Lecture 2: Optimality Conditions and Consequences

Lénaïc Chizat

February 26, 2020

The material of today's lecture comes from [2, 3] and – for the most part of it – the lecture notes of Q. Mérigot.

Announcements. Register on the course webpage[1]. Bring your own laptops next week, with a running version of Python 3 and Jupyter notebooks.

1 Introduction

Let X and Y be compact metric spaces, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and $c : X \times Y \to \mathbb{R}$ a continuous cost function. In Lecture 1, we have defined the Kantorovich problem

$$\mathcal{T}_c(\mu,\nu) := \inf_{\gamma} \left\{ \int_{X\times Y} c(x,y)\mathrm{d}\gamma(x,y) \mid \gamma \in \Pi(\mu,\nu) \right\}. \tag{KP}$$

where $\Pi(\mu,\nu) := \{\gamma \in \mathcal{M}_+(X\times Y) \mid (\pi_X)_\# \gamma = \mu \text{ and } (\pi_Y)_\#\gamma = \nu\}$ is the set of *transport plans* between μ and ν. Rewriting the marginal constraints leads to the problem

$$\inf_{\gamma \geqslant 0} \sup_{\varphi,\psi} \left\{ \int_X \varphi(x)\mathrm{d}\mu(x) + \int_Y \psi(y)\mathrm{d}\nu(y) + \int_{X\times Y} \Big(c(x,y) - \varphi(x) - \psi(y)\Big) \mathrm{d}\gamma(x,y)\right\}.$$

After *formally* inverting the inf-sup, and minimizing over γ, we get the *dual* problem

$$\mathcal{T}_c^{dual}(\mu,\nu) := \sup_{\varphi,\psi} \left\{ \int_X \varphi \mathrm{d}\mu + \int_Y \psi \mathrm{d}\nu \mid \varphi(x) + \psi(y) \leqslant c(x,y),\ \forall (x,y) \in X\times Y \right\}. \tag{DP}$$

Let us recall some results from Lecture 1:

- There exists minimizers to (KP) in $\mathcal{P}(X\times Y)$.
- There exists maximizers to (DP) in $\mathcal{C}(X)\times\mathcal{C}(Y)$.
- It holds $\mathcal{T}_c^{dual}(\mu,\nu) \leqslant \mathcal{T}_c(\mu,\nu)$.
- We also recall the definition of c-transforms for $\varphi : X \to \mathbb{R}$ and $\psi : Y \to \mathbb{R}$:

  $$\varphi^c(y) = \inf_{x\in X} c(x,y) - \varphi(x) \qquad\qquad \psi^{\bar c}(x) = \inf_{y\in Y} c(x,y) - \psi(y).$$

  It always holds $\varphi^{c\bar c} \geqslant \varphi$. If $\varphi(x) = \psi^{\bar c}(y)$ for some ψ, then φ is said *c-concave* and it holds $\varphi^{c\bar c} = \varphi$ (exercise, or see [2, Prop. 1.3.4]).

Today, we will show *strong duality*, derive primal-dual optimality conditions and explore their consequences. We assume that X and Y are compact for the sake of simplicity, but most statement have their counterpart in non-compact spaces.

[1] http://lchizat.github.io/ot2020orsay.html

2 Strong duality

2.1 The case of discrete optimal transport

We start with the case of finite discrete probability measures, which is important because:

- It often comes up in applications (e.g. optimal matching in economy).
- Numerical methods for the continuous case often resort to discretization.
- It is a convenient way to study the general case, through density arguments.

Proposition 2.1 (Duality, discrete case)**.** *If μ and ν are finitely supported, then $\mathcal{T}_c^{dual}(\mu,\nu) = \mathcal{T}_c(\mu,\nu)$.*

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

$$\mathcal{T}_c^{lp}(\mu,\nu) := \min\Big\{ \sum_{i,j} c(x_i,y_j)\gamma_{i,j} \mid \gamma_{i,j} \geqslant 0,\ \sum_j \gamma_{i,j} = \mu_i,\ \sum_i \gamma_{i,j} = \nu_j \Big\}.$$

which admits a solution that we denote γ. By linear programming duality (which is standard in the finite dimensional case, see e.g. [1]), we have strong duality

$$\mathcal{T}_c^{lp}(\mu,\nu) = \max\Big\{ \sum_i \varphi_i\mu_i + \sum_j \psi_j\nu_j \mid \varphi_i + \psi_j \leqslant c(x_i,y_j) \Big\}$$

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

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

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 Lemma 2.2. □

Lemma 2.2 (Duality criterion)**.** *Let $\gamma \in \Pi(\mu,\nu)$ and (φ,ψ) satisfying $\varphi(x) + \psi(y) \leqslant c(x,y)$. It $\varphi(x) + \psi(y) = c(x,y)$ for γ-almost every (x,y) then $\mathcal{T}_c^{dual}(\mu,\nu) = \mathcal{T}_c(\mu,\nu)$ and γ and (φ,ψ) are optimal for the primal and dual problem respectively.*

Proof. Observe that

$$\mathcal{T}_c(\mu,\nu) \leqslant \int c\mathrm{d}\gamma = \int \Big(\varphi(x) + \psi(y)\Big)\mathrm{d}\gamma(x,y) = \int \varphi \mathrm{d}\mu + \int \psi \mathrm{d}\nu \leqslant \mathcal{T}_c^{dual}(\mu,\nu)$$

Since we know that $\mathcal{T}_c^{dual}(\mu,\nu) \leqslant \mathcal{T}_c(\mu,\nu)$ this is sufficient to conclude. □

2.2 Density of discrete measures

In order the 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 2.3 (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 $\epsilon > 0$, there exists N points $x_1, \dots, x_n$ such that $X \subset \bigcup_i B(x_i, \epsilon)$. We introduce the partition $K_1, \dots, K_n$ of X defined recursively by $K_i = B(x_i, \epsilon) \setminus K_1 \cup \dots K_{i-1}$ and

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

To prove weak convergence of μ_ϵ to μ as $\epsilon \to 0$, take $\varphi \in \mathcal{C}(X)$. By compactness of X, φ admits a modulus of continuity ω, i.e. an increasing function satisfying $\lim_{t\to 0} \omega(t) = 0$ and $|\varphi(x) - \varphi(y)| \leqslant w(\mathrm{dist}(x, y))$. Using that $\mathrm{diam}(K_i) \leqslant \epsilon$, we get

$$\left| \int \varphi \mathrm{d}\mu - \int \varphi \mathrm{d}\mu_\epsilon \right| \leqslant \sum_{i=1}^{n} \int_{K_i} |\varphi(x) - \varphi(x_i)| \mathrm{d}\mu(x) \leqslant \omega(\epsilon).$$

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

Note that we even have weak density in $\mathcal{P}(X)$ of empirical measures, that is measures of the form $\frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$ for $n \in \mathbb{N}^*$ and $x_i \in X$. Indeed, take $x_1, \dots, x_n$ independent random variables with distribution μ. Then, by the uniform law of large numbers (a.k.a. Varadarajan's theorem) states that $\frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$ weakly converges to μ with probability 1.

2.3 Strong duality for the general case

Theorem 2.4 (Duality, general case)**.** *Let X, Y be compact metric spaces and $c \in \mathcal{C}(X \times Y)$. Then $\mathcal{T}_c(\mu, \nu) = \mathcal{T}_c^{dual}(\mu, \nu)$.*

Proof. By Lemma 2.3, 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 Proposition 2.1 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 $\mathcal{T}_c(\mu_k, \nu_k)$ and such that γ_k is supported on the set

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

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 $\{\varphi_k\}$ and $\{\varphi_k^c\}$ are uniformly continuous and bounded so by Ascoli-Arzelà 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 2.5, every pair $(x, y) \in \mathrm{spt}(\gamma)$ can be approximated by a sequence of pairs $(x_k, y_k) \in \mathrm{spt}(\gamma_k)$ with $\lim_{k\to\infty}(x_k, y_k) = (x, y)$. 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 2.2. □

Lemma 2.5. *If μ_n converges weakly to μ, then for any point $x \in \operatorname{spt}(\mu)$ there exists a sequence $x_n \in \operatorname{spt}(\mu_n)$ converging to x.*

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

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

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

3 Optimality conditions and stability

Let us write down three important properties that follow from our previous results. First, remark that the proof of Theorem 2.4 can be used to prove the following stability property (the modifications are left as an exercise).

Proposition 3.1 (Stability)**.** *Let X, Y be compact metric spaces. Consider $(\mu_k)_{k\in\mathbb{N}}$ and $(\nu_k)_{k\in\mathbb{N}}$ in $\mathcal{P}(X)$ and $\mathcal{P}(Y)$ converging weakly to μ and ν respectively and $(c_k)_{k\in\mathbb{N}}$ in $\mathcal{C}(X \times Y)$ converging uniformly to c.*

- *If γ_k is a minimizer for $\mathcal{T}_{c_k}(\mu_k, \nu_k)$ then, up to subsequences, (γ_k) converges weakly to a minimizer for $\mathcal{T}_c(\mu, \nu)$.*
- *Let $(\varphi_k, \varphi_k^{c_k})$ be a maximizer for $\mathcal{T}_{c_k}^{dual}(\mu_k, \nu_k)$ 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 maximizer for $\mathcal{T}_c^{dual}(\mu, \nu)$ with φ c-concave satisfying $\varphi(x_0) = 0$.*

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

Proposition 3.2 (Optimality conditions)**.** *For $\gamma \in \Pi(\mu, \nu)$ and $(\varphi, \psi) \in \mathcal{C}(X) \times \mathcal{C}(Y)$ satisfying $\varphi \oplus \psi \leqslant c$, the following are equivalent:*

(i) $\varphi(x) + \psi(y) = c(x, y)$ *holds γ-almost everywhere.*

(ii) γ *is a minimizer of* (KP), (φ, ψ) *is a maximizer of* (DP).

Proof. The proof of $(i) \Rightarrow (ii)$, is given by Lemma 2.2. To show $(ii) \Rightarrow (i)$, notice that Theorem 2.4 and (ii) imply

$$0 = \int c(x, y)\mathrm{d}\gamma(x, y) - \int \varphi(x) + \psi(y)\mathrm{d}\gamma(x, y) = \int \Big(c(x, y) - \varphi(x) - \psi(y)\Big)\mathrm{d}\gamma(x, y).$$

Since the last integrand is nonnegative, it must vanish γ-almost everywhere. □

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

Definition 3.3 (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

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

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

Proposition 3.4. *Let X, Y be compact metric spaces, $c \in \mathcal{C}(X \times Y)$ and $\gamma \in \Pi(\mu, \nu)$ an optimal transport plan between $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. Then* $\mathrm{spt}(\gamma)$ *is c-cyclically monotone.*

This result is rather direct in the discrete case and can also be proved without duality in the general case but our duality results lead to a straighforward proof.

Proof. Let $(x_i, y_i)_{i=1}^n$ be n points in $\mathrm{spt}(\gamma)$. By Proposition 3.2, we know that there exists (φ, ψ) such that $\varphi(x_i) + \psi(y_j) \leqslant c(x_i, y_j)$ for all i, j and such that $\varphi(x_i) + \psi(y_i) = c(x_i, y_i)$ for all i. Thus

$$\sum c(x_i, y_{i+1}) - \sum c(x_i, y_i) \geqslant \sum_i \Big(\varphi(x_i) + \psi(y_{i+1})\Big) - \sum_i \Big(\varphi(x_i) + \psi(y_i)\Big) = 0.$$

□

Remark 3.5. A stronger property holds: any c-cyclically monotonous set is 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 any $\gamma \in \Pi(\mu, \nu)$ such that $\mathrm{spt}(\gamma)$ is c-cyclically monotone is optimal.

4 Applications

Let us exploit the optimality conditions and duality results to describe the behavior of optimal transport in specific situations.

4.1 Optimal transport on the real line

Theorem 4.1 (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*

(i) γ is optimal for the Kantorovich problem;

(ii) $\mathrm{spt}(\gamma)$ *is* monotone *in the sense*

$$\forall (x, y), (x', y') \in \mathrm{spt}(\gamma), (x' - x) \cdot (y' - y) \geqslant 0.$$

Proof. We first prove that there exists at most one transport plan satisfying (ii). Recall that a probability measure on $\mathbb{R}^2$ is uniquely defined from the values $\gamma((-\infty, a] \times (-\infty, b])$ for any $a, b \in \mathbb{R}$. This follows from the fact that such sets generate the Borel σ-algebra. Consider $A = (-\infty, a] \times (b, +\infty)$ and $B = (a, +\infty) \times (-\infty, b]$. Then, by monotonicity of $\mathrm{spt}(\gamma)$ one cannot have $\gamma(A) > 0$ and $\gamma(B) > 0$ at the same time. Hence,

$$\begin{aligned}\gamma(] - \infty, a] \times] - \infty, b]) &= \min(\gamma((] - \infty, a] \times] - \infty, b]) \cup A), \gamma((] - \infty, a] \times] - \infty, b]) \cup B)) \\ &= \min(\mu(] - \infty, a]), \nu(] - \infty, b])).\end{aligned}$$

This shows that $\gamma(] - \infty, a] \times (] - \infty, b])$ is uniquely defined from μ, ν, so that γ is unique.

Now by Proposition 3.4, we know that for an optimal transport plan γ and $(x_i, y_i)_{i=1}^2 \in \mathrm{spt}(\gamma)^2$, it holds

$$c(x_0, y_0) + c(x_1, y_1) \leqslant c(x_0, y_1) + c(x_1, y_0).$$

We conclude with $c(x, y) = |x - y|^2$, the case of a general strictly convex function can be found in Chapter 2 of [2]. Expanding the squares and simplifying, the above inequality can be rewritten as

$$-x_0 y_0 - x_1 y_1 \leqslant -x_0 y_1 - x_1 y_0,$$

giving exactly $(x_0 - x_1)(y_0 - y_1) \geqslant 0$ as desired. □

While in this proof cyclical monotonicity of order 2 was enough to conclude, we warn the reader that this is in general not the case in higher dimension.

Remark 4.2 (Book-shifting). If $c(x, y) = |x - y|$ with the Euclidean norm, the solution to the optimal transport problem might be non-unique. Take for instance $\mu = \lambda|_{[0,1]}$ and $\nu = \lambda|_{[\varepsilon,1+\varepsilon]}$ for some $\varepsilon > 0$. Then, the maps $T : x \mapsto x + \varepsilon$ and $T'(x) = x$ if $x \in [\varepsilon, 1]$ and $T'(x) = x + 1$ if $x \in [0, \varepsilon]$ are both optimal with the same cost. (NB: proving the optimality of a transport map is in general a difficult matter, to which Kantorovich duality provides an answer.)

It turns out that the unique monotone transport map can be built using *quantile* functions. Given $\mu \in \mathcal{P}(\mathbb{R})$, define its cumulative distribution function $F_\mu : \mathbb{R} \to [0, 1]$ and its quantile function $Q_\mu : [0, 1] \to \mathbb{R}$ by:

$$F_\mu(x) = \mu((-\infty, x]) \quad \text{and} \quad Q_\mu(t) = \inf\{x \in \mathbb{R} \mid F_\mu(x) \geqslant t\}.$$

As a simple consequence of these definitions, we have

$$Q_\mu(t) \leqslant x \Leftrightarrow F_\mu(x) \geqslant t \quad \text{and} \quad Q_\mu(t) > x \Leftrightarrow F_\mu(x) < t. \tag{4.1}$$

Proposition 4.3 (Characterization of the monotone transport plan). *The unique monotone transport plan in $\Pi(\mu, \nu)$ is given by $\gamma_Q = (Q_\mu, Q_\nu)_\# \lambda$. In particular, for $c(x, y) = h(y - x)$ with h strictly convex, we have the following explicit optimal transport cost*

$$\mathcal{T}_c(\mu, \nu) = \int_0^1 h(Q_\nu(t) - Q_\mu(t))\mathrm{d}t$$

Proof. First, let us prove that Q_μ is a transport map between the Lebesgue measure on $[0, 1]$ (denoted λ) and μ. Using Eq. (4.1), we write

$$(Q_\mu)_\# \lambda|_{[0,1]}(] - \infty, a]) = \lambda\Big(\{t \in [0, 1] \mid Q_\mu(t) \leqslant a\}\Big) = \lambda\Big(\{t \in [0, 1] \mid F_\mu(a) \geqslant t\}\Big) = F_\mu(a),$$

which proves that $(Q_\mu)_\# \lambda|_{[0,1]} = \mu$ using the characterization of a measure through its CFD. It directly follows that $\gamma_Q \in \Pi(\mu, \nu)$. Then, let us compute

$$\begin{aligned}\gamma_Q(] - \infty, a] \times] - \infty, b]) &= \lambda\Big(\{t \in [0, 1] \mid Q_\mu(t) \leqslant x,\ Q_\nu(t) \leqslant x\}\Big)\\ &= \lambda\Big(\{t \in [0, 1] \mid F_\mu(x) \geqslant t,\ F_\nu(x) \geqslant t\}\Big)\\ &= \min\{F_\mu(x), F_\nu(x)\}\end{aligned}$$

and we recover the characterization of the monotone transport plan in the proof of Theorem 4.1. □

4.2 Duality formula for the distance cost

The dual problem takes a particularly simple form when the cost is of the form $c(x, y) = \mathrm{dist}(x, y)$.

Proposition 4.4 (Kantorovich-Rubinstein). *Let (X, dist) be a compact metric space and $\mu, \nu \in \mathcal{P}(X)$. Then*

$$\mathcal{T}_{\mathrm{dist}}(\mu, \nu) = \max_{\varphi : X \to \mathbb{R}} \Big\{ \int \varphi \,\mathrm{d}(\mu - \nu) \mid \varphi \text{ is 1-Lipschitz}\Big\}.$$

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

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

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

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

□

4.3 Optimal transport map for twisted costs

We recall the following characterization of solutions to Monge's problem from Lecture 1.

Lemma 4.5. *Let $\gamma \in \Pi(\mu,\nu)$ and $T : X \to Y$ measurable be such that $\gamma(\{(x,y) \in X \times Y \mid T(x) \neq y\}) = 0$. Then, $\gamma = \gamma_T := (\mathrm{id}, T)_\# \mu$.*

If γ is a minimizer for (KP) and (φ, φ^c) is a maximizer for (DP), we know that $\varphi \oplus \varphi^c = c$ γ-almost everywhere. To build a solution to Monge's problem, it is therefore 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 4.6 (Twisted cost)**.** A cost function $c \in \mathcal{C}^1(\mathbb{R}^d \times \mathbb{R}^d)$ is said to satisfy the *twist condition* if

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

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

Theorem 4.7. *Let $c \in \mathrm{C}^1(\mathbb{R}^d \times \mathbb{R}^d)$ be a 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 $\operatorname{spt}(\mu)$ is contained in the interior of X. First note that by compactness of $X \times Y$ and since c is $\mathcal{C}^1$, 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 $B \subset \operatorname{spt}(\mu)$ with $\mu(B) = 1$.

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

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

with equality at $x = x_0$, so that x_0 minimizes the function $x \mapsto c(x, y_0) - \varphi(x)$. Since $x_0 \in \operatorname{spt}(\mu)$ and 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 4.5. □

4.4 Square-norm cost and link with convexity

When the cost is given by $c(x,y) := \frac{1}{2}\|y-x\|_2^2$ there is a connection between c-concavity and the usual notion of convexity.

Proposition 4.8. *Given a function $\xi : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$, let us define $u_\xi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ through $u_\xi(x) = \frac{1}{2}\|x\|_2^2 - \xi(x)$. 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.*

Proof. Observe that

$$u_{\xi^c}(x) = \frac{1}{2}\|x\|_2^2 - \xi^c(x) = \sup_y \frac{1}{2}\|x\|_2^2 - \frac{1}{2}\|x-y\|_2^2 + \xi(x) = \sup_y \langle x, y\rangle - \Big(\frac{1}{2}\|y\|_2^2 - \xi(y)\Big).$$

This proves the first part of the statement. The second part follows from the fact that convex l.s.c. functions are characterized by the fact that they are sup of affine functions. □

Theorem 4.9. *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 there 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 $\tilde{\varphi} : \mathbb{R}^d \to \mathbb{R}$.*

Proof. Consider two compact subsets $X, Y \subset \mathbb{R}^d$ that contain $\mathrm{spt}(\mu)$ and $\mathrm{spt}(\nu)$) in their respective interior. Then apply of Theorem 4.7. 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 and define $\tilde{\varphi}(x) = \frac{1}{2}\|x\|^2 - \varphi(x)$ which is convex and l.s.c. by Proposition 4.8, with gradient $\nabla\tilde{\varphi}(x) = x - \nabla\varphi(x)$. □

References

[1] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe, *Convex optimization*, Cambridge university press, 2004.

[2] Filippo Santambrogio, *Optimal transport for applied mathematicians*, Springer, 2015.

[3] Cédric Villani, *Optimal transport: old and new*, vol. 338, Springer Science & Business Media, 2008.