Optimal transport for economists

A. Blanchet

April 2, 2025

Contents

Chapter 1

Introduction

1.1 Reminders

1.1.1 Probability

Definition 1 (σ-algebra)**.** *Let X be a set. We call* σ-algebra *on X, a set $\mathcal{A}$ of all the parties of X satisfying*

- *$\mathcal{A}$ is not empty,*
- *$\mathcal{A}$ is stable by complement,*
- *$\mathcal{A}$ is stable by countable unions.*

Exercice 1. 1. Describe possible σ-algebra of $X = \{a, b, c, d\}$.

2. Describe possible σ-algebra of a set X.

▶

1. A possible σ-algebra is $\mathcal{A} = \{\emptyset, \{a\}, \{b, c, d\}, X\}$. Another is $\mathcal{A} = \{\emptyset, \{a, b\}, \{c, d\}, X\}$.
2. A possible σ-algebra is $\mathcal{A} = \{\emptyset, X\}$. Another is $\mathcal{A} = \mathcal{P}(X)$ where $\mathcal{P}(X)$ represents all the parties of X (the set of all the subset of X).

Definition 2 (Borel sets)**.** *A* Borel σ-algebra *is the smallest σ-algebra Σ which contains all the open sets. The couple (X, Σ) is called a Borel space.*

A Borel set is an element of a Borel σ-algebra.

Exercice 2. Prove that

1. $\mathbb{R}$ is Borel σ-algebra.
2. $\mathbb{R}^N$ is a Borel σ-algebra.

▶

1. Spanned by $\{(a, +\infty), a \in \mathbb{R}\}$, by $\{[a, +\infty), a \in \mathbb{R}\}$, $\{[a, b), (a, b) \in \mathbb{R}^2\}$
2. Spanned by the open Euclidian balls, the closed Euclidian balls, the cuboid, etc...

Definition 3 (Measure)**.** *Let X be a set and Σ a σ-algebra over X. A set function $\mu : \Sigma \to \bar{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ is called a* measure *if the following hold:*

- *Non-negativity:* $\forall E \in \Sigma, \quad \mu(E) \geq 0$.
- $\mu(\emptyset) = 0$.

- *Countable additivity (or σ-additivity): For all countable collections $\{E_k\}_{k=1}^{\infty}$ of pairwise disjoint sets in Σ*

$$\mu\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} \mu(E_k).$$

A triple (X, Σ, μ) is called a measure space.

A probability measure *is a measure with total measure one that is, $\mu(X) = 1$.*

Definition 4 (Borel measure)**.** *Let X be a set and Σ a σ-algebra over X. A* Borel measure *is any measure μ defined on the σ-algebra of Borel sets.*

Example 1. Let μ be the measure which to any interval $(a, b]$ assigns $\mu((a, b]) = |b - a|$. This measure is often called «the» Borel measure on $\mathbb{R}$.

Definition 5 (Borel map)**.** *Let (X, Σ) and (Y, Σ') be Borel spaces. A function $f : (X, \Sigma) \to (Y, \Sigma')$ is called a* Borel map *if it is a measurable function i.e. for every $E \in \Sigma'$ the pre-image of E under f is in Σ; that is, for all $E \in \Sigma'$,*

$$f^{-1}(E) := \{x \in X \mid f(x) \in E\} \in \Sigma.$$

Example 2. Continuous functions are Borel maps (but not the reverse).

1.1.2 Weak and weak-* topologies

Definition 6 (Dual)**.** *Let E be a vector space. The continous dual space of E, denoted E^* is the set of the continous linear forms on E.*

Exercice 3. Determine $(\mathbb{R}^3)^*$.

▶ It is the set of all the forms:

$$(x, y, z) \mapsto ax + by + cz$$

where a, b and c are given real numbers.

Definition 7 (Strong convergence)**.** *Let $(X, \|\cdot\|)$ a normed vector space. A sequence $(x_n)_n$ in X* converges strongly *to $x \in X$ if*

$$\lim_{n\to\infty} \|x_n - x\| = 0\ .$$

Exercice 4. For all $p \in [1, \infty)$ and a general index set I, we define

$$l^p := \left\{(x_n)_{n\in I}\ :\ \sum_{n\in I} |x_n|^p < \infty\right\}$$

Let p' be such that $1/p + 1/p' = 1$. We also define

$$\begin{array}{rccl} H : & l^{p'} \times l^p & \to & \mathbb{R} \\ & (x_n), (y_n) & \mapsto & \sum_{n\in\mathbb{N}} x_n\, y_n \end{array}$$

1. Prove that H is bilinear and continuous.
2. Prove that we can define a linear and continous map $a_H : l^{p'} \to (l^p)^*$ such as
$$\langle a_H(y), x\rangle = H(y, x)$$
3. Prove that a_H is an isomorphism so that $l^{p'}$ identifies to $(l^p)^*$.

▶

1. Obvious.
2. By Riesz' representation theorem: For every continuous linear functional $\varphi \in H^*$, there exists a unique vector $f_\varphi \in H$, called the Riesz representation of φ, such that

$$\varphi(x) = \langle x, f_\varphi \rangle \qquad \text{for all } x \in H.$$

3. Obvious by uniqueness.

Exercice 5. In the space $(l^2, \|\cdot\|_2)$, consider the sequence $(e_n)_n$ of general term $e_n = (\delta_n(i))_i = (0, \cdots, 0, 1, 0, \cdots, 0, \cdots)$.

1. Compute $\|e_n - e_m\|_2$.
2. Prove that the sequence $(e_n)_n$ is not a Cauchy sequence.
3. Prove that the sequence $(e_n)_n$ does not converge strongly.

▶

1. $\|e_n - e_m\| = \sqrt{2}$.
2. By definition of a Cauchy sequence.
3. Any converging sequence is a Cauchy sequence.

Definition 8 (Weak convergence)**.** *Let $(X, \|\cdot\|)$ be a normed vector space. A sequence $(x_n)_n$ in X* converges weakly *to $x \in X$ if*

$$\forall f \in X^*, \qquad \lim_{n\to\infty} \langle f, x_n \rangle = \langle f, x \rangle \ .$$

We denote $x_n \rightharpoonup x$.

Exercice 6. In the previous exercise prove that $e_n \rightharpoonup 0$.

▶ For every $(x_n) \in l^2$,

$$\langle x, e_n \rangle = x_n \to 0$$

Definition 9. *Let X be a normed vector space. A sequence $(f_n)_n$ weak*-converges to $f \in X^*$ if*

$$\forall x \in X, \qquad \lim_{n\to\infty} \langle f_n, x \rangle = \langle f, x \rangle \ .$$

Exercice 7. Let

$$c_0 := \{(x_n) \in l^\infty \ : \ x_n \to 0\} \ .$$

Consider a sequence $(e_n)_n$ in $c_0^* \sim l^1$ defined by $e_n = (\delta_n(i))_i = (0, \cdots, 0, 1, 0, \cdots, 0, \cdots)$

1. Prove that $(e_n)_n$ weak*-converges to 0.
2. Prove that $(e_n)_n$ does not weak-converge to 0.

▶

1. We have, for all $(x_i)_i$ in c_0

$$\langle e_n, (x_i)_i \rangle = \sum_{i\in\mathbb{N}} e_i \, x_i = x_n \to 0 \ .$$

Hence $(e_n)_n$ weakly*-converges to 0.

2. Now consider $\phi = (\phi_i)_i \in l^\infty = (l^1)^*$. We have

$$\langle \phi, e_n \rangle = \sum_{i\in\mathbb{N}} \phi_i\, e_i = \phi_n \ .$$

But there is no reason why ϕ_n would go to 0. Take for instance $\phi = (1, \cdots, 1, \cdots) \in l^\infty$.

Theorem 1 (Banach-Alaoglu)**.** *Let E be a vector space. The closed ball $\{f \in E^* : \|f\| \le 1\}$ is compact for the weak*-topology.*

Cf. [6][Theorem 3.15, p. 42].

Theorem 2 (Kakutani)**.** *Let E be a vector space. The closed ball $\{x \in E : \|x\| \le 1\}$ is compact for the weak-topology if and only if $(E^*)^* = E$.*

Cf. [6][Theorem 3.16, p. 44].

Hence, from any sequence bounded in L^p for $p \in (1, \infty)$ we can extract a subsequence weakly converging in L^p. While from any sequence bounded in L^∞ we can extract a subsequence weakly-* converging in L^∞.

Note this kind of result cannot be true for L^1 as prove the following exercise:

Exercice 8. Consider the sequence $(g_n)_n$ defined in $[0, 1]$ by

$$g_n(x) := \begin{cases} -2n^2x + 2n & \text{if } x \in [0, 1/n) \\ 0 & \text{otherwise} \end{cases}$$

1. Prove that $\|g_n\|_1 = 1$.

2. Prove that $(g_n)_n$ converges neither weakly no weakly-* in $\mathcal{L}^1(0, 1)$.

We can decide to work in a space which is much bigger that the space of L^1 functions :

Definition 10 (Finite signed measure)**.** *Let X be a metric space. A map which associate to every Boret set $A \subset X$ a value $\lambda(A) \in \mathbb{R}$ such that, for every countable disjoint union we have*

$$\sum_i |\lambda(A_i)| < \infty \qquad \textit{and} \qquad \sum_i \cup_i \lambda(A_i) = \sum_i \lambda(A_i)$$

is called a finite signed measure*. We denote $\mathcal{M}(X)$ the set of finite measures on X.*

Theorem 3. *Let X be a compact set. Consider the space $E := \mathcal{C}(X)$ of continuous functions on X endowed with the sup-norm. E^* is isomorphic to $\mathcal{M}(X)$ endowed with the norm $\|\lambda\| = |\lambda|(X)$.*

Remark 1. If X is not bounded, we may consider the space of continous function which vanish at infinity, which dual is $\mathcal{M}(X)$. The convergence in finite signed measures $\mathcal{M}(X)$ in the duality $\mathcal{C}_0(X)$ is the weak* convergence. There is another notion of convergence, called the narrow convergence, in the duality $\mathcal{C}_b(X)$, of bounded continuous functions:

$$\forall \phi \in \mathcal{C}_b(X), \qquad \int_X \phi \,\mathrm{d}\mu_n \to \int_X \phi \,\mathrm{d}\mu \ .$$

But all those notions coincide when X is compact.

1.2 The Monge and the Kantorovich problems

In 1781, the french mathematician Gaspard Monge, first considered the problem of «remblais et déblais» which asks what is the most efficent (that is work minimizing) way to move a pile of soil or rubble to an excavation or fill. Imagine that the soil initially occupies the bounded region $A \subset \mathbb{R}^3$ and that the excavation is the region B, assume also that A and B have the same volume. One then looks for a map $T : A \to B$, (here $T(x) \in B$ represents the destination of the element of mass initially located at $x \in A$), the total work involved is

$$\int_A |x - T(x)| \, \mathrm{d}x$$

and one has to minimize it in the set of volume preserving maps $T : A \to B$. It is this constraint (of incompressibility) that makes the problem difficult (and in fact, the first rigorous existence proofs for a minimizer were given in the mid 90's!). We will come back to the original Monge problem in more details.

More generally, assume that X and Y are two compact sets (to make things as simple as possible) metric spaces and that μ and ν are two Borel measures with the same total mass (which we shall of course normalize to 1) and that we are also given a continuous transportation cost function $c : X \times Y \to \mathbb{R}$. Transport maps are then maps that fulfill some mass conservation requirement that is naturally defined as follows:

Definition 11. *Let $T : X \to Y$ be a Borel map. The* push-forward *or (image measure) of μ through T is the Borel measure, denoted $T_{\#}\mu$ defined on Y by*

$$\forall \varphi \in \mathcal{C}(Y), \qquad \int_Y \varphi(y) \, \mathrm{d}T_{\#}\mu(y) = \int_X \varphi(T(x)) \, \mathrm{d}\mu(x)$$

A Borel map $T : X \to Y$ is said to be a transport map (from μ to ν) if $T_{\#}\mu = \nu$.

Remark 2. The map T is a transport from μ to ν is equivalent to

$$\forall \varphi \in \mathcal{C}(Y), \qquad \int_Y \varphi(y) \, \mathrm{d}\nu(y) = \int_X \varphi(T(x)) \, \mathrm{d}\mu(x)$$

Remark 3. The image measure $T_{\#}\mu$ can equivalently be defined by

$$T_{\#}\mu(B) = \mu\left(T^{-1}(B)\right), \qquad \text{for all Borel subest } B \text{ of } Y.$$

Definition 12. *Monge's problem consists in finding a cost minimising transport from μ to ν, it reads:*

$$\inf_{T \,:\, T_{\#}\mu = \nu} \int_X c(x, T(x)) \, \mathrm{d}\mu(x) \, . \tag{MP}$$

If a solution to this problem exists it is called an optimal transport map.

We should now remark, that Monge's problem presents serious difficulties and in the first place the fact that there may be no transport map : if $\mu = \delta_0$ then it is impossible to transport μ on the Lebesgue measure! This example is somehow extreme and can be ruled out when one assumes that μ is nonatomic. One should also remark that the Monge's formulation is rather rigid in the sense that it requires that all the mass that is at x should be associated to the same target $T(x)$.

In the 1940's, Kantorovich proposed a relaxed formulation that allows mass splitting. More precisely, he introduced the problem which is by now known as the (Monge-)Kantorovich problem and reads as:

$$\inf_{\gamma \in \Pi(\mu,\nu)} \iint_{X \times Y} c(x, y) \, \mathrm{d}\gamma(x, y) \tag{KP}$$

where $\Pi(\mu, \nu)$ is the set of transport plans *i.e.* the set of Borel probability measures on $X \times Y$ which have μ and ν as marginals:

$$\gamma(A \times Y) = \mu(A);, \quad \gamma(X \times B) = \nu(B), \qquad \text{for every Borel set } A \subset X, \text{ and } B \subset Y.$$

Or equivalently

$$\forall(\varphi,\psi)\in\mathcal{C}(X)\times\mathcal{C}(Y),\qquad \iint_{X\times Y}(\varphi(x)+\psi(y))\,\mathrm{d}\gamma(x,y)=\int_X\varphi(x)\,\mathrm{d}\mu(x)+\int_Y\psi(y)\,\mathrm{d}\nu(y)\,.$$

Exercice 9. Prove that the set $\Pi(\mu,\nu)$ is never empty.

▶ For instance $\mu\otimes\nu$ belongs to $\Pi(\mu,\nu)$ where

$$(\mu\otimes\nu)(A\times B)=\mu(A)\,\nu(B),\qquad \text{for every Borel set } A\subset X, \text{ and } B\subset Y.$$

Kantorovich's problem is much simpler. We have

Theorem 4. *Let $c: X\times Y\to\mathbb{R}$ be a continous map. The Kantorovich problem admits a solution. Such solutions are called* optimal transport plans*.*

Proof.

Exercice 10. 1. Prove that

$$F:\gamma\mapsto\iint_{X\times Y}c(x,y)\,\mathrm{d}\gamma(x,y)$$

is linear.

2. Prove that $\Pi(\mu,\nu)$ is weakly-* compact.

3. Prove that F is continuous for the weak-* topology.

▶

1. It is obvious.

2. Obvious as the set of probability measures on a compact set is compact (*i.e.* every sequence of points in X has a convergent subsequence converging to a point in X) and for all $(\varphi,\psi)\in\mathcal{C}(X)\times\mathcal{C}(Y)$,

$$\int_X\varphi(x)\,\mathrm{d}\mu(x)+\int_Y\psi(y)\,\mathrm{d}\nu(y)=\iint_{X\times Y}(\varphi(x)+\psi(y))\,\mathrm{d}\gamma_n(x,y)$$
$$\to\iint_{X\times Y}(\varphi(x)+\psi(y))\,\mathrm{d}\gamma(x,y)\,.$$

3. Since c is continuous on a compact set, which is the predual of the set of the probability measures as $\mathcal{P}(X)^*\sim\mathcal{C}_0(X)$, we have

$$\langle c,\gamma_n\rangle\to\langle c,\gamma\rangle$$

By Weierstrass' theorem there exists a minimum. □

Existence of transport plans is therefore a straightforward fact but it does not say much about existence of optimal transport maps in general. However, let us remark that if T is a transport map then it induces a transport plan γ_T by $\gamma_T:=(\mathrm{id},T)_{\#}\mu$ *i.e.*

$$\iint_{X\times Y}\varphi(x,y)\,\mathrm{d}\gamma_T(x,y)=\int_X\varphi\left(x,T(x)\right)\,\mathrm{d}\mu(x),\qquad\forall\varphi\in\mathcal{C}(X\times Y)\,.$$

As for the cost one has

$$\int_X c\left(x,T(x)\right)\,\mathrm{d}\mu(x)=\iint_{X\times Y}c(x,y)\,\mathrm{d}\gamma_T(x,y)\,.$$

This proves that the minimum in the Kantorovich problem is smaller than the infimum of the Monge problem. It also means that if we are lucky enough to find an optimal plan that is of the form γ_T (which roughly speaking means that it is supported by the graph of T) then T is actually an optimal transport map.

A related question is whether the Kantorovich problem has a minimum. The answer is positive when μ is non-atomic:

Theorem 5. *If μ is non-atomic then*

$$\inf_{T\,:\,T_{\#}\mu=\nu} \int_X c\left(x, T(x)\right)\, \mathrm{d}\mu(x) = \min_{\gamma\in\Pi(\mu,\nu)} \iint_{X\times Y} c(x,y)\, \mathrm{d}\gamma(x,y)\ .$$

The proof uses Lyaponov's convexity theorem. We refer to [25, 22, 1] for details. As we already mentioned if the source measure has atoms there may be no such transport map, the previous theorem says in particular that without atoms, transport maps exist (and actually form a set that is large enough to be dense in transport plans). The presence of atoms actually is therefore the only serious source of nonexistence of transport maps.

Chapter 2

Existence of solutions

2.1 Duality of the Kantorovich problem

A key feature of the linear Monge-Kantorovich formulation is that it has a nice dual formulation that we are going to describe here in an informal way (a rigorous proof via the Fenchel-Rockafellar duality theorem can be found in [9]). As we shall see later, the dual problem is an essential tool to understand the geometry of optimal transport as well as for establishing existence of optimal transport maps for certain cost functions.

Theorem 6 (Kantorovich's duality)**.** *The Kantorovich problem* (KP) *is equivalent to*

$$\sup_{\substack{(\varphi,\psi)\in\mathcal{C}(X)\times\mathcal{C}(Y)\\ \varphi(x)+\psi(y)\leq c(x,y)}} \left\{ \int_X \varphi(x)\ \mathrm{d}\mu(x) + \int_Y \psi(y)\ \mathrm{d}\nu(y) \right\}. \tag{DP}$$

This problem is called the dual formulation of the transport problem (KP)*.*

Proof. The idea, which is standard in problems of this kind, is to rewrite the constrained infimum problem as an «inf sup» problem and exchange the two operation «inf sup» and «sup inf» by formally applying a min-max principle. We will not give a full proof of this result in this lecture but some heuristics as follows.

The intuition behind the Kantorovich duality is to view the functions φ and ψ as Lagrange multipliers associated to the constraints on the marginals:

Exercice 11. Prove that the problem (KP) can be rewriten in the form:

$$\inf_{\gamma\geq 0} \sup_{(\varphi,\psi)\in\mathcal{C}(X)\times\mathcal{C}(Y)} \left\{ \int_X \varphi(x)\ \mathrm{d}\mu(x) + \int_Y \psi(y)\ \mathrm{d}\nu(y) + \iint_{X\times Y} (c(x,y)-\varphi(x)-\psi(y))\ \mathrm{d}\gamma(x,y) \right\}.$$

- for $\gamma \geq 0$, $\gamma \in \Pi(\mu,\nu)$ if and only if

$$\forall(\varphi,\psi)\in\mathcal{C}(X)\times\mathcal{C}(Y) \qquad \iint_{X\times Y} (\varphi(x)+\psi(y))\ \mathrm{d}\gamma(x,y) = \int_X \varphi(x)\ \mathrm{d}\mu(x) + \int_Y \psi(y)\ \mathrm{d}\nu(y)\,.$$

or

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

Hence, if the contraint is satisfied nothing has been added and if not we get $+\infty$ (which will be avoided by the minimisation).

We would like this optimization problems «inf sup» to be equivalent to the following «sup inf» problem:

$$\sup_{(\varphi,\psi)\in\mathcal{C}(X)\times\mathcal{C}(Y)}\left\{\int_X \varphi(x)\,\mathrm{d}\mu(x)+\int_Y \psi(y)\,\mathrm{d}\nu(y)\,:\,c(x,y)\geq\varphi(x)+\psi(y)\text{ on }X\times Y\right\}. \quad \text{(DP)}$$

This equivalence is not always possible, and the main tool to do it is a theorem by Rockafellar requiring concavity in one variable, convexity in the other one, and some compactness assumption. Yet, Rockafellar's statement concerns finite-dimensional spaces, which is not the case here. To handle infinite-dimensional situations one needs to use a more general mini-max theorems. In this lecture, we prefer not to investigate anymore the question of obtaining the duality equality. For the moment, let us accept it as true and see the intuition behind it.

Exercice 12. Prove (KP) $\geq$ (DP)

▶ Consider then the «sup inf» problem:

$$\sup_{(\varphi,\psi)\in\mathcal{C}(X)\times\mathcal{C}(Y)}\inf_{\gamma\geq 0}\left\{\int_X \varphi(x)\,\mathrm{d}\mu(x)+\int_Y \psi(y)\,\mathrm{d}\nu(y)+\iint_{X\times Y}(c(x,y)-\varphi(x)-\psi(y))\,\mathrm{d}\gamma(x,y)\right\}$$
$$=\sup_{(\varphi,\psi)\in\mathcal{C}(X)\times\mathcal{C}(Y)}\left\{\int_X \varphi(x)\,\mathrm{d}\mu(x)+\int_Y \psi(y)\,\mathrm{d}\nu(y)+\inf_{\gamma\geq 0}\iint_{X\times Y}(c(x,y)-\varphi(x)-\psi(y))\,\mathrm{d}\gamma(x,y)\right\}.$$

Note that

$$\inf_{\gamma\geq 0}\left\{\iint_{X\times Y}(c(x,y)-\varphi(x)-\psi(y))\,\mathrm{d}\gamma(x,y)\right\},$$

is 0 whenever the inequality $c(x,y)\geq\varphi(x)+\psi(y)$ holds everywhere (take $\gamma=0$) and $-\infty$ otherwise (take $\gamma=\lambda\delta_{x_0,y_0}$ is a point where $c(x_0,y_0)\geq\varphi(x_0)+\psi(y_0)$ and let λ go to infinity). Hence, while looking for the sup, if a solution to the «sup inf» problem, (φ,ψ) atomatically satisfies the constraints of (DP). As a consequence the «supp inf» problem is equivalent to (DP). □

The shipper problem: here is an informal interpretation which is due to Caffarelli. Suppose that you are a mathematician and an industrial willing to transfer a huge amount of coal from your mines to your factories. Both the amount of coal which you can extract from each mine and the amount of coal which should be received by each factory are fixed. You can hire trucks to do this transport, but you have to pay them $c(x,y)$ for each ton of coal which is transported from a place x to a place y. As you are trying to solve a Kantorovich problem in order to minimise the price you have to pay, another mathematician comes to you and tells you «my friend, let me handle this for you: I will ship all your coal with my own trucks and you will not have to care of what goes where. I will just set a price $\varphi(x)$ for loading one ton of coal at place x, and a price $\psi(y)$ for unloading it at destination y. I will set the price in such a way that your financial interest will be to let me handle all your transport. Indeed, you can check easily that for all x and y, the sum $\varphi(x)+\psi(y)\leq c(x,y)$.»

What Kantorovich's duality tells you is that if the shipper is clever enough, then he can arrange the prices is such a way that you will pay him as much as you would have been ready to spend for the other method.

Theorem 7 (Existence of minimumum to (DP))**.** *The supremum is attained in* (DP).

We will give the main ingredient of the proof, which is the double-convexification trick which is based on

Definition 13 (c-concave transform)**.** *Let $g\in\mathcal{C}(Y)$. The* c-concave transform *of g, denoted g^c, is defined as for all $x\in X$ by*

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

Proof. For $(\varphi, \psi) \in \mathcal{C}(X) \times \mathcal{C}(Y)$, let us denote by J the criterion in (DP):

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

Exercice 13. 1. Prove that $\varphi \leq \psi^c$.

2. Prove that $J(\psi^c, \psi) \geq J(\varphi, \psi)$.

3. Prove that $\psi^{cc} \geq \psi$.

4. Prove

$$(\mathrm{DP}) = \sup_{\psi \in \mathcal{C}(Y)} J(\psi^c, \psi^{cc}) \, .$$

▶

1. The constraint on (φ, ψ) my be rewritten as

$$\forall (x, y) \in X \times Y, \qquad \varphi(x) \leq c(x, y) - \psi(y) \, .$$

   or taking the infimum in y

$$\forall x \in X, \qquad \varphi(x) \leq \inf_{y \in Y} \Big\{ c(x, y) - \psi(y) \Big\} \, .$$

2. By construction (ψ^c, ψ) is admissible and for an admissible pair (φ, ψ), since $\psi^c \geq \varphi$, we have

$$J(\psi^c, \psi) \geq J(\varphi, \psi) \, .$$

3. As

$$\forall (x, y) \in X \times Y \qquad c(x, y) - \psi^c(x) \geq \psi(y)$$

   we have

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

4. Remark again that (ψ^c, ψ^{cc}) is admissible (it is easy to check that it is continuous) and we have

$$J(\psi^c, \psi^{cc}) \geq J(\psi^c, \psi) \geq J(\varphi, \psi)$$

It turns out that the pairs (ψ^c, ψ^{cc}) form a sufficiently rigid set (because their modulus of continuity is controlled by that of c) to get compactness and thus prove the result. □

2.2 Brenier's theorem

Under suitable assumption, we can prove the existance of an optimal map. We will present it for a general cost c as it was later done in [19] but all these ideas where just adapatation of the original ideas of Yann Brenier in [5].

Theorem 8 (Existence of an optimal map)**.** *Let B be a ball of $\mathbb{R}^N$. Consider μ and ν be two probability measures on B. Assume that μ is absolutely continuous with respect to the Lebesgue measure. Assume that $c : B \to \mathbb{R}$ is a strictly convex function of class $\mathcal{C}^1$ such that $c(x, y) = c(x - y)$. There exists a unique transport map T in Problem* (MP)*. Moreover T is of the form*

$$T(x) = x - \nabla c^* \left(\nabla \varphi(x) \right), \qquad \textit{for almost every } x \tag{2.1}$$

where f^* is the inverse of f, for some c-concave potential φ. Moreover $(\mathrm{id}, T)\#\mu$ *is the only optimal transport plan in* (KP).

For refinements of this result, we refer to [12].

Proof. Let γ be an optimal transport plan *i.e.* a solution to

$$\min_{\gamma\in\Pi(\mu,\nu)} \iint_{B\times B} c(x-y)\,\mathrm{d}\gamma(x,y)$$

We will prove that γ is induced by a transport map *i.e.* is of the form $\gamma = (\mathrm{id}, T)_{\#}\mu$ for some transport map T.

Let (φ, ψ) be an optimal solution to the dual problem (DP). We know by Kantorich duality that

$$\iint_{X\times Y} c(x,y)\,\mathrm{d}\gamma(x,y) = \int_X \varphi(x)\,\mathrm{d}\mu(x) + \int_Y \psi(y)\,\mathrm{d}\nu(y) = \iint_{X\times Y} [\varphi(x)+\psi(y)]\,\mathrm{d}\gamma(x,y)$$

or equivalently

$$\iint_{X\times Y} [\varphi(x)+\psi(y)-c(x,y)]\,\mathrm{d}\gamma(x,y) = 0$$

As the function in the bracket is nonnegative it has to vanish $\mathrm{d}\gamma$-a.e.

We already know that φ can be chosen such that $\varphi = \psi^c$. We can deduce that φ is Lipschitz continuous with a constat less than $\|\nabla c\|_{L^\infty(2B)}$. By Rademacher's theorem φ is therefore differential $\mathcal{L}^d$-almost everywhere. Let us denote S the negligible set where φ fails to be continuous.

Exercice 14. Prove that for all $x \in B\setminus(S\cup\partial B)$ and $y \in B$ be such that $\varphi(x)+\psi(y) = c(x-y)$ we have

$$\nabla\varphi(x) = \nabla c(x-y)$$

▶ For h small enough we have on one hand

$$\varphi(x+h) = \varphi(x) + \nabla\varphi(x)\cdot h + o(h)$$

and on the other hand

$$\begin{aligned}\varphi(x+h) = c(x+h-y) - \psi(y) = c(x-y) + \nabla c(x-y)\cdot h - \psi(y) + o(h)\\ = \varphi(x) + \nabla c(x-y)\cdot h + o(h)\,.\end{aligned}$$

We obtain de result by combining the two.

Exercice 15. Prove that, for every x in $B\setminus(S\cup\partial B)$, the set

$$\{y\in B\ ,\ \varphi(x)+\psi(y) = c(x-y)\}$$

consists in the single element

$$T(x) = x - \nabla c^*\left(\nabla\varphi(x)\right)\,.$$

▶ By the previous exercice, $x \in B\setminus(S\cup\partial B)$ and $y\in B$, $\nabla\varphi(x) = \nabla c(x-y)$. Since c is strictly convex, ∇c is injective and we have

$$(\nabla c)^{-1}\nabla\varphi(x) = x-y\,.$$

Which we can write

$$y = x - (\nabla c)^{-1}\nabla\varphi(x)\,.$$

As the boundary points of B and the set S are μ-negligible, $\gamma = (\mathrm{id}, T)\#\mu$. We thus deduce that T is an optimal map. □

2.2.1 The quadratic case

Let us focus on the quadratic case which was initially solved by Yann Brenier in his pathbreaking article [5].

Theorem 9 (Brenier's theorem). *There exists a unique, up to a μ-negligible set, map of the form $T = \nabla u$ with u convex which transports μ onto ν. This map T is the optimal tranport between μ and ν for the quadratic cost.*

In other words, the optimal transport is the *gradient of a convex function*. The optimal map ∇u is called the Brenier map between μ and ν.

Proof. We already know from the previous section that there is a unique optimal transport T which is characterised by

$$\varphi(x) + \psi\,(T(x)) = \frac{|x - T(x)|^2}{2}$$

where φ and ψ are related by the conjugacy relations

$$\varphi(x) = \inf_y \left\{ \frac{|x-y|^2}{2} - \psi(y) \right\} \quad \text{and} \quad \psi(x) = \inf_x \left\{ \frac{|x-y|^2}{2} - \varphi(y) \right\}$$

This relation can be rewritten

$$\frac{|x|^2}{2} - \varphi(x) = \sup_y \left\{ x \cdot y - \frac{|y|^2}{2} + \psi(y) \right\}$$

and

$$\frac{|y|^2}{2} - \psi(y) = \sup_x \left\{ x \cdot y - \frac{|x|^2}{2} + \varphi(y) \right\} .$$

Define

$$u : x \mapsto \frac{|x|^2}{2} - \varphi(x) \quad \text{and} \quad v : y \mapsto \frac{|y|^2}{2} - \psi(y)$$

In the quadratic case Formula (2.1) can thus be rewritten

$$T(x) = x - \nabla\varphi(x) = \nabla u(x) .$$

□

2.2.2 The Monge-Ampère equation

Consider that μ and ν are absolutely continuous with respect to the Lebesgue measure: $\mu = \rho_0.\mathcal{L}^d$ and $\nu = \rho_1.\mathcal{L}^d$. Brenier's map ∇u then satisfied, at least formally:

By definition of a transport map

$$\forall \varphi \in \mathcal{C}(Y), \qquad \int_Y \varphi(y)\ \mathrm{d}\rho_1(y) = \int_X \varphi\,(\nabla u(x))\ \mathrm{d}\rho_0(x)$$

By taking the change of variable $y = \nabla u(x)$ on the left hand side we obtain

$$\int_Y \varphi(y)\ \mathrm{d}\rho_1(y) = \int_X \varphi\,(\nabla u(x)) \det(D^2 u(x))\ \mathrm{d}\rho_1\,(\nabla u(x))$$

so that, equalling the two, we obtain the Monge-Ampère equation

$$\det(D^2 u)\rho_1(\nabla u) = \rho_0 .$$

This Monge-Ampère equation could give a way to determine the optimal transport. Unfortunately this partial differential equation is extremely difficult to handle. A deep regularity theory due to Luis Caffarelli, [7] establishes conditions under which the Brenier map is actually smooth.

2.2.3 The one-dimensional case

Let f and g be two probability densities. Denote F and G their cumulative distribution functions. If there is an optimal transport map T transporting $f \cdot \mathcal{L}$ onto $g \cdot \mathcal{L}$ then by definition of a transport map

$$\forall \varphi \in \mathcal{C}(Y), \qquad \int_Y \varphi(y) g(y)\, \mathrm{d}y = \int_X \varphi\left(T(x)\right) f(x)\, \mathrm{d}x \ .$$

Let $a \in R$. Consider $\varphi = \mathbf{1}_{(\infty, T(a)]}$, we have

$$\int_{-\infty}^{T(a)} g(y)\, \mathrm{d}y = \int_{-\infty}^{a} f(x)\, \mathrm{d}x \ .$$

Hence for all $a \in \mathbb{R}$,

$$G(a) = F \circ T(a) \ .$$

As a consequence, the optimal transport is explicitely given by

$$T = F^{-1} \circ G \ ,$$

where F^{-1} is the generalized inverse of F:

$$F^{-1}(t) := \inf_{x \in \mathbb{R}} \{F(x) > t\} \ .$$

This expresses that the solution to the transport problem is given by the monotone rearrangement of μ onto ν: one has to proceed by transfering the sand into the hole starting from the left. Note that discountinuity points of G correspond to atoms for ν. Whenever ν has an atom, $G^{-1} \circ F$ is constant on some interval: when encountering an atom in the filling process, one hs to keep putting mass in this atom for some time.

Chapter 3

Geometry of optimal transport

In this chapter we will consider only the quadratic case $c(x,y) = |x-y|^2/2$. Whenever ρ is absolutely continuous with respect to le Lebesgue measure in $\mathbb{R}^N$, space which we denote $\mathcal{P}_{\mathrm{ac}}(\mathbb{R}^N)$, we shall identify it with its Lebesgue density. and write $\mathrm{d}\rho(x) = \rho(x)\,\mathrm{d}x$.

Introduce the Monge-Kantorovich distance, also called the Wasserstein distance, by

$$\mathcal{W}_2(\mu,\nu) = \sqrt{\inf_{T\,:\,T\#\mu=\nu} \int_X \frac{|x-T(x)|^2}{2}\,\mathrm{d}\mu(x)}\,.$$

$\mathcal{W}_2$ defines indeed a distance on the set of probability measures.

3.1 The Benamou-Brenier dynamic formulation: geodesics

Jean-David Benamou and Yann Brenier, in [3], gave a very interesting and fruitful dynamic formulation of the quadratic optimal transport problem. We aim to describe these ideas in an informal way.

Consider ρ_0 and ρ_1 two probabilty densities. The starting point of the Brenier-Benamou formulation is that we are now looking for a curve of measures connecting ρ_0 and ρ_1 while minimising some action functional.

In a Lagrangian point-of-view, classical in fluid dynamics, we will focus of the trajectories of the particles: we label each particle according to their initial position and study the trajectory of each labelled particle: $X_t(x)$ designs the position at time t of the particle which intial position was x. The dynamics can be described by the following:

$$\begin{aligned} &X_0(x) = x \\ &\partial_t X_t(x) = v\left(t, X_t(x)\right) \end{aligned}$$

If v is continous and Lipschitz with respect to the second variable, the Cauchy-Lipschitz theorem ensures that there is a unique solution.

Let us denote ρ_0 the initial distribution of the particles. How does the spatial distribution starting from ρ_0 and following the flow of v evolves with time? In other words, how is the initial distribution ρ_0 transported by the flow of v? In transport terms, this amounts to characterise teh curve of measures $t \mapsto \rho_t = X_t\#\rho_0$. In fluid dynamics, it is classical to realise that can be described by a *continuity equation*:

$$\begin{aligned} &\partial_t \rho + \nabla\cdot(\rho\, v) = 0 \\ &\rho(0,x) = \rho_0 \end{aligned} \tag{3.1}$$

Consider now the kinetic energy

$$\int_{\mathbb{R}^d} |v(t,x)|^2\,\mathrm{d}\rho_t(x)$$

and the average kinetic energy:

$$E(\rho, v) = \int_0^1 \int_{\mathbb{R}^d} |v(t,x)|^2 \, \mathrm{d}\rho_t(x) \, \mathrm{d}t$$

The Benamou-Brenier problem reads as the average kinetic energy minimisation over the set of all the curves/velocity following the continuity equation strating in ρ_0 and ending in ρ_1:

$$\inf_{\rho, v}\{E(\rho, v) \, : \, \text{The continuity equation (3.1) holds and } \rho(0,\cdot) = \rho_0 \text{ and } \rho(1,\cdot) = \rho_1\} \, . \qquad \text{(BBP)}$$

Theorem 10 (The Brenier-Benamou dynamic formulation)**.** *The Benamou-Brenier Problem* (BBP) *coincides with the square of the Monge-Kantorovich distance:* $\mathcal{W}_2(\rho_0, \rho_1)^2$*.*

In the mean time we will determine the optimal curve.

Proof. Let (ρ, v) be admissible for Problem (BBP) and let X_t be the flow associated to v.

1. Prove

$$\int_{\mathbb{R}^d} |v(t,x)|^2 \, \mathrm{d}\rho_t(x) = \int_{\mathbb{R}^d} |v(t, X_t(x))|^2 \, \mathrm{d}\rho_0(x)$$

2. Prove

$$E(\rho, v) = \int_{\mathbb{R}^d} \int_0^1 |\partial_t X_t(x)|^2 \, \mathrm{d}t \, \mathrm{d}\rho_0(x)$$

3. Prove

$$E(\rho, v) \geq \int_{\mathbb{R}^d} |X_1(x) - x|^2 \, \mathrm{d}\rho_0(x)$$

4. Conclude that

$$E(\rho, v) \geq \mathcal{W}_2(\rho_0, \rho_1)^2 \, .$$

▶

1. As $\rho_t = X_y \# \rho_0$, we have the desired resut.
2. By Fubini's theorem we obtain

$$\begin{aligned} E(\rho, v) &= \int_0^1 \int_{\mathbb{R}^d} |v(t,x)|^2 \, \mathrm{d}\rho_t(x) \, \mathrm{d}t = \int_0^1 \int_{\mathbb{R}^d} |v(t, X_t(x))|^2 \, \mathrm{d}\rho_0(x) \, \mathrm{d}t \\ &= \int_{\mathbb{R}^d} \int_0^1 |v(t, X_t(x))|^2 \, \mathrm{d}t \, \mathrm{d}\rho_0(x) = \int_{\mathbb{R}^d} \int_0^1 |\partial_t X_t(x)|^2 \, \mathrm{d}t \, \mathrm{d}\rho_0(x) \end{aligned}$$

3. By Jensen's inequality

$$\begin{aligned} E(\rho, v) = \int_{\mathbb{R}^d} \int_0^1 |\partial_t X_t(x)|^2 \, \mathrm{d}t \, \mathrm{d}\rho_0(x) &\geq \int_{\mathbb{R}^d} \left| \int_0^1 \partial_t X_t(x) \, \mathrm{d}t \right|^2 \, \mathrm{d}\rho_0(x) \\ &= \int_{\mathbb{R}^d} |X_1(x) - x|^2 \, \mathrm{d}\rho_0(x) \end{aligned}$$

4. As X_1 is a transport from ρ_0 to ρ_1 we obtain the result.

To prove that converse inequality let ∇u be the Brenier map between ρ_0 and ρ_1. Define

$$X_t(x) = \nabla u_t(x) \qquad \text{where} \qquad u_t(x) := (1-t)\frac{|x|^2}{2} + t u(x) \, .$$

Exercice 16. 1. Prove

$$\partial_t X_t(x) = \nabla u(x) - x \, .$$

2. Using that the associated velocity field is given by $\partial_t X_t(x) = v(t, X_t(x))$, deduce

$$v(t, \nabla u_t(x)) = \nabla u(x) - x$$

3. Conclude that

$$v(t,x) = \nabla u\left(\nabla u_t^*(x)\right) - \nabla u_t^*(x)\ .$$

4. We set $\rho_t = X_t \# \rho_0$. Prove that

$$E(\rho, v) = \mathcal{W}_2(\rho_0, \rho_1)^2\ .$$

▶

1. We obviously have

$$X_t(x) = \nabla u_t(x) = (1-t)x + t\nabla u(x)\ .$$

The result is readily obtained by differentiating in t.

2. The associated velocity field is defined by

$$\partial_t X_t(x) = v(t, \nabla u_t(x))\ .$$

The previous question gives the desired result.

3. Since $X_t = \nabla u_t$ is the gradient of a strictly convex function it is invertible with inverse ∇u_t^*. This yields the result.

4. By construction (ρ, v) is admissible for the Benamou-Brenier Problem (BBP) and

$$E(\rho, v) = \int_{\mathbb{R}^d} |\nabla u(x) - x|^2 \, \mathrm{d}\rho_0 = \mathcal{W}_2(\rho_0, \rho_1)^2\ .$$

□

This prove that the value of the Benamou-Brenier Problem (BBP) equals $\mathcal{W}_2(\rho_0, \rho_1)^2$ but also that (ρ, v) is optimal for the Benamou-Brenier Problem (BBP). Note the special form of the optimal curve of measures:

$$\rho_t = \nabla u_t \# \rho_0 = \left((1-t)\mathrm{id} + t\nabla u\right) \# \rho_0\ .$$

This optimal ρ_t is thus obtained by interpolating linearly the optimal transport.

We introduce

Definition 14 (Displacement interpolation)**.** *Let ρ_0 and ρ_1 be two measures in $\mathcal{P}_{\mathrm{ac}}(\mathbb{R}^N)$. Let ∇u be the Brenier map between ρ_0 and ρ_1. We define the* displacement interpolation *$(\rho_t)_{t\geq 0}$ between ρ_0 and ρ_1, for all $t \in [0,1]$ by*

$$\rho_t := \left[(1-t)\mathrm{id} + t\nabla u\right] \# \rho_0\ .$$

Moreover, since

$$(1-t)\mathrm{id} + t\nabla u = \nabla\left[(1-t)\frac{|\cdot|^2}{2} + t\,u\right]$$

is always the gradient of a convex function and pushes ρ_0 onto ρ_t, it is the optimal map between ρ_0 and ρ_t.

Exercice 17. 1. Compute $\rho_t|_{t=0}$ and $\rho_t|_{t=1}$.

2. Prove that $\mathcal{W}_2(\rho_0, \rho_t) = t\mathcal{W}_2(\rho_0, \rho_1)$.

▶

1. Note that $\rho_t|_{t=0} = \rho_0$ and $\rho_t|_{t=1} = \rho_1$.

2. As a consequence of Brenier's theorem

$$\begin{aligned}\mathcal{W}_2(\rho_0,\rho_t)^2 &= \int_{\mathbb{R}^n} |x-[(1-t)x+t\nabla u(x)]|^2 \, \mathrm{d}\rho_0 \\ &= t^2\int_{\mathbb{R}^n} |x-\nabla u(x)|^2 \, \mathrm{d}\rho_0 \\ &= t^2\mathcal{W}_2(\rho_0,\rho_1)^2\end{aligned}$$

3.2 Otto's interpretation: Metric structure

This section is out of the scope of this lecture but I would like to formally give the idea why the Benamou-Brenier formulation looks very much like a geodesic formula in Riemaniann geometry. Let us dig on this idea.

To develop the analogy, we would like to define a metric structure $\langle\cdot,\cdot\rangle_\rho$ on each tangent space $T_\rho\mathcal{P}$ depending smoothly on ρ. How to think of an element of the tangent space? This is the time derivative at time 0 of some trajectory $t\mapsto\rho(t)$ starting on ρ at time 0. As we have already seen, this path $(\rho(t))_t$ can be seen as the time-evolving density of a set of particles moving continuously with a velocity v_t and satisfies the continuity equation:

$$\frac{d\rho}{dt}+\nabla\cdot(\rho\, v)=0\,.$$

So that the tangent space is the space of all the probability densities of the form $-\nabla\cdot(\rho\, v)$. Denote $\|\cdot\|_\rho$ the norm associated to $\langle\cdot,\cdot\rangle_\rho$. It is natural to define

$$\left\|\frac{d\rho}{dt}\right\|^2_{\rho(t)} = \inf\left\{\int\rho|v|^2 \;:\; v\in\mathcal{L}^2(\,\mathrm{d}\rho),\ \frac{d\rho}{dt}+\nabla\cdot(\rho\, v)=0\right\}$$

The square of the Monge-Kantorovich distance between ρ_0 and ρ_1 should be equal to

$$\inf\left\{\int_0^1\left\|\frac{d\rho}{dt}\right\|^2_{\rho(t)}\mathrm{d}t \;:\; \text{for all path } \rho(t) \text{ connecting } \rho_0 \text{ and } \rho_1\right\}.$$

The intuition behind this formula is the following: let a density of particles ρ be given ; let $d\rho/dt$ be an infinitesimal variation of this probability density, *i.e.* an element of the tangent space $T_\rho\mathcal{P}$. We can assume that this infinitesimal variation corresponds to particles moving around. We have no idea on the way these particles move but we see the effect of the probability density. Let us try to guess the velocity field os partcicles, there are many possibilities, actually all the vector fields $v:\mathbb{R}^d\to\mathbb{R}$ solving the continuity equation are compatible with the observed variation of ρ. Among all these possible vector fields we try to select the one whose kinetic energy is the lowest.

Let us know prove that v_0 should be a gradient. Assume ρ is smooth and positive. Let v_0 be a minimiser. Consider w a vector field with zero divergence. For any $\varepsilon\neq 0$,

$$v_0+\varepsilon\frac{w}{\rho}$$

is still admissible in the sense that

$$-\nabla\cdot\left[\rho\left(v_0+\varepsilon\frac{w}{\rho}\right)\right]=\frac{d\rho}{dt}$$

Since v_0 is minimising, we should have

$$\int\rho|v^2|\le\int\rho\left|v_0+\varepsilon\frac{w}{\rho}\right|^2$$

Expanding the square, simplifying out term of order 0 in ε, dividing by ε and letting ε go to 0, we find

$$\int v_0 \cdot w = 0$$

In other words, v_0 should be orthogonal in $\mathcal{L}^2$-sense, to the set of divergence-free vector fields *i.e.* v_0 should be a gradient: $v_0 = \nabla u_0$.

To sum up, we can write, at least very formally,

$$\mathcal{W}_2(\rho_0,\rho_1)^2 = \inf \left\{ \int_0^1 \left\| \frac{\partial \rho}{\partial t} \right\|_{\rho(t)}^2 \mathrm{d}t \ : \ \text{for all path } \rho(t) \text{ connecting } \rho_0 \text{ and } \rho_1 \right\} .$$

with

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \nabla u) = 0 .$$

where

$$\left\| \frac{\partial \rho}{\partial t} \right\|_{\rho(t)}^2 = \int \rho |\nabla u|^2 .$$

This interpretation endows the set of probabilities with a Riemannian metric structure: indeed, by polarisation we can define the scalar product of two tangent vectors (elements of the tangent space):

$$\left\langle \left(\frac{\partial \rho}{\partial t}\right)_1, \left(\frac{\partial \rho}{\partial t}\right)_2 \right\rangle_\rho = \int \rho \langle \nabla u_1, \nabla u_2 \rangle$$

where u_1 and u_2 solve

$$\left(\frac{\partial \rho}{\partial t}\right)_1 = -\nabla \cdot (\rho \nabla u_1) \quad \text{and} \quad \left(\frac{\partial \rho}{\partial t}\right)_2 = -\nabla \cdot (\rho \nabla u_2) .$$

With the Riemaniannian structure come basic calculus rules for function defined on $\mathcal{P}(\mathbb{R}^d)$. In particular we can define the gradient operator, as usual using Riesz' representation theorem, denoted grad_W by:

$$\left\langle \mathrm{grad}_W F(\rho), \frac{\partial \rho}{\partial t} \right\rangle = DF(\rho) \cdot \frac{\partial \rho}{\partial t} .$$

3.3 Displacement convexity

3.3.1 Definition

A natural question, which will arrise in many applications which we will see later, is the following: let F be a functional on the space of probability measures, what can be said about the behaviour of $F(\rho_t)$ as t varies in $[0,1]$? Among properties to study, convexity is of course in the first places. This motivates the following definition:

Definition 15 (Displacement convexity). *Let $\mathcal{P}$ be a subset of the set of probability measures which are absolutely continuous with respect to the Lebesgue measure. Assume that $\mathcal{P}$ is stable under displacement interpolation. Let F be a function defined on $\mathcal{P}$ with values in $\mathbb{R} \cup +\infty$.*

- *The functional F is said to be* displacement convex *on $\mathcal{P}$ if for all ρ_0, ρ_1 in $\mathcal{P}$ and (ρ_t) is the displacement interpolation between ρ_0 and ρ_1 the function*

$$t \mapsto F(\rho_t)$$

  is convex on $[0,1]$.

- *The functional F is strictly convex on $\mathcal{P}$ if*

$$\rho_0 \neq \rho_1 \Rightarrow t \mapsto F(\rho_t) \quad \textit{is strictly convex on } [0,1].$$

- *For $\lambda > 0$, the function F is λ-uniformly convex on $\mathcal{P}$ if*

$$\frac{\mathrm{d}^2}{\mathrm{d}t} F(\rho_t) \geq \lambda \mathcal{W}_2(\rho_0,\rho_1)^2 .$$

3.3.2 Basic examples

Potential energy

Let $V : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$. We define the potential energy by

$$\mathcal{V}(\rho) = \int_{\mathbb{R}^d} V(x)\, \mathrm{d}\rho(x)\ .$$

Theorem 11. *Let $\mathcal{P}$ be a subset of $\mathcal{P}(\mathbb{R}^d)$ of probability measure with finite second moment. Assume that $\mathcal{P}$ is stable under displacement interpolation. If V is convex (resp. strictly convex, resp. λ-uniformly convex) then $\mathcal{V}$ is convex then (resp. strictly convex, resp. λ-uniformly convex).*

Proof. Let ρ_0 and ρ_1 be two measures in $\mathcal{P}$. We consider displacement interpolation of the form

$$\rho_t = (\mathrm{id} - t\,\theta)\#\rho_0$$

where $\theta = \mathrm{id} - \nabla\varphi$ with φ convex.

Exercice 18. 1. Prove that $\mathcal{V}$ is displacement convex on $\mathcal{P}$.

2. If V is striclty convex, prove that $\mathcal{V}$ is stricty displacement convex.

3. If V is λ-uniformly convex, prove that $\mathcal{V}$ is λ-uniformly displacement convex.

▶

1. By definition of the image-measure:

$$\mathcal{V}(\rho_t) = \int_{\mathbb{R}^d} V(x)\, \mathrm{d}\rho_t(x) = \int_{\mathbb{R}^d} V(x - t\,\theta(x))\, \mathrm{d}\rho_0(x)\ .$$

   The convexity of $t \mapsto \mathcal{V}(\rho_t)$ is a direct consequence of the convexity of V.

2. If $\rho_0 \neq \rho_1$ then $\theta \neq 0$ and the strict convexity of $t \mapsto \mathcal{V}(\rho_t)$ is a direct consequence of the strict convexity of V.

3. For all s, t and σ in $[0,1]$,

$$\begin{aligned}\sigma\mathcal{V}(\rho_t) + (1-\sigma)\mathcal{V}(\rho_s) - \mathcal{V}(\rho_{\sigma t+(1-\sigma)s}) &= \int_{\mathbb{R}^d} \Big[\sigma V\left(x - t\,\theta(x)\right) + (1-\theta)V\left(x - s\theta(x)\right) \\ &\qquad - \mathcal{V}\left[\sigma(x - t\,\theta(x)) + (1 - \sigma(x - s\,\theta(x)))\right]\Big]\, \mathrm{d}\rho_0(x) \\ &\geq \lambda\frac{\sigma(1-\sigma)}{2}\int_{\mathbb{R}^d} \left|(x - t\,\theta(x)) - (x - s\,\theta(x))\right|^2\, \mathrm{d}\rho_0 \\ &\geq \lambda\frac{\sigma(1-\sigma)}{2}\int_{\mathbb{R}^d} |\theta(x)|^2\, \mathrm{d}\rho_0 (t-s)^2 \\ &= \mathcal{W}_2(\rho_0,\rho_1)^2\end{aligned}$$

□

Internal energy

Let U be a measurable map $\mathbb{R}^+ \to \mathbb{R} \cup +\infty$. Consider the internal energiy of the form

$$\mathcal{U}(\rho) := \int_{\mathbb{R}^d} U\left(\rho(x)\right)\, \mathrm{d}x\ .$$

Theorem 12. *Let $\mathcal{P}$ be a subset of $\mathcal{P}_{\mathrm{ac}}(\mathbb{R}^d)$ stable under displacement interpolation. If U satisfies $U(0) = 0$ and*

$$r \mapsto r^d U(r^{-d})$$

is convex increasing on $(0,+\infty)$ then $\mathcal{U}$ is displacement convex on $\mathcal{P}$.

Among the most classical internal energies satisfying the assumption of the Theorem:

- $U(\rho) = \rho^\gamma$ with $\gamma \geq 1$,
- $U(\rho) = \rho \log \rho$,
- $U(\rho) = -\rho^\gamma$ with $(1 - 1/d) \leq \gamma \leq 1$.

Proof. This proof is formal. Let ρ_0 and ρ_1 be two measures in $\mathcal{P}$. We consider displacement interpolation of the form

$$\rho_t = (\mathrm{id} - t\,\theta)\#\rho_0$$

where $\theta = \mathrm{id} - \nabla\varphi$ with φ convex. We will denote $\nabla\theta$ the Jacobian matrix of u_t.

Exercice 19. 1. Prove

$$\mathcal{U}(\rho_t) = \int_{\mathbb{R}^d} U\left(\rho_t\left(x - t\,\theta(x)\right)\right) \det\left(I_d - t\,\nabla\theta(x)\right)\, \mathrm{d}x$$

2. Deduce

$$\mathcal{U}(\rho_t) = \int_{\mathbb{R}^d} U\left(\frac{\rho_0(x)}{\det\left(I_d - t\,\nabla\theta(x)\right)}\right) \det\left(I_d - t\,\nabla\theta(x)\right)\, \mathrm{d}x$$

▶

1. We have

$$\mathcal{U}(\rho_t) = \int_{\mathbb{R}^d} U\left(\rho_t(y)\right)\, \mathrm{d}y$$

By change of variable $y = x - t\,\theta(x)$ we obtain

$$\mathcal{U}(\rho_t) = \int_{\mathbb{R}^d} U\left(\rho_t\left(x - t\,\theta(x)\right)\right) \det\left(I_d - t\,\nabla\theta(x)\right)\, \mathrm{d}x$$

2. As $\rho_t = (\mathrm{id} - \theta)\#\rho_0$, the Monge-Ampère equation gives

$$\det\left(I_d - t\,\nabla\theta(x)\right) \rho_t\left(x - t\,\theta(x)\right) = \rho_0\,.$$

Hence

$$\mathcal{U}(\rho_t) = \int_{\mathbb{R}^d} U\left(\frac{\rho_0(x)}{\det\left(I_d - t\,\nabla\theta(x)\right)}\right) \det\left(I_d - t\,\nabla\theta(x)\right)\, \mathrm{d}x$$

Now we will use

Lemma 1 (Concavity of $\det^{1/N}$). *Let S be a symmetric matrix. If $S \leq I_d$ is not a scalar matrix then the function*

$$t \mapsto \det(I_d - t\,S)^{1/n}$$

is strictly concave.

Setting $S = \nabla\theta$ and $r = \rho_0(x)$, we have the two mappings

$$t \mapsto \lambda = \det\left(I_d - t\nabla\theta(x)\right)^{1/n} \qquad \text{and} \qquad \lambda \mapsto U\left(\frac{r}{\lambda^n}\right)\lambda^n$$

As the two functions are respectively convex non-increasing and concave, their composition is convex. □

Interaction energy

Consider the interaction potential W and

$$\mathcal{W}(\rho) = \frac{1}{2} \iint_{\mathbb{R}^d \times \mathbb{R}^d} W(x-y)\, \mathrm{d}\rho(x)\, \mathrm{d}\rho(y)\ .$$

Theorem 13. *Let $\mathcal{P}$ be a subset of $\mathcal{P}_{\mathrm{ac}}(\mathbb{R}^d)$ stable under displacement interpolation. If W is convex then $\mathcal{W}$ is convex increasing on $(0,+\infty)$ then $\mathcal{W}$ is displacement convex on $\mathcal{P}$. If W is strictly convex then for all $m \in \mathbb{R}^d$, $\mathcal{W}$ is strictly displacement convex on the subspace of probability measure having mean m.*

The mean or centre of mass of a probability measure ρ is

$$\int x\, \mathrm{d}\rho(x)\ .$$

Remark 4. The functional $\mathcal{W}$ is invariant under translation: if $\tau_a : x \mapsto x + a$ for some $a \in \mathbb{R}^d$, then $\mathcal{W}(\tau_a \# \rho) = \mathcal{W}(\rho)$. This is why strict displacement convexity can hold only if we rule out translations, for instance by fixing the mean.

Proof.

Exercice 20.

1. If W is convex. Prove that $\mathcal{W}$ is displacement convex.
2. If W is strictly convex. Prove that $\mathcal{W}$ is stricty displacement convex on the subspace of probability measure having mean m.

▶

1. We have
$$\mathcal{W}(\rho_t) = \frac{1}{2} \iint_{\mathbb{R}^d \times \mathbb{R}^d} W\left([x-y] - t[\theta(x) - \theta(y)]\right)\, \mathrm{d}\rho_0(x)\, \mathrm{d}\rho_0(y)\ .$$
It follows immediately that $t \mapsto \mathcal{W}(\rho_t)$ also is.
2. If w is strictly convex then equality occurs in the convex inequality if and only if $\theta(x) = a$ for $\mathrm{d}\rho_0$- almost all $x \in \mathbb{R}^d$, where a is some element of $\mathbb{R}^d$. This condition means that ρ_0 and ρ_1 are translates of each other:
$$\mathrm{d}\rho_1(x) = \mathrm{d}\rho_0(x+a)\ .$$
This is ruled out by the same mean condition.

□

3.4 Application to the minimum

Theorem 14. *Let U satisfy the conditions of the previous section, W be convex and V be strictly convex. There is at most a minimiser on $\mathcal{P}_{\mathrm{ac}}(\mathbb{R}^d)$ of*

$$\int_{\mathbb{R}^d} U(\rho(x))\, \mathrm{d}x + \int_{\mathbb{R}^d} V(x)\, \mathrm{d}\rho(x) + \frac{1}{2} \iint_{\mathbb{R}^d \times \mathbb{R}^d} W(x-y)\, \mathrm{d}\rho(x)\, \mathrm{d}\rho(y)\ .$$

Proof. Let ρ_1 and ρ_2 be two minimisers. Consider the displacement interpolant

$$\rho_t = ((1-t)\mathrm{id} + t\nabla u)\#\rho_1$$

where ∇u is the Brenier map between ρ_1 and ρ_2.

Exercice 21.

- Prove that
$$t \mapsto F(\rho_t)$$
is strictly convex.

- Prove
$$F(\rho_{1/2}) < \frac{F(\rho_1) + F(\rho_2)}{2}$$
- Conclude.

▶

1. This is the definition if strict convexity.
2. This is a direct consequence of the strict convexity.
3. Hence there is at most a minimiser.

□

3.5 Above the tangent formulation

It is a general fact that if a function $\Phi : [0,1] \to \mathbb{R} \cup +\infty$ is λ-uniformly convex then

$$\Phi(1) \geq \Phi(0) + \frac{d}{dt}\Big|_{t=0}^{+} \Phi(t) + \frac{\lambda}{2}$$

where

$$\frac{d}{dt}\Big|_{t=0}^{+} \Phi(t) = \limsup_{t \to 0^+} \frac{\Phi(t) - \Phi(0)}{t}$$

Let us give an explicit expression of

$$\frac{d}{dt}\Big|_{t=0}^{+} \Phi(t)$$

for the energy considered before.

Theorem 15. *Let ρ_0 and ρ_1 be absolutely continuous probability measures in $\mathbb{R}^d$. Let $\nabla\varphi$ be the unique Brenier map which pushes ρ_0 onto ρ_1. We have*

- $$\frac{d}{dt}\Big|_{t=0}^{+} \mathcal{V}(\rho_t) = \int_{\mathbb{R}^d} \nabla V(x) \cdot (\nabla\varphi(x) - x)\rho_0 \, \mathrm{d}x \, .$$
- $$\frac{d}{dt}\Big|_{t=0}^{+} \mathcal{U}(\rho_t) = \int_{\mathbb{R}^d} \left[U(\rho_0) - \rho_0 U'(\rho_0)\right] \cdot (\Delta_A \varphi(x) - d) \, \mathrm{d}x \, .$$
  where Δ_A stands for the Laplace operator in the Alexandrov sense,
- *and*
$$\frac{d}{dt}\Big|_{t=0}^{+} \mathcal{W}(\rho_t) = \iint_{\mathbb{R}^d \times \mathbb{R}^d} \nabla W(x-y) \cdot \left[(\varphi(x) - x) - (\varphi(y) - y)\right] \, \mathrm{d}\rho_0(x) \, \mathrm{d}\rho_0(y)$$

Proof.

Exercice 22. 1. Determine
$$\frac{\mathcal{V}(\rho_t) - \mathcal{V}(\rho_0)}{t} \, .$$

2. What is the limit of this quantity when t goes to 0?
3. Conclude for $\mathcal{V}$.
4. Determine
$$\frac{\mathcal{U}(\rho_t) - \mathcal{U}(\rho_0)}{t} \, .$$

5. What is the limit of this quantity when t goes to 0?
6. Conclude for $\mathcal{U}$.

▶

1. By definition of the push-forward
$$\frac{\mathcal{V}(\rho_t)-\mathcal{V}(\rho_0)}{t}=\int_{\mathbb{R}^d}\frac{V\left((1-t)x+t\nabla\varphi(x)-V(x)\right)}{t}\,\mathrm{d}\rho_0$$
2. The intergrand on the right hand side converges to
$$\nabla V(x)\cdot(\varphi(x)-x)\qquad a.e.$$
As V is convex the dominated convergence theorem allows to pass to the limit in the integral.
3. The conclusion follows from the monotone convergence theorem.
4. Using Theorem 12 we have
$$\frac{\mathcal{U}(\rho_t)-\mathcal{U}(\rho_0)}{t}=\int_{\mathbb{R}^d}\frac{u(t,x)-u(t,0)}{t}\,\mathrm{d}x$$
with
$$u(t,x)=U\left(\frac{\rho_0(x)}{\det\left(I_d-t\,\nabla\theta(x)\right)}\right)\det\left(I_d-t\,\nabla\theta(x)\right)$$
5. Like before the fonction $u(\cdot,t)$ is convex and integrable. Therefore its slope
$$\frac{u(t,x)-u(t,0)}{t}$$
is non-increasing as t goes to 0 and converges to
$$\frac{\partial u}{\partial t}(0,x)\ .$$
Let us now compute this limit: denote
$$\alpha(t)=\det\left(I_d-t\,\nabla\theta(x)\right)$$
We have
$$\frac{\partial}{\partial t}\left[U\left(\frac{\rho_0}{\alpha(t)}\right)\alpha(t)\right]=\alpha'(t)U\left(\frac{\rho_0}{\alpha(t)}\right)-U'\left(\frac{\rho_0}{\alpha(t)}\right)\rho_0\frac{\alpha'(t)}{\alpha(t)}$$
As $\det(I+\varepsilon A)=1+\varepsilon\mathrm{Tr}(A)+o(\varepsilon)$ we have
$$\alpha'(t)=\Delta\theta(x)-d$$
Hence in $t=0$:
$$\frac{\partial}{\partial t}\left[U\left(\frac{\rho_0}{\alpha(t)}\right)\alpha(t)\right](0,x)=\left(\Delta\theta(x)-d\right)U\left(\rho_0\right)-U'\left(\rho_0\right)\rho_0\left(\Delta\theta(x)-d\right)$$
6. The conclusion follows from the monotone convergence theorem.

The last case $\mathcal{W}$ is left as an exercise. □

Chapter 4

Gradient flows in the Monge-Kantorovich metric

4.1 Gradient flows

Many differential systems arising on physics can be described by an energy functional E and a gradient flow system:

$$\frac{dX}{dt}(t) = -\nabla E(X(t)) \ .$$

These gradient flows appear in all branches of physics, computation sceince, engineering, IA, etc. We intend to describe it in this chapter and to develop an application in economics in the next chapter.

It is known that a gradient flow can be descretised using the following scheme: let $\tau > 0$ be a time step so that $\{i\tau\}_{i\in\{0,\cdots,n\}}$ is a uniform discretisation of $[0, n\tau = T]$. For any $i \in \{0, \cdots, n\}$, we define a sequence $(X_\tau^n)_n$ by

$$X_\tau^{n+1} = \text{Argmin}\left\{E(X) + \frac{\text{dist}(X_\tau^n - X)^2}{2\,\tau}\right\}$$

In this Euclidean case, the Euler-Lagrange equation associated to this problem is

$$\frac{X_\tau^{n+1} - X_\tau^n}{\tau} = -\nabla E(X_\tau^{n+1})$$

which is exactly the time discretisation of the gradient flow. This scheme is called the (implicit Euler) scheme. Note that we could have choosen to evaluate ∇E is ρ_i, in which case the scheme is called the explicit Euler scheme. In the case of a explicit Euler scheme, the scheme is easier to implement we it is known to be far less stable than the implicit Euler scheme.

Next, we define for instance X_τ on $\mathbb{R}^+$ as the piecewise constant function with value X_τ^n on $[n\tau, (n+1)\tau]$. Remains then to pass to the limit $\tau \to 0$.

4.2 The JKO scheme

This strategy was adapted to the framework of probability measures, using the Monge-Kantorovich distance in a very inspiring article by Jordan, Kinderlherer and Otto, in [14]. In this chapter we will follow the main steps of this article for the linear Fokker-Planck equation

$$\frac{\partial \rho}{\partial t}(x) = \Delta\rho(x) + \nabla\cdot(x\rho(x))$$

But this strategy can be readily adapted to displacement convex functionals.

Let us consider

$$E(\rho) = \int \log(\rho)\ \mathrm{d}\rho + \int \frac{|x|^2}{2}\ \mathrm{d}\rho$$

As above, let τ be a time-step. Let $\rho^0 \in \mathcal{P}(\mathbb{R}^d)$ be an initial datum such that $E(\rho^0) < \infty$. We introduce the sequence $(\rho_\tau^n)_n$ defined by: $\rho_\tau^0 = \rho^0$ and recursively by

$$\rho_\tau^{n+1} = \text{Argmin}\left\{ E(\rho) + \frac{\mathcal{W}_2(\rho_\tau^n, \rho)^2}{2\,\tau} \right\} . \tag{4.1}$$

4.2.1 The Euler-Lagrange equation

Let $(\rho_\tau^n)_n$ be a sequence of minimisers. We will determine the equation satisfied by the minimisers. For this purpose we will a suitable variation of the minimiser ρ_τ^{n+1}. In the context of mass transport it is natural to proceed as follows: let $T_\varepsilon := \text{id} + \varepsilon\nabla\zeta$ with ζ smooth with compact support, we define ρ_ε the image measure by the transport map T_ε of ρ_τ^{n+1}

$$\rho_\varepsilon = T_\varepsilon \# \rho_\tau^{n+1} .$$

Exercice 23. Consider $\nabla\varphi$ be the Brenier map between ρ_τ^n and ρ_τ^{n+1}.

1. Prove that
$$\mathcal{W}_2\left(\rho_\tau^n, \rho_\varepsilon\right)^2 \leq \frac{1}{2}\int \left|x - \nabla\varphi(x) - \varepsilon\nabla\zeta\circ\nabla\varphi(x)\right|^2 \,\mathrm{d}\rho_\tau^n(x) .$$

2. Prove that,
$$\lim_{\varepsilon\to 0} \frac{\mathcal{W}_2(\rho_\tau^n, \rho_\varepsilon)^2 - \mathcal{W}_2(\rho_\tau^n, \rho_\tau^{n+1})^2}{\varepsilon} = \int \zeta(x)\,\mathrm{d}\left(\rho_\tau^{n+1}(x) - \rho_\tau^n(x)\right) + O\left(\mathcal{W}_2(\rho_\tau^n, \rho_\tau^{n+1})^2\right) .$$

▶

1. As $(\text{id}+\varepsilon\nabla\zeta)\circ\nabla\varphi$ pushes forward ρ_τ^n onto ρ_ε, eventhough it is not the Brenier map between the two, we have the expected result.

2. As $\nabla\varphi$ be the Brenier map between ρ_τ^n and ρ_τ^{n+1}, we have
$$\mathcal{W}_2(\rho_\tau^n, \rho_\tau^{n+1})^2 = \int |x - \nabla\varphi(x)|^2 \,\mathrm{d}\rho_\tau^n(x) .$$
Hence
$$\begin{aligned}\mathcal{W}_2(\rho_\tau^n, \rho_\varepsilon)^2 - \mathcal{W}_2(\rho_\tau^n, \rho_\tau^{n+1})^2 &= \int \left[|x - \nabla\varphi(x) - \varepsilon\nabla\zeta\circ\nabla\varphi(x)|^2 - |x - \nabla\varphi(x)|^2\right] \mathrm{d}\rho_\tau^n(x) \\ &= -\int \left[|x - \nabla\varphi(x)|^2 - |x - \nabla\varphi(x) - \varepsilon\nabla\zeta\circ\nabla\varphi(x)|^2\right] \mathrm{d}\rho_\tau^n(x) .\end{aligned}$$
Dividing by ε and letting ε going to 0, we obtain
$$\lim_{\varepsilon\to 0} \frac{\mathcal{W}_2(\rho_\tau^n, \rho_\varepsilon)^2 - \mathcal{W}_2(\rho_\tau^n, \rho_\tau^{n+1})^2}{\varepsilon} = \int \langle \nabla\varphi(x) - x, \nabla\zeta\circ\nabla\varphi(x)\rangle \,\mathrm{d}\rho_\tau^n(x) .$$
By the Taylor expansion $f(x+h) - f(x) = \nabla f(x)\cdot h + O(h^2)$, we have
$$\zeta\left[\nabla\varphi(x) + x - \nabla\varphi(x)\right] - \zeta\left[\nabla\varphi(x)\right] = \langle \nabla\zeta\left[\nabla\varphi(x)\right], x - \nabla\varphi(x)\rangle + O\left(|x - \nabla\varphi(x)|^2\right)$$
which we can recast
$$\zeta\left[\nabla\varphi(x)\right] - \zeta(x) = \langle \nabla\varphi(x) - x, \nabla\zeta\left[\nabla\varphi(x)\right]\rangle + O\left(|x - \nabla\varphi(x)|^2\right)$$
Hence the above term can be recast as
$$\begin{aligned}&\int \zeta\circ\nabla\varphi(x)\,\mathrm{d}\rho_\tau^n(x) - \int \zeta(x)\,\mathrm{d}\rho_\tau^n(x) + O\left(\int |x - \nabla\varphi(x)|^2 \,\mathrm{d}\rho_\tau^n(x)\right) \\ &\qquad = \int \zeta(x)\,\mathrm{d}\rho_\tau^{n+1}(x) - \int \zeta(x)\,\mathrm{d}\rho_\tau^n(x) + O\left(\mathcal{W}_2(\rho_\tau^n, \rho_\tau^{n+1})^2\right)\end{aligned}$$

Exercice 24. 1. Prove that

$$E(\rho_\varepsilon) = \int \log\left(\frac{\rho_\tau^{n+1}}{\det(I+\varepsilon D^2\zeta)}\right)\,\mathrm{d}\rho_\tau^{n+1} + \int V\left(x+\varepsilon\nabla\zeta(x)\right)\,\mathrm{d}\rho_\tau^{n+1}$$

2. Deduce

$$\lim_{\varepsilon\to 0}\frac{E(\rho_\varepsilon)-E(\rho_\tau^{n+1})}{\varepsilon} = -\int \Delta\zeta(x)\,\mathrm{d}\rho_\tau^{n+1} + \int \langle\nabla V(x)\,,\nabla\zeta(x)\rangle\,\mathrm{d}\rho_\tau^{n+1}\,.$$

▶

1. This is a direct consequence of the definition of the image-measure and of the Monge-Ampère equation as it was done in the previous chapter.
2. We have

$$\frac{E(\rho_\varepsilon)-E(\rho_\tau^{n+1})}{\varepsilon} = -\int \frac{\log\left(\det(I+\varepsilon D^2\zeta)\right)}{\varepsilon}\,\mathrm{d}\rho_\tau^{n+1} + \int \frac{V\left(x+\varepsilon\nabla\zeta(x)\right)-V(x)}{\varepsilon}\,\mathrm{d}\rho_\tau^{n+1}\,.$$

As

$$\lim_{\varepsilon\to 0}\frac{\log\left(\det(I+\varepsilon D^2\zeta)\right)-\log\left(\det(I)\right)}{\varepsilon} = \mathrm{Tr}(D^2\zeta) = \Delta\zeta\,,$$

we obtain the desired result.

Exercice 25. 1. Deduce from the two previous exercises that

$$\int \zeta(x)\,\mathrm{d}\frac{\rho_\tau^{n+1}(x)-\rho_\tau^n(x)}{\tau} + O\left(\mathcal{W}_2(\rho_\tau^n,\rho_\tau^{n+1})^2\right) = \int \left[(\Delta\zeta(x)+\langle\nabla V(x)\,,\nabla\zeta(x)\rangle\right]\,\mathrm{d}\rho_\tau^{n+1}\,.$$

▶

1. As ρ_τ^{n+1} is a minimiser we have

$$\frac{\mathcal{W}_2(\rho_\tau^n,\rho_\varepsilon)^2}{2\tau} + E(\rho_\varepsilon) - \frac{\mathcal{W}_2(\rho_\tau^n,\rho_\tau^{n+1})^2}{2\tau} - E(\rho_\tau^{n+1}) \geq 0\,.$$

Dividing by ε and letting ε going to 0 we obtain

$$\int \zeta(x)\,\mathrm{d}\frac{\rho_\tau^{n+1}(x)-\rho_\tau^n(x)}{\tau} + O\left(\mathcal{W}_2(\rho_\tau^n,\rho_\tau^{n+1})^2\right) - \int \left[\Delta\zeta(x)+\langle\nabla V(x)\,,\nabla\zeta(x)\rangle\right]\,\mathrm{d}\rho_\tau^{n+1} \geq 0\,.$$

Changing $\nabla\zeta$ to $-\nabla\zeta$ we obtain

$$\int \zeta(x)\,\mathrm{d}\frac{\rho_\tau^{n+1}(x)-\rho_\tau^n(x)}{\tau} + O\left(\mathcal{W}_2(\rho_\tau^n,\rho_\tau^{n+1})^2\right) = \int \left[\Delta\zeta(x)+\langle\nabla V(x)\,,\nabla\zeta(x)\rangle\right]\,\mathrm{d}\rho_\tau^{n+1}\,.$$

4.2.2 A priori estimates

Define the function ρ_τ by

$$\rho_\tau(x) = \rho_\tau^n \qquad \text{on } [n\tau,(n+1)\tau)\,.$$

Let us first prove that E is bounded from below by a constant on $\mathcal{P}_2(\mathbb{R}^d)$:

Exercice 26. 1. Prove that

$$E(\rho) \geq \int \log(\rho)\,\mathrm{d}\rho\,.$$

2. Prove that $E(\rho) \geq 0$.

▶

1. The second moment is non-negative.
2. As $x \mapsto x \log x$ is convex and ρ is a probability measure. We use Jensen's inequality:
$$E(\rho) \geq \int \log(\rho)\, \mathrm{d}\rho \geq \int \mathrm{d}\rho \log\left(\int \mathrm{d}\rho\right) = 0\,.$$

We can now prove the estimates:

Exercice 27. 1. Prove the energy estimate:
$$\sup_n E(\rho_\tau^n) \leq E(\rho_0)$$

2. Prove the total square distance estimate:
$$\sum_{n\geq 0} \mathcal{W}_2(\rho_\tau^n, \rho_\tau^{n+1})^2 \leq 2\tau\left[E(\rho_0) - \inf E\right]$$

3. Deduce a Hölder estimate: there exists $C > 0$ such that
$$\mathcal{W}_2(\rho_\tau(s), \rho_\tau(t))^2 \leq C\left[t - s + \tau\right]\,.$$

▶

1. As ρ_τ^{n+1} is a minimiser of
$$E(\rho) + \frac{\mathcal{W}_2(\rho_\tau^n, \rho)^2}{2\,\tau}\,,$$
we obviously have
$$E(\rho_\tau^{n+1}) + \frac{\mathcal{W}_2(\rho_\tau^n, \rho_\tau^{n+1})^2}{2\,\tau} \leq E(\rho_\tau^n)$$
Hence for any n, $E(\rho_\tau^{n+1}) \leq E(\rho_\tau^n)$ and taking the supremum we obtain the desired result.
2. By summing up the above inequality we obtain the total square distance estimate.
3. By the triangular and the Cauchy-Schwarz inequality:
$$\begin{aligned}\mathcal{W}_2(\rho_\tau(s), \rho_\tau(t))^2 &\leq \sum_{\mathbb{E}(s/\tau)\leq n\leq \mathbb{E}(t/\tau)+1} \mathcal{W}_2(\rho_\tau^n, \rho_\tau^{n+1})^2\\ &\leq \left[\frac{t-s}{\tau} + 1\right]\sum_{n\geq 0}\mathcal{W}_2(\rho_\tau^n, \rho_\tau^{n+1})^2\\ &\leq 2\left[E(\rho_0) - \inf E\right]\left[t - s + \tau\right]\,.\end{aligned}$$
which leads to the desired estimate with $C := 2\left[E(\rho_0) - \inf E\right]$.

4.2.3 Compactness

To pass to the limit when $\tau \to 0$ we can rely on the energy estimates. Indeed, the energy estimates ensures
$$E(\rho_\tau) = \int \rho_\tau \log\, \mathrm{d}\rho_\tau + \int |x|^2\, \mathrm{d}\rho_\tau < C =: E(\rho_0)\,.$$
As a consequence (ρ_τ) cannot vanish as
$$\int_{|x|>R} \mathrm{d}\rho_\tau \leq \frac{1}{R^2}\int |x|^2\, \mathrm{d}\rho \leq \frac{C}{R^2} \to_{R\to\infty} 0\,.$$
On the other hand (ρ_τ) cannot concentrate as $\rho_\tau \log \rho\tau$ is uniformly bounded by Carleman's estimate, Lemma 2 and
$$\int_{\rho_\tau>\alpha} \mathrm{d}\rho_\tau \leq \frac{1}{\log\alpha}\int \log\rho_\tau\, \mathrm{d}\rho_\tau \leq \frac{1}{\log\alpha}\int |\log\rho_\tau|\, \mathrm{d}\rho_\tau\,.$$

Lemma 2 (Carleman's estimate). *For any $u \in \mathcal{L}^1_+(\mathbb{R}^d)$, if $\int_{\mathbb{R}^d} |x|^2 \,\mathrm{d}u$ and $\int_{\mathbb{R}^d} \log u \,\mathrm{d}u$ are bounded from above, then $u \log u$ is uniformly bounded:*

$$\int_{\mathbb{R}^d} |\log u| \,\mathrm{d}u \leq \int_{\mathbb{R}^d} \Big(\log u + |x|^2\Big) \,\mathrm{d}u + 2\log(2\pi) \int_{\mathbb{R}^d} \mathrm{d}u + \frac{2}{\mathrm{e}}\,.$$

Proof. Let $\bar{u} := u \mathbb{1}_{\{u \leq 1\}}$ and $m = \int_{\mathbb{R}^d} \bar{u} \,\mathrm{d}x \leq M$. Then

$$\int_{\mathbb{R}^d} \bar{u}\Big(\log \bar{u} + \frac{1}{2}|x|^2\Big) \,\mathrm{d}x = \int_{\mathbb{R}^d} U \log U \,\mathrm{d}\mu - m \log(2\pi)$$

where $U := \bar{u}/\mu$, $\mathrm{d}\mu(x) = \mu(x)\,\mathrm{d}x$ and $\mu(x) = (2\pi)^{-d/2} e^{-|x|^2/2}$. By Jensen's inequality,

$$\int_{\mathbb{R}^d} U \log U \,\mathrm{d}\mu \geq \Big(\int_{\mathbb{R}^d} U \,d\mu\Big) \log\Big(\int_{\mathbb{R}^d} U \,\mathrm{d}\mu\Big) = m \log m\,,$$

$$\int_{\mathbb{R}^d} \bar{u} \log \bar{u} \,\mathrm{d}x \geq m \log\Big(\frac{m}{2\pi}\Big) - \frac{1}{2}\int_{\mathbb{R}^d} |x|^2 \bar{u} \,\mathrm{d}x \geq -\frac{1}{e} - M\log(2\pi) - \frac{1}{2}\int_{\mathbb{R}^d} |x|^2 \bar{u} \,\mathrm{d}x\,.$$

Using

$$\int_{\mathbb{R}^d} u\,|\log u| \,\mathrm{d}x = \int_{\mathbb{R}^d} u \log u \,\mathrm{d}x - 2\int_{\mathbb{R}^d} \bar{u} \log \bar{u} \,\mathrm{d}x\,,$$

this completes the proof. □

4.2.4 Limit equation

Let t_1 and t_2 be two arbitrary times. Summing up from $n_1 = [t_1/\tau]$ to $n_2 = [t_2/\tau]+1$ we obtained in Section 4.2.1

$$\int [\rho_\tau(t_2,x) - \rho_\tau(t_1,x)]\,\zeta(x) + O\left(\sum_{n=n_1}^{n_2} \mathcal{W}_2(\rho_\tau^n, \rho_\tau^{n+1})^2\right) = \int_{t_1}^{t_2} \int [\Delta\zeta(x) + \langle \nabla V(x)\,, \nabla\zeta(x)\rangle] \,\mathrm{d}\rho_\tau(t)$$

Using the Hölder continuity estimate we deduce

$$\int [\rho_\tau(t_2,x) - \rho_\tau(t_1,x)]\,\zeta(x) + O(\tau) = \int_{t_1}^{t_2} \int [\Delta\zeta(x) + \langle \nabla V(x)\,, \nabla\zeta(x)\rangle] \,\mathrm{d}\rho_\tau(t) \tag{4.2}$$

As a consequence of the previous section, $(\rho_\tau)_\tau$ is weakly-L^1 compact and Hölder-1/2 continuous in time. All this is sufficient to guarantee that there exists a subsequence of $(\rho_\tau)_\tau$, which we will stille denote $(\rho_\tau)_\tau$, which converges to some function ρ in the weak-L^1 topology. As a conseunce we can pass to the limit when τ goes to 0 in (4.2) to obtain

$$\int [\rho(t_2,x) - \rho(t_1,x)]\,\zeta(x) = \int_{t_1}^{t_2} \int [\Delta\zeta(x) + \langle \nabla V(x)\,, \nabla\zeta(x)\rangle] \,\mathrm{d}\rho(t)$$

which is a weak formulation of the so called linear Fokker-Planck equation with potential V:

$$\frac{\partial\rho}{\partial t} = \Delta\rho + \nabla\cdot(\rho\nabla V)\,.$$

We do not aim, in this lecture, to study further this partial differential equation but its study in very classical. We refer to the lecture on functional analysis on Fourrier transform or to the lecture of Pr. Miclo. Let us still note that the stationary solution has to satisfy

$$\Delta\bar{\rho} + \nabla\cdot(\bar{\rho}\nabla V) = 0$$

or equivalently that there exists C such that

$$\nabla\bar{\rho} + \bar{\rho}\nabla V = C$$

which we can integrate to:

$$\bar{\rho} = \lambda \mathrm{e}^{-V} \qquad \text{where} \qquad \lambda := \int \mathrm{e}^{-V(x)}\,\mathrm{d}x$$

to ensure that $\bar{\rho}$ is a probability densiy. This is Gibbs measure and in the classical case when $V(x) = |x|^2/2$ the stationary state is a Gassian measure. Using functional inequalities we can prove that the solution ρ converges exponentially fast to this Gaussian measure but this is another story...

Chapter 5

An application to economics

There are many appications of optimal transports to economics, including nice connexion with the matching-problem, see [10]. We refer to the book [11] for different application. Here we focus on an application to urban equilibrium that we developped with my colleagues Carlier, Mossay, and Santambroggio.

5.1 Game with a continuum of players

There are two main type of non-cooperative games

- *conflicts* among a small group of agents each of whom can make unilateral decisions which may significantly affect the welfare of the others as well as his own welfare (ex.: Card games, battles between opposing generals, etc.).
- the *individualistic* but not deliberately adversary behaviour of a large number of agents, none of whom alone is able to affect the circumstances of anyone except himself but whose actions in the aggregate determine the environment in which all must live (ex.: competitive markets).

In individualistic game it is often relevant to consider that only the distribution of the actions of the players matters rather than specifying the actions of each single individual. Such games are called *anonymous*.

Nash's result provides an existence result for all these games but such equilibria are more and more difficult to describe when the number of agents increases, see [13]:

> Microeconomics is full of elegant and persuasive arguments about the behaviour of representative firms and representative consumers in competitive markets in general, but in contrast it requires a great deal of elaborate computation to show that even a simple model of non-cooperative exchange yields competitive outcomes when there are many traders.

However, as was noticed by [24] when we deal with individualistics-type non-cooperative games:

> Institutions having a large number of competing participants are common in political and economic life (...) game theory has not yet been able so far to produce much in the way of fundamental principles of mass competition that might help to explain how they operate in practice. (...) it might be worth while to spend a little effort looking at the behaviour of existing n-person solution concepts, as n becomes very large.

Such phenomenon is perfectly known in the physics literature and was already pointed out in [26] in very explicit terms:

> An almost exact theory of a gas, containing about 10^{25} freely moving particles, is incomparably easier than that of the solar system, made up of 9 major bodies. (...)

> It is a well known phenomenon in many branches of the exact and physical sciences that very great numbers are often easier to handle than those of medium size. This is of course due to the excellent possibility of applying the laws of statistics and probabilities in the first case. (...)
>
> When the number of participants becomes really great, some hope emerges that the influence of every particular participant will become negligible, and that the above difficulties may recede and a more conventional theory become possible

Such a notion of games with a *continuum of players* was formalised in [2] as

> The most natural model for this purpose contains a continuum of participants, similar to the continuum of points on a line or the continuum of particles in a fluid. (...)
>
> The continuum can be considered an approximation to the "true" situation in which there is a large but finite number of particles.(...)
>
> The purpose of adopting the continuous approximation is to make available the powerful and elegant methods of a branch of mathematics called 'analysis' in a situation where treatment by finite methods would be much more difficult or hopeless.

Aumann even points that

> The choice of the unit interval as a model for the set of traders is of no particular significance. In technical terms, T can be any measure space without atoms. The condition that T have no atoms is precisely what is needed to ensure that each individual trader have no influence.

First results for this kind of game were obtained for *non-atomic games* described in [23] by

> Non-atomic games enable us to analyse a conflict situation where the single player has no influence on the situation but the aggregative behaviour of "large" sets of players can change the payoffs. The examples are numerous: Elections, many small buyers from a few competing firms, drivers that can choose among several roads, and so on.

[23] proves the existence of an equilibria in a non-atomic game with an arbitrary finite number of pure strategies when the payoff of the player only depend on the mean of distribution of all the payers actions. Note that such a result is wrong for a finite number of players as can be seen in the matching penny example. See [16] for more references.

Mathematically, we use here the formalism of [17]: given a space of players types X endowed with a probability measure $\mu \in \mathcal{M}(X)$ (which gives the exogenous distribution of the type of the agents), an action space Y and a cost Γ: $X \times Y \times \mathcal{M}(Y) \to \mathbb{R}$. The θ-type agents taking action x pay the cost $\Gamma(\theta, x, \nu)$ where ν is the distribution of the players' actions. A *Cournot-Nash equilibrium* is a joint probability measure $\gamma \in \mathcal{M}(X \times Y)$ with first marginal μ such that

$$\gamma(\{(\theta, x) \in X \times Y \ : \ \Gamma(\theta, x, \nu) = \min_{z \in Y} \Gamma(\theta, z, \nu)\}) = 1$$

where ν represents γ's second marginal.

Concerning the externalities exerted by the action of all the players we will consider two types

- Rivalry/Congestion: *The utility of the agent decreases when the number of players who choose the same action increases.* Examples:
  - Consumption of the same public good (motorway game),
  - Food supply in an habitat decreases with the number of its users (ex. Sticklebacks),
  - More populated areas lead to higher *competition for land.*
- Social interactions *The utility of the agents increases because some other agents play a similar action.* Examples:

- Location to go spend holidays,
- Quality of a product in a differentiated industry (technological choice),
- The agents benefit from *social interactions* but there is a cost to access to distant agents,

The results of [23, 17] and extensions see [15] assume that $\nu \mapsto \Gamma(\cdot,\cdot,\nu)$ is, in some sense, continuous. However this continuity assumption excludes the case of a purely local dependence which is relevant to capture congestion effects.

5.2 Beckman's model

We consider that agents have a type $x \in X$ and that the cost depend on the type:

$$\begin{array}{rcl} \mathcal{V} : X \times Y \times \mathcal{P}(Y) & \to & \mathbb{R} \\ (x,y,\nu) & \mapsto & \mathcal{V}(x,y,\nu) \end{array}$$

However in order to develop our optimal transport approach it seems necessary to assume that the costs $\mathcal{V}$ is on the additive separable form

$$\mathcal{V}(x,y,\nu) := c(x,y) + \mathcal{V}[\nu](y)\ .$$

5.2.1 The equilibrium model

For simplicity we consider X and Y to be compact metric spaces but unbounded domains could be considered too. To give a sense to the local congestion term we take m_0 be a reference probability measure. The definition of equilibrium reads:

Definition 16 (Cournot-Nash equilibrium). *Let μ be a Borel probability measure on X giving the exogenous distribution of the type of the agents. A Cournot-Nash equilibrium is a joint probability measure $\gamma \in \mathcal{P}(X \times Y)$ such that*

- $\gamma \in \Pi(\mu,\nu) := \{\gamma \in \mathcal{P}(X \times Y) \ : \ \pi_{X\#}\gamma = \mu,\ \pi_{Y\#}\gamma = \nu\}$
- *and there exists $\varphi \in \mathcal{C}(X)$ such that for all $x \in X$ and a.e. y*

$$\begin{cases} c(x,y) + \mathcal{V}[\nu](y) \geq \varphi(x) \\ c(x,y) + \mathcal{V}[\nu](y) = \varphi(x) \qquad \text{for } \gamma\text{-a.e. } (x,y). \end{cases} \tag{5.1}$$

To fix the idea let us give two examples where such equilibria arise:

Holiday choice

Let us consider a population of agents whose location is distributed according to some probability distribution $\mu \in \mathcal{P}(X)$ where X is some compact subset of $\mathbb{R}^2$ (say). These agents have to choose their holidays destination (possibly in mixed strategy). The set of possible holiday destinations is some compact subset of the plane Y (it can be X, a finite set, ...). The commuting cost from x to y is $c(x,y)$. In addition to the commuting cost, agents incur costs resulting from interactions with other agents, this is captured by a map $\nu \mapsto \mathcal{V}[\nu]$ that can be modelled as follows. A natural effect that has to be taken into account is congestion, *i.e.* the fact that more crowded location results in more disutility for the agents. Congestion thus requires to consider local effects and actually imposes that ν is not too concentrated; a way to capture this is to impose that ν is absolutely continuous with respect to some reference probability measure m_0. Still denoting by ν the Radon-Nikodym derivative of ν, a natural congestion cost is of the form $y \mapsto f(\nu(y))$ with f non-decreasing. In addition to the negative externality due to congestion effect, there may be a positive externality effect due to the positive social interactions between agents which can be captured through a non-local term of the form $y \mapsto \int_Y \phi(y,z)\, \mathrm{d}\nu(z)$ where for instance $\phi(y,.)$ is minimal for $z = y$ so that the previous term represents a cost for being far from the rest of the population. Finally, the presence of purely geographical factors (e.g. distance to the sea) can be reflected by a term of the form $y \mapsto v(y)$. The total externality cost generated by the distribution ν combines the three effects of congestion, positive interactions and geographical factors and can then be taken of the form

$$\mathcal{V}[\nu](y) = f(\nu(y)) + \int_Y \phi(y,z)\, \mathrm{d}\nu(z) + v(y).$$

Technological choice

Consider now a simple model of technological choice in the presence of externalities. There is a set of consumers indexed by a type $x \in X$ drawn according to the probability μ, and a set of technologies Y for a certain good (cell-phone, computer, tablet...). On the supply side, assume there is a single profit maximising profit firm with convex production cost $F(y,.)$ producing technology y, the supply (equals demand at equilibrium) of this firm is thus determined by the marginal pricing rule $p(y) = \partial_\nu F(y, \nu(y))$. Agents aim to minimise with respect to y a total cost which is the sum of their individual purchasing cost $c(x,y)+p(y) = c(x,y)+\partial_\nu F(y,\nu(y))$ and an additional usage/maintenance or accessibility cost which is positively affected by the number of consumers having purchased similar technologies *i.e.* a term of the form $\int_Y \phi(y,z)\, \mathrm{d}\nu(z)$ where ϕ is increasing in the distance between technologies y and z.

5.2.2 Connexion with optimal transport

Let $\mathcal{W}_c$ be the value of the Monge-Kantorovich optimal transport problem:

$$\mathcal{W}_c(\mu,\nu) := \inf_{\gamma \in \Pi(\mu,\nu)} \iint_{X\times Y} c(x,y)\, \mathrm{d}\gamma(x,y)$$

and let us also denote by $\Pi_o(\mu,\nu)$ the set of optimal transport plans *i.e.*

$$\Pi_o(\mu,\nu) := \{\gamma \in \Pi(\mu,\nu) \ : \ \iint_{X\times Y} c(x,y)\, \mathrm{d}\gamma(x,y) = \mathcal{W}_c(\mu,\nu)\}.$$

Proposition 1 (Connexion with optimal transport). *If γ is a Cournot-Nash equilibrium and ν denotes its second marginal then $\gamma \in \Pi_o(\mu,\nu)$.*

Proof. Indeed, let $\varphi \in \mathcal{C}(X)$ be such that (5.1) holds and let $\eta \in \Pi(\mu,\nu)$ then we have

$$\begin{aligned}\iint_{X\times Y} c(x,y)\, \mathrm{d}\eta(x,y) &\geq \iint_{X\times Y} (\varphi(x) - V[\nu](y))\, \mathrm{d}\eta(x,y)\\ &= \int_X \varphi(x)\, \mathrm{d}\mu(x) - \int_Y V[\nu](y)\, \mathrm{d}\nu(y) = \iint_{X\times Y} c(x,y)\, \mathrm{d}\gamma(x,y)\end{aligned}$$

so that $\gamma \in \Pi_o(\mu,\nu)$. □

In an euclidean setting, there are well-known conditions on c and μ which guarantee that such an optimal γ necessarily is pure whatever ν is. It is the case for instance if μ is absolutely continuous with respect to the Lebesgue measure, $c(x,y)$ is a smooth and strictly convex function of $x - y$ (see [20] who extended the seminal results of [5] in the quadratic cost case), or more generally, when it satisfies a generalised Spence-Mirrlees condition (see [8] for details):

Proposition 2 (Purification of equilibria). *Assume that $X = \overline{\Omega}$ where Ω is some open connected bounded subset of $\mathbb{R}^d$ with negligible boundary, that μ is absolutely continuous with respect to the Lebesgue measure, that c is differentiable with respect to its first argument, that $\nabla_x c$ is continuous on $\mathbb{R}^d \times Y$ and that it satisfies the generalised Spence-Mirrlees condition:*

$$\text{for every } x \in X, \text{ the map } y \in Y \mapsto \nabla_x c(x,y) \text{ is injective,}$$

then for every $\nu \in \mathcal{P}(Y)$, $\Pi_o(\mu,\nu)$ consists of a single element and the latter is of the form $\gamma = (\mathrm{id}, T)_\# \mu$ hence every Cournot-Nash equilibrium is pure.

5.2.3 Variational approach

In this section, we will see that in many relevant cases, one may obtain equilibria by the minimisation of some functional over a set of probability measures[1]. The main assumption for this variational approach to be valid is that the interaction map $\mathcal{V}[\nu]$ has the structure of a differential *i.e.* that $\mathcal{V}[\nu]$ can be seen as the first variation of some function $\nu \mapsto \mathcal{E}[\nu]$. In this case, the variational approach is based on the observation that the equilibrium condition is the first-order optimality condition for the minimisation of $\mathcal{W}_c(\mu, \nu) + \mathcal{E}[\nu]$.

Let $\mathcal{D}$ be defined by

$$\mathcal{D} := \{\nu \in \mathcal{L}^1(m_0) \ : \ V[\nu] \in \mathcal{L}^1(\nu)\} = \{\nu \in \mathcal{L}^1(m_0) \ : \ \int_Y |V[\nu]| \, \mathrm{d}\nu < +\infty\}.$$

Assume $a(\nu^\alpha - 1) \leq f(y, \nu) \leq b(\nu^\alpha + 1)$, for $\alpha \geq 0$ and that ϕ is symmetric.

$$\mathcal{V}[\nu](y) := f(\nu(y)) + \int_Y \phi(y, z) \, \mathrm{d}\nu(z) + v(y) \ .$$

is the differential of $\mathcal{E}$ in $\mathcal{D} := \mathcal{P}(Y) \cap \mathcal{L}^{\alpha+1}$ where

$$\mathcal{E}[\nu] = \int_Y F(\nu(y)) \, \mathrm{d}y + \frac{1}{2} \iint_{Y \times Y} \phi(y, z) \, \mathrm{d}\nu(y) \, \mathrm{d}\nu(z) + \int_Y v(y) \, \mathrm{d}\nu(y) \ ,$$

with $F' = f$.

Define the variational problem:

$$\inf_{\nu \in \mathcal{D}} \mathcal{J}_\mu[\nu] \quad \text{where} \quad \mathcal{J}_\mu[\nu] := \mathcal{W}_c(\mu, \nu) + \mathcal{E}[\nu]. \tag{5.2}$$

Proposition 3 (Existence of equilibria, [4]). *If*

- ν *solves* (5.2),
- *and* γ *is an optimal transport between* μ *and* ν

then γ *is a Cournot-Nash equilibrium.*

Moreover, (5.2) *admits minimisers in* $\mathcal{P}(Y) \cap \mathcal{L}^{\alpha+1}$ *so that there exists Cournot-Nash equilibria.*

The proof of this result uses a usual vertical perturbation of the functional. Let us mention that the optimality condition for (5.2) is the following: there is a constant M such that

$$\begin{cases} \varphi^c + \mathcal{V}[\nu] \geq M \\ \varphi^c + \mathcal{V}[\nu] = M \qquad \nu\text{-a.e.} \ , \end{cases} \tag{5.3}$$

where φ^c is the c-transform of φ.

5.2.4 Hidden convexity and further uniqueness results

So far, our variational approach has enabled us to prove the existence of equilibria by the minimisation problem (5.2). However, the previous results are not totally satisfying since in general there might exist equilibria that are not minimisers and even if we are only interested in the special equilibria obtained by minimisation, optimality conditions:

$$\nu(y) = f^{-1} \left(M - \varphi^c(y) - \int_Y \phi(y, z) \, \mathrm{d}\nu(z) \right). \tag{5.4}$$

are not tractable enough to provide a full characterisation. In the case where

$$\mathcal{E}[\nu] = \int_Y F(y, \nu(y)) \, \mathrm{d}m_0(y) + \frac{1}{2} \iint_{Y \times Y} \phi(y, z) \, \mathrm{d}\nu(y) \, \mathrm{d}\nu(z)$$

[1] Note the analogy with the variational approach of [21] for potential games, *i.e.* games for which the equilibria can be obtained by minimising some potential function.

there is a competition between the convexity of the congestion term that favours dispersion and the non-convexity of the interaction term so that in general nothing can be said about the convexity of $\mathcal{E}$ in the usual sense. However, for people familiar with optimal transport there is an hidden convexity due to R. McCann [18] which would restore the equivalence between being a minimiser and being an equilibrium. It would also give new uniqueness and characterisation results.

Throughout this section, we will assume the following:

- $X = Y = \overline{\Omega}$ where Ω is some open bounded convex subset of $\mathbb{R}^d$,
- μ is absolutely continuous with respect to the Lebesgue measure (that will be the reference measure m_0 from now on) and has a positive density on Ω,
- c is quadratic *i.e.*
$$c(x,y) := \frac{1}{2}|x-y|^2, \ (x,y) \in \mathbb{R}^d \times \mathbb{R}^d,$$
- $\mathcal{V}$ again takes the form
$$\mathcal{V}[\nu](y) = f(\nu(y)) + v(y) + \int_Y \phi(y,z)\, \mathrm{d}\nu(z)$$
where v is convex, f satisfies the Inada condition and $\phi \in \mathbb{C}(\mathbb{R}^d \times \mathbb{R}^d)$ is symmetric and $\mathcal{C}^{1,1}_{\text{loc}}$ (*i.e.* $\mathcal{C}^1$ with a locally Lipschitz gradient).

The variational problem (5.2) then takes the form

$$\inf_{\nu \in \mathcal{P}(\overline{\Omega})} \mathcal{J}_\mu[\nu] \quad \text{where} \quad \mathcal{J}_\mu[\nu] := \frac{1}{2}\mathcal{W}_2^2(\mu,\nu) + \mathcal{E}[\nu] \tag{5.5}$$

with $\mathcal{W}_2^2(\mu,\nu)$ is the squared-2-Wasserstein distance between μ and ν i.e.:

$$\mathcal{W}_2^2(\mu,\nu) := \inf_{\gamma \in \Pi(\mu,\nu)} \int_{X^2} |x-y|^2\, \mathrm{d}\gamma(x,y).$$

Two more structural assumptions are needed to guarantee the strict convexity of $\mathcal{J}_\mu$ along generalised geodesics with base μ, namely McCann's condition:

$$\nu \mapsto \nu^d F(\nu^{-d}) \text{ is convex non-increasing on } (0,+\infty) \tag{5.6}$$

and that ϕ is convex. Note that McCann's condition is satisfied for standard utility function as the power functions ν^m with an exponent larger than 1 as well as by the entropy $\log(\nu)$.

Proposition 4 (Uniqueness of the equilibrium, [4])**.** *Under the above assumptions there is a unique equilibrium (which is actually pure).*

5.2.5 Welfare analysis

It would be tempting to interpret the above results as a kind of welfare theorem. If a planner would decide where to allocate the players, he would do it in order to maximise the total social cost $\mathrm{SC}[\nu]$ defined by

$$\begin{aligned}
\mathrm{SC}[\nu] &= \int_{X\times Y} (c(x,y) + \mathcal{V}[\nu](y))\, \mathrm{d}\gamma(x,y) \\
&= \iint_{X\times Y} c(x,y)\, \mathrm{d}\gamma(x,y) + \int_Y \mathcal{V}[\nu](y)\, \mathrm{d}\nu(y) \\
&= \frac{1}{2}\mathcal{W}_2^2(\mu,\nu) + \int_Y \mathcal{V}[\nu](y)\, \mathrm{d}\nu(y) \\
&= \frac{1}{2}\mathcal{W}_2^2(\mu,\nu) + \int_Y f(\nu(y))\, \mathrm{d}\nu(y) + \int_Y V\, \mathrm{d}\nu + \int_Y \phi\, \mathrm{d}\nu,
\end{aligned}$$

However such a program leads to a result which differs from the equilibrium described above. Indeed, the second term $f(\nu)\nu$ is replaced in the functional minimised by the players by $F(\nu)$ (with $F' = f$) and the interaction term is divided by 2. This individual minimisation has of course no reason to correctly estimate the marginal effect of individual behaviour on the total social cost. In other words, there is some gap between the equilibrium and the efficient (social-cost minimising) configurations, and, since we are dealing with a situation with externalities, this is actually not surprising. The computation of the equilibrium and the optimum can be done numerically in dimension 1 by using the same kind of numerical computations, see Figure 5.1. The natural way to restore efficiency of the equilibrium is the design by some social planner of a

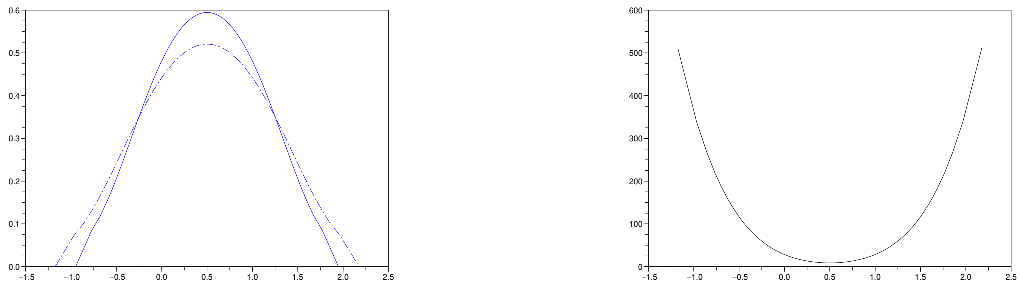

Figure 5.1: The optimum in continuous line and the equilibrium in dash line on the left. The corresponding taxes on the right.

proper system of tax/subsidies which, added to $\mathcal{V}[\nu]$, will implement the efficient configuration (or at least a stationary point of the social cost). Thanks to our variational approach, a tax system that restores the efficiency is easy to compute (up to an additive constant):

$$\mathrm{Tax}[\nu](y) = f(\nu(y))\,\nu(y) - F(\nu(y)) + \int_Y \phi(y,z)\,\mathrm{d}\nu(z).$$

The two terms in $\mathrm{Tax}[\nu]$ represent respectively a correction to the individual estimation of congestion cost and to the individual estimation of interaction cost.

5.3 Idea of the proof

5.3.1 A minimiser is an equilibrium

Let ν maximise $\mathcal{F}$ in $\mathcal{P}^{ac}(\mathcal{C})$. We consider some admissible density $\widetilde{\nu}$ and a family of perturbations indexed by $0 \leq \varepsilon \leq 1$,

$$\nu_\varepsilon = (1-\varepsilon)\nu + \varepsilon\widetilde{\nu}\ .$$

Given that ν maximises $\mathcal{F}$, we have

$$0 \geq \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathcal{F}\left[\nu_\varepsilon\right]_{|\varepsilon=0} = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathcal{U}[\nu_\varepsilon]_{|\varepsilon=0} + \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathcal{V}[\nu_\varepsilon]_{|\varepsilon=0} + \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathcal{W}[\nu_\varepsilon]_{|\varepsilon=0} \tag{5.7}$$

As $U' = u$, the derivative of the internal energy is given by

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathcal{U}[\nu_\varepsilon]_{|\varepsilon=0} = -\int U'(\nu(x))\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\nu_\varepsilon(x)\,\mathrm{d}x_{|\varepsilon=0} = -\int u(\nu(x))[\widetilde{\nu}(x) - \nu(x)]\,\mathrm{d}x\ .$$

The potential energy is easy to compute and its derivative is given by

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathcal{V}[\nu_\varepsilon]_{|\varepsilon=0} = -\int v(x)(\widetilde{\nu}(x) - \nu(x))\,\mathrm{d}x.$$

By using the symmetry of ϕ, the derivative of the interaction energy is given by

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathcal{W}[\nu_\varepsilon]_{|\varepsilon=0} \\
=& -\frac{1}{2}\iint \phi(x-y)\,(\nu(x)[\widetilde{\nu}(y)-\nu(y)]+[\widetilde{\nu}(x)-\nu(x)]\nu(y)) \\
=& -\iint \phi(x-y)\nu(y)[\widetilde{\nu}(x)-\nu(x)]\,\mathrm{d}x\,\mathrm{d}y \\
=& -\int \phi{*}\nu(x)(\widetilde{\nu}(x)-\nu(x))\,\mathrm{d}x\,.
\end{aligned}$$

By plugging the expressions into (5.7) we obtain

$$\int [u(\nu(x))+v(x)+\phi{*}\nu(x)]\,\widetilde{\nu}(x)\,\mathrm{d}x \geq \int [u(\nu(x)+v(x)+\phi{*}\nu(x)]\,\nu(x)\,\mathrm{d}x\,.$$

As this inequality holds for any admissible density $\widetilde{\nu}$, this implies that the density ν is concentrated on the set where the function $u(\nu)+\phi{*}\nu+v$ realises its minimal value.

5.3.2 Equivalence between equilibrium and minimiser

Assume now that ν is a solution to (5.5). Let $\tilde{\nu}$ be some admissible density and $\nabla\psi$ the optimal transport from ν to $\tilde{\nu}$. Consider $\psi_\varepsilon(x) := (1-\varepsilon)x^2/2+\varepsilon\psi$ and the family of perturbations defined for $0\leq\varepsilon\leq 1$ by $\nu_\varepsilon := \nabla\psi_{\varepsilon\#}\nu$. The Monge-Ampère equation: $\nu(x)=\nu_\varepsilon(\nabla\psi_\varepsilon(x))\det(D^2\psi_\varepsilon(x))$ is equivalent to

$$\nu_\varepsilon(y) = \frac{\nu(\nabla\psi_\varepsilon^{-1}(y))}{\det(D^2\psi_\varepsilon(\nabla\psi_\varepsilon^{-1}(y))}\,. \tag{5.8}$$

We will give an idea of the proof in the case when $\mathcal{E}$ is of the form

$$\mathcal{E}[\nu] = \int_Y F(\nu)\,\mathrm{d}y + \frac{1}{m+1}\int_{Y^{m+1}}\phi\,\mathrm{d}\nu^{\otimes(m+1)}$$

but a generalisation to any sum of symmetric interaction terms of different orders m is straightforward.

As ν is a minimiser, the derivative of $\mathcal{J}_\mu[\nu_\varepsilon]$ in ε is 0. Let us compute

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}_{|\varepsilon=0}\mathcal{J}_\mu[\nu_\varepsilon] = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}_{|\varepsilon=0}\mathcal{W}_2^2(\mu,\nu_\varepsilon) + \frac{\mathrm{d}}{\mathrm{d}\varepsilon}_{|\varepsilon=0}\int_Y F(\nu_\varepsilon)\,\mathrm{d}y + \frac{1}{m+1}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}_{|\varepsilon=0}\int_{Y^{m+1}}\phi\,\mathrm{d}\nu_\varepsilon^{\otimes(m+1)}.$$

• By (5.8) and the change of variable $x=\nabla\psi_\varepsilon^{-1}(y)$, the differential of the second term formally gives

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\varepsilon}_{|\varepsilon=0}\int_Y F(\nu_\varepsilon)\,\mathrm{d}m_0 &= \frac{\mathrm{d}}{\mathrm{d}\varepsilon}_{|\varepsilon=0}\int_Y F\left(\frac{\nu(\nabla\psi_\varepsilon^{-1}(y))}{\det(D^2\psi_\varepsilon(\nabla\psi_\varepsilon^{-1}(y)))}\right)\mathrm{d}y \\
&= \frac{\mathrm{d}}{\mathrm{d}\varepsilon}_{|\varepsilon=0}\int_Y F\left(\frac{\nu(y)}{\det(D^2\psi_\varepsilon(y))}\right)\det(D^2\psi_\varepsilon(y))\,\mathrm{d}y \\
&= -\int_Y \nu\,[\Delta\psi-d]\,F'(\nu)\,\mathrm{d}y + \int_Y F(\nu)\,[\Delta\psi-d]\,\mathrm{d}y \\
&= \int_Y \big[F(\nu)-\nu F'(\nu)\big]\,[\Delta\psi-d]\,\mathrm{d}y\,.
\end{aligned} \tag{5.9}$$

Where, as $\det(I+H)=1+\mathrm{tr}(H)+o(\|H\|)$, we have used

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}_{|\varepsilon=0}\det(D^2\psi_\varepsilon(y)) = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}_{|\varepsilon=0}\det(I+\varepsilon(D^2\psi-I)) = \Delta\psi-d\,.$$

By integrating by parts (5.9) we obtain

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}_{|\varepsilon=0} \int_Y F(\nu_\varepsilon)\,\mathrm{d}y = -\int_Y \nabla\left[F(\nu) - \nu F'(\nu)\right][\nabla\psi - \mathrm{id}]\ \mathrm{d}y \\ + \int_{\partial Y} \left[F(\nu) - \nu F'(\nu)\right][\nabla\psi - \mathrm{id}]\cdot n\ \mathrm{d}\sigma\ .$$

By convexity of Y, $(T - \mathrm{id})\cdot n \le 0$. By convexity of F, $x \mapsto F(x) - xF'(x)$ is non-increasing from $F(0) = 0$. So that the boundary term is non-positive and

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}_{|\varepsilon=0} \int_Y F(\nu_\varepsilon)\,\mathrm{d}m_0 \le -\int_Y \nabla\left[F(\nu) - \nu F'(\nu)\right][\nabla\psi - \mathrm{id}]\ \mathrm{d}y\ .$$

As $\nabla\left[F(\nu) - \nu F'(\nu)\right] = -\nu\nabla\left[F'(\nu)\right] = \nu\nabla\left[f(\nu)\right]$, we have

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}_{|\varepsilon=0} \int_Y F(\nu_\varepsilon)\,\mathrm{d}y \le -\int_Y \nu\nabla\left[f(\nu)\right][\nabla\psi - \mathrm{id}]\ \mathrm{d}y\ .$$

• By symmetry of ϕ and definition of the push-forward, the last term formally gives

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}\varepsilon}_{|\varepsilon=0} \int_{Y^{m+1}} \phi(y, z_1, \dots, z_m)\ \mathrm{d}\nu_\varepsilon(y)\ \mathrm{d}\nu_\varepsilon(z_1)\cdots\ \mathrm{d}\nu_\varepsilon(z_m) \\ &= \frac{\mathrm{d}}{\mathrm{d}\varepsilon}_{|\varepsilon=0} \int_{Y^{m+1}} \phi(\nabla\psi_\varepsilon(y), \nabla\psi_\varepsilon(z_1), \dots, \nabla\psi_\varepsilon(z_m))\ \mathrm{d}\nu^{\otimes(m+1)} \\ &= (m+1)\int_{Y^{m+1}} \nabla\phi(y, z_1, \dots, z_m)(\nabla\psi(y) - y)\ \mathrm{d}\nu^{\otimes(m+1)} \end{aligned}$$

• By the horizontal differentiability of the Monge-Kantorovich distance with $v = \nabla\psi - \mathrm{id}$, the differential of the first term is

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathcal{W}_2^2[\mu, \nu_\varepsilon]_{|\varepsilon=0} = \int_Y \langle y - \nabla\varphi^*(y), \nabla\psi - \mathrm{id}\rangle\ \mathrm{d}\nu(y)\ .$$

• Collecting the above computations we obtain

$$0 = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}_{|\varepsilon=0}\mathcal{J}_\mu[\nu_\varepsilon] \le \int_Y \nu\,\nabla\left[-f(\nu) + \int_{Y^m} \phi\nu^{\otimes m} + y - \nabla\varphi^*\right][\nabla\psi - \mathrm{id}]\ \mathrm{d}y\ .$$

So that, on the support of ν, there exists a constant M such that

$$-f(\nu) + \int_{Y^m} \phi\ \mathrm{d}\nu^{\otimes m} + \mathrm{id} - \nabla\varphi^* = M \quad m_0\text{-a.e.}\ .$$

This is exactly the optimality condition (5.3) as

$$\nabla\varphi^c(y) = \nabla\left(\frac{|y|^2}{2} - \varphi^*\right) = y - \nabla\varphi^*(y)\ .$$

In case $\nabla\varphi_\varepsilon$ is not regular enough div is the distributional divergence and a rigorous justification of this computation can be found in [25, Theorem 5.30] and [1, Theorem 10.4.13].

Appendix A

The continuity equation

Since we do not make any regularity assumptions on ρ_0, we have to understand the continuity equation in some appropriate weak sense, in the sense of distributions. We shall say that th family of probability measures $t \mapsto \rho_t$ is a measure valued solution of the continuity equation if for every $T > 0$, and every test function $\varphi \in \mathcal{C}^1([0,T] \times \mathbb{R}^d)$ supported in B_r such that $\varphi(T, \cdot) = 0$ we have

$$\int_0^T \int_{\mathbb{R}^d} [\partial_t \varphi(t,x) + v(t,x) \cdot \nabla \varphi(t,x)] \; \mathrm{d}\rho_t(x) \; \mathrm{d}t = -\int_{\mathbb{R}^d} \varphi(0,x) \; \mathrm{d}\rho_0(x) \; .$$

Let us prove that the measure-valued curve $t \mapsto X_t \# \rho_0$ is the only measure-valued solution of the continuity equation.

We have by Fubini's theorem

$$\int_0^T \int_{\mathbb{R}^d} [\partial_t \varphi(t,x) + v(t,x) \cdot \nabla \varphi(t,x)] \; \mathrm{d}\rho_t(x) \; \mathrm{d}t = \int_{\mathbb{R}^d} \int_0^T [\partial_t \varphi(t,x) + v(t,x) \cdot \nabla \varphi(t,x)] \; \mathrm{d}t \; \mathrm{d}\rho_t(x) \; .$$

Using that $\rho_t = X_t \# \rho_0$ we obtain

$$\begin{aligned} \int_0^T \int_{\mathbb{R}^d} [\partial_t \varphi(t,x) + v(t,x) \cdot \nabla \varphi(t,x)] \; \mathrm{d}\rho_t(x) \; \mathrm{d}t &= \int_{\mathbb{R}^d} \int_0^T [\partial_t \varphi(t, X_t(x)) + v(t, X_t(x)) \cdot \nabla \varphi(t, X_t(x))] \; \mathrm{d}t \; \mathrm{d}\rho_0(x) \\ &= \int_{\mathbb{R}^d} \int_0^T \frac{\mathrm{d}}{\mathrm{d}t} \varphi(t, X_t(x)) \; \mathrm{d}\rho_0 \\ &= -\int_{\mathbb{R}^d} \varphi(0,x) \; \mathrm{d}\rho_0(x) \; . \end{aligned}$$

We can even prove that $t \mapsto X_t \# \rho_0$ is the unique solution to the continuity equation.

Appendix B

Concavity of $t \mapsto \det(I_d - tS)^{1/d}$

Exercice 28. Let $x \geq 0$ and $N \in \mathbb{N}^*$. Consider

$$V_N(x) = \sup \left\{ \prod_{i=1}^{N} x_i \ : \ x_i \geq 0 \, , \sum_{i=1}^{N} x_i = x \right\} .$$

1. Compute V_1.
2. Prove that

$$V_N(x) = \sup_{y \in [0,x]} \left\{ y \, V_{N-1}(x-y) \right\}$$

3. Prove that

$$V_N(x) = \frac{x^N}{N^N}$$

4. Deduce the arithmetico-geometric inequality:

$$\left(\prod |x_i| \right)^{1/N} \leq \frac{1}{N} \sum_{i=1}^{N} |x_i| \, .$$

▶

1. We have obviously

$$V_1(x) = x$$

2. Let us consider the sequence solution to

$$V_N(x) = \sup_{y \in [0,x]} \left\{ y \, V_{N-1}(x-y) \right\}$$

We can rewrite it

$$V_N(x) = \sup_{x_N \in [0,x]} \left\{ x_N \, V_{N-1}(x - x_N) \right\}$$

or

$$V_N(x) = \sup_{x_N \in [0,x]} \left\{ x_N \sup_{x_{N-1} \in [0, x - x_N]} \left\{ x_{N-1} V_{N-2}(x - x_N - x_{N-1}) \right\} \right\}$$

Iterating the procedure we obtain

$$V_N(x) = \sup_{x_N \in [0,x]} \sup_{x_{N-1} \in [0, x - x_N]} \cdots \sup_{x_1 \in [0, x - \sum_{i=2}^{N} x_i]} \left\{ x_N \, x_{N-1} \cdots x_1 \right\}$$

The constraints require $x_i \geq 0$ for all $i \in \{1, \cdots, N\}$. For the bound from above, the most restrictive contraint is $x_1 \in [0, x - \sum_{i=2}^{N} x_i$ or $\sum_{i=1}^{N} x_i = x$, we can rewrite the problem as

$$V_N(x) = \left\{ \prod_{i=1}^{N} x_i \ : \ x_i \geq 0 \, , \sum_{i=1}^{N} x_i = x \right\} .$$

Which is the desired result.

3. Let $\mathcal{P}(n)$ be the proposition

$$V_n(x) = \frac{x^n}{n^n}$$

   - By the first question $\mathcal{P}(1)$ is true.
   - Assume that $\mathcal{P}(n)$ is true. We have

$$V_{n+1}(x) = \sup_{y\in[0,x]} \left\{ \frac{y^n}{n^n}(x-y) \right\} = \frac{x^{n+1}}{(n+1)^{n+1}} \, .$$

     Hence $\mathcal{P}(n)$ true implies that $\mathcal{P}(n+1)$ is true.
   - As a consquence $\mathcal{P}(n)$ is true for any $n \geq 1$.

4. We have for all positive x_i:

$$\frac{x^N}{N^N} \geq \left\{ \prod_{i=1}^{N} x_i \ : \ x_i \geq 0 \, , \sum_{i=1}^{N} x_i = x \right\}$$

   So that

$$\frac{1}{N}\sum_{i=1}^{N} x_i \geq \left(\prod_{i=1}^{N} x_i \right)^{1/N}$$

   The inequality is true for $|x_i|$.

Bibliography

[1] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2005.

[2] R. Aumann, *Existence of competitive equilibria in markets with a continuum of traders*, Econometrica, 32 (1964), pp. 39–50.

[3] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the monge-kantorovich mass transfer problem*, Numerische Mathematik, 84 (2000), pp. 375–393.

[4] A. Blanchet and G. Carlier, *Optimal transport and cournot-nash equilibria*, Mathematics of Operations research, 41 (2016), pp. 125–145.

[5] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math., 44 (1991), pp. 375–417.

[6] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland mathematics studies, North Holland Publishing Company, 1973.

[7] L. A. Caffarelli, *The regularity of mappings with a convex potential*, Journal of the American Mathematical Society, 5 (1992), pp. 99–104.

[8] G. Carlier, *Duality and existence for a class of mass transportation problems and economic applications*, Adv. in Math. Econ., 5 (2003), pp. 1–21.

[9] G. Carlier, *Optimal transportation and economic applications*, Lecture Notes, 18 (2012).

[10] P.-A. Chiappori, R. McCann, and B. Pass, *Multidimensional matching*, arXiv preprint arXiv:1604.05771, (2016).

[11] A. Galichon, *Optimal transport methods in economics*, Princeton University Press, 2018.

[12] W. Gangbo and R. J. McCann, *The geometry of optimal transportation*, (1996).

[13] E. J. Green and R. H. Porter, *Noncooperative collusion under imperfect price information*, Econometrica, 52 (1984).

[14] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal., 29 (1998), pp. 1–17.

[15] M. A. Kahn and Y. Sun, *Non-cooperative games with many players*, Handbook of Game Theory with Economic Applications, 3 (2002), pp. 1761–1808.

[16] M. Khan and Y. Sun, *Non-cooperative games with many players*, Handbook of Game Theory with Economic Applications, 3 (2002), pp. 1761–1808.

[17] A. Mas-Colell, *On a theorem of Schmeidler*, J. Math. Econ., 3 (1984), pp. 201–206.

[18] R. J. McCann, *Existence and uniqueness of monotone measure-preserving maps*, Duke Math. J., 80 (1995), pp. 309–323.

[19] ——, *A convexity principle for interacting cases*, Adv. Math., 128 (1997), pp. 153–179.

[20] R. J. McCann and W. Gangbo, *The geometry of optimal transportation*, Acta Math., 177 (1996), pp. 113–161.

[21] D. Monderer and L. Shapley, *Potential games*, Games and Economic Behavior, 14 (1996), pp. 124–143.

[22] F. Santambrogio, *Optimal transport for applied mathematicians*, vol. 87, Springer, 2015.

[23] D. Schmeidler, *Equilibrium points of nonatomic games*, J. Stat. Phys., 7 (1973), pp. 295–300.

[24] N. Z. Shapiro and L. S. Shapley, *Values of large games, i: A limit theorem*, Mathematics of Operations Research, 3 (1978).

[25] C. Villani, *Topics in optimal transportation*, vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2003.

[26] J. Von Neumann and O. Morgenstern, *Theory of games and economic behavior*, Princeton, NJ, US: Princeton University Press, 1944.