Chapter 4

Divergence-constrained problems and transport density

4.1 Eulerian and Lagrangian points of view

4.1.1 Statical and Dynamical models

This section presents a very informal introduction to the physical interpretation of dynamical models in optimal transport.

In fluid mechanics - and in many other topics with similar modelizations - it is classical to consider two complementary ways of describing motions, which are called Lagrangian and Eulerian.

When we describe a motion via Lagrangian formalism we give "names" to particles (using either a specific label, or the initial position they had, for instance) and then describe, for every time t and every label, what happens to that particle. "What happens" means providing its position and/or its speed. Hence we could for instance give some functions $y_x(t)$ standing for the position at time t of particle originally located at x. Other possibility, instead of giving names we could consider bundles of particles with the same behavior and indicate how many are they. This amounts to giving a measure on possible behaviors.

The description may be more or less refined. For instance if one only considers two different times $t = 0$ and $t = 1$, the behavior of a particle is only given by its initial and final positions. A measure on those pairs (x, y) is exactly a transport plan. This explains how the Kantorovitch problem is expressed in Lagrangian coordinates. The Monge problem is also Lagrangian, where particles are labelled by their initial position.

More refined models can be easily conceived, since it quite evident that reducing a movement to the initial and final positions is embarrassingly poor. Measures on the set of paths (curves $\omega : [0,1] \to \Omega$, with possible assumptions on their regularity) have been used in many modelizations, and in particolar in traffic issues, branched transport (see the Discussion Section for both these subjects), or in Brenier's variational formulation of the incompressible Euler equations for fluids (see Section 1.7.4 and [36, 37, 25]).

On the other hand, in the Eulerian formalism we describe, for every time t and every point x, what happens at such a point at such a time. "What happens" usually means what are the velocity, the density and/or the flow rate (both in intensity and in direction) of particles located at time t at point x.

Eulerian models may be distinguished into statical and dynamical ones. In a dynamical model we usually use two variables, i.e. the density $\rho(t,x)$ and the velocity $v(t,x)$. It is possible to write the equation satisfied by the density of a family of particles moving according to the velocity field v. This means that we prescribe the initial density ρ_0, and that the position of the particle originally located at x will be given by the solution of the ODE

$$\begin{cases} y_x'(t) = v(t, y_x(t)) \\ y_x(0) = x \end{cases}, \tag{1.1}$$

we define the map T_t through $T_t(x) = y_x(t)$, and we look for the measure $\rho_t := (T_t)_{\#}\rho_0$. It is well known that ρ_t and v_t solve together the so-called continuity equation

$$\partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0$$

that is briefly addressed here below.

The statical framework is a bit harder to understand, since it is maybe not clear what "statical" means when we want to describe movement. One has to think to a permanent, cyclical movement, where some mass is constantly injected into the motion at some points and constantly withdrawn somewhere else. We can also think at a time average of some dynamical model: suppose that you observe the traffic in a city and you wonder what happens at each point, but you do not want an answer depending on time. You could for instance consider as a traffic intensity at every point the average traffic intensity at such a point on the whole day. In this case we usually use a unique variable v standing for the mass flow rate (which equals density times speed), and the divergence $\nabla \cdot v$ stands for the excess of mass which is injected into the motion at every point. More precisely, if particles are injected into the motion according to a density μ and then exit with density ν, the vector fields v standing for flows connecting these two measures must satisfy

$$\nabla \cdot v = \mu - \nu.$$

4.1.2 The continuity equation

This section is devoted to the equation

$$\partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0,$$

its meaning, formulations, and uniqueness results. Even if most of the Chapter will be devoted to the statical divergence equation, we will see later that the dynamical case can be useful to produce transport plans given a vector field, and we need to develop some tools.

First, let us spend some time on the notion of solution for this equation.

Definition 11. We say that a family of pairs measures/vector fields (ρ_t, v_t) with $v_t \in L^1(\rho_t; \mathbb{R}^d)$ and $\int_0^1 ||v_t||_{L^1(\rho_t)} dt = \int_0^1 \int_\Omega |v_t| d\rho_t \, dt < +\infty$ solves the continuity equation in the distributional sense if for any test function $\phi \in C^1_c(]0,1[\times \overline{\Omega})$, compactly supported in time but not necessarily in space, we have

$$\int_0^1 \int_\Omega \partial_t \phi \, d\rho_t \, dt + \int_0^1 \int_\Omega \nabla \phi \cdot v_t \, d\rho_t \, dt = 0. \tag{1.2}$$

Obviously this formulation includes Neumann boundary conditions on $\partial\Omega$ for v_t. If we want to impose the initial and final measures we can say that (ρ_t, v_t) solves the same equation, in the sense of distribution, with initial and final data ρ_0 and ρ_1, respectively, if for any test function $\phi \in C^1([0,1] \times \overline{\Omega})$ (no compact support assumptions), we have

$$\int_0^1 \int_\Omega \partial_t \phi \, d\rho_t \, dt + \int_0^1 \int_\Omega \nabla \phi \cdot v_t \, d\rho_t \, dt = \int_\Omega \phi(1,x) \, d\rho_1(x) - \int_\Omega \phi(0,x) \, d\rho_0(x). \tag{1.3}$$

On the other hand we can define a weak solution of the continuity equation through the following condition: we say that (ρ_t, v_t) solves the continuity equation in the weak sense if for any test function $\psi \in C^1([0,1] \times \Omega)$, the function $t \mapsto \int \psi \, d\rho_t$ is absolutely continuous and, for a.e. t, we have

$$\frac{d}{dt} \int_\Omega \psi \, d\rho_t = \int_\Omega \nabla \psi \cdot v_t \, d\rho_t.$$

Notice that in this case $t \mapsto \rho_t$ is automatically continuous for the weak convergence, and imposing the values of ρ_0 and ρ_1 may be done pointwisely.

Proposition 4.1.1. *The two notions of solutions are actually equivalent: every weak solution is actually a distributional solution and every distributional solution admits a representative (another family $\tilde{\mu}_t = \mu_t$ for a.e. t) which is weakly continuous and is a weak solution.*

Proof. To prove the equivalence, take a distributional solution, and test it against functions ϕ of the form $\phi(t,x) = a(t)\psi(x)$. We get

$$\int_0^1 a'(t) \int_\Omega \psi(x)\, d\rho_t\, dt + \int_0^1 a(t) \int_\Omega \nabla\psi \cdot v_t\, d\rho_t\, dt = 0.$$

The arbitrariness of a shows that the distributional derivative (in time) of $\int_\Omega \psi(x)\, d\rho_t$ is $\int_\Omega \nabla\psi \cdot v_t\, d\rho_t$. This last function is L^1 in time since $\int_0^1 \left|\int_\Omega \nabla\psi \cdot v_t\, d\rho_t\right| dt \leq \mathrm{Lip}\,\psi \int_0^1 ||v_t||_{L^1(\rho_t)} dt < +\infty$. This implies that (ρ, v) is a weak solution.

Conversely, the same computations shows that weak solution satisfy (1.2) for any ϕ of the form $\phi(t,x) = a(t)\psi(x)$. It is then enough to prove that finite linear combination of these functions are dense in $C^1([0,1]\times\mathbb{R}^n)$ (this is true, but is a non-trivial exercise!). □

It is also evident that smooth functions satisfy the equation in the classical sense if and only if they are weak (or distributional) solutions.

The main way to produce solutions to the continuity equation is to use the flow of the vector field v_t. Let us check the validity of the equation when ρ_t is obtained from such a flow through (1.1). Let us suppose that $\mathrm{spt}(\rho_t) \subset \Omega$ (which is satisfied if ρ_0 is concentrated on Ω and v satisfies suitable Neumann boundary conditions). We will check that we have a weak solution. Fix a test function $\phi : \Omega \to \mathbb{R}$ and compute

$$\begin{aligned}\frac{d}{dt}\int \phi\, d\rho_t &= \frac{d}{dt}\int \phi(y_x(t))\, d\rho_0(x) = \int \nabla\phi(y_x(t)) \cdot y_x'(t)\, d\rho_0(x) \\ &= \int \nabla\phi(y_x(t)) \cdot v(t, y_x(t))\, d\rho_0(x) = \int \nabla\phi(y) \cdot v(t,y)\, d\rho_t(y),\end{aligned}$$

which proves that we have $\partial_t\rho_t = -\nabla\cdot(\rho_t v_t)$, in the weak sense.

Then, we would like to give at least a uniqueness result on ρ if v is Lipschitz continuous. This is true in a very general framework (see [5], Proposition 8.2.7) for a proof of the fact that the solution in the space of measures is unique for given v), but we prefer to give an easier proof which requires to consider smooth solutions.

Theorem 4.1.2. *Suppose that Ω is a compact domain and $v : [0,T]\times\Omega \to \mathbb{R}^d$ is Lipschitz continuous in x uniformly in t, and consider two smooth solutions $\rho^{(1)}$ and $\rho^{(2)}$ of $\partial_t\rho_t + \nabla\cdot(\rho_t v_t) = 0$, with $\rho^{(1)}(0,x) = \rho^{(2)}(0,x)$. Then $\rho^{(1)} = \rho^{(2)}$.*

Proof. The equation being linear, we only need to consider a solution $\rho = \rho^{(1)}-\rho^{(2)}$ with $\rho(0,x) = 0$ and prove that it vanishes for every time. Consider $E(t) = \frac{1}{2}\int_\Omega \rho(t,x)^2dx$. We have

$$E'(t) = \int \rho_t(\partial_t\rho_t) = \int \nabla\rho_t\cdot v_t\rho_t = \int \nabla\big(\frac{1}{2}\rho_t^2\big)\cdot v_t = -\int \frac{1}{2}\rho_t^2\nabla\cdot v_t \le CE(t),$$

where we used $-\nabla\cdot v_t \le C$ as a consequence of the Lipschitz continuity assumption. A simple application fo Gronwall's lemma allows to prove $E(t) = 0$ for every t, since $E(0) = 0$, and gives the thesis. □

Later (Chapter 5) we will give a variant of this theorem to adapt to the case of unbounded domains.

We finish this section by proving that this result may be applied to the solution produced by the flow, which is actually smooth thanks to change-of variable formula allowing to reconstruct its density.

Proposition 4.1.3. *If ρ_0 is smooth and v is smooth, then ρ_t is smooth in t and x.*

Proof. If v_t is Lipschitz, the flow map is injective (as a well-known consequence of the uniqueness of the solution of the ODE). Hence, the density of the image measure is obtained from the initial density through a simple change-of-variable involving the Jacobian factor. This means that the regularity of $\rho(t,x)$ only depends on the regularity of the Jacobian $a(t,x) = \det(A(t,x))$ and $A(t,x) = D_xy(t,x)$ where $y(t,x) = y_x(t)$ is defined through (1.1).

Notice that we have $A(0,x) = Id$, $a(0,x) = 1$ and

$$A'(t,x) = \partial_t D_x y(t,x)) = D_x\big(\partial_t y(t,x)\big) = D_x\big(v_t(y(t,x))\big) = Dv_t(y(t,x))\cdot A(t,x),$$

which implies, thanks to usual matrix calculus

$$\begin{aligned} a'(t,x) &= a(t,x)\mathrm{trace}(A(t,x)^{-1}A'(t,x)) \\ &= a(t,x)\mathrm{trace}(A(t,x)^{-1}Dv_t(y(t,x))A(t,x)) \\ &= a(t,x)\mathrm{trace}(Dv_t(y(t,x)))) = a(t,x)\nabla\cdot v_t(y(t,x)). \end{aligned}$$

This means that if $\nabla\cdot v_t$ is bounded from below, then $a(t,x)$ never vanishes, and if $\nabla\cdot v_t$ is smooth in x, so is $a(t,x)$. The considerations below allow to deduce the regularity of $\rho(t,x)$. □

Memo – *Change-of-variable and image measures*

Proposition. Suppose that $\rho \in L^1(\Omega)$ is a positive density on $\Omega \subset \mathbb{R}^d$ and $T : \Omega \to \mathbb{R}^d$ is a Lipschitz injective map, which is thus differentiable a.e.. We suppose that $\det(DT) \neq 0$a.e. on $\{\rho > 0\}$. Then the image measure $T_\#\rho$ is absolutely continuous and its density u is given by

$$u(y) = \frac{\rho(T^{-1}(y))}{\det(DT(T^{-1}(y)))}.$$

If T is non-injective, the formula becomes $T_\#\rho = u \cdot \mathcal{L}^d$ with u given by

$$u(y) = \sum_{x:T(x)=y} \frac{\rho(x)}{\det(DT(x))}.$$

The same formulae stay true id T is only countably Lipschitz, with the differential DT which is actually the differential of the restriction of T to each set where it is Lipschitz continuous (and coincides thus with the approximate differential of T).

4.2 Beckmann's problem

4.2.1 Introduction, formal equivalences and variants

The problem that has been proposed by Beckmann as a model for optimal transport in the '50s, without knowing Kantorovitch's works and the possible links between the two theories, is the following.

Beckmann's minimal flow problem Consider the minimization

$$(PB) \quad \min \left\{ \int |v(x)|\, dx \;\middle|\; v : \Omega \to \mathbb{R}^n,\, \nabla \cdot v = \mu - \nu \right\}, \qquad (2.1)$$

where the divergence condition is to be read in the weak sense, with Neumann boundary conditions, i.e. $-\int \nabla\phi d\lambda = \int \phi\, d(\mu-\nu)$ for any $\phi \in C^1(\overline{\Omega})$.

This proposition links the Monge-Kantorovich problem to the minimal flow problem first proposed by Beckmann in [11], under the name of *continuous transportation model*. He did not know this link, as Kantorovich's theory was being developed independently almost in the same years.

We will see now that an equivalence between (PB) and (PK) holds true. To do that, we can look at the following considerations and formal computations.

We take the problem (PB) and re-write the constraint on v by means of the quantity

$$\sup_\phi \int -\nabla\phi \cdot v\, dx + \int \phi\, d(\mu - \nu) = \begin{cases} 0 & \text{if } \nabla \cdot v = \mu - \nu \\ +\infty & \text{otherwise} \end{cases}.$$

Hence one can write (PB) as

$$\min_v \int |v(x)|\, dx + \sup_\phi \int -\nabla\phi \cdot\, v\, dx + \int \phi\, d(\mu - \nu)$$
$$= \sup_\phi \int \phi\, d(\mu - \nu) + \inf_v \int |v(x)|\, dx - \int \nabla\phi \cdot\, v\, dx, \quad (2.2)$$

where inf and sup have been exchanged formally as in the previous computations. After that one notices that

$$\inf_v \int |v(x)|\, dx - \int \nabla\phi \cdot\, v\, dx = \begin{cases} 0 & \text{if } |\nabla\phi| \le 1 \\ -\infty & \text{otherwise} \end{cases}$$

and this leads to the dual formulation for (PB) which gives

$$\sup_{\phi\,:\,|\nabla\phi|\le 1} \int_\Omega \phi\ d(\mu - \nu).$$

Since this problem is exactly the same as (PD) (a consequence of the fact that Lip_1 functions are exactly those functions whose gradient is smaller than 1), this a formal equivalence between (PB) and (PK). The reason for saying that it is only formal lies in the fact that we did not prove the equality in (2.2). Notice that we need to suppose that Ω is convex, otherwise functions with gradient smaller than 1 are only Lip_1 according to the geodesic distance in Ω.

Most of the considerations above, especially those on the problem (PB) do not hold for costs other than the distance $|x-y|$. The only possible generalizations which are known concern a cost c which comes from a Riemannian distance $k(x)$.

The simplest possible generalization of Problem (PB) is the following:

$$\min \int k(x)|v(x)| dx \ :\ \nabla \cdot v = \mu - \nu$$

that corresponds, by duality with the functions u such that $|\nabla u| \le k$, to

$$\min \int d_k(x, y) d\gamma \ :\ \gamma \in \Pi(\mu, \nu),$$

where $d_k(x,y) = \inf_{\omega(0)=x,\,\omega(1)=y} L_k(\omega) := \int_0^1 k(\omega(t))|\omega'(t)|dt$ is the distance associated to the Riemannian metric k.

This generalization above comes from the modelization of a non-uniform cost for the movement (due to geographical obstacles or configurations). It can be applied to several situation but in some other t is not satisfying, for instance in urban transport, where we want to consider the fact that the metric k is usually not a priori known, but it depends on the traffic distribution itself. We will develop this aspect in the discussion section at the end of this chapter, together with a completely different problem which is somehow "opposite": instead of looking at transport problems where concentration of the mass is penalized because it stands for traffic congestion, looking at problemns where it is encouraged because of the so-called "economy of scale" (i.e. the biggest the mass you transport, the cheapest the individual cost).

4.2.2 Producing a minimizer for PB

The first remark on Problem (PB) is that it is probably not well-posed, in the sense that there could not exist an L^1 vector field minimizing the L^1 norm under divergence constraints. This is easy to understand if we think at using the direct method in Calculus of Variations to prove existence : we take a minimizing sequence v_n and we would like to extract a converging subsequence. If we could, and we had $v_n \rightharpoonup v$, then it would be easy to prove that v still satisfies $\nabla \cdot v = \mu - \nu$, since the relation

$$-\int \nabla\phi \cdot v_n \, dx = \int \phi \, d(\mu - \nu)$$

would pass to the limit as $n \to \infty$. Yet, the information that $\int |v(x)|dx \leq C$ is not enough to extract a converging sequence, even weakly. Indeed, the space L^1 being non-reflexive, bounded sequences are not guaranteed to have weakly converging subsequences. This is on the contrary the case for dual spaces (and for reflexive spaces, which are roughly speaking the dual of their dual).

Notice that the strictly convex version that is proposed for traffic purposes in the discussion section is much better to handle: if for instance we minimize $\int |v|^2 \, dx$ then we can use compactness in L^2, which is a Hilbert space, and hence reflexive.

To avoid this difficulty, one needs to set (PB) in the framework of vector measures.

Memo – *Vector measures*

Definition: A finite vector measure λ on a set Ω is a map associating to every Borel subset $A \subset \Omega$ a value $\lambda(A) \in \mathbb{R}^d$ such that, for every disjoint union $A = \bigcup_i A_i$ with $A_i \cap A_j = \emptyset$ for $i \neq j$, we have

$$\sum_i |\lambda(A_i)| < +\infty \quad \text{and} \quad \lambda(A) = \sum_i \lambda(A_i).$$

Here the norm that we use on $\mathbb{R}^d$ to evaluate the series above is arbitrary, the finiteness of the result does not depend on this choice, since all norms on $\mathbb{R}^d$ are equivalent.

We denote by $\mathcal{M}^d(\Omega)$ the set of finite vector measures on Ω. To such measures we can associate a positive scalar measure $\lambda \in \mathcal{M}_+(\Omega)$ through

$$||\lambda||(A) := \sup \left\{ \sum_i ||\lambda(A_i)|| \ : \ A = \bigcup_i A_i \ \text{ with } A_i \cap A_j = \emptyset \text{ for } i \neq j \right\}.$$

This measure depends on the choice of the norm $|| \cdot ||$ on $\mathbb{R}^d$. Let us suppose for simplicity that it is the Euclidean norm (in such a case, we will often write $|\lambda|$.

The integral of a Borel function $f : \Omega \to \mathbb{R}^d$ w.r.t. λ is well-defined if $|f| \in L^1(\Omega, ||\lambda||)$ (again, this does not depend on the choice of the norm), is denoted $\int f \cdot d\lambda$ and can be computed as $\sum_{i=1}^d \int f_i \, d\lambda_i$, thus reducing to integrals of scalar functions according to scalar measures. It could also be defined as a limit of integral of piecewise constant functions.

Functional analysis facts The quantity $||\lambda||(\Omega)$ is a norm on $\mathcal{M}^d(\Omega)$, and this normed space is the dual of $C_0(\Omega; \mathbb{R}^d)$, the space of continuous function on Ω vanishing at infinity,through the duality $(f, \lambda) \mapsto \int f \cdot d\lambda$. This gives a notion of $\overset{*}{\rightharpoonup}$ convergence for which bounded sets in $\mathcal{M}^d(\Omega)$ are compact.

A last clarifying fact is the following.

Proposition : For every $\lambda \in \mathcal{M}^d(\Omega)$ and every norm $|| \cdot ||$ there exists a Borel function $\xi : \Omega \to \mathbb{R}^d$ such that $\lambda = \xi \cdot ||\lambda||$ and $||\xi|| = 1$ a.e. (for the measure $||\lambda||$).

Sketch of proof: The existence of a function ξ is a consequence, via Radon-Nikodym Theorem, of $\lambda \ll ||\lambda||$ (every A set such that $||\lambda(A)|| = 0$ obviously satisfies $\lambda(A) = 0$), possibly applied componentwise. The condition $||\xi|| = 1$ may be proven by considering the sets $\{||\xi|| < 1-\varepsilon\}$ and $\{\xi \cdot e > a+\varepsilon\}$ for all hyperplane such that the unit ball $B_1 := \{x \in \mathbb{R}^d \ : \ ||x|| \leq 1\}$ is contained in $\{x \in \mathbb{R}^d \ : \ x \cdot e \leq a\}$ (and, actually, we have $B_1 = \bigcap_{e,a} \{x \in \mathbb{R}^d \ : \ x \cdot e \leq a\}$, the intersection being possibly reduced to a countable intersection). These sets must be negligible otherwise we have a contradiction on the definition of $||\lambda||$.

Theorem 4.2.1. *Suppose that Ω is a compact convex domain in $\mathbb{R}^d$. Then, the problem*

$$(PB) \quad \min \{ |v|(\Omega) \ : \ v \in \mathcal{M}^n(\Omega) \, , \, \nabla \cdot v = \mu - \nu \}$$

(with divergence imposed in the weak sense, i.e. for every $\phi \in C^1(\overline{\Omega})$ *we impose* $-\int \nabla\phi \cdot dv = \int \phi\, d(\mu - \nu)$*, which also includes Neumann boundary conditions) admits a solution. Moreover, its minimal value equals the minimal value of (PK) and a solution of (PB) can be built from a solution of (PK). The two problems are hence equivalent.*

Proof. The first point that we want to prove is the equality of the minimal values $(PB) = (PK)$ and we start from $(PB) \geq (PK)$. In order to do so, take an arbitrary function $\phi \in \mathrm{Lip}_1 \cap C^1$ and consider that for any v with $\nabla \cdot v = \mu - \nu$, we have

$$|v|(\Omega) = \int 1\, d|v| \geq \int (-\nabla\phi) \cdot dv = \int \phi\, d(\mu - \nu)$$

(where we used the fact that $\phi \in \mathrm{Lip}_1 \Rightarrow |\nabla\phi| \leq 1$). If one takes a sequence of $\mathrm{Lip}_1 \cap C^1$ functions converging to the Kantorovitch potential u such that $\int u\, d(\mu - \nu) = \max(PD) = \min(PK)$ (for instance take convolutions $\phi_k = \rho_k * u$) then he gets

$$\int d|v| \geq (PK)$$

for any admissible v, i.e. $(PB) \geq (PK)$.

We will show at the same time the reverse inequality and how to construct an optimal v from an optimal γ for (PK).

Actually, one way to produce a solution to this divergence-constrained problem, is the following: take an optimal transport plan γ and build a vector measure v_γ defined through

$$< v_\gamma, \phi >:= \int_{\Omega\times\Omega} \int_0^1 \omega'_{x,y}(t) \cdot \phi(\omega_{x,y}(t)) dt\, d\gamma,$$

for every $\phi \in C^0(\Omega; \mathbb{R}^d)$, $\omega_{x,y}$ being a parametrization of the segment $[x, y]$. Even if for this proof it would not be important, we will fix the constant speed parametrization, i.e. $\omega_{x,y}(t) = (1-t)x + ty$. It is clear that this is the point where convexity of Ω is needed.

It is not difficult to check that this measure satisfies the divergence constraint, since if one takes $\phi = \nabla\psi$ then

$$\int_0^1 \omega'_{x,y}(t) \cdot \phi(\omega_{x,y}(t)) = \int_0^1 \frac{d}{dt} (\psi(\omega_{x,y}(t))\, dt = \psi(y) - \psi(x)$$

and hence $< v_\gamma, \nabla\psi >= \int \psi\, d(\nu - \mu)$.

To estimate its mass we can see that $|v_\gamma| \le \sigma_\gamma$, where the scalar measure σ_γ is defined through

$$< \sigma_\gamma, \phi >:= \int_{\Omega\times\Omega} \int_0^1 |\omega'_{x,y}(t)|\phi(\omega_{x,y}(t))dt\, d\gamma, \quad \forall \phi \in C^0(\Omega; \mathbb{R})$$

and it is called *transport density*. Actually, we can even say more, since we can use

$$\omega'_{x,y}(t) = -|x-y|\frac{x-y}{|x-y|} = -|x-y|\nabla u(\omega_{x,y}(t)),$$

which is valid for every $t \in]0,1[$ and every $x, y \in \mathrm{spt}(\gamma)$ (so that $\omega_{x,y}(t)$ is in the interior of the transport ray $[x,y]$, if $x \neq y$; anyway for $x = y$, both expression vanish).

This allow to write, for every $\phi \in C^0(\Omega; \mathbb{R}^d)$

$$\begin{aligned} < v_\gamma, \phi >= \int_{\Omega\times\Omega} \int_0^1 -|x-y|\nabla u(\omega_{x,y}(t)) \cdot \phi(\omega_{x,y}(t))dt\, d\gamma \\ = -\int_0^1 dt \int \nabla u(\omega_{x,y}(t)) \cdot \phi(\omega_{x,y}(t))|x-y|d\gamma \end{aligned}$$

If we introduce the function $\pi_t : \Omega \times \Omega \to \Omega$ given by $\pi_t(x,y) = \omega_{x,y}(t)$, we get

$$< v_\gamma, \phi >= -\int_0^1 dt \int \nabla u(z) \cdot \phi(z) d\big((\pi_t)_\#(c\cdot\gamma)\big),$$

where $c \cdot \gamma$ is the measure on $\Omega \times \Omega$ with density $c(x,y) = |x-y|$ w.r.t. γ.

Since on the other hand the same kind of computations give

$$< \sigma_\gamma, \psi >= \int_0^1 dt \int \psi(z) d\big((\pi_t)_\#(c\cdot\gamma)\big),$$

we get $< v_\gamma, \phi >=< \sigma_\gamma, -\phi \cdot \nabla u >$, which shows

$$v_\gamma = -\nabla u \cdot \sigma_\gamma.$$

This gives the density of v_γ with respect to σ_γ and proves $|v_\gamma| \le \sigma_\gamma$.

The mass of σ_γ is obviously

$$\int d\sigma_\gamma = \int \int_0^1 |\omega'_{x,y}(t)|dt\, d\gamma = \int |x-y|d\gamma = \min(PK),$$

which proves the optimality of v_γ since no other v may do better than this, and also proves $\min(PB) = \min(PK)$. □

It is interesting to investigate whether $\sigma_\gamma \ll \mathcal{L}^d$, since this would imply that Problem (B) is well-posed in L^1 instead of the space of vector measure. For the sake of the variants that we will see later on, it would be interesting to give conditions so that $\sigma_\gamma \in L^p$ as well. All these subjects have been widely studied by De Pascale, Pratelli (see [58, 59, 60]) but there is a more recent (and shorter) proof of the same estimates in [93]. It is in particular true that $\mu, \nu \in L^p$ implies that $\sigma_\gamma \in L^p$ and that it is sufficient that one of the two measures is absolutely continuous in order to get the same on σ_γ.

Notice that it would be possible to prove, at least under some absolute continuity assumptions on μ or ν, (see Theorem 7.3 in [2]) that

- any minimizer of (PB) is given by v_γ for a suitable optimal transport plan ;
- all the optimal transport plans γ provide the same v_γ.

This induces in particular a uniqueness results for (PB) which is not obvious, since it a convex but not strictly convex problem.

4.2.3 Traffic intensity and traffic flows for measures on curves

We introduce in this section some objects that generalize both v_γ and σ_γ and hat will be useful both for proving the characterization of the optimal v as coming from an optimal plan γ and for the modelization issues of the Discussion Section.

Let us introduce some notations.

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

$$L_\varphi(\sigma) := \int_0^1 \varphi(\omega(t))|\dot{\omega}(t)|dt. \tag{2.3}$$

This quantity is the length of the curve weighted with the weight φ. When we take $\varphi = 1$ we get the usual length of ω and we denote it by $L(\omega)$ instead of $L_1(\omega)$.

We consider probability measures Q on $C := \mathrm{Lip}([0,1],\Omega)$. The convergence that we use on C is the uniform convergence with bounds on the Lipschitz constants, i.e. we say that $\omega_n \to \omega$ if $\mathrm{Lip}(\omega_n)$ is bounded and $\omega_n \to \omega$ uniformly. This is the same as the weak convergence in the space of Lipschitz curves. Notice that Ascoli-Arzelà's theorem guarantees that the sets $\{\omega \in C : \mathrm{Lip}(\omega) \leq c\}$ are compact for this convergence for every c. We will associate two measures on Ω to such a Q. The first is a scalar one, called traffic intensity and denoted by $i_Q \in \mathcal{M}(\Omega)$; it is defined by

$$\int \varphi \, di_Q := \int_C \Big(\int_0^1 \varphi(\omega(t))|\dot{\omega}(t)|dt \Big) dQ(\omega) = \int_C L_\varphi(\omega) dQ(\omega).$$

for all $\varphi \in C(\Omega, \mathbb{R}_+)$. 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.

We also associate a vector measure to this probability Q, in the same spirit as what we did in order to define v_γ. Let us consider the vector-field θ_Q defined through

$$\forall X \in C(\Omega, \mathbb{R}^d) \quad \int_\Omega X \cdot d\theta_Q := \int_C \left(\int_0^1 X(\omega(t)) \cdot \dot{\omega}(t) dt \right) dQ(\omega).$$

Since this is a kind of vectorial traffic intensity, we will call it *traffic flow*. Taking a gradient field $X = \nabla\psi$ in the previous definition yields

$$\int_\Omega \nabla\psi \cdot d\theta_Q = \int_{C([0,1],\Omega)} [\psi(\theta(1)) - \psi(\theta(0))] dQ(\gamma) = \int_\Omega \psi(\mu_1 - \mu_0)$$

where $\mu_i := (e_i)_\# Q$ for $i = 0, 1$. This means that

$$\nabla \cdot \theta_Q = \mu_0 - \mu_1.$$

Moreover it is easy to check that

$$|\theta_Q| \leq i_Q.$$

This last inequality is not in general an equality, since the curves of Q could produce some cancellations (imagine a non-negligible amount of curves passing through the same point with opposite directions, so that $\theta_Q = 0$ and $i_Q > 0$).

It is straightforward that the constructions of v_γ and σ_γ given in the previous section are just a particular case of this one, more precisely they are obtained in the case where Q is the image through the map associating to every pair (x, y) the segment $\omega_{x,y}$ of the measure $\gamma \in \mathcal{P}(\Omega \times \Omega)$, optimal transport plan for the Euclidean cost.

We need some properties of the traffic intensity and traffic flow.

Proposition 4.2.2. *Both θ_Q and i_Q are invariant under reparametrization (i.e., if $T : C \to C$ is a map such that for every ω the curve $T(\omega)$ is just a reparametrization in time of ω, then $\theta_{T_\# Q} = \theta_Q$ and $i_{T_\# Q} = i_Q$).*

For every Q, the total mass $i_Q(\Omega)$ equals the average length of the curves according to Q, i.e. $\int_C L(\omega)\,dQ(\omega) = i_Q(\Omega)$.

If $Q_n \rightharpoonup Q$ and $i_{Q_n} \rightharpoonup i$, then $i \geq i_Q$.

If $Q_n \rightharpoonup Q$ and $i_{Q_n} \rightharpoonup i_Q$ (i.e. if there is equality above), then $\theta_{Q_n} \rightharpoonup \theta_Q$.

Proof. The invariance by reparametrization comes from the fact that both $L_\varphi(\omega)$ and $\int_0^1 X(\omega(t))\cdot\omega'(t)dt$ do not change under reparametrization.

The formula $\int_C L(\omega)\,dQ(\omega) = i_Q(\Omega)$ is obtained from the definition of i_Q by testing with the function 1.

To check the inequality $i \geq i_Q$, fix a positive test function $\phi \in C(\Omega)$ and write

$$\int \phi\,di_{Q_n} = \int_C \left(\int_0^1 \phi(\omega(t))|\dot\omega(t)|dt\right) dQ_n(\omega). \tag{2.4}$$

Notice that the function $C \ni \omega \mapsto \int_0^1 \phi(\omega(t))|\dot\omega(t)|dt$ is positive and lower-semi-continuous w.r.t. ω. Indeed, if $\omega_n \to \omega$, then $\omega_n' \rightharpoonup \omega$ weakly-* in L^∞, which implies, up to subsequences, the existence of an L^∞ function $\xi \geq |\omega'|$ such that $|\omega_n'| \rightharpoonup \xi$; moreover, $\phi(\omega_n(t)) \to \phi(\omega(t))$ uniformly, which gives $\int \phi(\omega_n(t))|\omega_n'(t)|dt \to \int \phi(\omega(t))\xi(t)dt \geq \int \phi(\omega(t))|\omega'(t)|dt$.

This allows to pass to the limit in (2.4), thus obtaining

$$\begin{aligned}\int \phi\,di = \lim_n \int \phi\,di_{Q_n} &= \liminf_n \int_C \left(\int_0^1 \phi(\omega(t))|\dot\omega(t)|dt\right) dQ_n(\omega)\\ &\geq \int_C \left(\int_0^1 \phi(\omega(t))|\dot\omega(t)|dt\right) dQ(\omega) = \int \phi\,di_Q,\end{aligned}$$

which proves the claim.

To check the last property, fix a bounded vector test function X and look at

$$\begin{aligned}&\int X\cdot d\theta_{Q_n} = \int_C \left(\int_0^1 X(\omega(t))\cdot\dot\omega(t)dt\right) dQ_n(\omega)\\ &= \int_C \left(\int_0^1 X(\omega(t))\cdot\dot\omega(t)dt + ||X||_\infty L(\omega)\right) dQ_n(\omega) - ||X||_\infty i_{Q_n}(\Omega),\end{aligned} \tag{2.5}$$

where we just added and subtracted the total mass of i_{Q_n}, which is equal to the average length of ω according to Q_n.

Now notice that $C \ni \omega \mapsto \int_0^1 X(\omega(t))\cdot\dot\omega(t)dt + ||X||_\infty L(\omega)$ is a positive quantity and it is l.s.c. in ω (it is a consequence of what we proved above, by taking $\phi = 1$). This means that if we pass to the limit in (2.5) we get

$$\liminf_n \int X \cdot d\theta_{Q_n}$$
$$\geq \int_C \left(\int_0^1 X(\omega(t)) \cdot \dot{\omega}(t) dt + ||X||_\infty L(\omega) \right) dQ(\omega) - ||X||_\infty i_Q(\Omega)$$
$$= \int_C \left(\int_0^1 X(\omega(t)) \cdot \dot{\omega}(t) dt \right) dQ(\omega) = \int X \cdot d\theta_Q.$$

By replacing X with $-X$ we also get the opposite inequality and we have proven $\theta_{Q_n} \rightharpoonup \theta_Q$. □

Good to know! – *Dacorogna-Moser transport*

A particular case of the construction in [54] (first used in optimal transport by [63]):

Construction : Suppose that $w : \Omega \to \mathbb{R}^d$ is a Lipschitz vector field parallel to the boundary (i.e.$w \cdot n_\Omega = 0$ on $\partial\Omega$) with $\nabla \cdot w = f_0 - f_1$, where f_0, f_1 are positive probability densities which are Lipschitz continuous and bounded from below. Then we can define the non-autonomous vector field $\tilde{w}(t,x)$ via

$$\tilde{w}(t,x) = \frac{w(x)}{f_t(x)} \quad \text{where } f_t = (1-t)f_0 + tf_1$$

and consider the Cauchy problem

$$\begin{cases} y_x'(t) = \tilde{w}(t, y_x(t)) \\ y_x(0) = x \end{cases},$$

We define the map $Y : \Omega \to C$ through $Y(x) = y_x(\cdot)$, and we look for the measure $Q = Y_\# f_0$ and $\rho_t := (e_t)_\# Q$. Thanks to the consideration in Section 4.1.2, ρ_t solves the continuity equation $\partial_t \rho_t + \nabla \cdot (\rho_t \tilde{w}_t) = 0$. Yet, it is easy to check that f_t also solves the same equation since $\partial_t f_t = f_1 - f_0$ and $\nabla \cdot (\tilde{w} f_t) = \nabla \cdot w = f_0 - f_1$. By the uniqueness result of Section 4.1.2, from $\rho_0 = f_0$ we infer $\rho_t = f_t$.

In particular, $x \mapsto y_x(1)$ is a transport map from f_0 to f_1.

It it interesting to check what are the traffic intensity and the traffic flow associated to the measure Q in Dacorogna-Moser construction. Fix a scalar test function φ:

$$\int_\Omega \varphi di_Q = \int_\Omega \int_0^1 \varphi(y_x(t)) |\tilde{w}(t, y_x(t))| dt f_0(x) dx$$
$$= \int_0^1 \int_\Omega \varphi(y) |\tilde{w}(t,y)| f_t(y) dy dt = \int_\Omega \varphi(y) |w(y)| dy$$

so that $i_Q = |v|$. Analogously, fix a vector test function X

$$\int_\Omega X \cdot d\theta_Q = \int_\Omega \int_0^1 X(y_x(t)) \cdot \tilde{w}(t, y_x(t)) dt f_0(x) dx$$
$$= \int_0^1 \int_\Omega X(y) \cdot \tilde{w}(t, y) f_t(y) dy dt = \int_\Omega X(y) \cdot w(y) dy,$$

which shows $\theta_Q = w$ (indeed, in this case we have $|\theta_Q| = i_Q$ and this is due to the fact that no cancellation is possible, since all the curves share the same direction at every given point).

With these tools it is possible to prove that every admissible vector field v in Beckmann problem is of the form $v = \theta_Q$.

Lemma 4.2.3. *Consider two probabilities $\mu, \nu \in \mathcal{P}(\Omega)$ and a vector measure v satisfying $\nabla \cdot v = \mu - \nu$ in distributional sense (with Neumann boundary conditions). Then, for every domain Ω' containing Ω is its interios, there exist a family of vector fields $w^\varepsilon \in C^\infty(\Omega')$ with $w^\varepsilon \cdot n_{\Omega'} = 0$, and two families of densities $\mu^\varepsilon, \nu^\varepsilon \in C^\infty(\Omega')$, with $\nabla \cdot w^\varepsilon = \mu^\varepsilon - \nu^\varepsilon$ and $\int_{\Omega'} \mu^\varepsilon = \int_{\Omega'} \nu^\varepsilon = 1$, weakly converging to w, μ and ν as measures, respectively and satisfying $|w^\varepsilon| \rightharpoonup |v|$.*

Proof. First, take convolutions (in the whole space $\mathbb{R}^d$) with a gaussian kernel η_ε, so that we get $\hat{v}^\varepsilon := v * \eta_\varepsilon$ and $\hat{\mu}^\varepsilon := \mu * \eta_\varepsilon$ $\hat{\nu}^\varepsilon := \nu * \eta_\varepsilon$, still satisfying $\nabla \cdot \hat{v}^\varepsilon = \mu^\varepsilon - \nu^\varepsilon$. Since the Gaussian Kernel is strictly positive, we also have strictly positive densities for $\hat{\mu}^\varepsilon$ and $\hat{\nu}^\varepsilon$. These convolved densities and vector field would do the job required by the theorem, but we have to take care of the support (which is not Ω') and of the boundary behavior.

Let us set $\int_{\Omega'} \hat{\mu}^\varepsilon = 1 - a_\varepsilon$ and $\int_{\Omega'} \hat{\nu}^\varepsilon = 1 - b_\varepsilon$. It is clear that $a_\varepsilon, b_\varepsilon \to 0$ as $\varepsilon \to 0$. Consider also $\hat{v}^\varepsilon \cdot n_{\Omega'}$: due to $d(\Omega, \partial\Omega') > 0$ and to the fact that η_ε goes uniformly to 0 locally outside the origin, we also have $|\hat{v}^\varepsilon \cdot n_{\Omega'}| \leq c_\varepsilon$, with $c_\varepsilon \to 0$.

Consider u^ε the solution to

$$\begin{cases} \Delta u^\varepsilon = \frac{a_\varepsilon - b_\varepsilon}{|\Omega'|} & \text{inside } \Omega' \\ \frac{\partial u^\varepsilon}{\partial n} = -\hat{v}^\varepsilon \cdot n_{\Omega'} & \text{on } \partial\Omega', \\ \int_{\Omega'} u = 0 \end{cases}$$

and the vector field $\delta^\varepsilon = \nabla u^\varepsilon$. Notice that a solution exists thanks to $\int_{\partial\Omega'} \hat{v}^\varepsilon \cdot n_{\Omega'} = a_\varepsilon - b_\varepsilon$. Notice also that an integration by parts shows

$$\int_{\Omega'} |\nabla u^\varepsilon|^2 = -\int_{\partial\Omega'} u^\varepsilon (\hat{v}^\varepsilon \cdot n_{\Omega'}) - \int_{\Omega'} u^\varepsilon \left(\frac{a_\varepsilon - b_\varepsilon}{|\Omega'|}\right) \leq C ||\nabla u^\varepsilon||_{L^2} (c_\varepsilon + a_\varepsilon + b_\varepsilon),$$

and provides an estimate on $\int_{\Omega'} |\nabla u^\varepsilon|^2 \to 0$ (since we have a dependence of order two at the left hand side and of order one at the right hand side). This shows $||\delta^\varepsilon||_{L^2} \to 0$.

Now take

$$\mu^\varepsilon = \hat{\mu}^\varepsilon + \frac{a_\varepsilon}{|\Omega'|}; \quad \nu^\varepsilon = \hat{\nu}^\varepsilon + \frac{b_\varepsilon}{|\Omega'|}; \quad w^\varepsilon = \hat{v}^\varepsilon + \delta^\varepsilon,$$

and check that all the requirements are satisfied. In particular, the last one is satisfied since $||\delta^\varepsilon||_{L^1} \to 0$ and $|\hat{v}^\varepsilon| \rightharpoonup |v|$ by general properties of the convolutions. □

Remark 10. Notice that considering explicitly the dependence on Ω' it is also possible to obtain the same statement with a sequence of domains Ω'_ε converging to Ω (for instance in the Hausdorff topology). It is just necessary to choose them so that, setting $t_\varepsilon := d(\Omega, \partial\Omega'_\varepsilon)$, we have $||\eta_\varepsilon||_{L^\infty(B(0,t_\varepsilon)^c)} \to 0$.

With these tools we can now prove

Proposition 4.2.4. *For every finite vector measure $v \in \mathcal{M}^d(\Omega)$ and $\mu, \nu \in \mathcal{P}(\Omega)$ with $\nabla \cdot v = \mu - \nu$ there exist a measure $Q \in \mathcal{P}(C)$ with $(e_0)_\# Q = \mu$ and $(e_1)_\# Q = \nu$ such that $|\theta_Q| \le i_Q \le |v|$, with $|\theta_Q| \ne |v|$ unless $\theta_Q = v$.*

Proof. By means of Lemma 4.2.3 and Remark 10 we can produce an approximating sequence $(w^\varepsilon, \mu^\varepsilon, \nu^\varepsilon) \rightharpoonup (w, \mu, \nu)$ of C^∞ functions supported on domains Ω_ε converging to Ω. We apply Dacorogna-Moser's construction to this sequence of vector fields, thus obtaining a sequence of measures Q_ε. We can consider these measures as probability measures on $\mathrm{Lip}([0,1];\Omega')$, where $\Omega \subset \Omega_\varepsilon \subset \Omega'$ which are, each, concentrated on curves valued in Ω_ε. They satisfy $i_{Q_\varepsilon} = |w^\varepsilon|$ and $\theta_{Q_\varepsilon} = w^\varepsilon$. We can reparametrize (without changing their names) by constant speed the curves on which Q_ε is supported, without changing traffic intensities and traffic flows. This means using curves ω such that $L(\omega) = \mathrm{Lip}(\omega)$. The equalities

$$\int_C \mathrm{Lip}(\omega)\, dQ_\varepsilon(\omega) = \int_C L(\omega)\, dQ_\varepsilon(\omega) = \int_{\Omega'} i_{Q_\varepsilon} = \int_{\omega'} |w^\varepsilon| \to |v|(\Omega') = |v|(\Omega)$$

show that $\int_C \mathrm{Lip}(\omega)\, dQ_\varepsilon(\omega)$ is bounded and hence Q_ε is tight. Hence, up to subsequences, we can assume $Q_\varepsilon \to Q$. The measure Q is obviously concentrated on curves valued in Ω. The measures Q_ε were constructed so that $(e_0)_\# Q_\varepsilon = \mu^\varepsilon$ and $(e_1)_\# Q_\varepsilon = \nu^\varepsilon$, which implies, at the limit, $(e_0)_\# Q = \mu$ and $(e_1)_\# Q = \nu$. Moreover, thanks to Proposition 4.2.2, since $i_{Q_\varepsilon} =$

$|w_\varepsilon| \rightharpoonup |v|$, we get $|v| \geq i_Q \geq |\theta_Q|$. The same Proposition 4.2.2 also states that, if $|v| = i_Q$, then $\theta_{Q_\varepsilon} \rightharpoonup \theta_Q$. Yet, we also know $\theta_{Q_\varepsilon} = w_\varepsilon \rightharpoonup v$ and we deduce $v = \theta_Q$. □

CYCLES

4.2.4 Beckman problem in one dimension

The one-dimensional case is very easy in what concerns Beckmann's formulation of the optimal transport problem, but it is interesting to analyze it both for checking the consistency with the Monge's formulaltion and for using the results throughout next sections. We will take $\Omega = [a, b] \subset \mathbb{R}$.

First of all, notice that the condition $\nabla \cdot v = \mu - \nu$ is much stronger in dimension one than in higher dimension. Indeed, the divergence is the trace of the Jacobian matrix, and hence prescribing it only gives one constraint on a matrix which has a priori $d \times d$ degrees of freedom. On the contrary, in dimension one there is only one partial derivative for the vector field v (which is actually a scalar), and this completely prescribes the behavior of v. Indeed, the condition $\nabla \cdot v = \mu - \nu$ with Neumann boundary conditions implies that v must be the primitive of $\mu - \nu$ with $v(a) = 0$ (the fact that μ and ν have the same mass also implies $v(b) = 0$). Notice that the fact that its derivative is a measure gives $v \in BV([a, b])$.

Memo – *Bounded variation functions in one variable*

BV functions are generally defined as L^1 functions whose distributional derivatives are measures. In dimension one this has a lot of consequences. In particular these functions coincide a.e. with functions which have bounded total variation in a pointwise sense: for each $f : [a, b] \to \mathbb{R}$ define

$$TV(f; [a, b]) := \sup\{\sum_{i=0}^{N-1} |f(t_{i+1} - f_i| \ : \ a = t_0 < t_1 < t_2 < \cdots < t_N = b\}.$$

Functions of bounded total variation are defined as those f such that $TV(f; [a, b]) < \infty$. It is easy to check that BV functions are a vector space, and that monotone functions are BV (indeed, if f is monotone we have $TV(f; [a, b]) = |f(b)-f(a)|$.Lipschitz functions are also BV and $TV(f; [a, b]) \leq \mathrm{Lip}(f)(b - a)$. On the other hand, continuous functions are not necessarily BV, neither it is the case for differentiable functions (obviously, C^1 functions, which are Lipschitz on bounded intervals, are BV). As an example one can consider

$$f(x) = \begin{cases} \frac{x}{\log x} \sin\left(\frac{1}{x}\right) & \text{if } 0 < x \leq \frac{1}{2} \\ 0 & \text{for } x = 0, \end{cases}$$

which is differentiable everywhere but not BV (consider a partition using points of the form $x = 2/(k\pi)$ and use $\sum_k \frac{1}{k \log k} = +\infty$).

On the other hand, BV functions have several properties:

Properties of BV functions in $\mathbb{R}$ If $TV(f;[a,b]) < \infty$ then f is the difference of two monotone functions (in particular we can write $f(x) = TV(f;[a,x]) - (TV(f;[a,x]) - f(x))$, both terms being non-decreasing functions); it is a bounded function and $\sup f - \inf f \leq TV(f;[a,b])$; it has the same continuity and differentiability properties of monotone functions (it admits left and right limits at every point, it is continuous up to a countable set of points and differentiable a.e.).

In particular in dimension one we have $BV \subset L^\infty$ which is not the case in higher dimension (in general, we have $BV \subset L^{d/(d-1)}$).

We finish by stressing the connections with measures: for every positive measure μ on $[a,b]$ we can build a monotone function by taking its cumulative distribution function, i.e. $F(x) = \mu([a,x])$ and the distributional derivative of this function is exactly the measure μ. Conversely, every monotone increasing function on a compact interval is the cumulative distribution function of a (uique) positive measure, and every BV function is the cumulative distribution function of a (unique) signed measure.

As a consequence, we have the following facts:

- In dimension one, there is only one competitor v which is given by $v(x) = F(x) - G(x)$ with $F(x) = \mu([a,x])$ and $G(x) = \nu([a,x])$.
- This field v belongs to $BV([a,b])$ and hence to every L^p space, including L^∞.
- The minimal cost in Beckmann's problem is given by $||F-G||_{L^1}$, which is consistent with Proposition 2.2.2.
- The transport density σ, characterized by $v = -u' \cdot \sigma$ is given by $\sigma = |v|$ and shares the same summability properties of v; it also belongs to BV as a composition of a BV function with the absolute value function.

4.2.5 Characterization and uniqueness of the optimal v

In this section we will show two facts: first we prove that the optimal v in the Beckmann's problem always comes from an optimal transport plan γ and then we prove that all the optimal γs give the same v_γ and the same σ_γ, provided one of the two measures is absolutely continuous.

Theorem 4.2.5. *Let v be optimal in (PB): then there is an optimal transport plan γ such that $v = v_\gamma$.*

Proof. Thanks to Proposition 4.2.4, we can find a measure $Q \in \mathcal{P}(C)$ with $(e_0)_\# Q = \mu$ and $(e_1)_\# Q = \nu$ such that $|v| \geq |\theta_Q|$ Yet, the optimality of v implies the equality $|v| = |\theta_Q|$ and the same Proposition 4.2.4 gives in such a case $v = \theta_Q$, as well as $|v| = i_Q$. We assume Q to be concentrated on curves parametrized by constant speed. Define $S : \Omega \times \Omega \to C$ the map associating to every pair (x, y) the segment $\omega_{x,y}$ parametrized with constant speed: $\omega_{x,y}(t) = (1-t)x + ty$. The statement is proven if we can prove that $Q = S_\# \gamma$ wit γ an optimal transport plan.

Indeed, using again the optimality of v and Proposition 4.2.4, we get

$$\begin{aligned}\min(PB) = |v|(\Omega) = i_Q(\Omega) = \int_C L(\omega)\, dQ(\omega \geq \int_C |\omega(0) - \omega(1)| dQ(\omega)\\ = \int_{\Omega\times\Omega} |x-y|\, d((e_0, e_1)_\# Q)(x, y) \geq \min(PK).\end{aligned}$$

The equality $\min(PB) = \min(PK)$ implies that all these inequalities are equalities. In particular Q must be concentrated on curves such that $L(\omega) = |\omega(0) - \omega(1)|$, i.e. segments. Also, the measure $(e_0, e_1)_\# Q$, which belongs to $\Pi(\mu, \nu)$, must be optimal in (PK). This concludes the proof. □

The proof of the following result is essentially taken from [2].

Theorem 4.2.6. *If $\mu \ll \mathcal{L}^d$, then the vector field v_γ does not depend on the choice of the optimal plan γ.*

Proof. Let us fix a Kantorovich potential u for the transport between μ and ν. This potential does not depend on the choice of γ. It determines a partition into transport rays: Corollary 3.1.4 reminds us that the only points of Ω which belong to several transport rays are non-differentiability points for u, and are hence Lebesgue-negligible. Let us call S the set of points which belong to several transport rays: we have $\mu(S) = 0$ but we do not suppose $\nu(S) = 0$ (ν is not supposed to be absolutely continuous). However, γ is concentrated on $(\pi_x)^{-1}(S^c)$. We can then disintegrate (see Section 2.3) γ according to the transport ray containing the point x. More precisely, we define a map $R : \Omega\times\Omega \to \mathcal{R}$, valued in the set $\mathcal{R}$ of all transport rays, sending each pair (x, y) into the ray containing x. This is well-defined $\gamma-$a.e. and we can write $\gamma = \gamma^r \otimes \lambda$, where $\lambda = R_\# \gamma$ and we denote by r the variable related to transport rays. Notice that, for a.e. $r \in \mathcal{R}$, the plan γ_r is optimal between its own marginals (otherwise we could replace it with an optimal plan, do it in a measurable way, and improve the cost of γ).

The measure v_γ may also be obtained through this disintegration, and we have $v_\gamma = v_{\gamma^r} \otimes \lambda$. This means that, in order to prove that v_γ does not depend on γ, we just need to prove that each v_{γ^r} and the measure λ do not depend on it. For the measure λ this is easy: it has been obtained as an image measure through a map only depending on x, and hence only depends on μ. Concerning v_{γ^r}, notice that it is obtained in the standard Beckmann way from an optimal plan, γ^r. Hence, thanks to the considerations in Section 4.2.4, it uniquely depends on the marginal measures of this plan.

This means that we only need to prove that $(\pi_x)_\#\gamma^r$ and $(\pi_y)_\#\gamma^r$ do not depend on γ. Again, this is easy for $(\pi_x)_\#\gamma^r$, since it must coincide with the disintegration of μ according to the map R (by uniqueness of the disintegration). It is more delicate for the second marginal.

The second marginal $\nu^r := (\pi_y)_\#\gamma^r$ will be decomposed in two parts: $(\pi_y)_\#(\gamma^r_{|\Omega\times S})$ and $(\pi_y)_\#(\gamma^r_{|\Omega\times S^c})$. This second part coincides with the disintegration of $\nu_{|S^c}$, which obviously does not depend on γ (since it only depends on the set S, which is built upon u).

We need now to prove that $\nu^r_{|S} = (\pi_y)_\#(\gamma^r_{|\Omega\times S})$ does not depend on γ. Yet, this measure can only be concentrated on the two endpoints of the transport ray r, since these are the only points where different transport rays can meet. This means that this measure is purely atomic and composed by at most two Dirac masses. Not only, the endpoint where u is maximal cannot contain some mass of ν: indeed the transport must follow a precise direction on each transport ray (as a consequence of $u(x)-u(y) = |x-y|$ on $\mathrm{spt}(\gamma)$), and the only way to have some mass of the target measure at the "beginning" of the transport ray would be to have an atom for the source measure as well. Yet, μ is absolutely continuous and Property N holds (see Section 3.1.4 and Theorem 3.1.7, which means that the set of rays r where μ^r has an atom is negligible. Hence $\nu^r_{|S}$ is a single Dirac mass. The mass equilibrium condition between μ^r and ν^r implies that the value of this mass must be equal to the difference $1-\nu^r_{|S^c}(r)$, and this last quantity does not depend on γ but only on μ and ν.

Finally, this proves that each v_{γ^r} does not depend on the choice of γ. □

Corollary 4.2.7. *If $\mu \ll \mathcal{L}^d$, then the optimal solution of (PB) is unique.*

Proof. We have seen in Theorem 4.2.5 that any optimal v is of the form v_γ and in Theorem 4.2.6 that all the fields v_γ coincide. □

4.3 Summability of the transport density

The analysis of Beckmann problem performed in the previous sections was mainly made in a measure setting, and the optimal v, as well as the transport density σ, where just measures on Ω. We investigate here the question whether they have extra regularity properties supposing extra assumptions on μ and/or ν.

We will give summability results, proving that σ is in some cases absolutely continuous and proving L^p estimates. The proofs are essentially taken from [93]: previous results, through very different techniques, were first presented in [58, 59, 60]. In these papers, different estimates on the "dimension" of σ are aslo presented, thus giving interesting information should σ fail to be absolutely continuous.

Notice that higher order questions, such as whether σ is continuous or Lipschitz or more regular provided μ and ν have smooth densities are completely open up to now (with the exception of a partial result in dimension 2, see [**?**], where a continuity result is given if μ and ν have Lipschitz densities on disjoint convex domains).

In all that follows Ω is a compact and convex domain in $\mathbb{R}^d$, and two probability measures are given on it. Since we will need to interpolate between them, we will rather call them μ_0 and μ_1 (and the interpolation will be called μ_t). At least one of them will be absolutely continuous, which implies uniqueness for σ (see Theorem 4.2.6).

Theorem 4.3.1. *Suppose $\mu_0 \ll \mathcal{L}^d$ and let σ be the transport density associated to the transport of μ_0 onto μ_1. Then $\sigma \ll \mathcal{L}^d$.*

Proof. Let γ be an optimal transport from μ_0 to μ_1 and take $\sigma = \sigma_\gamma$; call μ_t the standard interpolation between the two measures: $\mu_t = (\pi_t)_\# \gamma$ where $\pi_t(x, y) = (1 - t)x + ty$.

We have already seen that the transport density σ may be written as

$$\sigma = \int_0^1 (\pi_t)_\sharp (c \cdot \gamma) dt,$$

where $c : \Omega \times \Omega \to \mathbb{R}$ is the cost function $c(x, y) = |x - y|$ (hence $c \cdot \gamma$ is a positive measure on $\Omega \times \Omega$).

Since Ω is bounded it is evident that we have

$$\sigma \le C \int_0^1 \mu_t \, dt. \tag{3.1}$$

To prove that σ is absolutely continuous, it is sufficient to prove that almost every measure μ_t is absolutely continuous, so that, whenever $|A| = 0$, we have $\sigma(A) \le C \int_0^1 \mu_t(A) dt = 0$.

We will prove $\mu_t \ll \mathcal{L}^d$ for $t < 1$. First, we will suppose that μ_1 is finitely atomic (the point $(x_i)_{i=1,\dots,N}$ being its atoms). In this case we will choose

γ to be any optimal transport plan induced by a transport map T (which exists, since $\mu_0 \ll \mathcal{L}^d$). Notice that the absolute continuity of σ is an easy consequence of the behavior of the optimal transport from μ_0 to μ_1 (which is composed by N homothecies), but we also want to quantify this absolute continuity, in order to go on with an approximation procedure.

Remember that μ_0 is absolutely continuous and hence there exists a correspondence $\varepsilon \mapsto \delta = \delta(\varepsilon)$ such that

$$|A| < \delta(\varepsilon) \Rightarrow \mu_0(A) < \varepsilon. \tag{3.2}$$

Take now a Borel set A and look at $\mu_t(A)$. The domain Ω is the disjoint union of a finite number of sets $\Omega_i = T^{-1}(\{x_i\})$. We call $\Omega_i(t)$ the images of Ω_i through the map $x \mapsto (1-t)x + tT(x)$. These sets are essentially disjoint. Why? because if a point z belongs to $\Omega_i(t)$ and $\Omega_j(t)$, then two transport rays cross at z, the one going from $x_i' \in \Omega_i$ to x_i and the one from $x_j' \in \Omega_j$ to x_j. The only possibility is that these two rays are actually the same, i.e. that the five points x_i', x_j', z, x_i, x_j are aligned. But this implies that z belongs to one of the lines connecting two atoms x_i and x_j. Since we have finitely many of these lines this set is negligible. Notice that this argument only works for $d > 1$ (we will not waste time on the case $d = 1$, since the transport density is always a BV and hence bounded function). Moreover, if we sticked to the optimal transport which is monotone on transport rays, we could have actually proved that these sets are truly disjoint, with no negligible intersection.

Hence we have

$$\mu_t(A) = \sum_i \mu_t(A \cap \Omega_i(t)) = \sum_i \mu_0\left(\frac{A \cap \Omega_i(t) - tx_i}{1-t}\right) = \mu_0\left(\bigcup_i \frac{A \cap \Omega_i(t) - tx_i}{1-t}\right).$$

Since for every i we have

$$\left|\frac{A \cap \Omega_i(t) - tx_i}{1-t}\right| = \frac{1}{(1-t)^d}\,|A \cap \Omega_i(t)|$$

we have

$$\left|\bigcup_i \frac{A \cap \Omega_i(t) - tx_i}{1-t}\right| \leq \frac{1}{(1-t)^d}|A|.$$

Hence it is sufficient to suppose $|A| < (1-t)^d \delta(\varepsilon)$ to get $\mu_t(A) < \varepsilon$. This confirms $\mu_t \ll \mathcal{L}^d$ and gives an estimate that may pass to the limit.

Take a sequence $(\mu_1^n)_n$ of atomic measures converging to μ_1. The corresponding optimal transport plans γ^n converge to an optimal transport plan γ

and μ_t^n converge to the corresponding μ_t (see Theorem 1.6.10 in Chapter 1). Hence, to prove absolute continuity for the transport density σ associated to such a γ it is sufficient to prove that these μ_t are absolutely continuous.

Take a set A such that $|A| < (1-t)^d\delta(\varepsilon)$. Since the Lebesgue measure is regular, A is included in an open set B such that $|B| < (1-t)^d\delta(\varepsilon)$. Hence $\mu_t^n(B) < \varepsilon$. Passing to the limit, thanks to weak convergence and semicontinuity on open sets, we have

$$\mu_t(A) \leq \mu_t(B) \leq \liminf_n \mu_t^n(B) \leq \varepsilon.$$

This proves $\mu_t \ll \mathcal{L}^d$ and hence $\sigma \ll \mathcal{L}^d$. □

Remark 11. Where did we use the optimality of γ? we did it when we said that the $\Omega_i(t)$ are disjoint. For a discrete measure μ_1, it is always true that the measures μ_t corresponding to any transport plan γ are absolutely continuous for $t < 1$, but their absolute continuity may degenerate at the limit if we allow the sets $\Omega_i(t)$ to superpose (since in this case densities sum up and the estimates may depend on the number of atoms).

Remark 12. Notice that we strongly used the equivalence between the two different definitions of absolute continuity, i.e. the $\varepsilon \leftrightarrow \delta$ correspondence on the one hand and the condition on negligible sets on the other. Indeed, to prove that the condition $\mu_t \ll \mathcal{L}^d$ passes to the limit we need the first one, while to deduce $\sigma \ll \mathcal{L}^d$ we need the second one, since if we deal with non-negligible sets we have some $(1-t)^d$ factor to deal with...

Remark 13. As a byproduct of the proof we can see that any optimal transport pla from μ_0 to μ_1 which is approximable through optimal transport plans from μ_0 to atomic measures must be such that all the interpolating measures μ_t are absolutely continuous. This property is not satisfied by any optimal transport plan, since for instance the plan γ which sends $\mu_0 = \mathcal{L}^2_{|[-2,-1]\times[0,1]}$ onto $\mu_1 = \mathcal{L}^2_{|[1,2]\times[0,1]}$ moving (x,y) to $(-x,y)$ is optimal but is such that $\mu_{1/2} = \mathcal{H}^1_{|\{0\}\times[0,1]}$. Hence, this plan cannot be approximated by optimal plans sending μ_0 onto atomic measures. On the other hand, we proved in Lemma 3.1.12 that the monotone optimal transport can indeed be approximated in a similar way.

In the previous theorem we did not treat the one dimensional case, which is highly detailed in Section 4.2.4.

From now on we will often confuse absolutely continuous measures with their densities and write $||\mu||_p$ for $||f||_{L^p(\Omega)}$ when $\mu = f \cdot \mathcal{L}$.

Theorem 4.3.2. *Suppose* $\mu_0 = f \cdot \mathcal{L}^d$, *with* $f \in L^p(\Omega)$. *The, if* $p < d' := d/(d-1)$, *the unique transport density* σ *associated to the transport of* μ_0 *onto* μ_1 *belongs to* $L^p(\Omega)$ *as well, and if* $p \geq d'$ *it belongs to any space* $L^q(\Omega)$ *for* $q < d'$.

Proof. Start from the case $p < d'$: following the same strategy (and the same notations) as before, it is sufficient to prove that each measure μ_t (for $t \in [0,1[$) is in L^p and to estimate their L^p norm. Then we will use

$$||\sigma||_p \leq C \int_0^1 ||\mu_t||_p dt,$$

(which is a consequence of (3.1) and of Minkowski inequality), the conditions on p being chosen exactly so that this integral converges.

Consider first the discrete case: we know that μ_t is absolutely continuous and that its density coincides on each set $\Omega_i(t)$ with the density of an homothetic image of μ_0 on Ω_i, the homothecy ratio being $(1-t)$. Hence, if f_t is the density of μ_t, we have

$$\begin{aligned}\int_\Omega f_t(x)^p dx &= \sum_i \int_{\Omega_i(t)} f_t(x)^p dx = \sum_i \int_{\Omega_i} \left(\frac{f(x)}{(1-t)^d}\right)^p (1-t)^d dx \\ &= (1-t)^{d(1-p)} \sum_i \int_{\Omega_i} f(x)^p dx = (1-t)^{d(1-p)} \int_\Omega f(x)^p dx.\end{aligned}$$

We get $||\mu_t||_p = (1-t)^{-d/p'} ||\mu_0||_p$, where $p' = p/(p-1)$ is the conjugate exponent of p.

This inequality, which is true in the discrete case, stays true at the limit as well. If μ_1 is not atomic, approximate it through a sequence μ_1^n and take optimal plans γ^n and interpolating measures μ_t^n. Up to subsequences we have $\gamma^n \rightharpoonup \gamma$ (for an optimal transport plan γ) and $\mu_t^n \rightharpoonup \mu_t$ (for the corresponding interpolation); by semicontinuity we have

$$||\mu_t||_p \leq \liminf_n ||\mu_t^n||_p \leq (1-t)^{-d/p'} ||\mu_0||_p$$

and we deduce

$$||\sigma||_p \leq C \int_0^1 ||\mu_t||_p dt \leq C ||\mu_0||_p \int_0^1 (1-t)^{-d/p'} dt.$$

The last integral is finite whenever $p' > d$, i.e. $p < d' = d/(d-1)$.

The second part of the statement (the case $p \geq d'$) is straightforward once one considers that any density in L^p also belongs to any L^q space for $q < p$. □

EXAMPLE

What we just saw in the previous theorems is that the measures μ_t inherit some regularity (absolute continuity or L^p summability) from μ_0 exactly as it happens for homotheties of ratio $1-t$. This regularity degenerates as $t \to 1$, but we saw two cases where this degeneracy produced no problem: for proving absolute continuity, where the separate absolute continuous behavior of almost all the μ_t was sufficient, and for L^p estimates, provided the degeneracy stays integrable.

It is natural to try to exploit another strategy: suppose both μ_0 and μ_1 share some regularity assumption (e.g., they belong to L^p). Then we can give estimate on μ_t for $t \leq 1/2$ starting from μ_0 and for $t \geq 1/2$ starting from μ_1. In this way we have no degeneracy!

This strategy works quite well, but it has an extra difficulty: in our previous estimates we didn't know a priori that μ_t shared the same behavior of piecewise homotheties of μ_0, we got it as a limit from discrete approximations. And, when we pass to the limit, we do not know which optimal transport γ will be selected as a limit of the optimal plans γ^n. This was not important in the previous section, since any optimal γ induces the same transport density σ. Yet, here we would like to glue together estimates on μ_t for $t \leq 1/2$ which have been obtained by approximating μ_1, and estimates on μ_t for $t \geq 1/2$ which come from the approximation of μ_0. Should the two approximations converge to two different transport plans, we could not put together the two estimates and deduce anything on σ.

Hence, the main technical issue which we need to consider is proving that one particular optimal transport plan, i.e. the one which is monotone on transport rays, will be approximable in both directions. Lemma 3.1.12 exactly does the job (and, indeed, it was proven in [93] exactly for this purpose). Yet, the transport plans γ_ε we build in the approximation are not optimal for the cost $\int |x-y| d\gamma$ but for some costs $\int (|x-y| + \varepsilon |x-y|^2) d\gamma$. We need to do this in order to force the selected limit optimal transport to be the monotone one (through a secondary variational problem, say). Anyway, this will not be an issue since these approximating optimal transport will share the same geometric properties that will imply disjointness for the sets $\Omega_i(t)$ will allow for density estimates.

The first tool we need is a uniform L^p estimates of the measures μ_t in terms of the norm of μ_0, when μ_t is an interpolation from μ_0 to μ_1 corresponding to a transport plan γ which is optimal for another cost, different from $|x-y|$. In this case we do not have any transport ray argument, but the result is somehow even stronger under strict convexity assumptions.

Even if not precisely stated, the reader will be easily be able to check that all the results of this section stay true for $p = +\infty$ as well.

Lemma 4.3.3. *Let γ be an optimal transport plan between μ_0 and an atomic measure μ_1 for a transport cost $c(x,y) = \phi(y-x)$ where $\phi : \mathbb{R}^d \to \mathbb{R}$ is a strictly convex function. Set as usual $\mu_t = (\pi_t)_\#\gamma$. Then we have $||\mu_t||_p \leq (1-t)^{-d/p'}||\mu_0||_p$.*

Proof. The result is exactly the same as in Theorem 4.3.2, where the key tool is the fact that μ_t coincides on every set $\Omega_i(t)$ with an homothety of μ_0. The only fact that must be checked again is the disjointness of the sets $\Omega_i(t)$.

To do so, take a point $x \in \Omega_i(t)\cap\Omega_j(t)$. Hence there exist x_i, x_j belonging to Ω_i and Ω_j, respectively, so that $x = (1-t)x_i + ty_i = (1-t)x_j + ty_j$, being y_i and y_j atoms of μ_1. Set $a = y_i - x_i$ and $b = y_j - x_j$.

The $c-$cyclical monotonicity of the support of the optimal γ implies

$$\phi(a) + \phi(b) \leq \phi(y_j - x_i) + \phi(y_i - x_j) = \phi(tb + (1-t)a) + \phi(ta + (1-t)b).$$

Yet, if $y_j \neq y_i$ we have $a \neq b$, and strict convexity implies

$$\phi(tb+(1-t)a)+\phi(ta+(1-t)b) < t\phi(b)+(1-t)\phi(a)+t\phi(a)+(1-t)\phi(b) = \phi(a)+\phi(b),$$

which is a contradiction. Hence the sets $\Omega_i(t)$ are disjoint and this implies the bound on μ_t. □

Remark 14. Disjointness of the sets $\Omega_i(t)$ is easier in this strictly convex setting. If the cost is $|x-y|$ this is no more true, but it is anyway true that the two vector a and b should be parallel, i.e. all the points should be aligned, as we pointed out in Theorem 4.3.1. If μ does not give mass to lines, than the sets are essentially disjoint. Otherwise one can say that they are truly disjoint if one only looks at the optimal transport which is monotone on transport rays.

Theorem 4.3.4. *Suppose that μ_0 and μ_1 are probability measures on Ω, both belonging to $L^p(\Omega)$, and σ the unique transport density associated to the transport of μ_0 onto μ_1. Then σ belongs to $L^p(\Omega)$ as well.*

Proof. Let us consider the optimal transport plan $\bar{\gamma}$ from μ_0 to μ_1 defined by (1.2). We know that this transport plan may be approximated by plans γ_ε which are optimal for the cost $|x-y| + \varepsilon|x-y|^2$ from μ_0 to some discrete atomic measures ν_ε. The corresponding interpolation measures $\mu_t(\varepsilon)$ satisfy the L^p estimate from Lemma 4.3.3 and, at the limit, we have

$$||\mu_t||_p \leq \liminf_{\varepsilon\to 0} ||\mu_t(\varepsilon)||_p \leq (1-t)^{-d/p'}||\mu_0||_p.$$

The same estimate may be performed from the other direction, since the same transport plan $\bar{\gamma}$ may be approximated by optimal plans for the cost $|x-y| + \varepsilon|x-y|^2$ from atomic measures to μ_1. Putting together the two estimates we have

$$||\mu_t||_p \leq \min\left\{(1-t)^{-d/p'}||\mu_0||_p,\, t^{-d/p'}||\mu_1||_p\right\} \leq 2^{d/p'} \max\left\{||\mu_0||_p, ||\mu_1||_p\right\}.$$

Integrating these L^p norms we get the bound on $||\sigma||_p$. □

EXAMPLE

Theorem 4.3.5. *Suppose $\mu_0 \in L^p(\Omega)$ and $\mu_1 \in L^q(\Omega)$. For notational simplicity take $p > q$. Then, if $p < d/(d-1)$, the transport density σ belongs to L^p and, if $p \geq d/(d-1)$, it belongs to $L^r(\Omega)$ for all the exponents r satisfying*

$$r < r(p,q,d) := \frac{dq(p-1)}{d(p-1)-(p-q)}.$$

Proof. The first part of the statement (the case $p < d/(d-1)$ is a consequence of Theorem 4.3.2. For the second one, using exactly the same argument as before (Theorem 4.3.4) we get

$$||\mu_t||_p \leq (1-t)^{-d/p'}||\mu_0||_p; \quad ||\mu_t||_q \leq t^{-d/q'}||\mu_1||_q.$$

We then apply standard Hölder inequality to derive the usual interpolation estimate for any exponent $q < r < p$:

$$||f||_r \leq ||f||_p^{\alpha}||f||_q^{1-\alpha} \quad \text{with } \alpha = \frac{p(r-q)}{r(p-q)}, \text{ and } 1-\alpha = \frac{q(p-r)}{r(p-q)}.$$

This implies

$$||\mu_t||_r \leq C||\mu_t||_p \leq C||\mu_0||_p \text{ for } t < \frac{1}{2}; \quad ||\mu_t||_r \leq C(1-t)^{-\alpha d/p'}||\mu_0||_p^{\alpha}||\mu_1||_p^{1-\alpha} \text{ for } t > \frac{1}{2}.$$

Then, take $r < r(p,q,d)$, so that $\alpha d/p' < 1$ is ensured and hence the L^r norm is integrable, thus giving a bound on $||\sigma||_r$. □

Remark 15. We do not know whether this exponent $r(p,q,d)$ is sharp or not and whether σ belongs or not to $L^{r(p,q,d)}$.

On the contrary, Example 4.15 in [58] shows the sharpness of the bound on p that we set in Theorem 4.3.2.

4.4 Discussion

4.4.1 Congested transport

As we saw in Section 4.2, Beckmann's problem can admit an easy variant if we prescribe a positive function $k : \Omega \to \mathbb{R}_+$, where $k(x)$ stands for the local cost at x per unit length of a path passing through x. This models the possibility that the metric is non-homogeneous, due to geographical obstacles given a priori. Yet, it happens in many situation, in particular in urban traffic as everybody knows, that this metric k is indeed non-homogeneous, but is not given a priori: it depends on the traffic, i.e. it depends on the choice of all the commuters. In Beckmann's language, we must look for a vector field v optimizing a transport cost depending on v itself!

The easiest modelization, chosen by Beckmann [11] and later in [46] is to consider the same framework as (PB) but supposing that $k(x) = g(|v(x|)$ is a function of the modulus of the vector field v. This is quite formal for the moment (for instance it is not meaningful if v is a measure, but we will not set this problem in the class of measures, indeed). In this case we would like to solve

$$\min \int \mathcal{H}(v(x))dx \quad : \nabla \cdot v = \mu - \nu, \tag{4.1}$$

where $\mathcal{H}(v) = H(|v|)$ and $H(t) = g(t)t$. Notice that if H is superlinear (if $g(t) \to \infty$ as $t/to\infty$, i.e. if the congestion effect becomes bigger and bigger when the traffic increases) this problem is well posed in the class of vector fields $v \in L^1$ (or of absolutely continuous vector measures). For instance, if $g(t) = t$, which is the easiest case one can imagine, we must minimize the L^2 norm under divergence constraints:

$$\min \int |v(x)|^2 dx \quad : v \in L^2(\Omega; \mathbb{R}^d),\ \nabla \cdot v = \mu - \nu.$$

This problem is easily solvable since one can see that the optimal v must be a gradient (we will develop this computation in a more general framework below), and setting $v = \nabla u$ one gets $\Delta u = \mu - \nu$. This is complemented with Neumann boundary conditions and allow to find u, and then v.

We want now to discuss the meaning and pertinence of this model, keeping into account the following natural questions:

- is this the good modelization, or the coefficient k should rather depend on other traffic quantities, in particular a traffic intensity like i_Q? (notice that v can have cancellations);

- what is the connection with equilibrium issues? in traffic congestion, typically every agent decides alone which path to choose, and the final traffic intensity is rather the output of a collection of individual choices, rather than the result of a global optimization made by a single planner;
- is the example $g(t) = t$ a good choice in the modelization ? this implies $g(0) = 0$, i.e. no cost where there is not traffic, but we know that we cannot move at infinite speed even if there is no traffic.

To start our analysis we would like to present first an equilibrium model developed by Wardrop, [101], on a discrete network.

Traffic equilibria on a finite network The main data of the model are a finite oriented connected graph $G = (N, E)$ modeling the network, and edge travel times functions $g_e : w \in \mathbb{R}_+ \mapsto g_e(w)$ giving, for each edge $e \in E$, the travel time on arc e when the flow on this edge is w. The functions g_e are all nonnegative, continuous, nondecreasing and they are meant to capture the congestion effects (which may be different on the different edges, since 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: $\gamma_{x,y}$ 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 ω and we will use the notation $e \in \omega$ to indicate that the path ω uses the edge e.

The unknown of the problem is the flow configuration. The edge flows are denoted by $i = (i_e)_{e\in E}$ and the path flows are denoted by $q = (q_\omega)_{\omega\in C}$: this means that i_e is the total flow on edge e and q_ω is the mass traveling on the path ω. Of course the i_e's and q_ω's are nonnegative and constrained by the mass conservation conditions:

$$\gamma_{x,y} = \sum_{\omega\in C_{x,y}} q_\omega, \quad \forall (x, y) \in N^2 \tag{4.2}$$

and

$$i_e = \sum_{\omega\in C\,:\, e\in\omega} q_\omega, \quad \forall e \in E, \tag{4.3}$$

which means that i is a function of q. Given the edge flows $i = (i_e)_{e\in E}$, the total travel-time of the path $\omega \in C$ is

$$T_i(\omega) = \sum_{e \in \omega} g_e(i_e). \tag{4.4}$$

In [101], 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 (4.4). This leads to

Definition 12. A Wardrop equlibrium is a flow configuration $i = (i_e)_{e \in E}$, $q = (q_\omega)_{\omega \in C}$ (all nonnegative of course), satisfying the mass conservation constraints (4.2) and (4.3), such that, in addition, for every $(x, y) \in N^2$ and every $\omega \in C_{x,y}$, if $q_\omega > 0$ then

$$T_i(\omega) = \min_{\omega' \in C_{x,y}} T_i(\omega').$$

A few years after Wardrop introduced his equilibrium concept, Beckmann, McGuire and Winsten [12] realized that Wardrop equilibria can be characterized by the following variational principle:

Theorem 4.4.1. *The flow configuration $i = (i_e)_{e \in E}$, $q = (q_\omega)_{\omega \in C}$ is a Wardrop equilibrium if and only if it solves the convex minimization problem*

$$\inf_{(i,q)} \sum_{e \in E} H_e(i_e) \text{ s.t. nonnegativity and } (4.2)\ (4.3) \tag{4.5}$$

where, for each e, we take H_e to be the primitive of g_e.

Proof. Assume that $q = (q_\omega)_{\omega \in C}$ (with associated edge flows $(i_e)_{e \in E}$) is optimal for (4.5) then for every admissible $\eta = (\eta_\omega)_{\omega \in C}$ with associated (through (4.3)) edge-flows $(u_e)_{e \in E}$, one has

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

so that

$$\sum_{\omega \in C} q_\omega T_i(\omega) \leq \sum_{\omega \in C} \eta_\omega T_i(\omega)$$

minimizing the right-hand side thus yields

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

which exactly says that (q, i) is a Wardrop equilibrium. To prove the converse, it is enough to see that problem (4.5) 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).

Remark 16. It would be very tempting to deduce from theorem 4.4.1 that equilibria are efficient since they are minimizers of (4.5). One has to be cautious with this quick interpretation since the quantity $\sum_{e\in E} H_e(i_e)$ does not represent the natural total social cost measured by the total time lost in commuting which reads as

$$\sum_{e\in E} i_e g_e(i_e). \tag{4.6}$$

The efficient transport patterns are minimizers of (4.6) and thus are different from equilibria in general. Efficient and equilibria configurations coincide in the special case of power functions where $g_e(w) = a_e w^\alpha$, but this case is not realistic since it implies that traveling times vanish if there is no traffic... Moreover, a famous counter-example due to Braess shows that it may be the case that adding an extra road on which the travelling time is aways zero leads to an equilibrium where the total commuting time is increased! This illustrates the striking difference between efficiency and equilibrium, a topic which is very well-documented in the finite-dimensional network setting where it is frequently associated to the literature on the so-called *price of anarchy* (see [**?**]).

Remark 17. In the problem presented in this paragraph, 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

$$\sum \gamma_{x,y} d_i(x,y) \text{ with } d_i(x,y) := \min_{\omega\in C_{x,y}} T_i(\omega).$$

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.

Optimization and equlibrium in a continuous framework We want now to generalize the previous analysis to a continuous framework. In the continuous setting, 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 in a similar way as one defines transport density in the Monge's problem (this is the continuous analogue of the arc flows $(i_e)_e$ of the previous paragraph). The last and main idea will be in modelling the congestion effect through a metric that is monotone increasing in the traffic intensity (the analogue of $g_e(i_e)$).

We will deliberately avoid to enter into technicalities so the following description will be pretty informal (see [45] 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 two probability measures μ and ν (distribution of sources and destinations) on Ω 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 $\Omega \times \Omega$) in the short-term case,
- or more generally a convex and closed subset $\Gamma \subset \Pi(\mu, \nu)$ and we accept any $\gamma \in \Gamma$ (this is just a common mathematical framework for the two previous cases, where we can take $\Gamma = \{\gamma\}$ or $\Gamma = \Pi(\mu, \nu)$.

We will use the notations of Section 4.2.3, and use probability measures Q on $C := \mathrm{Lip}([0,1], \Omega)$, compatible with mass conservation, i.e. such that

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

We shall denote by $\mathcal{Q}(\Gamma)$ the set of admissible transport patterns. We are interested in finding an equilibrium i.e. a $Q \in \mathcal{Q}(\Gamma)$ that is supported on geodesics for a metric ξ_Q depending on Q itself (congestion).

The intensity of traffic associated to $Q \in \mathcal{Q}(\Gamma)$ is by definition the measure i_Q defined in Section 4.2.3.

The congestion effect is then captured by the *metric* associated to Q: suppose $i_Q \ll \mathcal{L}^2$ and set

$$\xi_Q(x) := g(x, i_Q(x))$$

for a given increasing function $g(x,.) : \mathbb{R}_+ \to \mathbb{R}_+$. The fact that there exists at least one $Q \in \mathcal{Q}(\Gamma)$ such that $i_Q \ll \mathcal{L}^2$ is not always true and depends on Γ but, for instance, it is true when $\Gamma = \Pi(\mu, \nu)$ and μ and ν are such that the transport density is absolutely continuous. Notice also that, for $\Gamma = \{\gamma\}$ (which is the most restrictive case) and $\mu, \nu \in L^\infty$, considerations from incompressible fluid mechanics in [36] allow to build a Q such that $i_Q \in L^\infty$. 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 $\omega \in C_{x,y}$ (i.e. an absolutely continuous curve ω such that $\omega(0) = x$ and $\omega(1) = y$) spends time

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

She will then try to minimize this time i.e. to achieve the corresponding geodesic distance

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

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

We can define

Definition 13. A Wardrop equilbrium is a $Q \in \mathcal{Q}(\Gamma)$ such that

$$Q(\{\omega \;:\; L_{\xi_Q}(\omega) = c_{\xi_Q}(\omega(0), \omega(1)) = 1. \tag{4.7}$$

Existence, and even well-posedness (what does it mean $L_\xi(\omega)$ if ξ is only measurable and ω is a Lipschitz curve?) of these equilibria are not straightforward. Again, we will characterize equilibria as solutions of a given minimal traffic problem.

Let us consider the (convex) variational problem

$$\inf_{Q \in \mathcal{Q}(\Gamma)} \int_\Omega H(x, i_Q(x))dx \tag{4.8}$$

where $H'(x,.) = g(x,.)$, $H(x,0) = 0$. 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 [45] can be summarized by:

Theorem 4.4.2. *Problem* (4.8) *admits at least one minimizer. Moreover* $\overline{Q} \in \mathcal{Q}(\Gamma)$ *solves* (4.8) *if and only if it is a Wardrop equilibrium and* $\gamma_Q := (e_0, e_1)_{\#}Q$ *solves the optimization problem*

$$\min \int_{\Omega\times\Omega} c_{\xi_Q}(x,y)\, d\gamma(x,y) \;:\; \gamma \in \Gamma.$$

In particular, if $\Gamma = \{\gamma\}$ *this last condition does not play any role (there is only one competitor) and we show existence of a Wardrop equilibrium corresponding to any given transport plan* γ. *If, on the contrary,* $\Gamma = \Pi(\mu, \nu)$, *then the second condition means that* γ *solves a Monge-Kantorovich problem for a distance cost depending on* Q *itself, which is a new equilibrium condition.*

The full proof is quite involved since it requires to take care of some regularity issues in details. In particular, the use of the weighted length functional $L_{\overline{\xi}}$ and thus also the geodesic distance $c_{\overline{\xi}}$ require some attention since defining these quantities actually makes sense only if $\overline{\xi}$ is continuous or at least l.s.c.. In [45] a possible construction when $\overline{\xi}$ is just an L^q function is given. Let us also mention that recent regularity results (see below) actually prove that $\overline{\xi}$ is in fact a continuous function, under reasonable assumptions on the data.

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 (**??**) may be difficult to solve since they involve measures on sets of curves that is two layers of infinite dimensions! We will not deal here with the numerical strategies, bases on convex optimization duality, and on the so-called Fast Marching Method to compute c_ξ for given ξ (and later to compute variations of c_ξ when ξ varies), and we refer to [17, 18]. These numerical methods are quite efficient and generalize what already done on finite networks, and are better suited for the short-term case.

On the contrary, in the next paragraph we develop an interesting feature of the long-term problem.

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.

As we already did in Section 4.2.3, for any $Q \in \mathcal{Q}(\Gamma)$ we can take the vector-field θ_Q.

If we consider the scalar problem (4.8), it is easy to see that its value is larger than that of the minimal flow problem à la Beckmann:

$$\min_{\sigma\,:\,\nabla\cdot\sigma=\mu-\nu} \int_\Omega \mathcal{H}(\sigma(x))dx \tag{4.9}$$

where $\mathcal{H}(\sigma) = H(|\sigma|)$ and H is taken independent of x only for simplicity. The inequality is justified by two facts: minimizing over all vector fields v with prescribed divergence gives a smaller result than minimizing over the vector fields θ_Q, and then we use $|\theta_Q| \leq i_Q$ and the fact that H is increasing.

We would like to understand if the two problems are equivalent.

Proposition 4.2.4 does the job: if we take a minimizer v for this minimal flow problem, then we are able to build a measure Q and, as we did in Theorem 4.2.5, the optimality of v gives $v = \theta_Q$ and $|\theta_Q| = i_Q$, thus proving that the minimal values are the same and that we can build a minimizer v from a minimizer Q (just take $v = \theta_Q$) and conversely a minimizer Q from v (use Proposition 4.2.4).

The connection between the two problems would be stronger should the Q that we build from v be somehow canonical and unique, instead of being obtained through an approximation and compactness argument. This means that we would like to have regularity results on the minimizer v, so that we can directly apply to it the construction by Dacorogna and Moser, without approximating and extracting a subsequence. Notice that, if H is strictly convex, the minimizer v is unique.

To be able to solve the Cauchy problem

$$\begin{cases} y_x'(t) = \tilde{v}(t, y_x(t)) \\ y_x(0) = x \end{cases},$$

with

$$\tilde{v}(t,x) = \frac{v(x)}{f_t(x)} \quad \text{where } f_t = (1-t)\mu + t\nu$$

one would need $\tilde{v}$ to be regular enough (say, Lipschitz continuous). Obviously, we can decide to add some assumptions on μ and ν, which will be supposed to be absolutely continuous with regular densities (at least Lipschitz continuous and bounded from below).

However, one needs to prove regularity for the optimal v, and for this one needs to look at the optimality conditions satisfied by v as a minimizer of (4.9). **PROOF OPTIMALITY** By duality, the solution of (4.9) is

$v = \nabla \mathcal{H}^*(\nabla u)$ where $\mathcal{H}^*$ is the Legendre transform of $\mathcal{H}$ and u solves the PDE:

$$\begin{cases} \nabla \cdot (\nabla \mathcal{H}^*(\nabla u)) & = & \mu_0 - \mu_1, & \text{in } \Omega, \\ \nabla \mathcal{H}^*(\nabla u) \cdot n_\Omega & = & 0, & \text{on } \partial\Omega, \end{cases} \tag{4.10}$$

This equation turns out to be a standard Laplace equation if $\mathcal{H}$ is quadratic, or it becomes a $p-$Laplace equation for other power functions. In these cases, regularity results are well-known, under regularity assumptions on μ_0 and μ_1. Yet, 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! This means 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 (4.10) very degenerate, even worse than the $p-$Laplacian. For instance, a reasonable model of congestion is $g(t) = 1 + t^{p-1}$ for $t \geq 0$, with $p > 1$, so that

$$\mathcal{H}(\sigma) = \frac{1}{p}|\sigma|^p + |\sigma|,\ \mathcal{H}^*(z) = \frac{1}{q}(|z| - 1)_+^q,\ \text{with } q = \frac{p}{p-1} \tag{4.11}$$

so that the optimal σ is

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

where u solves the very degenerate PDE:

$$\nabla \cdot \left(\Big(|\nabla u| - 1\Big)_+^{q-1} \frac{\nabla u}{|\nabla u|}\right) = \mu_0 - \mu_1, \tag{4.12}$$

with Neumann boundary condition

$$\Big(|\nabla u| - 1\Big)_+^{q-1} \frac{\nabla u}{|\nabla u|} \cdot n_\Omega = 0.$$

Note that there is no uniqueness for u but there is for v.

For this degenerate equation (more degenerate than the p-laplacian since the diffusion coefficient identically vanishes in the zone where $|\nabla u| \leq 1$), getting Lipschitz continuity on v is not reasonable. Yet, Sobolev regularity of v and Lipschitz regularity results for solutions of this PDE can be found in [32]. This enables one to build a flow *à la* DiPerna-Lions [62] and then to justify rigorously the construction above, even without a Cauchy-Lipschitz flow. Interestingly, recent continuity results are also available (see [97] in dimension 2, and then [53], with a different technique in arbitrary dimension), obtained as a consequence of a fine analysis of this degenerate elliptic PDE.

Besides the interest for this regularity result in itself, we also stress that continuity for v implies continuity for the optimal i_Q, and this exactly gives the regularity which is required in the proof of Theorem 4.4.2 (the main difficulty being defining $c_{\bar{\xi}}$ for a non-continuous $\bar{\xi}$, and this is the reason why our proof in Section 3 is only formal).

4.4.2 Branched transport

Opposite from what we saw in the previous section about congested transport, in many other practical issues we would like to look for a way of transporting the mass so that it moves as much jointly as possible, favoring particles to share the same displacement instead of spreading all around the domain and using as many different paths as possible. This comes from a very different modelization, which is more suitable for other purposes than studying traffic congestion: supose for instance that you have to build the network system to transport the mass; in this case you do not want to build infinitely many small roads, each one meant to transport a unique particle from its starting point to its destination, but you prefer to build one unique bigger road. This is usually due to "economy of scale" principles, something that we can experience everyday (exactly as it happens for traffic congestion, but on different phenomena): the idea is that buy, or building, something bigger will cost more, but proportionally less. In particular costs are supposed to be sub-additive (the cost of the sum of two objects must be less than the sum of the two costs), and in many cases in economy they have "decreasing marginal costs" (i.e. the cost for adding a unit to a given background quantity is a decreasing function of the background, which means that the cost is actually concave).

Notice that modeling this kind of effects require, either in Lagrangian or Eulerian language, to look at the paths actually followed by each particles, and it could not be done with the only use of a transport plan $\gamma \in \Pi(\mu, \nu)$. But, once we choose the good formulation via the tools developed in this chapter, we ca guess the shape of the optimal solution for this kind of problem: particles are collected at some points, move together as much as possible, and then branch towards their different destinations. This is why this class of problems is nowadays known as "branched transport".

As we did for congested transport, we start from the discrete framework to give a presentation of the problem and then move to the continuous models. Notice that also in this case the discrete framework is somehow classical in optimization and operational research, and the continuous one is much more recent. Anyway, it has been investigated by different schools