Optimal Transport for Applied Mathematicians
Reading notes

Alex Delalande

May 2019

Introduction

Original Monge Problem:

$$\min\{M(T) := \int_X c(x, T(x))d\mu(x) : T_\#\mu = \nu\} \qquad \text{(MP)}$$

$\implies$ Non linear in T & x cannot be sent to different places in Y through $T(x)$

Kantorovich formulation:

$$\min\{K(\gamma) := \int_{X\times Y} c d\gamma : (\pi_x)_\#\gamma = \mu, (\pi_y)_\#\gamma = \nu\} \qquad \text{(KP)}$$

$\implies$ Linear in γ & $\gamma(x, y) =$ "number of particles going from x to y"

1 Primal and dual problems

1.1 Kantorovich & Monge problem

Theorem 1.4 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 function. Then (KP) admits a solution.
Theorem 1.5 Let X and Y be compact metric spaces, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ a lower semi-continuous and bounded from below function. Then (KP) admits a solution.
Theorem 1.7 Let X and Y be Polish spaces, i.e. complete and separable metric spaces, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and $c : X \times Y \to [0, +\infty]$ a lower semi-continuous function. Then (KP) admits a solution.

When cost functions are not continuous:
Lemma 1.8 Let $\gamma_n, \gamma \in \Pi(\mu, \nu)$ be probabilities on $X \times Y$ and $a : X \to \tilde{X}$ and $b : Y \to \tilde{Y}$ be measurable maps valued in two separable metric spaces $\tilde{X}$ and $\tilde{Y}$. Let $c : \tilde{X} \times \tilde{Y} \to \mathbb{R}_+$ be a continuous function with $c(a, b) \leq f(a) + g(b)$ with f, g continuous and $\int (f \circ a) d\mu$, $\int (g \circ b) d\nu < +\infty$. Then:

$$\gamma_n \rightharpoonup \gamma \implies \int_{X \times Y} c(a(x), b(y)) d\gamma_n \to \int_{X \times Y} c(a(x), b(y)) d\gamma$$

1.2 Duality

(KP) can be rewritten as:

$$\min_{\gamma} \int_{X \times Y} c d\gamma + \sup_{\phi, \psi} \int_X \phi d\mu + \int_Y \psi d\nu - \int_{X \times Y} (\phi(x) + \psi(y)) d\gamma$$

with ϕ, ψ continuous bounded functions. inf and sup can actually be interchanged (not classical Rockafellar duality since not finite-dimensional spaces). $\implies$ dual problem:

$$\sup_{\phi, \psi} \int_X \phi d\mu + \int_Y \psi d\nu + \inf_{\gamma} \int_{X \times Y} (c(x, y) - \phi(x) - \psi(y)) d\gamma$$

Let $\phi \oplus \psi(x, y) = \phi(x) + \psi(y)$. Then

$$\inf_{\gamma} \int_{X \times Y} (c - \phi \oplus \psi) d\gamma = \begin{cases} 0 & \text{if} \quad \phi \oplus \psi \leq c \quad \text{on} \quad X \times Y \\ -\infty & \text{otherwise} \end{cases}$$

Thus we get the following Dual Problem:

$$\max\{\int_X \phi d\mu + \int_Y \psi d\nu : \phi \in C_b(X), \psi \in C_b(Y) : \phi \oplus \psi \leq c\} \qquad \text{(DP)}$$

Obviously:

$$\sup \text{(DP)} \leq \inf \text{(KP)}$$

Definition

- *c-transform* For $\chi : X \to \bar{\mathbb{R}}$, $\chi^c : Y \to \bar{\mathbb{R}}$ with $\chi^c(y) = \inf_{x \in X} c(x, y) - \chi(x)$
- *$\bar{c}$-transform* For $\zeta : Y \to \bar{\mathbb{R}}$, $\zeta^{\bar{c}} : X \to \bar{\mathbb{R}}$ with $\zeta^c(x) = \inf_{y \in Y} c(x, y) - \zeta(y)$
- *$\bar{c}$-concavity* ψ defined on Y is $\bar{c}$-concave if there exists χ such that $\psi = \chi^c$
- *c-concavity* ψ defined on Y is c-concave if there exists χ such that $\psi = \chi^{\bar{c}}$

Proposition 1.11 Suppose that X and Y are compact and c is continuous. Then there exists a solution (ϕ, ψ) to problem (DP) and it has the form $\phi \in c - conc(X)$, $\psi \in \bar{c} - conc(X)$, and $\psi = \phi^c$. In particular

$$\max(\mathrm{DP}) = \max_{\phi \in c-conc(X)} \int_X \phi d\mu + \int_Y \phi^c d\nu$$

By admitting min(KP) = max(DP) we have:

$$\min(KP) = \max_{\phi \in c-conc(X)} \int_X \phi d\mu + \int_Y \phi^c d\nu$$

which also shows that the minimum value of (KP) is a convex function of (μ, ν), as it is a supremum of linear functionals. The functions ϕ realizing this maximum are called *Kantorovich potentials*.

1.3 The case $c(x, y) = h(x - y)$ for h strictly convex, and the existence of an optimal T

Here, $X = Y = \Omega \subset \mathbb{R}^d$ and the cost c is of the form $c(x, y) = h(x - y)$, for a strictly convex function h. We will also assume Ω to be compact for simplicity.

Proposition 1.15 If c is C^1, ϕ is a Kantorovich potential for the cost c in the transport from μ to ν, and (x_0, y_0) belongs to the support of an optimal transport plan γ, then $\nabla\phi(x_0) = \nabla_x c(x_0, y_0)$, provided ϕ is differentiable at x_0. In particular, the gradients of two different Kantorovich potentials coincide on every point $x_0 \in spt(\mu)$ where both the potentials are differentiable.
Twist condition $c : \Omega \times \Omega \to \mathbb{R}$ differentiable w.r.t. x at every point and map $y \mapsto \nabla_x c(x_0, y)$ is injective for every x_0.
$\implies$ goal of this condition = deduce from $(x_0, y_0) \in spt(\gamma)$ that y_0 is indeed uniquely defined from x_0 $\implies$ γ is concentrated on a graph, that of the map associating y_0 to each x_0: this map will be the optimal transport plan! Since this map has been constructed using ϕ and c only, and not γ, it also provides uniqueness for the optimal γ. For $c(x, y) = h(x - y)$ and h differentiable, $\nabla\phi(x_0) = \nabla h(x_0 - y_0)$ (use subdifferential if not differentiable). For h strictly convex, analytic transport map defined by:

$$x_0 - y_0 = (\nabla h)^{-1}(\nabla\phi(x_0))$$

$\implies$ **Theorem 1.17** Given μ and ν probability measures on a compact domain $\Omega \subset \mathbb{R}^d$ there exists an optimal transport plan γ for the cost $c(x, y) = h(x - y)$ with h strictly convex. It is unique and of the form $(id, T)_{\#}\mu$, provided μ is absolutely continuous and $\partial\Omega$ is negligible. Moreover, there exists a Kantorovich potential ϕ, and T and the potentials ϕ are linked by

$$T(x) = x - (\nabla h)^{-1}(\nabla\phi(x))$$

Quadratic case $c(x, y) = \frac{1}{2}|x - y|^2 \implies T(x) = \nabla u(x)$ with u a convex (and l.s.c.) function. Actually, u exists and is unique (Brenier Theorem).
Theorem 1.22 Let μ, ν be probabilities over $\mathbb{R}^d$ and $c(x, y) = \frac{1}{2}|x - y|^2$. Suppose $\int |x|^2 dx, \int |y|^2 dy < +\infty$, which implies $\min(\mathrm{KP}) < +\infty$ and suppose that μ gives no mass to $(d - 1)$-surfaces of class C^2. Then there exists, unique, an optimal transport map T from μ to ν, and it is of the form $T = \nabla u$ for a convex function u. See *Remark 1.23* for nice consequence in 1D.
Quadratic case on the flat torus - Theorem 1.25 Take $\mu, \nu \in \mathcal{P}(\mathbb{T}^d)$, with $\mu << \mathcal{L}^d$ and $c(x, y) = \frac{1}{2}|[x - y]|^2$. Then there exists a unique optimal transport plan $\gamma \in \Pi(\mu, \nu)$, it has the form $\gamma = \gamma_T$ and the optimal map T is given by $T(x) = x - \nabla\phi(x)$ for a.e. x, where the sum $x - \nabla\phi(x)$ is to be intended modulo $\mathbb{Z}^d$. Here the function ϕ is a Kantorovich potential, solution to the dual problem, in the transport from μ to ν for the cost c. Moreover, for a.e. $x \in \mathbb{T}^d$, the point $T(x)$ does not belong to $cut(x)$.

1.4 Counter-examples to existence

(Counter examples to Theorem 1.17)

- If $\mu = \delta_a, a \in X$ and ν is not a Dirac mass. Then with $T_\# \delta_a = \delta_{T(a)}$, no transport map $T : X \to Y$ can exist if ν is not of the form δ_b for some $b \in Y$. In fact, $\forall$ atom a of μ, there exist an atom of ν of mass *at least* $\mu(\{a\})$. $\implies$ "absence of atom for source measure μ" = typical assumption before solving (MP). But for (KP), atomistic case = OK and solution is unique: $\gamma = \mu \otimes \nu = \delta_a \otimes \nu$
- Optimal transport map/plan may not exist while transportation is possible (see pages 19/20)

1.5 Kantorovich as a relaxation of Monge

Let $K(\gamma) = \int_{\Omega \times \Omega} c d\gamma$ and $J(\gamma) = \begin{cases} K(\gamma) = M(T) & \text{if} \quad \gamma = \gamma_T \\ +\infty & \text{otherwise} \end{cases}$. (MP) is equivalent to:

$$\min\{J(\gamma) : \gamma \in \Pi(\mu, \nu)\}$$

We will see here that K can be seen as the *relaxation* of J (see Memo Box 1.10 page 20 for relaxation definition).
Lemma 1.27 If μ, ν are two probability measures on the real line $\mathbb{R}$ and μ is atomless, then there exists at least a transport map T such that $T_\# \mu = \nu$. This is actually also true for measures on $\mathbb{R}^d$ (corollary 1.29).
Theorem 1.32 On a compact subset $\Omega \subset \mathbb{R}^d$, the set of plans γ_T induced by a transport is dense in the set of plans $\Pi(\mu, \nu)$ whenever μ is atomless.
Theorem 1.33 For $\Omega \subset \mathbb{R}^d$ compact, K is the relaxation of J. In particular, $\inf J = \min K$, and hence Monge and Kantorovich problems have the same infimum.

1.6 Convexity, c-concavity, cyclical monotonicity, duality and optimality

Theorem 1.37. If $\Gamma \neq \emptyset$ is a c-CM set in $X \times Y$ and $c : X \times Y \to \mathbb{R}$ (note that c is required not to take the value $+\infty$), then there exists a c-concave function $\phi : X \to \mathbb{R} \cup \{-\infty\}$ (different from the constant $-\infty$ function) such that $\Gamma \subset \{(x, y) \in X \times Y : \phi(x) + \phi^c(y) = c(x, y)\}$.
Theorem 1.38 If γ is an optimal transport plan for the cost c and c is continuous, then $spt(\gamma)$ is a c-CM set.
$\implies$ from the previous two theorems, we get the following **duality result**:
Theorem 1.39 Suppose that X and Y are Polish spaces and that $c : X \times Y \to \mathbb{R}$ is uniformly continuous and bounded. Then the problem (DP) admits a solution (ϕ, ϕ^c) and we have max(DP) = min(KP) (ϕ and ϕ^c are continuous and bounded).
Extension to quadratic case ($c(x, y) = \frac{1}{2}|x - y|^2$ is not bounded neither uniformly continuous):
Theorem 1.40 Let μ, ν be probabilities over $\mathbb{R}$ and $c(x, y) = \frac{1}{2}|x - y|^2$. Suppose $\int |x|^2 dx$, $\int |y|^2 dy < +\infty$. Consider the following variant of (DP):

$$\sup\{\int_{\mathbb{R}^d} \phi d\mu + \int_{\mathbb{R}^d} \psi d\nu : \phi \in L^1(\mu), \psi \in L^1(\nu), \phi \oplus \psi \leq c\} \qquad \text{(DP - var)}$$

Then (DP-var) admits a solution (ϕ, ψ), and the functions $x \mapsto \frac{1}{2}|x|^2 - \phi(x)$ and $y \mapsto \frac{1}{2}|y|^2 - \psi(y)$ are convex and conjugate to each other for the Legendre transform. Moreover, we have max(DP - var) = min(KP).
Extension of duality in the l.s.c. case
Theorem 1.42 If X, Y are Polish spaces and $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ is l.s.c. and bounded from below, then the duality formula min(KP) = sup(DP) holds. *Note that for the cost c we cannot guarantee the existence of a maximizing pair (ϕ, ψ).*

Theorem 1.43 If c is l.s.c. and γ is an optimal transport plan, then γ is concentrated on a c-CM set Γ (which will not be closed in general).

Sufficient conditions for optimality and stability
Theorem 1.47 Let $\Omega \subset \mathbb{R}^d$ be compact, and c be a C^1 cost function satisfying the twist condition on $\Omega \times \Omega$. Suppose that $\mu \in \mathcal{P}(\Omega)$ and $\phi \in c - conc(\Omega)$ are given, that ϕ is differentiable μ-a.e. and that $\mu(\partial\Omega) = 0$. Suppose that the map T satisfies $\nabla_x c(x, T(x)) = \nabla\phi(x)$. Then T is optimal for the transport cost c between the measures μ and $\nu := T_{\#}\mu$.
Quadratic case - Theorem 1.48 Suppose that $\mu \in \mathcal{P}(\mathbb{R}^d)$ is such that $\int |x|^2 d\mu(x) < +\infty$, that $u : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is convex and differentiable $\mu-a.e.$, and set $T = \nabla u$ and suppose $\int |T(x)|^2 d\mu(x) < +\infty$. Then T is optimal for the transport cost $c(x, y) := \frac{1}{2}|x - y|^2$ between the measures μ and $\nu := T_{\#}\mu$.
We know that the support of optimal plan are c-CM. Actually, the reverse is true: plan that are c-CM are optimal:
Theorem 1.49 Suppose that $\gamma \in \mathcal{P}(X \times Y)$ is given, that X and Y are Polish spaces, that $c : X \times Y \to \mathbb{R}$ is uniformly continuous and bounded, and that $spt(\gamma)$ is c-CM. Then γ is an optimal transport plan between its marginals $\mu = (\pi_x)_{\#}\gamma$ and $\nu = (\pi_y)_{\#}\gamma$ for the cost c.
Stability result:
Theorem 1.50 Suppose that X and Y are compact metric spaces and that $c : X \times Y \to \mathbb{R}$ is continuous. Suppose that $(\gamma_n) \in \mathcal{P}(X \times Y)$ is a sequence of transport plan which are optimal for the cost c between their own marginals $\mu_n = (\pi_x)_{\#}\gamma_n$ and $\nu_n = (\pi_y)_{\#}\gamma_n$, and suppose $\gamma_n \rightharpoonup \gamma$. Then $\mu_n \rightharpoonup \mu := (\pi_x)_{\#}\gamma$ and $\nu_n \rightharpoonup \nu := (\pi_y)_{\#}\gamma$ and γ is optimal in the transport between μ and ν.
$\implies$ Useful consequence: let's define for the cost $c : X \times Y \to \mathbb{R}$ and $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, $\mathcal{T}_c(\mu, \nu) := \min\{\int_{X \times Y} c d\gamma : \gamma \in \Pi(\mu, \nu)$. Then:
Theorem 1.51 Suppose that X and Y are compact metric spaces and that $c : X \times Y \to \mathbb{R}$ is continuous. Suppose that $\mu_n \in \mathcal{P}(X)$ and $\nu_n \in \mathcal{P}(Y)$ are two sequences of probability measures, with $\mu_n \rightharpoonup \mu$ and $\nu_n \rightharpoonup \nu$. Then we have $\mathcal{T}_c(\mu_n, \nu_n) \to \mathcal{T}_c(\mu, \nu)$.
Stability of the Kantorovich potentials:
Theorem 1.52 Suppose that X and Y are compact metric spaces and that $c : X \times Y \to \mathbb{R}$ is continuous. Suppose that $\mu_n \in \mathcal{P}(X)$ and $\nu_n \in \mathcal{P}(Y)$ are two sequences of probability measures, with $\mu_n \rightharpoonup \mu$ and $\nu_n \rightharpoonup \nu$. Let (ϕ_n, ψ_n) be, for each n, a pair of c-concave Kantorovich potentials for the cost c in the transport from μ_n to ν_n. Then, up to subsequences, we have $\phi_n \to \phi$, $\psi_n \to \psi$, where the convergence is uniform and (ϕ, ψ) is a pair of Kantorovich potentials for μ and ν.

1.7 Discussion

1.7.1 Probabilistic interpretation

(MP) can be seen as the following optimization problem:

$$\min\{\mathbb{E}[c(X, Y)] : X \sim \mu, Y \sim \nu\}$$

For the case $c(X, Y) = |X - Y|^2$, this problem becomes:

$$\max\{\mathbb{E}[(X - x_0).(Y - y_0)] : X \sim \mu, Y \sim \nu\}$$

where $x_0 = \mathbb{E}[X]$ and $y_0 = \mathbb{E}[Y]$. $\implies$ covariance maximization.

1.7.2 Polar Factorization

A classical result in linear algebra states that every matrix $A \in \mathcal{M}^{N \times N}$ can be decomposed as a product $A = SU$, where S is symmetric and positive-semidefinite, and U is a unitary matrix, i.e. $UU^t = I$. The decomposition is unique if A is non-singular (otherwise U is not uniquely defined), and in such a case S is positive definite. Also, one can see that the matrix U of this decomposition is also a solution (the unique one if A is non singular) of

$$\max\{A : R : RR^t = I\}$$

where $A : R$ stands for the scalar product between matrices, defined as $A : R := Tr(AR^t)$.
Analogously, in his first works about the quadratic optimal transport, Y. Brenier noted that Monge-Kantorovich theory allowed to provide a similar decomposition for vector fields instead of linear maps:
Theorem 1.53 Given a vector map $\xi : \Omega \to \mathbb{R}^d$ with $\Omega \subset \mathbb{R}^d$, consider the rescaled Lebesgue measure $\mathcal{L}_\Omega$ on Ω and suppose that $\xi_\#\mathcal{L}_\Omega$ is absolutely continuous; then, one can find a convex function $u : \Omega \to \mathbb{R}$ and a measure-preserving map $s : \Omega \to \Omega$ (i.e. such that $s_\#\mathcal{L}_\Omega = \mathcal{L}_\Omega$) such that $\xi = (\nabla u) \circ s$. Moreover, both s and ∇u are uniquely defined a.e. and s solves

$$\max\{\int \xi(x).r(x)dx : r_\#\mathcal{L}_\Omega = \mathcal{L}_\Omega\}$$

2 One-dimensional issues

2.1 Monotone transport maps and plans in 1D