Introduction to Optimal Transport

Lectured by Matthew Thorpe

Lent 2019

(20th February, Wednesday)

5 Semi-Discrete Optimal Transport

Assume

1. $\nu = \sum_{j=1}^n m_j \delta_{y_j}$, $\{m_i\}_{i=1}^n \subset [0,1]$, $\sum_{j=1}^n m_j = 1$ and $\{y_j\}_{j=1}^n \subset \mathbb{R}^d$.
2. $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ has density ρ.

Definition 5.1) The *Laguerre diagram (power diagram)* for a set of points $\{y_j\}_{j=1}^n$ and weight $\{w_j\}_{j=1}^n \subset \mathbb{R}$ is the collection of sets

$$L_j = \{x \in \mathbb{R}^d : |x-y|^2 - w_j < |x-y_i|^2 - w_i \ \forall i \neq j\}$$

for $j = 1, \cdots, n$.

Comments :

- If $w_j = 0$, then the *Laguerre diagram* are Voronoi cells.
- Each L_i is open.

Aims :

- Show there exists optimal $T : \mathbb{R}^d \to \mathbb{R}^d$ that defines a Laguerre diagram.
- and the weights $\{w_j\}_{j=1}^n$ for the optimal Laguerre diagram solve a concave varaiational problem

$$\max g(w) \quad \text{where } w = (w_1, \cdots w_n)$$

where g is as defined in the next lemma.

Lemma 5.2) Let $\rho \in L^1(\mathbb{R}^d)$ be a probability density, $\{m_j\}_{j=1}^m \subset [0,1]$ satisfy $\sum_{j=1}^m m_j = 1$ and $\{y_j\}_{j=1}^m \subset \mathbb{R}^d$. Then $g : \mathbb{R}^n \to \mathbb{R}$ defined by

$$g(w) = \int_{\mathbb{R}^d} \inf_j (|x-y_j|^2 - w_j)\rho(x)dx + \sum_{j=1}^n w_j m_j \quad \cdots\cdots\cdots \ (\star)$$

is concave.

Idea of proof : Introduce $\gamma : \mathbb{R}^n \to \{1, \cdots, n\}$. Define

$$G(\gamma, w) = \int_{\mathbb{R}^d} \Big(|x - y_{\gamma(x)}|^2 - w_{\gamma(x)}\Big)\rho(x)dx + \sum_{j=1}^{n} w_j m_j$$

Note (1) $w \mapsto G(\gamma, w)$ is an affine function, so concave and (2) $g(w) = \inf_\gamma G(\gamma, w)$. So g is also concave.

Lemma 5.3) Define g by $(\star)$ for $\rho \in L^1(\mathbb{R}^d)$, $\{y_j\}_{j=1}^n \subset \mathbb{R}^n$, $\{m_j\}_{j=1}^n \subset \mathbb{R}$. Let $\{L_i(w)\}_{i=1}^n$ be a *Laugerre diagram* with weights w and points $\{y_j\}_{j=1}^n$. Then

$$\frac{\partial g}{\partial w_i}(w) = -\int_{L_i(w)} \rho(x)dx + m_i$$

sketch proof) Let $\alpha_j(x, w) = \chi_{L_j(w)}(x)(|x - y_j|^2 - w_j)\rho(x)$ so

$$g(w) = \sum_{j=1}^{n} \Big(\int_{\mathbb{R}^d} \alpha_j(x, w)dx + w_j m_j \Big)$$

For any $x \in L_i(w)$, we have $\chi_{L_i(w+\pm te_i)}(x) = \chi_{L_j(w)}(x)$ for $t > 0$ sufficiently small, where e_i is the unit vector in i-direction. Moreover,

$$\frac{1}{t}\Big(\alpha_j(x, w + te_i) - \alpha_j(x, w)\Big) = -\chi_{L_j(w)}(x)\delta_{ij}\rho(x)$$

Hence

$$\begin{aligned}
\frac{\partial g}{\partial w_i}(w) &= \lim_{t\to 0^+} \frac{1}{t}\Big(g(w + te_i) - g(w)\Big) \\
&= \sum_{j=1}^{n} \lim_{t\to 0^+} \Big[\int_{\mathbb{R}^d} \frac{1}{t}\Big(\alpha_j(x, w + te_i) - \alpha_j(x, w)\Big)dx + (w_j + \delta_{ij}t)m_j - w_j m_j \Big] \\
&= -\int_{\mathbb{R}^d} \chi_{L_i(w)}(x)\rho(x)dx + m_i
\end{aligned}$$

(End of proof) $\square$

Then comes the main result of the section.

Theorem 5.4) Assume $\{y_j\}_{j=1}^m \subset \mathbb{R}$, $\{m_j\}_{j=1}^m \subset [0, 1]$, $\sum_{j=1}^n m_j = 1$ and $\nu = \sum_{j=1}^n m_j \delta_{y_j}$. Let $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ have density ρ and $g(w)$ is defined by $(\star)$, maximised by $w = (w_1, \cdots, w_n)$, and let $\{L_j\}_{j=1}^n$ be the corresponding *Laguerre diagram.* Now define

$$\begin{aligned}
&T^\dagger(x) = y_j \quad \text{if } x \in L_j \quad \text{(which defines } \mu\text{-a.e.)} \\
&\psi^\dagger(y_j) = w_j, \quad \varphi^\dagger(x) = \inf_j(|x - y_j|^2 - w_j)
\end{aligned}$$

Then,

1. $T^\dagger$ is a solution to the MOT problem with cost $c(x, y) = |x - y|^2$.
2. $(\varphi^\dagger, \psi^\dagger)$ are an optimal pair for the Kantorovich dual problem with cost $c(x, y) = |x - y|^2$.

   proof) We first assume that $T^\dagger$, $\varphi^\dagger$ and $\psi^\dagger$ are admissible for the optimisation problem:

   (a) $\varphi^\dagger \in L^1(\mu)$, (b) $T^\dagger_{\#}\mu = \nu$, (c) $\int_{L_j} \rho(x)dx = m_j$

Assume (a), (b), (c) then

$$\varphi^\dagger(x) + \psi^\dagger(y_i) = \inf_j \left(|x - y_j|^2 - w_j\right) + w_i \le |x - y_i|^2 = c(x, y_i)$$

So $(\varphi^\dagger, \psi^\dagger) \in \Phi_c$. Now we have

$$\begin{array}{rll} \mathbb{M}(T^\dagger) \ge & \inf_{T_\# \mu = \nu} \mathbb{M}(T) & \text{(by (b))} \\ \ge & \min_{\pi \in \Pi(\mu,\nu)} \mathbb{K}(\pi) & \\ = & \sup_{(\varphi,\psi)\in\Phi)c} \mathbb{J}(\varphi, \psi) & \text{by Theorem 4.1} \\ \ge & (\varphi^\dagger, \psi^\dagger) & \end{array} \qquad \cdots\cdots\cdots \ (\oplus)$$

and

$$\begin{aligned} \mathbb{J}(\varphi^\dagger, \psi^\dagger) &= \int_{\mathbb{R}^d} \varphi^\dagger(x)\rho(x)dx + \sum_{j=1}^n m_j \psi^\dagger(y) \\ &= \sum_{j=1}^n \left(\int_{L_j} \varphi^\dagger(x)\rho(x)dx + m_j \psi^\dagger(y_j) \right) \\ &= \sum_{j=1}^n \int_{L_j} \left(\left(|x - y_j|^2 - w_j\right)\rho(x)dx + m_j w_j \right) \\ &= \sum_{j=1}^n \int_{L_j} |x - y_j|^2 \rho(x)dx \\ &= \sum_{j=1}^n \int_{L-j} |x - T^\dagger(x)|^2 \rho(x)dx = \mathbb{M}(T^\dagger) \end{aligned}$$

Hence all inequalities in $(\oplus)$ are all equalities and in particular

$$\begin{aligned} \mathbb{M}(T^\dagger) &= \min_{T : T_\# \mu = \nu} \mathbb{M}(T) \\ \mathbb{J}(\varphi^\dagger, \psi^\dagger) &= \sup_{(\varphi,\psi)\in\Phi_c} \mathbb{J}(\varphi, \psi) \end{aligned}$$

Now we are left to prove assumptions (a), (b), (c).

(a) Has

$$-\sup_j w_j \le \varphi^\dagger(x) \le |x - y_i|^2 - w_i$$

for any i so $|\varphi^\dagger(x)| \le 2|x|^2 + C$ for a constant $C = 2|y_1|^2 - w_1 + \sup_j w_j$ and

$$\left\|\varphi\right\|_{L^1(\mu)} \le 2 \int_{\mathbb{R}^d} |x|^2 d\mu(x) + C < +\infty$$

(b) Pick $i \in \{1, \cdots, n\}$. Then

$$\begin{aligned} \mu((T^\dagger)^{-1}(y_i)) =& \mu(\{x : T^\dagger(x) = y_i\}) = \mu(L_i) = m_i \quad \text{(by (c))} \\ =& \nu(\{y_i\}) \end{aligned}$$

So $T^\dagger_\# \mu = \nu$ as required.

(c) Since w maximises g, has $\frac{\partial g}{\partial w_i}(w) = 0$ for each i, but $\frac{\partial g}{\partial w_i}(w) = -\int_{L_j} \rho(x)dx + m_j$ by **Lemma 5.3** so we have the result

(End of proof) □

(22nd February, Friday)

A proof for yesterday Example Class : when showing $\sup_{Ax\leq b} c\cdot x = \min_{y\geq 0, A^Ty=c} b\cdot y$. We first assume that $\exists y_0$ such that $y_0 \geq A^T y_0 = c$ (if not, can just define that the minimum takes value ∞)

Correction 1 : We defined $\Xi(x) = c\cdot x$, $\Xi : E \to \mathbb{R}$ by $A^T y_0 \cdot x = y_0 \cdot (Ax)$ so Ξ is indeed well-defined.
Correction 2 : We should have $E^* \cong \mathrm{Range}(A) \subset \mathbb{R}^m$.

6 Existence and Characterisation of Transport Maps

Aims in this chapter are the following.

(1) To find a sufficient condition for the existence of Optimal Transport mapos for the MOT problem.

(2) To find a sufficient condition for the Monge cost to equal the Kantorovich cost.

(3) To characterise Optimal Transport maps and plans.

The structure would look like

6.1 Sate main results

6.2 Background on convex analysis

6.3 Prove main results

6.1 Knott-Smith Optimality and Beiner's Theorem

Definition) The **subdifferential** of a convex function φ is defined to be

$$\partial\varphi(z) = \{y : \varphi(z) \geq \varphi(x) + y\cdot(z-x) \quad \forall z \in \mathbb{R}^d\}$$

This is a *set of slopes*.

Comments :

(1) Subdifferential always exists for convex lower semi-continuous functions.

(2) if φ is differentiable at x, then $\partial\varphi(x) = \{\nabla\varphi(x)\}$.

Theorem 6.1) *(Knott-Simith Optimality Criterion)* Let $\mu \in \mathcal{P}_2(X)$, $\nu \in \mathcal{P}_2(Y)$, $X, Y \subset \mathbb{R}^d$, $c(x,y) = \frac{1}{2}|x-y|^2$. Then $\pi^\dagger \in \Pi(\mu,\nu)$ minimisies the KOT problem *iff* $\exists \tilde{\varphi}^\dagger \in L^1(\mu)$ convex and lower semi-continuous such that $\mathrm{supp}(\pi^\dagger) \subset \mathrm{Gra}(\partial\tilde{\varphi}^\dagger)$ (equivalent to having $y \in \partial\tilde{\varphi}^\dagger(x)$ for $\pi^\dagger$-a.e. (x,y)). Moreover the pair $(\tilde{\varphi}^\dagger, (\tilde{\varphi}^\dagger)^c)$ is a minimiser of $\inf_{\tilde{\Phi}} \mathbb{J}$, where $\tilde{\Phi} = \{(\tilde{\varphi}, \tilde{\psi}) \in L^1(\mu)\times L^1(\nu) : \tilde{\varphi}(x) + \tilde{\psi}(y) \geq x\cdot y\}$.

(Previously, $\sup_{(\varphi,\psi)\in\Phi} \mathbb{J}(\varphi,\psi)$ was the dual problem. Here we are interested in the probelms $\inf_{(\tilde{\varphi},\tilde{\psi})\in\tilde{\Phi}} \mathbb{J}(\tilde{\varphi},\tilde{\psi})$. The two problems can be related by : $(\tilde{\varphi},\tilde{\psi}) \in \tilde{\Phi}$ minimises $\mathbb{J}$ over $\tilde{\Phi}$ *iff* $(\varphi,\psi) \in \Phi$ maximises $\mathbb{J}$ over Φ where $\tilde{\varphi}(x) = \frac{1}{2}|x|^2 - \varphi(x)$, $\tilde{\psi}(y) = \frac{1}{2}|y|^2 - \psi(y)$.)

In 1D we expect monotonicity, and the theorem is almost equivalent to monotonicity : if $x_1 \le x_2$ then $T^\dagger(x_1) \le T^\dagger(x_2)$. Hence $\mathrm{supp}(\pi^\dagger) = \{(x,y) : x \in X, y = T(x)\}$ is **cyclically monotone** - if $(x_1,y_1),(x_2,y_2) \in \mathrm{supp}(\pi^\dagger)$ and $x_1 \le x_2$ then $y_1 \le y_2$. Any cyclically monotone set can be written as a subdifferential of a convex function. This argument holds for higher dimensions and if T is “set valued”.

Theorem 6.2) *(Brenier’s Theorem)* Let $\mu \in \mathcal{P}_2(X)$, $\nu \in \mathcal{P}_2(Y)$, $X, Y \subset \mathbb{R}^d$ and $c(x,y) = \frac{1}{2}|x-y|^2$. Assume that μ does not give mass to small sets (a small set is any set with Hausdorff dimenstion at most $d-1$). Then there is a unique $\pi^\dagger \in \Pi(\mu,\nu)$ that minimises the KOT problem.

Moreover, $\pi^\dagger$ satisfies $\pi^\dagger = (id \times \nabla\tilde{\varphi})_{\#}\mu$ where $\nabla\tilde{\varphi}$ is the unique gradient of a convex function that pushes μ forward to ν (that is, $(\nabla\tilde{\varphi})_{\#}\mu = \nu$) and $(\tilde{\varphi},\tilde{\varphi}^c)$ minimise $\mathbb{J}$ over $\tilde{\Phi}$.

Comments :

(1) $\pi^\dagger = (id \times \nabla\tilde{\varphi})_{\#}\mu \Leftrightarrow d\pi^\dagger(x,y) = \delta_{\nabla\tilde{\varphi}(x)}(y) \times d\mu(x)$.

(2) We will show that, in **Proposition 6.5**, convex functions are differentiable Lebesgue almost everywhere. Since μ gives zero mass to ses of Lebesue measure 0, then any convex function is differentiable μ almost everywhere.

Corollary 6.3) Under the same assumptions as **Theorem 6.2**, $\nabla\tilde{\varphi}$ is the unique solution to the MOT problem, *i.e.*

$$\frac{1}{2}\int_X |x - \nabla\tilde{\varphi}(x)|^2 d\mu(x) = \inf_{T:T_{\#}\mu=\nu} \frac{1}{2}\int_X |x - T(x)|^2 d\mu(x)$$

proof) Since $\min \mathbb{K} \le \inf \mathbb{M}$, and $T^\dagger_{\#}\mu = \nu$ by **Theorem 6.2**, it is enough to show that $T^\dagger = \nabla\tilde{\varphi}$ satisfies $\mathbb{M}(T^\dagger) = \min \mathbb{K} = \mathbb{K}(\pi^\dagger)$. Indeed,

$$\begin{aligned}
\mathbb{M}(T^\dagger) =& \frac{1}{2}\int_X |x - T^\dagger(x)|^2 d\mu(x) \\
=& \frac{1}{2}\int_{X\times Y} |x - T^\dagger(x)|^2 d\pi^\dagger(x,y) \\
=& \frac{1}{2}\int_{X\times Y} |x - y|^2 d\pi^\dagger(x,y) \quad \text{(since } y = T^\dagger(x),\ \pi^\dagger\text{-a.e.)} \\
=& \mathbb{K}(\pi^\dagger)
\end{aligned}$$

(End of proof) $\square$

6.2 Preliminary Results for Convex Analysis

Just in this seciton, we will write φ rather than $\tilde{\varphi}$.

Recall the *Legendre-Fenchel transform*, or the convex conjugate defined by

$$\varphi^*(y) = \sup_{x\in\mathbb{R}^d} (x \cdot y - \varphi(x))$$

Proposition 6.4) Let $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ be proper (not identically $+\infty$), lower semi-continuous and convex fuction. Then $\forall x, y \in \mathbb{R}^d$,

$$xy = \varphi(x) + \varphi^*(y) \quad \Leftrightarrow \quad y \in \partial\varphi(x)$$

proof) Note, by definition, $\varphi^*(y) \geq x \cdot y - \varphi(x)$ for all $x \in \mathbb{R}^d$. So

$$\begin{aligned} x \cdot y = \varphi(x) + \varphi^*(y) \quad &\Leftrightarrow \quad x \cdot y \geq \varphi(x) + \varphi^*(y) \\ &\Leftrightarrow \quad x \cdot y \geq \varphi(x) + y \cdot z - \varphi(z) \quad \forall z \in \mathbb{R}^d \\ &\Leftrightarrow \quad \varphi(z) \geq \varphi(x) + y \cdot (z - x) \quad \forall z \in \mathbb{R}^d \\ &\Leftrightarrow \quad y \in \partial\varphi(x) \end{aligned}$$

(End of proof) $\square$

Proposition 6.5) Let $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be convex, then

(1) φ is differential Lebesgue-almost everywhere on the interior of its domain and

(2) Whenever φ is differentiable, has $\partial\varphi(x) = \{\nabla\varphi(x)\}$.

(25th February, Monday)

We use the following theorem for the proof.

Rademacher's Theorem) If $U \subset \mathbb{R}^d$ is open and $f : U \to \mathbb{R}$ is Lipschitz continuous then f is differentiable a.e.

We do not prove this results.

Proposition 6.5) Let $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be convex, then

(1) φ is differential Lebesgue-almost everywhere on the interior of its domain and

(2) Whenever φ is differentiable, has $\partial\varphi(x) = \{\nabla\varphi(x)\}$.

proof)

(1) Let $x \in \text{int}(\text{Dom}(\varphi))$ and δ^* be such that $\overline{B(x, \delta^*)} \subset \text{int}(\text{Dom}(\varphi))$. We show that φ is Lipschitz continuous on $B(x, \delta^*/4)$. Then by *Rademacher's Theorem*, φ is a.e. differentiable in $B(x, \delta^*/4)$. Then φ is differentiable a.e. on $\text{int}(\text{Dom}(\varphi))$.
(a) We show φ is uniformly bounded on $\overline{B(x, \delta^*/2)}$. Let Q be a cuboid centred at x with sides of legth $\sqrt{\frac{4}{d}}\delta^*$. Let $\{x_i\}_{i=1}^n$ be the corners of Q. Note $x \in \partial B(x, \delta^*)$ and that the set of exterme points of Q is $\underline{\{x_i\}_{i=1}^d}$. by the *Minkowski-Carethéodory theorem* (Theorem 3.5), for each $y \in \overline{B(x, \delta^*/\sqrt{d})}$, $\exists \{\lambda_i\}_{i=1}^{2^d} \subset [0,1]$ such that $\sum_{i=1}^{2^n} \lambda_i = 1$ and $y = \sum_{i=1}^{2^d} \lambda_i x_i$. So

$$\varphi(y) = \varphi\Big(\sum_{i=1}^{2^d} \lambda_i x_i\Big) \leq \sum_{i=1}^{2^d} \lambda_i \varphi(x_i) \leq \max_i |\varphi(x_i)| = C$$

where the first inequality follows from convexity of φ. Now define $y' = (x - (y - x)) = 2x - y \in B(x, \delta^*/2)$. Since $x = \frac{1}{2}y' + \frac{1}{2}y$, has

$$\begin{aligned} &\varphi(y) \geq 2\varphi(x) - \varphi(y') \geq 2\varphi(x) - C \\ \Rightarrow \quad &2\varphi(x) - C \leq \varphi(y) \leq C, \quad \forall y \in \overline{B(x, \delta^*/\sqrt{d})} \end{aligned}$$

so

$$\left\|\varphi\right\|_{L^\infty(\overline{B(x,\delta^*/\sqrt{d})})} \leq \max\{C-2\varphi(x), C\} = M < \infty$$

(b) We show φ is Lipschitz continuous on $B(x,\delta^*/2\sqrt{d})$. Let $x_1, x_2 \in B(x,\delta^*/2\sqrt{d})$ where $x_1 \neq x_2$. Take x_3 to be the intersectin of the line throgh x_1 and x_2 with $\partial B(x,\delta^*/\sqrt{d})$ and choose x_3 such that x_2 lies between x_1 and x_3. Define $\lambda = \frac{|x_2-x_3|}{|x_1-x_3|} \in (0,1)$. Now,

$$\begin{aligned}
\lambda x_1 + (1-\lambda)x_3 =& \lambda x_2 + \lambda(x_1-x_2) + (1-\lambda)x_2 + (1-\lambda)(x_3-x_2) \\
=& x_2 + \frac{|x_2-x_3|}{|x_1-x_3|}(x_1-x_2) + \frac{|x_1-x_3|-|x_2-x_3|}{|x_2-x_3|}(x_3-x_2) \\
=& x_2 + \frac{1}{|x_1-x_3|}\Big(|x_2-x_3|(x_1-x_2) + |x_1-x_2|(x_3-x_2)\Big) \\
=& x_2
\end{aligned}$$

By convexity of φ,

$$\begin{aligned}
&\varphi(x_2) \leq (1-\lambda)\varphi(x_3) + \lambda\varphi(x_1) \\
\Rightarrow \quad &\varphi(x_2) - \varphi(x_1) \leq (1-\lambda)(\varphi(x_3)-\varphi(x_1)) = \frac{|x_1-x_2|}{|x_1-x_3|}(\varphi(x_3)-\varphi(x_1)) \\
&\leq \frac{2\sqrt{d}\times 2M}{\delta^*}|x_1-x_2| = L|x_1-x_2|
\end{aligned}$$

with $L = 4\sqrt{d}M/\delta^*$. So φ is Lipschitz. *[This proof looks very complicated but the idea is very simple!]*

(2) Let φ be differentiable at x. Then

$$\begin{aligned}
\varphi(x) + \nabla\varphi(x)\cdot(z-x) =& \varphi(x) + \lim_{h\to 0^+}\left(\frac{\varphi(x+(z-x)h)-\varphi(x)}{h}\right) \\
=& \varphi(x) + \lim_{h\to 0^+}\left(\frac{\varphi((1-h)x+zh)-\varphi(x)}{h}\right) \\
\leq& \varphi(x) + \lim_{h\to 0}\frac{(1-h)\varphi(x)+h\varphi(z)-\varphi(x)}{h} = \varphi(z)
\end{aligned}$$

so $\nabla\varphi(x) \in \partial\varphi(x)$.

On the other hand, if $y \in \partial\varphi(x)$, then $\varphi(x) + y\cdot(z-x) \leq \varphi(x)$ for all $z \in \mathbb{R}^d$ so by letting $z = x+hw$ with $h>0$, we get

$$y\cdot w \leq \frac{\varphi(x+hw)-\varphi(x)}{h}$$

Let $h \to 0^+$, then $y\cdot w \leq \nabla\varphi(x)\cdot w$. By symmetry ($w \mapsto -w$), we also have $y\cdot w = \nabla\varphi(x)\cdot w$ for all $w \in \mathbb{R}^d$. Hence $y = \nabla\varphi(x)$.

(End of proof) □

Proposition 6.6) Let $\varphi : \mathbb{R}\cup\{+\infty\}$ be proper. then the following are equivalent.

(1) φ is convex and lower semi-continuous.

(2) $\varphi = \psi^*$ for some proper function ψ.

(3) $\varphi^{**} = \varphi$.

proof) The implications from (3) to (2) is immediate.

We are left to show implication (2) $\Rightarrow$ (1). Assume (2), so $\varphi = \psi^*$ for some proper function ψ. Let $x_1, x_2 \in \mathbb{R}^d$, $t \in [0,1]$, then

$$\begin{aligned}\varphi(tx_1 + (1-t)x_2) &= \sup_y \Big((tx_1 + (1-t)x_2)\cdot y - \psi(y)\Big) \\ &\leq \sup_y \Big(t(x_1\cdot y) - \psi(y)\Big) + \sup_y \Big((1-t)(x_2 \cdot y - \psi(y)\Big) \\ &= t\varphi(x_1) + (1-t)\varphi(x_2)\end{aligned}$$

so φ is convex. To show lower semi-continuity, let $x_m \to x$. Then

$$\begin{aligned}\liminf_{m\to\infty} \varphi(x_m) =& \liminf_{m\to\infty} \sup_y (x_m \cdot y - \psi(y)) \geq \lim_{m\to\infty} (x_m \cdot y - \psi(y)) \\ =& x\cdot y - \psi(y) \quad \text{for any } y \in \mathbb{R}^d\end{aligned}$$

so $\liminf_{m\to\infty} \varphi(x_m) \geq \sup_y (x\cdot y - \psi(y)) = \varphi(x)$, so φ is lower semi-continuous.

(27th February, Wednesday)

proof continued) Next we prove implication (1) to (3). Suppose φ is convex and lower semi-continuous. Assume $x \in \mathrm{int}(\mathrm{Dom}(\varphi))$ and we show $\varphi^{**}(x) = \varphi(x)$. Since φ can be bounded below by an affine function passing through $(x, \varphi(x))$, we have $\partial\varphi(x) \neq \phi$. Let $y_0 \in \partial\varphi(x)$. By **Proposition 6.4**, $x\cdot y_0 = \varphi(x) + \varphi^*(y_0)$, so

$$\varphi(x) = x\cdot y_0 - \varphi^*(y_0) \leq \sup_{y\in\mathbb{R}^n} (x\cdot y - \varphi^c(y)) = \varphi^{**}(x)$$

On the other hand, since $\varphi^*(y) \geq x\cdot y - \varphi(x)$ for all $y \in \mathbb{R}^d$, has

$$\varphi(x) \geq \sup_y (x\cdot y - \varphi^*(y)) = \varphi^{**}(x)$$

so $\varphi(x) = \varphi^{**}(x)$ on $\mathrm{int}(\mathrm{Dom}(\varphi))$. See note for the case $x \notin \mathrm{int}(\mathrm{Dom}(\varphi))$.

(End of proof) $\square$

6.3 Proof of the Knott-Smith Optimality Criterion

Let $c(x,y) = \frac{1}{2}|x-y|^2$, $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$. We first make few observations.

(A) Let $(\varphi, \psi) \in \Phi_c$. Define $\tilde{\varphi}(x) = \frac{1}{2}|x|^2 - \varphi(x)$ and $\tilde{\psi}(y) = \frac{1}{2}|y|^2 - \psi(y)$. One can see that $\tilde{\varphi} \in L^1(\mu)$, $\tilde{\psi} \in L^1(\nu)$. and

$$\begin{aligned}\tilde{\varphi}(x) + \tilde{\psi}(y) =& \frac{1}{2}|x|^2 + \frac{1}{2}|y|^2 - \varphi(x) - \varphi(y) \\ \geq& \frac{1}{2}|x|^2 + \frac{1}{2}|y|^2 - \frac{1}{2}|x-y|^2 = x\cdot y\end{aligned}$$

So $(\tilde{\varphi}, \tilde{\psi}) \in \tilde{\Phi}$ where $\tilde{\Phi} = \{(\tilde{\varphi}, \tilde{\psi}) \in L^1(\mu)\times L^1(\nu) : \tilde{\varphi}(x) + \tilde{\psi}(y) \geq x\cdot y\}$.

Similarly if $(\tilde{\varphi}, \tilde{\psi}) \in \tilde{\Phi}$, then $(\varphi, \psi) \in \Phi_c$.

(B) If we let $M = \frac{1}{2}\int_X |x|^2 d\mu(x) + \frac{1}{2}\int_Y |y|^2 d\nu(y)$, then $\mathbb{J}(\tilde{\varphi}, \tilde{\psi}) = M - \mathbb{J}(\varphi, \psi)$ and for $\pi \in \Pi(\mu, \nu)$, has $\mathbb{K}(\pi) = M - \int_{X\times Y} x \cdot y d\pi(x, y)$. Kantorovich duality (**Theorem 4.1**) implies that

$$\min_{(\tilde{\varphi},\tilde{\psi})\in\tilde{\Phi}} \mathbb{J}(\tilde{\varphi}, \tilde{\psi}) = \max_{\pi\in\Pi(\mu,\nu)} \int_{X\times Y} x \cdot y d\pi(x, y)$$

Also

$$\pi^\dagger \in \Pi(\mu, \nu) \text{ minimises } \mathbb{K} \text{ over } \Pi(\mu, \nu) \;\Leftrightarrow\; \pi^\dagger \text{ maximises } \int_{X\times Y} x \cdot y d\pi(x, y)$$
$$(\varphi, \psi) \in \Phi_c \text{ maximises } \mathbb{J} \text{ over } \Phi_c \;\Leftrightarrow\; (\tilde{\varphi}, \tilde{\psi}) \in \tilde{\Phi} \text{ minimise } \mathbb{J} \text{ over } \tilde{\Phi}.$$

(C) Recall that there exists maximiser $(\varphi^{cc}, \varphi^c) \in \Phi_c$ of $\mathbb{J}$. So $(\tilde{\varphi}, \tilde{\psi}) := (\frac{1}{2}|\cdot|^2 - \varphi^{cc}, \frac{1}{2}|\cdot|^2 - \varphi^c)$ minimise $\mathbb{J}$ over $\tilde{\Phi}$. Furthermore

$$\begin{aligned}\tilde{\psi}(y) =& \frac{1}{2}|y|^2 - \varphi^c(y) = \sup_{x\in X}\Big(\frac{1}{2}|y|^2 - \frac{1}{2}|x - y|^2 + \varphi(y)\Big)\\ =& \sup_{x\in X}\Big(x \cdot y - \frac{1}{2}|x|^2 + \varphi(x)\Big) = \sup_{x\in X}(x \cdot y - \tilde{\varphi}(x)) = \tilde{\varphi}^*(y)\end{aligned}$$

And

$$\begin{aligned}\tilde{\varphi}(x) =& \frac{1}{2}|x|^2 - \varphi^{cc}(x) = \sup_{y\in Y}\Big(\frac{1}{2}|x|^2 - \frac{1}{2}|x - y|^2 + \varphi^c(y)\Big)\\ =& \sup_{y\in Y}\Big(\frac{1}{2}|x|^2 - \frac{1}{2}|x - y|^2 + \frac{1}{2}|y|^2 - \tilde{\varphi}^*(y)\Big) \quad \text{(used previous computation)}\\ =& \sup_{y\in Y}(x \cdot y - \tilde{\varphi}^*(y))\\ =& \tilde{\varphi}^{**}(x)\end{aligned}$$

By **Proposition 6.6**, $\tilde{\eta} := \tilde{\varphi}^{**}$ is convex and lower semi-continuous and $\tilde{\eta}^* = \tilde{\varphi}^{***} = \tilde{\varphi}^*$.

(D) For $(\tilde{\varphi}, \tilde{\varphi}^*)$ with $\tilde{\varphi} \in L^1(\mu)$, we have

$$\begin{aligned}\int_{X\times Y} \tilde{\varphi}(x) + \tilde{\varphi}^*(y) d\pi^\dagger(x, y) =& \int_{X\times Y} x \cdot y d\pi^\dagger(x, y)\\ \le& \frac{1}{2}\int_{X\times Y} |x|^2 + |y|^2 d\pi^\dagger(x,) = \frac{1}{2}\int_X |x|^2 d\mu(x) + \frac{1}{2}\int_Y |y|^2 d\nu(y)\end{aligned}$$

so $\mathbb{J}(\tilde{\varphi}, \tilde{\varphi}^*) < \infty$.

(E) From the result of (C), if we just prove that $\tilde{\varphi}^* \in L^1(\nu)$ whenever $\tilde{\varphi} \in L^1(\mu)$, then $(\eta, \eta^*) \in L^1(\mu) \times L^1(\nu)$ so there is a minimiser of $\mathbb{J}$ and $\tilde{\Phi}$ that takes the form $(\tilde{\eta}, \tilde{\eta}^*)$ where $\tilde{\eta}$ is convex, lower semi-continuous and is proper.

To see this, assume $\tilde{\varphi} \in L^1(\mu)$. First note that $\exists x_0 \in X$ and $b_0 = \tilde{\varphi}(x_0) + 1 \in \mathbb{R}$ such that

$$\tilde{\varphi}^*(y) \ge x_0 \cdot y - \tilde{\varphi}(x_0) - 1 =: x_0 \cdot y - b_0 =: f(y)$$

Then we have $\tilde{\varphi}^* - f(y) \ge 0$, so $\left\|\tilde{\varphi}^* - f\right\|_{L^1(\mu)} = \int_Y \Big(\tilde{\varphi}^*(y) - f(y)\Big) d\nu(y)$. Hence

$$\begin{aligned}\left\|\tilde{\varphi}^* - f\right\|_{L^1(\mu)} =& \int_Y \Big(\tilde{\varphi}^*(y) - f(y)\Big) d\nu(y)\\ \le& \mathbb{J}(\tilde{\varphi}, \tilde{\varphi}^*) + \left\|\tilde{\varphi}\right\|_{L^1(\mu)} + \frac{1}{2}|x_0| + \frac{1}{2}\int_Y |y|^2 d\nu(y) + b_0\\ <& +\infty\end{aligned}$$

where the second line is implied by Cauchy-Schwarz ienquality. So $\tilde{\varphi}^* - f \in L^1(\nu)$ and since $f \in L^1(\nu)$, we conclude $\tilde{\varphi}^* \in L^1(\nu)$ as required.

We now come back to the one of the two main theorems of the chapter.

Theorem 6.1) *(Knott-Smith(KS) optimality criterion)* Let $\mu \in \mathcal{P}_2(X)$, $\nu \in \mathcal{P}_2(Y)$, $X, Y \in \mathbb{R}^d$, $c(x,y) = \frac{1}{2}|x-y|^2$. Then $\pi^\dagger \in \Pi(\mu,\nu)$ minimises $\mathbb{K}$ over $\Pi(\mu,\nu)$ *iff* there exists $\tilde{\varphi} \in L^1(\mu)$ convex, lower-semicontinuous such that $y \in \partial\tilde{\varphi}(x)$ for $\pi^\dagger$-a.e. (x,y).

Moreover $(\tilde{\varphi}, \tilde{\varphi}^*)$ maximises $\mathbb{J}$ over $\tilde{\Phi}$.

proof) Let $\pi^\dagger \in \Pi(\mu,\nu)$ minimise $\mathbb{K}$ over $\Pi(\mu,\nu)$ and $(\tilde{\varphi}, \tilde{\varphi}^*) \in \tilde{\Phi}$ minimise $\mathbb{J}$ over $\tilde{\Phi}$. By *Kantorovich duality*,

$$\begin{aligned}
&\int_X \tilde{\varphi}(x)d\mu(x) + \int_Y \tilde{\varphi}^*(y)d\nu(y) = \int_{X\times Y} x\cdot y d\pi^\dagger(x,y)\\
\Rightarrow\quad &\int_{X\times Y} \Big(\tilde{\varphi}(x) + \tilde{\varphi}^*(y) - x\cdot y\Big)d\pi^\dagger(x,y) = 0\\
\Rightarrow\quad &\tilde{\varphi}(x) + \tilde{\varphi}^*(y) = x\cdot y \quad \text{for } \pi^\dagger\text{-a.e. } (x,y) \quad (\text{by property of } \tilde{Phi})\\
\Rightarrow\quad &y \in \partial\tilde{\varphi}(x) \quad \text{for } \pi^\dagger\text{-a.e. } (x,y)
\end{aligned}$$

Conversely suppose $\pi^\dagger \in \Pi(\mu,\nu)$ and $\tilde{\varphi} \in L^1(\mu)$ and satisfy $y \in \partial\tilde{\varphi}(x)$ for $\pi^\dagger$-a.e. (x,y) where $\tilde{\varphi}$ is lower semi-continuos and convex. We want to show that $\pi^\dagger$ is opimal for KOT problem and $(\tilde{\varphi}, \tilde{\varphi}^*)$ optimal for "KD" problem.

By **Proposition 6.4**,

$$\int_{X\times Y} \Big(\tilde{\varphi}(x) + \tilde{\varphi}^*(y) - x\cdot y\Big)d\pi^\dagger(x,y) = 0$$

By point (D) above, we have $\tilde{\varphi}^* \in L^1(\nu)$, so $(\tilde{\varphi}, \tilde{\varphi}^*) \in \tilde{\Phi}$. Then

$$\min_{\tilde{\Phi}} \mathbb{J} \leq \mathbb{J}(\tilde{\varphi}, \tilde{\varphi}^*) = \int_{X\times Y} x\cdot y d\pi^\dagger(x,y) \leq \max_{\pi\in\Pi(\mu,\nu)} \int_{X\times Y} x\cdot y d\pi(x,y)$$

By duality, LHS=RHS, so all inequalities above are in fact eqaulities. So $(\tilde{\varphi}, \tilde{\varphi}^*)$ minimise $\mathbb{J}$ over $\tilde{\Phi}$ and $\pi^\dagger$ maximises $\int_{X\times Y} x\cdot y d\pi(x,y)$ over $\Pi(\mu,\nu)$.

(End of proof) $\square$

6.4 Proof of Brenier's Theorem

In this section, we prove the second one of the two main theorems of the chapter.

Theorem 6.2) *(Brenier's Theorem)* Let $\mu \in \mathcal{P}_2(X)$, $\nu \in \mathcal{P}_2(Y)$, $X, Y \subset \mathbb{R}^d$ and $c(x,y) = \frac{1}{2}|x-y|^2$. Assume that μ does not give mass to small sets (a small set is any set with Hausdorff dimension at most $d-1$). Then there is a unique $\pi^\dagger \in \Pi(\mu,\nu)$ that minimises the KOT problem.

Moreover, $\pi^\dagger$ satisfies $\pi^\dagger = (id \times \nabla\tilde{\varphi})_\# \mu$ where $\nabla\tilde{\varphi}$ is the unique gradient of a convex function that pushes μ forward to ν (that is, $(\nabla\tilde{\varphi})_\#\mu = \nu$) and $(\tilde{\varphi}, \tilde{\varphi}^c)$ minimise $\mathbb{J}$ over $\tilde{\Phi}$.

(1st March, Friday)

(We have a lecture on the 15th March)
(Exercies Class 3 is at Tuesday, 5th March, 4pm, MR13)

proof of Brenier's theorem) Let $\pi^\dagger \in \Pi(\mu,\nu)$ minimise KOT problem and $\{\pi^\dagger(\cdot|x)\}_{x\in X}$ be the disinteration of measure,

$$\pi^\dagger(A\times B) = \int_A \pi^\dagger(B|x)d\mu(x)$$

By **Theorem 6.1**, for μ-a.e. $x \in X$ and $\pi^\dagger(\cdot|x)$-a.e. $y \in Y$ we have $y \in \partial\tilde{\varphi}(x)$ where $\tilde{\varphi}$ minimises $\mathbb{J}$ over $\tilde{\Phi}$. Since $\partial\tilde{\varphi}(x) = \{\nabla\tilde{\varphi}(x)\}$ for μ-a.e. $x \in X$, has $y = \nabla\tilde{\varphi}(x)$ for μ-a.e. $x \in X$ and $\pi^\dagger(\cdot|x)$-a.e. $y \in Y$. Hence $\pi^\dagger(\cdot|x) = \delta_{\nabla\tilde{\varphi}(x)}$. So

$$\begin{aligned}\pi^\dagger(A\times B) = \int_A \mathbb{1}_{\nabla\tilde{\varphi}(x)\in B}\, d\mu(x) =& \mu\Big(\{x : x\in A \text{ and } \nabla\tilde{\varphi}(x)\in B\}\Big)\\ =&(\mathrm{Id}\times\nabla\tilde{\varphi})_{\#}\mu(A\times B)\end{aligned}$$

so $\pi^\dagger = (Id\times\nabla\tilde{\varphi})_{\#}\mu$. Also

$$\begin{aligned}\nu(B) = \pi^\dagger(X\times B) &= (\mathrm{Id}\times\nabla\tilde{\varphi})_{\#}\mu(X\times B)\\ &= \mu\Big(\{x : (x,\nabla\tilde{\varphi}(x))\in X\times B\}\Big)\\ &= \mu\Big(\{x : \nabla\tilde{\varphi}(x)\in B\}\Big) = (\nabla\tilde{\varphi})_{\#}\mu(B)\end{aligned}$$

For uniqueness, suppose $\overline{\varphi}$ is convex and satisfies $(\nabla\overline{\varphi})_{\#}\mu = \nu$. We show $\nabla\tilde{\varphi} = \nabla\overline{\varphi}$ a.e. By **Thoerem 6.1**,

$$\overline{\pi} = (Id\times\nabla\overline{\varphi})_{\#}\mu$$

is an optimal transport plan and $(\overline{\varphi},\overline{\varphi}^c)$ minimise $\mathbb{J}$ over $\tilde{\Phi}$. So

$$\begin{aligned}&\int_X \overline{\varphi}d\mu + \int_Y \overline{\varphi}^c d\nu = \int_X \tilde{\varphi}d\mu + \int_Y \tilde{\varphi}^c d\nu\\ \Rightarrow\quad &\int_{X\times Y}(\overline{\varphi}(x)+\overline{\varphi}^c(y))d\pi^\dagger(x,y) = \int_{X\times Y}(\tilde{\varphi}(x)+\tilde{\varphi}^c(y))d\pi^\dagger(x,y)\\ &= \int_{X\times Y} x\cdot y\, d\pi^\dagger(x,y) \quad \text{by } \textbf{Proposition 6.4}\\ &= \int_{X\times Y} x\cdot y\, d(\mathrm{Id}\times\nabla\tilde{\varphi})_{\#}\mu(x,y)\\ &= \int_X x\cdot\nabla\tilde{\varphi}(x)d\mu(x)\end{aligned}$$

Also,

$$\int_{X\times Y}\Big(\overline{\varphi}(x)+\overline{\varphi}^*(y)\Big)d\pi^\dagger(x,y) = \int_X\Big(\overline{\varphi}(x)+\overline{\varphi}^*(\nabla\tilde{\varphi}(x))\Big)d\mu(x)$$

so

$$\int_X\Big(\overline{\varphi}(x)+\overline{\varphi}^*(\nabla\tilde{\varphi}(x)) - x\cdot\nabla\tilde{\varphi}(x)\Big)d\mu(x) = 0$$

but $\overline{\varphi}(x)+\overline{\varphi}^*(y)\geq x\cdot y$, so

$$\overline{\varphi}(x)+\overline{\varphi}^*(\nabla\varphi\tilde{(x)}) - x\cdot\nabla\tilde{\varphi}(x) = 0 \quad \mu\text{-a.e. } x\in X$$

By **Proposition 6.4**, $\nabla\tilde{\varphi}(x)\in\partial\overline{\varphi}(x) = \{\nabla\overline{\varphi}(x)\}$, so by **Propositioin 6.5**, $\nabla\tilde{\varphi}(x) = \nabla\overline{\varphi}(x)$ for μ-a.e. $x\in X$.

(End of proof) □

7 Wasserstein Distances

In this chapter, we assum $c(x,y) = |x-y|^p$ with $p \in [1,+\infty)$ with $X = Y \subset \mathbb{R}^d$.

The objectives in this chapter are

(1) Define the *Wasserstein distnace* d_{W^p} and show it is a metric in $\mathcal{P}_p(X)$.

(2) Show equivalence of d_{W^p} and d_{W^q} when X is bounded.

(3) Relaionship with weak-* topology.

(4) Show that $(\mathcal{P}_p(X), d_{W^p})$ is a *geodesic space* (to be defined later).

Another interesting cost function that does not fit into the Wasserstein framework is $c(x,y) = 1_{x\neq y}$.

Proposition 7.1) Let $\mu,\nu \in \mathcal{P}(X)$, $X \subset \mathbb{R}^d$, $c(x,y) = \mathbb{1}_{x\neq y}$. Then

$$\inf_{\pi\in\Pi(\mu,\nu)} \mathbb{K}(\pi) = \frac{1}{2}\|\mu-\nu\|_{TV}$$

where $\|\mu\|_{TV} = 2\sup_A |\mu(A)|$.

proof) By the *KR theorem* (**Theorem 4.13**),

$$\begin{aligned}\min_{\pi\in\Pi(\mu,\nu)} \mathbb{K}(\pi) &= \sup\Big\{\int_X f d(\mu-\nu) : f \in L^1(|\mu-\nu|), \big\|f\big\|_{Lip} \le 1\Big\} \\ &= \sup\Big\{\int_X f d(\mu-\nu) : 0 \le f(x) \le 1 \ \forall x \in X\Big\}\end{aligned}$$

where $\|f\|_{Lip}$ is given in terms of c rather than a usual metric. Write $\mu-\nu = (\mu-\nu)_+ - (\mu-\nu)_-$ where $(\mu-\nu)_\pm \in \mathcal{M}_+(X)$ and are singular (*i.e.* has disjoint support). We may achieve the supremum when we choose $f(x) = 1$ if $x \in \operatorname{supp}(\mu-\nu)_+$ and $f(x) = 0$ otherwise. Then

$$\min_{\pi\in\Pi(\mu,\nu)} \mathbb{K}(\pi) = (\mu-\nu)_+(X)$$

For TV, the optimal choice is $A = \operatorname{supp}((\mu-\nu)_+)$ and $\|\mu-\nu\|_{TV} = 2(\mu-\nu)_+(X)$. So $\min \mathbb{K} = \frac{1}{2}\|\mu-\nu\|_{TV}$.

(End of proof) $\square$

7.1 Wasserstein Distances

We work on the space of measures with bounded p^{th} moment,

$$\mathcal{P}_p(X) = \{\mu \in \mathcal{P}(X) : \int_X |x|^p d\mu(x) < +\infty\}$$

If X is bounded, then $\mathcal{P}_p(X) = \mathcal{P}(X)$.

Definition 7.2) Let $\mu,\nu \in \mathcal{P}_p(X)$, then the p **-Wasserstien distance** is defined to be

$$d_{W^p}(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} \Big(\int_{X\times Y} |x-y|^p d\pi(x,y)\Big)^{1/p}$$

If $\mu \in \mathcal{P}_p(X)$, $\nu \in \mathcal{P}_p(X)$, then

$$d_{W^p}(\mu,\nu)^p = \min_{\pi\in\Pi(\mu,\nu)} p\int_{X\times Y} |x|^p + |y|^p d\pi(x,y) = p\int |x|^p d\mu(x) + p\int |y|^p d\nu(y) < +\infty$$

(4th March, Monday)

Recall, we defined
Definition 7.2) For $\mu,\nu \in \mathcal{P}_p(X)$, the **Wasserstein distance** is defined by

$$d_{W^p}(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} \left(\int_{X\times Y} |x-y|^p d\pi(x,y)\right)^{1/p}$$

Proposition 7.3) Let $X \subset \mathbb{R}^d$. Then the distnace $d_{W^p} : \mathcal{P}_p(X)\times\mathcal{P}_p(X) \to [0,\infty)$ is a metric.

proof)

(1) It is easy to see that $d_{W^p}(\mu,\nu) \geq 0$ for any μ, ν.

(2) If $\mu=\nu$, then $\pi(x,y) = \delta_y(x)\mu(x)$ satisfies $\pi\in\Pi(\mu,\nu)$ and

$$d_{W^p}(\mu,\nu) \leq \int_{X\times Y} |x-y|^p d\pi(x,y) = 0$$

If $d_{W^p}(\mu,\nu)=0$, then $\exists\pi\in\Pi(\mu,\nu)$ such that $\int_{X\times Y}|x-y|^p d\pi(x,y) = 0$ so $x=y$ for π-a.e. x,y. So for any function $f : X\to\mathbb{R}$,

$$\int_X f(x)d\mu(x) = \int_{X\times X} f(x)d\pi(x,y) = \int_{X\times X} f(y)d\pi(x,y) = \int_X f(y)d\nu(y) \quad \cdots\cdots\cdots \ (*)$$

Since $(*)$ holds for all $f : X\to\mathbb{R}$ integrable w.r.t μ and ν (or just for f continuous and bounded is sufficient), we have $\mu=\nu$.

(3) Clearly $d_{W^p}(\mu,\nu) = d_{W^p}(\nu,\mu)$. (To write out formally, define $s : (x,y)\mapsto(y,x)$ and observe that whenever $\pi\in\Pi(\mu,\nu)$, has $s_\#\Pi(\nu,\mu)$)

(4) To see the triangular inequality, we make use of *glueing lemma* (**Lemma 7.4**). Let $\mu,\nu,\omega\in\mathcal{P}_p(X)$ and $\pi_{XY}\in\pi(\mu,\nu)$ and $\pi_{YZ}\in\pi(\nu,\omega)$ be the optimal plans, *i.e.*

$$d_{W^p}(\mu,\nu) = \left(\int_{X\times X} |x-y|^p d\pi_{XY}(x,y)\right)^{1/p}$$
$$d_{W^p}(\nu,\omega) = \left(\int_{X\times X} |y-z|^p d\pi_{YZ}(y,z)\right)^{1/p}$$

Take $\gamma\in\mathcal{P}(X\times Y\times Z)$ such that $P^{X,Y}_\#\gamma = \pi_{X\times Y}$ and $P^Y_\#\gamma = \pi_{YZ}$ using the *glueing lemma*. Define $\pi_{XZ} = P^{X,Z}_\#\gamma$. Since $\pi_{XZ}(A\times X) = \gamma(A\times X\times X) = \pi_{XY}(A\times X) = \mu(A)$ and similarly $\pi_{XZ}(X\times C) = \omega(C)$, we see that $\pi_{XZ}\in\Pi(\mu,\omega)$. Now, by *Minkowski's inequality*,

$$\begin{aligned}
d_{W^p}(\mu,\omega) &\leq \left(\int_{X\times X}|x-z|^p d\pi_{XZ}(x,z)\right)^{1/p} = \left(\int_{X\times X\times X}|x-z|^p d\gamma(x,y,z)\right)^{1/p} \\
&\leq \left(\int_{X\times X\times X}|x-y|^p d\gamma(x,y,z)\right)^{1/p} + \left(\int_{X\times X\times X}|y-z|^p d\gamma(x,y,z)\right)^{1/p} \\
&= \left(\int_{X\times X}|x-y|^p d\pi_{XY}(x,y)\right)^{1/p} + \left(\int_{X\times X}|y-z|^p d\pi_{YZ}(y,z)\right)^{1/p} \\
&= d_{W^p}(\mu,\nu) + d_{W^p}(\nu,\omega)
\end{aligned}$$

(End of proof) $\square$

For the triangular inequality, we needed the following *glueing lemma*.

Lemma 7.4) Let $X, Y, Z \subset \mathbb{R}^d$, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, $\omega \in \mathcal{P}(Z)$, $\pi_1 \in \Pi(\mu, \nu)$, $\pi_2 \in \Pi(\nu, \omega)$. Then there is a measure $\gamma \in \mathcal{P}(X \times Y \times Z)$ such that $P_{\#}^{X \times Y}\gamma = \pi_1$ and $P_{\#}^{Y,Z} = \pi_{12}$ where $P^{X,Y}(x, y, z) = (x, y)$, $P^{Y,Z}(x, y, z) = (y, z)$.

proof) By disintegration of measures, we can write

$$\pi_1(A \times B) = \int_B \pi_1(A|y) d\nu(y), \quad \pi_2(B \times C) = \int_B \pi_2(C|y) d\nu(y)$$

for families $\{\pi_1(\cdot|y)\}_{y \in Y} \subset \mathcal{P}(X)$ and $\{\pi_2(\cdot|y)\}_{y \in Y} \subset \mathcal{P}(Z)$. Define $\gamma \in \mathcal{M}(X \times Y \times Z)$ by

$$\gamma(A \times B \times C) = \int_B \pi_1(A|y)\pi_2(C|y) d\nu(y)$$

We can check

$$\begin{aligned} \gamma(A \times B \times Z) &= \int_B \pi_1(A|y)\pi_2(Z|y) d\nu(y) = \pi_1(A \times B) \\ \gamma(X \times B \times C) &= \pi_2(B \times C) \end{aligned}$$

So $P_{\#}^{X,Y}(\gamma) = \pi_1$ and $P_{\#}^{Y,Z}\gamma = \pi_2$. It also follows that $\gamma \in \mathcal{P}(X \times Y \times Z)$.

(End of proof) $\square$

Proposition 7.5) Let $X \subset \mathbb{R}^d$. For every $p, q \in [1, +\infty)$, $q \leq p$ and any $\mu, \nu \in \mathcal{P}_p(X)$, we have $d_{W^p}(\mu, \nu) \geq d_{W^q}(\mu, \nu)$.

Furthermore, if X is bounded then $d_{W^p}^p(\mu, \nu) \leq \operatorname{diam}(X)^{p-1} d_{W^1}(\mu, \nu)$ (where $\operatorname{diam}(X) = \sup_{w,z \in X} |w - z|$)

proof) The first part uses Jensen's inequality. The second part makes use of Hölder inequality.

(End of proof) $\square$

7.2 The Wasserstein Topology

Aim : show $\mu_n \xrightarrow{w-*} \mu$ *iff* $d_{W^p}(\mu_n, \mu) \to 0$. We start with the case when $X \subset \mathbb{R}^d$ is compact.

Theorem 7.6) Let $X \subset \mathbb{R}^d$ be *compact*, $\{\mu_n\}_{n \in \mathbb{N}} \subset \mathcal{P}(X)$ and $\mu \in \mathcal{P}(X)$ and $p \in [1, \infty)$. Then

$$\mu_m \xrightarrow{w-*} \mu \quad \textit{iff} \quad d_{W^p}(\mu_m, \mu) \to 0$$

proof) By **Proposition 7.5**, it is enough to prove the result for $p = 1$.

Assume $d_{W^1}(\mu_m, \mu) \to 0$. by the *Kantorovich-Rubinstein Theorem* (**Theorem 4.13**),

$$d_{W^1}(\mu, \nu) = \sup\left\{ \int_X \varphi d(\mu - \nu) : \varphi \in L^1(|\mu - \nu|), |\varphi(x) - \varphi(y)| \leq |x - y| \right\}$$

Let φ be Lipschitz with $\|\varphi\|_{Lip} > 0$. Then define $\tilde{\varphi} = \frac{1}{\|\varphi\|_{Lip}}\varphi$ so that $\tilde{\varphi}$ is 1-Lipschitz. So

$$\frac{1}{\|\varphi\|_{Lip}} \int \varphi d(\mu_m - \mu) = \int \tilde{\varphi}(\mu_m - \mu) \leq d_{W^1}(\mu_m, \mu) \to 0$$

Then $\limsup_{m\to\infty} \int \varphi d\mu_m \leq \int \varphi d\mu$. Substituiting $\varphi \mapsto -\varphi$ gives the inverse inequality, so together we have

$$\lim_{m\to\infty} \int_X \varphi d\mu_n = \int \varphi d\mu$$

By *Portmanteau Theorem* (**Theorem 1.2**), has $\mu_m \xrightarrow{w-*} \mu$.

Conversely, assume $\mu_n \xrightarrow{w-*} \mu$ and let $(m_k)_k \subset \mathbb{N}$ be the subsequence such that

$$\lim_{k\to\infty} d_{W^1}(\mu_{m_k}, \mu) = \limsup_{m\to\infty} d_{W^1}(\mu_m, \mu)$$

Let $\tilde{\varphi}_{m_k}$ be 1-Lipschitz and such that $d_{W^1}(\mu_{m_k}, \mu) \leq \int_X \tilde{\varphi}_{m_k} d(\mu_{m_k} - \mu) + \frac{1}{k}$. Pick $x_0 \in \text{supp}(\mu)$ and define $\varphi_{m_k}(x) = \tilde{\varphi}_{m_k}(x) - \tilde{\varphi}_{m_k}(x_0)$. So,

$$\begin{aligned} d_{W^1}(\mu_{m_k}, \mu) &\leq \int_X \varphi_{m_k} d(\mu_{m_k} - \mu) - \tilde{\varphi}_{m_k}(x_0) \int_X d(\mu_{m_k} - \mu) + \frac{1}{k} \\ &= \int_X \varphi_{m_k} d(\mu_{m_k} - \mu) + \frac{1}{k} \end{aligned}$$

Since φ_{m_k} are 1-Lipschitz and $\varphi_{m_k}(x_0) = 0$, we have $\{\varphi_{m_k}\}_{k\in\mathbb{N}}$ uniformly bounded and equi-continuous. By the *Arzelà-Ascoli Theorem*, there is a subsequence (relabelled) such that $\varphi_{m_k} \to \varphi$ uniformly and φ is 1-Lipschitz. Hence,

$$\begin{aligned} \limsup_{m\to\infty} d_{W^1}(\mu_m, \mu) &\leq \limsup_{k\to\infty} \left(\int_X \varphi_{m_k} d(\mu_{m_k} - \mu) + \frac{1}{k} \right) \\ &= \limsup_{k\to\infty} (\int_X (\varphi_{m_k} - \varphi) d(\mu_{m_k} - \mu) + \int_X \varphi d(\mu_{m_k} - \mu)) \\ &\leq \limsup_{k\to\infty} \left\| \varphi_{m_k} - \varphi \right\|_\infty + \limsup_{k\to\infty} \int_X \varphi d(\mu_{m_k} - \mu) = 0 \end{aligned}$$

(End of proof) $\square$

(6th March, Wednesday)

Theorem 7.7) $\mu_n, \mu \in \mathcal{P}_p(\mathbb{R}^d)$. Then

$$d_{W^p}(\mu_n, \mu) \to 0 \quad \textit{iff} \quad \mu_n \xrightarrow{w^*} \mu \text{ and } \int_{\mathbb{R}^d} |x|^p d\mu_n \to \int_{\mathbb{R}^d} |x|^p d\mu$$

proof) Let $d_{W^p}(\mu_n, \mu) \to 0$. By **Proposition 7.5**, $d_{W^1}(\mu_n, \mu) \to 0$. Analogous to the proof of **Theorem 7.6**, we have $\int \varphi d(\mu_n - \mu) \to 0$ for all Lipschitz functions φ. So $\mu_n \xrightarrow{w^*} \mu$.

To show $\int |x|^p d\mu_n(x) \to \int |x|^p d\mu(x)$, we write

$$\int_{\mathbb{R}^d} |x|^p d\mu_n(x) = d^p_{W^p}(\mu_n, \delta_0), \quad \int_{\mathbb{R}^d} |x|^0 d\mu(x) = d^p_{W^p}(\mu, \delta_0)$$

Also by triangular inequality,

$$d_{W^p}(\mu, \delta_0) - d_{W^p}(\mu, \mu_n) \leq d_{W^p}(\mu_n, \delta_0) = \left(\int |x|^p d\mu_n(x) \right)^{1/p} \leq d_{W^p}(\mu_n, \mu) + d_{W^p}(\mu, \delta_0) \to d_{W^p}(\mu, \delta_0)$$

but both very LHS and the RHS converges to $d_{W^p}(\mu, \delta_0)$, so $d_{W^p}(\mu_n, \delta_0) \to d_{W^p}(\mu, \delta_0)$ as desired.

To show the converse, let $\mu_n \xrightarrow{w^*} \mu$ and $\int |x|^p d\mu \to \int |x|^p d\mu$. For any $R > 0$, define $\phi_R(x) = (\min\{|x|, R\})^p$ which is continuous and bounded. So

$$\int_{\mathbb{R}^d} |x|^p - \phi_R(x) d\mu_n \to \int_{\mathbb{R}^d} |x|^p - \phi_R(x) d\mu \quad \cdots\cdots\cdots \ (*)$$

Now

$$\int_{\mathbb{R}^d} (|x|^p - \phi_R(x)) d\mu(x) = \int_{|x|>R} (|x|^p - R^p) d\mu \leq \int_{|x|>R} |x|^p d\mu$$

For any $\epsilon > 0$, we can find $R > 0$ such that $\int_{\mathbb{R}^d} (|x|^p - \phi_R(x)) d\mu(x) < \epsilon/2$. By $(*)$, for m sufficiently large, $\int_{\mathbb{R}^d} (|x|^p - \phi_R(x)) d\mu_m(x) < \epsilon$. Since $(a+b)^p \geq (a^p + b^p)$ for any $a, b \geq 0$, for $|x| > R$, we have $(|x| - R)^p \leq |x|^p - R^p = |x|^p - \phi_R(x)$. So

$$\int_{|x|>R} (|x| - R)^p d\mu_m < \epsilon, \quad \int_{|x|>R} (|x| - R)^p d\mu < \epsilon$$

Let $P_R : \mathbb{R}^d \to \overline{B(0,R)}$ be the projection onto $\overline{B(0,R)}$, *i.e.* $P_R = x$ if $x \in \overline{B(0,R)}$ and $P_R(x) = xR/|x|$ is otherwise.Then P_R is continuous, and $P_R = id$ on $\overline{B(0,R)}$, and for $x \notin \overline{B(0,R)}$, we have $|x - P_R(x)| = |x| - R$. Hence

$$\begin{aligned} d_{W^p}(\mu, (P_R)_\# \mu) &\leq \Big(\int_{\mathbb{R}^d} |x - P_R(x)|^p d\mu(x) \Big)^{1/p} \quad \text{since } d_{W^p}(\cdot,\cdot) \text{ is optimal for MOT} \\ &= \Big(\int_{|x|>R} |x - P_R(x)|^p d\mu(x) \Big)^{1/p} \\ &= \Big(\int_{|x|>R} (|x| - R)^p d\mu(x) \Big)^{1/p} \leq \epsilon^{1/p} \end{aligned}$$

and by same principles, $d_{W^p}(\mu, (P_R)_\# \mu) \leq \epsilon^{1/p}$.

Meanwhile, for any $\varphi \in C_b^0(\mathbb{R}^d)$, we have

$$\begin{aligned} \int \varphi d((P_R)_\# \mu_m) &= \int \varphi(P_R(x)) d\mu_m(x) \\ &\to \int \varphi(P_R(x)) d\mu(x) \quad \text{by weak convergence of } (\mu_m) \\ &= \int \varphi d((P_R)_\# \mu_m) \end{aligned}$$

So $(P_R)_\# \mu_m \xrightarrow{w^*} (P_R)_\# \mu$. Then by **Theorem 7.6**, and $\overline{B(0,R)}$ is compact, has $d_{W^p}((P_R)_\# \mu_m, (P_R)_\# \mu) \to 0$.

Putting these toghther,

$$\begin{aligned} \limsup_{m\to\infty} d_{W^p}(\mu_m, \mu) &\leq \limsup_{m\to\infty} \Big(d_{W^p}(\mu_m, (P_R)_\# \mu_m) + d_{W^p}((P_R)_\# \mu_m, (P_R)_\# \mu) \Big) + d_{W^p}((P_R)_\# \mu, \mu) \\ &\leq 2\epsilon^{1/p} \end{aligned}$$

Let $\epsilon \to 0$ then $d_{W^p}(\mu_m, \mu) \to 0$ as requried.

(End of proof) $\square$

7.3 Geodesics in the Wasserstien Space

Definition 7.8) Let $p \in [1, +\infty]$ and (Z, d) be a metric space and $\omega : (a, b)(\subset \mathbb{R}) \to Z$ a curve in Z. We say $\omega \in \mathrm{AC}^p((a, b), Z)$ if $\exists g \in L^p((a, b))$ such that

$$d(\omega(t_0), \omega(t_1)) \leq \int_{t_0}^{t_1} g(s)ds, \quad \forall a < t_0 < t_1 < b$$

If $p = 1$, we say ω is a **absolutely continuous** curve.

If we only have $g \in L^p_{loc}((a, b))$, then we say $w \in \mathrm{AC}^p_{loc}((a, b), Z)$ and curves $\omega \in \mathrm{AC}^1_{loc}((a, b), Z)$ are called **locally absolutely continuous**.

Definition 7.9)

(1) Let (Z, d) be a metric space and $\omega : [0, 1] \to Z$ a curve. We define the **length** of ω by

$$\mathrm{Len}(\omega) := \sup\Big\{ \sum_{k=0}^{n-1} d(\omega(t_k), \omega(t_{k+1})) : n \in \mathbb{N}, 0 = t_0 < t_1 < \cdots < t_n = 1 \Big\}$$

(2) A curve $\omega : [0, 1] \to Z$ is said to be a **geodesic** between $z_0 \in Z$ and $z_1 \in Z$ if

$$\omega \in \mathrm{argmin}\Big\{ \mathrm{Len}(\overline{\omega}) \,\Big|\, \overline{\omega} : [0, 1] \to Z, \overline{\omega} = z_0, \overline{\omega}(1) = z_1 \Big\}.$$

(3) A curve $\omega : [0, 1] \to z$ is said to be a **constant speed geodesic** between $z_0 \in z$ and $z_1 \in z$ if

$$d(\omega(t), \omega(s)) = |t - s| d(\omega(0), \omega(1))$$

Notes :

(1) If ω is a constant speed geodesic, then it is a geodesic.

(2) If $d(\omega(t), \omega(s)) = |t - s| d(z_0, z_1)$ for all $s, t \in (0, 1)$ then $\omega \in \mathrm{AC}^1((0, 1), Z)$ with $g(s) = d(z_0, z_1)$.

Definition 7.10) Let (Z, d) be a metric space.

(1) We say (Z, d) is a **length space** if

$$\forall x, y \in z, \quad d(x, y) = \inf\{\mathrm{Len}(\omega) : \omega \in \mathrm{AC}^1((0, 1), Z), \omega(0) = x, \omega(1) = y\}$$

(2) We say (Z, d) is a **geodesic space** if

$$\forall x, y \in z, \quad d(x, y) = \min\{\mathrm{Len}(\omega) : \omega \in \mathrm{AC}^1((0, 1), Z), \omega(0) = x, \omega(1) = y\}$$

In particular, the minimum is achieved.

Theorem 7.11) Let $p \in [1, +\infty)$, $X \subset \mathbb{R}^d$ convex. Define $P_t : X \times X \to X$ by $P_t(x, y) = (1 - t)x + ty$. Let $\mu, \nu \in \mathcal{P}_p(X)$ and assume that $\pi \in \Pi(\mu, \nu)$ minimize $\mathbb{K}$ with cost $c(x, y) = |x - y|^p$. Then, the curve $\mu_t = (P_t)_\# \pi$ is a constant speed geodesic in $(\mathcal{P}_p(X), d_{W^p})$ connecting μ and ν.

Furthermore, if $\pi = (id \times T)_\# \mu$ where $T : X \to X$ (so in particular T is an optimal transport map), then $\mu_t = ((1 - t)id + tT)_\# \mu$.

proof) Note $P_0 = P^X$ and $P_1 = P^Y$, so $\mu_0 = (P_0)_\#\pi = (P^X)_\#\pi = \mu$ and similarly $\mu_1 = \nu$, so μ_+ connects μ_0 and μ_1. It is enough to show

$$d_{W^p}(\mu_s, \mu_t) = |t-s| d_{W^p}(\mu, \nu) \quad \forall s, t \in [0,1]$$

Suppose $d_{W^p}(\mu_s, \mu_t) \leq |t-s| d_{W^p}(\mu, \nu)$ for all $s, t \in [0,1]$. If we can find s, t such that $0 \leq s < t \leq 1$ such that $d_{W^p}(\mu_s, \mu_t) < (t-s) d_{W^p}(\mu, \nu)$ then

$$\begin{aligned} d_{W^p}(\mu, \nu) &\leq d_{W^p}(\mu, \nu_s) + d_{W^p}(\mu_s, \mu_t) + d_{W^p}(\mu_t, \nu) \\ &< d_{W^p}(\mu, \nu), \end{aligned}$$

a contradiction. So there are no such s, t.

So we are just left to show that $d_{W^p}(\mu_s, \mu_t) \leq |t-s| d_{W^p}(\mu, \nu)$ for all $s, t \in [0,1]$. To see this, let $\pi_{s,t} = (P_s \times P_t)_\#\pi$. Then for $A \subset X$,

$$\begin{aligned} \pi_{s,t}(A \times X) &= \pi\Big(\Big\{(x,y) : (P_s \times P_t)(x,y) \in A \times X\Big\}\Big) \\ &= \pi\Big(\Big\{(x,y) : P_s(x,y) \in A\Big\}\Big) \\ &= (P_s)_\#\pi(A) \end{aligned}$$

Similarly, has $\pi_{s,t}(X \times B) = \mu_t(B)$. Hence $\pi_{s,t} \in \Pi(\mu_s, \mu_t)$. Now

$$\begin{aligned} d_{W^p}(\mu_s, \mu_t) &\leq \Big(\int_{X \times X} |x-y|^p d\pi_{s,t}(x,y)\Big)^{1/p} \\ &= \Big(\int_{X \times X} |P_s(x,y) - P_t(x,y)|^p d\pi(x,y)\Big)^{1/p} \\ &= |t-s| \Big(\int_{X \times X} |x-y|^p d\pi(x,y)\Big)^{1/p} \\ &= |t-s| d_{W^p}(\mu, \nu) \end{aligned}$$

The furthermore part of theorem follows from

$$\begin{aligned} \mu_t &= (P_t)_\#\pi = (P_t)_\#(id \times T)_\#\mu \\ &= (P_t \circ (id \times T))_\mu \mu \end{aligned}$$

but $P_t \circ (id \times T)(x) = (1-t)x + tT(x)$ so

$$\mu_t = ((1-t)id + tT)_\#\mu$$

(End of proof) $\square$

(8th March, Friday)

8 Gradient Flows in Wasserstein Spaces

We first study the gradient Flow of function defined on an Euclidean space. The theory developed in this lecture can be generalized to the setting in a measure space with a metric.

In Euclidean space, the gradient flow is given by

$$\frac{d}{dt} u(t) = -\nabla \phi(u(t))$$

for $u : t \mapsto \mathbb{R}^d$.

8.1 Gradient Flows for Convex Functions in $\mathbb{R}^d$

Recall, we had

- $\phi : \mathbb{R}^d \to \mathbb{R}$ is **convex** if for all $u_0, u_1 \in \mathbb{R}^d$, $\theta \in [0,1]$

$$\phi(\theta u_0 + (1-\theta)u_1) \leq \theta\phi(u_0) + (1-\theta)\phi(u_1)$$

- If ϕ is differentiable, then it is convex *iff* it has monotonicity of $\nabla\phi$, i.e.

$$\langle \nabla\phi(u_0) - \nabla\phi(u_1), u_0 - u_1 \rangle \geq 0$$

- If $\nabla\phi$ is **L-Lipschitz** then

$$\phi(u_0) \leq \phi(u_1) + \langle \nabla\phi(u_1), u_0 - u_1 \rangle + \frac{L}{2}\left\|u_0 - u_1\right\|^2$$

Definition) ϕ is **λ-convex** if it satisfies, for $u_\theta = (1-\theta)u_0 + \theta u_1$, $\theta \in [0,1]$,

$$\phi(u_\theta) \leq (1-\theta)\phi(u_0) + \theta\phi(u_1) - \frac{\lambda}{2}\theta(1-\theta)\left\|u_0 - u_1\right\|^2$$

If $\phi \in C^2(\mathbb{R}^d)$, this is equivalent to having

(1) $\nabla^2\phi(u) \geq \lambda \mathrm{Id}$. **(Hessian inequality / curvature condition)**

(2) $\langle \nabla\phi(u_0) - \nabla\phi(u_1), u_0 - u_1 \rangle \geq \lambda\|u_0 - u_1\|^2$. **($\lambda$-monotonicity of $\nabla\phi$)**

(3) $\phi(u_1) \geq \phi(u_0) + \langle \nabla\phi(u_0), u_1 - u_0 \rangle + \frac{\lambda}{2}\|u_1 - u_0\|^2$. **(subgradient inequality)**

Remark :

- ϕ is λ-convex *iff* $f(u) := \phi(u) - \frac{\lambda}{2}\|u\|^2$ is convex. Hence whenever ϕ is λ-convex, there are constants a, b such that

$$\phi(x) \geq a + b \cdot x + \frac{\lambda}{2}\|x\|^2$$

- It is direct from the definition of being λ-convexity that whenever ϕ is λ-convex and has a minimizer, then the minimizer is unique.

Consider in the *discrete case*, the problem

$$\min_{u \in \mathbb{R}^d} \phi(u)$$

(I could not copy down) then u should solve $u_{k+1} = u_k - \tau\nabla\phi(u_k)$.

Now let us transfer this to the continuous case. So

$$u_{k+1} = u_k - \tau\nabla\phi(u_k) \quad \Leftrightarrow \quad \frac{u_{k+1} - u_k}{\tau} = -\nabla\phi(u_k)$$

so if we put $u_{k+1} = u_{(k+1)\tau}$, then the analogous equation in the continuous setting will be

$$\frac{d}{dt}u(t) = -\nabla\phi(u(t))$$

Definition) A **Gradient Flow** of $\phi : \mathbb{R}^d \to \mathbb{R}$ starting from an initial pint $u_0 \in \mathbb{R}^d$ is a curve $u : (0, +\infty) \to \mathbb{R}^d$, $t \mapsto u(t) \in \mathbb{R}^d$ that solves (uniquely) the Cauchy problem

$$\begin{cases} \frac{d}{dt}u(t) = -\nabla\phi(u(t)) \\ \lim_{t\to 0^+} u(t) = u_0 \end{cases} \qquad \cdots\cdots\cdots \ (\mathrm{P_{GI}})$$

Proposition 1) Suupose ϕ is a convex function. Let u_1, u_2 be two solutions of $(\mathrm{P_{GI}})$, then we have

$$\frac{d}{dt}\phi(u(t)) = -\left\|\nabla\phi(u(t))\right\|^2$$

and

$$\|u_1(t) - u_2(t)\| \leq \|u_1(0) - u_2(0)\|$$

i.e. gradient flow is a contraction.

In particular, $(\mathrm{P_{GI}})$ has a unique solution with initial condition given.

proof) The first equality is a direct consequence of $(\mathrm{P_{GI}})$.

To see the second inequality, let $g(t) = \frac{1}{2}\|u_1(t) - u_2(t)\|^2$. Then since $u_j(t) = -\nabla\phi(u_j(t))$ $(j = 1, 2)$,

$$\begin{aligned} g'(t) &= \langle u_1(t) - u_2(t), u_1'(t) - u_2'(t)\rangle \\ &= -\langle u_1(t) - u_2(t), \nabla\phi(u_1(t)) - \nabla\phi(u_2(t))\rangle \leq 0 \end{aligned}$$

by convexity of ϕ. So $g(t) : [0, +\infty) \to \mathbb{R}_+$ has $g(t) \leq g(0)$.

(End of proof) □

We can draw a stonger result if ϕ is λ-convex. Suppose ϕ is λ-convex. Then

$$-\langle u_1(t) - u_2(t), \nabla\phi(u_1(t)) - \nabla\phi(u_2(t))\rangle \leq -\lambda\|u_1(t) - u_2(t)\|^2$$

so

$$g'(t) \leq -\lambda\|u_1(t) - u_2(t)\|^2 = -2\lambda g(t)$$

The *Gronwall's lemma* now implies

$$g(t) \leq e^{-2\lambda t}g(0)$$

and in other words,

$$\|u_1(t) - u_2(t)\|^2 \leq e^{-2\lambda t}\|u_1(0) - u_2(0)\|^2,$$

the *exponential convergence* (also called linear convergence).

Moreover, we have that if ϕ is λ-convex, then $\mathrm{argmin}_u\phi$ is a singleton. So if $u_1(t)$ is a curve and $u_2(t) \equiv \overline{u} = \mathrm{argmin}(\phi)$, then

$$\left\|u_1(t) - \overline{u}\right\|^2 \leq e^{-2\lambda t}\left\|u_1(0) - \overline{u}\right\|^2 \to 0$$

Definition) An **explicit Euler scheme** solves for

$$\frac{u_{k+1}^\tau - u_k^\tau}{\tau} = -\nabla\phi(u_k^\tau)$$

This is very easy to implement. However, if we choose $\tau(1-\frac{\tau L}{2}) > 0$, we get a stability issue.
An **implicit Euler scheme** solves for

$$\frac{u_{k+1}^\tau - u_k^\tau}{\tau} = -\nabla\phi(u_{k+1}^\tau)$$

This corresponds to

$$u_{k+1}^\tau = \operatorname{argmin}\Big\{\phi(u) + \frac{1}{2\tau}\|u-u_k^\tau\|^2\Big\} \quad \Leftrightarrow \quad 0 = \nabla\phi(u_{k+1}^\tau) + \frac{1}{\tau}(u_{k+1}^\tau - u_k^\tau)$$

With such choice of u_k^τ strating from $u_0^\tau = u_0$,

$$\phi(u_{k+1}^\tau) + \frac{1}{2\tau}\left\|u_{k+1}^\tau - u_k^\tau\right\|^2 \leq \phi(u_k^\tau)$$

so if we let $\Delta_k = \frac{1}{2\tau}\left\|u_{k+1}^\tau - u_k^\tau\right\|^2$, then $\Delta_k \leq \phi(u_k^\tau) - \phi(u_{k+1}^\tau)$, and so

$$\sum_{k+1}^{K} \Delta_k \leq \phi(u_0^\tau) - \phi(u_{K+1}^\tau) =: C < \infty$$

If we have defined

$$u^\tau(t) = u_k^\tau \quad \text{for } t = k\tau$$

$$\tilde{u}^\tau(t) = u_k^\tau + (t-k\tau)\nu_{k+1}^\tau \quad \text{for } t\in[k\tau,(k+1)\tau) \text{ and } v_{k+1}^\tau \qquad\qquad = \frac{u_{k+1}^\tau - u_k^\tau}{\tau}$$

then $(\tilde{u}^\tau)'(t) = v^\tau(t)$ and

$$\frac{\|u_{k+1}^\tau - u_k^\tau\|^2}{2\tau} = \frac{\tau}{2}\Big(\frac{\|u_{k+1}^\tau - u_k^\tau\|^2}{\tau^2}\Big)^2 = \frac{\tau}{2}\left\|v_{k+1}^\tau\right\|^2 = \frac{\tau}{2}\int_{k\tau}^{(k+1)\tau}\frac{1}{2}\left\|(\tilde{u}^\tau)'(t)\right\|^2 dt$$

and because of summability ($\sum_{k+1}^{K}\Delta_k = C < \infty$),

$$\frac{\tau}{2}\int_0^T \frac{1}{2}\left\|(\tilde{u}^\tau)'(t)\right\|^2 < C$$

Using this, one can show that $u^\tau(t)$ converges uniformly to u on every compact set $[0,T]$.

(11th March, Monday)

In Euclidean spaces, we defined the solution u of

$$\frac{du}{dt}(t) = -\nabla\phi(u(t)) \quad \cdots\cdots\cdots \ (*)$$

to be the gradient flow of ϕ. Note,

(a)

$$\frac{d}{dt}\phi(u(t)) = \langle \nabla\phi(u(t)), \frac{du}{dt}(t)\rangle = -\left\|\nabla\phi(u(t))\right\|^2$$

If ϕ is λ-convex, then

(b) $(*)$ is equivalent in $\mathbb{R}^d$ to **Propsition 8.8**

$$\frac{d}{dt}\phi(u(t)) = -\frac{1}{2}\left|\frac{du}{dt}(t)\right|^2 - \frac{1}{2}|\nabla\phi(u(t))|^2 \quad \forall t \in (0,\infty) \quad \cdots\cdots\cdots \text{(EDE)}$$

(c) $(*)$ is equivalent in $\mathbb{R}^d$ to

$$\frac{d}{dt}\left(\frac{1}{2}|u(t)-v|^2\right) + \frac{\lambda}{2}|u(t)-v|^2 \leq \phi(v) - \phi(u(t)) \quad \forall t \in (0,\infty),\ \forall v \in \mathbb{R}^d \quad \cdots\cdots\cdots \text{(EVI)}$$

8.2 Gradient Flows in Metric Spaces

Recall : A curve $w : (a,b) \to Z$ is **absolutely continuous** in a metric space (Z,d) if $\exists g \in L^1((a,b))$ such that

$$d(w(t_0), w(t_1)) \leq \int_{t_0}^{t_1} g(s)ds \quad \forall t_0, t_1 \in (a,b) \text{ with } t_0 < t_1 \quad \cdots\cdots\cdots (**)$$

And if $g \in L^p((a,b))$ we write $w \in \mathrm{AC}^p((a,b), Z)$.

Definition) We define the **metric derivative** of w by

$$|w'|(t) = \lim_{s\to t}\frac{d(w(s),w(t))}{|t-s|} \quad \cdots\cdots\cdots (\oplus)$$

if the limit exists.

Theorem 8.10) Let (Z,d) be a complete and separable metric space. If $w : (a,b) \to Z$ is absolutely continuous then the limit $(\oplus)$ exists for Lebesgue-a.e. $t \in (a,b)$. Moreover, the function

$$(a,b) \to \mathbb{R}, \quad t \mapsto |w'|(t)$$

is an $L^1(a,b)$-function and one can choose $g = |w'|$ in $(**)$.

If $\tilde{g}$ is any other function satisfying $(**)$, then $|w'|(t) \leq \tilde{g}(t)$ for Lebesgue-a.e. t.

Definition 8.11) Let (Z,d) be a metric space then the **metric slope** of $\phi : Z \to \mathbb{R}$ at $v \in \mathbb{R}$ is defined by

$$|\partial\phi|(v) = \limsup_{w\to v}\frac{(\phi(v)-\phi(w))_+}{d(v,w)}$$

8.2.1 GVI Gradient Flows

Definition 8.12) Given a metric space (Z, d) and a function $\phi : Z \to \mathbb{R}$ an **evolution variational inequality ($\mathbf{EVI}_\lambda$) gradient flow** is a locally absolutely continuous curve

$$(a, b) \to \mathbb{R}, \quad t \mapsto u(t) \in \mathrm{Dom}(\phi) \subset Z$$

satisfying

$$\frac{1}{2}\frac{d}{dt}\Big(d^2(u(t), v)\Big) + \frac{\lambda}{2}d^2(u(t), v) \leq \phi(v) - \phi(u(t)) \quad \text{for a.e. } t \in (0, \infty)$$

Proposition 8.13) Let (Z, d) be a complete and separable metric space and $\phi : Z \to \mathbb{R} \cup \{+\infty\}$ a proper function. If u and v are EVI_λ gradient flows with initial condition $u(0) = u_0$ and $v(0) = v_0$. Then

$$d(u(t), v(t)) \leq e^{-\lambda t} d(u_0, v_0)$$

Remark : If $\lambda > 0$ then $\exists$ at most one EVI_λ gradient flow and $u(t)$ converges exponentially to u^*, the minimiser of ϕ, *i.e.*

$$d(u(t), u^*) \leq e^{-\lambda t} d(u_0, u^*)$$

We have not mentioned that ϕ is λ-convex, but if the EVI_λ gradient flow exists for any initial condition, then ϕ is λ-convex, and this implies that the minimiser is unique.

(13th March, Wednesday)

Proposition 8.13) Let (Z, d) be a complete and separable metric space and $\phi : Z \to \mathbb{R} \cup \{+\infty\}$ a proper function. If u and v are EVI_λ gradient flows with initial condition $u(0) = u_0$ and $v(0) = v_0$. Then

$$d(u(t), v(t)) \leq e^{-\lambda t} d(u_0, v_0)$$

proof) We have, for any $w \in \mathbb{R}$,

$$\frac{1}{2}\frac{dt}{d}d^2(u(t), w) + \frac{\lambda}{2}d^2(u(t), w) \leq \phi(w) - \phi(u(t)) \quad \cdots\cdots\cdots \ (1)$$
$$\frac{1}{2}\frac{dt}{d}d^2(v(t), w) + \frac{\lambda}{2}d^2(v(t), w) \leq \phi(w) - \phi(v(t)) \quad \cdots\cdots\cdots \ (2)$$

Choose $w = v(s)$ in (1) and $w = u(s)$ in (2), then (1), (2) and λ-convexity of u together gives

$$\frac{1}{2}\frac{d}{dt}d^2(u(t), v(s))\Big|_{s=t} + \frac{1}{2}\frac{d}{dt}d^2(v(t), u(s))\Big|_{s=t} \leq -\lambda d^2(u(t), v(t))$$

Since

$$\frac{d}{dt}d^2(u(t), v(t)) = \frac{d}{dt}d^2(u(t), v(s))\Big|_{s=t} + \frac{d}{dt}d^2(u(s), v(t))\Big|_{s=t}$$

we have

$$\frac{1}{2}\frac{d}{dt}d^2(u(t),v(t)) \leq -\lambda d^2(u(t),v(t))$$

Multiply by $e^{2\lambda t}$, then

$$\frac{d}{dt}\Big(e^{2\lambda t}d^2(u(t),v(t))\Big) = 2\lambda e^{2\lambda t}d^2(u(t),v(t)) + e^{2\lambda t}\frac{d}{dt}d^2(u(t),v(t)) \leq 0$$

Hence we have the result.

(End of proof) □

(skipping some materials in the section.)

8.3 Gradient Flows in the Wasserstein Space

We will need the continuity equation (conservation of mass). Assume we have density $\rho(x,t)$ at times t. So each set $A \subset \mathbb{R}^d$ has mass $\int_A \rho(x,t)dx$. We assume that mass is only lost on the boundary of A. So

$$\int_A \frac{\partial\rho}{\partial t}(x,t)dt = \frac{d}{dt}\int \rho(x,t)dx = \int_{\partial A} v(x,t)\cdot n(x,t)dS$$

where the mass is moving with velocity v and n is the unit normal to ∂A. By the divergence theorem,

$$\int_A \frac{\partial\rho}{\partial t}(x,t)dx = -\int_A \nabla\cdot(v(x,t),\rho(x,t))dx$$

This is true for any suitable A, so we have

$$\frac{\partial\rho}{\partial t}(x,t) = -\nabla\cdot(v(x,t)\rho(x,t)),$$

called the **Continuity equation**.

We now make change of notation, $u(t)\to\mu_t$.

8.3.1 Wasserstein Tangent Space

(see [4] Ambrosio, Gigli, Savaré, 2008, "Gradient Flows in Metric Spaces" for more information)

Theorem 8.23) *(Abolutely continous curves / the continuity equation)*

(1) Let $(0,+\infty)\to\mathcal{P}_2(\mathbb{R}^d)$, $t\mapsto\mu_t$ be absolutely continuous and $|\mu'|\in L^1((0,+\infty))$ be its metric derivative. Then $\exists$ a vector field $v_t\in L^2(\mu_t,\mathbb{R}^d)$ such that

$$\|v_t\|_{L^2(\mu_t;\mathbb{R}^d)} \leq |\mu'|(t) \quad \text{for a.e. } t\in(0,+\infty)$$

and

$$\frac{\partial}{\partial t}\mu_t + \nabla\cdot(v_t\mu_t) = 0 \quad \text{in } \mathbb{R}^d\times(0,+\infty) \quad \cdots\cdots\cdots \ (*)$$

holds in the sense of distributions.

(2) Let $(0,+\infty) \to \mathcal{P}_2(\mathbb{R}^d)$, $t \mapsto \mu_t$ be continuous with respect to weak* topology on $\mathcal{P}(\mathbb{R}^d)$ and satisfies $(*)$ for some vector field v_t with

$$\int_0^\infty \|v_t\|_{L^2(\mu,\mathbb{R}^d)} dt < +\infty$$

Then μ_t is absolutely continuous and $|\mu'(t)| \leq \|v_t\|_{L^2(\mu_t;\mathbb{R}^d)}$ for a.e. $t \in (0,\infty)$

[We define $(*)$ to hold in the sense of distributions iff

$$\int_0^\infty \int_{\mathbb{R}^d} \Big(\frac{\partial f}{\partial t}(x,t) + v_t(x) \cdot \nabla f(x,t)\Big) d\mu_(x) dt = 0 \quad \forall f \in C_c^\infty(\mathbb{R}^d \times (0,+\infty))$$

]

The following proposition motivates the tangent space. The tangent direction can be thought of as the direction in which a variation in the direction stays in the probability space.

Proposition 8.26) Let $(0,+\infty) \to \mathcal{P}_2(\mathbb{R}^d)$, $t \mapsto \mu_t$ be absolutely continous. Suppose $v_t \in \overline{\{\nabla\varphi : \varphi \in C_c^\infty(\mathbb{R})\}}^{L^2(\mu_t;\mathbb{R}^d)}$ satisfies $(*)$. Assume there is an optimal transport map $T^{(t,s)}$ between μ_t and ν_s for each s, t, *i.e.* $T^{(t,s)}_{\#}\mu_t = \mu_s$. Then

$$\lim_{h\to 0} \frac{T^{(t,t_h)} - id}{h} = v_t$$

where this limit is taken in $L^2(\mu_t;\mathbb{R}^d)$ (for all t? check)

Definition 8.27) For $\mu \in \mathcal{P}_2(\mathbb{R}^2)$, the **tangent space to $\mathcal{P}_2(\mathbb{R}^d)$ at the point** μ is

$$\mathrm{Tan}_\mu \mathcal{P}_2(\mathbb{R}^d) = \overline{\{\nabla\varphi; \varphi \in C_c^\infty(\mathbb{R}^d)\}}^{L^2(\mu;\mathbb{R}^d)}$$

We can use **Proposition 8.26** to differentiate $t \mapsto \frac{1}{2} d^2_{W^2}(\mu_t, \sigma)$.

Theorem 8.28) Let $\sigma \in \mathcal{P}_2(\mathbb{R}^d)$, $(0,+\infty) \to \mathcal{P}_2(\mathbb{R}^d)$, $t \mapsto \mu_t$ be absolutely continuous and $v_t \in \mathrm{Tan}_{\mu_t}\mathcal{P}_2(\mathbb{R}^d)$ satisfy $(*)$. Assume there is an optimal transport map T^t between σ and μ_t for all t, *i.e.* $T^{(t)}_{\#}\mu_t = \sigma$. Then

$$\frac{1}{2}\frac{d}{dt} d^2_{W^2}(\mu_t, \sigma) = \int_{\mathbb{R}^d} (x - T^{(t)}(x)) \cdot v_t(x) d\mu_t(x)$$

sketch proof) Maps $T^{(t)}$, $T^{(t,t+h)}$, $T^{(t+h)}$ are all optimal maps. We suppose $T^{(t)} = T^{(t+h)} \circ T^{(t,t+h)}$. *[Of course, this is not true in general, but there is a reasonable justification for this : when h is small, then composition of optimal maps would be close to an optimal map, so we expect the equality to hold up to a $o(h)$ correction.]* Then

$$\begin{aligned}
\frac{d}{dt} d^2_{W^2}(\mu_t,\sigma) &= \lim_{h\to 0^+} \frac{1}{h}\Big[\int_{\mathbb{R}^d} |T^{(t+h)}(x) - x|^2 d\mu_{t+h}(x) - \int_{\mathbb{R}^d} |T^{(t)}(x) - x|^2 d\mu_t(x)\Big] \\
&= \lim_{h\to 0^+} \frac{1}{h} \int_{\mathbb{R}^d} \Big[|T^{(t+h)}(T^{(t,t+h)}(x)) - T^{(t,t+h)}(x)|^2 - |T^{(t)}(x) - x|^2\Big] d\mu_t(x) \\
&= \lim_{h\to 0^+} \int_{\mathbb{R}^d} (x - T^{(t,t+h)}(x)) \cdot (2T^{(t)}(x) - x - T^{(t,t+h)}(x)) d\mu_t(x) \\
&= \int_{\mathbb{R}^d} (-v_t(x)) \cdot (2T^{(t)}(x) - x - x) d\mu_t(x)
\end{aligned}$$

(End of proof) $\square$