AN INTRODUCTION TO OPTIMAL TRANSPORT
AND WASSERSTEIN GRADIENT FLOWS

ALESSIO FIGALLI

ABSTRACT. These short notes summarize a series of lectures given by the author during the School "Optimal Transport on Quantum Structures", which took place on September 19th-23rd, 2022, at the Erdös Center - Alfréd Rényi Institute of Mathematics. The lectures aimed to introduce the classical optimal transport problem and the theory of Wasserstein gradient flows.

CONTENTS

1. The origin of optimal transport and its modern formulation

1.1. **Monge's formulation.** In his celebrated work in 1781, Gaspard Monge introduced the concept of transport maps starting from the following practical question: Assume one extracts soil from the ground to build fortifications. What is the cheapest possible way to transport the soil? To formulate this question rigorously, one needs to specify the transportation cost, namely how much one pays to move a unit of mass from a point x to a point y. In Monge's case, the ambient space was $\mathbb{R}^3$, and the cost was the Euclidean distance $c(x, y) = |x - y|$.

1.2. **Kantorovich's formulation.** After 150 years, in the 1940s, Leonid Kantorovich revisited Monge's problem from a different viewpoint. To explain this, consider:
- N bakeries located at positions $(x_i)_{i=1,\ldots,N}$;
- M coffee shops located at $(y_j)_{j=1,\ldots,M}$;
- the ith bakery produces an amount $\alpha_i \geq 0$ of bread;
- the jth coffee shop needs an amount $\beta_j \geq 0$.
Also, assume that supply is equal to demand, and (without loss of generality) normalize them to be equal to 1, i.e.,

$$\sum_{i=1}^{N} \alpha_i = \sum_{j=1}^{M} \beta_j = 1.$$

In Monge's formulation, the transport is deterministic: the mass located at x can be sent to a unique destination $T(x)$. Unfortunately this formulation is incompatible with the problem above, since one bakery may supply bread to multiple coffee shops, and one coffee shop may buy bread from multiple bakeries. For this reason Kantorovich introduced a new formulation: given $c(x_i, y_j)$ the cost to move one unit of mass from x_i to y_j, he looked for matrices $(\gamma_{ij})_{\substack{i=1,\ldots,N \\ j=1,\ldots,M}}$ such that

(i) $\gamma_{ij} \geq 0$ (the amount of bread going from x_i to y_j is a nonnegative quantity);

(ii) for all i, $\alpha_i = \sum_{j=1}^{M} \gamma_{ij}$ (the total amount of bread sent to the different coffee shops is equal to the production);

(iii) for all j, $\beta_j = \sum_{i=1}^{N} \gamma_{ij}$ (the total amount of bread bought from the different bakeries is equal to the demand);

(iv) γ_{ij} minimizes the cost $\sum_{i,j} \gamma_{ij} c(x_i, y_j)$ (the total transportation cost is minimized).

It is interesting to observe that, as functions of the matrix γ_{ij}:
- constraint (i) is convex;
- constraints (ii) and (iii) are linear;
- the objective function in (iv) is linear.
In other words, Kantorovich's formulation corresponds to minimizing a linear function with convex constraints.

1.3. **Notation.** In these notes we will use standard notations. For instance, when we write $\mathcal{P}(Z)$, we mean the space of (Borel) probability measures over sole locally compact separable complete metric space Z. Given a set E, we use $\mathbf{1}_E$ to denote its indicator functions. We refer the interested reader to [4] for more details and notation.

1.4. **Transport maps and couplings.**

Definition 1.1. *Given* $\mu \in \mathcal{P}(X)$ *and* $\nu \in \mathcal{P}(Y)$, *a measurable map* $T \colon X \to Y$ *is called a* transport map *from* μ *to* ν *if* $T_{\#}\mu = \nu$, *that is,*

(1.1) $$\nu(A) = \mu(T^{-1}(A)) \qquad \forall A \subset Y \ Borel.$$

Remark 1.2. *Condition* (1.1) *can also be rewritten in terms of test functions as follows (see for instance* [4, Corollary 1.2.6]*):*

$$T_{\#}\mu = \nu \qquad \Longleftrightarrow \qquad \int_Y \psi \, d\nu = \int_X \psi \circ T \, d\mu \quad \forall \, \psi : Y \to \mathbb{R} \text{ Borel and bounded.}$$

In these notes we will often make use of this fact.

Remark 1.3. *Given μ and ν, the set of transport maps from μ to ν may be empty. For instance, given $\mu = \delta_{x_0}$ with $x_0 \in X$ and a map $T : X \to Y$, then $T_{\#}\mu = \delta_{T(x_0)}$. Hence, unless ν is a Dirac delta, for any map T we have $T_{\#}\mu \neq \nu$ and the set $\{T : T_{\#}\mu = \nu\}$ is empty.*

Definition 1.4. *We call $\gamma \in \mathcal{P}(X \times Y)$ a* coupling *or* transport plan *between μ and ν if*

$$(\pi_X)_{\#}\gamma = \mu \quad \text{and} \quad (\pi_Y)_{\#}\gamma = \nu,$$

where $\pi_X(x, y) = x$ and $\pi_Y(x, y) = y$ for every $(x, y) \in X \times Y$, namely

$$\mu(A) = \gamma(A \times Y), \qquad \nu(B) = \gamma(X \times B), \qquad \forall \, A \subset X, \, B \subset Y.$$

We denote by $\Gamma(\mu, \nu)$ the set of couplings of μ and ν.

Remark 1.5. *Given μ and ν, the set $\Gamma(\mu, \nu)$ is always nonempty. Indeed, the product measure $\gamma = \mu \otimes \nu$ is a coupling (see* [4, Remark 1.4.4]*).*

Remark 1.6. *Let $T : X \to Y$ satisfy $T_{\#}\mu = \nu$. Consider the map $\mathrm{id} \times T : X \to X \times Y$, i.e., $x \mapsto (x, T(x))$, and define*

$$\gamma_T := (\mathrm{id} \times T)_{\#}\mu \in \mathcal{P}(X \times Y).$$

We claim that $\gamma_T \in \Gamma(\mu, \nu)$. Indeed,

$$(\pi_X)_{\#}\gamma_T = (\pi_X)_{\#}(\mathrm{id} \times T)_{\#}\mu = (\pi_X \circ (\mathrm{id} \times T))_{\#}\mu = \mathrm{id}_{\#}\mu = \mu,$$
$$(\pi_Y)_{\#}\gamma_T = (\pi_Y)_{\#}(\mathrm{id} \times T)_{\#}\mu = (\pi_Y \circ (\mathrm{id} \times T))_{\#}\mu = T_{\#}\mu = \nu.$$

This proves that any transport map T induces a coupling γ_T.

Vice versa, if $\gamma \in \Gamma(\mu, \nu)$ and $\gamma = (\mathrm{id} \times T)_{\#}\mu$ for some map $T : X \to Y$, then $T_{\#}\mu = \nu$. In other words, if a coupling is induced by a map, then this map is a transport.

1.5. **Monge and Kantorovich's problems.** Fix $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, and $c : X \times Y \to [0, +\infty]$ lower semicontinuous. The Monge and Kantorovich's problems can be stated as follows (recall Definition 1.4):

$$(1.2) \qquad C_M(\mu, \nu) := \inf \left\{ \int_X c(x, T(x)) \, d\mu(x) \, \middle| \, T_{\#}\mu = \nu \right\},$$

$$(1.3) \qquad C_K(\mu, \nu) := \inf \left\{ \int_{X \times Y} c(x, y) \, d\gamma(x, y) \, \middle| \, \gamma \in \Gamma(\mu, \nu) \right\}.$$

In other words, Monge's problem (1.2) consists in minimizing the transportation cost among all transport maps, while Kantorovich's problem (1.3) consists in minimizing the transportation cost among all couplings.

Remark 1.7. *Recalling Remark 1.6, we also notice that*

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

In other words, any transport map T induces a coupling γ_T with the same cost. In particular,

$$C_M(\mu, \nu) \geq C_K(\mu, \nu).$$

Thanks to this inequality we deduce that, if $\gamma \in \Gamma(\mu, \nu)$ is optimal in (1.3) and $\gamma = (\mathrm{id} \times T)_{\#}\mu$ for some map $T : X \to Y$, then T solves (1.2).

As we shall see in the next section, while Kantorovich's problem can be solved under minimal assumptions on the cost, the solution of Monge's problem is considerably more complicated.

2. Solving Monge and Kantorovich's problems

Motivated by Remark 1.7, a natural strategy to solve Monge and Kantorovich's problems is the following:
(1) show that a solution of Kantorovich's problem exists under minimal assumptions;
(2) analyze the structure of an optimal coupling, and try to understand under which assumptions this coupling is induced by a map.

2.1. **Existence of optimal couplings.** As mentioned above, optimal couplings exist under minimal assumptions on the cost. We refer the reader to [4, Theorem 2.3.2] for a detailed proof of the following:

Theorem 2.1. *Let $c\colon X \times Y \to [0, +\infty)$ be lower semicontinuous, $\mu \in \mathcal{P}(X)$, and $\nu \in \mathcal{P}(Y)$. Then there exists a coupling $\bar\gamma \in \Gamma(\mu, \nu)$ that is a minimizer for* (1.3).

Sketch of the proof. Let $(\gamma_k)_{k\in\mathbb{N}} \subset \Gamma(\mu, \nu)$ be a minimizing sequence, namely

$$\int_{X \times Y} c \, d\gamma_k \to \alpha := \inf_{\gamma \in \Gamma(\mu,\nu)} \int_{X \times Y} c \, d\gamma \qquad \text{as } k \to \infty.$$

Thanks to the marginal constraints on the measures γ_k, one can prove that there exists a subsequence $(\gamma_{k_j})_{j\in\mathbb{N}}$ such that $\gamma_{k_j} \rightharpoonup \bar\gamma \in \Gamma(\mu, \nu)$. Also, thanks to the nonnegativity and lower semicontinuity of c, one can prove that

$$\alpha = \liminf_{j\to\infty} \int_{X \times Y} c \, d\gamma_{k_j} \geq \int_{X \times Y} c \, d\bar\gamma.$$

Since the opposite inequality $\int_{X \times Y} c \, d\bar\gamma \geq \alpha$ holds (by definition of α), this proves that $\bar\gamma$ is a minimizer. □

Remark 2.2. *In general, the optimal coupling is not unique. To see this, let $X = Y = \mathbb{R}^2$, $c(x, y) = |x - y|^2$, consider the points in $\mathbb{R}^2$ given by*

$$x_1 := (0, 0), \quad x_2 := (1, 1), \quad y_1 := (1, 0), \quad y_2 := (0, 1),$$

and define the measures

$$\mu = \tfrac{1}{2}\delta_{x_1} + \tfrac{1}{2}\delta_{x_2}, \quad \nu = \tfrac{1}{2}\delta_{y_1} + \tfrac{1}{2}\delta_{y_2}.$$

In this case, the set of all couplings $\Gamma(\mu, \nu)$ is given by

$$\gamma_\alpha = \alpha\delta_{(x_1,y_1)} + \left(\tfrac{1}{2} - \alpha\right)\delta_{(x_1,y_2)} + \left(\tfrac{1}{2} - \alpha\right)\delta_{(x_2,y_1)} + \alpha\delta_{(x_2,y_2)}, \quad \alpha \in \left[0, \tfrac{1}{2}\right],$$

and one can check that all these couplings have the same cost, so they are all optimal.

2.2. **The structure of optimal couplings: Kantorovich duality.** To study the structure of optimal couplings, it is useful to relate Kantorovich's problem to a dual problem. Formally, the

argument is based on general abstract results in convex analysis and goes as follows:

$$\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{X \times Y} c(x,y)\,d\gamma \overset{(i)}{=} \inf_{\gamma \geq 0} \sup_{\varphi,\psi} \left\{ \int_{X \times Y} c(x,y)\,d\gamma + \overbrace{\left(\int_{X \times Y} \varphi(x)\,d\gamma - \int_X \varphi(x)\,d\mu \right)}^{\text{Lagrange multiplier for } (\pi_X)_{\#}\gamma = \mu} \right.$$

$$\left. + \overbrace{\left(\int_{X \times Y} \psi(y)\,d\gamma - \int_Y \psi(y)\,d\nu \right)}^{\text{Lagrange multiplier for } (\pi_Y)_{\#}\gamma = \nu} \right\}$$

$$\overset{(ii)}{=} \inf_{\gamma \geq 0} \sup_{\varphi,\psi} \left\{ \int_X -\varphi\,d\mu + \int_Y -\psi\,d\nu + \int_{X \times Y} \big(c(x,y) + \varphi(x) + \psi(y)\big)d\gamma \right\}$$

$$\overset{(iii)}{=} \sup_{\varphi,\psi} \inf_{\gamma \geq 0} \left\{ \int_X -\varphi\,d\mu + \int_Y -\psi\,d\nu + \int_{X \times Y} \big(c(x,y) + \varphi(x) + \psi(y)\big)d\gamma \right\}$$

$$\overset{(iv)}{=} \sup_{\varphi,\psi} \left\{ \left(\int_X -\varphi\,d\mu + \int_Y -\psi\,d\nu \right) + \inf_{\gamma \geq 0} \int_{X \times Y} \big(c(x,y) + \varphi(x) + \psi(y)\big)d\gamma \right\}$$

$$\overset{(v)}{=} \sup_{\varphi(x)+\psi(y)+c(x,y) \geq 0} \int_X -\varphi\,d\mu + \int_Y -\psi\,d\nu,$$

where the equalities are justified as follows:

(i) While we keep the sign constraint $\gamma \geq 0$, we remove the coupling constraints on γ.[1] These constraints are now "hidden" in the Lagrange multipliers. Indeed, unless $(\pi_X)_{\#}\gamma \neq \mu$ (resp. if $(\pi_Y)_{\#}\gamma \neq \nu$), the supremum over φ (resp. over ψ) is $+\infty$.

(ii) We simply rearranged the terms.

(iii) We used [9, Theorem 1.9] to exchange inf and sup.

(iv) We note that the infimum over γ only affects the last integral.

(v) We have the following two possible situations:

• If $c(x,y) + \varphi(x) + \psi(y) \geq 0$ for all (x,y), then

$$\inf_{\gamma \geq 0} \int_{X \times Y} \big(c(x,y) + \varphi(x) + \psi(y)\big)d\gamma = 0.$$

Indeed, the integrand is nonnegative and we can choose $\gamma \equiv 0$ to deduce that the infimum is zero.

• If there exists $(\bar{x}, \bar{y})$ such that $c(\bar{x}, \bar{y}) + \varphi(\bar{x}) + \psi(\bar{x}) < 0$, then we can choose $\gamma = M\delta_{(\bar{x},\bar{y})}$ to find that

$$\inf_{\gamma \geq 0} \int_{X \times Y} \big(c(x,y) + \varphi(x) + \psi(y)\big)d\gamma \leq M\big(c(\bar{x}, \bar{y}) + \varphi(\bar{x}) + \psi(\bar{x})\big).$$

Hence, letting $M \to +\infty$, we conclude that in this case the infimum is $-\infty$.

---

[1] Keeping the sign constraint $\gamma \geq 0$ is convenient because of the following simple remark: if $\gamma \geq 0$ has one of the two marginals equal to μ or ν, then it is a probability measure. Indeed, if for instance $(\pi_X)_{\#}\gamma = \mu$, then

$$\gamma(X \times Y) = \mu(X) = 1.$$

In other words, in the space of nonnegative measures, the condition of being probability measures is automatically implied by the marginal constraints.

The argument above shows that the infimum in Kantorovich's problem is equal to the supremum over a dual problem. Actually, again under minimal assumptions on the cost, it is possible to prove that the infimum and the supremum are respectively a minimum and a maximum, see [4, Theorem 2.6.5]. Hence, the following general Kantorovich duality result holds:

Theorem 2.3. *Let $c \in C^0(X \times Y)$ be bounded from below, and assume that $\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{X \times Y} c \, d\gamma < +\infty$. Then*

$$\min_{\gamma \in \Gamma(\mu,\nu)} \int_{X \times Y} c \, d\gamma = \max_{\varphi(x)+\psi(y)+c(x,y) \geq 0} \int_X -\varphi \, d\mu + \int_Y -\psi \, d\nu.$$

2.3. **From Kantorovich duality to Brenier's Theorem.** As a consequence of Theorem 2.3, we can now study the structure of optimal plans.

Let $\bar{\gamma}$ be an optimal plan, and let $(\bar{\varphi}, \bar{\psi})$ be a couple of functions for which equality holds in Theorem 2.3, namely

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

where the second equality follows from the marginal constraints on $\bar{\gamma}$. Hence, this proves that

$$\int_{X \times Y} \left[c(x,y) + \bar{\varphi}(x) + \bar{\psi}(y) \right] d\bar{\gamma} = 0.$$

Since the function $\Psi(x,y) := c(x,y) + \bar{\varphi}(x) + \bar{\psi}(y)$ is nonnegative everywhere (by the assumption on $(\bar{\varphi}, \bar{\psi})$ in Theorem 2.3), we deduce that $\Psi = 0$ $\bar{\gamma}$-a.e. In other words, Ψ attains its minimum at $\bar{\gamma}$-a.e. point. In particular, if we knew that Ψ is differentiable in x at $\bar{\gamma}$-a.e. point, then we would deduce that

$$(2.1) \qquad 0 = \nabla_x \Psi(x,y) = \nabla_x c(x,y) + \nabla\varphi(x) \qquad \text{for } \bar{\gamma}\text{-a.e. } (x,y).$$

To understand what this relation entails, we consider the classical "quadratic case".

Let $X = Y = \mathbb{R}^d$ and $c(x,y) = \frac{|x-y|^2}{2}$. Then, in this case, (2.1) becomes

$$0 = x - y + \nabla\bar{\varphi}(x) \qquad \text{for } \bar{\gamma}\text{-a.e. } (x,y),$$

or equivalently

$$y = x + \nabla\bar{\varphi}(x) = \nabla\left(\frac{|x|^2}{2} + \bar{\varphi} \right)(x) \qquad \text{for } \bar{\gamma}\text{-a.e. } (x,y).$$

In other words, if $\bar{\varphi}$ is differentiable μ-a.e.[2] then we deduce that

$$y = T(x), \quad T = \nabla\left(\frac{|x|^2}{2} + \bar{\varphi} \right) \qquad \text{for } \bar{\gamma}\text{-a.e. } (x,y),$$

namely $\bar{\gamma}$ is contained inside the graph of T. By Remark 1.7, this map T would then be a solution to Monge's problem.

Making this argument rigorous leads to the proof of the following celebrated theorem of Brenier [3] (see [4, Theorem 2.5.10] for a detailed proof):

Theorem 2.4. *Let $X = Y = \mathbb{R}^d$ and $c(x,y) = \frac{|x-y|^2}{2}$. Suppose that*

$$\int_{\mathbb{R}^d} |x|^2 \, d\mu + \int_{\mathbb{R}^d} |y|^2 \, d\nu < +\infty$$

---

[2]Note that, since $\bar{\varphi}$ depends only on x, asking that $\bar{\varphi}$ is differentiable $\bar{\gamma}$-a.e. is equivalent to asking that $\bar{\varphi}$ is differentiable μ-a.e.

and that $\mu \ll dx$ (i.e., μ is absolutely continuous with respect to the Lebesgue measure). Then there exists a unique optimal plan $\bar{\gamma}$. In addition, $\bar{\gamma} = (\mathrm{id} \times T)_{\#}\mu$ and $T = \nabla\phi$ for some convex function ϕ.

Sketch of proof. We first prove existence, and then discuss the uniqueness.

- *Step 1: Existence.* Given $\gamma \in \Gamma(\mu,\nu)$ it holds

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \, d\gamma \leq 2 \int_{\mathbb{R}^d \times \mathbb{R}^d} \left(|x|^2 + |y|^2\right) d\gamma = 2 \int_{\mathbb{R}^d} |x|^2 \, d\mu + 2 \int_{\mathbb{R}^d} |y|^2 \, d\nu < +\infty,$$

thus Theorem 2.1 ensures the existence of a nontrivial optimal transport plan $\bar{\gamma}$. By the argument above, we can find a pair $(\bar{\varphi}, \bar{\psi})$ such that

$$\Psi(x,y) := \frac{|x-y|^2}{2} + \bar{\varphi}(x) + \bar{\psi}(y) \geq 0, \qquad \Psi = 0 \quad \bar{\gamma}\text{-a.e.}$$

It is also possible to show that $\Psi \geq 0$ implies that $\phi(x) := \frac{|x|^2}{2} + \bar{\varphi}(x)$ coincides μ-a.e. with a convex function. Hence, assuming without loss of generality that ϕ is convex, and recalling that convex functions are differentiable a.e. with respect to the Lebesgue measure, since $\mu \ll dx$ we deduce that the function ϕ (and so also $\bar{\varphi}$) is differentiable μ-a.e. Thus, the argument presented above shows that

$$y = T(x), \quad T = \nabla\phi \qquad \text{for } \bar{\gamma}\text{-a.e. } (x,y).$$

This implies that $\bar{\gamma}$ is induced by the map T and that $T = \nabla\phi$ is a solution to Monge's problem (see Remark 1.7).

- *Step 2: Uniqueness.* To show uniqueness, one observes that if $\bar{\gamma}_1$ and $\bar{\gamma}_2$ are optimal for the Kantorrovich's problem, by linearity of the problem (and convexity of the constraints) also $\frac{\bar{\gamma}_1 + \bar{\gamma}_2}{2}$ is optimal. Hence, by Step 1, there exist three convex functions ϕ_1, ϕ_2, $\bar{\phi}$ such that

$$(x,y) = (x, \nabla\phi_1(x)) \; \bar{\gamma}_1\text{-a.e.}, \qquad (x,y) = (x, \nabla\phi_2(x)) \; \bar{\gamma}_2\text{-a.e.}, \qquad (x,y) = (x, \nabla\bar{\phi}(x)) \; \frac{\bar{\gamma}_1 + \bar{\gamma}_2}{2}\text{-a.e.}$$

This implies

$$(x, \nabla\phi_1(x)) = (x, \nabla\bar{\phi}(x)) \quad \bar{\gamma}_1\text{-a.e.} \qquad \Rightarrow \qquad \nabla\phi_1(x) = \nabla\bar{\phi}(x) \quad \mu\text{-a.e.}$$

$$(x, \nabla\phi_2(x)) = (x, \nabla\bar{\varphi}(x)) \quad \bar{\gamma}_2\text{-a.e.} \qquad \Rightarrow \qquad \nabla\phi_2(x) = \nabla\bar{\varphi}(x) \quad \mu\text{-a.e.}$$

Thus $\nabla\phi_1 = \nabla\phi_2$ μ-a.e., and therefore $\bar{\gamma}_1 = \bar{\gamma}_2$, as desired. $\qquad\square$

Remark 2.5. *The argument above can be extended to more general costs. Indeed, the key point behind the proof of Theorem 2.4 is the fact that the relation (2.1), namely*

$$0 = \nabla_x \Psi(x,y) = \nabla_x c(x,y) + \nabla\varphi(x),$$

uniquely identifies y in terms of x and $\nabla\varphi(x)$. For instance, this happens when $c(x,y) = |x-y|^p$ with $p > 1$.

To see this, set $v := -\nabla\varphi(x)$ and assume that (2.1) holds. Then, since $\nabla_x c(x,y) = p|x-y|^{p-2}(x-y)$, we have

$$(2.2) \qquad\qquad p|x-y|^{p-2}(x-y) = v.$$

Since $|x-y|^{p-2} > 0$, we deduce that the vectors v and $(x-y)$ are parallel and point in the same direction, hence

$$\frac{x-y}{|x-y|} = \frac{v}{|v|}.$$

In addition, taking the modulus in relation (2.2) we get

$$p|x-y|^{p-1} = |v|.$$

Combining these two facts we deduce that

$$x - y = \frac{v}{|v|}|x - y| = \frac{v}{|v|}\left(\frac{|v|}{p}\right)^{\frac{1}{p-1}},$$

or equivalently, recalling that $v = -\nabla\varphi(x)$,

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

This proves that (2.1) uniquely identifies y in terms of x and $\nabla\varphi(x)$, and Brenier's Theorem can indeed be extended to the family of costs $c(x,y) = |x - y|^p$ when $p > 1$ (see [4, Theorem 2.7.1]).

It is worth observing that this argument fails for $p = 1$, since (2.2) becomes

$$\frac{x - y}{|x - y|} = v,$$

and this relation does not uniquely identify y in terms of x and v. We refer the interested reader to [4, Chapter 2.7] for more details.

2.4. **p-Wasserstein distances and geodesics.** In the space of probability measures with finite p-moment, optimal transport can be used to introduce the so-called p-Wasserstein distance. Although one can perform this construction in arbitrary metric spaces, here we restrict ourselves to $\mathbb{R}^d$.

Definition 2.6. *Given $1 \le p < \infty$, let*

$$(2.3) \qquad \mathcal{P}_p(\mathbb{R}^d) := \left\{\sigma \in \mathcal{P}(X) : \int_{\mathbb{R}^d}|x|^p\,d\sigma(x) < +\infty\right\}$$

be the set of probability measures with finite p-moment.

Definition 2.7. *Given $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, we define their p-Wasserstein distance as*

$$W_p(\mu, \nu) := \left(\inf_{\gamma \in \Gamma(\mu,\nu)}\int_{\mathbb{R}^d \times \mathbb{R}^d}d(x,y)^p\,d\gamma(x,y)\right)^{\frac{1}{p}}.$$

Remark 2.8. *If $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, then for all $\gamma \in \Gamma(\mu,\nu)$ it holds*

$$\int_{\mathbb{R}^d \times \mathbb{R}^d}|x - y|^p\,d\gamma \le 2^{p-1}\int_{\mathbb{R}^d \times \mathbb{R}^d}\left(|x|^p + |y|^p\right)d\gamma = 2^{p-1}\left(\int_{\mathbb{R}^d}|x|^p\,d\mu + \int_{\mathbb{R}^d}|y|^p\,d\nu\right) < \infty.$$

Hence, W_p is finite on $\mathcal{P}_p(\mathbb{R}^d) \times \mathcal{P}_p(\mathbb{R}^d)$.

The terminology "p-Wasserstein distance" is justified by the following result (see [4, Theorem 3.1.5] and [1, Proposition 7.1.5] for a proof):

Theorem 2.9. *W_p is a distance on the space $\mathcal{P}_p(\mathbb{R}^d)$, and $(\mathcal{P}_p(\mathbb{R}^d), W_p)$ is a complete metric space.*

In addition, one can prove that Wasserstein distances metrize the weak* convergence of measures on compact sets, see [4, Corollary 3.1.7] for a proof (note that, on compact sets, all probability measures have finite p-moment).

Proposition 2.10. *Let $K \subset \mathbb{R}^d$ be compact, $p \ge 1$, $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(K)$ a sequence of probability measures, and $\mu \in \mathcal{P}(K)$. Then*

$$\mu_n \rightharpoonup^* \mu \qquad \Longleftrightarrow \qquad W_p(\mu_n, \mu) \to 0.$$

Since $(\mathcal{P}_p(\mathbb{R}^d), W_p)$ is a complete metric space, we can also talk about geodesics. To describe them, let $\mu_0, \mu_1 \in \mathcal{P}_p(\mathbb{R}^d)$, and assume first, for simplicity, that there exists an optimal transport map T from μ_0 to μ_1 for the cost $c(x, y) = |x - y|^p$. Then the geodesic between μ_0 and μ_1 takes the form

$$\mu_t := (T_t)_\# \mu_0, \qquad T_t(x) := (1 - t)x + tT(x).$$

One can indeed check that, with this definition,

$$W_p(\mu_t, \mu_s) = |t - s| W_p(\mu_0, \mu_1), \qquad \text{hence } t \mapsto \mu_t \text{ is a constant-speed geodesics.}$$

In the general case when an optimal map may not exist, one considers $\gamma \in \Gamma(\mu_0, \mu_1)$ an optimal coupling for W_p, set $\pi_t(x, y) := (1 - t)x + ty$, and define $\mu_t := (\pi_t)_\# \gamma$. Note that, if the optimal coupling is not unique, then also the geodesic is not unique (since each optimal coupling induces a different geodesic). We refer to [4, Section 3.1.1] for more details.

3. AN APPLICATION OF OPTIMAL TRANSPORT: THE ISOPERIMETRIC INEQUALITY

In this section we show how one can use optimal transport to prove the sharp isoperimetric inequality in $\mathbb{R}^d$. We use $|E|$ to denote the Lebesgue measure of a set E.

Theorem 3.1. *Let $E \subset \mathbb{R}^d$ be a bounded set with smooth boundary. Then*

$$\text{Area}(\partial E) \geq d|B_1|^{\frac{1}{d}} |E|^{\frac{d-1}{d}},$$

where $|B_1|$ is the volume of the unit ball.

To prove this result, we need the following result.

Lemma 3.2. *Consider the probability measures $\mu := \frac{\mathbf{1}_E}{|E|} dx$ and $\nu := \frac{\mathbf{1}_{B_1}}{|B_1|} dy$, and let $T = \nabla\phi$ denote the Brenier map from μ to ν (see Theorem 2.4). Assume T to be smooth.[3] Then the following hold:*
(a) $|T(x)| \leq 1$ for every $x \in \overline{E}$;
(b) $\det \nabla T = \frac{|B_1|}{|E|}$ in E;
(c) $\operatorname{div} T \geq d \left(\det \nabla T \right)^{\frac{1}{d}}$.

Proof. We prove the three properties.
(a) If $x \in E$, then $T(x) \in B_1$ and thus $|T(x)| \leq 1$. By continuity, the estimate also holds for every $x \in \overline{E}$.
(b) Let $A \subset B_1$, so that $T^{-1}(A) \subset E$. Since $T_\# \mu = \nu$ and $\mu = \frac{\mathbf{1}_E}{|E|} dx$, we have

$$\nu(A) = \mu(T^{-1}(A)) = \int_{T^{-1}(A)} \frac{dx}{|E|}.$$

On the other hand, by the classical change of variable formulas, setting $y = T(x)$ we have $dy = |\det \nabla T| dx$, therefore

$$\nu(A) = \int_A \frac{dy}{|B_1|} = \int_{T^{-1}(A)} \frac{1}{|B_1|} |\det \nabla T(x)| \, dx.$$

Furthermore, since $\nabla T = D^2\phi$ is nonnegative definite (recall that ϕ is convex), it follows that $\det \nabla T \geq 0$, hence

$$\int_{T^{-1}(A)} \frac{dx}{|E|} = \nu(A) = \int_{T^{-1}(A)} \frac{1}{|B_1|} \det \nabla T(x) \, dx.$$

---

[3]The smoothness assumption can be dropped with some fine analytic arguments relying on the theory of functions of bounded variations, see [5, Section 2.2] for more details.

Since $A \subset B_1$ is arbitrary, we deduce that

$$\frac{\det \nabla T}{|B_1|} = \frac{1}{|E|} \quad \text{inside } E.$$

(c) Note that, since the matrix $\nabla T = D^2 \varphi$ is symmetric and nonnegative definite, fixed a point $x \in E$ we can choose a system of coordinates so that $\nabla T(x)$ is diagonal with nonnegative entries:

$$\nabla T(x) = \begin{pmatrix} \lambda_1(x) & 0 & 0 & 0 \\ 0 & \lambda_2(x) & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_d(x) \end{pmatrix}, \qquad \lambda_i(x) \geq 0.$$

Hence, by the arithmetic-geometric mean inequality,

$$\operatorname{div} T(x) = \sum_{i=1}^{d} \lambda_i(x) = d\left(\frac{1}{d} \sum_{i=1}^{d} \lambda_i(x)\right) \geq d\left(\prod_{i=1}^{d} \lambda_i(x)\right)^{\frac{1}{d}} = d(\det \nabla T(x))^{\frac{1}{d}}.$$

$\square$

Proof of Theorem 3.1. Thanks to properties (a), (b), (c) in Lemma 3.2, denoting by ν_E the outer unit normal to ∂E and by $d\sigma$ the surface measure on ∂E, we have

$$\operatorname{Area}(\partial E) = \int_{\partial E} 1 \, d\sigma \overset{(a)}{\geq} \int_{\partial E} |T| \, d\sigma \geq \int_{\partial E} T \cdot \nu_E \, d\sigma = \int_E \operatorname{div} T \, dx$$

$$\overset{(c)}{\geq} d \int_E \left(\det \nabla T\right)^{\frac{1}{d}} dx \overset{(b)}{=} d \int_E \left(\frac{|B_1|}{|E|}\right)^{\frac{1}{d}} dx = d|B_1|^{\frac{1}{d}} |E|^{\frac{d-1}{d}},$$

where the last equality in the first line follows from the divergence theorem. $\square$

4. Gradient flows in Hilbert spaces

4.1. **An informal introduction to gradient flows.** Let $\mathcal{H}$ be a Hilbert space (think, as a first example, $\mathcal{H} = \mathbb{R}^N$) and let $\phi \colon \mathcal{H} \to \mathbb{R}$ be of class C^1. Given $x_0 \in \mathcal{H}$, the gradient flow of ϕ starting at x_0 is given by the ordinary differential equation

$$(4.1) \qquad \begin{cases} x(0) = x_0, \\ \dot{x}(t) = -\nabla \phi(x(t)). \end{cases}$$

Note that, if $x(t)$ solves (4.1), then

$$(4.2) \qquad \frac{d}{dt} \phi(x(t)) = \nabla \phi(x(t)) \cdot \dot{x}(t) = -|\nabla \phi(x(t))|^2 \leq 0.$$

In other words, ϕ decreases along the curve $x(t)$, and $\frac{d}{dt}\phi(x(t)) = 0$ if and only if $|\nabla \phi(x(t))| = 0$ (i.e., $x(t)$ is a critical point of ϕ). In particular, if ϕ has a unique stationary point that coincides with the global minimizer (this is for instance the case if ϕ is strictly convex), then one expects $x(t)$ to converge to the minimizer as $t \to +\infty$.

Remark 4.1. *To define a gradient flow, one needs a scalar product. Indeed, as a general fact, given a function $\phi \colon \mathcal{H} \to \mathbb{R}$ one defines its differential $d\phi(x) \colon \mathcal{H} \to \mathbb{R}$ as*

$$d\phi(x)[v] = \lim_{\varepsilon \to 0} \frac{\phi(x + \varepsilon v) - \phi(x)}{\varepsilon} \qquad \forall \, v \in \mathcal{H}.$$

If ϕ is of class C^1, then the map $d\phi(x) \colon \mathcal{H} \to \mathbb{R}$ is linear and continuous, which means that $d\phi(x) \in \mathcal{H}^*$ (the dual space of $\mathcal{H}$). On the other hand, if $t \mapsto x(t) \in \mathcal{H}$ is a curve, then

$$\dot{x}(t) = \lim_{\varepsilon \to 0} \frac{x(t + \varepsilon) - x(t)}{\varepsilon} \in \mathcal{H}.$$

This shows that $\dot{x}(t) \in \mathcal{H}$ and $d\phi(x(t)) \in \mathcal{H}^*$ live in different spaces. Hence, to define a gradient flow, we need a way to identify $\mathcal{H}$ and $\mathcal{H}^*$.

This can be done if we introduce a scalar product. Indeed, if $\langle \cdot, \cdot \rangle$ is a scalar product on $\mathcal{H} \times \mathcal{H}$, we can define the gradient of f at x as the unique element of $\mathcal{H}$ such that

$$\langle \nabla \phi(x), v \rangle := d\phi(x)[v] \qquad \forall\, v \in \mathcal{H}.$$

In other words, the scalar product allows us to identify the gradient and the differential, and thanks to this identification we can now make sense of $\dot{x}(t) = -\nabla \phi(x(t))$.

Note that, if one changes the scalar product, then the gradient (and therefore the gradient flow) will be different.

4.2. Gradient flows of convex functions. Let $\phi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a convex function.

Definition 4.2. *Given $x \in \mathcal{H}$ such that $\phi(x) < +\infty$, we define the* subdifferential *of ϕ at x as*

$$\partial \phi(x) := \big\{ p \in \mathcal{H} \,:\, \phi(z) \geq \phi(x) + \langle p, z - x \rangle \quad \forall z \in \mathcal{H} \big\}.$$

Note that, if ϕ is differentiable at x, then $\partial\phi(x) = \{\nabla\phi(x)\}$.

With this definition, we can define the gradient flow of ϕ starting at x_0 as

$$(4.3) \qquad \begin{cases} x(0) = x_0, \\ \dot{x}(t) \in -\partial\phi(x(t)) & \text{for almost every } t > 0. \end{cases}$$

4.3. An example of gradient flow on $\mathcal{H} = L^2(\mathbb{R}^d)$: the heat equation.

Proposition 4.3. *Let $\mathcal{H} = L^2(\mathbb{R}^d)$ and*

$$\phi(u) = \begin{cases} \dfrac{1}{2} \displaystyle\int_{\mathbb{R}^d} |\nabla u|^2 \, dx & \text{if } u \in W^{1,2}(\mathbb{R}^d), \\ +\infty & \text{otherwise.} \end{cases}$$

Then

$$\partial\phi(u) \neq \emptyset \qquad \Longleftrightarrow \qquad \Delta u \in L^2(\mathbb{R}^d),$$

and in that case $\partial\phi(u) = \{-\Delta u\}$.

Proof. Even though the proofs are quite similar, we prove the two implications separately.
$\Rightarrow$) Let $p \in L^2(\mathbb{R}^d)$ with $p \in \partial\phi(u)$. Then, by definition, for any $v \in L^2(\mathbb{R}^d)$ we have

$$\phi(v) \geq \phi(u) + \langle p, v - u \rangle_{L^2},$$

or equivalently

$$\int_{\mathbb{R}^d} \frac{|\nabla v|^2}{2} \, dx - \int_{\mathbb{R}^d} \frac{|\nabla u|^2}{2} \, dx \geq \int_{\mathbb{R}^d} p(v - u) \, dx.$$

Take $v = u + \varepsilon w$ with $w \in W^{1,2}(\mathbb{R}^d)$ and $\varepsilon > 0$. Then, rearranging the terms and dividing by ε yields

$$\int_{\mathbb{R}^d} \nabla u \cdot \nabla w \, dx + \frac{\varepsilon}{2} \int_{\mathbb{R}^d} |\nabla w|^2 \, dx \geq \int_{\mathbb{R}^d} pw \, dx,$$

so by letting $\varepsilon \to 0$ we obtain

$$\int_{\mathbb{R}^d} \nabla u \cdot \nabla w \geq \int_{\mathbb{R}^d} pw \, dx \qquad \forall\, w \in W^{1,2}(\mathbb{R}^d).$$

Replacing w with $-w$ in the inequality above, we conclude that

$$\int_{\mathbb{R}^d} \underbrace{-\Delta u}_{\substack{\text{as a} \\ \text{distribution}}} w = \int_{\mathbb{R}^d} \nabla u \cdot \nabla w \, dx = \int_{\mathbb{R}^d} pw \, dx \qquad \forall \, w \in W^{1,2}(\mathbb{R}^d),$$

that is, $-\Delta u = p \in L^2(\mathbb{R}^d)$.

$\Leftarrow$) Assume that the distributional Laplacian Δu belongs to $L^2(\mathbb{R}^d)$. By definition of ϕ, for any $w \in W^{1,2}(\mathbb{R}^d)$ we have

$$\phi(u+w) - \phi(u) = \int_{\mathbb{R}^d} \frac{|\nabla u + \nabla w|^2}{2} \, dx - \int_{\mathbb{R}^d} \frac{|\nabla u|^2}{2} \, dx$$
$$= \int_{\mathbb{R}^d} \nabla u \cdot \nabla w \, dx + \frac{1}{2} \int_{\mathbb{R}^d} |\nabla w|^2 \, dx \geq \int_{\mathbb{R}^d} \nabla u \cdot \nabla w \, dx = \int_{\mathbb{R}^d} -\Delta u \, w \, dx.$$

On the other hand, if $u \in W^{1,2}(\mathbb{R}^d)$ (so that $\phi(u) < +\infty$) while $w \notin W^{1,2}(\mathbb{R}^d)$, then $u+w \notin W^{1,2}(\mathbb{R}^d)$ and therefore

$$\phi(u+w) = +\infty > \phi(u) + \int_{\mathbb{R}^d} -\Delta u \, w \, dx.$$

This proves that $-\Delta u \in \partial\phi(u)$. $\square$

As a consequence of this discussion, we obtain the following:

Corollary 4.4 (Heat equation as gradient flow). *Let $\mathcal{H}$ and ϕ be as in Proposition 4.3. Then the gradient flow of ϕ with respect to the L^2-scalar product is the* heat equation, *i.e.,*

$$\partial_t u(t) \in -\partial\phi(u(t)) \qquad \Longleftrightarrow \qquad \partial_t u(t,x) = \Delta u(t,x).$$

5. Continuity equation and Benamou–Brenier

Let $\Omega \subset \mathbb{R}^d$ be a convex set[4] ($\Omega = \mathbb{R}^d$ is admissible), let $\bar\rho_0 \in \mathcal{P}_2(\Omega)$ be a probability density with finite second moment, and let $v \colon [0,T] \times \Omega \to \mathbb{R}^d$ be a smooth bounded vector field tangent to the boundary of Ω. Let $X(t,x)$ denote the flow of v, namely

$$(5.1) \qquad \begin{cases} \dot{X}(t,x) = v(t, X(t,x)), \\ X(0,x) = x, \end{cases}$$

and set $\rho_t = (X(t))_\# \bar\rho_0$. Note that, since v is tangent to the boundary, the flow remains inside Ω, hence $\rho_t \in \mathcal{P}(\Omega)$.

Lemma 5.1. *Let $v_t(\cdot) := v(t, \cdot)$. Then (ρ_t, v_t) solves* continuity equation

$$(5.2) \qquad \partial_t \rho_t + \mathrm{div}(v_t \rho_t) = 0$$

in the distributional sense.

---

[4]The convexity of Ω guarantees that if $\rho_0, \rho_1 \in P_2(\Omega)$, then also the Wasserstein geodesic between them is contained inside Ω, *cf.* Section 2.4.

Proof. Let $\psi \in C_c^\infty(\Omega)$, and consider the function $t \mapsto \int_\Omega \rho_t(x)\psi(x)\,dx$. Then, using the definitions of X and ρ_t, we get

$$\int_\Omega \partial_t\rho_t(x)\psi(x)\,dx = \frac{d}{dt}\int_\Omega \rho_t(x)\psi(x)\,dx \overset{(i)}{=} \frac{d}{dt}\int_\Omega \psi(X(t,x))\bar\rho_0(x)\,dx$$

$$= \int_\Omega \nabla\psi(X(t,x))\cdot\dot X(t,x)\bar\rho_0(x)\,dx$$

$$\overset{(ii)}{=} \int_\Omega \nabla\psi(X(t,x))\cdot v_t(X(t,x))\bar\rho_0(x)\,dx$$

$$\overset{(iii)}{=} \int_\Omega \nabla\psi(x)\cdot v_t(x)\rho_t(x)\,dx = -\int_\Omega \psi(x)\mathrm{div}(v_t\rho_t)\,dx,$$

where (i) and (iii) follow from $\rho_t = (X(t))_{\#}\bar\rho_0$, while (ii) follows from (5.1). □

Definition 5.2. *Given a pair (ρ_t, v_t) solving the continuity equation (5.2) with $\rho_t v_t \cdot \nu|_{\partial\Omega} = 0$,[5] we define its action as*

$$A[\rho_t, v_t] := \int_0^1 \int_\Omega |v_t(x)|^2 \rho_t(x)\,dx\,dt.$$

The following formula, due to Benamou and Brenier [2], shows a deep link between the continuity equation and the W_2-distance.

Theorem 5.3 (Benamou–Brenier formula). *Given two probability measures $\bar\rho_0 \in \mathcal{P}_2(\Omega)$ and $\bar\rho_1 \in \mathcal{P}_2(\Omega)$, it holds that*

$$W_2(\bar\rho_0, \bar\rho_1)^2 = \inf\big\{A[\rho_t, v_t] \,:\, \rho_0 = \bar\rho_0,\ \rho_1 = \bar\rho_1,\ \partial_t\rho_t + \mathrm{div}(v_t\rho_t) = 0,\ \rho_t v_t \cdot \nu|_{\partial\Omega} = 0\big\}.$$

Proof. We give only a formal proof, referring to [1, Chapter 8] for a rigorous argument. By approximation we shall assume that $\bar\rho_0$ and $\bar\rho_1$ are absolutely continuous.

Let (ρ_t, v_t) satisfy $\rho_0 = \bar\rho_0$, $\rho_1 = \bar\rho_1$, $\partial_t\rho_t + \mathrm{div}(v_t\rho_t) = 0$, $\rho_t v_t \cdot \nu|_{\partial\Omega} = 0$, and denote by $X(t,x)$ the flow of v_t, so that $\rho_t = (X(t))_{\#}\bar\rho_0$. In particular $X(1)_{\#}\bar\rho_0 = \bar\rho_1$, which implies that $X(1)$ is a transport map from $\bar\rho_0$ to $\bar\rho_1$. Then

(5.3)
$$\begin{aligned}
A[\rho_t, v_t] &= \int_0^1 \int_\Omega |v_t|^2 \rho_t\,dx\,dt \overset{(i)}{=} \int_0^1 \int_\Omega |v_t(X(t,x))|^2 \bar\rho_0(x)\,dx\,dt \\
&\overset{(ii)}{=} \int_0^1 \int_\Omega |\dot X(t,x)|^2 \bar\rho_0(x)\,dt\,dx = \int_\Omega \bar\rho_0(x) \int_0^1 |\dot X(t,x)|^2\,dt\,dx \\
&\overset{(iii)}{\geq} \int_\Omega \bar\rho_0(x)\left|\int_0^1 \dot X(t,x)\,dt\right|^2 dx = \int_\Omega \bar\rho_0(x)|X(1,x) - x|^2\,dx \overset{(iv)}{\geq} W_2^2(\bar\rho_0, \bar\rho_1),
\end{aligned}$$

where (i) follows from $\rho_t = (X(t))_{\#}\bar\rho_0$, (ii) is just the definition of X, (iii) follows from Hölder inequality, and for (iv) we used that $X(1)$ is a transport map from $\bar\rho_0$ to $\bar\rho_1$. This proves that $W_2^2(\bar\rho_0, \bar\rho_1)$ is always less than or equal to the infimum appearing in the statement.

To show equality, take $X(t,x) = x + t(T(x) - x)$ where $T = \nabla\varphi$ is optimal from $\bar\rho_0$ to $\bar\rho_1$ (see Theorem 2.4), set $\rho_t := X(t)_{\#}\bar\rho_0$, and let v_t be such that $\dot X(t) = v_t \circ X(t)$. With these choices we

---

[5]The constraint $\rho_t v_t \cdot \nu|_{\partial\Omega} = 0$ guarantees that, whenever there is some mass ρ_t near the boundary, the flow of v_t does not allow it to go outside Ω. In particular

$$\frac{d}{dt}\int_\Omega \rho_t\,dx = \int_\Omega \partial_t\rho_t\,dx = -\int_\Omega \mathrm{div}(v_t\rho_t)\,dx = -\int_{\partial\Omega} \rho_t v_t \cdot \nu = 0,$$

hence $\rho_t \in \mathcal{P}(\Omega)$ for all t.

have $(T(x) - x) = \dot{X}(t,x) = v_t(X(t,x))$, and looking at the computations above one can easily check that all inequalities in (5.3) become equalities, proving that

$$A[\rho_t, v_t] = W_2^2(\bar\rho_0, \bar\rho_1).$$

$\square$

6. A DIFFERENTIAL VIEWPOINT OF OPTIMAL TRANSPORT

As in the previous section, we assume that $\Omega \subset \mathbb{R}^d$ is convex.

6.1. **From Benamou-Brenier to the Wasserstein scalar product.** Starting from the Benamou-Brenier formula, we will see how this motivates Otto's interpretation of the Wasserstein space as a Riemannian manifold [8].

Thanks to Theorem 5.3, we can write

$$W_2(\bar\rho_0, \bar\rho_1)^2 = \inf_{\rho_t, v_t}\left\{ \int_0^1\left(\int_\Omega |v_t|^2 \rho_t\, dx \right) dt \,\middle|\, \partial_t \rho + \mathrm{div}(v_t \rho_t) = 0,\ \rho_t v_t \cdot \nu|_{\partial\Omega} = 0,\ \rho_0 = \bar\rho_0,\ \rho_1 = \bar\rho_1 \right\}$$

$$= \inf_{\rho_t}\left\{ \inf_{v_t} \int_0^1 \int_\Omega |v_t|^2 \rho_t\, dx\, dt \,\middle|\, \partial_t \rho + \mathrm{div}(v_t \rho_t) = 0,\ \rho_t v_t \cdot \nu|_{\partial\Omega} = 0,\ \rho_0 = \bar\rho_0,\ \rho_1 = \bar\rho_1 \right\}$$

$$= \inf_{\rho_t}\left\{ \int_0^1 \inf_{v_t}\left\{ \int_\Omega |v_t|^2 \rho_t\, dx \,\middle|\, \mathrm{div}(v_t \rho_t) = -\partial_t \rho_t,\ \rho_t v_t \cdot \nu|_{\partial\Omega} = 0 \right\} dt \,\middle|\, \rho_0 = \bar\rho_0,\ \rho_1 = \bar\rho_1 \right\},$$

where in the last equality we used that, for each time $t \in [0,1]$, and for ρ_t and $\partial_t \rho_t$ given, one can minimize with respect to all vector fields v_t satisfying the constraints $\mathrm{div}(v_t \rho_t) = -\partial_t \rho_t$ and $\rho_t v_t \cdot \nu|_{\partial\Omega} = 0$.

In analogy with the formula for the Riemannian distance on a manifold[6], it is natural to define the *Wasserstein norm* of the derivative $\partial_t \rho_t$ at ρ_t as

$$(6.1) \qquad \|\partial_t \rho_t\|_{\rho_t}^2 := \inf_{v_t}\left\{ \int_\Omega |v_t|^2 \rho_t\, dx \,\middle|\, \mathrm{div}(v_t \rho_t) = -\partial_t \rho_t,\ \rho_t v_t \cdot \nu|_{\partial\Omega} = 0 \right\}.$$

In other words, at each time t, the continuity equation gives a constraint on the divergence of $v_t \rho_t$. So, with this definition, we get

$$W_2(\bar\rho_0, \bar\rho_1)^2 = \inf_{\rho_t}\left\{ \int_0^1 \|\partial_t \rho_t\|_{\rho_t}^2\, dt \,\middle|\, \rho_0 = \bar\rho_0,\ \rho_1 = \bar\rho_1 \right\}.$$

To find a better formula for the Wasserstein norm of $\partial_t \rho_t$ we want to understand the properties of the vector field v_t that realizes the infimum in (6.1). Hence, given ρ_t and $\partial_t \rho_t$, let v_t be a minimizer, and let w be a vector field such that $\mathrm{div}(w) = 0$ and $w \cdot \nu|_{\partial\Omega} = 0$. Then, for every ε we have

$$\mathrm{div}\left(\left(v_t + \varepsilon \frac{w}{\rho_t}\right)\rho_t\right) = -\partial_t \rho_t, \qquad \rho_t\left(v_t + \varepsilon \frac{w}{\rho_t}\right) \cdot \nu|_{\partial\Omega} = 0.$$

Thus $v_t + \varepsilon \frac{w}{\rho_t}$ is an admissible vector field in the minimization problem (6.1), and so by minimality of v_t we get

$$\int_\Omega |v_t|^2 \rho_t\, dx \le \int_\Omega \left| v_t + \varepsilon \frac{w}{\rho_t} \right|^2 \rho_t\, dx = \int_\Omega |v_t|^2 \rho_t\, dx + 2\varepsilon \int_\Omega \langle v_t, w \rangle\, dx + \varepsilon^2 \int_\Omega \frac{|w|^2}{\rho_t}\, dx.$$

---

[6]Given (M, g) a Riemannian manifold, one can define the distance d_M between two points $x, y \in M$ as

$$d_M(x,y)^2 = \inf\left\{ \int_0^1 |\dot\gamma(t)|^2\, dt \,\middle|\, \gamma\colon [0,1] \to M,\ \gamma(0) = x,\ \gamma(1) = y \right\}.$$

Dividing by ε and letting it go to zero yields

$$\int_\Omega \langle v_t, w \rangle = 0$$

for every w such that $\mathrm{div}(w) \equiv 0$. By Helmholtz decomposition, this implies that

$$v_t \in \{w \,:\, \mathrm{div}(w) = 0, \, w \cdot \nu|_{\partial\Omega} = 0\}^\perp = \{\nabla q \,:\, q\colon \Omega \to \mathbb{R}\}.$$

In other words, if v_t realizes the infimum in (6.1), then there exists a function ψ_t such that $v_t = \nabla \psi_t$. Also, since $\mathrm{div}(v_t \rho_t) = -\partial_t \rho_t$ and $\rho_t v_t \cdot \nu|_{\partial\Omega} = 0$, then ψ_t is a solution of

(6.2)
$$\begin{cases} \mathrm{div}(\rho_t \nabla \psi_t) = -\partial_t \rho_t & \text{in } \Omega, \\ \rho_t \dfrac{\partial \psi_t}{\partial \nu} = 0 & \text{on } \partial\Omega. \end{cases}$$

Note that if ρ_t is a smooth curve of positive probability densities, then (6.2) is a uniformly elliptic equation with Neumann boundary conditions for ψ_t, and the solution ψ_t is unique up to a constant.[7] So, instead of using (6.1), one can define

$$\|\partial_t \rho_t\|_{\rho_t}^2 = \int_\Omega |\nabla \psi_t|^2 \rho_t \, dx,$$

where ψ_t solves (6.2).

More generally, given $\rho \in \mathcal{P}_2(\Omega)$, we can construct a scalar product (compatible with the norm defined above) as follows:

Definition 6.1. *Given two functions* $h_1, h_2\colon \Omega \to \mathbb{R}$ *with* $\int_\Omega h_1 = \int_\Omega h_2 = 0$,[8] *we define their Wasserstein scalar product at* ρ *as*

$$\langle h_1, h_2 \rangle_\rho := \int_\Omega \nabla \psi_1 \cdot \nabla \psi_2 \, \rho \, dx, \qquad where \quad \begin{cases} \mathrm{div}(\rho \nabla \psi_i) = -h_i & in \ \Omega, \\ \rho \dfrac{\partial \psi_i}{\partial \nu} = 0 & on \ \partial\Omega. \end{cases}$$

6.2. **Wasserstein gradient flows.** Now that we have a scalar product (see Definition 6.1), we can define the gradient of a functional in the Wasserstein space (cf. Remark 4.1).

Definition 6.2. *Given a functional* $\mathcal{F}\colon \mathcal{P}_2(\Omega) \to \mathbb{R}$, *its* Wasserstein gradient *at* $\bar\rho \in \mathcal{P}_2(\Omega)$ *is the unique function* $\mathrm{grad}_{W_2} \mathcal{F}[\bar\rho]$ *(if it exists) such that*

$$\left\langle \mathrm{grad}_{W_2} \mathcal{F}[\bar\rho], \frac{\partial \rho_\varepsilon}{\partial \varepsilon}\Big|_{\varepsilon=0} \right\rangle_{\bar\rho} = \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \mathcal{F}[\rho_\varepsilon]$$

for any smooth curve $(-\varepsilon_0, \varepsilon_0) \in \varepsilon \mapsto \rho_\varepsilon \in \mathcal{P}(\Omega)$ *with* $\rho_0 = \bar\rho$.

---

[7]Note that, since by assumption $\int_\Omega \rho_t \, dx = 1$, then

$$\int_\Omega \partial_t \rho_t \, dx = \frac{d}{dt} \int_\Omega \rho_t \, dx = 0.$$

It is a classical fact (at least in the smooth setting when $\rho_t > 0$) that the zero-integral condition above is necessary and sufficient for the solvability of (6.2).

[8]The condition $\int_\Omega h = 0$ is needed for the solvability of the elliptic equation

$$\begin{cases} \mathrm{div}(\rho \nabla \psi) = -h & \text{in } \Omega, \\ \rho \frac{\partial \psi}{\partial \nu} = 0 & \text{on } \partial\Omega, \end{cases}$$

since

$$\int_\Omega h \, dx = -\int_\Omega \mathrm{div}(\rho \nabla \psi) \, dx = -\int_{\partial\Omega} \rho \frac{\partial \psi}{\partial \nu} = 0.$$

We now aim to obtain an explicit formula for the Wasserstein gradient of a functional. Given a functional $\mathcal{F}\colon \mathcal{P}_2(\Omega) \to \mathbb{R}$ and a probability measure $\bar{\rho} \in \mathcal{P}_2(\Omega)$, let us denote by $\frac{\delta \mathcal{F}[\bar{\rho}]}{\delta \rho}$ its first L^2-variation, i.e., the function such that

$$\frac{d}{d\epsilon}\Big|_{\epsilon=0} \mathcal{F}[\rho_\epsilon] = \int_\Omega \frac{\delta \mathcal{F}[\bar{\rho}]}{\delta \rho}(x) \frac{\partial \rho_\epsilon(x)}{\partial \epsilon}\Big|_{\epsilon=0} dx$$

for any smooth curve $(-\varepsilon_0, \varepsilon_0) \in \varepsilon \mapsto \rho_\varepsilon \in \mathcal{P}_2(\Omega)$ such that $\rho_0 = \bar{\rho}$. Hence, by Definition 6.2,

$$\Big\langle \mathrm{grad}_{W_2} \mathcal{F}[\bar{\rho}], \frac{\partial \rho_\varepsilon}{\partial \varepsilon}\Big|_{\varepsilon=0} \Big\rangle_{\bar{\rho}} = \int_\Omega \frac{\delta \mathcal{F}[\bar{\rho}]}{\delta \rho} \frac{\partial \rho_\varepsilon}{\partial \varepsilon}\Big|_{\varepsilon=0} dx.$$

Thus, denoting by ψ the solution of

$$\begin{cases} \mathrm{div}(\nabla \psi\, \bar{\rho}) = -\frac{\partial \rho_\varepsilon}{\partial \varepsilon}\big|_{\varepsilon=0} & \text{in } \Omega, \\ \bar{\rho}\frac{\partial \psi}{\partial \nu} = 0 & \text{on } \partial\Omega, \end{cases}$$

we have

$$\Big\langle \mathrm{grad}_{W_2} \mathcal{F}[\bar{\rho}], \frac{\partial \rho_\varepsilon}{\partial \varepsilon}\Big|_{\varepsilon=0} \Big\rangle_{\bar{\rho}} = -\int_\Omega \frac{\delta \mathcal{F}[\bar{\rho}]}{\delta \rho} \mathrm{div}(\nabla\psi\bar{\rho})\, dx = \int_\Omega \nabla \frac{\delta \mathcal{F}[\bar{\rho}]}{\delta \rho} \cdot \nabla \psi \bar{\rho}\, dx$$

and therefore, by Definition 6.1, it follows that

$$(6.3) \qquad \mathrm{grad}_{W_2} \mathcal{F}[\bar{\rho}] = -\mathrm{div}\Big(\nabla\Big(\frac{\delta \mathcal{F}[\bar{\rho}]}{\delta \rho}\Big)\bar{\rho}\Big), \qquad \text{with} \quad \bar{\rho}\frac{\partial \Big(\frac{\delta \mathcal{F}[\bar{\rho}]}{\delta \rho}\Big)}{\partial \nu}\Big|_{\partial\Omega} = 0.$$

Example 6.3. *If $\mathcal{F}[\rho] = \int_\Omega U(\rho(x))\, dx$ with $U\colon \mathbb{R} \to \mathbb{R}$, then for any smooth variation $\epsilon \mapsto \rho_\epsilon$ it holds that*

$$\frac{d}{d\epsilon}\Big|_{\epsilon=0} \int_\Omega U(\rho_\epsilon(x))\, dx = \int_\Omega U'(\bar{\rho}(x))\frac{\partial \rho_\epsilon(x)}{\partial \epsilon}\Big|_{\epsilon=0} dx,$$

and therefore the first L^2-variation of $\mathcal{F}[\rho]$ at $\bar{\rho} \in \mathcal{P}_2(\Omega)$ is given by

$$\frac{\delta \mathcal{F}[\bar{\rho}]}{\delta \rho}(x) = U'(\bar{\rho}(x)).$$

Using (6.3), this implies that the Wasserstein gradient of $\mathcal{F}$ is

$$\mathrm{grad}_{W_2} \mathcal{F}[\bar{\rho}] = -\mathrm{div}\big(\bar{\rho}\nabla[U'(\bar{\rho})]\big) = -\mathrm{div}\big(\bar{\rho}U''(\bar{\rho})\nabla\bar{\rho}\big), \qquad \bar{\rho}U''(\bar{\rho})\frac{\partial \bar{\rho}}{\partial \nu}\Big|_{\partial\Omega} = 0.$$

In the special case $U(s) = s\log(s)$ (hence, $\mathcal{F}$ is the entropy functional) one has $U''(s) = \frac{1}{s}$, thus

$$\mathrm{grad}_{W_2} \mathcal{F}[\bar{\rho}] = -\Delta\bar{\rho}, \qquad \frac{\partial \bar{\rho}}{\partial \nu}\Big|_{\partial\Omega} = 0.$$

If instead $U(s) = \frac{s^m}{m-1}$ for some $m \neq 1$, then we get

$$\mathrm{grad}_{W_2} \mathcal{F}[\bar{\rho}] = -m\,\mathrm{div}\big(\bar{\rho}^{m-1}\nabla\bar{\rho}\big) = -\Delta(\bar{\rho}^m), \qquad \frac{\partial \bar{\rho}^m}{\partial \nu}\Big|_{\partial\Omega} = 0.$$

Example 6.4. *Let $\mathcal{F}[\rho] = \int_\Omega V(x)\rho(x)\, dx$ with $V\colon \Omega \to \mathbb{R}$. Then*

$$\frac{\delta \mathcal{F}[\bar{\rho}]}{\delta \rho}(x) = V(x),$$

and therefore the Wasserstein gradient of $\mathcal{F}$ is

$$\mathrm{grad}_{W_2} \mathcal{F}[\bar{\rho}] = -\mathrm{div}\big(\nabla V\, \bar{\rho}\big), \qquad \bar{\rho}\frac{\partial V}{\partial \nu}\Big|_{\partial\Omega} = 0.$$

Example 6.5. *Let* $\mathcal{F}[\rho] = \frac{1}{2} \iint_{\Omega \times \Omega} W(x - y)\rho(x)\rho(y)\,dx\,dy$ *with* $W \colon \mathbb{R}^d \to \mathbb{R}$ *such that* $W(z) = W(-z)$. *Then*

$$\frac{\delta \mathcal{F}[\bar{\rho}]}{\delta \rho}(x) = W * \bar{\rho}(x) = \int_\Omega W(x - y)\rho(y)\,dy,$$

where $*$ *denotes the convolution, and therefore the Wasserstein gradient of* $\mathcal{F}$ *is*

$$\mathrm{grad}_{W_2}\mathcal{F}[\bar{\rho}] = -\mathrm{div}\big((\nabla W * \bar{\rho})\bar{\rho}\big), \qquad \bar{\rho}\frac{\partial W * \bar{\rho}}{\partial \nu}\Big|_{\partial \Omega} = 0.$$

Finally, the definition of gradient flow in the Wasserstein space is the expected one.

Definition 6.6. *Given a functional* $\mathcal{F} \colon \mathcal{P}_2(\Omega) \to \mathbb{R}$, *a curve of probability measure* $[0, T) \ni t \mapsto \rho_t \in \mathcal{P}_2(\Omega)$ *is a* W_2-*gradient flow of* $\mathcal{F}$ *starting from* $\bar{\rho}_0$ *if*

$$\begin{cases} \partial_t \rho_t = -\mathrm{grad}_{W_2}\mathcal{F}[\rho_t], \\ \rho_0 = \bar{\rho}_0. \end{cases}$$

By the computations in Example 6.3, the W_2-gradient flow of the entropy functional $\mathcal{F}[\rho] = \int_\Omega \rho \log(\rho)\,dx$ is the heat equation with Neumann boundary conditions:

$$\partial_t \rho_t = \Delta \rho_t, \qquad \frac{\partial \rho_t}{\partial \nu}\Big|_{\partial \Omega} = 0.$$

On the other hand, if $\mathcal{F}[\rho] = \frac{1}{m-1}\int_\Omega \rho^m$ for $m \neq 1$ with $m > 0$, then the gradient flow is

$$\partial_t \rho_t = \Delta(\rho_t^m), \qquad \frac{\partial \rho_t^m}{\partial \nu}\Big|_{\partial \Omega} = 0.$$

that is, the porous medium equation (if $m > 1$) or the fast diffusion equation (if $m \in (0, 1)$) .

6.3. **Implicit Euler and JKO schemes.** Given $\phi \colon \mathcal{H} \to \mathbb{R}$ of class C^1, a classical way to construct solutions of (4.1) is by discretizing the ODE in time, via the so-called *implicit Euler scheme*. More precisely, with a small fixed time step $\tau > 0$, we discretize the time derivative $\dot{x}(t)$ as $\frac{x(t+\tau)-x(t)}{\tau}$, so that (4.1) becomes

$$\frac{x(t + \tau) - x(t)}{\tau} = -\nabla\phi(y)$$

for a suitable choice of the point y. A natural idea would be to choose $y = x(t)$ (as in the *explicit* Euler scheme), but for our purposes the choice $y = x(t + \tau)$ (as in the *implicit* Euler scheme) works better. Thus, given $x(t)$, one looks for a point $x(t + \tau) \in \mathcal{H}$ solving the relation

$$\frac{x(t + \tau) - x(t)}{\tau} = -\nabla\phi(x(t + \tau)).$$

With this idea in mind, we set $x_0^\tau = x_0$. Then, given $k \geq 0$ and x_k^τ, we want to find x_{k+1}^τ by solving

$$\frac{x_{k+1}^\tau - x_k^\tau}{\tau} = -\nabla\phi(x_{k+1}^\tau),$$

or equivalently

$$\nabla_x\left(\frac{\|x - x_k^\tau\|^2}{2\tau} + \phi(x)\right)\Big|_{x=x_{k+1}^\tau} = \frac{x_{k+1}^\tau - x_k^\tau}{\tau} + \nabla\phi(x_{k+1}^\tau) = 0.$$

In other words, x_{k+1}^τ is a critical point of the function $\psi_k^\tau(x) := \frac{\|x-x_k^\tau\|^2}{2\tau} + \phi(x)$. Therefore, a natural way to construct x_{k+1}^τ is by looking for a global minimizer of ψ_k^τ:

$$x_{k+1}^\tau := \mathrm{argmin}\left\{x \mapsto \frac{\|x - x_k^\tau\|^2}{2\tau} + \phi(x)\right\}.$$

Once the sequence $(x_k^\tau)_{k\geq 0}$ has been constructed, setting $x^\tau(0) := x_0$ and $x^\tau(t) := x_k^\tau$ for $t \in ((k-1)\tau, k\tau]$, one obtains an approximate solution $t \mapsto x^\tau(t)$. Then, the main challenge is to let $\tau \to 0$ and prove that there exists a limit curve $x(t)$ that indeed solves (4.1).

In many situations, this approach works and allows one to construct gradient flows (for instance, when ϕ is convex, it allows one to construct solutions of (4.3)). Hence, motivated by this, one can mimic the same strategy in the Wasserstein space: given an initial measure $\bar{\rho}_0$ and a functional $\mathcal{F}$, one inductively defines

$$\rho_{k+1}^\tau \text{ is the minimizer in } \mathcal{P}(\mathbb{R}^d) \text{ of } \rho \mapsto \frac{W_2^2(\rho, \rho_k^\tau)}{2\tau} + \mathcal{F}[\rho].$$

This approach is called JKO scheme, as it was first introduced by Jordan, Kinderlehrer, and Otto in the case $\mathcal{F}[\rho] = \int \rho \log(\rho)$ to construct solutions to the heat equation [6], and it has become by now a very versatile tool to construct solutions of evolution PDEs.

6.4. **Displacement convexity.** When considering gradient flows, the convexity of the energy plays a crucial role. Indeed, let ϕ be a convex function, and let $x(t)$, $y(t)$ be solutions of (4.3) with initial conditions x_0 and y_0 respectively. If ϕ is of class C^1 then

$$\frac{d}{dt}\frac{\|x(t) - y(t)\|^2}{2} = \langle x(t) - y(t), \dot{x}(t) - \dot{y}(t)\rangle = -\langle x(t) - y(t), \nabla\phi(x(t)) - \nabla\phi(y(t))\rangle \leq 0,$$

where the last inequality follows from the convexity of ϕ. This argument can easily be extended to non-smooth convex functions (see [4, Remark 3.2.3]), yielding

$$\|x(t) - y(t)\| \leq \|x_0 - y_0\| \qquad \forall t \geq 0,$$

which implies both uniqueness and stability of gradient flows.

A natural question is understanding the analogue of convexity in the Wasserstein space. The correct answer has been found by McCann [7]:

Definition 6.7. *We say that a functional $\mathcal{F}: \mathcal{P}_2(\Omega) \to \mathbb{R}$ is W_2-convex, or* displacement convex, *if the one-dimensional function*

$$[0,1] \ni t \mapsto \mathcal{F}[\rho_t]$$

is convex for any W_2-geodesic $[0,1] \ni t \mapsto \rho_t \in \mathcal{P}_2(\Omega)$.

It turns out that, under this assumption, uniqueness and stability hold. More precisely, let $\mathcal{F}$ be displacement convex, and let ρ_t and σ_t be two gradient flows of $\mathcal{F}$. Then

$$W_2(\rho_t, \sigma_t) \leq W_2(\rho_0, \sigma_0) \qquad \forall t \geq 0,$$

see [1, Theorem 11.1.4]. Motivated by these results, it is natural to look for sufficient conditions that guarantee the displacement convexity of a functional.

We summarize here some important examples (see [4, Chapter 4.3] for a proof):

Proposition 6.8. *Let $\Omega \subset \mathbb{R}^d$ be a convex set. The following functionals are displacement convex on $\mathcal{P}_2(\Omega)$.*

(1) $\mathcal{F}[\rho] := \int_\Omega U(\rho(x))\,dx$, where $U: [0,\infty) \to \mathbb{R}$ satisfy

$$(0,\infty) \ni s \mapsto U\left(\frac{1}{s^d}\right)s^d \quad \text{is convex and nonincreasing.}$$

(2) $\mathcal{F}[\rho] := \int_\Omega V(x)\rho(x)\,dx$ with $V: \Omega \to \mathbb{R}$ convex.
(3) $\mathcal{F}[\rho] := \iint_{\Omega\times\Omega} W(x-y)\rho(x)\rho(y)\,dx\,dy$ with $W: \mathbb{R}^d \to \mathbb{R}$ convex.

Corollary 6.9. *Let $\mathcal{F}[\rho] = \int_\Omega U(\rho(x))\,dx$. The following choices of U induce displacement convex functionals:*

$$U(s) := \begin{cases} s\log(s) & \rightsquigarrow & \partial_t\rho_t = \Delta\rho & \text{(heat equation)}, \\ \frac{1}{m-1}s^m & \text{for } m > 1 & \rightsquigarrow & \partial_t\rho_t = \Delta(\rho^m) & \text{(porous medium equation)}, \\ \frac{1}{m-1}s^m & \text{for } m \in [1-\frac{1}{d},1) & \rightsquigarrow & \partial_t\rho_t = \Delta(\rho^m) & \text{(fast diffusion equation)}. \end{cases}$$

7. Conclusion

In these notes, we have introduced the optimal transport problem, the Wasserstein distances, and the theory of Wasserstein gradient flows. Of course, the material contained here is far from complete. We invite the interested reader to consult the recent monograph [4] for more details, an extended bibliography, and further readings.

References

[1] L. Ambrosio, N. Gigli, G. Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Second edition. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2008. x+334 pp.

[2] J.-D. Benamou, Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *umer. Math.* 84 (2000), no. 3, 375-393.

[3] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* 44 (1991), no. 4, 375-417.

[4] A. Figalli, F. Glaudo. *An invitation to optimal transport, Wasserstein distances, and gradient flows.* EMS Text-books in Mathematics. EMS Press, Berlin, 2021. vi+136 pp.

[5] A. Figalli, F. Maggi, A. Pratelli. A mass transportation approach to quantitative isoperimetric inequalities. *Invent. Math.* 182 (2010), no. 1, 167-211.

[6] R. Jordan, D. Kinderlehrer, F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.* 29 (1998), no. 1, 1-17

[7] R. J. McCann. A convexity principle for interacting gases. *Adv. Math.* 128 (1997), no. 1, 153-179.

[8] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations* 26 (2001), no. 1-2, 101-174.

[9] C. Villani. *Topics in optimal transportation.* Graduate Studies in Mathematics, 58. American Mathematical Society, Providence, RI, 2003. xvi+370 pp.

Email address: `alessio.figalli@math.ethz.ch`

Department of Mathematics, ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland