CS2910 - Optimization Summer 2023

Lecture A. Optimal Transport

Lecturer: Yunwei Ren *Scribed by Zhidan Li*

Contents

1 Overview **1**

2 Optimal Transport **2**

2.1 Kantorovich duality 3

1 Overview

In this (advanced) lecture, we introduce a type of optimization problems called the *optimal transport*. In this lecture, we focus on the optimal 'cost' of transporting one probability measure to another. For a state space Ω, let $\mathcal{B}(\Omega)$ be the collection of all Borel sets in Ω.

Definition 1 (probability measure)**.** *Given the state space $\Omega = \mathbb{R}^d$ equipped with the Borel algebra $\mathcal{B}(\mathbb{R}^d)$, $\mu : \mathcal{B}(\mathbb{R}^d) \to [0, 1]$ is said to be a* ***probability measure*** *if*

- $\mu(\varnothing) = 0$.
- $\mu(\mathbb{R}^d) = 1$.
- *For any disjoint measurable set $A_1, A_2, \ldots$,*

$$\mu(A_1 \cup A_2 \cup \ldots) = \sum_{i=1}^{\infty} \mu(A_i).$$

We use $P(\mathbb{R}^d)$ to denote the space of all probability measures over $\mathbb{R}^d$.

Also, for more concise and precise description, we introduce the *coupling* of two probability measures.

Definition 2 (coupling)**.** *For two probability measures $\mu, \nu \in P(\mathbb{R}^d)$, we say $\gamma : \mathbb{R}^d \times \mathbb{R}^d \to [0, 1]$ is a* ***coupling*** *of μ and ν if:*

- *For all measurable $A \subseteq \mathbb{R}^d$,*

$$\mu(A) = \int_{y \in \mathbb{R}^d} d\gamma(A, y).$$

- *For all measurable $B \subseteq \mathbb{R}^d$,*

$$\nu(B) = \int_{x \in \mathbb{R}^d} d\gamma(x, B).$$

2 Optimal Transport

Now we describe the optimal transport in details. For a transport map $T : \mathcal{B}(\mathbb{R}^d) \to \mathcal{B}(\mathbb{R}^d)$, define the operator $T_\# : P(\mathbb{R}^d) \to P(\mathbb{R}^d)$ as: for all measurable set $E \subseteq \mathbb{R}^d$,

$$T_\#\mu(E) = \mu(T^{-1}(E)).$$

Definition 3 (Monge's problem)**.** *Given two probability measures $\mu, \nu \in \mathbb{R}^d$, a cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ meaning the cost of moving 1 unit of particles. Then, the* ***Monge's problem*** *is the following optimization problem:*

$$\begin{aligned} &\min_T \int_{x\in\mathbb{R}^d} c(x, T(x))\, d\mu(x) \\ &s.t.\ T_\#\mu = \nu. \end{aligned} \tag{1}$$

We also call this optimization problem (MP).

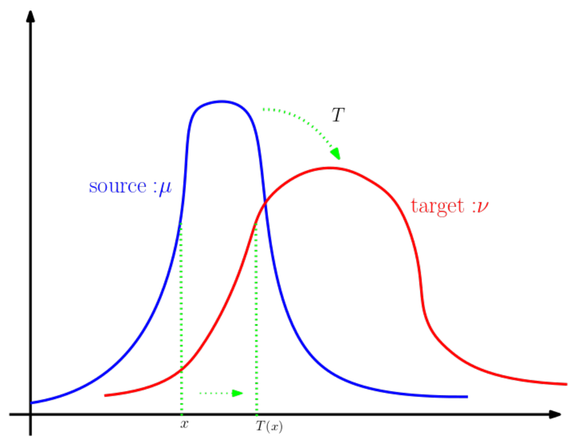

Figure 1: an illustration of Monge's problem

However, note that, for general μ and ν, the feasible transport map T does not necessarily exist. Here we give an example.

Let $\mu = \delta_0$ and $\nu = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$, where δ_x is the Dirac measure at x (only has point mass at x). Since the transport map cannot move particles from one position to multiple positions, it is impossible to transport μ to ν.

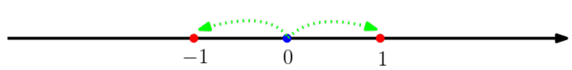

Figure 2: an example that there is no feasible transport map.

To handle this case, we use the coupling of two probability measures instead of transport map.

Definition 4 (Kantorovich's problem)**.** *Given two probability measures $\mu, \nu \in \mathbb{R}^d$, a cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ meaning the cost of moving 1 unit of particles. Then, the* ***Kantorovich's problem*** *is the following optimization problem:*

$$\begin{aligned} &\min_\gamma \int_{x\in\mathbb{R}^d} \int_{y\in\mathbb{R}^d} c(x, y)\, d\gamma(x, y) \\ &s.t.\ \gamma \text{ is a coupling of } \mu, \nu. \end{aligned} \tag{2}$$

For convenience we call this optimization problem (KP).

The coupling γ is also named as a *transport plan*. Note that, under the assumptions in Definition 4, there always exists a feasible coupling, since $\gamma = \mu \otimes \nu$ is always a coupling ($\mu \otimes \nu$ is defined as, for all measurable $A, B \subseteq \mathbb{R}^d$, $(\mu \otimes \nu)(A, B) = \mu(A)\nu(B)$).

Though γ does always exist, it's unfortunate that (KP) does not always admit a solution. We note that, under some regular conditions, (KP) admits a solution.

2.1 Kantorovich duality

To establish a solution to (KP), similarly to the things we do in Lecture 3, we introduce a kind of dual property called *Kantorovich duality*.

Let $\Pi(\mu, \nu)$ be the collection of all couplings of μ and ν, and $\mathcal{M}_+(\Omega)$ be the collection of all measures on Ω. Now we define the indicator function $\iota_{\Pi(\mu,\nu)} : \mathcal{M}_+(\mathbb{R}^d \times \mathbb{R}^d) \to \mathbb{R} \cup \{\infty\}$,

$$\iota_{\Pi(\mu,\nu)}(\gamma) = \begin{cases} 0, & \gamma \in \Pi(\mu, \nu) \\ \infty & \gamma \notin \Pi(\mu, \nu). \end{cases}$$

Additionally, let $C_b(\mathbb{R}^d)$ be the collection of continuous and bounded functions on $\mathbb{R}^d$. Now we use affine functions to express $\iota_{\Pi(\mu,\nu)}$.

Lemma 5. *It holds that*

$$\iota_{\Pi(\mu,\nu)}(\gamma) = \sup_{\varphi,\psi \in C_b(\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} \varphi(x)\, d\mu(x) + \int_{\mathbb{R}^d} \psi(y)\, d\nu(y) - \int_{\mathbb{R}^d \times \mathbb{R}^d} (\varphi(x) + \psi(y))\, d\gamma(x, y) \right\}.$$

Proof. Note that

$$\begin{aligned} L(\gamma, \varphi, \psi) &\triangleq \int_{\mathbb{R}^d} \varphi(x)\, d\mu(x) + \int_{\mathbb{R}^d} \psi(y)\, d\nu(y) - \int_{\mathbb{R}^d \times \mathbb{R}^d} (\varphi(x) + \psi(y))\, d\gamma(x, y) \\ &= \int_{\mathbb{R}^d} \varphi(x)\, d\left(\mu(x) - \int_{y \in \mathbb{R}^d} \gamma(x,\, dy)\right) + \int_{\mathbb{R}^d} \psi(y)\, d\left(\nu(y) - \int_{x \in \mathbb{R}^d} \gamma(\,dx, y)\right). \end{aligned}$$

When $\gamma \in \Pi(\mu, \nu)$, it is clear that by definition $L(\gamma, \varphi, \psi) = 0$. When $\gamma \notin \Pi(\mu, \nu)$, since we can put all 'weights' of function φ or ψ on any point $x \in \mathbb{R}^d$, $\iota_{\Pi(\mu,\nu)}(\gamma)$ is larger than any arbitrary positive real. Thus $\iota_{\Pi(\mu,\nu)}(\gamma) = \infty$. $\square$

Now it is equivalent to solve the following optimization problem.

$$\min_{\gamma \in \mathcal{M}_+(\mathbb{R}^d \times \mathbb{R}^d)} \sup_{\varphi,\psi \in C_b(\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} (c(x, y) - \varphi(x) - \psi(y))\, d\gamma(x, y) + \int_{\mathbb{R}^d} \varphi(x)\, d\mu(x) + \int_{\mathbb{R}^d} \psi(y)\, d\nu(y) \right\}.$$

Interchanging the infimum and supremum, we obtain the following optimization problem

$$\sup_{\varphi,\psi \in C_b(\mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d} \varphi(x)\, d\mu(x) + \int_{\mathbb{R}^d} \psi(y)\, d\nu(y) + \min_{\gamma \in \mathcal{M}_+(\mathbb{R}^d \times \mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} (c(x, y) - \varphi(x) - \psi(y))\, d\gamma(x, y) \right\} \right\}. \tag{3}$$

Now we focus on the term

$$\min_{\gamma \in \mathcal{M}_+(\mathbb{R}^d \times \mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} (c(x,y) - \varphi(x) - \psi(y)) \, d\gamma(x,y) \right\}.$$

Let $(\varphi \oplus \psi)(x,y) = \varphi(x) + \varphi(y)$. It is clear that

$$\min_{\gamma \in \mathcal{M}_+(\mathbb{R}^d \times \mathbb{R}^d)} \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} (c(x,y) - \varphi(x) - \psi(y)) \, d\gamma(x,y) \right\} = \begin{cases} 0 & \varphi \oplus \psi \le c, \\ -\infty & \text{otherwise.} \end{cases}$$

Then we derive the dual problem

$$\begin{aligned} \sup_{\varphi, \psi \in C_b(\mathbb{R}^d)} & \left\{ \int_{\mathbb{R}^d} \varphi(x) \, d\mu(x) + \int_{\mathbb{R}^d} \psi(y) \, d\nu(y) \right\} \\ & s.t. \ \varphi \oplus \psi \le c. \end{aligned} \tag{4}$$

Theorem 6. *Under some mild conditions the strong duality holds. When the strong duality holds, the optimum can be attained by some c-concave pair.*