Diffeomorphic flows, optimal transport and geometry [1]

François-Xavier Vialard

Université Gustave Eiffel
Laboratoire d'informatique Gaspard Monge.

November 2023

1. Right-invariant metrics vs optimal transport

Outline

1. Motivations from imaging and learning
2. How to build Riemannian metric in infinite dimensions?
3. Right-invariant metrics
4. Dynamic formulation of optimal transport
5. Unbalanced Optimal Transport
6. Gradient flows
7. On global convergence of ResNets
8. The Camassa-Holm equation as an incompressible Euler equation

Contents

1 Motivations from imaging and learning

2 How to build Riemannian metric in infinite dimensions?

3 Right-invariant metrics

4 Dynamic formulation of optimal transport

5 Unbalanced Optimal Transport

6 Gradient flows

7 On global convergence of ResNets

8 The Camassa-Holm equation as an incompressible Euler equation

A diffeomorphic deformation

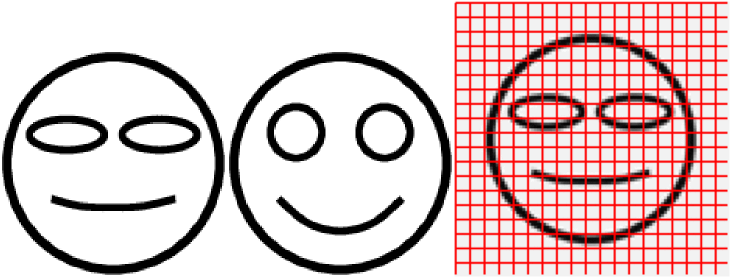


Figure – Interpolation of happiness

A diffeomorphic transformation is a smooth map which is smoothly invertible.

Computational Anatomy

Old questions:

- to find a framework for registration of biological shapes,
- to develop statistical analysis in this framework.

Computational Anatomy

Old questions:

- to find a framework for registration of biological shapes,
- to develop statistical analysis in this framework.

Action of a transformation group on shapes or images
Idea pioneered by Grenander and al. (80's), then developed by M.Miller, A.Trouvé, L.Younes,...

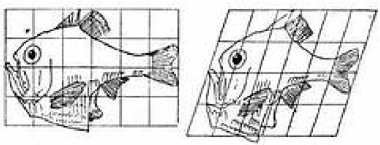


Figure – Deforming the shape of a fish by D'Arcy Thompson, author of *On Growth and Forms* (1917)

Variety of shapes

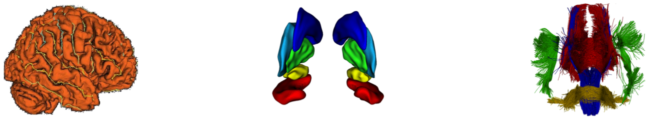


Figure – Different anatomical structures extracted from MRI data

Variety of shapes

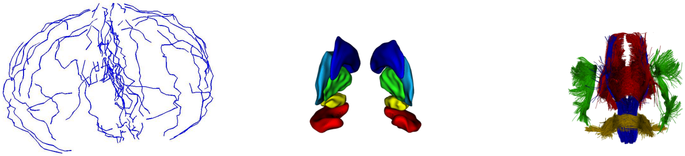


Figure – Different anatomical structures extracted from MRI data

Idea of Riemannian metrics on shapes

Generalizations of statistical tools in Euclidean space:

- Distance often given by a Riemannian metric.
- Straight lines $\to$ 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 $\to$ Fréchet/Karcher mean.

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

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

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

Idea of Riemannian metrics on shapes

Generalizations of statistical tools in Euclidean space:

- Distance often given by a Riemannian metric.
- Straight lines $\to$ 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 $\to$ Fréchet/Karcher mean.

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

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

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

Riemannian metric needed, or at least a connection.

Idea of Riemannian metrics on shapes

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).

Paradigm shift with deep learning

High-dimensions + large data

- Regression (analysis): find maps $F : \mathbb{R}^d \to \mathbb{R}^n$ with $d >> n$, from large number of points.
- Generative modeling (synthesis): find maps $F : \mathbb{R}^n \to \mathbb{R}^d$ with $d >> n$, from large number of points.

- Shapes are replaced by probability distributions in high-dimension.
- Some mathematical tools can be re-used to understand learning dynamics.

Contents

1. Motivations from imaging and learning
2. How to build Riemannian metric in infinite dimensions?
3. Right-invariant metrics
4. Dynamic formulation of optimal transport
5. Unbalanced Optimal Transport
6. Gradient flows
7. On global convergence of ResNets
8. The Camassa-Holm equation as an incompressible Euler equation

Infinite dimensional spaces

- The space of maps $M \to N$.
- The space of diffeomorphisms $\mathrm{Diff}(M)$.
- The space of densities $\mathrm{Dens}(M)$.

We will use different Hilbert/Riemannian metrics: for instance:

- Example: $L^2(M,\mathbb{R}) = \{f : M \to N\,; \int_M \|f(x)\|^2 d\mu(x)\}$
- Riemannian metric on $L^2(M,N)$, $g(f)(\delta f,\delta f) = \int_M g_N(f(x))(\delta f(x),\delta f(x)) d\mu(x)$.
- $H^s(M,N)$.
- Shape spaces: $\mathrm{Emb}(S_1,\mathbb{R}^2) / \mathrm{Diff}(S_1)$.

Some definitions

Definition (Natural operations on groups)

Let G be a Lie group.

$$\mathrm{Ad}_g(m) = gmg^{-1}\,. \tag{1}$$

Note that

$$\mathrm{Ad}_{g_1}\mathrm{Ad}_{g_2} = \mathrm{Ad}_{g_1 g_2}\,. \tag{2}$$

and $\mathrm{Ad}_g : T_{Id}G \mapsto T_{Id}G$.

Definition (Lie algebra and Lie bracket)

The tangent space at $Id \in G$, is called the Lie algebra. It is equipped with the Lie bracket:

$$[\zeta, \xi] := \mathrm{ad}_\zeta\,\xi := \frac{d}{dt}\mathrm{Ad}_{g(t)}(\xi)\,. \tag{3}$$

where $\frac{d}{dt}_{t=0} g(t) = \zeta$ at $g(t=0) = \mathrm{Id}$.

Left action

Definition (Left action)

A left action of the group G on a space M is a map $G \times M \to M$ satisfying

1. $Id \cdot q = q$ *for* $q \in M$.
2. $g_2 \cdot (g_1 \cdot M) = (g_2 g_1) \cdot q$.

- The group on itself by multiplication, left multiplication or right multiplication by the inverse.
- $GL_n(\mathbb{R})$ acting by multiplication on $\mathbb{R}^n$: $(M, v) \mapsto Mv$.
- The group of diffeomorphisms Diff of $\mathbb{R}^d$ on $n-$points (landmarks) $(\mathbb{R}^d)^n$: $(\varphi, (x_1, \dots, x_n)) \mapsto (\varphi(x_1), \dots, \varphi(x_n))$.
- Diff on functions (0-forms): $(\varphi, I) \mapsto I \circ \varphi^{-1}$.
- Diff on densities (n-forms): $(\varphi, \rho) \mapsto \mathrm{Jac}(\varphi^{-1}) \rho \circ \varphi^{-1}$.

Some definitions

Definition (Infinitesimal action)

Let $\xi \in T_{\mathrm{Id}} G$ and $q \in M$. Consider $g(t) \in G$ such that $g(0) = \mathrm{Id}$ and $\dot{g}(0) = \xi$ and define:

$$\xi \cdot q \overset{\text{def.}}{=} \left.\frac{d}{dt}\right|_{t=0} g(t) \cdot q \,. \tag{4}$$

Examples

1. For matrices, $\xi \cdot M = \xi M$.
2. For $SO(n)$ acting on the sphere, the Lie algebra is skew-symmetric matrices. $\xi \cdot v = \xi v$.
3. For densities: $v \cdot \rho = -\operatorname{div}(\rho v)$.
4. For functions: $v \cdot I = -\langle \nabla I, v \rangle$.

A concrete example

$G = \mathrm{Diff}(\mathbb{R}^2)$ and $M = \mathrm{Emb}(S_1, \mathbb{R}^2) / \mathrm{Diff}(S_1)$.

A concrete example

$G = \mathrm{Diff}(\mathbb{R}^2)$ and $M = \mathrm{Emb}(S_1, \mathbb{R}^2) \big/ \mathrm{Diff}(S_1)$.

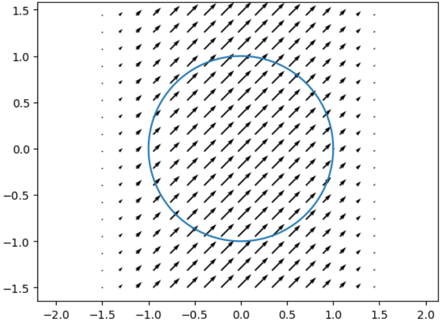


Figure – A vector field and a curve

A concrete example

$G = \mathrm{Diff}(\mathbb{R}^2)$ and $M = \mathrm{Emb}(S_1, \mathbb{R}^2) / \mathrm{Diff}(S_1)$.

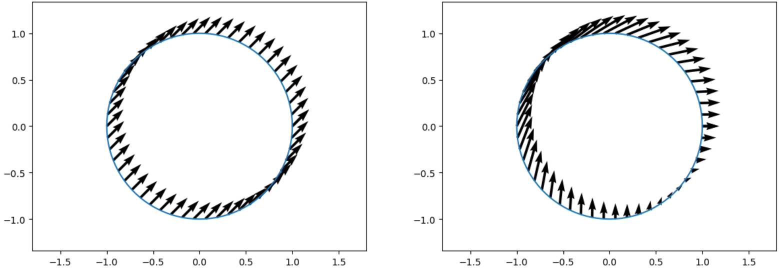


Figure – Two different vector fields lead to the same infinitesimal action

Momentum maps

Moment/Momentum map

Let G act on the left on M, then G acts on TM and T^*M by

$$g \cdot (x, v) = (g \cdot x, d[g \cdot x](v)) \tag{5}$$

$$g^{-1*} \cdot p = d[g \cdot x]^{-\top}(p) \,. \tag{6}$$

Given $(x, p) \in T^*M$, define $v \mapsto \langle p, v \cdot x \rangle$.Riesz theorem:

$$\begin{cases} J: & T^*M \to (T_e G)^* \\ & (x, p) \mapsto \langle J(p, x), v \rangle := \langle p, v \cdot x \rangle \,. \end{cases} \tag{7}$$

$$J(g^{-1*} \cdot p, g \cdot x) = \mathrm{Ad}^*_{g^{-1}} J(p, x) \,. \tag{8}$$

- $GL_n(\mathbb{R})$ acting by multiplication on $\mathbb{R}^n$: $J(p, v) = pv^\top$.
- Action on densities: $J(P, \rho) = \nabla P \rho$.
- Action on images: $J(\rho, I) = \nabla I \rho$.

Riemannian submersion

Let (M, g_M) and (N, g_N) be two Riemannian manifolds and $f : M \mapsto N$.

Definition

The map f is a Riemannian submersion if f is a submersion and for any $x \in M$, the map $df_x : \mathrm{Ker}(df_x)^\perp \mapsto T_{f(x)}N$ is an isometry.

- $\mathrm{Vert}_x := \mathrm{Ker}(df(x))$ is the vertical space.
- $\mathrm{Hor}_x \stackrel{\text{def.}}{=} \mathrm{Ker}(df(x))^\perp$ is the horizontal space.
- Geodesics on N can be lifted "horizontally" to geodesics on M.

Horizontal lift of geodesics

Consider $\pi : M \mapsto N$ a Riemannian submersion. Then, for every geodesic $y(t) \in N$, there exists a unique geodesic $x(t) \in M$ of minimal length such that $\pi(x(t)) = y(t)$.

Other property

Example of application

- Some operations are stable under submersions.
- Polar factorisation like in the case of groups.

Theorem (O'Neill's formula)

Let X, Y be two orthonormal vector fields on M with horizontal lifts $\tilde{X}$ and $\tilde{Y}$, then

$$K_N(X, Y) = K_M(\tilde{X}, \tilde{Y}) + \frac{3}{4} \| \operatorname{vert}([\tilde{X}, \tilde{Y}]) \|_M^2 \,. \tag{9}$$

How to build such submersions ?

1. Consider G left acting on M transitively with a surjective infinitesimal action.
2. For each $q \in M$, choose a metric on $T_{\mathrm{Id}}G$: $g(q)(\xi, \xi)$.
3. Define

$$g_M(q)(v, v) \stackrel{\text{def.}}{=} \min_{\xi \in T_{Id}G} g(q)(\xi, \xi) \text{ under the constraint } v = \xi \cdot q \,, \quad (10)$$

Proposition

1. *Choose a point q_0 and the map $g \mapsto g \cdot q_0$ is a Riemannian submersion. Work out the metric on the group?*
2. *On the isotropy subgroup G_{q_0}, the induced metric is right-invariant (w.r.t. G_{q_0}).*

Principal fiber bundle

Definition (Principal fiber bundle)

A principal fiber bundle is the data of P, B two manifolds and G a group acting freely (and transitively on the fibers of π) on P and a map $\pi : P \to B$ which satisfy these relations:

- $\pi(g \cdot x) = \pi(x)$.
- *Locally, $P \approx G \times U$ with $U \subset B$ and π is the projection on U.*

Example: $P = \mathbb{C} \times \mathbb{C}$ and the group is $S_1 = \{z \in \mathbb{C}\,;\, |z| = 1\}$. The action is $(z_1, z_2), e^{i\theta} \mapsto (e^{i\theta} z_1, e^{i\theta} z_2)$.

Principal fiber bundle

Assume that the action of G is via isometries for the metric g on P. Then, the projection π can be made into a Riemannian submersion: for a given $y = \pi(x)$ and $v_y \in T_y B$

$$\|v_y\|^2 = \inf_{v_x \in T_x P} \{\|v_x\|^2\,;\, d\pi(x)(v_x) = v_y\}\,, \qquad (11)$$

which does not depend on the choice of x in the fiber.

Contents

1. Motivations from imaging and learning
2. How to build Riemannian metric in infinite dimensions?
3. Right-invariant metrics
4. Dynamic formulation of optimal transport
5. Unbalanced Optimal Transport
6. Gradient flows
7. On global convergence of ResNets
8. The Camassa-Holm equation as an incompressible Euler equation

Right-invariant metric on groups

Use G a Lie group or at least a group with continuous composition and manifold structure.

Right-invariant metric on groups

Choose an inner product $\langle \cdot, \cdot \rangle$ on $T_{\mathrm{Id}} G$ and $g(q)(\xi, \xi) \stackrel{\text{def.}}{=} \langle \xi, \xi \rangle$.
The metric on G is the associated right-invariant metric.

The distance is given by:

$$d(\varphi_0, \varphi_1)^2 = \inf_{\xi} \{ \int_0^1 \|\xi(t)\|^2 dt \, ; \, \psi(1) \circ \varphi_0 = \varphi_1 \, ; \, \partial_t \psi = \xi(t, v(t,x)) \} \, . \tag{12}$$

Geodesic flow for right-invariant metric

Calculus of variations:

- Write variations for φ: $\delta\varphi$ vanishing at $t = 0, 1$.
- In Eulerian coordinates, $\delta\varphi \circ \varphi^{-1} = w$.
- Define v as $\partial_t \varphi = v \circ \varphi$.
- Compute

$$\partial_{ts}\varphi = \delta v \circ \varphi + Dv(\varphi)(w \circ \varphi) = \partial_t w \circ \varphi + Dw(\varphi)(v \circ \varphi)\,.$$

- We have $\delta v = \partial_t w + Dw(v) - Dv(w) = \partial_t w - \mathrm{ad}_v\, w$.
- Insert into the Lagrangian: $\int \langle Lv, \partial_t w - \mathrm{ad}_v\, w\rangle dt$
- Integrate by parts, obtain Euler-Arnold-Poincaré equation:

$$\partial_t m + \mathrm{ad}_v^*\, m = 0 \quad \text{where } m = Lv\,.$$

Reference: *Sur la géométrie différentielle des groupes de lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits*, Arnold, 1966.

Examples of Euler-Arnold-Poincaré

1. Incompressible Euler equation, L^2 norm on divergence free vector fields.

$$\partial_t v + \nabla_v v = -\nabla p \,. \tag{13}$$

2. Camassa-Holm equation, 1D, H^1 norm on vector fields. Model for shallow water equation, wave breaking, blow-up in finite time.

$$\partial_t u - \frac{1}{4}\partial_{txx} u\, u + 3\partial_x u\, u - \frac{1}{2}\partial_{xx} u\, \partial_x u - \frac{1}{4}\partial_{xxx} u\, u = 0 \tag{14}$$

3. Korteweg-de-Vries equation: L^2 metric on the group $\mathrm{Diff}(S_1) \ltimes \mathbb{R}$.

$$\partial_t u = -3u\partial_x u - a\partial_{xxx} u \,. \tag{15}$$

Khesin, Misiolek,...

Conserved quantities

Noether's theorem

For $L(q,\dot{q})$ invariant wrt symmetries $\varphi(s) = \mathrm{Id} + s\xi$:

$$L(\varphi(s)q, \frac{d}{dt}\varphi(s)q) = cste \implies \langle \frac{\delta L}{\delta \dot{q}}, \xi(q)\rangle = cste \tag{16}$$

For Euler-Poincaré, $\langle \frac{\delta \ell}{\delta v}(\dot{\varphi}\circ\varphi^{-1}), \dot{\varphi}\circ\varphi^{-1}\rangle_{L^2}$. and the symmetries generate $\xi(\varphi) = d\varphi(\xi)$ with $\xi \in \mathrm{Vec}(M)$.

$$\implies \langle \frac{\delta \ell}{\delta v}(\dot{\varphi}\circ\varphi^{-1}), d\varphi(\xi)\circ\varphi^{-1}\rangle_{L^2}$$

Euler-Poincaré-Arnold integrated form:

$$\mathrm{Ad}^*_{\varphi}\, m(t) = m(0) \text{ or } \mathrm{Ad}^*_{\varphi^{-1}}(m(0)) = m(t)\,. \tag{17}$$

Regularity of composition

Definition

H^s Diffeomorphisms: M is a closed manifold or Euclidean space

$$\mathrm{Diff}^s(\mathbb{R}^d) := \{\mathrm{id} + h\,,\; h \in H^s \textit{ and } \mathrm{id} + h \textit{ invertible.}\} \tag{18}$$

On M it is even simpler: set of C^1 diffeomorphisms (orientation preserving) in H^s.

The following maps are regular for $s > d/2 + 1$, $\varphi \in \mathrm{Diff}^s(M)$, $k \geq 0$:

1. $(f, g) \in H^{s-1} \times H^{s-1} \mapsto fg$ is smooth.
2. $f \in H^{s+k}, \varphi \in \mathrm{Diff}^s(M) \mapsto f \circ \varphi$ are C^k.
3. $\varphi \in \mathrm{Diff}^s(M) \mapsto f \circ \varphi$ is C^k for $f \in H^{s+k}$.
4. $\varphi \in \mathrm{Diff}^{s+k}(M) \mapsto \varphi^{-1}$ is C^k.
5. $(\delta\varphi, \varphi) \in H^s \times \mathrm{Diff}^s(M) \mapsto \delta\varphi \circ \varphi^{-1}$ only C^0!

Reference: *On the regularity of the composition of diffeomorphisms*, Inci, Kappeler, Topalov, 2012.

Local existence in time

Example of incompressible Euler: Rewrite in Lagrangian coordinates.

$$\ddot{\varphi} = (\partial_t v + \nabla_v v) \circ \varphi = -\nabla p \circ \varphi \,. \tag{19}$$

Taking divergence, we get

$$\operatorname{div}(\nabla p) = -\operatorname{div}(\nabla_v v) \tag{20}$$

$$-\nabla p = \nabla \Delta^{-1} \nabla_v v \tag{21}$$

$$\tag{22}$$

In the end:

$$\ddot{\varphi} = Q_\varphi(\dot{\varphi}) \text{ with } Q_\varphi(f) = Q(f \circ \varphi^{-1}) \circ \varphi$$

An ODE in Hilbert spaces!

Smoothness of Q_φ

Smoothness wrt φ

The map $\varphi \mapsto Q_\varphi$ for $\varphi \in H^s$ is smooth.

Proof.

Set $\partial_t \varphi = w \circ \varphi$, compute

$$\frac{d}{dt} Q_\varphi = (\nabla_w (Q(\cdot)) - Q(\nabla_w(\cdot)))_\varphi = [\nabla_w, Q]_\varphi \,. \tag{23}$$

Two important points:

1. Same order than Q.
2. Can iterate this formula: note that $w = \delta\varphi \circ \varphi^{-1} \implies$ second term is also a conjugated differential operator applied to $\delta\varphi$.

□

Reference: *Groups of Diffeomorphisms and Fluid Motion: Reprise*, D. Ebin., 2015

No loss of regularity

What about if the initial data is more regular, local existence in H^s.

No loss, no gain

Let $T^*(n)$ the possible blow-up time for the geodesic. Then,

$$T^*(k) = T^*(n) \quad \text{for } d/2 + 1 < k \leq n\,. \tag{24}$$

Proof.

Use right-invariance to get information on the derivative: $\varphi(t) \circ \psi(s)$ is a geodesic in t for fixed s. Then,

$$\mathrm{Fl}_t(\varphi(0), v(0)) \circ \psi(s) = \mathrm{Fl}_t(\varphi(0) \circ \psi(s), v(0) \circ \psi(s))\,, \tag{25}$$

Differentiating in s leads to

$$d\varphi(t)(w) = d\,\mathrm{Fl}(\varphi(0), v(0))(d\varphi(0)w, dv(0)w)\,. \tag{26}$$

□

Ref: *Semi-invariant Riemannian metrics in hydrodynamics*, Bauer, Modin, 2020

Reproducing kernel Hilbert spaces

Reproducing Kernel Hilbert Spaces (RKHS)

Consider $H \subset \mathcal{F}(\Omega, \mathbb{R})$ Hilbert Space such that $H \hookrightarrow C^0(\Omega)$.

- $\delta_x \in H^*$: Map $\mu \mapsto \int \delta_x d\mu(x) \in H^*$.
- $\langle \delta_x, v \rangle = v(x) =: \langle k(x, \cdot), v \rangle_H$.

- $k(x, y) = e^{-\|x-y\|^2/\sigma^2}$ Gaussian kernel or sums of it.
- Used also to compare probability measures: called Maximum Mean Discrepancy (MMD)

$$\|\mu - \nu\|_{H^*} = \sup_{f \in B_H(0,1)} \langle f, \mu - \nu \rangle . \tag{27}$$

- For Sobolev spaces order $k > d/2$, so-called Matern kernel. The differential operator is similar to $(\mathrm{id} - \alpha\Delta)^k$.

Trouvé's group: start from vector fields

Data:

1. D a domain in $\mathbb{R}^d$.
2. $V \hookrightarrow C^1(D, \mathbb{R}^d)$ a Hilbert space of vector fields.
3. Let $\xi \in L^2([0,1], V)$. Denote $\mathrm{Fl}_1(\xi)$

$$\partial_t \varphi(t,x) = \xi(t, \varphi(t,x)) \tag{28}$$

$$\varphi(0,x) = x \;\; \forall x \in D\,. \tag{29}$$

Definition

The group $\mathcal{G}_V$ is defined by

$$\mathcal{G}_V \stackrel{\text{def.}}{=} \{\varphi(1) \,:\, \exists\, \xi \in L^2([0,1], V) \text{ s.t. } \mathrm{Fl}_1(\xi)\}\,, \tag{30}$$

and is a complete metric space with

$$\mathrm{dist}(\psi_1, \psi_0)^2 = \inf \left\{ \int_0^1 \|\xi\|_V^2 \,\mathrm{d}t \,:\, \xi \in L^2([0,1], V) \text{ s.t. } \psi_1 = \mathrm{Fl}_1(\xi) \circ \psi_0 \right\} \tag{31}$$

Sobolev group of diffeomorphisms

Question

When is $\mathcal{G}_V$ a Riemannian manifold (of infinite dimensions) ?
(Locally modelled on a Hilbert space + smoothness of the metric)

Theorem

Let $n > d/2 + 1$. Define $(G_{H^n(\mathbb{R}^d)}) = \{Id + f \mid f \in H^n(\mathbb{R}^d)\} \cap C^1_{Diff(\mathbb{R}^d)}$ with the right-invariant metric H^n. Then,

1. $(G_{H^n(\mathbb{R}^d)})_0 = \mathcal{G}_{H^n(\mathbb{R}^d)}$,
2. $(G_{H^n(\mathbb{R}^d)})_0$ *is a complete Riemannian manifold: Initial and boundary value problems have (global) solutions.*

Open question: are there other RKHS for which the group is a manifold?

Ideas of proof

Two ingredients:

1. Smoothness of the metric (Ebin-Marsden 1970).

$$TG^n \times_{G^n} TG^n \to H^n(\mathbb{R}^d, \mathbb{R}^d) \times H^n(\mathbb{R}^d, \mathbb{R}^d) \to \mathbb{R}$$

$$\underbrace{\underbrace{(\varphi, X, Y) \mapsto (X \circ \varphi^{-1}, Y \circ \varphi^{-1})}_{\text{only continuous}} \mapsto \langle X \circ \varphi^{-1}, Y \circ \varphi^{-1} \rangle_{H^n}}_{\textit{smooth}!}$$

2. Direct method of calculus of variations. Minimize a lower semicontinuous functional under weakly closed constraints.

- *Fredholm properties of Riemannian exponential maps on diffeomorphism groups* (Misiolek and Preston), Inventiones Math.
- On Completeness of Groups of Diffeomorphisms. [JEMS, with M. Bruveris]

Right-invariant metric and left action

Recall $G \times Q \to Q$ a left action + right-invariant metric $\implies$ metric on Q.

Write geodesic equation using Lagrange multipliers:

$$\mathcal{L}(v,p,q) = \int_0^1 \frac{1}{2}\|v\|_V^2 - \langle p, v\cdot q\rangle \, dt \tag{32}$$

Variations in (p,q) implies

$$\begin{cases} \dot q = v\cdot q \\ \dot p = -dv^\top(q)(p) \\ Lv = J(p,q)\,. \end{cases} \tag{33}$$

where $J(p,q)$ denotes the linear form on V induced by $v \mapsto \langle p, v(q)\rangle$.

$$\frac{d}{dt} J(p,q) + \mathrm{ad}_v^*(J(p,q)) = 0\,. \tag{34}$$

Example on points

- $Q = (\mathbb{R}^d)^N$.
- $G = \mathrm{Diff}(M)$.
- The action is by composition. $(\varphi(q_i))_{i=1,\ldots,N}$.
- Recall k is the inverse of L.

Geodesic equations are

$$\begin{cases} \dot{q}_i = \sum_{j=1}^N k(q_i, q_j) p_j \\ \dot{p}_i = -(\sum_{j=1}^N D_1 k(q_i, q_j) p_j)^\top p_i \, . \end{cases} \tag{35}$$

Hamiltonian equations for

$$H(p, q) = \frac{1}{2} \sum_{i,j=1}^N \langle p_i, k(q_i, q_j) p_j \rangle \, .$$

Example on densities

- $Q = \mathrm{Dens}(M)$.
- $G = \mathrm{Diff}(M)$.
- The action is by pushforward $\varphi_\sharp(\rho)$.

The metric is

$$g_R(\rho)(\delta\rho, \delta\rho) = \frac{1}{2} \iint \langle \nabla P(x), k(x,y) \nabla P(y) \rangle d\rho(x) d\rho(y)\,, \tag{36}$$

with

$$\delta\rho = \mathrm{div}(\rho K \star (\rho \nabla P))\,.$$

Geodesic equations are

$$\begin{cases} \partial_t \rho + \mathrm{div}(v\rho) = 0 \\ \partial_t P + \langle \nabla P, v \rangle = 0 \\ v = K \star (\nabla P \rho)\,. \end{cases} \tag{37}$$

Hamiltonian equations for

$$H(\rho, P) = \frac{1}{2} \iint \langle \nabla P(x), k(x,y) \nabla P(y) \rangle d\rho(x) d\rho(y)\,.$$

$\Longrightarrow$ the order of the metric is $\deg(L) - 1$.

Beware of the order

Non-degenerate/degenerate right-invariant metrics on group of diffeomorphisms

Theorem (Mumford, Michor, 2005)

For $d \geq 1$, the H^{div} right-invariant distance on the group of diffeomorphisms is non-degenerate.

$$\|v\|^2_{\mathrm{div}} = \int_{\mathbb{R}^d} \|v(x)\|^2 + |\operatorname{div}(v)(x)|^2 dx\,. \tag{38}$$

Theorem (Maor, Jerrard, 2018)

For $d \geq 1$, the H^s right-invariant distance on the group of diffeomorphisms is degenerate if $s < 1$.
That is, between any two connected diffeomorphisms, there exists a path of arbitrarily small length.

Contents

1. Motivations from imaging and learning
2. How to build Riemannian metric in infinite dimensions?
3. Right-invariant metrics
4. **Dynamic formulation of optimal transport**
5. Unbalanced Optimal Transport
6. Gradient flows
7. On global convergence of ResNets
8. The Camassa-Holm equation as an incompressible Euler equation

Optimal transport

Wasserstein metric tensor

- Consider $G = \mathrm{Diff}(\mathbb{R}^d)$ endowed with the L^2 metric $\int_{\mathbb{R}^d} \|\delta\varphi(x)\|^2 dx$.
- Consider $S = \mathrm{SDiff}(\mathbb{R}^d)$ the subgroup of volume preserving diffeomorphisms.
- Remark that S is the isometry subgroup of G.

$\Longrightarrow$ $\mathrm{Diff}(\mathbb{R}^d) \to \mathrm{Diff}(\mathbb{R}^d)/\mathrm{SDiff}(\mathbb{R}^d)$ is a Riemannian submersion with:

$$g(\rho)(v,v) \stackrel{\text{def.}}{=} \int_{\mathbb{R}^d} |v(x)|^2 \rho(x)\,\mathrm{d}x\,. \tag{39}$$

- $\mathrm{Diff}(\mathbb{R}^d)/\mathrm{SDiff}(\mathbb{R}^d) \approx \mathrm{Dens}_1(\mathbb{R}^d)$.
- A more computational way is to set the metric on $\mathrm{Dens}_1(\mathbb{R}^d)$ as

$$g_W(\rho)(\delta\rho,\delta\rho) = \inf_v \left\{ \int_{\mathbb{R}^d} |v(x)|^2 \rho(x)\,\mathrm{d}x\,;\ \delta\rho = -\operatorname{div}(\rho v) \right\}\,. \tag{40}$$

Optimal transport

The Wasserstein metric tensor

The Riemannian(-like) metric tensor at a density ρ is

$$\int_M \|\nabla \Delta_\rho^{-1} \delta\rho\|^2 d\rho(x)\,, \tag{41}$$

where $\Delta_\rho^{-1}(\delta\rho)$ is the unique solution to the elliptic equation:

$$\operatorname{div}(\rho \nabla P) = \delta\rho\,. \tag{42}$$

It is also equal to

$$\int_M \Delta_\rho^{-1}(\delta\rho)\delta\rho dx\,. \tag{43}$$

$\Longrightarrow$ it is an H^{-1} type metric on the space of densities.

A geometric picture: Otto's Riemannian submersion

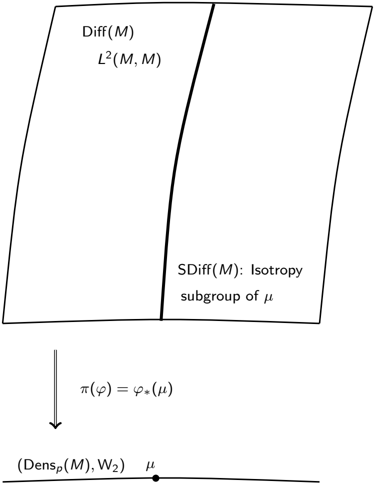


Figure – A Riemannian submersion: SDiff(M) as a Riemannian submanifold of $L^2(M, M)$: Incompressible Euler equation on SDiff(M)

A pre-formulation of the polar factorization

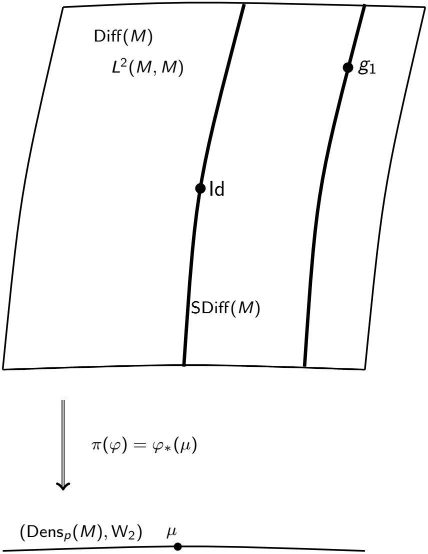


Figure – A Riemannian submersion: SDiff(M) as a Riemannian submanifold of $L^2(M, M)$: Incompressible Euler equation on SDiff(M)

A pre-formulation of the polar factorization

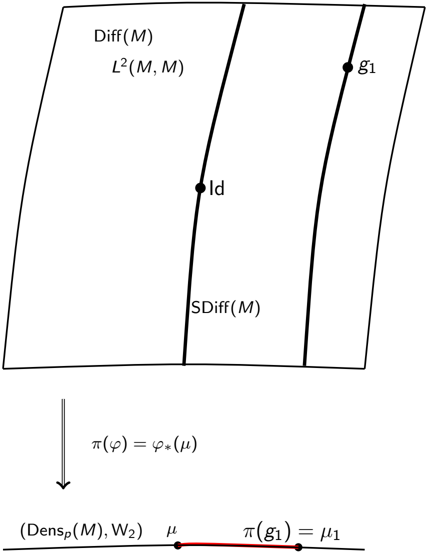


Figure – A pre polar factorization

A pre-formulation of the polar factorization

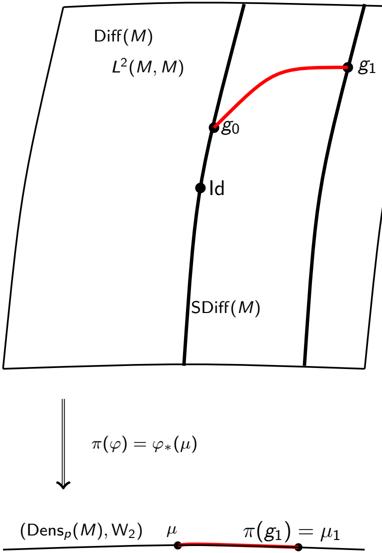


Figure – Polar factorization: $g_0 = \arg\min_{g \in \mathrm{SDiff}} \|g_1 - g\|_{L^2}$

Reminders: Dynamic formulation (Benamou-Brenier)

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

$$\inf \mathcal{E}(v) = \frac{1}{2}\int_0^1 \int_M |v(x)|^2 \rho(x)\,\mathrm{d}x\,\mathrm{d}t \;, \tag{44}$$

s.t.

$$\begin{cases} \dot\rho + \nabla\cdot(v\rho) = 0 \\ \rho(0) = \mu_0 \text{ and } \rho(1) = \mu_1 \,. \end{cases} \tag{45}$$

Reminders: Dynamic formulation (Benamou-Brenier)

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

$$\inf \mathcal{E}(v) = \frac{1}{2}\int_0^1 \int_M |v(x)|^2 \rho(x)\, \mathrm{d}x\, \mathrm{d}t \;, \tag{44}$$

s.t.

$$\begin{cases} \dot\rho + \nabla\cdot(v\rho) = 0 \\ \rho(0) = \mu_0 \text{ and } \rho(1) = \mu_1 \,. \end{cases} \tag{45}$$

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

$$\inf \mathcal{E}(m) = \frac{1}{2}\int_0^1 \int_M \frac{|m(x)|^2}{\rho(x)}\, \mathrm{d}x\, \mathrm{d}t \,, \tag{46}$$

s.t.

$$\begin{cases} \dot\rho + \nabla\cdot m = 0 \\ \rho(0) = \mu_0 \text{ and } \rho(1) = \mu_1 \,. \end{cases} \tag{47}$$

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

Reminders: Dynamic formulation (Benamou-Brenier)

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

$$\inf \mathcal{E}(v) = \frac{1}{2}\int_0^1 \int_M |v(x)|^2 \rho(x)\, \mathrm{d}x\, \mathrm{d}t \ , \tag{44}$$

s.t.

$$\begin{cases} \dot\rho + \nabla \cdot (v\rho) = 0 \\ \rho(0) = \mu_0 \text{ and } \rho(1) = \mu_1 \,. \end{cases} \tag{45}$$

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

$$\inf \mathcal{E}(m) = \frac{1}{2}\int_0^1 \int_M \frac{|m(x)|^2}{\rho(x)}\, \mathrm{d}x\, \mathrm{d}t \,, \tag{46}$$

s.t.

$$\begin{cases} \dot\rho + \nabla \cdot m = 0 \\ \rho(0) = \mu_0 \text{ and } \rho(1) = \mu_1 \,. \end{cases} \tag{47}$$

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.

Geodesic flow

The geodesic are

$$\begin{cases} \partial_t \rho + \operatorname{div}(\rho \nabla P) = 0 \\ \partial_t P + \frac{1}{2}\|\nabla P\|^2 = 0\,. \end{cases} \tag{48}$$

Second equation does not depend on ρ!
Hamiltonian equation for

$$H(\rho, P) = \frac{1}{2}\int_{\mathbb{R}^d} \|\nabla P\|^2 \rho(x) dx\,.$$

Note that it is linear in ρ.

Contents

1. Motivations from imaging and learning
2. How to build Riemannian metric in infinite dimensions?
3. Right-invariant metrics
4. Dynamic formulation of optimal transport
5. Unbalanced Optimal Transport
6. Gradient flows
7. On global convergence of ResNets
8. The Camassa-Holm equation as an incompressible Euler equation

Unbalanced optimal transport

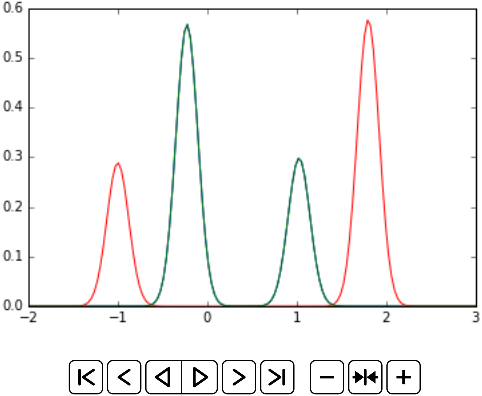


Figure – Optimal transport between bimodal densities

Unbalanced optimal transport

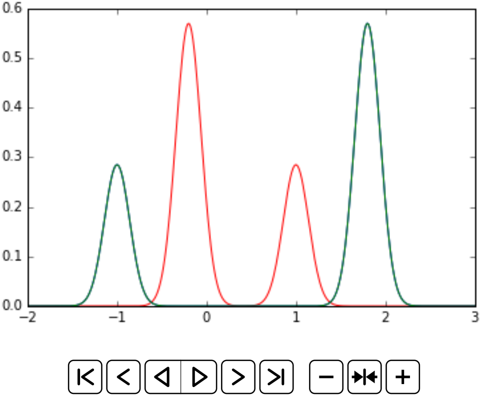


Figure – Another transformation

Two possible directions

Pros and cons:

- Extend static formulation: Frogner et al.

$$\min \lambda KL(\mathrm{Proj}^1_* \gamma, \rho_1) + \lambda KL(\mathrm{Proj}^2_* \gamma, \rho_2) + \int_{M^2} \gamma(x,y) d(x,y)^2 \,\mathrm{d}x\,\mathrm{d}y \quad (49)$$

  Good for numerics, but is it a distance ?

- Extend dynamic formulation: on the tangent space of a density, choose a metric on the transverse direction.
  Built-in metric property but does there exist a static formulation ?

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.

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

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.

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

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:

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

Convex reformulation

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

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

The Wasserstein-Fisher-Rao metric:

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

- 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

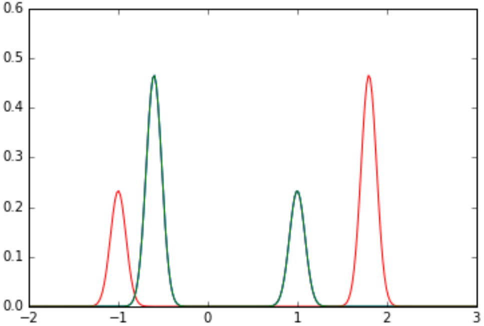


Figure – WFR geodesic between bimodal densities

An other Hamilton-Jacobi equation

Fenchel-Rockafellar theorem implies

$$\mathrm{WF}(\mu,\nu) = \sup_{\varphi} \langle \varphi(1), \mu \rangle - \langle \varphi(0), \nu \rangle \tag{50}$$

where $\varphi \in C^1(M)$ satisfies

$$\partial_t \varphi + \frac{1}{2}\left(|\nabla \varphi|^2 + \frac{\varphi^2}{\delta^2} \right) \leq 0\,,$$

together with the generalized continuity equation

$$\partial_t \mu + \mathrm{div}(\mu \nabla \varphi) = \mu \varphi\,. \tag{51}$$

Question: Can we remove time from the optimization by finding a Kantorovich formulation?

From dynamic to static

Group action

Mass can be moved and changed: consider $m(t)\delta_{x(t)}$. Lagrangian coordinates: $(\varphi(t, x_0), \lambda(t, x_0)m_0)$

From dynamic to static

Group action

Mass can be moved and changed: consider $m(t)\delta_{x(t)}$. Lagrangian coordinates: $(\varphi(t, x_0), \lambda(t, x_0)m_0)$

Infinitesimal action

$$\dot{\rho} = -\nabla \cdot (v\rho) + \alpha\rho \Leftrightarrow \begin{cases} \dot{x}(t) = v(t, x(t)) \\ \dot{m}(t) = \mu(t, x(t)) \end{cases}$$

From dynamic to static

Group action

Mass can be moved and changed: consider $m(t)\delta_{x(t)}$. Lagrangian coordinates: $(\varphi(t,x_0), \lambda(t,x_0)m_0)$

Infinitesimal action

$$\dot\rho = -\nabla\cdot(v\rho) + \alpha\rho \Leftrightarrow \begin{cases} \dot x(t) = v(t,x(t)) \\ \dot m(t) = \mu(t,x(t)) \end{cases}$$

A cone metric

$$\mathrm{WF}^2(x,m)\left((\dot x,\dot m),(\dot x,\dot m)\right) = \frac{1}{2}\left(m\dot x^2 + \frac{\dot m^2}{m}\right),$$

Change of variable: $r^2 = m$...

Riemannian cone

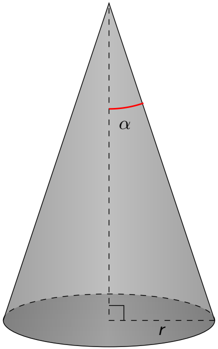
Definition

Let (M, g) be a Riemannian manifold. The cone over (M, g) is the Riemannian manifold $(M \times \mathbb{R}_+^*, r^2 g + \mathrm{d}r^2)$.

Riemannian cone

Definition

Let (M, g) be a Riemannian manifold. The cone over (M, g) is the Riemannian manifold $(M \times \mathbb{R}_+^*, r^2 g + \mathrm{d}r^2)$.



For $M = S_1(r)$, radius $r \leq 1$. One has $\sin(\alpha) = r$.

Geometry of a cone

- The distance:

$$d((x_1, m_1), (x_2, m_2))^2 =$$
$$m_2 + m_1 - 2\sqrt{m_1 m_2}\cos\left(\frac{1}{2}d_M(x_1, x_2) \wedge \pi\right) . \quad (52)$$

- $M = \mathbb{R}$ then $(x, m) \mapsto \sqrt{m}e^{ix/2} \in \mathbb{C}$ local isometry.

Corollary

If (M, g) has sectional curvature greater than 1, then $(M \times \mathbb{R}_+^*, m\, g + \frac{1}{4m}\,\mathrm{d}m^2)$ has non-negative sectional curvature. For X, Y two orthornormal vector fields on M,

$$K(\tilde{X}, \tilde{Y}) = (K_g(X, Y) - 1) \quad (53)$$

where K and K_g denote respectively the sectional curvatures of $M \times \mathbb{R}_+^*$ and M.

Visualize geodesics for $r^2 g + \mathrm{d}r^2$

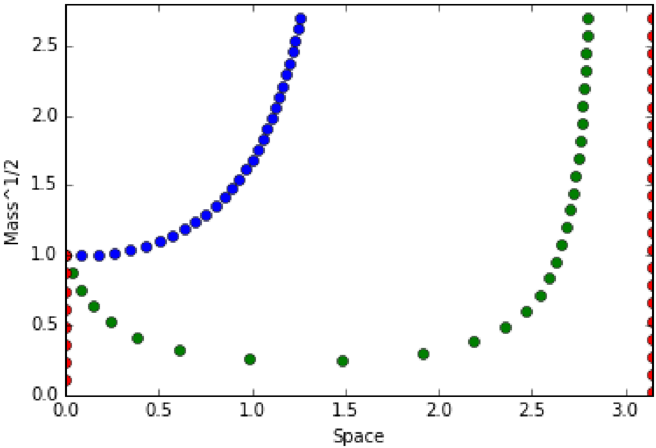


Figure – Geodesics on the cone

Generalization of Otto's Riemannian submersion

The left group action:

$$\pi : \left(\mathrm{Diff}(M) \ltimes C^\infty(M, \mathbb{R}_+^*)\right) \times \mathrm{Dens}(M) \mapsto \mathrm{Dens}(M)$$
$$\pi\left((\varphi, \lambda), \rho\right) := \varphi_*(\lambda^2 \rho)$$

Group law:

$$(\varphi_1, \lambda_1) \cdot (\varphi_2, \lambda_2) = (\varphi_1 \circ \varphi_2, (\lambda_1 \circ \varphi_2)\lambda_2) \tag{54}$$

Generalization of Otto's Riemannian submersion

The left group action:

$$\pi : \left(\mathrm{Diff}(M) \ltimes C^\infty(M, \mathbb{R}_+^*)\right) \times \mathrm{Dens}(M) \mapsto \mathrm{Dens}(M)$$
$$\pi\left((\varphi, \lambda), \rho\right) := \varphi_*(\lambda^2 \rho)$$

Group law:

$$(\varphi_1, \lambda_1) \cdot (\varphi_2, \lambda_2) = (\varphi_1 \circ \varphi_2, (\lambda_1 \circ \varphi_2)\lambda_2) \tag{54}$$

Theorem

Let $\rho_0 \in \mathrm{Dens}(M)$ *and* $\pi_0 : \mathrm{Diff}(M) \ltimes C^\infty(M, \mathbb{R}_+^*) \mapsto \mathrm{Dens}(M)$ *defined by* $\pi_0(\varphi, \lambda) := \varphi_*(\lambda^2 \rho_0)$*. It is a Riemannian submersion*

$$(\mathrm{Diff}(M) \ltimes C^\infty(M, \mathbb{R}_+^*), L^2(M, M \times \mathbb{R}_+^*)) \xrightarrow{\pi_0} (\mathrm{Dens}(M), \mathrm{WF})$$

(where $M \times \mathbb{R}_+^*$ *is endowed with the cone metric).*

O'Neill's formula: sectional curvature of $(\mathrm{Dens}(M), \mathrm{WF})$ nonnegative if $K_M \geq 1$.

Monge formulation

Monge formulation

$$WF(\rho_0,\rho_1) = \inf_{(\varphi,\lambda)} \left\{ \|(\varphi,\lambda) - (Id,1)\|_{L^2(\rho_0)} \,:\, \varphi_*(\lambda\rho_0) = \rho_1 \right\} \quad (55)$$

Under existence and smoothness of the minimizer, there exists a function $p \in C^\infty(M,\mathbb{R})$ such that

$$(\varphi(x),\lambda(x)) = \exp_x^{\mathcal{C}(M)}\left(\frac{1}{2}\nabla p(x), p(x)\right), \quad (56)$$

Equivalent to Monge-Ampère equation

With $z \stackrel{\text{def.}}{=} \log(1+p)$ one has

$$(1+|\nabla z|^2)e^{2z}\rho_0 = \det(D\varphi)\rho_1 \circ \varphi \quad (57)$$

and

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

A relaxed static OT formulation

Define

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

Theorem (Dual formulation)

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

subject to $\forall (x,y) \in M^2$, $\phi(x) \leq 1\,, \quad \psi(y) \leq 1$ *and*

$$(1-\phi(x))(1-\psi(y)) \geq \cos^2\left(|x-y|/2 \wedge \pi/2\right)$$

The corresponding primal formulation

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

Theorem

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

Contents

1. Motivations from imaging and learning
2. How to build Riemannian metric in infinite dimensions?
3. Right-invariant metrics
4. Dynamic formulation of optimal transport
5. Unbalanced Optimal Transport
6. Gradient flows
7. On global convergence of ResNets
8. The Camassa-Holm equation as an incompressible Euler equation

Gradient flow of the entropy

With respect to the Wasserstein metric:

Minimize w.r.t. v the Lagrangian:

$$\frac{1}{2}\int_{\mathbb{R}^d} \|v(x)\|^2 \rho(x) dx + \frac{\delta \mathcal{F}(\rho)}{\delta \rho}(-\operatorname{div}(\rho v)) \tag{58}$$

Therefore,

$$v\rho = -\rho \nabla \frac{\delta \mathcal{F}}{\delta \rho}$$

For $\mathcal{F}(\rho) = \int_{\mathbb{R}^d} \rho \log(\rho / e^{-V}) dx$

$$\text{Fokker-Planck } \partial_t \rho = \operatorname{div}(\nabla \rho + \rho \nabla V)) = \Delta \rho + \operatorname{div}(\rho \nabla V)\,. \tag{59}$$

Convergence to the equilibrium measure

Fokker-Planck describes the law of the Langevin diffusion $X(t)$

$$dX(t) = -\nabla V(X(t)) + \sqrt{2}dB(t)\,. \tag{60}$$

Since it is a gradient flow, it might converge to a steady state:

$$0 = +\operatorname{div}(\rho(\nabla \log(\rho) + \nabla V)) \implies \rho = e^{-V} \tag{61}$$

Main tool: Geodesic convexity

- Entropy is convex along geodesic. (In fact $\mathrm{Ric}(M) \geq 0$.)
- $\int_M V(x)\rho(x)dx$ is convex along geodesic iif V is convex (along geodesic).

Key tool for convergence

Polyak-Lojasiewicz condition

$$\exists\lambda > 0 \quad \text{s.t. } \lambda(f(x) - f_*) \leq \frac{1}{2}\|\nabla f(x)\|^2 \,, \tag{62}$$

implies exponential convergence.

Example: $\dot{x} = -\nabla f(x)$.

$$\frac{d}{dt}(f(x) - f_*) = -\|\nabla f(x)\|^2 \leq 2\lambda(f(x) - f_*) \,. \tag{63}$$

$$f(x(t)) - f_* \leq (f(x(0)) - f_*)e^{-2\lambda t} \,, \tag{64}$$

- No need for convexity, applies to Riemannian manifolds.
- Only need to measure the length of the gradient.

Reference: *Local conditions for global convergence of gradient flows and proximal point sequences in metric spaces*, Dello Schiavo, Maas, Pedrotti, 2023.

How to prove PL inequality?

Proposition

If $F : H \mapsto \mathbb{R}$ is λ-strongly convex, then F satisfies the PL condition for the constant $\frac{1}{2\lambda}$.

Stability of PL

Let $\varphi : \Omega \to \Omega$ be a C^1 diffeomorphism of the definition domain of f, then $\varphi^* f(y) \triangleq f \circ \varphi(y)$ satisfies $PL(\lambda/M^2)$ if f satisfies $PL(\lambda)$ for $M = \sup_{x \in \Omega} \|d\varphi(x)^{-1}\|$

PL says nothing on convergence of $x(t)$. Add regularity condition such as $\|\nabla f(x)\|^2 \leq \beta(f(x) - f_*)\,, \implies$ convergence towards $x_\star \in \arg\min f$.
Local PL is a much more mild condition:

Local PL

On a metric space (X, d), $F : X \to \mathbb{R}$ satisfies a local PL condition if on any bounded set $B \subset \Omega$, there exists $\lambda(B)$ such that

$$\lambda(B)(f(x) - f_*) \leq \frac{1}{2}\|\nabla f(x)\|^2 \quad \forall x \in B\,, \tag{65}$$

Gradient flow of the entropy

With respect to a right-invariant metric:

Minimize w.r.t. v the Lagrangian:

$$\frac{1}{2}\|v\|_V^2 + \frac{\delta \mathcal{F}(\rho)}{\delta \rho}(-\operatorname{div}(\rho v)) \tag{66}$$

Therefore,

$$Lv = -\rho \nabla \frac{\delta \mathcal{F}}{\delta \rho}$$

For $\mathcal{F}(\rho) = \int_{\mathbb{R}^d} \rho \log(\rho / e^{-V}) dx$

$$\text{SVGD } \partial_t \rho = \operatorname{div}(\rho K \star (\nabla \rho + \rho \nabla V)) \tag{67}$$

Stein variational gradient descent

In general $\nabla\rho$ not defined for Dirac masses!

$$\partial_t \rho = \mathrm{div}(\rho \nabla K \star \rho + K \rho \nabla V) \tag{68}$$

For empirical measures: $\rho = \frac{1}{N}\sum_{i=1}^N \delta_{q_i}$,

$$\dot{q}_i = -\sum_{j=1}^N \nabla K(q_i, q_j) + K(q_i, q_j)\nabla V(q_j)\,. \tag{69}$$

What is the speed of convergence to the equilibrium measure?

References:

- *Scaling Limit of the Stein Variational Gradient Descent: The Mean Field Regime*, Lu, Lu, Nolen, 2019.
- *On the geometry of Stein variational gradient descent*, Duncan, Nüsken, Szpruch, 2022.

Gradient flow of MMD

Interaction energy on $\mathcal{P}(\mathbb{R}^d)$

On $\mathbb{R}^d$, interaction function

$$\mathcal{F}_\nu(\mu) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} (\mu(x) - \nu(x))k(x,y)(\mu(y) - \nu(y))dxdy\,. \tag{70}$$

where $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, $k(x,y) = k(y,x)$ and k is a conditionally positive definite kernel (CPD).

Study the PDE

Gradient flow PDE

$$\partial_t \mu(x) = -\operatorname{div}(\mu(x)\nabla \int_{\mathbb{R}^d} k(x,y)(\mu(y) - \nu(y))dy)\,. \tag{71}$$

Question: long time behaviour of this PDE?

Minimization of a positive definite function

Definition

A kernel k is CPD if it satisfies for all $n \geq 1$, $p_i \in \mathbb{R}$,

$$\sum_{i,j=1,\ldots,n} p_i k(x_i, x_j) p_j \geq 0 \text{ for } x_i \in \mathbb{R}^d \text{ and } \sum_{i=1}^{n} p_j = 0\,. \tag{72}$$

Consequence

$\mathcal{F}_\nu(\mu)$ is nonnegative and 0 iff $\mu = \nu$:

$$\mathcal{F}_\nu(\mu) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (\mu(x) - \nu(x)) k(x,y) (\mu(y) - \nu(y)) dxdy\,. \tag{73}$$

1. Positive definite kernels: $k(x,y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$.
2. Energy distance: metric space of negative type: $k(x,y) = -d(x,y)$ is CPD.
   True for Euclidean spaces, sphere, hyperbolic space.
3. Coulomb kernel: $k(x,y) = \frac{cste}{\|x-y\|^{d-2}}$.

Motivation 1: Relaxation of shallow neural networks

Shallow neural network

$$f_{(\theta_1,\theta_2)}(x) = \theta_1(\sigma(\theta_2(x)))\,, \tag{74}$$

where $\theta_1 \in L(\mathbb{R}^d, \mathbb{R}^{kd})$, $\theta_2 \in L(\mathbb{R}^{kd}, \mathbb{R}^d)$, σ a pointwise nonlinearity.
Optimization via gradient descent on parameters

$$\sum_{i=1}^{n} \|y_i - f_{(\theta_1,\theta_2)}(x_i)\|^2\,. \tag{75}$$

Convex relaxation (Barron 90's), overparametrize by a measure:
$f_\mu(x) = \int \psi_\theta(x) d\mu(\theta)$.

The problem becomes quadratic in μ

$$\arg\min_{\mu} \sum_{i=1}^{n} \|y_i - f_\mu(x_i)\|^2 \tag{76}$$

Corresponding gradient descent: Wasserstein gradient flow.

Example of use in imaging/machine learning

Diffeomorphic surface matching/normalizing flows:
Model:

$$\begin{cases} \partial_t \varphi(t,x) = v(t,\varphi(t,x))\,, \\ \varphi(t=0,x) = x\,. \end{cases} \tag{77}$$

Loss: $D(\mu,\nu)$, relative entropy, MMD...

Minimize the loss without regularization (greedy approach).

- $\varphi(t=1)$ such that $D(\varphi(t=1)_\sharp(\mu_0),\nu)$ is small.
- Done by gradient descent, need to specify a norm on v.
- Simplification: $\psi = \mathrm{Id} + \varepsilon v$ then for $\|v\|^2_{L^2}$, optimizing on v leads formally to:

$$\partial_t \mu = -\operatorname{div}\left(\mu \nabla \frac{\delta D}{\delta \mu}(\mu,\nu)\right) \tag{78}$$

Properties of the energy

Standard convexity

The function $\mathcal{F}_\nu(\mu)$ is convex along standard linear interpolations:

$$t\mu_0 + (1-t)\mu_1\,, \tag{79}$$

but in general rather nonconvex in the Wasserstein geometry!

Exception: $-|x-y|$ in $1d$. Consequence: global convergence in $1d$.

How to extend this result in any dimension?

Deeper question: understand the gradient flow dynamic in terms of the kernel.

Comparison with optimal transport

Let H be a RKHS of smooth functions (e.g. Gaussian kernels),

$$D_H(\mu,\nu) = \sup_{f \in B_1(H)} \langle f, \nu - \mu \rangle = \| K^{1/2}(\mu - \nu) \|_{L^2} \,, \tag{80}$$

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

$$W_1(\mu,\nu) = \sup_{f \in B_1(\mathrm{Lip})} \langle f, \nu - \mu \rangle \,. \tag{81}$$

1. L^2 Wasserstein obtains "global" solutions.
2. Still expensive, entropic regularization...

Comparison with optimal transport

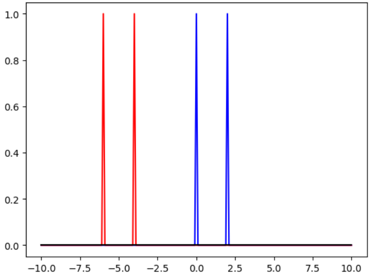


Figure – $\mu = \delta_0 + \delta_2$ to $[T_x]_*(\mu)$ in function of x and OT (green).

Comparison with optimal transport

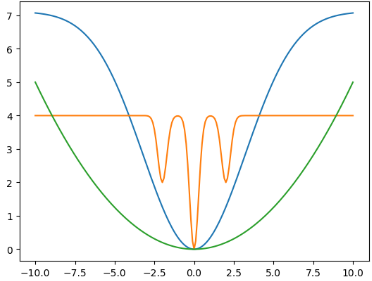


Figure – Kernels distances squared of $\mu = \delta_0 + \delta_2$ to $[T_x]_*(\mu)$ in function of x and OT (green).

Simulations: Choice of kernel matters!

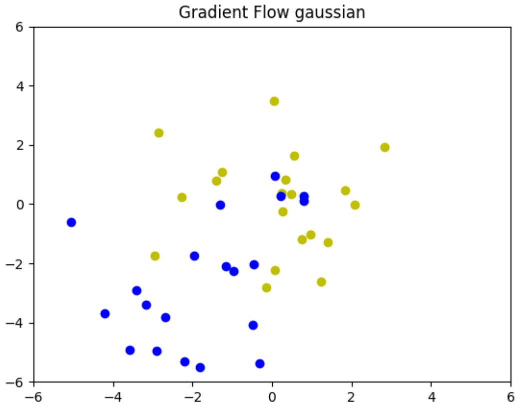


Figure – Gaussian kernel

Simulations: Choice of kernel matters!

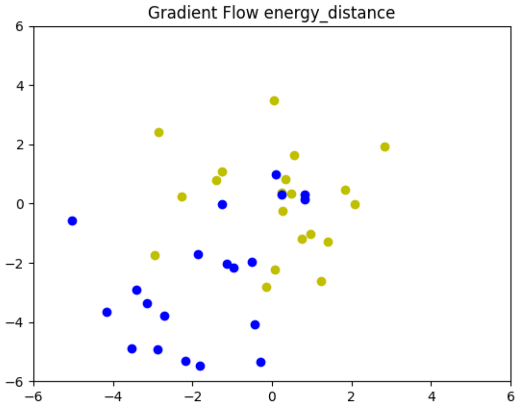


Figure – Energy distance kernel: $-d(x, y)$

Simulations: Choice of kernel matters!

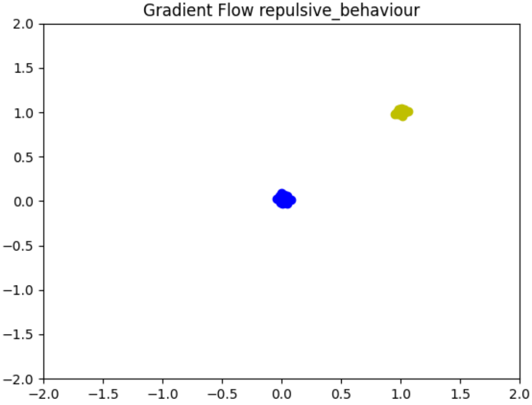


Figure – Diffusion due to repulsivity of the kernel $-d(x, y)$

A first example where PL is local

Study of the Coulomb kernel case Δ^{-1} on a closed connected Riemannian manifold.

$$\mathcal{F}_\nu(\mu) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} (\mu(x) - \nu(x))\Delta^{-1}(\mu(y) - \nu(y))dxdy\,. \tag{80}$$

$$\mathcal{F}_\nu(\mu) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Delta^{-1}(\mu(x) - \nu(x))\Delta\Delta^{-1}(\mu(y) - \nu(y))dxdy\,. \tag{81}$$

PL estimate

Notation: $\varphi_{\mu-\nu}$ the solution of $\Delta\varphi_{\mu-\nu} = \mu - \nu$.

On a closed Riemannian manifold M

If μ is bounded below by $\underline{\mu}$.

- Using the "carré du champ":

$$E_\nu(\mu) = \int_M \|\nabla\varphi_{\mu-\nu}(x)\|^2 d\mathrm{vol}(x)\,. \tag{82}$$

- The Wasserstein gradient norm:

$$\|\nabla_{W_2} E_\nu(\mu)\|^2_{W_2} = \int_M \|\nabla\varphi_{\mu-\nu}(x)\|^2 d\mu(x)\,. \tag{83}$$

PL inequality holds if μ is bounded below.

$$\int_M \|\nabla\varphi_{\mu-\nu}(x)\|^2 d\mathrm{vol}(x) \leq \frac{1}{\underline{\mu}} \int_M \|\nabla\varphi_{\mu-\nu}(x)\|^2 d\mu(x) \tag{84}$$

Stability of the lower bound?

$$\partial_t \mu(t,x) = -\operatorname{div}(\mu \nabla \varphi_{\mu-\nu}) = -\langle \nabla \mu, \nabla \varphi_{\mu-\nu} \rangle - \mu(\Delta \varphi_{\mu-\nu}) \tag{85}$$

$$= \mu(\nu - \mu) \text{ if } \nabla \mu(t,x) = 0\,. \tag{86}$$

The right-hand side is driving to equilibrium at a minimum/maximum.
Analogy: gradient flow of $F(\mu) = \frac{1}{2}\|\mu - \nu\|^2$ wrt Fisher-Rao metric: $\int_M \frac{(\delta\mu)^2}{\mu}$:

$$\dot{\mu} = \mu(\nu - \mu)\,. \tag{87}$$

Stability of the lower bound?

$$\partial_t \mu(t,x) = -\operatorname{div}(\mu \nabla \varphi_{\mu-\nu}) = -\langle \nabla \mu, \nabla \varphi_{\mu-\nu} \rangle - \mu(\Delta \varphi_{\mu-\nu}) \tag{85}$$

$$= \mu(\nu - \mu) \text{ if } \nabla \mu(t,x) = 0\,. \tag{86}$$

The right-hand side is driving to equilibrium at a minimum/maximum. Analogy: gradient flow of $F(\mu) = \frac{1}{2}\|\mu - \nu\|^2$ wrt Fisher-Rao metric: $\int_M \frac{(\delta\mu)^2}{\mu}$:

$$\dot{\mu} = \mu(\nu - \mu)\,. \tag{87}$$

$\Longrightarrow$ if $\mu(t,x)$ is sufficiently smooth, lower and upper bound are stable.

Conclusion

Question of global convergence is reduced to regularity and long time existence of the PDE.

Local existence in Hölder/Sobolev space

Rewrite the system in Lagrangian coordinates as

$$\begin{cases} \partial_t \psi(t,x) = \Delta^{-1}(\mu(t) - \nu) \circ \psi(t) \\ \partial_t \mu(t,x) = -\operatorname{div}(\mu \nabla \Delta^{-1}(\mu(t) - \nu)) \end{cases} \tag{88}$$

Local existence in Hölder/Sobolev space

Rewrite the system in Lagrangian coordinates as

$$\begin{cases} \partial_t \psi(t,x) = \Delta^{-1}(\mu(t) - \nu) \circ \psi(t) \\ \partial_t \mu(t,x) = -\operatorname{div}(\mu \nabla \Delta^{-1}(\mu(t) - \nu)) \end{cases} \tag{88}$$

Introduce $f(t) \coloneqq \mu(t) \circ \psi(t)$, we get

$$\begin{cases} \partial_t \psi(t,x) = \Delta^{-1}_{\psi(t)}(f - \nu \circ \psi(t)) \\ \partial_t f(t,x) = -f(t,x)^2 + f(t,x)\nu \circ \psi(t) \end{cases} \tag{89}$$

where $\Delta^{-1}_{\psi(t)}(h) = \Delta^{-1}(h \circ \psi^{-1}) \circ \psi$.

Local existence in Hölder/Sobolev space

Rewrite the system in Lagrangian coordinates as

$$\begin{cases} \partial_t \psi(t,x) = \Delta^{-1}(\mu(t) - \nu) \circ \psi(t) \\ \partial_t \mu(t,x) = -\operatorname{div}(\mu \nabla \Delta^{-1}(\mu(t) - \nu)) \end{cases} \tag{88}$$

Introduce $f(t) := \mu(t) \circ \psi(t)$, we get

$$\begin{cases} \partial_t \psi(t,x) = \Delta^{-1}_{\psi(t)}(f - \nu \circ \psi(t)) \\ \partial_t f(t,x) = -f(t,x)^2 + f(t,x)\nu \circ \psi(t) \end{cases} \tag{89}$$

where $\Delta^{-1}_{\psi(t)}(h) = \Delta^{-1}(h \circ \psi^{-1}) \circ \psi$.

As shown in the first part, it is smooth in ψ... $\implies$ local existence.
Using potential theory, it is possible to prove long time existence.

Reference: *On the global convergence of Wasserstein gradient flow of the Coulomb discrepancy*, Vialard, Boufadène, 2023.

Contents

1. Motivations from imaging and learning
2. How to build Riemannian metric in infinite dimensions?
3. Right-invariant metrics
4. Dynamic formulation of optimal transport
5. Unbalanced Optimal Transport
6. Gradient flows
7. On global convergence of ResNets
8. The Camassa-Holm equation as an incompressible Euler equation

Supervised learning

Setting: supervised learning.
Goal:

$$G_\star = \min_\theta \mathcal{G}(\theta) := \mathbb{E}[\| f_\theta(X) - Y \|^2] \,, \tag{90}$$

but only acess X, Y through samples: (x_i, y_i).

$$\Longrightarrow \mathcal{L}(\theta) := \min_\theta \frac{1}{N} \sum_{i=1}^{N} \| f_\theta(x_i) - y_i \|^2 \,. \tag{91}$$

- Global convergence of gradient descent on $\mathcal{L}(\theta)$, find $\theta_\star$.
- Generalization, i.e. measure $G(\theta_\star) - G_\star$.

Structure of f_θ.

Define *Single Hidden Layer*

$$\mathrm{SHL}_\theta(x) = \theta_1(\sigma(\theta_2(x)))\,, \tag{92}$$

with $\sigma(x)$ entrywise nonlinearity ($\max(0,x)$).
Deep networks

$$f_\theta(x) = \mathrm{SHL}_{\theta_n} \circ \ldots \circ \mathrm{SHL}_{\theta_1}(x)\,. \tag{93}$$

ResNets, encode residuals

$$f_\theta(x) = (\mathrm{Id} + f_{\theta_n}) \circ \ldots \circ (\mathrm{Id} + f_{\theta_1})(x)\,. \tag{94}$$

Structure of f_θ.

Define *Single Hidden Layer*

$$\mathrm{SHL}_\theta(x) = \theta_1(\sigma(\theta_2(x)))\,, \tag{92}$$

with $\sigma(x)$ entrywise nonlinearity ($\max(0, x)$).
Deep networks

$$f_\theta(x) = \mathrm{SHL}_{\theta_n} \circ \ldots \circ \mathrm{SHL}_{\theta_1}(x)\,. \tag{93}$$

ResNets, encode residuals

$$f_\theta(x) = (\mathrm{Id} + f_{\theta_n}) \circ \ldots \circ (\mathrm{Id} + f_{\theta_1})(x)\,. \tag{94}$$

- Very successful architecture (*Deep Residual Learning for Image Recognition*, [He et al.] 10^5 citations)
- Resembles to an Euler integration scheme for ODE.

Model and loss

Neural ODE:

$$\begin{cases} \partial_t \varphi(t,x) = v(t,\varphi(t,x))\,, \\ \varphi(t=0,x) = x\,. \end{cases} \tag{95}$$

Minimize:

$$\mathcal{L}(v) = \frac{1}{N}\sum_{i=1}^{N} |\varphi(1)(x_i) - y_i|^2\,. \tag{96}$$

Remark: no regularization on v (aka weight-decay).
Parameter is $v \in L^2([0,1],V)$ with this Hilbert metric structure.

Compute the gradient

Gradient of $\mathcal{L}$

$$D\mathcal{L}(\xi)(\eta) = \int_0^1 \langle J(p,q), \eta \rangle \, \mathrm{d}t \,, \tag{97}$$

whith p, q satisfying

$$\begin{cases} \dot{p} = -\, \mathrm{d}\xi^\top(q)(p) \\ \dot{q} = \xi(q)\,, \end{cases} \tag{98}$$

and initial conditions $p(1) = -\partial_q \ell(q(1))$. [a]

a. $\ell(q) = \sum \|B(q_i(1)) - y_i\|^2$.

$\Longrightarrow$ possible to integrate: $J(p(t), q(t)) = \mathrm{Ad}^*_{g(t)\cdot g(1)^{-1}}(J(p(1), q(1)))$.
But, $p(1) = B^*(B(q(1)) - y)$ and therefore, $\ell(q) = \frac{1}{2}\|p(1)\|^2_{[BB^*]^{-1}}$.

Conclusion: The gradient is a diffeomorphic deformation of the gradient of the L^2 loss.

Local PL condition

Local PL

Set $\delta = \min_{i \neq j} |x_i - x_j| > 0$ and $D = \max_{i \neq j} |x_i - x_j|$.
Assuming K the kernel of V satisfies
$\lambda(D,\delta)\,\mathrm{Id} \preccurlyeq (K(x_i,x_j)) \preccurlyeq \Lambda(D,\delta)\,\mathrm{Id}$. Then, a local PL is satisfied, on $B(R)$ in $L^2([0,1],V)$, one has

$$c\ell(\xi) \leq 2MRe^R \|\nabla \ell(\xi)\|^2 \tag{99}$$

$$\|\nabla \ell(\xi)\|^2 \leq 2MCRe^R \ell(\xi)\,. \tag{100}$$

- All critical points are global.
- If loss is small enough, global convergence.
- If iterates are bounded, then global convergence.
- Global convergence is false: symmetries are preserved.

Open question: global convergence with random initialization

Reference: *Global convergence of ResNets: From finite to infinite width using linear parameterization*, Barboni, Peyré, Vialard, 2022.

Contents

1 Motivations from imaging and learning

2 How to build Riemannian metric in infinite dimensions?

3 Right-invariant metrics

4 Dynamic formulation of optimal transport

5 Unbalanced Optimal Transport

6 Gradient flows

7 On global convergence of ResNets

8 The Camassa-Holm equation as an incompressible Euler equation

Arnold's remark on incompressible Euler

Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits, Ann. Inst. Fourier, 1966.

Theorem

The incompressible Euler equation is the geodesic flow of the (right-invariant) L^2 Riemannian metric on $\mathrm{SDiff}(M)$ *(volume preserving diffeomorphisms).*

Arnold's remark on incompressible Euler

Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits, Ann. Inst. Fourier, 1966.

Theorem

The incompressible Euler equation is the geodesic flow of the (right-invariant) L^2 Riemannian metric on $\mathrm{SDiff}(M)$ *(volume preserving diffeomorphisms).*

- An intrinsic point of view by Ebin and Marsden, *Groups of diffeomorphisms and the motion of an incompressible fluid*, Ann. of Math., 1970. Short time existence results for smooth initial conditions.
- An extrinsic point of view by Brenier, relaxation of the variational problem, optimal transport, polar factorization.

Arnold's remark continued

Rewritten in terms of the flow φ, the action reads

$$\int_0^1 \int_M |\partial_t \varphi(t,x)|^2 \, \mathrm{d}x \, \mathrm{d}t \,, \tag{101}$$

under the constraint

$$\varphi(t) \in \mathrm{SDiff}(M) \text{ for all } t \in [0,1] \,. \tag{102}$$

Arnold's remark continued

Rewritten in terms of the flow φ, the action reads

$$\int_0^1 \int_M |\partial_t \varphi(t,x)|^2 \, \mathrm{d}x \, \mathrm{d}t \,, \tag{101}$$

under the constraint

$$\varphi(t) \in \mathrm{SDiff}(M) \text{ for all } t \in [0,1] \,. \tag{102}$$

Riemannian submanifold point of view:

Let $M \hookrightarrow \mathbb{R}^d$ be isometrically embedded: A smooth curve $c(t) \in M$ is a geodesic if and only if $\ddot{c} \perp T_c M$.

Incompressible Euler in Lagrangian form:

$$\begin{cases} \ddot{\varphi} = -\nabla p \circ \varphi \\ \varphi(t) \in \mathrm{SDiff}(M) \,. \end{cases} \tag{103}$$

A geometric picture: Otto's Riemannian submersion

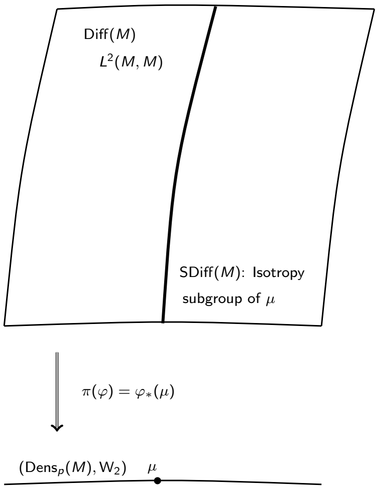


Figure – A Riemannian submersion: SDiff(M) as a Riemannian submanifold of $L^2(M, M)$: Incompressible Euler equation on SDiff(M)

Incompressible Euler and optimal transport

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

What is the corresponding fluid dynamic equation for *WFR*?

The Riemannian submersion for WFR

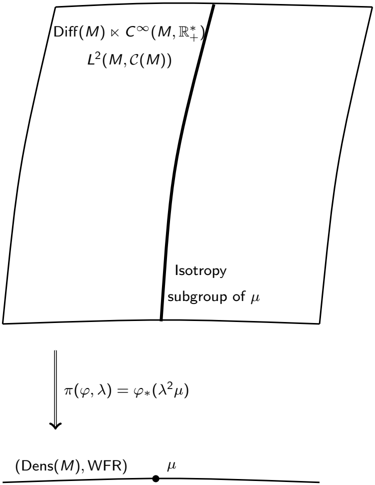


Figure – The same picture in our case: what is the corresponding equation to Euler?

The isotropy subgroup for unbalanced optimal transport

Recall that

$$\pi_0^{-1}(\{\rho_0\}) = \{(\varphi, \lambda) \in \mathrm{Diff}(M) \ltimes C^\infty(M, \mathbb{R}_+^*) \, : \, \varphi_*(\lambda^2 \rho_0) = \rho_0\}$$

The isotropy subgroup for unbalanced optimal transport

Recall that

$$\pi_0^{-1}(\{\rho_0\}) = \{(\varphi, \lambda) \in \mathrm{Diff}(M) \ltimes C^\infty(M, \mathbb{R}_+^*) \,:\, \varphi_*(\lambda^2 \rho_0) = \rho_0\}$$

$$\pi_0^{-1}(\{\rho_0\}) = \{(\varphi, \sqrt{\mathrm{Jac}(\varphi)}) \in \mathrm{Diff}(M) \ltimes C^\infty(M, \mathbb{R}_+^*) \,:\, \varphi \in \mathrm{Diff}(M)\}\,.$$

The vertical space is

$$\mathrm{Vert}_{(\varphi,\lambda)} = \{(v, \alpha) \circ (\varphi, \lambda)\,;\, \mathrm{div}(\rho v) = 2\alpha\rho\}\,, \tag{104}$$

where $(v, \alpha) \in \mathrm{Vect}(M) \times C^\infty(M, \mathbb{R})$. The horizontal space is

$$\mathrm{Hor}_{(\varphi,\lambda)} = \left\{ \left(\frac{1}{2} \nabla p, p \right) \circ (\varphi, \lambda)\,;\, p \in C^\infty(M, \mathbb{R}) \right\}. \tag{105}$$

The isotropy subgroup for unbalanced optimal transport

Recall that

$$\pi_0^{-1}(\{\rho_0\}) = \{(\varphi, \lambda) \in \mathrm{Diff}(M) \ltimes C^\infty(M, \mathbb{R}_+^*) \,:\, \varphi_*(\lambda^2 \rho_0) = \rho_0\}$$

$$\pi_0^{-1}(\{\rho_0\}) = \{(\varphi, \sqrt{\mathrm{Jac}(\varphi)}) \in \mathrm{Diff}(M) \ltimes C^\infty(M, \mathbb{R}_+^*) \,:\, \varphi \in \mathrm{Diff}(M)\}\,.$$

The vertical space is

$$\mathrm{Vert}_{(\varphi,\lambda)} = \{(v, \alpha) \circ (\varphi, \lambda)\,;\, \mathrm{div}(\rho v) = 2\alpha\rho\}\ , \tag{104}$$

where $(v, \alpha) \in \mathrm{Vect}(M) \times C^\infty(M, \mathbb{R})$. The horizontal space is

$$\mathrm{Hor}_{(\varphi,\lambda)} = \left\{\left(\frac{1}{2}\nabla p, p\right) \circ (\varphi, \lambda)\,;\, p \in C^\infty(M, \mathbb{R})\right\}\ . \tag{105}$$

The induced metric is

$$G(v, \mathrm{div}\, v) = \int_M |v|^2 \,\mathrm{d}\mu + \frac{1}{4}\int_M |\,\mathrm{div}\, v|^2 \,\mathrm{d}\mu\,. \tag{106}$$

The H^{div} right-invariant metric on the group of diffeomorphisms.

Toward the incompressible Euler equation

Why? Unbalanced OT is linked to standard OT on the cone (Liero, Mielke, Savaré).

Toward the incompressible Euler equation

Why? Unbalanced OT is linked to standard OT on the cone (Liero, Mielke, Savaré).

Question

Understand $\mathrm{Diff}(M) \ltimes C^\infty(M, \mathbb{R}_+^*)$ *as a subgroup of* $\mathrm{Diff}(\mathcal{C}(M))$*?*

Answer

The cone $\mathcal{C}(M)$ is a trivial principal fibre bundle over M.
The automorphism group $\mathrm{Aut}(\mathcal{C}(M)) \subset \mathrm{Diff}(\mathcal{C}(M))$ can be identified with $\mathrm{Diff}(M) \ltimes C^\infty(M, \mathbb{R}_+^*)$. One has $(\varphi, \lambda) : (x, r) \mapsto (\varphi(x), \lambda(x)r)$.

Recall that $\psi \in \mathrm{Aut}(\mathcal{C}(M))$ if $\psi \in \mathrm{Diff}(\mathcal{C}(M))$ and $\forall \lambda \in \mathbb{R}_+^*$ one has $\psi(\lambda \cdot (x, r)) = \lambda \cdot \psi(x, r)$ where $\lambda \cdot (x, r) \stackrel{\text{def.}}{=} (x, \lambda r)$.

CH as an incompressible Euler equation

The geodesic equation on $\mathrm{Diff}(M) \ltimes C^\infty(M, \mathbb{R}_+^*)$

$$\begin{cases} \frac{D}{Dt}\dot{\varphi} + 2\frac{\dot{\lambda}}{\lambda}\dot{\varphi} = -\nabla^g P \circ \varphi \\ \ddot{\lambda} r - \lambda r g(\dot{\varphi}, \dot{\varphi}) = -2\lambda r P \circ \varphi \,. \end{cases} \tag{107}$$

can be extended to $\mathrm{Aut}(\mathcal{C}(M))$ as

$$\frac{D}{Dt}(\dot{\varphi}, \dot{\lambda r}) = -\nabla \Psi_P \circ (\varphi, \lambda r)\,, \tag{108}$$

where $\Psi_P(x, r) \overset{\text{def.}}{=} r^2 P(x)$.

CH as an incompressible Euler equation

The geodesic equation on $\mathrm{Diff}(M) \ltimes C^\infty(M, \mathbb{R}_+^*)$

$$\begin{cases} \frac{D}{Dt}\dot{\varphi} + 2\frac{\dot{\lambda}}{\lambda}\dot{\varphi} = -\nabla^g P \circ \varphi \\ \ddot{\lambda} r - \lambda r g(\dot{\varphi}, \dot{\varphi}) = -2\lambda r P \circ \varphi \,. \end{cases} \tag{107}$$

can be extended to $\mathrm{Aut}(\mathcal{C}(M))$ as

$$\frac{D}{Dt}(\dot{\varphi}, \dot{\lambda r}) = -\nabla \Psi_P \circ (\varphi, \lambda r)\,, \tag{108}$$

where $\Psi_P(x, r) \stackrel{\text{def.}}{=} r^2 P(x)$.

Question

Does there exist a density $\tilde{\mu}$ on the cone such that $\mathrm{inj}(\mathrm{Diff}(M)) \subset \mathrm{SDiff}_{\tilde{\mu}}(\mathcal{C}(M))$*? (answer: yes)*

CH as an incompressible Euler equation

The geodesic equation on $\mathrm{Diff}(M) \ltimes C^\infty(M, \mathbb{R}_+^*)$

$$\begin{cases} \frac{D}{Dt}\dot{\varphi} + 2\frac{\dot{\lambda}}{\lambda}\dot{\varphi} = -\nabla^g P \circ \varphi \\ \ddot{\lambda} r - \lambda r g(\dot{\varphi}, \dot{\varphi}) = -2\lambda r P \circ \varphi \,. \end{cases} \tag{107}$$

can be extended to $\mathrm{Aut}(\mathcal{C}(M))$ as

$$\frac{D}{Dt}(\dot{\varphi}, \dot{\lambda r}) = -\nabla \Psi_P \circ (\varphi, \lambda r)\,, \tag{108}$$

where $\Psi_P(x, r) \overset{\text{def.}}{=} r^2 P(x)$.

Question

Does there exist a density $\tilde{\mu}$ on the cone such that $\mathrm{inj}(\mathrm{Diff}(M)) \subset \mathrm{SDiff}_{\tilde{\mu}}(\mathcal{C}(M))$*? (answer: yes)*

Proof.

The measure $\tilde{\mu} \overset{\text{def.}}{=} r^{-3}\,\mathrm{d}r\,\mathrm{d}\mu$ where μ denotes the volume form on M. □

Results

Theorem

Let φ be the flow of a smooth solution to the Camassa-Holm equation then $\Psi(\theta, r) \overset{\text{def.}}{=} (\varphi(\theta), \sqrt{\text{Jac}(\varphi(\theta))}r)$ 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 $\Psi(\theta, r) \stackrel{\text{def.}}{=} (\varphi(\theta), \sqrt{\operatorname{Jac}(\varphi(\theta))}r)$ 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

$$\begin{cases} \partial_t u - \frac{1}{4}\partial_{txx} u \, u + 3\partial_x u \, u - \frac{1}{2}\partial_{xx} u \, \partial_x u - \frac{1}{4}\partial_{xxx} u \, u = 0 \\ \partial_t \varphi(t, x) = u(t, \varphi(t, x)) \, . \end{cases} \tag{109}$$

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

$$\begin{cases} \dot{v} + \nabla_v v = -\nabla p \, , \\ \nabla \cdot (\rho v) = 0 \, . \end{cases} \tag{110}$$

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