Notes on Optimal Transport

Shuailong Zhu

Contents

1 Some Preliminaries for Optimal Transport

1.1 About Push-Forward

Proposition 1.1.1. *Let $T : X \to Y$, $\mu \in \mathcal{P}(X)$, and $v \in \mathcal{P}(X)$. Then*

$$v = T\mu$$

if and only if for anly $\varphi : Y \to \mathbb{R}$ Borel and bounded, we have that

$$\int_Y \varphi(y)dv(y) = \int_X \varphi(T(x)d\mu(x)$$

Proof. For any Borel set $A \subset Y$, it holds

$$\int_Y \mathbb{1}_A dv = \mu(T^{-1}(A)) = \int_Y \mathbb{1}_{T^{-1}(A)} d\mu = \int_Y \mathbb{1}_A \circ T d\mu$$

Thus, for any simple function $\varphi : Y \to \mathbb{R}$,

$$\int_Y \varphi dv = \int_Y \varphi \circ T d\mu$$

Consider a fixed bounded Borel function, we can have a sequence of simple functions $(\varphi_k)_{k\in\mathbb{N}}$ such that $|\varphi_k - \varphi| \to 0$ uniformly. Then,

$$\int_Y \varphi dv = \lim_{k\to\infty} \int_X \varphi_k dv = \lim_{k\to\infty} \int_X \varphi_k \circ T d\mu = \int_X \varphi \circ T d\mu$$

where the last equality comes from dominated convergence theorem because $\varphi \circ T$ can still bound $\varphi_k \circ T$, but how to show it is measurable and absolutely integrable? Absolutely integrable can be deduced from the boundedness and finite measure μ; measurable should be related to borel measurability of φ, which is very similar to Example **??**. □

Proposition 1.1.2 (Change of Variable, Theorem 6.1.7 [1])**.** *Assume that T is a diffeomorphism between open sets X and Y of $\mathbb{R}^d$, and assume probability measures μ, v are absolutely continuous with respect to Lebesgue measure. Then,*

$$\int_Y \varphi(y)\sigma(y)dy = \int_X \varphi(T(x))\sigma(T(x))det(DT(x))dx$$

1.2 Weak-$*$ Topology and Narrow Topology

Proposition 1.2.1 (Riesz Representation Theorem for $C_c(X)$)**.**

$$\begin{aligned} \mathcal{M}(X) &:= \{\textit{finite signed measures on } X\} && (1.2.1)\\ &= C_c(X)^* := \{\textit{continuous compactly supported functions}\}^* && (1.2.2)\\ &= C_0(X)^* := \{\textit{continuous functions vanishing at } \infty\}^* && (1.2.3) \end{aligned}$$

Precisely, $(\mathcal{M}(X), \|\cdot\|_{TV})$ is the dual space of $(C_0(X), \|\cdot\|_{L^\infty})$ or $(C_c(X), \|\cdot\|_{L^\infty})$. Notice $C_c(X)$ not closed, $C_0(X)$ closed.

Weak-$\star$ convergence. By Banach-Alaoglu's Theorem, if $(\mu_k)_{k\in\mathbb{N}}$ is a sequence of probability measures, then $\exists$ a subsequence that weakly-* converges to a measure $\mu \in M(X)$.

Narrow Topology. We say $\mu_k \overset{*}{\rightharpoonup} \mu$ if

$$\int \varphi d\mu_k \to \int \varphi d\mu, \text{ for all } \varphi \in C_b(X) \tag{1.2.4}$$

This is equivalent to: if

$$\liminf_{k\to\infty} \int \varphi d\mu_k \geq \int \varphi d\mu \tag{1.2.5}$$

for all φ that is lower semi-continuous and lower bounded.

2 Existence and Optimal Condition in (KP)

2.1 Existence and Optimal Condition

Proposition 2.1.1 (Existence of Optimal Plan/Coupling γ in (KP) problem)**.** *Let $c : X \times Y \to [0, infty]$ be lower semicontinuous, $\mu, v \in \mathcal{P}(X)$. Then there exists a coulping $\gamma \in \Gamma(\mu, v)$ which is a minimizer for (KP).*

The strategy we use is: (1) Compactness in a certain topology, i.e. narrow topology of $\Gamma(\mu, v)$ deduced from tightness; (2) lower semi-continuity. We might use similar strategy for the space of transport map (T_k) for (MP).

Remark. However, this strategy doesn't work for (MP). Using Proposition 3.8 and 3.9 in my notes of functional analysis, we can check the "admitting of a weakly convergent sequence" in L^2 and the weak limit of the operator.

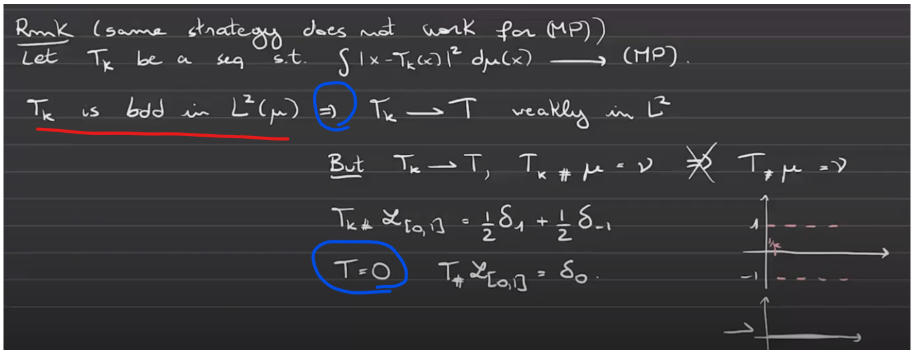

Figure 1: Same strategy doesn't work for (MP) [2]

Theorem 2.1.2 (Optimal Condition)**.** *Let $\bar{\gamma}$ be optimal, and $c : X \times Y \to \mathbb{R}$ is continuous. Then supp($\bar{\gamma}$) is c-cyclically monotone. Note: assume $\mu = \sum \frac{1}{2^i}\delta_{q_i} \in \mathcal{P}(\mathbb{R})$, where q_i is a rational number. Then supp(μ) $= \mathbb{R}$.*

2.2 Some Convex Analysis Tools

A function $\varphi : \mathbb{R}^d \to \bar{\mathbb{R}}$ is convex if

$$\varphi(x) = \sup_{y\in\mathbb{R}^d} \{x \cdot y + \lambda_y\} \tag{2.2.1}$$

Definition 2.2.1 (c-convex)**.** *Using the idea of the supremum of affine functions for convex, given X and Y metric spaces, $c : X \times Y \to \mathbb{R}$, we define that $\varphi : X \to \bar{\mathbb{R}}$ is c-convex if*

$$\varphi(x) = \sup_{y\in\mathbb{R}^d} \{-c(x, y) + \lambda_y\} \tag{2.2.2}$$

Theorem 2.2.1. *[Rockafellar, from convex analysis] A set $S \subset \mathbb{R}^d \times \mathbb{R}^d$ is (c-)cyclically monotone iff there exists a (c-)convex function $\varphi : \mathbb{R}^d \to \bar{\mathbb{R}}$ such that $S \subset \partial\varphi$. Rockafellar theorem provides a more clear description of c-cyclically monotone.*

2.3 General Kantorovich Duality

Definition 2.3.1 (c-Legendre transform)**.** *Given a c-convex function $\varphi : X \to \bar{\mathbb{R}}$, we define its c-Legendre transform $\varphi^c : Y \to \bar{\mathbb{R}}$ as*

$$\varphi^c(y) = \sup_{x\in X}\{-c(x, y) - \varphi(x)\} \tag{2.3.1}$$

Properties.

$$\varphi(x) + \varphi^c(y) + c(x, y) \geq 0 \text{ for all } x \in X, y \in Y \tag{a}$$

$$\varphi(x) + \varphi^c(y) + c(x, y) = 0 \text{ iff } y \in \partial_c\varphi(x) \tag{b}$$

Theorem 2.3.1 (Kantorovich duality)**.** *Let $c(\cdot, \cdot)$ be continuous and bounded from below, and assume that $\inf_{\gamma\in\Gamma(\mu,v)} \int_{X\times Y} c d\gamma < +\infty$, then*

$$\min_{\gamma\in\Gamma(\mu,v)cd\gamma} \int_{X\times Y} cd\gamma = \max_{\varphi(x)+\psi(y)+c(x,y)\geq 0} \int_X -id\mu + \int_Y -\psi dv$$

Proof. . (1) (KP)$\geq$ (DP).

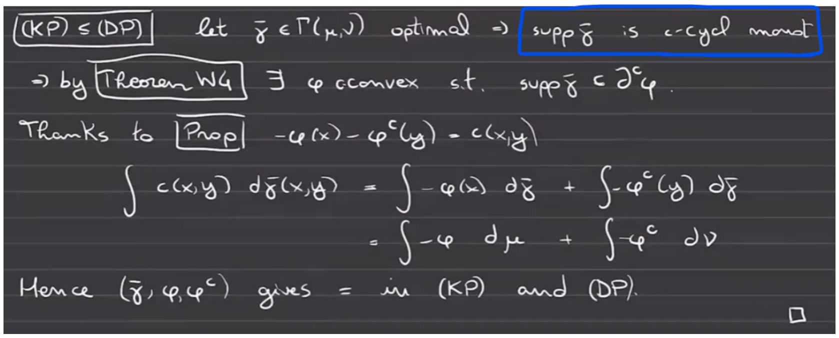

Figure 2: (KP) $\leq$ (DP)

$\square$

Remark. The existence of minimizer of LHS only requires $c(\cdot,\cdot)$ to be lower-continuous in Proposition 2.1.1. We can obtain some corollaries that are a bit tricky due to the proof's specialty, i.e. it only requires that supp(γ) to be c-cyclically monotone.

Corollary 2.3.2 (It might be inaccurate. We might need to put this corollary under the setting of c continuous.)**.** *Given a γ with supp(γ) is c-cyclically monotone, then $\exists$ φ c-convex such that*

$$(KP) \leq \int c d\gamma = \int -\varphi d\mu + \int -\varphi^c dv \leq (DP)$$

Well, (KP)$\geq$(DP) is general. Thus, we can obtain (KP)=(DP) and γ is the optimal.

Interestingly, Theorem 2.1.2 is for continuous cost, while Theorem 2.2.1 is for "any" cost.

Corollary 2.3.3. *if $c(\cdot,\cdot)$ is continuous, the following are equivalent:*

- *γ is optimal;*
- *supp(γ) is c-cyclically monotone;*
- *there exists a convex map φ such that supp$(\gamma) \subset \partial_c \varphi$*

Corollary 2.3.4 (Theorem 2.3.2, [3])**.** *If $c(\cdot,\cdot)$ is lower semi-continuous, then we still have (KP)=(DP), which can be shown by approximation.*

We can find an excellent example that is c-cyclically monotone but not optimal, which is complementary to Corollary 2.3.3. However, this exmaple seems a bit controversial to Corollary 2.3.2.

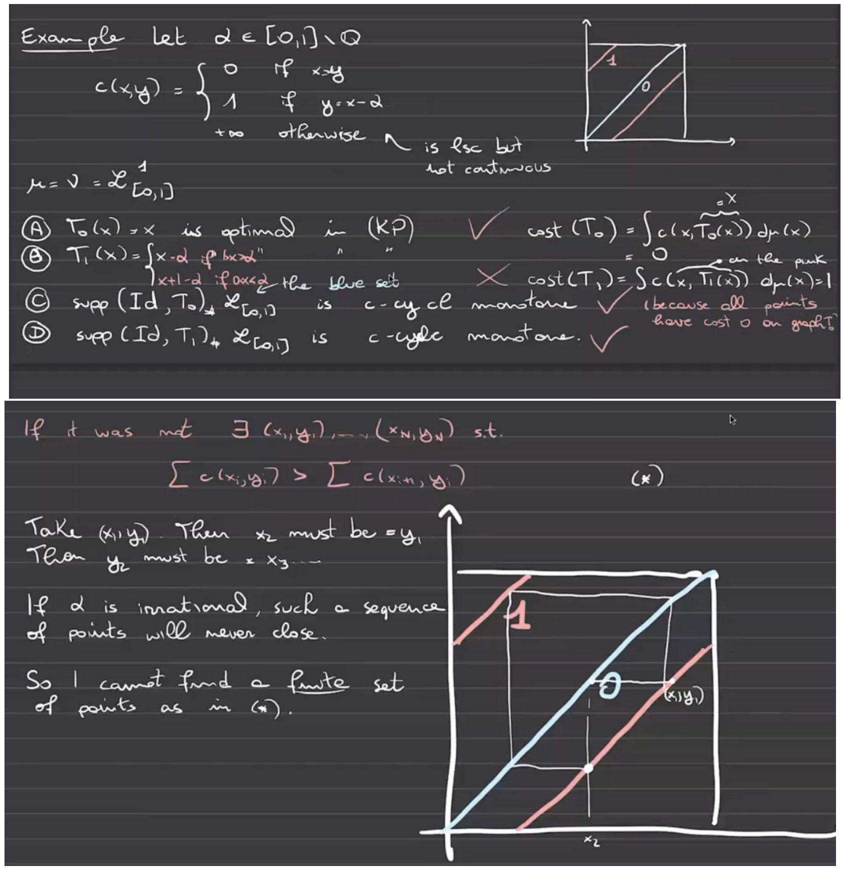

Figure 3: A beautiful counter example.

2.4 From Convex Geometry to (KP)=(DP)

$$
\begin{aligned}
\inf_{\gamma\in\Gamma(X,Y)} c(x,y)d\gamma &= \inf_{\gamma\geq 0}\left\{\int c d\gamma + \sup_{\varphi}\{\int_{X\times Y}\varphi(x)d\gamma - \int_X \varphi(x)d\mu\} + \sup_{\varphi}\{\int_{X\times Y}\psi(y)d\gamma - \int_Y \psi(y)dv\}\right\} \\
&= \inf_{\gamma\geq 0}\sup_{\varphi,\psi}\left\{\int c d\gamma + \int_{X\times Y}\varphi(x)d\gamma - \int_X\varphi(x)d\mu + \int_{X\times Y}\psi(y)d\gamma - \int_Y\psi(y)dv\right\} \\
&\qquad\qquad \text{(Simon Minimax Theorem)} \\
&= \sup_{\varphi,\psi}\left\{-\int_X\varphi(x)d\mu - \int_Y\psi(y)dv + \inf_{\gamma\geq 0}\int_{\times Y}[c+\varphi(x)+\psi(y)]d\gamma\right\} \\
&= \sup_{\varphi(x)+\psi(y)+c(x,y)\geq 0}\left\{-\int_X\varphi(x)d\mu - \int_Y\psi(y)dv\right\}
\end{aligned}
$$

3 Existence and Characterization of Transport Maps in (MP)

3.1 Brenier's Theorem

Theorem 3.1.1 (Brenier's Theorem)**.** *Let $X = Y = \mathbb{R}^d$, $c(x,y) = \dfrac{|x-y|^2}{2}$. Suppose that*

$$
\int_X |x|^2 dx + \int_Y |y|^2 dy < +\infty
$$

and $\mu \ll dx$. Then there exists a unique optimal plan γ. In addition, $\gamma = (Id \times T)_{\#}\mu$ and $T = \nabla\varphi$ for some convex function φ.

Theorem 3.1.2 (General Brenier's Theorem)**.** *Let $X = Y = \mathbb{R}^d$, $\mu \ll dx$, and supp(v) compact. Let c be continuous and bounded from below, and assume that $inf_{\gamma} \int_{X\times Y} < \infty$. Also, suppose that:*

- *for every $y \in supp(v)$, the map $x \to c(x,y)$ is differentiable;*
- *for every $x \in \mathbb{R}^d$, the map $y \to \nabla_x c(x,y)$ is injective.*
- *for every $y \in supp(v)$ and $R > 0$, $|\nabla_x c(x,y)| \le C_R$ for every $x \in \mathbb{B}_R$*

Then there exists a unique optimal plan γ with $\gamma = (Id \times T)_{\#}\mu$ and T satisfying

$$\nabla_x c(x,y)|_{y=T(x)} + \nabla\varphi(x) = \nabla_x c(x, T(x)) + \nabla\varphi(x) = 0$$

for some c-convex function φ.

3.2 Stability and Regularity of Optimal Transport Plans/Maps

Before going to any details, it seems that we refer to "the same γ" in the topology of $\Gamma(\mu, v)$, we are talking about almost everywhere equal.

Theorem 3.2.1 (Stability)**.** *Let $\{\mu_k\}, \{v_k\} \subset \mathcal{P}(X)$ with $supp(\mu_k), supp(v_k) \subset K$ compact,*

$$\mu_k \rightharpoonup \mu, v_k \rightharpoonup v$$

Let $c : X \times X \to [0, \infty]$ Then,

- *Any weak limit point of π_k optimal in $\Gamma(\mu_k, v_k)$ is optimal.*
- *if $X = \mathbb{R}^d$, $c(x,y) = \dfrac{\|x-y\|^2}{2}$, $\mu \ll \mathscr{L}^d$ then*

  $$\pi_h \to (Id, \nabla\varphi)_{\#}\mu$$

  If $\mu_h \ll \mathscr{L}^d$, then

  $$(Id, \nabla\varphi_h)_{\#}\mu_h \to (Id, \nabla\varphi)_{\#}\mu$$

  The above is just a simple consequence of the first point and Brenier's theorem. More interestingly, if $\mu \equiv \mu_h$

  $$\nabla\varphi_k \to \nabla\varphi \text{ in } L^p(\mu)$$

Example 3.2.1 (Regularity)**.** *Even if the $supp(\mu_h)$ and $supp(v_h)$ is connected, there exists h_0, such that T_h for $h \ge h_0$ might not be continuous.*

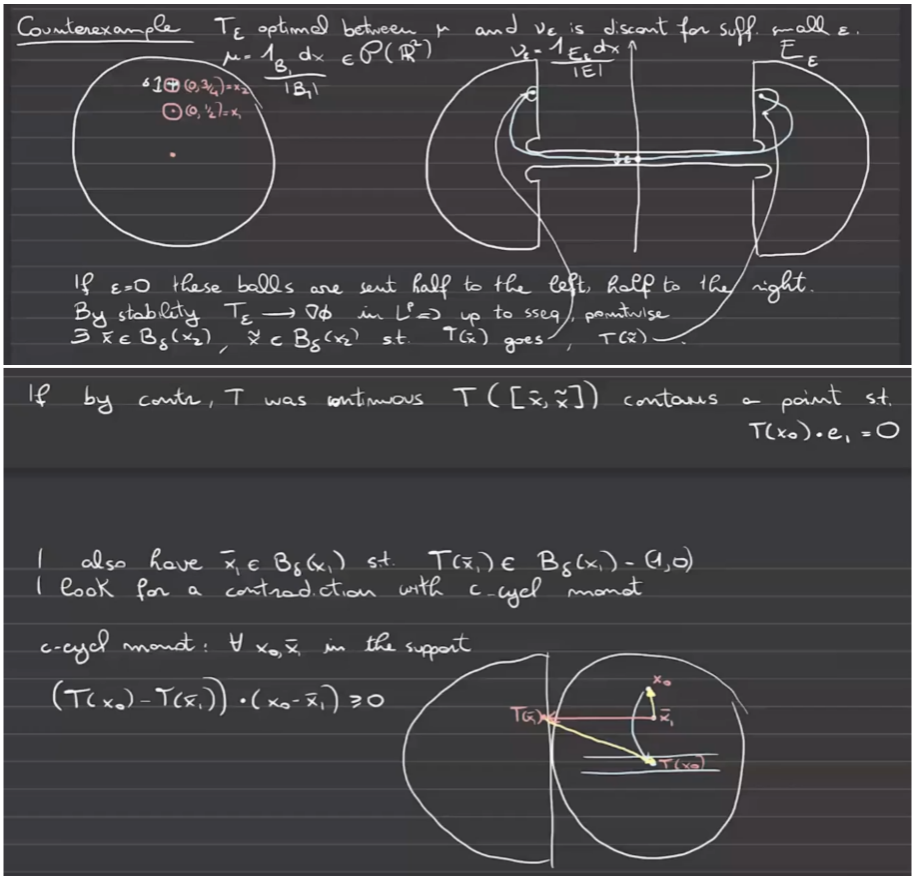

Figure 4: A example for regularity.

4 Wasserstein Space and Geodesic

We will leave the tasks of organizing this part of note to future.

References

[1] Donald L Cohn. *Measure theory*. Vol. 5. Springer, 2013.

[2] David Mandel. *Notes on Modes of Convergence in Probability Theory*. URL: https://www.math.fsu.edu/~dmandel/Primers/Convergence.pdf.

[3] Alessio Figalli and Federico Glaudo. *An invitation to optimal transport, Wasserstein distances, and gradient flows*. 2021.