the lungs and outside the body. In the center we use the density action, but without a local-varying penalty (i.e. $f(x) = \sigma$). On the right is the Jacobian determinant using LDDMM. Without the local-varying penalty, there is contraction and expansion outside of the lungs. In LDDMM, the contraction and expansion outside of the lungs is even more severe.



Energy $\text{sig}(x)$

Fig. 7: Energy plot and the logistic function used for the penalty.

5 Discussion

In this paper, we introduced a computationally efficient method for estimating registration maps between thoracic CT images. The proposed solution accurately incorporates the fundamental property of mass conservation and the spatially varying compressibility of thoracic anatomy. We conserve mass by viewing the images as densities and applying the density action of a diffeomorphism instead of the typical L^2 action. We limit the volume change in incompressible organs by placing a space-varying penalty on the Jacobian determinant of the diffeomorphism. While any non-negative function $f(x)$ can be used, we simply use a soft-thresholding function on the initial image. This choice is based on the assumption that low CT image values (such as the lungs and air) exhibit a large

amount of volume change whereas high images values (such as other soft tissue and bone) are quite incompressible.

Acknowledgments

The authors thank Rick Jacob at the Pacific Northwest National Laboratory for the imaging data which was funded by a grant from the National Heart, Lung, and Blood Institute of the National Institutes of Health (R01 HL073598). The works was partially supported by the grant NIH R01 CA169102-01A13, the Swedish Foundation for Strategic Research (ICA12-0052), an EU Horizon 2020 Marie Sklodowska-Curie Individual Fellowship (661482) and by the Erwin Schrödinger Institute programme: Infinite-Dimensional Riemannian Geometry with Applications to Image Matching and Shape Analysis. M. Bauer was supported by the European Research Council (ERC), within the project 306445 (Isoperimetric Inequalities and Integral Geometry) and by the FWF-project P24625 (Geometry of Shape spaces).

References

1. Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C.: A reproducible evaluation of {ANTs} similarity metric performance in brain image registration. NeuroImage 54(3), 2033 – 2044 (2011), http://www.sciencedirect.com/science/article/pii/S1053811910012061
2. Ay, N., Jost, J., Le, H.V., Schwachhöfer, L.: Information geometry and sufficient statistics. The annals of statistics (2014)
3. Bauer, M., Bruveris, M., Michor, P.W.: Uniqueness of the Fisher–Rao metric on the space of smooth densities. submitted (2015)
4. Bauer, M., Joshi, S., Modin, K.: Diffeomorphic density matching by optimal information transport. accepted in SIIMS (2015)
5. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International Journal of Computer Vision 61(2), 139–157 (2005)
6. Cao, K., Du, K., Ding, K., Reinhardt, J., Christensen, G.: Regularized nonrigid registration of lung CT images by preserving tissue volume and vesselness measure. Medical Image Analysis for the Clinic-A Grand Challenge pp. 43–54 (2010)
7. De Vos, W., Casselman, J., Swennen, G.: Cone-beam computerized tomography (cbct) imaging of the oral and maxillofacial region: a systematic review of the literature. International journal of oral and maxillofacial surgery 38(6), 609–625 (2009)
8. Geneser, S.E., Hinkle, J., Kirby, R.M., Wang, B., Salter, B., Joshi, S.: Quantifying variability in radiation dose due to respiratory-induced tumor motion. Medical image analysis 15(4), 640–649 (2011)
9. Gorbunova, V., Sporring, J., Lo, P.: Mass preserving image registration: Results of evaluation of methods for pulmonary image registration 2010 challenge. : A Grand Challenge pp. 155–164 (2010), http://www.diagnijmegen.nl/~bram/grandchallenge2010/155.pdf

10. Hinkle, J., Joshi, S.: Idiff: irrotational diffeomorphisms for computational anatomy. In: Information Processing in Medical Imaging. pp. 754–765. Springer (2013)
11. Jacob, R.E., Lamm, W.J.: Stable small animal ventilation for dynamic lung imaging to support computational fluid dynamics models. PLoS ONE 6(11) (2011)
12. Keall, P.J., Joshi, S., Vedam, S.S., Siebers, J.V., Kini, V.R., Mohan, R.: Four-dimensional radiotherapy planning for dmlc-based respiratory motion tracking. Medical physics 32(4), 942–951 (2005)
13. Khesin, B., Lenells, J., Misiołek, G., Preston, S.C.: Geometry of Diffeomorphism Groups, Complete integrability and Geometric statistics. Geom. Funct. Anal. 23(1), 334–366 (2013)
14. Modin, K.: Generalized Hunter–Saxton equations, optimal information transport, and factorization of diffeomorphisms. J. Geom. Anal. 25(2), 1306–1334 (2015)
15. Murphy, K., van Ginneken, B., Reinhardt, J., Kabus, S., Ding, K., Deng, X., Cao, K., Du, K., Christensen, G., Garcia, V., Vercauteren, T., Ayache, N., Commowick, O., Malandain, G., Glocker, B., Paragios, N., Navab, N., Gorbunova, V., Sporring, J., de Bruijne, M., Han, X., Heinrich, M., Schnabel, J., Jenkinson, M., Lorenz, C., Modat, M., McClelland, J., Ourselin, S., Muenzing, S., Viergever, M., De Nigris, D., Collins, D., Arbel, T., Peroni, M., Li, R., Sharp, G., Schmidt-Richberg, A., Ehrhardt, J., Werner, R., Smeets, D., Loeckx, D., Song, G., Tustison, N., Avants, B., Gee, J., Staring, M., Klein, S., Stoel, B., Urschler, M., Werlberger, M., Vandemeulebroucke, J., Rit, S., Sarrut, D., Pluim, J.: Evaluation of registration methods on thoracic ct: The empire10 challenge. Medical Imaging, IEEE Transactions on 30(11), 1901–1920 (Nov 2011)
16. Nakanishi, N.: A survey of Nambu-Poisson geometry. Lobachevskii J. Math. 4, 5–11 (electronic) (1999)
17. Preston, J., Hinkle, J., Singh, N., Rottman, C., Joshi, S.: Pyca - python for computational anatomy. `https://bitbucket.org/scicompanat/pyca` (2012-2015)
18. Sawant, A., Keall, P., Pauly, K.B., Alley, M., Vasanawala, S., Loo Jr, B.W., Hinkle, J., Joshi, S.: Investigating the feasibility of rapid mri for image-guided motion management in lung cancer radiotherapy. BioMed research international 2014 (2014)
19. Seo, D., Ho, J., Vemuri, B.: Computing diffeomorphic paths for large motion interpolation. In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. pp. 1227–1232 (June 2013)
20. Seo, D., Ho, J., Traverse, J., Forder, J., Vemuri, B.: Computing diffeomorphic paths with application to cardiac motion analysis. In: 4th MICCAI Workshop on Mathematical Foundations of Computational Anatomy. pp. 83–94 (2013)
21. Song, G., Tustison, N., Avants, B., Gee, J.C.: Lung ct image registration using diffeomorphic transformation models. Medical image analysis for the clinic: a grand challenge pp. 23–32 (2010)
22. Suh, Y., Murray, W., Keall, P.J.: Imrt treatment planning on 4d geometries for the era of dynamic mlc tracking. Technology in cancer research & treatment 13(6), 505–515 (2014)
23. Younes, L., Arrate, F., Miller, M.I.: Evolutions equations in computational anatomy. NeuroImage 45(1, Supplement 1), S40 – S50 (2009), `http://www.sciencedirect.com/science/article/pii/S105381190801166X`, mathematics in Brain Imaging

Stochastic EPDiff Landmark Dynamics

Darryl D. Holm and Tomasz M. Tyranowski

Department of Mathematics
Imperial College London
SW7 2AZ London, UK

Abstract. We develop a variational method of deriving stochastic partial differential equations whose solutions follow the flow of a stochastic vector field. As an example in one spatial dimension we numerically simulate singular solutions (landmarks) of the stochastically perturbed EPDiff equation derived using this method. These numerical simulations show that singular solutions of the stochastically perturbed EPDiff equation persist, and some choices of stochastic perturbations allow landmarks to interpenetrate and exchange order on the real line in overtaking collisions, although this behaviour does not occur for singular solutions of the unperturbed deterministic EPDiff equation. This solution behaviour introduces the possibility of a topological change and may be of importance in registration of noisy images in computational anatomy.

Keywords: Geometric mechanics, cylindrical stochastic processes, stochastic soliton dynamics, symmetry reduced variational principles

1 Introduction

Trouvé and Vialard [14, 15] study the stochastic evolution of landmarks in LDDMM [12] as a stochastic perturbation of the canonical Hamiltonian system arising from the singular reduction to a finite dimensional system of Lagrangian particles of a solution of the EPDiff equation for the geodesics on the group of diffeomorphisms, which arises from the LDDMM variational principle [4]. From this viewpoint, the variational principle for shape analysis using LDDMM has a natural analogue in particle dynamics. In particular, papers [14, 15] suggest adding white noise to the Hamiltonian evolution equation for the landmark canonical "momentum", as though the noise were a random force acting on a system of particles. However, there exist many ways of introducing stochasticity into particle dynamics. Here, we will explore an alternative approach for including noise in Hamilton equations which is still consistent with the LDDMM variational principle for landmark evolution. For brevity, and to simplify matters, we will take the viewpoint of particle dynamics and defer its potential applications in landmark dynamics for LDDMM until later work.

Our approach is based on a generalisation in [7] of earlier work by Bismut [1], Lázaro-Camí and Ortega [11], and Bou-Rabee and Owhadi [2] for stochastic ordinary differential equations (SDE). The *parametric stochastic deformation*

(P-SD) approach of [7] unifies the Hamiltonian and Lagrangian approaches to temporal stochastic dynamics, and extends them to stochastic partial differential equations (SPDE) in the case of cylindrical noise in which the spatial dependence is *parametric*, while temporal dependence is stochastic.

Objectives. This paper has two main objectives. The first objective is the inclusion of parametric stochastic deformation (P-SD) in the variational principle for the EPDiff partial differential equation. The second objective is the numerical study of the statistical effects of parametric and canonically Hamiltonian stochastic deformations (CH-SD) on the soliton-like solutions of deterministic EPDiff in one spatial dimension, when the Lagrangian in Hamilton's principle is a Sobolev norm on the continuous vector fields. When the Lagrangian is the H^1 norm, the deterministic equation is the completely integrable CH equation [3] and the solutions are true solitons (peakons).

2 Stochastic variational perturbations in one spatial dimension

2.1 Singular peakon solutions of the EPDiff equations

Let $\mathrm{Diff}(\mathbb{R}^n)$ be the diffeomorphism group of $\mathbb{R}^n$, and $\mathfrak{X}(\mathbb{R}^n)$ its Lie algebra, i.e., the set of all smooth vector fields on $\mathbb{R}^n$. The EPDiff equation is obtained from the variational principle $\delta S = 0$ for the action functional $S = \int \ell(u)\,dt$ with the restricted variations $\delta u = \dot{v} - [u, v]$ (see [8]). The EPDiff(H^1) equation in the one-dimensional case when $\ell(u) = \frac{1}{2}\|u\|^2_{H^1} = \frac{1}{2}\int u^2 + \alpha^2 u_x^2\,\mathrm{d}x$ is called the Camassa-Holm (CH) equation for $m = \delta\ell/\delta u = u - \alpha^2 u_{xx}$ with positive constant α^2; namely [3],

$$m_t + (um)_x + mu_x = 0 \quad \text{with} \quad m = u - \alpha^2 u_{xx}\,. \tag{1}$$

This equation has singular peaked soliton (peakon) solutions, given by

$$m(x,t) = \sum_{a=1}^{N} p_a(t)\delta(x - q_a(t))\,, \quad \text{so that} \quad u(x,t) := \sum_{b=1}^{N} p_b(t)K(x - q_b(t)), \tag{2}$$

where $K(x-y) = \exp(-|x-y|/\alpha)$ is the Green's function for the Helmholtz operator $1-\alpha^2\partial_x^2$. The peaked shape of the velocity profile of the soliton solution of the CH equation $u(x,t) := p(t)\exp(-|x-q(t)|/\alpha)$ provided the name, peakon.

Peakons are emergent singular solutions which dominate the initial value problem, since an initially confined smooth velocity distribution will decompose into peakon solutions and, in fact, *only* peakon solutions. The main point to notice is that the distance between any two peaks never passes through zero. That is, the peakons keep their order, even after any number of overtaking collisions (the taller peakons travel faster). Substituting the (weak) solution Ansatz (2)

into the CH equation (1) and integrating against a smooth test function yields the following dynamical equations for the $2N$ solution parameters $q_a(t)$ and $p_a(t)$

$$\frac{dq_a}{dt} = u(q_a(t),t) \quad \text{and} \quad \frac{dp_a}{dt} = -\,p_a(t)\frac{\partial u(q_a(t),t)}{\partial q_a}\,. \tag{3}$$

The system of equations for the peakon parameters comprises a completely integrable canonical Hamiltonian system, whose solutions determine the positions $q_a(t)$ and amplitudes $p_a(t)$, for all N solitons, $a = 1, \ldots, N$, and also describe the dynamics of their multi soliton interactions.

2.2 Singular momentum map version of the Stratonovich stochastic EPDiff equations

The objective of the remainder of the paper is to introduce stochasticity into the EPDiff equation and study its effects on the interactions of the peakon solutions of the CH equation. We consider the canonical Hamiltonian stochastic deformation (CH-SD), and also its special case, the parametric stochastic deformation (P-SD). To achieve our objective, we propose an action functional which contains a Stratonovich stochastic term, and treats q as an advected quantity, where the advection condition (the first equation in (3)) is enforced as a constraint with the help of the Lagrange multiplier p, and then prove the following theorem.

Theorem 1 (Canonical Hamiltonian Stochastic Deformation (CH-SD) of EPDiff).

The action $S(u,p,q)$ for the stochastic variational principle $\delta S = 0$ given by

$$S(u,p,q) = \underbrace{\int \left(\ell(u) + \sum_a \left\langle p_a\,,\, \frac{dq_a}{dt} - u(q_a,t) \right\rangle \right) dt}_{\text{Lebesgue integral}} - \underbrace{\int \sum_i h_i(q,p) \circ dW_i(t)\,,}_{\text{Stratonovich integral}} \tag{4}$$

leads to the following Stratonovich form of the *stochastic* EPDiff equation

$$\begin{aligned} dm &= -\pounds_u m\, dt + \sum_i \big\{ m\,,\, h_i(q,p) \big\} \circ dW_i(t)\,, \\ dq_a &= u(q_a,t)\, dt + \sum_i \big\{ q_a\,,\, h_i(q,p) \big\} \circ dW_i(t)\,, \\ dp_a &= -p_a(t)\frac{\partial u}{\partial x}(q_a,t)\, dt + \sum_i \big\{ p_a\,,\, h_i(q,p) \big\} \circ dW_i(t)\,, \end{aligned} \tag{5}$$

where the momentum density m and velocity u are given by

$$m(x,t) := \frac{\delta \ell}{\delta u} = \sum_{a=1}^{N} p_a \delta(x - q_a(t))\,, \quad \text{and} \quad u(x,t) := \sum_{b=1}^{N} p_b K(x - q_b(t)). \tag{6}$$

Proof. Take the variations of the action integral (4), to find

$$\begin{aligned}
\delta u: &\quad \frac{\delta \ell}{\delta u} - \sum_{a=1}^{N} p_a \delta(x - q_a(t)) = 0\,,\\
\delta p: &\quad dq_a - u(q_a,t)\,dt - \sum_i \frac{\partial h_i}{\partial p_a}(q,p) \circ dW_i(t) = 0\,,\\
\delta q: &\quad -dp_a - p_a(t)\frac{\partial u}{\partial x}(q_a,t)\,dt - \sum_i \frac{\partial h_i}{\partial q_a}(q,p) \circ dW_i(t) = 0\,,
\end{aligned} \tag{7}$$

after integrations by parts with vanishing endpoint and boundary conditions. The first variational equation captures the relation (6), and latter two equations in (7) produce the corresponding equations in (5). Substituting the latter two equations in (7) into the time derivative of the first one yields the first equation in (5).

The particular choice of the functions $h_i(q,p) = \sum_{a=1}^{N} p_a \xi_i(q_a)$ lead to the parameterised stochastic deformation (P-SD) of the peakon solutions. We summarise this observation in the following Corollary.

Corollary 1. [P-SD is a special case of CH-SD for EPDiff] Given the set of diffusivities $\xi_i(x)$, $i = 1, \ldots, M$, let $h_i(q,p) = \sum_{a=1}^{N} p_a \xi_i(q_a)$. Then the momentum density $m(x,t)$ satisfies the equation

$$dm + \pounds_{dx_t} m = 0\,, \tag{8}$$

where the stochastic vector field $dx_t(x)$ is given by the P-SD formula,

$$dx_t(x) = u(x,t)\,dt + \sum_i \xi_i(x) \circ dW_i(t)\,. \tag{9}$$

Proof. Specialise to $h_i(q,p) = \sum_{a=1}^{N} p_a \xi_i(q_a)$ in the first line of equation (5) in Theorem 1.

Remark 1 (Outlook: Comparing results for P-SD and CH-SD). In Section 3 and Section 5 we will investigate the effects of choosing between two slightly different stochastic potentials on the interaction of two peakons, $N = 2$, corresponding to P-SD and CH-SD. The two options are $h_i^{(1)}(q,p) = \sum_{a=1}^{N} p_a \xi_i(q_a)$ and $h_i^{(2)}(q,p) = \sum_{a=1}^{N} p_a \varphi_{ia}(q)$, respectively. These are both linear in the peakon momenta and in the simplest case they have constant coefficients. Although these two choices are very similar, they will produce quite different solution behaviour in our numerical simulations of peakon-peakon overtaking collisions in Section 5.

Remark 2 (Stratonovich stochastic EPDiff equations in one dimension).

1. In one spatial dimension, equation (8) becomes

$$\big(\partial_t m + u m_x + 2 m u_x\big)dt + m_x \sum_i \xi_i(x) \circ dW_i(t) + 2m \sum_i \xi_i'(x) \circ dW_i(t) = 0\,. \tag{10}$$

Importantly, the multiplicative noise multiplies both the solution and its gradient. The latter is not a common form for stochastic PDEs. In addition, both the spatial correlations $\xi_i(x)$ and their derivatives $\xi_i'(x)$ are involved.

2. The equations for dq_a and dp_a in (5) are stochastic canonical Hamiltonian equations (SCHEs) in the sense of Bismut [1, 11]. These equations for dq_a and dp_a may be rewritten as

$$\begin{aligned} dq_a &= \frac{\partial H}{\partial p_a}(q,p)\,dt + \sum_i \frac{\partial h_i}{\partial p_a}(q,p) \circ dW_i(t)\,, \\ dp_a &= -\frac{\partial H}{\partial q_a}(q,p)\,dt - \sum_i \frac{\partial h_i}{\partial q_a}(q,p) \circ dW_i(t)\,, \end{aligned} \tag{11}$$

with the deterministic Hamiltonian

$$H(q,p) = \frac{1}{2}\sum_{a,b} p_a p_b K(q_a - q_b)\,. \tag{12}$$

The stochastic canonical Hamilton equations in (11) can also be obtained by extremising the *phase-space action functional*

$$S\big[q(t),p(t)\big] = \int_0^T \Big(\sum_{a=1}^N p_a \dot{q}_a - H(q,p)\Big)\,dt - \int_0^T \sum_{i=1}^M h_i(q,p) \circ dW_i(t)\,. \tag{13}$$

This is the restriction of (4) to the submanifold defined by the Ansatz (6).

3 The Fokker-Planck equation

The stochastic process in (11) for $(q(t),p(t))$ can be described with the help of a transition density function $\rho(t,q,p;\bar{q},\bar{p})$ which represents the probability density that the process, initially in the state $(\bar{q},\bar{p})$, will reach the state (q,p) at time t. The transition density function satisfies the Fokker-Planck equation corresponding to (11) (see [6], [9]). Let us examine the form of this equation in the case of $h_i(q,p) = \sum_{a=1}^N p_a \beta_{ia}$, where $\beta_{ia} = \text{const}$. In that case the noise in (11) is additive, and the Stratonovich and Itô calculus yield the same equations of motion.

3.1 Single-pulson dynamics

Consider a single pulson ($N = 1$) subject to one-dimensional (i.e., $M = 1$) Wiener process, with the stochastic potential $h(q,p) = \beta p$, where β is a nonnegative real parameter. The stochastic Hamiltonian equations (11) take the form $dq = p\,dt + \beta \circ dW(t)$, $dp = 0$. The corresponding Fokker-Planck equation takes the form

$$\frac{\partial \rho}{\partial t} + p\frac{\partial \rho}{\partial q} - \frac{1}{2}\beta^2 \frac{\partial^2 \rho}{\partial q^2} = 0 \tag{14}$$

with the initial condition $\rho(0,q,p;\bar{q},\bar{p}) = \delta(q-\bar{q})\delta(p-\bar{p})$. This advection-diffusion equation is easily solved with the help of the fundamental solution for the heat equation, and the solution yields

$$\rho_\beta(t,q,p;\bar{q},\bar{p}) = \frac{1}{\beta\sqrt{2\pi t}} e^{-\frac{(q-\bar{q}-pt)^2}{2\beta^2 t}} \delta(p-\bar{p}). \tag{15}$$

This solution means that the initial momentum $\bar{p}$ is preserved. The position has a Gaussian distribution which widens with time, and whose maximum is advected with velocity $\bar{p}$.

3.2 Two-pulson dynamics

The dynamics of two interacting pulsons has been thoroughly studied and possesses interesting features (see [5], [8]). It is therefore intriguing to see how this dynamics is affected by the presence of noise. Consider $N = 2$ pulsons subject to a two-dimensional (i.e., $M = 2$) Wiener process, with the stochastic potentials $h_1(q,p) = \beta_1 p_1$ and $h_2(q,p) = \beta_2 p_2$, where $q = (q_1, q_2)$, $p = (p_1, p_2)$. The corresponding Fokker-Planck equation takes the form

$$\frac{\partial\rho}{\partial t} + \frac{\partial}{\partial q_1}\big[a_1\rho\big] + \frac{\partial}{\partial q_2}\big[a_2\rho\big] + \frac{\partial}{\partial p_1}\big[a_3\rho\big] + \frac{\partial}{\partial p_2}\big[a_4\rho\big] - \frac{1}{2}\beta_1^2\frac{\partial^2\rho}{\partial q_1^2} - \frac{1}{2}\beta_2^2\frac{\partial^2\rho}{\partial q_2^2} = 0 \tag{16}$$

with the initial condition $\rho\big(0,q,p;\bar{q},\bar{p}\big) = \delta(q_1-\bar{q}_1)\delta(p_1-\bar{p}_1)+\delta(q_2-\bar{q}_2)\delta(p_2-\bar{p}_2)$, where

$$\begin{aligned} a_1(q,p) &= p_1 + p_2 K(q_1-q_2), & a_3(q,p) &= -p_1 p_2 K'(q_1-q_2), \\ a_2(q,p) &= p_2 + p_1 K(q_1-q_2), & a_4(q,p) &= p_1 p_2 K'(q_1-q_2). \end{aligned} \tag{17}$$

Despite its relatively simple structure, it does not appear to be possible to solve this equation analytically. It is nevertheless an elementary exercise to verify that the function

$$\rho(t,q_1,q_2,p_1,p_2;\bar{q}_1,\bar{q}_2,\bar{p}_1,\bar{p}_2) = \rho_{\beta_1}(t,q_1,p_1;\bar{q}_1,\bar{p}_1) + \rho_{\beta_2}(t,q_2,p_2;\bar{q}_2,\bar{p}_2), \tag{18}$$

where ρ_{β_i} is given by (15), satisfies (16) asymptotically as $q_1 - q_2 \longrightarrow \pm\infty$, assuming the Green's function and its derivative decay in that limit. This simple observation gives us an intuition that stochastic pulsons should behave like individual particles when they are far from each other, just like in the deterministic case. In order to study the stochastic dynamics of the collision of pulsons, we need to resort to Monte Carlo simulations.

In Section 4 we discuss our numerical algorithm, and in Section 5 we present the results of our numerical studies.

4 Stochastic variational integrator

Given the variational structure of the problem we have formulated in Theorem 1, it is natural to employ variational integrators for numerical simulations. For an extensive review of variational integrators we refer the reader to Marsden & West [13] and the references therein. Stochastic variational integrators were first introduced in Bou-Rabee & Owhadi [2]. These integrators were derived for Lagrangian systems using the Hamilton-Pontryagin variational principle. In our case, however, we find it more convenient to stay on the Hamiltonian side and use the discrete variational Hamiltonian mechanics introduced in Lall & West [10]. We combine the ideas of [2] and [10], and propose the following discretization of the phase-space action functional (13):

$$\begin{aligned} S_d = &\sum_{k=0}^{K-1} \Bigg(\sum_{i=1}^{N} p_i^k \frac{q_i^{k+1} - q_i^k}{\Delta t} - H(q^{k+1}, p^k) \Bigg) \Delta t \\ &- \sum_{k=0}^{K-1} \sum_{m=1}^{M} \frac{h_m(q^k, p^k) + h_m(q^{k+1}, p^{k+1})}{2} \Delta W_k^m, \end{aligned} \tag{19}$$

where $\Delta t = T/K$ is the time step, (q^k, p^k) denote the position and momentum at time $t_k = k\Delta t$, and $\Delta W_k^m \sim N(0, \Delta t)$ are independent normally distributed random variables for $m = 1, \dots, M$ and $k = 0, \dots, K-1$. Extremizing (19) with respect to q^k and p^k yields the following implicit stochastic variational integrator:

$$\begin{aligned} \frac{q_i^{k+1} - q_i^k}{\Delta t} &= \frac{\partial H}{\partial p_i}\big(q^{k+1}, p^k\big) + \sum_{m=1}^{M} \frac{\partial h_m}{\partial p_i}\big(q^k, p^k\big) \frac{\Delta W_{k-1}^m + \Delta W_k^m}{2\Delta t}, \\ \frac{p_i^{k+1} - p_i^k}{\Delta t} &= -\frac{\partial H}{\partial q_i}\big(q^{k+1}, p^k\big) - \sum_{m=1}^{M} \frac{\partial h_m}{\partial q_i}\big(q^{k+1}, p^{k+1}\big) \frac{\Delta W_k^m + \Delta W_{k+1}^m}{2\Delta t}, \end{aligned} \tag{20}$$

for $i = 1, \dots, N$. Knowing (q^k, p^k) at time t_k, the system above allows to solve for the position q^{k+1} and momentum p^{k+1} at the next time step. For increased computational efficiency, it is advisable to solve the first (nonlinear) equation for q^{k+1} first, and then the second equation for p^{k+1}. Assuming the stochastic potentials are of the form $h_i(q,p) = \sum_{a=1}^{N} p_a \varphi_{ia}(q)$, the second equation is a linear system for p^{k+1}, and in case $\varphi_{ia} = \text{const}$, it becomes an explicit update rule.

The integrator (20) is symplectic, and preserves momentum maps corresponding to (discrete) symmetries of the discrete Hamiltonian—for instance, if $H(q,p)$ and all $h_i(q,p)$ are translationally invariant, as in our simulations in Section 5, then the total momentum $\sum_{i=1}^{N} p_i$ is numerically preserved. The proof of these facts trivially follows from [2], keeping in mind that the momenta p_i and velocities $\dot{q}_i$ are related via the Legendre transform. By a straightforward application of the Stratonovich-Taylor expansion (see [9]), one can show that the integrator (20) has strong order of convergence 0.5, and weak order of convergence 1.

5 Numerical experiments

We performed numerical simulations of the rear-end collision of two pulsons for two different Green's functions, namely $K(q_1-q_2) = e^{-(q_1-q_2)^2}$ and $K(q_1-q_2) = e^{-2|q_1-q_2|}$. In the latter case, the corresponding pulsons are commonly called 'peakons'. We investigated the initial conditions $\bar{q}_1 = 0$, $\bar{q}_2 = 10$, $\bar{p}_2 = 1$ together with the following four initial values: $\bar{p}_1 = 8$, $\bar{p}_1 = 4$, $\bar{p}_1 = 2$, $\bar{p}_1 = 1$.

That is, we varied the initial momentum of the faster pulson. We perturbed the slower pulson by introducing a one-dimensional Wiener process with the stochastic potential $h(q,p) = \beta p_2$ (this corresponds to $\beta_1 = 0$, $\beta_2 = \beta$ in Section 3.2). The pulsons were initially well-separated, so their initial evolution was described by (18). The parameter β was varied in the range $[0, 6.5]$. We used the time step $\Delta t = 0.02$, and for each choice of the parameters 50000 sample solutions were computed until the time $T = 100$.

5.1 Sample paths and mean solutions

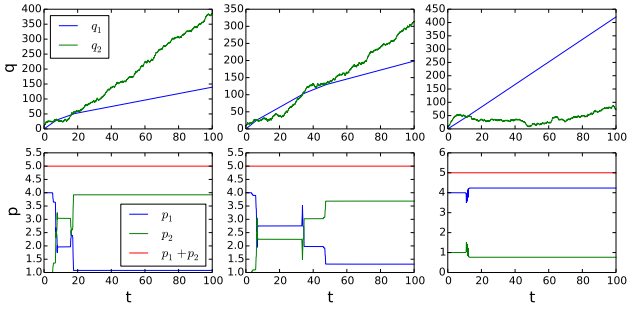


Fig. 1. Example numerical sample paths for Gaussian pulsons for the simulations with $\bar{p}_1 = 4$ and $\beta = 4$. The positions are depicted in the plots in the upper row, and the corresponding momenta are shown in the plots in the lower row.

Figure 1 shows a few sample paths from the simulations of the interaction of Gaussian pulsons for the case with $\bar{p}_1 = 4$ and $\beta = 4$. The simulations for $\bar{p}_1 = 8$ and $\bar{p}_1 = 2$, as well as the simulations for peakons, gave qualitatively similar results. The most striking feature is that the faster pulson/peakon may in fact cross the slower one. In the deterministic case one can show that the faster pulson can never pass the slower one—they just exchange their momenta. The proof relies on the fact that both the Hamiltonian and total momentum are preserved (see [5], [8]). In our case, however, the Hamiltonian (12) is not preserved due to the presence of the time-dependent noise, which allows much richer dynamics of the interactions. This may find interesting applications in landmark matching—see the discussion in Section 6.

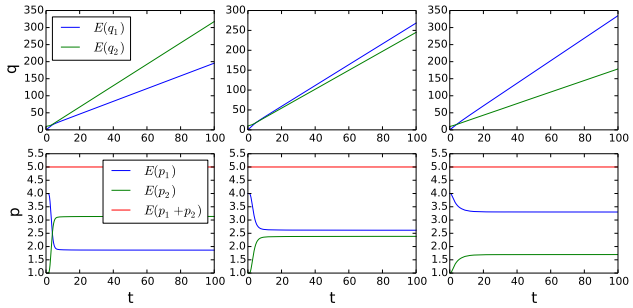


Fig. 2. Numerical mean paths for Gaussian pulsons for the simulations with $\bar{p}_1 = 4$. Results for three example choices of the parameter β are presented: $\beta = 1.5$ (*left*), $\beta = 2.5$ (*middle*), and $\beta = 4.5$ (*right*). The positions are depicted in the plots in the upper row, and the corresponding momenta are shown in the plots in the lower row.

Looking at Figure 1 we also note that our variational integrator exactly preserves the total momentum, as expected. Figure 2 depicts the mean solution for Gaussian pulsons with the initial condition $\bar{p}_1 = 4$ for different values of the noise intensity β. We see that for small noise the mean solution resembles the deterministic one, but as the parameter β is increased, the mean solution represents two pulsons passing through each other with increasingly less interaction. We study the probability of crossing in more detail in Section 5.2.

We observed that pulsons may cross even when they have the same initial momentum ($\bar{p}_1 = 1$). In the deterministic case they would just propagate in the same direction, retaining their relative distance.

5.2 Probability of crossing

We studied in more detail the distance between the pulsons $\Delta q(t) = q_2(t) - q_1(t)$ at the end of the simulation, that is, at time $t = 100$. The probability of crossing as a function of the noise intensity β is depicted in Figure 3. We see that this probability seems to approach unity for the simulations with $\bar{p}_1 > 1$, and 0.5 for $\bar{p}_1 = 1$.

5.3 Noise screening

In the numerical experiments described above we observed that the presence of noise causes pulsons to cross with a non-zero probability. The functions $q_1(t)$, $p_1(t)$, $q_2(t)$ and $p_2(t)$ define a transformation of the real line through (6). In the deterministic case this transformation is a diffeomorphism, but not when noise is added, since the crossing of pulsons introduces topological changes in the image of the real line under this transformation. This may be of interest in image matching, as in [14], when one would like to construct a deformation between two images which are not exactly diffeomorphic. However, with that application

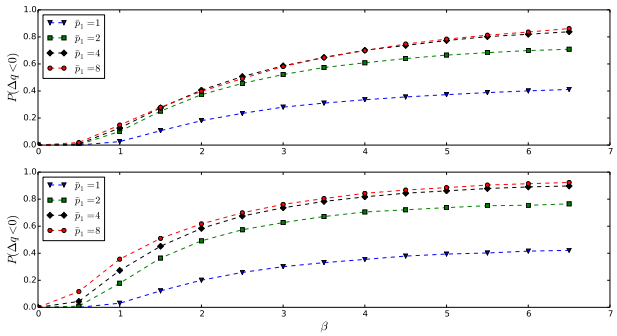


Fig. 3. The probability of crossing, that is, the probability that $q_2(t) < q_1(t)$ at time $t = 100$, as a function of the parameter β for Gaussian pulsons (*top*) and peakons (*bottom*).

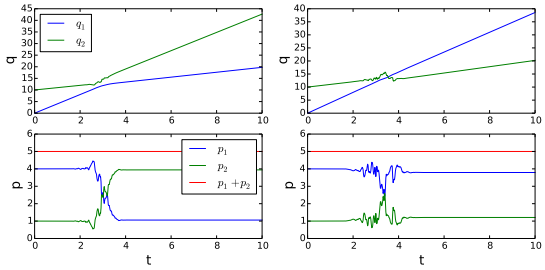


Fig. 4. Example numerical sample paths for Gaussian pulsons for the simulations with the initial conditions $\bar{q}_1 = 0$, $\bar{p}_1 = 4$, $\bar{q}_2 = 10$, and $\bar{p}_2 = 1$, and the stochastic potential $h(q,p) = \beta p_2 \exp(-(q_2 - q_1)^2)/\gamma$, with the parameters $\beta = 4$ and $\gamma = 4$. The positions are depicted in the plots in the upper row, and the corresponding momenta are shown in the plots in the lower row.

in mind, one may want to restrict the stochastic effects only to the situation when two pulsons get close to each other. This can be obtained by applying the stochastic potential $h(q,p) = \beta p_2 \exp(-(q_2 - q_1)^2)/\gamma$. The parameter β adjusts the noise intensity, just as before, while the parameter γ controls the range over which the stochastic effects are non-negligible. We performed a few simulations with this stochastic potential. A few sample paths are depicted in Figure 4. Note that this stochastic potential is translation-invariant, so the total momentum is preserved. As expected, our variational integrator preserves the total momentum exactly.

5.4 Restriction to parametric noise and additive noise in the momentum equation

Interestingly, crossing of pulsons does not seem to occur for the case of parametric stochastic deformation with the restriction $\varphi_{ai}(q) = \xi_i(q_a)$ as in Lemma 1. We ran numerical experiments for the potential $h(q,p) = \beta(p_1 + p_2)$, which has the form as in Lemma 1 with $\xi(x) = \beta$, but observed no interpenetration. We also did not observe crossing when the stochastic potential is independent of p. For instance, we performed simulations with the potential $h(q,p) = \beta q_2$. Such a potential results in additive noise in the momentum equation in (11) only, as in [14]. In many cases the pulsons would asymptotically approach each other, but never pass. We observed similar behavior for the (translationally invariant) potential $h(q,p) = \beta \exp(-(q_1 - q_2)^2/\gamma)$.

6 Summary and prospects

We have seen in Section 2 that the finite-dimensional peakon solutions for the EPDiff partial differential equation in one spatial dimension persist under both parametric stochastic deformation (P-SD) and canonical Hamiltonian stochastic deformations (CH-SD) of the EPDiff variational principle. We took advantage of the flexibility of CH-SD to study stochastic peakon-peakon collisions in which noise was introduced into *only one* of the peakon position equations (rather than symmetrically into both of the canonical position equations, as occurs with P-SD), while at the same time not introducing any noise into either of the corresponding canonical momentum equations. Our numerical experiments in Section 5 revealed that this type of noise allows the soliton-like singular peakon and pulson solutions of EPDiff to interpenetrate and change order on the real line, although this is not possible for the diffeomorphic flow represented by the solutions of the unperturbed deterministic EPDiff equation. This crossing of peakon paths was observed and its statistics were studied in detail. In contrast, crossing of peakon paths was *not* observed for the corresponding P-SD simulations in which the noise enters symmetrically in both position equations. Crossing of peakon paths was also not observed when stochasticity was added only in the canonical momentum equations, as studied in Trouvé and Vialard [14], where the authors considered the equations

$$dq_a = u(q_a,t)\,dt \quad \text{and} \quad dp_a = -p_a(t)\frac{\partial u}{\partial x}(q_a,t)\,dt + \sigma dW(t) \tag{21}$$

for stochastic landmark matching in computational anatomy. This perturbation corresponds to (11) with $h_1(q,p) = \sigma \sum_a q_a$, and enforces a stochastic Brownian force on the particles, rather than making particle paths stochastic. Trouvé and Vialard showed that this simple additive noise in the momentum equation is in general enough to account for correlations between points on the curve during landmark evolution under stochastic forcing. Our results in Section 5 demonstrated that noise in the position equations may additionally allow the

landmarks to change their order on the line, thus allowing matching of two images which are not diffeomorphic.

6.1 Acknowledgements

We are very grateful for the encouragement of the many people who took the time to discuss these matters with us, or comment on drafts, especially our friends and colleagues N. Bou-Rabee, A. Castro, C. J. Cotter, D. Crisan, H. Owhadi and J. P. Ortega. This work was partially supported by the European Research Council Advanced Grant 267382 FCCA.

References

1. J. M. Bismut [1981] *Mécanique aléatoire*, Berlin: Springer.
2. N. Bou-Rabee and H. Owhadi [2009] Stochastic variational integrators, *IMA Journal of Numerical Analysis* 29: 421–443.
3. R. Camassa and D. D. Holm [1993] An integrable shallow water equation with peaked solitons, *Phys. Rev. Lett.* **71** 1661–1664. `http://xxx.lanl.gov/abs/patt-sol/9305002`
4. CJ Cotter and DD Holm [2006] Singular solutions, momentum maps and computational anatomy. Preprint at `http://arxiv.org/abs/nlin.SI/0605020`. Paper at `http://hal.inria.fr/docs/00/63/58/75/PDF/Cotter_MFCA06.pdf`
5. O. B. Fringer and D. D. Holm [2001], Integrable vs. nonintegrable geodesic soliton behavior, *Physica D: Nonlinear Phenomena*, 150(3-4):237-263.
6. C. Gardiner [2009], *Stochastic Methods: A Handbook for the Natural and Social Sciences*, Springer Series in Synergetics, Springer Berlin Heidelberg
7. D. D. Holm [2015] Variational principles for stochastic fluid dynamics, *Proc Roy Soc A*, online at `http://rspa.royalsocietypublishing.org/content/471/2176/20140963`, arXiv e-print available at `http://arxiv.org/pdf/1410.8311.pdf`
8. D. D. Holm, T. Schmah and C. Stoica [2009], *Geometric Mechanics and Symmetry: From Finite to Infinite Dimensions*, Oxford Texts in Applied and Engineering Mathematics, Oxford University Press.
9. P. E. Kloeden and E. Platen [1995], *Numerical Solution of Stochastic Differential Equations*, Applications of Mathematics: Stochastic Modelling and Applied Probability, Springer.
10. S. Lall and M. West [2006], Discrete variational Hamiltonian mechanics, *Journal of Physics A: Mathematical and General*, 39(19):5509-5519.
11. J. A. Lázaro-Camí and J. P. Ortega [2008] Stochastic Hamiltonian dynamical systems, *Rep. Math. Phys.*, 61 (1): 65–122.
12. M. I. Miller, A. Trouvé, and L. Younes [2003] On the metrics, Euler equations and normal geodesic image motions of computational anatomy. In Proceedings of the 2003 International Conference on Image Processing, volume 2, pages 635–638. IEEE.
13. J. E. Marsden and M. West [2001], Discrete mechanics and variational integrators, *Acta Numerica*, 10(1):357-514.
14. A. Trouvé and F. X. Vialard [2012] Shape splines and stochastic shape evolutions: a second order point of view, *Quart. of Appl. Math.* 70: 219–251.
15. F.-X. Vialard [2013] Extension to infinite dimensions of a stochastic second-order model associated with the shape splines, *Stoch. Proc. App.* 123 (6) 2110–2157.

Reconstructing Karcher Means of Shapes on a Riemannian Manifold of Metrics and Curvatures

Boris A. Gutman[1], P. Thomas Fletcher[2], Greg Fleishman[3], Paul M. Thompson[1]

1. Imaging Genetics Center, INI, KSOM, University of Southern California
2. School of Computing, University of Utah
3. Department of Bioengineering, University of California Los Angeles

Abstract. In a recent paper [1], the authors suggest a novel Riemannian framework for comparing shapes. In this framework, a simple closed surface is represented by a field of metric tensors and curvatures. A product Riemannian metric is developed based on the L^2 norm on symmetric positive definite matrices and scalar fields. Taken as a quotient space under the group of volume-preserving diffeomorphisms, the space becomes a proper metric manifold of shapes. In this work, we simplify this representation, showing that only mean curvature and metric tensor fields are needed for a complete surface representation. In this simplified framework, we develop an algorithm for computing Karcher means, and compare the results to standard Euclidean averages of surface embeddings.

Keywords: Shape Analysis, Riemannian Metric, Surface Registration, Cortical Surface, Karcher Mean

1 Introduction

Comparison of simple 3D shapes remains one of the staples of medical image analysis. As in any population analysis, statistical comparison of a group of shapes typically requires a group template, the average shape. However, computing means of shapes requires a metric which respects the invariance to Euclidean motion that is inherent in proper shape analysis. In the absence of such a metric, the mean shape is often approximated as a Euclidean average of coordinates after registration and affine alignment. Many non-linear registration tools for shapes have been developed, of which we name a few below.

Gu et al., developed a conformal mapping algorithm [2] for spherical mapping and formulated a landmark-matching energy as a Mobius transform. A relaxation of the conformal energy, the quasi-conformal mapping of Zeng et al. [3] simultaneously solves the Beltrami equations and minimizes curvature mismatch. Shi et al. [4] applies fluid registration to the flat 2D domain after conformally mapping a surface with prescribed boundaries. Spherical Demons [5] adapts the diffeomorphic demons algorithm [6] to the sphere, matching curvature-derived intensity functions to register cortical surfaces. A similar approach is taken in [7], adapting fluid registration [8] to the sphere. In [9], the authors compute high-dimensional embeddings of surfaces based on eigenfunctions of the Laplace-Beltrami Operator (LBO). In this approach,

metric tensors are scaled to match the resulting LBO representations in the Euclidean sense.

A number of manifolds of shape representations are possible, see for example [10]. In general, these fall into one of two categories: metrics on spatial diffeomorphisms to be applied to a known surface embedding, and metrics on representations from which the surface can be reconstructed directly. In the first category, the authors in [11] apply the large deformation framework to compute the length of the path in the space of diffeomorphism resulting from morphing one boundary onto another. An improvement on this is suggested in [12], measuring distances on the deformation of the surface itself rather than in the ambient space as done in [11]. Using [11], the authors in [13] develop an EM approach to estimate the shape mean based on the initial momentum describing the set of deformations.

In the second category, Kurtek et al. [14] developed a Riemannian framework for simple surfaces, using q-maps. The L^2 distance on q-maps, or simply the surface embedding weighted by the root of the volume form, remains invariant under reparameterizations. Q-maps can be used to directly reconstruct the surface, a significant advantage over previous methods. Computing averages of a group of shapes reduces to estimating the mean q-map under spherical diffeomorphisms. However, the representation is still of the surface *embedding*, with all the resulting nuisances. Some standard heuristics are applied to the initial surfaces, namely centering each shape at the origin. This implies that a local change in the surface has a global effect on the representation. Further, the metric is on the space $\mathcal{S}$ of smooth functions from the 2-sphere to $\mathbb{R}^n$, which ignores the surface metric structure.

Finally, in [1] the authors proposed a metric on a surface representation which is completely independent of the surface embedding. Applying the Ebin metric to a field of pullback metric tensors on the sphere, and the L^2 metric to mean and Gaussian curvatures, the authors develop a Riemannian space of shapes. The representation can then be used to reconstruct the surface purely from intrinsic surface properties, with no need to normalize for Euclidean motion. Further, the mapping between surfaces resulting from removing the action of volumorphisms leads to an equiareal surface-to-surface mapping that is as-conformal-as-possible. Building on this framework, we make the following contributions: First, we show that the conformal equivalence between genus-zero shapes implies that the shapes can be uniquely represented with a field of spherical tensor metrics and mean curvatures, as shown in [15]. Thus, Gaussian curvature is no longer required. Second, we develop an algorithm for computing Karcher means of these representations from a population of shapes. Our modified reconstruction algorithm produces plausible reconstructed averages for subcortical and cortical surface models.

The remainder of the paper is organized as follows. Section 2 introduces the Riemannian metric on metric tensors, describing briefly its invariance properties. Section 3 describes the full metric plus curvature framework, showing that our representation is sufficient to reconstruct a surface. Section 4 shows how the framework can be used to compute intrinsic means of metric plus curvature maps. Section 5 gives some implementation aspects. Sections 6 and 7 present some experimental results and conclude the paper.

2 A Metric on Metrics

Given a set of surfaces with a mapping from the 2-sphere $\mathbb{S}^2$ to space, $\mathcal{S} = \{S: \mathbb{S}^2 \to \mathbb{R}^3 | S \in C^\infty\}$, we represent the metric structure of our shapes as the pull-back metric tensor g, $g_{i,j} = S_i^T S_j$. The field of these tensors lives in the space of positive definite tensors $\mathcal{M}(\mathbb{S}^2) = \{h: T\mathbb{S}^2 \times T\mathbb{S}^2 \to \mathbb{R} | h \in SPD(2)\}$. A metric on this space must be invariant to reparameterizations of a pair of tensor fields to be an intrinsic distance on metric structures. More formally, for a given metric on $\mathcal{M}$, the group of diffeomorphisms on $\mathbb{S}^2$ must act on $\mathcal{M}$ by isometry. Ebin et al. [16] showed that the L^2 Riemannian metric on the tangent bundle of $\mathcal{M}$, satisfies this criteria: given $g \in \mathcal{M}$, $h, k \in \Sigma \cong T_g\mathcal{M}$, the metric can be written as:

$$(h,k)_g = \int_M \langle h,k\rangle_g d\mu_g, \tag{1}$$

where $\langle h,k\rangle_g$ is the inner product induced by g, $\langle h,k\rangle_g = tr(g^{-1}hg^{-1}k)$, and μ_g is the volume form also induced by g. This metric produces geodesics on $\mathcal{M}$ whose length can be computed point-wise and in closed form. A reparameterization $\varphi \in \Phi = \{\phi: \mathbb{S}^2 \to \mathbb{S}^2 | \phi, \phi^{-1} \in C^2\}$ acts on g by conjugation with the pushforward (Jacobian) $D\varphi: T_x\mathbb{S}^2 \to T_{\varphi(x)}\mathbb{S}^2$, $\varphi \circ g = D\varphi^T g\, D\varphi$. Given two parameterized surfaces $A, B \in \mathcal{S}$, a closed-form solution for the geodesic distance between g_A and g_B at a point x is [17]

$$D(g_A[x], g_B[x]) = \sqrt{\int_0^1 \langle g_t'(x), g_t'(x)\rangle_{g_{t(x)}} dt} = \left\| \mathrm{Log}\left[g_A^{-1/2} g_B g_A^{-1/2}\right] \right\|_F. \tag{2}$$

This metric can be shown to be invariant under simultaneous spherical re-mappings of A and B [17], since $D(g_A, g_B) = D(D\varphi^T g_A D\varphi, D\varphi^T g_B D\varphi)$.

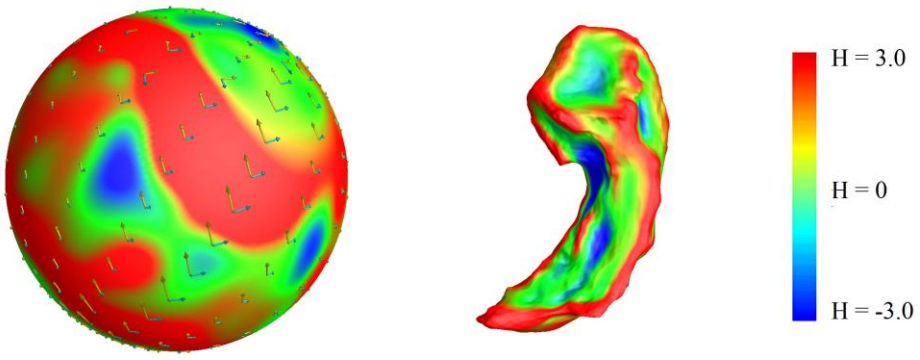


Fig. 1. Metric tensor fields and mean curvature – a complete surface representation. Tensors are displayed as their eigenvectors in $T\mathbb{S}^2$ with magnitude corresponding to the eigenvalues. The color map indicates mean curvature. The scale bar indicates mean curvature values.

3 Metrics on the Space of Surfaces

The change in the volume form due to reparameterization prevents a straightforward generalization of $(\cdot\ ,\ \cdot)_g$ to the quotient space $\mathcal{M}\backslash\Phi$. Instead, the authors in [1] consider the submanifold $\mathcal{M}_\mu$ of metrics which correspond to a fixed measure μ. $\mathcal{M}_\mu$ is a metric space under $(\cdot\ ,\ \cdot)_g$. The restriction of Φ to its subgroup of maps with a unitary pushforward $\Phi_U = \{\phi \in \Phi | det(D\phi) = 1\}$, leads to a quotient space $\mathcal{M}_\mu\backslash\Phi_U$ that is a metric space under the metric

$$\mathfrak{D}(A,B) = \min_{\varphi\in\Phi_U}\left(\int_{\mathbb{S}^2}\left\|\mathrm{Log}\left[g_A^{-1/2}\varphi\circ g_{\varphi\circ B}g_A^{-1/2}\right]\right\|_F^2 d\mathbb{S}^2\right)^{1/2}\quad A,B\in\mathcal{S}_1. \tag{3}$$

Here, $\mathcal{S}_1$ is the restriction of $\mathcal{S}$ to equiareal spherical parameterizations of shapes with normalized surface area.

The metric $\mathfrak{D}(A,B)$ allows us to compute intrinsic distances between metric structures. However, metric structure must be augmented with curvature information in order to represent a surface uniquely. The following theorem given by Gu, et al. [15], shows how this may be done in the case of conformal parameterization:

Theorem 1. *A closed conformally parameterized surface S in $\mathbb{R}^3$ is determined by its conformal factor λ and its mean curvature H uniquely up to Euclidean motion, where the metric tensor $g = \lambda I$.*

As our spherical parameterization is equiareal, we cannot expect them to be conformal except in the trivial case where $S = \mathbb{S}^2$. However, it is known that for all genus-zero surfaces there exists a conformal equivalence. Further, a conformal reparameterization can be found using only the metric structure, without knowing the surface [15]. Thus, we have the following result:

Theorem 2. *A closed parameterized surface $S \in \mathcal{S}$ is determined by its metric tensor g and its mean curvature H uniquely up to Euclidean motion.*

We note also that a generalization of this result is shown in [18]. An illustrative example of such a representation is shown in **Figure 1** above. From this result, it is clear that we only need augment the space $\{\mathcal{M}(\mathcal{S}_1)\backslash\Phi_U,\ \mathfrak{D}(\cdot,\cdot)\}$ with a corresponding metric on H. We now define the space of shapes as

$$\mathfrak{S} = \left\{\mathcal{M}(\mathcal{S}_1)\times C^2(\mathbb{S}^2)\backslash\Phi_U,\ \ \mathfrak{D}(\cdot,\cdot)\times D_{L^2\backslash\Phi_U}(\cdot,\cdot)\right\} \tag{4}$$

Here, $C^2(\mathbb{S}^2) = \{f\colon \mathbb{S}^2\to\mathbb{R} | f\in C^2\}$, and the usual L^2 distance modified by Φ_U, $D_{L^2\backslash\Phi_U}(a,b) = \min_{\phi\in\Phi_U}\sqrt{\int_{\mathbb{S}^2}(a-\phi\circ b)^2 d\mathbb{S}^2}$. For brevity, we will call the 2-product metric $\mathcal{L}(\cdot,\cdot) = \mathfrak{D}(\cdot,\cdot)\times D_{L^2\backslash\Phi_U}(\cdot,\cdot)$, defined explicitly as

$$\mathcal{L}(A,B) = \sqrt{\mathfrak{D}^2(A,B) + D^2_{L^2\backslash\Phi_U}(H_A,H_B)}. \tag{5}$$

4 Computing the Karcher Mean on $\mathfrak{S}$

Given a set of parametric surfaces representations $\mathfrak{s} = \{S_i | S_i \in \mathfrak{S}\}$, we would now like to find their intrinsic average. Note that by intrinsic we mean invariant to affine transformations, i.e. in our sense mean curvature is "intrinsic." Finding the Karcher mean $\nu(\mathfrak{s}) = \{g_{\mathcal{S}}, H_{\mathcal{S}}\}$ under $\mathcal{L}(\cdot,\cdot)$ requires simultaneous estimation of geodesic lengths to each shape's orbit in $\mathfrak{S}$. In other words, we must find several reparameterizations $\phi_i \in \Phi_U$, while estimating $\nu(\mathfrak{s})$:

$$\nu(\mathfrak{s}) = \underset{\mathcal{S}\in\mathfrak{S}}{\operatorname{argmin}}\left(\min_{\phi_i\in\Phi_U} \sum_i \int_{\mathbb{S}^2} \left\| \operatorname{Log}\left[g_{S_i}^{-\frac{1}{2}} \phi_i \circ g_{\phi_i \circ \mathcal{S}} g_{S_i}^{-\frac{1}{2}} \right] \right\|_F^2 + \left[H_{S_i} - \phi_i \circ H_{\mathcal{S}} \right]^2 d\mathbb{S}^2 \right) \quad (6)$$

Estimating $\nu(\mathfrak{s})$ can be done with a two-step optimization: first, holding the estimates of the ϕ_i constant to update the current $\nu(\mathfrak{s})$ estimate; and second, holding $\nu(\mathfrak{s})$ constant to update all remappings ϕ_i simultaneously. These two steps are repeated until an optimality condition is met.

The first step in the optimization of (6) requires repeated point-wise estimates of the mean metric structure and the "mean" mean curvature. While the curvature term is trivially computed, the first term has no closed-form solution. Iterative approximation is required. Under the log metric, the mean metric is a 2x2 matrix satisfying

$$G(g_{\mathcal{S}}) = \arg \min_{g_{\mathcal{S}}\in SPD(2)} \sum_i \left\| \operatorname{Log}\left[g_{S_i}^{-\frac{1}{2}} g_{\mathcal{S}} g_{S_i}^{-\frac{1}{2}} \right] \right\|_F^2 . \quad (7)$$

The gradient of the above expression can be shown to be

$$\nabla G_{jk} = \sum_i 2tr\left[\operatorname{Log} X_i \, X_i^{-1} \frac{d}{d g_{\mathcal{S},jk}} X_i \right], \qquad X_i = g_{S_i}^{-\frac{1}{2}} g_{\mathcal{S}} g_{S_i}^{-\frac{1}{2}} . \quad (8)$$

This formulation differs slightly from [17], but the means are in fact equivalent.

The second step in the optimization scheme of (6) requires an additional term to ensure that the spherical remappings ϕ_i remain in Φ_U, i.e. that they remain area-preserving. We use the same term as was done in [1]. The second optimization step then becomes very similar to the optimization problem in [1] summed over the surfaces in $\mathfrak{s}$. The only difference is the absence of the Gaussian curvature term, which, as we have shown, is not needed for a unique representation.

The overall optimization problem for finding Karcher means in the intrinsic shape framework can now be stated briefly as finding the optimal metric and curvature structure ν and spherical reparameterizations φ_i to minimize the following cost:

$$\mathcal{C}(\mathfrak{s}, \nu, \{\varphi_i\}) = \sum_{S_i\in\mathfrak{s}} \mathcal{L}^2(S_i, \nu) + R \int_{\mathbb{S}^2} (\log det[D\varphi_i])^2 d\mathbb{S}^2 . \quad (9)$$

5 Implementation Details

5.1 Optimization

We follow [1], using the spherical fluid framework to optimize (9) and compute the Karcher mean. As an initial step, all surfaces are spherically registered to a single target shape, exactly as in [1]. The point-wise metric and curvature mean map serves as the initial guess before group-wise registration. From this point, the only difference between [1] and this work is that the moving template metric + curvature field is updated at every iteration as the current point-wise mean. The average mean curvature is straight-forward, while the average metric tensors can be estimated quickly with a backtracking line search using the gradient in (8).

5.2 Surface Reconstruction

In [19], the authors propose to integrate the Gauss-Codazzi equations directly to reconstruct a surface from its conformal parameterization. As we do not compute explicit conformal maps, using this approach on the general metric tensor may be computationally challenging. Instead, we use a least-squares estimate that is a modification of the approach in [1]. As in [1], we rely on discrete differential geometry operators described in [20]. Suppose we have a spherical mesh $m = \langle V, E\rangle,\ |x| = 1\ \forall x \in V$, and g, H defined at each vertex in V, with an edge set E. The mesh representing an embedding in space, $\langle S(V), E\rangle$ minimizes the least-squares problems:

$$E_g = \sum_{xy\in E} \sqrt{\boldsymbol{A}(x)\boldsymbol{A}(y)}\left(\|Sx - Sy\| - L_{xy}\right)^2,$$
$$L_{xy} = \frac{1}{2}[(x-y)^T g(x)(x-y)]^{1/2} + \frac{1}{2}[(x-y)^T g(y)(x-y)]^{1/2} \tag{10}$$

$$E_H = \sum_{x\in V} \boldsymbol{A}(x)\left(\left\langle\left[\sum_{y\in N_1(x)} \frac{(\cot a_{xy} + \cot b_{xy})(Sx - Sy)}{4\boldsymbol{A}(x)}\right], \boldsymbol{n}(x)\right\rangle - H(x)\right)^2.$$

Here, $\boldsymbol{n}$ is the surface normal, $\boldsymbol{A}(x)$ is the area element, and a_{xy}, b_{xy} are angles opposite edge xy. The area element can be estimated from the spherical area element and $det(g)^{\frac{1}{2}}$. The cotangent weights in the estimated curvature operator themselves vary with the evolving mesh. However, when the initial guess is sufficiently close, e.g. when it is the Euclidean mean, fixing $\cot a_{xy} + \cot b_{xy}$ according to the metric generally does not affect the quality of the solution. To solve the system in (10), we must define a fixed frame. This can be done by computing the shape of an individual triangle based on the metric tensor alone, and fixing it in space. In practice, avoiding this step when the initial guess is sufficiently close does not affect the behavior of the optimization.

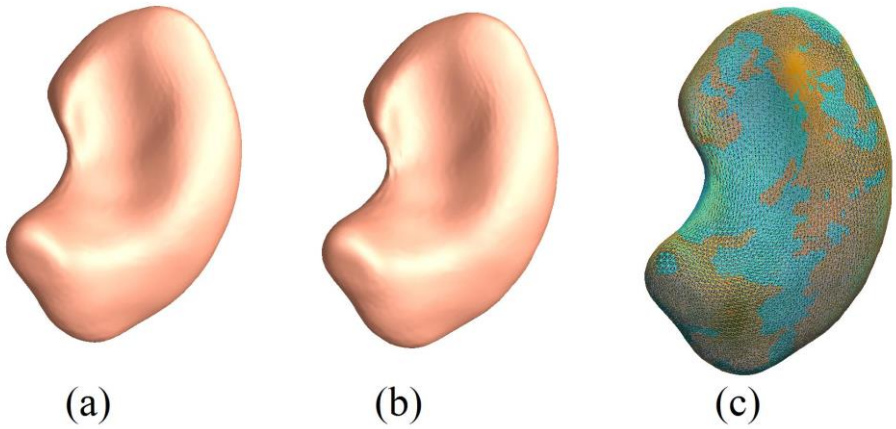


Fig. 2. Putamen (a) Euclidean mean; (b) Karcher mean; (c) overlay of (a) – blue and (b) – orange. Extreme curvatures are better preserved in Karcher means compared to Euclidean means consistently across shape types.

6 Experiments

We compute the Karcher means of 100 white matter surface models from healthy elderly participants in the ADNI 1 study. In **Figure 3**, we show the comparison to the Euclidean average. In general, more geometric detail is preserved over the Karcher mean, as Euclidean averaging tends to erode sharp features. Additional experiments were run on 400 subcortical shapes of typically developing children and young adults. Results for hippocampal shape are displayed in **Figure 4** and for putamen shape in **Figure 2**. Sum of squared distances $\sum_{S_i \in \mathbb{S}} \mathcal{L}^2(S_i, \nu)$ is displayed in **Table 1** for the Euclidean and Karcher means. Compute times for the Karcher mean scale linearly with the number of subjects, since the reparameterization step is an order of magnitude costlier than the point-wise metric mean step of the optimization. When compared to the performance of pairwise registration in [1], the analogous computation here – a mean of two shapes – is not significantly different (on the order of a few minutes for a spherical harmonic bandwidth of 128).

Table 1. Sum of geodesic squares for Euclidean and Karcher means.

	Cortex	Hippocampus	Putamen
$\sum_{S_i \in \mathbb{S}} \mathcal{L}^2(S_i, \nu)$ Euclidean	2.1 x 10^4	8.5 x 10^2	2.4 x 10^3
$\sum_{S_i \in \mathbb{S}} \mathcal{L}^2(S_i, \nu)$ Karcher	5.4 x 10^3	3.2 x 10^2	1.4 x 10^3

7 Conclusion

We have presented a novel intrinsic shape representation in Riemannian setting, based on metric tensors and mean curvatures. In this setting, we show that it is possible to efficiently compute manifold mean representations and reconstruct their surfaces embedded in space. The mean estimation method is efficient due to the closed form solution for the geodesic length on the shape manifold. Our method is capable of group-wise registering complex shapes such as the cortical surface, and efficiently estimating sample means. When reconstructed into real surfaces, the realization of our mean estimate consistently preserves high-curvature and fine geometric features better than Euclidean averaging of coordinates. This provides some empirical proof that the suggested shape framework captures intrinsic properties of surfaces better than simpler methods based on the surface embedding.

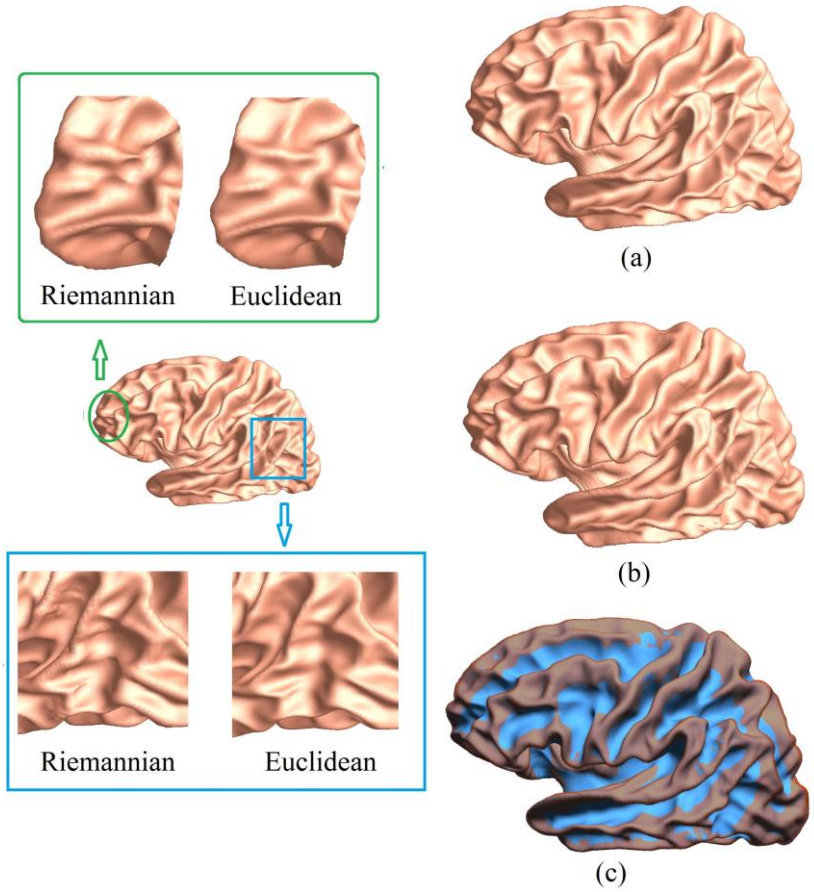


Fig. 3. (a) Euclidean Average; (b) Riemannian average – Karcher mean; (c) Overlay of (a) and (b), Karcher mean in orange, Euclidean mean in blue. Deeper sulci and taller gyri are prominent in the Riemannian average when compared to the Euclidean approximation.

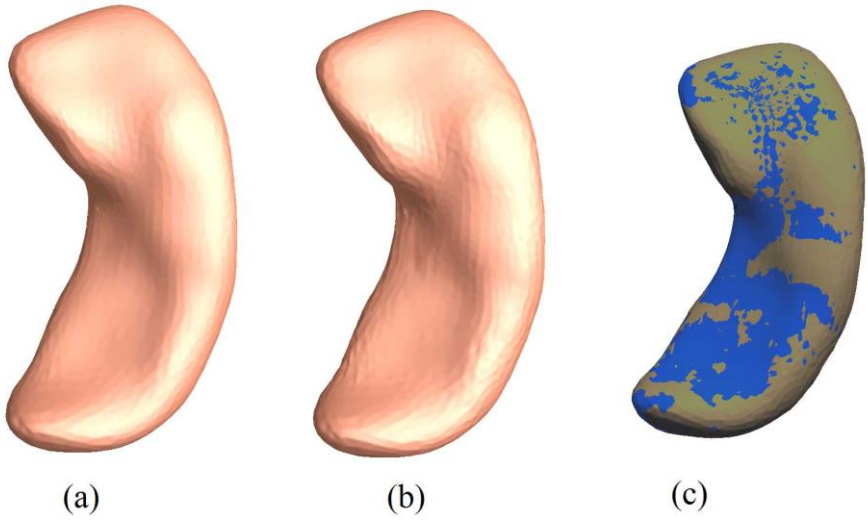


Fig. 4. Hippocampal (a) Euclidean mean; (b) Karcher mean; (c) overlay of (a) – blue and (b) – orange

References

1. Gutman, B.A., Fletcher, T.P., Cardoso, M.J., Fleishman, G., Lorenzi, M., Thompson, P.T., Ourselin, S.: A Riemannian Framework for Intrinsic Comparison of Closed Genus-Zero Shapes. Lect Notes Comput Sc (2015)
2. Gu, X., Wang, Y., Chan, T.F., Thompson, P.M., Yau, S.T.: Genus zero surface conformal mapping and its application to brain surface mapping. Ieee T Med Imaging 23, 949-958 (2004)
3. Zeng, W., Lui, L.M., Luo, F., Chan, T.F.-C., Yau, S.-T., Gu, D.X.: Computing quasiconformal maps using an auxiliary metric and discrete curvature flow. Numer. Math. 121, 671-703 (2012)
4. Shi, J., Thompson, P.M., Gutman, B., Wang, Y.: Surface fluid registration of conformal representation: Application to detect disease burden and genetic influence on hippocampus. Neuroimage 78, 111-134 (2013)
5. Yeo, B.T.T., Sabuncu, M.R., Vercauteren, T., Ayache, N., Fischl, B., Golland, P.: Spherical Demons: Fast Diffeomorphic Landmark-Free Surface Registration. Ieee T Med Imaging 29, 650-668 (2010)
6. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. Neuroimage 45, S61-S72 (2009)
7. Gutman, B.A., Madsen, S.K., Toga, A.W., Thompson, P.M.: A Family of Fast Spherical Registration Algorithms for Cortical Shapes. In: Shen, L., Liu, T., Yap, P.-T., Huang, H., Shen, D., Westin, C.-F. (eds.) Multimodal Brain Image Analysis, vol. 8159, pp. 246-257. Springer International Publishing (2013)

8. Christensen, G.E., Rabbitt, R.D., Miller, M.I.: Deformable templates using large deformation kinematics. Image Processing, IEEE Transactions on 5, 1435-1447 (1996)
9. Yonggang, S., Rongjie, L., Wang, D.J.J., Pelletier, D., Mohr, D., Sicotte, N., Toga, A.W.: Metric Optimization for Surface Analysis in the Laplace-Beltrami Embedding Space. Medical Imaging, IEEE Transactions on 33, 1447-1463 (2014)
10. Bauer, M., Bruveris, M., Michor, P.: Overview of the Geometries of Shape Spaces and Diffeomorphism Groups. Journal of Mathematical Imaging and Vision 50, 60-97 (2014)
11. Qiu, A., Younes, L., Miller, M.I., Csernansky, J.G.: Parallel transport in diffeomorphisms distinguishes the time-dependent pattern of hippocampal surface deformation due to healthy aging and the dementia of the Alzheimer's type. Neuroimage 40, 68-76 (2008)
12. Bauer, M., Bruveris, M.: A new Riemannian Setting for Surface Registration. 3rd MICCAI Workshop on Mathematical Foundations of Computational Anatomy 182-194 (2011)
13. Ma, J., Miller, M.I., Younes, L.: A Bayesian Generative Model for Surface Template Estimation. International Journal of Biomedical Imaging 2010, 974957 (2010)
14. Kurtek, S., Klassen, E., Zhaohua, D., Jacobson, S.W., Jacobson, J.B., Avison, M.J., Srivastava, A.: Parameterization-Invariant Shape Comparisons of Anatomical Surfaces. Medical Imaging, IEEE Transactions on 30, 849-858 (2011)
15. Gu, X., Wang, Y., Yau, S.-T.: Geometric Compression Using Riemann Surface Structure. 171-182 (2003)
16. Ebin, D.G.: On the space of Riemannian metrics. 1001-1003 (1968)
17. Moakher, M.: A Differential Geometric Approach to the Geometric Mean of Symmetric Positive-Definite Matrices. SIAM Journal on Matrix Analysis and Applications 26, 735-747 (2005)
18. Lawson, J.H.B., Tribuzy, R.d.A.: On the mean curvature function for compact surfaces. 179-183 (1981)
19. Lui, L.M., Wen, C.F., Gu, X.F.: A Conformal Approach for Surface Inpainting. Journal of Inverse Problems and Imaging 7, 863-884 (2013)
20. Meyer, M., Desbrun, M., Schröder, P., Barr, A.: Discrete Differential-Geometry Operators for Triangulated 2-Manifolds. In: Hege, H.-C., Polthier, K. (eds.) Visualization and Mathematics III, pp. 35-57. Springer Berlin Heidelberg (2003)

Adaptive time-stepping in diffeomorphic image registration with bounded inverse consistency error

Akshay Pai[1], Stefan Klein[3], Stefan Sommer[1], Lauge Sørensen[1,2], Sune Darkner[1], Jon Sporring[1] and Mads Nielsen[1,2]

[1] DIKU, University of Copenhagen, Copenhagen,
[2] Biomediq A/S, Copenhagen
[3] Biomedical Imaging Group Rotterdam, Depts. of Medical Informatics & Radiology, Erasmus MC, Rotterdam, Netherlands

Abstract. In a continuous setting, diffeomorphisms generated by stationary velocity fields (SVF) are invertible transformations with differentiable inverses. However, due to the numerical integration of the velocity field, inverse consistency is not achieved in practice. In SVF based image registration, inverse consistency is therefore often enforced through a penalty term. Existing penalty terms penalize the inverse consistency error generated by the composition of the forward and backward transformations. However, in such terms, a higher consistency requirement pushes the transformation towards linearity due to the discretization involved and fixed number of integration time-steps. In this paper, we propose a method to both penalize inverse consistency error and to adaptively set the number of integration time-steps required, so that the predicted maximum inverse consistency error is bounded, taking into account discretization errors. This formulation allows more flexibility in the transformation model to realize complex deformations while still achieving the desired level of inverse consistency. Using synthetic examples, we show that the measured inverse consistency and the predicted inverse consistency match. Also, the proposed method is able to achieve more accurate image registration. On the MGH10 dataset, the Jaccard index of the proposed method on inter-subject registration reaches the same level as the registration scheme using a fixed-time step and the conventional penalty term while using a lower number of integration time-steps, thus saving on the computational time.

1 Introduction

Image registration plays a very important role in the field of medical research and clinical applications. It has found utilities in both the longitudinal and cross-sectional characterization of human anatomy. It is particularly useful because it provides localized transformations that can be used to study deformation at an organ level. For instance, image registration in the form of tensor-based morphometry is used to measure atrophy in various brain regions which is then

used to quantify diseases such as the Alzheimer's disease. However, to effectively use transformations from an image registration, they need to be free of any bias. One common bias that has been found to have severe implications on bio-marker quantification like longitudinal atrophy estimation, is the inverse consistency error in transformations [1].

Ideally in image registration, the transformation between two images is expected to be invariant to the order of the choice of source and target. In practice however, such a transformation is not possible due to several reasons such as discrete image information, finite degrees of freedom, and discretization errors due to numerical integration of flow fields. Therefore, inverse consistency is often enforced as a penalty term. The existing inverse consistent methods, particularly pertaining to diffeomorphic approaches, either look at the forward and backward transformations simultaneously [2] or the velocity field by making an assumption that the forward and backward velocity fields are exact negatives of each other [3].

In a continuous setting, the path generated by a velocity field can be exactly retraced. However, in a discrete setting, the path is approximated using a set of piecewise linear steps. Most often retracing these steps will yield an inverse inconsistent transformation. The fewer the time-steps, the higher is the inverse consistency. Also, the higher the curvature of the path, the more steps are required to be close to being inverse consistent. Therefore, if one needs control over the inverse consistency error, the effect of the discretization error needs to be accounted for.

In this paper, we propose a method to both penalize and account for discretization errors. Instead of pushing the inverse consistency error to zero which tends the transformation to linearity, we propose to bound the error by adjusting the number of integration time-steps used to integrate the stationary velocity field (SVF). This lends flexibility in the transformation model to reach more complex deformations while still being reasonably inverse consistent. The main contributions of the paper are as follows,

- We propose to penalize the inverse consistency error based only on the velocity fields. This is achieved using the Baker-Campbell-Hausdorff (BCH) [4] expansion to formulate the theoretical inverse consistency in terms of the velocity field.
- We propose to adjust the required number of integration time-steps prior to each optimization iteration by predicting and thereby bounding the inverse consistency error based on the properties of the velocity field.

Note that through out the paper, we use two sets of parameterizations of the velocity field - one for the forward transformation and one for the backward.

2 Background and Outline

The most commonly used inverse consistency term follows a similar framework as proposed by *Christensen et al* [2] with variations. The method in essence involves

jointly estimating the forward and backward transformations while minimizing the inverse consistency error. While the earlier approach [2] involved estimating both the transformations and its inverses, the current approaches [5], just maintain two transformations and through the penalty term, push them to be inverse consistent. In a different approach [6], the proposed model incorporated stochastic errors in the inverse consistent constraints as a post-processing step. Further in [7], the gradient descent updates of the forward and backward transformations were symmetrized by estimating a linear Taylor series expansion of the inverse consistency condition. Another class of inverse consistent registration scheme warps the image to a mid point [8], median [9] or to a mean shape [10]. The popular logDemons approach on the other hand assumes the forward velocity field is the negative of the backward velocity field to maintain inverse consistency [3]. In the method we propose to adjust the number of integration time steps during the registration optimization to bound the inverse consistency error. And, in order to ensure the number of time-steps is in a reasonable range, we also add a penalty term.

We begin with introducing the stationary velocity field based image registration method followed by presenting the Wendland kernel bundle framework used as a transformation model. We will then discuss our contribution where the inverse consistency term and the predictive inverse consistency term are proposed. We will then present some experiments on synthetic examples and also on the publicly available MGH10 dataset.

3 Registration

Given a floating image S_1 and a reference image S_2 with a spatial domain $\Omega \in \mathbb{R}^d$, image registration involves finding a transformation $\varphi : \Omega \to \Omega$ that aligns these images. We maintain two warps φ_f, φ_b: one for a forward transformation and the second for backward. The transformation is a result of minimizing the dissimilarity between the images under certain constraints encoded in the regularization term. A general cost function is of the form:

$$\begin{aligned} E(\varphi_f, \varphi_b) = \arg\min_{\varphi_f, \varphi_b} & E_D(S_1(\varphi_f), S_2) + E_D(S_2(\varphi_b), S_1) + \lambda_R E_{R_f}(\varphi_f, \varphi_b) \\ & + \lambda_{\text{ICC}} E_{\text{ICC}_f}(\varphi_f, \varphi_b) + \lambda_R E_{R_b}(\varphi_b, \varphi_f) \\ & + \lambda_{\text{ICC}} E_{\text{ICC}_b}(\varphi_b, \varphi_f). \end{aligned} \tag{1}$$

where $\lambda_R, \lambda_{\text{ICC}}$ are the user-specified constants, E_D is a dissimilarity measure that allows comparison of the floating image to the reference image, E_R is a regularization term that encodes desired properties of φ, and E_{ICC} is an additional penalty term to enforce inverse consistency which is the focus of this paper. Normalized mutual information (NMI) [11] is used as the similarity measure in this paper.

Let Diff(Ω) be the space containing the diffeomorphic transformation $\varphi : \Omega \to \Omega$, $\phi : \Omega \times \mathbb{R} \to \Omega$ and finally, $\mathbf{v} : \Omega \to \mathbb{R}^d$ be the velocity field belonging

to the tangent space of Diff(Ω) at identity Id. In SVFs, the governing differential equation can be written as:

$$\frac{\partial \phi(\mathbf{x}, t)}{\partial t} = \mathbf{v}(\phi(\mathbf{x}, t)), \quad \varphi = \phi(\mathbf{x}, 1) = \mathrm{Exp}(\mathbf{v}), \tag{2}$$

ϕ, φ represent both forward and backward transformations. The final transformation is then defined as the Lie group exponential map $\mathrm{Exp}(\mathbf{v})$. This Lie group exponential is realized as an Euler integration.

3.1 Wendland kernels

Instead of the usual approach of first choosing a operator and then constructing a reproducing kernel Hilbert space (RKHS), the reproducing kernels with an associated RKHS was directly chosen [12, 13]. One example of such a reproducing kernel is the class of Wendland basis functions [14]. They are particularly interesting because of their finite support and smoothness properties similar to popularly used B-spline. In addition, they are norm-minimizing.

Wendland kernels are positive definite functions with positive Fourier transforms. They are minimal degree polynomials on $[0, 1]$ and yield C^{2k} (k is the desired degree of smoothness) smooth radial basis functions on $\mathbb{R}^d$. We choose the Wendland kernels $\psi(r) = (1 - r)_+$ and $\psi(r) = (1 - r)_+^4(4r + 1)$ where $(\cdot)_+$ denotes semi-positive definiteness. The velocity field defined in (2) may now be parameterized using Wendland kernels as follows,

$$\mathbf{v}(\mathbf{x}) = \sum_i^N \psi(r(\mathbf{x}_i, \mathbf{x}))\mathbf{p}_i. \tag{3}$$

Note that $r(\mathbf{x}, \mathbf{y}) = \frac{||\mathbf{x}-\mathbf{y}||_2}{a}$, where a is a scaling parameter, $\mathbf{p}_i \in \mathbb{R}^d$ is the coefficient attached with every kernel center $\mathbf{x}_i$, and N is the number of kernels having an influence on $\mathbf{x}$. Due to the reproducing property of the kernel parameterizing the velocity fields, the regularization term E_R is chosen to be the Hilbert norm on the velocity field given by $||\mathbf{v}||_V^2 = \sum_{i,j} \mathbf{p}_i^T \psi(r(\mathbf{x}_i, \mathbf{x}_j))\mathbf{p}_j$.

3.2 Kernel bundle framework

The reasoning behind the need for a multi-scale representation of a deformation has been well discussed in a previous work [15]. In short, the velocity fields are linear sums of individual kernels at L levels. It is represented as,

$$\mathbf{v}(\mathbf{x}) = \sum_{m=1}^{L} \mathbf{v}_m(\mathbf{x}) = \sum_{m=1}^{L} \sum_i^{N_m} \psi_m(r(\mathbf{x}_i^m, \mathbf{x}))\mathbf{p}_i^m. \tag{4}$$

The expression of the optimization to (1) in a kernel bundle framework can be written as,

$$\begin{cases} \arg\min_{\mathbf{v}_{f_1},\mathbf{v}_{b_1}} E(\mathrm{Exp}(\mathbf{v}_{f_1}),\mathrm{Exp}(\mathbf{v}_{b_1})), \\ \arg\min_{\mathbf{v}_{f_2},\mathbf{v}_{b_2}} E(\mathrm{Exp}(\mathbf{v}_{f_1}+\mathbf{v}_{f_2}),\mathrm{Exp}(\mathbf{v}_{b_1}+\mathbf{v}_{b_2})), \\ \vdots \\ \arg\min_{\mathbf{v}_{f_L},\mathbf{v}_{b_L}} E(\mathrm{Exp}(\sum_{m=1}^{L}\mathbf{v}_{f_m}),\mathrm{Exp}(\sum_{m=1}^{L}\mathbf{v}_{b_m})) \end{cases}$$

We sequentially optimize for each space of the velocity fields. The kernels at each level can be of any support. For instance, one can have infinitely supported Gaussian kernels in a coarser registration scale and have compactly supported kernels handle finer resolutions in the registration.

4 Inverse consistency

Usually, inverse consistency is addressed as a penalization of the offset generated by a composition of the forward and backward transformations. If this offset is pushed towards zero while having a fixed number of time-steps, the transformations tend towards linearity. In order to achieve larger deformations in finite steps, one needs to allow some degree of inverse inconsistency due to the discrete nature of the image registration problem. While the existing methods penalize the error, in this paper we propose to both penalize the error and also adjust the number of integration time-steps such that the inverse consistency error is bounded. We will, through the relationship between Lie brackets and inverse consistency, show that the proposed scheme allows for some inverse consistency error in a symmetric fashion.

In the following section, we will address a number of inverse consistency terms. For reading ease, we define the abbreviations early and as follows: I_t is the theoretical inverse consistency, I_p is the commonly used inverse consistency term, $I_p^{\frac{1}{n}}$ is the inverse consistency error for the first composition of the Euler's integration of the velocity field and I_{pd} is the predicted inverse consistency term. We will use the subscripts f, b for forward and backward registration. In addition, the formulation of the backward inverse consistency error will follow the forward unless stated otherwise.

Theoretically, the forward inverse consistency term I_{t_f}, can be expressed as a composition of the forward and backward exponentials,

$$I_{t_f} = \mathrm{Exp}(\mathbf{v}_b)\mathrm{Exp}(\mathbf{v}_f) - \mathbf{x}$$

where $\mathbf{v}_f, \mathbf{v}_b$ are the forward and backward velocity fields respectively. The BCH formula is used to efficiently express this composition in terms of the velocity fields as follows,

$$\begin{aligned} I_{t_f} &= \mathrm{Exp}(\mathbf{v}_f + \mathbf{v}_b + \frac{1}{2}[\mathbf{v}_f, \mathbf{v}_b]) - \mathbf{x}, \\ &= \mathbf{v}_f + \mathbf{v}_b + \frac{1}{2}[\mathbf{v}_f, \mathbf{v}_b] + \mathcal{O}(||\mathbf{v}_f + \mathbf{v}_b + \frac{1}{2}[\mathbf{v}_f, \mathbf{v}_b]||^2). \end{aligned} \tag{5}$$

where $[,]$ is the Lie bracket or the commutator that quantifies the amount of change of $\mathbf{v}_b$ in the direction of $\mathbf{v}_f$. It is expressed as,

$$[\mathbf{v}_f, \mathbf{v}_b] = \nabla \mathbf{v}_b \ \mathbf{v}_f - \nabla \mathbf{v}_f \ \mathbf{v}_b$$

The second line in (5) follows from a linear approximation of the exponential. The commonly used inverse consistency term may be defined as,

$$I_{p_f} = \phi_b(\phi_f(\mathbf{x})) - \mathbf{x}. \tag{6}$$

where ϕ_b and ϕ_f are the results of numerical integration of $\mathbf{v}_b$ and $\mathbf{v}_f$, respectively. This term may also be approximated in terms of the velocity field similar to I_{t_f}. It was shown that the final inverse consistency can be expressed in terms of the step-wise inverse consistency error [5] as follows; assuming n time-steps, we get,

$$\begin{aligned} I_{p_f} &\approx n(I_{p_f}^{\frac{1}{n}}), && (7)\\ &= n(\phi_b^{\frac{1}{n}}(\phi_f^{\frac{1}{n}}(\mathbf{x})) - \mathbf{x}), && (8)\\ &= n(\mathbf{x} + \frac{1}{n}\mathbf{v}_f(\mathbf{x}) + \frac{1}{n}\mathbf{v}_b(\mathbf{x} + \frac{1}{n}\mathbf{v}_f(\mathbf{x})) - \mathbf{x}), && (9)\\ &= \mathbf{v}_f(\mathbf{x}) + \mathbf{v}_b(\mathbf{x}) + \frac{1}{n}\nabla \mathbf{v}_b \ \mathbf{v}_f(\mathbf{x}) + \mathcal{O}(||\frac{1}{n}\mathbf{v}_f||^2). && (10) \end{aligned}$$

Note that (9) was obtained using the Taylor expansion of $\mathbf{v}_b(\mathbf{x} + \frac{1}{n}\mathbf{v}_f(\mathbf{x}))$.

In theory, I_{t_f} and I_{p_f} are equivalent as $n \to \infty$, however in practice they are not since n is finite. This implies that the inverse consistency error term we expect to minimize I_{t_f}/I_{t_b} is different than the inverse consistency error we practically minimize I_{p_f}/I_{p_b}.

In this work, we will propose a new way of handling inverse consistency. What we propose is to use the I_{p_f} for penalization and use a prediction term I_{pd} to adjust the number of integration time-steps such that the inverse consistency is bounded.

Given that we maintain two warps, the predicted inverse consistency error term may be derived as follows,

$$I_{pd} = n(\frac{1}{2}(\phi_b^{\frac{1}{n}}(\phi_f^{\frac{1}{n}}(\mathbf{x})) - \mathbf{x} + \phi_f^{\frac{1}{n}}(\phi_b^{\frac{1}{n}}(\mathbf{x})) - \mathbf{x})) \tag{11}$$

Using (9), we get,

$$\begin{aligned} I_{pd} &= \mathbf{v}_f(\mathbf{x}) + \mathbf{v}_b(\mathbf{x}) + \frac{1}{2n}\nabla \mathbf{v}_b(\mathbf{x}) \ \mathbf{v}_f(\mathbf{x}) + \frac{1}{2n}\nabla \mathbf{v}_f(\mathbf{x}) \ \mathbf{v}_b(\mathbf{x}), && (12)\\ &= \mathbf{v}_f + \mathbf{v}_b + \frac{1}{2n}\{\mathbf{v}_f, \mathbf{v}_b\}. && (13) \end{aligned}$$

We use $\{\}$ since it resembles an anti-commutator. It is sometimes used to construct a Jordan algebra, but that is not our purpose. Since $E_{ICC_f} = ||I_{t_f}||^2$ will be penalized during registration, $\mathbf{v}_f + \mathbf{v}_b$ will already be pushed towards zero.

It is thus fair to discard that term in (13) for prediction purposes. We can then derive the following formula to estimate the number of time steps required to achieve a user-specified inverse consistency I_{max} as:

$$n = \frac{\max_{\mathbf{x}} ||\{\mathbf{v}_f, \mathbf{v}_b\}||}{2I_{max}}. \tag{14}$$

n will therefore give us the number of integration time-steps required to achieve a maximum inverse consistency of I_{max}.

In summary, the proposed method involves using (5) as a penalty term and (14) to predict the number integration time-steps required at each iteration of the registration optimization but setting a user-defined value, I_{max} .

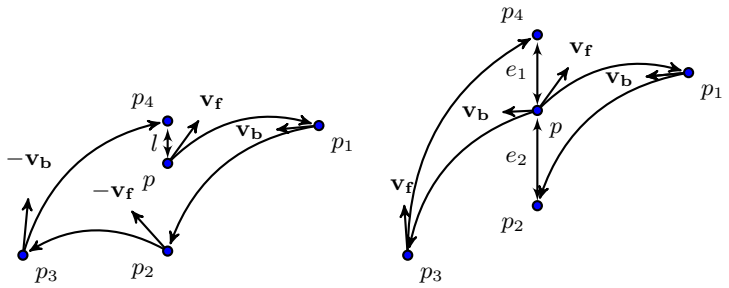


Fig. 1: a) Representation of the Lie Bracket; $l = [\mathbf{v}_f, \mathbf{v}_b]$; b) Representation of $I^{\frac{1}{n}}_{p_f}$ and $I^{\frac{1}{n}}_{p_b}$; $e_1 = I^{\frac{1}{n}}_{p_f}$, $e_2 = I^{\frac{1}{n}}_{p_b}$

The geometric interpretation of a Lie bracket may be seen in Figure 1a. What it represents is that, there is no difference in starting at one point p, traveling a time t over the flow of $\mathbf{v}_f$ and then a time t over the flow of $\mathbf{v}_b$, or, instead, traveling first t over the flow of $\mathbf{v}_b$ and then over the flow of $\mathbf{v}_f$. This is similar to looking at the difference in the step-wise inverse consistency term ($I^{\frac{1}{n}}_{p}$) shown in Figure 1b. We can see that, the Lie Bracket and inverse consistency errors are related since both capture the second order information of the deformation. The formal relation can be shown as follows,

$$I^{\frac{1}{n}}_{p_f} = \frac{1}{n}\mathbf{v}_f(\mathbf{x}) + \frac{1}{n}(\mathbf{v}_b(\mathbf{x}) + \frac{1}{n}\nabla\mathbf{v}_b\mathbf{v}_f(\mathbf{x})). \tag{15}$$

Similarly,

$$\begin{aligned} I^{\frac{1}{n}}_{p_b} &= \phi_f^{\frac{1}{n}}(\phi_b^{\frac{1}{n}}) - \mathbf{x}, \\ &= \frac{1}{n}\mathbf{v}_b(\mathbf{x}) + \frac{1}{n}(\mathbf{v}_f(\mathbf{x}) + \frac{1}{n}\nabla\mathbf{v}_f\mathbf{v}_b(\mathbf{x})),. \end{aligned} \tag{16}$$

Now subtracting (15) and (16):

$$I_{p_f}^{\frac{1}{n}} - I_{p_b}^{\frac{1}{n}} = \frac{1}{n^2}(\nabla \mathbf{v}_b \mathbf{v}_f(\mathbf{x}) - \nabla \mathbf{v}_f \mathbf{v}_b(\mathbf{x})) = \frac{1}{n^2}[\mathbf{v}_f, \mathbf{v}_b]. \tag{17}$$

Using this equivalence, we can redefine our penalty term E_{ICC} as follows:

$$\begin{aligned} E_{ICC_f} &= ||\mathbf{v}_f + \mathbf{v}_b + \frac{1}{2}[\mathbf{v}_f, \mathbf{v}_b]||^2, \\ &= ||\mathbf{v}_f + \mathbf{v}_b + \frac{n^2}{2}(I_{p_f}^{\frac{1}{n}} - I_{p_b}^{\frac{1}{n}})||^2. \end{aligned} \tag{18}$$

and E_{ICC_b} follows the same formulation. Note that the representation of (18) in terms of step-wise inverse consistency error is particularly useful because the gradient of (18) with respect to the transformation parameters yields the terms $\nabla \mathbf{v}_f$ and $\nabla \mathbf{v}_b$ which are required to compute the number of time-steps in (14).

5 Experiments and Results

In this section, we will conduct three experiments to evaluate the performance of the proposed method against the fixed-time step with the conventional inverse consistency regularization method. For evaluation purposes, we will use the following conventional definition of inverse consistency error term, i.e.,

$$\text{ICC} = \frac{1}{k}\sum_{i=1}^{k}\frac{1}{2}(||\varphi_b(\varphi_f(\mathbf{x}_i)) - \mathbf{x}_i||^2 + ||\varphi_f(\varphi_b(\mathbf{x}_i)) - \mathbf{x}_i||^2). \tag{19}$$

where φ_b, φ_f are the backward and forward transformations obtained from integrating the velocity fields $\mathbf{v}_b$ and $\mathbf{v}_f$ respectively and k is the number of voxels.

5.1 Prediction test

In the first of the synthetic experiments, we will compare the predicted inverse consistency error (I_{pd}, (13)) by the proposed method and the measured inverse consistency error (ICC, (19)). We perform this test since (13) is a key to computing the number of integration time-steps. We pick a random scan from the

Mag of CC	I_{pd}(std)	ICC(std)	Max I_{pd}	Max ICC
0.25	0.55 (0.34)	0.53 (0.46)	1.89	2.35
1.0	0.55 (0.43)	0.58 (0.60)	3.68	4.86
2.5	0.63 (0.43)	0.59 (0.67)	2.45	4.79

Table 1: Synthetic example results; Mag of CC: Std of Gaussian used to control the magnitude of the control point coefficient (CC); Units in mm

MGH10 dataset and deform it using a B-spline (control point spacing of 5 mm) whose control coefficients are chosen randomly with a monotonic increase in magnitude. The original images and the deformed images are registered using the Wendland kernel bundle framework together with the common inverse consistency error term (19) and a fixed time-step of 16. Once the optimum velocity fields are estimated, the predicted inverse consistency error (I_{pd}, (13)) is computed. This is compared to the measured inverse consistency error using (19). As Table 1 illustrates, the mean and standard deviation of the inverse consistency error predicted by the proposed method (13) matches that of the measured inverse consistency error (19) indicating that (14) is a fair formulation for estimating the required number of integration time-steps.

5.2 Registration on synthetic example

To test the registration proposed in this paper, we generated synthetic data where the ground truth deformation is known. The synthetic data was generated by deforming the faces of the cube (50^3 mm^3 placed in a 256^3 mm^3 image) using a Gaussian of standard deviation 0.07 (of the cube side length). The deformed and undeformed cubes were registered using the proposed adaptive time-stepping scheme and a conventional scheme with an inverse consistency error penalty as (19) and a fixed number of time-steps of 16 . For both the schemes, the regularization constant was set to 0.1 and two levels of the kernel bundle were used. The first level was parameterized with a kernel of support 4 and the next by a kernel of support 2. I_{max} was set to 1. We set $\lambda_R = 0.3$ and $\lambda_{ICC} = 0.01$. Figure 2 illustrates the source and target images. Following this, the accuracy of the overlap is assessed using the Jaccard index which is given by,

$$\mathrm{JI} = \frac{|S(\phi) \cap T|}{|S(\phi) \cup T|}. \tag{20}$$

where S is the source segmentation, ϕ is the transformation and T is the target segmentation. In addition to this, we measure ICC errors for both the methods using (19). In Table 2, we see that the adaptive scheme showed a better mean overlap (forward and backward) and a very similar inverse consistency error when compared to the fixed-time stepping scheme. However, in the first level of registration the adaptive scheme used only a maximum of 8 time-steps and in the second level it used a maximum of 12 time-steps.

Method	JI	ICC(std)	Max ICC	Max. time steps
Adaptive	0.96	0.01 (0.14)	0.90	8,12
Fixed	0.89	0.01 (0.06)	2.30	16

Table 2: Adaptive scheme versus fixed time-step scheme. Max. time steps: Maximum number of time-steps on level 1 and 2 of the kernel bundle framework.

Fig. 2: Target cube, deformed target cube

5.3 Application on MGH10

We perform an evaluation on the MGH10 dataset[4]. Here the ability to match a set of manually segmented regions of interest via pair-wise registration is evaluated. Each scan is initially reformatted to isotropic voxels (voxel size 1 mm^3 and dimensions of 256^3. The pair-wise images are linearly aligned using 9 DOF. Both the images were mapped to the mid-point of the affine space using the square root of the affine transformation. After the linear alignment, images are non-linearly aligned using the presented registration scheme. Once the registrations are performed, the manual labeled segmentation from the floating image are warped to the reference image using a nearest neighbor interpolation. One

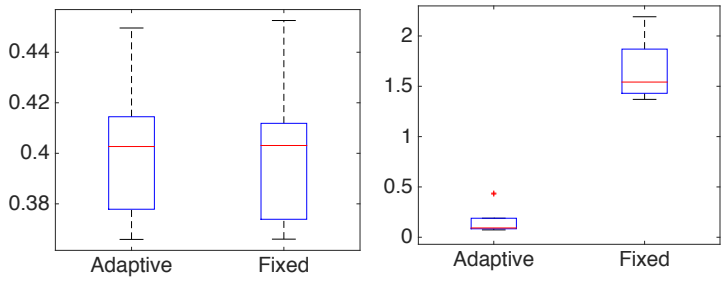


Fig. 3: Box plot of the jaccard index and mean inverse consistency error.

subject is randomly chosen and the other subjects are registered to it. Three levels are used in the kernel bundle. Each level is made of kernels of support 8, 4 and 2 respectively. NMI was used as a similarity measure with 64 bins for the histogram. Images were smoothed with a Gaussian of 0.2 mm. The number of time-steps for the fixed time-step registration scheme was set to 16. The desired maximum inverse consistency error (I_{max}) was set to 1 voxel. We set $\lambda_R = 0.3$ and $\lambda_{ICC} = 0.1$. The comparison is made based in Jaccard Index (20) and ICC error (19). Table 3 indicates that both the methods achieve similar overlaps and inverse consistency errors. However, the adaptive time-stepping scheme

[4] www.mindboggle.info

took fewer number of time-steps for integration. Figure 3 illustrates the overlaps based on both the registration schemes and the mean inverse consistency error.

Method	JI	ICC(std)	Max ICC	Max. time steps
Adaptive	0.39	0.16 (0.21)	1.86	3,6,12
Fixed	0.38	1.65 (1.55)	6.78	16

Table 3: Adaptive scheme versus fixed time-step scheme on MGH10 dataset; Max: mean maximum error.

6 Discussion

In this paper, we presented a way of adjusting the time-steps required to integrate a velocity field such that the maximum inverse consistency error is bounded. We proposed a way to adaptively estimate the number of time-steps based on only the velocity field i.e., without having to realize the entire deformation. We used two terms to handle inverse consistency: one for penalization purposes and the other for time-step prediction. Through the first synthetic experiment we showed that the proposed prediction term produced good estimates of the actual inverse consistency error. The second of the synthetic experiments showed that a higher degree of match and a similar inverse consistency can be achieved using the adaptive time-stepping scheme when compared to the fixed-time step version of registration. On the MGH10 dataset, we showed that the proposed registration scheme reaches the same accuracy as the fixed time-step registration scheme with fewer number of time-steps and a similar inverse consistency.

In a recent version of logDemons [3], the zeroth order of BCH was used to compose two velocity fields. However, in this work since we use $\nabla \mathbf{v}_f \mathbf{v}_b$ and $\nabla \mathbf{v}_b \mathbf{v}_f$ to set the number of time-steps, it makes sense to also take them into account in the penalty term. By using the 1^{st} order BCH term, we keep $\mathbf{v}_f$ more close to $-\mathbf{v}_b$ in regions where $\nabla \mathbf{v}_f \mathbf{v}_b$ and/or $\nabla \mathbf{v}_b \mathbf{v}_f$ are large. In that way we limit the magnitude of $\{\mathbf{v}_f, \mathbf{v}_b\}$.

Two other ways of maintaining inverse consistency are: a) To use a higher order integration scheme. Since higher order integration schemes can be computationally expensive, we restricted our analyses to a forward Euler's scheme. b) By solving for the inverse transformation subject to the inverse consistent condition. This can be computationally expensive since it involves solving a high-dimensional linear system of equations and the transformations need to be estimated sequentially.

The proposed method also saves on computational time since it only takes a relevant number of steps based on the desired inverse consistency level and the properties of the velocity field. This implies, the closer the transformation is to being represented as a small deformation, the lower number of time-steps

are taken. That is, a low $\nabla \mathbf{v}_f \mathbf{v}_b$ and/or $\nabla \mathbf{v}_b \mathbf{v}_f$ implies a low n. This is usually the case in the first few iterations of the optimization and the adaptive scheme helps in reducing the computational time by only taking a relevant number of time-steps.

7 Conclusion

In conclusion, we presented a method that lends more flexibility to the transformation model by adaptively setting the time-steps required to account for the discretization error in the numerical integration scheme. The required number of time-steps is chosen based on the properties of the velocity field prior to every optimization iteration. We showed that the Lie brackets can be conveniently represented in terms of the inverse consistency term. We showed that the proposed registration scheme solves the intra-subject registration problem equally well as the inverse consistent fixed time-step scheme by using a fewer number of time-steps.

References

1. Paul A. Yushkevich, Brian B .Avants, Sandhitsu R. Das, John Pluta, Murat Altinay, Caryne Craige, and the Alzheimer's Disease Neuroimaging Initiative. Bias in estimation of hippocampal atrophy using deformation-based morphometry arises from asymmetric global normalization: an illustration in ADNI 3T MRI data. *Neuroimage*, 50(2):434–445, April 2010.
2. G. E. Christensen and H. J. Johnson. Consistent image registration. *IEEE Transactions on Medical Imaging*, 20:568–582, 2001.
3. M. Lorenzi, N. Ayache, G. B. Frisoni, X. Pennec, and Adni. LCC-Demons: A robust and accurate symmetric diffeomorphic registration algorithm. *NeuroImage*, 81:470–483, 2013.
4. Wulf Rossmann. *Lie Groups – An Introduction Through Linear Groups*. Oxford Science Publications, 2002.
5. Akshay Pai, Stefan Sommer, Lauge Sørensen, Sune Darkner, Jon Sporring, and Mads Nielsen. Stepwise inverse consistent euler's scheme for diffeomorphic image registration. *In proceeding of workshop on Biomedical Image Registration*, pages 223–230, 2014.
6. Sai Kit Yeung and Pengcheng Shi. Stochastic inverse consistency in medical image registration. *MICCAI*, 3750:188–196, 2005.
7. A Leow, SC Huang, A Geng, J Becker, S Davis, A Toga, and P Thompson. Inverse consistent mapping in 3d deformable image registration: its construction and statistical properties. In *Information Processing in Medical Imaging*, volume 19, pages 493–503, 2005.
8. Mirza Faisal Beg and Ali Khan. Symmetric data attachment terms for large deformation image registration. *IEEE Transactions on Medical Imaging*, 26(9):1179–1189, 2007.
9. Martin Reuter, H. Diana Rosas, and Bruce Fischl. Highly accurate inverse consistent registration: A robust approach. *Neuroimage*, 53(4):1181–1196, 2010.

10. B. Avants and JC Gee. Geodesic estimation for large deformation anatomical shape averaging and interpolation. *Neuroimage*, 23(1):S139–S150, 2004.
11. Sune Darkner and Jon Sporring. Locally orderless registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012.
12. Laurent Younes. *Shapes and Diffeomorphisms*, volume 171. Springer, 2010.
13. Akshay Pai, Stefan Sommer, Lauge Sørensen, Sune Darkner, Jon Sporring, and Mads Nielsen. Image registration using stationary velocity fields parameterized by norm-minimizing wendland kernel. In *SPIE Medical Imaging*, 2015, (In press).
14. Holger Wendland. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in Computational Mathematics*, 4(1):389–396, 1995.
15. S Sommer, F Lauze, M Nielsen, and X Pennec. Sparse multi-scale dieomorphic registration: the kernel bundle framework. *J. of Mathematical Imaging and Vision*, 46(3):292–308, 2012.

Mixed-effects model for the spatiotemporal analysis of longitudinal manifold-valued data

J.-B. Schiratti[3,1], S. Allassonnière[3], the Alzheimer's Disease Neuroimaging Initiative, O. Colliot[1,2], S. Durrleman[1]

[1] Inria Paris-Rocquencourt, Inserm U1127, CNRS UMR 7225, Sorbonne Universités, UPMC Univ Paris 06 UMRS 1127, Institut du Cerveau et de la Moelle épinière, ICM, F-75013, Paris, France
[2] AP-HP, Pitié-Salpêtrière Hospital, Departments of Neurology and Neuroradiology, F-75013, Paris, France
[3] Ecole Polytechnique, Palaiseau, France

Abstract. In this work, we propose a generic hierarchical spatiotemporal model for longitudinal manifold-valued data, which consist in repeated measurements over time for a group of individuals. This model allows us to estimate a group-average trajectory of progression, considered as a geodesic of a given Riemannian manifold. Individual trajectories of progression are obtained as random variations, which consist in parallel shifting and time reparametrization, of the average trajectory. These spatiotemporal tranformations allow us to characterize changes in the direction and in the pace at which trajectories are followed. We propose to estimate the parameters of the model using a stochastic approximation of the expectation-maximization (EM) algorithm, the Monte Carlo Markov Chain Stochastic Approximation EM (MCMC SAEM) algorithm.
This generic spatiotemporal model is used to analyze the temporal progression of a family of biomarkers. This progression model estimates a normative scenario of the progressive impairments of several cognitive functions, considered here as biomarkers, during the course of Alzheimer's disease. The estimated average trajectory provides a normative scenario of disease progression. Random effects provide unique insights into the variations in the ordering and timing of the succession of cognitive impairments across different individuals.

1 Introduction

Neurodegenerative diseases such as Alzheimer's disease (AD) or Parkinson's disease are known to affect the metabolism, brain structure and cognitive functions. The effect of the disease can be quantified by observing cerebrospinal fluid (CSF), neuroimaging or neuropsychological biomarkers. In [9], Clifford R. Jack et al. proposed an hypothetical model to describe the temporal progression of these biomarkers during the course of the disease. However, there is still a need for data-driven models which could give experimental evidence of such patterns of disease progression. Statistical models for longitudinal data have been subject

to a growing interest in the last few years. In particular, mixed-effects models, which include fixed and random effects, have a hierarchical structure which allows us to describe the model at the group and subjects level.

Still, a statistical model of disease progression should take into account the fact that the age of a given individual is not an indicator of his stage of disease progression. Two individuals, observed at the same age, might actually be at very different stages of disease progression. As a consequence, trajectories should be registered in time to account for this variability in stages of disease progression. In [6], the concept of "time-warps" was introduced to allow for the registration in time of trajectories of shape changes. However, in order to combine the time-warps with the variability of shapes accross individuals, the authors assumed that the variance of shapes does not depend on time whereas it should adapt to the average trajectory of shape changes. The set of the measurements of an individual at a given time-point is often a high-dimensional and nonlinear space. Building a model of disease progression therefore consists in estimating continuous subject-specific trajectories and an average trajectory in this space. At a given time point, the disease progression of two individuals will probably be described by two different trajectories. To construct the average trajectory, the individual trajectories need to be registered in *space*, where *space* may refer to the 2D or 3D space of spatial objects, or more generally to the space of measurements. In [15], time-warps were also used to define a metric between curves which has the property of being invariant under time-reparametrization. The authors did not spatially register the curves because of the small variability of the trajectories.

The framework of mixed-effects models provides tools to deal with this hierarchical problem. Mixed-effects models for longitudinal measurements were introduced in the seminal paper of Laird and Ware [11] and have been widely developed since then (see [4], [13] for instance). It should be pointed out that this kind of models suffers from two main drawbacks regarding our problem. These models describe the distribution of the measurements at a given reference time. In many situations, this reference time is given by the experimental setting: in plant growth studies, the point in time at which the plant was seeded is a natural choice, as well as the date of birth in developmental studies. However, in studies on neurodegenerative diseases, there is no natural choice of reference time as the disease-onset time is most probably different for each individual. Another limitation of usual mixed-effects models is that they are defined for data lying in Euclidean spaces. Although the development of statistical models for manifold-valued data is a blooming topic (see [14], [15]), the construction of statistical models for longitudinal data on a manifold remains an open problem.

In this paper, we propose a statistical model to describe the temporal progression of a family of biomarkers. This progression model can be seen as a particular case of a more general spatiotemporal model for longitudinal manifold-valued data. The Riemannian manifold and its metric are choosen a priori, which allows us to introduce anatomical, physiological constraints into the model. The definition of the generic spatiotemporal model requires no other choice. The models

which we introduce herein are based on the concept of parallel curves on a manifold. The random effects of the model allow to spatially and temporally register individual trajectories of progression. The generic spatiotemporal model belongs to a class of statistical models for which maximum likelihood estimates cannot be obtained in closed form. We address this issue by using a stochastic version of the Expectation-Maximization algorithm [3], namely the MCMC SAEM [1], for which theoretical results regarding the convergence have been proven in [2], [1].

In section 2, we will introduce our propagation model for a family of biomarkers and explain how this model appears as a particular case of a more general mixed-effects model for longitudinal manifold-valued data. We explain how the MCMC-SAEM was used in section 3. The last section consists of experimental results obtained on neuropsychological test scores.

2 Propagation model for a family of biomarkers

2.1 Riemannian geometry prerequisites

Let $(\mathbb{M}, g^{\mathbb{M}})$ be a Riemannian manifold of dimension N equipped with a Riemannian metric $g^{\mathbb{M}}$, which we assume to be geodesically complete. Meaning that the geodesics of $\mathbb{M}$ are defined on $\mathbb{R}$. The Riemannian metric $g^{\mathbb{M}}$ defines a unique affine connexion on $\mathbb{M}$, namely the Levi-Civita connexion, denoted by $\nabla^{\mathbb{M}}$. Let $\boldsymbol{\gamma}$ denote a geodesic of $\mathbb{M}$ and $t_0 \in \mathbb{R}$. We recall that, given a tangent vector $\boldsymbol{\xi}$ in $\mathrm{T}_{\boldsymbol{\gamma}(t_0)}\mathbb{M}$, the parallel transport of $\boldsymbol{\xi}$ along $\boldsymbol{\gamma}$, denoted by $X(s) = \mathrm{P}_{\boldsymbol{\gamma},t_0,s}(\boldsymbol{\xi})$, is a vector field along $\boldsymbol{\gamma}$ which satisfies : $X(t_0) = \boldsymbol{\xi}$ and $\nabla^{\mathbb{M}} X(s) = 0$. Let $\mathbf{p} \in \mathbb{M}$. The Riemannian exponential in $\mathbb{M}$ at $\mathbf{p}$ is denoted by $\mathrm{Exp}^{\mathbb{M}}_{\mathbf{p}}$. For $\mathbf{v} \in \mathrm{T}_{\mathbf{p}}\mathbb{M}$, $\mathrm{Exp}^{\mathbb{M}}_{\mathbf{p}}(\mathbf{v})$ denotes the value at time 1 of the geodesic in $\mathbb{M}$ issued from $\mathbf{p}$ with initial velocity $\mathbf{v}$.

2.2 Model description

We are interested in the temporal progression of a family of N ($N \geq 2$) scalar biomarkers. We consider a longitudinal dataset of the form $(\mathbf{y}_{i,j}, t_{i,j})_{i,j}$, obtained by observing p individuals at repeated time points. The vector $\mathbf{y}_{i,j}$ denotes the j-th observation ($1 \leq j \leq n_i$) of the i-th individual. The k-th coordinate of $\mathbf{y}_{i,j}$, denoted by $y_{i,j,k}$, corresponds to the measurement of the k-th biomarker, at time $t_{i,j}$. We will assume that each measurement $y_{i,j,k}$ belongs to a one dimensional Riemannian manifold (M, g) which is geodesically complete. In this setting, the observations $\mathbf{y}_{i,j} = (y_{i,j,1}, \ldots, y_{i,j,N})$ can be considered as points in the product manifold $\mathbb{M} = M^N$. The average progression of this family of biomarkers is modeled by a geodesic trajectory on the manifold $\mathbb{M}$, which is equipped with the product metric, denoted by $g^{\mathbb{M}}$.

The statistical model is described for observations on a manifold wich is a product of one-dimensional manifolds. This framework is particularly convenient to analyze the temporal progression of a family of biomarkers. The model can be seen as a particular case of a more generic spatiotemporal model (2).

Choice of the average trajectory

In order to determine relative progression of the biomarkers among themselves, the average trajectory is choosen among the parametric family of geodesics : $\left(t \mapsto (\gamma_0(t), \gamma_0(t+\delta_1), \ldots, \gamma_0(t+\delta_{N-1}))\right)_{\boldsymbol{\delta}}$, where $\boldsymbol{\delta} = (0, \delta_1, \ldots, \delta_{N-1})$ and γ_0 is a geodesic, of the one-dimensional manifold M, parametrized by a point p_0 in M, a time t_0 and a velocity v_0 in $\mathrm{T}_{p_0}M$. This parametrization of the geodesic γ_0 is the natural parametrization : $\gamma_0(t_0) = p_0$ and $\dot{\gamma}_0(t_0) = v_0$. By choosing the average trajectory among this parametrized family of geodesics, we assume that, on average, the biomarkers follow the same trajectory but shifted in time. The delay between the progression of the different biomarkers is measured by the vector $\boldsymbol{\delta} = (0, \delta_1, \ldots, \delta_{N-1}) \in \mathbb{R}^N$. The parameters δ_i $(1 \leq i \leq N-1)$ measure a relative delay between two consecutive biomarkers. The parameter t_0 plays the role of reference time as the trajectory of the first biomarker will reach the value p_0 at time t_0 whereas the other trajectories will reach the same value p_0 at different points in time, shifted with respect to t_0.

Construction of subject-specific trajectories of disease progression with time reparametrization

The model proposed herein is a hierarchical model : data points are assumed to be sampled from subject-specific trajectories of progression. These individual trajectories are derived from the average trajectory $\boldsymbol{\gamma_\delta}$. The subject-specific trajectory of the i-th individual is constructed by considering a non-zero tangent vector $\mathbf{w}_i$ in $\mathrm{T}_{\boldsymbol{\gamma_\delta}(t_0)}\mathbb{M}$, orthogonal to $\dot{\boldsymbol{\gamma}}_{\boldsymbol{\delta}}(t_0)$ for the inner product defined by the metric $(\langle \cdot, \cdot \rangle_{\boldsymbol{\gamma_\delta}(t_0)} = g^{\mathbb{M}}_{\boldsymbol{\gamma_\delta}(t_0)})$. This tangent vector $\mathbf{w}_i = (w_{1,i}, \ldots, w_{N,i})$ is a *space shift* which allows us to register the individual trajectories in the space of measurements. The tangent vector $\mathbf{w}_i$ is transported along the geodesic $\boldsymbol{\gamma_\delta}$ from time t_0 to time s using parallel transport. This transported tangent vector is denoted by $\mathrm{P}_{\boldsymbol{\gamma_\delta}, t_0, s}(\mathbf{w}_i)$. At the point $\boldsymbol{\gamma_\delta}(s)$, a new point in $\mathbb{M}$ is obtained by taking the Riemannian exponential of $\mathrm{P}_{\boldsymbol{\gamma_\delta}, t_0, s}(\mathbf{w}_i)$. This new point is denoted by $\eta^{\mathbf{w}_i}(\boldsymbol{\gamma_\delta}, s)$. As s varies, this point describes the curve $s \mapsto \eta^{\mathbf{w}_i}(\boldsymbol{\gamma_\delta}, s)$, which is considered as a "parallel" to the curve $\boldsymbol{\gamma_\delta}$ (Fig. 1). The orthogonality condition on the tangent vectors $\mathbf{w}_i$ is an important hypothesis which ensures that a point $\eta^{\mathbf{w}_i}(\boldsymbol{\gamma_\delta}, s)$ on a parallel moves at the same pace on this parallel than on the average trajectory. This hypothesis ensures the uniqueness of the decomposition between spatial and temporal components.

The trajectory γ_i of the i-th individual is obtained by reparametrizing the parallel $\eta^{\mathbf{w}_i}(\boldsymbol{\gamma_\delta}, \cdot) : \gamma_i(s) = \eta^{\mathbf{w}_i}(\boldsymbol{\gamma_\delta}, \psi_i(s))$, where the mapping $\psi_i(s) = \alpha_i(s - t_0 - \tau_i) + t_0$ is a subject-specific affine reparametrization which allows us to register in time the different individual trajectories of progression. This time-warp was introduced in [12] to define subject-specific trajectories of progression from an average trajectory, in the case of univariate manifold-valued data. In this univariate work, because the manifold is one-dimensional, no random effect is associated to the fixed effect p_0. Here, the tangent vector $\mathbf{w}_i$ can be considered, in the light of the univariate model, as a random effect associated to the point p_0. The effect α_i is an *acceleration factor* which encodes whether the i-th individual

is progressing faster or slower than the average individual. Whereas the effect τ_i encodes the advance or delay of the i-th individual with respect to the average. Both are assumed to be random, non observed, variables as are also the space shifts $\mathbf{w}_i$.

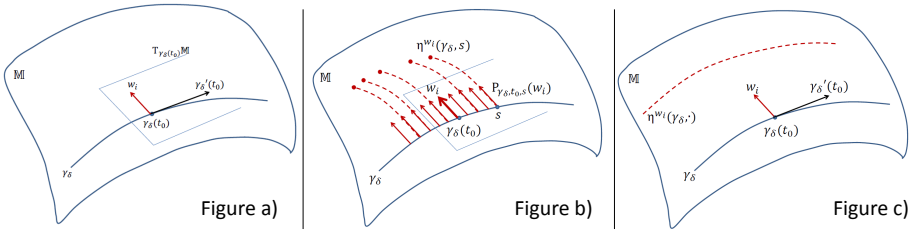


Fig. 1: Schematic description of parallel shifting. Figure a) (left) : A non-zero tangent vector $\mathbf{w}_i$ is choosen in $\mathrm{T}_{\boldsymbol{\gamma}(t_0)}\mathbb{M}$. Figure b) (middle) : the tangent vector $\mathbf{w}_i$ is transported along $\boldsymbol{\gamma}$ from $\boldsymbol{\gamma}(t_0)$ to $\boldsymbol{\gamma}(s)$ using parallel transport on $\mathbb{M}$. A new point, $\eta^{\mathbf{w}_i}(\boldsymbol{\gamma}, s)$ is obtained at time s by shooting with the Riemannian exponential. Figure c) (right) : the curve $\eta^{\mathbf{w}_i}(\boldsymbol{\gamma}, \cdot)$ is the "parallel" to $\boldsymbol{\gamma}$ constructed from $\mathbf{w}_i$.

Because $\mathbb{M}$ is equipped with the product metric, the parallel transport of the tangent vector $\mathbf{w}_i \in \mathrm{T}_{\boldsymbol{\gamma_\delta}(t_0)}\mathbb{M}$ is a N-dimensional vector whose k-th ($1 \leq k \leq N$) component is equal to the parallel transport of the tangent vector $w_{k,i} \in \mathrm{T}_{\gamma_0(\cdot+\delta_{k-1})}M$ along the curve $s \mapsto \gamma_0(s+\delta_{k-1})$ in the one-dimensional manifold M. It follows that $\mathrm{P}_{\boldsymbol{\gamma_\delta},t_0,s}(\mathbf{w}_i) = \big(\frac{w_{1,i}}{\dot{\gamma}_0(t_0)}\dot{\gamma}_0(s), \ldots, \frac{w_{N,i}}{\dot{\gamma}_0(t+\delta_{N-1})}\dot{\gamma}_0(s+\delta_{N-1})\big)$. Taking the Riemanniann exponential, in $\mathbb{M}$, of the tangent vector $\mathrm{P}_{\boldsymbol{\gamma_\delta},t_0,s}(\mathbf{w}_i)$ boils down to taking the Riemannian exponential, in M, of each component of the vector. If Exp^M denotes the Riemannian exponential map in M, the k-th ($1 \leq k \leq N$) component of $\eta^{\mathbf{w}_i}(\boldsymbol{\gamma_\delta}, s)$ is given by : $\mathrm{Exp}^M_{\gamma_0(s+\delta_{k-1})}\big[\frac{w_{k,i}}{\dot{\gamma}_0(t_0+\delta_{k-1})}\dot{\gamma}_0(s+\delta_{k-1})\big] = \gamma_0\big(s+\delta_{k-1}+\frac{w_{k,i}}{\dot{\gamma}_0(t_0+\delta_{k-1})}\big)$. For the longitudinal dataset $(\mathbf{y}_{i,j}, t_{i,j})$ ($1 \leq i \leq p$, $1 \leq j \leq n_i$), our hierarchical model writes : $\mathbf{y}_{i,j} = \gamma_i(t_{i,j}) + \boldsymbol{\varepsilon}_{i,j}$. In particular, for the k-th biomarker, this hierarchical model writes :

$$y_{i,j,k} = \gamma_0\left(\frac{w_{k,i}}{\dot{\gamma}_0(t_0+\delta_{k-1})} + t_0 + \alpha_i(t_{i,j} - t_0 - \tau_i) + \delta_{k-1}\right) + \varepsilon_{i,j,k}. \qquad (1)$$

with $\alpha_i = \exp(\eta_i)$, $\mathbf{w}_i = (w_{1,i}, \ldots, w_{N,i})$, $\mathbf{w}_i = A\mathbf{s}_i$ and :

$$\eta_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\eta^2),\ \tau_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\tau^2),\ \boldsymbol{\varepsilon}_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2 \mathrm{I}_{n_i}) \text{ and } s_{j,i} \overset{\text{i.i.d.}}{\sim} \text{Laplace}(1/2).$$

The parameters of the model are $\boldsymbol{\theta} = (p_0, t_0, v_0, \boldsymbol{\delta}, \sigma_\eta, \sigma_\tau, \sigma, \mathrm{vec}(A))$ and the random effects of the model are $(\alpha_i, \tau_i, \mathbf{w}_i)$ ($1 \leq i \leq p$). Note that the first two random effects are scalars. The acceleration factor is assumed to follow a

log-normal distribution to ensure its positiveness (the affine reparametrization must not reverse time). The time shifts follow a Gaussian distribution with zero mean. The space shifts are vectors of dimension $N-1$ which belong to the vector space $\{\dot{\boldsymbol{\gamma}}_{\boldsymbol{\delta}}(t_0)\}^{\perp}$. In the spirit of independent component analysis [8], we assume that the tangent vectors $\mathbf{w}_i$ are a linear combination of $N_s < N$ statistically independent components. This writes $\mathbf{w}_i = A\mathbf{s}_i$ where A is a $N \times N_s$ matrix of rank N_s whose columns are vectors in $\mathrm{T}_{\boldsymbol{\gamma}_{\boldsymbol{\delta}}(t_0)}\mathbb{M}$ and $\mathbf{s}_i$ is a vector of N_s independent sources following a Laplace distribution with parameter $1/2$. To ensure the orthogonality condition on the tangent vectors $\mathbf{w}_i$, we assume that, for all $j \in \{1, \ldots, N_s\}$, $\langle \mathbf{A}_j, \dot{\boldsymbol{\gamma}}_{\boldsymbol{\delta}}(t_0)\rangle_{\boldsymbol{\gamma}_{\boldsymbol{\delta}}(t_0)} = 0$, where $\mathbf{A}_j$ denotes the j-th column of A.

The model given in (1) can be used to analyze longitudinal observations on any geodesically complete Riemannian manifold. The *generic spatiotemporal model* writes:

$$\mathbf{y}_{i,j} = \eta^{\mathbf{w}_i}(\boldsymbol{\gamma}_{\boldsymbol{\delta}}, \psi_i(t_{i,j})) + \boldsymbol{\varepsilon}_{i,j}. \tag{2}$$

where the parallel $s \mapsto \eta^{\mathbf{w}_i}(\boldsymbol{\gamma}_{\boldsymbol{\delta}}, s)$ is given by:

$$\eta^{\mathbf{w}_i}(\boldsymbol{\gamma}_{\boldsymbol{\delta}}, s) = \mathrm{Exp}^{\mathbb{M}}_{\boldsymbol{\gamma}_{\boldsymbol{\delta}}(s)}\big(\mathrm{P}_{\boldsymbol{\gamma}_{\boldsymbol{\delta}}, t_0, s}(\mathbf{w}_i)\big),\ s \in \mathbb{R}.$$

It should be pointed out that a parallel $s \mapsto \eta^{\mathbf{w}_i}(\boldsymbol{\gamma}_{\boldsymbol{\delta}}, s)$ to the geodesic $\boldsymbol{\gamma}_{\boldsymbol{\delta}}$ is not, in general a geodesic. In the Euclidean case, a parallel to $\boldsymbol{\gamma}_{\boldsymbol{\delta}}$ is just a translation of $\boldsymbol{\gamma}_{\boldsymbol{\delta}}$: $\eta^{\mathbf{w}_i}(\boldsymbol{\gamma}_{\boldsymbol{\delta}}, s) = \boldsymbol{\gamma}_{\boldsymbol{\delta}}(s) + \mathbf{w}_i$.

2.3 The logistic propagation model

If the measurements of the biomarkers can be *normalized*, we can consider these measurements as points in the one-dimensional manifold $M =]0, 1[$. For example, neuropsychological test scores are bounded above by a maximum score and can therefore be normalized to produce measurements in $]0, 1[$. In this case, the model given in (1) can be used to analyze these measurements. We consider that $M =]0, 1[$ is equipped with the Riemannian metric g given by : for $p \in]0, 1[$, $(u, v) \in \mathrm{T}_pM \times \mathrm{T}_pM$, $g_p(u, v) = u\mathrm{G}(p)v$ with $\mathrm{G}(p) = 1/(p^2(1-p)^2)$. For this Riemannian metric, the geodesics are the logistic curves of the form : $\gamma_0(t) = \big(1 + (\frac{1}{p_0} - 1)\exp\big(-\frac{v_0}{p_0(1-p_0)}(t - t_0)\big)\big)^{-1}$. In this setting, (1) writes:

$$y_{i,j,k} = \left(1 + \Big(\frac{1}{p_0} - 1\Big)\exp\left(-\frac{v_0\alpha_i(t_{i,j} - t_0 - \tau_i) + v_0\delta_k + v_0\frac{(A\mathbf{s}_i)_k}{\dot{\gamma}_0(t_0+\delta_k)}}{p_0(1-p_0)}\right)\right)^{-1} + \varepsilon_{i,j,k}, \tag{3}$$

where $(A\mathbf{s}_i)_k$ denotes the k-th component of the vector $\mathbf{w}_i = A\mathbf{s}_i$.

3 Parameters estimation

A stochastic version of the Expectation-Maximization (EM) algorithm [3] is used to estimate the parameters $\boldsymbol{\theta} = (p_0, t_0, v_0, \boldsymbol{\delta}, \sigma_\eta, \sigma_\tau, \sigma, \mathrm{vec}(A))$ of the model.

Because of the nonlinearity of the model, the E step of the EM algorithm is intractable. As a consequence, we consider a stochastic version of the EM algorithm, namely the Monte-Carlo Markov Chain Stochastic Approximation Expectation-Maximization (MCMC-SAEM) algorithm [1], based on [2].

In order to ensure the theoretical convergence of the MCMC SAEM algorithm, the model must belong to the curved exponential family. Equivalently, the complete log-likelihood of the model shall write : $\log q(\mathbf{y}, \mathbf{z} \,|\, \boldsymbol{\theta}) = -\phi(\boldsymbol{\theta}) + S(\mathbf{y}, \mathbf{z})^\top \psi(\boldsymbol{\theta})$, where $S(\mathbf{y}, \mathbf{z})$ is a sufficent statistic of the model. Note that the *logistic propagation model* does not belong to the curved exponential family. A usual workaround consists in regarding the parameters of the model as realizations of independents Gaussian random variables ([10]) : $\boldsymbol{\theta} \sim \mathcal{N}(\overline{\boldsymbol{\theta}}, \mathbf{D})$ where $\mathbf{D}$ is a diagonal matrix with very small diagonal entries and the estimation now targets $\overline{\boldsymbol{\theta}}$. This yields: $p_0 \sim \mathcal{N}(\overline{p_0}, \sigma^2_{p_0})$, $t_0 \sim \mathcal{N}(\overline{t_0}, \sigma^2_{t_0})$, $v_0 \sim \mathcal{N}(\overline{v_0}, \sigma^2_{v_0})$ and, for all k, $\delta_k \sim \mathcal{N}(\overline{\delta}_k, \sigma^2_\delta)$. The matrix A is also considered as a realization of a Gaussian random variable and, in order to ensure the orthogonality condition on the columns of A, we assume that A follows a normal distribution on the space $\boldsymbol{\Sigma} = \{A = (\mathbf{A}_1, \ldots, \mathbf{A}_{N_s}) \in \left(\mathrm{T}_{\gamma_\delta(t_0)}\mathbb{M}\right)^{N_s} ; \forall j, \langle \mathbf{A}_j, \dot{\gamma_\delta}(t_0) \rangle_{\gamma_\delta(t_0)} = 0\}$. Equivalently, we assume that the matrix A writes : $A = \sum_{k=1}^{(N-1)N_s} c_k \mathcal{A}_k$ where, for all k, $c_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\overline{c}_k, \sigma^2_c)$ and $(\mathcal{A}_1, \ldots, \mathcal{A}_{(N-1)N_s})$ is an orthonormal basis of $\boldsymbol{\Sigma}$ obtained by application of the Gram-Schmidt process to a basis of $\boldsymbol{\Sigma}$. The random variables $c_1, \ldots, c_{(N-1)N_s}$ are considered as new hidden variables of the model. The parameters of the model are $\boldsymbol{\theta} = (\overline{p_0}, \overline{t_0}, \overline{v_0}, (\overline{\delta}_k)_{1 \leq k \leq N-1}, (\overline{c}_k)_{1 \leq k \leq (N-1)N_s}, \sigma_\eta, \sigma_\tau, \sigma)$ whereas the hidden variables of the model are $\mathbf{z} = (p_0, t_0, v_0, (\delta_k)_{1 \leq k \leq N-1}, (c_k)_{1 \leq k \leq (N-1)N_s}, (\eta_i)_{1 \leq i \leq p}, (\tau_i)_{1 \leq i \leq p}, (s_{j,i})_{1 \leq j \leq N_s,\ 1 \leq i \leq p})$.

Overview of the MCMC-SAEM algorithm

The MCMC-SAEM itterates, until convergence, between three steps : simulation, stochastic approximation and maximization. Let k be an integer greater than 1 and $\boldsymbol{\theta}^{(k-1)}$ (respectively $\mathbf{z}^{(k-1)}$) denote the parameters (respectively the hidden variables) at the $k-1$-th iteration of the algorithm. The k-th iteration is summarized as follows :

- Simulation : $\mathbf{z}^{(k)}$ is sampled from the transition kernel of an ergodic Markov Chain whose stationary distribution is the conditional distribution of the hidden variables knowing the observations $\mathbf{y} = (\mathbf{y}_{i,j})_{i,j}$ and the current estimates of the parameters $\boldsymbol{\theta}^{(k-1)}$. This sampling is done by using the Hasting-Metropolis within Gibbs sampler.
- Compute the sufficent statistics : we compute the sufficent statistics $\mathbf{S}^{(k)}$.
- Stochastic approximation : because the model belongs to the curved exponential family, the stochastic approximation is done on the sufficent statistics as follows : $\mathbf{S}^{(k+1)} \leftarrow \mathbf{S}^{(k)} + \varepsilon_k(\mathbf{S}(\mathbf{y}, \mathbf{z}^{(k)}) - \mathbf{S}^{(k)})$., where $(\varepsilon_k)_k$ is a decreasing sequence of positive step sizes.
- Maximization : parameters updates are obtained in closed form from the stochastic approximation on the sufficent statistics.

4 Experiments

4.1 Data

We use the neuropsychological assessment test "ADAS-Cog 13" from the ADNI1, ADNIGO or ADNI2 cohorts of the Alzheimer's Disease Neuroimaging Initiative (ADNI). The "ADAS-Cog 13" consists of 13 questions, which allow to test the impairment of several cognitive functions. To each of the 13 items of the test a cognitive function was associated as follows : memory (items $1, 4, 7, 8$ and 9), language (items $2, 5, 10, 11$ and 12), praxies (items 3 and 6), concentration (item 13). The score of each item was normalized by the maximum possible score. Consequently, each data point of each individual consists in thirteen normalized scores, which can be seen as a point on the manifold $\mathbb{M} =]0, 1[^{13}$.

We choose to consider 248 individuals who were diagnosed with mild cognitive impairment (MCI) at their first visit and whose diagnosis changed to AD before their last visit. Among this population, we have an average of 6 visits per individual (with a minimum of 3 and a maximum of 11) and a typical duration of either 6 or 12 months between consecutive visits.

4.2 Experimental results

In this situation where $\mathbb{M} =]0, 1[^{13}$, the number of independent sources N_s can be any integer between 1 and 12. The choice of the number of independent sources influences the number of parameters to be estimated, which equals $9 + 12N_s$. In order to keep a reasonable runtime, we conducted 3 experiments with N_s equal to 1, 2 and 3. For each experiment, the MCMC-SAEM was run five times with different initial parameters. Only the experiment which returned the smallest residual noise variance was kept. Increasing the number of sources allowed to decrease the residual noise among the experiments : $\sigma^2 = 0.02$ for $N_s = 1$, $\sigma^2 = 0.0162$ for $N_s = 2$ and $\sigma^2 = 0.0159$ for $N_s = 3$. Because the residual noise was almost similar for $N_s = 2$ and $N_s = 3$ sources, we choose to report here the results obtained with the less complex model. As a consequence, we report the results obtained with 2 independent sources.

The average trajectory $\boldsymbol{\gamma_\delta}$ is given in Fig. 2, where each curve represents the temporal progression of one specific item of the ADAS-Cog test. The estimated fixed effects are $p_0 = 0.74$, $t_0 = 79.88$ years, $v_0 = 0.047$ unit per year, and $\boldsymbol{\delta} = [0; -14; -11; 4.6; -13; -14; -7.7; -0.9; -14.4; -14.05;$
$-11.80; -15.3292]$ years. This means that, on average, the memory-related items (items $1, 4, 7, 8, 9$) reach the value $p_0 = 0.74$ at respectively $t_0, t_0 - \delta_4, t_0 - \delta_7, t_0 - \delta_8$ and $t_0 - \delta_9$ years, which correponds to respectively $79.88, 75.2, 87.6, 80.7$ and 94.3 years. The concentration item reaches the same value at $t_0 - \delta_{13} = 86.1$ years. The progression of the concentration item is followed by praxis and language items.

Random effects show the variability of this average trajectory within the studied population. The standard deviation of the time-shift equals $\sigma_\tau = 8.3$ years, meaning that the disease progression model in Fig. 2 is shifted in time by

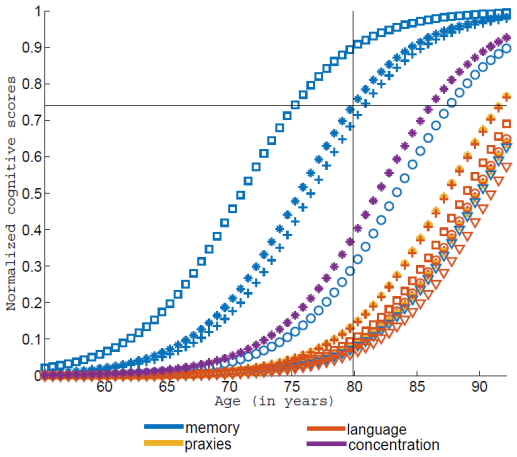


Fig. 2: The estimated average trajectory. In blue: the average trajectory of progression for the 5 memory-related items (item 1:∗, item 4:□, item 7:○, item 8:+ and item 9:△). In orange: average trajectory for the 5 language-related items (item 2:∗, item 5:□, item 10:○, item 11:+ and item 12:△). In yellow: average progression trajectory for the 2 praxies-related items (item 3:∗ and item 6:□). In purple: average progression trajectory for the concentration-related item (item 13:∗).

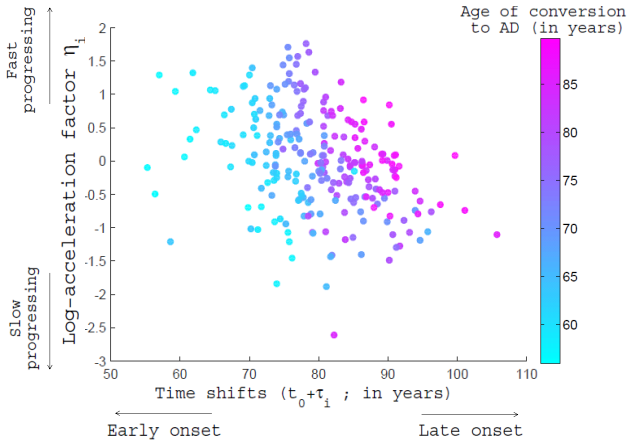


Fig. 3: Plot of the subject-specific random effects: the log-acceleration factor η_i is plotted against the time-shifts $t_0 + \tau_i$. Each point is colored according to the age of conversion to AD.

at most ±8.3 years for 95% of the population. This accounts for the variability in the age of disease onset among the population. The effects of the variance of the acceleration factors, and the two independent components of the space-shifts are illustrated in Fig. 4. The first column of Fig. 4 illustrates the variability in pace of disease progression (the time-shifts are assumed to be zero in order to illustrate the effect of acceleration factor only). This variability is encoded by the variance

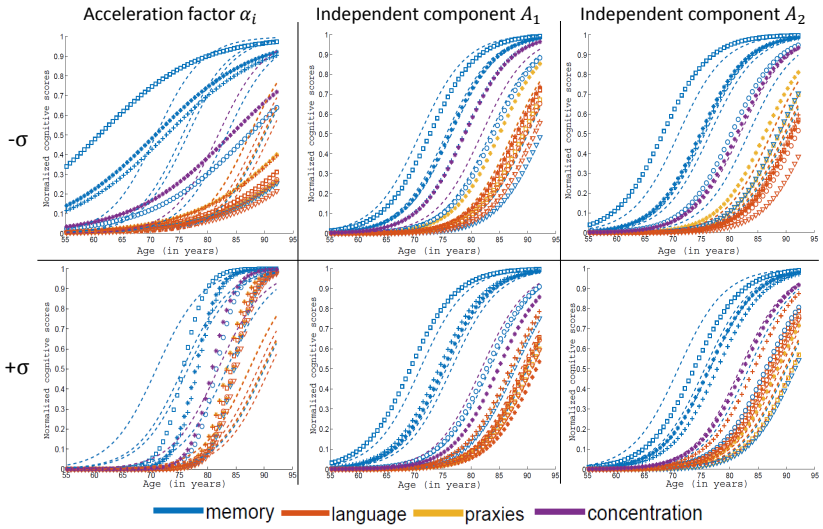


Fig. 4: First column : plot of $t \mapsto \boldsymbol{\gamma_\delta}(\alpha(t - t_0) + t_0)$ with $\alpha = \exp(\pm\sigma_\eta)$ with $\sigma_\eta = 0.8$. Second (resp. third) column : plots of parallels $\eta^{\pm\sigma_{s_i}\mathbf{A}_i}(\boldsymbol{\gamma_\delta}, \cdot)$ in the direction given by the independent components.

$\sigma_\eta = 0.8$ of the acceleration factor. The first and second independent components illustrates the variability in the relative timing of the cognitive impairments. The first independent direction shows that some memory items and language items are shifted in time with respect to the other ones, especially for memory item 4 (□) and item 7 (○). The ordering of the memory item 7 (○) and the concentration item is inverted for individuals with a space shift $\mathbf{w}_i = -\sigma_{s_{i,1}} A_1$. For those individuals, praxies items are impaired later, after the language items 2 (∗), items 12 (△) and item 5 (□). The second independent component shows a greater variability for the memory-related items than for the first independent components, in particular for memory item 9 (△) and item 4 (□). For individuals with a space shift $\mathbf{w}_i = \sigma_{s_{i,2}} A_2$, language-related items might be impaired later than the average individual, especially for the language item 12 (△).

The subject-specific random effects estimated for each individual are obtained from the sampling step of the last iteration of the MCMC-SAEM and are plotted in Fig. 3. The figure shows that the individuals who have a positive (respectively negative) time shift (they are evolving ahead, respectively behind, the average trajectory) are the individuals who converted late (respectively early) to AD. This means that the individual time-shifts correspond well to the age at which a given individual was diagnosed with AD. We also note that there is a negative correlation, equal to -0.4, between the estimated log-acceleration factors and time shifts. There is a tendancy for early onset patients to be fast progressers.

Through its subject-specific affine reparametrization, the age of a given individual is registered to the common timeline of the average scenario. In figure 5 (left), $t \mapsto \sum_i |t_i^{\text{diag}} - \psi_i^{-1}(t)|$ (where t_i^{diag} corresponds to the age at which the

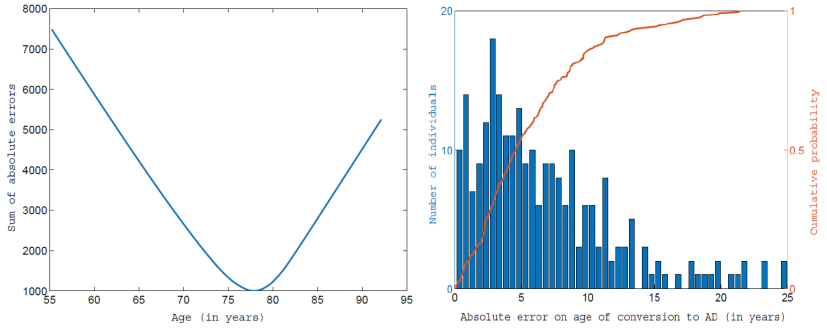


Fig. 5: Left : Sum of $|t_i^{\text{diag}} - \psi_i^{-1}(t)|$ across the 248 individuals as a function of the variable t (time). Right : histogram of the absolute errors $|t_i^{\text{diag}} - 77.45|$.

i-th individual converted to AD) shows a unique minimum at $t^* = 77.45$ years. This age can be understood as the age of symptoms onset in the timeline of the normative scenario of disease progression. The histogram in 5 shows that the age t^* is a prediction of the true age of conversion : the error of prediction is less than 5 years for 50% of the population. This prediction is obtained by analyzing cognitive scores, which are inherently noisy and the reproducibility of these scores is questionable. We believe that the prediction can be improved by analyzing other types of biomarkers which are more objective and indicative of the progression of the disease.

4.3 Discussion and perspectives

We proposed a mixed-effects model to analyze the temporal progression of a family of biomarkers. This model appears as a particular case of a generic spatiotemporal model which can be used to analyze longitudinal manifold-valued measurements. These two models allow to estimate an average trajectory of disease progression. Individual trajectories of disease progression are constructed from the average trajectory by using subject-specific space shifts, acceleration factors and time shifts, which allow to spatially and temporally register the individual trajectories of progression.

The model for a family of biomarkers was used to estimate a scenario of Alzheimer's disease progression from neuropsychological tests. We validated the estimates of the spatiotemporal registration between individual trajectories by the fact that they put into correspondence the age at which patients were diagnosed with Alzheimer's disease. Alternatives to estimate model of disease progression include the event-based model proposed in [7], which estimates the ordering of categorical variables. The combination of spatial and temporal sources of variations in longitudinal data can be futher improved by use of methods such as in [5]. In this work, we introduced the methodolocical background to construct models of disease progression based on longitudinal manifold-valued measurements. Improvements to the model we introduced above consist in an-

alyzing multimodal biomarkers. By doing so, we could experimental evidence, based on a data-driven model, of temporal progression of biomarkers as in [9].

Acknowledgments. The research leading to these results has received funding from the program Investissements davenir ANR-10-IAIHU-06.

References

1. Allassonnière, S., Kuhn, E., Trouvé, A.: Construction of bayesian deformable models via a stochastic approximation algorithm: a convergence study. Bernoulli 16(3), 641–678 (2010)
2. Delyon, B., Lavielle, M., Moulines, E.: Convergence of a stochastic approximation version of the em algorithm. Annals of statistics pp. 94–128 (1999)
3. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the em algorithm. Journal of the royal statistical society. Series B (methodological) pp. 1–38 (1977)
4. Diggle, P., Heagerty, P., Liang, K.Y., Zeger, S.: Analysis of longitudinal data. Oxford University Press (2002)
5. Donohue, M.C., Jacqmin-Gadda, H., Le Goff, M., Thomas, R.G., Raman, R., Gamst, A.C., Beckett, L.A., Jack, C.R., Weiner, M.W., Dartigues, J.F., Aisen, P.S., the Alzheimer's Disease Neuroimaging Initiative: Estimating long-term multivariate progression from short-term data. Alzheimer's & Dementia 10(5), 400–410 (2014)
6. Durrleman, S., Pennec, X., Trouvé, A., Braga, J., Gerig, G., Ayache, N.: Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data. International Journal of Computer Vision 103(1), 22–59 (2013)
7. Fonteijn, H.M., Modat, M., Clarkson, M.J., Barnes, J., Lehmann, M., Hobbs, N.Z., Scahill, R.I., Tabrizi, S.J., Ourselin, S., Fox, N.C., et al.: An event-based model for disease progression and its application in familial alzheimer's disease and huntington's disease. NeuroImage 60(3), 1880–1889 (2012)
8. Hyvärinen, A., Karhunen, J., Oja, E.: Independent component analysis, vol. 46. John Wiley & Sons (2004)
9. Jack, C.R., Knopman, D.S., Jagust, W.J., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., Trojanowski, J.Q.: Hypothetical model of dynamic biomarkers of the alzheimer's pathological cascade. The Lancet Neurology 9(1), 119–128 (2010)
10. Kuhn, E., Lavielle, M.: Maximum likelihood estimation in nonlinear mixed effects models. Computational Statistics & Data Analysis 49(4), 1020–1038 (2005)
11. Laird, N.M., Ware, J.H.: Random-effects models for longitudinal data. Biometrics pp. 963–974 (1982)
12. Schiratti, J.B., Allassonnière, S., Routier, A., the ADNI, Colliot, O., Durrleman, S.: A mixed-effcts model with time reparametrization for longitudinal univariate manifold-valued data. In: Information Processing in Medical Imaging. pp. 564–576. Springer (2015)
13. Singer, J.D., Willett, J.B.: Applied longitudinal data analysis: Modeling change and event occurrence. Oxford university press (2003)
14. Singh, N., Hinkle, J., Joshi, S., Fletcher, P.T.: A hierarchical geodesic model for diffeomorphic longitudinal shape analysis. In: Information Processing in Medical Imaging. pp. 560–571. Springer (2013)
15. Su, J., Kurtek, S., Klassen, E., Srivastava, A., et al.: Statistical analysis of trajectories on riemannian manifolds: Bird migration, hurricane tracking and video surveillance. The Annals of Applied Statistics 8(1), 530–552 (2014)

Geodesic Refinement Using James-Stein Estimators

Greg M. Fleishman[1,2], P. Thomas Fletcher[3], Boris A. Gutman[2], Gautam Prasad[2], Yingnian Wu[4], and Paul M. Thompson[2]

[1] UC Los Angeles, Department of Bioengineering
[2] Imaging Genetics Center, University of Southern California
[3] Scientific Computing and Imaging Institute, University of Utah
[4] UC Los Angeles, Department of Statistics

Abstract. In longitudinal imaging studies, geodesic regression in the space of diffeomorphisms [9] can be used to fit a generative model to images over time. The parameters of the model, primarily its initial direction or momentum, are important objects for study that contain biologically meaningful information about the dynamics occurring in the underlying anatomy. Unfortunately, it is common for any given subject to have a very limited number of longitudinal images available, the acquisition of which is corrupted by noise and variability due to scanning conditions. Furthermore, the underlying anatomy is subject to many entangled biological processes, the effect of which on images is in many cases poorly characterized. Hence, the approach must fit a model to few data points with uncharacterized variability. Here, we propose supplementing the lack of longitudinal information for an individual patient with information that can be extracted cross-sectionally from a population of time series to improve the model fit for the individual. To that effect, we propose a probabilistic model that leads to a well established technique from classical statistics: James-Stein estimators. We show that recent work on groupwise registration for improved geodesic estimation is a sub-optimal special case of our proposed model. Finally, we validate the model by showing geodesics refined by the James-Stein estimator extrapolate more accurately on average than raw geodesic estimates.

1 Introduction

In the large deformation diffeomorphic metric mapping (LDDMM) framework for nonlinear image registration [1], interpolation and extrapolation of longitudinal image time series can be accomplished with geodesic regression [9]. In this setting, a geodesic on a manifold of diffeomorphisms is estimated such that it passes maximally close to transformations that optimally map the initial image to all subsequent images in the time series. The geodesic is parameterized by an initial transformation (here fixed at the identity for simplicity) and a single vector field (tangent to the manifold at the identity), which specifies the direction of the geodesic. If one assumes this vector field is everywhere proportional

to the initial image gradient [8], then the geodesic is fully specified by a single scalar-valued image, henceforth referred to as the momentum. The task of geodesic regression can then be formulated as: given the time series of images $I_1(x), ..., I_N(x)$, find the momentum $p(x)$ such that the geodesic parameterized by $p(x)$ passes through $\phi_2(x), ..., \phi_N(x)$ and $\sum_{i=2}^{N} d(I_1 \circ \phi_i, I_i)^2$ is minimal; where $d(I, J)$ is some quantitative assessment of similarity between images I and J.

As in any learning task, our confidence in the ability of the geodesic model to make accurate predictions at unobserved time points increases with the number of observations. Unfortunately however, due to the high cost of collecting anatomical images, many longitudinal studies of brain structure collect images at fewer than 5 time-points per individual, and often at relatively small time intervals. The short time intervals are particularly problematic considering the slow dynamics of many neurodegenerative diseases. Such a small number of observations, prone to noise, over a short time interval may be insufficient to fit a geodesic that generalizes to unobserved time points with an acceptable level of confidence. We address the challenge of improving geodesic model generalization for an individual time series by pooling information from multiple time series cross-sectionally, and using it to regularize the individual geodesic models. Such an approach may have practical implications on study design, wherein a researcher may choose to acquire fewer images over a shorter time period from more individuals, and yet achieve similar confidence in the accuracy of individual geodesics had they collected more images over a longer period of time from fewer individuals.

We find a natural mathematical setting to implement this in the James-Stein estimator. The James-Stein estimator is a classical statistical model that improves upon the maximum-likelihood estimate for the mean of a Gaussian random variable. That is, the James-Stein estimator is closer in Euclidean distance on average to the unobserved ground truth value of the mean than its maximum-likelihood estimate. James-Stein estimators are commonly used for massively parallel data sets where the same inference must be made for many samples. Information is pooled across the samples and used to regularize the inference of each individual sample. This model reflects the case in neuroimaging where only short sparsely sampled time series are available but for many patients. Using James-Stein estimates as opposed to maximum-likelihood estimates can offer substantial improvements on model accuracy on average [3]. We utilize the James-Stein estimator to leverage the information contained cross-sectionally in a population of time series to improve the geodesic fit for each individual time series.

A necessary first step for James-Stein estimators is to estimate a groupwise representation of the samples. Several recent works have proposed methods for constructing a groupwise representation of image time series data, any of which is compatible with our proposal. In [2] the authors proposed a method to register time series of images in both space and time simultaneously; a groupwise representation of the time series, or spatiotemporal atlas, can then be found in the common spatiotemporal coordinate system. In [10] the authors propose a

hierarchical geodesic model in which individual geodesics are estimated, then used to construct a groupwise geodesic. Their proposed probabilistic model allows an extension that is not fully explored in [10], which is to re-estimate the individual geodesics after the groupwise representation has been constructed. If the groupwise representation is used as a prior (which is suggested by the probabilistic model), the new estimates are similar to the James-Stein estimates for the individual trajectories. The James-Stein estimator shows how to do this second inference optimally.

After a groupwise representation is obtained, James-Stein estimators shrink individual estimates toward the groupwise representation. We show below that this is in fact a maximum *a posteriori* (MAP) estimate, where the shape of the prior distribution is inferred from the data itself. This can also be viewed as a groupwise consistency constraint. Other recent works have proposed groupwise consistency to cope with difficulty in estimation of individual models. In [12], the authors propose a hierarchical Markov random field (hMRF) for segmentation of structural MRI images into functional networks based on fMRI time series. Individual segmentations are constrained to be smooth and consistent with the fMRI data for that individual. They are also constrained to be similar to a grouplevel representation of the network which is jointly estimated with the individual networks. The authors show that this cross-sectional constraint improves the recovery of networks in ficticious data and results in smoother networks with more anatomical meaning in real data.

Similarly, in [4], pairs of longitudinal brain images from a population of individuals diagnosed with Alzheimer's disease (AD) were registered simultaneously. The optimal set of transformations was defined not only to map the template images to their references, but also to satisfy a groupwise consistency constraint. The authors showed that the resulting geodesics predicted a third time point image not used in the learning step more accurately on average than geodesics learned without the groupwise consistency constraint. We demonstrate below that their approach is in fact a special case of James-Stein estimators. Establishing the connection with James-Stein estimators grounds that work in a probabilistic model from classical statistics that provides better intuition for the meaning of parameters and how to find their optimal values.

2 Methods

2.1 Derivation of the multivariate James-Stein estimator for momenta: For simplicity, we consider time series with two images. Because the derivation of James-Stein estimators will deal exclusively with momenta, the generalization to time series of arbitrary length is trivial. Let I_i and J_i for $i \in \{1, ..., N\}$ be initial and follow up image acquisitions of the same anatomy for N patients. In order to share information cross sectionally we must have a common coordinate system. So we also assume we're given transformations ψ_i such that $I_i(\psi_i) \sim I_j(\psi_j)$ for all i and j. This can be accomplished by finding a study specific atlas, or minimal deformation template (MDT), for the images I_i. All further formula are assumed

to be in the common coordinate system. (That is, all momenta have been moved to the common coordinate system by co-adjoint transport, which for the scalar field p_i is $D\psi_i \cdot p_i(\psi_i)$, where D is the Jacobian operator.)

Now, suppose p_i specifies a geodesic beginning at identity and passing through an optimal ϕ_i such that $I_i(\phi_i) \sim J_i$ for all i. The true values of the p_i are unknown, but let β_i be a noisy estimate of p_i acquired via geodesic regression. Now, suppose the following probabilistic model:

$$p_i \sim \mathcal{N}(p^*, A), \tag{1}$$

$$\beta_i | p_i \sim \mathcal{N}(p_i, \sigma_0^2 \cdot \mathrm{Id}). \tag{2}$$

Equation (1) indicates the unobservable p_i are independent samples from a normal distribution with mean p^* and covariance A. This distribution models the variability in time series trajectory across individuals due to differing contributions of the underlying processes that affect the dynamics of aging and disease. The mean momentum parameterizes a geodesic representing the average dynamics over time for images in the population. (Hence, any one of the previously discussed methods for construction of a groupwise representation of time series [2, 10] can be taken as a definition for p^*.)

Equation (2) indicates the observable β_i are independent samples from a normal distribution with mean p_i and covariance $\sigma_0^2 \cdot \mathrm{Id}$, where Id is the matrix identity of the appropriate size. This distribution models the variability of the momentum measurement β_i due to image noise and registration inaccuracies. Hence, each β_i is distributed about its (unobserved) ground truth value of p_i with isotropic variability, the extent of which is given by σ_0^2. This is consistent with standard noise assumptions in much of the image registration literature.

These distributions have the form of a prior and likelihood, which enables us to write the posterior distribution for the p_i:

$$P(p_i|\beta_i) = \frac{P(\beta_i|p_i)P(p_i)}{\int P(\beta_i|p_i)P(p_i)dp_i} = \mathcal{N}(\beta_i - \sigma_0^2 B(\beta_i - p^*),\ \sigma_0^2 B), \tag{3}$$

where $B = (A + \sigma_0^2 \cdot \mathrm{Id})^{-1}$. We see from (3) that the MAP estimate of p_i is:

$$p_i^{map} = \beta_i - \sigma_0^2 B(\beta_i - p^*). \tag{4}$$

Equation (4) reveals what we gain by incorporating (1) as a prior to regularize β_i. We see that p_i^{map} is equal to the measurement β_i minus an adjustment: $\sigma_0^2 B(\beta_i - p^*)$. The adjustment is a linear transformation of the difference vector $\beta_i - p^*$. If that transformation were the identity, this would simply move β_i toward p^*. However, the linear transformation is actually the covariance matrix of the posterior distribution: $\sigma_0^2 B$. Hence, (4) begins with the idea of moving β_i toward p^*, but takes into account the shapes of the prior and likelihood distributions and adjusts the direction in which we move the estimate accordingly. The net affect is the rearrangment of the observations β_i such that the scatter of the p_i^{map} is more consistent with the prior covariance structure A. Assuming the prior (1)

is correct, p_i^{map} is guaranteed to be a better estimate of p_i on average than the original measurement β_i [7, 3].

Unfortunately, we cannot use (4) directly, as σ_0^2, p^* and A are unknown. However, with N independent parallel time series at our disposal, we can estimate them directly from the data. First we observe the marginal distribution for β_i:

$$P(\beta_i) = \int P(\beta_i|p_i)P(p_i)dp_i = \mathcal{N}(p^*,\ A + \sigma_0^2 \cdot \mathrm{Id}) \tag{5}$$

The maximum likelihood estimate for the mean of a Gaussian random variable is the sample mean. Hence, the maximum likelihood estimate for p^* is simply: $p^* \sim \hat{\beta} = \frac{1}{N}\sum_{i=1}^{N}\beta_i$. Next, we define the sample covariance matrix for the β_i as: $S = \sum_{i=1}^{N}(\beta_i - \hat{\beta})(\beta_i - \hat{\beta})^T$. Because β_i is a random variable, so too is S; which hence must have a corresponding distribution. In fact, the sample covariance matrix of a multivariate normal random variable (such as β_i) is distributed by the Wishart distribution, a multivariate analog of the χ^2 distribution. We now observe:

$$E\{(N-d-1)\sigma_0^2 S^{-1}\} = \sigma_0^2 B \tag{6}$$

where d is the dimensionality of β_i and the expectation is taken with respect to the Wishart distribution. From (6) then, we see that $(N-d-1)\sigma_0^2 S^{-1}$ is the maximum likelihood estimate for $\sigma_0^2 B$. Combining this with $\hat{\beta}$ (the maximum likelihood estimate for p^*) and equation (4) we arrive at the James-Stein estimator for image time series momenta:

$$p_i^{js} = \beta_i - (N-d-1)\sigma_0^2 S^{-1}(\beta_i - \hat{\beta}). \tag{7}$$

The final ingredient is to estimate σ_0^2. Recall, in this model σ_0^2 does not model any biological variability, which is entirely captured by the prior covariance A in (1). σ_0^2 is the noise in the β_i estimates exclusively due to image noise and registration inaccuracy. Hence, any method for estimating the variability due to noise and registration inaccuracy can be used to estimate σ_0^2.

We note here that if we let d be the number of image voxels (the naive dimensionality of β_i), it is almost certain for image analysis applications that $d >> N$, which is generally prohibited if equation (8) is to be useful. Furthermore if $d >> N$, S is certain to be singular and therefore the estimation of S^{-1} becomes problematic. This is the crux issue to be dealt with if one wants to use p_i^{js} for the proposed application. Below, we make the simplest (and least informative) assumption to contend with this issue and then discuss alternatives that might improve the framework.

2.2 Connection to groupwise registration with similarity constraint: To incorporate cross sectional information into the registration of a population of N time series, recent works [4] proposed an objective function of the form:

$$\alpha\mathcal{P}[\phi_1, ..., \phi_N] + \sum_{i=1}^{N}\mathcal{D}[J_i, I_i[\phi_i]] + \gamma\mathcal{S}[\phi_i] = \min \tag{8}$$

Here, the typical image similarity term $\mathcal{D}$ and smoothing prior $\mathcal{S}$ are summed over the N pairs of images. The objective is augmented by a new term $\mathcal{P}$ that is a function of the full set of N transformations, or in the diffeomorphic case, of the estimated transformation momenta in MDT coordinates β_i. Specifically, for $\mathcal{P}$ those works proposed:

$$\mathcal{P}[\beta_1, ..., \beta_N] = \sum_{i=1}^{N} \|\beta_i - \hat{\beta}\|^2 \tag{9}$$

which is the sum of squared difference of the N momenta from their sample average. The Euler-Lagrange equations for this term are: $\nabla_{\beta_i}\mathcal{P}[\beta_1, ..., \beta_N] = 2\alpha(\beta_i - \hat{\beta})$ such that at every iteration the estimate for β_i is updated according to:

$$\beta_i^{t+1} = \beta_i^t - 2\alpha(\beta_i - \hat{\beta}) - \nabla_{\beta_i}\mathcal{D} - \nabla_{\beta_i}\mathcal{S} \tag{10}$$

The first two terms in equation (10) are very similar to equation (4). In fact, if B in (4) were proportional to the identity matrix then the first two terms in (10) would be identical to (4): a shrinkage of the estimate β_i directly toward $\hat{\beta}$ proportional to some scalar value. B is proportional to the identity if and only if A in (1) is proportional to the identity. This reveals two things: the simultaneous registration with groupwise consistency is equivalent to using p_i^{js} with an isotropic prior distribution instead of β_i at every iteration, and that the parameter α in (10) is a function of A and σ_0^2. The perspective of James-Stein estimators thus enables us to generalize the groupwise consistency to anisotropic prior structures and provides an interpretation of the groupwise consistency parameter α.

3 Experiments and Results

3.1 Images: We downloaded screening, 1 year follow up, and 2 year follow up 1.5 Tesla T1-weighted images for 57 participants in the Alzheimer's Disease Neuroimaging Initiative (ADNI). All 57 participants had been diagnosed with Alzheimer's Disease (AD) prior to the acquisition of their screening image. The population consisted of 32 males mean age 75.91 +/- 7.85 years and 25 females mean age 75.08 +/- 8.15 years. This was the maximum number of individuals we could download from the ADNI 1 cohort that were in the AD group and had screening, year 1, and year 2 follow up images available. All images were corrected for geometric distortion and bias in the static field with GradWarp and N3 before downloading as part of the ADNI preprocessing protocol. Subsequent to downloading, the images were linearly registered to the ICBM template and skull stripped using ROBEX [6]. Transformations ψ^i mapping the template images I_i into a MDT coordinate system were computed using a preexisting implementation of [13]. We then registered the screening (I_i) to the follow up images (J_i) to acquire the β_i using our own implementation of the geodesic

shooting algorithm proposed in [11].

3.2 Experimental design: The multivariate James-Stein estimator, equation (7), presents some computational challenges for image data. The full image resolution for most image data sets (a total of d voxels) is very large. Hence S and S^{-1} may be computationally intractable to compute or store. The easiest way to avoid this problem is to assume A and thus S and S^{-1} are proportional to the identity. In that case, the coefficient in front of the second term in (7) reduces to a scalar value:

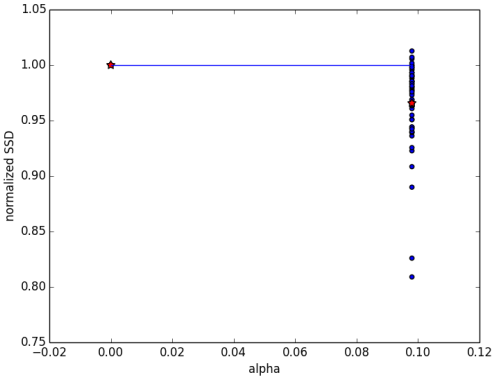
$$p_i^{js} = \beta_i - \alpha(\beta_i - \hat{\beta}) \tag{11}$$

The scalar α can then be estimated empirically using cross-validation, which is what we've done for our first tier validation experiments. This assumption is permitted in the context of James-Stein estimators, and more accurate assumptions about the prior structure can only improve results. More elegant solutions that would allow for anisotropic prior densities are explored in the discussion.

3.3 Results: Using the empirically determined value $\alpha = 0.098$, we computed p_i^{js} according to equation (11). We then compared the ability of the β_i and the p_i^{js} to predict the year 2 follow up images (K_i) by extrapolating their geodesics forward to the year 2 time point and composing the initial image I_i with the resulting transformations. This produced two predictions for each K_i, which we label K_i^{β} and K_i^{js} respectively. We calculated the square Euclidean distances $d(K_i, K_i^{\beta})^2$ and $d(K_i, K_i^{js})^2$ between the ground truth year 2 images and those predictions. In Figure 1 we present $\frac{d(K_i,K_i^{js})^2}{d(K_i,K_i^{\beta})^2}$ for all 57 patients.

Figure 1 shows that by measure of sum of squared differences, the p_i^{js} make better predictions of the third time point image for nearly all patients by about 5% on average. In the best case, an improvement of 20% is achieved. We also subjected the differences $d(K_i, K_i^{\beta})^2 - d(K_i, K_i^{js})^2$ to a pairwise one sided Student's t-test to evaluate the likelihood of achieving these improvements by chance. The p-value of 0.0002 suggests that these results are significant, and that the improvements are due to the use of the James-Stein estimates.

We also inspected the predicted images K_i^{β} and K_i^{js} for any qualitative differences. While the majority of gains due to p_i^{js} are spread thinly throughout the whole image, some improvements clearly correspond to an anatomical interpretation. Figure 2 shows one such case, where β overestimates the expansion of the posterior horn of the left ventricle. The top row is the time series of images I, J, and K from left to right. The bottom row are the predictions corresponding to β and p^{js}. The heat map shows $|\frac{K-\mu}{\sigma} - \frac{K^{js}-\mu^{js}}{\sigma^{js}}| - |\frac{K-\mu}{\sigma} - \frac{K^{\beta}-\mu^{\beta}}{\sigma^{\beta}}|$. That is, it is the difference of the absolute values of the difference images, normalized to their own intensity distributions. This reveals, in cool colors, the locations where p^{js} provided a better estimate of K. The boxed areas show β overestimates the expansion of the ventricle more severely than p^{js}.



Student T test results for $d(K_i, K_i^{\beta})^2 - d(K_i, K_i^{js})^2$

μ	187.23
σ	341.84
T	4.135
p	**.0002**

Fig. 1. Square euclidean distance between ground truth year 2 images and predictions made with p_i^{js} for $\alpha = 0.098$. For each i the distance is normalized by the distance between the ground truth year 2 image and the prediction made with the unrefined β_i. This reveals (by the distance under the red line) the percent improvement earned by using p_i^{js} instead of β_i. The pairwise one sided student's T test shows the improved predictions are due to the use of p_i^{js}.

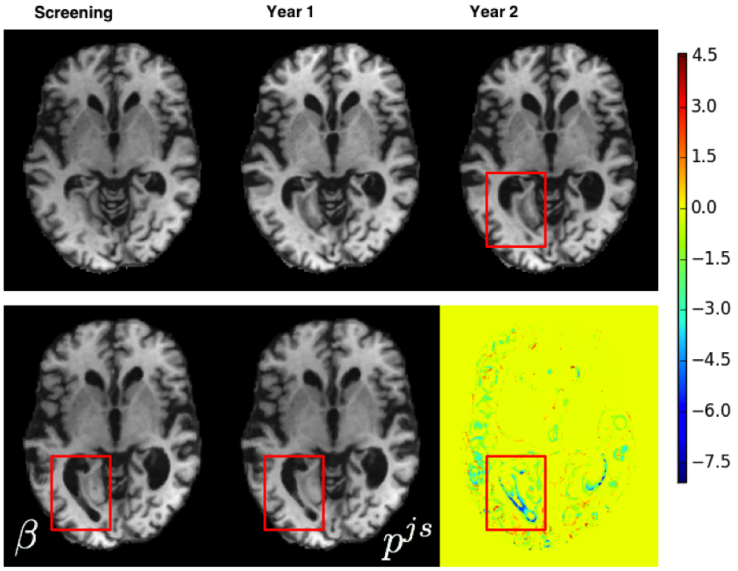


Fig. 2. A time series of images from one patient is shown in the top row. The predictions for the year 2 image derived from β and p^{js} are in the bottom row. The heat map shows in cool colors areas where the p^{js} improved the prediction over β. For this patient, p^{js} reduced an over estimation of ventrical expansion.

4 Discussion

Consistent with expectations, the results indicate the James-Stein estimates p_i^{js} provide geodesics that extrapolate more accurately on average. Hence, our choice of an isotropic prior covariance (that is, $A = a \cdot \mathrm{Id}$ for some scalar a) to cope with the high dimensionality of the β_i is sufficient to gain some improvement in trajectory estimates. A more accurate prior model can only provide more information to improve results.

The simplest relaxation is to let A be diagonal but not necessarily proportional to the identity. In that case, we only have to estimate d variables, an independent variance at every voxel. Different parts of the brain are more or less likely to change over time depending on age and pathology, hence this is more biologically plausible than $A = a \cdot \mathrm{Id}$. More plausible still is to allow A to be non diagonal, but assume that it is sparse. The spatial dependence between voxels is likely to fall off after some appropriate distance, hence many entries in A are likely to be zero or near zero. In that case, many recent methods for learning with sparsity constraints may be brought to bear.

Possibly the most elegant solution would be to use a low dimensional parameterization for the β_i. One option would be to use a subset of the principal components. First one would have to determine an optimal number of components that retains the fine scale variability inherent to longitudinal deformations while reducing the dimension to an acceptable level. A second possibility is to use a band limited Fourier basis. It was recently shown that geodesics for cross-sectional image registration can be parameterized with as few as eight Fourier coefficients per spatial dimension without compromising registration accuracy [14].

Above, we used the estimate $p^* \sim \hat{\beta} = \frac{1}{N}\sum_{i=1}^{N} \beta_i$, which is the maximum likelihood estimate of p^* under the marginal distribution for β_i. However, for many groupwise representations of time series, p^* is a function of time. Hence, the β_i would need to be normalized in time (as well as in space) before averaging. Similarly when computing p_i^{js}, which involves a term $(\beta_i - p^*)$, p^* should be normalized in the time domain to β_i. Propagating p^* along a geodesic is a simple matter of parallel transport, however finding the appropriate correspondence in time between subjects is not trivial. The naive solution is to use nominal time, however aging and pathological effects may not have constant velocity in time. Also, the age of onset of pathological affects is not known for most patients. Hence, a method that infers temporal correspondence directly from the data independent of the acquisition times of the images such as those in [5, 2] would be needed.

5 Conclusion

We have presented the derivation of multivariate James-Stein estimators in the context of image time series regression. We have established a previously pub-

lished method as a sub-optimal special case of the current model. Further, we have demonstrated that the use of James-Stein estimators can improve the extrapolation of individual geodesics in a population of time series, even with the most naive prior structure. We conclude that for the purpose of interpolation and extrapolation of individual time series within a population, the James-Stein estimate of the geodesic is a more accurate representation of the underlying biological dynamics than the raw measurement.

References

1. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. Int. J. Comput. Vision 61(2), 139–157 (Feb 2005)
2. Durrleman, S., Pennec, X., Trouvé, A., Gerig, G., Ayache, N.: Spatiotemporal atlas estimation for developmental delay detection in longitudinal datasets. pp. 297–304. MICCAI '09, Springer-Verlag, Berlin, Heidelberg (2009)
3. Efron, B.: Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction. Institute of Mathematical Statistics Monographs, Cambridge University Press, Leiden (2010)
4. Fleishman, G.M., Gutman, B.A., Fletcher, P., Thompson, P.M.: Simultaneous longitudinal registration with group-wise similarity prior. In: Ourselin, S., Alexander, D.C., Westin, C.F., Cardoso, M.J. (eds.) Information Processing in Medical Imaging, Lecture Notes in Computer Science, vol. 9123, pp. 746–757 (2015)
5. Hong, Y., Singh, N., Kwitt, R., Niethammer, M.: Time-warped geodesic regression. In: Medical Image Computing and Computer-Assisted Intervention MICCAI 2014, Lecture Notes in Computer Science, vol. 8674, pp. 105–112. Springer International Publishing (2014)
6. Iglesias, J., Liu, C., Thompson, P., Tu, Z.: Robust brain extraction across datasets and comparison with publicly available methods. IEEE Transactions on Medical Imaging 30(9), 1617–1634 (2011)
7. James, W., Stein, C.: Estimation with quadratic loss. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics. pp. 361–379. University of California Press, Berkeley, Calif. (1961)
8. Miller, M.I., Trouvé, A., Younes, L.: Geodesic shooting for computational anatomy. Journal of Mathematical Imaging and Vision 24(2), 209–228 (2006)
9. Niethammer, M., Huang, Y., Vialard, F.X.: Geodesic regression for image time-series. In: Fichtinger, G., Martel, A.L., Peters, T.M. (eds.) MICCAI (2). Lecture Notes in Computer Science, vol. 6892, pp. 655–662. Springer (2011)
10. Singh, N., Hinkle, J., Joshi, S., Fletcher, P.: A hierarchical geodesic model for diffeomorphic longitudinal shape analysis. In: Proceedings of the International Conference on Information Processing in Medical Imaging (IPMI). Lecture Notes in Computer Science (LNCS) (2013)
11. Vialard, F.X., Risser, L., Rueckert, D., Cotter, C.J.: Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation. Int. J. Comput. Vision 97(2), 229–241 (Apr 2012)
12. Wei, L., Awate, S., Anderson, J., Fletcher, T.: A functional network estimation method of resting-state fMRI using a hierarchical Markov random field. Neuroimage In press (June, 2014)

13. Yanovsky, I., Thompson, P.M., Osher, S., Leow, A.D.: Topology preserving log-unbiased nonlinear image registration: Theory and implementation. In: CVPR. IEEE Computer Society (2007)
14. Zhang, M., Fletcher, P.: Finite-dimensional lie algebras for fast diffeomorphic image registration. In: Proceedings of the International Conference on Information Processing in Medical Imaging (IPMI). Lecture Notes in Computer Science (LNCS) (2015)

Barycentric Subspaces Analysis on Spheres

Xavier Pennec[1]

Inria Sophia-Antipolis and Côte d'Azur University (UCA), France
xavier.pennec@inria.fr

Abstract. This paper addresses the generalization of Principal Component Analysis (PCA) to Riemannian manifolds. Current methods like Principal Geodesic Analysis (PGA) and Geodesic PCA (GPCA) minimize the distance to a "Geodesic subspace". This allows to build sequences of nested subspaces which are consistent with a forward component analysis approach. However, these methods cannot easily be adapted to a backward analysis and alck symmetric in the parametrization of the subspaces. We propose in [10] a new and more general type of family of subspaces in manifolds, *barycentric subspaces*, which are implicitly defined as the locus of points which are weighted means of $k+1$ reference points. Depending on the generalization of the mean that we use, we obtain the Fréchet / Karcher / Exponential barycentric subspaces (FBS / KBS / EBS). The completion of the last one is called the affine span. These definitions were shown to define locally submanifolds of dimension k.
In this paper, we investigate barycentric subspaces in one of the simplest manifold: the sphere. We show that the affine span is a great subsphere in generic conditions, i.e. also a geodesic subspace. This coincidence of spaces is due to the very high symmetry of the sphere. For second order jets, we show that we obtain subspheres of different radii as in the case of principal nested spheres (PNS) analysis. Among the points of the affine span, determining which ones belong to the Karcher barycentric subspaces (KBS) turns out to be a surprisingly difficult problem. Practical experiments show that the KBS covers in general only a small portion of the subsphere containing the reference points. This suggests that the affine span might be a much more interesting definition to work with for subspace definition purposes. We finally discuss the use to these barycentric subspaces to generalize PCA on manifolds, a procedure that we name Barycentric Subpsace Analysis (BSA). Like PGA, barycentric subspaces allow the construction of a forward nested sequence of subspaces which contains the Fréchet mean. However, the definition also allows the construction of backward nested sequence which may not contain the mean.

1 Introduction

In a Euclidean space, the principal k-dimensional affine subspace of the Principal Component Analysis (PCA) procedure is equivalently defined by minimizing the variance of the residuals (the projection of the data point to the subspace) or by maximizing the explained variance within that affine subspace. This is due to

Pythagoras' theorem, which does not hold in more general manifolds. A second important observation is that principal components or different orders are nested, which allows to build forward and backward estimation methods by iteratively adding or removing principal components.

Generalizing PCA to manifolds first requires to define the equivalent of affine subspaces in manifolds. For the zero-dimensional subspace, intrinsic generalization of the mean on manifolds naturally comes into mind: the Fréchet mean is the set of global minima of the variance, as defined by Fréchet in general metric spaces [3]. The set of local minima of the variance was named Karcher mean by W.S Kendall [8] after the work of Karcher et al. [6] on Riemannian centers of mass (see [7] for a discussion of the naming and earlier works). The one-dimensional component is then quite naturally a geodesic which should pass through the mean point. Higher-order components are more difficult to define. The simplest generalization is tangent PCA (tPCA), which amounts to unfold the whole distribution in the tangent space at the mean, and to compute the principal components of the covariance matrix in the tangent space. The method is thus based on the maximization of the explained variance. tPCA was used implicitly or explicitly in a lot of statistical works on shape spaces and Riemannian manifolds because it is simple and efficient. However, if tPCA is good for analyzing data which are sufficiently centered around a central value (unimodal or Gaussian-like data), it is often not sufficient for multimodal or large support distributions (e.g. uniform on close compact subspaces).

Instead of an analysis of the covariance matrix, Fletcher et al. [2] proposed to rely on the least square distance to subspaces which are totally geodesic at a point. These *Geodesic Subspaces (GS)* are spanned by the geodesics going through a point with tangent vector restricted to belong to a linear subspace of the tangent space. The procedure was coined Principal Geodesic Analysis (PGA). However, the least-square procedure was computationally expensive, so that it was approximated in practice with tPCA. A complete implementation of the original PGA procedure was only provided recently by Sommer et al. [14]. PGA is intrinsic and allows to build a flag (sequences of embedded subspaces) of principal geodesic subspaces which is consistent with a forward component analysis approach: we build iteratively the components from dimension 0 (the mean point), dimension 1 (a geodesic) and higher dimensions by iteratively selecting the direction in the tangent space at the mean that optimally reduce the square distance of data point to the geodesic subspace. In this procedure, the mean always belong to geodesic subspaces even when they are not part of the support of the distribution.

To alleviate this problem, Huckemann et al. [12] proposed to relax the fact that the base-point of the geodesic subspace has to be the Fréchet mean: they start at the first order component directly with the geodesic that best fits the data, which is not necessarily going through the mean. The second principal geodesic is chosen orthogonally to the first one, and higher order components are added orthogonally at the crossing point of the first two components. The method was named Geodesic PCA (GPCA). Further relaxing the assumption

that second and higher order components should cross at a single point, Sommers [13] proposed to parallel transport the second direction along the first principal geodesic to define the second coordinates, and iteratively define higher order coordinates through horizontal development along the previous modes.

All the cited methods are intrinsically forward methods that build successively larger and larger approximation spaces. A notable exception is the concept of Principal Nested Spheres (PNS), proposed by Jung, et al. [4] as a general framework for non-geodesic decomposition of high-dimensional spheres or high-dimensional planar landmarks shape spaces. Here, subsphere can be viewed as a slicing of a higher dimensional sphere by an affine hyperplane. In this process, the nested subsphere is not of radius one, unless the hyperplane is passing through the origin. The backward analysis approach, determining a decreasing family of subspace, has been recently generalized to more general manifold with the help of a nested sequence of relations [1]. However, up to know, such a sequence of relationships was only known for spheres or Euclidean spaces.

In [10], we proposed to replace geodesic subspaces with a new and more general type of family of subspaces in manifolds: barycentric subspaces (BS). They are implicitly defined as the locus of points which are weighted means of $k+1$ reference points. Depending on the generalization of the mean that we use on manifolds, Fréchet mean, Karcher mean or exponential barycenter, we obtain the Fréchet / Karcher / Exponential barycentric subspaces (FBS / KBS / EBS). These definition were shown to be included into each other. Here we call affine span the metric completion of the largest barycentric subspace. In generic conditions, barycentric subspaces are stratified spaces that are locally submanifolds of dimension k. Their singular set of dimension $k-l$ corresponds to the case where l of the reference point belongs to the barycentric subspace defined by the $k-l$ other reference points. In non-generic conditions, points may coalesce along certain directions, defining non local jets instead of a regular $k+1$-tuple. Geodesic subspaces (in a restricted sense), which are defined by k tangent vectors at a point, correspond to the limit of the affine span when the k-tuple converges towards that jet.

In this paper, we derive the equations of barycentric subspaces in one of the simplest manifold: the sphere. We show that the affine span of $k+1$ different reference points on the n-dimensional sphere is the k-dimensional great subsphere that contains the reference points. In fact, any $k+1$-tuple of points of a great k-dimensional subsphere generates the same affine span, which is also a geodesic subspace. This coincidence of spaces is due to the very high symmetry of the sphere. For second order jets, we show that we obtain subspheres of different radii as used in PNS. Among the points of the affine span, determining which ones belong to the Karcher barycentric subspaces (KBS) turns out to be a surprisingly difficult problem. Practical experiments show that the index of the variance at critical points can be arbitrary, thus subdividing the affine span into many regions. As a result, the KBS covers in general only a small portion of the subsphere containing the reference points. This suggests that the affine span might be a much more interesting definition for subspace definition pur-

poses. Finally, we discuss the use to these barycentric subspaces to generalize PCA on manifolds, a procedure that we name Barycentric Subspace Analysis (BSA). Barycentric subspaces can be naturally nested, by defining an ordering of the reference points. Like for PGA, this allows the construction of a forward nested sequence of subspaces which contains the Fréchet mean. However, BSA also allows the construction of backward nested sequence which may not contain the mean.

2 Introduction

In this section, we recall the main notations and results from of [10]. We consider an embedding Riemannian manifold $\mathcal{M}$ of dimension n. The Riemannian metric is denoted $\langle \, . \mid . \, \rangle_x$ on each tangent space $T_x\mathcal{M}$ of the manifold. The expression of the the underlying norm in a chart is $\|v\|_x^2 = v^{\mathrm{T}} \, G(x) \, v = v^i v^j g_{ij}(x)$ using Einstein notations for tensor contractions. We assume the manifold to be geodesically complete (no boundary nor any singular point that we can reach in a finite time). As an important consequence, the Hopf-Rinow-De Rham theorem states that there always exists at least one minimizing geodesic between any two points of the manifold.

We denote by $\exp_x(v)$ the *exponential map* at point x which associate to each tangent vector $v \in T_x\mathcal{M}$ the point of $\mathcal{M}$ reached by the geodesic starting at x with this tangent vector after a unit time. This map is a local diffeomorphism from $0 \in T_x\mathcal{M}$ to $\mathcal{M}$, and we denote $\overrightarrow{xy} = \log_x(y)$ its inverse: it may be defined as the smallest vector of $T_x\mathcal{M}$ that allows to shoot a geodesic from x to y. When the tangent space is provided with an orthonormal basis, the Riemannian exponential and logarithmic maps provide *a normal coordinate systems at* x. A set of normal coordinate systems at each point of the manifold realize an atlas which allows to work very easily on the manifold. The implementation of exp and log maps is the basis of programming on Riemannian manifolds, and most the geometric operations needed for statistics or image processing can be rephrased based on them [9, 11].

2.1 $(k+1)$-pointed Riemannian manifold

Let $\{x_0, \dots x_k\} \in \mathcal{M}^{k+1}$ be a set of $k+1$ distinct points in the Riemannian manifold $\mathcal{M}$ and $C(x_0, \dots x_k) = \cup_{i=0}^{k} C(x_i)$ be the union of the cut loci of these points. We call $(k+1)$-pointed manifold $\mathcal{M}^*(x_0, \dots x_k) = \mathcal{M} / C(x_0, \dots x_k)$ the submanifold of the non-cut points of the points.

Since the cut locus of each point is closed and has null measure, $\mathcal{M}^*(x_0, \dots x_k)$ is open and dense in $\mathcal{M}$. Thus, it is a submanifold of $\mathcal{M}$ (not necessarily connected). On this submanifold $\mathcal{M}^*(x_0, \dots x_k)$, the distance to the points x_i and the Riemannian log function $\overrightarrow{xx_i} = \log_x(x_i)$ are smooth.

2.2 Weighted moments of a $(k+1)$-pointed manifold

Let $(\lambda_0, \dots \lambda_k) \in \mathbb{R}^{k+1}$ be weights such that $\sum_i \lambda_i \neq 0$. We call such weights barycentric coordinates. They are elements of projective space $\mathcal{P}_k$ minus the orthogonal of the line element $\mathbb{1} = (1 : 1 : \dots 1)$: $\mathcal{P}_k^* = \{(\lambda_0 : \dots : \lambda_k) \in \mathbb{R}^{k+1}$ such that $\sum_i \lambda_i \neq 0\}$. Standard charts of this space are given either by the intersection of the line elements with the "upper" unit sphere S_k of $\mathbb{R}^{k+1}$ with north pole $\mathbb{1}/\sqrt{k}$ (unit weights) or by the k-plane of $\mathbb{R}^{k+1}$ passing through the point $\mathbb{1}/k$ and orthogonal to this vector. We call normalized weights $\underline{\lambda}_i = \lambda_i / (\sum_{j=0}^k \lambda_j)$ this last projection.

Given barycentric coordinates $\lambda \in \mathcal{P}_k^*$, we can consider the distribution (or 0-current) $\mu(x) = \sum_i \lambda_i \delta_{x_i}(x)$ on $\mathcal{M}$. As it is not normalized and weights can be negative, it is generally not a probability distribution. To define the the weighted n-order moment of that distribution, we have to restrict to the $(k+1)$-pointed Riemannian manifold $\mathcal{M}^*(x_0, \dots x_k)$ because the Riemannian log and distance functions are not defined (resp. smooth) at the cut-locus of the points $\{x_i\}$:

$$\mathfrak{M}_n(x, \lambda) = \sum_i \lambda_i \underbrace{\overrightarrow{xx_i} \otimes \overrightarrow{xx_i} \dots \otimes \overrightarrow{xx_i}}_{n \text{ times}} \tag{1}$$

The 0-th order moment (the mass) $\mathfrak{M}_0(\lambda) = \sum_i \lambda_i = \mathbb{1}^{\mathrm{T}} \lambda$ is constant. All other moment are homogeneous of degree 1 in λ and can be normalized by dividing by the mass $\mathfrak{M}_0(\lambda)$. The first order moment $\mathfrak{M}_1(x, \lambda) = \sum_i \lambda_i \overrightarrow{xx_i}$ is a smooth vector field on the manifold $\mathcal{M}^*(x_0, \dots x_k)$. The second and higher order moments are smooth $(n, 0)$ tensor fields that will be used through their contraction with the Riemannian curvature tensor.

2.3 Barycentric subspaces of $k+1$ points

Let $(\mathcal{M}, \mathrm{dist})$ be a metric space and $(x_0, \dots x_k) \in \mathcal{M}^k$ be $k+1$ distinct reference points. The (normalized) weighted variance at point x with weight $\lambda \in \mathcal{P}_k^*$ is: $\sigma^2(x, \lambda) = \frac{1}{2} \sum_{i=0}^k \underline{\lambda}_i \, \mathrm{dist}^2(x, x_i) = \frac{1}{2} \sum_{i=0}^k \lambda_i \, \mathrm{dist}^2(x, x_i) / (\sum_{j=0}^k \lambda_j)$. The **Fréchet barycentric subspace** is the locus of weighted Fréchet means of these points, i.e. the set of absolute minima of the weighted variance:

$$FBS(x_0, \dots x_k) = \{\arg\min_{x \in \mathcal{M}} \sigma^2(x, \lambda), \ \lambda \in \mathcal{P}_k^*\}.$$

The **Karcher barycentric subspace** $KBS(x_0, \dots x_k)$ is defined similarly with local minima instead of global ones.

A point $x \in \mathcal{M}^*(x_0, \dots x_k)$ is a weighted exponential barycenters of the reference points if we can find barycentric coordinates $\lambda \in \mathcal{P}_k^*$ such that

$$\mathfrak{M}_1(x, \lambda) = \sum_{i=0}^k \lambda_i \overrightarrow{xx_i} = 0. \tag{2}$$

The **Exponential barycentric subspace** $EBS(x_0, \dots x_k)$ is the set of weighted exponential barycenters of the reference points in $\mathcal{M}^*(x_0, \dots x_k)$:

$$EBS(x_0, \dots x_k) = \{x \in \mathcal{M}^*(x_0, \dots x_k) | \exists \lambda \in \mathcal{P}_k^* : \mathfrak{M}_1(x, \lambda) = 0\}.$$

This definition is only valid on $\mathcal{M}^*(x_0, \dots x_k)$ and may hide some discontinuities or continuity on the union of the cut locus of the reference points. In order to ensure the completeness of the subspace and potentially reconnect components, we define consider the closure of this set: we call **affine span** of the points $(x_0, \dots x_k) \in \mathcal{M}^k$ the closure of the EBS in $\mathcal{M}$:

$$\mathrm{Aff}(x_0, \dots x_k) = \overline{\mathrm{EBS}}(x_0, \dots x_k).$$

Because we assumed that $\mathcal{M}$ is geodesically complete, this is equivalent to the metric completion of the EBS.

Outside the cut locus of the reference points, which is of null measure, the gradient of the squared distance $d^2_{x_i}(x) = \mathrm{dist}^2(x, x_i)$ is well defined and is equal to $\nabla d^2_{x_i}(x) = -2 \log_x(x_i)$. Thus, one recognizes that Eq.(2) defines nothing else than the critical points of the variance $\sigma^2(x, \lambda) = \frac{1}{2} \sum_i \lambda_i \, \mathrm{dist}^2(x, x_i)$. The local minima of the variance which are potnetially located on the cut-locus of the reference points are not part of the EBS but they are recovered in the affine span thanks to the metric completion. FBS and KBS are thus included in the affine span, and the affine span is the largest of the barycentric subspaces.

2.4 SVD Characterization of the EBS

Let us consider field of $n \times (k+1)$ matrices $Z(x) = [\overrightarrow{xx_0}, \dots \overrightarrow{xx_k}]$ on $\mathcal{M}^*(x_0, \dots x_k)$. We can rewrite Eq.(2) in matrix form: $\mathfrak{M}_1(x, \lambda) = Z(x)\lambda = 0$. Thus, we see that the EBS is controled by the kernel of the matrix field $Z(x)$. Let now $Z(x) = U(x).S(x).V(x)^{\mathrm{T}}$ be a singular decomposition with singular values sorted in decreasing order. The barycentric subspace $EBS(x_0, \dots x_k)$ is the zero level-set of the $k+1$ singular value $s_{k+1}(x)$ and the subspace of valid barycentric weights is spanned by the right singular vectors corresponding to the l vanishing singular values: $Span(v_{k-l}, \dots v_k)$ (it is empty if $l = 0$).

2.5 Karcher barycentric subspace and positive span

A critical point of the variance $x \in EBS(x_0, \dots x_k)$ is said non-degenerated (resp. positive) if the Hessian matrix $H(x, \lambda) = -\sum_{i=0}^{k} \lambda_i D_x \log_x(x_i)$ is invertible (resp. positive definite) for all λ in the right singular space of the zero singular value of $Z(x)$. The set of degenerate (resp. non-degenerate or positive) exponential barycenter is called the degenerate EBS and denoted $EBS^0(x_0, \dots, x_k)$ (resp. non-degenerate $EBS^*(x_0, \dots, x_k)$ or positive $EBS^+(x_0, \dots x_k)$). In Euclidean spaces, all the points are positive and non-degenerated. However, in Riemannian manifolds, we generally have degenerated points and non-degenerated points which are non-positive, as we will see with with the example of spheres. Thus, we can conclude that the KBS is the positive EBS plus potentially some degenerate points of the affine span and some points of the cut locus of the reference points.

2.6 Geodesic subspaces as limit case of the affine span

The usual definition of the geodesic subspaces $GS(W_x) = \{\exp_x(w), w \in W_x\}$ is too large in certain cases to be useful (e.g. in torus). We call restricted geodesic submanifold $GS^*(W_x) = \{\exp_x(w), w \in W_x \cap D(x)\}$ its restriction to the points that are reached without going through the cut locus of x. This is a well defined submanifold of $\mathcal{M}$ whose points are described by homogeneous coordinates at infinity (or on the equator of $\mathbb{1}/\sqrt{k}$ depending of the chart we chose for $\mathcal{P}_k^*$) of the affine span $\mathrm{Aff}(x, x_1, \dots x_k)$ when the points $x_i = \exp_x(\eta w_i)$ are converging to x at first order along the tangent vectors w_i defining the k-dimensional subspace $W_x \subset T_x\mathcal{M}$.

3 Example on spheres

Let us consider the unit sphere as our base manifold. We represent points of $\mathcal{M} = \mathcal{S}_n$ as unit vectors in $\mathbb{R}^{n+1}$. The tangent space at x is naturally represented by the linear space of vectors orthogonal to x: $T_x\mathcal{S}_n = \{v \in \mathbb{R}^{n+1}, v^{\mathrm{T}}x = 0\}$. The natural Riemannian metric on the unit sphere is inherited from the Euclidean metric of the embedding space $\mathbb{R}^{n+1}$. With these conventions, the Riemannian distance is the arc-length $d(x, y) = \arccos(x^{\mathrm{T}}y)$. Let us denote $f(\theta) = 1/\mathrm{sinc}(\theta) = \theta/\sin(\theta)$. The spherical exp and log maps are

$$\exp_x(v) = \cos(\|v\|)x + \frac{\sin(\|v\|)}{\|v\|}v = \cos(\|v\|)x + \mathrm{sinc}(\|v\|)v$$

$$\log_x(y) = \frac{\theta}{\sin(\theta)}(y - \cos(\theta)x) = f(\theta)(y - \cos(\theta)x) \quad \text{with} \quad \theta = \arccos(x^{\mathrm{T}}y).$$

Notice that $f(\theta)$ is a smooth function from $]-\pi;\pi[$ to $\mathbb{R}$ that is always greater than one and is locally quadratic at zero: $f(\theta) = 1 + \theta^2/6 + O(\theta^4)$.

Let us pick $k+1$ points on the sphere that we put in a matrix $X = [x_0; \dots x_k]$. In the sequel, we use the same notation for the matrix and the set of points. The cut locus of x_i is its antipodal point $-x_i$ so that $\mathcal{M}^*(X) = \mathcal{S}_n/-X$. Let us denote $\theta_i = \arccos(x_i^{\mathrm{T}}x)$. The log at a point x is $\overrightarrow{xx_i} = (\mathrm{Id} - xx^{\mathrm{T}})f(\theta_i)x_i$, so that the first weighted moment is

$$\mathfrak{M}_1(x, \lambda) = (\mathrm{Id} - xx^{\mathrm{T}})\sum_i \lambda_i f(\theta_i)x_i = (\mathrm{Id} - xx^{\mathrm{T}})XF(X, x)\lambda$$

where $F(X, x) = \mathrm{Diag}(f(\theta_i))$ is a diagonal matrix with entries that are always greater than one for $x \in \mathcal{M}^*(X)$. Thus the matrix $Z(x) = (\mathrm{Id} - xx^{\mathrm{T}})XF(X, x)$ has the same kernel as the matrix $\tilde{Z}(x) = (\mathrm{Id} - xx^{\mathrm{T}})X$. This corresponds to a renormalized $\tilde{\lambda} = F(X, x)\lambda$ of the weights which is linear in λ but non-linear in x and X through the function $F(X, x)$. The solutions of the equation $\tilde{Z}(x)\tilde{\lambda} = 0$ under the constraint $\|x\| = 1$ are given by $(x^{\mathrm{T}}X\tilde{\lambda})x = X\tilde{\lambda}$ or more explicitly $x = \pm X\tilde{\lambda}/\|X\tilde{\lambda}\|$. This means that the point $x \in \mathcal{M}^*(X)$ has to belong to the Euclidean span of the reference vectors. Conversely, any unit vector $x = X\tilde{\lambda}$ of

the Euclidean span of X verifies $\tilde{Z}(x)\tilde{\lambda} = 0$. Thus, the unit vectors $x = \pm X\tilde{\lambda}/\|\tilde{\lambda}\|$ have barycentric coordinates $\lambda = F(X,x)^{(\text{-}1)}\tilde{\lambda}$ if x is not at the cut locus of the reference points. This shows that

$$\text{Aff}(X) = \text{Span}\{x_0, \dots x_k\} \cap \mathcal{S}_n. \tag{3}$$

Following the same principle, we can orthogonalize the reference points x_i: let us denote by $X = USV^{\text{T}}$ a singular value decomposition of the matrix of reference vectors. All the singular values s_i are positive since the vectors x_i are assumed to be linearly independent. Thus, $Z(x)$ has the same kernel as $Y(x) = (\text{ Id} - xx^{\text{T}})U$. This shows that the exponential barycentric subspace generated by the original and orthogonalized points is the same, except at the cut locus of all these points which however belongs to the closure: the affine span. Thus, for spherical data as for Euclidean data, the affine span only depend on the reference points through the point of the Grassmanian that they define.

When the reference points x_i are not linearly independent, the matrix X has $l \geq 1$ vanishing singular values. A singular value decomposition $X = USV^{\text{T}}$ shows that the value of $\tilde{\lambda}$ (and thus of λ) is in that case unconstrained in the vector space generated by the right singular vectors $v_{k-l}, \dots v_k$ associated to the l vanishing singular values $s_{k-l}, \dots s_k$. Thus, the space of admissible weights at each point of the EBS is of dimension l, and the affine span itself is still the subsphere generated the Euclidean span of the reference points (minus their cut-locus), which is of dimension $k - l$.

In conclusion, the affine span $\text{Aff}(X)$ of $k+1$ different reference unit points $X = [x_0; \dots x_k]$ on the n-dimensional sphere $\mathcal{S}_n$ provided with the canonical Euclidean metric of the embedding space $\mathbb{R}^{n+1}$ is the great subsphere of dimension $rank(X) - 1$ that contains the reference points.

3.1 Reference points coalescing at order 1

Assume now all the reference points coalesce to a single point $x_i = \exp_{x_0}(\epsilon w_i)$ along the tangent vectors w_i which are satisfying $x_0^{\text{T}} w_i = 0$ (to belong to the tangent space at x_0) and $\sum_i w_i = 0$. This amounts to say that we are following the curve $X_\epsilon = X_0 + \epsilon W$ in the space of barycentric subspaces, with $X_0^{\text{T}} W = 0$ and $W X_0^{\text{T}} = 0$, where here $X_0 = x_0 \mathbb{1}^{\text{T}}$.

As previously, the points of the EBS of X_ϵ are solution of the equation $Z(x)\lambda = 0$ and since $Z(x)$ has the same kernel as the matrix $Z'(x) = (\text{ Id} - xx^{\text{T}})X_\epsilon$, x is solution if we can find some α_ϵ and λ such that $\alpha_\epsilon x = X_\epsilon \lambda$. The additional constraint $\|x\|^2 = 1$ tells us that $\alpha_\epsilon^2 = \|X_\epsilon \lambda\|^2$, which gives $x = X_\epsilon \lambda / \alpha_\epsilon$ when we take the positive root and reintegrate the sign into λ.

In our case, thanks to the orthogonality of X_0 and W, we have

$$\alpha_\epsilon = \|X_0\lambda + \epsilon W\lambda\| = \sqrt{\|X_0\lambda\|^2 + \epsilon^2 \|W\lambda\|^2},$$

Assuming that $\|X_0\lambda\| \neq 0$ (which is in particular the case when $X_0 = x_0\mathbb{1}^{\text{T}}$ since $\mathbb{1}^{\text{T}}\lambda$ is assumed not to vanish), we have $\alpha_\epsilon^{-1} = \frac{1}{\|X_0\lambda\|} - \frac{\epsilon^2}{2}\frac{\|W\lambda\|^2}{\|X_0\lambda\|^3} + O(\epsilon^4)$,

so that finally

$$x = \left(1 - \frac{\epsilon^2}{2}\frac{\|W\lambda\|^2}{\|X_0\lambda\|^2} + O(\epsilon^4)\right)\left(\frac{X_0\lambda}{\|X_0\lambda\|} + \epsilon\frac{W\lambda}{\|X_0\lambda\|}\right) = \frac{X_0\lambda}{\|X_0\lambda\|} + W\frac{\epsilon\lambda}{\|X_0\lambda\|} + O(\epsilon^2)$$

Thus, we see that the space $Aff(x)$ is the intersection of the sphere with the Euclidean hyperplane generated by X_0 and W, which is once again the geodesic subspace $GS(X_0, W)$.

3.2 Coalescence at order 2 and link with principal nested spheres

Principal nested spheres were proposed by Jung, Dryden and Marron as a general framework for non-geodesic decomposition of high-dimensional spheres or high-dimensional planar landmarks shape spaces [5, 4]. A subsphere $\mathcal{A}_{n-1}$ of $\mathcal{S}_n$ is defined as the set of points which are at a fixed distance $\theta \in (0, \pi/2]$ of a point $x \in \mathcal{S}_n$: $\mathcal{A}_{n-1}(x, \theta) = \{y \in \mathcal{S}_n \,/\, d(x, y) = \theta\}$. The subsphere $\mathcal{A}_{n-1}(x, \theta)$ can be viewed as a slicing of $\mathcal{S}_n$ by the n-dimensional affine hyperplane $P(x, \theta) = \{y \in \mathbb{R}^{n+1} \,/\, y^{\mathrm{T}}x = \cos(\theta)\}$. Notice that the coordinates $(x, \cos(\theta))$ of the affine hyperplane parametrize all the possible subspheres of dimension $n - 1$. In this process, the nested subsphere is not of radius one, unless one takes $\theta = \pi/2$, in which case the hyperplane is passing through the origin.

In order to figure out how nested subspheres and barycentric subspaces are related, we consider the top circle of radius $r \in [0; 1]$ around the axis e_3 on the 3-sphere implicitly described by the equation $x^{\mathrm{T}}e_3 = \sqrt{1 - r^2}$. The computations would be exactly the same for the bottom circle $x^{\mathrm{T}}e_3 = -\sqrt{1 - r^2}$. The explicit equation of our circle is: $x(\psi) = r\cos(\psi)e_1 + r\sin(\psi)e_2 + \sqrt{1 - r^2}e_3$. We consider the three points on this cicle at angle $\psi = 0$, $\psi = \epsilon$ and $\psi = -\epsilon$: $X = [x(0), x(\epsilon), x(-\epsilon)]$. The spherical affine span consists of the points $x = X\lambda/\|X\lambda\|$ for $\lambda \in \mathcal{S}_3$, with:

$$\begin{aligned} X\lambda &= r\left(\lambda_0 + \cos(\epsilon)(\lambda_1 + \lambda_2)\right)e_1 + r(\lambda_1 - \lambda_2)\sin(\epsilon)e_2 + \sqrt{1 - r^2}(\lambda_0 + \lambda_1 + \lambda_2)e_3 \\ &= (\lambda_0 + \lambda_1 + \lambda_2)\left(re_1 + \sqrt{1 - r^2}e_3\right) + \epsilon r(\lambda_1 - \lambda_2)e_2 - \frac{r\epsilon^2}{2}(\lambda_1 + \lambda_2)e_1 + O(\epsilon^3). \end{aligned}$$

Using new coordinates $s = (\lambda_0 + \lambda_1 + \lambda_2)$, $u = (\lambda_1 - \lambda_2)\epsilon/s$ and $v = (\lambda_1 + \lambda_2)\epsilon^2/(2s)$, we get the equation

$$X\frac{\lambda}{s} = r(1 - v)e_1 + rue_2 + \sqrt{1 - r^2}e_3 + O(\epsilon^3).$$

Thus, the equation $x = X\lambda$ can only describe the hyperplane $x^{\mathrm{T}}e_3 = \sqrt{1 - r^2}$ when ϵ goes to zero (up to a scaling factor s that we can freely choose to be 1 thanks to the homogeneous coordinates), and its intersection with the sphere can only describe a circle of radius r.

Iterating the process, one can generalize the above construction to subspheres of arbitrary dimensions. Thus, we see that Nested Spheres as a limit case of the affine span when the k reference points tend to a 2-jet. It would be interesting to determine which types of subspaces could be obtained by such limits for more general non-local and higher order jets.

4 Barycentric subspace analysis

We turn in this section to the generalization of principal component analysis itself. PCA can be viewed as the search for a sequence of nested linear spaces that best approximate the data at each level, for instance by minimizing the variance of the residues. In a Euclidean space, this process boils down to an independent optimization of orthogonal subspaces at each level of approximation, thanks to the Pythagorean theorem. This allows to build each subspace of the sequence by adding (resp. subtracting) the optimal one-dimensional subspace iteratively in a forward (resp. backward) analysis. Of course, this property does not scale up to manifolds, for which subspaces have no reason to be orthogonal (even this notion is not well defined).

4.1 Flags of barycentric subspaces in manifolds

Nestedness of approximation spaces has been argued to be one of the most important characteristics for generalizing PCA to more general spaces [1]. Barycentric submanifolds can easily be nested, for instance by adding or removing one or several points at a time, which corresponds to put the barycentric weight of this (or these) point(s) to zero. One obtains in that case a family of embedded submanifolds which we call a flag because this generalizes flags of vector spaces. Indeed a flag of a vector space V is a filtration of subspaces (an increasing sequence of subspaces, where "increasing" means each subspace is a proper subspace of the next): $\{0\} = V_0 \subset V_1 \subset V_2 \subset \cdots \subset V_k = V$. Denoting by $d_i = \dim(V_i)$ the dimension of the subspaces, we have $0 = d_0 < d_1 < d_2 < \cdots < d_k = n$, where n is the dimension of V. Hence, we must have $k \leq n$. A flag is called a complete flag if $d_i = i$, otherwise it is called a partial flag.

With barycentric subspaces of an n-dimensional manifold $\mathcal{M}$, an ordering of $n+1$ distinct points $x_0, \dots x_n$ defines a complete flag of barycentric subspaces in the sense that: $BS(x_0) = \{x_0\} \subset \cdots BS(x_0, x_1, x_k) \cdots \subset BS(x_0, \dots x_n)$.

Grouping points together in the addition/removal process generates a partial flag of barycentric subspaces. Among the barycentric subspaces, the affine span seems to be the most interesting definition to use because the affine span of $n+1$ distinct points covers the full manifold: $\mathrm{Aff}(x_0, \dots x_n) = \mathcal{M}$ while we only generate a submanifold with the Fréchet or Karcher barycentric subspaces, as we have seen with the example of spheres.

4.2 Forward and backward barycentric subspaces analysis

In the classical PCA, the flag of linear subspaces can be built in a forward way, by computing the best 0-th order approximation (the mean), then the best first order approximation (the first mode), etc. It can also be built backward, by removing the direction with the minimal residue from the current affine subspace. In a manifold, we can use similar forward and backward analysis, but they have no reason to give the same result.

With a forward analysis, we compute iteratively the flag of affine spans by adding one point at a time and keeping the previous ones fixed. Thus, we begin by computing the optimal barycentric subspace of dimension 0: $\mathrm{Aff}(x_0) = \{x_0\}$. Since there is only one weight and it should be unit, the optimal point x_0 found by minimizing the unexplained variance is a Karcher mean. Adding one more point amounts to compute the geodesic passing through the mean that best approximate the data. Adding a second points now differ from PGA, unless the three points coalesce to a single one. The procedure is continued point by point, which mean that the Fréchet mean always belong to the barycentric subspace. In practice, the forward analysis should be stopped when the variance of the residues reaches the noise level of the data, hopefully with k much lower than the dimension n of the embedding manifold, which allows to have an efficient dimension reduction.

The backward analysis consists in iteratively removing one dimension, thus one point in our case. One theoretically should start with a full set of points $x_0, \dots x_n$ which generates the full manifold and chose which one to remove. However, as all the sets of $n+1$ distinct points generate the full manifold, the optimization really begin with the set of n points $x_0, \dots x_{n-1}$. Actually this should normally be the only time when we perform an optimization for the point positions, since one should afterward only test for which of the n points we should remove, and this optimization is particularly ill-posed and inefficient in very large dimensional spaces! In order to get around this problem, we propose to run a non-nested forward analysis until we reach the noise level of the data for a dimension $k \gg n$. Since the goal is only to characterize the optimal k-dimensional subspace, we may optimize the point positions at each step to better fit the data. Then, a backward sweep at the end only reorders the points if necessary by iteratively selecting the one that least increase the unexplained variance. With this process, there is no reason why the Fréchet mean should belong to the reference points (and even to any of the barycentric subspaces). For instance, if we have clusters of points, one expects the reference points to localize within these clusters rather than at the Fréchet mean.

5 Discussion

We have first investigated in the paper barycentric subspaces in spheres and shown that they encompass both principal geodesic subspaces and nested subspheres as limit cases. It would be interesting to see if we can obtain other types of subspaces with higher order and non-local jets.

The second study point of this paper concerns the generalization of PCA to manifolds using Barycentric Subspace Analysis (BSA). We showed that an ordering of the reference points naturally defines a flag of nested barycentric subspaces. We proposed a first optimization procedure, but the lack of symmetry between the forward and the backward estimations calls for a proper global criterion to be optimized by all k-tuple for $k = 0 \dots n$ together and not just a greedy approach as done by the classical forward and backward approaches.

Other potential practical issues include the fact that the optimization on k-tuple might have multiple solutions, as in the case of spheres. Here, we need to find a suitable quotient space similar to the quotient definition of Grassmanians. The optimization might also converge towards a non-local jet instead on a k-tuple, and good renormalization techniques need to be designed to guaranty the numerical stability.

References

1. J. Damon and J. S. Marron. Backwards Principal Component Analysis and Principal Nested Relations. *Journal of Mathematical Imaging and Vision*, 50(1-2):107–114, Oct. 2013.
2. P. Fletcher, C. Lu, S. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23(8):995–1005, Aug. 2004.
3. M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut Henri Poincaré*, 10:215–310, 1948.
4. S. Jung, I. L. Dryden, and J. S. Marron. Analysis of principal nested spheres. *Biometrika*, 99(3):551–568, Sept. 2012.
5. S. Jung, X. Liu, J. S. Marron, and S. M. Pizer. Generalized PCA via the Backward Stepwise Approach in Image Analysis. In J. Angeles, B. Boulet, J. J. Clark, J. Kvecses, and K. Siddiqi, editors, *Brain, Body and Machine*, number 83 in Advances in Intelligent and Soft Computing, pages 111–123. Springer Berlin Heidelberg, 2010.
6. H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications in Pure and Applied Mathematics*, 30:509–541, 1977.
7. H. Karcher. Riemannian Center of Mass and so called karcher mean. *arXiv:1407.2087 [math]*, July 2014. arXiv: 1407.2087.
8. W. Kendall. Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence. *Proc. London Math. Soc.*, 61(2):371–406, 1990.
9. X. Pennec. Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006. A preliminary appeared as INRIA RR-5093, January 2004.
10. X. Pennec. Barycentric Subspaces and Affine Spans in Manifolds. In *Geometric Science of Information GSI'2015*, Proceedings of Geometric Science of Information GSI'2015, Palaiseau, France, Oct. 2015.
11. X. Pennec, P. Fillard, and N. Ayache. A Riemannian Framework for Tensor Computing. *International Journal of Computer Vision*, 66(1):41–66, 2006. A preliminary version appeared as INRIA Research Report 5255, July 2004.
12. A. M. S. Huckemann, T. Hotz. Intrinsic shape analysis: Geodesic principal component analysis for Riemannian manifolds modulo Lie group actions. *Statistica Sinica*, pages 1–100, 2010.
13. S. Sommer. Horizontal Dimensionality Reduction and Iterated Frame Bundle Development. In F. Nielsen and F. Barbaresco, editors, *Geometric Science of Information*, number 8085 in Lecture Notes in Computer Science, pages 76–83. Springer Berlin Heidelberg, 2013.
14. S. Sommer, F. Lauze, and M. Nielsen. Optimization over geodesics for exact principal geodesic analysis. *Advances in Computational Mathematics*, 40(2):283–313, June 2013.

Curve Matching with Applications in Medical Imaging

Martin Bauer[1], Martins Bruveris[2], Philipp Harms[3], Jakob Møller-Andersen[4]* **

[1] Institut für Diskrete Mathematik und Geometrie, TU Wien
[2] Department of Mathematics, Brunel University London
[3] Department of Mathematics, ETH Zürich
[4] Department of Applied Mathematics and Computer Science, Technical University of Denmark

Abstract. In the recent years, Riemannian shape analysis of curves and surfaces has found several applications in medical image analysis. In this paper we present a numerical discretization of second order Sobolev metrics on the space of regular curves in Euclidean space. This class of metrics has several desirable mathematical properties. We propose numerical solutions for the initial and boundary value problems of finding geodesics. These two methods are combined in a Riemannian gradient-based optimization scheme to compute the Karcher mean. We apply this to a study of the shape variation in HeLa cell nuclei and cycles of cardiac deformations, by computing means and principal modes of variations.

Keywords: Curve matching, Sobolev metrics, Riemannian shape analysis, discrete geodesics, minimizing geodesics

1 Introduction

The comparison and analysis of geometric shapes plays an important role in medical imaging [27], biology [9] as well as many other fields [11]. Spaces of geometric shapes are inherently nonlinear. To make standard methods of statistical analysis applicable, one can linearize the space locally around each shape. This can be achieved by introducing a Riemannian structure, which is able to describe jointly the global nonlinearity of the space as well as its local linearity.

Over the past decade, Riemannian shape analysis has become an active area of research in pure and applied mathematics. Driven by applications, a variety of spaces, equipped with different Riemannian metrics, have been used to represent geometrical shapes and their attributes. Ideally, the particular choice of metric should be dictated by the data at hand rather than by mathematical or numerical

* All authors were partially supported by the Erwin Schrödinger Institute programme: Infinite-Dimensional Riemannian Geometry with Applications to Image Matching and Shape Analysis. M. Bauer was supported by the European Research Council (ERC), within the project 306445 (Isoperimetric Inequalities and Integral Geometry) and by the FWF-project P24625 (Geometry of Shape spaces).

** All authors contributed equally to the article.

convenience. This leads to the task of developing efficient numerical methods for the statistical analysis of shapes for general and flexible classes of metrics.

The topic of this paper are second order Sobolev metrics on the space of regular, planar curves. This space and its quotients by translations, rotations, scalings, and reparametrizations are important and widely used spaces in shape analysis. Second order Sobolev metrics are mathematically well-behaved: the geodesic equation is globally well-posed, any two curves in the same connected component can be connected by a minimizing geodesic, the metric completion consists of all H^2-immersions, and the metric extends to a strong Riemannian metric on the metric completion [7, 8].

We provide the first numerical implementation of the initial and boundary value problems for geodesics with respect to these metrics and apply them to medical data.[5] Our implementation allows us to compute Karcher means, principal components, and clusters of curves in reasonable time. The parameters in the metric can be chosen freely because we do not rely on transforms which exist only for special choices of parameters [23, 26]. We are also able to factor out the action of the finite-dimensional translation and rotation groups.

We illustrate the behaviour of the metrics in two medical applications. In Sect. 4.1 we use images of HeLa cell nuclei from [5] and compute the mean shape of the nucleus and the principal modes of shape variation; in Sect. 4.2 we use traces of cardiac image sequences from [12, 24] and compare the mean shape of patients with Tetralogy of Fallot with the mean shape of a control group.

The paper is structured as follows: Sect. 2 contains the definition of second order Sobolev metrics and some relevant mathematical background. Section 3 describes our discretization of the geodesic equation and the Riemannian energy functional. Finally, in Sect. 4, we apply the metrics to medical imaging data.

2 Mathematical Background

In this article we center our attention on the space of smooth, regular curves with values in $\mathbb{R}^d$,

$$\operatorname{Imm}(S^1, \mathbb{R}^d) = \left\{ c \in C^\infty(S^1, \mathbb{R}^d) \colon \forall \theta \in S^1, c_\theta(\theta) \neq 0 \right\} \ , \tag{1}$$

where Imm stands for *immersion*. As an open subset of the Fréchet space $C^\infty(S^1, \mathbb{R}^d)$, it is a Fréchet manifold. Its tangent space, $T_c \operatorname{Imm}(S^1, \mathbb{R}^d)$, at any curve c is the vector space $C^\infty(S^1, \mathbb{R}^d)$.

We denote the Euclidean inner product on $\mathbb{R}^d$ by $\langle \cdot, \cdot \rangle$. Moreover, for any fixed curve c, we denote differentiation and integration with respect to arc length by $D_s = \frac{1}{|c_\theta|}\partial_\theta$ and $ds = |c_\theta| d\theta$ respectively.

Definition 1. *A second order Sobolev metric on* $\operatorname{Imm}(S^1, \mathbb{R}^d)$ *is given by*

$$G_c(h, k) = \int_{S^1} a_0 \langle h, k \rangle + a_1 \langle D_s h, D_s k \rangle + a_2 \langle D_s^2 h, D_s^2 k \rangle \, ds \ , \tag{2}$$

[5] Implementations of second order Sobolev metrics can be found in [15] and [3], but none of the previous implementations seem to have been used in applications.

where $a_0, a_2 > 0$*,* $a_1 \geq 0$*, and* $h, k \in T_c \operatorname{Imm}(S^1, \mathbb{R}^d)$ *are tangent vectors.*

Remark 2 (Invariance of the metric). The metric G is invariant with respect to translations, rotations, and reparametrizations by diffeomorphisms of S^1, but not with respect to scalings. It can be made scale-invariant by introducing weights that depend on the length ℓ_c of the curve c. A scale-invariant metric is given by

$$\widetilde{G}_c(h,k) = \int_{S^1} \frac{a_0}{\ell_c^3}\langle h,k\rangle + \frac{a_1}{\ell_c}\langle D_s h, D_s k\rangle + a_2\ell_c\langle D_s^2 h, D_s^2 k\rangle \, ds \ . \tag{3}$$

Depending on the application, it can be desirable to factor out some of these isometry groups. For example, we consider the HeLa cell nuclei in Sect. 4.1 as curves modulo translations and rotations. For the traces of cardiac images in Sect. 4.2, however, both the absolute position of the curve and its orientation have intrinsic meaning and should therefore not be factored out.

Remark 3 (Choosing the constants). The constants a_0, a_1 and a_2 in the definition of the metric determine the relative weight of the L^2-, H^1-, and H^2-parts. Their choice is a non-trivial and important task, and it should, in each application, be informed by the data at hand. The influence of different parameter values on geodesics and Karcher means can be seen in Fig. 1 and Fig. 2.

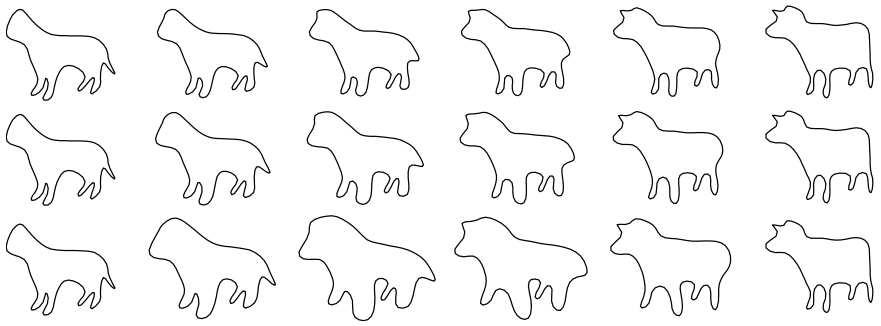


Fig. 1: Geodesics between a cat and a cow. The metric parameter a_2 in the first row is increased by a factor 10 in the second and a factor 100 in the third row.

Remark 4 (Generalizations to higher-dimensions). Sobolev metrics have natural generalizations to manifold-valued curves, embedded surfaces, and more generally spaces of immersions of a compact manifold M into a Riemannian manifold N; see [1, 4] for details and [2] for a general overview.

Deformations of curves can be seen as smooth paths $c\colon [0,1] \to \operatorname{Imm}(S^1, \mathbb{R}^d)$. Their velocity is c_t, the subscript t denoting differentiation.

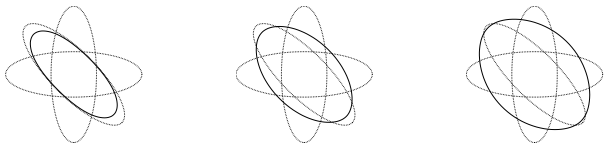


Fig. 2: Karcher means (bold) of three rotated ellipses (dashed). The parameter a_2 of the metric used in the first figure is increased by a factor 10 in the second and by a factor 100 in the third figure.

Definition 5 (Geodesic distance). *The length of a path of curves c is*

$$L(c) = \int_0^1 \sqrt{G_{c(t)}(c_t(t), c_t(t))}\, dt \ . \tag{4}$$

The distance between two curves in $\operatorname{Imm}(S^1, \mathbb{R}^d)$ *(with respect to the metric G) is the infimum of the lengths of all paths connecting these curves.*

$$\operatorname{dist}(c_0, c_1) = \inf_{c(0)=c_0, c(1)=c_1} L(c)\,.$$

Geodesics are locally distance-minimizing paths.

Geodesics can be described by a partial differential equation, called the geodesic equation. It is the first order condition for minima of the energy

$$E(c) = \frac{1}{2} \int_0^1 G_{c(t)}\big(c_t(t), c_t(t)\big)\, dt \ . \tag{5}$$

Recently, some local and global existence results for geodesics of Sobolev metrics were shown in [7, 8, 13]. Since they provide the theoretical underpinnings for the numerical methods presented in this paper, we summarize them here.

Theorem 6 (Geodesic equation). *Let $a_0, a_2 > 0$ and $a_1 \geq 0$. The geodesic equation of the metric G, written in terms of the momentum $p = |c'|(a_0 c_t - a_1 D_s^2 c_t + a_2 D_s^4 c_t)$, is given by*

$$\begin{aligned} \partial_t p = &- \frac{a_0}{2} |c_\theta| D_s(\langle c_t, c_t \rangle D_s c) + \frac{a_1}{2} |c_\theta| D_s(\langle D_s c_t, D_s c_t \rangle D_s c) \\ &- \frac{a_2}{2} |c_\theta| D_s(\langle D_s^3 c_t, D_s c_t \rangle D_s c) + \frac{a_2}{2} |c_\theta| D_s(\langle D_s^2 c_t, D_s^2 c_t \rangle D_s c)\,. \end{aligned} \tag{6}$$

Given any initial condition $(c_0, u_0) \in T \operatorname{Imm}(S^1, \mathbb{R}^2)$, the solution of the geodesic equation exists for all time.

If, however, $a_0, a_1 > 0$ and $a_2 = 0$, then G is a first order Sobolev metric. Its geodesic equation is locally, but not globally, well-posed.

Remark 7 (Comparison to elastic metrics). Closely related are elastic metrics [14], which in the planar case are given by

$$G_c(h, k) = \int_{S^1} a^2 \langle D_s h, n \rangle \langle D_s k, n \rangle + b^2 \langle D_s h, v \rangle \langle D_s k, v \rangle ds \ . \tag{7}$$

Here a, b are constants and v, n denote the unit tangent and normal vectors to c. Two special cases deserve to be highlighted: for $a = 1$, $b = \frac{1}{2}$ [23] and $a = b$ [26] there exist nonlinear transforms, the square root velocity transform and the basic mapping, that greatly simplify numerics. Both of these metrics have been applied to a variety of problems in shape analysis.

We note that the elastic metric with $a = b$ corresponds to a first order Sobolev metric as in Def. 1 with $a_0 = a_2 = 0$ and $a_1 = a^2 = b^2$. As it has no L^2-part, it is a Riemannian metric only on the space of curves modulo translations.

3 Discretization

We propose to represent paths of curves as tensor product B-splines

$$c(t, \theta) = \sum_{i=1}^{N_t} \sum_{j=1}^{N_\theta} c_{i,j} B_i(t) C_j(\theta) \tag{8}$$

of degree n_t in time (t) and n_θ in space (θ), respectively, using N_t and N_θ basis functions and controls $c_{i,j} \in \mathbb{R}^d$. The spatial basis functions C_j are defined on the uniform knot sequence $\theta_j = \frac{j-n}{2\pi \cdot N_\theta}$, $0 \le j \le 2n + N_\theta$, and satisfy periodic boundary conditions. The time basis functions B_i are defined using uniform knots on the interior of the interval $[0, 1]$ and with full multiplicity on the boundary. Full multiplicity of the boundary knots ensures that fixing the initial and final curves is equivalent to fixing the control points $c_{1,j}$ and $c_{N_t,j}$. Regarding smoothness, we have $B_i \in C^{n_t - 1}([0, 1])$ and $C_j \in C^{n_\theta - 1}([0, 2\pi])$.

The proposed discretization scheme is independent of the special form of the metric, and can be applied equally to the scale-invariant versions as well as to the family of elastic metrics.

3.1 Boundary Value Problem for Geodesics

To evaluate the energy (5) on discretized paths (8), we use Gaussian quadrature on each interval between subsequent knots. The resulting discretized energy $E_{\text{discr}}(c_{1,1}, \ldots, c_{N_t,N_\theta})$ is minimized over all discrete paths (8) with fixed initial and final control points $c_{1,j}$ and $c_{N_t,j}$. This is a finite-dimensional, nonlinear, unconstrained[6] optimization problem.

We solved this problem using either Matlab's `fminunc` function with gradients computed by finite differences, or using AMPL [10] and the Ipopt solver [25] with gradients computed by automatic differentiation. By classical approximation results [22], the discrete energy of discrete paths converges to the continuous energy of smooth paths as the number of control points tends to infinity. Hence we expect to obtain close approximations of geodesics. Establishing rigorous convergence results for our discretization will be the subject of future work.

[6] We neglect the condition $c_\theta(t, \theta) \neq 0$. All smooth paths violating this condition have infinite energy because of the completeness of the metric. By approximation, spline paths violating this condition have very high energy and will, in practice, be avoided during optimization.

3.2 Computing the Karcher Mean

The Karcher mean $\bar{c}$ of a set $\{c_1, \ldots, c_n\}$ of curves is the minimizer of

$$F(c) = \frac{1}{n} \sum_{j=1}^{n} \operatorname{dist}(c, c_j)^2 \ . \tag{9}$$

It can be calculated by a gradient descent on $(\operatorname{Imm}(S^1, \mathbb{R}^d), G)$. Letting $\operatorname{Log}_c c_j$ denote the Riemannian logarithm, the gradient of F with respect to G is [17]

$$\operatorname{grad}^G F(c) = \frac{1}{n} \sum_{j=1}^{n} \operatorname{Log}_c c_j \ . \tag{10}$$

3.3 Initial Value Problem for Geodesics

To calculate the Karcher mean by gradient descent, one has to repeatedly solve the geodesic equation (6). To this aim, we use the time-discrete variational geodesic calculus [21]. Given three curves c_0, c_1, c_2, one defines the discrete energy

$$E_2(c_0, c_1, c_2) = G_{c_0}(c_1 - c_0, c_1 - c_0) + G_{c_1}(c_2 - c_1, c_2 - c_1) \ . \tag{11}$$

A 3-tuple (c_0, c_1, c_2) is a discrete geodesic if it is a minimizer of the discrete energy E_2 with fixed endpoints c_0, c_2. The discrete exponential map is defined as follows: $c_2 = \operatorname{Exp}_{c_0} c_1$, if (c_0, c_1, c_2) is a discrete geodesic, in other words, if $c_1 = \operatorname{argmin} E_2(c_0, \cdot, c_2)$. To find c_2, we differentiate the discrete energy E_2 with respect to c_1 and solve the resulting system of nonlinear equations,

$$2G_{c_0}(c_1 - c_0, \cdot) - 2G_{c_1}(c_2 - c_1, \cdot) + D_{c_1} G_{\cdot}(c_2 - c_1, c_2 - c_1) = 0 \ . \tag{12}$$

We use the solver `fsolve` in Matlab to solve this system of equations.

Given a time resolution Δt and an initial velocity v, we choose $c_1 = c_0 + v\Delta t$ and compute $c_2 = \operatorname{Exp}_{c_0} c_1$. The procedure is repeated with c_1, c_2 in order to compute $c_3 = \operatorname{Exp}_{c_1} c_2$, which is iterated until we reach the required final time.

While the convergence results in [21] do not apply to our setting, we found good experimental agreement with the solutions of the boundary value problem.

4 Applications

4.1 Hela Cells

In our first example we want to characterize nuclear shape variation in HeLa cells. We use fluorescence microscope images of HeLa cell nuclei[7] (87 images in total). The acquisition of the cells is described in [5]. A similar study on this dataset was performed using the LDDMM framework in [19, 20]; applying intrinsically

[7] The dataset was downloaded from http://murphylab.web.cmu.edu/data.

defined Sobolev metrics to the same problem will provide a complimentary point of view.

To extract the boundary of the nucleus, we apply a thresholding method [16] to obtain a binary image and then fit – using least squares – a spline with $N_\theta = 12$ and $n_\theta = 4$ to the longest 4-connected component of the thresholded image. Then we reparametrize the boundary to approximately constant speed. The remaining degree of freedom is the starting point of the parametrization; when computing minimizing geodesics, we minimize over this parameter as well. The shapes of cell nuclei are thus represented by curves modulo translations, rotations, and constant shifts of the parametrisation, i.e. the two curves c and $e^{i\beta}c(\cdot - \alpha) + \lambda$ are considered equivalent. Examples of the extracted curves are depicted in Fig. 3.

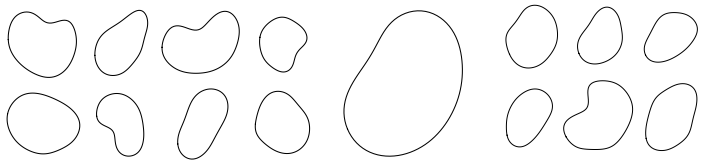


Fig. 3: Eight boundaries of HeLa cell nuclei, the Karcher mean of all nuclei (enlarged), and six randomly sampled cells using a Gaussian distribution in normal coordinates with the same covariance as the data.

The parameters a_0, a_1, and a_2 of the Riemannian metric are chosen as follows: we compute the average L^2-, H^1- and H^2-contributions $\overline{E}_{L^2}$, $\overline{E}_{H^1}$, $\overline{E}_{H^2}$ to the energy of linear paths between each pair of curves in the dataset. As the L^2- and H^2-contributions scale differently, we rescale all curves such that $\overline{E}_{L^2} = \overline{E}_{H^2}$. Then we choose constants a_0, a_1, and a_2 such that

$$a_0\overline{E}_{L^2} : a_1\overline{E}_{H^1} : a_2\overline{E}_{H^2} = 3 : 1 : 6 \quad \text{and} \quad \overline{E} = a_0\overline{E}_{L^2} + a_1\overline{E}_{H_1} + a_2\overline{E}_{H^2} = 100 \ ,$$

resulting in an average length $\overline{\ell}_c = 12.45$ and $a_0 = 3.36$, $a_1 = 2.20$, and $a_2 = 6.73$.

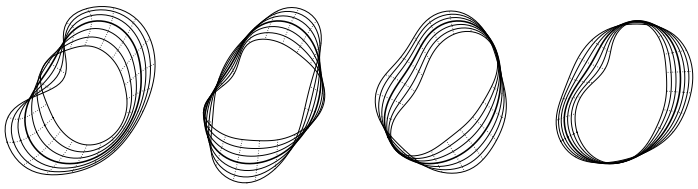


Fig. 4: Geodesics from the mean in the first four principal directions. The curves show the geodesic at times $-3, -2, \ldots, 2, 3$; the bold curve is the mean. One can see four different characteristic deformations of the cell: bending, stretching along the long axis, stretching along the short axis, and a combination of stretching with partial bending.

The average shape of the nucleus can be captured by the Karcher mean $\overline{c}$. To compute the Karcher mean of the 87 nuclei, we use a conjugate gradient method on the Riemannian manifold of curves, as implemented in the Manopt library [6], to solve the minimization problem (9). We obtain convergence in 28 steps; the final value of the objective function (9) is $F(\overline{c}) = 10.55$ and the norm of the gradient is $\| \operatorname{grad}^G F(\overline{c})\|_{\overline{c}} < 10^{-3}$. The mean shape can be seen in Fig. 3.

Having computed the mean $\overline{c}$, we represent each nuclear shape c_j by the initial velocity $v_j = \operatorname{Log}_{\overline{c}} c_j$ of the minimal geodesic connecting $\overline{c}$ and c_j. We perform principal component analysis with respect to the inner product $G_{\overline{c}}$ on the set of initial velocities $\{v_j : j = 1, \ldots, 87\}$. The first four eigenvalues are $4.10, 2.39, 1.68$ and 1.00, and they explain $38.04\%, 60.21\%, 75.78\%$, and 85.07% of the total variance. Geodesics from the mean in the directions of the first four principal directions can be seen in Fig. 4. Fig. 5 shows the data projected to the subspace spanned by the first two eigenvectors.

Finally, we sample from a normal distribution with the same covariance matrix as the data and project the sampled velocities $\hat{v}$ back to the space of curves using the exponential map $\hat{c} = \operatorname{Exp}_{\overline{c}} \hat{v}$; some examples can be seen in Fig. 3.

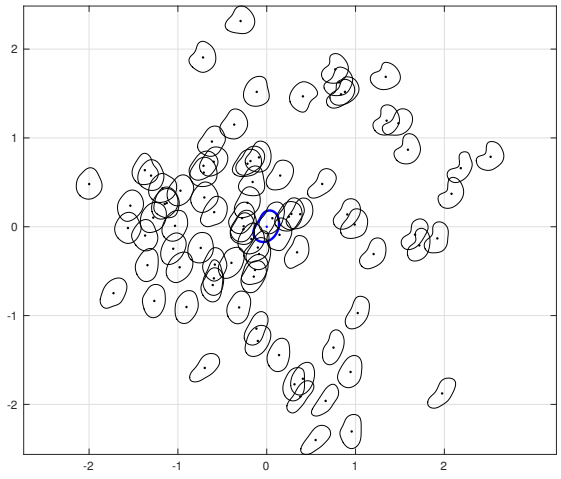


Fig. 5: All 87 cell nuclei projected to the plane in the tangent space which is spanned by the first two principal directions. The mean (in blue) is situated at the origin. The units on the coordinate axes are standard deviations. We see that the first coordinate is related to the bending of the nucleus, while the second coordinate is related to its elongation.

4.2 Traces of Cardiac Images

In our second application we study curves that are obtained from images of the cardiac cycle. More precisely, we consider a sequence of 30 cardiac images,

taken at equispaced time points along the cardiac cycle. Each image is projected to a barycentric subspace of dimension two, yielding a closed curve in the two-dimensional space of barycentric coordinates. After normalizing the coordinates [18, Sect. 3] we obtain a closed, plane curve – with the curve parameter representing time – to which we can apply the methods presented in Sect. 2. Details regarding the acquisition and projection of the images can be found in [12, 24]; barycentric subspaces on manifolds are described in [18].

The data consists of 10 cardiac cycles of patients with Tetralogy of Fallot and 9 patients from a control group. Each cardiac cycle is originally represented by three-dimensional homogeneous coordinates $x_1 : x_2 : x_3$, sampled at 30 time points. We project the homogeneous coordinates onto the plane $x_1 + x_2 + x_3 = 1$ and choose a two-dimensional coordinate system for this plane. Then we use spline interpolation with degree $n_\theta = 3$ and $N_\theta = 30$ control points to reconstruct the planar curves from the data points; see Fig. 6.

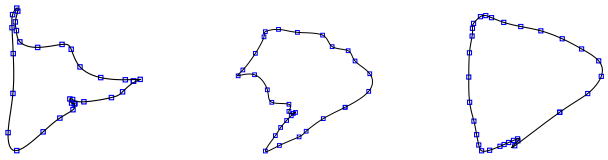
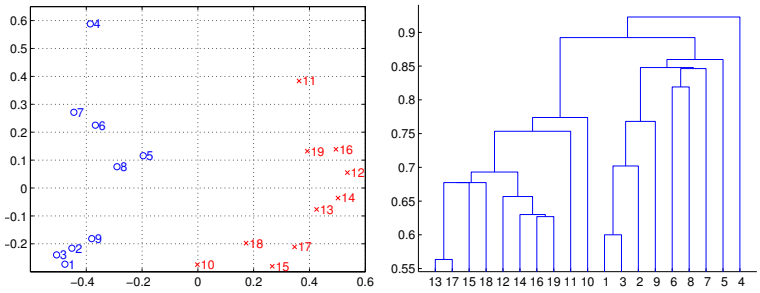


Fig. 6: Projections to a two-dimensional barycentric subspace of 30 images from the cardiac cycles of three patients. Cubic splines interpolation of degree $n_\theta = 3$ with $N_\theta = 30$ control points is used.

The parameters a_0, a_1, and a_2 in the metric are chosen similarly to Sect. 4.1; however, the scale of the curves is not changed and we use equal weighting between the L^2-, H^1- and H^2-parts of the average energy for linear paths. This leads to parameters $a_0 = 1$, $a_1 = 0.1$, and $a_2 = 10^{-9}$. To see if the metric structure derived from the Sobolev metric enables us to distinguish between diseased patients and the control group, we compute all 171 pairwise distances between the 19 curves; this takes about 15 minutes on a 2 GHz single core processor. Multi-dimensional scaling of the distance matrix shows that the metric separates healthy and diseased patients quite well (Fig. 7a). Indeed, a cluster analysis based on the distance matrix recovers exactly – with exception of one outlier (patient 4) – the subgroups of healthy and diseased patients (Fig. 7b).

The Karcher means of the healthy and diseased subgroups as well as of the entire population are depicted in Fig. 8. The mean was computed using a gradient descent method as described in Sect. 3.2 with a threshold of 10^{-4} for the norm of the gradient. The average distance from the mean for the diseased group is 0.6853 with a variance of 0.0149, and for the control group the distance is 0.7708 with a variance of 0.0083.

To investigate the variability of the observed data, we performe principal component analysis on the initial velocities of the minimal geodesics connecting



(a) (b)

Fig. 7: (a) Two dimensional representation of the data using multi dimensional scaling of the pairwise distance matrix. (b) A dendrogram of clusters computed from the pairwise distance matrix using the single linkage criterion. Healthy patients are labelled 1–9 and diseased ones 10–19.

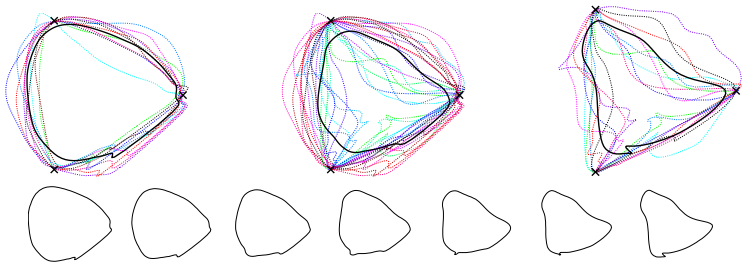


Fig. 8: First row: Karcher means of pathological cardiac cycles (left), all cycles (middle), and healthy cycles (right). Second row: geodesic connecting the Karcher mean of pathological cycles to the Karcher mean of healthy cycles. The crosses denote the position of images, with respect to whom the barycentric projection was computed.

curves to the respective means (c.f. Sect. 4.1). Fig. 9 shows the initial velocities projected to the subspace spanned by the first two principal directions. Within the healthy and sick subgroups, less then 30% of the principal components are needed to explain 90% of the shape variation. If, in contrast, principal components are analyzed for the entire dataset based on the global Karcher mean, then 35% of the principal components are needed to explain 90% of the shape variation.

5 Conclusions

In this article we numerically solved the initial and boundary value problems for geodesics on the space of regular curves under second order Sobolev metrics. We analyzed two medical datasets using our approach. In future work we plan to

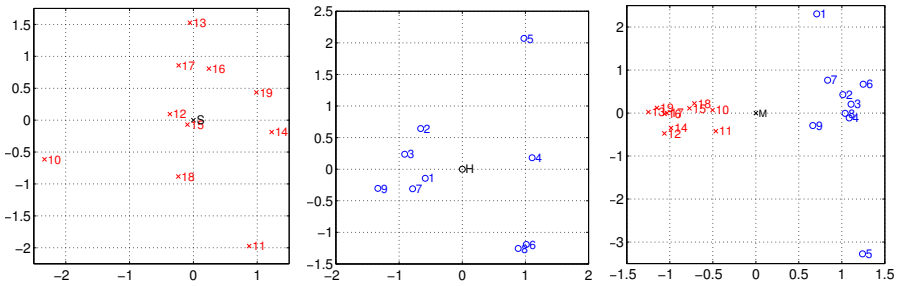


Fig. 9: (a) Initial velocities of minimizing geodesics projected to the subspace spanned by first two principal components for the diseased group. (b) The same picture for the control group. (c) The same picture for the whole population.

prove rigorous convergence results for our discretizations, further investigate the impact of the constants in the metric, and treat unparametrized curves.

Acknowledgments

We would like to thank Xavier Pennec and Marc-Michel Rohé for providing us the cardiac image data and Hermann Schichl for his invaluable help with AMPL.

References

1. Bauer, M., Bruveris, M.: A new Riemannian setting for surface registration. In: 3nd MICCAI Workshop on Mathematical Foundations of Computational Anatomy. pp. 182–194 (2011)
2. Bauer, M., Bruveris, M., Michor, P.W.: Overview of the geometries of shape spaces and diffeomorphism groups. J. Math. Imaging Vis. 50, 60–97 (2014)
3. Bauer, M., Bruveris, M., Michor, P.W.: R-transforms for Sobolev H^2-metrics on spaces of plane curves. Geom. Imaging Comput. 1(1), 1–56 (2014)
4. Bauer, M., Harms, P., Michor, P.W.: Sobolev metrics on shape space of surfaces. J. Geom. Mech. 3(4), 389–438 (2011)
5. Boland, M.V., Murphy, R.F.: A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells. Bioinformatics 17(12), 1213–1223 (2001)
6. Boumal, N., Mishra, B., Absil, P.A., Sepulchre, R.: Manopt, a Matlab toolbox for optimization on manifolds. J. Mach. Learn. Res. 15, 1455–1459 (2014)
7. Bruveris, M.: Completeness properties of Sobolev metrics on the space of curves. J. Geom. Mech. 7(2) (2015)
8. Bruveris, M., Michor, P.W., Mumford, D.: Geodesic completeness for Sobolev metrics on the space of immersed plane curves. Forum Math. Sigma 2, e19 (2014)
9. Dryden, I.L., Mardia, K.V.: Statistical Shape Analysis. Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester (1998)

10. Fourer, R., Gay, D., Kernighan, B.W.: The AMPL book (2002)
11. Krim, H., Yezzi, Jr., A. (eds.): Statistics and Analysis of Shapes. Modeling and Simulation in Science, Engineering and Technology, Birkhäuser Boston (2006)
12. McLeod, K., Sermesant, M., Beerbaum, P., Pennec, X.: Spatio-temporal tensor decomposition of a polyaffine motion model for a better analysis of pathological left ventricular dynamics. IEEE Trans. Med. Imaging (2015)
13. Michor, P.W., Mumford, D.: An overview of the Riemannian metrics on spaces of curves using the Hamiltonian approach. Appl. Comput. Harmon. Anal. 23(1), 74–113 (2007)
14. Mio, W., Srivastava, A., Joshi, S.: On shape of plane elastic curves. Int. J. Comput. Vision 73(3), 307–324 (Jul 2007)
15. Nardi, G., Peyré, G., Vialard, F.X.: Geodesics on shape spaces with bounded variation and Sobolev metrics. http://arxiv.org/abs/1402.6504 (2014)
16. Otsu, N.: A threshold selection method from gray-level histograms. IEEE T. Syst. Man Cyb. 9(1), 62–66 (1979)
17. Pennec, X.: Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements. J. Math. Imaging Vision 25(1), 127–154 (2006)
18. Pennec, X.: Barycentric subspaces and affine spans in manifolds (2015), to appear in the proceeding of Geometric Science of Information, 2015
19. Rohde, G.K., Ribeiro, A.J.S., Dahl, K.N., Murphy, R.F.: Deformation-based nuclear morphometry: capturing nuclear shape variation in HeLa cells. Cytometry Part A 73A(4), 341–350 (2008)
20. Rohde, G.K., Wang, W., Peng, T., Murphy, R.F.: Deformation-based nonlinear dimension reduction: applications to nuclear morphometry. In: 5th IEEE Int. Symposium on Biomedical Imaging: From Nano to Macro. pp. 500–503 (2008)
21. Rumpf, M., Wirth, B.: Variational time discretization of geodesic calculus. IMA Journal of Numerical Analysis (2014)
22. Schumaker, L.L.: Spline Functions: Basic Theory. Cambridge Mathematical Library, Cambridge University Press, Cambridge, third edn. (2007)
23. Srivastava, A., Klassen, E., Joshi, S.H., Jermyn, I.H.: Shape analysis of elastic curves in Euclidean spaces. IEEE T. Pattern Anal. 33(7), 1415–1428 (2011)
24. Tobon-Gomez, C., De Craene, M., Mcleod, K., Tautz, L., Shi, W., Hennemuth, A., Prakosa, A., Wang, H., Carr-White, G., Kapetanakis, S., Lutz, A., Rasche, V., Schaeffter, T., Butakoff, C., Friman, O., Mansi, T., Sermesant, M., Zhuang, X., Ourselin, S., Peitgen, H.O., Pennec, X., Razavi, R., Rueckert, D., Frangi, A.F., Rhode, K.: Benchmarking framework for myocardial tracking and deformation algorithms: an open access database. Medical Image Analysis 17(6), 632–648 (2013)
25. Wächter, A., Biegler, L.T.: On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Math. Program. 106(1), 25–57 (2006)
26. Younes, L., Michor, P.W., Shah, J., Mumford, D.: A metric on shape space with explicit geodesics. Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl. 19(1), 25–57 (2008)
27. Younes, L.: Spaces and manifolds of shapes in computer vision: an overview. Image Vision Comput. 30(6), 389–397 (2012)

Covariant un-reduction for curve matching

A. Arnaudon[1], M. Castrillón López[2], D.D. Holm[1]

[1] Dept. of Mathematics
Imperial College London
SW7 2AZ London, UK
[2] Dept. Geometría y Topología, Facultad de Ciencias Matemáticas
Universidad Complutense de Madrid
28040 Madrid, Spain

Abstract. The process of un-reduction, a sort of reversal of reduction by the Lie group symmetries of a variational problem, is explored in the setting of field theories. This process is applied to the problem of curve matching in the plane, when the curves depend on more than one independent variable. This situation occurs in a variety of instances such as matching of surfaces or comparison of evolution between species. A discussion of the appropriate Lagrangian involved in the variational principle is given, as well as some initial numerical investigations.

Keywords: shape analysis, curve matching, Lagrange-Poincaré reduction, covariant field theory

1 Introduction

The idea of un-reduction was introduced in [9] for the purpose of using reparametrisation by the action of the group $G = \mathrm{Diff}(S^1)$ to improve resolution of selected features in dynamics and optimal control problems, particularly in matching closed planar curves, whose configuration space Q is the space of embeddings $Q = \mathrm{Emb}(S^1, \mathbb{R}^2)$, by redistributing grid points in S^1 along the curve. The un-reduction process used in [9] was based on *reconstruction*, the inverse of Lagrangian reduction by symmetry [13], which relates the solutions on TQ/G to solutions on TQ. This version of the un-reduction process was applied in the outer metric setting in [9, 8]. In this setting, the deformation of the shape of a curve in $\mathrm{Emb}(S^1, \mathbb{R}^2)$ was applied to the *embedding* space, $\mathbb{R}^2$, completely independently of any reparametrisation of the embedded space, S^1.

In contrast, the un-reduction technique introduced in [4] seeks a family of equations, called the un-reduced equations, on TQ, whose solution projects onto those of a set of Euler-Lagrange equations on $T(Q/G)$. Thus, the un-reduction process used in [4] is distinct from the reconstruction process used in [9, 8]. Moreover, the un-reduction approach introduced in [4] raised an important issue for numerical applications in curve matching by optimal control, since it intertwined the reduction and reconstruction processes. Namely, the geodesic distance between two curves should be independent of their parametrisations. In particular,

the measure of the deformation of the shape of a curve should be independent of its parametrisation.

Here we address this issue for curve matching from the viewpoint of a reformulation of the un-reduction scheme of [4], which gives a framework in the setting of inner metric discussed in [3], in which the shape deformation is applied to the *embedded* space, rather than the *embedding* space.

In the process of addressing this issue, we will also generalise the un-reduction scheme by formulating it as a covariant space-time field theory. This generalisation gives us the freedom to introduce additional independent 'time' or 'space' variables for the purpose of coordinating comparisons among shapes. Introducing additional independent variables allows more flexibility in making shape comparisons than, for instance, the time warp approach of [10], which does not increase the number of independent variables. For example, one could imagine making comparisons of cylindrical surfaces by assembling closed curves resolved on two-dimensional slices. In this case, the additional space variable would be transverse to the slices, and one would make comparisons of surfaces as single entities, rather than comparing the evolution in time alone of each slice independently. In addition, the covariant field theory generalisation of the un-reduction framework in the inner metric setting could lead to a variety of other applications, a few of which are mentioned in Section 4.

This work may be summarized as an extension of [4] in the following three directions:

(1) We promote the un-reduction formulation of [4] in classical mechanics to a covariant field theory by following the same reasoning. Namely, we derive the Lagrange-Poincaré reduction of the shape space (Section 3.1) and un-reduce it by including an independent parametrisation (Section 3.2).
(2) Instead of the curvature weighted metric used in [4], we implement Sobolev metrics, which avoid the issue of arbitrary small geodesic distances (Section 3.3).
(3) We finish by illustrating this approach and assessing its validity with a few numerical experiments in the classical mechanical setting (Section 4.1).

A complete exposition of covariant un-reduction containing the proofs and applications in other areas, such as in theoretical physics, can be found in [1].

The main topic of the present paper is the first point listed above. We shall focus our discussion on the covariant Lagrange-Poincaré (LP) reduction by symmetry in the context of curve matching. The symmetry will be $G = \mathrm{Diff}(S^1)$, the diffeomorphism group which acts on the configuration space of planar curves, $Q = \mathrm{Emb}(S^1, \mathbb{R}^2)$, as a reparametrisation of S^1. LP reduction is a general method to deal with noncanonical reductions, in which the configuration space is not the symmetry group. LP reduction allows the explicit derivation of the dynamical equations on the quotient space, Q/G, which, in our case, is the space of shapes, $\mathrm{Emb}(S^1, \mathbb{R}^2)/\mathrm{Diff}(S^1)$. The field theoretic version of LP reduction that we will use here was developed in [5, 11], based on the classical reduction theory introduced in [7]. We will use a simplified version of this theory applied directly to curve matching.

The second improvement relative to [4] on the list above involves the Riemannian metric used to derive the un-reduction equations. As recently pointed out in [3, 2], use of Sobolev inner metrics avoids the problem of having arbitrarily small geodesic distances between two curves. In addition, the un-reduction equations take a simpler form with the Sobolev metric than with the curvature weighted metrics used in [4].

Finally, we tested the un-reduction approach in a few numerical experiments where we considered an initial value problem with reparametrisation. Even though we simply chose the forward Euler method for this initial value problem, without any further modifications, it still converged to the expected solution. The success of these simple numerical simulations motivates us to go further in future work to consider boundary value problems in the full field theoretical framework as explained in Section 4.2.

2 The geometry of curve matching

We start by recalling some basic facts about the geometry of curve matching that we will use throughout the text. We refer to earlier works such as [14, 3, 2, 4] for more details.

2.1 Reduction structure.

Let $Q = \mathrm{Emb}^+(S^1, \mathbb{R}^2)$ be the manifold of positive oriented embeddings from S^1 to $\mathbb{R}^2$. Elements in Q are maps $c(\theta) \in \mathbb{R}^2$ for $\theta \in S^1$ and elements in the tangent space T_cQ are pairs (c, v) with $c \in \mathrm{Emb}^+(S^1, \mathbb{R}^2)$ and $u \in C^\infty(S^1, \mathbb{R}^2)$ a parametrized vector field along the curve c, thus $TQ = Q \times C^\infty(S^1, \mathbb{R}^2)$.

We then consider the group $G = \mathrm{Diff}^+(S^1)$ of orientation preserving diffeomorphisms of S^1 and its Lie algebra $\mathfrak{g} = \mathfrak{X}(S^1)$. The group G acts on the right in $\mathrm{Emb}^+(S^1, \mathbb{R}^2)$ as reparametrisation of curves c and the reduced space is the space of shapes in $\mathbb{R}^2$

$$\Sigma := \frac{Q}{G} = \frac{\mathrm{Emb}^+(S^1, \mathbb{R}^2)}{\mathrm{Diff}^+(S^1)}. \tag{1}$$

The reduction of the phase space TQ by the group of symmetry G is a complicated space but can be decomposed via the introduction of a principal connection $\mathcal{A} : TQ \to \mathfrak{g}$ as follows

$$\frac{TQ}{G} \xrightarrow{\mathcal{A}} T\Sigma \oplus \tilde{\mathfrak{g}} = T\left(\frac{Q}{G}\right) \oplus \frac{Q \times \mathfrak{g}}{G}. \tag{2}$$

The space $T\Sigma$ is then the tangent space to the space of shapes, and the adjoint bundle $\tilde{\mathfrak{g}}$ will encode the parametrisation velocity of the curve. This space seems rather abstract but it corresponds to having a Lie algebra attached to each point of the base space Σ, which means that for each shape we have the freedom to attach an arbitrary parametrisation velocity. The construction of the connection

is straightforward for curve matching. Namely, given the velocity $u \in T_cQ$, we consider its tangent and normal decomposition,

$$u(\theta) = v(\theta)\mathbf{t}(\theta) + h(\theta)\mathbf{n}(\theta)\,, \tag{3}$$

where $(\mathbf{t}, \mathbf{n})$ is the orthonormal Frenet frame along c and $v(\theta), h(\theta) \in \mathfrak{X}(S^1)$ are scalar functions along the curve, parametrised by $\theta \in S^1$. We clearly have that $v(\theta)\mathbf{t}(\theta)$ is a vector tangent to the orbits of $G = \mathrm{Diff}^+(S^1)$. This decomposition defines the principal connection $\mathcal{A} : TQ \to \mathfrak{g}$, which, when applied to u, gives $\mathcal{A}(u) = v$, the reparametrisation velocity. The horizontal part of u is then $h(\theta)\mathbf{n}(\theta)$ and we have a decomposition $TQ = HQ \oplus VQ$. We will also need the curvature of $\mathcal{A}$, defined as $\mathcal{B} := d^{\mathcal{A}}\mathcal{A} = d\mathcal{A} + [\mathcal{A}, \mathcal{A}]$, but its exact form will not be needed here. In fact, we shall skip any technical details which are not directly useful for the present work, and refer the interested reader to [7, 5, 11] for the full discussion of this construction.

2.2 Field theoretical structure.

We can now extend this reduction structure by promoting the classical system to a covariant field theory, see [6, 11]. In order to do this, we consider an open domain $N \subset \mathbb{R} \times \mathbb{R}$ endowed with the Euclidean metric, the associated coordinates (t, x) and a volume form $\mathbf{v} = dtdx$. For simplicity we only consider a two dimensional space-time manifold N, but more dimensions can be added in a straightforward way. The tangent space TQ is then promoted to the jet bundle in order to capture the space-time direction used to compute a tangent vector. In this simple setting, the jet bundle has a simple geometric meaning, given by $J^1(N, Q) \simeq T^*N \otimes TQ$ and a generic element will be written

$$j^1c = c_t(\theta)(t, x)dt + c_x(\theta)(t, x)dx, \tag{4}$$

that is, $c_t, c_x \in TQ$ are the derivatives of a map $c : N \to Q$ along t and x respectively. The generalisation of the time derivative is the divergence operator, given in this simple case by $\mathrm{div}\, j^1c = \frac{d}{dt}c_t + \frac{d}{dx}c_x$.

2.3 Riemannian metrics

The choice of a convenient Riemannian metric on Q which is invariant with respect to the action of G is an interesting topic. See, for example, [2] and [3] and the references therein for more discussion. In our case, invariance under reparametrisation is achieved by considering arclength integrations and derivatives $ds = |c_\theta|d\theta$ and $D_\theta = \frac{1}{|c_\theta|}\partial_\theta$. However, the main difficulty lies in the geometrical properties of the metric, as we deal with an infinite dimensional space. The natural L^2 metric

$$g_{L^2}(u, v) = \int_{S^1} \langle u(\theta), v(\theta)\rangle ds, \tag{5}$$

with $u, v \in T_cQ$ such that $\langle \cdot, \cdot \rangle$ is a dot product in $\mathbb{R}^2$, is not very useful as it can lead to arbitrarily small geodesic distances in both Q and Q/G, see [14]. The problem can be overcome in the shape space Q/G by the metric

$$g_\kappa(u, v) = \int_{S^1} (1 + A\kappa(\theta)^2)\langle u(\theta), v(\theta)\rangle ds, \tag{6}$$

where $A > 0$ and κ denotes the Frenet curvature of the curve c, defined as

$$\kappa := (D_\theta D_\theta c) \cdot J(D_\theta c) = D_\theta \mathbf{t} \cdot \mathbf{n}\,. \tag{7}$$

In fact, the weighted metric in (6) can still have arbitrarily small geodesic distance in Q along the fibres of the fibration $Q \to Q/G$. A metric with a well defined Riemannian distance in both Q and Q/G may be obtained by adding higher order derivatives of u and v in a Sobolev-type expression as

$$g_{H^1}(u, v) = \int_{S^1} \left(\langle u(\theta), v(\theta)\rangle + A^2\langle D_\theta u(\theta), D_\theta v(\theta)\rangle\right) ds\,. \tag{8}$$

We can collect these three cases (as well as others, see [2]) by using

$$g_{\mathcal{P}}(u, v) = \int_{S^1} \langle u(\theta), \mathcal{P}v(\theta)\rangle, \tag{9}$$

for a convenient choice of a G-invariant self-adjoint pseudo-differential operator $\mathcal{P}$ which can depend on the curve and its derivatives. In particular, the operator for (6) is $\mathcal{P} = 1 + A\kappa^2$ and for (8) we have $\mathcal{P} = 1 - A^2 D_\theta^2$. One additional advantage of the operator associated to (8) is that it does not depend on the curve, whereas the operator for (6) depends on the curvature of the curve where it is evaluated. This represents a great simplification in the expression of the un-reduced equations.

3 Reduction and un-reduction

We are now ready to perform reduction by symmetry from the space of embeddings to the shape space using the covariant Lagrange-Poincaré reduction.

3.1 Lagrange-Poincaré reduction

Let's begin by recalling the original problem of curve matching. The matching problem is a boundary value problem in $\Sigma = Q/G$ with Lagrangian $l : T\Sigma \to \mathbb{R}$. Hamilton's principle states that the Euler-Lagrange equations associated with this Lagrangian yield the solution which minimises the action functional given by l. In practice, the matching problem is solved by using a shooting method for determining the initial momentum such that the curve at the final time matches the target curve within some specified tolerance. Instead, we will start in the larger space Q, where the numerical experiments can be done easily and, as a

first step, reduce this system such that we recover the Euler-Lagrange equations on Σ.

We project the variational principle defined for L from $J^1(N,Q)$ to its quotient $J^1(N,Q)/G = J^1(N,\Sigma) \oplus (T^*N \otimes \tilde{\mathfrak{g}})$, where $\tilde{\mathfrak{g}} = (Q \times \mathfrak{g})/G$, as in equation (2). Critical solutions are maps $\sigma : N \to T^*N \otimes \tilde{\mathfrak{g}}$ which, moreover, project to maps $\rho : N \to \Sigma = Q/G$ as $\rho = \pi_{\tilde{\mathfrak{g}}} \circ \sigma$ according to the diagram

$$\begin{array}{ccc} & & T^*N \otimes \tilde{\mathfrak{g}} \\ & \overset{\sigma}{\nearrow} & \downarrow{\scriptstyle \pi_{\tilde{\mathfrak{g}}}} \\ N & \xrightarrow{\rho} & \Sigma \end{array} \tag{10}$$

where $\pi_{\tilde{\mathfrak{g}}} : T^*N \otimes \tilde{\mathfrak{g}} \to \Sigma$ is the projection of the adjoint bundle neglecting the T^*N factor. The free variations of the initial problem provide a family of constrained variations that define a new type of variational equations, called Lagrange-Poincaré equations, [5], [11]. The next theorem gives the Lagrange-Poincaré reduction which includes forces $F : T^*N \otimes TQ \to TQ$. Before stating this theorem without proof, we will make another important assumption which is satisfied by most of the Lagrangians used in curve matching. Namely, we assume our Lagrangian decomposes as a sum of two Lagrangians taking values from the vertical and horizontal space that will be denoted $L = L^h + L^v$ and $\ell = \ell^h + \ell^v$.

Theorem 1 (Covariant Lagrange-Poincaré reduction with forces). *Given a map $c : N \to Q$, let $\sigma : N \to T^*N \otimes \tilde{\mathfrak{g}}$ be defined as*

$$\sigma(x)(\omega) = [s(x), \mathcal{A}(Ts \cdot (\omega))]_G \,, \tag{11}$$

with $\omega \in T_xN, x \in N$ and where $[\cdot]_G$ stands for the quotient by G; $\rho : N \to \Sigma$, $\rho(x) = [s(x)]_G = \pi_{\tilde{\mathfrak{g}}} \circ \sigma$. With the previous definitions, the following points are equivalent

(1) s is a critical mapping of the variational principle

$$\delta \int_N L(s, j^1 s)\mathbf{v} + \int_N \langle F(s, j^1 s), \delta s\rangle \mathbf{v} = 0 \tag{12}$$

with free variations δs.

(2) The Euler-Lagrange form of L satisfies the relation

$$\mathcal{EL}(L\mathbf{v})(j^2 s) = F,$$

where $\mathcal{EL}$ is the Euler-Lagrange operator acting on the second jet bundle (second order field theoretical tangent space), which gives the usual Euler-Lagrange equations.

(3) $\sigma : N \to T^*N \otimes \tilde{\mathfrak{g}}$ is a critical mapping of the variational principle

$$\delta \int_N \ell(j^1\rho, \sigma)\mathbf{v} + \int_N \langle f^h(j^1\rho, \sigma), \delta\rho\rangle \mathbf{v} + \int_N \langle f^v(j^1\rho, \sigma), \eta\rangle \mathbf{v} = 0,$$

for variations of the form $\delta\sigma = \nabla\eta - [\sigma, \eta] + \mathcal{B}(\delta\rho, T\rho) \in \tilde{\mathfrak{g}}$, where $\delta\rho \in T_\rho\Sigma$ is a free variation of ρ and η is a free section of $\tilde{\mathfrak{g}} \to \Sigma$.

(4) σ satisfies the Lagrange-Poincaré equations, written if L and ℓ decomposes according to the vertical/horizontal decomposition

$$\left.\begin{aligned} \operatorname{div}\left(\frac{\delta \ell^h}{\delta j^1 \rho}\right) - \frac{\delta \ell^h}{\delta \rho} = f^h + \frac{\delta \ell^v}{\delta \rho} - \left\langle \frac{\delta \ell^v}{\delta \sigma}, \mathcal{B}(T\rho, \cdot) \right\rangle, \\ \operatorname{div} \frac{\delta \ell^v}{\delta \sigma} + \operatorname{ad}^*_\sigma \frac{\delta \ell^v}{\delta \sigma} = f^v. \end{aligned}\right\} \tag{13}$$

One recognises left hand side of the first equation in (13) as an Euler-Lagrange equation and the second one as an Euler-Poincaré equation. The right hand side of both equations are either forces, or coupling with between them. The solution σ of the Euler-Poincaré equation in (13) will influence the Euler-Lagrange equation via the term involving the curvature of the connection $\mathcal{A}$. An additional coupling arises because σ is in the adjoint bundle and therefore depends implicitly on the base curve in Q.

3.2 Un-reduction

The particular form of the equations in (13), based on the decomposition of the Lagrangian and the inclusion of the force term will allow us to decouple these equations in the sense that the right hand side of the EL equation will vanish; so the feedback of the EP equation to the EL equation will disappear. Before stating the un-reduction theorem we must recall that the canonical momentum map $\mathbf{J} : T^*Q \to \mathfrak{g}^*$ for the natural lift action of G on T^*Q, is defined by

$$\langle \mathbf{J}(\alpha_q), \xi \rangle_{\mathfrak{g} \times \mathfrak{g}^*} = \langle \alpha_q, \xi_Q \rangle_{TQ \times TQ^*},$$

where $\alpha_q \in T^*Q$, $\xi \in \mathfrak{g}$, and $\xi_Q \in TQ$ is the infinitesimal transformation of the action of G on Q at the point $q \in Q$. The map $\mathbf{J}$ extends to a map $\mathbf{J} : TN \otimes T^*Q \to TN \otimes \mathfrak{g}^*$, trivially in the factor TN.

We can finally state the covariant un-reduction theorem. We refer to [1] for the proof and more details about this theorem.

Theorem 2. *We consider a G-equivariant force $F : J^1(N, Q) \to T^*Q$ such that $F^v = p^v \circ F$ is arbitrary and $F^h = p^h \circ F$ is given by the condition*

$$f^h = -\frac{\nabla \ell^v}{\delta \rho} + \left\langle \frac{\delta \ell^v}{\delta \sigma}, \mathcal{B}(T\rho, \cdot) \right\rangle, \tag{14}$$

for its projection $f^h : J^1(N, \Sigma) \times (T^*N \otimes \tilde{\mathfrak{g}}) \to T^*\Sigma$. Then, the variational equations of the problem defined by L and F read

$$\left.\begin{aligned} \mathcal{EL}(L^h)(j^2 s) = 0 \\ \mathcal{A}^* \operatorname{div}\left(\mathbf{J}\left(\frac{\delta L^v}{\delta j^1 s}\right)\right) = F^v(j^1 s), \end{aligned}\right\} \tag{15}$$

where $\mathcal{A}^* : \mathfrak{g}^* \to V^*Q$ is the dual of the connection form. Finally, critical solutions $s : N \to Q$ of (15) project to critical solutions $\rho = [s]_G$ of the Euler-Lagrange equations $\mathcal{EL}(l)(j^2 \rho) = 0$.

Remark 1. For $N = \mathbb{R}$, $\mathbf{v} = dt$, that is, in the case of classical mechanics, we have $\mathrm{div} = d/dt$ and we recover the results and equations of [4].

The first equation in (15) is the usual Euler-Lagrange equation for the horizontal Lagrangian L^h, needed for solving the matching problem. Regarding the interpretation of the second equation, the definition of $\mathbf{J}$ above shows that $\mathbf{J}(\delta L^v/\delta j^1 s)$ is a covariant momentum map, so that $\mathrm{div}\,\mathbf{J}(\delta L^v/\delta j^1 s)$ is the expression of a conservation law with respect to the group of symmetries. If one set $F^v = 0$, the conservation law is complete. However, sometimes it may be interesting to keep this vertical force, as it might be used to externally control the dynamics along the vertical space; that is, the reparametrisation.

3.3 Un-reduction with Sobolev metric

We consider the $\mathrm{Diff}^+(S^1)$-invariant Lagrangian $L : J^1(N,Q) \simeq T^*N \otimes TQ \to \mathbb{R}$ which can be decomposed as $L = L^h + L^v$ with respect to the connection $\mathcal{A}$ as

$$L^h(j^1 c) = \frac{1}{2}\int_{S^1} (h_t \mathcal{P} h_t + h_x \mathcal{P} h_x)\, ds, \quad L^v(j^1 c) = \frac{1}{2}\int_{S^1} (v_t \mathcal{P} v_t + v_x \mathcal{P} v_x)\, ds,$$

where

$$c_t = v_t \mathbf{t} + h_t \mathbf{n} \quad \text{and} \quad c_x = v_x \mathbf{t} + h_x \mathbf{n}.$$

The un-reduction equations (15) are then computed in Proposition 1 below in the case when $\mathcal{P}$ is independent of the curve, that is, Sobolev metrics.

Proposition 1. *The un-reduced equations* (15) *for the bi-dimensional problem of planar simple curves defined by the Lagrangian defined above and the metric* (9) *with* $\mathcal{P}$ *being the Sobolev operator, read*

$$\begin{aligned} \partial_x \mathcal{P} h_x + \partial_t \mathcal{P} h_t &= D_\theta (h_x \mathcal{P} v_x + h_t \mathcal{P} v_t) - \kappa H \\ \partial_x \mathcal{P} v_x + \partial_t \mathcal{P} v_t &= F^v \end{aligned} \tag{16}$$

for any choice of vertical force F^v*, where*

$$H = \frac{1}{2}\left(h_x \mathcal{P} h_x + h_t \mathcal{P} h_t\right). \tag{17}$$

In (16), the function H is the shape kinetic energy density and κ is the Frenet curvature of the curve c, defined in (7). This term can be interpreted as a penalty term in deforming curved regions. The sign of this term would depend on the concavity or convexity of the curve at this point, and thus this force would try to prevent the curve to be deformed too fast in these regions. Equation (16) shows that the dynamics in (x,t) is governed by the coupling between h_t and v_t required for the shape deformation to be independent of the reparametrisation. In fact, it also contains a derivative with respect to θ and vertical vectors. This is the only term which couples with the vertical equation, and it also gives the corrections necessary for the curve deformation to be independent of the reparametrisation.

Remark 2. In the classical mechanics setting, the un-reduction equations with Sobolev metric would be very similar to (16), but without the x-dependent terms. They will be the equations used for the numerics in section 4.1. Owing to the simplicity of the Sobolev metric compared to the curvature weighted metric, the derivation of this equation is directly done from the un-reduction equations, not from the variational principle, as in [4]. We refer to [1] for the details of this calculation.

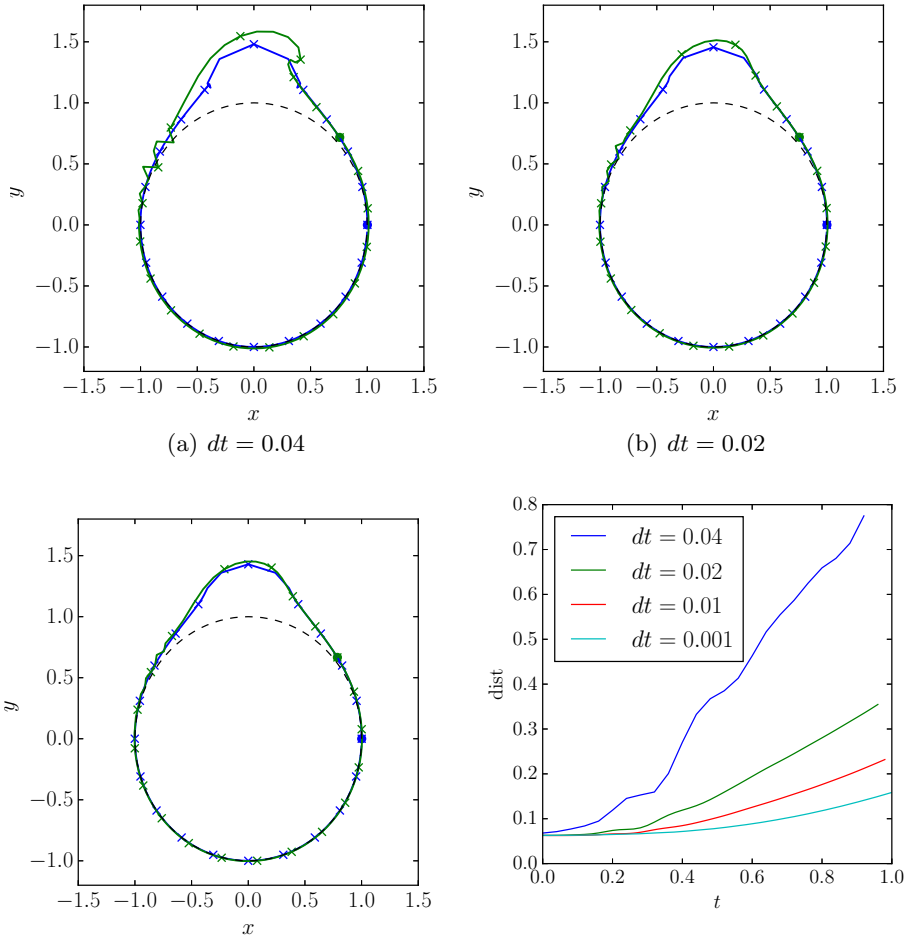
4 Applications

Before discussing the possible applications of the covariant un-reduction scheme in curve matching, we shall present a short numerical study of the un-reduction in classical mechanics using the Sobolev metric H^1.

4.1 Numerical validation

In order to test and illustrate the un-reduction scheme, we performed a few simple numerical experiments. We restricted ourselves to the classical case, already done theoretically in [4], but with the H^1 norm instead of the curvature weighted norm. The only effect of the Sobolev norm that which interested us is that it regularises the curve deformation and prevents large bending of the curve, smaller that the scale given by A. A deeper analysis of the effect of the Sobolev norm in the matching process is not the aim of this paper but is important for applications. Our main goal here was to check the decoupling between the shape and the reparametrisation dynamics for a simple initial value problem.

Our numerical scheme used the Euler explicit scheme in time and 2nd order centred finite difference approximation for D_θ in order to have a symmetric space discretisation in $\theta \leftrightarrow -\theta$. The application of the Sobolev operator $\mathcal{P} = 1 - A^2 D_\theta^2$ was done in Fourier space with $A = 0.3$. Our initial condition was a circle and the initial horizontal velocity was a bump function, so the curve deformed as in Fig. 1. The curve was discretised with 100 points and we used a set of time steps ($dt = 0.04, 0.02, 0.01, 0.001$) to study the convergence of the scheme, and especially the decoupling between the reparametrisation and the shape deformation. In order to do this, we ran two initial value problems where one of them also had a vertical initial velocity, taken to be constant such that the parametrisation would rotate during the evolution of the curve. We then computed the distance between the two curves using the methods of currents [12] at each time to make a parametrisation-free comparison of the shapes of the curve. The results are displayed in Fig. 1 together with the distance between the curves as a function of time. Even with the simple numerics we used (the Euler scheme and finite difference), the un-reduction feature was verified. The example we studied is simple and did not require high resolutions and robustness tests as for more realistic shapes. Further numerical studies using the un-reduction scheme would thus include improvements of the current implementation and a shooting method in order to solve the correct matching, or boundary value problem.



(a) $dt = 0.04$

(b) $dt = 0.02$

(c) $dt = 0.01$

(d) Distance between fixed and reparametrised curves with the method of currents

Fig. 1. This figure illustrates the effect of the time step of S^1-reparameterisation on the quality of the deformation of a curve using the explicit Euler scheme in the un-reduction approach. The blue curve has horizontal initial velocity h_t for shape deformation, while the green curve has an additional constant vertical initial velocity v_t, for reparameterisation. The blue and green dots begin at the same initial point, but then the green one shifts along the green curve as the reparameterisation proceeds. The black dashed curve is the initial condition of the simulation. Upon decreasing the time steps, the coupling between the vertical and horizontal dynamics decreases and the quality of the deformation improves, even when using an explicit Euler scheme. On the bottom right panel, we plotted the distance between the two curves, as a function different time resolutions. The distance is computed by the method of currents, which is independent of the parametrisation.

4.2 Covariant matching

In the simplest case of field theory treated here, namely a two dimensional space-time, two main applications for boundary value problems (BVP) present themselves for further discussion. Besides BVPs, initial value problems (IVP) could also be considered, but IVP are not of great interest for curve matching. We will thus forego discussing them here, although a possible application would be to predict the evolution of a particular model, knowing that it should "roughly" follow a generic model. In this case, the initial value problem must have carefully chosen initial values, a subject which is out of the scope of this work. The two applications for BVP that we will discuss are the following:

(1) Matching between cylindrical surfaces, and
(2) Spatio-temporal analysis.

Case (1) In the first case, a set of slices along a cylindrical surface (a typical example would be a bone) are given, where x is the parameter along the main axis of the surface. For the sake of simplicity, we will just consider two slices, but more could be added without to much trouble. The first step would be to generate the initial and final conditions, namely use un-reduction for the IVP to interpolate between the two curves and generate the initial and final surfaces. Once this is done, covariant un-reduction could be applied using a shooting method in time, such that the solution is a critical point of the action functional $\int L(c, c_t, c_x)dtdx$. In our simple case, where the Lagrangian is purely quadratic, the solution would be a harmonic map, or a minimal surface, and would then require more advanced mathematics, beyond the present discussion. This model would compute the distance between two surfaces, taking into account that the interpolation between the slices in space should be imposed simultaneously with the matching in time. The resulting distance will be different than a naive model, which would compute the matching in time, slice by slice. For an illustration of matching slice by slice, we refer to the last example in [8] where a surface representing a nasal cavity is reconstructed out of a set of slices. The step done there corresponds to the generation of initial and final surfaces, whereas covariant un-reduction would compute the distance between these two surfaces, during the temporal deformation of the entire nasal cavity.

Case (2) Spatio-temporal analysis has recently been reviewed in [10], from yet another viewpoint. Indeed, the matching in space done in [10] does not depend on a space parameter, but is *instantaneous*, namely given by a single map between the two curves. They also included a "time warp" which account for the change of pace of the evolution of the two models to be compared. In our case, the spatial variable comes into play on the same footing as time, and, thus, brings more flexibility into the comparison. Again, the theory of harmonic maps could help in understanding the properties of the solutions, and it is possible that the time warp reparameterisation could be recovered as well.

We finally want to mention the freedom to choose the vertical force in the un-reduction equations. This force could be used to control the parametrisation during the matching. Different types of forces could be considered, such as a

force which would optimally redistribute the parametrisation in different regions of the curve, such that the number of points for discretising the curve could be reduced. Another force could be used to match the paramerisation of the target curve, such that the computationally more expensive method of currents could be avoided.

Acknowledgements We are grateful to M. Bauer, M. Bruveris, L. Younes, S. Durrleman, R. Montgomery and T. Ratiu for valuable discussions during the course of this work. AA acknowledges partial support from an Imperial College London Roth Award, AA and DH from the European Research Council Advanced Grant 267382 FCCA. MCL has been partially funded by MINECO (Spain) under projects MTM2011-22528 and MTM2010-19111. MCL wants to thank Imperial College for its hospitality during the visit in which the main ideas of this work were developed.

References

1. A. Arnaudon, M. Castrillón López, D.D. Holm, *Covariant un-reduction and applications*, in preparation.
2. M. Bauer, M. Bruveris, S. Marsland, P.W. Michor, *Constructing reparameterization invariant metrics on spaces of plane curves*, Differential Geom. Appl. **34** (2014), 139–165.
3. M. Bauer, M. Bruveris, P.W. Michor, *Overview of the Geometries of Shape Spaces and Diffeomorphism Groups*, J. Math. Imaging and Vision **50** (2014), 67–90.
4. M. Bruveris, D.C.P. Ellis, F. Gay-Balmaz, D.D. Holm, *Un-reduction*, Journal of Geometric Mechanics **3** (2011), 363–387.
5. M. Castrillón López, T.S. Ratiu, *Reduction in Principal Bundles: Covariant Lagrange-Poincaré Equations*, Comm. Math. Phys. **236** (2003), 223-250.
6. M. Castrillón López, P.L. García, T.S. Ratiu, *Euler–Poincaré reduction on principal bundles*, Lett. Math. Phys. **58** (2001), 167–180.
7. Cendra, Hernán and Marsden, Jerrold E and Ratiu, Tudor S *Lagrangian reduction by stages*, American Mathematical Soc. **722**, (2001), x+108
8. C.J. Cotter, A. Clark, J. Peiró, *A reparameterisation based approach to geodesic constrained solvers for curve matching*, Int. J. Comput. Vis. **99**, (2012), 103121
9. C.J. Cotter and D.D. Holm, *Geodesic boundary value problems with symmetry*, J. Geom. Mech. **2**, no. 1 (2010), 417–444.
10. S. Durrleman, X. Pennec, A. Trouvé, J. Braga, G. Gerig, N. Ayache, *Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data*, Int J Comput Vis. **103**, no. 1 (2013), 22–59.
11. D.C. Ellis, F. Gay-Balmaz, D.D. Holm, T.S. Ratiu, *Lagrange-Poincaré field equations*, Journal of Geometry and Physics **61**, no. 11, (2011), 2120–2146.
12. J. Glaunes, A. Trouve, L. Younes, *Modeling planar shape variation via hamiltonian flows of curves*, in H. Krim Jr. and Y.A (Eds.), Statistics and analysis of shapes (2006), Basel Birkhäuser.
13. D.D. Holm, J.E. Marsden and T.S. Ratiu, The Euler–Poincaré equations and semidirect products with applications to continuum theories, *Adv. in Math.*, **137**, (1998), 1–81.
14. P.W. Michor, D. Mumford, *An overview of the Riemannian metrics on spaces of curves using the Hamiltonian approach*, Appl. Comput. Harmon. Anal. **23** (1) (2007), 74–113.

Kernel Metrics on Normal Cycles and Application to Curve Matching

Pierre Roussillon and Joan Alexis Glaunès

MAP5, UMR 8145 CNRS, Université Paris Descartes, Sorbonne Paris Cité, France
pierre.roussillon@parisdescartes.fr alexis.glaunes@parisdescartes.fr

Abstract. In this work we introduce a new dissimilarity measure for shape registration using the notion of normal cycles, a concept from geometric measure theory which allows to generalize curvature for non smooth subsets of the euclidean space. Our construction is based on the definition of kernel metrics on the space of normal cycles which take explicit expressions in a discrete setting. This approach is closely similar to previous works based on currents and varifolds [13,5]. We derive the computational setting for discrete curves in $\mathbb{R}^3$, using the Large Deformation Diffeomorphic Metric Mapping framework as model for deformations. We present synthetic experiments and compare with the currents and varifolds approaches.

Introduction

Many applications in medical image analysis require a coherent alignment of images as a pre-processing step, using efficient rigid or non-rigid registration algorithms. Moreover, in the field of computational anatomy, the estimation of optimal deformations between images, or geometric structures segmented from the images, is a building block for any statistical analysis of the anatomical variability of organs. Non-rigid registration is classically tackled down by minimizing a functional composed of two terms, one enforcing regularity of the mapping, and the data-attachment term which evaluates dissimilarity between shapes. Defining good data-attachment terms is important, as it may improve the minimization process, and focus the registration on the important features of the shapes to be matched.

In [13,9] a new framework for dissimilarity measures between sub-manifolds was proposed using kernel metrics defined on spaces of *currents*. This setting is now commonly used in computational anatomy ; its advantages lie in its simple implementation and the fact that it provides a common framework for continuous and discrete shapes (see [7] for a computational analysis of currents and their numerical implementation). However, currents are oriented objects and thus a consistent orientation of shapes is needed for a coherent matching. Moreover, due to this orientation property, artificial cancellation can occur with shapes with high local variations. To deal with this problem, a more advanced model based on *varifolds* has been introduced recently [4]. Varifolds are measures over fields of non-oriented linear subspaces. See [4], chap. 3 for an exhaustive analysis.

In this work, we propose to use a second-order model called *normal cycle* for defining shape dissimilarities. The normal cycle of a submanifold X is the current associated with its normal bundle $\mathcal{N}_X$. The normal cycle encodes second order, i.e. curvature information of X; more precisely one can compute integrals of curvatures by evaluating the normal cycle over simple differential forms. Moreover, it has a canonical orientation which is independent of the orientation of X (in fact X does not need to be oriented).

Our approach is closely related to the currents and varifolds models in that it is based on the definition of kernel metrics that take explicit form in a discrete setting. This paper is organized as follows : in Sect. 1 and 2 we introduce the mathematical notions of currents and normal cycles and define the kernel metric in a general setting. In Sect. 3 we derive explicit formulas for the metric in the case of discrete curves in $\mathbb{R}^3$. In Sect. 4 we introduce the general curve matching problem and recall some basic facts about the diffeomorphic model. Finally we present two sets of synthetic experiments in Sect. 5.

1 Currents and Normal Cycles

1.1 Currents

We recall here the definition of current used in [7], definition 1.2.

Definition 1 (Currents).
The space of m-currents in $\mathbb{R}^d$ is defined as the topological dual $\Omega_0^m(\mathbb{R}^d)'$, where $\Omega_0^m(\mathbb{R}^d) := \mathcal{C}_0^0(\mathbb{R}^d, (\Lambda^m\mathbb{R}^d)^*)$ is the space of continuous m-differential forms vanishing at infinity, with the supremum norm : $\|\omega\|_\infty = \sup_{x\in\mathbb{R}^d} |\omega_x|$. A m-current is thus a linear map $T : \Omega_0^m(\mathbb{R}^d) \to \mathbb{R}$ such that there exists $C_T > 0$ such that for every differential form ω,

$$T(\omega) \leq C_T \|\omega\|_\infty$$

Example 1. A fundamental example of current (which will be useful when dealing with discrete shape) is the "Dirac" current. Let $x \in \mathbb{R}^d$, $\alpha \in \Lambda^m(\mathbb{R}^d)$. For $\omega \in \Omega_0^m(\mathbb{R}^d)$, we define $\delta_x^\alpha(\omega) := \omega_x(\alpha)$.

Any sufficiently regular shape in $\mathbb{R}^d$ can be seen as a current. Let $\mathcal{H}^m$ be the m-dimensional Hausdorff measure in $\mathbb{R}^d$. If X is a regular m-dimensional submanifold, $\mathcal{H}^m$ coincides on X with the volume form of X. Assume X is a compact, oriented, m-rectifiable set (see definition in [8]). Then the tangent space T_xX exists for $\mathcal{H}^m$-almost every $x \in X$, and one can associate to X a current $[X]$, defined as :

$$[X](\omega) := \int_X \omega_x(\tau_X(x)) d\mathcal{H}^m(x) \tag{1}$$

where $\tau_X(x) = \tau_1(x) \wedge \cdots \wedge \tau_m(x)$, with $(\tau_i(x))_{1\leq i\leq m}$ an orthonormal basis of T_xX. If we consider the opposite orientation of $X : \tilde{X}$, we have $[\tilde{X}] = -[X]$.

1.2 Normal Cycles

Normal cycles find their roots in the seminal work of Federer. In [8], he proved that for a set with positive reach (see definition below), the volume of the ε-parallel neighbour of $X \cap B$ (where B is a borelian) can be expressed as a polynomial of ε; and more importantly that the coefficients of this polynomial can be interpreted as *curvature measures* of the set X. These measures have integral representation, and Zähle in [15] introduced $d-1$ generalized principal curvatures for sets with positive reach, and retrieved Federer's curvature measures by integrating functions of these principal curvatures over the unit normal bundle. She showed that this can be done by integrating adequate differential forms on the associated current : this is exactly the normal cycle. This work was pushed forward in [16]. The book of Morvan ([12]) is a self-sufficient reference for normal cycles as they will be used in this paper. Note that normal cycles have already been applied for computational analysis of discrete surfaces in [6]. Cohen-Steiner and Morvan derive a definition of discrete curvature and discrete curvature tensor for polyhedral surfaces based on the normal cycle.

We follow Federer [8] to give a definition of the normal bundle and the normal cycle in the context of sets with positive reach.

Definition 2 (Sets with positive reach). *Let $X \subset \mathbb{R}^d$ and for any $\varepsilon > 0$, denote $X_\varepsilon = \{x \in \mathbb{R}^d, d(x, X) \leq \varepsilon\}$ the ε tube around X. The reach of X is the supremum of $r > 0$ such that for every $0 < \varepsilon \leq r$, there exists a unique projection of $x \in X_\varepsilon$ onto X. X is said to be a* positive reach set *if $r > 0$.*

On a set with positive reach R, one can roll a ball of radius less than R. Thus, a set with positive reach can be seen heuristically as a set with a bounded below curvature.

Definition 3 (Unit Normal bundle). *Let X be a set with positive reach. We define the* unit normal cone *at $x \in X$ and the* unit normal bundle *repectively as*

$$Nor(X, x) = \{u \in \mathbb{R}^d \mid \exists \varepsilon > 0 \mid \forall y \in B(x, \varepsilon) \cap X, \langle y - x, u \rangle \leq 0\} \cap S^{d-1},$$
$$\mathcal{N}_X = \{(x, n),\ x \in X,\ n \in Nor(X, x)\}.$$

For a $\mathcal{C}^2$-submanifold, the unit normal bundle defined here coincides with the classical one, which is a $(d-1)$-submanifold in the $(2d-1)$ dimensional manifold $\mathbb{R}^d \times S^{d-1}$. More generally, $\mathcal{N}_X$ is a $(d-1)$-rectifiable set in $\mathbb{R}^d \times S^{d-1}$ when X has positive reach and ∂X_ε is a $(d-1)$-dimensional differentiable submanifold (hypersurface), with Lipschitzian unit normal vector field (see again [8], 4.8).

Example 2 (Unit normal bundle of a curve in $\mathbb{R}^3$). Let $\gamma : [0, L] \to \mathbb{R}^3$ be a parametrized curve in $\mathbb{R}^3$, and suppose γ is $\mathcal{C}^2$ on $[0, L]$, with $\gamma'(t) \neq 0$, $\forall t \in [0, L]$. On a regular part of the curve (i.e. $\gamma(t)$, $0 < t < L$), the normal cone is simply $\gamma'(t)^\perp \cap S^2$ (note that for a segment, the normal bundle is thus a cylinder). For the singular part (i.e. the two endpoints), we denote $S_v^+ := \{u \in S^2, \langle u, v \rangle \geq 0\}$. One can easily show that the normal cone at $\gamma(0)$ and $\gamma(1)$ are $\{\gamma(0)\} \times S^+_{-\gamma'(0)}$ and $\{\gamma(1)\} \times S^+_{\gamma'(1)}$ respectively. These are two half spheres with a coherent orientation with respect to the normal bundle (independent of the parametrization).

Canonical orientation of $\mathcal{N}_X$. ∂X_ε being a closed hypersurface, it is canonically oriented and this induces a canonical orientation on the unit normal bundle of X as follows : let $(e_1, \dots, e_d)$ be an orthonormal basis of $\mathbb{R}^d$, and let $(\tau_1(x,n), \dots, \tau_{d-1}(x,n))$ an orthonormal basis of $T_{(x,n)}\mathcal{N}_X$ such that

$$\left\langle (\pi_p + \varepsilon\pi_n)(\tau_1(x,n)) \wedge \cdots \wedge (\pi_p + \varepsilon\pi_n)(\tau_{d-1}(x,n)) \wedge n \, , \, e_1 \wedge \cdots \wedge e_d \right\rangle > 0, \quad (2)$$

where π_p denotes the projection on the spatial space, and π_n the projection on the normal space : $\pi_p : (x,n) \in \mathbb{R}^d \times \mathbb{R}^d \mapsto x$, $\pi_n : (x,n) \in \mathbb{R}^d \times \mathbb{R}^d \mapsto n$. Then $\tau_{\mathcal{N}_X}(x,n) := \tau_1(x,n) \wedge \cdots \wedge \tau_{d-1}(x,n)$ is a simple $(d-1)$-vectorfield orienting $\mathcal{N}_X$. Expression (2) is independent of $0 < \varepsilon < \operatorname{Reach} X$. Besides, the orientation of the normal bundle of X does not depend on any orientation of X.

Definition 4 (Normal cycle). *The normal cycle of a positive reach set $X \subset \mathbb{R}^d$ is the $(d-1)$-current associated with $\mathcal{N}_X$ with its canonical orientation. For any differential form $\omega \in \Omega_0^{d-1}(\mathbb{R}^d \times \mathbb{R}^d)$, one has :*

$$N(X)(\omega) := [\mathcal{N}_X](\omega) = \int_{\mathcal{N}_X} \omega_{(x,n)}(\tau_{\mathcal{N}_X}(x,n)) d\mathcal{H}^{d-1}(x,n) \quad (3)$$

Normal cycles for unions of sets with positive reach The theory of normal cycles can be extended to the case of finite unions of sets with positive reach, as done in [16]. This allows to define normal cycles for a very large class of subsets. In particular it allows to define normal cycles for unions of segments, which will be used as our discrete models for curves. This extension can be stated as follows :

Theorem 1 (Additive property). *There is a unique extension of normal cycles to finite unions of sets with positive reach such that the following property holds for any two such sets C, S :*

$$N(C \cup S) := N(C) + N(S) - N(C \cap S) \quad (4)$$

In fact this additive property holds for sets with positive reach, and is used recursively as definition in the case of unions of such sets. It can be shown that this definition is fully coherent.

2 Metrics on Normal Cycles

For our numerical purpose, we need a computable expression for the dissimilarity between shapes. In the very same spirit of [9,7,4], we will use a dual kernel metric on normal cycles as dissimilarity measure. This can be done by considering a kernel $K : (\mathbb{R}^d \times S^{d-1})^2 \to \mathcal{B}(\Lambda^{d-1}(\mathbb{R}^d \times \mathbb{R}^d), \mathbb{R})$ of the form

$$K((x,u),(y,v))(\alpha,\beta) := k_p(x,y) k_n(u,v) \langle \alpha, \beta \rangle$$

where $x, y \in \mathbb{R}^d$, $u, v \in S^{d-1}$, $\alpha, \beta \in \Lambda^{d-1}(\mathbb{R}^d \times \mathbb{R}^d)$, and:

- $k_p(x,y)$ is a positive definite kernel on $\mathbb{R}^d$. In our experiments we used $k_p(x,y) = \exp\left(\frac{-|x-y|^2}{\sigma_W^2}\right)$, where σ_W is a scale parameter.

- $k_n(u,v)$ is a positive definite kernel on S^{d-1}. We used the reproducing kernel of a Sobolev space $H^s(S^{d-1})$ of order s. This kernel is in fact a scalar function of the distance between two vectors of S^{d-1}. Therefore it is rotation invariant and can be expanded in spherical harmonics, which will be useful for the numerical aspect.

- $\langle \cdot, \cdot \rangle$ is a scalar product between $d-1$-vectors in $\mathbb{R}^d \times \mathbb{R}^d$. We used a modification of the canonical scalar product between $d-1$-vectors that takes into account the parameter σ_W to ensure invariance of the kernel metric when a rescaling is applied jointly to the data coordinates and σ_W. It is defined by

$$\langle u_1 \wedge \cdots \wedge u_{d-1}, v_1 \wedge \cdots \wedge v_{d-1} \rangle := \det((\langle u_i, v_j \rangle_{\sigma_V})_{1 \leq i,j \leq d-1})$$

where $\langle u, v \rangle_{\sigma_V} := \frac{1}{\sigma_V^2} \langle \pi_p(u), \pi_p(v) \rangle_{\mathbb{R}^d} + \langle \pi_n(u), \pi_n(v) \rangle_{\mathbb{R}^d}$.

This kernel K defines a Reproducing Kernel Hilbert Space (RKHS) W of $(d-1)$-differential forms in $\mathbb{R}^d \times S^{d-1}$ and $\iota : W \hookrightarrow \Omega_0^{d-1}(\mathbb{R}^d \times S^{d-1})$ is a continuous injection under some regularity conditions on the kernel (see [9], prop 18, basically we require the continuity of the kernel, bounded, and vanishing at infinity). W' has also a structure of Hilbert space, and $\Omega_0^{d-1}(\mathbb{R}^d \times S^{d-1})' \subset W'$.

The key point of the model is the formula for the scalar product between two normal cycles in the dual space W' :

$$\langle N(C), N(S) \rangle_{W'} = \int_{\mathcal{N}_C} \int_{\mathcal{N}_S} k_p(x,y) k_n(u,v) \langle \tau_{\mathcal{N}_S}(x,u), \tau_{\mathcal{N}_C}(y,v) \rangle \, d\mathcal{H}^{d-1}(x,u) d\mathcal{H}^{d-1}(y,v) \tag{5}$$

The dissimilarity between two shapes S and C is then defined as

$$\|N(S) - N(C)\|_{W'}^2 = \langle N(S), N(S) \rangle_{W'} - 2 \langle N(S), N(C) \rangle_{W'} + \langle N(C), N(C) \rangle_{W'} \tag{6}$$

3 Computational Framework

The aim of this section is to derive the expression of the kernel metric on normal cycles (5) for unions of segments in $\mathbb{R}^3$, which we will use as approximations of real curves.

3.1 Decomposition of the Normal Cycle for Unions of Segments

Let $a, b \in \mathbb{R}^3$ and $C = [a,b]$ be the segment with extremities a and b. We denote $\tilde{C} = C \setminus \{a,b\}$. As noticed in example 2, the normal bundle of C is composed of two parts, a cylindrical part and a spherical part. More precisely, $\mathcal{N}_C = \mathcal{N}_C^{cyl} \cup \mathcal{N}_C^{sph}$ with $\mathcal{N}_C^{cyl} := \tilde{C} \times ((b-a)^\perp \cap S^2)$ and $\mathcal{N}_C^{sph} := (\{a\} \times S_{a-b}^+) \cup (\{b\} \times S_{b-a}^+)$. The normal cycle $N(C)$ thus satisfies $N(C) = N(C)^{cyl} + N(C)^{sph}$ with $N(C)^{cyl} := [\mathcal{N}_C^{cyl}]$ and $N(C)^{sph} := [\mathcal{N}_C^{sph}]$.

In order to get a nice decomposition in the case of unions of segments, it is convenient to define the normal cycle associated to the "open" segment $\tilde{C}$ as: $N(\tilde{C}) := N(C) - N(\{a\}) - N(\{b\})$. This definition is made on purpose to allow to extend the additive property (4) to such open segments. Since the normal bundles of $\{a\}$ and $\{b\}$ are entire spheres, we see that $N(\tilde{C})$ expresses also as a sum of a cylindrical part and a spherical part: $N(\tilde{C}) = N(C)^{cyl} + N(\tilde{C})^{sph}$ with $N(\tilde{C})^{sph} := -[\{a\} \times S^+_{b-a}] - [\{b\} \times S^+_{a-b}]$.

Now let $C = C_1 \cup \cdots \cup C_n$ be a union of n segments in $\mathbb{R}^3$. We can consider without loss of generality that the intersection of two segments $C_i \cap C_j$ is either empty or composed of a single point. If we denote $\{v_1, \ldots, v_N\}$ the set of end points of all segments C_i, $1 \leq i \leq n$, we can rewrite C as the disjoint union of the $\tilde{C}_i$, $1 \leq i \leq n$, and the $\{v_j\}$, $1 \leq j \leq N$. The additive property (4) then becomes straightforward and we get

$$N(C_1 \cup \cdots \cup C_n) = \sum_{i=1}^{n} N(\tilde{C}_i) + \sum_{j=1}^{N} N(\{v_j\}) \tag{7}$$

which we can further decompose into cylindrical and spherical parts as follows :

$$N(C_1 \cup \cdots \cup C_n) = \left(\sum_{i=1}^{n} N(C_i)^{cyl} \right) + \left(\sum_{i=1}^{n} N(\tilde{C}_i)^{sph} + \sum_{j=1}^{N} N(\{v_i\}) \right)$$

3.2 Computation of the Kernel Metric for Unions of Segments

Let $C = C_1 \cup \cdots \cup C_{n_C}$, $S = S_1 \cup \cdots \cup S_{n_S}$ be two unions of segments. The calculation of the expression of (5) in this case is simplified by the following property:

Theorem 2. *The cylindrical part and the spherical part are orthogonal with respect to the kernel metric presented in Sect. 2.*

This means we only need to compute scalar products between spherical parts, and scalar products between cylindrical parts. We do not give full details of this computations here and only sketch the main arguments. We denote $x_1, \ldots, x_{N_C}$ (resp. $y_1, \ldots, y_{N_S}$) the vertices of C (resp. of S) and $f_i = x_{f_i^2} - x_{f_i^1}, 1 \leq i \leq n_C$ (resp. $g_j = y_{g_j^2} - y_{g_j^1}, 1 \leq j \leq n_S$) the edges of C (resp. S). For an edge f_i, $x_{f_i^1}$ and $x_{f_i^2}$ are its two vertices. Moreover, we define $c_i = \frac{1}{2}(x_{f_i^1} + x_{f_i^2})$, $d_j = \frac{1}{2}(y_{g_j^1} + y_{g_j^2})$ and, $\theta_{ij} = \arccos\left(\left\langle \frac{f_i}{|f_i|}, \frac{g_j}{|g_j|} \right\rangle\right)$.

For the cylindrical part, in (5), the point kernel integrated over the segment is approximated by its value at the center, with a coefficient taking into account the length of the edge. For the normal part, we use an expansion in spherical harmonics of k_n (which is valid since k_n is rotation invariant), and pre integrate the kernel on the normal parts. The result is a quantity depending on the angle

between the edges. For the spherical part, the evaluation of the point kernel is immediate and we use the same argument for the normal kernel. Finally we get

$$\left\langle N(C)^{cyl}, N(S)^{cyl} \right\rangle_{W'} \simeq \frac{1}{\sigma_W^2} \sum_{i=1}^{n_C} \sum_{j=1}^{n_S} k_p(c_i, d_j) \left\langle f_i, g_j \right\rangle \sum_{m \geq 0} a_m \cos(m\theta_{ij}) \quad (8)$$

and

$$\begin{aligned} \left\langle N(C)^{sph}, N(S)^{sph} \right\rangle_{W'} = & \sum_{k=1}^{N_C} \sum_{l=1}^{N_S} k_p(x_k, y_l) \left(1 - \frac{n_{x_k} + n_{y_l}}{2}\right) \beta \\ & + \sum_{i=1}^{n_C} \sum_{j=1}^{n_S} \sum_{a,b=1}^{2} \left(b_0 + (-1)^{a+b} \sum_{m \geq 0} b_m \cos(m\theta_{ij})\right) k_p(x_{f_i^a}, y_{g_j^b}) \end{aligned} \quad (9)$$

where n_{x_k} (resp. n_{y_l}) is the number of edges adjacent to the vertex x_k (resp. y_l).

The constant β and the a_m and b_m coefficients have explicit expansions in spherical harmonics, and are pre-computationable. Even though they are not detailed here, we just precise the fact that they vanish for m even. This ensures that if we invert the orientation of the edges (i.e. if we invert $x_{f_i^1}$ and $x_{f_i^2}$), the scalar product remains unchanged. We retrieve here the fact that normal cycles are unoriented objects. With these two scalar products, we have all we need to implement an algorithm which computes dissimilarity between two discrete curves. This is the first step to have a matching algorithm.

4 Curve Matching via Normal Cycles

Given two curves C, S in $\mathbb{R}^3$, we define the curve matching problem as the minimization of a functional over a given set of deformations G. This functional takes the form

$$\varphi_0 = \arg\min_{\varphi \in G} E(\varphi) + A(N(\varphi(C))) \quad (10)$$

where $A(N(\varphi(C))) = \|N(\varphi(C)) - N(S)\|^2_{W'}$ is the data attachment term evaluating the dissimilarity between the deformed curve $\varphi(C)$ and the target S, and $E(\varphi)$ is an energy which ensures regularity of the mapping. In our experiments we chose to use the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework for defining the space G of deformations and the energy E, but of course other frameworks for non-rigid registration could be used, such as for example Thin Plate Splines ([3]).

In a discrete setting, curves are given as unions of segments. As a first approximation, we replace $\varphi(C)$ by the union of segments corresponding to moving only the vertices of C (we do not consider the deformation of all the curve). Note that it is possible to define a geometric action of a diffeomorphism φ of $\mathbb{R}^d$ on normal cycles, by considering the diffeomorphism induced by φ on $\mathbb{R}^d \times S^{d-1}$ and the standard push forward action on currents. However we do not use this action in this work.

4.1 Gradient of the Data Attachment Term Associated with Normal Cycle

As explained previously, we suppose that the deformation acts only on the vertices. Thus, we can consider that A is a function of the current position of the vertices of C : $A(\varphi(x_k)_{1\leq k\leq N})$. This function can be computed explicitely using (6) and the expressions for the scalar products (8), (9). Then a numerical implementation of the minimization of (10) requires the computation of $\nabla A((x_k)_{1\leq k\leq N}))$, which takes an explicit form by deriving these expressions. We do not detail this calculation here.

4.2 Large Deformation Diffeomorphic Metric Mapping (LDDMM)

The classical LDDMM framework as explained for example in [2] is a variational problem :

$$\begin{cases} \min\limits_{v\in L^2([0,1],V)} J(v) := \gamma\left(\displaystyle\int_0^1 \|v_t\|_V^2\, dt\right) + A(\varphi_1(C)) \\ \dfrac{\partial \varphi_t}{\partial t} = v_t \circ \varphi_t \end{cases} \tag{11}$$

where V, the space of deformation is chosen to be a RKHS with kernel K_V. In this article, K_V will be a Cauchy kernel, with width σ_V : $K_V(x,y) = \frac{1}{1+\frac{|x-y|^2}{\sigma_V^2}}$.

We can show in the same spirit as [9] (Chap. 1, Prop. 9 and Chap. 5, Prop 34) that if $V \hookrightarrow \mathcal{C}_0^3(\mathbb{R}^d, \mathbb{R}^d)$, then there exists a minimum for problem (11).

As explained previously, we only consider at the discrete level the deformation of N points $(x_i)_{1\leq i\leq N}$ in $\mathbb{R}^3$ (the vertices of discrete curves). As shown in [9], if we denote by $q_i(t) = \varphi_t(x_i)$ the points trajectories, the optimal vector field minimizing (11) is necessarily of the form

$$v_t = \sum_{i=1}^{N} K_V(\cdot, q_i(t)) p_i(t)$$

where the $p_i(t) \in \mathbb{R}^3$ are called momentum vectors. Further, it was shown in [11] that the problem can be written in Hamiltonian form and that $q_i(t)$ and $p_i(t)$ must satisfy geodesic equations which write

$$\begin{cases} \dot{q}_i(t) = \displaystyle\sum_{j=1}^{N} K_V(q_i(t), q_j(t)) p_j(t) \\ \dot{p}_i(t) = -(d_{q_i(t)} v_t)^* p_i(t). \end{cases} \tag{12}$$

Initial positions $q_i(0)$ being fixed, we can consider the mapping φ and further functional J as function of the $p_i(0)$ only. This property allows to derive an algorithm which optimizes only on these initial momentum vectors, which reduces significantly the dimensionality of the problem. This algorithm is called geodesic shooting ([11,1]).

5 Application to 3D Curve Matching

5.1 Algorithm

We use the shooting algorithm and optimize the functional depending on p_0 with a quasi Newton Broyden Fletcher Goldfarb Shanno algorithm with limited memory (L-BFGS) [10]. The step in the descent direction is fixed by a Wolfe line search. For the numerical integrations, a Runge-Kutta (4,5) scheme is used (function ode45 in Matlab). For the normal cycles, the point kernel k_p is a Gaussian kernel, with width σ_W, and the normal kernel k_n is a Sobolev kernel, associated with the operator $L = (I - \Delta)^2$. We used a spherical harmonics expansion of this kernel truncated at order 5 for the numerical purpose. All the numerical computations have been done on a laptop using Matlab.

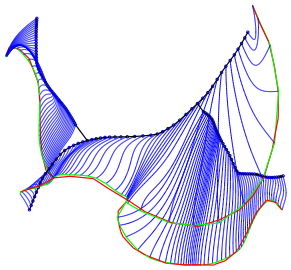
5.2 Numerical Results on Synthetic Data

In this section, we show some of our results on synthetic data and compare them with the varifolds method and currents method. The point kernel chosen for the varifolds is a Gaussian kernel, with the same width σ_W as for normal cycles. The kernel associated with the Grassmanian is chosen linear (see [4]), so that no parameter is involved as for the normal kernel with normal cycles. Lastly, a Gaussian kernel is used as well for currents, again with width σ_W. The trade-off parameter γ is fixed for normal cycles : $\gamma = 0.1$ (since a factor $\frac{1}{\sigma_W^2}$ appears in the cylindrical scalar product, the balance between the two terms is also modified with the kernel width σ_W). For currents and varifolds, we set the parameter γ to be consistent with the metric on normal cycles : we set $\gamma = 0.1 \times \sigma_W^2$.

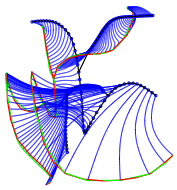
Registrations of branching curves (Fig. 1) The first example of registration is two 3D curves with branching. These curves were chosen because the distance between them is large compared to their typical sizes, the curves have some high local curvature and the size of the corresponding branches implies high local deformations. Besides, we would like to see the behaviour of normal cycles with respect to connecting points.

The two curves are enclosed in a cubic box of size one. Both curves have 150 vertices. In Fig. 1, we show two views of a matching using normal cycles, varifolds and currents. The kernel K_V associated to the deformation space is chosen to be a Cauchy kernel, with width $\sigma_V = 0.2$. Computation time for registration with currents and varifolds were 37 and 120 seconds respectively. Computations with normal cycles are more expensive and took 580 seconds.

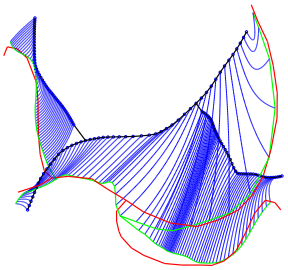
As we can see in Fig. 1, we get a nearly perfect registration with normal cycles. The connecting points of the two curves are well matched, as well as the end points. This is not the case with varifolds and currents, which give less accurate matchings on this example.



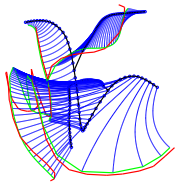
(a) Normal cycles, view 1



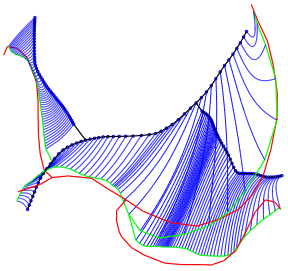
(b) Normal cycles, view 2



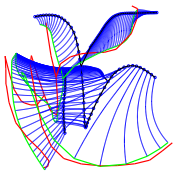
(c) Varifolds, view 1



(d) Varifolds, view 2



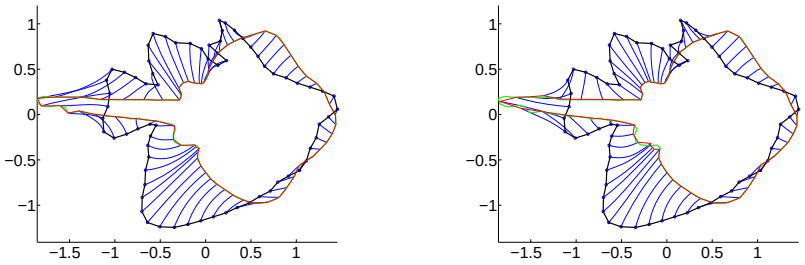
(e) Currents, view 1



(f) Currents, view 2

Fig. 1: Registration of two 3D curves with different data attachment terms. Initial curve is in black, target curve in red, and deformed curve in green. Trajectories of vertices along the flow are displayed in blue. Parameters are $\sigma_V = 0.2$ and $\sigma_W = 0.3$

Registration of fishes contours (Fig. 2) Here a registration between two fishes contours is performed (see [14] for the original data). Even if they are 2D objects, we consider them as 3D objects with no z variation. In this example, fishes have around 100 vertices. A first optimization of the momenta was performed with parameters $\sigma_W = 0.75$ and $\sigma_V = 0.2$. This can be seen as an initialization step to avoid local minima. Then minimization was done with $\sigma_W = 0.2$ and $\sigma_V = 0.2$. Computation time was 310 seconds for normal cycles and 60 seconds for varifolds. The main difficulty here is the trade off to find between the matching of the long tail of the stingray (in green in Fig. 2) and the high local curvature in the upper part of the fish in dark. The results in Fig. 2 show that a perfect matching with normal cycles can be achieved, even with $\sigma_W = 0.2$ which is quite large compared to the local feature in the upper part of the fish. With varifolds, one can see that this local feature still remains in the green matched curve. To avoid this behaviour, one can decrease the size of σ_W, but it would lead to a bad matching of the tail.



(a) Normal cycles (b) Varifolds

Fig. 2: Registration of a dark fish to a red fish. In green the dark deformed fish matching the red one. We used normal cycles and varifolds with the same parameters $\sigma_V = 0.2$ and $\sigma_W = 0.2$. The registration with currents is worse than with varifolds.

6 Perspectives

In this article, we have presented the first application of normal cycles in the context of 3D curve registration. As for currents, a kernel metric is used to provide a closed form for the distance between two curves, and a numerical derivation is done for curves approximated by unions of segments. The first results on synthetic data are promising and suggest that the normal cycles metric improves matchings between connection points and regions with high curvature. Of course, wether such a property is desirable in real applications remains an open debate and highly depends on the type of data in use. Moreover, more exhaustive studies on synthetic and real data are necessary to validate the method.

The next stage will be the registration using normal cycles for surfaces. This case is more intricate, at least numerically since the decomposition of normal bundle as seen in Sect. 3 is more complex. We also would like to investigate the link between varifolds and normal cycles, as we believe that varifolds can be seen in our context as a projection of normal cycles, by ignoring variation in S^{d-1}.

References

1. Arguillere, S., Trélat, E., Trouvé, A., Younes, L.: Shape deformation analysis from the optimal control viewpoint. arXiv:1401.0661 [math] (Jan 2014)
2. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms. International Journal of Computer Vision 61(2), 139–157 (Feb 2005)
3. Bookstein, F.L.: Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence 11(6), 567–585 (1989)
4. Charon, N.: Analysis of geometric and functionnal shapes with extension of currents. Application to registration and atlas estimation. Ph.D. thesis, École Normale Supérieure de Cachan (2013)
5. Charon, N., Trouvé, A.: The varifold representation of nonoriented shapes for diffeomorphic registration. SIAM J. Imaging Sciences 6(4), 2547–2580 (2013)
6. Cohen-Steiner, D., Morvan, J.M.: Restricted Delaunay Triangulations And Normal Cycle. SoCG'03 (2003)
7. Durrleman, S.: Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution. Ph.D. thesis, Université Nice - Sophia Antipolis (2010)
8. Federer, H.: Curvature measures. Trans. Amer. Maths. Soc. 93 (1959)
9. Glaunès, J.: Transport par difféomorphismes de points, de mesures et de courants pour la comparaison de formes et l'anatomie numérique. Ph.D. thesis, Université Paris 13 (2005)
10. Liu, D.C., Nocedal, J.: On the limited memory BFGS method for large scale optimization. Mathematical Programming 45(1-3), 503–528 (Aug 1989)
11. Miller, M.I., Trouvé, A., Younes, L.: Geodesic Shooting for Computational Anatomy. Journal of Mathematical Imaging and Vision 24(2), 209–228 (Jan 2006)
12. Morvan, J.M.: generalized curvatures. Springer (2008)
13. Vaillant, M., Glaunès, J.: Surface Matching via Currents. In: Christensen, G.E., Sonka, M. (eds.) Information Processing in Medical Imaging, pp. 381–392. No. 3565 in Lecture Notes in Computer Science, Springer Berlin Heidelberg (2005)
14. US Dept of the Interior Fish and Wildlife Service: Fishes of the gulf of maine (1953), `http://www.nefsc.noaa.gov/lineart/`
15. Zähle, M.: Integral and current representation of Federer's curvature measure. Arch. Maths. 23, 557–567 (1986)
16. Zähle, M.: Curvatures and currents for unions of set with positive reach. Geometriae Dedicata 23, 155–171 (1987)

Efficient Metamorphosis Computation for Classifying Embryonic Cardiac Action Potentials

Giann Gorospe[1], Renjun Zhu[1], Jia-Qiang He[2], Leslie Tung[1], Laurent Younes[1], and René Vidal[1]

[1] Johns Hopkins University, Baltimore, USA
[2] Virginia Tech University, Blacksburg, USA

Abstract. As purification methods for obtaining cardiomyocytes from stem cells continue to improve, the need for automated methods for high-throughput classification of these cells is becoming extremely important. Since the shape of the action potential of an adult cell is discriminative of its phenotype, a promising classification approach is to use the metamorphosis distance between the action potentials of embryonic and adult cells. However, current gradient descent methods for computing the metamorphosis distance are extremely slow, hence unsuitable for large scale classification. In this paper, we show that the metamorphosis path can be computed in closed form given the velocity field, which leads to an efficient alternating minimization approach for computing the metamorphosis distance. We test this algorithm on heart cell datasets varying from 100 to 7,000 cells.

Keywords: Cardiac electrophysiology, shape analysis, machine learning

1 Introduction

Ever since Kehat et al. [1] pioneered cardiomyocyte differentiation back in 2001, there has been a lot of hope in the potential of stem cell based cardiology. This dream was bolstered further by the work of [2], which showcased the medical potential for human embryonic stem cell derived cardiomyocytes in infarcted rat hearts. The goal of utilizing stem-cell derived cardiomyocytes for cardiac regeneration, as well as disease models [3] could prove instrumental to the future of cardiology. However, as expressed in multiple texts [4, 5], there is still a need for methods to identify cardiomyocyte phenotype (nodal, atrial, or ventricular) in order to prevent potential arrhythmias, improve sustained cardiac regeneration, or to provide consistent models for therapeutic study.

In 2003, [6] showed that stem cell derived cardiomyocytes could be divided into several phenotypes based on their electrophysiological signature, called an action potential (AP). They labelled the three phenotypes embryonic nodal-like, embryonic atrial-like, and embryonic ventricular-like. Examples of the embryonic atrial-like and embryonic ventricular-like APs are provided in Figure 1. They determined this classification manually, and verified it by obtaining measurements

of features of the action potentials, such as action potential amplitude and duration, and showing statistically significant differences between the phenotypes. Other works [7, 8] have used similar processes for classification. The problem with classification methods of this type is that they are subjective, which makes them difficult to transfer across datasets, where acquisition protocols may be different.

We believe that automated, objective methods for embryonic heart cell classification are integral to the study of embryonic cardiomyocytes. However, the development of classification methods faces several fundamental challenges. First, the selection of shape features for classification is only well understood in the case of adult cells, and continues to be an art. Second, the phenotypes of immature cells need not coincide with the phenotype of adult cells, and even if they do, the shape of the action potential may change throughout the maturation process. Understanding the morphological changes of the AP during maturation may prove insightful to the underlying cellular processes.

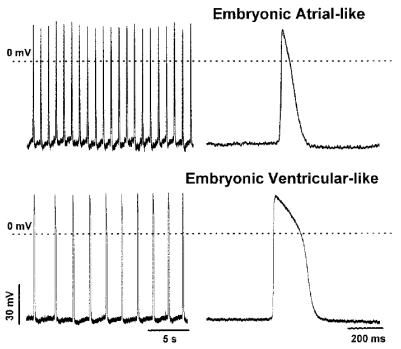


Fig. 1. Sample embryonic action potentials as described by [6]

As these embryonic cardiomyocytes will eventually become one of the mature phenotypes of interest to clinicians, determining the fate and maturation process of an embryonic cardiomyocyte from the shape of the AP will help reduce the potential risks in future studies. We would like a model that not only provides a way to determine the mature fate of the embryonic cardiomyocyte, but also provides insight into the modification of the AP as it matures from infancy to adulthood. While there is limited evidence of the maturation process, the general hypothesis is that the action potential evolves smoothly from infancy to maturity. Thus, smooth deformation models, like the metamorphosis distance, introduced in [9–11] and applied in [12] to cardiomyocyte classification, is a promising approach to address these challenges. However, while [12] showed very promising results on using the metamorphosis distance for action potential classification, the method used for computing the distance is computationally inefficient. As cardiomyocyte differentiation methods mature, and larger populations of cells become available, there is a need for faster approaches for computing the metamorphosis distance.

The main contribution of this paper is to propose an alternative method for computing the discrete metamorphosis model. Given two action potentials, the metamorphosis is obtained by finding the optimal interpolant and deformation paths that interpolate the two action potentials, warping one into the other. Rather than solving the optimization problem by gradient descent, we show that the optimal interpolant path for a given velocity field can be computed in closed form. Our experiments show that this leads to a slightly more efficient

alternating minimization approach for computing the metamorphosis distance, which requires fewer iterations and perfoms better compared to gradient descent methods. We demonstrate the performance of the proposed algorithm on a small microelectrode recording dataset of about 100 action potentials and, for the first time, on a large optical mapping dataset of about 7,000 action potentials.

2 Review of Metamorphosis

The metamorphosis distance, proposed in [10] and [11], is an interpolation scheme used for defining a Riemannian distance between two shapes. In this section, we will review this scheme in the context of cardiac action potentials, which are one dimensional shapes. However, this scheme may be applicable to other shapes.

Let $I_0(\tau)$ and $I_1(\tau)$ be two action potentials, called, respectively, template and target. We assume that these signals are periodic, continuously differentiable and square integrable, i.e., $I_0, I_1 \in L^2(\mathcal{S}^1)$, where $\mathcal{S}^1$ is the unit circle. A metamorphosis is a family of action potentials $\{I(\cdot,t) \in L^2(\mathcal{S}^1)), t \in [0,1]\}$ that interpolates between the template $(I(\tau,0) = I_0(\tau))$ and the target $(I(\tau,1) = I_1(\tau))$. Each element of this family can be further decomposed in terms of a diffeomorphism $\phi(\cdot,t) \in \mathrm{Diff}(\mathcal{S}^1)$ acting on an evolving template $i(\tau,t)$ as

$$I(\tau,t) = \phi(\tau,t) \cdot i(\tau,t) = i(\phi^{-1}(\tau,t),t). \tag{1}$$

To define a distance between two action potentials using this model, we need to define an energy that depends on the infinitesimal change in the deformation $(\frac{\partial \phi}{\partial t})$ and the infinitesimal change in the template $(\frac{\partial i}{\partial t})$. Because ϕ is a diffeomorphism, we can define the infinitesimal change in the deformation $\frac{\partial \phi}{\partial t}$ by a smooth flow field v $(\frac{\partial \phi}{\partial t} = v(\phi(\tau,t),t))$ and penalize its smoothness with a Sobolev norm (a norm on a function and its derivatives). For example, let $L(\cdot)$ is a linear differential operator acting on v (for example, $Lv = v - \alpha \frac{\partial^2}{\partial \tau^2} v$). We can use a Sobolev norm that can be expressed in terms of the Euclidean norm as:

$$\left\| \frac{\partial \phi}{\partial t} \right\|_V^2 = \|v\|_V^2 = \langle v, Lv \rangle. \tag{2}$$

On the other hand, we can compute the infinitesimal change in the template by taking the derivative of $I(\tau,t)$ with respect to t, which gives (see [11] for the details):

$$\frac{\partial i}{\partial t} = \frac{\partial I}{\partial t} + \frac{\partial I}{\partial \tau} v, \tag{3}$$

and then penalize this change using its Euclidean distance. Combining the two penalties with a balancing parameter σ allows us to define an energy on the family, which can be minimized over all families to define a distance between template and target as:

$$d^2_{\mathcal{M}}(I_0,I_1) = \inf_{\substack{v,I \\ I(\tau,0)=I_0(\tau) \\ I(\tau,1)=I_1(\tau)}} \int_0^1 \left(\frac{1}{2}\|v(\tau,t)\|_V^2 + \frac{1}{2\sigma^2} \left\| \frac{\partial I}{\partial t}(\tau,t) + \frac{\partial I}{\partial \tau}(\tau,t)\, v(\tau,t) \right\|_{L^2}^2 \right) dt. \tag{4}$$

One approach to minimizing this distance is to discretize the energy, and then develop methods to minimize the resulting discretized energy. For example, [12] and [13], take advantage of the following approximation to discretize τ:

$$\int_0^1 \left\| \frac{\partial I}{\partial t}(\tau,t) + \frac{\partial I}{\partial \tau}(\tau,t)v(\tau,t) \right\|_{L_2}^2 dt \approx \sum_{k=0}^{S-1} \left\| \frac{I(t+\delta t v(\tau,t_k),t_{k+1}) - I(\tau,t_k)}{\delta t} \right\|_{L_2}^2. \tag{5}$$

After combining this with a discretization for τ, the energy becomes:

$$E(v(\tau_i,t_k),I(\tau_i,t_k)) = \sum_{k=0}^{S-1} \frac{1}{2}\|v(\tau_i,t_k)\|_V^2 + \frac{1}{2\sigma^2}\|I(\tau_i+v(\tau_i,t_k),t_{k+1})) - I(\tau_i,t_k)\|_{l_2}^2. \tag{6}$$

Both [12] and [13] minimize this discrete energy via alternating gradient descent. In particular, following [12], let L_d be the discretized version of the linear operator in (2), and let $K = L_d^{-1}$ be the corresponding smoothing kernel for the derivative operator L_d. Now, using $w = L_d^{1/2}v$, the gradient of E with respect to this new variable w at each interpolation step can be calculated via the chain rule as:

$$\frac{\partial E}{\partial w(\tau_i,t_k)} = w(\tau_i,t_k) + \frac{1}{\sigma^2}K^{1/2}(I(\bar{\tau}_i,t_{k+1}) - I(\tau_i,t_k))\frac{\partial I(\bar{\tau}_i,t_{k+1})}{\partial \tau_i}, \tag{7}$$

where $\bar{\tau}_i = \tau_i + K^{1/2}w(\tau_i,t_k)$.

Now, given w, and as a result, v, the gradient update of the metamorphosis interpolants, $I(\tau,t_k)$ can be determined by making an approximation. Since we are discretizing in the "spatial" domain τ, we sample each $I(\tau,t_k)$ and specific points τ_i. Thus, when we look at $I(\tau+v(\tau,t_k),t_{k+1})$, it is likely that $\tau_i+v(\tau_i,t_k)$ does not coincide with the original samples τ_i. Therefore we have to approximate the value of $I(\tau_i+v(\tau_i,t_k),t_{k+1})$ in terms of the samples original discretization $I(\tau_i,t_{k+1})$ so that they can be compared to the samples of $I(\tau_i,t_k)$. The action of approximating the sampling of $I(\tau+v(\tau,t_k),t_{k+1})$ using the samples of $I(\tau,t_{k+1})$ is denoted by the operator N_{v_k} and is realized by the linear interpolation matrix:

$$N_{v_k}I(\tau,t_{k+1}) \approx I(\tau+v(\tau,t_k),t_{k+1}). \tag{8}$$

This leads to:

$$E(I(\tau,t)) \approx \sum_{k=0}^{S-1} \frac{1}{2\sigma^2}\|N_{v_k}I(\tau,t_{k+1}) - I(\tau,t_k)\|_{L^2}^2. \tag{9}$$

The gradient with respect to $I(\tau,t_k), k = 1,\ldots,S-1$, is given by:

$$\frac{\partial E}{\partial I(\tau,t_k)} = \frac{1}{\sigma^2}(N_{v_{k-1}}^T(N_{v_{k-1}}I(\tau,t_k) - I(\tau,t_{k-1})) - (N_{v_k}I(\tau,t_{k+1}) - I(\tau,t_k))). \tag{10}$$

While alternating between updates of v and I does find a local minimum to the optimization problem, the gradient descent steps are handcuffed by the choice of the step size. This can be addressed by either finding an appropriate adaptive step size, or by a large number of iterations at a fixed, but stably small step size. However, both schemes require additional computational effort.

3 A Closed Form Update for Continuous Metamorphosis

If the goal is a model of maturation and classification of large populations of cardiomyocytes, it will be difficult to scale gradient descent methods for solving tens of thousands of optimization problems. In this paper, we propose an alternative approach in which the solution for some variables can be computed in closed form given the other variables. To motivate the proposed approach, in this section we present a formulation derived in [14] for solving the alternating minimization updates of the continuous energy in (4). In the next section, we will show how this approach can be extended to the discrete energy.

We begin by applying the chain rule to compute the variation of the energy in (4) with respect to v. This leads to an update for $v(\tau,t)$ with fixed $I(\tau,t)$ that is based on solving the following non-homogeneous differential equation:

$$\Big(2L + \frac{1}{\sigma^2}\Big(\frac{\partial I}{\partial \tau}(\tau,t)\Big)^2\Big)v(\tau,t) = \frac{1}{\sigma^2}\Big(-\frac{\partial I}{\partial t}(\tau,t)\frac{\partial I}{\partial \tau}(\tau,t)\Big), \tag{11}$$

where $Lv = v - \alpha\frac{\partial^2}{\partial \tau^2}v$. Now, alternatively, minimizing this distance with respect to $I(\tau,t)$, with $v(\tau,t)$ fixed, leads to minimizing the following energy:

$$E(I(\tau,t)) = \int_0^1 \frac{1}{2\sigma^2}\Big\|\frac{\partial I}{\partial t}(\tau,t) + \frac{\partial I}{\partial \tau}(\tau,t)\ v(\tau,t)\Big\|_{L^2}^2 dt. \tag{12}$$

Taking the variation with respect to $I(\tau,t)$ leads to a complicated differential equation to be solved. However, if we instead let $J(\tau,t) = I(\phi(\tau,t),t)$ and $u = \phi^{-1}(\tau,t)$, then it is not difficult to show that differentiating $J(\tau,t)$ with respect to t leads to the following result:

$$\frac{\partial J(\phi^{-1}(\tau,t),t)}{\partial t} = \frac{\partial I(\tau,t)}{\partial t} + \frac{\partial I(\tau,t)}{\partial \tau}v(\tau,t). \tag{13}$$

After making the change of variables $u = \phi^{-1}(\tau,t)$ and substituting the above relationship, the energy becomes:

$$E(J(u,t)) = \int_0^1 \|\frac{\partial J(u,t)}{\partial t}\|_{L^2}^2 \frac{\partial \phi(u,t)}{\partial u}dt. \tag{14}$$

The Gateaux variation of E with respect to J, and setting it to 0 leads to:

$$J(u,t) = J(u,0)\frac{\int_t^1 \frac{1}{\frac{\partial \phi(u,t')}{\partial u}}dt'}{\int_0^1 \frac{1}{\frac{\partial \phi(u,t')}{\partial u}}dt'} + J(u,1)\frac{\int_0^t \frac{1}{\frac{\partial \phi(u,t')}{\partial u}}dt'}{\int_0^1 \frac{1}{\frac{\partial \phi(u,t')}{\partial u}}dt'}. \tag{15}$$

After computing $J(u,t)$, one may obtain $I(\tau,t)$ by back-substitution. Thus, not only it is possible to compute the optimal $I(\tau,t)$ given $v(\tau,t)$ in closed form, the optimal solution is essentially a weighted combination of the source and the target. In practice, however, the update for v given I and vice versa cannot be implemented without first discretizing the equations, as discussed next.

4 A Closed Form Update for Discrete Metamorphosis

One approach to implementing the updates for v and I described in the previous section is to simply discretize the updates. However, a naive discretization of the continuous updates need not coincide with the updates for a discretization of the original objective, such as that in (6). In this section, we derive closed form updates for the minimization of the discrete energy in (6).

4.1 Formulation

Let $N_k = N_{v_k}$ and $I_k = I(\tau, t_k)$. If $Z_k = N_k I_{k+1} - I_k$, then at the optimum, (10) simplifies to:

$$Z_k = N_{k-1}^T Z_{k-1} \tag{16}$$

If we let $R_{i,j} = N_i N_{i+1} \dots N_{j-1}$, it follows that $Z_k = R_{0,k}^T Z_0$. Using this, and the original definition of Z_k, we can write an equation for I_k:

$$I_k = N_k I_{k+1} - Z_k = N_k I_{k+1} - R_{0,k}^T Z_0. \tag{17}$$

Iterating backwards from $k = S-1$, we can write these equations using I_S as:

$$I_k = R_{k,S} I_S - (\sum_{i=k}^{S-1} R_{k,i} R_{0,i}^T) Z_0 = R_{k,S} I_S - A_{k,S} Z_0, \tag{18}$$

where $A_{l,m} = \sum_{i=l}^{m-1} R_{l,i} R_{0,i}^T$. To determine Z_0, we look at I_0:

$$I_0 = R_{0,S} I_S - A_{0,S} Z_0 \implies Z_0 = A_{0,S}^{-1}(R_{0,S} I_S - I_0). \tag{19}$$

So, after replacing Z_0, we find an update for I_k that depends only on I_0 and I_S:

$$I_k = A_{k,S} A_{0,S}^{-1} I_0 + (R_{k,S} - A_{k,S} A_{0,S}^{-1} R_{0,S}) I_S. \tag{20}$$

4.2 Computation

From (20), we get a closed form update for the interpolants I_k in terms of the template I_0 and target I_S. The next major question to solve is how to efficiently compute this update. Looking at the equation, we need $R_{k,S}$ and $A_{k,S}$ for all k. But given the current definition, $A_{k,S}$ requires knowledge of all $R_{i,j}$. This is

a large computational storage overhead, but it can be avoided by noticing that $A_{k,S}$ can also be determined using backwards iteration:

$$A_{k,S} = \sum_{i=k}^{S-1} R_{k,i} R_{0,i}^T = R_{0,k}^T + N_k A_{k+1,S}. \tag{21}$$

Thus we have the following system of forward and backwards updates:

$$I_k = B_{k,S} I_0 + (R_{k,S} - B_{k,S} R_{0,S}) I_S \tag{22}$$

$$R_{k,S} = N_k R_{k+1,S} \tag{23}$$

$$C_{0,k}^T = N_k^T C_{0,k-1}^T \tag{24}$$

$$B_{k,S} = C_{0,k}^T + N_k B_{k+1,S}, \tag{25}$$

where $B_{k,S} = A_{k,S} A_{0,S}^{-1}, C_{0,k}^T = R_{0,k}^T A_{0,S}^{-1}$. The initial conditions for the updates are: $R_{S,S} = Id$, $C_{0,0}^T = A_{0,S}^{-1}$, and $B_{S,S} = 0$. Here, $A_{0,S}^{-1}$ can be computed by using the original definition and the storage of $R_{0,k}$ which then can be used to generate $C_{0,k}$. Since the biggest computational task in this update is computing $A_{0,S}^{-1}$, the update of $A_{0,S}^{-1}$ can be done once every n iterations if a faster approximation is required. Our update for velocity follows that of [12]. The overall algorithm is presented in Algorithm 1. We acknowledge that this step is still gradient descent, and may be a limiting step in performing the overall algorithm. Finding a closed form solution for this optimization problem is difficult given the nonlinearity in v of the template evolution summand of the energy. Alternative approaches to performing this update more efficiently are a future research direction.

4.3 Convergence to continuous formulation

We have derived a closed form update for the metamorphosis interpolants, and provided a way to efficiently compute this optimum using forward-backward schemes. We comment now on how this proposed update relates to the continuous formulation. We omit many of the details here, but they can be found in the Supplementary Material. The convergence result can be summarized by the following theorem:

Theorem 1. *Let $I_k, k = 0, \ldots, S$, be the metamorphosis interpolants derived from* (20)*, and let $I(\tau, t)$ be the family of interpolants derived from $J(u, t)$ given by* (13)*. Then, as $S \to \infty$, $I_k \to I(\tau, t)$.*

Sketch of the Proof: Given that J represents the interpolants in the template domain, it follows that the discrete version J_k of I_k is $J_k = R_{0,k} I_k$. Proceeding from $I_k = A_{k,S} A_{0,S}^{-1} I_0 + (R_{k,S} - A_{k,S} A_{0,S}^{-1} R_{0,S}) I_S$, it is not difficult to show that $J_k = (Id - A_{0,k} A_{0,S}^{-1}) J_0 + A_{0,k} A_{0,S}^{-1} J_S$. From here, after some analysis one can show $R_{0,k}(J(\cdot, t)) \to J(\phi(\cdot, \frac{k}{S}), t)$, which leads to the result.

Algorithm 1 Discrete Metamorphosis via Direct Image Computation

Given a Template Signal $I_0(\tau)$, a Target Signal $I_1(\tau)$, a balance parameter σ, the number of evolution time steps S, and a Sobolev Operator L_d, update frequency n.

1. Initialization.
   (a) Set $m = -1$, $d_{-1} = \infty$. Calculate $K = L_d^{-1}$.
   (b) Set $w(\tau_i, t_k) \equiv 0$, $v(\tau_i, t_k) = K^{1/2} w(\tau_i, t_k) \equiv 0$, for all t_k and τ_i.
   (c) For $k = 0, \ldots, S$: Set $I(\tau_i, t_k) = \frac{S-k}{S} I_0(\tau_i) + \frac{k}{S} I_1(\tau_i)$
   (d) Calculate $d_0^2 = \sum_{k=0}^{S-1} \frac{1}{2} \|w(\tau_i, t_k)\|_{l_2}^2 + \frac{1}{2\sigma^2} \|N_{v_k} I(\tau_i, t_{k+1})) - I(\tau_i, t_k)\|_{l_2}^2$
2. Until $d_{m-1} - d_m$ converges
   (a) Set $d_m \to d_{m-1}, m + 1 \to m$.
   (b) For $k = 0, \ldots, S-1$, Update $w(\tau_i, t_k)$ using (7).
   Calculate $v(\tau_i, t_k) = real(K^{1/2} w(\tau_i, t_k))$, Update N_{v_k}.
   Compute $R_{0,k}$ and $R_{k,S}$ for $k = 0, \ldots, S$.
   (c) if $\mathrm{mod}(m, n) = 0$,
   Calculate $A_{0,S} = \sum_{i=0}^{S-1} R_{0,i} R_{0,i}^T$, Calculate $A_{0,S}^{-1}$.
   (d) Initialize $C_{0,0}^T = A_{0,S}^{-1}$, $B_{S,S} = 0$. Calculate:
   $C_{0,k}^T = N_k^T C_{0,k-1}^T$
   $B_{k,S} = C_{0,k}^T + N_k B_{k+1,S}$
   $I_k = B_{k,S} I_0 + (R_{k,S} - B_{k,S} R_{0,S}) I_S$
   (e) Calculate $d_m^2 = \sum_{k=0}^{S-1} \frac{1}{2} \|w(\tau_i, t_k)\|_{l_2}^2 + \frac{1}{2\sigma^2} \|N_{v_k} I(\tau_i, t_{k+1})) - I(\tau_i, t_k)\|_{l_2}^2$

5 Experiments

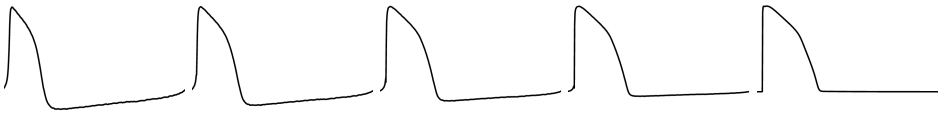
5.1 Patch Clamp Data

In this section, we evaluate the efficiency of our algorithm by comparing it to the gradient descent based method proposed in [12] on the dataset generated by [6]. The dataset contains 16 embryonic atrial-like and 36 embryonic ventricular-like cardiomyocytes, manually labeled according to [6] based on AP features. The data was pre-processed using the protocol described in [12]. Namely, we fixed the cycle length to 1 second using the algorithm presented by [15]. For classification purposes, we generated 10 mature atrial and 10 mature ventricular protoype action potentials using the atrial model of [16] and the ventricular model of [17], respectively. All signals were then normalized so that the resting membrane potential has voltage 0, and the amplitude has voltage 1.

We computed the metamorphosis distance from each of the embryonic cardiomyocytes to each one of the mature prototypes. We used the linear operator $L_d(\cdot) = id(\cdot) - \alpha\Delta(\cdot)$, with $\alpha = 8$, and set the parameterσ to 0.3 and the number of interpolants between template and target to 3 ($S = 4$). We iterated our algorithm and that of [12] until they reached convergence or 300 iterations.

Figure 2 compares the method of [12] and two variants of our method (with $A_{0,S}^{-1}$ updated each iteration or every 10 iterations) in terms of the final interpolations and the distances they produce. We see that there is very little difference between the three interpolations, and that the three distances are approximately equal. The main difference is that our method with $A_{0,S}^{-1}$ updated every 10 iterations requires about half the number of iterations than the other methods.

(a) Gradient Descent Metamorphosis [12]: $d^2_{\mathcal{M}} = 206.8240$, Iterations: 74



(b) Proposed method, $A^{-1}_{0,S}$ updated every iteration: $d^2_{\mathcal{M}} = 208.2602$, Iterations: 72

(c) Proposed method, $A^{-1}_{0,S}$ updated every 10 iterations: $d^2_{\mathcal{M}} = 207.6308$, Iterations: 38

Fig. 2. Comparison of the metamorphosis method in [12] and two variants of the proposed method in terms of the interpolants and the distances they compute.

Table 1 compares the the Euclidean distance, the metamorphosis distance computed with the method in [12], and the metamorphosis distance computed using the two variants of our method in terms of their classification performance and computation time on the entire dataset. For this purpose, we use the 20 mature prototypes as our training set, and the entire dataset as the test set. Classification is done with the 1 nearest neighbor (NN) and 3 NN classifiers, meaning we classify an AP based on the class of the closest 1 or 3 mature prototypes. While the Euclidean distance is the fastest to compute, the classification performance is better using the metamorphosis distance. Moreover, we see that our method provides improved classification rate relative to the current state-of-the-art method at reduced computation time. In fact, when $n = 10$, the computation time is almost completely in the gradient descent update for v, suggesting that the computational limit in the interpolant update has been reached. While improving the speed of the velocity updates is one of our future research goals, the findings on this dataset suggest that our algorithm provides an improvement over the current standard.

5.2 Optical Data

We also tested our algorithm on a much larger dataset consisting of 9 cell clusters with APs recorded using the optical mapping technique of [18]. The number of APs in each cell cluster ranges from 400 to 1000, and the total number of APs in the dataset equals 6940. Mature prototypes were generated using the same computational models as in the previous experiment. The signals were paced at

Table 1. Comparison of the metamorphosis method in [12] and two variants of the proposed method in terms of classification performance and computation time on a patch clamp dataset.

	Euclidean Distance	Gorospe et al. [12]	Our Method (n = 1)	Our Method (n = 10)
1 NN Atrial Scoring	**16**/16	13/16	14/16	14/16
1 NN Ventricular Scoring	29/36	**36**/36	**36**/36	**36**/36
3 NN Atrial Scoring	**16**/16	13/16	14/16	14/16
3 NN Ventricular Scoring	29/36	**36**/36	**36**/36	**36**/36
Computation Time (in seconds)	$< \mathbf{1}$	17.0181	15.8265	12.0815

a rate of 1.5 Hz (cycle length of $\frac{2}{3}$ seconds), and also normalized to have resting potential voltage 0, and maximum voltage amplitude of 1.

We computed the metamorphosis distance using our new formulation with $n = 1$ from each AP in the dataset to each one of the mature prototypes using the same parameters as in the patch clamp experiment. The algorithms were run in 2 8-core computer nodes with 8 hyperthreaded 2.3 GHz CPUs per node. The total time to complete the analysis on the entire dataset was 13 hours, with individual cell clusters taking between 50 and 80 minutes.

Figure 3 compares the classification results obtained by a 1-NN classifier with the Euclidean distance versus the metamorphosis distance for each one of the 9 cell clusters. We omit the 3 NN results as the results are identical. The blue color indicates areas where the APs were classified as atrial, and red indicates areas that were classified as ventricular. While some of the cell clusters present with only 1 phenotype, the majority of the cell clusters present with both phenotypes in varying concentrations, affirming recent work [19].

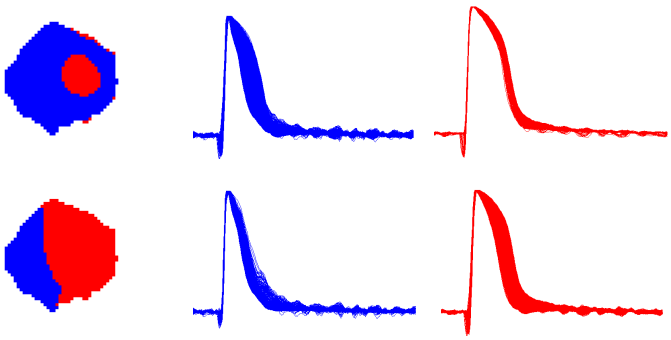
Figure 3(c) compares the action potentials of the cell clusters obtained by a 1-NN classifier with the Euclidean and Metamorphosis distances for a pair of heterogeneous clusters. For the metamorphosis classification, the two classes show distinct shapes, and they are similar to those described by [6] for embryonic atrial-like and embryonic ventricular-like. In comparison, the Euclidean classification fails to capture the distinction between the phenotypes. This affirms that the metamorphosis distance is a suitable automated counterpart to manual classification by biologists. More importantly, it suggests that the metamorphosis model could be used to reliably assess the phenotype statistics of populations of APs. As a consequence, the metamorphosis model may prove insightful to a growing collection of methods that have been derived to isolate a particular phenotype of embryonic cardiomyocytes [20, 21].

6 Conclusion

We have presented an algorithm for computing the metamorphosis that performs comparably with the current state of the art, but at approximately two-thirds the run time. We presented a closed form update for the interpolants that can be computed via a series of forward and backwards updates, as well as demon-

(a) Euclidean 1-NN

(b) Metamorphosis 1-NN



Classified Cell Cluster Atrial Signals Ventricular Signals

(c) Cell cluster classifications and their corresponding action potentials.

Fig. 3. Comparison of the Euclidean distance and the metamorphosis distance for 1-NN classification of the optical dataset.

strated its convergence to the continuous evolution metamorphosis updates. We demonstrated its effectiveness on a studied microelectrode recording dataset, as well as a much larger scale optical mapping dataset. We believe that the new method could lead to advances in stem cell cardiology, as well as lead to potential new frontiers in computational shape analysis.

References

1. Kehat, I., Kenyagin-Karsenti, D., Snir, M., Segev, H., Amit, M., Gepstein, A., Livne, E., Binah, O., Itskovitz-Eldor, J., Gepstein, L.: Human embryonic stem cells can differentiate into myocytes with structural and functional properties of cardiomyocytes. Journal of Clinical Investigation **108**(3) (2001) 407–414
2. Laflamme, M.A., Chen, K.Y., Naumova, A.V., Muskheli, V., Fugate, J.A., Dupras, S.K., Reinecke, H., Xu, C., Hassanipour, M., Police, S., O'Sullivan, C., Collins, L., Chen, Y., Minami, E., Gill, E.A., Ueno, S., Yuan, C., Gold, J., Murry, C.E.: Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. Nature Biotech. **25**(9) (2007) 1015–24
3. Hoekstra, M., Mummery, C.L., Wilde, A.A., Bezzina, C.R., Verkerk, A.O.: Induced pluripotent stem cell derived cardiomyocytes as models for cardiac arrhythmias. Frontiers in Physiology **3**(346) (2012)

4. Maher, Kevin O. andi Xu, C.: Marching Towards Regenerative Cardiac Therapy with Human Pluripotent Stem Cells. Discovery Medicine **15** (2013) 349–356
5. Xu, X.Q., Sun, W.: Perspective from the heart: The potential of human pluripotent stem cell-derived cardiomyocytes. Journal of Cellular Biochemistry **114**(1) (2013) 39–46
6. He, J.Q., Ma, Y., Lee, Y., Thomson, J.A., Kamp, T.J.: Human embryonic stem cells develop into multiple types of cardiac myocytes: action potential characterization. Circulation Research **93**(1) (2003) 32–9
7. Moore, J.C., Fu, J., Chan, Y.C., Lin, D., Tran, H., Tse, H.F., Li, R.A.: Distinct cardiogenic preferences of two human embryonic stem cell (hESC) lines are imprinted in their proteomes in the pluripotent state. Biochemical and biophysical research communications **372**(4) (2008) 553–8
8. Zhang, J., Wilson, G.F., Soerens, A.G., Koonce, C.H., Yu, J., Palecek, S.P., Thomson, J.A., Kamp, T.J.: Functional cardiomyocytes derived from human induced pluripotent stem cells. Circulation Research **104**(4) (February 2009) e30–41
9. Miller, M., Younes, L.: Group actions, homeomorphisms, and matching: A general framework. International Journal of Computer Vision **41**(1-2) (2001) 61–84
10. Trouvé, A., Younes, L.: Metamorphoses Through Lie Group Action. Foundations of Computational Mathematics **5**(2) (2005) 173–198
11. Younes, L.: Shapes and Diffeomorphisms. Volume 171 of Applied Mathematical Sciences. Springer (2010)
12. Gorospe, G., Younes, L., Tung, L., Vidal, R.: A metamorphosis distance for embryonic cardiac action potential interpolation and classification. In: Medical Image Computing and Computer Assisted Intervention. (2013) 469–476
13. Garcin, L., Younes, L.: Geodesic image matching: A wavelet based energy minimization scheme. Energy Minimization Methods in Computer Vision and Pattern Recognition (2005) 349–364
14. Trouvé, A., Younes, L.: Local geometry of deformable templates. SIAM Journal on Mathematical Analysis **37**(1) (2005) 17–59
15. Iravanian, S., Tung, L.: A novel algorithm for cardiac biosignal filtering based on filtered residue method. IEEE Transactions on Biomedical Engineering **49**(11) (2002) 1310–7
16. Nygren, A., Fiset, C., Firek, L., Clark, J., Lindblad, D., Clark, R., Giles, W.: Mathematical model of an adult human atrial cell: The role of K+ currents in repolarization. Circulation Research **82**(1) (1998) 63–81
17. O'Hara, T., Virág, L., Varró, A., Rudy, Y.: Simulation of the undiseased human cardiac ventricular action potential: model formulation and experimental validation. PLoS computational biology **7**(5) (2011)
18. Weinberg, S., Lipke, E.A., Tung, L.: In Vitro Electrophysiological Mapping of Stem Cells. Methods in Molecular Biology **660** (2010) 215–237
19. Gorospe, G., Zhu, R., Millrod, M., Zambidis, E., Tung, L., Vidal, R.: Automated grouping of action potentials of human embryonic stem cell-derived cardiomyocytes. IEEE Transactions on Biomedical Engineering (2014)
20. Zhang, Q., Jiang, J., Han, P., Yuan, Q., Zhang, J., Zhang, X., Xu, Y., Cao, H., Meng, Q., Chen, L., Tian, T., Wang, X., Li, P., Hescheler, J., Ji, G., Ma, Y.: Direct differentiation of atrial and ventricular myocytes from human embryonic stem cells by alternating retinoid signals. Cell Research **21**(4) (2011) 579–87
21. Zhang, X., Guo, J.P., Chi, Y.L., Liu, Y.C., Zhang, C.S., Yang, X.Q., Lin, H.Y., Jiang, E.P., Xiong, S.H., Zhang, Z.Y., Liu, B.H.: Endothelin-induced differentiation of Nkx2.5 cardiac progenitor cells into pacemaking cells. Molecular and Cellular Biochemistry **366**(1-2) (2012) 309–18

Estimating the Template in the Total Space with the Fréchet Mean on Quotient Spaces may have a Bias: a Case Study on Vector Spaces Quotiented by the Group of Translations

Stéphanie Allassonnière[1], Loïc Devilliers[1,2], and Xavier Pennec[3]

[1] École Polytechnique, Palaiseau, France.
[2] École Normale Supérieure de Cachan, France. loic.devilliers@ens-cachan.fr
[3] Asclepios project-team, INRIA Sophia Antipolis, France.

Abstract. When we have a deformation group acting on a vector space of observations, these data are not anymore elements of our space but rather orbits for the group action we consider. If the data are generated from an unknown template with noise, to estimate this template, one may want to minimise the variance in the quotient set. In this article we study statistics on a particular quotient space. We prove that the expected value of a random variable in our vector space mapped in the quotient space is different from the Fréchet mean in the quotient space when the observations are noisy.

Introduction

In the theory of shape introduced by Kendall [6], in Computational anatomy [4] or in image analysis, one often aims at estimating a template (which stands for the mean of the data) of shapes (for instance an average shape of an organ from a population of subject scans). To understand the observations, one assumes that these data follow a statistical model. A very popular one is that the observations are random deformations of the template with additional noise. This is the model proposed in [4] which is the foundation of Computational Anatomy. This introduces the notion of group action where the deformations we considered are elements of a group which acts on the set of objects, namely the images. In this particular setting, the template estimation is most of the time based on the minimization of the empirical variance in the quotient space (called the empirical Fréchet mean) (see for instance [7,5,9] among many others).

More precisely here, we consider M a finite dimensional vector space with a euclidean norm, G a finite group acting on M, such that the action is isometric with the respect of the euclidean norm on M. Thus the quotient M/G is equipped with a quotient distant, noted here ρ, moreover we call $[m]$ the orbit of $m \in M$.

We consider the following generative model: Y is a random variable on M with a density h for the Lebesgue measure. Y is not a random variable constant: it is the sum of a template and a white noise. And we aim to estimate this template. If Y was an observed variable, this question will be meaningless it would suffice to compute: $\int_M yh(y)dy$ which corresponds at the case where we have an infinite number of observations. Instead, here the random variable Y which lives in the total space M is not an observable variable, only $[Y]$ is observable.

In the following, the deformations of the group G will be restricted to translations and the noise will follow a general distribution. This kind of action is a simplified setting for image registration, for instance medical images can be obtained by translation of one scan to another due to different poses. More precisely, we work in the vector space $M = \mathbb{R}^{\mathbb{T}}$ where $\mathbb{T} = (\mathbb{Z}/N\mathbb{Z})^D$ is a discrete torus in D-dimension, an element of $\mathbb{R}^{\mathbb{T}}$ is seen as a function $y : \mathbb{T} \to \mathbb{R}$, $y(\tau)$ is the value at the pixel τ. When $D = 1$, y can be seen like a discretised signal with N pixels, when $D = 2$, we can see y like a picture with $N \times N$ pixels etc. We then define the group action of $G = \mathbb{T}$ on $\mathbb{R}^{\mathbb{T}}$ by:

$$\tau \in \mathbb{T},\ y \in \mathbb{R}^{\mathbb{T}} \quad \tau \cdot y : \sigma \mapsto y(\sigma + \tau). \tag{1}$$

We note $||\quad||$ the canonical Euclidean norm over $\mathbb{R}^{\mathbb{T}}$. We define a distance in the quotient space by:

$$\rho([y],[z]) = \inf_{\tau,\sigma\in\mathbb{T}} ||\tau \cdot y - \sigma \cdot z|| = \inf_{\tau\in\mathbb{T}} ||\tau \cdot y - z||. \tag{2}$$

Now the fact that Y has a density for the Lebesgue measure, implies that $[Y]$ has a density in M/G for the image measure noted ν. This density is given by: $\tilde{h}([y]) = \frac{1}{|G|}\sum_{g\in G} h(g\cdot y)$, therefore we can write the variance of $[Y]$ at the point $[\mu] \in M/G$ by:

$$\begin{aligned} F([\mu]) &= \mathbb{E}(\rho([\mu],[Y])^2) && (3)\\ &= \int_{M/G} \rho([\mu],z)^2 \tilde{h}(z)\nu(dz) && (4)\\ &= \frac{1}{|G|}\int_{M/G} \rho([\mu],[y])^2 \sum_{g\in G} h(g\cdot y)\nu(d[y])) && (5)\\ &= \int_M \rho([\mu],[y])^2 h(y)\lambda(dy) && (6)\\ &= \int_{\mathbb{R}^{\mathbb{T}}} \inf_{\tau\in\mathbb{T}} ||\tau y - \mu||^2 h(y)\lambda(dy) = J(\mu). && (7) \end{aligned}$$

J is non-negative, continuous, $\lim_{||\mu||\to+\infty} J(\mu) = +\infty$, therefore J reaches its minimum. The points in $\mathbb{R}^{\mathbb{T}}/\mathbb{T}$ which minimises F are the Fréchet means of $[Y]$. In this article, the central question is: is the template - which generates the random variable Y in the total space - mapped in the quotient space, a Fréchet mean of $[Y]$ or not?

About this kind of questions, previous works have been done before: for instance Allassonnière, Yali and Trouvé in [1] show an example of translated step function. They compared the iterative algorithm which numerically estimates the empirical Frechet mean in the quotient space to the Expectation-Maximization [3] algorithm which approximates the maximum likelihood estimator. In this example, even with a large number of observations, estimating the empirical Fréchet mean did not succeed to estimate well the template (the step function from which the synthetic samples were generated) when the noise on the observation was large enough.

To understand the example found in [1], different algorithms and theorems have been proposed (for instance in [2,7] or [10]), to improve or ensure the convergence of the empirical Fréchet mean in a more general case than presented in this article. A first contribution to provide a clue to know if even with an infinite number of observations, we could estimate the template has been given by Miolane and Pennec in [8]. They show that the presence of noise may imply that the template mapped in the quotient space is not a Fréchet mean in the quotient space. Then estimating the template in the total space with the Fréchet mean in the quotient space produces a bias. Considering the action of rotations on an euclidean space, they highlight the influence of dimension of the considered vector space and the influence of the ratio signal over noise on the bias. Although they showed a general result with a finite dimensional manifold and an isometric Lie group action, they made the assumption of a Gaussian noise. Here we do not make this assumption to show the presence of bias. For instance, here even with a bounded support of the density, under some condition we may have a bias. Moreover the method proposed here is different from [8], which can provide another explanation to the presence of bias in this context.

This paper is organised as follows. In Section 1, we show that the expected value of Y mapped in the quotient space is not a Fréchet mean of $[Y]$ as soon as the density of Y satisfies a certain condition. In Section 2, we compute the bias in a special case of torus with a Gaussian noise. This trivial example aims to give us an intuition of which parameters the bias depends on.

1 Existence of a bias for any discrete torus

In this section, we show that under some conditions of the density, the expected value of the random variable Y is not a minimum of J (defined in Equation (7)). To show that, we first study the differentiability of the integrand of J. Then we justify that the gradient of the variance J is the integral of the gradient's integrand. Finally we show that the gradient of the variance J at the expected value of the random variable Y is not zero. It will imply that the expected value of Y mapped in the quotient space in not a Fréchet mean of $[Y]$.

1.1 Study of the Differentiate of the Integrand

In this sub-part, we search to see when the integrand is differentiable, and to compute its gradient. In order to do that we defined:

$$\forall\, \mu,\, y \in \mathbb{R}^{\mathbb{T}} \quad f(\mu, y) = \inf_{\tau \in \mathbb{T}} ||\tau \cdot y - \mu||^2 h(y) dy. \tag{8}$$

Then we have: $J(\mu) = \int_{\mathbb{R}^{\mathbb{T}}} f(\mu, y) dy$, we will see that the differentiability of $\mu \mapsto f(\mu, y)$ at the point μ_0 depends on y and μ_0, more precisely the question on differentiability is related to the isotropy group of μ_0 and to the distances between y and $\tau \cdot \mu_0$ for $\tau \in \mathbb{T}$. Indeed one difficulty appears here: the inf of several differentiable functions is not necessary differentiable.

Remark 1. Let $f_1, \ldots, f_r : \mathbb{R}^n \to \mathbb{R}$ be differentiable functions at a point x_0, $f = \inf_{1 \leq i \leq r} f_i$ is differentiable at x_0 if: $\forall i,\, j \in [\![1, r]\!]^2,\, i \neq j \Longrightarrow f_i(x_0) \neq f_j(x_0)$. Indeed in this case, we take $k = \text{argmin}\{f_i(x_0), i \in [\![1, r]\!]\}$, we have locally around x_0: $f = f_k$. Then f is differentiable at x_0, and $\nabla f(x_0) = \nabla f_k(x_0)$ (where $\nabla f(x)$ is the gradient of f at the point x).

In Equation (8), let τ, τ' be two distinct elements of $\mathbb{T}$ then: $||\tau y - \mu|| = ||\tau' y - \mu||$ is equivalent to $||y - (-\tau)\mu|| = ||y - (-\tau')\mu||$,[4] there are two cases:

- If $(-\tau)\mu = (-\tau')\mu$ then $\forall y \in \mathbb{R}^{\mathbb{T}}\ ||\tau y - \mu|| = ||\tau' y - \mu||$.
- If for all $\tau \neq \tau'$ we have $\tau\mu \neq \tau'\mu$, i.e. the isotropy group is reduced to $\{0\}$, (the isotropy group is defined by: $\text{Iso}(\mu) = \{\tau \in \mathbb{T},\, \tau\mu = \mu\}$). We call such a μ a regular point, otherwise we say that μ is a singular point. We note

$$A_\mu = \bigcup_{\tau, \tau' \in \mathbb{T},\, \tau \neq \tau'} \{x \in \mathbb{R}^{\mathbb{T}},\, ||x - \tau \cdot \mu|| = ||x - \tau' \cdot \mu||\}. \tag{9}$$

  Foj μ regual, A_μ is the set of points equally distant from two points of the orbit of μ, A_μ is a finite union of hyperplanes, therefore the Lebesgue's measure of A_μ is null. In this case for every regular point μ and for all most every y (y does not belong to A_μ), the infimum in Equation (8) is reached at a unique $\tau \in \mathbb{T}$. When the infimum in Equation (8) is reached at a unique $\tau \in T$, we note this τ by:

$$\tau(y, \mu) = \text{argmin}\{||\tau \cdot y - \mu||,\, \tau \in \mathbb{T}\}. \tag{10}$$

We note $\text{Sing} = \{\mu \in \mathbb{R}^{\mathbb{T}},\ \text{such that } \text{Iso}(\mu) \neq \{0\}\}$ the set of singular points. Notice that: $\text{Sing} = \bigcup_{\tau \neq 0} \ker(x \mapsto \tau \cdot x - x)$ is a finite union of strict linear subspaces of $\mathbb{R}^{\mathbb{T}}$, then Sing is a null set for the Lebesgue's measure. For $\mu \notin \text{Sing}$ we have then for almost all y:

$$f(\mu, y) = \inf_{\tau \in \mathbb{T}} ||\mu - \tau y||^2 h(y) = ||\mu - \tau(y, \mu) y||^2 h(y). \tag{11}$$

[4] Because $||x|| = ||\tau x||$, and $\tau(x + y) = \tau x + \tau y$.

We can now apply the remark 1 to differentiate the integrand f defined in (8). But first we need to see how τ variates. Let μ be a regular point and $y \notin A_\mu$ therefore: $||\mu - \tau(y,\mu)\cdot y|| < \inf_{\alpha \neq \tau(y,\mu)} ||\mu - \alpha y||$. For continuity reason we have the existence of $\alpha_{\mu,y} > 0$, $\beta_{\mu,y} > 0$ such that for $\nu, z \in \mathbb{R}^{\mathbb{T}}$ verifying $||\mu - \nu|| < \alpha_{\mu,y}$, $||y - z|| < \beta_{\mu,y}$ we still have:

$$||\nu - \tau(y,\mu)\cdot z|| < \inf_{\alpha \neq \tau(y,\mu)} ||\nu - \alpha \cdot z||. \tag{12}$$

And then we have:

$$\forall \nu, z \in \mathbb{R}^{\mathbb{T}},\ ||\mu - \nu|| < \alpha_{\mu,y},\ ||y - z|| < \beta_{\mu,y} \Longrightarrow \tau(z,\nu) = \tau(y,\mu). \tag{13}$$

Finally, we can differentiate $\mu \mapsto f(\mu, y)$ with respect to μ in $\mu_0 \notin \text{Sing}$ and $y \notin A_{\mu_0}$, Equation (13) allow us to differentiate $\mu \mapsto \tau(y,\mu)$ (which is locally constant) which yields:

$$\frac{\partial f}{\partial \mu}(\mu_0, y) = 2(\mu_0 - \tau(y,\mu_0)y)h(y). \tag{14}$$

Now that we have seen the differentiability of the integrand, we justify in the next part that we can permute the differentiation and the integral sign.

1.2 Justification of the Differentiation of the Integral

In order to differentiate the variance in the quotient space (noted J), we propose to do the following things:

- Showing that $\mu \mapsto f(\mu, y)$ is weakly differentiable for almost all y, and computing its weak gradient.
- Deducing that J is weakly differentiable and finding its weak gradient ∇J.
- Showing that ∇J is continuous at some point, therefore by integration J is differentiable at these points, and ∇J is its strong gradient.

Remark 2. We can not apply here the theorem of differentiation under the integral sign, because $\mu \mapsto f(\mu, y)$ is differentiable at μ_0 for almost all y, but "the allmost y" is $\mathbb{R}^{\mathbb{T}} \setminus A_{\mu_0}$ depends of μ_0.

Weak differentiation of *f(,y)* for almost all *y*. First we define $\mathcal{C}_c^\infty(\mathbb{R}^{\mathbb{T}}, \mathbb{R})$ as the set of functions of infinite class whose support is a compact set. We want here to show that for almost all y ($y \notin \text{Sing}$):

$$\forall \varphi \in \mathcal{C}_c^\infty(\mathbb{R}^{\mathbb{T}}, \mathbb{R}) \quad \int f(\mu, y)\nabla\varphi(\mu)d\mu = -\int \varphi(\mu)\frac{\partial f}{\partial \mu}(\mu, y)d\mu. \tag{15}$$

Let $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}^{\mathbb{T}}, \mathbb{R})$ and by linearity it is sufficient to show:

$$\int f(\mu, y)\frac{\partial \varphi(\mu)}{\partial \mu_1}d\mu = -\int \frac{\partial f}{\partial \mu_1}(\mu, y)\varphi(\mu)d\mu.$$

Which is equivalent to:[5]

$$\int \left(\int f(\mu, y) \frac{\partial \varphi(\mu)}{\partial \mu_1} d\mu_1 \right) d\mu_2 \dots d\mu_{|\mathbb{T}|} = - \int \left(\int \frac{\partial f}{\partial \mu_1}(\mu, y) \varphi(\mu) d\mu_1 \right) d\mu_2 \dots d\mu_{|\mathbb{T}|}.$$

Hence, it is sufficient to show that for almost all $(\mu_2, \dots, \mu_{|T|})$ we have:

$$\int f(\mu, y) \frac{\partial \varphi(\mu)}{\partial \mu_1} d\mu_1 = - \int \frac{\partial f}{\partial \mu_1}(\mu, y) \varphi(\mu) d\mu_1. \tag{16}$$

There are two cases: $L = \{(x, \mu_2, \dots, \mu_{|\mathbb{T}|}),\ x \in \mathbb{R}\}$ is included in A_y, or not.
If L is not included in A_y then $L \cap A_y$ is finite, so in a connected component of $L \setminus A_y$, $f_L : x \mapsto f(x, \mu_2, \dots, \mu_{\mathbb{T}}, y)$ is derivable with a strong derivative:[6]

$$f'_L(x) = \langle (\mu - \tau(y, \mu) y) h(y) \,|\, e \rangle \ \text{ with } \mu = (x, \mu_2, \dots, \mu_{|\mathbb{T}|}),\ e = (1, 0, \dots, 0).$$

Therefore by cutting the integral by pieces, where each piece is a connected component on $L \setminus A_y$ and by integrating by part on each piece, we get[7] in this case Equation (16).
We note $B = \{(\mu_2, \dots \mu_{|\mathbb{T}|}) \in \mathbb{R}^{|\mathbb{T}|-1}, \forall x \in \mathbb{R}\, (x, \mu_2, \dots \mu_{|\mathbb{T}|}) \in A_y\}$ and we define for all $(\tau, \alpha) \in \mathbb{T}^2$ with $\tau \neq \alpha$:

$$\Psi_{\tau,\alpha} : \begin{pmatrix} \mathbb{R}^{|\mathbb{T}|-1} & \to & \text{Aff} \\ (\mu_2, \dots \mu_{|\mathbb{T}|}) & \mapsto & x \mapsto \Psi_{\tau,\alpha}(\mu_2, \dots, \mu_{|\mathbb{T}|})(x) \end{pmatrix}$$

$$\begin{aligned} \Psi_{\tau,\alpha}(\mu_2, \dots, \mu_{|\mathbb{T}|})(x) &= ||(x, \mu_2, \dots \mu_{|\mathbb{T}|}) - \tau y||^2 - ||(x, \mu_2, \dots \mu_{|\mathbb{T}|}) - \alpha y||^2 \\ &= 2 \left\langle (x, \mu_2, \dots, \mu_{|\mathbb{T}|}) \,|\, \alpha y - \tau y \right\rangle. \end{aligned}$$

where Aff is set of real affine maps, $\Psi_{\tau,\alpha}$ is well defined, affine, non zero (because $y \notin$ Sing), so $\Psi^{-1}_{\tau,\alpha}(\{0\})$ is a strict affine subspace of $\mathbb{R}^{|\mathbb{T}|-1}$ therefore: $B = \bigcup_{\tau \neq \alpha} \Psi^{-1}_{\tau,\alpha}(\{0\})$ is a null set.

To conclude, we have for almost all $(\mu_2, \dots \mu_{|\mathbb{T}|})$ the equation (16) which proves (15).

Remark 3. We did not show here that f belongs to a Sobolev space, because generally a Sobolev space is defined as the set of L^2 (or L^p) functions whose weak derivative exist and are in L^2 (or L^p), here $f(\,, y) \notin L^p$ for every $p > 1$, because $f(\mu, y) \to +\infty$ when $||\mu|| \to +\infty$. Instead we have shown that the derivative of the distribution associated to $\mu \mapsto f(\mu, y)$ is a distribution associated to another function (namely $\mu \mapsto \frac{\partial f}{\partial \mu}(\mu, y)$). The only thing we need in order to speak about a distribution associated to a function is that the function is integrable over each compact set of $\mathbb{R}^{\mathbb{T}}$ which is the case here.

[5] For writing $\mu \in \mathbb{R}^{\mathbb{T}}$ $\mu = (\mu_1, \dots, \mu_{|\mathbb{T}|})$ we suppose that we have chosen (once for all) an arbitrary order between the $|\mathbb{T}|$ real variables.

[6] By using the result in (14) by permuting the role of μ and y to ensure that the inf in (8) is unique.

[7] In fact, this is a particular case of the theorem of derivation of a distribution represented by a function with jumps, the derivative of a jump at the position a is obtained by a Dirac distribution function in a, here there is no Dirac because the function f is continuous.

Weakly differentiability of the variance in the quotient space We now prove that $J(\mu) = \int f(\mu, y)dy$ is weakly differentiable: Let $\varphi \in \mathcal{C}_c(\mathbb{R}^{\mathbb{T}}, \mathbb{R})$ then (by permutation of integrals thanks to Fubini's theorem):

$$\begin{aligned}\int J(\mu)\varphi'(\mu)d\mu &= \int \left(\left[\int f(\mu, y)dy \right] \varphi'(\mu) \right) d\mu \\ &= \int \left(\int f(\mu, y)\varphi'(\mu)d\mu \right) dy \\ &= -\int \int \varphi(\mu)\frac{\partial f}{\partial \mu}(\mu, y)d\mu dy \\ &= -\int \left(\int \frac{\partial f}{\partial \mu}(\mu, y)dy \right) \varphi(\mu)d\mu.\end{aligned}$$

Thus J is weakly differentiable, and its weakly derivative is:

$$\nabla J(\mu) = \int \frac{\partial f}{\partial \mu}(\mu, y)dy = 2\left(\mu - \int \tau(y, \mu)yh(y)dy\right). \tag{17}$$

Continuity of the weak gradient at the regular points. We show the continuity of the weak gradient at the regular points by simply applying the continuity under integral sign:
Let $\mu_0 \in \mathbb{R}^{\mathbb{T}} \setminus \text{Sing}$, then for $y \notin A_{\mu_0}$, $\mu \mapsto f(\mu, y)$ is continuous at μ_0 by Equation (13), moreover $||\tau(y, \mu) \cdot y||h(y) \leq ||y||h(y)$ with $y \mapsto ||y||h(y)$ an integrable function independent of μ. Therefore by the continuity under integral sign:

$$\nabla J(\mu) = 2\left(\mu - \int \tau(y, \mu) \cdot yh(y)dy\right), \tag{18}$$

is continuous at μ_0. This implies that the variance in the quotient space, (noted J) is differentiable over $\mathbb{R} \setminus \text{Sing}$ and its strong gradient over $\mathbb{R} \setminus \text{Sing}$ is:

$$\nabla J(\mu) = 2\left(\mu - \int_{\mathbb{R}^{\mathbb{T}}} \tau(y, \mu) \cdot yh(y)dy\right). \tag{19}$$

1.3 The expected value of Y mapped in the quotient space is not necessarily a Fréchet mean of $[Y]$

We suppose that $\mathbb{E}(Y)$ (noted y_0) the expected value of the random variable Y is a regular point (to ensure that J is differentiable at y_0) and verifies $\nabla J(y_0) = 0$, and we want to find a contradiction, we know from Equation (19) that:

$$\frac{1}{2}\nabla J(y_0) = \int_{\mathbb{R}^{\mathbb{T}}} yh(y)dy - \int_{\mathbb{R}^{\mathbb{T}}} \tau(y, y_0) \cdot yh(y)dy. \tag{20}$$

Therefore:

$$\frac{1}{2}\langle \nabla J(y_0) \,|y_0\rangle = \int_{\mathbb{R}^{\mathbb{T}}} \left(\langle y \,|y_0\rangle - \langle \tau(y, y_0) \cdot y \,|y_0\rangle\right) h(y)dy = 0. \tag{21}$$

We shall remember that $\tau(y, y_0)$ minimises $\{||\tau \cdot y - y_0||,\ \tau \in \mathbb{T}\}$ for almost all y. Then it minimises for almost all y:

$$\{||\tau \cdot y - y_0||^2 = ||y||^2 + ||y_0||^2 - 2\left\langle \tau y \,|y_0\right\rangle,\ \tau \in \mathbb{T}\},$$

and then almost surely $\tau(y, y_0)$ maximises:

$$\{\left\langle \tau \cdot y \,|y_0\right\rangle, \tau \in \mathbb{T}\},$$

This leads to:

$$\left\langle y \,|y_0\right\rangle - \left\langle \tau(y, y_0) \cdot y \,|y_0\right\rangle \leq 0 \text{ almost surely.}$$

So the integral of a non-positive function is null, so if we note $\mathrm{Supp}(h)$ the support of h we have then:

$$\forall y \in \mathrm{Supp}(h),\ \left\langle y \,|y_0\right\rangle = \left\langle \tau(y, y_0) \cdot y \,|y_0\right\rangle \text{ almost surely.} \tag{22}$$

Then $\tau = 0$ maximises the dot product almost surely. Therefore (as we know that $\tau(y, y_0)$ is unique almost surely, since y_0 is regular):

$$\forall y \in \mathrm{Supp}(h),\ \tau(y, y_0) = 0 \text{ almost surely.} \tag{23}$$

Let us suppose that the support of h contains a neighbourhood of $y \in \mathbb{R}^{\mathbb{T}}$ such that $\tau(y, y_0)$ is unique and $\tau(y, y_0) = \alpha \neq 0$, therefore: $||\alpha y - y_0|| < ||\tau y - y_0|| \quad \forall \tau \in \mathbb{T} \setminus \{\alpha\}$, and like in Equation (12), we have the existence of $r > 0$ such that:

$$\forall z \in B(y, r) \quad ||\alpha \cdot z - y_0|| < \inf_{\tau \in \mathbb{T}, \tau \neq \alpha} ||\tau z - y_0||. \tag{24}$$

Then for $z \in B(y, r)$ $\tau(z, y_0)$ is unique and $\tau(z, y_0) = \alpha \neq 0$, which is a contradiction with Equation (23). We have therefore proved the following theorem:

Theorem 1. *Let Y be a random variable of density h, whose expected value has a isotropy group reduced to $\{0\}$. If Supp(h) contains a neighbourhood of a point y such that $\tau(y, \mathbb{E}(Y)) \neq 0$ (which means that y is strictly closer to $\tau \cdot \mathbb{E}(Y)$ with some $\tau \neq 0$ than $\mathbb{E}(Y)$ itself), then we can say that $[\mathbb{E}(Y)]$ is not a Fréchet mean of $[Y]$ in the quotient of the space quotiented by the action of translations.*

2 Example in a very simple torus

In the previous part, we have shown that the expected value of Y can not be estimated by the Fréchet mean estimator. But we did not say how far this expected value of Y was from the set of all the Fréchet means in the quotient space. In this section, we take a very simple example: we take only two pixels: in other words here we work with $\mathbb{R}^{\mathbb{T}}$ where $\mathbb{T} = \mathbb{Z}/2\mathbb{Z}$, we can identify $\mathbb{R}^{\mathbb{T}}$ with $\mathbb{R}^2$ and work with the canonical basis of $\mathbb{R}^2$, we note by (u, v) the coordinates of an element of $\mathbb{R}^{\mathbb{T}}$. $0 \cdot (u, v) = (u, v)$ and $1 \cdot (u, v) = (v, u)$,. We note $L = \{(u, u),\ u \in \mathbb{R}\}$, and $\mathrm{HP} = \{(u, v),\ v > u\}$ the half-plane above the line L. Here we suppose that Y follows a Gaussian law of variance σ^2 and expected value $\mathbb{E}(Y) \in \mathrm{HP}$.

2.1 Graphical Interpretation of the Presence of the Bias

In this subpart we explain why there is a biasin this situation. Let $\mu \in \mathrm{HP}$ [8] (as it is the case for $\mathbb{E}(Y)$). We remind that $\tau(y,\mu)$ is an element of $\mathbb{T}$ which minimises $||\tau \cdot y - \mu||$ see (10)):

- If $y \in \mathrm{HP}$ then $\tau(y,\mu)\cdot y = y$, because μ, y are in the same half-plane delimited by L, and L is the perpendicular bisector of y and $1\cdot y$.
- If $y \notin \overline{\mathrm{HP}}$ then $\tau(y,\mu)\cdot y = 1\cdot y \in \mathrm{HP}$.

For $\mu \in \mathrm{HP}$, we define $Z = \tau(Y,\mu)\cdot Y$ (Z do not depend of $\mu \in \mathrm{HP}$ see above) we have $J(\mu) = \mathbb{E}(||Z-\mu||^2)$.

Lemma 1. *The global minimums of J are exactly: $\mathbb{E}(Z)$ and $1\cdot\mathbb{E}(Z)$.*

Proof. Let μ_0 be a global minimum of J, we know that $J(\mu) = J(1\cdot\mu)$. Without loss of generality we can assume that $\mu_0 \in \overline{\mathrm{HP}}$. Now as $\mathbb{E}(Z)$ is the expected value of Z we know that $\mathbb{E}(Z)$ is the only point where the variance of Z:

$$\begin{pmatrix} \mathbb{R}^{\mathbb{T}} \to & \mathbb{R}^{+} \\ \mu \mapsto & \mathbb{E}(||Z-\mu||^2) \end{pmatrix} \tag{25}$$

is minimal, moreover we know that Z takes value in $\overline{\mathrm{HP}}$, then for convexity reason $\mathbb{E}(Z) \in \overline{\mathrm{HP}}$. Then by restriction to $\overline{\mathrm{HP}}$, $\mathbb{E}(Z)$ is still the unique minimum of:

$$\begin{pmatrix} \overline{\mathrm{HP}} \to & \mathbb{R}^{+} \\ \mu \mapsto & \mathbb{E}(||Z-\mu||^2) \end{pmatrix}. \tag{26}$$

As a conclusion we have $\mu_0 = \mathbb{E}(Z)$. □

When we represent graphically these two random variables Y, Z, we can see that Y, Z have different means. This case shows graphically the bias. On the Fig. 1(a) and Fig. 1(b), we see the noise's influence: more the noise is important, more the mass under the line L is big and more the mean of Z is far from the expected value of Y. On the Fig. 1 we understand the condition of the density in the theorem 1, if the density's support is too small, then there are no mass under the line L, therefore $Y = Z$ and in this case there is no bias.

2.2 Localize the Fréchet mean

Thanks to the lemma 1, we can compute a Fréchet mean by computing $\mathbb{E}(Z)$, which is the sum of the area of the grey part (for the density h) and of the area of the black part (for the density h) in the Fig. 1(b), we have:

$$\mathbb{E}(Z) = \int_{v>u} (u,v)h(u,v)dudv + \int_{v<u} (v,u)h(u,v)dudv, \tag{27}$$

[8] For symmetry reason, because $J(\mu) = J(\tau\mu)$.

where (u, v) are the coordinates of a point in $\mathbb{R}^{\mathbb{T}} \simeq \mathbb{R}^2$. To compute (27) we convert to polar coordinates: $(u, v) = \mathbb{E}(Y) + (r\cos\theta, r\sin\theta)$ where $r > 0$ et $\theta \in [0, 2\pi]$. We also define: $d = \text{dist}(\mathbb{E}(Y), L)$. We get:

$$\mathbb{E}(Z) = \mathbb{E}(Y) + \int_d^{+\infty} \frac{r^2 \exp(-\frac{r^2}{2\sigma^2})}{\pi\sigma^2} \sqrt{2} g\left(\frac{d}{r}\right) dr \times (-1, 1), \tag{28}$$

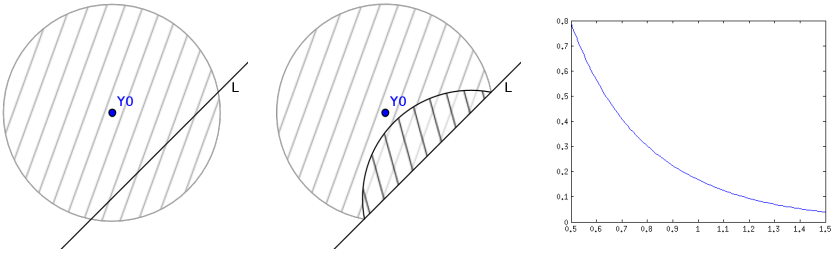
where g is a non-negative function on $[0, 1]$ defined by $g(x) = \sin(\arccos(x)) - x\arccos(x)$. Here we want to compute $\tilde{\rho} = \rho([\mathbb{E}(Y)], [\mathbb{E}(Z)])$ where ρ is the distance in the quotient space defined in (2). As we know that $\mathbb{E}(Y)$, $\mathbb{E}(Z)$ are in the same half-plane delimited by L, we have:

$$\tilde{\rho} = \rho([\mathbb{E}(Y)], [\mathbb{E}(Z)]) = ||\mathbb{E}(Y) - \mathbb{E}(Z)|| = 2\int_d^{+\infty} \frac{r^2 \exp(-\frac{r^2}{2\sigma^2})}{\pi\sigma^2} g\left(\frac{d}{r}\right) dr. \tag{29}$$

We can conclude that:

Theorem 2. *The Fréchet mean of* $[Y]$ *in the quotient space is an orbit of two points which are on the line passing through the expected value of* Y *and perpendicular to* L *and we compute the relative gap between the bias* $\tilde{\rho}$ *and* $d = dist(\mathbb{E}(Y), L)$ *by:*

$$\frac{\tilde{\rho}}{d} = \frac{\sigma}{d}\frac{2}{\pi}\int_{\frac{d}{\sigma}}^{+\infty} r^2 \exp\left(-\frac{r^2}{2}\right) g\left(\frac{d}{r\sigma}\right) dr. \tag{30}$$



(a) Graphic representation of Y, $y_0 = \mathbb{E}(Y)$ (b) Graphic representation of Z, $y_0 \neq \mathbb{E}(Z)$ (c) Variation of $\frac{\tilde{\rho}}{d}$ in function of $\frac{d}{\sigma}$.

Fig. 1. Z and Y have not the same mean, therefore there is a bias.

Remark 4. Here, contrarily to [8], it is not the ratio $||\mathbb{E}(Y)||$ over the noise which matters to estimate the bias, but the ratio $\text{dist}(\mathbb{E}(Y), L)$ over the noise which matters. But in fact, there it is not so different, in both case we measure the distance between the signal and the singularities (which is $\{0\}$ in [8] for the action of rotations, L in this case).

Discussion

In this article we have compared two notions of mean, one is the expectation of our random variable in our linear space, the other is the Fréchet mean in the space quotiented by translations. By differencing the variance in the quotient space we managed to show that when our random variable has a density whose support is large enough due to noise, the template in the total space mapped in the quotient space is not a Fréchet mean. But is the template mapped in the quotient space close to the Fréchet mean in the quotient space? We have answered to this question only in a special case of torus by computing the bias with a Gaussian noise. In this case, the bias depends on the scale of the noise and on the regularity of the signal, (measured here by how far our signal is from the set of singularity). In future work, we will generalise this estimation of the bias for a general torus $\mathbb{T} = (\mathbb{Z}/N\mathbb{Z})^D$, in order to see the influence of N (the number of pixels for each side of the picture) and D (the dimension of the picture) on the size of the bias. We have also seen the role played by the nature of the isotropy group for the presence of bias. This was already observed in [8], restricted to Gaussian noise.

In the section 1, we showed a bias for the Fréchet mean estimator with a particular group action defined in (1). But we have never used that definition. We have only used some properties of this group action: a finite group acts isometrically and effectively on a finite dimensional vector space. Therefore the theorem 1 generalises to any group with these properties.

What if the isotropy group is not reduced to $\{0\}$? Suppose now that G is a finite group acting isometrically and effectively on $\mathbb{R}^n$. Let suppose that $\mathbb{E}(Y)$ is singular, and that $\mathrm{Supp}(h)$ contains a neighbourhood of a point y such that y is strictly closer to $g_0 \cdot \mathbb{E}(Y)$ than $\mathbb{E}(Y)$ with some $g_0 \in G$. We make the extra assumption that it exists K a subgroup of G with $G = \{k \times i,\ k \in K,\ i \in Iso(\mathbb{E}(Y))\}$ and $K \cap \mathrm{Iso}(\mathbb{E}(Y)) = \{e_G\}$, then we have the same result: Indeed, by noting $y_0 = \mathbb{E}(Y)$, we have:

$$[y_0] = \{g \cdot y_0,\ g \in G\} = \{g \cdot y_0,\ g \in K\} \tag{31}$$

Therefore if we note:

$$\tilde{J}(\mu) = \int_{\mathbb{R}^n} \inf_{g \in K} ||y - g \cdot \mu||^2 h(y) dy \tag{32}$$

Then we have: $J(y_0) = \tilde{J}(y_0)$ and for all $\mu \in \mathbb{R}^n$ $\tilde{J}(\mu) \geq J(\mu)$, now if we consider the action of G but restricted to K (the action is still isometric, and effective: since no element leaves y_0, moreover by writing $g_0 = k \times i$ with $iy_0 = y_0$, and $k \in K$, y is strictly closer to $k \cdot \mathbb{E}(Y)$ than $\mathbb{E}(Y)$), then we know (by the theorem (1), with y_0 a regular point for the action of the group K) that it exists $\mu \in \mathbb{R}^n$ such that $\tilde{J}(\mu) < \tilde{J}(y_0)$, therefore $J(y_0) = J(\tilde{y}_0) > \tilde{J}(\mu) \geq J(\mu)$, then y_0 is still not a minimum of the variance J, therefore the expected value

of Y mapped in the quotient space is not a Frechet mean of $[Y]$ even in this case.

In a more general case: when we take an infinite-dimensional vector space quotiented by a group action, for instance when the group is a subgroup of the group of smooth diffeomorphism, is there always a bias? And when it does, can we measure the bias in function of the scale of the noise and the distance between the template and the singularities? Figure 1(c) shows us that the bias is not so important in favourable cases: when the noise is low and the signal far from the singularities. Then we can hope that it will be also the case in a more general case. If so, one could keep using the Fréchet mean in the quotient space in order to estimate the template in the total space.

References

1. Allassonnière, S., Amit, Y., Trouvé, A.: Towards a coherent statistical framework for dense deformable template estimation. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69(1), 3–29 (2007)
2. Allassonnière, S., Kuhn, E., Trouvé, A.: Construction of bayesian deformable models via a stochastic approximation algorithm: a convergence study. Bernoulli 16(3), 641–678 (2010)
3. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society 1, 1–22 (1977)
4. Grenander, U., Miller, M.I.: Computational anatomy: An emerging discipline. Q. Appl. Math. LVI(4), 617–694 (Dec 1998)
5. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. Neuroimage 23, S151–S160 (2004)
6. Kendall, D.G.: A survey of the statistical theory of shape. Statistical Science pp. 87–99 (1989)
7. Kurtek, S.A., Srivastava, A., Wu, W.: Signal estimation under random time-warpings and nonlinear signal alignment. In: Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., Weinberger, K. (eds.) Advances in Neural Information Processing Systems 24, pp. 675–683. Curran Associates, Inc. (2011)
8. Miolane, N., Pennec, X.: Biased estimators on quotient spaces. Geometric Science of Information (2015)
9. Sabuncu, M., Balci, S.K., Golland, P.: Discovering modes of an image population through mixture modeling. Proceeding of the MICCAI conference LNCS(5242), 381–389 (2008)
10. Zhang, M., Singh, N., Fletcher, P.: Bayesian estimation of regularization and atlas building in diffeomorphic image registration. In: Gee, J., Joshi, S., Pohl, K., Wells, W., Zöllei, L. (eds.) Information Processing in Medical Imaging, Lecture Notes in Computer Science, vol. 7917, pp. 37–48. Springer Berlin Heidelberg (2013)

An efficient recursive estimator of the Fréchet mean on a hypersphere with applications to Medical Image Analysis

Hesamoddin Salehian[1], Rudrasis Chakraborty[1], Edward Ofori[2], David Vaillancourt[2], and Baba C. Vemuri[1] ⋆

[1] Department of CISE, University of Florida, Gainesville, Florida, USA ⋆⋆
{salehian, rudrasis, vemuri}@cise.ufl.edu
[2] Department of Applied Physiology and Kinesiology, University of Florida, Florida, USA
{eofori,vcourt}@ufl.edu

Abstract. Finding the Riemannian center of mass or the Fréchet mean (FM) of manifold-valued data sets is a commonly encountered problem in a variety of fields of Science and Engineering including but not limited to, Medical Image Computing, Machine Learning, and Computer Vision. For instance, it is encountered in tasks such as, atlas construction, clustering, principal geodesic analysis etc. Traditionally, algorithms for computing the FM of the manifold-valued data require that the entire data pool be available apriori and not incrementally. When encountered with new data, the FM needs to be recomputed over the entire pool, which can be computationally as well as storage inefficient. A computational and storage efficient alternative is to consider a recursive algorithm for computing the FM which simply updates the previously computed FM when presented with a new data set. In this paper, we present such an alternative called the incremental Fréchet mean estimator (*iFME*) for data on the hypersphere. We prove the asymptotic convergence of *iFME* to the true FM of the underlying distribution from which the data samples were drawn. Further, we present several experiments demonstrating the performance on synthetic and real data sets.

1 Introduction

With the advent of sophisticated sensing technologies, manifold-valued data sets have become pervasive in many fields of applied sciences and Engineering including Medical Image Computing, Machine Learning and Computer Vision. Among these data, the most widely encountered are those that lie on a $k-$sphere, $k \geq 2$. To mention a few, the directional data which are often encountered in Image Processing and Computer Vision are points on the unit $2-$sphere $\mathbb{S}^2$ [15]. Further, 3×3 rotation matrices can be parameterized by unit quaternions which can be represented by points on the 3-dimensional unit sphere $\mathbb{S}^3$ [9]. Also, any probability density function, e.g., Orientation Distribution Function (ODF) in diffusion magnetic resonance imaging (MRI) [23], can be represented as points on a unit Hilbert sphere [4,20].

⋆ Corresponding author.

⋆⋆ This research was supported in part by the NIH grant NS066340 to BCV

In most of these applications, mean computation is a key ingredient. Examples include, the interpolation and smoothing of ODF fields [5,4,8], estimation of the mean rotation from several corresponding pairs of points in multi-view geometry [9] and statistical analysis of directional data [15]. Given a set of samples on $\mathbb{S}^k$, the *Fréchet mean* (FM), is defined as the minimizer of the sum of squared geodesic distances. In general, the minimizer is non-unique and this issue has been well studied in literature and we refer the reader to [1,16] and references therein for details. It is also known that for a set of more than two samples on a hypersphere, the FM cannot in general be computed in closed form, and iterative schemes like the gradient descent must be employed [1,16] which for very large data sets can prove to be computationally quite expensive. Further, in many real-world applications the entire input data are not available all at once, and the population is usually augmented over time. Hence, in this context the standard gradient descent based iterative computation of the FM suffers from two major drawbacks: (1) for each new sample, it has to compute the new FM from scratch, and (2) it requires the entire input data to be stored, in order to estimate the new FM. Instead, an *incremental* i.e., a recursive technique can address this problem more efficiently with respect to time/space utility.

Recently, several incremental mean estimators for manifold-valued data have been reported. In [21], Sturm presented an incremental mean, the so called *inductive mean*, and proved its convergence to the true FM for all non-positively curved (NPC) spaces. In [7], authors showed several algorithms (including a recursive algorithm) for FM computation for data residing in CAT(0) spaces, which are NPC. They also demonstrated several applications of the same to Computer Vision and Medical Imaging. Further, in [10] an incremental FM computation algorithm along with its convergence and applications was presented for a population of Symmetric Positive Definite (SPD) matrices. Recently, in [14], Lim presented an inductive FM to estimate the weighted FM of SPD matrices. The convergence analysis in all of these works is applicable only to the samples belonging to NPC spaces and hence, their convergence analysis does not apply to the case of the hypersphere which is a positively curved Riemannian manifold with constant sectional curvature [11]. In [3], Arnaudon et al. present a stochastic gradient descent algorithm for barycenter computation of probability measures on Riemannian manifolds under some conditions. They also proved that their algorithm almost surely converges to the true Riemannian barycenter. Their algorithm is a stochastic version of ours as well as that of Sturm [21].

In this paper, we present a novel incremental FM estimator (*iFME*) of a set of samples on the hypersphere. When encountered with a new sample data set, an incremental update of the previously estimated FM is more computationally efficient compared to the non-incremental counterpart (*henceforth denoted by nFM*), because the update problem involves just the weighted FM of two items (previously computed mean and the new sample) and no optimization method is needed for its computation. This leads to significant efficiency in time and space (storage) consumption. Further, we will analytically show that in the limit (over the number of samples), our estimator converges to the true FM of the distribution from which the samples are drawn. To the best of our knowledge, this is the first convergence analysis for an incremental FM estimator on a positively curved Riemannian manifold. Finally, we will present examples of recursive

FM computation on several synthetic and real data sets along with its application to an incremental principal geodesic analysis *iPGA* algorithm which is used in the classification of movement disorder patients from their diffusion MR scans.

2 Riemannian Geometry of the Hypersphere

The hypersphere is the simplest of the constant positive curvature Riemannian manifolds encountered in numerous application problems. Its geometry is well known and here we will simply present the closed form expressions for the Riemannian Exponential and Log maps as well as the geodesic between two points on it. Further, we also present the square root parametrization of probability density functions, which allows one to identify them with points on the unit Hilbert sphere. This will be needed in representing the probability density functions namely, the ensemble average propagators (EAPs) in diffusion MRI, as points on the unit Hilbert sphere.

Without loss of generality we restrict the analysis to PDFs defined on the interval $[0, T]$ for simplicity: $\mathcal{P} = \{p : [0, T] \rightarrow \mathbb{R} | \forall s, p(s) \geq 0, \int_0^T p(s)ds = 1\}$. In [17], the Fisher-Rao metric was introduced to study the Riemannian structure of a statistical manifold (the manifod of probability densities). For a PDF $p_i \in \mathcal{P}$, the Fisher-Rao metric is defined as $\langle v_j, v_k \rangle = \int_0^T v_j(s)v_k(s)p_i(s)ds$, where $v_j, v_k \in T_{p_i}\mathcal{P}$. The Fisher-Rao metric is invariant to reparameterizations of the functions. In order to facilitate easy computations when using Riemannian operations, the square root density representation $\psi = \sqrt{p}$ was used in [20]. The space of square root density functions is defined as $\Psi = \{\psi : [0, T] \rightarrow \mathbb{R} | \forall s, \psi(s) \geq 0, \int_0^T \psi^2(s)ds = 1\}$. As we can see, Ψ forms a convex subset of the unit sphere in a Hilbert space. Then, the Fisher-Rao metric can be written as $\langle v_j, v_k \rangle = \int_0^T v_j(s)v_k(s)ds$ where, $v_j, v_k \in T_{\psi_i}\Psi$ are tangent vectors. Given any two functions $\psi_i, \psi_j \in \Psi$, the geodesic distance between these two points is given in closed form by $d(\psi_i, \psi_j) = \cos^{-1}(\langle \psi_i, \psi_j \rangle)$ The geodesic at ψ_i with a direction $v \in T_{\psi_i}\Psi$ is defined as $\gamma(t) = \cos(t)\psi_i + \sin(t)\frac{v}{|v|}$. Then, the Riemannian exponential map can be expressed as $exp_{\psi_i}(v) = \cos(|v|)\psi_i + \sin(|v|)\frac{v}{|v|}$, where, $|v| \in [0, \pi)$. The Riemannian logarithmic map is then given by $log_{\psi_i}(\psi_j) = u \cos^{-1}(\langle \psi_i, \psi_j \rangle)/\sqrt{\langle u, u \rangle}$ where, $u = \psi_j - \langle \psi_i, \psi_j \rangle \psi_i$.

Using the geodesic distance provided above, one can define the Fréchet mean (FM) of a set of points on the hypersphere as the minimizer of the sum of squared geodesic distances (so called Fréchet functional). Let $B(C, \rho)$, be the geodesic ball centered at C with radius ρ, i.e., $B(C, \rho) = \{Q \in \mathbb{S}^k | d(C, Q) < \rho\}$. Authors in [13] showed that for any $C \in \mathbb{S}^k$ and for data samples in $B(C, \frac{\pi}{2})$, the minimizer of the Fréchet functional exists and is unique. Therefore, in the rest of the paper, we assume that this condition is satisfied for any set of given points, $X_i \in \mathbb{S}^k$. For more details on Riemannian geometry of the sphere, reader is referred to chapter 2 of [11] and references therein.

3 Weak Consistency of *iFME* on the Sphere

In this section, we present the detailed proof of convergence of our recursive estimator on $\mathbb{S}^k$. The proposed method is similar in "spirit" to the incremental arithmetic mean

update in the Euclidean space; given the old mean, M_{n-1}, and the new sample, X_n, we define the new mean, M_n, as the weighted mean of M_{n-1} and X_n with the weights being $\frac{n-1}{n}$ and $\frac{1}{n}$, respectively. From a geometric viewpoint, this corresponds to the choice of the point on geodesic curve between M_{n-1} and X_n, with the parameter $t = \frac{1}{n}$.

Formally, let $X_1, X_2, ..., X_N$ be a set of N samples on hypersphere $\mathbb{S}^k$, all of which belong to the geodesic ball of radius $\frac{\pi}{2}$). The *iFME* estimate M_n of the FM with the n^{th} given sample X_n is defined by:

$$M_1 = X_1 \quad (1)$$
$$M_n = M_{n-1} \#_{\frac{1}{n}} X_n \quad (2)$$

where $A\#_t B$ is the point on the shortest geodesic path from A to B ($\in \mathbb{S}^k$) for a parameter value of t, and $\frac{1}{n}$ is the weight assigned to the new sample point (in this case the n^{th} sample), which is henceforth called the *Euclidean weight*. In the rest of this section, we will show that if the number of given samples, N, tends to infinity, the *iFME* estimates will converge to the FM of the distribution from which the samples are drawn. *Note that the proof steps given below are not needed to compute the iFME, these steps are needed only to prove the weak consistency of iFME.* Our proof is based on the idea of projecting the samples on the sphere, X_i, to the tangent plane using the Gnomonic Projection [9], and perform the convergence analysis on the projected samples in this linear space, i.e., $\mathbf{x_i}$, instead of doing the analysis on the hypersphere. We take advantage of the fact that the geodesic curve between any pair of points on the hemisphere, is projected to a straight line in the tangent space at the anchor point (in this case, without loss of generality, assumed to be the north pole), via the gnomonic projection. A figure depicting the Gnomonic Projection is shown in Fig. 1.

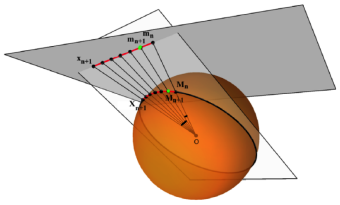


Fig. 1: Gnomonic Projection

Despite the simplifications used in the statistical analysis of the *iFME* estimates on the hypersphere using the gnomonic projection, there is one important obstacle that must be considered. Without loss of generality, suppose the true FM of the input samples, X_i, is the north pole. Then, it can be shown through counter examples that:

– The use of Euclidean weights, $\frac{1}{n}$, to update the *iFME* estimates on $\mathbb{S}^k$, does **not** necessarily correspond to the same weighting scheme between the old arithmetic mean and the new sample, in the projection space i.e., the tangent space.

The above fact can be illustrated using two sample points on a unit circle ($\mathbb{S}^1$), $X_1 = \pi/6$ and $X_2 = \pi/3$, whose intrinsic mean is $M = \pi/4$. Then, the midpoint of the gnomonic projections of X_1 and X_2, which are denoted by $\mathbf{x_1}$ and $\mathbf{x_2}$, is $\hat{\mathbf{m}} = \frac{tan(\pi/3)+tan(\pi/6)}{2} = 1.1547 \neq tan(\pi/4) = \mathbf{m}$ (see Fig. 2).

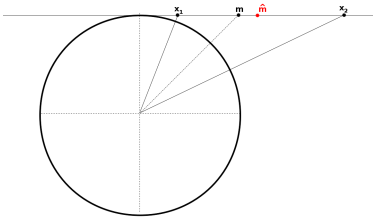


Fig. 2: Illustration of the counterexample

For the rest of this section, without loss of generality, we assume that the true FM of N given samples is located at the

north pole. Since the gnomonic projection space is anchored at the north pole, this assumption leads to significant simplifications in our convergence analysis. However, a similar convergence proof can be developed for any arbitrary location of the FM, with the tangent (projection) space anchored at the location of this mean.

In what follows, we prove that the use of Euclidean weights, i.e., $w_n = \frac{1}{n}$, to update the incremental FM on the hypersphere, corresponds to a set of weights in the projection space, denoted henceforth by t_n, for which the weighted incremental mean in the tangent plane, converges to the true FM on the hypersphere, which in this case is the point of tangency.

Theorem 1 (Angle Bisector Theorem). *[2] Let M_n and M_{n+1} denote the iFME estimates for n and $n+1$ given samples, respectively, and X_{n+1} denotes the $(n+1)^{st}$ sample. Further, let $\mathbf{m_n}, \mathbf{m_{n+1}}, \mathbf{x_{n+1}}$ be the corresponding points in the projection space, then*

$$t_n = \frac{||\mathbf{m_n} - \mathbf{m_{n+1}}||}{||\mathbf{x_{n+1}} - \mathbf{m_{n+1}}||} = \frac{||O - \mathbf{m_n}||}{||O - \mathbf{x_{n+1}}||} \times \frac{\sin(d(M_n, M_{n+1}))}{\sin(d(M_{n+1}, X_{n+1}))} \tag{3}$$

where, $d(.)$ is the geodesic distance on the hypersphere.

In the rest of this section, we assume that the input samples, X_i, are within the geodesic ball, $B(C, \phi)$, where $0 < \phi < \pi/2$. This is needed for the uniqueness of the FM on the hypersphere (see [1]). Then, we bound t_n with respect to the radius ϕ.

Lemma 1 (Lower and Upper Bounds for t_n). *With the same assumptions made as in Theorem 1, the following inequality holds:*

$$\frac{\cos(\phi)}{n} \leq t_n \leq \frac{1}{\cos(\phi)^3 n} \tag{4}$$

Lower Bound. To prove this lower bound for t_n, we find the lower bounds for each fraction on the right hand side of Eq. 3. The first term reaches its minimum value, if M_n is located at the north pole, and X_{n+1} is located on the boundary of the geodesic ball, $B(C, \phi)$. In this case, $||O - \mathbf{m_n}|| = 1$ and $||O - \mathbf{x_{n+1}}|| = \frac{1}{\cos(\phi)}$. This implies that:

$$\frac{||O - \mathbf{m_n}||}{||O - \mathbf{x_{n+1}}||} \geq \cos(\phi) \tag{5}$$

Next, note that based on the definition of *iFME*, this second fraction in 3 can be rewritten as $\frac{\sin(d(M_n, M_{n+1}))}{\sin(d(M_{n+1}, X_{n+1}))} = \frac{\sin(d(M_n, M_{n+1}))}{\sin(n \times d(M_n, M_{n+1}))} = \frac{1}{U_{n-1}(\cos(d(M_n, M_{n+1})))}$ where, $U_{n-1}(x)$ is the Chebyshev polynomial of the second kind [19]. For any $x \in [-1, 1]$, the maximum of $U_{n-1}(x)$ is reached when $x = 1$, for which $U_{n-1}(1) = n$. Therefore, $U_{n-1}(x) \leq n$ and $\frac{1}{U_{n-1}(x)} \geq \frac{1}{n}$. This implies that:

$$\begin{aligned} \frac{\sin(d(M_n, M_{n+1}))}{\sin(n \times d(M_{n+1}, M_{n+1}))} &= \frac{1}{U_{n-1}(\cos(d(M_n, M_{n+1})))} \\ &\geq \frac{1}{n} \end{aligned} \tag{6}$$

The inequalities 5 and 6, complete the proof. ■

Note that when ϕ tends to zero, $\cos(\phi)$ converges to one, and this lower bound tends to $\frac{1}{n}$, which is the case in Euclidean space.

Upper Bound. First, the upper bound for the first term in 3 is reached when M_n is on the boundary of geodesic ball, and X_{n+1} is given at the north pole. Therefore,

$$\frac{||O - \mathbf{m_n}||}{||O - \mathbf{x_{n+1}}||} \leq \frac{1}{\cos(\phi)} \tag{7}$$

Finding the upper bound for the $\sin$ term however is quite involved. Note that the maximum of the angle between OM_n and OX_{n+1}, denoted by α, is reached when M_n and X_{n+1} are both on the boundary of the geodesic ball, i.e., $\alpha \leq 2\phi$. Therefore, $\phi \in [0, \frac{\pi}{2})$ implies that $\alpha \in [0, \pi)$. Further, we show in the Appendix that the following inequality holds for any $\alpha \in (0, \pi)$.

$$\frac{\sin(\frac{n\alpha}{n+1})}{\sin(\frac{\alpha}{n+1})} \geq n\cos^2(\frac{\alpha}{2}) = n\cos^2(\phi) \tag{8}$$

From 7 and 6, the result follows. ■

Thus far, we have shown analytical bounds for the sequence of weights, t_n, in the projection space, corresponding to Euclidean weights on sphere (Eq. 4). We now prove the convergence of *iFME* estimates to the true FM of distribution from which the samples are drawn, when the number of samples tend to infinity.

Theorem 2 (Unbiasedness). *Let (σ, ω) denote a probability space with probability measure ω. A vector valued random variable, $\mathbf{x}$ is a measurable function on σ taking values in $\mathbb{R}^k$, i.e., $\mathbf{x} : \sigma \to \mathbb{R}^k$. The distribution of $\mathbf{x}$ is the push-forward probability measure, $dP(\mathbf{x}) = \mathbf{x}^*(\sigma)$ on $\mathbb{R}^k$. The expectation is defined by $E[\mathbf{x}] = \int_\sigma \mathbf{x} d\omega$. Let $\mathbf{x_1}, \mathbf{x_2}, ...$ be i.i.d. samples from the distribution of $\mathbf{x}$. Also, let $\mathbf{m_n}$ be the incremental mean estimate corresponding to n^{th} given sample, $\mathbf{x_n}$, which is defined by: (i) $\mathbf{m_1} = \mathbf{x_1}$, (ii) $\mathbf{m_n} = t_n\mathbf{x_n} + (1 - t_n)\mathbf{m_{n-1}}$. Then, $\mathbf{m_n}$ is an unbiased estimator of the expectation $E[\mathbf{x}]$.*

Proof. For $n = 2$; $\mathbf{m_2} = t_2\mathbf{x_2} + (1 - t_2)\mathbf{x_1}$, hence $E[\mathbf{m_2}] = t_2E[\mathbf{x}] + (1 - t_2)E[\mathbf{x}] = E[\mathbf{x}]$.

By induction hypothesis we have, $E[\mathbf{m_{n-1}}] = E[\mathbf{x}]$. Then, $E[\mathbf{m_n}] = t_nE[\mathbf{x}] + (1 - t_n)E[\mathbf{x}] = E[\mathbf{x}]$, hence the result. ■

Theorem 3 (Weak Consistency). *Let $var[\mathbf{m_n}]$ denotes the variance of the n^{th} incremental mean estimate (which is defined in Theorem 2), with $\frac{\cos(\phi)}{n} \leq t_n \leq \frac{1}{\cos(\phi)^3 n}, \forall \phi \in [0, \pi/2)$. Then, $\exists p \in (0, 1]$, such that $\frac{var[\mathbf{m_n}]}{var[\mathbf{x}]} \leq (n^p \cos^6(\phi))^{-1}$.*

First note that $var[\mathbf{m_n}] = t_n^2 var[\mathbf{x}] + (1 - t_n)^2 var[\mathbf{m_{n-1}}]$. Since, $0 \leq t_n \leq 1$, one can see that $var[\mathbf{m_n}] \leq var[\mathbf{x}]$ for all n. Besides, for each n, the maximum of the right hand side is achieved, when t_n attains either its minimum or its maximum value. Therefore, we need to prove the theorem for the following two values of t_n, (i) $t_n = \frac{\cos(\phi)}{n}$ and (ii) $t_n = \frac{1}{n\cos^3(\phi)}$. These two cases will be proved in the Lemmas 2 and 3 respectively.

Lemma 2. *Suppose the same assumptions as in Theorem 2 are made. Further,* $t_n = \frac{1}{n\cos^3(\phi)}$, $\forall n$ *and* $\forall \phi \in [0, \pi/2)$, *then* $\frac{var[\mathbf{m_n}]}{var[\mathbf{x}]} \leq (n\cos^6(\phi))^{-1}$.

Proof. For $n = 1$, $var[\mathbf{m_1}] = var[\mathbf{x}]$ which yields the result, since $\cos(\phi) \leq 1$. Now, assume by induction that $\frac{var[\mathbf{m_{n-1}}]}{var[\mathbf{x}]} \leq (n-1)\cos^6(\phi))^{-1}$. Then,

$$\begin{aligned}\frac{var[\mathbf{m_n}]}{var[\mathbf{x}]} = t_n^2 + (1-t_n)^2\frac{var[\mathbf{m_{n-1}}]}{var[\mathbf{x}]} &\leq t_n^2 + (1-t_n)^2\frac{1}{(n-1)\cos^6(\phi)} \\ &\leq \frac{1}{\cos^6(\phi)n^2} + (1-\frac{1}{\cos^3(\phi)n})^2 \times \frac{1}{(n-1)\cos^6(\phi)} \\ &\leq \frac{1}{\cos^6(\phi)n^2} + (1-\frac{1}{n})^2 \times \frac{1}{(n-1)\cos^6(\phi)} \\ &= \frac{1}{\cos^6(\phi)n^2} + \frac{n-1}{n^2\cos^6(\phi)} = \frac{1}{n\cos^6(\phi)}\end{aligned} \tag{9}$$

■

Lemma 3. *Suppose the same assumptions as in Theorem 2 hold. Further,* $t_n = \frac{\cos(\phi)}{n}$, $\forall n$ *and* $\forall \phi \in [0, \pi/2)$, *then,* $\frac{var[\mathbf{m_n}]}{var[\mathbf{x}]} \leq n^{-p}$ *for some* $0 < p \leq 1$..

Proof. For $n = 1$, $var[\mathbf{m_n}] = var[\mathbf{x}]$ which yields the result, since $\cos(\phi) \leq 1$. Now, assume by induction that $\frac{var[\mathbf{m_{n-1}}]}{var[\mathbf{x}]} \leq (n-1)^{-p}$. Then,

$$\begin{aligned}\frac{var[\mathbf{m_n}]}{var[\mathbf{x}]} = t_n^2 + (1-t_n)^2\frac{var[\mathbf{m_{n-1}}]}{var[\mathbf{x}]} &\leq t_n^2 + (1-t_n)^2\frac{1}{(n-1)^p} \\ &\leq \frac{\cos^2(\phi)}{n^2} + \frac{(n-\cos(\phi))^2}{n^2} \times \frac{1}{(n-1)^p} \\ &= \frac{(n-1)^p\cos^2(\phi) + \cos^2(\phi) - 2n\cos(\phi) + n^2}{n^2(n-1)^p}\end{aligned} \tag{10}$$

Now, it suffices to show that the numerator of the above expression is not greater than $n^{2-p}(n-1)^p$. In other words:

$$(n-1)^p\cos^2(\phi) + \cos^2(\phi) - 2n\cos(\phi) + n^2 - n^{2-p}(n-1)^p \leq 0 \tag{11}$$

The above quadratic function in $\cos(\phi)$ is less than zero, when

$$n(\frac{1-(n-1)^{p/2}\sqrt{(\frac{n-1}{n})^p + \frac{1}{n^p} - 1}}{1+(n-1)^p}) \leq \cos(\phi) \leq n(\frac{1+(n-1)^{p/2}\sqrt{(\frac{n-1}{n})^p + \frac{1}{n^p} - 1}}{1+(n-1)^p}) \tag{12}$$

The inequality on the right is satisfied for all values of the cos function. Besides, it is easy to see that the function on the left hand side is increasing w.r.t. $n > 1$, and hence attains its minimum when $n = 2$. This implies that:

$$\begin{aligned}1 - \sqrt{2^{1-p} - 1} &\leq \cos(\phi) \\ \rightarrow \phi &\leq \cos^{-1}(1 - \sqrt{2^{1-p} - 1}) \\ \rightarrow 0 < p &\leq 1 - log_2[(1-\cos(\phi))^2 + 1]\end{aligned} \tag{13}$$

Note that $p > 0$, for all $\phi < \pi/2$. ■

Convergence. Armed with the above two results, it is easy to see that $\forall \phi \in [0, \pi/2)$, there exists a p satisfying $0 < p \leq 1$, such that

- If $t_n = \frac{\cos(\phi)}{n}$, then $\frac{var[\mathbf{m_n}]}{var[\mathbf{x}]} \leq \frac{1}{n^p} \leq \frac{1}{n^p \cos^6(\phi)}$, because $\cos(\phi) \leq 1$.
- If $t_n = \frac{1}{n \cos^3(\phi)}$, then $\frac{var[\mathbf{m_n}]}{var[\mathbf{x}]} \leq \frac{1}{n \cos^6(\phi)} \leq \frac{1}{n^p \cos^6(\phi)}$, because $p \leq 1$.

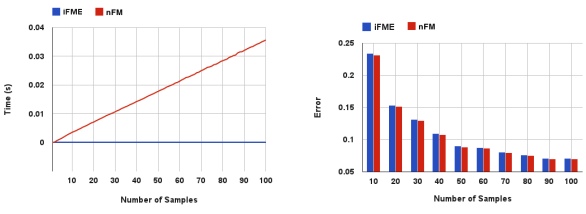
These two pieces together complete the proof of convergence. ■

The inequality in Theorem 3 implies that when $n \to \infty$, for any $\phi \in [0, \pi/2)$ the variance of *iFME* estimates in the projection space tends to zero. Besides, when ϕ approaches $\pi/2$, the corresponding power of n, as well as $\cos(\phi)$, become very small, hence the rate of convergence gets slower. Note that instead of the weights scheme used here (i.e., in spirit of incremental mean in Euclidean space), one can choose different weights scheme inherent to the manifold (i.e., as a function of curvature) to speed up the convergence rate.

4 Experimental Results

We now evaluate the effectiveness of the *iFME* algorithm, compared to the non-incremental counterpart, *nFM*, for computing the FM of a finite set of samples on the sphere (northern hemi-sphere not including the equator). As mentioned earlier, *nFM* for computing the FM uses a gradient descent technique to minimize the sum of squared geodesic distances cost function. We report the results for samples drawn from a mixture of Log-Normal distribution on the upper hemi-sphere. A set of random samples are drawn from the distribution and fed to both the *iFME* and the *nFM* algorithms, incrementally. The computation time needed by each method for computing the sample FM, and the error was recorded, for each new sample incrementally introduced. The error is defined by the geodesic distance between the estimated mean (using either *iFME* or the *nFM*) and the true expected value of the input distribution. Because of the randomness in generating the samples, we repeated this experiment 100 times for each case, and the mean time consumption and the error for each method are shown. All the computation time required for various algorithms reported in this paper, were measured on an Intel-7 quad-core processor, 25GB RAM desktop.

A set of samples are drawn from a mixture of Log-Normal distributions on the sphere. The mean of each Log-Normal component is set randomly, and the covariance matrices are set to $0.1I$ and $0.2I$ respectively, where I is the identity matrix. Similar to the previous experiment, the



(a) (b)

Fig. 3: Time and error comparisons of *iFME* and *nFM*

performances of *iFME* and *nFM* are evaluated with respect to the time consumption and accuracy, and are illustrated in Fig. 3a and Fig. 3b respectively.

Though *iFME*'s accuracy is still very similar to that of *nFM*, it estimates the intrinsic mean significantly faster. Now, we compare the performance of *iFME* with *nFM* and *eFME* (the extrinsic FM estimator). The *eFME* is defined as $\Pi(\sum_i X_i)$, where Π is the projection operator from $\mathbb{R}^{k+1}$ to $\mathbb{S}^k$, i.e., $\Pi(\mathbf{x}) = \mathbf{x}/|\mathbf{x}|$. We randomly generated 50 samples on northern hemi-sphere with varying data variance. From Fig. 4, it is evident that for high data variance, *eFME* exhibits high computationally efficiency but a rather poor accuracy in its estimate of the FM.

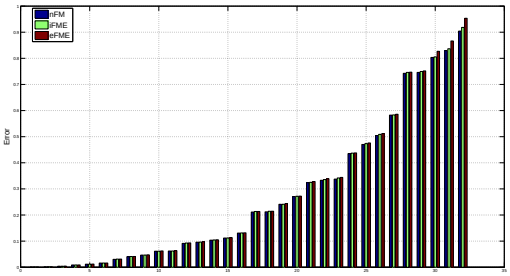


Fig. 4: Comparison between *iFME*, *nFM*, *eFME*

5 Application to the classification of movement disorders

In this section we first present a novel incremental version of the PGA algorithm in [25] applicable to data lying on a sphere. We will call this the *iPGA* algorithm. Then, we present an application of *iPGA* to real data sets. The (batch-mode) PGA proposed in [25] for diffusion tensor fields consists of (1) computing the FM of the input data, (2) projecting each data point to the tangent space at the FM using the Riemannian log-map, (3) performing standard PCA in the tangent plane and (4) projecting the result (principal vectors) back to the manifold, using the Riemannian exp-map.

An incremental form of this PGA technique was proposed recently in [18]. However, their technique was limited to manifolds with non-positive sectional curvatures. Equipped with the *iFME* on the sphere presented in the previous section, we can now extend the *iPGA* technique in [18] to the case when data lie on a hypersphere. For this, we need to use *iFME* for the FM computation and use the parallel transport operation on the hypersphere. The parallel transport operation on the hypersphere can be expressed in a closed form expression. The formula for parallel transporting $\mathbf{p} \in T_{\mathbf{n}}\mathbb{S}^k$ from $\mathbf{n}$ to $\mathbf{m}$ is given by

$$\begin{aligned}\mathbf{q} &= \Gamma_{\mathbf{n}\to\mathbf{m}}(\mathbf{p}) \\ &= \left(\mathbf{p} - \mathbf{v}\left(\frac{\mathbf{v}^t\mathbf{p}}{||\mathbf{v}||^2}\right)\right) + \frac{\mathbf{v}^t\mathbf{p}}{||\mathbf{v}||^2}\left(\mathbf{n}\left(-\sin(||\mathbf{v}||)||\mathbf{v}||\right) + \mathbf{v}\cos(||\mathbf{v}||)\right)\end{aligned}$$

where, $\mathbf{v} = Log_{\mathbf{n}}\mathbf{m}$. We now present the *iPGA* method in an algorithm form summarized in Table 1 and refer the reader for details to [18].

The dataset for classification contains HARDI scans from (1) healthy controls, and patients with, (2) Parkinson's disease (PD), and (3) essential tremor (ET). We aim to automatically discriminate between these three classes using features derived from the HARDI data. This dataset consists of 25 controls, 24 PD, and 15 ET images. The

HARDI data were acquired using a 3T Phillips MR scanner with the following parameters: $TR = 7748\,ms$, $TE = 86\,ms$, $b-$values: 0, 1000 $\frac{s}{mm^2}$, 64 gradient directions and voxel size = $2 \times 2 \times 2\,mm^3$.

Authors in [24] employed DTI based analysis, using scalar-valued features to address the problem of movement disorder classification. Later in [18], a PGA-based classification algorithm was proposed, using Cauchy deformation tensors (computed from a non-rigid registration of patient scans to a HARDI atlas) which are SPD matrices. In the next subsection, we develop classification method based on (1) Ensemble Average Propagators (EAPs) derived from HARDI data within an ROI, and (2) shapes of the ROI over the input population. Using a square root density parameterization [22], both features can be mapped to points on an unit Hilbert sphere, where the proposed *iFME* in conjunction with the *iPGA* method is applicable.

1: **Input** the data matrix $\mathbb{A}_k = [\mathbf{v}_1, ..., \mathbf{v}_k]$ for k samples
the new sample $\mathbf{x}_{k+1}$, and the old mean $\mathbf{m}_k$
2: Compute $\mathbf{m}_{k+1}$ from $\mathbf{x}_{k+1}$ and $\mathbf{m}_k$, using Eq. 2
3: $\mathbf{y}_{k+1} = Log_{\mathbf{m}_{k+1}}(\mathbf{x}_{k+1})$
4: Parallel Transport $\mathbf{z}_{k+1} = \Gamma_{\mathbf{m}_{k+1} \rightarrow \mathbf{n}}(\mathbf{y}_{k+1})$
5: Compute $\mathbf{r}_{k+1} = Log_{\mathbf{m}_{k+1}}(\mathbf{m}_k)$ and $\mathbf{t}_{k+1} = \Gamma_{\mathbf{m}_{k+1} \rightarrow \mathbf{n}}(\mathbf{r}_{k+1})$
6: Add $\mathbf{t}_{k+1}$ to every column of $\mathbb{A}_k$ to obtain $\hat{\mathbb{A}}_k = [\hat{\mathbf{v}}_1, ..., \hat{\mathbf{v}}_\mathbf{k}]$
7: Perform standard PCA on $\mathbb{A}_{k+1} = \left[\, \hat{\mathbb{A}}_k \vdots \mathbf{z}_{k+1} \,\right]$
8: Parallel transport the j^{th} principal component, $\mathbf{p}_j$,
back to $T_{\mathbf{m}_{k+1}}\mathbb{S}^\mathbf{k}$, via $\mathbf{q}_j = \Gamma_{\mathbf{n} \rightarrow \mathbf{m_{k+1}}}(\mathbf{p}_j)$

Table 1: The Incremental PGA Algorithm on a Unit Hypersphere

Classification Results using the Ensemble Average Propagator as Features: To capture the full diffusional information, we chose to use the ensemble average propagator (EAP) at each voxel as our feature in the classification. We compute the EAPs using the method described in [12] and use the square root density parameterization of each EAP. This way the full diffusion information at each voxel is represented as a point on the unit Hilbert sphere.

We now present the classification algorithm which is a combination of *iPGA*-based reduced representation and the nearest neighbor classifier. The input to the *iPGA* algorithm is EAP features in this case. The input HARDI data are first rigidly aligned to the atlas computed from the normal group, then a 3-D box surrounding the ROI, i.e., the midbrain, is placed on each image, and the EAPs within this box are computed. Finally, the EAP field extracted from each ROI image is identified with a point on the product manifold (the number of elements in the product is equal to the number of voxels in the ROI) of unit Hilbert spheres. This is in spirit similar to the case of the product manifold formalism in [25,18].

A set of 10 Control, 10 PD and 5 ET images are randomly picked as the test set, and the rest of the images are used for training. Also, classification is performed using *iPGA*, PGA and the standard PCA, and is repeated 300 times to report the average accuracy. The results using EAP features are summarized in Table 2. It is evident that the accuracy of *iPGA* is roughly the same as that of the non-incremental PGA, while both methods are considerably more accurate than the standard PCA, as they account for the non-linear geometry of the sphere. Further, the savings in computation time for *iPGA* are significant in comparison to PGA as evident from the table.

	Results using Shape Features									Results using EAP Features								
	Control vs. PD			Control vs. ET			PD vs. ET			Control vs. PD			Control vs. ET			PD vs. ET		
	iPGA	PGA	PCA	iPGA	PGA	PCA	iPGA	PGA	PCA	iPGA	PGA	PCA	iPGA	PGA	PCA	iPGA	PGA	PCA
Accuracy	91.5	93.0	67.3	88.3	90.1	75.7	86.1	87.6	64.6	92.7	93.5	59.8	90.1	91.3	70.2	89.7	90.9	66.0
Sensitivity	88.0	91.0	52.0	84.4	86.2	80.1	80.5	82.4	58.4	90.7	91.8	48.3	87.5	89.7	79.8	84.0	84.7	56.3
Specificity	95.0	95.0	82.7	92.2	94.1	71.3	91.7	92.8	70.8	94.7	95.2	71.3	92.7	92.9	60.6	95.5	97.1	75.7
Time (s)	4.1	18.5		4.0	14.2		3.5	14.7		11.6	30.9		9.8	27.3		10.8	28.0	

Table 2: Classification results from *iPGA*, PGA and PCA respectively.

Classification Results using the Shapes as Features: In this section, we evaluated the *iPGA* algorithm based on shape of the Substantia Nigra region in the brain images, for the task of movement disorder classification. We first collected random samples (point) on the boundary of each 3-D shape, and applied the Schrodinger distance transform (SDT) technique in [6] to represent each shape as a point on the unit hyper-sphere. The size of the ROI for the 3-D shape of interest was set to $28 \times 28 \times 15$, the resulting samples lie on a $\mathbb{S}^{11759}$ manifold. Then, we used *iPGA* for classification. The results given in Table 2 show significant time gains for *iPGA* over PGA but with similar accuracy.

6 Conclusion

In this paper, we presented a novel incremental Fréchet mean estimator (*iFME*), for data lying on a hypersphere. We proved the asymptotic convergence of *iFME* to the true FM. Significant time efficiency of *iFME* compared to *nFM* was shown via synthetic and real data experiments. Further, we also presented an incremental PGA (*iPGA*) algorithm that entailed the use of *iFME*. We used the *iPGA* in conjunction with a nearest neighbor classifier to classify movement disorder patients using diffusion MR brain scans. Our classification demonstrated significant gains in computation time compared to batch mode *PGA* (in conjunction with the nearest neighbor classifier), and as expected achieved the same accuracy as batch-mode PGA. In our future work, we will focus on providing an upper bound on the distance between iFME and the FM for finite set of samples.

Appendix

In this appendix, we show that $\frac{\sin(\frac{n\alpha}{n+1})}{\sin(\frac{\alpha}{n+1})} \geq n\cos^2(\frac{\alpha}{2})$ for any $\alpha \in (0, \pi)$.

Proof. Let, $f = sin(n\theta) - ncos^2(\frac{n+1}{2}\theta)sin(\theta), \theta \in (0, \alpha/(n+1)), \alpha \in (0, \pi), n \geq 1$. and $f_\theta = n\,cos(n\theta) + 2n\,cos(\frac{n+1}{2}\theta)\,sin(\theta)\,sin(\frac{n+1}{2}\theta)\,(\frac{n+1}{2}) - n\,cos^2(\frac{n+1}{2}\theta)\,cos(\theta)$ Solving this eqn., as $\theta \in (0, \pi/(n+1))$, we get, $\theta = 0$. But, $f_{\theta\theta}|_{\theta=0} = 0$. Hence, we check $f_{\theta\theta\theta}$.

$f_{\theta\theta\theta}|_{\theta=0} = -n^3 + 1.5n\,(n+1)^2 + n \quad > 0, \;\; n \geq 1$ So, at $\theta = 0$, f has a minimum where, $\theta \in (0, \alpha/(n+1))$. $f|_{\theta=0} = 0$ Thus, $f \geq 0$ as $n \geq 1$. As for $\theta \in (0, \alpha/(n+1))$, $sin(\theta) > 0$, $\frac{f}{sin(\theta)} \geq 0$ $\frac{f}{sin(\theta)} = \frac{sin(n\theta)}{sin(\theta)} - ncos^2(\frac{n+1}{2}\theta)$ Hence, $\frac{sin(n\theta)}{sin(\theta)} - ncos^2(\frac{n+1}{2}\theta) \geq 0$ Then, by substituting $\theta = \alpha/(n+1)$, we get $\frac{\sin(\frac{n\alpha}{n+1})}{\sin(\frac{\alpha}{n+1})} \geq n\cos^2(\frac{\alpha}{2})$

■

References

1. B. Afsari, R. Tron, , et al. On the convergence of gradient descent for finding the riemannian center of mass. *SIAM J. on Control and Optimization*, 51(3):2230–2260, 2013.
2. G. Amarasinghe. On the standard lengths of angle bisectors and the angle bisector theorem. *Global Journal of Advanced Research on Classical and Modern Geometries*, 1(1), 2012.
3. M. Arnaudon, C. Dombry, A. Phan, and L. Yang. Stochastic algorithms for computing means of probability measures. *Stochastic Processes and their Applications*, 122(4):1437–1455, 2012.
4. H. E. Cetingul, B. Afsari, et al. Group action induced averaging for HARDI processing. In *ISBI*, pages 1389–1392, 2012.
5. J. Cheng, A. Ghosh, et al. A riemannian framework for orientation distribution function computing. In *MICCAI*, pages 911–918. Springer, 2009.
6. Y. Deng, A. Rangarajan, et al. A riemannian framework for matching point clouds represented by the schrodinger distance transform. In *CVPR*, 2014.
7. A. Feragen, S. Hauberg, M. Nielsen, and F. Lauze. Means in spaces of tree-like shapes. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 736–746. IEEE, 2011.
8. A. Goh, C. Lenglet, et al. A nonparametric riemannian framework for processing high angular resolution diffusion images and its applications to odf-based morphometry. *NeuroImage*, pages 1181–1201, 2011.
9. R. Hartley, J. Trumpf, et al. Rotation averaging. *IJCV*, 103(3):267–305, 2013.
10. J. Ho, G. Cheng, et al. Recursive karcher expectation estimators and geometric law of large numbers. In *AISTATS*, pages 325–332, 2013.
11. B. Iversen. *Hyperbolic geometry*, volume 25. Cambridge University Press, 1992.
12. B. Jian and B. C. Vemuri. A unified computational framework for deconvolution to reconstruct multiple fibers from DWMRI. *IEEE TMI*, 26:1464–1471, 2007.
13. W. S. Kendall. Probability, convexity, and harmonic maps with small image i: uniqueness and fine existence. *Proceedings of the London Mathematical Society*, 3(2):371–406, 1990.
14. Y. Lim and M. Pálfia. Weighted inductive means. *LAA*, 453:59–83, 2014.
15. K. V. Mardia and P. E. Jupp. *Directional statistics*, volume 494. John Wiley & Sons, 2009.
16. X. Pennec. Probabilities and statistics on riemannian manifolds: Basic tools for geometric measurements. In *NSIP*, pages 194–198, 1999.
17. C. R. Rao. Differential metrics in probability spaces. *Differential geometry in statistical inference*, 10:217–240, 1987.
18. H. Salehian, D. Vaillancourt, et al. iPGA: Incremental principal geodesic analysis with applications to movement disorder classification. In *MICCAI*, pages 765–772. 2014.
19. N. J. Sloane et al. The on-line encyclopedia of integer sequences, 2003.
20. A. Srivastava, I. Jermyn, and S. Joshi. Riemannian analysis of probability density functions with applications in vision. In *CVPR*, pages 1–8, 2007.
21. K. T. Sturm. Probability measures on metric spaces of nonpositive curvature. In *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces*, 2003.
22. J. Sun, Y. Xie, et al. Dictionary learning on the manifold of square root densities and application to reconstruction of diffusion propagator fields. In *IPMI*, pages 619–631, 2013.
23. D. S. Tuch, T. G. Reese, et al. Diffusion mri of complex neural architecture. *Neuron*, 40(5):885–895, 2003.
24. D. Vaillancourt, M. Spraker, et al. High-resolution diffusion tensor imaging in the substantia nigra of de novo parkinson disease. *Neurology*, pages 1378–1384, 2009.
25. Y. Xie, B. C. Vemuri, and J. Ho. Statistical analysis of tensor fields. In *MICCAI*, pages 682–689. Springer, 2010.

An Efficient Recursive Algorithm for Atlas Construction

Rudrasis Chakraborty[1], Monami Banerjee[1], Dohyung Seo[2], Sara Turner[3], David Fuller[3], John Forder[4], and Baba C. Vemuri[1] ⋆

[1] Department of CISE, University of Florida, Gainesville, Florida, USA ⋆⋆
{rudrasis, monami, vemuri}@cise.ufl.edu
[2] Mitsubishi Electric Research Laboratories, USA
dhseo.118@gmail.com
[3] Department of Physical Therapy, University of Florida, Florida, USA
{smturner, ddf}@phhp.ufl.edu
[4] Department of Radiology, University of Florida, Florida, USA
jforder@mbi.ufl.edu

Abstract. Atlas construction is a fundamental problem in Medical Image Computing. Every disease assessment task requires a template "reference" to compare to in order to assess the amount of changes in the anatomy or function. The key then is to define this "reference" in a meaningful fashion. The "reference" also commonly referred to as an atlas is normally defined as the most representative of the population of given data. Statistically, this is chosen to be an average or a weighted average of the given data set. Since the control population consists of distinct subjects, the task of estimating an unbiased atlas is posed as a groupwise non-rigid diffeomorphic registration of the given data to this unknown average defined as the minimizer of the sum of squared geodesic distances cost function. This is a hard joint minimization over the space of diffeomorphisms and atlases, and is computationally very expensive. In this paper, we present an efficient alternative which involves arbitrarily choosing one of the given data sets as a reference and estimating the diffeomorphisms from the given pool of data to this reference. Then, efficiently estimating the Fréchet mean (FM) of these diffeomorphisms and applying this FM-diffeomorphism to the chosen reference yields the desired unbiased atlas. We prove that the atlas obtained in this manner is the same as the one obtained using the conventional groupwise registration approach mentioned above. The key advantage of our approach over conventional groupwise registration approach is that we do not require any optimization over the space of atlases, thereby reducing computational cost dramatically. Further, our approach is a recursive approach and thus is amenable to updates when the data pool is augmented with new data without the need to compute the atlas from scratch. We present several real data experiments demonstrating the computational advantages of our proposed approach over state-of-the-art.

⋆ This research was in part supported by the NIH grant NS066340 to BCV.
⋆⋆ This research was supported in part by the National Institutes of Health under Grant NS066340 to BCV

1 Introduction

An atlas is an informative representative of a population of "objects" (images in our context). Constructing an atlas is a key ingredient to many applications including but not limited to image alignment [22], image segmentation [23,27] and statistical analysis [2,10]. Hence, over the past decade, algorithms for atlas construction have attracted substantial attention in Medical Image Analysis research. In particular, an unbiased diffeomorphic atlas construction algorithm was first proposed in [13]. This algorithm spawned a flurry of activity in the area of atlas construction resulting in many variants being proposed in the recent past. Some of the recent methods for constructing multiple atlases on heterogeneous data are [19,26]. A popular approach for atlas construction is to compute the arithmetic mean of group-wise registered images. But, these algorithms [13,16,19] often generate a blurred atlas. Recently Xie et al. [26] proposed a multiple atlas construction framework which yields "sharp" atlases. Other methods for generating "sharp" atlases also exist in literature and we refer the reader to [4,25]

Since most of the popular atlas construction (sequential) algorithms (implemented on standard multi-core desktops) are computationally intensive (typically taking tens of hours to days of CPU time), there is a dire need to develop a time efficient atlas construction algorithm. In this work, we propose a computationally efficient atlas construction algorithm. We further assume that the images in the given data pool can be diffeomorphically registered to each other and to atlas being sought. This is not an uncommon assumption and was made in Younnes [28] for tackling the diffeomorphic registration problem. Now, instead of doing averaging on the image space, we compute the average over the space of diffeomorphisms. By using methods in [15], we can map a subset of "interesting" diffeomorphisms to the Hilbert sphere. We propose an efficient mean estimator on the hypersphere and consistency of this estimator is shown in an accompanying manuscript published in this workshop[20]. Using this fast estimator, we obtain an efficient way to compute the atlas of an image population. In the experiments section, we have shown the significant time gain of our proposed atlas construction algorithm over state-of-the-art. Now, we present a brief survey of existing atlas construction algorithms.

A popular way to construct an atlas is based on an unbiased group-wise registration method. This type of formulation [13] requires the solution to a hard non-convex optimization problem involving two high dimensional unknowns namely, the atlas and the non-rigid transformations between the unknown atlas and input image data. It is solved using an alternating strategy, involving, fixing the average and estimating the transformations required to transform all the input data sets to this average, then fixing the transformations and estimating the average. This is a hard optimization problem and local solutions can be unsatisfactory at times. Another, popular atlas construction algorithm is $iCluster$ [19], which computes the atlas by fitting Gaussian mixture model to the input images. Then, they use an Expectation-Maximization (EM) algorithm to construct the atlas. In this algorithm, they also compute the arithmetic mean of the groupwise registered images. Both of these algorithms [13,19] lead to a blurred atlas which is the result of arithmetic mean of images. Note that, the arithmetic mean of the images need not necessarily lie on the underlying manifold (from which the images were sampled), hence if the images are not tightly clustered on the manifold, arithmetic mean is a *poor*

choice. Hence, in Xie et al. [26] they proposed a multiple atlas construction method which produces a "sharp" atlas. In this method, they use a graph representation of the underlying manifold. Then, they used a graph partitioning followed by computing the FM of images belonging to each cluster. Though this algorithm produced a "sharp" atlas, like other atlas construction algorithm it is computationally expensive. For other recent image atlas construction methods we refer the reader to [8,9,11,3,24,5] and for shape atlases to [6] and the references therein.

The rest of the paper is organized as follows. In section 2, we describe our proposed method in detail. Several real data experimental results along with comparisons are presented in section 3. Finally, we conclude in section 4.

2 Methodology

In this section, we present our proposed method of incremental atlas construction from multiple images. *We will use the term incremental, recursive and inductive interchangeably throughout the paper*. We would like to emphasize that our proposed method is applicable to scalar, vector and tensor-valued fields. However, for simplicity, here we present a formulation for the case of scalar-valued fields, i.e., intensity images. Let $\mathcal{I}$ denote the space of images, we define images in $\mathcal{I}$ as $\mathbf{L}^2$ functions on a image domain $\sigma \subset \mathbf{R}^d$. Given a set of images $\mathcal{C} = \{I_1, \cdots, I_n\} \subset \mathcal{I}$, our goal is to construct an atlas from these n images. We will assume that the set of images are rigidly registered by randomly choosing a reference image and rigidly registering the rest to this reference image. Now, given the set of images in the same coordinate system, an atlas can be defined as the minimizer of the sum of squared (geodesic) distances from it to the rest of the given images in the data pool. Let, $d_{\mathcal{I}}(I_i, I_j)$ denote the "distance" between two images I_i and I_j. Note that here we use "distance" loosely without having defined the underlying metric or even the underlying manifold. In the spirit of Xie et al. [26], we will assume that the images are samples from an unknown manifold. Now, let I^* be the atlas of the n input images, then $I^* = \arg\min_{\mu} \sum_{i=1}^{n} d^2_{\mathcal{I}}(\mu, I_i)$. Note that this is primarily the Fréchet mean (FM) [7] of the n images. Though this formulation is simple, it is computationally expensive as the space of images $\mathcal{I}$ is huge, and the minimization of the sum of squared distances formulation involves searching over the entire space $\mathcal{I}$. This provides us sufficient motivation to seek an alternative time-efficient algorithm to compute the atlas. Below, we present a simple yet illustrative example, that captures the essence of our proposed approach to atlas construction.

Motivating Example: Given the following real numbers -1, -5, 0, 2, 8 and 8, the Fréchet mean (FM) (arithmetic mean in this simple case) is 2. But instead of averaging on the numbers, we can instead do the following. We randomly choose one of the numbers as the reference number. Then compute the FM (arithmetic mean in this case) of the differences between each of the numbers and the reference. For example, let us choose -5 as the reference number, then the differences between -5 and each of the numbers are 4, 0, 5, 7, 13 and 13 respectively. The average of these six numbers is 7. Then, if we add this mean-difference to our chosen reference, i.e., -5 we get the mean of the numbers, i.e., 2.

At the outset, this approach seems more complicated than simply using the standard arithmetic mean of six numbers. But, if instead of an input of real numbers, our input is a set of images, then developing a time-efficient algorithm for computing the FM of the "differences", will make this second approach more time-efficient. But, two important questions that surface naturally are, first, like in our toy example above, will the second approach still yield the true FM of images? Second, what is the "difference" term in context of images? In the context of image registration and atlas construction, it is natural to interpret this "difference" as the transformation required to register the two images. *In this work, we will assume that the images in the given data pool whose atlas is to be constructed can all be diffeomorphically registered with each other and with the atlas to be constructed.* The answer to the first question is given by the theorem given below. *Using the second approach, the key advantage that we gain in atlas construction over conventional approaches such as [13] and variants thereof is that the hard joint optimization over the space of diffeomorphisms and atlases is now transformed to one over the space of diffeomorphisms. This leads to an enormous savings in computation time as evidenced through the experiments in section 3*. We now state and prove the aforementioned theorem.

Let σ be an open subset of $\mathbf{R}^d$ and G a group of diffeomorphisms on σ. Consider a set of images $\mathcal{J} \in \mathcal{I}$ on which G has an action, i.e., for every $I \in \mathcal{J}$ and every $\phi \in G$, the result of the action of ϕ on $\mathcal{J}$ is denoted by $\phi . I \in \mathcal{J}$, where . is the group operator. Let $\mathcal{C} \subset \mathcal{J}$ be a set of n images, i.e., $\mathcal{C} = \{I_1, \cdots, I_n\}$. Let $I_{ref} \in \mathcal{C}$ be an arbitrarily chosen reference image. Let, T_i be the diffeomorphism from I_{ref} to I_i, i.e., $I_{ref}(x) = I_i(T_i(x))$, $\forall i$. Further, given an arbitrary image $\mu \in \mathcal{J}$, let T_μ be the diffeomorphism from I_{ref} to μ. Let $d_{\mathcal{I}}$ and $d_{\mathcal{T}}$ be the geodesic distance functions on the space of images and transformations respectively. Let us define the relation $\sim$ between two objective functions f_1 and f_2 iff $\exists$ a bijection between $\mathfrak{F}(f_1)$ and $\mathfrak{F}(f_2)$, where $\mathfrak{F}(.)$ is the set of solutions of the corresponding objective function. It is easy to show that $\sim$ is an equivalence relation. Then we have,

Theorem 1.

$$arg\min_{\mu} \sum_{i=1}^{n} d^2_{\mathcal{I}}(\mu, I_i) \sim \arg\min_{T_\mu} \sum_{i=1}^{n} d^2_{\mathcal{T}}(T_i, T_\mu) \tag{1}$$

Proof.

$$\begin{aligned}
\arg\min_{\mu} \sum_{i=1}^{n} d^2_{\mathcal{I}}(\mu, I_i) &= \arg\min_{\mu} \sum_{i=1}^{n} d^2_{\mathcal{I}}(I_i(T_i \circ T_\mu^{-1}(x), I_i(x)) \\
&\sim \arg\min_{T_\mu} \sum_{i=1}^{n} d^2_{\mathcal{I}}(I_i(T_i \circ T_\mu^{-1}(x), I_i(x)) \\
&= \arg\min_{T_\mu} \sum_{i=1}^{n} d^2_{\mathcal{I}}(I_i(T_i \circ T_\mu^{-1}(x), I_i(T_\mu \circ T_\mu^{-1}(x)) \\
&\sim \arg\min_{T_\mu} \sum_{i=1}^{n} d^2_{\mathcal{T}}(T_i, T_\mu)
\end{aligned}$$

The above equalities hold based on the following two claims.

Claim 1: $\arg\min_{\mu} \sum_{i=1}^{n} d^2_{\mathcal{I}}(I_i(T_i \circ T_\mu^{-1}(x), I_i(x)) \sim \arg\min_{T_\mu} \sum_{i=1}^{n} d^2_{\mathcal{I}}(I_i(T_i \circ T_\mu^{-1}(x), I_i(x))$

Proof. Let $S_1 = \arg\min_{\mu} \sum_{i=1}^{n} d^2_{\mathcal{I}}(I_i(T_i \circ T_\mu^{-1}(x), I_i(x))$ and $S_2 = \arg\min_{T_\mu} \sum_{i=1}^{n} d^2_{\mathcal{I}}(I_i(T_i \circ T_\mu^{-1}(x), I_i(x))$. We have to prove that $\exists$ a bijection between S_1 and S_2. Let us first prove the cardinality of S_1 and S_2 are same. Let $I \in S_1$, then, it is easy to see that $T_I \in S_2$, which proves $S_1 \subset S_2$. The other way of double containment is similar to prove. Hence, the cardinalities are same. And given $I \in \mathcal{I}$, $\exists!$ a T_I such that $I(x) = I_{ref}((T_I)^{-1}(x))$. And for a given T_I, the choice of I is also unique. Hence, $\exists$ a bijection between S_1 and S_2. This proves the claim. ■

Claim 2: $\arg\min_{T_\mu} \sum_{i=1}^{n} d^2_{\mathcal{I}}(I_i(T_i \circ T_\mu^{-1}(x), I_i(T_\mu \circ T_\mu^{-1}(x)) \sim \arg\min_{T_\mu} \sum_{i=1}^{n} d^2_{\mathcal{T}}(T_i, T_\mu)$

Proof. It's easy to see that $\arg\min_{T_\mu} \sum_{i=1}^{n} d^2_{\mathcal{I}}(I_i(T_i(y), I_i(T_\mu(y)) \sim \arg\min_{T_\mu} \sum_{i=1}^{n} d^2_{\mathcal{T}}(T_i, T_\mu)$. And by transitivity of the relation, $\sim$, our claim holds. ■

Now, from the proofs of the two claims, the proof of the theorem follows. ■

So by Theorem 1, in order to compute the FM of the given images, we first compute the FM of the diffeomorphisms (between the images and an arbitrarily chosen reference I_{ref}) and apply this mean diffeomorphism on the arbitrarily chosen reference. Hence, if T^* is the FM of the diffeomorphisms and I^* is the atlas, $I^*(x) = I_{ref}((T^*)^{-1}(x))$.

Note that the above hypothesis that, $\exists$ a diffeomorphism between any two images in $\mathcal{I}$ simply means that members of $\mathcal{I}$ are of same topology and diffeomorphically related. In practise, for the atlas construction problem, the given image data pool from which the atlas is being constructed can be assumed to have the same structures of interest, since, it is meaningful to construct atlas from a population of say,"normal" human brains but it is not meaningful to construct an atlas from a population consisting of "normal" human brains and brains with pathology as they maybe of a different topology.

Corollary 1. *Given the hypothesis as above, let a set of n images $\mathcal{C} = \{I_1, \cdots, I_n\}$, then, eqn. 1 in Thm. 1 holds for any transformations $GL(m)$ (with appropriate m), where $GL(m)$ denotes the general linear group consisting of $m \times m$ invertible matrices.*

Proof. The proof follows from Thm. 1. ■

Corollary 2. *Given the hypothesis as above, let a set of n images $\mathcal{C} = \{I_1, \cdots, I_n\}$, then, eqn. 1 in Thm. 1 holds for all local affine transformations.*

Proof. For local affine transformations, the transformation N_i for each image is a product of $GL(m)$ matrices (with appropriate m). Hence, the proof follows from Cor. 1. ■

Now, our next concern is how to *efficiently* compute FM on the space of diffeomorphisms. Let M be the image domain and let $d\mu$ be an associated Riemannian volume form on M. Let $Diff(M)$ and $Diff_\mu(M)$ denote the infinite dimensional group of

diffeomorphisms on M and its infinite dimensional subgroup of volume preserving diffeomorphisms on M. Khesin et al. in [15] showed that the right invariant $\dot{H}^1$ metric (see [15] for definition of this metric) on $Diff(M)$ descends to a non-degenerate Riemannian metric on the homogeneous space of densities on M, $Dens(M) = Diff(M)/Diff_\mu(M)$. Further, they proved that equipped with the $\dot{H}^1$ metric, the space $Dens(M)$ is isometric to a subset of an infinite dimensional sphere in the Hilbert space. We use this result in our work here and compute the FM of this class of diffeomorphisms (points in $Dens(M)$) identified with those on the Hilbert sphere.

Now, we can use the square root of the density parameterization to map a point in $Dens(M)$ to the infinite dimensional unit Hilbert Sphere. But, note that our goal was to compute FM on the space of diffeomorphisms, i.e., $Diff(M)$. Hence, we need to justify why working on the quotient space $Diff(M)/Diff_\mu(M)$ instead of $Diff(M)$ is an acceptable choice. The atlas construction problem is normally posed as follows: Given a population of images acquired from distinct subjects, the goal is to construct a representative image or an atlas. It is reasonable to assume that in constructing the atlas from distinct subject scans, volume preserving diffeomorphisms are highly unlikely and ought to be treated as nuisance transformations. Hence, they ought to be quotiented out. Moreover, by quotienting out the volume preserving diffeomorphisms (volumorphisms), the computed atlas becomes invariant to any volumorphisms i.e., rigid transformations etc. This is an additional advantage of our proposed atlas construction scheme. Now, given that we have mapped the points from $Dens(M)$ to the unit Hilbert Sphere, we propose an efficient scheme to compute FM on the unit Hilbert sphere.

A common approach to computing the FM of a finite sample set of points on a Riemannian manifold is to find the global optimum (if it exists) of the sum of squared geodesic distances cost function. A popular approach to solve this problem involves the use of the gradient descent method. An alternative way to compute the FM is to develop a recursive/inductive definition that does not involve optimizing the aforementioned cost function. Where applicable, a recursive algorithm can take advantage of the closed form solution to compute the FM of two points as the base of the recursion and recurse through the number of points in the given set. This will yield a much faster way to compute the FM if and when the convergence of the algorithm can be proved.

Note that in Euclidean space, the recursive form of computing the arithmetic mean (which yields the same solution as the minimization of sum of squared distances) involves only two points in each recursion step and can be geometrically interpreted as moving an appropriate distance away from the already computed mean (old mean) towards the new-mean on the straight line joining the old mean and the new data point. This geometric procedure can be readily extended to any Riemannian manifold using geodesics. To this end, we make use of the closed form expression – derived using the sphere metric on the hypersphere – for the geodesic between two points on the hypersphere. More precisely, after computing the estimate of the FM of k points, denoted by M_k, the $k+1^{th}$ estimate lies on the geodesic between M_k and the $k+1^{th}$ point S_{k+1}. This readily yields an algorithm for computing the FM that does not require any function optimization, a considerable advantage often realized as gains in computation time of several orders in magnitude over non-incremental algorithms based on minimization of sum of squared geodesic distances. *Since, the FM on a sphere is unique only when all*

the data lie with in an injectivity radius of $\pi/2$ [18,14,1], we will make this assumption through the rest of this paper.

Let $\{S_1, \cdots, S_n\} \subset S^\infty$, then we define the inductive (recursive) estimator of the FM by the recursion in Eqn. 2. A proof of convergence of this algorithm is presented in [20].

Here, $\Gamma\Big(X, Y, .\Big)$ denotes the geodesic between X and Y. The geodesic $\Gamma(X, Y, t)$ on S^∞ is defined as follows:

$$\Gamma(X, Y, t) = Exp_X(t\ Exp_X^{-1}(Y)) \quad (4)$$

$$M_1 = S_1 \quad (2)$$

$$M_{k+1} = \Gamma\Big(M_k, S_{k+1}, \omega_{k+1}\Big) \quad (3)$$

where, Exp and Exp^{-1} are the Riemannian exponential and inverse exponential mapping as defined below.

- **Exponential Map**: Given a vector $\mathbf{v} \in T_X S^\infty$, the Riemannian Exponential map on S^N is defined as $Exp_X(\mathbf{v}) = \cos(|\mathbf{v}|)X + \sin(|\mathbf{v}|)\mathbf{v}/|\mathbf{v}|$. The Exponential map gives the point which is located on the great circle along the direction defined by the tangent vector $\mathbf{v}$ at a distance $|\mathbf{v}|$ from X.M
- **Inverse Exponential Map**: The tangent vector $\mathbf{v} \in T_X S^\infty$ directed from X to Y is given by, $Log_X(Y) = \frac{\theta}{\sin(\theta)}(Y - X\cos(\theta))$ where, $\theta = \arccos(X^T Y)$.

After computing the FM on hypersphere, we lift it back to the space of diffeomorphisms, $Diff(M)$ using the formulation proposed in [21] to get the mean diffeomorphism. Note that this mean diffeomorphism is unique upto volume preserving transformations. And finally we apply this mean diffeomorphism to the reference image to get the atlas image. As claimed above, this formulation is easily generalized to tensor field data. We call this atlas construction procedure as the *incremental atlas construction algorithm*, $iAcA$. We summarize the steps of our algorithm in the following:

Algorithm 1 Algorithm for *incremental atlas construction*

1: *Input*: a population of n images $\{I_1, \cdots, I_n\}$.
2: *Output*: an atlas image, I^* of the population.
3: Step 1. Arbitrarily choose any one of the given images as the reference, denoted by I_{ref} .
4: Step 2. Compute the diffeomorphisms $\{T_i\}$ from I_{ref} to I_i, where T_i is the diffeomorphism to I_i.
5: Step 3. Map each of these diffeomorphisms, T_i, to the hypersphere (of appropriate dimension) using the scheme proposed in [21,15]. Let the points on the hypersphere be $\{S_i\}$.
6: Step 4. Compute the inductive FM, S^*, of $\{S_i\}$ using Eqn. 2.
7: Step 5. Map S^* onto $Diff(M)$ using the method in [21]. Let the FM diffeomorphism be T^*.
8: Step 6. Apply T^* on I_{ref} to get the atlas image I^*.

We now list a few advantages of $iAcA$.

- The key advantage of $iAcA$ over popular atlas construction methods such as the one in [13] and variants thereof perform a hard multi-variate optimization involving

joint search over very large spaces of diffeomorphisms and atlases respectively. In contrast, $iAcA$ involves a search only over the space of diffeomorphisms. This leads to a much simpler and more time efficient alternative.
- The $iAcA$ is very time efficient as demonstrated via the experimental results in section 3.
- $iAcA$ yields an atlas that is invariant to volume preserving transformations.
- In many medical imaging applications, it is customary to augment the data pool as and when new scans are acquired. In such situations, it is more time efficient to update the already computed atlas rather than to compute the atlas from scratch. Due to it's recursive nature, $iAcA$ achieves this optimally.

3 Experimental Results

In this section, we present experimental results of our atlas construction algorithm, $iAcA$, and compare its performance with two other atlas construction algorithms, one for constructing atlases from fields of ensemble average propagators (EAPs) derived from diffusion MR scans acquired from rat spinal cords [5] and another for 2D shapes from the MPEG-7 database [17]. For the MPEG-7 data, we used the atlas construction algorithm in the well known *ANTS* [3] software. We report the time taken by both of these algorithms. All the computation time required for various algorithms reported in this paper, were measured on an Intel-7 quad-core processor, 16GB RAM desktop. Accuracy is hard to assess on real data sets and will be focus of our future work. *We would like to point out that though ANTS is a highly optimized toolbox written in C++, our $iAcA$ code was written in MATLAB, which is not efficient for non-matrix operations. Code optimization to achieve further time savings with $iAcA$ implementation will be the focus of our future work.*

3.1 Atlas construction from diffusion MR scans of rat spinal cords

In this section, we used HARDI data acquired from several rat spinal cords. The HARDI scans were acquired using a 3T Phillips MR scanner with the following parameters: $b-$values: 0, $1500\,\frac{s}{mm^2}$, 22 gradient directions and voxel size = $2 \times 2 \times 2\,mm^3$. We have constructed the EAP field atlas from EAP fields derived from 7 control rat data sets. The EAP fields from each HARDI data set was first estimated using the approach in [12]. Sample slices from the HARDI scans of the rat spinal cords as well as the atlases constructed by the method in [5] and $iAcA$ respectively are shown in Fig. 1. In this figure, the top row (left to right) depicts the zero gradient image, S_0 from the HARDI scans of three rat spinal cords. Second row, left to right, depicts a slice from the estimated EAP fields superposed on the corresponding slice of the S_0 images. Last row, left to right, depicts a slice from the atlas EAP fields (superimposed on the corresponding S_0 images) estimated by $iAcA$ and the approach in [5].

Class	Time (s)	
	$iAcA$	ANTS
Spinal_Cord	109800.0	302400.0
apple	2092.0	27280.0
heart	2272.9	34800.0
car	2304.6	37134.0

Table 1: Computational time for atlas construction

From a visual inspection view point, the atlas computed from $iAcA$ appears to be much sharper and better than that from the method in [5]. The time required by these

two atlas construction methods are presented in Table 1. From this table, it is evident that computationally, $iAcA$ is significantly faster compared to the approach in [5].

3.2 Atlas construction on MPEG-7 data

We randomly chose 3 classes of objects from the MPEG-7 database namely, heart, apple and car shapes. Each of these classes contains 20 two-dimensional images from which we construct an atlas for each shape class. In Fig. 2, we present 10 random images of each of these three classes and the atlases constructed by *iAcA* and *ANTS* respectively. For *ANTS*, we have used *Greedy Symmetric Normalization* for non-rigid registration. Further, the initial atlas is chosen to be the default i.e., the arithmetic mean of the population of the class. In the figure, for all these three subjects, the leftmost image in the bottom row is the atlas constructed by $iAcA$ and this is followed by the atlas constructed using ANTS. The first, second, fourth and fifth rows consist of images of 10 sample data from the respective classes. The computation time required for these two algorithms is reported in Table 1, which clearly depicts the superior time efficiency of $iAcA$ over *ANTS*. Further, from a visual inspection view point, the atlas constructed by $iAcA$ appears to be of higher quality (sharper). Our future work will focus quantitatively validating the accuracy of the constructed atlases.

4 Conclusions

In this paper, we presented a novel incremental atlas construction algorithm called *iAcA*. The key advantage of this algorithm over conventional unbiased groupwise registration based atlas construction approach (which requires a joint optimization over the space of diffeomorphisms and atlases) is that, it needs an optimization only over the space of diffeomorphisms to register $(n-1)$ pairs of data sets from the given pool of n data sets. A reference data set is arbitrarily chosen from the given pool and all the other data are diffeomorphically registered to this reference. We then compute the FM of these diffeomorphisms after quotienting out the volumorphisms. The FM computation is achieved

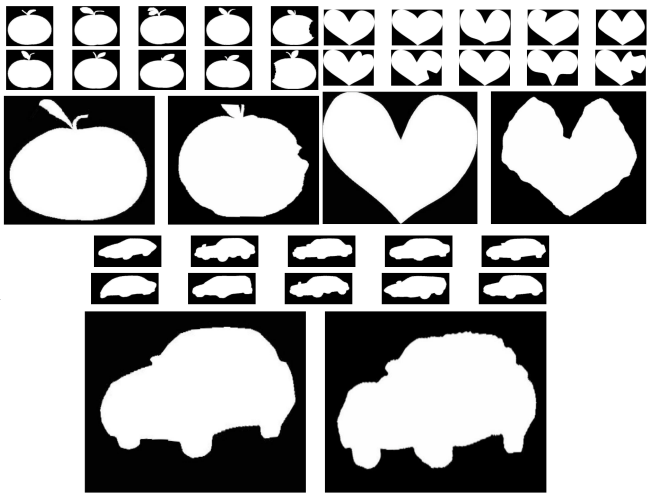


Fig. 2: Apple, heart and car shape atlas construction. Rows 1,2,4 & 5 depict samples from the data pool. Rows 3 & 6 depict the atlas obtained using $iAcA$ and *ANTS* repectively.

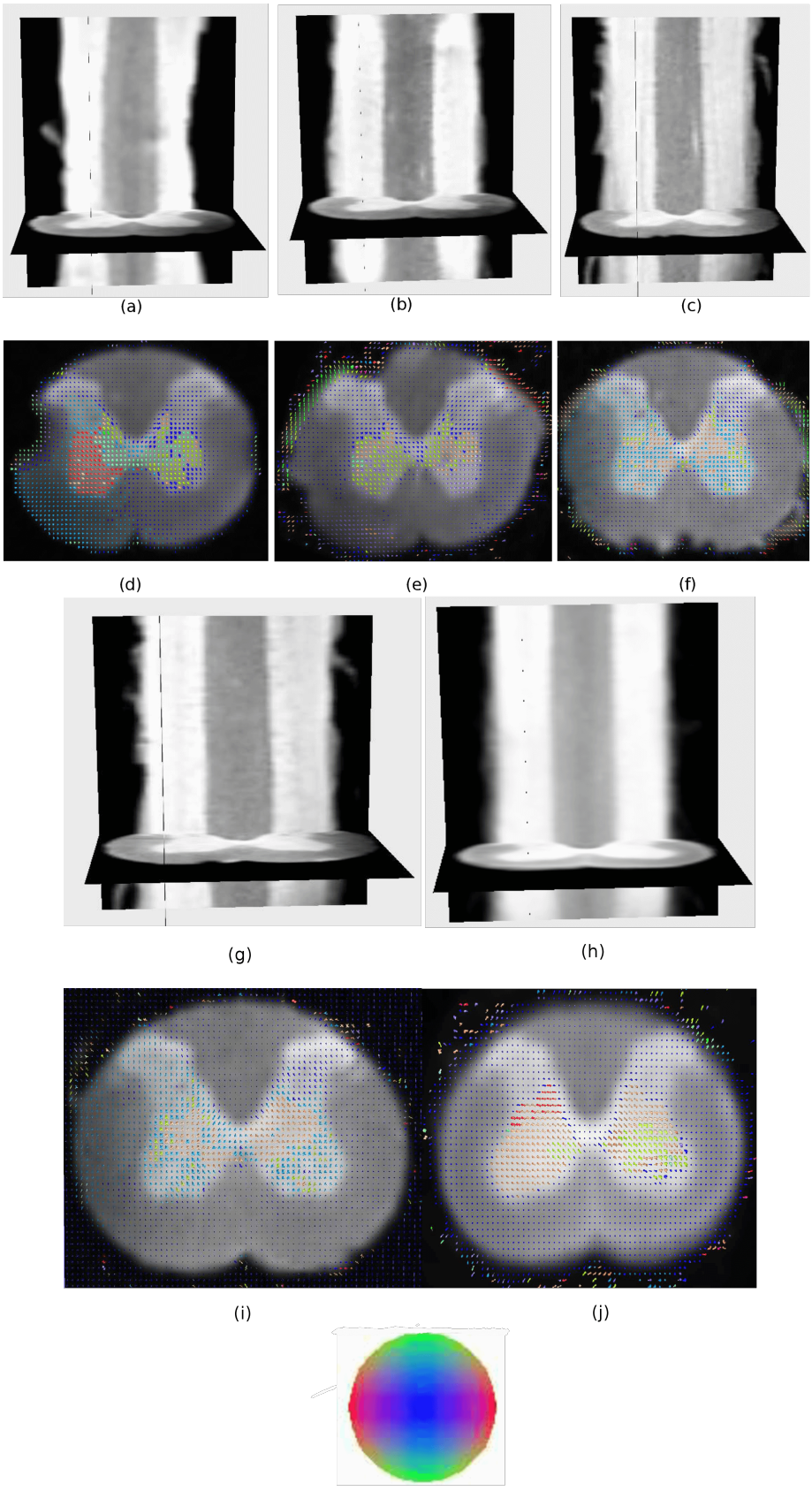


Fig. 1: Constructed Atlases from Spinal Cord 3D data. [a-c]: sample S_0 images from the population of controls, [d-f]: corresponding sample EAP fields superposed on respective S_0 images, [g,i]: S_0 and EAP atlas using $iAcA$, [h,j]: S_0 and EAP atlas using [5]. Last row shows the color ball used to color the EAPs.

recursively and does not require any optimization. This Fm is then applied to the chosen reference to obtain the desired atlas. We demonstrated dramatic savings in computational cost using our approach (over state-of-the-art) for the task of atlas construction from diffusion MR scans of rat spinal cords and MPEG-7 shape data sets. Our future efforts will focus on a thorough quantitative validation of the constructed atlases.

References

1. B. Afsari. Riemannian lp center of mass: Existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(2):655–673, 2011.
2. J. Ashburner, C. Hutton, R. Frackowiak, I. Johnsrude, C. Price, K. Friston, et al. Identifying global anatomical differences: deformation-based morphometry. *Human brain mapping*, 6(5-6):348–357, 1998.
3. B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee. A reproducible evaluation of ants similarity metric performance in brain image registration. *Neuroimage*, 54(3):2033–2044, 2011.
4. D. J. Blezek and J. V. Miller. Atlas stratification. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pages 712–719. Springer, 2006.
5. G. Cheng, B. C. Vemuri, M.-S. Hwang, D. Howland, and J. R. Forder. Atlas construction from high angular resolution diffusion imaging data represented by gaussian mixture fields. In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pages 549–552. IEEE, 2011.
6. S. Durrleman. *Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution*. PhD thesis, Université Nice Sophia Antipolis, 2010.
7. M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. In *Annales de l'institut Henri Poincaré*, volume 10, pages 215–310. Presses universitaires de France, 1948.
8. S. Gerber, T. Tasdizen, P. T. Fletcher, S. Joshi, R. Whitaker, A. D. N. Initiative, et al. Manifold modeling for brain population analysis. *Medical image analysis*, 14(5):643–653, 2010.
9. J. Hamm, D. H. Ye, R. Verma, and C. Davatzikos. Gram: A framework for geodesic registration on anatomical manifolds. *Medical image analysis*, 14(5):633–642, 2010.
10. X. Hua, A. D. Leow, N. Parikshak, S. Lee, M.-C. Chiang, A. W. Toga, C. R. Jack, M. W. Weiner, and P. M. Thompson. Tensor-based morphometry as a neuroimaging biomarker for alzheimer's disease: an mri study of 676 ad, mci, and normal subjects. *Neuroimage*, 43(3):458–469, 2008.
11. H. Jia, G. Wu, Q. Wang, and D. Shen. Absorb: Atlas building by self-organized registration and bundling. *NeuroImage*, 51(3):1057–1070, 2010.
12. B. Jian and B. C. Vemuri. A unified computational framework for deconvolution to reconstruct multiple fibers from DWMRI. *IEEE TMI*, 26:1464–1471, 2007.
13. S. Joshi, B. Davis, M. Jomier, and G. Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23:S151–S160, 2004.
14. Kendall, Wilfrid S. A survey of riemannian centres of mass for data. *Proceedings 59th ISI World Statistics Congress*, 2010.
15. B. Khesin, J. Lenells, G. Misiołek, and S. Preston. Geometry of diffeomorphism groups, complete integrability and geometric statistics. *Geometric and Functional Analysis*, 23(1):334–366, 2013.
16. P. Lorenzen, B. C. Davis, and S. Joshi. Unbiased atlas formation via large deformations metric mapping. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2005*, pages 411–418. Springer, 2005.
17. B. S. Manjunath, P. Salembier, and T. Sikora. *Introduction to MPEG-7: multimedia content description interface*, volume 1. John Wiley & Sons, 2002.

18. X. Pennec. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *JMIV*, 25(1):127–154, 2006.
19. M. R. Sabuncu, S. K. Balci, M. E. Shenton, and P. Golland. Image-driven population analysis through mixture modeling. *Medical Imaging, IEEE Transactions on*, 28(9):1473–1487, 2009.
20. H. Salehian, R. Chakraborty, E. Ofori, D. Vaillancourt, and B. C. Vemuri. An efficient recursive estimator of the Fréchet mean on a hypersphere with applications to medical image analysis. In *Mathematical Foundations of Computational Anatomy*, 2015.
21. D. Seo, J. Ho, and B. C. Vemuri. Computing diffeomorphic paths for large motion interpolation. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1227–1232. IEEE, 2013.
22. J. Talairach and P. Tournoux. *Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: an approach to cerebral imaging*. Thieme, 1988.
23. B. C. Vemuri, J. Ye, Y. Chen, and C. M. Leonard. Image registration via level-set motion: Applications to atlas-based segmentation. *Medical image analysis*, 7(1):1–20, 2003.
24. T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1):S61–S72, 2009.
25. G. Wu, H. Jia, Q. Wang, and D. Shen. Sharpmean: groupwise registration guided by sharp mean image and tree-based registration. *NeuroImage*, 56(4):1968–1981, 2011.
26. Y. Xie, J. Ho, and B. C. Vemuri. Multiple atlas construction from a heterogeneous brain mr image collection. *Medical Imaging, IEEE Transactions on*, 32(3):628–635, 2013.
27. B. T. Yeo, M. R. Sabuncu, R. Desikan, B. Fischl, and P. Golland. Effects of registration regularization and atlas sharpness on segmentation accuracy. *Medical image analysis*, 12(5):603–615, 2008.
28. L. Younes. Shapes and diffeomorphisms, volume 171 of applied mathematical sciences, 2010.

Author Index