Optimal transport: introduction, applications and derivation
Lecture 1: the Brenier theorem

Contents

As the name suggests, the fundamental question of optimal transport theory is to compute the cheapest way to transfer mass from one location (say for instance a quarry) to another (say a construction site).

Figure 1: The first question in optimal transport: how to move the mass from a quarry to a construction site with minimal effort.

When we say that we want to transfer mass, we mean that the data of the problem are two distributions of mass, modelled by nonnegative measures μ and ν, of same total mass. Up to normalizing them, we will always assume that they are probability measures. When we say that we want to minimize something, we imply that there is a notion of cost, a function $c = c(x, y)$ of transporting a unite of mass from location x to location y. To finally set a problem, it remains to explain how we model the transportation of mass. Historically, it has been done in two steps.

1 The Monge problem

In what follows, if X is a measurable space (whose σ-algebra is not specified, but always assumed to be the Borel σ-algebra when dealing with metric spaces), $\mathcal{M}(X)$ stands for the set of nonnegative measures on X and $\mathcal{P}(X)$ stands for the set of probability measures on X.

The first idea is to model the transport by a map $T = T(x)$, meaning that all the mass located at x initially will be sent at location $T(x)$. Doing so, if the mass is initially distributed according to the measure μ, it will end up being distributed according to the measure $T_{\#}\mu$ defined as follows.

Definition 1 (Push-forward operation)**.** Let X, Y be two measurable 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 $T_{\#}\mu \in \mathcal{M}(Y)$ defined for all measurable set $A \subset Y$ by

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

The map T is an admissible transport for the problem of sending μ onto ν provided $T_{\#}\mu = \nu$. If so, we say that T is a transport map from μ to ν.

Definition 2 (Transport map)**.** Let X, Y be two measurable spaces, $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. We say that $T : X \to Y$ is a transport map from μ to ν if T is measurable and $T_{\#}\mu = \nu$.

We have now all the ingredients for setting the Monge problem of optimal transport.

Definition 3 (Monge problem)**.** Let X, Y be two measurable spaces, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and let us choose a cost function $c : X \times Y \to \mathbb{R}$, measurable.

The Monge problem of transporting μ onto ν with cost c is the minimization problem

$$C_M(\mu, \nu) := \inf \left\{ \int c(x, T(x))\, \mathrm{d}\mu(x) \ : \ T_{\#}\mu = \nu \right\}. \tag{1}$$

In plain words, the Monge problem consists in minimizing the transport cost

$$\int c(x, T(x))\, \mathrm{d}\mu(x)$$

among all transport maps T from μ to ν.

Remark 4. Uniqueness?

A major issue of this optimization problem is that it does not always admit a solution, or even a competitor. Indeed, for some μ and ν, no transport map from μ to ν exists.

Exemple 5. Let us assume that μ is a Dirac mass. Then for all $T : X \to Y$ measurable, $T_{\#}\mu$ is a Dirac mass as well. Hence, whenever ν is not a Dirac mass, there cannot exist any transport map from μ to ν.

Therefore, to get solutions in general, it is necessary to allow concentrated mass to split along the transport, and hence to change the modelling of the transport. This is done as follows.

2 The Kantorovic problem

To model this more general kind of transports, we use probability measures on $X \times Y$. The idea is that if γ is such a probability measure, and if $A \subset X$ and $B \subset Y$ are measurable, $\gamma(A \times B)$ describes the mass sent along the transport from A to B. This γ is compatible with the data μ and ν provided it is a *coupling*, or a *transport plan* between μ and ν as defined below.

Definition 6 (Transport plan)**.** Let X, Y be two measurable spaces, $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. We denote by $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ the canonical projections. The probability measure $\gamma \in \mathcal{P}(X \times Y)$ is a transport plan between μ and ν if $\pi_{X\#}\gamma = \mu$ and if $\pi_{Y\#}\gamma = \nu$. We denote by $\Pi(\mu, \nu)$ the set of transport plans from μ to ν.

A very good property of transport plans is that they always exist: the product measure $\mu \otimes \nu$ is always a transport plan from μ to ν. The Kantorovic problem is the following optimization problem.

Definition 7 (Kantorovic)**.** Let X, Y be two measurable spaces, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and let us choose a cost function $c : X \times Y \to \mathbb{R}$, measurable.

The Kantorovic problem of transporting μ onto ν with cost c is the minimization problem

$$C_K(\mu, \nu) := \inf \left\{ \int c(x, y) \, \mathrm{d}\gamma(x, y) \ : \ \gamma \in \Pi(\mu, \nu) \right\}. \tag{2}$$

In plain words, the Kantorovic problem consists in minimizing the transport cost

$$\int c(x, y) \, \mathrm{d}\gamma(x, y)$$

among all transport plans from μ to ν

The Kantorovic problem is a generalization of the Monge problem, as explained in the next proposition whose proof is left as an exercise.

Proposition 8. *Let X, Y be two measurable spaces, $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and $c : X \times Y \to \mathbb{R}$. If T is a transport map from μ to ν, then $\gamma_T := (\mathrm{Id}, T)_{\#}\mu$ is a transport plan from μ to ν, and*

$$\int c \, \mathrm{d}\gamma_T = \int c(x, T(x)) \, \mathrm{d}\mu(x).$$

In particular, $C_K(\mu, \nu) \le C_M(\mu, \nu)$.

Here, we are in a very good situation: the Kantorovic problem consists in minimizing an affine functional under affine constraints. This is a linear program! And indeed, with very little structure, we are able to prove the existence of a solution.

Proposition 9. *Let us assume that*

- *X and Y are separable and complete metric spaces;*
- *c is lower semi-continuous, below bounded, with possibly values $+\infty$;*
- $\int c \,\mathrm{d}\mu \otimes \nu < +\infty$.

Then the Kantorovic problem (2) *admits a minimizer.*

The proof relies on the Prokhorov theorem, that we recall here without a proof.

Theorem 10 (Prokhorov)**.** *Let X be a separable metric space. A subset $\mathcal{K}$ of $\mathcal{P}(X)$ is relatively sequentially compact for the topology of narrow convergence (*i.e.* in duality with the set $C_b(X)$ of bounded continuous functions) if and only if it is tight, that is*

$$\forall \varepsilon > 0,\ \exists K_\varepsilon \subset X \text{ compact s.t. } \forall \mu \in \mathcal{K},\ \mu(X \backslash K_\varepsilon) \leq \varepsilon.$$

We are no ready to prove Proposition 9.

Proof of Proposition 9. We want to minimize the functional

$$F : \gamma \in \mathcal{P}(X \times Y) \mapsto \int c \,\mathrm{d}\gamma \in \mathbb{R} \cup \{+\infty\}.$$

over the set $\Pi(\mu, \nu)$. Hence, to prove existence, it suffices to show that the problem admits a competitor, that F is lower semi-continuous, and that $\Pi(\mu, \nu)$ is sequentially compact for some topology. It will be done in the narrow topology. The existence of a competitor is given by assumption (take $\mu \otimes \nu$).

F is narrowly l.s.c. Take $(\gamma_n) \in \mathcal{P}(X \times Y)^{\mathbb{N}}$ a sequence narrowly converging to some $\gamma \in \mathcal{P}(X \times Y)$. As $X \times Y$ is metric, the cost function c being l.s.c. and bounded below, it is the supremum of a sequence of bounded continuous functions. Let us take $(c_p) \in C_b(X \times Y)^{\mathbb{N}}$ a sequence of bounded continuous function increasing towards c. By the monotone convergence theorem,

$$F(\gamma) = \lim_{p \to +\infty} \int c_p \,\mathrm{d}\gamma.$$

So for all $\varepsilon > 0$ there exists p such that $\int c_p \,\mathrm{d}\gamma \geq F(\gamma) - \varepsilon$. But we have

$$\liminf_{n \to +\infty} \int c \,\mathrm{d}\gamma_n \geq \lim_{n \to +\infty} \int c_p \,\mathrm{d}\gamma_n = \int c_p \,\mathrm{d}\gamma \geq F(\gamma) - \varepsilon.$$

So we get the result by letting $\varepsilon \to 0$.

$\Pi(\mu, \nu)$ is narrowly sequentially compact. This is a direct consequence of the Prokhorov theorem. As $\Pi(\mu, \nu)$ is clearly closed, we only need to show that it is tight. So we give ourselves $\varepsilon > 0$ and we want to find a compact K_ε such that for all $\gamma \in \Pi(\mu, \nu)$, $\gamma((X \times Y) \backslash K_\varepsilon) \leq \varepsilon$. But as X ans Y are separable and complete, μ and ν are tight. So we can find $A_\varepsilon \subset X$ and $B_\varepsilon \subset Y$ compact such that $\mu(X \backslash A_\varepsilon) \leq \varepsilon/2$ and $\nu(Y \backslash B_\varepsilon) \leq \varepsilon/2$. Choosing $K_\varepsilon = A_\varepsilon \times B_\varepsilon$, as

$$(X \times Y) \backslash K_\varepsilon = \Big((X \backslash A_\varepsilon) \times Y \Big) \cup \Big(X \times (Y \backslash B_\varepsilon) \Big),$$

we find that

$$\gamma((X \times Y) \backslash K_\varepsilon) \leq \gamma((X \backslash A_\varepsilon) \times Y) + \gamma(X \times (Y \backslash B_\varepsilon)) = \mu(X \backslash A_\varepsilon) + \nu(Y \backslash B_\varepsilon) \leq \varepsilon.$$

The result is therefore proven. □

3 The Brenier theorem, statement

The Kantorovic problem has the advantage of barely always admitting solutions. The next question is hence to study these solutions:

- Do we know something about these solutions, their support?
- Are these solutions unique?

To have a picture in mind of the kind of answer we can get, let us give ourselves as a target the following result due to Yann Brenier [1].

Theorem 11 (Brenier). *Let us set $X = Y = \mathbb{R}^d$, for some $d \in \mathbb{N}^*$, and for all $x, y \in \mathbb{R}^d$, $c(x,y) = |y-x|^2$. Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ satisfy*

$$\int |x|^2 \,\mathrm{d}\mu(x) < +\infty \quad \textit{and} \quad \int |y|^2 \,\mathrm{d}\nu(y) < +\infty. \tag{3}$$

Under the assumption that $\mu \ll \mathrm{d}x$, the Kantorovic problem (2) *admits a unique solution γ, and there exists a* convex *function $\alpha : \mathbb{R}^d \to \mathbb{R}^d$ such that $\gamma = (\mathrm{Id}, \nabla\alpha)_\# \mu$. In particular, $C_M(\mu,\nu) = C_K(\mu,\nu)$ and the Monge problem* (1) *admits a solution, unique up to a μ-negligible set.*

Remark 12. By a famous result, convex functions are differentiable $\mathrm{d}x$-almost everywhere, so under the assumption $\mu \ll \mathrm{d}x$, $\nabla\alpha$ is well defined μ almost everywhere.

So in the quadratic case and under few assumptions the (unique) solution of the Kantorovic problem is concentrated on a graph: we say that it is of Monge type. Relaxing the Monge problem does not add any new solution. It is a standard trick in calculus of variations when studying a problem for which existence is not clear, to relax it in order to get solutions, possibly in a wider class, and then to show that solutions of the relaxed problem are actually solutions of the original one.

We will prove almost everything in Theorem 11, only adding the assumption that μ and ν are concentrated on a compact. Along the way, we will discuss several general aspects of the solutions of optimal transport problems for general (continuous) costs.

4 Duality

We said that the Kantorovic problem consists in minimizing an affine (hence convex) functional under affine constraints. For such problems, a robust method for getting optimality conditions is to use the concept of duality. This concept starts with a formal trick. Starting from (2), assuming for instance that X and Y are separable and complete metric spaces, we notice the following identity valid for all $\gamma \in \mathcal{M}(X \times Y)$:

$$\sup_{\varphi \in C_b(X),\, \psi \in C_b(Y)} \int \varphi \,\mathrm{d}\mu + \int \psi \,\mathrm{d}\nu - \int \varphi \oplus \psi \,\mathrm{d}\gamma = \begin{cases} 0 & \text{if } \gamma \in \Pi(\mu,\nu), \\ +\infty & \text{otherwise,} \end{cases}$$

where we used the notation $\varphi \oplus \psi$ for the function $(x,y) \mapsto \varphi(x) + \psi(y)$. Using this identity, we observe:

$$C_K(\mu,\nu) := \inf_{\gamma \in \Pi(\mu,\nu)} \int c \,\mathrm{d}\gamma = \inf_{\gamma \in \mathcal{M}(X\times Y)} \sup_{\varphi,\psi \in C_b} \int c\,\mathrm{d}\gamma + \int \varphi \,\mathrm{d}\mu + \int \psi \,\mathrm{d}\nu - \int \varphi \oplus \psi \,\mathrm{d}\gamma.$$

Now, let us imagine that we can invert the "inf" and the "sup". We get

$$C_K(\mu,\nu) = \sup_{\varphi,\psi \in C_b} \int \varphi \,\mathrm{d}\mu + \int \psi \,\mathrm{d}\nu + \inf_{\gamma \in \mathcal{M}(X\times Y)} \int \{c - \varphi \oplus \psi\} \,\mathrm{d}\gamma.$$

But a quick analysis shows

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

Therefore, when computing the supremum in φ, ψ, the infimum needs to not be $-\infty$, and so the supremum is necessarily achieved for test functions satisfying $\varphi \oplus \psi \leq c$, that is

$$\inf_{\gamma \in \Pi(\mu,\nu)} \int c \,\mathrm{d}\gamma = \sup_{\substack{\varphi,\psi \in C_b \\ \varphi \oplus \psi \leq c}} \int \varphi \,\mathrm{d}\mu + \int \psi \,\mathrm{d}\nu. \tag{4}$$

This *duality* formula is at the core of the analysis of optimal transport problems.

Let us emphasize that because $\inf\sup \geq \sup\inf$, the inequality "$\geq$" is always true, even allowing φ and ψ to be in the largest space of $\mathcal{L}^1$ functions: for all $\gamma \in \Pi(\mu,\nu)$, $\varphi \in \mathcal{L}^1(\mu)$, $\psi \in \mathcal{L}^1(\nu)$, if $\varphi \oplus \psi \leq c$ everywhere, then

$$\int c \,\mathrm{d}\gamma \geq \int \varphi \oplus \psi \,\mathrm{d}\gamma = \int \varphi \,\mathrm{d}\mu + \int \psi \,\mathrm{d}\nu. \tag{5}$$

So the nontrivial part in equality (4) is "$\leq$".

In order to give a first insight of the power of duality, let us give a first very general result which does not need any intricate idea and which only use the easy part of (4).

Proposition 13 (Sufficient conditions for optimality)**.** *Let X, Y be any measurable spaces, $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. Consider the following assertions.*

1. *γ is a solution of the Kantorovic problem* (2)*.*
2. *There exists $\varphi \in \mathcal{L}^1(\mu)$ and $\psi \in \mathcal{L}^1(\nu)$ such that $\varphi \oplus \psi \leq c$ everywhere, and the following identity holds*
$$\int c \,\mathrm{d}\gamma = \int \varphi \,\mathrm{d}\mu + \int \psi \,\mathrm{d}\nu. \tag{6}$$
3. *There exists $\varphi \in \mathcal{L}^1(\mu)$ and $\psi \in \mathcal{L}^1(\nu)$ such that $\varphi \oplus \psi \leq c$ everywhere, and for γ-almost all $(x,y) \in X \times Y$,*
$$\varphi(x) + \psi(y) = c(x,y). \tag{7}$$

We have $2 \Leftrightarrow 3 \Rightarrow 1$. Otherwise stated, 2 *and* 3 *are sufficient conditions for optimality.*

Remark 14.
- We will see later on contexts when these conditions are also necessary, that is, when there is no *duality gap*. But this result needs more assumptions, and is more difficult.
- Observe that this very easy proposition gives a way to build solutions of optimal transport problems with unknown marginals μ and ν. Indeed, take φ and ψ two measurable functions with $\varphi \oplus \psi \leq c$ (say bounded to avoid integrability issues). Then build γ a probability measure concentrated on the subset of $X \times Y$ where equality (7) holds. Then γ is a solution of the optimal transport problem with its own marginals.
- If condition 2 and 3 hold, then because of (5) and (6), (φ, ψ) is a maximizer of $(\varphi,\psi) \mapsto \int \varphi \,\mathrm{d}\mu + \int \psi \,\mathrm{d}\nu$ under constraint $\varphi \oplus \psi \leq c$.

Proof. Let us first prove that $2 \Rightarrow 1$. Assuming 2, let us consider φ and ψ as given by this assertion, and let us choose $\gamma' \in \Pi(\mu,\nu)$. Because $\varphi \oplus \psi \leq c$ everywhere

$$\int c \,\mathrm{d}\gamma' \geq \int \varphi \oplus \psi \,\mathrm{d}\gamma' = \int \varphi \,\mathrm{d}\mu + \int \psi \,\mathrm{d}\nu = \int c \,\mathrm{d}\gamma.$$

So γ minimizes $\int c \,\mathrm{d}\gamma'$ over $\gamma' \in \Pi(\mu,\nu)$.

Now, let us prove that $2 \Rightarrow 3$. Assuming 2, considering φ and ψ as given by the assertion, identity (6) rewrites

$$\int \{c - \varphi \oplus \psi\} \,\mathrm{d}\gamma = 0.$$

As $c \geq \varphi \oplus \psi$, the result follows.

Let us prove $3 \Rightarrow 2$. Assuming 3, considering φ and ψ as given by the assertion, we have

$$\int c \, \mathrm{d}\gamma = \int \varphi \oplus \psi \, \mathrm{d}\gamma = \int \varphi \, \mathrm{d}\mu + \int \psi \, \mathrm{d}\nu.$$

□

5 Proof of the absence of duality gap

In this section, we prove the absence of duality gap in the case when X and Y are compact metric spaces and c is continuous. Actually, the result would be true even in separable and complete spaces, with lower-semi continuous costs, working in duality with bounded continuous functions, see [5, Theorem 1.3]. Our proof recovers the existence of a solution of the optimal transport problem.

Theorem 15. *Let X and Y be compact metric spaces, $c : X \times Y \to \mathbb{R}$ be continuous, $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$. We have*

$$\inf_{\gamma \in \Pi(\mu,\nu)} \int c \, \mathrm{d}\gamma = \sup_{\substack{\varphi \in \mathcal{L}^1(\mu), \psi \in \mathcal{L}^1(\nu) \\ \varphi \oplus \psi \leq c}} \int \varphi \, \mathrm{d}\mu + \int \psi \, \mathrm{d}\nu = \sup_{\substack{\varphi \in C(X), \psi \in C(Y) \\ \varphi \oplus \psi \leq c}} \int \varphi \, \mathrm{d}\mu + \int \psi \, \mathrm{d}\nu.$$

Moreover, the infimum in the l.h.s. is achieved.

The proof uses the following elementary lemma.

Lemma 16. *Let Z be a metric compact space and $\lambda : C(Z) \to \mathbb{R}$ be a linear functional (not necessarily continuous). The two following assertions are equivalent:*

1. *There exists $m \in \mathcal{P}(Z)$ such that for all $f \in C(Z)$, $\lambda(f) = \int f \, \mathrm{d}m$.*
2. *For all $f \in C(Z)$, $\lambda(f) \leq \sup f$.*

Proof of the Lemma. $1 \Rightarrow 2$ is trivial. Let us prove $2 \Rightarrow 1$, that is, let us assume that for all $f \in C(Z)$, $\lambda(f) \leq \sup f$. First of all, we have $\lambda(f) \leq \sup f \leq \|f\|_\infty$ and $-\lambda(f) = \lambda(-f) \leq -\inf f \leq \|f\|_\infty$, so that $|\lambda(f)| \leq \|f\|_\infty$. λ is therefore a continuous linear functional on $C(Z)$, that is, in virtue of the Riesz theorem, it can be represented by a Radon measure.

Moreover, if $f \geq 0$, $\lambda(-f) \leq \sup f \leq 0$, so $\lambda(f) \geq 0$, so λ can be represented by a nonnegative Radon measure $m \in \mathcal{M}(Z)$.

It remains to prove that $\lambda(1) = 1$. This is because $\lambda(1) \leq 1$ and $\lambda(-1) \leq -1$. □

We are now ready to prove Theorem 15.

Proof of Theorem 15. We already observed that

$$\inf_{\gamma \in \Pi(\mu,\nu)} \int c \, \mathrm{d}\gamma \geq \sup_{\substack{\varphi \in \mathcal{L}^1(\mu), \psi \in \mathcal{L}^1(\nu) \\ \varphi \oplus \psi \leq c}} \int \varphi \, \mathrm{d}\mu + \int \psi \, \mathrm{d}\nu \geq \sup_{\substack{\varphi \in C(X), \psi \in C(Y) \\ \varphi \oplus \psi \leq c}} \int \varphi \, \mathrm{d}\mu + \int \psi \, \mathrm{d}\nu.$$

So we just need to prove that there exists $\gamma \in \Pi(\mu, \nu)$ such that

$$\int c \, \mathrm{d}\gamma = \sup_{\substack{\varphi \in C(X), \psi \in C(Y) \\ \varphi \oplus \psi \leq c}} \int \varphi \, \mathrm{d}\mu + \int \psi \, \mathrm{d}\nu.$$

Let us call K the quantity in the r.h.s., and $\mathcal{E} := \{\varphi \oplus \psi \,|\, \varphi \in C(X),\, \psi \in C(Y)\} \subset C(X \times Y)$. From Lemma 16, we deduce that it suffices to find a linear functional λ on $C(X, Y)$ such that:

1. For all $f \in C(X \times Y)$, $\lambda(f) \leq \sup f$.
2. For all $\varphi \oplus \psi \in \mathcal{E}$, $\lambda(\varphi \oplus \psi) = \int \varphi \, \mathrm{d}\mu + \int \psi \, \mathrm{d}\nu$.
3. The value $\lambda(c)$ of λ at $f = c$ is K.

This is exactly the kind of things that the Hahn-Banach theorem (see for instance [2, Theorem I.1]) can do! As $f \mapsto \sup f$ is real valued, positively homogeneous and subadditive, if we can prove that

$$\begin{aligned} \lambda: \quad & \mathcal{F} := \mathrm{Vect}(\mathcal{E}, c) \to \mathbb{R} \\ & f = \varphi \oplus \psi + sc \mapsto \int \varphi \, \mathrm{d}\mu + \int \psi \, \mathrm{d}\nu + sK \end{aligned}$$

satisfies for all $f \in \mathcal{F}$ the inequality $\lambda(f) \leq \sup f$, then the Hahn-Banach theorem allows to extend λ to the whole $C(X \times Y)$ while keeping the upper-bound $\lambda(f) \leq \sup f$, which provides the γ we are looking for.

So let us prove that for all $\varphi \in C(X)$, $\psi \in C(Y)$ and $s \in \mathbb{R}$

$$\int \varphi \, \mathrm{d}\mu + \int \psi \, \mathrm{d}\nu + sK \leq \sup\{\varphi \oplus \psi + sc\}.$$

<u>First case</u>: $s \geq 0$.

In this case, we just use the fact that $K \leq \int c \, \mathrm{d}\mu \otimes \nu$. Therefore,

$$\int \varphi \, \mathrm{d}\mu + \int \psi \, \mathrm{d}\nu + sK \leq \int \{\varphi \oplus \psi + sc\} \, \mathrm{d}\mu \otimes \nu \leq \sup\{\varphi \oplus \psi + sc\}.$$

<u>Second case</u>: $s < 0$.

Calling $t = -s$, we need to prove that

$$\frac{1}{t}\left(\int \varphi \, \mathrm{d}\mu + \int \psi \, \mathrm{d}\nu - \sup\{\varphi \oplus \psi - tc\}\right) \leq K.$$

But calling $\varphi' := (\varphi - \sup\{\varphi \oplus \psi - tc\})/t$ and $\psi' := \psi/t$, we have $\varphi' \oplus \psi' \leq c$. So by definition of K,

$$K \geq \int \varphi' \, \mathrm{d}\mu + \int \psi' \, \mathrm{d}\nu = \frac{1}{t}\left(\int \varphi \, \mathrm{d}\mu + \int \psi \, \mathrm{d}\nu - \sup\{\varphi \oplus \psi - tc\}\right),$$

as announced. □

6 Existence in the dual problem

In this section, we assume that X and Y are compact metric spaces, and that c is continuous. Let us set a few notations. We call

$$\begin{gathered} \mathrm{Comp} := \{(\varphi, \psi) \in C(X) \times C(Y) \text{ s.t. } \varphi \oplus \psi \leq c\}, \\ \forall \varphi \in C(X),\, \psi \in C(Y), \quad J(\varphi, \psi) := \int \varphi \, \mathrm{d}\mu + \int \psi \, \mathrm{d}\nu. \end{gathered}$$

Therefore, our dual problem rewrites

$$\sup_{(\varphi,\psi) \in \mathrm{Comp}} J(\varphi, \psi). \tag{8}$$

The main result that we want to show is the existence of optimizer φ and ψ in this problem. A notion that will be crucial towards this perspective is the c-transform of a test function. To give a precise definition, let us use the following notations: we write $\varphi \in \mathcal{F}_X$ provided φ is a function from X to $\mathbb{R} \cup \{-\infty\}$, and there exists $x \in X$ such that $\varphi(x) \in \mathbb{R}$. We define the similar notation $\mathcal{F}_Y$ for functions defined on Y.

Definition 17 (c-transform)**.** Let $\varphi \in \mathcal{F}_X$. The c-transform of φ is the function $\varphi^c : Y \to \mathbb{R}$ defined for all $y \in Y$ by

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

Similarly, let $\psi \in \mathcal{F}_Y$. The c-transform of ψ is the function $\psi^c : X \to \mathbb{R}$ defined for all $x \in X$ by

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

Remark 18. We use the same notation for c-transform of functions of x and y, so be cautious! We will not always specify that when proving results about functions of x, they also apply for functions of y.

Let us gather some information about c-transforms of test functions.

Proposition 19. *1. Let $\varphi \in \mathcal{F}_X$, then $\varphi^c \in C(Y)$. Similarly, let $\psi \in \mathcal{F}_Y$, then $\psi^c \in C(X)$.*

2. The sets

$$\{\varphi^c \,|\, \varphi \in \mathcal{F}_X\} \subset C(Y) \quad \textit{and} \quad \{\psi^c \,|\, \psi \in \mathcal{F}_Y\} \subset C(X)$$

are uniformly equicontinuous.

3. Given $\varphi \in C(X)$, $(\varphi, \varphi^c) \in \mathrm{Comp}$ and we have

$$\sup_{\psi \text{ s.t. } (\varphi,\psi) \in \mathrm{Comp}} J(\varphi, \psi) = J(\varphi, \varphi^c).$$

Similarly, given $\psi \in C(Y)$, $(\psi^c, \psi) \in \mathrm{Comp}$ and we have

$$\sup_{\varphi \text{ s.t. } (\varphi,\psi) \in \mathrm{Comp}} J(\varphi, \psi) = J(\psi^c, \psi).$$

Proof. The first point is a consequence of the second one. So let us prove that $\{\varphi^c \,|\, \varphi \in \mathcal{F}_X\}$ is uniformly equicontinuous. The second set is treated in the same way. Let $\varepsilon > 0$ and $\delta > 0$ such that for all $x, x' \in X$, for all $y, y' \in Y$ such that $d_X(x, x') + d_Y(y, y') \leq \delta$, $|c(x', y') - c(x, y)| \leq \varepsilon$. This is possible since $X \times Y$ is compact and c is continuous. Now for $y, y' \in Y$ such that $d_Y(y, y') \leq \delta$, we have

$$\varphi^c(y') - \varphi^c(y) = \inf_{x' \in X} \sup_{x \in X} c(x', y') - \varphi(x') - c(x, y) + \varphi(x) \leq \sup_{x \in X} c(x, y') - c(x, y) \leq \varepsilon,$$

where we chose $x' = x$ in the infimum to get an upper bound, and $d_Y(y, y') \leq \delta$ to get the last inequality. Reversing the roles of y and y', we get $|\varphi^c(y') - \varphi^c(y)| \leq \varepsilon$, and hence, the result.

It remains to prove the third point. For $\varphi \in C(X)$, $x \in X$ and $y \in Y$,

$$\varphi^c(y) \leq c(x, y) - \varphi(x),$$

so that $\varphi \oplus \varphi^c \leq c$ and by continuity of φ^c, $(\varphi, \varphi^c) \in \mathrm{Comp}$. The last thing to prove is that for all $(\varphi, \psi) \in \mathrm{Comp}$, $J(\varphi, \psi) \leq J(\varphi, \varphi^c)$. But whenever $(\varphi, \psi) \in \mathrm{Comp}$, for all $x \in X$ and $y \in Y$

$$\psi(y) \leq c(x, y) - \varphi(x).$$

Taking the infimum in x in the r.h.s., we find $\psi \leq \varphi^c$. So the conclusion follows from the monotonicity of $J(\varphi, \psi)$ w.r.t. ψ. □

Therefore, if $(\varphi, \psi) \in \mathrm{Comp}$, then (φ, φ^c) provides a better competitor for our maximization problem. We can keep going and consider $(\varphi^{cc}, \varphi^c)$, $(\varphi^{cc}, \varphi^{ccc})$,... Actually, this iteration stops thanks to the following lemma.

Lemma 20. *For all $\varphi \in \mathcal{F}_X$, $\varphi^{ccc} = \varphi^c$. Similarly, for all $\psi \in \mathcal{F}_Y$, $\psi^{ccc} = \psi^c$.*

Proof. The two ingredients for this proof is that on the one hand, for all φ, $\varphi^{cc} \geq \varphi$, and on the other hand, if $\varphi \leq \varphi'$, then $\varphi^c \geq \varphi'^c$. Indeed, with these two properties we know that $\varphi^{ccc} = (\varphi^c)^{cc} \geq \varphi^c$ by the first point, and $\varphi^{ccc} = (\varphi^{cc})^c \leq \varphi^c$ by the first point together with $\varphi^{cc} \geq \varphi$.

The second point is a consequence of the minus sign in the definition of φ^c. Let us prove the first one. We have for all $x \in X$,

$$\varphi^{cc}(x) = \inf_{y\in Y} \sup_{x'\in X} c(x,y) - c(x',y) + \varphi(x') \geq \varphi(x),$$

by chosing $x' = x$ in the sup to get a lower bound. □

As a consequence, it is often useful to give a name to c-transforms of functions.

Definition 21 (c-concave functions)**.** A function $\varphi : X \to \mathbb{R}$ is said to be c-concave if it is the c-transform of a function $\psi \in \mathcal{F}_Y$. Similarly, a function $\psi : Y \to \mathbb{R}$ is said to be c-concave if it is the c-transform of a function $\varphi \in \mathcal{F}_Y$.

Remark 22. Due to the previous results, when X and Y are compact and c is continuous, if φ is c-concave, then φ is continuous, and is the c-transform of a continuous function.

Proposition 23. *A function φ is c-concave if and only if $\varphi^{cc} = \varphi$. In particular, if φ is c-concave, calling $\psi = \varphi^c$, we also have $\varphi = \psi^c$. In this case, we call the pair (φ, ψ) a pair of c-concave conjugates.*

We have all the ingredients needed to prove

Proposition 24. *Let X, Y be compact metric spaces, and $c : X \times Y \to \mathbb{R}$ be continuous. The maximization problem* (8) *admit some maximizers $(\bar\varphi, \bar\psi) \in \mathrm{Comp}$. Moreover, some of these maximizers are pairs of convex conjugates.*

Proof. Let $(\varphi_n, \psi_n) \in \mathrm{Comp}^{\mathbb{N}}$ be a maximizing sequence (this is possible since Comp is nonempty: choose $\varphi \equiv -M$ and $\psi \equiv 0$ for M sufficiently large). For all n, up to replacing (φ_n, ψ_n) by $(\varphi_n^{cc}, \varphi_n^c)$, which increases the objective functional, we can assume that $\varphi_n \in \{\psi^c \,|\, \psi \in \mathcal{F}_Y\}$ and $\psi_n \in \{\varphi^c \,|\, \varphi \in \mathcal{F}_X\}$, and hence that both (φ_n) and (ψ_n) are uniformly equicontinuous sequences.

Moreover, let us replace φ_n by $\varphi_n - \inf \varphi_n$ and ψ_n by $\psi_n + \inf \varphi_n$. This does not change the value of J, keeps the property $(\varphi_n, \psi_n) \in \mathrm{Comp}$, keeps the property of φ_n and ψ_n to be c-tranforms of each other and keeps the sequences to be uniformly equicontinuous. By this trick, we can assume that $\inf \varphi_n = 0$. By uniform equicontinuity, (φ_n) is therefore uniformly bounded, and by definition of the c-tranform, (ψ_n) is as well.

So the Ascoli-Arzela theorem applies: up to extraction, (φ_n) converges uniformly towards a certain function $\bar\varphi \in C(X)$, and (ψ_n) converges uniformly towards a function $\bar\psi \in C(Y)$. Passing to the pointwise limit in $\varphi_n \oplus \psi_n \leq c$ leads to $(\bar\varphi, \bar\psi) \in \mathrm{Comp}$, and we have

$$J(\bar\varphi, \bar\psi) = \lim_{n\to+\infty} J(\varphi_n, \psi_n) = \sup_{(\varphi,\psi)\in\mathrm{Comp}} J(\varphi, \psi),$$

and the optimization problem (8) admits optimizers.

Finally, up to replacing $(\bar\varphi, \bar\psi)$ by $(\bar\varphi^{cc}, \bar\varphi^c)$, we can assume that $\bar\varphi = \bar\psi^c$ and $\bar\psi = \bar\varphi^c$. □

7 Necessary condition for optimality

Let us still work in the case when X and Y are compact metric spaces, and c is continuous. By Propositions 9 and 24, and Theorem 15, we know that there is existence in both the primal and dual problems and that the corresponding values coincide. This is exactly the kind of situations where we can derive necessary and sufficient optimality conditions.

Proposition 25 (Optimality conditions for the Kantorovic problem)**.** *Let $\gamma \in \Pi(\mu, \nu)$. The following assertions are equivalent:*

1. *γ is a solution of the Kantorovic problem* (2).
2. *There exists $\varphi \in \mathcal{L}^1(\mu)$ and $\psi \in \mathcal{L}^1(\nu)$ such that $\varphi \oplus \psi \leq c$ everywhere, and for γ-almost all (x, y),*

$$\varphi(x) + \psi(y) = c(x, y).$$

3. *There exists a c-concave function $\varphi \in C(X)$ such that for all $(x, y) \in \mathrm{Supp}(\gamma)$*

$$\varphi(x) + \varphi^c(y) = c(x, y).$$

Remark 26. Notice that in case γ is a solution of the Kantorovic problem, *any* maximizer of the dual problem (8) satisfies point 2 and that *any* c-concave maximizer of the dual problem satisfies point 3.

Proof. $3 \Rightarrow 2$ is obvious setting $\psi := \varphi^c$, $2 \Rightarrow 1$ has already been done in Proposition 13. It remains to prove $1 \Rightarrow 3$. By Proposition 24, the dual problem (8) admits as a maximizer a pair of c-concave functions $(\varphi, \varphi^c) \in C(X) \times C(Y)$. By Theorem 15, there is no duality gap, that is

$$\int c \,\mathrm{d}\gamma = \int \varphi \,\mathrm{d}\mu + \int \varphi^c \,\mathrm{d}\nu.$$

So as proven in Proposition 13, for γ almost all (x, y), $\varphi(x) + \varphi^c(y) = c(x, y)$. As both c and $\varphi \oplus \psi$ are continuous, the set of points where they coincide is closed, and hence contains the support of γ. □

8 c-cyclical monotonicity and 1 dimensional case

This optimality condition has an important geometric consequence on the support of solutions of optimal transport problems: they are c-cyclically monotone.

Definition 27 (c-cyclical monotonicity)**.** A subset $S \subset X \times Y$ is said to be c-cyclically monotone whenever for all $N \in \mathbb{N}^*$ and $(x_1, y_1), \dots, (x_N, y_N) \in S$,

$$c(x_1, y_1) + \cdots + c(x_N, y_N) \leq c(x_1, y_2) + \cdots + c(x_{N-1}, y_N) + c(x_N, y_1).$$

Remark 28. This is the same as saying that for all $N \in \mathbb{N}^*$, $(x_1, y_1), \dots, (x_N, y_N) \in S$ and $\sigma \in \mathfrak{S}_N$,

$$c(x_1, y_1) + \cdots + c(x_N, y_N) \leq c(x_1, y_{\sigma(1)}) + \cdots + c(x_N, y_{\sigma(N)}).$$

(Just decompose σ as the composition of disjoint supports cycles.)

Proposition 29. *Let $\gamma \in \Pi(\mu, \nu)$ be a solution of the Kantorovic problem* (2)*. Then* $\mathrm{Supp}(\gamma)$ *is c-cyclically monotone.*

Remark 30. We will prove this result using duality, but it could be proven directly (see [3, Theorem 2.4.3]). It could be proven that this necessary condition is also sufficient, see [3, Corollary 2.6.8].

Proof. Let $N \in \mathbb{N}^*$ and $(x_1, y_1), \dots, (x_N, y_N) \in \mathrm{Supp}(\gamma)$. Let φ be given by point 3 of Proposition 25. We have:

$$\begin{aligned} c(x_1, y_1) + \cdots + c(x_N, y_N) &= \varphi(x_1) + \varphi^c(y_1) + \cdots + \varphi(x_N) + \varphi^c(y_N) \\ &= \varphi(x_1) + \varphi^c(y_2) + \cdots + \varphi(x_{N-1}) + \varphi^c(y_N) + \varphi(x_N) + \varphi^c(y_1) \\ &\leq c(x_1, y_2) + \cdots + c(x_{N-1}, y_N) + c(x_N, y_1). \end{aligned}$$

□

As a consequence of c-cyclical monotonicity, we can give results about optimal transport in one space dimension. We leave this as an exercise.

Exercice 31. Let $h : \mathbb{R} \to \mathbb{R}$ be a strictly convex function and $\mu, \nu \in \mathcal{P}(\mathbb{R})$ with compact support (we could do better but let us keep to the compact case). Set $c : (x, y) \in \mathbb{R}^2 \mapsto h(y - x)$, and let γ be a solution of the optimal transport problem between μ and ν with cost c.

1. Take $(x, y), (x', y') \in \mathrm{Supp}(\gamma)$ and assume that $x < x'$. Prove using c-cyclical monotonicity that $y \geq y'$.

2. Assume in addition for the rest of the exercise that μ has no atom. Take $x, y, y' \in \mathbb{R}$ with $y < y'$, $(x, y) \in \mathrm{Supp}(\gamma)$ and $(x, y') \in \mathrm{Supp}(\gamma)$. Prove with question 1 that $\gamma(\mathbb{R} \times (y, y')) = 0$. Conclude that for all $x \in \mathbb{R}$, $\{y \in \mathbb{R} \text{ s.t. } (x, y) \in \mathrm{Supp}(\gamma)\}$ has cardinality 0, 1 or 2.

3. Show that those x whose corresponding set of ys defined in the previous question has cardinality 2 are at most countable.

4. Conclude that γ is concentrated on the graph of a nondecreasing function f.

5. Using the convexity of the set of solutions, prove that γ is unique.

9 The quadratic case, proof of the Brenier theorem

We will prove a slight modification of the Brenier Theorem 11 replacing the assumption (3) by a compact support for μ and ν in order to be able to use Proposition 25. That is:

Theorem 32. *Let us set $X = Y = \overline{B}_R \subset \mathbb{R}^d$, for some $R > 0$ and $d \in \mathbb{N}^*$, and for all $x, y \in \mathbb{R}^d$, $c(x, y) = \frac{1}{2}|y - x|^2$. Let $\mu \in \mathcal{P}(\overline{B}_R)$ and $\nu \in \mathcal{P}(\overline{B}_R)$.*

Under the assumption that $\mu \ll \mathrm{d}x$, the Kantorovic problem (2) *admits a unique solution γ, and there exists a* convex *function $\alpha : \overline{B}_R \to \mathbb{R}$ such that $\gamma = (\mathrm{Id}, \nabla\alpha)_{\#}\mu$. In particular, $C_M(\mu, \nu) = C_K(\mu, \nu)$ and the Monge problem* (1) *admits a solution, unique up to a μ-negligible set.*

In the proof, we will make a crucial use of:

Theorem 33. *Let $U \subset \mathbb{R}^d$ be a open set and $\alpha : U \to \mathbb{R}$ convex, then α is $\mathrm{d}x$-almost everywhere differentiable on U.*

Proof. Let γ be a solution of the Kantorovic problem (2) as given by Proposition 9. We first show that there exists a convex function α on $\overline{B}_R$ such that $\gamma = (\mathrm{Id}, \nabla\alpha)_{\#}\mu$.

By Proposition 25, there exists $\varphi \in C(X)$ a c-concave function such that for every $(x, y) \in \mathrm{Supp}(\gamma)$,

$$\varphi(x) + \varphi^c(y) = \frac{1}{2}|y - x|^2.$$

Let us set for all $x, y \in \overline{B}_R$

$$\alpha(x) := \frac{|x|^2}{2} - \varphi(x) \quad \text{and} \quad \beta(y) := \frac{|y|^2}{2} - \varphi^c(y).$$

We have for all $x, y \in \overline{B}_R$:

$$\begin{aligned}\alpha(x) + \beta(y) &= \frac{|x|^2}{2} + \frac{|y|^2}{2} - \varphi(x) - \varphi^c(y) \\ &\geq \frac{|x|^2}{2} + \frac{|y|^2}{2} - \frac{1}{2}|y - x|^2 = \langle x, y\rangle,\end{aligned}$$

so that

$$\alpha(x) + \beta(y) \geq \langle x, y\rangle$$

with an equality sign on the support of γ. This is reminiscent of the theory of convex functions: (α, β) are convex conjugate functions. For instance, to see that α is convex, observe that as $\varphi^{cc} = \varphi$,

$$\begin{aligned}\alpha(x) = \frac{|x|^2}{2} - \varphi(x) &= \frac{|x|^2}{2} - \inf_{y \in \overline{B}_R} \frac{|y-x|^2}{2} - \varphi^c(y) \\ &= \sup_{y \in \overline{B}_R} \langle x, y \rangle - \left(\frac{|y|^2}{2} - \varphi^c(y) \right) \\ &= \sup_{y \in \overline{B}_R} \langle x, y \rangle - \beta(y),\end{aligned}$$

which is convex as a supremum of affine functions.

Let us prove that for γ almost all $(x, y) \in \overline{B}_R \times \overline{B}_R$, $x \in B_R$, α is differentiable at x and $y = \nabla\alpha(x)$. As α is convex and $\mu \ll \mathrm{d}x$, $\mu(B_R) = 1$ and α is differentiable μ-almost everywhere on B_R. Let $\mathcal{D} \subset B_R$ be measurable, such that $\mu(\mathcal{D}) = 1$, and for all $x \in \mathcal{D}$, α is differentiable at x. Let us call $\mathcal{S} := \mathrm{Supp}(\gamma) \cap (\mathcal{D} \times \overline{B}_R)$. Of course, $\gamma(\mathcal{S}) = 1$ (this is because $\gamma(\mathrm{Supp}(\gamma)) = 1$ and $\gamma(\mathcal{D} \times \overline{B}_R) = \mu(\mathcal{D}) = 1$). Moreover, for *all* $(x, y) \in \mathcal{S}$, $x \in B_R$, α is differentiable at x, and the function

$$x' \in B_R \mapsto \alpha(x') - \langle x', y \rangle$$

achieves its maximum $\beta(y)$ at $x' = x$. So differentiating at $x' = x$, we find

$$y = \nabla\alpha(x).$$

In conclusion, for all $f \in C(\overline{B}_R \times \overline{B}_R)$,

$$\begin{aligned}\int f \,\mathrm{d}\gamma = \int_{\mathcal{S}} f \,\mathrm{d}\gamma &= \int_{\mathcal{S}} f(x, \nabla\alpha(x)) \,\mathrm{d}\gamma(x, y) \\ &= \int_{\mathcal{D} \times \overline{B}_R} f(x, \nabla\alpha(x)) \,\mathrm{d}\gamma(x, y) \\ &= \int_{\mathcal{D}} f(x, \nabla\alpha(x)) \,\mathrm{d}\mu(x) = \int f(x, \nabla\alpha(x)) \,\mathrm{d}\mu(x) \\ &= \int f \,\mathrm{d}\,(\mathrm{Id}, \nabla\alpha)_{\#}\mu(x, y).\end{aligned}$$

We conclude that γ is of Monge type, and hence that $C_K(\mu, \nu) = C_M(\mu, \nu)$.

It remains to prove uniqueness. Assume that the Kantorovic problem admits two solutions γ_1 and γ_2. Then, $\gamma := (\gamma_1 + \gamma_2)/2$ is a solution as well. By the first part of the proof, there exists φ_1, φ_2 and φ three convex functions such that

$$\begin{gathered}\text{for } \gamma_1\text{-almost all } (x, y),\ \alpha_1 \text{ is differentiable at } x \text{ and } y = \nabla\alpha_1(x), \\ \text{for } \gamma_2\text{-almost all } (x, y),\ \alpha_2 \text{ is differentiable at } x \text{ and } y = \nabla\alpha_2(x), \\ \text{for } \gamma\text{-almost all } (x, y),\ \alpha \text{ is differentiable at } x \text{ and } y = \nabla\alpha(x).\end{gathered}$$

But as $\gamma_1 \ll \gamma$, the third point is also valid for γ_1-almost all (x, y), that is, for γ_1-almost all (x, y), α_1 and α are differentiable at x and $y = \nabla\alpha_1(x) = \nabla\alpha(x)$. In particular, for γ_1-almost all (x, y), α_1 and α are differentiable at x and $\nabla\alpha_1(x) = \nabla\alpha(x)$. This assertion does not depend on y, so it is true for μ-almost all x. Doing the same reasoning with γ_2, we find that for μ-almost all x, α_1 and α_2 are differentiable at x, and $\nabla\alpha_1(x) = \nabla\alpha_2(x)$. Uniqueness for the Kantorovic problem follows. Uniqueness for the Monge problem follows the same lines, observing that if T is a solution of the Monge problem, then $(\mathrm{Id}, T)_{\#}\mu$ is a solution of the Kantorovic problem (as $C_K(\mu, \nu) = C_M(\mu, \nu)$). Therefore, our argument shows that $T = \nabla\alpha$ μ-almost everywhere.

□

Remark 34. In the proof, we saw that if φ is c-concave with $c(x, y) := |y - x|^2/2$, then $\alpha : x \mapsto |x|^2/2 - \varphi(x)$ is convex. Actually, this is an equivalence, and $\beta : y \mapsto |y|^2/2 - \varphi^c(y)$ is the *Legendre transform* of α. This draws a link between c-transform and Legendre transform.

Another way to see this link is that using $c(x, y) := |y - x|^2/2$ or $c(x, y) := -\langle x, y\rangle$ does not affect the solutions of optimal transport. But with $c(x, y) := -\langle x, y\rangle$, the c-transform is exactly the Legendre transform of concave functions, and therefore, c-concavity coincide with standard concavity.

See [4, Section 1.6.1] for more material about this.

10 The Monge-Ampère equation

The purpose of this section is to draw a formal link between optimal transport problems and the Monge-Ampère equation. Let us place ourselves in the context of the Monge problem on $\mathbb{R}^d$: μ and ν are probability measures on $\mathbb{R}^d$. Assume in addition that $\mu \ll \mathrm{d}x$ and $\nu \ll \mathrm{d}y$. In this case, there exist f and g two nonnegative functions on $\mathbb{R}^d$ such that $\mathrm{d}\mu(x) = f(x)\,\mathrm{d}x$ and $\mathrm{d}\mu(y) = g(y)\,\mathrm{d}y$. Assume that they are continuous and positive. Let T be a transport map from μ to ν. Assuming that T is a diffeomorphism, T^{-1} is a transport map from ν to μ. Therefore, we have for all $\varphi \in C_b(\mathbb{R}^d)$

$$\int \varphi(x) f(x)\,\mathrm{d}x = \int \varphi(T^{-1}(y)) g(y)\,\mathrm{d}y.$$

But assuming in addition that $\det \mathrm{D}T > 0$, we can apply the change of variables $y = T(x)$ (and hence $\mathrm{d}y = \det \mathrm{D}T(x)\,\mathrm{d}x$), and find

$$\int \varphi(x) f(x)\,\mathrm{d}x = \int \varphi(x) g(T(x)) \det \mathrm{D}T(x)\,\mathrm{d}x,$$

and hence, for all x,

$$\det \mathrm{D}T(x) = \frac{f(x)}{g(T(x))}.$$

Now assume that T is the solution of the Monge problem with quadratic cost: in virtue of the Brenier theorem, there exists a convex function α such that $T\nabla\alpha$. Therefore, our identity rewrites

$$\det \mathrm{D}^2\alpha(x) = \frac{f(x)}{g(\nabla\alpha(x))}.$$

This is the Monge-Ampère equation, a fully nonlinear elliptic equation. Studying the regularity of optimal transport maps for the quadratic cost (that is, showing that whenever f and g are smooth, α is smooth) reduces to study the regularity of this elliptic PDE.

11 Bibliographical notes

We decided here to present results about the Kantorovic problem relying on duality. This is a robust approach that can be used for a large variety of contexts. To get finer results (in particular the generalization in Polish spaces), we refer to [5, Chapter 1].

Another approach leading to the same results in a more efficient but less robust way is to mainly rely on c-cyclical monotonicity. Once proven that the support of a solution γ of the Kantorovic problem is c-cyclically monotone, it is possible to build by hand a c-concave function φ such that for every (x, y) on the support of γ, $\varphi(x) + \varphi^c(y) = c(x, y)$. This is the approach chosen in [3, Chapter 2].

Let us also mention the book [4] which gives an overview of both approaches, with a lot of examples and illustrations of the results.

References

[1] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

[2] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations.* Universitext. Springer, New York, 2011.

[3] A. Figalli and F. Glaudo. *An invitation to optimal transport, Wasserstein distances, and gradient flows.* EMS Textbooks in Mathematics. EMS Press, Berlin, [2021] ©2021.

[4] F. Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 2015.

[5] C. Villani. *Topics in optimal transportation.* Number 58. American Mathematical Society, 2003.