Optimal transport

II Optimality Conditions and Consequences

Seongho, Joo

Seoul National University

Introduction

Let X and Y be compact metric spaces, $\mu \in \mathcal{P}(X), \mu \in \mathcal{P}(Y)$ and $c : X \times Y \to \mathbb{R}$ a continuous cost function. Let us recall some results from Chapter 2:

- There exists to minimizes to (KP) in $\mathcal{P}(X \times Y)$.
- There exists maximizes to (DP) in $C(X) \times C(Y)$.
- It holds $T_c^{dual}(\mu, \nu) \leq T_c(\mu, \nu)$.
- We also recall the definition of c-transforms for $\varphi : X \to \mathbb{R}$ and $\psi : Y \to \mathbb{R}$:
  $$\varphi^c(y) = \inf_{x \in X} c(x, y) - \varphi(x) \quad \psi^{\bar{c}}(x) = \inf_{y \in Y} c(x, y) - \psi(y).$$
  It always holds $\varphi^{c\bar{c}} \geq \varphi$. If $\varphi(x) = \psi^{\bar{c}}(y)$ for some ψ, then φ is said c-concave and it holds $\varphi^{c\bar{c}} = \varphi$.

We are aiming to show strong duality.

Strong duality

- We start with the case of finite discrete probability measures.

Proposition 1 (Duality, discrete case)

If μ, ν are finitely supported, then $T_c^{dual}(\mu, \nu) = T_c(\mu, \nu)$.

Proof. Let us write $\mu = \sum_{i=1}^{m} \mu_i \delta_i$ and $\mu = \sum_{j=1}^{n} \nu_j \delta_{y_j}$ where all μ_i and ν_j are strictly positive. Consider the linear program

$$T_c^{lp}(\mu, \nu) := \min \left\{ \sum_{i,j} c(x_i, y_j)\gamma_{i,j} \,|\, \gamma_{i,j} \geq 0, \sum_{j} \gamma_{i,j} = \mu_i, \sum_{i} \gamma_{i,j} = \nu_j \right\}$$

which admits a solution that we denote γ. By linear programming duality, we have strong duality,

$$T_c^{lp}(\mu, \nu) = \max \left\{ \sum_{i} \varphi_i \mu_i + \sum_{j} \psi_j \nu_j \,|\, \varphi_i + \psi_j \leq c(x_i, y_j) \right\}$$

and at optimality $\gamma_{i,j}(c_{i,j} - \varphi_i - \psi_j) = 0$ (the complementary slackness in KKT condition). Let us now build a c-concave function φ such that $\varphi(x) \oplus \varphi^c(y) = c(x, y)$ on the set $\{(x_i, y_j) \,|\, \gamma_{i,j} > 0\}$.

Cont.

For this purpose, we introduce

$$\psi(y) = \begin{cases} \psi_i & \text{if } y = y_i \\ +\infty & \text{otherwise} \end{cases}$$

and let $\varphi = \psi^{\bar{c}}$. For $i_0 \in [n]$, there exists $j_0 \in [n]$ such that $\gamma_{i_0,j_0} > 0$ and thus by complementary slackness, $\varphi_{i_0} + \psi_{j_0} = c(x_{i_0}, y_{j_0})$ and thus

$$\varphi(x_{i_0}) = \int_{y \in Y} (c(x_{i_0}, y) - \psi(y)) = \min_{j \in [n]} (c(x_{i_0}, y_j) - \psi_j) = c(x_{i_0}, y_{j_0}) - \psi_{j_0} = \varphi_{i_0}$$

Similarly, one can show that $\varphi^c(y_j) = \psi_j$ for all $j \in [n]$. Finally, we define $\gamma = \sum_{i,j} \gamma_{i,j} \delta_{(x_i, y_j)} \in \Pi(\mu, \nu)$. We conclude with the following Lemma.

Cont.

For this purpose, we introduce

$$\psi(y) = \begin{cases} \psi_i & \text{if } y = y_i \\ +\infty & \text{otherwise} \end{cases}$$

and let $\varphi = \psi^{\bar{c}}$. For $i_0 \in [n]$, there exists $j_0 \in [n]$ such that $\gamma_{i_0,j_0} > 0$ and thus by complementary slackness, $\varphi_{i_0} + \psi_{j_0} = c(x_{i_0}, y_{j_0})$ and thus

$$\varphi(x_{i_0}) = \int_{y \in Y} (c(x_{i_0}, y) - \psi(y)) = \min_{j \in [n]} (c(x_{i_0}, y_j) - \psi_j) = c(x_{i_0}, y_{j_0}) - \psi_{j_0} = \varphi_{i_0}$$

Similarly, one can show that $\varphi^c(y_j) = \psi_j$ for all $j \in [n]$. Finally, we define $\gamma = \sum_{i,j} \gamma_{i,j} \delta_{(x_i, y_j)} \in \Pi(\mu, \nu)$. We conclude with the following Lemma.

Lemma 1 (Duality criterion)

Let $\gamma \in \Pi(\mu, \nu)$ and (φ, ψ) satisfying $\varphi(x) + \psi(y) \leq c(x, y)$. If $\varphi(x) + \psi(y) = c(x, y)$ for γ-almost every (x, y) then $T_c^{dual}(\mu, \nu) = T_c(\mu, \nu)$ and γ and (φ, ψ) are optimal and dual problem respectively.

Proof. Observe that

$$T_c(\mu, \nu) \leq \int c \,\mathrm{d}\gamma = \int (\varphi(x) + \psi(y)) \,\mathrm{d}\gamma(x, y) = \int \varphi \,\mathrm{d}\mu + \int \psi \,\mathrm{d}\nu \leq T_c^{dual}(\mu, \nu)$$

Density of discrete measures

- In order to prove the general case, we will use the density of discrete measures for the weak topology and a stability property of optimal dual and primal solutions.

Lemma 2 (Density of discrete measures)

Let X be a compact space and $\mu \in \mathcal{P}(X)$. Then, there exists a sequence of finitely supported probability measures weakly converging to μ.

Proof.

Density of discrete measures

- In order to prove the general case, we will use the density of discrete measures for the weak topology and a stability property of optimal dual and primal solutions.

Lemma 2 (Density of discrete measures)

Let X be a compact space and $\mu \in \mathcal{P}(X)$. Then, there exists a sequence of finitely supported probability measures weakly converging to μ.

Proof. By compactness, for any $\varepsilon > 0$, there exists N points $x_1, \dots, x_n$ such that $X \subset \bigcup_i B(x_i, \varepsilon)$. We introduce partition $K_1, \dots K_n$ of X defined recursively by $K_i = B(x_i, \varepsilon) \setminus K_1 \cup \dots \cup K_{i-1}$ and

$$\mu_\varepsilon := \sum_{i=1}^{n} \mu(K_i)\delta_{x_i}$$

To prove weak convergence of μ_ε to μ as $\varepsilon \to 0$, take $\varphi \in C(X)$. By compactness of X, φ admits a modulus of continuity w, $\varphi(x) - \varphi(y) \leq w(\mathrm{dist}(x, y))$. Using that $\mathrm{diam}(K_i) \leq \varepsilon$, we get

$$\left| \int \varphi \, \mathrm{d}\mu - \int \varphi \, \mathrm{d}\mu_\varepsilon \right| \leq \sum_{i=1}^{n} \int_{K_i} |\varphi(x) - \varphi(x_i)| \, \mathrm{d}\mu(x) \leq w(\varepsilon)$$

We deduce that μ_ε weakly converges to μ (recall that for measures on a compact space, weak and weak* topologies are the same).

Strong duality for the general case

Theorem 1 (Duality, general case)

Let X, Y be compact metric spaces and $c \in C(X \times Y)$. Then $T_c(\mu, \nu) = T_c^{dual}(\mu, \nu)$.

Proof.

Strong duality for the general case

Theorem 1 (Duality, general case)

Let X, Y be compact metric spaces and $c \in C(X \times Y)$. Then $T_c(\mu,\nu) = T_c^{dual}(\mu,\nu)$.

Proof. By Lemma 2, there exists a sequence $\mu_k \in \mathcal{P}(X)$ (resp. $\nu_k \in \mathcal{P}(Y)$) of finitely supported measures which converge weakly to μ (resp. ν). By prop 1, and its proof, there exists for all k, γ_k and (φ_k, φ_k^c) with φ_k c-concave which are optimal primal-dual solutions to $T_c(\mu_k, \nu_k)$ and such that γ_k is supported on the set

$$S_k := \{(x,y) \in X \times Y \,|\, \varphi_k(x) + \varphi_k^c(y) = c(x,y)\}$$

Strong duality for the general case

Theorem 1 (Duality, general case)

Let X, Y be compact metric spaces and $c \in C(X \times Y)$. Then $T_c(\mu,\nu) = T_c^{dual}(\mu,\nu)$.

Proof. By Lemma 2, there exists a sequence $\mu_k \in \mathcal{P}(X)$ (resp. $\nu_k \in \mathcal{P}(Y)$) of finitely supported measures which converge weakly to μ (resp. ν). By prop 1, and its proof, there exists for all k, γ_k and (φ_k, φ_k^c) with φ_k c-concave which are optimal primal-dual solutions to $T_c(\mu_k,\nu_k)$ and such that γ_k is supported on the set

$$S_k := \{(x,y) \in X \times Y \,|\, \varphi_k(x) + \varphi_k^c(y) = c(x,y)\}$$

Adding a constant if necessary, we can also assume that $\varphi_k(x_0) = 0$ for some point $x_0 \in X$. As in the previous lecture, we see that $\{\varphi_k\}_{k=1}^{\infty}, \{\varphi_k^c\}_{k=1}^{\infty}$ are uniformly continuous and bounded so by Ascoli-Arzela theorem, converge uniformly to some (φ, ψ) up to a subsequence. We easily have that φ is c-concave and $\psi = \varphi^c$.

By weak compactness of $\mathcal{P}(X \times Y)$, we can assume that the sequence γ_k weakly converges to $\gamma \in \Pi(\mu,\nu)$. Moreover, by Lemma 3, every pair $(x,y) \in \text{spt}(\gamma)$ can be approximated by a sequence of pairs $(x_k, y_k) \in \text{spt}(\gamma_k)$ with $\lim_{k\to\infty}(x,y)$. Since γ_k is supported on S_k one has $c(x_k,y_k) = \varphi_k(x_k) + \varphi_k^c(y_k)$, which gives at the limit $c(x,y) = \varphi(x) + \varphi^c(y)$. We conclude with Lemma 1.

Cont.

Lemma 3

If μ_n converges weakly to μ, then for any point $x \in \text{spt}(\mu)$ there exists a sequence $x_n \in \text{spt}(\mu_n)$ converging to x.

Proof. Consider $x \in \text{spt}\mu$. For any $k \in \mathbb{N}$, consider the function $\varphi_k(z) = \max\{0, 1 - k\text{dist}(x, z)\}$ which is continuous. Then

$$\lim_{n\to\infty} \int \varphi_k \, \mathrm{d}\mu_n = \int \varphi_k \, \mathrm{d}\mu > 0$$

Thus, there exists n_k such that for any $n \geq n_k$, $\int \varphi_k \, \mathrm{d}\mu_n > 0$. This implies the existence of a sequence $(x_n^{(k)}) \in X$ such that $x_n^{(k)} \in \text{spt}(\mu_n)$ and $\text{dist}(x_n^{(k)}, x) \leq 1/k$ for $n \geq n_k$. By a diagonal argument, we build the sequence $x_n = x_n^{k_n}$ where $k_n = \max\{k \mid k = 0 \text{ or } n \geq n_k\}$. By construction, $k_n \to \infty$, we have $x_n \to x$.

Optimality conditions and stability

The proof of Theorem 1 can be used to prove the following results.

Proposition 2 (Stability)

Let X, Y be compact metric spaces. Consider $(\mu_k)_{k\in\mathbb{N}}$ and $(\nu_k)_{k\in\mathbb{N}}$ converging weakly to μ and ν respectively and $(c_k)_{k\in\mathbb{N}}$ in $C(X \times Y)$ converging uniformly to c.

- If γ_k is a minimizer for $T_{c_k}(\mu_k, \nu_k)$ then, up to subsequences, (γ_k) converges weakly to a minimizer for $T_c(\mu, \nu)$.
- Let $(\varphi_k, \varphi_k^{c_k})$ be a maximize for $T_{c_k}^{dual}(\mu_k, \nu_k)$ and be such that φ_k is c_k-concave and $\varphi_k(x_0) = 0$. Then, up to subsequences, $(\varphi_k, \varphi_k^{c_k}$ converges uniformly to (φ, φ^c) a maximize for T_c^{dual} with φ c-concave satisfying $\varphi(x_0) = 0$.

Optimality conditions and stability

- Let us emphasize on the optimality conditions, which are just a continuous version of complementary slackness.

Proposition 3 (Optimality conditions)

For $\gamma \in \Pi(\mu, \nu)$ and $(\varphi, \psi) \in C(X) \times C(Y)$ satisfying $\varphi \oplus \psi \leq c$, the following are equivalent:

1. $\varphi(x) + \psi(y) = c(x, y)$ holds γ-almost everywhere.
2. γ is a minimizer of (KP), (φ, ψ) is a maximizer of (DP).

Optimality conditions and stability

- Another useful notion attached to optimal transport solutions is that of cyclical monotonicity.

Definition 1 (Cyclical monotonicity)

A set $S \subset X \times Y$ is said c-cyclically monotone if for any $n \in \mathbb{N}^*$ and $(x_i, y_i)_{i=1}^n \in S^n$, it holds

$$\sum_{i=1}^{n} c(x_i, y_i) \leq \sum_{i=1}^{n} c(x_i, y_{i+1})$$

with the convention $y_{n+1} = y_1$

Optimality conditions and stability

- Another useful notion attached to optimal transport solutions is that of cyclical monotonicity.

Definition 1 (Cyclical monotonicity)

A set $S \subset X \times Y$ is said c-cyclically monotone if for any $n \in \mathbb{N}^*$ and $(x_i, y_i)_{i=1}^n \in S^n$, it holds

$$\sum_{i=1}^{n} c(x_i, y_i) \leq \sum_{i=1}^{n} c(x_i, y_{i+1})$$

with the convention $y_{n+1} = y_1$

Proposition 4

Let X, Y be compact metric spaces, $c \in C(X \times Y)$ and $\gamma \in \Pi(\mu, \nu)$ an optimal transport plan between μ on $\mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. Then spt(γ) is c-cyclically monotone.

Remark. A stronger property holds: any c-cyclically monotonous set si contained in a set of the form $\{(x, y) \in X \times Y \,|\, \varphi(x) + \varphi^c(y) = c(x, y)\}$ for some c-concave function φ. This implies that and $\gamma \in \Pi(\mu, \nu)$ such that spt(γ) is c-cyclically monotone is optimal.

Applications

Optimal transport on the real line

Theorem 2 (Optimality of the monotone transport plan)

Let μ, ν be two probability measures on $\mathbb{R}$, and $c(x,y) := h(x-y)$ where h is strictly convex. Then, there exists a unique $\gamma \in \Gamma(\mu,\nu)$ satisfying the two following statements, which are equivalent:

1. γ is the optimal for the Kantorovich problem.
2. $\mathsf{spt}(\gamma)$ is monotone in the sense

$$\forall (x,y),(x',y') \in \mathsf{spt}(\gamma),\ (x'-x)\cdot(y'-y) \geq 0$$

Duality formula for the distance cost

- The dual problem takes a particularly simple form $c(x, y) = \text{dist}(x, y)$.

Proposition 5 (Kantorovich-Rubinstein)

Let $(X, dist)$ be a compact metric space and $\mu, \nu \in \mathcal{P}(X)$. Then

$$T_{dist}(\mu, \nu) = \max_{\varphi : X \to \mathbb{R}} \left\{ \int \varphi \, \mathrm{d}(\mu - \nu) \,|\, \varphi \text{ is 1-Lipschitz} \right\}$$

Proof. Note that $\psi^{\bar{c}}(x) = \inf_y \text{dist}(x, y) - \psi(y)$ is 1-Lipschitz as a infimum of 1-Lipschtiz functions, and the same holds for $\psi^{\bar{c}c}$. Moreover, if ψ is 1-Lipschitz, then $\text{dist}(x, y) - \psi(y) \geq -\psi(x)$, so that

$$\psi^{\bar{c}}(x) = \inf_y \text{dist}(x, y) - \psi(y) = -\psi(x)$$

Thus, $\varphi = -\psi$ and any 1-Lipschitz function is c-concave. Thus

$$T_{dist}(\mu, \nu) = \sup_{\psi : Y \to \mathbb{R}} \int \psi^{\bar{c}} \, \mathrm{d}\mu + \int \psi^{\bar{c}c} \, \mathrm{d}\nu = \sup_{\varphi:\text{ 1-Lip}} \int \varphi \, \mathrm{d}\mu + \int \varphi^c \, \mathrm{d}\nu = \sup_{\varphi:\text{ 1-Lip}} \int \varphi \, \mathrm{d}(\mu - \nu).$$

Optimal transport map for twisted costs

- Recall the following characterization of solutions to Monge's problem from Lecture 1.

Lemma 4

Let $\gamma \in \Pi(\mu,\nu)$ and $T : X \to Y$ measurable be such that $\gamma(\{(x,y) \in X \times Y \,|\, T(x) \neq y\}) = 0$. Then, $\gamma = \gamma_T := (\mathrm{id}, T)_{\#}\mu$.

To build a solution to Monge's problem, it is sufficient to show that the set $\{\varphi \oplus \varphi^c = c\}$ is contained in the graph of a function. This will be possible for the following class of costs:

Definition 2 (Twisted cost)

A cost function $c \in C^1(\mathbb{R}^d \times \mathbb{R}^d)$ is said to satisfy the *twist* condition if

$$\forall x_0 \in \mathbb{R}^d, \text{ the map } y \mapsto \nabla_x c(x_0, y) \in \mathbb{R}^d \text{ is injective}$$

Given $x, y \in \mathbb{R}^d$, we denote $y_c(x_0, v)$ the unique point such that $\nabla_x c(x_0, y_c(x_0, v)) = v$.

Optimal transport map for twisted costs

Theorem 3

Let $c \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ be twisted cost, let $X, Y \subset \mathbb{R}^d$ be compact subsets and $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. Assume that μ is absolutely continuous with respect to the Lebesgue measure. Then there exists a c-concave function φ that is differentiable almost everywhere such that $\nu = T_{\#}\mu$ where $T(x) = y_c(x, \nabla\varphi(x))$. Moreover, the only optimal transport plan between μ and ν is γ_T.

Proof.

Optimal transport map for twisted costs

Theorem 3

Let $c \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ be twisted cost, let $X, Y \subset \mathbb{R}^d$ be compact subsets and $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. Assume that μ is absolutely continuous with respect to the Lebesgue measure. Then there exists a c-concave function φ that is differentiable almost everywhere such that $\nu = T_{\#}\mu$ where $T(x) = y_c(x, \nabla\varphi(x))$. Moreover, the only optimal transport plan between μ and ν is γ_T.

Proof. Enlarging X if necessary, we may assume that spt(μ) is contained in the interior of X. First note that by compactness of $X \times Y$ and since c is C^1, the cost c is Lipschitz continuous on $X \times Y$. Take (φ, φ^c) a maximizing pair for (DP) with φ c-concave. Since $\varphi(x) = \min_{y \in Y} c(x, y) + \varphi^c(y)$ we see that φ is Lipschitz. By Rademacher theorem, φ is thus differentiable Lebesgue almost everywhere and, since μ is assumed absolutely continuous, it is differentiable on a set $B \subset$ spt(μ) with $\mu(B)$ $= 1$.

Cont.

Consider an optimal transport plan $\gamma \in \Pi(\mu, \nu)$. For every pair of points $(x_0, y_0) \in \mathsf{spt}(\gamma) \cap (B \times Y)$, we have

$$\varphi^c(y_0) \leq c(x, y_0) - \varphi(x), \ \forall x \in X$$

with equality at $x = x_0$, so that x_0 minimizes the function $x \mapsto c(x, y_0) - \varphi(x)$. Since $x_0 \in \mathsf{spt}(\mu)$, x_0 belongs to the interior of X, one necessarily has $\nabla\varphi(x_0) = \nabla_x c(x_0, y_0)$. Then, by the twist condition, one necessarily has $y_0 = y_c(x_0, \nabla\varphi(x_0))$. This shows that any optimal transport plan γ is supported on the graph of the map $T : x \in B \mapsto y_c(x_0, \nabla\varphi(x_0))$, and $\gamma = \gamma_T$ by Lemma 4.

Squared-norm cost an link with convexity

- When the cost is given by $c(x,y) := \frac{1}{2}\|y-x\|$ there is a connection between c-concavity and the usual notion of convexity.

Proposition 6

Given a function $\xi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, let us define $u_\xi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ by $u_\xi(x) = \frac{1}{2}\|x\|^2 - \xi(x)$. Then for $c(x,y) = \frac{1}{2}\|y-x\|_2^2$, we have $u_{\xi^c} = (u_\xi)^*$. In particular, a function ξ is c-concave iff $x \mapsto \frac{1}{2}\|x\|_2^2 - \xi(x)$ is convex and lower-semicontinuous.

Theorem 4

Let $c(x,y) = \frac{1}{2}\|y-x\|^2$ and $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ be compactly supported. If μ is absolutely continuous then exists a unique optimal transport plan between μ and ν which is of the form $(\mathrm{id} \times \nabla\tilde{\varphi})_{\#}\mu$ for some convex function $\tilde{\varphi} : \mathbb{R}^d \to \mathbb{R}$.

Cont.

Theorem 4 Proof. Consider two compact subsets $X, Y \subset \mathbb{R}^d$ that contain spt(μ) and spt(ν) in their respective interior. Then apply of Theorem 3. It holds $\nabla_x c(x_0, y) = x_0 - y$, which is injective for all x_0, thus $y_x(x_0, v) = x_0 - v$ and the optimal transport map is $T(x) = x - \nabla\varphi(x)$ for some c-concave φ. Finally, extend φ by $-\infty$ outside of X an define $\tilde{\varphi}(x) = \frac{1}{2}||x||^2 - \varphi(x)$ which is convex and l.s.c by Prop 6, with gradient $\nabla\tilde{\varphi}(x) = x - \nabla\varphi(x)$.