Geometry of the Wasserstein Space

Nikolay Pogodaev

March 21, 2022

Contents

1 Optimal transport

1.1 Finite case

Imagine that we have N bakeries located at x_i and N cafés located at y_i. We would like to deliver bread form the bakeries to the cafés in an optimal way. To make the problem a bit easier we assume that each bakery must supply only one café. Finally, we assume that delivering bread from x_i to

y_j costs $|x_i - y_j|^2$. So if we have a delivering plan $\mathbf{s}$ the corresponding total cost is given by

$$\sum_{i=i}^{N} |x_i - \mathbf{s}(x_i)|^2.$$

Let us denote by $\bar{\mathbf{s}}$ an optimal plan and reorder x's and y's so that

$$\bar{\mathbf{s}}(x_i) = y_i, \quad i = 1, \ldots, N,$$

that is, $\bar{\mathbf{s}}$ now corresponds to the identity permutation.

Consider a plan $\mathbf{s}$ which is constructed from $\bar{\mathbf{s}}$ by a cyclic permutation of k elements:

$$\begin{cases} \mathbf{s}(x_i) = y_{i+1}, \quad i = 1, \ldots, k-1, \quad \mathbf{s}(x_k) = y_1, \\ \mathbf{s}(x_i) = y_i, \quad i = k+1, \ldots, N. \end{cases} \tag{1.1}$$ eq:cycle

Then, by the optimality of $\bar{\mathbf{s}}$, we have

$$\sum_{i=1}^{N} |x_i - y_i|^2 \leq \sum_{i=1}^{k-1} |x_i - y_{i+1}|^2 + |x_k - y_1|^2 + \sum_{i=k+1}^{N} |x_i - y_i|^2, \tag{1.2}$$ eq:opt1

which is equivalent to

$$\langle y_1, x_2 - x_1 \rangle + \langle y_2, x_3 - x_2 \rangle + \cdots + \langle y_k, x_1 - x_k \rangle \leq 0. \tag{1.3}$$ eq:opt2

The inequality (1.3) can be considered as a necessary optimality condition. Now we will show that it is also sufficient.

Proposition 1.1. *A plan $\bar{\mathbf{s}}$ is optimal if and only if for any cyclic perturbation* (1.1) *the inequality* (1.3) *hold.*

Proof. Any transport plan $\mathbf{s}$ can be written in the form $\mathbf{s} = \bar{\mathbf{s}} \circ \sigma$, where $\sigma \in S_N$ and the optimal plan corresponds to the identity permutation. We know that (1.3) (and thus (1.2)) holds for any k-cycle. It is known that every permutation σ can be expressed as a composition of disjoint cycles. Now (1.3) holds for any such cycle. The latter is equivalent to

$$\sum_{i=1}^{N} |x_i - y_i|^2 \leq \sum_{i=1}^{N} |x_i - y_{\sigma(i)}|^2,$$

as desired. $\square$

Definition 1.2. *A set $S \subset \mathbb{R}^d \times \mathbb{R}^d$ is called cyclic if any finite collection of its points (x_i, y_i) satisfies* (1.3)*.*

Now we may restate our optimality condition in a more pleasant way.

Corollary 1.3. *A plan $\bar{\mathbf{s}} \colon \mathbb{R}^d \to \mathbb{R}^d$ is optimal if and only if its graph is cyclic.*

Example 1. Now we are able to solve the transport problem on the line. It is easy to see that we get an optimal plan if we reorder x's and y's so that $x_1 \leq \cdots \leq x_N$ and $y_1 \leq \cdots \leq y_N$. Indeed, in this case the set $S = \{(x_1, y_1), \ldots, (x_N, y_N)\}$ is cyclic.

Example 2. The graph of any nondecreasing function $f \colon \mathbb{R} \to \mathbb{R}$ is cyclic.

1.2 Convex functions

Let $f \colon \mathbb{R}^d \to (-\infty, \infty]$ be a convex function. Recall that a pair (x, y) belongs to its subdifferential $\partial f \subset \mathbb{R}^d \times \mathbb{R}^d$ if and only if

$$f(z) \geq \langle y, z - x\rangle + f(x) \quad \forall z \in \mathbb{R}^d.$$

Proposition 1.4. *The set ∂f is cyclically monotone.*

Proof. Take $(x_i, y_i) \in \partial f$, $i = 1, \ldots, k$. Then

$$f(z_i) \geq \langle y_i, z_i - x_i\rangle + f(x_i) \quad \forall z_i.$$

Taking $z_i = x_{i+1}$, $z_k = x_1$ and summing the above inequalities, we obtain (1.3). ☐

Example 3. Let $f \colon \mathbb{R} \to \mathbb{R}$ be continuously differentiable and convex. Then its gradient $\nabla f \colon \mathbb{R} \to \mathbb{R}$ is nondecreasing. We already know that the graph of any nondecreasing function is cyclic.

Now we are going to prove a more surprising fact.

Proposition 1.5. *Let $S \subset \mathbb{R}^d \times \mathbb{R}^d$ by cyclic. Then there exists a proper convex function f such that $S \subset \partial f$.*

Proof. Fix some $(x_0, y_0) \in S$ and let

$$f(x) = \sup\{\langle y_k, x - x_k\rangle + \langle y_{k-1}, x_k - x_{k-1}\rangle + \cdots + \langle y_0, x_1 - x_0\rangle\}, \tag{1.4}$$

eq:f

where sup is taken over all finite collections $(x_i, y_i) \in S$, $i = 1, \ldots, k$. Note that

1. f is convex and $f(x) > -\infty$ as a supremum of affine maps;
2. $f(x_0) = 0$, since cyclic monotonicity of S implies that $f(x_0) \leq 0$ and choosing $k = 1$ we get zero in the right-hand side of (1.4).

It remains to show that $S \subset \partial f$. Let $(x, y) \in S$. It suffice to check that, for any z,

$$f(z) \geq \langle y, z - x\rangle + \alpha \quad \forall \alpha < f(x).$$

The identity (1.4) implies that there exists (x_i, y_i), $i = 1, \ldots, k$, such that

$$\alpha < \langle y_k, x - x_k\rangle + \langle y_{k-1}, x_k - x_{k-1}\rangle + \cdots + \langle y_0, x_1 - x_0\rangle.$$

Now setting $x_{k+1} = x$ and $y_{k+1} = y$, we get

$$\begin{aligned}\langle y, z - x\rangle + \alpha &< \langle y_{k+1}, z - x_{k+1}\rangle \\ &\quad + \langle y_k, x - x_k\rangle + \langle y_{k-1}, x_k - x_{k-1}\rangle + \cdots + \langle y_0, x_1 - x_0\rangle \leq f(z),\end{aligned}$$

by the definition of f. □

1.3 Infinite case

In the infinite case or optimization problem looks as follows:

$$\min\left\{C(\Pi) \,:\, \Pi \in \Gamma(\mu, \nu)\right\}, \quad \text{where} \quad C(\Pi) = \int |x - y|^2 \, \mathrm{d}\Pi\,(x, y).$$

We begin by pointing out two easy facts.

1. Let $(u \times v)(\omega) = (u(\omega), v(\omega))$. Consider the pushforward measure $(u \times v)_\sharp \eta$. We state that its projections are $u_\sharp \eta$ and $v_\sharp \eta$.

   Indeed,

   $$\begin{aligned}\pi^1_\sharp (u \times v)_\sharp \eta(A) &= (u \times v)_\sharp \eta(A \times \mathbb{R}^d) \\ &= \eta\left(\{\omega \,:\, u(\omega) \in A, v(\omega) \in \mathbb{R}^d\}\right) \\ &= \eta(u^{-1}(A)).\end{aligned}$$

2. A fact from probability theory. Let $\mu_j \in \mathcal{P}(X_j)$ be a finite collection of measures. Then there exists a probability space $(\Omega, \mathcal{B}, \eta)$ and Borel measurable maps $\pi^j \colon \Omega \to X_j$ such that $\mu_j = \pi^j_\sharp \eta$.

   It is enough to take $\Omega = X_1 \times \cdots \times X_n$, $\eta = \mu_1 \times \cdots \times \mu_n$, π^j are the usual projections.

thm:transportplan **Theorem 1.6.** *If Π is optimal and $C(\Pi)$ is finite then* $\operatorname{spt}\Pi$ *is cyclic.*

Proof. Suppose that $\operatorname{spt}\Pi$ contains a collection of points (x_i, y_i) satisfying

$$f(x_1, \ldots, x_n, y_1, \ldots, y_n) = \sum |x_i - y_{i+1}|^2 - \sum |x_i - y_i|^2 < 0.$$

Using these points, we will construct a plan which is better then Π. Since f is continuous there exist compact neighborhoods U_i and V_i of x_i and y_i such that $f < 0$ on $U_1 \times \cdots \times U_n \times V_1 \times \cdots \times V_n$. Since $(x_i, y_i) \in \operatorname{spt}\Pi$, we conclude that

$$\lambda \doteq \min_j \Pi(U_j \times V_j) > 0.$$

Let Π_j be the normalized restriction of Π on $U_j \times V_j$. We know that there exists η and Borel maps $\omega \mapsto (u_j(\omega), v_j(\omega))$ satisfying $(u_j \times v_j)_\sharp \eta = \Pi_j$.

Let us take

$$\Pi' \doteq \Pi + \frac{\lambda}{n} \sum_{j=1}^{n} \left[(u_j \times v_{j+1})_\sharp \eta - (u_j, v_j)_\sharp \eta \right].$$

This is a positive measure (it will be positive even if we remove the first term in the sum) and it has the same projections as Π. Finally, we get

$$\begin{aligned} C(\Pi') &= \int |x - y|^2 \, \mathrm{d}\Pi'(x, y) \\ &= C(\Pi) + \frac{\lambda}{n} \int f(u_1, \ldots, u_n, v_1, \ldots, v_n) \, \mathrm{d}\eta < C(\Pi). \end{aligned}$$

□

So we may say that every optimal plan is supported on the subdifferential of a convex function.

thm:transportmap **Theorem 1.7.** *Suppose that a measure $\Pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ is supported on the graph of the subgradient of a convex function ψ. Let μ and ν be the projections of Π. If μ vanishes on Borel sets of Hausdorff dimension $d - 1$ then $\nabla\psi$ pushes μ forward to ν; in fact, $\Pi = (\mathrm{id} \times \nabla\psi)_\sharp \mu$.*

Before passing to ther prove we need to establish two facts:

1. If f is convex and lsc then $\operatorname{graph}(\partial f)$ is closed.

   Let $(x_k, y_k) \in \partial f$ and $x_k \to x$, $y_k \to y$. Then for each z

   $$f(z) \geq \langle y_k, z - x_k \rangle + f(x_k).$$

   Taking $\liminf$ from both sides, we get the result.

2. If f is convex then there exists a convex lsc function $\bar{f}$ such that $\bar{f} = f$ on the interior of $\operatorname{dom} f$.

   One may try to take as $\bar{f}$ the lower-semicontinuous envelop of f. This is the greatest lsc function which is not greater then f and (see [Braides, p. 15]):
   $$\bar{f}(x) = \liminf_{y \to x} f(y),$$
   By [Dal Maso, Prop. 3.5(b)], $\operatorname{epi} \bar{f} = \overline{\operatorname{epi} f}$. This means that $\bar{f}$ is convex. It is known that f is locally Lipschitz on the interior of its domain. The formula from Braides implies that $\bar{f}$ and f coinsides on this set. In particular,
   $$\{x : \exists \nabla f\} = \{x : \exists \nabla \bar{f}\}.$$

Example 4. The function $f \colon [-1, 1] \to \mathbb{R}$ defined by
$$f(x) = \begin{cases} 0, & x \in (-1, 1), \\ 1, & x \in \{-1, 1\}. \end{cases}$$
is convex but not lsc. As a result $\partial f(-1) = \partial f(1) = \varnothing$. Its lsc envelop is given by $\bar{f} \equiv 0$.

Proof of Theorem 1.7. The above discussion implies that ψ can be taken so that $\partial\psi$ is closed. Since $\operatorname{spt} \Pi \subset \partial\psi \subset \operatorname{dom}\psi \times \mathbb{R}^d$, the left marginal of Π is supported on the closure of $\operatorname{dom}\psi$. First we must check that $\nabla\psi$ is defined on a set of full measure μ and is Borel measurable. Since $\operatorname{dom}\psi$ is convex, its boundary is $d-1$ dimensional. $\operatorname{dom}\nabla\psi$ coinsides with $\operatorname{dom}\psi$ everywhere except a $(d-1)$-dimensional set of singular points. Hence $\mu(\operatorname{dom}\nabla\psi) = 1$. Measurability of $\nabla\psi$ manifested since it coincides with the pointwise limit of a sequence of continuous approximants
$$\langle \nabla\psi(x), z \rangle = \lim_{n \to \infty} n(\psi(x + z/n) - \psi(z)).$$
If we show that $(\mathrm{id} \times \nabla\psi)_\sharp \mu = \Pi$, we are done by property 2 before Theorem 1.6.

Actually it suffices to show that these two measures coincide on products $M \times N$ of Borel sets $M, N \subset \mathbb{R}^d$. Define
$$S \doteq \{(x, \nabla\psi(x)) : x \in \operatorname{dom}\psi) = (\operatorname{dom}\nabla\psi \times \mathbb{R}^d) \cap \partial\psi.$$
The second formula and the fact that $\partial\psi$ is closed inplies that S is Borel. We know that Π is supported on $\partial\psi$ and $\mu(\operatorname{dom}\nabla\psi) = 1$, so S contains the

set $(\operatorname{dom}\nabla\psi \times \mathbb{R}^d) \cap \partial\psi$ of full measure for Π. Thus $\Pi(Z \cap S) = \Pi(Z)$ for any Borel $Z \subset \mathbb{R}^d \times \mathbb{R}^d$. Applied to

$$(M \times N) \cap S = \Big((M \times (\nabla\psi)^{-1}(N)) \times \mathbb{R}^d\Big) \cap S$$

(draw a picture), this yields

$$\begin{aligned}\Pi(M \times N) &= \Pi\Big((M \times (\nabla\psi)^{-1}(N)) \times \mathbb{R}^d\Big)\\ &= \mu(M \times (\nabla\psi)^{-1}(N))\\ &= (\mathrm{id} \times \nabla\psi)_\sharp \mu(M \times N).\end{aligned}$$

The latter equation follows from the following one

$$(u \times v)^{-1}(M \times N) = u^{-1}(M) \cap v^{-1}(N).$$

□

2 Duality

2.1 General duality theorem

We begin with, we prove a rather general duality theorem taken from the first chapter of [1].

Let E be a normed vector space and $f\colon E \to (-\infty, \infty]$ be a proper convex function ($\operatorname{dom} f \neq \varnothing$). Recall that

$$f^*(p) = \sup_{x \in E} \left[\langle p, x\rangle - f(x)\right], \quad p \in E^*,$$

is called the conjugate of f. The idea is simple: assume for some $p \in E^*$ there exists a supporting hyperplane $y = \langle p, x\rangle - \alpha$ of $\operatorname{epi} f$, that is,

$$\langle p, x\rangle - \alpha \leq f(x) \quad \forall x$$

and for some x^* the equality is attained. It is clear that $\alpha = f^*(p)$.

thm:duality **Theorem 2.1** (Fenchel-Rockafellar)**.** *Let $\varphi, \psi\colon E \to (-\infty, \infty]$ be two convex functions. Assume that there is some $x_0 \in \operatorname{dom}\varphi \cap \operatorname{dom}\psi$ such that φ is continuous at x_0. Then*

$$\inf_{x \in E}[\varphi(x) + \psi(x)] = \max_{p \in E^*}[-\varphi^*(-p) - \psi^*(p)] \tag{2.1}$$ eq:duality

Proof. We must show that

$$\sup_{p\in E^*} \inf_{x,y\in E} \left\{ \langle p, x-y\rangle + \varphi(x) + \psi(y) \right\} = \inf_{x\in E} \left\{ \varphi(x) + \psi(x) \right\}.$$

1. The choice $x = y$ shows that the left-hand side cannot be larger than the right hand side. So we just have to find a linear form $p \in E^*$ such that

$$\forall x, y \in E \qquad \langle p, x-y\rangle + \varphi(x) + \psi(y) \geq m \doteq \inf\{\varphi + \psi\}.$$

Since $\varphi(x_0) + \psi(x_0) < \infty$, the infimum is finite.

2. Then we consider two sets, one is the epigraph of φ and the other is the hypograph of $m - \psi$:

$$C = \operatorname{epi}\varphi, \quad C' = \operatorname{hypo}(m - \psi).$$

Since $(x_0, \varphi(x_0) + 1) \in C$ and φ is continuous at x_0, we conclude that $C^\circ \neq \varnothing$. Both C° and C' are convex and disjoint. Hence we can separate them by a nontrivial linear form $\langle p, x\rangle + \alpha\lambda = \beta$:

$$\langle p, x\rangle + \alpha\lambda \geq \langle p, y\rangle + \alpha\mu \quad \forall (x,\lambda) \in C^\circ \quad \forall (y,\mu) \in C'.$$

3. Note that $\alpha \geq 0$ because otherwise we may take the sequence (x_0, k), $k \to \infty$, and make the left-hand side as close to $-\infty$ as we want. Moreover, α cannot be zero because the hyperplane must separate the point $(x_0, m - \psi(x_0) - 1)$ and an open ball around $(x_0, \varphi(x_0) + 1)$. Hence, taking $q = p/\alpha$, we may rewrite the above inequality as

$$\langle q, x\rangle + \lambda \geq \langle q, y\rangle + \mu \quad \forall (x,\lambda) \in C \quad \forall (y,\mu) \in C',$$

where we took $C \subset \overline{C} = \overline{C^\circ}$ instead of C°, by continuity. In particular, we have

$$\langle q, x\rangle + \varphi(x) \geq \langle q, y\rangle + m - \psi(y) \quad \forall x, y \in E,$$

as desired. $\square$

2.2 Kantorovich duality

Before passing to the real theorem, let us go back to our bakery-café interpretation. Assume that we own this consortium and one day a man came to us and say: "Let's make a deal. I take bread from your bakeries and bring it to your cafés. You pay me $\varphi(x)$ dollars for loading 1 kilo of bread at bakery x, and $\psi(y)$ dollars for unloading 1 kilo of bread at café y. I guarantee that

$\varphi(x)+\psi(y)\le c(x,y)$. So you don't care about transporting anymore. Do you agree?" Of course, you do. This man will clearly try to maximize his profit and he will choose φ and ψ so that

$$\sup_{\varphi,\psi}\left\{\int \varphi\,\mathrm{d}\mu+\int \psi\,\mathrm{d}\nu \;:\; \varphi(x)+\psi(y)\le c(x,y)\right\}.$$

We will see this is exactly the dual problem for

$$\inf_{\pi\in\Gamma(\mu,\nu)}\int c(x,y)\,\mathrm{d}\pi\,(x,y).$$

In what follows, restrict ourself to the case where X and Y are compact and $c\colon X\times Y\to\mathbb{R}$ is nonnegative and continuous. Let us define two functionals

$$I[\pi]=\int c(x,y)\,\mathrm{d}\pi\,(x,y),\quad \pi\in P(X\times Y),$$
$$J[\varphi,\psi]=\int \varphi(x)\,\mathrm{d}\mu\,(x)+\int \psi(y)\,\mathrm{d}\mu\,(y),\quad (\varphi,\psi)\in\mathbf{C}^0(X)\times\mathbf{C}^0(Y),$$

and the set

$$\Phi_c=\left\{(\varphi,\psi)\in\mathbf{C}^0(X)\times\mathbf{C}^0(Y)\;:\;\varphi(x)+\psi(y)\le c(x,y)\right\}.$$

Clearly, the latter set is contained in the former.

thm:kdual **Theorem 2.2** (Kantorovich duality)**.** *Under the above assumptions*

$$\min_{\pi\in\Gamma(\mu,\nu)} I[\pi]=\max_{(\varphi,\psi)\in\Phi_c} J[\varphi,\psi].$$

Proof. 1. Note that

$$\sup_{\Phi_c} J[\varphi,\psi]\le\inf_{\Gamma(\mu,\nu)} I[\pi]. \tag{2.2}$$ eq:kdual1

Indeed, if $(\varphi,\psi)\in\Phi_c$ and $\pi\in\Gamma(\mu,\nu)$ then

$$\begin{aligned}J[\varphi,\psi]=\int\varphi\,\mathrm{d}\mu+\int\psi\,\mathrm{d}\nu&=\int[\varphi(x)+\psi(y)]\,\mathrm{d}\pi\,(x,y)\\&\le\int c(x,y)\,\mathrm{d}\pi\,(x,y)=I[\pi].\end{aligned}$$

2. Let us define two maps

$$F\colon u\in\mathbf{C}^0(X\times Y)\mapsto\begin{cases}0 & \text{if } u(x,y)\ge -c(x,y),\\ +\infty, & \text{otherwise.}\end{cases}$$

$$G\colon u \in \mathbf{C}^0(X\times Y) \mapsto \begin{cases} \int \varphi \,\mathrm{d}\mu + \int \psi \,\mathrm{d}\nu & \text{if } \varphi(x)+\psi(y) = u(x,y), \\ +\infty, & \text{otherwise.} \end{cases}$$

Note that G is well-defined: if $\varphi+\psi = \tilde\varphi+\tilde\psi$ then $\varphi = \tilde\varphi + a$ and $\psi = \tilde\psi - a$, for some c, and therefore $\int \varphi\,\mathrm{d}\mu + \int \psi\,\mathrm{d}\nu = \int \tilde\varphi\,\mathrm{d}\mu + \int \tilde\psi\,\mathrm{d}\nu$. Both matps F and G are convex and F is continuous at $u \equiv 1$. Hence we may apply Theorem 2.1. Let us compute both sides of (2.1). From the left-hand side we have

$$\begin{aligned} \inf\big\{F(u)+G(u)\big\} &= \inf\left\{\int \varphi\,\mathrm{d}\mu + \int \psi\,\mathrm{d}\nu \;:\; \varphi(x)+\psi(y) \geq -c(x,y)\right\} \\ &= -\sup\big\{J[\varphi,\psi] \;:\; (\varphi,\psi)\in\Phi_c\big\}\,. \end{aligned}$$

3. Next, we must compute Legendre transforms of F and G. First, for any $\pi \in M(X\times Y)$,

$$\begin{aligned} F^*(-\pi) &= \sup_{u\in\mathbf{C}^0(X\times Y)} \left\{-\int u(x,y)\,\mathrm{d}\pi\,(x,y) \;:\; u(x,y) \geq -c(x,y)\right\} \\ &= \sup_{u\in\mathbf{C}^0(X\times Y)} \left\{\int u(x,y)\,\mathrm{d}\pi\,(x,y) \;:\; u(x,y) \leq c(x,y)\right\} \end{aligned}$$

If π is not nonnegative then we may find a nonpositive function v such that $\int v\,\mathrm{d}\pi > 0$. The choise $u = \lambda v$, with $\lambda\to\infty$, shows that the supremum is $+\infty$. On the other hand, if π is nonegative then the supremum is $\int c\,\mathrm{d}\pi$. This means that

$$F^*(-\pi) = \begin{cases} \int c(x,y)\,\mathrm{d}\pi\,(x,y), & \pi \in M_+(X\times Y), \\ +\infty, & \text{otherwise.} \end{cases}$$

In the similar way we show that

$$G^*(\pi) = \begin{cases} 0, & \int(\varphi+\psi)\,\mathrm{d}\pi = \int\varphi\,\mathrm{d}\mu + \int\psi\,\mathrm{d}\nu\,, \\ +\infty, & \text{otherwise.} \end{cases}$$

So $F^*(-\pi) + G^*(\pi)$ is equal to the sum of I and indicator functions of $M_+(X\times Y)$ and $\Gamma(\mu,\nu)$, while the right-hand side of (2.1) is equal to

$$-\min_{\pi\in M(X\times Y)} \big[F^*(-\pi)+G^*(\pi)\big]\,.$$

$\square$

3 Geodesics

3.1 Basic metric geometry

Given a metric space (X, d) and a curve $\rho\colon I \to X$ in it, we say that ρ is *absolutely continuous* if there exists $g \in \mathbf{L}^1(I)$ such that

$$d(\rho_t, \rho_s) \leq \int_t^s g(r)\,\mathrm{d}r\,, \quad \forall t, s \in I,\ t \leq s.$$

It is well-known fact that for any absolutely continuous curve the limit

$$\lim_{h\to 0} \frac{d(\rho_{t+h}, \rho_t)}{h}$$

exists for a.e. $t \in I$ and defines a function $|\dot\rho| \in \mathbf{L}^1(I)$, witch is called *metric derivative.*

For a curve one may also introduce the notion of length:

$$\mathcal{L}(u) \doteq \sup \sum_{i=0}^{n-1} d(\rho_{t_i}, \rho_{t_{i+1}}),$$

where the supremum is taken over all finite partitions of I. If $\rho\colon I \to X$ happens to be absolutely continuous then

$$\mathcal{L}(\rho) = \int_I |\dot\rho_t|\,\mathrm{d}t\,.$$

Using a time change, any absolutely continuous curve may be converted to a constant speed curve $|\dot\rho_t| = C$.

3.2 Wasserstein space as a length space

A metric space X is called a *length space* if the distance between two points $x_0, x_1 \in X$ is the infimum of the lengths of all absolutely continuous curves connecting x_0 and x_1. The shortest curve is called a *geodesic.*

In the Wasserstein space $\mathcal{P}(X)$ constructing a geodesic is simple. Let Π be an optimal plan for a pair of measures ρ_0, ρ_1. Then the geodesic is given by the McCann interpolation:

$$\rho_t = ((1-t)\pi_1 + t\pi_2)_\sharp \Pi.$$

prop:geodesic **Proposition 3.1.** *Let Π be an optimal plan between ρ_0 and ρ_1. Then the curve ρ_t is a geodesic.*

Proof. For any time moments $0 \le s \le t \le 1$, consider the following transport plan

$$\Pi_{s,t} = \big((1-s)\pi_1 + s\pi_2, (1-t)\pi_1 + t\pi_2\big)_\sharp \Pi.$$

Note that $(\pi_1)_\sharp \Pi_{s,t} = \rho_s$ and $(\pi_2)_\sharp \Pi_{s,t} = \rho_t$, so $\Pi_{s,t}$ is a transport plan between ρ_s and ρ_t. This implies

$$\begin{aligned} W(\rho_s, \rho_t)^2 &\le \int |x-y|^2 \, \mathrm{d}(\,\Pi_{s,t})(x,y) \\ &= (t-s)^2 \int |x-y|^2 \, \mathrm{d}\Pi\,(x,y) \\ &= (t-s)^2 W(\rho_0, \rho_1)^2. \end{aligned}$$

In particular, ρ_t is 1-Lipschitz and $L(\rho) = W(\rho_0, \rho_1)$. □

Thus, the Wasserstein space is a length space. Actually, any geodesic takes this form (see Gigli's thesis for details).

Example 5. Consider two measures δ_0 and δ_1 on $\mathbb{R}$. Let us draw two curves joining them. One is $\gamma_t = (1-t)\delta_0 + t\delta_1$. It fails to be absolutely continuous because $W(\gamma_s, \gamma_t) = \sqrt{|t-s|}$. The other is given by the McCann's interpolation $\rho_t = ((1-t)\pi_1 + t\pi_2)_\sharp \Pi$, where $\Pi = \delta_{(0,1)}$ is the unique optimal transport plan between δ_0 and δ_1. Let us compute ρ_t. Given a measurable $E \subset \mathbb{R}$, we note that

$$\rho_t(E) = \delta_{(0,1)}\big(\{(x,y) \,:\, (1-t)x + ty \in E\}\big) = \begin{cases} 0, & t \notin E \\ 1, & t \in E \end{cases} = \delta_t(E).$$

3.3 AC curves satisfy the continuity equation

The central result of this section is the following theorem.

Theorem 3.2 (Characterisation of AC curves)**.** *Let $\rho_t \colon [0,1] \to \mathcal{P}(\mathbb{R}^d)$ be an absolutely continuous curve. Then there exists a time dependent vector field $v \in L^2(\rho)$ whose $L^2(\rho_t)$ norm, at a.e. time moment $t \in [0,1]$, is controlled by the metric derivative of ρ_t:*

$$\|v_t\|_{\rho_t} \le |\dot{\rho}_t|,$$

and such that the continuity equation

$$\partial_t \rho_t + \nabla \cdot (v_t \rho_t) = 0$$

holds in the sense of distributions. Here $\|v_t\|_{\rho_t} = \left(\int |v_t|^2 \, \mathrm{d}\rho_t\right)^{1/2}$.

The proof loosely follows Exercise 8.5 from [2]. This exercise is centered around the map defined through the expression $(d/dt)\int \varphi \,\mathrm{d}\rho_t$, where $\varphi \in \mathbf{C}_c^1(\mathbb{R}^d)$.

1. We begin by showing that the derivative above is well-defined. Fix two points on the curve, say ρ_s and ρ_t, and denote by $\Pi_{s,t}$ an optimal transport plan between them. Then

$$\begin{aligned}
\left|\int \varphi \,\mathrm{d}\rho_s - \int \varphi \,\mathrm{d}\rho_t\right| &= \left|\int [\varphi(x)-\varphi(y)] \,\mathrm{d}\Pi_{s,t}(x,y)\right| \\
&\leq \left(\int |\varphi(x)-\varphi(y)|^2 \,\mathrm{d}\Pi_{s,t}\right)^{1/2} \cdot \left(\int 1 \,\mathrm{d}\Pi_{s,t}\right)^{1/2} \\
&\leq \mathbf{Lip}\,(\varphi) W_2(\rho_s,\rho_t) \\
&\leq \mathbf{Lip}\,(\varphi)|\dot{\rho}_t|\,|s-t|.
\end{aligned}$$

Hence $t \mapsto \int \varphi \,\mathrm{d}\rho_t$ is Lipschitz and its derivative exists for a.e. t.

2. Now we are going to estimate $|(d/dt)\int \varphi \,\mathrm{d}\rho_t\,|$. For a.e. $t \in I$, we have

$$\begin{aligned}
\left|\frac{d}{dt}\int \varphi \,\mathrm{d}\rho_t\right| &= \lim_{h\downarrow 0} \frac{|\rho_{t+h}(\varphi)-\rho_t(\varphi)|}{h} \\
&\leq \lim_{h\downarrow 0} \frac{1}{h}\int |\varphi(x)-\varphi(y)| \,\mathrm{d}\Pi_{t,t+h}(x,y)\,. \qquad (3.1) \quad \boxed{\texttt{eq:dt1}}
\end{aligned}$$

To proceed, we shall require the following technical result.

Lemma 3.3. *For any $\varphi \in \mathbf{C}_c^1(\mathbb{R}^d)$, the map*

$$H_\varphi(x,y) \doteq \begin{cases} |\nabla\varphi(x)|, & x = y, \\ \frac{|\varphi(x)-\varphi(y)|}{|x-y|}, & \textit{otherwise.} \end{cases}$$

is bounded and upper semicontinuous.

Proof. The boundedness follows form the fact that φ is Lipschitz. Assume that $(x_k,y_k) \to (x,y)$ and $H_\varphi(x_k,y_k)$ converges. If $x \neq y$ then $x_k \neq y_k$, for all sufficiently large k, and therefore,

$$\lim_{k\to\infty} H_\varphi(x_k,y_k) = \lim_{k\to\infty} \frac{|\varphi(x_k)-\varphi(y_k)|}{|x_k-y_k|} = H_\varphi(x,y).$$

If $x = y$, there are two options. First: the sequence (x_k,y_k) contains a subsequence with $x_{k'} = y_{k'}$. In this case,

$$\lim_{k\to\infty} H_\varphi(x_k,y_k) = \lim_{k\to\infty} H_\varphi(x_{k'},y_{k'}) = \lim_{k\to\infty} |\nabla\varphi(x_k)| = H_\varphi(x,y).$$

Second: the sequence (x_k, y_k) contains a subseqence with $x_{k'} \neq y_{k'}$. Now,

$$\lim_{k\to\infty} H_\varphi(x_k, y_k) = \lim_{k\to\infty} H_\varphi(x_{k'}, y_{k'}) = \lim_{k\to\infty} \frac{|\varphi(x_{k'}) - \varphi(y_{k'})|}{|x_{k'} - y_{k'}|}.$$

Since

$$\varphi(a) - \varphi(b) = \int_0^1 \nabla\varphi((1-t)a + tb)\,\mathrm{d}t \cdot (a-b),$$

we conclude that

$$\frac{|\varphi(x_{k'}) - \varphi(y_{k'})|}{|x_{k'} - y_{k'}|} \leq \int_0^1 |\nabla\varphi((1-t)x_{k'} + ty_{k'})|\,\mathrm{d}t\,.$$

By passing to the limit, we finish the proof. □

Now, the right-hand side of (3.1) is smaller than

$$\lim_{h\downarrow 0} \frac{W_2(\rho_{t+h}, \rho_t)}{h} \left(\int H_\varphi^2(x,y)\,\mathrm{d}\Pi_{t,t+h}(x,y)\right)^{1/2}.$$

Note that $\Pi_{t,t+h}$ narrowly converge to $(\mathrm{id}\times\mathrm{id})_\sharp\rho_t$, which means that

$$\limsup_{h\downarrow 0} \int f(x,y)\,\mathrm{d}\Pi_{t,t+h}(x,y) \leq \int f(x,x)\,\mathrm{d}\rho_t(x)\,,$$

for any upper semicontinuous function f. Thus, for a.e. $t \in I$, we have

$$\left|\frac{d}{dt}\int \varphi\,\mathrm{d}\rho_t\right| = \lim_{h\downarrow 0} \frac{|\rho_{t+h}(\varphi) - \rho_t(\varphi)|}{h} \leq |\dot\rho_t|\,\|\nabla\varphi\|_{\rho_t}. \tag{3.2}$$ eq:bigestim

3. Consider the set V of all functions $\psi = \psi(t,x)$ such that $\psi(t,x) = \nabla\varphi(t,x)$ for some $\varphi \in \mathbf{C}_c^1([0,1]\times\mathbb{R}^d)$ and define on V the functional

$$L(\psi) = -\int \frac{\partial}{\partial t}\varphi(t,x)\,\mathrm{d}\rho\,.$$

To see that L is really a function of $\psi = \nabla\varphi$, we note that φ can be recovered from $\nabla\varphi$ in the following way:

$$\varphi(s,x) = \int_0^1 \frac{d}{dt}\varphi(s, y(t))\,\mathrm{d}t = \int_0^1 \nabla\varphi(s,y(t))\cdot\dot y(t)\,\mathrm{d}t\,,$$

where $y\colon I \to \mathbb{R}^d$ is a curve such that $y(0)$ lies outside $\operatorname{spt}\varphi$ and $y(1) = x$. It is obvious that L is linear.

4. Let us show that L is bounded. Clearly,

$$\int \frac{\partial}{\partial t}\psi(t,x)\,\mathrm{d}\rho = \lim_{h\downarrow 0}\int \frac{\varphi(t+h,x)-\varphi(t,x)}{h}\,\mathrm{d}\rho\,.$$

The integral in the limit can be rewritten using the disintegration of ρ as

$$\begin{aligned}\int \frac{\varphi(t+h,x)-\varphi(t,x)}{h}\,\mathrm{d}\rho = {} & \frac{1}{h}\int_0^1\int \varphi(t+h,x)\,\mathrm{d}\rho_t\,\mathrm{d}t \\ & -\frac{1}{h}\int_0^1\int \varphi(t,x)\,\mathrm{d}\rho_t\,\mathrm{d}t\,.\end{aligned}$$

After the change of the variables $t\mapsto t-h$, the first integral in the right-hand side takes the form

$$\frac{1}{h}\int_h^{1+h}\int \varphi(t,x)\,\mathrm{d}\rho_{t-h}\,\mathrm{d}t\,.$$

If h is sufficiently small then $\varphi(t,x)=0$ whenever $t\in(0,h)$, so the lower limit can be replaced by 0. Note that

$$\begin{aligned}\frac{1}{h}\int_1^{1+h}\int \varphi(t,x)\,\mathrm{d}\rho_{t-h}\,\mathrm{d}t = {} & \frac{1}{h}\int_1^{1+h}\int \varphi(t,x)\,\mathrm{d}\rho_t\,\mathrm{d}t \\ & +\frac{1}{h}\int_1^{1+h}\int \varphi(t,x)(\mathrm{d}\rho_{t-h}-\mathrm{d}\rho_t)\,\mathrm{d}t \\ & \le \frac{1}{h}\int_1^{1+h}\int \varphi(t,x)\,\mathrm{d}\rho_t\,\mathrm{d}t \\ & +\frac{1}{h}\int_1^{1+h}\int \mathbf{Lip}\,(\varphi)h\,\mathrm{d}t\,.\end{aligned}$$

Hence, after passing to the limit, we obtain

$$\lim_{h\downarrow 0}\frac{1}{h}\int_1^{1+h}\int \varphi(t,x)\,\mathrm{d}\rho_{t-h}\,\mathrm{d}t = \int \varphi(1,x)\,\mathrm{d}\rho_1 = 0,$$

and therefore

$$\frac{1}{h}\int_h^{1+h}\int \varphi(t,x)\,\mathrm{d}\rho_{t-h}\,\mathrm{d}t = \frac{1}{h}\int_0^1\int \varphi(t,x)\,\mathrm{d}\rho_{t-h}\,\mathrm{d}t + O(h).$$

This allows us to conclude that

$$\int \frac{\partial}{\partial t}\varphi(t,x)\,\mathrm{d}\rho = -\lim_{h\downarrow 0}\int_0^1 \frac{\rho_t(\varphi(t,\cdot))-\rho_{t-h}(\varphi(t,\cdot))}{h}\,\mathrm{d}t\,.$$

Now using the fact that $|\lim a_n| = \lim |a_n|$, then Fatou's lemma and equation (3.2), we get

$$\left|\int \frac{\partial}{\partial t}\varphi(t,x)\,\mathrm{d}\rho\right| \le \int_0^1 |\dot\rho_t|\,\|\nabla\varphi(t,x)\|_{\rho_t}\,\mathrm{d}t\,.$$

5. Let us denote by H the closure of V in $\mathbf{L}^2(\rho)$. By the previous estimate, we can extend L from V to H so that the norm of L is smaller than $\left(\int |\dot\rho_t|^2\right)^{1/2}$. Since H is a Hilbert space, we may use the Ritz representation theorem to show that there is a vector field $v \in H$ such that

$$\int \frac{\partial}{\partial t}\varphi(t,x)\,\mathrm{d}\rho = -L(\nabla\varphi) = -\int \nabla\varphi(t,x)\cdot v\,\mathrm{d}\rho\,.$$

It remains to show that $\|v_t\|_{\rho_t} \le |\dot\rho_t|$. We know that $\int_0^1 \|v_t\|_{\rho_t}^2 \le \int_0^1 |\dot\rho_t|^2$. Repeating the proof for the segment (s,t), we may construct w so that $\int_s^t \|v_\tau\|_{\rho_\tau}^2 \le \int_s^t |\dot\rho_\tau|^2$ and

$$\int_s^t \int \frac{\partial}{\partial\tau}\varphi(\tau,x)\,\mathrm{d}\rho_\tau\,\mathrm{d}\tau = -\int_s^t \int \nabla\varphi(\tau,x)\cdot w(\tau,x)\,\mathrm{d}\rho_\tau\,\mathrm{d}\tau\,.$$

Now it follows that $v = w$ ρ-a.e. and we are done.

3.4 Solutions of the continuity equation are AC

The converse is also true. If ρ_t is a solution of the continuity equation with a time dependent vector field $v \in L^2(\rho)$, then it is absolutely continuous.

3.5 Regular tangent space

Given an AC curve ρ_t, there are infinitely many families of vector fields $v_t \in \mathbf{L}^2(\rho_t)$ such that the corresponding continuity equation holds. Indeed, let us choose $w_t \in \mathbf{L}^2(\rho_t)$ so that $\nabla\cdot(w_t\rho_t) = 0$, then the pair $(\rho_t, v_t + w_t)$ also satisfies the continuity equation. Now, the idea is to select a unique v_t in some way.

The second part of the above theorem says that for any solution (ρ_t, v_t) of the continuity equation, we have

$$|\dot\rho_t| \le \|v_t\|_{\rho_t}.$$

On the other hand, according to the first part, one may choose v_t so that

$$|\dot\rho_t| = \|v_t\|_{\rho_t}.$$

Now, among all vector fields satisfying the continuity equation, let us choose the one with the minimal norm:

$$\|v\|_\vartheta \le \|v+w\|_\vartheta \quad \forall w \in \mathbf{L}^2(\vartheta) \ \text{ such that } \ \nabla\cdot(w\vartheta) = 0.$$

This may be equivalently rewritten as

$$\int v\cdot w \,\mathrm{d}\vartheta = 0 \quad \forall w \in \mathbf{L}^2(\vartheta) \ \text{ such that } \ \nabla\cdot(w\vartheta) = 0.$$

Hence v lies in the orthogonal subspace of

$$W = \{w \,:\, \nabla\cdot(w\vartheta) = 0\}.$$

But $\nabla\cdot(w\vartheta) = 0$ actually means that

$$\int \nabla\varphi\cdot w \,\mathrm{d}\vartheta = 0 \quad \forall \varphi \in C_c^\infty(\mathbb{R}^d).$$

Hence

$$W^\perp = \overline{\left\{\nabla\varphi \,:\, \varphi\in C_c^\infty(\mathbb{R}^d)\right\}},$$

where the closure is taken in $L^2(\vartheta)$.

A The Benamou-Bernier functional

We begin this section by recalling the Filippov lemma. It states that if $y(T)\subset f(X)$ then there exists x such that the diagram

$$\begin{tikzcd} Y & \\ T \arrow[u, "y \text{ meas.}"] \arrow[r, "x \text{ meas.}"'] & X \arrow[ul, "f \text{ cont.}"'] \end{tikzcd}$$

commutes.

lem:BB **Lemma A.1.** *Let*

$$K_2 = \left\{(a,b)\in\mathbb{R}\times\mathbb{R}^d \colon a + \frac{1}{2}|b|^2 \le 0\right\}.$$

Then, for all $(t,x)\in\mathbb{R}\times\mathbb{R}^d$, we have

$$\sup_{(a,b)\in K_2}\{at + b\cdot x\} = f_2(t,x) = \begin{cases} \frac{1}{2t}|x|^2, & t>0,\\ 0, & t=0, x=0,\\ +\infty, & t<0 \ \text{or}\ t=0, x\neq 0. \end{cases}$$

In particular, f_2 is convex and lsc.

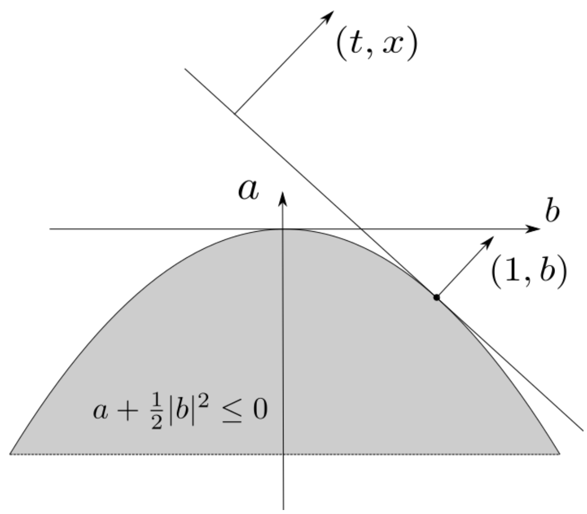

fig:BB

Figure 1: Illustration for Lemma A.1

Proof. We just have to maximize a linear function on a convex set. Figure 1 gives the solution. In the only nontrivial case ($t>0$), we have $(t,x) = \alpha(1,b)$, for some $\alpha > 0$. Hence $b = \frac{x}{t}$ and $a = -\frac{1}{2t^2}|x|^2$, and we get the result. Since f_2 is expressed as a supremum of linear maps, it is convex and lsc. $\square$

Now, for each couple (ρ, E), where $\rho \in \mathcal{M}(X)$ is a measure and $E \in \mathcal{M}(X;\mathbb{R}^d)$ is a vector measure, we correspond the number

$$\mathcal{B}_2(\rho,E) = \sup\left\{\int_X a(x)\,\mathrm{d}\rho\,(x) + \int_X b(x)\cdot \mathrm{d}E\,(x) \;:\; (a,b) \in C_b(X;K_2)\right\}.$$

Proposition A.2. *The map $\mathcal{B}_2$ is convex and lsc on $\mathcal{M}(X)\times\mathcal{M}(X;\mathbb{R}^d)$. Moreover,*

(i) $\mathcal{B}_2 \ge 0$,

(ii) $C_b(X;K_2)$ *can be replaced with* $L^\infty(X;K_2)$ *in the definition of* $\mathcal{B}_2$,

(iii) if ρ and E are absolutely continuous with respect to a positive measure λ then

$$\mathcal{B}_2(\rho,E) = \int f_2(\rho(x),E(x))\,\mathrm{d}\lambda\,(x).$$

(iv) $\mathcal{B}_2(\rho,E) < +\infty$ *only if* $\rho \ge 0$ *and* $E \ll \rho$,

(v) *for* $\rho \geq 0$ *and* $E \ll \rho$, *we have* $E = v\rho$ *and* $\mathcal{B}_2(\rho, E) = \frac{1}{2}\int |v|^2 \,\mathrm{d}\rho$,

(vi) *if* $X = \mathbb{R}^d$, $\rho^\varepsilon = \rho * \eta_\varepsilon$, $E^\varepsilon = E * \eta_\varepsilon$, *then* $\mathcal{B}_2(\rho^\varepsilon, E^\varepsilon) \leq \mathcal{B}_2(\rho, E)$.

Proof. Since $\mathcal{B}_2$ is a supremum of linear maps, its epigraph is convex and strongly closed. As a result it is weakly closed. Hence $\mathcal{B}_2$ is lsc on the space $\mathcal{M}(X) \times \mathcal{M}(X; \mathbb{R}^d)$ equipped with the weak topology.

Note that $(0, 0) \in C_b(X; K_2)$ implies (i). Now, assume for a moment that (ii) holds. If $\rho = \rho\lambda$ and $E = E\lambda$ then

$$\begin{aligned}\mathcal{B}_2(\rho, E) &= \sup\left\{\int_X \Big[a(x)\rho(x) + b(x)\cdot E(x)\Big]\,\mathrm{d}\lambda\,(x) \,:\, (a, b) \in L^\infty(X; K_2)\right\}\\ &\leq \int_X f_2(\rho(x), E(x))\,\mathrm{d}\lambda\,(x).\end{aligned}$$

To prove the opposite inequality, we must consider two cases.

(1) Assume that $f_2(\rho(x), E(x))$ is finite for λ-a.e. $x \in X$. Then, according to the Filippov lemma, there are measurable $a(\cdot)$ and $b(\cdot)$ such that

$$f_2(\rho(x), E(x)) = a(x)\rho(x) + b(x)\cdot E(x) \quad \text{for } \lambda\text{-a.e. } x \in X.$$

(2) Assume that $f_2(\rho(x), E(x))$ is infinite on a set of positive measure λ. There two more cases:

(a) There exists A such that $\lambda(A) > 0$ and $\rho(x) < 0$, for all $x \in A$. Then we take $a_k = -k1_A$, $b_k = 0$ and note that

$$\mathcal{B}_2(\rho, E) \geq \int (a_k\rho + b_k\cdot E)\,\mathrm{d}\lambda \to +\infty. \tag{A.1}$$ eq:toinf

(b) There exists A such that $\lambda(A) > 0$ and $\rho(x) = 0$, for all $x \in A$. Then $|E(x)| \neq 0$, for all $x \in A$. Hence by choosing $a_k = -\frac{k^2}{2}1_A$, $b_k = k\frac{E}{|E|}1_A$, we again arrive at (A.1).

Note that (iii) implies (v). Indeed, choosing $\rho = 1\lambda$ and $E = v\lambda$ gives

$$\mathcal{B}_2(\rho, v\rho) = \int f_2(1, v(x))\,\mathrm{d}\lambda\,(x) = \frac{1}{2}\int |v(x)|^2\,\mathrm{d}\lambda\,(x).$$

Let us prove (iv). If there is A such that $\rho(A) < 0$, we can take the sequence from (a) and obtain $\mathcal{B}_2(\rho, E) \geq k\rho(A)$. If there is A such that

$\rho(A) = 0$ and $E(A) \neq 0$, we can take the sequence $a_k = -\frac{k^2}{2} 1_A$, $b_k = ke1_A$, where e is a unit vector from X, and get $\mathcal{B}_2(\rho, E) \geq ke \cdot E(A)$. Since e is arbitrary, we get $\mathcal{B}_2(\rho, E) = +\infty$.

Finally, note that (ii) follows by the usual approximation arguments and (vi) is not so important for now. $\square$

References

[1] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations.* New York, NY: Springer, 2011.

[2] C. Villani. *Topics in optimal transportation.*, volume 58. Providence, RI: American Mathematical Society (AMS), 2003.