LECTURE 1
MONGE AND KANTOROVICH PROBLEMS: FROM PRIMAL TO DUAL

L. CHIZAT AND L. NENNA

These notes are based on the ones by Quentin Mérigot

Some motivations for studying optimal transport.

- Variational principles for (real) Monge-Ampère equations occuring in geometry (e.g. Gaussian curvature prescription) or optics.
- Wasserstein/Monge-Kantorovich distance between probability measures μ, ν on e.g. $\mathbb{R}^d$: how much kinetic energy does one require to move a distribution of mass described by μ to ν ?
  $\longrightarrow$ interpretation of some parabolic PDEs as Wasserstein gradient flows, construction of (weak) solutions, numerics, e.g.

$$\begin{cases} \partial_t \rho + \operatorname{div}(\rho v) = 0 \\ v = -\nabla \log \rho \end{cases} \quad \text{or} \quad \begin{cases} \partial_t \rho + \operatorname{div}(\rho v) = 0 \\ v = -\nabla p - \nabla V \\ p(1-\rho) = 0 \\ p \geqslant 0, \rho \leqslant 1 \end{cases}$$

  $\longrightarrow$ interesting geometry on $\mathcal{P}(X)$, with an embedding $X \hookrightarrow \mathcal{P}(X)$. Applications in geometry (synthetic notion of Ricci curvature for metric spaces), machine learning, inverse problems, etc.
- Quantum physics: electronic configuration in molecules and atoms.

References. Introduction to optimal transport, with applications to PDE and/or calculus of variations can be found in books by Villani [7] and Santambrogio [6]. Villani's second book [8] concentrates on the application of optimal transport to geometric questions (e.g. synthetic definition of Ricci curvature), but its first chapters might be useful. We also mention Gigli, Ambrosio and Savaré [2] for the study of gradient flows with respect to the Monge-Kantorovich/Wasserstein metric.

Notation. In the following, we assume that X is a *complete and separable metric space*. We denote $\mathcal{C}(X)$ the space of continuous functions, $\mathcal{C}_0(X)$ the space of continuous function vanishing at infinity $\mathcal{C}_b(X)$ the space of bounded continuous functions. We denote $\mathcal{M}(X)$ the space of Borel regular measures on X with finite total mass and

$$\mathcal{M}^+(X) := \{\mu \in \mathcal{M}(X) \mid \mu \geqslant 0\}$$

$$\mathcal{P}(X) := \{\mu \in \mathcal{M}^+(X) \mid \mu(X) = 1\}$$

Some reminders.

Definition 0.1 (Lower semi-continuous function)**.** On a metric space Ω, a function $f : \Omega \to \mathbb{R} \cup \{+\infty\}$ is said to be lower semi-continuous (l.s.c.) is for every sequence $x_n \to x$ we have $f(x) \leqslant \liminf_n f(x_n)$.

Definition 0.2. A metric space Ω is said to be compact if from any sequence x_n, we can extract a converging subsequence $x_{n_k} \to x \in \Omega$.

Theorem 0.3 (Weierstrass)**.** *If $f : \Omega \to \mathbb{R} \cup \{+\infty\}$ is l.s.c. and Ω is compact, then there exists $x^\star \in \Omega$ such that $f(x^\star) = \min\{f(x) \mid x \in \Omega\}$.*

Definition 0.4 (weak and weak$-\star$ convergence)**.** A sequence x_n in a Banach space $\mathcal{X}$ is said to be weakly converging to x and we write $x_n \rightharpoonup x$, if for every $\eta \in \mathcal{X}'$ ($\mathcal{X}'$ is the topological dual of $\mathcal{X}$ and $\langle \cdot, \cdot \rangle$ is the duality product) we have $\langle \eta, x_n \rangle \to \langle \eta, x \rangle$. A sequence $\eta_n \in \mathcal{X}'$ is said to be weakly-$\star$ converging to $\eta \in \mathcal{X}'$, and we write $\eta_n \stackrel{\star}{\rightharpoonup} \eta$, if for every $x \in \mathcal{X}$ we have $\langle \eta_n, x \rangle \to \langle \eta, x \rangle$.

Theorem 0.5 (Banach-Alaoglu)**.** *If $\mathcal{X}'$ is separable and φ_n is a bounded sequence in $\mathcal{X}'$, then there exists a subsequence φ_{n_k} weakly converging to some $\varphi \in \mathcal{X}'$*

Theorem 0.6 (Riesz)**.** *Let X be a compact metric space and $\mathcal{X} = \mathcal{C}(X)$ then every element of $\mathcal{X}$ is represented in a unique way as an element of $\mathcal{M}^+(X)$, that is for every $\eta \in \mathcal{X}$ there exists a unique $\lambda \in \mathcal{M}^+(X)$ such that $\langle \eta, \varphi \rangle = \int_X \varphi \mathrm{d}\lambda$ for every $\varphi \in \mathcal{X}$.*

Definition 0.7 (Narrow convergence)**.** A sequence of finite measures $(\mu_n)_{n \geqslant 1}$ on X narrowly converges to $\mu \in \mathcal{M}(X)$ if

$$\forall \varphi \in \mathcal{C}_b(X), \quad \lim_{n \to \infty} \int_X \varphi \mathrm{d}\mu_n = \int_X \varphi \mathrm{d}\mu.$$

With a slightly abuse of notation we will denote it by $\mu_n \rightharpoonup \mu$.

Remark 0.8. Since we will mostly work on compact set X, then $\mathcal{C}(X) = \mathcal{C}_0(X) = \mathcal{C}_b(X)$. This means that the narrow convergence of measures, that is the notion of convergence in duality with $\mathcal{C}_b(X)$, corresponds to the weak-$\star$ convergence (the convergence in duality with $\mathcal{C}_0(X)$).

1. The problems of Monge and Kantorovich

1.1. Monge problem.

Definition 1.1 (Push-forward and transport map)**.** Let X, Y be metric spaces, $\mu \in \mathcal{M}(X)$ and $T : X \to Y$ be a measurable map. The *push-forward* of μ by T is the measure $T_\# \mu$ on Y defined by

$$\forall B \subseteq Y, \; \mathrm{T}_\# \mu(B) = \mu(T^{-1}(B)).$$

or equivalently if the following change-of-variable formula holds for all measurable and bounded $\varphi : Y \to \mathbb{R}$:

$$\int_Y \varphi(y) \mathrm{d}T_\# \mu(y) = \int_X \varphi(T(x)) \mathrm{d}\mu(x).$$

A measurable map $T : X \to Y$ such that $T_\# \mu = \nu$ is also called a *transport map* between μ and ν.

Example 1.2. If $Y = \{y_1, \ldots, y_n\}$, then $T_\# \mu = \sum_{1 \leqslant i \leqslant n} \mu(T^{-1}(\{y_i\}))\delta_{y_i}$.

Example 1.3. Assume that T is a $\mathcal{C}^1$ diffeomorphism between open sets X, Y of $\mathbb{R}^d$, and assume also that the probability measures μ, ν have continuous densities ρ, σ with respect to the Lebesgue measure. Then,

$$\int_Y \varphi(y)\sigma(y)\mathrm{d}y = \int_X \varphi(T(x))\sigma(T(x)) \det(\mathrm{D}T(x))\mathrm{d}x.$$

Hence, T is a transport map between μ and ν iff

$$\forall \varphi \in \mathcal{C}_b(X), \int_X \varphi(T(x))\sigma(T(x)) \det(\mathrm{D}T(x))\mathrm{d}x = \int_X \varphi(T(x))\rho(x)\mathrm{d}x$$

Hence, T is a transport map iff the non-linear Jacobian equation holds

$$\rho(x) = \sigma(T(x)) \det(\mathrm{D}T(x)).$$

Definition 1.4 (Monge problem)**.** Consider two metric spaces X, Y, two probability measures $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and a *cost function* $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$. *Monge's problem* is the following optimization problem

$$(\mathrm{MP}) := \inf \left\{ \int_X c(x, T(x))\mathrm{d}\mu(x) \mid T : X \to Y \text{ and } T_\#\mu = \nu \right\} \tag{1.1}$$

This problem exhibits several difficulties, one of which is that both the constraint ($T_\#\mu = \nu$) and the functional are non-convex.

Example 1.5. There might exist no transport map between μ and ν. For instance, consider $\mu = \delta_x$ for some $x \in X$. Then, $\mathrm{T}_\#\mu(B) = \mu(T^{-1}(B)) = \delta_{T(x)}$. In particular, if $\mathrm{card}(\mathrm{spt}(\nu)) > 1$ (see Def. 1.15), there exists no transport map between μ and ν.

Example 1.6. The infimum might not be attained even if μ is atomless (i.e. for every point $x \in X$, $\mu(\{x\}) = 0$). Consider for instance $\mu = \frac{1}{2}\, \lambda|_{\{\pm 1\} \times [-1,1]}$ on $\mathbb{R}^2$ and $\nu = \lambda|_{\{0\} \times [-1,1]}$, where λ is the Lebesgue measure. One solution is to allow mass to split, leading to Kantorovich's relaxation of Monge's problem.

1.2. **Kantorovich problem.**

Definition 1.7 (Marginals)**.** The *marginals* of a measure γ on a product space $X \times Y$ are the measures $\pi_{X\#}\gamma$ and $\pi_{Y\#}\gamma$, where $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ are their projection maps.

Definition 1.8 (Transport plan)**.** A transport plan between two probabily measures μ, ν on two metric spaces X and Y is a probability measure γ on the product space $X \times Y$ whose marginals are μ and ν. The space of transport plans is denoted $\Pi(\mu, \nu)$, i.e.

$$\Pi(\mu, \nu) = \{\gamma \in \mathcal{P}(X \times Y) \mid \pi_{X\#}\gamma = \mu,\ \pi_{Y\#}\gamma = \nu\}.$$

Note that $\Pi(\mu, \nu)$ is a convex set.

Example 1.9 (Tensor product)**.** Note that the set of transport plans $\Pi(\mu, \nu)$ is never empty, as it contains the measure $\mu \otimes \nu$.

Example 1.10 (Transport plan associated to a map)**.** Let T be a transport map between μ and ν, and define $\gamma_T = (id, T)_\#\mu$. Then, γ_T is a transport plan between μ and ν.

Definition 1.11 (Kantorovich problem)**.** Consider two metric spaces X, Y, two probability measures $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and a *cost function* $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$. *Kantorovich's problem* is the following optimization problem

$$(\mathrm{KP}) := \inf \left\{ \int_{X\times Y} c(x,y)\mathrm{d}\gamma(x,y) \mid \gamma \in \Pi(\mu,\nu) \right\} \tag{1.2}$$

Remark 1.12. The infimum in Kantorovich problem is less than the infimum in Monge problem. Indeed, consider a transport map satisfying $T_{\#}\mu = \nu$ and the associated transport plan γ_T. Then, by the change of variable one has

$$\int_{X\times Y} c(x,y)\mathrm{d}(id,T)_{\#}\mu(x,y) = \int_X c(x,T(x))\mathrm{d}\mu,$$

thus proving the claim.

Example 1.13 (Finite support)**.** Assume that $X = Y = \{1,\dots,N\}$ and that μ,ν are the uniform probability measures over X and Y. Then, Monge's problem can be rewritten as a minimization problem over bijections between X and Y:

$$\min\{\frac{1}{N}\sum_{1\leqslant i\leqslant N} c(i,\sigma(i)) \mid \sigma \in \mathfrak{S}_N\}.$$

In Kantorovich's relaxation, the set of transport plans $\Pi(\mu,\nu)$ agrees with the set of bi-stochastic matrices :

$$\gamma \in \Pi(\mu,\nu) \iff \gamma \geqslant 0, \sum_i \gamma(i,j) = 1/N = \sum_j \gamma(i,j).$$

By Birkhoff's theorem, any extremal bi-stochastic matrix is induced by a permutation. This shows that, in this case, the solution to Monge's and Kantorovich's problems agree.

Remark 1.14. Proposition 1.16 shows that a transport plan concentrated on the graph of a function $T : X \to Y$ is actually induced by a transport map. One can prove that transport plans concentrated on graphs are extremal points in the convex set $\Pi(\mu,\nu)$, but the converse does not hold in general (the counter-examples are quite tricky to construct, see [1]). This means that one cannot resort to a simple argument such as Birkhoff's theorem to show that solutions to Kantorovich's problem (transport plans) are induced by transport maps.

Definition 1.15 (Support)**.** Let Ω be a separable metric space. The *support* of a non-negative measure μ is the smallest closed set on which μ is concentrated

$$\mathrm{spt}(\mu) := \bigcap\{A \subseteq \Omega \mid A \text{ closed and } \mu(X\setminus A) = 0\}.$$

A point x belongs to $\mathrm{spt}(\mu)$ iff for every $r > 0$ one has $\mu(\mathrm{B}(x,r)) > 0$.

Proposition 1.16. *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, $\gamma = \gamma_T$.*

Proof. By definition of γ_T one has $\gamma_T(A\times B)=\mu(T^{-1}(B)\cap A)$ for all Borel sets $A\subseteq X$ and $B\subseteq Y$. On the other hand,

$$\begin{aligned}\gamma(A\times B) &= \gamma(\{(x,y)\mid x\in A,\ \text{and } y\in B\})\\ &= \gamma(\{(x,y)\mid x\in A, y\in B \text{ and } y=T(x)\})\\ &= \gamma(\{(x,y)\mid x\in A\cap T^{-1}(B), y=T(x)\}\\ &= \mu(A\cap T^{-1}(B)),\end{aligned}$$

thus proving the claim. □

2. Existence of solutions to Kantorovich's problem

The proof of existence relies on the direct method in the calculus of variations, i.e. the fact that the minimized functional is lower semi-continuous and the set over which it is minimized is compact.

Theorem 2.1. *Let X,Y be two compact spaces, and $c:X\times Y\to\mathbb{R}\cup\{+\infty\}$ be a lower semi-continuous cost function, which is bounded from below. Then Kantorovich's problem admits a minimizer.*

Lemma 2.2. *Let $f:X\to\mathbb{R}\cup\{+\infty\}$ be a lower semi-continuous function, which is also bounded from below. Define $\mathcal{F}:\mathcal{P}(X)\to\mathbb{R}\cup\{+\infty\}$ through $\mathcal{F}(\mu)=\int_X f\mathrm{d}\mu$. Then, $\mathcal{F}$ is lower-semicontinuous for the narrow convergence, i.e.*

$$\forall\mu_n\rightharpoonup\mu,\ \liminf_{n\to\infty}\mathcal{F}(\mu_n)\geqslant\mathcal{F}(\mu).$$

Proof. **Step 1.** We show that there exists a family of bounded and continuous functions f^k such that $k\mapsto f^k$ is pointwise increasing and $f=\sup_k f^k$. We assume that there exists x_0 such that $f(x_0)<+\infty$ (if not, there is nothing to prove). Define $g^k(x)=\inf_{y\in X} f(y)+k\mathrm{d}(x,y)\leqslant f(x_0)+k\mathrm{d}(x,x_0)$. The function g^k is k-Lipschitz as a minimum of k-Lipschitz functions, and one obviously has $g^k\leqslant g^\ell\leqslant f$ for $k\leqslant\ell$. Let us prove that $\sup_k g^k(x)=f(x)$ for any $x\in X$. Given x, and for every k, there exists a point x_k such that

$$f(x_k)+k\mathrm{d}(x,x_k)\leqslant g^k(x)+1/k\leqslant f(x)+1/k. \tag{2.3}$$

Using that $f\geqslant M>-\infty$ we get

$$\mathrm{d}(x,x_k)\leqslant\frac{1}{k}(f(x)+1/k-f(x_k))\leqslant\frac{1}{k}(f(x)+1/k-M),$$

so that $x_k\to x$. Then, taking the limit in (2.3) and using the lower semi-continuity of f leads to $f(x)\leqslant\liminf_{k\to\infty}f(x_k)\leqslant\sup_{k\to\infty}g^k(x)$. Finally, set $f^k(x)=\min(g^k(x),k)$. Then f^k is k-Lipschitz, bounded by k and one has $\sup_k f^k=f$.

Step 2. Let $\mathcal{F}^k(\mu)=\int f^k\mathrm{d}\mu$. Since f^k is continuous and bounded, the linear form $\mathcal{F}^k$ is narrowly continuous. Thus, $\mathcal{F}=\sup_k\mathcal{F}^k$ is lower semi-continuous as a maximum of lower semi-continuous functions. □

Proof of Theorem 2.1. Define $\mathcal{F}(\gamma):=\int c\mathrm{d}\gamma$, then by Lemma 2.2 $\mathcal{F}$ is l.s.c. for the narrow convergence. We just need to show that the set $\Pi(\mu,\nu)$ is compact for narrow topology. Take a sequence $\gamma_n\in\Pi(\mu,\nu)$, since they are probability measures then they are bounded in the dual of $\mathcal{C}(X\times Y)$. Moreover, usual weak-$\star$ compactness of $\mathcal{P}(X\times Y)$ guarantees the existence of a converging subsequence $\gamma_{n_k} \rightharpoonup \gamma \in \mathcal{P}(X \times Y)$. We need to check that $\gamma \in \Pi(\mu, \nu)$. Fix $\varphi \in \mathcal{C}(X)$, then $\int \varphi(x) \mathrm{d}\gamma_{n_k} = \int \varphi \mathrm{d}\mu$ and by passing to the limit we have $\int \varphi(x) \mathrm{d}\gamma = \int \varphi \mathrm{d}\mu$. This shows that $\pi_{X\#}\gamma = \mu$. The same may be done for π_Y which concludes the proof.

□

3. Kantorovich as a relaxation of Monge

The question that we consider here is the equality between the infimum in Monge problem and the minimum in Kantorovich problem. This part is taken from Santambrogio [6].

Theorem 3.1. *Let $X = Y$ be a compact subset of $\mathbb{R}^d$, $c \in \mathcal{C}(X \times Y)$ and $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$. Assume that μ is atomless. Then,*

$$\inf(\mathrm{MP}) = \min(\mathrm{KP}).$$

This theorem was first proved on $\mathbb{R}$ by Gangbo [4]. The proof presented here is taken from Santambrogio's book [6]. The next two counter examples are due to an article of Pratelli [5], where he also proves an extension of this theorem.

Example 3.2. Take the same measures on $\mathbb{R}^2$ as in example 1.6, but take the discontinuous (but lsc) cost $c(x, y) = 1$ if $\|x - y\| \leqslant 1$ and 2 if not. Then, the value of the infimum in Monge's problem is 2, while the minimum in Kantorovich's problem is 1.

Proof. Take any transport map T between μ and ν. It suffices to show that $\mu(\{x \mid \|T(x) - x\| = 1\}) = 0$, or equivalently that $\mu(E_\pm) = 0$ where $E_\pm = \{x \mid T(x) = x \pm (1, 0)\}$. But, by definition of the measures, $\nu(T(E_+)) = 2\mu(E_+)$, which contradicts the property $T_\#\mu = \nu$ unless $\mu(E_+) = 0$. □

Example 3.3. Consider $\mu_i = \frac{1}{2}(\delta_{x_i} + \alpha\, \lambda|_{B(y_i,1)})$ with $\alpha = \frac{1}{\lambda(B(y_i,1))}$ on $\mathbb{R}^2$ with $c(x, y) = \|x - y\|$. Then, any transport map must transport the Dirac to the Dirac and the ball to the ball, so that its cost is $\|x_1 - x_2\| + \|y_1 - y_2\|$. On the other hand, a transport plan can transport δ_{x_1} to $\alpha\, \lambda|_{B(y_2,1)}$ with cost $\leqslant \|x_1 - y_2\| + 1$. The total cost of this transport plan is $2 + \|x_1 - y_2\| + \|x_2 - y_1\|$, which can be (much) lower that $\|x_1 - x_2\| + \|y_1 - y_2\|$ for suitable positions for these points.

We quote the following lemma without proof, see Corollary 1.28 in [6].

Lemma 3.4. *If $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and μ has no atoms, then $\exists T : \mathbb{R}^d \to \mathbb{R}^d$ measurable such that $T_\#\mu = \nu$.*

Lemma 3.5. *Let K be a compact metric space. For any $\varepsilon > 0$ there exists a (measurable) partition $K_1, \dots, K_N$ of K such that for every i, $\operatorname{diam}(K_i) \leqslant \varepsilon$.*

Proof. By compactness, there exists N points $x_1, \dots, x_N$ such that $K \subseteq \bigcup_i \mathrm{B}(x_i, \varepsilon)$. The partition $K_1, \dots, K_N$ of K defined recursively by $K_i = \{x \in K \setminus K_1 \cup \dots K_{i-1} \mid \forall j, d(x, x_i) \leqslant d(x, x_j)\}$ satisfies $K_i \subseteq \mathrm{B}(x_i, \varepsilon)$. □

Proof of Theorem 3.1. Using the continuity of the functional $\gamma \mapsto \int c \mathrm{d}\gamma$ (which uses the continuity of the cost), the statement will follow if we are able to prove that any transport plan $\gamma \in \Pi(\mu, \nu)$, there exists a sequence of transport maps $T^N : \mathbb{R}^d \to \mathbb{R}^d$ such that $T^N_\# \mu = \nu$ and γ_{T^N} narrowly converges to γ.

By Lemma 3.5, for any $\varepsilon > 0$ there exist a measurable partition $K_1, \dots, K_N$ of X such that $\mathrm{diam}(K_i) \leqslant \varepsilon$. Define $\gamma_i := \gamma|_{K_i \times \mathbb{R}^d}$. Now, let $\mu_i := \pi_{X\#}\gamma_i$ and $\nu_i := \pi_{Y\#}\gamma_i$. Since $\mu_i \leqslant \mu$, the measure μ_i has no atoms, so that by the previous Lemma, there exists a transport plan $S_i : K_i \to \mathbb{R}^d$ with $S_{i\#}\mu_i = \nu_i$. Then by gluing the transports S_i we get a transport T^N sending μ onto ν (here we use that $\mu = \sum_i \mu_i$ and $\nu = \sum_i \nu_i$) as the measures μ_i are concentrated on disjoint sets.

Since $\gamma_{S_i}, \gamma_i \in \mathcal{P}(K_i \times Y)$ both have marginals μ_i and ν_i, one has

$$\gamma_{S_i}(K_i \times K_j) = \nu_i(K_j) = \gamma_i(K_i \times K_j).$$

To prove narrow convergence, we consider a test function $\varphi \in \mathcal{C}_b(X \times Y)$. By compactness of $X \times Y$, this function has a uniform continuity modulus ω_φ with respect to the Euclidean norm on $\mathbb{R}^d \times \mathbb{R}^d$. Moreover,

$$\begin{aligned}
\int_{X \times Y} \varphi \mathrm{d}(\gamma - \gamma_{T^N}) &= \sum_{ij} \int_{K_i \times K_j} \varphi \mathrm{d}(\gamma_i - \gamma_{S_i}) \\
&\leqslant \sum_{ij} \gamma_i(K_i \times K_j) \max_{K_i \times K_j} \varphi - \gamma_{S_i}(K_i \times K_j) \min_{K_i \times K_j} \varphi \\
&\leqslant \sum_{ij} \gamma_i(K_i \times K_j) \omega_\varphi(\mathrm{diam}(K_i \times K_j)) = \mathrm{O}(\omega_\varphi(2\varepsilon)).
\end{aligned}$$

Since this holds for any function φ, one sees that γ_{T^N} converges to γ narrowly. In particular, if γ is the minimizer in Kantorovich's problem, then γ_{T^N} is a minimizing sequence. Then, $T^N_\# \mu = \nu$ and

$$\lim_{N \to \infty} \int_{X \times Y} c(x, T^N(x)) \mathrm{d}\mu(x) = \int_{X \times Y} c \mathrm{d}\gamma,$$

thus proving the statement. □

4. The dual problem

We now focus on duality theory. We firstly find a formal dual problem by exchanging $\inf - \sup$. Let write down the constraint $\gamma \in \Pi(\mu, \nu)$ as follows: if $\gamma \in \mathcal{M}^+(X \times Y)$ (we remind that X, Y are compact spaces) we have

$$\sup_{\varphi, \psi} \int_X \varphi \mathrm{d}\mu + \int_Y \psi \mathrm{d}\nu - \int_{X \times Y} (\varphi(x) + \psi(y)) \mathrm{d}\gamma = \begin{cases} 0 & \text{if } \gamma \in \Pi(\mu, \nu), \\ +\infty & \text{otherwise,} \end{cases}$$

where the supremum is taken on $\mathcal{C}_b(X) \times \mathcal{C}_b(Y)$. Thus we can now remove the constraint on γ in (KP)

$$\inf_{\gamma \in \mathcal{M}^+(X \times Y)} \int_{X \times Y} c \mathrm{d}\gamma + \sup_{\varphi, \psi} \int_X \varphi \mathrm{d}\mu + \int_Y \psi \mathrm{d}\nu - \int_{X \times Y} (\varphi(x) + \psi(y)) \mathrm{d}\gamma$$

and by interchanging sup and inf we get

$$\sup_{\varphi, \psi} \int_X \varphi \mathrm{d}\mu + \int_Y \psi \mathrm{d}\nu + \inf_{\gamma \in \mathcal{M}^+(X \times Y)} \int_{X \times Y} (c(x, y) - \varphi(x) - \psi(y)) \mathrm{d}\gamma.$$

One can now rewrite the inf in γ as constraint on φ and ψ as

$$\inf_{\gamma\in\mathcal{M}^+(X\times Y)}\int_{X\times Y}(c-\varphi\oplus\psi)\mathrm{d}\gamma=\begin{cases}0 & \text{if } \varphi\oplus\psi\leqslant c \text{ on } X\times Y\\ -\infty & \text{otherwise}\end{cases},$$

where $\varphi\oplus\psi(x,y):=\varphi(x)+\psi(y)$.

Definition 4.1 (Dual problem)**.** Given $\mu\in\mathcal{P}(X)$, $\nu\in\mathcal{P}(Y)$ and a *cost function* $c\in\mathcal{C}(X\times Y)$. The *dual problem* is the following optimization problem

$$(\mathrm{DP}):=\sup\left\{\int_X\varphi\mathrm{d}\mu+\int_Y\psi\mathrm{d}\nu \mid \varphi\in\mathcal{C}_b(X),\ \psi\in\mathcal{C}_b(Y),\ \varphi\oplus\psi\leqslant c\right\} \tag{4.4}$$

Remark 4.2. One trivially has the weak duality inequality (KP) $\geqslant$ (DP). Indeed, denoting

$$L(\gamma,\varphi,\psi)=\int_{X\times Y}(c-\varphi\oplus\psi)\mathrm{d}\gamma)+\int_X\varphi\mathrm{d}\mu+\int_Y\psi\mathrm{d}\nu,$$

one has for any $(\varphi,\psi,\gamma)\in\mathcal{C}_b(X)\times\mathcal{C}_b(Y)\times\mathcal{M}^+(X\times Y)$,

$$\inf_{\tilde\gamma\geqslant 0}L(\tilde\gamma,\varphi,\psi)\leqslant L(\gamma,\varphi,\psi)\leqslant\sup_{\tilde\varphi,\tilde\psi}L(\gamma,\tilde\varphi,\tilde\psi)$$

Taking the supremum with respect to (φ,ψ) on the left and the infimum with respect to γ on the right gives inf (KP) $\geqslant$ sup (DP). When sup (DP) = inf (KP), one talks of *strong duality.* Note that this is independent of whether the infimum and the supremum are attained.

Remark 4.3. As often, the Lagrange multipliers (or Kantorovich potentials) φ,ψ have an economic interpretation as prices. For instance, imagine that μ is the distribution of sand available at quarries, and ν describes the amount of sand required by construction work. Then, (KP) can be interpreted as finding the cheapest way of transporting the sand from μ to ν for a construction company. Imagine that this company wants to externalize the transport, by paying a loading coast $\varphi(x)$ at a point x (in a quarry) and an unloading coast $\psi(y)$ at a point y (at a construction place). Then, the constraint $\varphi(x)+\psi(y)\leqslant c(x,y)$ translates the fact that the construction company would not externalize if its cost is higher than the cost of transporting the sand by itself. Then, Kantorovich's dual problem (DP) describes the problem of a transporting company: maximizing its revenue $\int\varphi\mathrm{d}\mu+\int\psi\mathrm{d}\nu$ under the constraint $\varphi\oplus\psi\leqslant c$ imposed by the construction company. The economic interpretation of the strong duality (KP) = (DP) is that in this setting, externalization has exactly the same cost as doing the transport by oneself.

We now focus on the existence of a pair (ψ,ψ) which solves (DP) and postpone the proof of the strong duality to the next lecture.

Definition 4.4 (c-transform and $\overline{c}$-transform)**.** Given a function $f:x\to\overline{\mathbb{R}}$, we define its c-transform $f^c:Y\to\overline{\mathbb{R}}$ by

$$f^c(y)=\inf_{x\in X}c(x,y)-f(x).$$

We also define the $\overline{c}$-transform of $g : Y \to \overline{\mathbb{R}}$ by

$$g^{\overline{c}}(x) = \inf_{y \in Y} c(x,y) - g(y).$$

We also say that a function ψ on Y is $\overline{c}$-concave if there exists f such that $\psi = f^c$. Notice now that if c is continuous on a compact set, and hence uniformly continuous, then there exists am increasing function $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ with $\omega(0) = 0$ such that

$$|c(x,y) - c(x',y')| \leqslant \omega(d_X(x,x') + d_Y(y,y')).$$

If we consider f^c we have that $f^c(y) = \inf_x \tilde{f}_x(y)$ with $\tilde{f}_x(y) = c(x,y) - f(x)$, and the functions $\tilde{f}_x$ satisfy $|\tilde{f}_x(y) - \tilde{f}_x(y)| \leqslant \omega(d_Y(y,y'))$. This implies that f^c actually share the same continuity modulus of c. It snow quite easy to see that given an admissible pair (ψ,ψ) in (DP), one cas always replace it with (φ,φ^c) and then $(\varphi^{c\overline{c}},\varphi^c)$ and the constrained are preserved and the integrals increased. The underlying idea of these transformations is actually to improve a maximizing sequence to get a uniform bound on its continuity.

Theorem 4.5. *Suppose that X and Y are compact and $c \in \mathcal{C}(X \times Y)$. Then there exists a pair $(\varphi^{c\overline{c}},\varphi^c)$ which solves* (DP).

Proof. Let us first denote by $\mathcal{J}(\varphi,\psi)$ the following functional

$$J(\varphi,\psi) = \int_X \varphi, \mathrm{d}\mu + \int_Y \psi \mathrm{d}\nu,$$

then it is clear that for every constant λ we have $\mathcal{J}(\varphi - \lambda, \psi + \lambda) = \mathcal{J}(\varphi,\psi)$. Given now a maximising sequence (φ_n,ψ_n) we can improve it by means of the c- and $\overline{c}$-transform obtaining a new one $(\varphi_n^{c\overline{c}},\varphi_n^c)$. Notice that by the consideration above the sequences $\varphi_n^{c\overline{c}}$ and φ_n^c are uniformly equicontinuous. Since φ_n^c is continuous on a compact set we can always subtract its minimum and assume that $\min_Y \varphi_n^c = 0$. This implies that the sequence φ_n^c is also equibounded as $0 \leqslant \varphi_n^c \leqslant \omega(\mathrm{diam}(Y))$. We also deduce uniform bounds on $\varphi_n^{c\overline{c}}$ as $\varphi_n^{c\overline{c}} = \inf_Y c(x,y) - \varphi^c(y)$. This let us apply Ascoli-Arzela's theorem and extract two uniformly converging subsequences $\varphi_{n_k}^{c\overline{c}} \to \overline{\varphi}$ and $\varphi_{n_k}^c \to \overline{\psi}$ where the pair $(\overline{\varphi},\overline{\psi})$ satisfies the inequality constraint. Moreover, since $(\varphi_n^{c\overline{c}},\varphi_n^c)$ is a maximising sequence we get that the pair $(\overline{\varphi},\overline{\psi})$ is optimal. now one can apply again the c- and $\overline{c}$-transforms obtaining an optimal pair of the form $(\overline{\varphi}^{c\overline{c}},\overline{\varphi}^c)$. ☐

References

1. Najma Ahmad, Hwa Kil Kim, and Robert J McCann, *Extremal doubly stochastic measures and optimal transportation*, arXiv preprint arXiv:1004.4147 (2010).
2. Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Gradient flows: in metric spaces and in the space of probability measures*, Springer Science & Business Media, 2008.
3. Richard M Dudley, *Real analysis and probability*, vol. 74, Cambridge University Press, 2002.
4. Wilfrid Gangbo, *The monge mass transfer problem and its applications*, Contemporary Mathematics **226** (1999), 79–104.
5. Aldo Pratelli, *On the equality between monge's infimum and kantorovich's minimum in optimal mass transportation*, Annales de l'Institut Henri Poincare (B) Probability and Statistics **43** (2007), no. 1, 1–13.
6. Filippo Santambrogio, *Optimal transport for applied mathematicians*, Springer, 2015.

7. Cédric Villani, *Topics in optimal transportation*, no. 58, American Mathematical Soc., 2003.
8. ______, *Optimal transport: old and new*, vol. 338, Springer Science & Business Media, 2008.