

Lecture Notes

*Optimal Transportation and
Economic Applications*

GUILLAUME CARLIER

IMA, NEW MATHEMATICAL MODELS IN ECONOMICS AND
FINANCE, JUNE 2010

Contents

1	Introduction to optimal transportation	6
1.1	The Monge and the Monge-Kantorovich problems	6
1.2	Some examples of transport maps	10
1.3	Remarks on existence and uniqueness	13
1.4	Kantorovich duality, c -concave functions	14
1.5	The unidimensional case	18
1.6	The discrete case	21
2	The case of strictly convex costs	23
2.1	Existence by duality	23
2.2	The quadratic case, Brenier's Theorem	25
2.3	Related PDEs: Monge-Ampère, Hamilton-Jacobi and the continuity equations	26
2.4	The borderline case of the distance cost and its connection with Beckmann's problem	29
3	Transport, economics, optima and equilibria	33
3.1	Matching and equilibria	34
3.2	Equilibria in discrete choice models	39
3.3	Urban economics I : optimal planning	41
3.4	Urban economics II : equilibria	43
3.5	Multidimensional screening	46
4	Congested Transport	50
4.1	Wardrop equilibria in a simple congested network model	50
4.2	Optimal transport with congestion and equilibria in a continuous framework	53
4.3	Duality for the short-term problem	58
4.4	Beckmann-like reformulation of the long-term problem	61
4.5	Dynamic setting : perspectives from mean-field games theory	63

5	Appendix	65
5.1	Convex duality	65
5.2	A rigorous proof of the Kantorovich duality	67
5.3	The Disintegration Theorem	69
5.4	The continuity equation	70
5.5	Moser's deformation argument	73
5.6	Wasserstein distances	74

Chapter 1

Introduction to optimal transportation

In this introductory chapter, we introduce the Monge and Monge-Kantorovich optimal transport problems. In these notes, we will mainly stay at a rather elementary (or heuristic) level, so we refer the interested reader to the recent books of Villani [73], [74] for a much more exhaustive and detailed account of the theory and its most recent developments (see also [65] for a more probabilistic-oriented presentation and the lecture notes [5] and [42] for a very comprehensive treatment of both the quadratic and distance cases).

Throughout these notes, given a metric space X , we shall denote by $\mathcal{M}_1^+(X)$ the set of Borel probability measures on X .

1.1 The Monge and the Monge-Kantorovich problems

In 1781, the french mathematician Gaspard Monge, first considered the problem of "remblais et déblais" which asks what is the most efficient (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 $T : A \rightarrow B$ ($T(x) \in B$ represents the destination of the element of mass initially located at $x \in A$), the total work involved is

$$\int_A |x - T(x)| dx$$

and one has to minimize it in the set of *volume preserving* maps $T : A \rightarrow B$. 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'll come back to the original Monge problem in more details in paragraph 2.4.

More generally, assume that X and Y are two compact (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 $c : X \times Y \rightarrow \mathbb{R}$. Transport maps are then maps that fulfill some mass conservation requirement that is naturally defined as follows:

Definition 1.1 *Let T be a Borel map $: X \rightarrow Y$, the push forward (or image measure) of μ through T is the Borel measure, denoted $T_{\#}\mu$ defined on Y by*

$$T_{\#}\mu(B) = \mu(T^{-1}(B)), \text{ for every Borel subset } B \text{ of } Y.$$

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

Let us remark that $T_{\#}\mu$ can equivalently be defined by the *change of variables* formula:

$$\int_Y \varphi dT_{\#}\mu = \int_X \varphi(T(x))d\mu(x), \forall \varphi \in C(Y)$$

so that the requirement that T is a transport can be reformulated as

$$\int_X \varphi(T(x))d\mu(x) = \int_Y \varphi(y)d\nu(y), \forall \varphi \in C(Y).$$

The (generalized) Monge's problem then consists in finding a cost minimizing transport between μ and ν , it thus reads

$$\inf_{T : T_{\#}\mu = \nu} \int_X c(x, T(x))d\mu(x). \tag{1.1}$$

A solution to this problem (if any!) is called an *optimal transport map* or a *Monge solution*.

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 $\mu = \delta_a$ then it is impossible to transport μ on a target measure that is not itself a Dirac mass! This example is somehow extreme and can be ruled out when one assumes that μ is nonatomic. But even for very "well-behaved" measures

(say measures with smooth densities, ρ_0 and ρ_1 on \mathbb{R}^d) the requirement that a diffeomorphism T is a transport reads as the highly non linear equation

$$|\det DT| \rho_1(T) = \rho_0$$

and there is no reason in general to think that we should restrict ourselves to diffeomorphisms. 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:

$$\inf_{\gamma \in \Pi(\mu, \nu)} \int_{X \times Y} c(x, y) d\gamma(x, y) \quad (1.2)$$

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

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

which can also be formulated as

$$\int_{X \times Y} (\varphi(x) + \psi(y)) d\gamma(x, y) = \int_X \varphi(x) d\mu(x) + \int_Y \psi(y) d\nu(y),$$

for every $(\varphi, \psi) \in C(X) \times C(Y)$.

The Monge-Kantorovich problem is much simpler than the Monge's one: $\Pi(\mu, \nu)$ is never empty since it contains $\mu \otimes \nu$ and moreover it is a linear problem. In addition, $\Pi(\mu, \nu)$ is obviously weakly-* compact and since c is continuous the criterion $\gamma \mapsto \int_{X \times Y} c d\gamma$ is continuous for the weak-* topology, hence we get for free:

Theorem 1.1 *The Monge-Kantorovich problem (1.2) admits solutions : such solutions are called optimal transport plans.*

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 $\gamma_T := (\text{id}, T)_\# \mu$ i.e

$$\int_{X \times Y} \varphi(x, y) d\gamma_T(x, y) := \int_X \varphi(x, T(x)) d\mu(x), \quad \forall \varphi \in C(X, Y)$$

in other words, one can canonically imbed transport maps into the set of transport plan. As for the cost, of course one has

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

This proves that the minimum in the Monge-Kantorovich problem is smaller than the infimum of the Monge problem but 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. We will see two situations where, one may solve Monge's problem this way:

- the discrete case, where transport maps simply are permutations and transport plans bistochastic matrices, as we shall see, a celebrated result of Birkhoff says that the extreme points of bistochastic measures are permutation matrices,
- the case of strictly convex costs where a careful inspection of the optimality conditions obtained via duality theory enables one to prove that optimal plans are actually induced by transport maps.

A related question is whether the Monge-Kantorovich is a relaxation of the Monge problem in the usual sense i.e. is it true that the minimum in (1.2) coincides with the infimum in (1.1). The answer is positive when μ is nonatomic:

Theorem 1.2 *If μ is nonatomic then the set $\{\gamma_T : T_{\#}\mu = \nu\}$ is weak-* dense in $\Pi(\mu, \nu)$ and therefore one has the relaxation relation:*

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

The proof of the density statement uses Lyapunov's convexity theorem, we refer to the lecture notes of Ambrosio [5] 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.

Let us end this paragraph, by remarking that there is an obvious probabilistic interpretation of the Monge-Kantorovich in terms of optimal coupling.

Indeed, one can see the Monge-Kantorovich problem as finding a coupling (X, Y) which minimizes the average cost

$$\mathbb{E}(c(X, Y))$$

among all pairs (X, Y) with prescribed marginal laws respectively μ and ν .

1.2 Some examples of transport maps

Before, going further, we wish to abandon for a while the question of optimality and spend some time by describing some examples of transport maps.

The monotone rearrangement

Let μ and ν be Borel probability measures on the real line such that μ has no atom, then set for every $x \in \mathbb{R}$:

$$T(x) := \inf\{t \in \mathbb{R} : \nu((-\infty, t]) > \mu((-\infty, x])\}$$

T is obviously monotone and $T_{\#}\mu = \nu$. It is also easy to see that T is the only monotone map such that $T_{\#}\mu = \nu$. It turns out, as we shall explain later on that the monotone rearrangement is in fact an optimal transport for a wide class of transportation costs (but this is very peculiar to the one-dimensional case where the order of \mathbb{R} plays a crucial role). Of course, the statistic-orienter reader will naturally have made the connection with the notion of *quantile*.

Knothe's transport

Let us assume (for simplicity) that $d = 2$ and that μ is absolutely continuous, let μ_1 and ν_1 be respectively the first marginal of μ and ν respectively. By the disintegration theorem (see appendix), we may write

$$\mu = \mu_1 \otimes \mu_2^{x_1}, \nu := \nu_1 \otimes \nu_2^{y_1}$$

i.e. $\mu_2^{x_1}$ is the conditional probability distribution of x_2 given x_1 according to μ . Or, put differently, for every continuous φ one has

$$\int \varphi(x_1, x_2) d\mu(x_1, x_2) = \int \left(\int \varphi(x_1, x_2) d\mu_2^{x_1}(x_2) \right) d\mu_1(x_1)$$

and a similar interpretation holds for ν^{y_1} . Now let T_1 be the monotone rearrangement from μ_1 to ν_1 . Then for fixed x_1 let $T_2(x_1, \cdot)$ be the monotone rearrangement from $\mu_2^{x_1}$ to $\nu_2^{T_1(x_1)}$. Finally set $T(x_1, x_2) := (T_1(x_1), T_2(x_1, x_2))$, for any test function φ , we thus have

$$\begin{aligned} \int \varphi(T(x)) d\mu(x) &= \int \left(\int \varphi(T_1(x_1), T_2(x_1, x_2)) d\mu_2^{x_1}(x_2) \right) d\mu_1(x_1) \\ &= \int \left(\int \varphi(T_1(x_1), y_2) d\nu_2^{T_1(x_1)}(y_2) \right) d\mu_1(x_1) \\ &= \int \left(\int \varphi(y_1, y_2) d\nu_2^{y_1}(y_2) \right) d\nu_1(x_1) = \int \varphi d\nu \end{aligned}$$

so that T transport μ to ν . Note that by construction the Jacobian matrix DT is triangular and has nonnegative entries on the diagonal, T is called Knothe's transport (or sometimes conditional quantile transform by statisticians, see [69]). One can of course define the Knothe transport between measures in \mathbb{R}^d , since the construction uses monotone rearrangements at each step, it is necessary that the successive disintegrations of μ are nonatomic.

Knothe [52] realized that this transport map could be a simple and powerful tool for proving certain geometric inequalities. Let us give a "transport proof" of the isoperimetric inequality. Let B be the unit ball of \mathbb{R}^d and A be another (regular enough) domain. Let T be the Knothe transport between $|A|^{-1}\chi_A$ and $|B|^{-1}\chi_B$ so that

$$\det(DT) = \frac{|B|}{|A|}.$$

Since by construction DT is triangular with nonnegative eigenvalues, the arithmetico-geometric inequality gives

$$\det(DT)^{1/d} \leq \frac{1}{d} \operatorname{tr}(DT) = \frac{1}{d} \operatorname{div}(T)$$

integrating, we obtain

$$|B|^{1/d} |A|^{1-1/d} \leq \frac{1}{d} \int_A \operatorname{div}(T) = \frac{1}{d} \int_{\partial A} T \cdot n$$

and since $T \in B$, we get

$$|B|^{1/d} |A|^{1-1/d} \leq \frac{1}{d} \operatorname{Per}(A) = |B| \frac{\operatorname{Per}(A)}{\operatorname{Per}(B)}$$

that is

$$\frac{|A|^{1-1/d}}{\operatorname{Per}(A)} \leq \frac{|B|^{1-1/d}}{\operatorname{Per}(B)}.$$

Moser's transport

We now consider the case of probability measures having smooth densities denoted respectively ρ_0 and ρ_1 defined on some smooth open bounded subset Ω of \mathbb{R}^d , we also assume that these densities are uniformly bounded away from 0 on Ω . It is then a natural question to try to construct a smooth diffeomorphism transporting ρ_0 to ρ_1 . In [64], Moser gave such a construction by a very elegant flow argument (see also Dacorogna and Moser [34] for sharp regularity results and Evans and Gangbo [43] who used a similar construction in the framework of the Monge problem). The construction is as follows, first solve the Laplace equation with Neumann boundary condition:

$$\Delta u = \rho_0 - \rho_1 \text{ in } \Omega, \quad \frac{\partial u}{\partial n} = 0 \text{ on } \partial\Omega.$$

This Neumann problem is solvable because $\rho_0 - \rho_1$ has zero mean and by standard elliptic regularity u is a smooth function. Then consider the interpolation

$$\rho_t := (1 - t)\rho_0 + t\rho_1, \quad t \in [0, 1]$$

as well as the smooth vector-field

$$v(t, x) := \frac{\nabla u(x)}{\rho_t(x)}.$$

And let X_t denote the flow of this vector field i.e. the solution of

$$\partial_t X_t(x) = v(t, X_t(x)), \quad X_0(x) = x.$$

This defines a family of diffeomorphisms $x \mapsto X_t(x)$, and one can prove that

$$\rho_t = X_{t\#}\rho_0$$

so that in particular $X_{1\#}\rho_0 = \rho_1$: X_1 is therefore a smooth diffeomorphism transporting ρ_0 to ρ_1 . In the appendix, we give two different proofs of the fact that $\rho_t = X_{t\#}\rho_0$. The first one is based on the fact, that by construction, ρ_t solves the *continuity equation*:

$$\partial_t \rho + \operatorname{div}(\rho v) = \rho_1 - \rho_0 + \Delta u = 0.$$

The second proof is more direct and follows Moser's initial deformation arguments. It is important to understand this flow construction, since we will use it again in chapter 6 devoted to congested transport. Let us finally remark that Moser's construction does much more than just constructing a smooth

and invertible transport between ρ_0 and ρ_1 : the flow at time t , X_t transports ρ_0 to the linear interpolation ρ_t for every $t \in [0, 1]$.

We end this paragraph devoted to examples by mentioning that it is possible to find a continuous curve $T : [0, 1] \rightarrow [0, 1] \times [0, 1]$ that is continuous and transports the unidimensional Lebesgue measure on the segment to the two-dimensional Lebesgue measure on the square. This may seem strange at first glance, but the existence and construction of such curves is a special case of *space-filling curves* that may be traced back to Peano in the early 1890's.

1.3 Remarks on existence and uniqueness

We have seen in theorem 1.2 that when the source measure is nonatomic, the infimum of the Monge problem coincides with the minimum of the Monge-Kantorovich problem. If we then take a minimizing sequence of transport maps (T_n) , by compactness of $\Pi(\mu, \nu)$, up to a subsequence one may assume that γ_{T_n} weakly $*$ converges to some optimal transport plan γ , now in general there is no reason why γ should be of the form γ_T . Nonexistence of an optimal transport typically happens when minimizing sequences exhibit strong oscillations. behavior. Let us now give a simple example: let μ be uniformly distributed on $[0, 1]$, ν be uniformly distributed on $[-1, 1]$ and take as cost function:

$$c(x, y) := (x^2 - y^2)^2.$$

Now let $n \in \mathbb{N}^*$ and divide $[0, 1]$ into the $2n$ intervals $[k/(2n), (k+1)/(2n)]$, $k = 0, \dots, 2n-1$ and then define

$$T_n(x) = \begin{cases} (2x - k/n) & \text{if } x \in [k/(2n), (k+1)/2n] \text{ with } k \text{ even} \\ (-2x + k/n) & \text{if } x \in [k/(2n), (k+1)/2n] \text{ with } k \text{ odd.} \end{cases}$$

It is easy to check $T_n\#\mu = \nu$ and that $\int_0^1 c(x, T_n(x))d\mu(x)$ tends to 0, hence the infimum of the Monge problem is zero. Now if there was an optimal transport T then since it has zero cost, one would have $T(x) \in \{-x; x\}$ a.e. and one could thus write

$$T(x) = (\chi_A(x) - \chi_{[0,1]\setminus A})x$$

for some measurable $A \subset [0, 1]$ with measure $1/2$ (note that $A = \{T \geq 0\}$) but since $T^{-1}(A) = A$, we should also have $\nu(A) = \mu(T^{-1}(A)) = 1/4$ which yields a contradiction. An optimal plan is given by

$$\gamma = \mu \otimes \left(\frac{1}{2}\delta_x + \frac{1}{2}\delta_{-x} \right)$$

that is the optimal plan consists in splitting the mass at x , sending half of it at $-x$ leaving the remaining mass at x . In this example, the cost is a polynomial, this suggests that existence or nonexistence is not only a matter of regularity but is also related to some structure of the cost.

As for nonuniqueness, let us consider the so-called *book shifting example*. Again in dimension 1 take, $c(x, y) = |x - y|$, μ uniform on $[0, 1]$ and ν uniform on $[1/2, 3/2]$, then one can check that the translation $T(x) = x + 1/2$ is optimal but the transport that consists in translating the mass on $[0, 1/2]$ to $[1, 3/2]$ and leaving the common mass yields the same total cost 1 (the translation is twice longer but the mass transported is half of the total mass).

1.4 Kantorovich duality, c -concave functions

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 is given in the appendix). 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.

Let γ be a nonnegative measure on $X \times Y$, then we have already remarked that $\gamma \in \Pi(\mu, \nu)$ if and only if for every $(\varphi, \psi) \in C(X) \times C(Y)$ one has:

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

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

If $\gamma \geq 0$, the quantity

$$\sup_{(\varphi, \psi) \in C(X) \times C(Y)} \int_X \varphi d\mu + \int_Y \psi d\nu - \int_{X \times Y} (\varphi(x) + \psi(y)) d\gamma(x, y)$$

is either $+\infty$ or 0 and is 0 exactly when γ is a transport plan. We may then rewrite (1.2) in the "inf-sup" form:

$$\inf_{\gamma \geq 0} \sup_{(\varphi, \psi) \in C(X) \times C(Y)} \left\{ \int_X \varphi d\mu + \int_Y \psi d\nu + \int_{X \times Y} (c(x, y) - \varphi(x) - \psi(y)) d\gamma(x, y) \right\}. \quad (1.3)$$

In an informal way, the dual problem is obtained by switching the inf and the sup in the previous program. We then remark that

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

is 0 whenever the inequality $c(x, y) \geq \varphi(x) + \psi(y)$ holds everywhere and $+\infty$ otherwise. The "sup-inf" problem then reads as

$$\sup_{(\varphi, \psi) \in C(X) \times C(Y)} \left\{ \int_X \varphi d\mu + \int_Y \psi d\nu : c(x, y) \geq \varphi(x) + \psi(y) \text{ on } X \times Y \right\} \quad (1.4)$$

This problem is called the dual¹ formulation of the transport problem (1.2). Let $\gamma \in \Pi(\mu, \nu)$ and (φ, ψ) be a pair of continuous functions that satisfy the constraint of (1.4), we then have

$$\int_{X \times Y} c(x, y) d\gamma(x, y) \geq \int_{X \times Y} (\varphi(x) + \psi(y)) d\gamma(x, y) = \int_X \varphi d\mu + \int_Y \psi d\nu$$

we therefore have the *weak duality* inequality

$$\min(1.2) \geq \sup(1.4)$$

and if we are lucky enough to find $\gamma \in \Pi(\mu, \nu)$ and an admissible pair (φ, ψ) such that

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

then γ is an optimal transport plan and (φ, ψ) solves (1.4). Kantorovich duality formula asserts (see the appendix for a proof) that there is no gap between the value of (1.2) and that (1.4) i.e. that

$$\min(1.2) = \sup(1.4).$$

Now, let us have a closer look at (1.4), first remark that the constraint on (φ, ψ) may be rewritten as

$$\varphi(x) \leq \inf_{y \in Y} \{c(x, y) - \psi(y)\}$$

i.e. $\varphi \leq \psi^c$ where ψ^c is the c -concave transform of ψ i.e. the rightmost member of the previous inequality:

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

Let us denote by J the criterion in (1.4):

$$J(\varphi, \psi) := \int_X \varphi d\mu + \int_Y \psi d\nu, \quad (\varphi, \psi) \in C(X) \times C(Y).$$

¹This is a slight abuse of language, since in view of the usual convex duality theory, it is (1.2) which naturally appears as the dual of (1.4): indeed (1.2) is posed on a space of measures i.e. in the dual of a space of continuous functions. See the appendix for details.

For an admissible pair (φ, ψ) since $\psi^c \geq \varphi$ we have

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

and by construction, (ψ^c, ψ) is also admissible. Remarking that for every $(x, y) \in X \times Y$

$$c(x, y) - \psi^c(x) \geq \psi(y)$$

we get

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

so that

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

and again (ψ^c, ψ^{cc}) is admissible (the fact that it remains continuous is easy to see). This "double c -concavification" trick enables one to reduce the optimization in (1.4) to pairs of the form (ψ^c, ψ) or even (ψ^c, ψ^{cc}) :

$$\sup(1.4) = \sup_{\psi \in C(Y)} J(\psi^c, \psi) = \sup_{\psi \in C(Y)} J(\psi^c, \psi^{cc}).$$

Pairs of the form (ψ^c, ψ^{cc}) are called conjugate pairs of c -concave functions. Indeed, we have by definition

$$\psi^{cc}(y) = \inf_{x \in X} \{c(x, y) - \psi^c(x)\} \tag{1.5}$$

and thus

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

but since $\psi \leq \psi^{cc}$ we also get

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

So that in addition to the conjugacy formula (1.5) we have a kind of symmetric one

$$\psi^c(x) = \inf_{y \in Y} \{c(x, y) - \psi^{cc}(y)\}. \tag{1.6}$$

It turns out that such pairs form a sufficiently rigid set (because their modulus of continuity is controlled by that of c) to get compactness and thus prove existence of a maximizer in (1.4):

Theorem 1.3 *The supremum is attained in (1.4) : there exists a pair of conjugate c -concave functions (ψ^{cc}, ψ^c) that solves (1.4).*

Proof:

Let us first remark that since μ and ν have the same total mass, for every constant λ and every pair (ψ, φ) we have $J(\varphi - \lambda, \psi + \lambda) = J(\varphi, \psi)$. We may therefore find a maximizing sequence of the form (ψ_n^{cc}, ψ_n^c) such that in addition

$$\min_X \psi_n^c = 0. \quad (1.7)$$

Let us denote by d_X and d_Y the distances on X and Y respectively and define the modulus of continuity of c by:

$$\omega_c(t) := \sup\{|c(x', y') - c(x, y)| : d_X(x, x') + d_Y(y, y') \leq t\}$$

since c is uniformly continuous $\omega_c(t)$ tends to 0 as $t \rightarrow 0^+$. Now it is easy to check that for every $(x, x') \in X^2$ and $(y, y') \in Y^2$ one has

$$|\psi_n^c(x) - \psi_n^c(x')| \leq \omega_c(d_X(x, x')), \quad |\psi_n^{cc}(y) - \psi_n^{cc}(y')| \leq \omega_c(d_Y(y, y'))$$

so that both sequences (ψ_n^c) and (ψ_n^{cc}) are uniformly equicontinuous. Moreover, by our normalization condition (1.7), we have the bounds

$$0 \leq \psi_n^c \leq \omega_c(\text{diam}(K))$$

from which we also deduce uniform bounds on ψ_n^{cc} . Thanks to Ascoli-Arzelà theorem, we may then assume that some (not relabeled) subsequence converges uniformly to some pair $(\bar{\varphi}, \bar{\psi})$, obviously $(\bar{\varphi}, \bar{\psi})$ satisfies the inequality constraint of (1.4) and since (ψ_n^c, ψ_n^{cc}) is a maximizing sequence $(\bar{\varphi}, \bar{\psi})$ solves (1.4). Now if one wants a solution that is a pair of c -concave functions, it is enough to consider $(\bar{\psi}^{cc}, \bar{\psi}^c)$.

□

To sum up, we now know that : the infimum in (1.2) and the supremum in (1.4) are attained and that both values coincide. Now if γ is *any* solution of (1.2) and if (ψ^c, ψ^{cc}) is *any* solution of (1.4), we could have

$$\int_{X \times Y} (c(x, y) - \varphi(x) - \psi(y)) d\gamma(x, y) = 0$$

in other words, one should have $c(x, y) - \psi^c(x) - \psi^{cc}(y) = 0$ γ -a.e. (and in fact, by continuity, this identity should hold on the support of γ). It means that γ is concentrated on the set of pairs (x, y) for which

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

If, one can prove that for μ -a.e. $x \in X$ there is a single $y \in Y$ that solves the minimization problem above, then this will prove that in fact γ has to be

induced by a map and thus prove the existence of an optimal transport map. We will see in the next chapter a class of costs for which this strategy works and actually provides existence of a Monge solution. Finally, let us remark that if $\gamma \in \Pi(\mu, \nu)$ is supported on the set of pairs (x, y) such that

$$\psi^c(x) + \psi^{cc}(y) = c(x, y)$$

then γ is in fact an optimal transport plan (and (ψ^c, ψ^{cc}) is an optimal pair for the dual).

Let us finally illustrate, what Kantorovich duality formula tells us in the case where $X = Y$ and $c(x, y) = d(x, y)$, the distance of X (which was the case originally considered by Monge himself). The constraint in the dual reads as

$$\varphi(x) + \psi(y) \leq d(x, y)$$

and we have seen that this can be replaced by

$$\varphi(x) = \inf_{y \in X} \{d(x, y) - \psi(y)\}$$

from which we immediately deduce that φ is 1-Lipschitz. But for such a 1-Lipschitz function, one has

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

This means that in the problem, it is enough to maximize over pairs of the form $(u, -u)$ with u 1-Lipschitz. Kantorovich duality formula then takes the form

$$\inf_{\gamma \in \Pi(\mu, \nu)} \int_{X \times X} d(x, y) d\gamma(x, y) = \sup \left\{ \int_X u d(\mu - \nu) : u \text{ 1-Lipschitz} \right\}. \quad (1.8)$$

We thus see an interesting feature of this particular case : the value of the optimal transport problem with d as cost is the dual of the Lipschitz seminorm (in particular it only depends on the difference $\mu - \nu$).

1.5 The unidimensional case

In the scalar case, one may take advantage of the natural order structure which leads to *rearrangement inequalities* which give a simple solution to a large class of transport problems on the real line. For the sake of simplicity, we will restrict ourselves here to the case $X = [0, 1]$ equipped with the Lebesgue measure denoted μ . Given a measurable, and bounded (say)

function $x : [0, 1] \rightarrow \mathbb{R}$, the rearrangement of x , denoted \tilde{x} is given by the (generalized) inverse of the distribution function of x :

$$\tilde{x}(t) := \inf\{\alpha \in \mathbb{R} : \mu(\{x \leq \alpha\}) > t\}.$$

By construction \tilde{x} is nondecreasing and x and \tilde{x} are *equimeasurable* i.e. $x_{\#}\mu = \tilde{x}_{\#}\mu$ (because x and \tilde{x} have the same cumulative distribution function) and in fact, \tilde{x} is the only such function (up to μ -a.e.). It turns out, that among equimeasurable functions to x , \tilde{x} is optimal for a wide class of costs. Let us start with the following which is nothing but a particular case of the celebrated Hardy-Littlewood inequality:

Proposition 1.1 *Let $\phi : [0, 1] \rightarrow \mathbb{R}_+$ be a nondecreasing bounded function and x be bounded and measurable, then*

$$\int_0^1 \phi(t)\tilde{x}(t)dt \geq \int_0^1 \phi(t)x(t)dt \quad (1.9)$$

in particular, \tilde{x} solves:

$$\sup \left\{ \int_0^1 \phi(t)z(t)dt : z_{\#}\mu = x_{\#}\mu \right\}.$$

Proof:

Adding a constant to x if necessary, we may assume that $x \geq 0$. Let us assume first that ϕ is of the form $\phi = \chi_{[a,1]}$ for some $a \in [0, 1]$, then by Fubini's theorem we have

$$\int_0^1 \phi(t)\tilde{x}(t)dt = \int_0^\infty \mu([a, 1] \cap \{x \geq \alpha\})d\alpha$$

and

$$\int_0^1 \phi(t)x(t)dt = \int_0^\infty \mu([a, 1] \cap \{\tilde{x} \geq \alpha\})d\alpha$$

Now, we remark that $\{x \geq \alpha\}$ and $\{\tilde{x} \geq \alpha\}$ have same measure, the latter being (for a.e. α) an interval of the form $[a(\alpha), 1]$, we thus have

$$\begin{aligned} \mu([a, 1] \cap \{x \geq \alpha\}) &\leq \min(1 - a, \mu(\{x \geq \alpha\})) \\ &= 1 - \max(a, a(\alpha)) = \mu([a, 1] \cap \{\tilde{x} \geq \alpha\}) \end{aligned}$$

which proves that (1.9) holds when $\phi = \chi_{[0,a]}$ and thus also when ϕ is any linear combination with nonnegative coefficients of such elementary functions. We conclude the general case by a standard approximation argument.

□

The previous result means that \tilde{x} is the optimal rearrangement of x as far as the correlation with a nondecreasing function is concerned. This is actually quite clear intuitively that if one wants to maximize the integral in (1.9), one should associate large values of x to large values of ϕ which is precisely what the monotone rearrangement does. The previous result extends to a more general class of integrands, namely those for which, this kind of complementarity property holds. Such integrands are sometimes referred to as *supermodular*. Here, we shall restrict ourselves to smooth functions and say that $L: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ is supermodular if its mixed second derivative is always nonnegative:

$$\frac{\partial^2 L}{\partial t \partial x} \geq 0.$$

The interpretation of this condition is again clear in terms of complementarity: larger values of x should be associated to larger values of t . Such a condition plays an important role in adverse selection problems in economics where it is usually referred to as the Spence-Mirrlees condition.

Proposition 1.2 *Let L be a supermodular function and x be bounded, and measurable, then*

$$\int_0^1 L(t, \tilde{x}(t)) dt \geq \int_0^1 L(t, x(t)) dt$$

in particular, \tilde{x} solves:

$$\sup \left\{ \int_0^1 L(t, z(t)) dt : z_{\#} \mu = x_{\#} \mu \right\}.$$

Proof:

Replacing $L(t, x)$ by $L(t, x) - L(0, x)$ if necessary, we may assume $L(0, x) = 0$ (by equimeasurability, adding a function of x only to L does not affect the inequality to be proven). We thus have:

$$\begin{aligned} \int_0^1 L(t, \tilde{x}(t)) dt - \int_0^1 L(t, x(t)) dt &= \int_0^1 \left(\int_0^t [\partial_t L(s, \tilde{x}(t)) - \partial_t L(s, x(t))] ds \right) dt \\ &= \int_0^1 \left(\int_s^1 [\partial_t L(s, \tilde{x}(t)) - \partial_t L(s, x(t))] dt \right) ds \\ &= \int_0^1 \left(\int_0^1 \chi_{[s,1]} [\partial_t L(s, \tilde{x}(t)) - \partial_t L(s, x(t))] dt \right) ds \end{aligned}$$

where we have used Fubini's theorem in the second line. Now, for fixed s , we remark that $t \mapsto \partial_t L(s, x(t))$ and $t \mapsto \partial_t L(s, \tilde{x}(t))$ are equimeasurable

and that the latter is nondecreasing since L is supermodular and \tilde{x} is nondecreasing. This implies that $t \mapsto \partial_t L(s, \tilde{x}(t))$ is the monotone rearrangement of $t \mapsto \partial_t L(s, x(t))$. Using (1.9) (with $\phi = \chi_{[s,1]}$), we obtain the desired inequality.

□

The requirement that L is smooth was only for conveniency and is by no means a restriction (supermodularity is preserved by convolution so that usual regularization arguments enable one to treat the general case). There is no difficulty to check that the previous rearrangement inequalities also hold when one replaces the Lebesgue measure by a general nonatomic measure on the real line. Let us finally remark that if c is convex, then $(t, x) \mapsto -c(t - x)$ is supermodular. In particular, one therefore has

$$\int_0^1 c(t - \tilde{x}(t))dt \leq \int_0^1 c(t - x(t))dt$$

which proves the optimality of the monotone rearrangement for *any* convex cost. In particular \tilde{x} solves:

$$\inf \left\{ \int_0^1 |t - z(t)|^p dt : z_{\#}\mu = x_{\#}\mu \right\}.$$

whatever the exponent $p \in [1, \infty)$ is! This is extremely specific to the one-dimensional case.

1.6 The discrete case

Let us consider now the case of finitely supported measures and equal masses:

$$\mu = \sum_{i=1}^N \delta_{x_i}, \quad \nu = \sum_{j=1}^N \delta_{y_j}$$

and denote by c_{ij} the cost of transporting x_i to y_j . The Monge-Kantorovich problem then takes the form of the linear program

$$\inf \left\{ \sum_{ij} c_{ij} \gamma_{ij} : \gamma_{ij} \geq 0, \sum_j \gamma_{ij} = 1, \sum_i \gamma_{ij} = 1 \right\}$$

this problem is often referred to as the assignment or Hitchcock's problem. The admissible set is the set of $N \times N$ matrices with nonnegative entries, such the sum of entries on each row and each column is 1, such matrices are

called *bistochastic*. Clearly, the set of bistochastic matrices form a convex and compact set, so in the assignment problem the minimal cost is attained at at least one extreme point. A permutation matrix is a matrix of the form $\gamma_{ij} = \delta_{j\sigma(i)}$ where σ is a permutation of $\{1, \dots, N\}$. In other words a permutation matrix is a bistochastic matrix that contains exactly one entry equal to 1 on each line and each column. The following result due to Birkhoff identifies the set of extremal points of the set of bistochastic matrices to the permutation matrices:

Theorem 1.4 *The set of extreme points of the set of bistochastic matrices coincides with the set of permutation matrices. In particular, the set of bistochastic measures is a polyhedron with $N!$ vertices and every bistochastic matrix is a convex combination of permutation matrices (as a consequence of the Krein-Milman's theorem).*

Proof:

First, it is easy to see that permutation matrices are extremal points. Let $\gamma = [\gamma_{ij}]$ be a bistochastic matrix and assume that it is not a permutation matrix: then there is some row index i_1 and two different column indices j_1 and j_2 such that γ_{i_1, j_1} and γ_{i_1, j_2} are in $(0, 1)$, then there is a $i_2 \neq i_1$ such that $\gamma_{i_2, j_2} > 0$ and $j_3 \neq j_2$ such that $\gamma_{i_2, j_3} > 0$. If $j_3 = j_1$ then $(i_1, j_1), (i_1, j_2), (i_2, j_2), (i_2, j_3)$ is a cycle, if $j_3 \neq j_1$, one repeats the argument. One easily finds that after a finite number of steps, there necessarily is a *cycle* in the matrix that is a succession of positions $(i'_1, j'_1), (i'_1, j'_2), \dots, (i'_k, j'_k), (i'_k, j'_{k+1})$ with $j'_{k+1} = j_1$ all with entries in $(0, 1)$ and such that i'_1, \dots, i'_k are all different as well as j'_1, \dots, j'_k , let us then set $i'_{k+1} = i'_1$ and let Δ be the matrix with zero entries outside the cycle, $+1$ at position (i'_l, j'_l) and -1 at position (i'_l, j'_{l+1}) . Then, for $\varepsilon > 0$ small enough $\gamma \pm \varepsilon \Delta$ is bistochastic and $\gamma = (\gamma + \varepsilon \Delta)/2 + (\gamma - \varepsilon \Delta)/2$ which contradicts the extremality of γ .

□

As a consequence, we deduce that there is a permutation matrix that solves the assignment problem. In other words, the discrete problem that consists in finding an optimal permutation can be reformulated as a linear programming problem.

A natural question, at this point is whether the previous arguments could give a general strategy to prove existence of optimal transport maps i.e. is it true, under quite general assumptions, that extreme points of the set of transport plans are in fact given by transport maps? Unfortunately, it is not the case in the continuous setting, for more on extremal transport plans, we refer for instance to the recent paper [4] and the references therein.

Chapter 2

The case of strictly convex costs

In this chapter, we shall prove existence of an optimal transport map when the cost is strictly convex and smooth. The first result in this vein is in the seminal paper of Yann Brenier [14] that solved the quadratic case, the generalization to more general strictly convex costs is due to Robert McCann and Wilfrid Gangbo [51]. In the final paragraph, we will discuss the borderline case of the distance as cost and emphasize its connections with a problem posed by Martin Beckmann in the 50's.

2.1 Existence by duality

We shall prove existence (and uniqueness) of an optimal transport map for the problem

$$\inf_{T_{\#}\mu=\nu} \int_B c(x - T(x))d\mu(x). \quad (2.1)$$

under the following (not optimal) assumptions:

- μ and ν are Borel probability measures supported on a ball B of \mathbb{R}^d and μ is absolutely continuous with respect to the d -dimensional Lebesgue measure, \mathcal{L}^d ,
- c is a C^1 and strictly convex function : $\mathbb{R}^d \rightarrow \mathbb{R}$.

Let $\gamma \in \Pi(\mu, \nu)$ be an optimal transport plan i.e. a solution of :

$$\min_{\gamma \in \Pi(\mu, \nu)} \int_{B \times B} c(x - y)d\gamma(x, y). \quad (2.2)$$

We shall prove that γ is necessarily induced by a transport map i.e. is of the form $\gamma = \gamma_T$ for some transport T . Let (φ, ψ) be an optimal solution to the dual problem as introduced in the previous chapter:

$$\sup \left\{ \int_B \varphi d\mu + \int_B \psi d\nu : \varphi(x) + \psi(y) \leq c(x - y) \right\}. \quad (2.3)$$

We already know that φ can be chosen as the c -transform of ψ i.e.

$$\varphi(x) := \min_{y \in B} \{c(x - y) - \psi(y)\} \quad (2.4)$$

and that γ is supported by the closed set

$$\{(x, y) \in B^2 : \varphi(x) = c(x, y) - \psi(y)\}.$$

One deduces from (2.4) that φ is Lipschitz continuous (with Lipschitz constant less than $\|\nabla c\|_{L^\infty(2B)}$). By a well-known theorem of Rademacher, φ is therefore differentiable \mathcal{L}^d -a.e. hence also μ -a.e., let us denote by S the negligible set where φ fails to be continuous, we then first have

Lemma 2.1 *Let $x \in B \setminus (S \cup \partial B)$ and $y \in B$ be such that*

$$\varphi(x) = c(x, y) - \psi(y)$$

then one has

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

Proof:

For h small enough, we have:

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

from which we deduce the desired result.

□

Now since c is strictly convex, ∇c is injective and we therefore deduce that for every $x \in B \setminus (S \cup \partial B)$, the set

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

consists of the single element:

$$T(x) = x - \nabla c^*(\nabla \varphi(x)). \quad (2.5)$$

Since boundary points and points where φ fails to be differentiable are μ -negligible, it is easy to check that $\gamma = \gamma_T$ and thus to deduce that T is an optimal transport map. We have also proved uniqueness since we have started with an arbitrary optimal plan γ and constructed T depending only on a solution to the dual problem. We therefore have:

Theorem 2.1 *There exists an optimal transport map T in problem (2.1). Moreover T is of the form:*

$$T(x) = x - \nabla c^*(\nabla \varphi(x)), \text{ for a.e. } x$$

for some c -concave potential φ . Uniqueness also holds : γ_T is the only optimal transport plan in (2.2), in particular if R is another optimal transport map then $T = R$ μ -a.e..

For refinements of this result, we refer to [51], see also [25] for other costs.

2.2 The quadratic case, Brenier's Theorem

Let us now have a closer look at the quadratic case of the quadratic cost which was initially solved by Yann Brenier in his pathbreaking article [14] i.e. let us consider

$$\inf_{T \# \mu = \nu} \int_B \frac{1}{2} |x - T(x)|^2 d\mu(x). \quad (2.6)$$

We already know that there is a unique optimal transport T that is characterized by the fact that

$$\varphi(x) + \psi(T(x)) = \frac{1}{2} |x - T(x)|^2$$

where φ and ψ are related by the conjugacy relations:

$$\varphi(x) = \inf_y \left\{ \frac{1}{2} |x - y|^2 - \psi(y) \right\}, \quad \psi(y) = \inf_x \left\{ \frac{1}{2} |x - y|^2 - \varphi(x) \right\}$$

which can be rewritten as

$$\frac{1}{2} |x|^2 - \varphi(x) = \sup_y \left\{ x \cdot y - \frac{1}{2} |y|^2 + \psi(y) \right\}$$

and

$$\frac{1}{2} |y|^2 - \psi(y) = \sup_x \left\{ x \cdot y - \frac{1}{2} |x|^2 + \varphi(x) \right\}$$

which means that the functions:

$$u := \frac{1}{2} |\cdot|^2 - \varphi, \quad v := \frac{1}{2} |\cdot|^2 - \psi$$

are convex and conjugate to each other in the usual sense of convex analysis i.e. $u = v^*$ and $v = u^*$. Now, in the quadratic case, formula (2.5) gives that T has the form

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

in other words, the optimal transport is *the gradient of a convex function* and it is actually a characterization of optimality. The optimal map ∇u is called the Brenier's map between μ and ν , it generalizes in some sense the notion of monotone rearrangement to the multidimensional setting. Brenier's theorem reads as the following

Theorem 2.2 *There exists a unique (up to μ negligible sets) map of the form $T = \nabla u$ with u convex that transports μ to ν , this map is also the optimal transport between μ to ν for the quadratic cost.*

If $\mu = \rho_0 \cdot \mathcal{L}^d$, $\nu = \rho_1 \cdot \mathcal{L}^d$, then the Brenier's map ∇u , at least formally satisfies the Monge-Ampère equation:

$$\det(D^2u)\rho_1(\nabla u) = \rho_0 \tag{2.7}$$

as for boundary condition, is the requirement that ∇u maps the support of ρ_0 onto that of ρ_1 . A deep regularity theory due to Luis Caffarelli ([22], [23]) establishes conditions under which the Brenier's map is in fact smooth.

2.3 Related PDEs: Monge-Ampère, Hamilton-Jacobi and the continuity equations

The Monge-Ampère equation

If $\mu = \rho_0 \cdot \mathcal{L}^d$, $\nu = \rho_1 \cdot \mathcal{L}^d$, then the Brenier's map ∇u , at least formally satisfies the Monge-Ampère equation:

$$\det(D^2u)\rho_1(\nabla u) = \rho_0 \tag{2.8}$$

as for boundary condition, is the requirement that ∇u maps the support of ρ_0 onto that of ρ_1 . A deep regularity theory due to Luis Caffarelli ([22], [23], [24]) establishes conditions under which the Brenier's map is in fact smooth.

The Brenier-Benamou dynamic formulation

So far, we have only considered a static framework, Brenier and Benamou in [9], [10] gave a very interesting and fruitful dynamic formulation of the quadratic optimal transport problem (but this can be generalized to more general convex costs) that we aim now to describe in an informal way. Let us assume that we have as source and target measures two probability measures

with densities ρ_0 and ρ_1 , the value of the corresponding quadratic optimal transportation is the squared 2-Wasserstein distance between ρ_0 and ρ_1 :

$$W_2^2(\rho_0, \rho_1) := \inf_{\gamma \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\gamma(x, y).$$

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 and minimizing some action functional. This curve of measure will be given by the solution of the *continuity equation*:

$$\partial_t \rho + \operatorname{div}(\rho v) = 0, \quad \rho|_{t=0} = \rho_0. \quad (2.9)$$

As explained in the appendix, when v is smooth, this equation captures the evolution of the spatial distribution of particles that are initially distributed according to ρ_0 and whose velocity is v . In other words, $\rho_t = X_t \# \rho_0$ where X_t is the flow of v :

$$X_0(x) = x, \quad \partial_t X_t(x) = v(t, X_t(x)).$$

The Brenier-Benamou problem then reads as the average kinetic energy minimization:

$$\inf_{(\rho, v)} \left\{ E(\rho, v) = \int_0^1 \int_{\mathbb{R}^d} |v(t, x)|^2 \rho_t(x) dx dt : (2.9) \text{ holds, } \rho|_{t=0} = \rho_0, \rho|_{t=1} = \rho_1 \right\}. \quad (2.10)$$

Let us now prove that the value of the Brenier-Benamou problem (2.10) coincides with $W_2^2(\rho_0, \rho_1)$. Let (ρ, v) be admissible for (2.10) (with v smooth) and let X_t be the flow of v so that $\rho_t = X_t \# \rho_0$, we then have

$$\int_{\mathbb{R}^d} |v(t, x)|^2 \rho_t(x) dx = \int_{\mathbb{R}^d} |v(t, X_t(x))|^2 \rho_0(x) dx. \quad (2.11)$$

By Fubini's theorem and Jensen's inequality, we get

$$\begin{aligned} E(\rho, v) &= \int_{\mathbb{R}^d} \left(\int_0^1 |\partial_t X_t(x)|^2 dt \right) \rho_0(x) dx \\ &\geq \int_{\mathbb{R}^d} \left| \int_0^1 \partial_t X_t(x) dt \right|^2 \rho_0(x) dx \\ &= \int_{\mathbb{R}^d} |X_1(x) - x|^2 \rho_0(x) dx \\ &\geq W_2^2(\rho_0, \rho_1) \end{aligned}$$

where in the last line, we have used the fact that X_1 is a transport. This proves that the value of (2.10) is larger than $W_2^2(\rho_0, \rho_1)$. To prove the converse inequality let ∇u be the Brenier's map and define:

$$X_t(x) = (1-t)x + t\nabla u(x) = \nabla u_t(x), \quad u_t := \frac{(1-t)}{2}|x|^2 + tu(x), \quad \rho_t = \nabla u_t \# \rho_0,$$

the associated velocity field is defined through:

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

and since $X_t = \nabla u_t$ is the gradient of a strictly convex function, it is at least formally invertible with inverse ∇u_t^* . This yields

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

By construction (ρ, v) is admissible for the Brenier-Benamou problem, and using (2.11) directly yields

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

which proves not only that the value of (2.10) equals $W_2^2(\rho_0, \rho_1)$, but also that (ρ, v) constructed above is optimal for the Brenier Benamou problem. Note the special form of the optimal curve of measures:

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

This optimal ρ_t is then obtained by interpolating linearly the optimal transport, this interpolation is called the McCann's interpolation.

If in (2.10), we make the change of variables $(\rho, v) \mapsto (\rho, \rho v)$, the continuity equation becomes linear and E becomes a convex functional (because $(\rho, m) \mapsto |m|^2/\rho$ is convex on $(0, +\infty) \times \mathbb{R}^d$). There is then a dual formulation to (2.10) that consists in maximizing the linear criterion:

$$\int_{\mathbb{R}^d} \varphi(1, x) \rho_1(x) dx - \int_{\mathbb{R}^d} \varphi(0, x) \rho_0(x) dx$$

among subsolutions of the Hamilton-Jacobi equation:

$$\partial_t \varphi + \frac{1}{2} |\nabla \varphi|^2 \leq 0.$$

We shall not give details here, but simply mention why Hamilton-Jacobi equations naturally enter the picture: we have used above the change of variables ∇u_t^* and it is well known that u_t^* actually solves an Hamilton-Jacobi equation of the type above.

2.4 The borderline case of the distance cost and its connection with Beckmann's problem

In this final paragraph, we discuss the case, originally considered by Monge himself, where the cost is $|x - y|$, $|\cdot|$ denoting the euclidean norm (for other norms we refer the interested reader to [32], [2]). It is a borderline case since the cost is convex but not strictly convex (note also that it is not differentiable on the diagonal). Existence of an optimal transport map is a very delicate issue that was only solved recently (see Evans and Gangbo [43], Caffarelli, Feldman and McCann [21], Ambrosio and Pratelli [1], Champion and De Pascale [31], [32]). The optimal transport problem (originally considered by Monge himself) reads:

$$\inf_{T: T\#\mu=\nu} \int_{\mathbb{R}^d} |x - T(x)| d\mu(x). \quad (2.12)$$

As usual, it is natural to introduce its Monge-Kantorovich relaxation, whose value is the 1-Wasserstein distance between μ and ν :

$$W_1(\mu, \nu) := \inf_{\gamma \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y| d\gamma(x, y). \quad (2.13)$$

As we have seen in the first chapter, the dual problem in the case where the cost is the distance reads as:

$$\sup \left\{ \int_{\mathbb{R}^d} u(x) d\mu(x) - \int_{\mathbb{R}^d} u(y) d\nu(y) : u \text{ 1-Lipschitz} \right\} \quad (2.14)$$

Let u be a Kantorovich potential i.e. a solution to this dual problem and let γ be an optimal transport plan, then, Kantorovich duality tells us that the extremality relation

$$u(x) - u(y) = |x - y|$$

holds on the support of γ , $\text{supp}(\gamma)$, and that it is actually a sufficient optimality condition. This means that if $(x, y) \in \text{supp}(\gamma)$ then u decreases with its maximal rate on the segment $[x, y]$; (maximal) segments on which u has slope -1 are called (maximal) *transport rays*. If, in addition, $x \neq y$ and if x is a point of differentiability of u , arguing as in the strictly convex case, one has

$$\nabla u(x) = \frac{x - y}{|x - y|}$$

i.e. the gradient of the Kantorovich potential gives the direction of transport rays. In other words, duality tells us in which direction x should be

transported but due to lack of strict convexity in the radial direction it does not give any information about the length of the displacement $|x - y|$. To recover this missing piece of information, extra efforts have to be spent. Note that, as the book-shifting counter-example indicates, there is no hope to have uniqueness, so extra selection principles have to be carefully designed in order to recover the missing piece of information about the displacement length. We refer to the papers cited in the beginning of this paragraph, for different such selection strategies and a proof of the following:

Theorem 2.3 *If μ is absolutely continuous with respect to \mathcal{L}^d then there exists an optimal transport map for the Monge problem (2.12).*

In [7], Beckmann proposed what he called a *continuous model of transportation* that can be described as follows. Assume that we are given an urban area Ω , which is an open bounded connected subset of \mathbb{R}^d with a smooth boundary, μ and ν the respective distributions of residents and services in the city. As a normalization, we may assume that μ and ν are probability measures on $\bar{\Omega}$ and that μ (respectively ν) also gives the distribution of consumption (respectively of production) so that the signed measure $\mu - \nu$ represents the local measure of excess demand. Following [7], we assume that the consumers' traffic is given by a traffic flow field, i.e. a vector field $\mathbf{y} : \Omega \rightarrow \mathbb{R}^d$ whose direction indicates the consumers' travel direction and whose modulus $|\mathbf{y}|$ is the intensity of traffic. The relationship between the excess demand and the traffic flow is obtained from an equilibrium condition as follows. There is equilibrium in a subregion $K \subset \Omega$ if the outflow of consumers equals the excess demand of K :

$$\int_{\partial K} \mathbf{y} \cdot n \, dS = (\mu - \nu)(K).$$

Since the previous has to hold for arbitrary K , this formally yields:

$$\operatorname{div} \mathbf{y} = \mu - \nu. \tag{2.15}$$

It is also assumed that the urban area is isolated, i.e. no traffic flow should cross the boundary of the city, hence:

$$\mathbf{y} \cdot n = 0 \text{ on } \partial\Omega. \tag{2.16}$$

One has to understand the conditions (2.15) and (2.16) in the weak sense (or in the sense of distributions) i.e.:

$$\int_{\Omega} \nabla \varphi \cdot \mathbf{y} = \int_{\bar{\Omega}} \varphi d(\nu - \mu), \quad \forall \varphi \in C^1(\bar{\Omega}).$$

If the transportation cost per consumer is assumed to be uniform, then the total transportation cost is given by the L^1 norm of \mathbf{y} . Beckmann therefore argued that one may define the transportation cost between μ and ν as the infimum of $\|\mathbf{y}\|_{L^1}$ subject to the equilibrium conditions (2.15)-(2.16). There is no reason why an L^1 solution should exist since L^1 is not reflexive. One therefore has to relax the problem in the space of measures which leads to minimize the total variation in the set of vector-valued measures \mathbf{y} that satisfy (2.15) and (2.16) in the sense of distributions. This gives the *minimal flow* problem:

$$\inf \{ \|\mathbf{y}\|_{\mathcal{M}} : \mathbf{y} \text{ satisfies (2.15)-(2.16)} \}. \quad (2.17)$$

What was not realized by Beckmann is that, problem (2.17) is tightly related to the Monge-Kantorovich (2.13) in the following sense:

Theorem 2.4 *The value of the minimal flow problem (2.17) coincides with $W_1(\mu, \nu)$. Moreover, if γ is an optimal transport plan for (2.13) then the vector-valued measure \mathbf{y} defined by*

$$\langle \mathbf{y}_\gamma, X \rangle := \int_{\bar{\Omega} \times \bar{\Omega}} \left(\int_0^1 X(x + t(y-x)) \cdot (y-x) dt \right) d\gamma(x, y), \quad \forall X \in C(\bar{\Omega}, \mathbb{R}^d)$$

is a solution of Beckmann's problem (2.17).

Proof:

First, let us check that \mathbf{y}_γ satisfies (2.15)-(2.16), let $\varphi \in C^1(\bar{\Omega})$, by definition

$$\begin{aligned} \langle \mathbf{y}_\gamma, \nabla \varphi \rangle &= \int_{\bar{\Omega} \times \bar{\Omega}} \left(\int_0^1 \nabla \varphi(x + t(y-x)) \cdot (y-x) dt \right) d\gamma(x, y) \\ &= \int_{\bar{\Omega} \times \bar{\Omega}} (\varphi(y) - \varphi(x)) d\gamma(x, y) = \int_{\bar{\Omega}} \varphi d(\nu - \mu). \end{aligned}$$

Now let u be a Kantorovich potential, since it is 1-Lipschitz function, for every \mathbf{y} that is admissible for (2.17), we have

$$\|\mathbf{y}\|_{\mathcal{M}} \geq \langle \mathbf{y}, -\nabla u \rangle = \int_{\bar{\Omega}} u d(\mu - \nu) = W_1(\mu, \nu)$$

which proves that the value of (2.17) is larger than that of (2.13). Next, we remark that

$$\|\mathbf{y}_\gamma\|_{\mathcal{M}} \leq \int_{\bar{\Omega} \times \bar{\Omega}} |y-x| d\gamma(x, y) = W_1(\mu, \nu)$$

which proves that the value of (2.17) is $W_1(\mu, \nu)$ and that \mathbf{y}_γ is optimal for (2.17). \square

Remark 1. One can directly deduce that the values of (2.17) and (2.14) coincide. We leave as an instructive exercise the proof of this fact as an application of the Fenchel-Rockafellar duality theorem. In other words, both Monge-Kantorovich problem (2.13) and the min-flow problem (2.17) have (2.14) as dual.

Remark 2. The total variation measure $|\mathbf{y}_\gamma|$ can be alternatively be defined as

$$|\mathbf{y}_\gamma|(A) = \int_{\overline{\Omega} \times \overline{\Omega}} \mathcal{H}^1([x, y] \cap A) d\gamma(x, y), \quad \forall A \text{ Borel}$$

so that it represents the total cumulated traffic in region A associated with the transport plan γ when consumers travel along straight lines (that are geodesics here). In the L^1 theory of optimal transport, $|\mathbf{y}_\gamma|$ is called the *transport density*. Regularity results for the transport density (which implies the existence of L^1 solutions to Beckmann's problem) were established by De Pascale and Pratelli [37] and Santambrogio [70].

Chapter 3

Transport, economics, optima and equilibria

This chapter is intended to illustrate how optimal transportation theory (as presented in chapter 1 or in the disguise of slight variants) may be a powerful and natural tool to attack a variety of problems in economics in particular equilibrium problems but also urban planning and multidimensional screening.

Before going further, let us indicate that the Kantorovich duality formula has a strong "economic flavour" and can be interpreted either as a decentralization principle or in equilibrium terms. This will be useful in the sequel to keep these interpretation in mind.

Let us first recall the Kantorovich duality formula in the form

$$\inf_{\gamma \in \Pi(\mu, \nu)} \int_{X \times Y} c(x, y) d\gamma(x, y) = \sup \left\{ \int_X \varphi d\mu + \int_Y \psi d\nu : \varphi(x) + \psi(y) \leq c(x, y) \right\} \quad (3.1)$$

The left hand-side is the least cost for a planner that chooses a total transport cost minimizing plan (planning which amount of coal should be sent from different mines to different steel factories in the USSR being a typical example). The right-hand side may be interpreted as a decentralized solution, indeed, interpret $\varphi(x)$ as a price paid per unit of mass taken out from location x and $\psi(y)$ represents the price paid per unit of mass delivered at location y . The right hand-side then represents the maximal profit from transport companies, their tariffs being constrained by the *reservation* constraint $\varphi(x) + \psi(y) \leq c(x, y)$ meaning that the total cost paid to the transport sector for shipping one unit from x to y should not exceed the "if I do it myself" cost $c(x, y)$. In other words, the Kantorovich duality formula expresses a decentralization principle according to which the planning solution and the

solution where transport is delegated to a decentralized transport sector have the same cost.

Let us now consider the discrete optimal transport problem:

$$\sup_{\gamma_{ij}} \sum_{ij} u_{ij} \gamma_{ij} : \sum_j \gamma_{ij} = \mu_i, \sum_i \gamma_{ij} = \nu_j, \gamma_{ij} \geq 0 \quad (3.2)$$

where the marginals satisfy the obvious compatibility constraint:

$$\sum_i \mu_i = \sum_j \nu_j.$$

Now interpret j as an indivisible object available in quantity ν_j , i as a type of customer with μ_i the number of customer of type i and think of u_{ij} as the valuation or utility of agent i consuming object j . Let us further assume that utility is transferable in the sense that agent i paying p for object j has net utility $u_{ij} - p$. Given a price system p_j , agent i of course chooses an object j that maximizes $u_{ij} - p_j$. The dual to (3.2) reads

$$\inf_{p_j} \left\{ \sum_j \nu_j p_j + \sum_i \mu_i \max_j (u_{ij} - p_j) \right\}. \quad (3.3)$$

and the equality of values of these two problems says that if γ_{ij} is optimal for (3.2) then

$$\gamma_{ij} > 0 \Rightarrow u_{ij} - p_j = \max_{j'} (u_{ij'} - p_{j'})$$

which means that j is a maximizing utility object given the price system p_j solving (3.3). This captures rational behavior of agents and the fact that γ_{ij} satisfy the constraints of (3.2) express that offer=demand for each type of object. A price system that solves (3.3) is therefore an equilibrium price system.

3.1 Matching and equilibria

Ivar Ekeland in [40], [41] showed that some equilibrium issues in matching situations as well as in hedonic models can be attacked by duality techniques that present great similarities with the Monge-Kantorovich problem. In an independent but related article, Chiappori, McCann and Nesheim [33] introduced a Monge-Kantorovich formulation and obtained new uniqueness results for the optimal transportation problem. In this paragraph, we briefly describe the *matching for teams* situations studied in Carlier Ekeland [27]

which generalizes the one-to-one matching model to the case of several populations. We consider a market where there is a single, indivisible good which comes in different qualities $z \in Z$. In the sequel, we will refer to Z as the *quality space*. The quality good z requires the formation of a *team* (say, one buyer, and a set of producers that have to gather to make the quality good available, a typical example is the market for houses). The different populations that constitute the teams are heterogeneous in the sense that each agent in these populations has her own utility (for a consumer) or marginal cost (for a producer) for the quality z . The data of the model are:

- a compact metric quality good Z ,
- compact metric spaces X_j , $j = 0, \dots, N$, modeling the different populations, a generic element $x_j \in X_j$ has to be interpreted as an agent type affecting her utility or cost/disutility function, each X_j is equipped with a Borel probability μ_j measure capturing the distribution of type x_j in population j ,
- (continuous) cost functions $c_j : X_j \times Z \rightarrow \mathbb{R}$, with the interpretation that $c_j(x_j, z)$ is the cost or disutility of an agent of population j with type x_j when she participates to a team that produces $z \in Z$,
- disutilities are all quasi-linear, which means that an agent of population j with type x_j who participates a team that produces z and gets monetary transfer w_j has total disutility:

$$c_j(x_j, z) - w_j.$$

One can think for instance that $j = 0$ corresponds to buyers and $j = 1, \dots, N$ to producers (mason, plumber etc... in the case of houses). In this case, one naturally has to interpret c_0 as $-u_0$ where u_0 is the (type and quality dependent) consumer utility function. In this case, for $j \geq 1$, w_j is the wage received by member j of the team and w_0 is minus the total price paid by the consumer, at equilibrium we shall require that are self-financing which will be expressed by the fact that the sum of all transfers is 0. Note that the quasi-linear (dis)utility assumption is a strong one that restricts the subsequent analysis to the so-called *transferable utility* case.

We are now looking for a system of (quality dependent) monetary transfers that clears the market for the quality good. A system of price transfers is a family of function $\varphi_j : Z \rightarrow \mathbb{R}$, and we shall require that it is balanced i.e.

fulfills the self-financing condition:

$$\sum_{j=0}^N \varphi_j(z) = 0, \quad \forall z \in Z.$$

For given transfers, optimal qualities for type x_j are determined by

$$\varphi_j^{c_j}(x_j) := \inf_{z \in Z} \{c_j(x_j, z) - \varphi_j(z)\}. \quad (3.4)$$

which is the indirect disutility which type x_j derives from the transfer φ_j . Note that $\varphi_j^{c_j}$ is nothing but the c_j -concave transform of the transfer function φ_j as defined in paragraph 1.4. Note that for every $(x_j, z) \in X_j \times Z$, one has the so-called *Young's inequality*

$$\varphi_j^{c_j}(x_j) + \varphi_j(z) \leq c_j(x_j, z)$$

and cost-minimizing qualities are characterized by

$$\varphi_j^{c_j}(x_j) + \varphi_j(z) = c_j(x_j, z). \quad (3.5)$$

This induces for each j , a coupling $\gamma_j \in \mathcal{M}_1^+(X_j \times Z)$ such that (3.5) holds γ_j a.e., the interpretation of $\gamma_j(A_j \times B)$ is the probability that an agent with type in A_j has an optimal quality choice in B (given the transfer scheme φ_j). Of course, the first marginal of the coupling γ_j , $\pi_{X_j} \gamma_j$ should be μ_j . The last equilibrium requirement is that the demand distribution for the quality good should be the same for any population. In other words the marginal on Z of the coupling γ_j should be independent of j (this common distribution is an equilibrium quality line). Putting everything together, this leads to

Definition 3.1 *A matching equilibrium consists of a family of transfers $\varphi_j \in C(Z, \mathbb{R})$, a family of probabilities $\gamma_j \in \mathcal{M}_1^+(X_j \times Z)$, $j = 0, \dots, N$ and a quality line $\nu \in \mathcal{M}_1^+(Z)$ such that:*

1. For all $z \in Z$:

$$\sum_{j=0}^N \varphi_j(z) = 0, \quad (3.6)$$

2. $\gamma_j \in \Pi(\mu_j, \nu)$ for every $j = 0, \dots, N$,

3. for every $j = 0, \dots, N$, one has:

$$\varphi_j^{c_j}(x_j) + \varphi_j(z) = c_j(x_j, z) \quad \gamma_j\text{-a.e. on } X_j \times Z.$$

Thanks to the Kantorovich duality formula

$$\begin{aligned} W_{c_j}(\mu, \nu) &:= \inf_{\gamma_j \in \Pi(\mu_j, \nu)} \int_{X_j \times Z} c_j(x_j, z) d\gamma_j(x, z) \\ &= \sup_{\varphi_j \in C(Z)} \left\{ \int_{X_j} \varphi_j^{c_j} d\mu_j + \int_Z \varphi_j d\nu \right\} \end{aligned}$$

we see that requirements 2 and 3 exactly mean that γ_j is an optimal plan for the optimal transport problem $W_{c_j}(\mu_j, \nu)$ and that φ_j solves its dual. Let us assume now that $(\varphi_j, \gamma_j, \nu)$ is a matching equilibrium. We thus have

$$W_{c_j}(\mu_j, \nu) = \int_{X_j \times Z} c_j(x_j, z) d\gamma_j(x_j, z) = \int_{X_j} \varphi_j^{c_j} d\mu_j + \int_Z \varphi_j d\nu.$$

Summing these equalities and using the balance condition (3.6) then yields:

$$\sum_{j=0}^d W_{c_j}(\mu_j, \nu) = \sum_{j=0}^d \int_{X_j} \varphi_j^{c_j} d\mu_j \quad (3.7)$$

Now let $\psi_j \in C(Z, \mathbb{R})$ be another balanced family of transfers:

$$\sum_{j=0}^N \psi_j(z) = 0, \quad \forall z \in Z. \quad (3.8)$$

The Monge-Kantorovich duality formula yields:

$$W_{c_j}(\mu_j, \nu) \geq \int_{X_j} \psi_j^{c_j} d\mu_j + \int_Z \psi_j d\nu \quad (3.9)$$

summing these inequalities and using (3.8) we then get:

$$\sum_{j=0}^d W_{c_j}(\mu_j, \nu) \geq \sum_{j=0}^d \int_{X_j} \psi_j^{c_j} d\mu_j. \quad (3.10)$$

With (3.7), we deduce that the transfers φ_j 's solve the following (concave) program:

$$(\mathcal{P}) \sup \left\{ \sum_{j=0}^d \int_{X_j} \varphi_j^{c_j} d\mu_j : \sum_{j=0}^d \varphi_j = 0 \right\}.$$

Take now some $\eta \in \mathcal{M}_1^+(Z)$. With the Monge-Kantorovich duality formula, the balance condition (3.6) and (3.7), we get

$$\begin{aligned} \sum_{j=0}^d W_{c_j}(\mu_j, \eta) &\geq \sum_{j=0}^d \left(\int_{X_j} \varphi_j^{c_j} d\mu_j + \int_Z \varphi_j d\eta \right) \\ &= \sum_{j=0}^d \int_{X_j} \varphi_j^{c_j} d\mu_j = \sum_{j=0}^d W_{c_j}(\mu_j, \nu) \end{aligned}$$

So that ν solves

$$(\mathcal{P}^*) \inf \left\{ \sum_{j=0}^d W_{c_j}(\mu_j, \nu) : \nu \in \mathcal{M}_1^+(Z) \right\}.$$

At this point, we haven't proven anything about the existence of equilibria, but have discovered that if $(\varphi_j, \gamma_j, \nu)$ is a matching equilibrium then: the transfers φ_j 's solve (\mathcal{P}) , the quality line ν solves (\mathcal{P}^*) , and for each j , γ_j solves $W_{c_j}(\mu_j, \nu)$.

It turns out that in fact, we have much more:

Theorem 3.1 *The supremum in (\mathcal{P}) and the supremum in (\mathcal{P}^*) are attained and the two values are equal. Moreover $(\varphi_j, \gamma_j, \nu)$ is a matching equilibrium if and only if:*

- *the transfers φ_j 's solve (\mathcal{P}) ,*
- *the quality line ν solves (\mathcal{P}^*) ,*
- *for each j , γ_j solves $W_{c_j}(\mu_j, \nu)$.*

In particular matching equilibria exist.

For a complete proof of the previous result, we refer to [27]. It is a quite remarkable fact that equilibrium prices and equilibrium quality lines can be determined by solving independently the optimization problems (\mathcal{P}) and (\mathcal{P}^*) respectively. Also remark that (\mathcal{P}) is a concave maximization program and (\mathcal{P}^*) is a convex minimization program. We will see in the sequel two other situations (discrete choice models and Wardrop equilibria in congested transport) where equilibria are characterized by a convex minimization problem (where optimal transport in some broad sense plays an important role).

Remark 3. Interestingly, there is an alternative linear programming formulation due to Chiappori, McCann and Nesheim [33] based on the opposite of the *surplus* function

$$S(x_0, \dots, x_N) := \inf_{z \in Z} \sum_{j=0}^N c_j(x_j, z)$$

as well as the *multi marginals* optimal transport problem which consists in minimizing

$$\int_{X_0 \times \dots \times X_N} S(x_0, \dots, x_N) d\gamma(x_0, \dots, x_N)$$

among probability measures γ on $X_0 \times \dots \times X_N$ having μ_0, \dots, μ_N as marginals. Exercise: guess what the dual form of the multi-marginals Monge-Kantorovich problem looks like, then prove it rigorously.

3.2 Equilibria in discrete choice models

Discrete choice models have been very popular among econometricians since they lead to tractable and testable models with possibly qualitative data. The celebrated Logit and Probit models rely on discrete choice models. The article of McFadden [59] gives a very comprehensive exposition of the theory, and we also refer the reader to Mc Fadden [60] for applications to residential choice. It is fair, especially in view of the general subject of these notes, to quote here the contributions of Daganzo (see in particular [35]), strongly motivated by route choice in urban traffic. These models also have an appealing flavour of statistical mechanics : arguing that the way an agent ranks a finite number of alternatives inherently involves some randomness one is naturally led to replace brute force exact optimization by a probability over all alternatives that is a clear analogue of Gibbs measures.

The purpose of this paragraph is to show that discrete choice models are a kind of stochastic perturbation of the Monge-Kantorovich optimal transport problem in its dual form and to give a variational argument for the existence and equilibria of equilibrium prices in such models. I am grateful to Roberto Cominetti for showing me this variational characterization.

There is a finite number of agents types denoted $i = 1, \dots, N$ and a finite number of different indivisible goods or objects denoted $j = 1, \dots, M$ (typical examples being apartments, houses, paintings etc...). Utility of agent i for good j is again quasi-linear but now contains a random component, it is of the form

$$u_{ij} + \varepsilon_{ij} - p_j$$

where $u_{ij} \in \mathbb{R}$ is the deterministic part of utility, ε_{ij} is a random variable and p_j is the price of object j . We shall assume that the random variables ε_{ij} are integrable, i.i.d., centered, with a continuous distribution and have the whole real line as support. We assume that the number of agents of the different types is given by μ_i , $i = 1, \dots, N$ and the different objects are available in quantity ν_j , $j = 1, \dots, M$, we assume also that total demand equals total offer

$$\sum_{i=1}^N \mu_i = \sum_{j=1}^M \nu_j.$$

In this setting, any object has a positive probability to be chosen by any agent so that demand is described by probabilities for the different objects to be chosen. Given prices $p = (p_1, \dots, p_M)$ define for every $i \in \{1, \dots, N\}$

$$U_i(p) := \max_{j=1, \dots, M} (u_{ij} + \varepsilon_{ij} - p_j), \quad V_i(p) := \mathbb{E}(U_i(p))$$

the probability of i choosing j is then

$$\mathbb{P}(U_i(p) = u_{ij} + \varepsilon_{ij} - p_j).$$

so that the price system p is an equilibrium if

$$\nu_j = \sum_{i=1}^N \mu_i \mathbb{P}(U_i(p) = u_{ij} + \varepsilon_{ij} - p_j), \quad \forall j = 1, \dots, M. \quad (3.11)$$

It is easy to see that V_i is a C^1 convex function and that

$$\frac{\partial V_i}{\partial p_j}(p) = -\mathbb{P}(U_i(p) = u_{ij} + \varepsilon_{ij} - p_j)$$

so that (3.11) is nothing but the Euler-Lagrange equation for the convex minimization problem:

$$\inf_p \left\{ \sum_{j=1}^M \nu_j p_j + \sum_{i=1}^N \mu_i V_i(p) \right\}. \quad (3.12)$$

This proves that p is an equilibrium if and only if it solves (3.12). We let the reader check that (3.12) admits minimizers so that there exists equilibria.

Note that when there is no noise i.e. $\varepsilon_{ij} = 0$ the previous problem reads as

$$\inf_p \left\{ \sum_{j=1}^M \nu_j p_j + \sum_{i=1}^N \mu_i \left(\max_{j=1, \dots, M} (u_{ij} - p_j) \right) \right\}$$

which is the dual formulation of the discrete Monge-Kantorovich problem

$$\sup \left\{ \sum_{i,j} u_{ij} \gamma_{ij} : \gamma_{ij} \geq 0, \sum_i \gamma_{ij} = \nu_j, \sum_j \gamma_{ij} = \mu_i \right\}.$$

3.3 Urban economics I : optimal planning

We aim to show now that optimal transport may be a powerful tool to study (toy but instructive) urban planning problems. Imagine the situation where some planner has to design a city from scratch, meaning that he has at disposal some land domain and has to choose the structure of the city in this domain. In a very idealized situation, this structure is simply given by two probability measures : μ the distribution of residents and ν the distribution of services. It is often argued in urban economics that:

- ν should be concentrated (because of production externalities: if there are 50 mathematicians in the room, each of them is presumed to be more productive than if he/she was just by him/herself),
- μ should be spread (because residents like to live at large, for instance),
- there is a force that balances the two previous ones : residents have to commute from their home to the services (to shop, to work or both) : this is of course where transport will naturally come into play.

The city shape is given by Ω , a bounded open connected subset of \mathbb{R}^2 and the structure of the city will be determined by two measures on $\overline{\Omega}$. A reasonable way for the planner to take into account these three effects is to minimize with respect to μ and ν (of same total mass normalized to 1 as usual) a criterion of the form:

$$T(\mu, \nu) + G(\mu) + V(\nu)$$

where the coupling term T measures in some sense how close from each other μ and ν are, G is a term that favours dispersion or uniformity and V favours concentration. Following Santambrogio [71], let us make the following choices for the three terms T , G and V (see [30] or [20] for other specifications):

- T is given by the value of some optimal transport problem, for simplicity let's take the quadratic cost which gives

$$T(\mu, \nu) = \frac{1}{2} W_2^2(\mu, \nu) := \frac{1}{2} \inf_{\gamma \in \Pi(\mu, \nu)} \int |x - y|^2 d\gamma(x, y)$$

- G is given by a convex integral functional forcing absolute continuity, for simplicity we take a quadratic functional:

$$G(\mu) = \begin{cases} \frac{1}{2} \int_{\Omega} u^2 & \text{if } \mu = u \cdot \mathcal{L}^2, u \in L^2(\Omega), \\ +\infty & \text{otherwise,} \end{cases}$$

- V is an interaction term

$$V(\nu) := \int_{\bar{\Omega} \times \bar{\Omega}} F(|x - y|^2) d\nu(x) d\nu(y)$$

where F is a smooth function such that $F' > 0$ on \mathbb{R}_+ .

The planner's problem then is

$$\inf_{(\mu, \nu)} \{W_2^2(\mu, \nu) + G(\mu) + V(\nu)\}.$$

It is easy to see that this problem admits solutions but since the problem is nonconvex (due to the interaction term) uniqueness is not guaranteed. Minimizing with respect to μ for fixed ν (this is a convex problem), the optimal μ is characterized by $\mu = u \cdot \mathcal{L}^2$ with

$$u(x) = \left(\varphi(x) - \frac{1}{2}|x|^2 + c_1 \right)_+$$

where $\nabla\varphi$ is the Brenier's map between μ and ν and c is some constant (that is a multiplier associated to the constraint that μ has mass 1). Now the first-order optimality condition with respect to ν reads as: there exists a constant c_2 such that

$$\left(\varphi^*(y) - \frac{1}{2}|y|^2 \right) \geq 2 \int F(|x - y|^2) d\nu(x) + c_2$$

and this equality is in fact an equality ν -a.e., these optimality conditions formally have to be complemented with the Monge-Ampère equation:

$$\det(D^2\varphi)\nu(\nabla\varphi) = u.$$

In general, these conditions cannot be solved explicitly, but they enabled Santambrogio to prove some regularity of minimizers : $u \in C(\Omega)$ (it is even Lipschitz) and $\nu \in L^\infty$. Moreover when $F(t) = t$ (i.e. V is the variance), Santambrogio proved that the optimal structure consists of measures with the same center of mass, having as densities two truncated quadratic functions (and ν is more concentrated than μ in the sense that it is the homothetic image of μ with an homothety factor < 1).

Let us finally mention that other problems in urban planning can also be considered in the framework of optimal mass transport, like the design of a transport network [18] or optimal pricing policies (see [19] for tariffication of public transport or [17] for location dependendent tariffication of a consumption good).

3.4 Urban economics II : equilibria

Another important issue in urban economics is whether one can deduce the structure of the city as an equilibrium solution from the competition between firms and individuals for land use. As before, the structure of the city captures the way land is shared between those uses in terms of two densities μ (number of residents per land unit) and ν (number of jobs per land unit).

Classical references for competitive equilibrium models where the structure of the city results from the rational behaviour of firms and residents can be found in Fujita and Ogawa [48], [49], Fujita [46] and Lucas and Rossi-Hansberg [58]. Most of the literature on the topic is restricted to the unidimensional (or radially symmetric) case, in [28] Carlier and Ekeland attacked the two-dimensional case without assuming radial symmetry thanks to optimal transport arguments. The model considered in [28] is directly inspired by that of Lucas and Rossi-Hansberg [58], the main departure from the Lucas-Rossi-Hansberg model is in the form of the transport cost assumed to be monetary as in Berliant et al. [13]. More generally, we refer the reader interested in urban economics to the textbooks [46], [50] and [47].

As in the previous paragraph, the city is given by some, bounded, open and connected subset of \mathbb{R}^2 . There are three kinds of actors: agents, firms and landowners. A single good is consumed and produced in Ω .

Agents are assumed to be identical and to have a utility function $(c, S) \mapsto U(c, S)$, where c denotes consumption of the good and S denotes land consumption i.e. surface occupied. Firms are also identical, with production function $(z, n) \mapsto f(z, n)$ where z is a productivity parameter (which may vary from one location to another due to production externalities), and n is level of employment. Landowners play no role in consumption or production (absentee landlords) but they extract all the surplus and rent the land to highest bidder.

An important ingredient is production externalities. Given employment density $\nu(y)dy$ in the city, the productivity function is:

$$z(x) = Z_\nu(x) := \chi \left(\int_\Omega \rho(x, y) \nu(y) dy \right) \text{ for all } x \in \Omega \quad (3.13)$$

With ρ a continuous positive kernel and χ a continuous increasing function such that $\chi(\mathbb{R}_+) \subset [\underline{z}, \bar{z}] \subset (0, +\infty)$. There are finally monetary commuting¹ costs given by $(x, y) \mapsto c(x, y)$. Finally, we consider an open city

¹In Lucas and Rossi-Hansberg [58], transport costs are taken into account in terms of time lost in commuting: an employee that leaves home has one unit of working time to

model in the sense that population size is not fixed (but the utility of agents is).

Before defining equilibria, we have to spend some time describing the actors' rational behavior. There are two aspects in this behavior: a local one and a nonlocal one (free mobility of labor).

Agents At equilibrium all agents have the same utility \bar{u} . If available revenue at $x \in \Omega$ is $\varphi = \varphi(x)$, and denoting Q the rent, one gets:

$$\varphi = V(Q) := \min \{c + QS : U(c, S) \geq \bar{u}\} \quad (3.14)$$

Using $Q = V^{-1}(\varphi)$ one gets $c(\varphi)$ and $S(\varphi)$.

The number of residents per unit of surface used for residential use then is

$$N(\varphi) = \frac{1}{S(\varphi)}$$

note that $Q(\varphi)$ is the rent for residential use.

Firms If, at $y \in \Omega$, productivity is z and wage is ψ the firm solves

$$q(z, \psi) := \max_{n \geq 0} f(z, n) - \psi \cdot n \quad (3.15)$$

$q(z, \psi)$ is then the rent for business use. The employment level then is $n(z, \psi)$: the solution of (3.15).

Landowners At $x \in \Omega$, if productivity is z , wage is ψ and residents' revenue is φ there are two rents: $q(z, \psi)$ (business) and $Q(\varphi)$ (residence). Landowners determine the fraction of surface devoted to business use i.e. a fraction $\theta \in [0, 1]$. The landowners being rational, they allocate land to the highest bidder i.e.

$$q(z(x), \psi(x)) > Q(\varphi(x)) \Rightarrow \theta(x) = 1, \quad (3.16)$$

$$q(z(x), \psi(x)) < Q(\varphi(x)) \Rightarrow \theta(x) = 0, \quad (3.17)$$

From agents, firms and landowners behaviors, we obtain the residents and employment densities:

$$\mu = (1 - \theta)N(\varphi) \text{ and } \nu = \theta n(z, \psi). \quad (3.18)$$

These two formulas capture pointwise rationality, another crucial (nonlocal) aspect that has to be taken into account in the definition of equilibria is in job/residential location choice, namely:

offer, some fraction of this time is lost in commuting so that the employee is not paid for one unit but for unit net of the commuting time. This leads to the so-called *iceberg*-like costs.

Free mobility of labor

An agent living at x chooses a job location that maximizes her wage net of transport cost which gives a first conjugacy relation:

$$\varphi(x) = \sup_{y \in \Omega} \{\psi(y) - c(x, y)\}, \quad \forall x \in \Omega \quad (3.19)$$

firms will also look for the cheapest employees which symmetrically yields

$$\psi(y) = \inf_{x \in \Omega} \{\varphi(x) + c(x, y)\}, \quad \forall y \in \Omega. \quad (3.20)$$

All this induces a residence location/job location transport plan $\gamma \in \Pi(\mu, \nu)$ ($\gamma(A \times B)$ represents the number of agents living in A and working in B) that is compatible with free-mobility of labor i.e.

$$\psi(y) - \varphi(x) = c(x, y) \quad \gamma\text{-a.e.} \quad (3.21)$$

Of course, free mobility of labor translates directly in the language of optimal transport by the requirement that γ solves the Monge-Kantorovich problem

$$\inf_{\gamma \in \Pi(\mu, \nu)} \int c(x, y) d\gamma(x, y), \quad (3.22)$$

and (φ, ψ) solves its dual in the form

$$\sup \left\{ \int \psi d\nu - \int \varphi d\mu : \psi(y) - \varphi(x) \leq c(x, y) \right\}. \quad (3.23)$$

An equilibrium then is a situation where there are as many jobs as agents that is consistent with pointwise rational behavior and free mobility of labor, which gives the formal definition:

Definition 3.2 *An equilibrium consists of $(\mu, \nu) \in (L^1(\Omega, \mathbb{R}_+))^2$, $z \in C(\Omega, \mathbb{R})$, $(\psi, \varphi) \in (C(\Omega, \mathbb{R}_{++}))^2$, $\theta \in L^\infty(\Omega, [0, 1])$, and γ a nonnegative measure on $\Omega \times \Omega$ such that*

1. $\int_\Omega \mu = \int_\Omega \nu > 0$,
2. $z = Z_\nu$,
3. for $\mathcal{L}^2 \otimes \mathcal{L}^2$ a.e. $(x, y) \in \Omega^2$:

$$\mu(x) = (1 - \theta(x))N(\varphi(x)), \quad \nu(y) = \theta(y)n(z(y), \psi(y)),$$

4. (3.16) and (3.17) hold,

5. (ψ, φ) are conjugate (in the sense of (3.19) and (3.20)),
 6. $\gamma \in \Pi(\mu, \nu)$ and:

$$\psi(y) - \varphi(x) = c(x, y) \quad \gamma\text{-a.e.}$$

Under some natural assumptions that we do not reproduce here, the existence of equilibria is proved in [28]. Without entering the details, let us very informally explain the strategy proof. First, we regularize θ (landowners's choice) by replacing it by $\theta_\varepsilon(z, \psi, \varphi)$ the solution of

$$\max_{\theta \in [0,1]} \{ \theta q(z, \psi) + (1 - \theta)Q(\varphi) - \frac{\varepsilon}{2}\theta^2 \}$$

where $\varepsilon > 0$ is a small regularization parameter. Then, we start with densities μ and ν of same total positive mass, we then compute $z = Z_\nu$ by formula (3.13), and (ψ, φ) as a solution of (3.23). We have one more degree of freedom since (ψ, φ) are defined up to a constant λ , this degree of freedom is used to solve the equal mass condition:

$$\int \theta_\varepsilon(z, \psi, \varphi)n(z, \psi + \lambda) = \int (1 - \theta_\varepsilon(z, \psi + \lambda, \varphi + \lambda))N(\varphi + \lambda)$$

we then obtain two densities

$$\mu' = (1 - \theta_\varepsilon(z, \psi + \lambda, \varphi + \lambda))N(\varphi + \lambda)$$

and

$$\nu' = \theta_\varepsilon(z, \psi, \varphi)n(z, \psi + \lambda)$$

and using Schauder's fixed-point theorem, we find a fixed-point of the map $(\mu, \nu) \mapsto (\mu', \nu')$. It remains to let $\varepsilon \rightarrow 0$ and prove that we recover an equilibrium this way. Again we refer to [28] for precise statements and proofs. A crucial step in the construction is to deduce wages and revenues from densities: this is possible thanks to the dual formulation of the optimal transport (3.22).

3.5 Multidimensional screening

A central paradigm in modern microeconomic theory is the principal-agent model with adverse selection which we shall exemplify by the optimal design of a nonlinear tariff by a monopolist (the principal) facing an heterogeneous population of customers (the agents) the preferences of whom he is not able

to observe (this is a typical case of adverse selection : i.e. asymmetric information with hidden *types*). The agents' type space is denoted X and the type distribution (known to the principal) is given by a probability μ , the product space is denoted Y and utility is quasi-linear once again i.e. of the form

$$u(x, y) - p$$

where $x \in X$ is the agent's type, $y \in Y$ the good attributes (that may be a quality, a vector of attributes or both) and p is the price. The principal has to design a contract that is a map $x \in X \mapsto (y(x), p(x))$ that associates to each type the product and price designed for her. Of course, for this specification to be consistent $(y(x), p(x))$ has to be a utility-maximizing contract for type x i.e.

$$u(x, y(x)) - p(x) := \max_{x' \in X} \{u(x, y(x')) - p(x')\}, \quad \forall x \in X \quad (3.24)$$

which is the so-called *incentive-compatibility* constraint. There should be some freedom left to consumers not to accept the contract and then get the *outside option* (get nothing and pay nothing say), let us denote by y_0 this outside option (and take as new product space $Y \cup \{y_0\}$ if necessary) and assume that $u(x, 0) = 0$ for every type x . Type x thus accepts the contract offered by the principal if

$$u(x, y(x)) - p(x) \geq 0, \quad \forall x \in X, \quad (3.25)$$

that is the *participation constraint*. There is a cost for producing y which is denoted $C(y)$ so that the monopolist's net average cost is

$$\int_X [C(y(x)) - p(x)] d\mu(x)$$

that he seeks to minimize subject to (3.24) and (3.25). In this form, this seems a quite intricate problem (because of the *nonlocal* constraint (3.25)). Let us transform it by changing variables (y, p) into (v, p) with

$$v(x) := u(x, y(x)) - p(x) := \max_{x' \in X} \{u(x, y(x')) - p(x')\}$$

and let's see what happens to the nasty constraint (3.24). Firstly, (3.24) implies that v is restricted to be u -convex that is v is a supremum of functions of the form $x \mapsto u(x, y) - p$ (up to some minus signs this is the notion of c -concavity we encountered so many times), secondly $y(x)$ is related to v by the condition that it belongs to the u -subgradient of v at x , i.e. the set

$$\partial^u v(x) := \{y \in Y : v(x') - v(x) \geq u(x', y) - u(x, y), \quad \forall x' \in X\}.$$

Conversely, it is easy to check that whenever v is u -convex and $y(x) \in \partial^u v(x)$ for all $x \in X$ then the contract $x \mapsto (y(x), p(x))$ satisfies (3.24) with $p(x) := u(x, y(x)) - v(x)$. So that the monopolist's problem can be reformulated as:

$$\inf \int_X [C(y(x)) - u(x, y(x)) + v(x)] d\mu(x) \quad (3.26)$$

$$\text{subject to : } v \geq 0, v \text{ is } u\text{-convex, and } y(x) \in \partial^u v(x), \forall x \in X. \quad (3.27)$$

This reformulation as an unusual calculus of variations problem looks slightly more elegant but it is still not clear how to solve it. There is one advantage though, in this reformulation, in terms of existence. Recall indeed that under some continuity/compactness assumptions, u -convex functions are equicontinuous, this enables one to prove quite general existence results (this is basically what is done in [26]) but...not much more. What about uniqueness issues, qualitative properties, tractable optimality conditions and numerical computations? Almost nothing is known except in some very specific cases. First of all the one-dimensional case is very well-understood at least when u satisfies Spence-Mirrlees condition ([72], [63]) for instance a fairly complete analysis of the solution can be found in Mussa and Rosen [62]. In the multidimensional case, the picture is far from being complete. Even in the case of the Rochet and Choné model [68] where X is some convex body in \mathbb{R}^d , $Y = \mathbb{R}_+^d$, $u(x, y) := x \cdot y$ (so that u -convexity essentially reduces to usual convexity and the u -subgradient is the usual subgradient from convex analysis) and a quadratic cost function, the analysis is very delicate. A very interesting variant due to Rochet [66] leads to the additional constraint that v solves a multi-time Hamilton-Jacobi equations (see Lions and Rochet [57] for Hopf-like representation formulas). This introduces nonconvexities that render the mathematical and numerical analysis of the corresponding multidimensional screening extremely challenging.

Recently, by arguments coming from optimal transport theory, Figalli, Kim and McCann in [44] identified structural conditions on u (too technical to reproduce here) under which the multidimensional screening problem (3.26)-(3.27) is a convex program. A remarkable and somehow intriguing fact is that these conditions are intimately linked to the so-called Ma-Trudinger-Wang condition (see [61]) which plays a crucial role in the regularity theory for optimal transport with a general cost function. More importantly, the results of [44] open the way for a totally new research program on qualitative properties (regularity in particular), characterization and numerical computational strategies for multidimensional screening. All these perspectives were totally out of reach before Figalli, Kim and McCann made this major breakthrough and we are convinced that [44] will be the starting point of an

exciting and totally new field of investigation combining deep mathematics and extremely important economic problems.

Chapter 4

Congested Transport

In the classical Monge-Kantorovich problem, the transportation cost only depends on the amount of mass sent from sources to destinations and not on the paths followed by each particle forming this mass. Thus, it does not allow for congestion effects. By congestion effect, we mean that the travelling cost (or time) of a path depends on "how crowded" this path is. Starting from a simple network model, we shall define equilibria in the presence of congestion. We will then extend this theory to the continuous setting.

4.1 Wardrop equilibria in a simple congested network model

The main data of the model are a finite oriented connected graph $G = (N, E)$ modelling the network, and edge travel times functions $w \in \mathbb{R}_+ \mapsto g_e(w)$ which for each edge $e \in E$ gives the travel time on arc e when the flow on this edge is w . The functions g_e are all nonnegative, continuous, nondecreasing and capture the congestion effect (which may differ on the different edges capturing for instance the idea that some roads may be longer or wider and may have different responses to congestion). The last ingredient of the problem is a transport plan on pairs of nodes $(x, y) \in N^2$ interpreted as pairs of sources/destinations, we denote by $(\gamma_{x,y})_{(x,y) \in N^2}$ this transport plan i.e. $\gamma_{x,y}$ represents the "mass" to be sent from x to y . We denote by $C_{x,y}$ the set of simple paths connecting x to y , so that $C := \cup_{(x,y) \in N^2} C_{x,y}$ is the set of all simple paths, a generic path will be denoted by σ and we will use the notation $e \in \sigma$ to indicate that the path σ uses the edge e .

The unknown of the problem is the flow configuration. The edge flows are denoted by $w = (w_e)_{e \in E}$ and the path flows are denoted by $q = (q_\sigma)_{\sigma \in C}$, this means that w_e is the total flow on edge e and q_σ is the mass traveling

on the path σ . Of course the w_e 's and q_σ 's are nonnegative and constrained by the mass conservation conditions:

$$\gamma_{x,y} = \sum_{\sigma \in C_{x,y}} q_\sigma, \quad \forall (x,y) \in N^2 \quad (4.1)$$

and

$$w_e = \sum_{\sigma \in C : e \in \sigma} q_\sigma, \quad \forall e \in E. \quad (4.2)$$

Given edge flows $w = (w_e)_{e \in E}$, the total travel-time of the path $\sigma \in C$ is

$$T_w(\sigma) = \sum_{e \in \sigma} g_e(w_e). \quad (4.3)$$

In [76], Wardrop defined a notion of noncooperative equilibrium that has been very popular since among engineers working in the field of congested transport and that may be described as follows. Roughly speaking, a Wardrop equilibrium is a flow configuration such that every actually used path should be a shortest path taking into account the congestion effect i.e. formula (4.3). This leads to

Definition 4.1 *A Wardrop equilibrium is a flow configuration $w = (w_e)_{e \in E}$, $q = (q_\sigma)_{\sigma \in C}$ (all nonnegative of course), satisfying the mass conservation constraints (4.1) and (4.2) such that in addition, for every $(x,y) \in N^2$ and every $\sigma \in C_{x,y}$, if $q_\sigma > 0$ then*

$$T_w(\sigma) = \min_{\sigma' \in C_{x,y}} T_w(\sigma').$$

A few years after Wardrop introduced his equilibrium concept, Beckmann, McGuire and Winsten [8] realized that Wardrop equilibria can be characterized by the following variational principle:

Theorem 4.1 *The flow configuration $w = (w_e)_{e \in E}$, $q = (q_\sigma)_{\sigma \in C}$ is a Wardrop equilibrium if and only if it solves the convex minimization problem*

$$\inf_{(w,q)} \sum_{e \in E} H_e(w_e) \text{ s.t. nonnegativity and (4.1) (4.2)} \quad (4.4)$$

where for each e , and $w \in \mathbb{R}_+$,

$$H_e(w) := \int_0^w g_e(s) ds.$$

Proof:

Note that due to (4.2), one can deduce w from q so that (4.4) is an optimization problem on $q = (q_\sigma)_{\sigma \in C}$ only. Assume that $q = (q_\sigma)_{\sigma \in C}$ (with associated edge flows $(w_e)_{e \in E}$) is optimal for (4.4) then for every admissible $\eta = (\eta_\sigma)_{\sigma \in C}$ with associated (through (4.2)) edge-flows $(u_e)_{e \in E}$, one has

$$\begin{aligned} 0 &\leq \sum_{e \in E} H'_e(w_e)(u_e - w_e) = \sum_{e \in E} g_e(w_e) \sum_{\sigma \in C : e \in \sigma} (\eta_\sigma - q_\sigma) \\ &= \sum_{\sigma \in C} (\eta_\sigma - q_\sigma) \sum_{e \in \sigma} g_e(w_e) \end{aligned}$$

so that

$$\sum_{\sigma \in C} q_\sigma T_w(\sigma) \leq \sum_{\sigma \in C} \eta_\sigma T_w(\sigma)$$

minimizing the right-hand side thus yields

$$\sum_{(x,y) \in N^2} \sum_{\sigma \in C_{x,y}} q_\sigma T_w(\sigma) = \sum_{(x,y) \in N^2} \gamma_{x,y} \min_{\sigma' \in C_{x,y}} T_w(\sigma')$$

which exactly says that (q, w) is a Wardrop equilibrium. To prove the converse, it is enough to see that problem (4.4) is convex so that the inequality above is indeed sufficient for a global minimum.

□

The previous characterization actually is the reason why Wardrop equilibria became so popular. Not only, one deduces for free existence results but also uniqueness for w (not for q) as soon as the functions g_e are increasing (so that H_e is strictly convex). The variational formulation (4.4) also admits a dual formulation. Another major advantage of (4.4) is that the techniques of numerical convex optimization can be used to compute Wardrop equilibria, however there are as many variables as the number of paths which obviously restricts computations to small networks, the dual formulation has much less variables but involves nonsmooth terms. Let us also mention an interesting extension of the model to a stochastic setting by Baillon and Cominetti [6].

Remark 4. In the problem above, the transport plan γ is fixed, this may be interpreted as a *short term problem*. Instead, we could consider the *long term* problem where only the distribution of sources μ_0 and the distribution of destinations μ_1 are fixed. In this case, one requires in addition, in the definition of an equilibrium that γ is efficient in the sense that it minimizes among transport plans between μ_0 and μ_1 the total cost

$$\sum \gamma_{x,y} d_w(x, y) \text{ with } d_w(x, y) := \min_{\sigma \in C_{x,y}} T_w(\sigma).$$

In the long term problem where one is allowed to change the assignment as well, equilibria still are characterized by a convex minimization problem where one also optimizes over γ .

4.2 Optimal transport with congestion and equilibria in a continuous framework

This paragraph aims to generalize the previous results to a continuous setting. This means that there will be no network, all paths in a certain given region will therefore be admissible. The first idea is to formulate the whole path dependent transport pattern in terms of a probability measure Q on the set of paths (this is the continuous analogue of the path flows $(q_\sigma)_\sigma$ of the previous paragraph). The second one is to measure the intensity traffic generated by Q through an analogue of the transport density we saw in paragraph 2.4 (except that the paths considered will be more general than straight lines). The last and main idea will be in modelling the congestion effect through a metric that is monotone increasing in the traffic intensity.

We will deliberately avoid to enter into technicalities so the following description will be pretty informal (see [29] for details). From now on, Ω denotes an open bounded connected subset of \mathbb{R}^2 (a city say), and we are also given :

- either probability measures μ_0 and μ_1 (distribution of sources and destinations) on $\bar{\Omega}$ in the case of the long-term problem,
- or a transport plan γ (joint distribution of sources and destinations) that is a joint probability on $\bar{\Omega} \times \bar{\Omega}$ in the case of the short-term problem.

Given an absolutely curve $\sigma ; [0, 1] \mapsto \bar{\Omega}$ and a continuous function φ , let us set

$$L_\varphi(\sigma) := \int_0^1 \varphi(\sigma(t)) |\dot{\sigma}(t)| dt. \quad (4.5)$$

A transport pattern is by definition a probability measure Q on $C := C([0, 1], \bar{\Omega})$ concentrated on absolutely continuous curves that is compatible with mass conservation, i.e. such that either

$$e_{0\#}Q = \mu_0, \quad e_{1\#}Q = \mu_1$$

in the case of the long-term problem, or

$$(e_0, e_1)\#Q = \gamma, \quad \text{with } e_t(\sigma) := \sigma(t), \quad \forall t \in [0, 1]$$

in the case of the short-term problem. We shall denote by $\mathcal{Q}(\mu_0, \mu_1)$ and $\mathcal{Q}(\gamma)$ the set of admissible transport patterns respectively for the long-term and for the short-term problem:

$$\mathcal{Q}(\mu_0, \mu_1) := \{Q : e_{0\#}Q = \mu_0, e_{1\#}Q = \mu_1\}$$

and

$$\mathcal{Q}(\gamma) := \{Q : (e_0, e_1)\#Q = \gamma\}.$$

In the remainder of this paragraph, we will focus on the long-term problem. We are interested in finding an equilibrium i.e. a $Q \in \mathcal{Q}(\mu_0, \mu_1)$ that is supported by geodesics for a metric ξ_Q depending on Q itself (congestion).

The intensity of traffic associated to $Q \in \mathcal{Q}(\mu_0, \mu_1)$ is by definition the measure $i_Q \in \mathcal{M}(\bar{\Omega})$, defined by

$$\int \varphi di_Q := \int_{C([0,1], \bar{\Omega})} \left(\int_0^1 \varphi(\gamma(t)) |\dot{\gamma}(t)| dt \right) dQ(\gamma) = \int_C L_\varphi(\sigma) dQ(\sigma).$$

for all $\varphi \in C(\bar{\Omega}, \mathbb{R}_+)$. This definition is a generalization of the notion of transport density and the interpretation is the following: for a subregion A , $i_Q(A)$ represents the total cumulated traffic in A induced by Q , it is indeed the average over all paths of the length of this path intersected with A .

The congestion effect is then captured by the *metric* associated to Q :

$$\xi_Q(x) := g(x, i_Q(x)), \text{ for } i_Q \ll \mathcal{L}^2 \text{ (+}\infty \text{ otherwise).}$$

for a given increasing function $g(x, \cdot) : \mathbb{R}_+ \rightarrow \mathbb{R}_+$. The fact that there exists at least one $Q \in \mathcal{Q}(\mu_0, \mu_1)$ such that $i_Q \ll \mathcal{L}^2$ is not always true and depends on μ_0 and μ_1 but again we do not wish to enter the details, let us only indicate that this condition is satisfied when μ_0 and μ_1 are "well behaved". Let us now describe, what a reasonable definition of an equilibrium should look like. If the overall transport pattern is Q , an agent commuting from x to y choosing a path $\sigma \in C_{x,y}$ (i.e. an absolutely continuous curve σ such that $\sigma(0) = x$ and $\sigma(1) = y$) spends time

$$L_{\xi_Q}(\sigma) = \int_0^1 g(\sigma(t), i_Q(\sigma(t)) |\dot{\sigma}(t)|) dt$$

she will of try to minimize this time i.e. to achieve the corresponding geodesic distance

$$c_{\xi_Q}(x, y) := \inf_{\sigma \in C_{x,y}} L_{\xi_Q}(\sigma)$$

paths in $C_{x,y}$ such that $c_{\xi_Q}(x, y) = L_{\xi_Q}(\sigma)$ are called geodesics (for the metric induced by the congestion effect generated by Q). A first requirement, in the

definition of an equilibrium therefore is that Q -a.e. path σ is a geodesic between its endpoints $\sigma(0)$ and $\sigma(1)$. The transportation pattern may be disintegrated with respect to $\gamma_Q := (e_0, e_1)_\#Q$:

$$Q = \gamma_Q \otimes (p^{x,y})$$

i.e.

$$\int_C \Phi(\sigma) dQ(\sigma) = \int_{\bar{\Omega} \times \bar{\Omega}} \left(\int_{C_{x,y}} \Phi(\sigma) dp^{x,y}(\sigma) \right) d\gamma_Q(x, y), \quad \forall \Phi.$$

In other words, $\gamma_Q(A \times B)$ is the probability that a path has starting point in A and a terminal point in B (so that $\gamma_Q \in \Pi(\mu_0, \mu_1)$ because $Q \in \mathcal{Q}(\mu_0, \mu_1)$) and given starting and terminal points (x, y) , $p^{x,y}$ is a probability on $C_{x,y}$ that represents the probability over paths conditional on (x, y) . The requirement that Q gives full mass to geodesics says that for γ_Q -a.e. (x, y) , $p^{x,y}$ is supported on the set of geodesics between x and y but this does require any particular property on the coupling γ_Q . We thus supplement the definition of an equilibrium by the additional requirement that γ_Q should solve the optimal transportation problem:

$$\inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{\bar{\Omega} \times \bar{\Omega}} c_{\xi_Q}(x, y) d\gamma(x, y). \quad (4.6)$$

This yields:

Definition 4.2 *A Wardrop equilibrium (for the long-term problem) is a $Q \in \mathcal{Q}(\mu_0, \mu_1)$ such that*

$$Q(\{\sigma : L_{\xi_Q}(\sigma) = c_{\xi_Q}(\sigma(0), \sigma(1)) = 1\}) = 1 \quad (4.7)$$

and $\gamma_Q := (e_0, e_1)_\#Q$ solves the optimal transport problem (4.6).

Of course in the short-term case, γ_Q is fixed equal to γ so that Wardrop equilibria are defined by condition (4.7) only.

Let us then consider the (convex) variational problem

$$\inf_{Q \in \mathcal{Q}(\mu_0, \mu_1)} \int_{\Omega} H(x, i_Q(x)) dx \quad (4.8)$$

where $H'(x, \cdot) = g(x, \cdot)$, $H(x, 0) = 0$. We shall refer to (4.8) as the congested optimal mass transportation problem for reasons that will be clarified later. Under some technical assumptions that we do not reproduce here, the main results of [29] can be summarized by

Theorem 4.2 *Problem 4.8 admits at least one minimizer. Moreover $\bar{Q} \in \mathcal{Q}(\mu_0, \mu_1)$ solves (4.8) if and only if it is a Wardrop equilibrium. In particular there exists Wardrop equilibria.*

The full proof is quite involved since it involves to take care of some regularity issues in details. But the intuition of why solutions of (4.8) are Wardrop equilibria can be understood easily from the following formal manipulations. By convexity arguments, it is easily seen that $\bar{Q} = \bar{\gamma} \otimes \bar{p}^{x,y} \in \mathcal{Q}(\mu_0, \mu_1)$ solves (4.8) if and only if it satisfies the variational inequalities

$$\int_{\Omega} \bar{\xi} i_{\bar{Q}} = \inf \left\{ \int_{\Omega} \bar{\xi} i_Q : Q \in \mathcal{Q}(\mu_0, \mu_1) \right\} \text{ with } \bar{\xi} := H'(\cdot, i_{\bar{Q}}(\cdot)) = g(\cdot, i_{\bar{Q}}(\cdot)). \quad (4.9)$$

which we may rewrite as

$$\begin{aligned} \int_{\Omega} \bar{\xi} i_{\bar{Q}} &= \int_C L_{\bar{\xi}}(\sigma) d\bar{Q}(\sigma) \\ &= \int_{\bar{\Omega} \times \bar{\Omega}} \left(\int_{C^{x,y}} L_{\bar{\xi}}(\sigma) d\bar{p}^{x,y}(\sigma) \right) d\bar{\gamma}(x, y) \\ &= \inf_{(\gamma, p)} \int_{\bar{\Omega} \times \bar{\Omega}} \left(\int_{C^{x,y}} L_{\bar{\xi}}(\sigma) dp^{x,y}(\sigma) \right) d\gamma(x, y) \\ &= \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{\bar{\Omega} \times \bar{\Omega}} \left(\inf_{p \in \mathcal{M}_+^1(C^{x,y})} \int_{C^{x,y}} L_{\bar{\xi}}(\sigma) dp(\sigma) \right) d\gamma(x, y) \\ &= \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{\bar{\Omega} \times \bar{\Omega}} \left(\inf_{\sigma \in C^{x,y}} L_{\bar{\xi}}(\sigma) \right) d\gamma(x, y) \end{aligned}$$

Let us then define $c_{\bar{\xi}} = c_{\xi_{\bar{Q}}}$ i.e.

$$c_{\bar{\xi}}(x, y) := \inf_{\sigma \in C^{x,y}} L_{\bar{\xi}}(\sigma),$$

we firstly get

$$\begin{aligned} \int_{\bar{\Omega} \times \bar{\Omega}} c_{\bar{\xi}}(x, y) d\bar{\gamma}(x, y) &\leq \int_C L_{\bar{\xi}} d\bar{Q} \\ &= \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{\bar{\Omega} \times \bar{\Omega}} c_{\bar{\xi}}(x, y) d\gamma(x, y) \end{aligned}$$

so that $\bar{\gamma}$ solves the Monge-Kantorovich problem:

$$\inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{\bar{\Omega} \times \bar{\Omega}} c_{\bar{\xi}}(x, y) d\gamma(x, y).$$

Secondly, we obtain

$$\begin{aligned} \int_C L_{\bar{\xi}}(\sigma) d\bar{Q}(\sigma) &= \int_{\bar{\Omega} \times \bar{\Omega}} c_{\bar{\xi}}(x, y) d\bar{\gamma}(x, y) \\ &= \int_C c_{\bar{\xi}}(\sigma(0), \sigma(1)) d\bar{Q}(\sigma) \end{aligned}$$

and since $L_{\bar{\xi}}(\sigma) \geq c_{\bar{\xi}}(\sigma(0), \sigma(1))$, we get

$$L_{\bar{\xi}}(\sigma) = c_{\bar{\xi}}(\sigma(0), \sigma(1)) \quad \text{for } \bar{Q}\text{-a.e. } \sigma.$$

or, in an equivalent way, for $\bar{\gamma}$ -a.e. (x, y) one has:

$$L_{\bar{\xi}}(\sigma) = c_{\bar{\xi}}(x, y) \quad \text{for } \bar{p}^{x,y}\text{-a.e. } \sigma$$

which exactly proves that \bar{Q} is a Wardrop equilibrium.

Remark 5. It would be very tempting to deduce from the previous result that equilibria are efficient since they are minimizers in (4.8). One has to be cautious with this quick interpretation since the quantity $\int_{\Omega} H(x, i_Q(x)) dx$ does not represent the natural total social cost measured by the total time lost in commuting which reads as

$$\int_{\Omega} g(x, i_Q(x)) i_Q(x) dx. \quad (4.10)$$

The efficient transport patterns are minimizers of (4.10) and thus are different from equilibria in general. This difference between efficiency and equilibrium is very well-documented in the finite-dimensional network setting where it is frequently associated to the literature on the so-called *cost of anarchy*. Interestingly, one can restore efficiency of equilibria by a well-chosen toll system, which, from a public economics perspective, justifies public intervention in the management of road traffic.

Remark 6. For the short-term problem, a similar variational characterization holds, namely that $\bar{Q} \in \mathcal{Q}(\gamma)$ is a (short-term) Wardrop equilibrium if and only if it solves

$$\inf_{Q \in \mathcal{Q}(\gamma)} \int_{\Omega} H(x, i_Q(x)) dx. \quad (4.11)$$

We have proved that, as in the finite-dimensional network case, Wardrop equilibria have a variational characterization which is in principle easier to deal with than the definition. Unfortunately, the convex problems (4.8) and (4.11) may be difficult to solve since they involve measures on sets of curves that is two layers of infinite dimensions! The next two paragraphs are precisely intended to consider different formulations that turn out to be much more tractable:

- for the short-term problem (4.11), we will see that the equilibrium metrics solve a kind of dual problem that can be solved numerically,
- for the long-term problem (4.8), we will deduce optimal Q 's from a minimal flow problem à la Beckmann and a construction à la Moser, in other words, the problem will amount to solve a certain nonlinear elliptic PDE (which turns out to be quite degenerate in realistic congestion models).

4.3 Duality for the short-term problem

The purpose of this paragraph, is to give a dual and tractable formulation of the variational problem for the short-term problem (4.11). For every $x \in \Omega$ and $\xi \geq 0$, let us define

$$H^*(x, \xi) := \sup\{\xi i - H(x, i), i \geq 0\}, \quad \xi_0(x) := g(x, 0).$$

By our assumptions on g , one has $H^*(x, \xi) = 0$ for every $x \in \Omega$ and $\xi \leq \xi_0(x)$; let us recall Young's inequality:

$$H(x, i) + H^*(x, \xi) \geq \xi i, \quad \forall i \geq 0, \forall \xi \geq \xi_0(x) \quad (4.12)$$

and that inequality (4.12) is strict unless $\xi = g(x, i) \geq \xi_0(x)$. In particular, for $Q \in \mathcal{Q}(\gamma)$, we have the identity

$$H(x, i_Q(x)) + H^*(x, \xi_Q(x)) = \xi_Q(x) i_Q(x) \quad (4.13)$$

and

$$H(x, i_Q(x)) + H^*(x, \xi) > \xi i_Q(x), \quad \forall \xi \geq \xi_0(x), \quad \xi \neq \xi_Q(x). \quad (4.14)$$

Let us now define the functional

$$J(\xi) = \int_{\Omega} H^*(x, \xi(x)) dx - \int_{\bar{\Omega} \times \bar{\Omega}} c_{\xi}(x, y) d\gamma(x, y) \quad (4.15)$$

where as usual c_{ξ} is the geodesic distance associated to the metric ξ i.e.

$$c_{\xi}(x, y) := \inf_{\sigma \in C_{x,y}} L_{\xi}(\sigma).$$

and consider:

$$\sup \{-J(\xi) : \xi \geq \xi_0\} \quad (4.16)$$

Theorem 4.3 *The following duality formula holds*

$$\min(4.11) = \max(4.16) \quad (4.17)$$

and ξ solves (4.16) if and only if $\xi = \xi_Q$ for some $Q \in \mathcal{Q}(\gamma)$ solving (4.11).

Proof:

Let $Q \in \mathcal{Q}(\gamma)$ (so that $\xi_Q \geq \xi_0$) and let $\xi \geq \xi_0$, from (4.12) and

$$\int_{\Omega} \xi(x) i_Q(x) dx = \int_C L_{\xi}(\sigma) dQ(\sigma). \quad (4.18)$$

we first get:

$$\begin{aligned} \int_{\Omega} H(x, i_Q(x)) dx &\geq \int_{\Omega} \xi i_Q - \int_{\Omega} H^*(x, \xi(x)) dx \\ &= \int_C L_{\xi}(\sigma) dQ(\sigma) - \int_{\Omega} H^*(x, \xi(x)) dx. \end{aligned}$$

Using the fact that

$$L_{\xi}(\sigma) \geq c_{\xi}(\sigma(0), \sigma(1)) \quad (4.19)$$

and $Q \in \mathcal{Q}(\gamma)$ we then have

$$\int_C L_{\xi}(\sigma) dQ(\sigma) \geq \int_C c_{\xi}(\sigma(0), \sigma(1)) dQ(\sigma) = \int_{\bar{\Omega} \times \bar{\Omega}} c_{\xi}(x, y) d\gamma(x, y).$$

Since $Q \in \mathcal{Q}(\gamma)$ and $\xi \geq \xi_0$ are arbitrary and since we already know that the infimum of (4.11) is attained we thus deduce

$$\min(4.11) \geq \sup(4.16). \quad (4.20)$$

Now let $Q \in \mathcal{Q}(\gamma)$ solve (4.11) and let $\xi := \xi_Q$ (recall that ξ_Q does not depend on the choice of the minimizer Q), from the equivalence between Wardrop equilibria and solutions of (4.11), we know that

$$L_{\xi}(\sigma) = c_{\xi}(\sigma(0), \sigma(1)) \text{ for } Q\text{-a.e. } \sigma \in C$$

with (4.18), integrating the previous identity and using $Q \in \mathcal{Q}(\gamma)$ we then get:

$$\int_{\Omega} \xi i_Q = \int_C L_{\xi}(\sigma) dQ(\sigma) = \int_{\bar{\Omega} \times \bar{\Omega}} c_{\xi}(x, y) d\gamma(x, y).$$

Using (4.13), (4.20) and the fact that $Q \in \mathcal{Q}(\gamma)$ solves (4.11) yields:

$$\begin{aligned} \sup(4.16) \leq \min(4.11) &= \int_{\Omega} H(x, i_Q(x)) dx = \int_{\Omega} \xi i_Q - \int_{\Omega} H^*(x, \xi(x)) dx \\ &= \int_{\bar{\Omega} \times \bar{\Omega}} c_{\xi}(x, y) d\gamma(x, y) - \int_{\Omega} H^*(x, \xi(x)) dx \end{aligned}$$

so that ξ solves (4.16) and (4.17) is satisfied. Finally if ξ solves (4.16) and $Q \in \mathcal{Q}(\gamma)$ solves (4.11), then with (4.18) and (4.19), one has

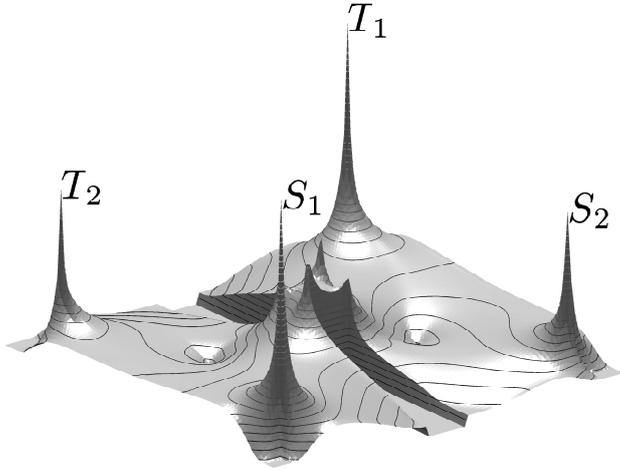
$$\begin{aligned} \int_{\Omega} \xi i_Q - \int_{\Omega} H^*(x, \xi(x)) dx &\geq \int_{\bar{\Omega} \times \bar{\Omega}} c_{\xi}(x, y) d\gamma(x, y) - \int_{\Omega} H^*(x, \xi(x)) dx \\ &= \max(4.16) = \min(4.11) = \int_{\Omega} H(x, i_Q(x)) dx \end{aligned}$$

and thus we deduce from (4.12) and (4.14) that $\xi = \xi_Q$.

□

Remark 7. Under reasonable continuity and strict monotonicity assumptions on the congestion function g , the dual problem (4.16) has a unique solution so that the equilibrium metric ξ_Q and the equilibrium intensity of traffic i_Q are unique although Wardrop equilibria Q might not be unique.

In [11], [12], we designed a consistent numerical scheme to approximate the equilibrium metric ξ_Q by a descent method on the dual which can be done in an efficient way by the *Fast Marching Algorithm*. One can recover the corresponding equilibrium intensity i_Q by inverting the relation $\xi(x) = g(x, i_Q(x))$. An example is given in the following figure:



In a symmetric configuration of two sources S_1 and S_2 , and two targets T_1 and T_2 ; we consider a river where there is no traffic and a bridge linking the two sides of the river (see figure 4.3 (a)). We chose the traffic weights such that $\gamma_{1,1} + \gamma_{1,2} = 2(\gamma_{2,1} + \gamma_{2,2})$ and $\frac{\gamma_{2,2}}{\gamma_{2,1}} = \frac{\gamma_{1,1}}{\gamma_{1,2}} = 2$. The traffic intensity going out from S_1 is twice S_2 's. One can note the two hollows on each side of the river appearing because of the inter-sides and intra-sides crossed traffics.

4.4 Beckmann-like reformulation of the long-term problem

In the long-term problem (4.8), we have one more degree of freedom since the transport plan is not fixed. This will enable us to reformulate the problem as a variational divergence constrained problem à la Beckmann and ultimately to reduce the equilibrium problem to solving some nonlinear PDE. For $Q \in \mathcal{Q}(\mu_0, \mu_1)$, let us define the vector-field σ_Q by, $\forall X \in C(\bar{\Omega}, \mathbb{R}^d)$:

$$\int_{\bar{\Omega}} X(x) \sigma_Q(x) dx := \int_{C([0,1], \bar{\Omega})} \left(\int_0^1 X(\gamma(t)) \cdot \dot{\gamma}(t) dt \right) dQ(\gamma)$$

which is a kind of vectorial traffic intensity. Taking a gradient field $X = \nabla u$ in the previous definition yields

$$\int_{\bar{\Omega}} \nabla u \sigma_Q = \int_{C([0,1], \bar{\Omega})} [u(\sigma(1)) - u(\sigma(0))] dQ(\gamma) = \int_{\Omega} u(\mu_1 - \mu_0)$$

which means that

$$\operatorname{div}(\sigma_Q) = \mu_0 - \mu_1,$$

moreover it is easy to check that

$$|\sigma_Q| \leq i_Q.$$

Since H is increasing, it proves that the value of the scalar problem (4.8) is larger than that of the minimal flow problem à la Beckmann:

$$\inf_{\sigma : \operatorname{div}(\sigma) = \mu_0 - \mu_1} \int_{\Omega} \mathcal{H}(\sigma(x)) dx \quad (4.21)$$

where $\mathcal{H}(\sigma) = H(|\sigma|)$ and H is taken independent of x for simplicity. Conversely, if σ is a minimizer of (4.21) and $Q \in \mathcal{Q}(\mu_0, \mu_1)$ is such that $i_Q = |\sigma|$ then Q solves the scalar problem (4.8) (i.e. is an equilibrium).

To build such a Q , we can formally use the following construction à la Moser (assuming σ smooth, μ_0, μ_1 have nice densities bounded away from 0). Consider the ODE

$$\dot{X}(t, x) = \frac{\sigma(X(t, x))}{(1-t)\mu_0(X(t, x)) + t\mu_1(X(t, x))}, \quad X(0, x) = x.$$

and define \bar{Q} by

$$\bar{Q} = \delta_{X(\cdot, x)} \otimes \mu_0$$

Set $\mu_t = (1 - t)\mu_0 + t\mu_1$ and

$$v(t, x) = \frac{\sigma(x)}{\mu_t(x)}$$

then by construction μ_t solves the continuity equation:

$$\partial_t \mu_t + \operatorname{div}(\mu_t v) = 0$$

By construction $e_{0\#}\overline{Q} = \mu_0$ and because of the continuity equation, $X(t, \cdot)\#\mu_0 = \mu_t = (1 - t)\mu_0 + t\mu_1$. In particular the image of μ_0 by the flow at time 1, $X(1, \cdot)$ is μ_1 , which proves that $e_{1\#}\overline{Q} = \mu_1$ hence $\overline{Q} \in \mathcal{Q}(\mu_0, \mu_1)$. Moreover for every test-function φ :

$$\begin{aligned} \int_{\Omega} \varphi di_{\overline{Q}} &= \int_{\Omega} \int_0^1 \varphi(X(t, x)) |v(t, X(t, x))| dt d\mu_0(x) \\ &= \int_0^1 \int_{\Omega} \varphi(x) |v(t, x)| \mu_t(x) dx dt \\ &= \int_{\Omega} \varphi(x) |\sigma(x)| dx \end{aligned}$$

so that $i_{\overline{Q}} = |\sigma|$ and then \overline{Q} is optimal.

The previous argument works as soon as σ is regular enough. By duality, the solution of (4.21) is $\sigma = \nabla \mathcal{H}^*(\nabla u)$ where \mathcal{H}^* is the Legendre transform of \mathcal{H} and u solves the PDE:

$$\begin{cases} \operatorname{div} \nabla \mathcal{H}^*(\nabla u) &= \mu_0 - \mu_1, & \text{in } \Omega, \\ \nabla \mathcal{H}^*(\nabla u) \cdot \nu &= 0, & \text{on } \partial\Omega, \end{cases} \quad (4.22)$$

Let us recall that $H' = g$ where g is the congestion function, so it is natural to have $g(0) > 0$: the metric is positive even if there is no traffic, so that the radial function \mathcal{H} is not differentiable at 0 and then its subdifferential at 0 contains a ball. By duality, this implies $\nabla \mathcal{H}^* = 0$ on this ball which makes (4.22) very degenerate. A reasonable model of congestion is $g(t) = 1 + t^{p-1}$ for $t \geq 0$, with $p > 1$, so that

$$\mathcal{H}(\sigma) = \frac{1}{p} |\sigma|^p + |\sigma|, \quad \mathcal{H}^*(z) = \frac{1}{q} (|z| - 1)_+^q, \quad \text{with } q = \frac{p}{p-1} \quad (4.23)$$

so that the optimal σ is

$$\sigma = \left(|\nabla u| - 1 \right)_+^{q-1} \frac{\nabla u}{|\nabla u|}$$

where u solves the very degenerate PDE:

$$\operatorname{div}\left(\left(|\nabla u| - 1\right)_+^{q-1} \frac{\nabla u}{|\nabla u|}\right) = \mu_0 - \mu_1, \quad (4.24)$$

with Neumann boundary condition

$$\left(|\nabla u| - 1\right)_+^{q-1} \frac{\nabla u}{|\nabla u|} \cdot \nu = 0.$$

Note that there is no uniqueness for u but there is for σ . Sobolev regularity of σ and Lipschitz regularity results for solutions of this PDE (more degenerate than the p -laplacian since the diffusion coefficient identically vanishes in the zone where $|\nabla u| \leq 1$) can be found in [15]. This enables one to build a flow à la DiPerna-Lions and then to justify rigorously the construction above

4.5 Dynamic setting : perspectives from mean-field games theory

The situation described of the previous paragraphs, is purely stationary (and this is reminiscent from Beckmann's formulation). It is not clear to us how to extend the previous analysis to the dynamic setting in a satisfactory way from the modeling point. Looking at learning dynamics might give some hints but we think that the most relevant answers come from mean-field games (MFG) theory introduced by Jean-Michel Lasry and Pierre-Louis Lions (see [53], [54], [55] and the inspiring lectures by Lions [56]). Concluding these notes is a good occasion, to say a few words on MFG theory that is, in our opinion, likely to become a new paradigm for analyzing rational expectations-like equilibria in the presence of general externalities and undoubtedly deserved to be called *New Mathematical Models in Economics Finance*.

Consider a certain time period $[0, T]$ and a population of players initially distributed according to a certain spatial density ρ_0 , denote by ρ_t the (unknown) players spatial distribution at time t . A generic player initially located at x , will typically minimize some cost that may depend on the repartition of the other players (for instance because of congestion) that in a first analysis stage, she takes as given, which leads to an individual program (for instance) of the form

$$\inf_{X(0,x)=x} \int_0^T L(\rho(s, X(s, x)), X(s, x), \dot{X}(s, x)) ds + \phi(X(T, x))$$

ignoring regularity issues, an optimal feedback velocity may be found by the dynamic programming approach i.e.

$$v = \nabla H(\rho, x, \nabla u) \quad (4.25)$$

where H is the usual Hamiltonian and u is the value function, that is characterized by the Hamilton-Jacobi equation:

$$\partial_t u + H(\rho, x, \nabla u) = 0, \quad u|_{t=T} = \phi. \quad (4.26)$$

This captures how players optimally move given ρ_t . At equilibrium, this should be consistent with the evolution of ρ_t induced by the velocity field (4.25) i.e. the continuity equation:

$$\partial_t \rho + \operatorname{div}(\rho \nabla H(\rho, x, \nabla u)) = 0, \quad \rho|_{t=0} = \rho_0. \quad (4.27)$$

An equilibrium is a situation where

- agents choose cost minimizing paths, given what they expect ρ_t will be
- this leads to an evolution which is indeed consistent with the previous expectations

it is therefore fully captured by the system (4.26)-(4.27). This system is a special case of MFG, note that it is nonlinear and it has the unusual feature that one has a terminal condition for the first equation and an initial condition for the second one. This forward-backward nature of the MFG system makes it mathematically very rich and captures the concept of rational expectation (in nature, a fixed-point problem) in a synthetic and elegant way.

Chapter 5

Appendix

5.1 Convex duality

Let E and F be two normed spaces, $\Lambda \in \mathcal{L}(E, F)$, f and g be two lsc convex functions, $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ and $g : F \rightarrow \mathbb{R} \cup \{+\infty\}$ further assumed to be *proper* i.e. not identically $+\infty$. The Legendre Transform of f denoted f^* is the function defined on E' by

$$f^*(q) := \sup_{x \in E} \{\langle q, x \rangle - f(x)\}, \quad \forall q \in E'$$

similarly, the Legendre Transform of g is defined by

$$g^*(p) := \sup_{y \in F} \{\langle p, y \rangle - g(y)\}, \quad \forall p \in F'.$$

Let us then consider the convex optimization problem:

$$\inf_{x \in E} \{f(x) + g(\Lambda x)\} \tag{5.1}$$

and let us define its *dual* as :

$$\sup_{p \in F'} \{-f^*(-\Lambda^* p) - g^*(p)\}. \tag{5.2}$$

Before going further and proving the Fenchel-Rockafellar theorem, we shall need a few preliminary observations. Given $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ a proper lsc and convex function and denoting by f^* its Legendre transform, we have by definition Young's inequality:

$$f(x) + f^*(q) \geq \langle q, x \rangle, \quad \forall (q, x) \in E' \times E, \tag{5.3}$$

so that for every $x \in E$:

$$f(x) \geq f^{**}(x) := \sup_{q \in E'} \{\langle q, x \rangle - f^*(q)\}. \tag{5.4}$$

Lemma 5.1 *Let $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper, lsc and convex function, then f^* is proper, lsc and convex on E^* .*

Proof:

The fact that f^* is lsc convex follows directly from the fact that by definition it is a supremum of affine and continuous functions. It remains to prove that f^* is not everywhere $+\infty$. Let $x_0 \in E$ be such that $f(x_0) < +\infty$ and $\lambda_0 < f(x_0)$, we then have $(\lambda_0, x_0) \notin \text{Epi}(f) := \{(\lambda, x) \in \mathbb{R} \times E : \lambda \geq f(x)\}$. Si f is lsc and convex, $\text{Epi}(f)$ is closed and convex, one can therefore separate strictly (λ_0, x_0) from $\text{Epi}(f)$: there exists $(k, p) \in \mathbb{R} \times E'$ and $\varepsilon > 0$ such that

$$k\lambda_0 - \langle p, x_0 \rangle \leq k\lambda - \langle p, x \rangle - \varepsilon, \quad \forall (\lambda, x) \in \text{Epi}(f) \quad (5.5)$$

which implies that $k > 0$; by homogeneity we may then as well assume $k = 1$. This yields in particular

$$f^*(p) = \sup_{x \in E} \{\langle p, x \rangle - f(x)\} \leq \langle p, x_0 \rangle - \lambda_0 - \varepsilon < +\infty.$$

□

The previous lemma exactly says that f possesses an affine continuous minorant $(x \mapsto \langle p, x \rangle - f^*(p))$ avec $f^*(p) < +\infty$.

Remark 8. As an exercise, we let the reader verify the following fact. Let $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$, $f \neq \infty$, f^{**} is the largest lsc convex function everywhere below f . In particular, f is lsc and convex if and only if $f = f^{**}$.

Theorem 5.1 (Fenchel-Rockafellar duality theorem) *If there exists $x_0 \in E$ such that $f(x_0) < +\infty$ and g is continuous at $\Lambda(x_0)$ and the infimum of (5.1) is finite, then :*

$$\inf_{x \in E} \{f(x) + g(\Lambda x)\} = \max_{p \in F'} \{-f^*(-\Lambda^* p) - g^*(p)\}.$$

(In particular, the supremum is attained in (5.2)).

Proof:

Let α and β be respectively the infimum in (5.1) and the supremum in (5.2). It follows from Young's inequalities that, for every $(x, p) \in E \times F'$ one has:

$$f(x) \geq \langle -\Lambda^* p, x \rangle - f^*(-\Lambda^* p), \quad g(\Lambda x) \geq \langle p, \Lambda x \rangle - g^*(p)$$

summing these inequalities and optimizing exactly yields $\alpha \geq \beta$.

Set

$$C := \{(\lambda, x, y) \in \mathbb{R} \times E \times Y : \lambda \geq g(\Lambda x - y)\}$$

and let A be the interior of C (A is nonempty since g is continuous at Λx_0). One can easily check that C is convex and dense in A . Also define:

$$B := \{(\mu, z, 0) : \mu \in \mathbb{R}, z \in E, \alpha - \mu \geq f(z)\},$$

B is nonempty, convex and by definition of α , $A \cap B = \emptyset$. One can therefore separate (in the large sense) B from A (and thus also from C by density): there exists $(k, q, p) \in \mathbb{R} \times E' \times F' \setminus \{(0, 0, 0)\}$ and $a \in \mathbb{R}$ such that

$$k\lambda + \langle q, x \rangle + \langle p, y \rangle \geq a \geq k\mu + \langle q, z \rangle, \quad \forall (\lambda, x, y) \in C, \quad \forall (\mu, z, 0) \in B. \quad (5.6)$$

We deduce from the previous that $k \geq 0$ (otherwise, the left hand side of (5.6) would not be bounded from below). Now, if $k = 0$, thanks to the continuity of g at Λx_0 , $u \in E$ and $v \in F$ small enough, one would have

$$\langle q, u \rangle + \langle p, v \rangle \geq 0$$

which would imply $p = 0$ and $q = 0$, which is the desired contradiction. We thus have $k > 0$ and may as well assume $k = 1$. We can therefore rewrite (5.6) as

$$\inf_{(x,y) \in E \times F} \{g(\Lambda x - y) + \langle q, x \rangle + \langle p, y \rangle\} \geq a \geq \alpha + \sup_{z \in E} \{\langle q, z \rangle - f(z)\} = \alpha + f^*(q). \quad (5.7)$$

In particular, for every $u \in E$

$$\langle q, u \rangle + \langle p, \Lambda u \rangle \geq a - g(\Lambda x_0)$$

so that $q = -\Lambda^* p$, and the left hand side of (5.7) can be rewritten as

$$\inf_{(x,y) \in E \times F} \{g(\Lambda x - y) - \langle p, \Lambda x - y \rangle\} = -g^*(p)$$

together with (5.7) this yields

$$-g^*(p) - f^*(-\Lambda^* p) \geq \alpha \geq \beta$$

which finally implies that $\alpha = \beta$ and that p solves (5.2).

□

5.2 A rigorous proof of the Kantorovich duality

Let us now come back to our Monge-Kantorovich problem (1.2) and let us prove that it may naturally be written as the dual of a certain convex minimization on $C(X) \times C(Y)$. Of course, thanks to Riesz's theorem, we identify, $\mathcal{M}(X)$, $\mathcal{M}(Y)$ and $\mathcal{M}(X \times Y)$ to the topological dual of $C(X)$, $C(Y)$ and $C(X \times Y)$. Let $\Lambda : C(X) \times C(Y)$ be defined by $\Lambda(\varphi, \psi) := \varphi \oplus \psi$ for every $(\varphi, \psi) \in C(X) \times C(Y)$ with

$$(\varphi \oplus \psi)(x, y) := \varphi(x) + \psi(y), \quad \forall (x, y) \in X \times Y.$$

The adjoint of Λ , Λ^* is then the linear and continuous operator $\mathcal{M}(X \times Y) \rightarrow \mathcal{M}(X) \times \mathcal{M}(Y)$ given by : for every $\gamma \in \mathcal{M}(X \times Y)$, $\Lambda^*\gamma = (\pi_X\gamma, \pi_Y\gamma)$ where for every $(\varphi, \psi) \in C(X) \times C(Y)$:

$$\begin{aligned} \int_{X \times Y} \varphi(x) d\gamma(x, y) &= \int_X \varphi(x) d(\pi_X\gamma)(x), \\ \int_{X \times Y} \psi(y) d\gamma(x, y) &= \int_Y \psi(y) d(\pi_Y\gamma)(y). \end{aligned}$$

This means that $\pi_X\gamma$ and $\pi_Y\gamma$ are the marginals of the (signed) measure γ .

Let us now consider:

$$\inf_{(\varphi, \psi) \in C(X) \times C(Y)} f(\Lambda(\varphi, \psi)) + g(\varphi, \psi) \quad (5.8)$$

where, for every $\theta \in C(X \times Y)$

$$g(\theta) := \begin{cases} 0 & \text{if } \theta \leq c \\ +\infty & \text{otherwise} \end{cases}$$

and

$$f(\varphi, \psi) := - \int_X \varphi d\mu - \int_Y \psi d\nu.$$

By direct computation, we get

$$f^*(-\Lambda^*\gamma) = \begin{cases} 0 & \text{if } (\pi_X\gamma, \pi_Y\gamma) = (\mu, \nu) \\ +\infty & \text{otherwise} \end{cases}$$

and

$$g^*(\gamma) = \begin{cases} \int_{X \times Y} cd\gamma & \text{si } \gamma \geq 0 \\ +\infty & \text{sinon} \end{cases}$$

so that the dual (in the sense of paragraph 5.1) of (5.8) is

$$\sup_{\gamma \in \Pi(\mu, \nu)} - \int_{X \times Y} cd\gamma = - \inf_{\gamma \in \Pi(\mu, \nu)} \int_{X \times Y} cd\gamma$$

by a direct application of the Fenchel-Rockafellar theorem, we obtain the existence of solutions to (1.2) (which we already knew!) and the Kantorovich duality formula:

Theorem 5.2 (Kantorovich duality formula)

$$\min_{\gamma \in \Pi(\mu, \nu)} \int_{X \times Y} c(x, y) d\gamma(x, y) = \sup_{(\varphi, \psi) \in C(X) \times C(Y) : \varphi \oplus \psi \leq c} \int_X \varphi d\mu + \int_Y \psi d\nu.$$

5.3 The Disintegration Theorem

A very useful consequence of the Radon-Nikodym theorem is the so-called disintegration theorem, a special case of which reads as:

Theorem 5.3 (Disintegration of a probability measure on a product with respect to one of its marginals) *Let X_1 and X_2 be two compact metric spaces equipped with their Borel σ -algebras, \mathcal{B}_{X_1} and \mathcal{B}_{X_2} , let γ be a Borel probability measure on $X_1 \times X_2$ and $\mu := \pi_{X_1} \gamma$ be its first marginal, then there exists a family of probability measures on X_2 , $(\gamma^{x_1})_{x_1 \in X_1}$ measurable in the sense that $x_1 \mapsto \gamma^{x_1}(A_2)$ is μ -measurable for every $A_2 \in \mathcal{B}_2$ and such that $\gamma = \gamma^{x_1} \otimes \mu$ i.e.*

$$\gamma(A_1 \times A_2) = \int_{A_1} \gamma^{x_1}(A_2) d\mu(x_1)$$

for every $A_1 \in \mathcal{B}_1$ and $A_2 \in \mathcal{B}_2$.

With the same notations as in the theorem, note that

$$\int_{X_1 \times X_2} \varphi(x_1, x_2) d\gamma(x_1, x_2) = \int_{X_1} \left(\int_{X_2} \varphi(x_1, x_2) d\gamma^{x_1}(x_2) \right) d\mu_1(x_1)$$

for every $\varphi \in C(X_1 \times X_2)$. The complete proof of the disintegration theorem is quite long and subtle at some points, we therefore do not reproduce it here and rather refer the interested reader to [36] or [75] for a proof. Of course, the notion of disintegration of a measure is tightly related to the notion of conditional law in probability. Let us now prove an easy consequence of the disintegration (which is useful, among other things, to prove the triangle inequality for Wasserstein distances)

Lemme 5.2 (Dudley's gluing Lemma) *Let X_i , $i = 1, 2, 3$ be compact metric spaces, equipped with their Borel σ -algebras. Let γ_{12} (resp. γ_{23}) be a Borel probability measure on $X_1 \times X_2$ (resp. $X_2 \times X_3$) with marginals μ_1, μ_2 (resp. μ_2, μ_3), then there exists a Borel probability measure γ on $X_1 \times X_2 \times X_3$ such that $\pi_{12}\gamma = \gamma_{12}$ and $\pi_{23}\gamma = \gamma_{23}$ (where obvious notations are used for the projections and corresponding marginals).*

Proof:

Disintegrate γ_{12} and γ_{13} with respect to their common marginal μ_2 :

$$\gamma_{12} = \eta^{x_2} \otimes \mu_2, \quad \gamma_{23} = \theta^{x_2} \otimes \mu_2$$

and define γ by

$$\gamma(A_1 \times A_2 \times A_3) := \int_{A_2} \eta^{x_2}(A_1) \theta^{x_2}(A_3) d\mu_2(x_2)$$

for every Borel sets A_1, A_2, A_3 . One easily verifies that γ satisfies the desired properties.

□

5.4 The continuity equation

Let v be a smooth vector-field $\mathbb{R}_+ \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that there is a constant C such that

$$|v(t, x)| \leq C(1 + |x|), \quad |v(t, x) - v(t, y)| \leq C|x - y|, \quad \forall (t, x, y).$$

Then, given $x \in \mathbb{R}^d$, let us define the flow map $t \mapsto X_t(x)$ as the value at time t of the solution of the nonautonomous ODE

$$\dot{y}(s) = v(s, y(s)), \quad y(0) = x.$$

In other words, X_t is characterized by

$$\partial_t X_t(x) = v(t, X_t(x)), \quad X_0(x) = x, \quad (t, x) \in \mathbb{R}_+ \times \mathbb{R}^d.$$

Our assumptions guarantee that the flow map is globally in time well defined and that for every $t > 0$, X_t is a diffeomorphism. Assume now, that we are given a probability measure ρ_0 on \mathbb{R}^d , that captures a certain initial spatial distribution of particles that follow the flow of v , a natural question is : how does this spatial distribution of particles evolve with time? In other words, how is the initial distribution ρ_0 transported by the flow of

v ? In transport terms, this amounts to characterize the curve of measures $t \mapsto X_{t\#}\rho_0$. We shall see that $X_{t\#}\rho_0$ is characterized by the following PDE called the *continuity* equation:

$$\partial_t \rho + \operatorname{div}(\rho v) = 0 \quad (5.9)$$

together, of course, with the initial condition:

$$\rho|_{t=0} = \rho_0. \quad (5.10)$$

Since we haven't made any regularity assumption on ρ_0 (ρ_0 could be a Dirac mass and then $X_{t\#}\rho_0$ would remain a Dirac mass for every $t > 0$), one has to understand the continuity equation in some appropriate weak sense i.e. in the sense of Distributions. We shall say that the family of probability measures $t \mapsto \rho_t$ is a measure-valued solution of (5.9)-(5.10) if :

- it is continuous in the sense that for every $\phi \in C_c(\mathbb{R}^d)$, the map

$$M_\phi : t \mapsto \int_{\mathbb{R}^d} \phi d\rho_t \text{ is continuous on } [0, \infty) \text{ and } M_\phi(0) = \int_{\mathbb{R}^d} \phi d\rho_0, \quad (5.11)$$

- for every $T > 0$, every $r > 0$ and every $\varphi \in C^1([0, T] \times \mathbb{R}^d)$ such that $\varphi(T, \cdot) = 0$ and $\varphi(t, \cdot)$ is supported by B_r for every $t \in [0, T]$ one has

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

Theorem 5.4 *The measure-valued curve $t \mapsto X_{t\#}\rho_0$ is the only measure-valued solution of (5.9)-(5.10).*

Proof:

First, it is clear that $t \mapsto \rho_t := X_{t\#}\rho_0$ satisfies the continuity requirement (5.11). Let us then check that $t \mapsto \rho_t := X_{t\#}\rho_0$ satisfies the continuity equation. Let φ be a test-function as above, then using the definition $\rho_t := X_{t\#}\rho_0$, Fubini's theorem and $\varphi(T, \cdot) = 0$, we have

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

so that $X_{t\#}\rho_0$ is a measure-valued solution of (5.9)-(5.10). Now, let us prove uniqueness, assume that $t \mapsto \rho_t$ and $t \mapsto \nu_t$ are two solutions and set $\mu_t := \rho_t - \nu_t$, then for every test-function φ as above, one has

$$\int_0^T \left(\int_{\mathbb{R}^d} (\partial_t \varphi(t, x) + v(t, x) \cdot \nabla \varphi(t, x)) d\mu_t(x) \right) dt = 0. \quad (5.12)$$

Let $\psi \in C_c((0, +\infty) \times \mathbb{R}^d)$ and let us consider the linear transport PDE (adjoint to the continuity equation, in some sense):

$$\partial_t \varphi + v \cdot \nabla \varphi = \psi \text{ on } (0, T) \times \mathbb{R}^d, \varphi(T, \cdot) = 0 \quad (5.13)$$

to solve this equation, we use the classical method of characteristics. Here it consists in rewriting the equation as

$$\frac{d}{dt} [\varphi(t, X_t(x))] = \psi(t, X_t(x)), \varphi(T, \cdot) = 0$$

which can be integrated directly as

$$\varphi(t, X_t(x)) = - \int_t^T \psi(s, X_s(x)) ds$$

which gives that the unique solution of (5.13) is given by

$$\varphi(t, x) = - \int_t^T \psi(s, X_s \circ X_t^{-1}(s)) ds.$$

This function is compactly supported in space uniformly in time $t \in [0, T]$, we can therefore use it as a test-function in (5.12), which gives that

$$\int_0^T \int_{\mathbb{R}^d} \psi(t, x) d\mu_t(x) dt = 0$$

since ψ is arbitrary this gives $\mu_t = 0$ for a.e. t and we conclude by continuity. \square

Remark 9. If, in addition Ω is some smooth bounded subset of \mathbb{R}^d , and v further satisfies the tangential condition

$$v(t, x) \cdot n(x) = 0, \forall t > 0, \forall x \in \partial\Omega \quad (5.14)$$

where $n(x)$ denotes the exterior normal vector to Ω at $x \in \partial\Omega$, then $\bar{\Omega}$ is invariant by the flow (trajectories starting in $\bar{\Omega}$ stay in $\bar{\Omega}$ forever).

Remark 10. The fact that the velocity field v is smooth and Lipschitz plays a crucial role through the regularity properties of the flow in the argument above. Understanding what happens precisely in the case of a nonsmooth v has been the subject of an intensive line of research that originated with the fundamental work of DiPerna and Lions [38].

5.5 Moser's deformation argument

We now aim to explain Moser's deformation argument (see [64], [34]) that is in fact direct and does not use the continuity equation. We are given smooth densities ρ_0 and ρ_1 on Ω , and we also assume that they are bounded away from zero. We then define the linear homotopy between ρ_0 and ρ_1 (or "teletransport deformation"):

$$\mu_t := (1 - t)\rho_0 + t\rho_1.$$

Now assume that σ is a smooth vector field that satisfies

$$\operatorname{div}(\sigma) = \rho_0 - \rho_1, \text{ in } \Omega, \quad \sigma \cdot n = 0 \text{ on } \partial\Omega. \quad (5.15)$$

For instance one can take $\sigma = \nabla u$ where u solves the Laplace equation with Neumann boundary condition. Then define the nonautonomous vector field v by

$$v(t, x) := \frac{\sigma(x)}{\mu_t(x)}$$

this vector field is smooth and we can define its flow X_t as well as the image of ρ_0 by the flow i.e. the measure

$$\rho_t := X_{t\#}\rho_0.$$

Our aim is to prove directly that $\rho_t = \mu_t$ (we already know that it is true since both μ_t and ρ_t solve the continuity equation with the same smooth velocity field and coincide at $t = 0$, one can then invoke the uniqueness result proved above). We then have to prove that $X_{t\#}^{-1}\mu_t = \rho_0$ i.e. that for every test-function φ , one has

$$\int \varphi(X_t^{-1}(y))\mu_t(y)dy = \int \varphi\rho_0$$

or equivalently, by the change of variables formula:

$$\int \varphi(x)\mu_t(X_t(x))|\det D_x X_t(x)|dx = \int \varphi\rho_0$$

this is obvious for $t = 0$, so that if we prove that

$$\frac{d}{dt} \left(\det(D_x X_t(x))((1 - t)\rho_0(X_t(x)) + t\rho_1(X_t(x))) \right) = 0 \quad (5.16)$$

then the desired result will follow. We first remark that $D_x X_t(x)$ solves

$$\partial_t D_x X_t(x) = D_x v(t, X_t(x))D_x X_t(x)$$

so that

$$D_x X_{t+h}(x) = (I + hD_x v(t, X_t(x)))D_x X_t(x) + o(h)$$

and thus

$$\begin{aligned} \det(D_x X_{t+h}(x)) &= \det(I + hD_x v(t, x)) \det(D_x X_t(x)) + o(h) \\ &= \det(D_x X_t(x))(1 + h\text{tr}(D_x v(t, X_t(x)))) + o(h) \end{aligned}$$

so that

$$\frac{d}{dt} \left(\det(D_x X_t(x)) \right) = \det(D_x X_t(x)) \text{div}(v(t, X_t(x)))$$

we thus have

$$\begin{aligned} \frac{d}{dt} \left(\det(D_x X_t(x))((1-t)\rho_0(X_t(x)) + t\rho_1(X_t(x))) \right) &= \det(D_x X_t(x)) \times \\ \left(\text{div}(v(t, X_t(x)))\mu_t(X_t(x)) + (\rho_1 - \rho_0)(X_t(x)) + \nabla \mu_t(X_t(x)) \cdot v(t, X_t(x)) \right) \\ &= \det(D_x X_t(x)) \left(\rho_1 - \rho_0 + \text{div}(\mu_t(X_t(x))v(t, X_t(x))) \right) \\ &= \det(D_x X_t(x)) \left(\rho_1 - \rho_0 + \text{div}(\sigma(X_t(x))) \right) = 0 \end{aligned}$$

In particular we have $\rho_1 = X_{1\#}\rho_0$ so that we have built a smooth and invertible transport from ρ_0 to ρ_1 .

5.6 Wasserstein distances

We end these notes, by introducing a class of distances on the space of probability measures on X , that are induced by some optimal transportation problems : the so-called Wasserstein distances. Again we assume that X is a compact metric space and denote by d its distance. Let us denote by $\mathcal{M}_1^+(X)$ the set of Borel probability measures on X . For $p \in [1, \infty)$, and given μ and ν in $\mathcal{M}_1^+(X)$ the p -Wasserstein distance between μ and ν is by definition

$$W_p(\mu, \nu) := \left(\inf_{\gamma \in \Pi(\mu, \nu)} \int_{X \times X} d(x, y)^p d\gamma(x, y) \right)^{1/p} \quad (5.17)$$

The fact that W_p is indeed a distance which metrizes the weak $*$ convergence is given by the following:

Théorème 5.1 *Let (X, d) be a compact metric space. For every $p \geq 1$, W_p is a distance on $\mathcal{M}_1^+(X)$. Moreover, if $(\mu_n)_n$ and μ belong to $\mathcal{M}_1^+(X)$, then (μ_n) converges weakly $*$ to μ if and only if $W_p(\mu_n, \mu) \rightarrow 0$ quand $n \rightarrow \infty$.*

Proof:

To prove that W_p is a distance, the only thing which really requires a proof is the triangle inequality. Let μ_1, μ_2 and μ_3 be $\mathcal{M}_1^+(X)$, let $\gamma_{12} \in \Pi(\mu_1, \mu_2)$ and $\gamma_{23} \in \Pi(\mu_2, \mu_3)$ be such that

$$\begin{aligned} W_p(\mu_1, \mu_2)^p &= \int_{X^2} d(x_1, x_2)^p d\gamma_{12}(x_1, x_2), \\ W_p(\mu_2, \mu_3)^p &= \int_{X^2} d(x_2, x_3)^p d\gamma_{23}(x_2, x_3). \end{aligned}$$

One deduces from lemma 5.2 the existence of a $\gamma \in \mathcal{M}_1^+(X^3)$ such that $\pi_{12}\gamma = \gamma_{12}$ and $\pi_{23}\gamma = \gamma_{23}$ so that $\gamma_{13} := \pi_{13}\gamma \in \Pi(\mu_1, \mu_3)$. Using the triangle and Minkowski's inequality, we thus get

$$\begin{aligned} W_p(\mu_1, \mu_3) &\leq \left(\int_{X \times X} d(x_1, x_3)^p d\gamma_{13}(x_1, x_3) \right)^{1/p} \\ &= \left(\int_{X^3} d(x_1, x_3)^p d\gamma(x_1, x_2, x_3) \right)^{1/p} \\ &\leq \left(\int_{X^3} (d(x_1, x_2) + d(x_2, x_3))^p d\gamma(x_1, x_2, x_3) \right)^{1/p} \\ &\leq \left(\int_{X^3} d(x_1, x_2)^p d\gamma(x_1, x_2, x_3) \right)^{1/p} \\ &\quad + \left(\int_{X^3} d(x_2, x_3)^p d\gamma(x_1, x_2, x_3) \right)^{1/p} \\ &= \left(\int_{X^2} d(x_1, x_2)^p d\gamma_{12}(x_1, x_2) \right)^{1/p} \\ &\quad + \left(\int_{X^2} d(x_2, x_3)^p d\gamma_{23}(x_2, x_3) \right)^{1/p} \\ &= W_p(\mu_1, \mu_2) + W_p(\mu_2, \mu_3). \end{aligned}$$

Now assume that $W_p(\mu_n, \mu)$ tends to 0 and let $\gamma_n \in \Pi(\mu_n, \mu)$ be such that

$$W_p(\mu_n, \mu)^p = \int_{X \times X} d(x, y)^p d\gamma_n.$$

Let $\varphi \in C(X)$ and let ω be the modulus of continuity of φ , we thus have

$$\begin{aligned} \left| \int_X \varphi d(\mu_n - \mu) \right| &= \left| \int_X (\varphi(x) - \varphi(y)) d\gamma_n(x, y) \right| \\ &\leq \int_{X \times X} \omega(d(x, y)) d\gamma_n(x, y) \end{aligned}$$

and thus

$$\limsup \left| \int_X \varphi d(\mu_n - \mu) \right| \leq \limsup \int_{X \times X} \omega(d(x, y)) d\gamma_n(x, y)$$

up to an extraction, we may assume that γ_n weakly $*$ converges to some γ , and that the limsup in the right hand side is in fact a limit. We then get

$$\int_{X \times X} d(x, y)^p d\gamma(x, y) = 0$$

and thus also

$$\limsup \int_{X \times X} \omega(d(x, y)) d\gamma_n(x, y) = \int_{X \times X} \omega(d(x, y)) d\gamma(x, y) = 0$$

so that (μ_n) weakly $*$ converges to μ . Conversely, assume now that (μ_n) weakly $-*$ converges to μ and let us prove that $W_p(\mu_n, \mu) \rightarrow 0$. First, up to dividing d by $\text{diam}(X)$, we can assume that $d \leq 1$ so that $W_p^p \leq W_1$. It is therefore enough to show that $W_1(\mu_n, \mu) \rightarrow 0$. Recalling the dual expression for W_1 given in (1.8), we easily deduce from Ascoli-Arzelà theorem (or directly from theorem 1.3) that there exists φ_n 1-Lipschitz such that

$$W_1(\mu_n, \nu) = \int_X \varphi_n d(\mu_n - \mu)$$

we may also assume that $\varphi_n(x_0) = 0$ where x_0 is a given point in X so that (φ_n) is bounded and equilipschitz. Using En appliquant Ascoli-Arzelà once again, up to a subsequence, we may therefore assume that (φ_n) converges uniformly to some φ and that $W_1(\mu_n, \mu)$ converges to $\limsup W_1(\mu_n, \mu)$. Thanks to the weak $*$ convergence of (μ_n) to (μ) , we get

$$\limsup W_1(\mu_n, \mu) = \lim \int_X \varphi_n d(\mu_n - \mu) = 0$$

which completes the proof.

□

Bibliography

- [1] L. Ambrosio, A. Pratelli, Existence and stability results in the L^1 theory of optimal transportation, Lecture Notes in Math. 2003.
- [2] L. Ambrosio, B. Kirchheim, A. Pratelli, Existence of optimal transport maps for crystalline norms, Duke Math. J., 125, 2004.
- [3] L. Ambrosio, N. Gigli, G. Savarè, Gradient flows in metric spaces and in the space of probability measure, Lectures in Mathematics ETH Zurich, Birkhäuser Verlag, Basel, 2005.
- [4] N. Ahmad, H. Kim, R.J. McCann, Extremal doubly stochastic measures and optimal transportation.
- [5] L. Ambrosio, Lecture Notes on Optimal Transport Problems, Lecture Notes in Mathematics (2000).
- [6] J.-B. Baillon, R. Cominetti, Markovian Traffic Equilibrium, Math. Prog., **111**, 2008.
- [7] M. Beckmann, A continuous model of transportation, *Econometrica* **20** (1952), 643-660.
- [8] M. Beckmann, C. McGuire, C. Winsten, Studies in Economics of Transportation. Yale University Press, New Haven, 1956.
- [9] J.D. Benamou, Y. Brenier, A numerical method for the optimal time-continuous mass transport problem and related problems, *Contemp. Math.*, 226, pp. 1–11, 1999.
- [10] J.D. Benamou, Y. Brenier, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem, *Numer. Math.*, 84, pp. 375–393, 2000.

- [11] F. Benmansour, G. Carlier, G. Peyré, F. Santambrogio, Numerical Approximation of Continuous Traffic Congestion Equilibria, *Netw. Heterog. Media* 4, no. 3, 605–623 (2009).
- [12] F. Benmansour, G. Carlier, G. Peyré, F. Santambrogio, Derivatives with Respect to Metrics and Applications: Subgradient Marching Algorithm, to appear in *Numerische. Mathematik*.
- [13] M. Berliant, S-K Peng, P. Wang, Production externalities and urban configuration, *J.Econ. Theory* 104, 2002.
- [14] Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions, *Communications on Pure and Applied Mathematics* 44, 375-417, 1991.
- [15] L. Brasco, G. Carlier, F. Santambrogio, Congested traffic dynamics, weak flows and very degenerate elliptic equations, *Journal de Math. Pures et Appliquées*, 2010.
- [16] A. Brancolini, G. Buttazzo, Optimal Networks For Mass Transportation Problems, *ESAIM Control Optim. Calc. Var.*, 2005.
- [17] G. Buttazzo, G. Carlier, Optimal spatial pricing strategies with transportation costs, *Proceedings subseries of the American Mathematical Society Contemporary Mathematics series*, 2010 .
- [18] G. Buttazzo, A. Pratelli, S. Solimini, E. Stepanov, Optimal urban networks via mass transportation, *Lecture Notes in Math.* 2009.
- [19] G. Buttazzo, A. Pratelli, E. Stepanov, Optimal pricing policies for public transportation networks, *SIAM J. Opt*, 2006.
- [20] G. Buttazzo, F. Santambrogio, A model for the optimal planning of an urban region, *SIAM J. Marh. Anal.*, 2005.
- [21] L.A. Caffarelli, R. J. McCann, M. Feldman, Constructing optimal maps in Monge’s transport problem as a limit of strictly convex costs, *J. Amer. Math. Soc.* 15, 2002.
- [22] L.A. Caffarelli, The regularity of mappings with a convex potential. *J. Amer. Math. Soc.*, 1992.
- [23] L.A. Caffarelli, Boundary regularity of maps with convex potentials. *Comm. Pure Appl. Math.* 45, 1992.

- [24] L.A. Caffarelli, Boundary regularity of maps with convex potentials II. *Ann. of Math. (2)* **144**, 1996.
- [25] G. Carlier, *Duality and existence for a class of mass transportation problems and economic applications*, *Adv. in Math. Econ.*, vol. 5 (2003), pp. 1–21.
- [26] G. Carlier, A general existence result for the principal-agent problem with adverse selection, *J. Math. Econom.* **35**, 2001.
- [27] G. Carlier, I. Ekeland, matching for teams, *Economic Theory*, 2010.
- [28] G. Carlier, I. Ekeland, Equilibrium structure of a bidimensional asymmetric city, *Nonlinear Analysis, Series B, Real World Applications*, Volume 8, Issue 3, 725-748 (2007).
- [29] G. Carlier, C. Jimenez, F. Santambrogio, Optimal transportation with traffic congestion and Wardrop equilibria, *SIAM J. Control Optim.* **47** (2008), 1330-1350.
- [30] G. Carlier, F. Santambrogio, A variational model for urban planning with traffic congestion, *ESAIM Control Optim. Calc. Var.* **11** (2007), 595-613.
- [31] T. Champion, L. De Pascale, The Monge problem for strictly convex norms in \mathbb{R}^d , to appear in *J. Eur. Math. Soc.*
- [32] T. Champion, L. De Pascale, The Monge problem in \mathbb{R}^d , to appear in *Duke Math. J.*
- [33] P.-A. Chiappori, R. J. McCann, L. Nesheim, Hedonic price equilibria, stable matching and optimal transport: equivalence, topology and uniqueness, *Economic Theory*, 2010.
- [34] B. Dacorogna, J. Moser, On a partial differential equation involving the Jacobian determinant, *Annales del'I.H.P.* **7** (1990), 1-26.
- [35] C.F. Daganzo, *Multinomial Probit: The theory and its application to demand forecasting*, Academic Press, New York, 1979.
- [36] C. Dellacherie, P.-A. Meyer, *Probabilities and Potential*, *Mathematical Studies* **29**, North-Holland (1978).
- [37] L. De Pascale and A. Pratelli, Regularity properties for Monge transport density and for solutions of some shape optimization problem. *Calc. Var. Partial Differential Equations* **14**, no. 3, 249–274, 2002.

- [38] R. J. DiPerna, P.-L. Lions, Ordinary differential equations, transport theory and Sobolev spaces, *Invent. Math.* **98** (1989), 511-547.
- [39] I. Ekeland, R. Temam, *Convex analysis and variational problems*. Classics in Mathematics, 28. SIAM, Philadelphia, PA, 1999.
- [40] I. Ekeland, An optimal matching problem, *ESAIM Contrôle Optimal et Calcul des Variations*, **11**, 1, 57-71 (2005).
- [41] I. Ekeland, Existence, uniqueness and efficiency of equilibrium in hedonic markets with multidimensional types, *Economic Theory*, 2010.
- [42] L.C. Evans, *Partial Differential Equations and Monge-Kantorovich Mass Transfer*, Current Developments in Mathematics, 1997.
- [43] L. C. Evans, W. Gangbo, Differential equations methods for the Monge-Kantorovich mass transfer problem, *Mem. Amer. Math. Soc.* **137**, 653 (1999).
- [44] A. Figalli, Y. Kim, R. J. McCann, When is multidimensional screening a convex program? preprint 2009.
- [45] Gangbo, W., An elementary proof of the polar factorization of vector-valued functions, *Arch. Rational Mech. Anal.*, 128, pp. 381–399, 1994.
- [46] M. Fujita, *Urban economic theory*, Cambridge University Press, 1989.
- [47] M. Fujita, P. Krugman, A.J. Venables, *The Spatial Economics*, MIT Press (1999).
- [48] M. Fujita and M. Ogawa, Equilibrium land use patterns in a monocentric city” *Journal of Regional Science* vol 20, 1980.
- [49] M. Fujita and M. Ogawa, Multiple equilibria and structural transition of non-monocentric urban configuration, *Regional Science and Urban Economics* vol 12, 1982.
- [50] M. Fujita and J. Thisse, *Economics of Agglomeration*, Cambridge University Press (2002).
- [51] Gangbo, W., McCann, R. J., The geometry of optimal transportation, *Acta Math.*, **177** (1996), 113–161.
- [52] H. Knothe, Contributions to the theory of convex bodies, *Michigan Mathematical Journal* 4, pp. 39–52, 1957.

- [53] J.-M. Lasry and P.-L. Lions, *Jeux à champ moyen. I. Le cas stationnaire. (French) [Mean field games. I. The stationary case]*, C. R. Math. Acad. Sci. Paris, **343**, no. 9, pp. 619–625 (2006).
- [54] J.-M. Lasry and P.-L. Lions, *Jeux à champ moyen. II. Horizon fini et contrôle optimal. (French) [Mean field games. II. Finite horizon and optimal control]*, C. R. Math. Acad. Sci. Paris, **343**, no. 10, pp. 679–684 (2006).
- [55] J.-M. Lasry and P.-L. Lions, *Mean field games*, Jpn. J. Math., **2**, no. 1, pp. 229–260 (2007).
- [56] P.-L. Lions, *Théorie des jeux de champ moyen et applications*, cours au Collège de France, videos can be viewed on the Collège de France website.
- [57] P.-L. Lions and J.-C. Rochet, Hopf formulas and multi-time Hamilton-Jacobi equations, Proc. Amer. Math. Soc. 96 (1986)
- [58] R. E. Lucas, Jr. and E. Rossi-Hansberg, On the Internal Structure of Cities, *Econometrica*, vol. 70 2002.
- [59] D. McFadden *Econometric Models of Probabilistic choice*, in *Structural Analysis of Discrete Data and Econometric Applications*, Charles F. Manski and Daniel L. McFadden, Editors Cambridge: The MIT Press, 1981.
- [60] D. McFadden, *Modelling the choice of residential location*, In A. Karlquist et al., editor, *Spatial interaction theory and residential location*, pages 75-96, Amsterdam, 1978. North-Holland.
- [61] X.-N. Ma, N. Trudinger and X.-J. Wang, Regularity of potential functions of the optimal transportation problem, *Arch. Rational Mech. Anal.* 177 (2005).
- [62] M. Mussa, S. Rosen, *Monopoly and product quality*, *J. Econom. Theory* 18, 1978.
- [63] J. Mirrlees, *Optimal tax theory : a synthesis*, *Journal of Public Economics*, 6, 1976.
- [64] J. Moser, *On the volume elements on a manifold*, *Trans. Am. Math. Soc.* **120** (1965), 286-294.

- [65] S.T. Rachev, L. Rüschendorf. Mass Transportation Problems. Vol. I: Theory; Vol. II : Applications, Springer-Verlag (1998).
- [66] J.-C. Rochet, The taxation principle and multi-time Hamilton-Jacobi equations, *Journal of Mathematical Economics*, 14:113–128, 1985.
- [67] J.-C. Rochet, A necessary and sufficient condition for rationalizability in a quasi-linear context. *Journal of Mathematical Economics*, 16, 1987.
- [68] J.-C. Rochet, P. Choné, Ironing, sweeping and multidimensional screening, *Econometrica*, 66, 1998.
- [69] M. Rosenblatt, Remarks on a multivariate transformation, *Annals of Mathematical Statistics* 23, pp. 470–472, 1952.
- [70] F. Santambrogio, Absolute continuity and summability of transport densities: simpler proofs and new estimates, *Calc. Var. PDEs*, 36, 2009.
- [71] F. Santambrogio, Transport and Concentration Problems with Interaction Effects, *J. Global Opt.*, 2007
- [72] M. Spence, Competitive and optimal responses to signals: An analysis of efficiency and distribution, *J. Econom. Theory* 7, 1974.
- [73] C. Villani, Topics in optimal transportation, AMS, 2004.
- [74] C. Villani, Optimal Transport, Old and New, Springer-Verlag, 2009.
- [75] C. Villani, Cours d'intégration et d'analyse de Fourier, ENS Lyon, 2008, <http://www.umpa.ens-lyon.fr/~cvillani/Cours/iaf-2007.html>
- [76] J. G. Wardrop, Some theoretical aspects of road traffic research, *Proc. Inst. Civ. Eng.* 2 (1952), 325-378.