A lecture on the interface between information geometry, optimization and optimal transport

Flavien Léger

INRIA & Université Paris Dauphine

flavienleger.github.io/slides

Online seminar on statistics and geometry
December 13, 2024

Outline

The Kim–McCann geometry

The Kim–McCann framework

X and Y are n-dimensional smooth manifolds, and $\hat{M} \subset X \times Y$ is an open subset. Consider a fonction $c \in C^2(\hat{M})$. $\hat{M}$ is the **ambient space** and c is a **cost function**. We always assume that the cost $c(x, y)$ is **nondegenerate** in the sense that for each $(x, y) \in \hat{M}$, the linear map $\nabla^2_{xy} c(x, y) \colon T_x X \to T^*_y Y$ is one-to-one.

Definition (Kim and McCann [KM10])

The **Kim–McCann metric** is the pseudo-Riemannian metric on $\hat{M}$ defined by

$$\hat{g}_{(x,y)} = \frac{1}{2} \begin{pmatrix} 0 & -\nabla^2_{xy} c \\ -\nabla^2_{xy} c & 0 \end{pmatrix}.$$

Recall: **pseudo-Riemannian** means: at each $z \in \hat{M}$, $\hat{g}_z$ is a symmetric nondegenerate bilinear form on $T_z \hat{M} \times T_z \hat{M}$.

Remark. The full (x, y) Hessian of c is not well-defined, but the product structure makes the cross terms $\nabla^2_{xy} c$ well-defined. Indeed for fixed $y \in Y$, $x \mapsto \nabla_y c(x, y)$ map to the same space $T^*_y Y$.

First properties

1. $\hat{M}$ open $\subset X \times Y$ makes $T_{(x,y)}\hat{M}$ split as $T_xX \oplus T_yY$. If $U = \xi \oplus \eta \in T_{(x,y)}\hat{M}$ with $\xi \in T_xX$ and $\eta \in T_yY$ then

$$\hat{g}(U,U) = -\nabla^2_{xy}c(x,y)(\xi,\eta) = -\frac{\partial^2 c}{\partial x^i \partial y^{\bar{\jmath}}}\xi^i\eta^{\bar{\jmath}}.$$

2. $\hat{g}$ has signature (n,n). (Use $K(\xi \oplus \eta) = (-\xi \oplus \eta)$.)

3. $-\nabla^2_{xy}c$ is unaffected by adding to $c(x,y)$ a function of x or a function of y. The Kim–McCann metric captures only the interaction between x and y.

Spacelike submanifolds

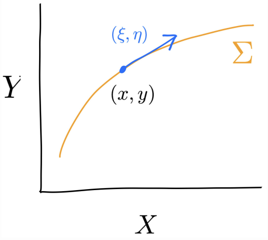
Let $f\colon M \to \hat{M}$ be an **embedding**. $\Sigma := f(M) \subset \hat{M}$ is called a **submanifold** of $\hat{M}$. We often identify $M \approx \Sigma$. If $\hat{g}$ is a pseudo-Riemannian metric on $\hat{M}$ then we may define the pulled back metric $g = f^*\hat{g}$ on M ($\approx$ the restriction of $\hat{g}$ to Σ).

Definition

Σ is **spacelike** if g is **Riemannian**.

This means
for any nonzero **tangential vector** U, $g(U,U) = \hat{g}(U,U) > 0$.

Most often, Σ is given as the graph of a map $T\colon X \to Y$.



Example

$X = Y = \mathbb{R}^n$, $\hat{M} = \mathbb{R}^n \times \mathbb{R}^n$, $c(x, y) = -\langle x, y \rangle$. Then

$$\hat{g}_{(x,y)} = \frac{1}{2} \begin{pmatrix} 0 & I_n \\ I_n & 0 \end{pmatrix}.$$

In other words,

$$\hat{g}(\xi \oplus \eta, \xi \oplus \eta) = \langle \xi, \eta \rangle.$$

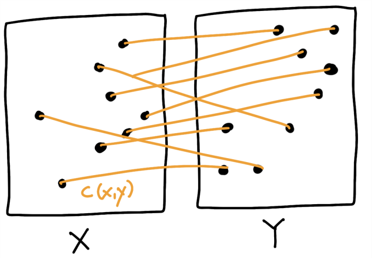
Same Kim–McCann metric $\hat{g}$ for the quadratic cost $c(x, y) = \frac{1}{2}|x - y|^2$ or more generally **Fenchel–Young costs**

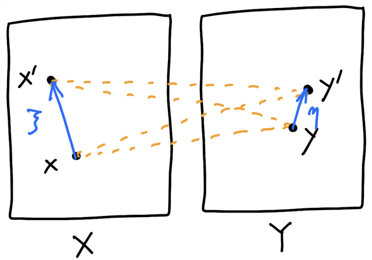
$$c(x, y) = u(x) + u^*(y) - \langle x, y \rangle,$$

where $u \colon \mathbb{R}^n \to \mathbb{R}$ is a convex function (more on this later).

Motivation from Optimal Transport

Optimal transport consists of matching a distribution of points in X with another distribution of points in Y minimizing the total cost.





The cross-difference introduced by McCann [McC99, McC14] is

$$\delta_c(x', y'; x, y) = [c(x, y') + c(x', y)] - [c(x, y) + c(x', y')]$$

Then

$$\delta_c(x + \xi, y + \eta; x, y) = -\nabla^2_{xy} c(x, y)(\xi, \eta) + o(|\xi|^2 + |\eta|^2)$$

Geodesics

Consider local coordinates x^i on X and $y^{\bar\imath}$ on Y. The only nonzero Christoffel symbols $\Gamma^{\gamma}_{\alpha\beta}$ are when α, β, γ are all non-barred or all barred. Then

$$\Gamma^k_{ij} = c^{k\bar m} c_{\bar m ij}, \quad \Gamma^{\bar k}_{\bar\imath\bar\jmath} = c^{\bar k m} c_{m\bar\imath\bar\jmath}.$$

In general a geodesic is of the form $(x(t), y(t))$. Those geodesics for which either the first on second component is constant in time are of particular interest. They are called c-segments and admit a "closed form" formula. Indeed $(x, y(t))$ is a geodesic if

$$\frac{d^2}{dt^2} \nabla_x c(x, y(t)) = 0,$$

while $(x(t), y)$ is a geodesic if

$$\frac{d^2}{dt^2} \nabla_y c(x(t), y) = 0.$$

For example, the geodesic joining (x, y_0) to (x, y_1) takes the form

$$\nabla_x c(x, y(t)) = (1-t)\nabla_x c(x, y_0) + t\nabla_x c(x, y_0).$$

Curvature

Let $\hat{R}$ denote the Riemann curvature of $\hat{g}$. In local coordinates x^i, $y^{\bar{\imath}}$, the only nonzero of $\hat{R}_{\alpha\beta\gamma\delta}$ are when two indices are barred and two unbarred and α, β (thus γ, δ) are of opposite type.

This can be rephrased as follows. Define $K_z : T_z\hat{M} \to T_z\hat{M}$ by $K(\xi \oplus \eta) = (-\xi) \oplus \eta$ and consider the quadrilinear form $Q(U) = \hat{R}(U, KU, U, KU)$.

Proposition

$\hat{R}$ is uniquely determined by Q.

Remark. In general pseudo-Riemannian geometry the Riemann tensor R is uniquely determined by the unnormalized sectional curvature $R(U, V, U, V)$.

Para-Kähler geometry

The Kim–McCann geometry is an instance of **para-Kähler geometry**.

Definition

A **para-Kähler** manifold $(\hat{M}, \hat{g}, K)$ consists of a pseudo-Riemannian manifold $(\hat{M}, \hat{g})$ together with a $(1,1)$ tensor field K parallel with respect to the Levi-Civita connection which is involutive and whose eigenbundles associated with the two eigenvalues $+1$ and -1 of K have the same rank.

In other words:

- $\hat{M}$ is a $2n$-dimensional smooth manifold;
- At each $z \in \hat{M}$, $\hat{g}_z$ is a symmetric nondegenerate bilinear form on $T_z\hat{M}$;
- At each $z \in \hat{M}$, K_z is a linear map from $T_z\hat{M}$ to $T_z\hat{M}$
- $\hat{\nabla} K = 0$ where $\hat{\nabla}$ denote the Levi-Civita connection of $\hat{g}$;
- At each $z \in \hat{M}$, $K_z^2 = \mathrm{Id}_{T_z\hat{M}}$. K_z is therefore diagonalizable with eigenvalues ± 1 and the corresponding eigenspaces $T_z^{\pm}\hat{M}$ have dimension n.

Remark

We also get for free a symplectic form $\omega = \hat{g}(K\cdot, \cdot)$.

Remark

The para-complex numbers (aka split-complex or hyperbolic numbers) are $z = x + ky$ with $k^2 = 1$. An algebra similar to complex numbers but not a field since numbers $x \pm kx$ are not invertible.

References

The Kim–McCann geometry. Kim and McCann's original papers [KM10, KM12]. McCann's review [McC14]. A recent exposition [LV23, Section 2].

Para-Kähler geometry. See the reviews [AiMT09] and [CFG96].

Information geometry

Introduction

In information geometry we consider finite-dimensional parametrized subspaces of measures

$$\{\mu_\theta : \theta \in \Theta\} \subset \mathcal{P}(\Omega).$$

Here Ω is say a domain of $\mathbb{R}^d$ or a smooth manifold. Information geometry assumes Θ to be an n-dimensional smooth manifold, called *statistical manifold*. A typical problem is to optimize over the μ_θ, for instance the maximum likelihood problem is related to minimizing the function

$$F(\theta) = \mathrm{KL}(\nu|\mu_\theta),$$

where $\nu \in \mathcal{P}(\Omega)$ is given.

Example (Gaussians)

Optimize over the spaces of Gaussians $\Theta = \mathbb{R}^d \times \mathrm{S}^d_{++}$ parametrized by $\theta = (\text{mean}, \text{covariance})$.

Example (Exponential families)

Given $s\colon \mathbb{R}^d \to \mathbb{R}^n$ consider the **exponential family** $\mu_\theta(dx) = e^{\langle s(x), \theta\rangle - A(\theta)}\nu(dx)$, with $\Theta \subset \mathbb{R}^n$. Here $A(\theta)$ ensures μ_θ has mass 1 and ν is a fixed reference measure on $\mathbb{R}^d$.

Submanifolds Σ from divergences

Consider a triple $(X \times Y, c)$, as in the Kim–McCann framework. A pair (ϕ, ψ) with $\phi\colon X \to \mathbb{R}$ and $\psi\colon Y \to \mathbb{R}$ is called c-conjugate if $\phi(x) = \psi^c(x) := -\inf_{y\in Y} c(x,y) + \psi(y)$ and $\psi(y) = \phi^c(y) := -\inf_{x\in X} c(x,y) + \phi(x)$. If (ϕ, ψ) is c-conjugate we have

$$D(x,y) := \phi(x) + \psi(y) + c(x,y) \geq 0,$$

with

$$\inf_{x\in X} D(x,y) = \inf_{y\in Y} D(x,y) = 0.$$

Definition

D is called a divergence.

Then a spacelike submanifold $\Sigma \subset X \times Y$ can be constructed, under additional mild assumptions, as the set where X vanishes,

$$\Sigma = \{(x,y) \in X \times Y : D(x,y) = 0\}.$$

Observations

1. The cross-differences $\delta_D = \delta_c$. Therefore the Kim–McCann metrics induced by D and c are the same.

2. Why the vanishing set Σ of $D(x,y)$ can be expected to be spacelike: if $(x,y) \in \Sigma$ and $(x+\xi, y+\eta) \in \Sigma$ then

$$\delta_D(x+\xi, y+\eta; x, y) = \underbrace{D(x+\xi, y)}_{\geq 0} + \underbrace{D(x, y+\eta)}_{\geq 0} - \underbrace{D(x,y)}_{=0} - \underbrace{D(x+\xi, y+\eta)}_{=0} \geq 0.$$

By 1., $\delta_D = \delta_c$. Take $\xi \to 0$ and $\eta \to 0$ then $-\nabla^2_{xy} c(x,y)(\xi,\eta) \geq 0$. ($\Sigma$ is c-monotone).

3. Oftentimes we consider costs $c(x,y)$ satisfying

$$\inf_{x \in X} c(x,y) = \inf_{y \in Y} c(x,y) = 0.$$

Then $(\phi = 0, \psi = 0)$ is a c-conjuate pair and c is directly a divergence, $D(x,y) = c(x,y)$.
Example. A squared Riemannian distance $c(x,y) = d^2(x,y)$.

Example: Fenchel–Young gap functions

Consider cost $c(x,y) = -\langle x, y\rangle$ on $\mathbb{R}^n \times \mathbb{R}^n$. The Kim–McCann metric is

$$\hat{g} = \frac{1}{2}\begin{pmatrix} 0 & I_n \\ I_n & 0 \end{pmatrix}.$$

Let $u \in C^2(\mathbb{R}^n)$ be a strictly convex function and consider the divergence

$$D(x,y) = u(x) + u^*(y) - \langle x, y\rangle.$$

D vanishes on $\Sigma = \{(x, \nabla u(x))\}$ and the induced Riemannian metric is Hessian,

$$g = \nabla^2 u.$$

Pulled back divergences

Back to statistical manifolds. In practice the divergence D on $\Theta \times \Theta$ is often pulled back from a "divergence" $\mathbb{D}$ on $\mathcal{P}(\Omega) \times \mathcal{P}(\Omega)$,

$$D(\theta, \theta') = \mathbb{D}(\mu_\theta, \mu_{\theta'}).$$

Example

The Kullback–Leibler divergence or relative entropy $\mathbb{D}(\mu, \mu') = \int_\Omega \log(d\mu/d\mu')\, d\mu$. Under certains assumptions the diagonal of $\Theta \times \Theta$ is spacelike and the Kim–McCann metric on Σ is the Fisher information

$$g_{ij}(\theta) = \int_\Omega \frac{\partial \ln \mu_\theta}{\partial \theta^i} \frac{\partial \ln \mu_\theta}{\partial \theta^j} \mu_\theta(dx).$$

Other examples. The Hellinger divergence $\mathbb{D}(\mu, \mu') = \int_\Omega (\sqrt{d\mu/d\nu} - \sqrt{d\mu'/d\nu})^2\, d\nu$. The squared Wasserstein distance $\mathbb{D}(\mu, \mu') = W_2^2(\mu, \mu')$.

Example: exponential families

Given $s\colon \mathbb{R}^d \to \mathbb{R}^n$, recall the exponential family $\mu_\theta(dx) = e^{\langle s(x),\theta\rangle - A(\theta)}\nu(dx)$, with $\Theta \subset \mathbb{R}^n$. The pullback of the KL divergence takes the form

$$D(\theta,\theta') = \int_{\mathbb{R}^d} \ln\left(\frac{d\mu_\theta}{d\mu_{\theta'}}\right) d\mu_\theta = A(\theta') - A(\theta) - \langle \nabla A(\theta), \theta' - \theta\rangle.$$

This is the **Bregman divergence** of A.
Here Σ is the diagonal and the Fisher information metric metric is the Hessian metric $\nabla^2 A(\theta)$.

Classical submanifold theory

Let $(\hat{M}, \hat{g})$ be a pseudo-Riemannian manifold and $f\colon M \to \hat{M}$ be an **embedding**. Define the submanifold $\Sigma = f(M) \subset \hat{M}$ and $g = f^*\hat{g}$ on M (or Σ).
Let U, V be tangential vector fields on Σ. To obtain an affine connection on Σ we want to project $\hat{\nabla}_U V$ onto $T\Sigma$. The classical way is to project orthogonally and obtain a connection $\nabla_U V$ on Σ. It turns out ∇ is nothing else than the Levi-Civita connection for g.

Classically there are then three notions of curvatures on Σ: R, $\hat{R}$ and **the second fundamental form** $I\!I\colon TM \times TM \to T^\perp M$ defined by

$$I\!I(U, V) = \hat{\nabla}_U V - \nabla_U V.$$

The **mean curvature** $H \in T^\perp\Sigma$ is then a normal vector field defined as the trace of $I\!I$ (with respect to g).
$\hat{R}$, R are intrinsic while $I\!I$, H are extrinsic.

Information geometry's dual connections

Information geometry takes a different approach. Due to the special product structure $X \times Y$ it defines instead two connections ∇^1, ∇^2 on Σ which are different from the Levi-Civita ∇ coming from g. Given tangential U, V, project onto TX and TY respectively,

$$\nabla^1_U V = \pi^1(\hat{\nabla}_U V),$$
$$\nabla^2_U V = \pi^2(\hat{\nabla}_U V).$$

It turns out that $\frac{1}{2}(\nabla^1 + \nabla^2) = \nabla$. The classical (∇, II) are replaced by (∇^1, ∇^2).

There are three notions of curvatures $\hat{R}$, R^1, R^2.

References

The presentation roughly follows Wong and Yang [WY22]. See also the nicely written review of Khan and Zhang [KZ22].

A classical reference for information geometry is the textbook of Amari [Ama16]. See also the review of Nielsen [Nie20] and Mishra, Kumar and Wong [MKW23].

Application to gradient descent-type schemes

Introduction

We want to iteratively minimize a differentiable function $f\colon X \to \mathbb{R}$, where X is a smooth manifold. Since there is no metric on X, we cannot follow the "direction of steepest descent". Indeed, the differential $\nabla f(x)$ is a covector (i.e. a one-form) rather than a tangent vector.

Motivating example. Let V be an n-dimensional real vector space, and $f \in C^1(V)$. Then an update of the type $x_{k+1} - x_k = -\nabla f(x_k)$ **doesn't make sense** since $x_{k+1} - x_k \in V$ while $\nabla f(x_k) \in V^*$. But if we choose a map $T\colon V \to V^*$ then we can go back-and-forth between V and V^* and have a working scheme. *Remark:* an inner product $\langle \cdot, \cdot \rangle$ induces a canonical map $V \to V^*$.

Basic setting

X and $f \in C^1(X)$ are given. Choose an n-dimensional manifold Y (the "dual space") and a nondegenerate cost $c \in C^2(X \times Y)$. Choose a n-dimensional submanifold $\Sigma \subset X \times Y$ which is the graph of a diffeomorphism $T\colon X \to Y$. Σ acts as our one-to-one correspondence between X and Y.

Define $F\colon X \times Y \to \mathbb{R}$ by

$$F(x, y) = f(x).$$

Note that $\hat{\nabla} F = \hat{\nabla} f \oplus 0$. The Kim–McCann metric provides **a gradient** $\operatorname{grad} F = \hat{g}^{-1} \hat{\nabla} F$. Due to the special structure of $\hat{g}$ and F, the gradient is of the form $\operatorname{grad} F = 0 \oplus \eta$, where (in coordinates)

$$\eta^{\bar{\imath}} = -c^{\bar{\imath} j} \frac{\partial f}{\partial x^j}.$$

Gradient descent with a general cost (GDGC)

This suggest the following iterative method (GDGC, [LAF23]).

Given $(x_k, y_k) \in \Sigma$.
y-update: compute $\hat{\exp}_{(x_k,y_k)}(-\operatorname{grad} F) =: (x_k, y_{k+1})$
x-update: $(x_{k+1}, y_{k+1}) \in \Sigma$.

Here $\hat{\exp}$ uses the ambient Kim–McCann connection $\hat{\nabla}$ on $X \times Y$, in particular it leaves Σ.
Then map back into Σ.
The exponential map $\hat{\exp}_{(x,y)}(0 \oplus \eta)$ admits a **closed-form formula**.

Under mild assumptions GDGC can be written as

$$
\begin{aligned}
-\nabla_x c(x_k, y_{k+1}) &= -\nabla f(x_k), \\
\nabla_x c(x_{k+1}, y_{k+1}) &= 0.
\end{aligned}
$$

Example: mirror descent

Suppose we are given an objective function $f\colon V \to \mathbb{R}$ where V is an n-dimensional vector space, **without inner product**. Let $u \in C^2(V)$ be a strictly convex function and consider the Fenchel–Young divergence $c(x,y) = u(x) + u^*(y) - \langle x, y\rangle$ on $V \times V^*$, vanishing on the subset $\Sigma = \{(x, \nabla u(x))\} \subset V \times V^*$.

The ambient Kim–McCann metric is $\hat{g} = \frac{1}{2}\begin{pmatrix} 0 & I_n \\ I_n & 0 \end{pmatrix}$, i.e. $\hat{g}(\xi \oplus \eta, \xi \oplus \eta) = \langle \xi, \eta\rangle$. The induced Riemannian metric on Σ can be written (in "x-coordinates") as the Hessian metric $g = \nabla^2 u(x)$, i.e. $g(\xi,\xi) = \nabla^2 u(x)(\xi,\xi)$. Indeed if $\xi \oplus \eta$ is tangent to Σ then $\eta = \nabla^2 u(x)\xi$.
Instantiate the GDGC method: Given $x_k \in V$ and $y_k = \nabla u(x_k)$:

- y-update: $y_{k+1} = y_k - \nabla f(x_k)$ (flat connection).
- x-update $x_{k+1} = (\nabla u)^{-1}(y_{k+1})$.

We obtain the mirror descent update

$$\nabla u(x_{k+1}) - \nabla u(x_k) = -\nabla f(x_k).$$

Other examples

$f \in C^1(V)$, strictly convex $u \in C^2(V)$, Bregman cost $c(x,y) = u(y) - u(x) - \langle \nabla u(x), y - x\rangle$ on $V \times V$, with Σ =diagonal. Then GDGC=natural gradient descent

$$x_{k+1} - x_k = -\nabla^2 u(x_k)^{-1} \nabla f(x_k).$$

(M,g) Riemannian manifold, $f \in C^1(M)$, squared geodesic cost $c(x,y) = \frac{1}{2\tau} d_M^2(x,y)$ on $M \times M$ with $\tau > 0$, and Σ =diagonal. Then GDGC=Riemannian gradient descent

$$x_{k+1} = \exp_{x_k}(-\tau \nabla f(x_k)).$$

More to explore

1. When $f(x) = \inf_{y \in Y} c(x,y) + h(y)$ (f is c-concave), GDGC can be formulated as the alternating minimization of
$$c(x,y) + h(y).$$
This is a **nonsmooth formulation** valid in infinite dimensions [LAF23].

2. There are implicit and forward–backward (explicit–implicit) extensions [LAF23].

3. The condition $\hat{R}(U, KU, U, KU) \geq 0$ is known as **nonnegative cross-curvature** (NNCC) [KM10, KM12]. Under NNCC convexity of the objective f along c-segments provides rates of convergence [LAF23]. Moreover NNCC admits a synthetic formulation applicable to infinite-dimensional spaces [LTV24].

Apriori estimate in optimal transport

Optimal transport setting

This section is based on [BLMR24].
X and Y are two n-dimensional smooth manifold, $\hat{M} \subset X \times Y$ is an open domain and $c \in C^4(\hat{M})$ is a nondegenerate cost. μ and ν are two smooth probability measures on X and Y respectively.
In the **optimal transport** problem we want to find a map $T\colon X \to Y$ **pushing** μ **to** ν, which minimizes the total cost

$$\int_X c(x, T(x))\, d\mu(x). \tag{4.1}$$

1. Problem (4.1) can be formulated as a ~~minimal~~ maximal surface problem, $\Sigma = \operatorname{gra} T \subset \hat{M}$.

2. New, geometric proof of the Pogorelov-style Ma–Trudinger–Wang estimates

$$|DT(x)| \leq C$$

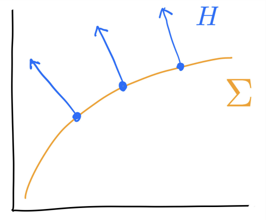
(Lipschitz bound on the transport map).

Background on submanifold theory

Let $f\colon M \to \hat{M}$ be an **embedding**. $\Sigma = f(M) \subset \hat{M}$ is called a submanifold of $\hat{M}$. Identify $M \approx \Sigma$. If $\hat{g}$ is a pseudo-Riemannian metric on $\hat{M}$ then we may define $g = f^*\hat{g}$ on M (or Σ).
There are three notions of curvatures on Σ: R, $\hat{R}$ and **the second fundamental form** $II\colon TM \times TM \to T^{\perp}M$ defined by

$$II(U,V) = \hat{\nabla}_U V - \nabla_U V.$$

The **mean curvature** $H \in T^{\perp}\Sigma$ is then
a normal vector field defined as the trace of II (with respect to g).
Recall
the first variation formula: for compact Riemannian submanifolds the mean curvature is “minus the gradient” of the area functional.



Kim–McCann–Warren

Define the **conformal factor** $\chi\colon \hat{M} \to \mathbb{R}$ by

$$\chi(x,y)^n = \frac{d(\mu \otimes \nu)}{d\,\mathrm{vol}_{\mathrm{KM}}} = \frac{d\mu/dx(x)d\nu/dy(y)}{|\det \nabla_{xy} c(x,y)|},$$

where $\mathrm{vol}_{\mathrm{KM}}$ denotes the volume form of the Kim–McCann metric and $d\mu/dx$, $d\nu/dy$ denote the densities of μ and ν in local coordinates.
In [?] Kim, McCann and Warren introduced the pseudo-Riemannian metric on $\hat{M}$

$$\hat{g} = \chi(x,y) \begin{pmatrix} 0 & -\nabla^2_{xy} c \\ -\nabla^2_{xy} c & 0 \end{pmatrix}. \tag{4.2}$$

Kim, McCann and Warren show

Theorem ([KMW10])

*For the Kim–McCann-Warren metric (4.2), the submanifold $\Sigma \subset \hat{M}$ is spacelike maximizing. In particular it has **zero mean curvature**.*

A priori estimate

Recall we are interested to show $|DT| \leq C$. This can be shown to be equivalent to an upper bound on g, where g denote the restriction of $\hat{g}$ to Σ. Geometrically, g to be compared to something. We therefore fix an ambient **Riemannian metric** $\hat{S}$ on $\hat{M}$ and denote by S its restriction to Σ. There are two main ingredients.

Ingredient 1. The Kim–McCann–Warren metric satisfies on Σ

$$\mathrm{vol}(g) = \mu,$$

therefore the product of the eigenvalues of g is bounded above and below.

Ingredient 2. the Ma–Trudinger–Wang condition $(\kappa > 0)$

$$\hat{R}^{\mathrm{KM}}(U, KU, U, KU) \geq \kappa\big(\hat{S}(U,U)\hat{S}(KU,KU) - \hat{S}(U,KU)^2\big)$$

for null U i.e. $\hat{g}(U,U) = 0$.

Remarks. $\hat{R}^{\mathrm{KM}}(U, KU, U, KU) = \hat{Q}(U)$ is the para-Kähler quadrilinear form.
$\hat{S}(U,U)\hat{S}(V,V) - \hat{S}(U,V)^2$: transforms like a curvature tensor.

Approach: bound $g \geq C^{-1}S$ on Σ, i.e.

$$S \leq Cg.$$

Proposition ([BLMR24])

At any point $p \in \Sigma$, let (e_i) denote a g-orthonormal basis of $T_p\Sigma$ that diagonalizes S. Then at p,

$$\sum_{l=1}^{n} \hat{R}(e_l, e_n, e_l, e_n) \leq \frac{1}{2}\frac{(\Delta S)(e_n, e_n)}{S(e_n, e_n)} + C\sum_{l=1}^{n} S(e_l, e_l).$$

Maximum principle. At a point $p_0 \in \Sigma$ where S maximizes its largest eigenvalue λ_n relative to g we have $(\Delta S)(e_n, e_n) \leq 0$, where (e_i) is g-orthonormal and $S(e_i, e_j) = \lambda_i \delta_{ij}$. Therefore at the point p_0,

$$\sum_{l=1}^{n-1} \hat{R}(e_l, e_n, e_l, e_n) \leq CS(e_n, e_n).$$

Using the MTW condition (**Ingredient 2**):

$$\sum_{l=1}^{n-1} \lambda_l \leq C.$$

Ingredient 1 (Monge–Ampère equation) ≈ product of the eigenvalues is bounded above and below. Conclusion:

$$\lambda_n \leq C.$$

Formal Taylor expansion of entropic optimal transport

Problem formulation

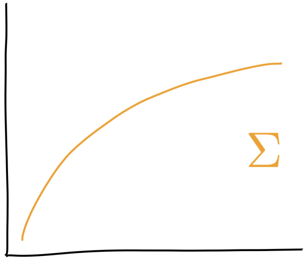
This section is based on work in progress.
X and Y are two n-dimensional smooth manifolds, and $c \in C^4(X \times Y)$ is a nondegenerate cost. μ and ν are two smooth probability measures on X and Y respectively. Finally $\varepsilon > 0$ is a temperature parameter.
The entropic optimal transport problem is

$$\mathcal{T}_{c,\varepsilon}(\mu,\nu) = \min_{\pi \in \Pi(\mu,\nu)} \iint_{X \times Y} c(x,y)\, d\pi(x,y) + \varepsilon\, \mathrm{KL}(\pi|\mu \times \nu).$$

Here $\Pi(\mu,\nu)$ consists of all the joint probability measures on $X \times Y$ with respective marginals μ and ν.
When $\varepsilon = 0$ we recover the optimal transport problem. Under some assumptions π is concentrated on a set Σ which is the graph of a map $T\colon X \to Y$.
When $\varepsilon > 0$, the support of π is all of $X \times Y$.

Kim–McCann geometry

Let $\hat{g}$ denote the Kim–McCann metric on $X \times Y$ with respect to $c(x, y)$ and Σ the graph of an optimal transport map (i.e. taking $\varepsilon = 0$).



$\varepsilon = 0$

$\Sigma =$ zero set of the divergence

$\phi_0(x) + \psi_0(y) + c(x, y) \geq 0.$

$\varepsilon > 0$

$$\pi_\varepsilon(dx, dy) = e^{-[\phi_\varepsilon(x)+\psi_\varepsilon(y)+c(x,y)]/\varepsilon}\mu(dx)\nu(dy)$$

Question: formal asymptotics as $\varepsilon \to 0$.

Formal results

Let H denote the mean curvature of Σ and K the para-complex structure. Then KH is a tangential vector field. Let π_0 denote the optimal transport plan (for $\varepsilon = 0$).
Solve for a potential $V\colon \Sigma \to \mathbb{R}$ the elliptic PDE $\Delta_\pi V = \mathrm{div}_\pi(KH)$, in the weak sense

$$\forall h \in C_c^\infty(\Sigma), \quad \int_\Sigma g(\nabla V, \nabla h)\, d\pi = \int_\Sigma g(KH, \nabla h)\, d\pi.$$

This is the natural projection of KH onto gradient vector fields.

Theorem (Formal)

Define $f\colon \Sigma \to \mathbb{R}$ *by* $e^{-V}\mu = e^{-2f}\,\mathrm{vol}(g)$. *Then*

$$\phi_\varepsilon = \phi_0 + \varepsilon f + o(\varepsilon).$$

Formal results

Theorem (Formal)

$$\mathcal{T}_{c,\varepsilon}(\mu,\nu) = \mathcal{T}_{c,0}(\mu,\nu) - \varepsilon \ln(2\pi\varepsilon)^{d/2} - \varepsilon H(\pi_0 | \operatorname{vol}(g))$$

$$+\frac{\varepsilon^2}{8}\int_\Sigma \left[|\nabla \ln(\pi_0/\operatorname{vol}(g))|^2 + \frac{1}{4}\hat{R} + R + \frac{5}{3}|II|^2 - |\nabla V|^2 \right] d\pi_0 + o(\varepsilon^2)$$

Remark: for the quadratic cost on Euclidean space $c(x,y) = |x-y|^2$, Conforti and Tamanini [CT21] show the ε^2 to be

$$\frac{\varepsilon^2}{8}\int_0^1 \mathrm{FI}(\rho_t)\, dt,$$

where FI is the Fisher information and ρ_t the McCann interpolation between μ and ν.

Bibliography I

- D. V. Alekseevskiĭ, K. Medori, and A. Tomassini, Homogeneous para-Kählerian Einstein manifolds, Uspekhi Mat. Nauk **64** (2009), no. 1(385), 3–50, `doi:10.1070/RM2009v064n01ABEH004591`.
- Shun-ichi Amari, Information geometry and its applications, Applied Mathematical Sciences, vol. 194, Springer, [Tokyo], 2016, `doi:10.1007/978-4-431-55978-8`.
- Simon Brendle, Flavien Léger, Robert J. McCann, and Cale Rankin, A geometric approach to apriori estimates for optimal transport maps, J. Reine Angew. Math. **817** (2024), 251–266, `doi:10.1515/crelle-2024-0071`.
- V. Cruceanu, P. Fortuny, and P. M. Gadea, A survey on paracomplex geometry, Rocky Mountain J. Math. **26** (1996), no. 1, 83–115, `doi:10.1216/rmjm/1181072105`.
- Giovanni Conforti and Luca Tamanini, A formula for the time derivative of the entropic cost and applications, J. Funct. Anal. **280** (2021), no. 11, Paper No. 108964, 48, `doi:10.1016/j.jfa.2021.108964`.

Bibliography II

- Young-Heon Kim and Robert J. McCann, Continuity, curvature, and the general covariance of optimal transportation, J. Eur. Math. Soc. (JEMS) **12** (2010), no. 4, 1009–1040, `doi:10.4171/JEMS/221`.
- ______, Towards the smoothness of optimal maps on Riemannian submersions and Riemannian products (of round spheres in particular), J. Reine Angew. Math. **664** (2012), 1–27, `doi:10.1515/CRELLE.2011.105`.
- Young-Heon Kim, Robert J. McCann, and Micah Warren, Pseudo-Riemannian geometry calibrates optimal transportation, Math. Res. Lett. **17** (2010), no. 6, 1183–1197, `doi:10.4310/MRL.2010.v17.n6.a16`.
- Gabriel Khan and Jun Zhang, When optimal transport meets information geometry, Inf. Geom. **5** (2022), no. 1, 47–78, `doi:10.1007/s41884-022-00066-w`.
- Flavien Léger and Pierre-Cyril Aubin-Frankowski, Gradient descent with a general cost, 2023, URL: `https://arxiv.org/abs/2305.04917`, `arXiv:2305.04917`.

Bibliography III

- Flavien Léger, Gabriele Todeschi, and François-Xavier Vialard, Nonnegative cross-curvature in infinite dimensions: synthetic definition and spaces of measures, URL: `https://arxiv.org/abs/2409.18112`, `arXiv:2409.18112`.
- Flavien Léger and François-Xavier Vialard, A geometric Laplace method, Pure Appl. Anal. **5** (2023), no. 4, 1041–1080, `doi:10.2140/paa.2023.5.1041`.
- Robert J. McCann, Exact solutions to the transportation problem on the line, R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci. **455** (1999), no. 1984, 1341–1380, `doi:10.1098/rspa.1999.0364`.
- ______, A glimpse into the differential topology and geometry of optimal transport, Discrete Contin. Dyn. Syst. **34** (2014), no. 4, 1605–1621, `doi:10.3934/dcds.2014.34.1605`.
- Kumar Vijay Mishra, M. Ashok Kumar, and Ting-Kam Leonard Wong, Information geometry for the working information theorist, 2023, URL: `https://arxiv.org/abs/2310.03884`, `arXiv:2310.03884`.

Bibliography IV

- Frank Nielsen, An elementary introduction to information geometry, Entropy **22** (2020), no. 10, Paper No. 1100, 61, `doi:10.3390/e22101100`.
- Ting-Kam Leonard Wong and Jiaowen Yang, Pseudo-Riemannian geometry encodes information geometry in optimal transport, Inf. Geom. **5** (2022), no. 1, 131–159, `doi:10.1007/s41884-021-00053-7`.