Diffeomorphisms, optimal transport and imaging applications

François-Xavier Vialard

LIGM, UPEM

Labex Bézout, Dec. 2018

Contents

Example of problems of interest

Given two shapes, find a diffeomorphism of $\mathbb{R}^3$ *that maps one shape onto the other*

Example of problems of interest

Given two shapes, find a diffeomorphism of $\mathbb{R}^3$ that maps one shape onto the other

Different data types and different way of representing them.

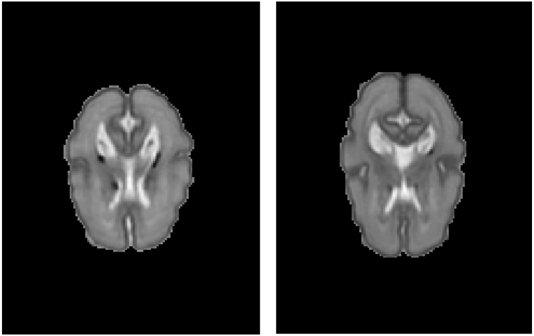


Figure – Two slices of 3D brain images of the same subject at different ages

Variety of shapes

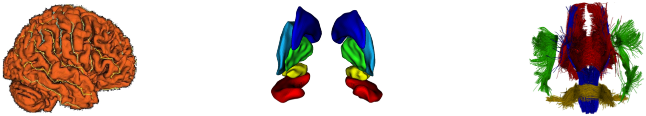


Figure – Different anatomical structures extracted from MRI data

Variety of shapes

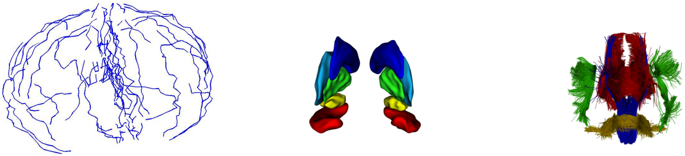


Figure – Different anatomical structures extracted from MRI data

Variational formulation

Find the best deformation, minimize

$$\mathcal{J}(\phi) = \inf_{\phi \in G_V} \underbrace{d(\phi.A, B)^2}_{\text{similarity measure}} \tag{1}$$

Variational formulation

Find the best deformation, minimize

$$\cancel{\mathcal{J}(\phi) = \inf_{\phi \in G_V} \underbrace{d(\phi.A, B)^2}_{\text{similarity measure}}} \tag{1}$$

Tychonov regularization:

$$\mathcal{J}(\phi) = \underbrace{R(\phi)}_{\text{Regularization}} + \underbrace{\frac{1}{2\sigma^2} d(\phi.A, B)^2}_{\text{similarity measure}} \,. \tag{2}$$

- Choice of similarity measure,
- Choice of the deformation regularization,
- Choice of the optimization method.

A Riemannian approach to diffeomorphic registration

Several diffeomorphic registration methods are available:

- Free-form deformations B-spline-based diffeomorphisms by D. Rueckert
- Log-demons (X.Pennec et al.)
- Large Deformations by Diffeomorphisms (M. Miller,A. Trouvé, L. Younes)
  Riemannian framework: Right-invariant metric on the group of diffeomorphisms and corresponding metric on objects.

Why does the Riemannian framework matter?

Generalizations of statistical tools in Euclidean space:

- Distance often given by a Riemannian metric.
- Straight lines $\rightarrow$ geodesic defined by

Variational definition: $\arg\min_{c(t)} \int_0^1 \|\dot{c}\|^2_{c(t)} \, dt = 0\,,$

Equivalent (local) definition: $\nabla_{\dot{c}} \dot{c} = \ddot{c} + \Gamma(c)(\dot{c}, \dot{c}) = 0\,.$

- Average $\rightarrow$ Fréchet/Karcher mean.

Variational definition: $\arg\min\{x \rightarrow E[d^2(x, y)] d\mu(y)\}$

Critical point definition: $E[\nabla_x d^2(x, y)] d\mu(y)] = 0\,.$

- PCA $\rightarrow$ Tangent PCA or PGA.
- Geodesic regression, cubic regression...(variational or algebraic)

Why does the Riemannian framework matter?

Generalizations of statistical tools in Euclidean space:

- Distance often given by a Riemannian metric.
- Straight lines $\rightarrow$ geodesic defined by

$$\text{Variational definition:} \quad \arg\min_{c(t)} \int_0^1 \|\dot{c}\|^2_{c(t)}\, dt = 0\,,$$

$$\text{Equivalent (local) definition:} \quad \nabla_{\dot{c}}\dot{c} = \ddot{c} + \Gamma(c)(\dot{c},\dot{c}) = 0\,.$$

- Average $\rightarrow$ Fréchet/Karcher mean.

$$\text{Variational definition:} \qquad \arg\min\{x \rightarrow E[d^2(x,y)]d\mu(y)\}$$

$$\text{Critical point definition:} \qquad E[\nabla_x d^2(x,y)]d\mu(y)] = 0\,.$$

- PCA $\rightarrow$ Tangent PCA or PGA.
- Geodesic regression, cubic regression...(variational or algebraic)

Riemannian metric needed, or at least a connection.

Why does the Riemannian framework matter?

Generalizations of statistical tools in Euclidean space:

- Distance often given by a Riemannian metric.
- Straight lines $\rightarrow$ geodesic defined by

$$\text{Variational definition: } \arg\min_{c(t)} \int_0^1 \|\dot{c}\|^2_{c(t)}\, dt = 0\,,$$

$$\text{Equivalent (local) definition: } \nabla_{\dot{c}}\dot{c} = \ddot{c} + \Gamma(c)(\dot{c},\dot{c}) = 0\,.$$

- Average $\rightarrow$ Fréchet/Karcher mean.

$$\text{Variational definition: } \quad \arg\min\{x \rightarrow E[d^2(x,y)]d\mu(y)\}$$

$$\text{Critical point definition: } \quad E[\nabla_x d^2(x,y)]d\mu(y)] = 0\,.$$

- PCA $\rightarrow$ Tangent PCA or PGA.
- Geodesic regression, cubic regression...(variational or algebraic)

Riemannian metric needed, or at least a connection.

Pitfalls:

- Loose uniqueness of geodesic or average (positive curvature).
- Equivalent definitions diverge (generalisation of PCA).

What needs to be implemented?

Solve the control problem:

$$\min \int_0^1 \|v_t\|_H^2 \,\mathrm{d}t + d(I(t=1,x),J)^2 \text{ under the constraint} \tag{3}$$

$$\begin{cases} \dot{I} + \langle \nabla I, v \rangle = 0\,, \\ I(t=0,x) = I_0(x)\,. \end{cases} \tag{4}$$

At a critical point, shooting equations on $I(t,x)$ the image, and $P(t,x)$ the momentum:

$$\begin{cases} \dot{I} + \langle \nabla I, v \rangle = 0\,, \\ \dot{P} + \mathrm{div}(Pv) = 0\,, \\ v + K \star (P\,\nabla I) = 0\,. \end{cases} \tag{5}$$

Optimization on the initial momentum using adjoint equations, equivalently, using automatic differentiation.

Interpolation, Extrapolation

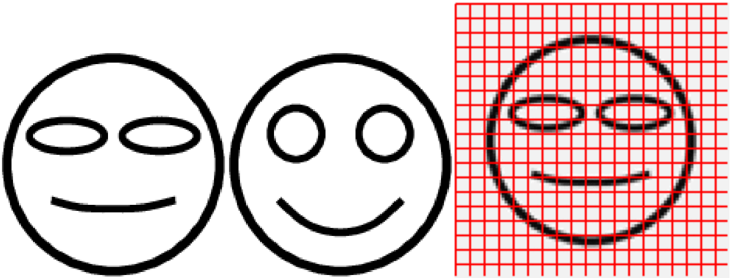


Figure – Geodesic regression (MICCAI 2011)

Interpolation, Extrapolation

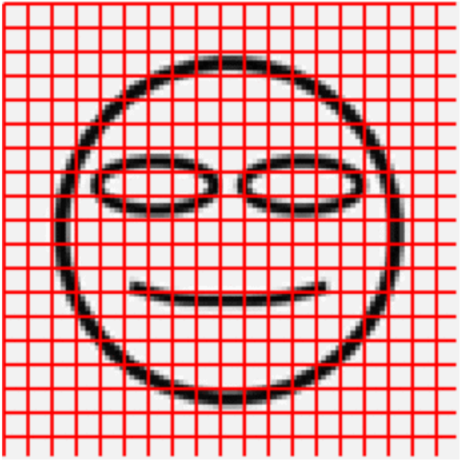
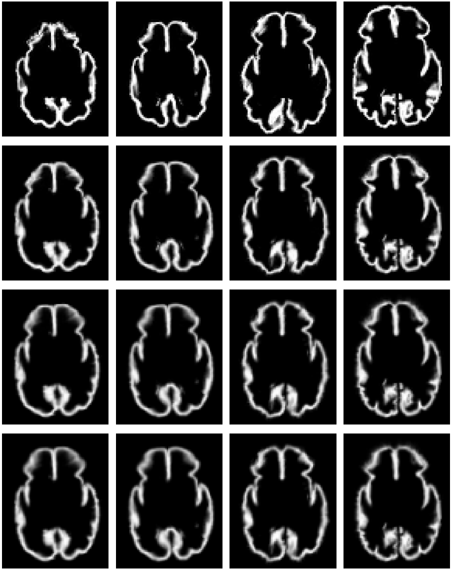


Figure – Extrapolation of happiness

Karcher mean on 3D images



Init. guesses

1 iteration

2 iterations

3 iterations

A_i^1 A_i^2 A_i^3 A_i^4

How to smoothly interpolate longitudinal data

In the Euclidean space:

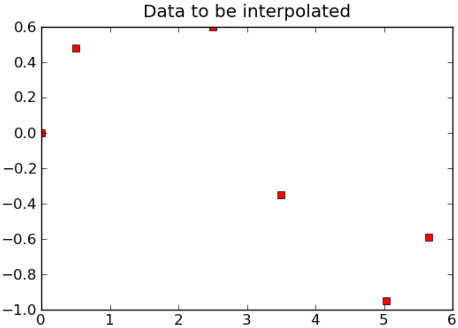


Figure – Sparse data from a sinus curve

How to smoothly interpolate longitudinal data

In the Euclidean space:

Minimizing the L^2 norm of the **speed** $\rightarrow$ piecewise linear interpolation

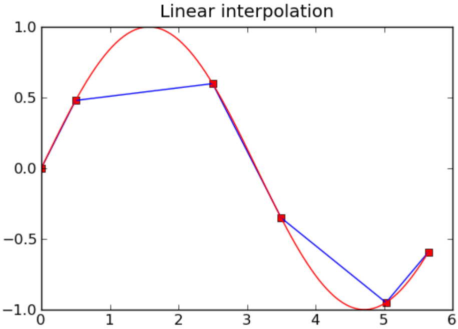


Figure – Linear interpolation of the data.

How to smoothly interpolate longitudinal data

In the Euclidean space:

Minimizing the L^2 norm of the **acceleration** $\rightarrow$ cubic spline interpolation

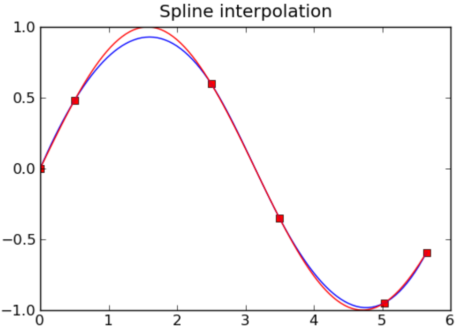


Figure – Cubic spline interpolation of the data.

Riemannian cubic splines

Acceleration on a Riemannian manifold M: let $c : I \to M$ be a C^2 curve. The notion of acceleration is:

$$\frac{D}{dt}\dot{c}(t) = \nabla_{\dot{c}}\dot{c}(= \ddot{c}_k + \sum_{i,j} \dot{c}_i \Gamma^k_{i,j} \dot{c}_j) \tag{6}$$

with ∇ the Levi-Civita connection.

Riemannian cubic splines

Acceleration on a Riemannian manifold M: let $c : I \to M$ be a C^2 curve. The notion of acceleration is:

$$\frac{D}{dt}\dot{c}(t) = \nabla_{\dot{c}}\dot{c}(= \ddot{c}_k + \sum_{i,j} \dot{c}_i \Gamma^k_{i,j} \dot{c}_j) \tag{6}$$

with ∇ the Levi-Civita connection.
Riemannian (variational) splines:

$$\inf_c \int_0^1 \frac{1}{2} |\nabla_{\dot{c}_t} \dot{c}_t|^2_M + \frac{\varepsilon}{2} |\dot{c}_t|^2_M dt\,. \tag{7}$$

subject to $c(i) = c_i$ and $\dot{c}(i) = v_i$ for $i = 0, 1$.

- Noakes, Crouch, Silva-Leite (SO_3 optimal control problem)
- Trouvé, Vialard (on landmarks, finite dimensional parametrization of the group of diffeomorphisms, QAM, 2012).
- Singh, Vialard, Niethammer (IEEE TMI, 2015).

What metric to choose?

Choosing the right-invariant metric

Right-invariant metric: Eulerian fluid dynamic viewpoint on regularization.

Space V of vector fields is defined equivalently by

- its kernel K such as Gaussian kernel,
- its differential operator, for instance $(\mathrm{Id} - \sigma\Delta)^n$ for Sobolev spaces.

Choosing the right-invariant metric

Right-invariant metric: Eulerian fluid dynamic viewpoint on regularization.

Space V of vector fields is defined equivalently by

- its kernel K such as Gaussian kernel,
- its differential operator, for instance $(\mathrm{Id} - \sigma\Delta)^n$ for Sobolev spaces.

The norm on V is simply

$$\|v\|_V^2 = \int_\Omega \langle v(x), (Lv)(x) \rangle \, dx = \int_\Omega (L^{1/2} v)^2(x) \, dx \,.$$

Choosing the right-invariant metric

Right-invariant metric: Eulerian fluid dynamic viewpoint on regularization.

Space V of vector fields is defined equivalently by

- its kernel K such as Gaussian kernel,
- its differential operator, for instance $(\mathrm{Id} - \sigma\Delta)^n$ for Sobolev spaces.

The norm on V is simply

$$\|v\|_V^2 = \int_\Omega \langle v(x), (Lv)(x)\rangle \, dx = \int_\Omega (L^{1/2} v)^2(x) \, dx \, .$$

Scale parameter important!

$$k_\sigma(x,y) = e^{-\frac{\|x-y\|^2}{\sigma^2}} \text{ kernel/operator } (\mathrm{Id} - \sigma\Delta)^n \tag{8}$$

- σ small: good matching but non regular deformations and more local minima.
- σ large: poor matching but regular deformations and more global minima.

Sum of kernels and multiscale

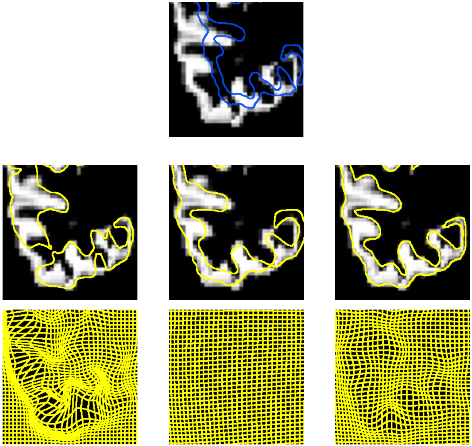
Choice of mixture of Gaussian kernels: (Risser, Vialard et al. 2011)

$$K(x, y) = \sum_{i=1}^{n} \alpha_i e^{-\frac{\|x-y\|^2}{\sigma_i^2}} \tag{9}$$

Sum of kernels and multiscale

Choice of mixture of Gaussian kernels: (Risser, Vialard et al. 2011)

$$K(x,y) = \sum_{i=1}^{n} \alpha_i e^{-\frac{\|x-y\|^2}{\sigma_i^2}} \tag{9}$$



From Eulerian to Lagrangian viewpoints

Spatial correlation of the deformation: need for local deformability on the tissues.
Toward a more Lagrangian point of view.

From Eulerian to Lagrangian viewpoints

Spatial correlation of the deformation: need for local deformability on the tissues.
Toward a more Lagrangian point of view.

How to introduce spatially varying metric?

Using kernels: χ_i being a partition of unity of the domain.

$$K = \sum_{i=1}^{n} \chi_i K_i \chi_i \,, . \tag{10}$$

This kernel is associated to the following variational interpretation:

$$\|v\|^2 = \min_{(v_1,\ldots,v_n)\in V_1\times\ldots\times V_n} \left\{ \sum_{i=1}^{n} \|v_i\|_{V_i}^2 \;\middle|\; \sum_{i=1}^{n} \chi_i v_i = v \right\} . \tag{11}$$

Learning the metric

- Choice of a simple model for diffeomorphisms (SVF).
- Optimize on the choice of the metric: partition of unity and weights.
- Make it data adaptive using NN parametrization.

More precisely,

$$v_0(x) \stackrel{\text{def.}}{=} \sum_{i=0}^{N-1} \sqrt{w_i(x)} \int_y G_i(|x-y|) \sqrt{w_i(y)} m_0(y)\, \mathrm{d}y \tag{12}$$

$$m^* = \underset{m_0}{\operatorname{argmin}}\ \lambda \langle m_0, v_0 \rangle + \mathrm{Sim}[I_0 \circ \Phi^{-1}(1), I_1] + \lambda_{\mathrm{OMT}} \int \widehat{\mathrm{OMT}}(w(x))\, \mathrm{dx} + \lambda_{\mathrm{TV}} \int \gamma(\|\nabla I_0(x)\|) \sqrt{\sum_{i=0}^{N-1} \|\nabla \omega_i(x)\|_2^2\, \mathrm{dx}}\ , \tag{13}$$

subject to the constraints $\Phi_t^{-1} + D\Phi^{-1} v = 0$ and $\Phi^{-1}(0) = \mathrm{id}$;
$\lambda_{TV}, \lambda_{OMT} \geq 0$

Learning the metric

Method	**mean**	**std**	**1%**	**5%**	**50%**	**95%**	**99%**
FLIRT	0.394	0.031	0.334	0.345	0.396	0.442	0.463
AIR	0.423	0.030	0.362	0.377	0.421	0.483	0.492
ANIMAL	0.426	0.037	0.328	0.367	0.425	0.483	0.498
ART	0.503	0.031	0.446	0.452	0.506	0.556	0.563
Demons	0.462	**0.029**	0.407	0.421	0.461	0.510	0.531
FNIRT	0.463	0.036	0.381	0.410	0.463	0.519	0.537
Fluid	0.462	0.031	0.401	0.410	0.462	0.516	0.532
SICLE	0.419	0.044	0.300	0.330	0.424	0.475	0.504
SyN	0.514	0.033	0.454	0.460	0.515	0.565	0.578
SPM5N8	0.365	0.045	0.257	0.293	0.370	0.426	0.455
SPM5N	0.420	0.031	0.361	0.376	0.418	0.471	0.494
SPM5U	0.438	0.029	0.373	0.394	0.437	0.489	0.502
SPM5D	0.512	0.056	0.262	0.445	0.523	0.570	0.579
m/c stage 0	0.423	0.029	0.363	0.381	0.423	0.470	0.490
m/c stage 1	0.444	0.031	0.387	0.399	0.446	0.493	0.506
m/c stage 2	0.518	0.035	0.454	0.461	0.523	0.570	0.583
c/c stage 0	0.423	0.029	0.363	0.381	0.423	0.470	0.490
c/c stage 1	0.439	0.031	0.382	0.394	0.441	0.488	0.501
c/c stage 2	**0.523**	0.034	**0.458**	**0.466**	**0.528**	**0.573**	**0.586**
i/c stage 0	0.423	0.029	0.363	0.381	0.423	0.470	0.490
i/c stage 1	0.445	0.031	0.388	0.400	0.447	0.494	0.507
i/c stage 2	0.520	0.035	0.456	0.462	0.525	0.571	0.584

Table – Mean target overlap ratios for different methods. Best results are in bold. Niethammer, Kwitt, Vialard.

Contents

A new similarity measure

$$\mathcal{J}(\phi) = \underbrace{R(\phi)}_{\text{Regularization}} + \underbrace{\frac{1}{2\sigma^2} d(\phi.A, B)^2}_{\text{similarity measure}} . \quad (14)$$

Mild constraints:

- Fast to evaluate
- Differentiable
- Convex

Examples: SSD, normalized cross correlation, LCC.

- Locality/Semi-locality.

More global similarity measure: Optimal Transport

Why? impressive numerical advances in the recent years.

Optimal transport

Given a metric space (X, d), optimal transport is a natural way to lift this metric to $\mathcal{P}(X)$ the space of probability measures (or nonnegative densities that integrates to 1).

Essentially, $W(\delta_x, \delta_y) = d(x, y)$.

Optimal transport

Given a metric space (X, d), optimal transport is a natural way to lift this metric to $\mathcal{P}(X)$ the space of probability measures (or nonnegative densities that integrates to 1).

$$\boxed{\text{Essentially, } W(\delta_x, \delta_y) = d(x, y).}$$

For data applications, NOT EVERYTHING IS A DENSITY !

Static Formulation

Monge formulation (1781)

Let $\mu, \nu \in \mathcal{P}_+(M)$,

$$\textbf{Minimize} \int_M c(x, \varphi(x)) d\mu \tag{15}$$

among the map s.t. $\varphi_*(\mu) = \nu$.

Static Formulation

Monge formulation (1781)

Let $\mu, \nu \in \mathcal{P}_+(M)$,

$$\textbf{Minimize} \int_M c(x, \varphi(x)) d\mu \tag{15}$$

among the map s.t. $\varphi_*(\mu) = \nu$.

1. ill posed problem, the constraint may not be satisfied.
2. the constraint can hardly be made weakly closed.

$\rightarrow$ Relaxation of the Monge problem.

Static Formulation

Kantorovich formulation (1942)

Let $\mu, \nu \in \mathcal{P}_+(M)$, define D by

$$D(\mu,\nu) = \inf_{\gamma \in \mathcal{P}(M^2)} \left\{ \int_{M^2} c(x,y)\, \mathrm{d}\gamma(x,y) : \pi^1_* \gamma = \mu \text{ and } \pi^2_* \gamma = \nu \right\}$$

1. Existence result: c lower semi-continuous and bounded from below.
2. Also valid in Polish spaces.
3. If $c(x,y) = \frac{1}{p}|x-y|^p$, $D^{1/p}$ is the Wasserstein distance denoted by W_p.

Static Formulation

Kantorovich formulation (1942)

Let $\mu, \nu \in \mathcal{P}_+(M)$, define D by

$$D(\mu,\nu) = \inf_{\gamma \in \mathcal{P}(M^2)} \left\{ \int_{M^2} c(x,y)\, \mathrm{d}\gamma(x,y) : \pi^1_* \gamma = \mu \text{ and } \pi^2_* \gamma = \nu \right\}$$

1. Existence result: c lower semi-continuous and bounded from below.
2. Also valid in Polish spaces.
3. If $c(x,y) = \frac{1}{p}|x-y|^p$, $D^{1/p}$ is the Wasserstein distance denoted by W_p.

Linear optimization problem and associated numerical methods.
Recently introduced, entropic regularization. (C. Léonard, M. Cuturi, J.C. Zambrini <- Schrödinger)

A simple example where Monge = Kantorovich

$$\mathcal{P}_+^n(\Omega) \stackrel{\text{def.}}{=} \left\{ \mu = \sum_{i=1}^{n} \frac{1}{n}\delta_{x_i} \,;\, m_i \geq 0,\, x_i \in \Omega \right\}$$

$$B_n \stackrel{\text{def.}}{=} \left\{ \gamma \in M_n(\mathbb{R}) \,;\, \gamma(i,j) \geq 0, \sum_{i=1}^{n} \gamma(i,j) = \frac{1}{n} \sum_{j=1}^{n} \gamma(i,j) = \frac{1}{n} \right\}$$

The Kantorovich problem on $\mathcal{P}_+^n(\Omega)$

Let $\nu = \sum_{j=1}^n \frac{1}{n}\delta_{y_j}$ and $\mu = \sum_{j=1}^n \frac{1}{n}\delta_{x_j}$, then

$$D(\mu,\nu) = \inf_{\gamma \in B_n} \sum_{i,j} c(x_i, y_j)\gamma(i,j) \text{ s.t. } \begin{cases} \sum_{j=1}^n \gamma(i,j) = \frac{1}{n} \\ \sum_{i=1}^n \gamma(i,j) = \frac{1}{n} \end{cases}$$

A simple example where Monge = Kantorovich

$$\mathcal{P}_+^n(\Omega) \stackrel{\text{def.}}{=} \left\{ \mu = \sum_{i=1}^n \frac{1}{n}\delta_{x_i} \,;\, m_i \geq 0,\ x_i \in \Omega \right\}$$

$$B_n \stackrel{\text{def.}}{=} \left\{ \gamma \in M_n(\mathbb{R}) \,;\, \gamma(i,j) \geq 0, \sum_{i=1}^n \gamma(i,j) = \frac{1}{n} \sum_{j=1}^n \gamma(i,j) = \frac{1}{n} \right\}$$

The Kantorovich problem on $\mathcal{P}_+^n(\Omega)$

Let $\nu = \sum_{j=1}^n \frac{1}{n}\delta_{y_j}$ and $\mu = \sum_{j=1}^n \frac{1}{n}\delta_{x_j}$, then

$$D(\mu,\nu) = \inf_{\gamma \in B_n} \sum_{i,j} c(x_i, y_j)\gamma(i,j) \text{ s.t. } \begin{cases} \sum_{j=1}^n \gamma(i,j) = \frac{1}{n} \\ \sum_{i=1}^n \gamma(i,j) = \frac{1}{n} \end{cases}$$

Linear optimization problem achieved at an extremal point of B_n: $(\frac{1}{n}P\ )$ with P a permutation matrix.

$\rightarrow$ A Monge solution to the Kantorovich problem.

Dynamic formulation (Benamou-Brenier)

For geodesic costs, for instance $c(x,y) = \frac{1}{2}|x-y|^2$

$$\inf \mathcal{E}(v) = \frac{1}{2}\int_0^1 \int_M |v(x)|^2 \rho(x) \, \mathrm{d}x \, \mathrm{d}t \; , \tag{16}$$

s.t.

$$\begin{cases} \dot{\rho} + \nabla \cdot (v\rho) = 0 \\ \rho(0) = \mu_0 \text{ and } \rho(1) = \mu_1 \, . \end{cases} \tag{17}$$

Dynamic formulation (Benamou-Brenier)

For geodesic costs, for instance $c(x,y) = \frac{1}{2}|x-y|^2$

$$\inf \mathcal{E}(v) = \frac{1}{2}\int_0^1 \int_M |v(x)|^2 \rho(x)\, \mathrm{d}x\, \mathrm{d}t \;, \tag{16}$$

s.t.

$$\begin{cases} \dot\rho + \nabla \cdot (v\rho) = 0 \\ \rho(0) = \mu_0 \text{ and } \rho(1) = \mu_1 \,. \end{cases} \tag{17}$$

Convex reformulation: Change of variable: momentum $m = \rho v$,

$$\inf \mathcal{E}(m) = \frac{1}{2}\int_0^1 \int_M \frac{|m(x)|^2}{\rho(x)}\, \mathrm{d}x\, \mathrm{d}t\,, \tag{18}$$

s.t.

$$\begin{cases} \dot\rho + \nabla \cdot m = 0 \\ \rho(0) = \mu_0 \text{ and } \rho(1) = \mu_1 \,. \end{cases} \tag{19}$$

where $(\rho, m) \in \mathcal{M}([0,1] \times M, \mathbb{R} \times \mathbb{R}^d)$.

Dynamic formulation (Benamou-Brenier)

For geodesic costs, for instance $c(x,y) = \frac{1}{2}|x-y|^2$

$$\inf \mathcal{E}(v) = \frac{1}{2}\int_0^1 \int_M |v(x)|^2 \rho(x)\, \mathrm{d}x\, \mathrm{d}t \ , \tag{16}$$

s.t.

$$\begin{cases} \dot\rho + \nabla \cdot (v\rho) = 0 \\ \rho(0) = \mu_0 \text{ and } \rho(1) = \mu_1 \, . \end{cases} \tag{17}$$

Convex reformulation: Change of variable: momentum $m = \rho v$,

$$\inf \mathcal{E}(m) = \frac{1}{2}\int_0^1 \int_M \frac{|m(x)|^2}{\rho(x)}\, \mathrm{d}x\, \mathrm{d}t \, , \tag{18}$$

s.t.

$$\begin{cases} \dot\rho + \nabla \cdot m = 0 \\ \rho(0) = \mu_0 \text{ and } \rho(1) = \mu_1 \, . \end{cases} \tag{19}$$

where $(\rho, m) \in \mathcal{M}([0,1] \times M, \mathbb{R} \times \mathbb{R}^d)$.

Existence of minimizers: Fenchel-Rockafellar.
Numerics: First-order splitting algorithm: Douglas-Rachford.

Available numerical methods

- Auction algorithm $O(N^3)$.
- Proximal methods on dynamic (PDE) formulation (Benamou, Brenier), 10^5 points, restricted to W_2.
- Newton method for semi-discrete methods (Mérigot, Levy), 10^7 points, restricted to W_2.
- Approximation by entropic regularization, apply to any cost c, 10^9 points.

Starting point (2015) and initial motivation

Optimal transport applications: Imaging, machine learning, gradient flows, ...

Bottleneck in optimal transport: data has fixed total mass.

- Relax the mass constraint to extend OT distance between positive measures of arbitrary mass.
- Develop associated numerical algorithms.

Unbalanced optimal transport

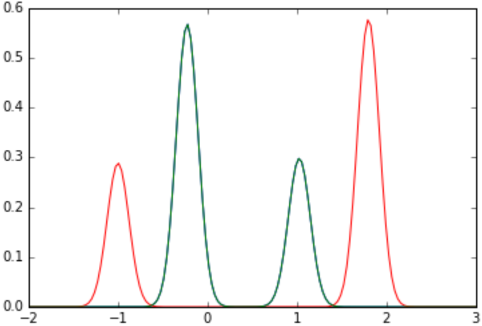


Figure – Optimal transport between bimodal densities

Unbalanced optimal transport

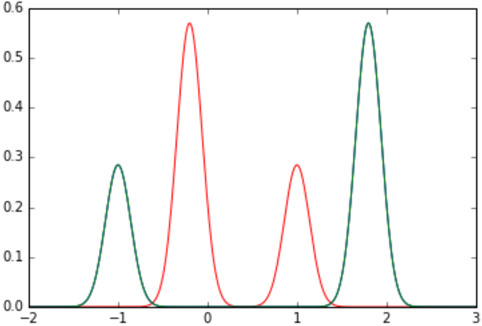


Figure – Another transformation

An extension of Benamou-Brenier formulation

Add a source term in the constraint: (weak sense)

$$\dot{\rho} = -\nabla \cdot (\rho v) + \alpha \rho \,,$$

where α can be understood as the growth rate.

$$\mathrm{WF}^2 \stackrel{\text{def.}}{=} \inf_{(v,\alpha)} \frac{1}{2} \int_0^1 \int_M |v(x,t)|^2 \rho(x,t) \,\mathrm{d}x\,\mathrm{d}t + \frac{\delta^2}{2} \int_0^1 \int_M \alpha(x,t)^2 \rho(x,t) \,\mathrm{d}x\,\mathrm{d}t \,.$$

where δ is a length parameter.

An extension of Benamou-Brenier formulation

Add a source term in the constraint: (weak sense)

$$\dot{\rho} = -\nabla \cdot (\rho v) + \alpha\rho\,,$$

where α can be understood as the growth rate.

$$\mathrm{WF}^2 \stackrel{\text{def.}}{=} \inf_{(v,\alpha)} \frac{1}{2}\int_0^1 \int_M |v(x,t)|^2 \rho(x,t)\, \mathrm{d}x\, \mathrm{d}t + \frac{\delta^2}{2}\int_0^1 \int_M \alpha(x,t)^2 \rho(x,t)\, \mathrm{d}x\, \mathrm{d}t\,.$$

where δ is a length parameter.
Remark: very natural and not studied before.

Convex reformulation

Add a source term in the constraint: (weak sense)

$$\dot{\rho} = -\nabla \cdot m + \mu \,.$$

The Wasserstein-Fisher-Rao metric:

$$\mathrm{WF}^2 \overset{\text{def.}}{=} \inf_{(v,\alpha)} \frac{1}{2} \int_0^1 \int_M \frac{|m(x,t)|^2}{\rho(x,t)} \,\mathrm{d}x\,\mathrm{d}t + \frac{\delta^2}{2} \int_0^1 \int_M \frac{\mu(x,t)^2}{\rho(x,t)} \,\mathrm{d}x\,\mathrm{d}t \,.$$

Convex reformulation

Add a source term in the constraint: (weak sense)

$$\dot{\rho} = -\nabla \cdot m + \mu \,.$$

The Wasserstein-Fisher-Rao metric:

$$\mathrm{WF}^2 \stackrel{\text{def.}}{=} \inf_{(v,\alpha)} \frac{1}{2} \int_0^1 \int_M \frac{|m(x,t)|^2}{\rho(x,t)} \,\mathrm{d}x \,\mathrm{d}t + \frac{\delta^2}{2} \int_0^1 \int_M \frac{\mu(x,t)^2}{\rho(x,t)} \,\mathrm{d}x \,\mathrm{d}t \,.$$

- Fisher-Rao metric: Hessian of the Boltzmann entropy/ Kullback-Leibler divergence and reparametrization invariant. Wasserstein metric on the space of variances in 1D.
- Convex and 1-homogeneous: convex analysis (existence and more)
- Numerics: First-order splitting algorithm: Douglas-Rachford.
- Code available at `https://github.com/lchizat/optimal-transport/`

Numerical simulations

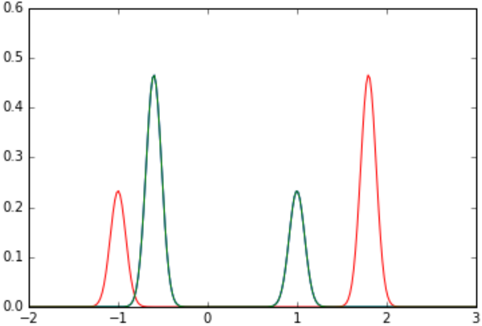


Figure – WFR geodesic between bimodal densities

Numerical simulations

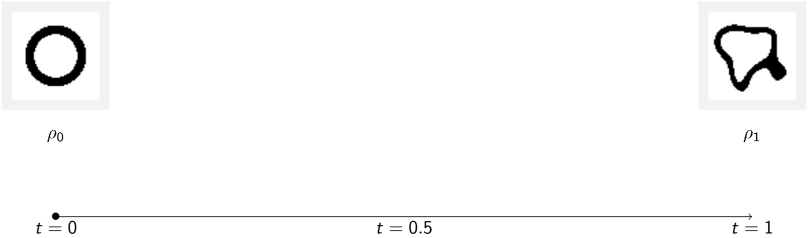


Figure – Geodesics between ρ_0 and ρ_1 for (1st row) Hellinger, (2nd row) W_2, (3rd row) partial OT, (4th row) WF.

An Interpolating Distance between Optimal Transport and Fisher-Rao, L. Chizat, B. Schmitzer, G. Peyré, and F.-X. Vialard, FoCM, 2016.

Numerical simulations

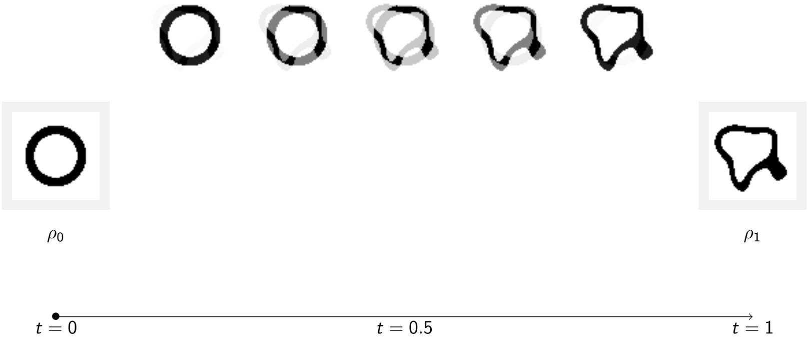


Figure – Geodesics between ρ_0 and ρ_1 for (1st row) Hellinger, (2nd row) W_2, (3rd row) partial OT, (4th row) WF.

An Interpolating Distance between Optimal Transport and Fisher-Rao, L. Chizat, B. Schmitzer, G. Peyré, and F.-X. Vialard, FoCM, 2016.

Numerical simulations

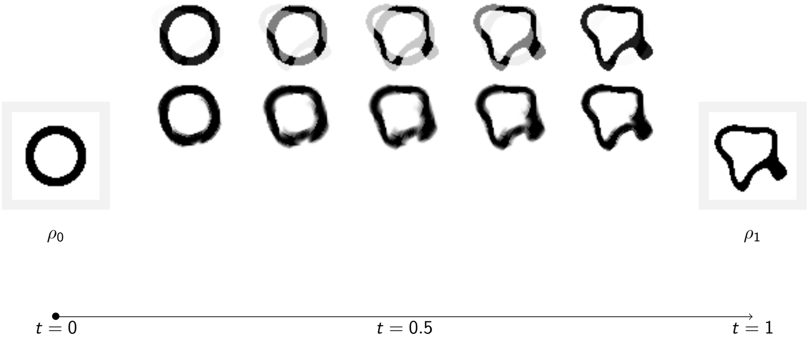


Figure – Geodesics between ρ_0 and ρ_1 for (1st row) Hellinger, (2nd row) W_2, (3rd row) partial OT, (4th row) WF.

An Interpolating Distance between Optimal Transport and Fisher-Rao, L. Chizat, B. Schmitzer, G. Peyré, and F.-X. Vialard, FoCM, 2016.

Numerical simulations

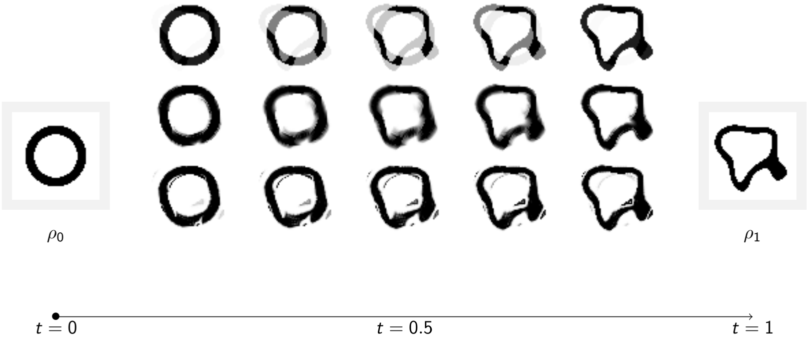


Figure – Geodesics between ρ_0 and ρ_1 for (1st row) Hellinger, (2nd row) W_2, (3rd row) partial OT, (4th row) WF.

An Interpolating Distance between Optimal Transport and Fisher-Rao, L. Chizat, B. Schmitzer, G. Peyré, and F.-X. Vialard, FoCM, 2016.

Numerical simulations

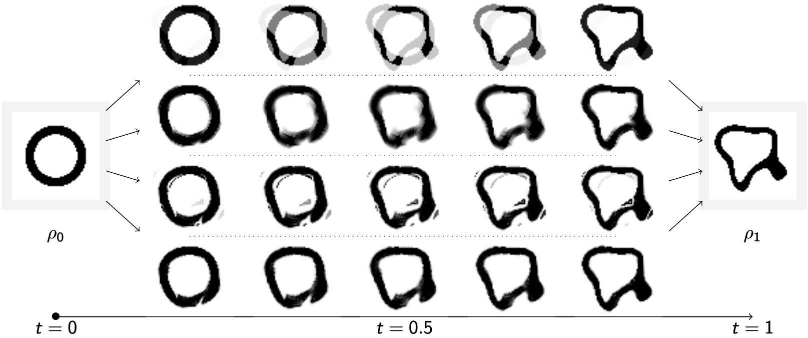


Figure – Geodesics between ρ_0 and ρ_1 for (1st row) Hellinger, (2nd row) W_2, (3rd row) partial OT, (4th row) WF.

An Interpolating Distance between Optimal Transport and Fisher-Rao, L. Chizat, B. Schmitzer, G. Peyré, and F.-X. Vialard, FoCM, 2016.

A relaxed static OT formulation

Define

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

$$\begin{aligned} WF^2(\rho_1,\rho_2) = \inf_{\gamma} KL(\mathrm{Proj}^1_* \gamma, \rho_1) + KL(\mathrm{Proj}^2_* \gamma, \rho_2) \\ - \int_{M^2} \gamma(x,y) \log(\cos^2(d(x,y)/2 \wedge \pi/2))\, \mathrm{d}x\, \mathrm{d}y \end{aligned}$$

Theorem (Gallouet - Vialard)

On a Riemannian manifold (compact without boundary), the static and dynamic formulations are equal.

Idea: Generalize Otto's Riemannian submersion

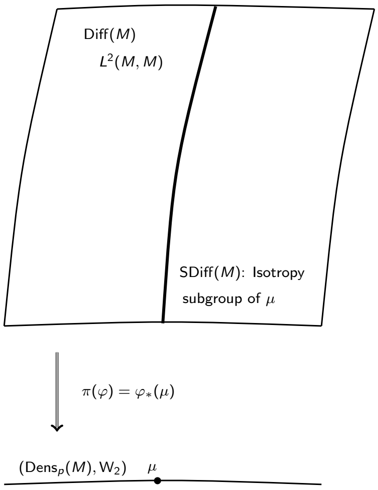


Figure – A Riemannian submersion: SDiff(M) as a Riemannian submanifold of $L^2(M, M)$: Incompressible Euler equation on SDiff(M)

Idea: Generalize Otto's Riemannian submersion

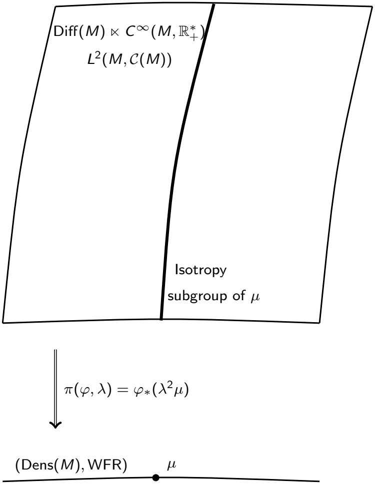


Figure – The same picture in our case: what is the corresponding equation to Euler?

New algorithm

Scaling Algorithms for Unbalanced Transport Problems, L. Chizat, G. Peyré, B. Schmitzer, F.-X. Vialard.

- Use of entropic regularization.

$$
\begin{aligned}
WF_{\varepsilon}{}^2(\rho_1, \rho_2) = \inf_{\gamma} \lambda KL(\mathrm{Proj}^1_* \gamma, \rho_1) + \lambda KL(\mathrm{Proj}^2_* \gamma, \rho_2) \\
- \int_{M^2} \gamma(x,y) \log(\cos^2(d(x,y)/2 \wedge \pi/2)) \,\mathrm{d}x \,\mathrm{d}y + \varepsilon KL(\gamma, \mu_0)\,.
\end{aligned}
$$

Optimality conditions imply $\gamma = a(x)b(y)K(x,y)$ with $K(x,y) = e^{-\frac{1}{\varepsilon}c(x,y)}$ and $c(x,y) = -\log(\cos^2(d(x,y)))$.

Scaling algorithms

Define

$$h(p, s) \stackrel{\text{def.}}{=} (p/s)^{\lambda/(\lambda+\varepsilon)}$$

and

$$\text{Marginal}_1(i) = \sum_j a_i K_{i,j} b_j \;\; \text{Marginal}_2(j) = \sum_i a_i K_{i,j} b_j \,.$$

function ScalingAlgo(K,μ, ν)
  $b \leftarrow \mathbb{1}_J$
  repeat
    $a \leftarrow h(\text{Marginal}_1, \mu)$
    $b \leftarrow h(\text{Marginal}_2, \nu)$
  until stopping criterion
  return $(a_i K_{i,j} b_j)_{i,j}$
end function

- Alternate projection algorithm (contraction for a Hilbert type metric, generalization of Sinkhorn's proof).
- Applications to color transfer, Fréchet-Karcher mean (barycenters).
- Simulations for gradient flows.

Application to color transfer

Figure – Transporting the color histograms: initial and final image

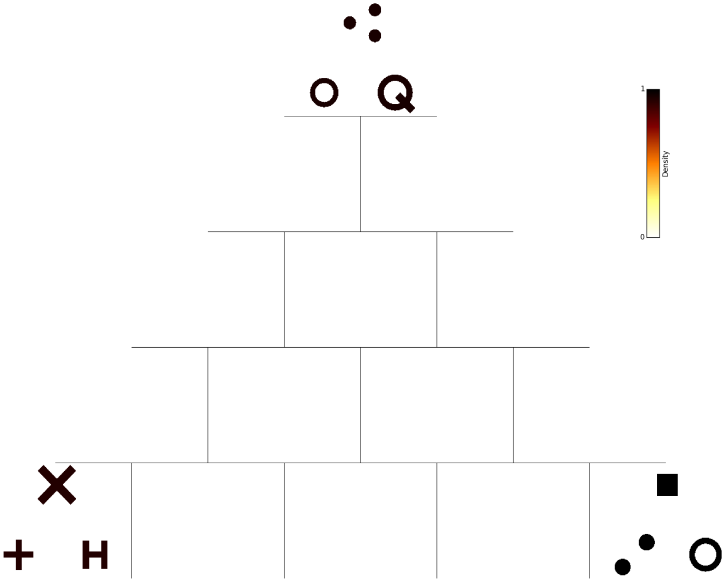
Optimal transport

Range constraint

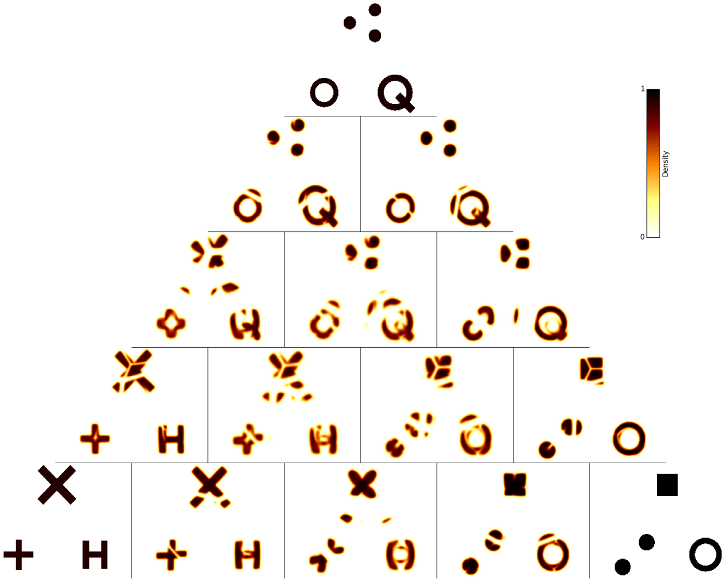
Kullback-Leibler

Total variation

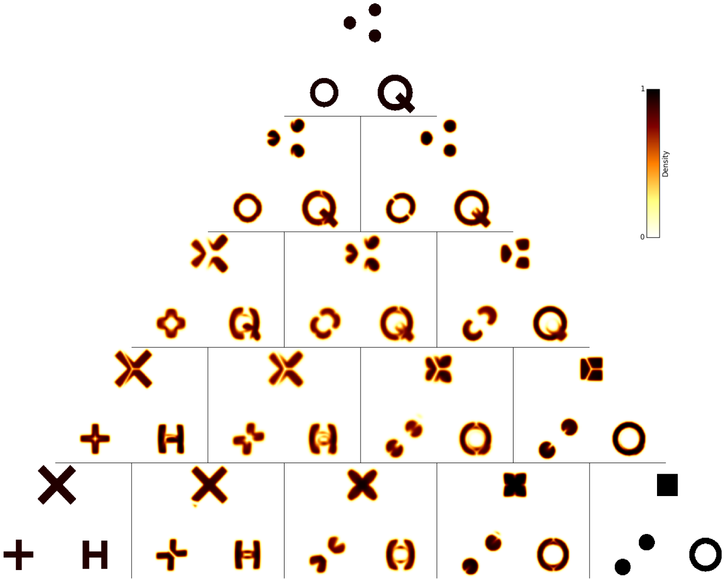
Barycenters : Wasserstein



Barycenters : Wasserstein



Barycenters : Unbalanced (GHK)



Contents

Incompressible Euler and optimal transport

Optimal transport appears in the projection onto *SDiff* in Brenier's work.

The Camassa-Holm equation is the corresponding fluid dynamic equation for *WF*.

Otto's Riemannian submersion for WFR

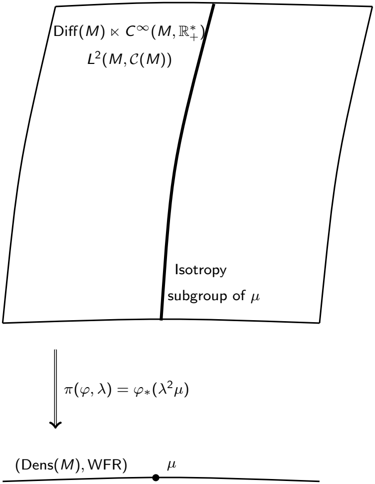


Figure – The same picture in our case: what is the corresponding equation to Euler?

The isotropy subgroup for unbalanced optimal transport

The induced metric on the isotropy subgroup is

$$G(v, \operatorname{div} v) = \int_M |v|^2 \,\mathrm{d}\mu + \frac{1}{4} \int_M |\operatorname{div} v|^2 \,\mathrm{d}\mu \,. \tag{20}$$

The isotropy subgroup for unbalanced optimal transport

The induced metric on the isotropy subgroup is

$$G(v, \operatorname{div} v) = \int_M |v|^2 \, \mathrm{d}\mu + \frac{1}{4} \int_M |\operatorname{div} v|^2 \, \mathrm{d}\mu \,. \tag{20}$$

Geodesics for $M = S_1$ is the Camassa-Holm equation 1981/1993.

- Model for waves in shallow water.
- Completely integrable system (bi-Hamiltonian).
- Exhibits particular solutions named as peakons. (geodesics as collective Hamiltonian).
- Blow-up of solutions which gives a model for wave breaking.

Consequences

1. Using Gauss-Codazzi formula, it generalizes a curvature formula by Khesin et al. obtained on $\mathrm{Diff}(S_1)$.
2. Smooth geodesics are length minimizing for a short enough time under mild conditions (generalization of Brenier's proof).
3. The Camassa-Holm equation as a particular case of incompressible Euler.
4. A new polar factorization theorem.
5. Generalized solutions to Camassa-Holm equation.

Results

Theorem

Let φ be the flow of a smooth solution to the Camassa-Holm equation then $\Psi(\theta, r) \stackrel{\text{def.}}{=} (\varphi(\theta), \sqrt{\mathrm{Jac}(\varphi(\theta))}\, r)$ is the flow of a solution to the incompressible Euler equation for the density $\frac{1}{r^4} r \,\mathrm{d}r\, \mathrm{d}\theta$.

Results

Theorem

Let φ be the flow of a smooth solution to the Camassa-Holm equation then $\Psi(\theta, r) \stackrel{\text{def.}}{=} (\varphi(\theta), \sqrt{\mathrm{Jac}(\varphi(\theta))}r)$ is the flow of a solution to the incompressible Euler equation for the density $\frac{1}{r^4} r \,\mathrm{d}r \,\mathrm{d}\theta$.

Case where $M = S_1$, $\mathcal{M}(\varphi) = [(\theta, r) \mapsto r\sqrt{\partial_x \varphi(\theta)} e^{i\varphi(\theta)}]$ then the CH equation is

$$\begin{cases} \partial_t u - \frac{1}{4}\partial_{txx} u\, u + 3\partial_x u\, u - \frac{1}{2}\partial_{xx} u\, \partial_x u - \frac{1}{4}\partial_{xxx} u\, u = 0 \\ \partial_t \varphi(t,x) = u(t, \varphi(t,x))\,. \end{cases} \tag{21}$$

The Euler equation on the cone, $\mathcal{C}(M) = \mathbb{R}^2 \setminus \{0\}$ for the density $\rho = \frac{1}{r^4}$ Leb is

$$\begin{cases} \dot{v} + \nabla_v v = -\nabla p\,, \\ \nabla \cdot (\rho v) = 0\,. \end{cases} \tag{22}$$

where $v(\theta, r) \stackrel{\text{def.}}{=} \left(u(\theta), \frac{r}{2}\partial_x u(\theta)\right)$.

Contents

Where to apply these efficient algorithms ?

Variational inverse problems

$$\mathcal{J}(\phi) = \underbrace{R(\phi)}_{\text{Regularization}} + \underbrace{\frac{1}{2\sigma^2}\mathcal{S}(\phi.A, B)^2}_{\text{similarity measure}} . \tag{23}$$

Mild constraints on $\mathcal{S}$:

- Fast to evaluate
- Differentiable
- Convex

Examples: SSD, normalized cross correlation, LCC.

- Locality/Semi-locality.

Entropic Wasserstein is smooth and convex.

A new divergence on probability measures

Recall that for $\epsilon > 0$, define

$$\mathrm{OT}_\varepsilon(\alpha,\beta) \stackrel{\text{def.}}{=} \min_{\pi_1=\alpha,\pi_2=\beta} \int_{X^2} C\, d\pi + \varepsilon\, \mathrm{KL}(\pi|\alpha\otimes\beta) \tag{24}$$

$$\text{where } \mathrm{KL}(\pi|\alpha\otimes\beta) \stackrel{\text{def.}}{=} \int_{X^2} \log \frac{d\pi}{d\alpha d\beta} d\pi\,.$$

A possible solution, (Feydy et al.)

The Sinkhorn divergence:

$$\mathrm{S}_\varepsilon(\alpha,\beta) \stackrel{\text{def.}}{=} \mathrm{OT}_\varepsilon(\alpha,\beta) - \frac{1}{2}\,\mathrm{OT}_\varepsilon(\alpha,\alpha) - \frac{1}{2}\,\mathrm{OT}_\varepsilon(\beta,\beta) \tag{25}$$

is non-negative, symmetric and convex in each variable.

Test on fiber bundle

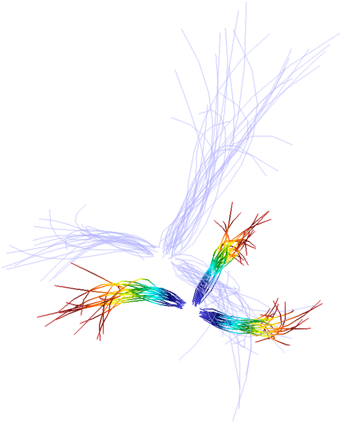


Figure – Registration with kernel metrics

Test on fiber bundle

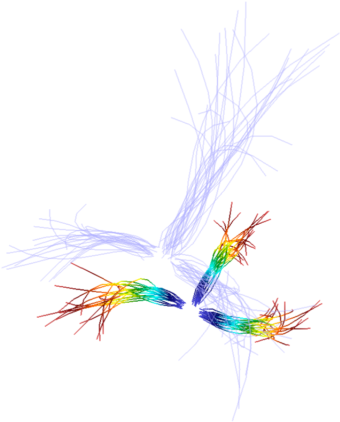


Figure – Registration with entropic optimal transport

With different scales

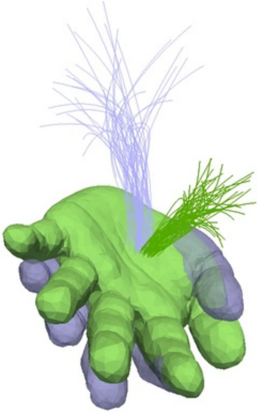


Figure – Registration with kernel metrics

With different scales

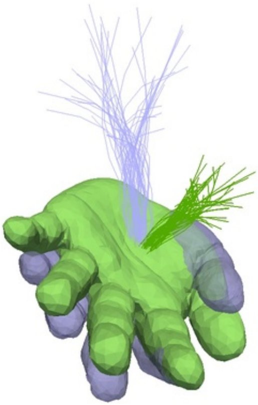


Figure – Registration with entropic optimal transport

References I