

Optimal transport: introduction, applications and derivation
Lecture 2: the Benamou-Brenier formula

Contents

1	The Monge-Kantorovic distance	2
2	The continuity equation	4
2.1	Definition, first example	5
2.2	In the framework of Cauchy-Lipschitz	6
2.3	Generalized flows	7
2.4	All solutions are induced by generalized flows	11
3	Benamou-Brenier formula	12
4	Interpretation, additional information	13
5	Extensions	16

Now that we have seen what optimal transport problems are, we can start studying the space of probability measures over a compact subspace of \mathbb{T}^d as a formal Riemannian manifold, meaning that:

- We can define a distance between such measures.
- This distance comes with a "bilinear form" on some "tangent space" allowing to define the notion of action, geodesics, gradient of functionals...

This is enough to perform some formal differential calculus of order 1.

1 The Monge-Kantorovic distance

Let (X, d) be a compact metric space.

Definition 1. Let $\mu, \nu \in \mathcal{P}(X)$ be two probability measures over X . The Wasserstein (or Monge-Kantorovic) distance of order 2 between μ and ν is defined by the formula

$$W_2^2(\mu, \nu) = \inf \left\{ \int d(x, y)^2 d\gamma(x, y) \mid \gamma \in \Pi(\mu, \nu) \right\}.$$

Remark 2. • We could extend this definition in the non-compact case to all probability measures μ having a finite second moment, that is, when for some (and hence any) $x_0 \in X$, $\int d(x_0, x)^2 d\mu(x) < +\infty$.

- We could also define more distances by replacing the 2 by any other power $p \in [1, \infty)$.
- Because of the results of the first lecture, this formula gives a proper definition, an optimal γ exists, and duality holds.

Proposition 3. *The Wasserstein distance is a distance on the space $\mathcal{P}(X)$.*

In order to prove the most difficult part of this proposition, that is, the triangle inequality, we will need the disintegration theorem, that we recall here.

Theorem 4 (Disintegration theorem). *Let X, Y be two complete separable metric spaces, $m \in \mathcal{P}(X)$ and $T : X \rightarrow Y$ a measurable map. Then there exists a measurable collection $(m^y)_{y \in Y}$ of probability measures on X such that for $T_{\#}m$ -almost all y , m^y is concentrated on the set $\{x \mid T(x) = y\}$, and such that for all measurable function φ on X nonnegative or bounded*

$$\int_X \varphi dm = \int_Y \left(\int_X \varphi dm^y \right) dT_{\#}m(y). \quad (1)$$

Moreover, this family is unique up to a $T_{\#}m$ -negligible set.

In the case when $X = Y \times Z$ is the product of two separable and complete metric spaces, and T is the projection on the first coordinate, up to identifying $\{y\} \times Z$ with Z , we can assume that (m^y) is a measurable family of probability measures on Z . In that case, formula (1) writes for all φ measurable and nonnegative or bounded on $Y \times Z$,

$$\int \varphi(y, z) dm(y, z) = \int_Y \left(\int_Z \varphi(y, z) dm^y(z) \right) dT_{\#}m(y).$$

Proof of Proposition 3. Symmetry. If $\gamma \in \mathcal{P}(X \times X)$, we denote by $\gamma^s \in \mathcal{P}(X \times X) := (\pi_2, \pi_1)_{\#}\gamma$, where π_1 and π_2 are the canonical projections from $X \times X$ to X . It is easy to check that for all $\mu, \nu \in \mathcal{P}(X)$, $\gamma \in \Pi(\mu, \nu)$ if and only if $\gamma^s \in \Pi(\nu, \mu)$. Therefore, if $\gamma \in \Pi(\mu, \nu)$ is optimal, then γ^s is a competitor between ν and μ , and hence

$$W_2^2(\nu, \mu) \leq \int d(x, y)^2 d\gamma^s(x, y) = \int d(y, x)^2 d\gamma(x, y) = \int d(x, y)^2 d\gamma(x, y) = W_2^2(\mu, \nu)^2.$$

Inverting the role of μ and ν , the result follows.

Separation. Let $\mu, \nu \in \mathcal{P}(X)$ be such that $W_2^2(\mu, \nu) = 0$, and $\gamma \in \Pi(\mu, \nu)$ optimal. We have

$$\int d(x, y)^2 d\gamma(x, y) = 0.$$

So γ -almost everywhere, $d(x, y) = 0$, that is, $x = y$. So for all $\varphi \in C(X)$,

$$\int \varphi(x) d\mu(x) = \int \varphi(x) d\gamma(x, y) = \int \varphi(y) d\gamma(x, y) = \int \varphi(y) d\nu(y).$$

Therefore, $\mu = \nu$.

Triangle inequality. Let $\mu_1, \mu_2, \mu_3 \in \mathcal{P}(X)$, and $\gamma_{12} \in \Pi(\mu_1, \mu_2)$ and $\gamma_{23} \in \Pi(\mu_2, \mu_3)$ be optimal transport plans. We use the disintegration theorem for disintegrating γ_{12} along the second coordinate, and γ_{23} along the first one. We define

$$\gamma_{13} := \int \gamma_{12}^y \otimes \gamma_{23}^y d\mu_2(y).$$

Let us check that $\gamma_{13} \in \Pi(\mu_1, \mu_3)$. For all $\varphi \in C(X)$,

$$\begin{aligned} \int \varphi(x) d\gamma_{13}(x, z) &= \int \left(\int \varphi(x) d\gamma_{12}^y \otimes \gamma_{23}^y(x, z) \right) d\mu_2(y) \\ &= \int \left(\int \varphi(x) d\gamma_{12}^y(x) \right) d\mu_2(y) \\ &= \int \varphi(x) d\gamma_{12}(x, y) = \int \varphi d\mu_1. \end{aligned}$$

The second marginal follows the same lines.

In particular,

$$W_2^2(\mu_1, \mu_3) \leq \int d(x, z)^2 d\gamma_{13}(x, z) = \int \left(\int d(x, z)^2 d\gamma_{12}^y \otimes \gamma_{23}^y(x, z) \right) d\mu_2(y).$$

But we can use the triangle inequality inside the integral, and find

$$W_2(\mu_1, \mu_3) \leq \sqrt{\int \left(\int (d(x, y) + d(y, z))^2 d\gamma_{12}^y \otimes \gamma_{23}^y(x, z) \right) d\mu_2(y)}.$$

Now by the Minkowski inequality,

$$\begin{aligned} W_2(\mu_1, \mu_3) &\leq \sqrt{\int \left(\int d(x, y)^2 d\gamma_{12}^y \otimes \gamma_{23}^y(x, z) \right) d\mu_2(y)} + \sqrt{\int \left(\int d(y, z)^2 d\gamma_{12}^y \otimes \gamma_{23}^y(x, z) \right) d\mu_2(y)} \\ &= \sqrt{\int \left(\int d(x, y)^2 d\gamma_{12}^y(x) \right) d\mu_2(y)} + \sqrt{\int \left(\int d(y, z)^2 d\gamma_{23}^y(z) \right) d\mu_2(y)} \\ &= \sqrt{\int d(x, y)^2 d\gamma_{12}(x, y)} + \sqrt{\int d(y, z)^2 d\gamma_{23}(y, z)} \\ &= W_2(\mu_1, \mu_2) + W_2(\mu_2, \mu_3). \end{aligned}$$

□

If we wonder which topology the Wasserstein distance metrizes, when X is compact, we recover the narrow topology. We do it for $p = 2$, but the result is true for any $p \in [1, +\infty)$.

Proposition 5. *The Wasserstein distance metrizes the narrow topology.*

Proof. Narrow convergence implies convergence in W^2 .

Let us assume that $(\mu_n) \in \mathcal{P}(X)^{\mathbb{N}}$ converges narrowly towards $\mu \in \mathcal{P}(X)$. We show by contradiction that $W_2(\mu_n, \mu) \rightarrow 0$. If it is not true, up to extraction, we can assume that $\inf_n W_2^2(\mu_n, \mu) > 0$. By duality, for all $n \in \mathbb{N}$, there exists (φ_n, ψ_n) a pair of d^2 -concave conjugates such that

$$W_2^2(\mu_n, \mu) = \int \varphi_n d\mu_n + \int \psi_n d\mu. \quad (2)$$

Of course without loss of generality, we can assume $\inf \varphi_n = 0$ for all $n \in \mathbb{N}$. But by uniform equiintegrability of the set of d^2 -concave functions, up to extraction, we can assume that (φ_n) and (ψ_n) uniformly converge towards $\varphi \in C(X)$ and $\psi \in C(X)$ respectively. By pointwise convergence, we have $\varphi \oplus \psi \leq d^2$. In particular, for all $x \in X$, $\varphi(x) + \psi(x) \leq 0$. Passing to the limit in (2) (which is valid since in the first integral, (φ_n) converges strongly and (μ_n) converges weakly), we find

$$0 \leq \lim_{n \rightarrow +\infty} W_2^2(\mu_n, \mu) = \int \varphi d\mu + \int \psi d\mu = \int (\varphi(x) + \psi(x)) d\mu(x) \leq 0.$$

So we reached a contradiction, and $W_2(\mu_n, \mu) \rightarrow 0$.

Convergence in W_2 implies narrow convergence.

Let us consider $(\mu_n) \in \mathcal{P}(X)^{\mathbb{N}}$ and $\mu \in \mathcal{P}(X)$ with $W_2(\mu_n, \mu) \rightarrow 0$. We need to show that for all $\varphi \in C(X)$, $\int \varphi d\mu_n \rightarrow \int \varphi d\mu$. Let $\varepsilon > 0$ and φ^ε be a Lipschitz continuous function such that $\|\varphi - \varphi^\varepsilon\|_\infty \leq \varepsilon/2$. We have

$$\begin{aligned} \left| \int \varphi d\mu_n - \int \varphi d\mu \right| &\leq \int |\varphi - \varphi^\varepsilon| d\mu_n + \left| \int \varphi^\varepsilon d\mu_n - \int \varphi^\varepsilon d\mu \right| + \int |\varphi - \varphi^\varepsilon| d\mu \\ &\leq \varepsilon + \left| \int \varphi^\varepsilon d\mu_n - \int \varphi^\varepsilon d\mu \right|. \end{aligned}$$

But let us call $L^\varepsilon > 0$ a Lipschitz bound for φ^ε , and for all $n \in \mathbb{N}$, $\gamma_n \in \Pi(\mu_n, \mu)$ optimal. We have

$$\begin{aligned} \left| \int \varphi^\varepsilon d\mu_n - \int \varphi^\varepsilon d\mu \right| &= \left| \int (\varphi^\varepsilon(x) - \varphi^\varepsilon(y)) d\gamma_n(x, y) \right| \leq \int |\varphi^\varepsilon(x) - \varphi^\varepsilon(y)| d\gamma_n(x, y) \\ &\leq L^\varepsilon \int d(x, y) d\gamma_n(x, y) \leq L^\varepsilon \sqrt{\int d(x, y)^2 d\gamma_n(x, y)} \rightarrow 0. \end{aligned}$$

Bringing everything together, we find

$$\limsup_{n \rightarrow +\infty} \left| \int \varphi d\mu_n - \int \varphi d\mu \right| \leq \varepsilon,$$

and the result is proved by letting $\varepsilon \rightarrow 0$. □

2 The continuity equation

In the following of this lecture, the goal is to reinterpret the Wasserstein distance as a geodesic distance with respect to some bilinear form on some space "tangent" to the space of measures. To do so, we need to define what is the "speed" of a curve in the space of measures. This will be done thanks to the distributional solutions of the continuity equation, which can be written

$$\partial_t \rho + \operatorname{div}(\rho v) = 0. \quad (3)$$

2.1 Definition, first example

We will give a definition of solutions on the d -dimensional torus \mathbb{T}^d to avoid technicalities, but the whole theory could be developed in the \mathbb{R}^d with only technical changes, see [1].

Definition 6. Let $\rho : [0, 1] \rightarrow \mathcal{P}(\mathbb{T}^d)$ be a measurable curve in the space of probability measures over \mathbb{T}^d and $v : [0, 1] \times \mathbb{T}^d \rightarrow \mathbb{R}^d$ be a measurable vector field. We say that the pair (ρ, v) is a solution of the continuity equation provided:

- Continuity: The map $t \in [0, 1] \mapsto \rho(t) \in \mathcal{P}(\mathbb{T}^d)$ is continuous for the topology of narrow convergence.
- Integrability: The following bound holds:

$$\int_0^1 \int |v(t, x)|^2 d\rho(t, x) dt < +\infty. \quad (4)$$

- Distributional solution: For all $\varphi \in C_c^1((0, 1) \times \mathbb{T}^d)$,

$$\int_0^1 \int \{\partial_t \varphi(t, x) + v(t, x) \cdot \nabla \varphi(t, x)\} d\rho(t, x) dt = 0. \quad (5)$$

Remark 7. • The continuity assumption could be removed. Indeed, it is possible to prove that whenever (ρ, v) satisfies the two other conditions of the previous definition, then $t \mapsto \rho(t)$ is continuous up to changing its value on a dt -negligible set. In that case, we always work with this version of ρ .

- The integrability condition rewrites $v \in L^2(dt \otimes \rho(t))$, and indeed, we see with the second condition that only its values $dt \otimes \rho(t)$ -almost everywhere matter. It is possible to have a more general theory only asking for L^1 integrability, but for our purpose of dealing with a formal Riemannian manifold, L^2 integrability will play a role later and will slightly simplify some proofs.
- We can check that under assumption (4), condition (5) rewrites as follows: for all $\varphi \in C^1(\mathbb{T}^d)$, the map $t \mapsto \int \varphi d\rho(t, x)$ has distributional derivative given by

$$\frac{d}{dt} \int \varphi d\rho(t, x) = \int \nabla \varphi(x) \cdot v(t, x) d\rho(t, x).$$

If so, due to the bounds we have on v , this map is in $H^1(0, 1)$, and this identity also holds for almost every t .

The whole theory we want to develop relies on the following crucial example of a solution to the continuity equation.

Example 8. Let $\gamma : [0, 1] \rightarrow \mathbb{T}^d$ be a solution to the classical ODE $\dot{\gamma}(t) = v(t, \gamma(t))$ for some smooth vector field v . Then calling $\rho(t) := \delta_{\gamma(t)}$, the pair (ρ, v) is a solution of the continuity equation.

Proof. The integrability is obvious assuming that v is smooth. Let us still write it to get familiar with the kind of quantities that we will deal with:

$$\int_0^1 |v(t, \gamma(t))|^2 dt < +\infty.$$

To check the distributional condition, let us consider $\varphi \in C_c^1((0, 1) \times \mathbb{T}^d)$. Then the function $t \mapsto \varphi(t, \gamma(t))$ is differentiable and

$$\frac{d}{dt} \varphi(t, \gamma(t)) = \partial_t \varphi(t, \gamma(t)) + v(t, \gamma(t)) \cdot \nabla \varphi(t, \gamma(t)).$$

Integrating from 0 to 1, we find indeed

$$\int_0^1 \{\partial_t \varphi(t, \gamma(t)) + v(t, \gamma(t)) \cdot \nabla \varphi(t, \gamma(t))\} dt = \int_0^1 \int \{\partial_t \varphi(t, x) + v(t, x) \cdot \nabla \varphi(t, x)\} d\rho(t, x) dt = 0.$$

□

2.2 In the framework of Cauchy-Lipschitz

In this subsection, we assume that v satisfies the conditions necessary to apply the Cauchy-Lipschitz theorem, that is, for almost all t , $v(t, \cdot)$ is Lipschitz continuous, and

$$\int_0^1 \|v(t, \cdot)\|_{W^{1,\infty}} dt < +\infty. \quad (6)$$

Under this assumption, we have:

Theorem 9 (Cauchy-Lipschitz). *Let v satisfy assumption (6). Then there exists a unique flow $\phi : [0, 1] \times [0, 1] \times \mathbb{T}^d \rightarrow \mathbb{T}^d$ that satisfies for all $s \in [0, 1]$ and $x \in \mathbb{T}^d$:*

- $\phi(s; s, x) = x$,
- *distributionally and for almost every t ,*

$$\frac{d}{dt}\phi(t; s, x) = v(t, \phi(t; s, x)).$$

In addition, for all $s, t, u \in [0, 1]$, the flow identity holds, that is, for all $x \in \mathbb{T}^d$, $\phi(u; t, \phi(t; s, x)) = \phi(u; s, x)$. In particular, for all $s, t \in [0, 1]$, $\phi(t; s, \cdot)$ is invertible, and its invert is $\phi(s; t, \cdot)$.

In this case, the following holds for the continuity equation.

Theorem 10. *Let v satisfy assumption (6) and $\rho_0 \in \mathcal{P}(\mathbb{T}^d)$. There exists a unique measurable curve $t \in [0, 1] \mapsto \rho(t) \in \mathcal{P}(\mathbb{T}^d)$ such that $\rho(0) = \rho_0$ and (ρ, v) is a solution of the continuity equation in the sense of Definition 6, and it is given for all $t \in [0, 1]$ by the formula*

$$\rho(t) = \phi(t; 0, \cdot) \# \rho_0, \quad (7)$$

where ϕ is the Cauchy-Lipschitz flow given by Theorem 9.

Proof. We first show that formula (7) gives rise to a curve ρ such that (ρ, v) is a solution of the continuity equation. The continuity of this ρ comes from the continuity of ϕ , and the integrability (4) is an easy consequence of (6). According to Remark 7, it suffices to show that for all $\varphi \in C^1(\mathbb{T}^d)$, the map $t \mapsto \int \varphi d\rho(t)$ has distributional derivative given by

$$\frac{d}{dt} \int \varphi d\rho(t) = \int \nabla \varphi(x) \cdot v(t, x) d\rho(t, x).$$

Due to the definition of ρ , this rewrites

$$\frac{d}{dt} \int \varphi(\phi(t; 0, x)) d\rho_0(x) = \int \nabla \varphi(\phi(t; 0, x)) \cdot v(t, \phi(t; 0, x)) d\rho_0(x).$$

But for all $x \in \mathbb{T}^d$, the map $t \mapsto \phi(t; 0, x)$ is of regularity $W^{1,1}$, so as φ is C^1 , by classical chain rules in Sobolev spaces (see [4, Corollary VIII.10]), $t \mapsto \varphi(\phi(t; 0, x))$ is of regularity $W^{1,1}$ as well, and distributionally

$$\frac{d}{dt} \varphi(\phi(t; 0, x)) = \nabla \varphi(\phi(t; 0, x)) \cdot \partial_t \phi(t; 0, x) = \nabla \varphi(\phi(t; 0, x)) \cdot v(t, \phi(t; 0, x)).$$

Our claim is just an integrated version of this identity, and hence is a consequence of the Fubini theorem.

It remains to prove uniqueness, which is the hardest part of the proof. So we give ourselves a curve ρ such that $\rho(0) = \rho_0$ and (ρ, v) is a solution of the continuity equation. We want to prove that for all $\varphi \in C(\mathbb{T}^d)$ and $t \in [0, 1]$, $\int \varphi d\rho(t) = \int \varphi(\phi(t; 0, x)) d\rho_0(x)$. Changing the variables by calling $\xi := \varphi \circ \phi(t; 0, \cdot)$ and using the invertibility of $\phi(t; 0, \cdot)$, this is the same as proving that for all $\xi \in C(\mathbb{T}^d)$ and for all $t \in [0, 1]$, $\int \xi(\phi(0; t, \cdot)) d\rho(t) = \int \xi d\rho_0$. At that point, we can obviously restrict ourselves to smooth $\xi \in C^1(\mathbb{T}^d)$,

and as the claim is clearly true at $t = 0$, the only thing we have to prove is that for all $\xi \in C^1(\mathbb{T}^d)$, the distributional derivative of the map $t \mapsto \int \xi(\phi(0; t, \cdot)) d\rho(t)$ is 0, that is, for all smooth $\chi \in C_c^\infty(0, 1)$,

$$\int_0^1 \chi'(t) \left(\int \xi(\phi(0; t, x)) d\rho(t, x) \right) dt = 0. \quad (8)$$

Observe the following estimate, valid for all v, v' satisfying (6), of associated flow ϕ and ϕ' , whose proof left to the reader as an exercise. For all t :

$$\|\phi'(0; t, \cdot) - \phi(0; t, \cdot)\|_\infty \leq \int_0^t \|v'(\tau, \cdot) - v(\tau, \cdot)\|_\infty d\tau \times \exp\left(\int_0^t \text{Lip}(v(\tau, \cdot)) d\tau\right).$$

With this estimate, we can clearly restrict ourselves to smooth v , and then use a density argument in $L_t^1(W_x^{1, \infty})$ to pass to the limit in (8).

So we are left to proving the result in the case when $v(t, x)$ is smooth w.r.t. both t and x . In that case, by classical results in the theory of ODEs, ϕ is smooth with respect to its three coordinates. Differentiating w.r.t. s the identity holding for all $x \in \mathbb{T}^d$

$$\phi(0; s, \phi(s; 0, x)),$$

we find

$$\partial_s \phi(0; s, \phi(s; 0, x)) + D_x \phi(0; s, \phi(s; 0, x)) \cdot v(s, \phi(s; 0, x)).$$

Therefore, for all $y \in \mathbb{T}^d$, applying this identity at $x = \phi(0; s, y)$, we find

$$\partial_s \phi(0; s, y) + D_x \phi(0; s, y) \cdot v(s, y). \quad (9)$$

In that case, for all $\xi \in C^1(\mathbb{T}^d)$ and all $x \in \mathbb{T}^d$, the map $\varphi : (t, x) \mapsto \xi(\phi(0; t, x))$ is an actual C^1 function, and for all $t \in [0, 1]$ and $x \in \mathbb{T}^d$, using (9), we find

$$\begin{aligned} & \partial_t \varphi(t, x) + v(t, x) \cdot \nabla \varphi(t, x) \\ &= -\nabla \xi(\phi(0; t, x)) \cdot \left(D_x \phi(0; t, x) \cdot v(t, x) \right) + v(t, x) \cdot \left({}^t D_x \phi(0; t, x) \cdot \nabla \xi(\phi(0; t, x)) \right) = 0. \end{aligned}$$

Now, for a given $\chi \in C_c^1((0, 1) \times \mathbb{T}^d)$, it remains to use (5) to the $C_c^1((0, 1) \times \mathbb{T}^d)$ given by $(t, x) \mapsto \chi(t)\varphi(t, x)$ in order to find

$$\int_0^1 \int \left\{ \chi'(t)\varphi(t, x) + \chi(t) \left(\partial_t \varphi(t, x) + v(t, x) \cdot \nabla \varphi(t, x) \right) \right\} d\rho(t, x) dt = 0,$$

and because of the formula above,

$$\int_0^1 \int \chi'(t)\varphi(t, x) d\rho(t, x) dt = 0,$$

which is a rewriting of (8). □

2.3 Generalized flows

An interpretation of the previous result is that at least when v has enough regularity, the unique curve ρ such that (ρ, v) is a solution of the continuity equation is the macroscopic distribution of a population of particles following the flow of the ODE induced by v . We would like to get such an interpretation for any pair (ρ, v) of solution of the continuity equation. As it is always possible to regularize v , this question reduces to know which notion of flow is stable when a family of smooth (v_n) converges towards a singular v , only satisfying (4). Actually, due to the lack of regularity of the solutions of an ODE associated with a singular velocity, convergence cannot hold in the space of flows, and we need to introduce a new notion, that we call generalized flows.

Definition 11 (Generalized flows). A generalized flow is a probability measure $P \in \mathcal{P}(C([0, 1]; \mathbb{T}^d))$. In order to lighten the notations, in the following, we call $\Omega := C([0, 1]; \mathbb{T}^d)$, and for all $t \in [0, 1]$, $X_t : \omega \in \Omega \mapsto \omega(t) \in \mathbb{T}^d$ the evaluation map.

These flows come with some action functionals, which is a quantity analogous to the action defined in (4) for solutions of the continuity equation, but in the space of generalized flows.

Definition 12 (Action of a curve, action of a flow). On the set Ω , we define the following functional:

$$A : \omega \in \Omega \mapsto \begin{cases} \frac{1}{2} \int_0^1 |\dot{\omega}(t)|^2 dt, & \text{if } \omega \in H^1([0, 1]; \mathbb{T}^d), \\ +\infty, & \text{otherwise.} \end{cases}$$

Then, we define the action of a generalized flow $P \in \mathcal{P}(\Omega)$ by

$$\mathcal{A}(P) := \int A(\omega) dP(\omega) \in [0, +\infty].$$

These quantities have good properties for minimization problems, that lets us hope for good stability results.

Lemma 13. *The functional A has compact sublevels in Ω (endowed with the topology of uniform convergence). Consequently, \mathcal{A} has compact sublevels for the topology of narrow convergence.*

Proof. For the first point, let us show that $\mathcal{K}_M := \{\omega \in \Omega \text{ s.t. } A(\omega) \leq M\}$ is compact in Ω , $M > 0$ being given. For all ω in this set and $0 \leq s \leq t \leq 1$, we have

$$d(\omega(s), \omega(t)) \leq \int_s^t |\dot{\omega}(\tau)| d\tau \leq 2M\sqrt{t-s}.$$

So our set \mathcal{K}_M is uniformly equicontinuous, and hence precompact for the topology of uniform convergence in virtue of the Ascoli-Arzelà theorem. Let us show that it is closed. Let $(\omega_n) \in \mathcal{K}_M^{\mathbb{N}}$ a sequence uniformly converging towards ω , and let us show that $\omega \in \mathcal{K}_M$. The family $(\dot{\omega}_n)$ is bounded in $L^2([0, 1]; \mathbb{R}^d)$, so it has a weak limit α in L^2 , and $\|\alpha\|_{L^2} \leq \liminf_n \|\dot{\omega}_n\|_{L^2}$, because the norm is lower-semicontinuous with respect to weak convergence. But of course, by continuity of the derivation in the topology of distribution, $\alpha = \dot{\omega}$, and hence $\omega \in H^1$ and $A(\omega) \leq \liminf_n A(\omega_n) \leq M$.

For the second point, given $C > 0$, let us show that the set $\tilde{\mathcal{K}}_C := \{P \in \mathcal{P}(\Omega) \text{ s.t. } \mathcal{A}(P) \leq C\}$ is compact for the topology of narrow convergence. First, it is tight. Indeed, if $P \in \tilde{\mathcal{K}}_C$, we have for all $M > 0$

$$P(\mathcal{K}_M^c) \leq \frac{C}{M}$$

which can be made arbitrarily small by taking M arbitrarily large. As \mathcal{K}_M is compact by the first point, the result follows. It remains to prove that $\tilde{\mathcal{K}}_C$ is closed for the topology of narrow convergence. So let us consider $(P_n) \in \tilde{\mathcal{K}}_C^{\mathbb{N}}$ a sequence narrowly converging towards $P \in \mathcal{P}(\Omega)$, and let us show that $P \in \tilde{\mathcal{K}}_C$. As we have seen that A is l.s.c., for all $\varepsilon > 0$, there exists $A' \leq A$ bounded and continuous on Ω such that

$$\int A' dP \geq \mathcal{A}(P) - \varepsilon.$$

But by narrow convergence,

$$\int A' dP = \lim_n \int A' dP_n \leq \mathcal{A}(P_n).$$

By letting ε tend to 0, we conclude that

$$\mathcal{A}(P) \leq \liminf_n \mathcal{A}(P_n),$$

and the result follows. □

ODE flows together with initial distributions induce generalized flows, as developed in the following example.

Example 14. Let v satisfy (6), ϕ be the associated flow given by Theorem 9, and $\rho_0 \in \mathcal{P}(\mathbb{T}^d)$. Let us call Ψ the following map

$$\begin{aligned}\Psi : \mathbb{T}^d &\rightarrow \Omega \\ x &\mapsto \left(t \mapsto \phi(t; 0, x) \right).\end{aligned}$$

Then $P := \Psi_{\#}\rho_0$ is a generalized flow. Moreover, for all $t \in [0, 1]$,

$$\rho(t) := X_{t\#}P = X_{t\#}(\Psi_{\#}\rho_0) = (X_t \circ \Psi)_{\#}\rho_0 = \phi(t; 0, \cdot)_{\#}\rho_0.$$

Therefore, the curve $(t \mapsto X_{t\#}P)$ is nothing but the curve ρ for which (ρ, v) is the unique solution of the continuity equation, as described in Theorem 10.

In that case, for P -almost every ω , ω satisfies

$$\dot{\omega}(t) = v(t, \omega(t)) \tag{10}$$

distributionally and for almost every t , and

$$\mathcal{A}(P) = \frac{1}{2} \int \left(\int_0^1 |v(t, \phi(t; 0, x))|^2 dt \right) d\rho_0(x) = \frac{1}{2} \int_0^1 \int |v(t, x)|^2 d\rho(t, x) dt.$$

Now, we will see how we can generalize this example when v has less regularity than (6). First of all, we can show that *any* generalized flows induces a solution of the continuity equation.

Proposition 15. *Let $P \in \mathcal{P}(\Omega)$ with $\mathcal{A}(P) < +\infty$. For all $t \in [0, 1]$, let us call $\rho(t) := X_{t\#}P$. The vector-valued measure m on $[0, 1] \times \mathbb{T}^d$ acting on all $\xi : [0, 1] \times \mathbb{T}^d \rightarrow \mathbb{R}^d$ measurable and bounded as*

$$\int_0^1 \int \xi \cdot dm = \int \left\{ \int_0^1 \xi(t, \omega(t)) \cdot \dot{\omega}(t) dt \right\} dP(\omega) \tag{11}$$

is well defined, and is absolutely continuous w.r.t. $dt \otimes \rho(t)$.

Calling v the corresponding Radon-Nikodym derivative, (ρ, v) is a solution of the continuity equation in the sense of Definition 6, and

$$\frac{1}{2} \int_0^1 \int |v(t, x)|^2 d\rho(t, x) dt \leq \mathcal{A}(P). \tag{12}$$

Remark 16. In this case, the definition (11) of m does not insure (10) to hold. Actually, this is not true in general. We could see that (11) has the consequence that for $dt \otimes \rho(t)$ -almost every (t, x) , $v(t, x)$ is the conditional expectation of $\omega \mapsto \dot{\omega}(t)$ (which is well defined for almost every t in $L^2(P)$ thanks to $\mathcal{A}(P)$) knowing $X_t(\omega) = x$. Hence, a way to understand (12) is through Jensen's inequality. Although, our proof will be different.

In order to prove Proposition 15, we will need the following lemma.

Lemma 17. *Let $\rho : [0, 1] \rightarrow \mathcal{P}(\mathbb{T}^d)$ be a measurable curve in the space of probability measures over \mathbb{T}^d and m a vector-valued Radon measure on $[0, 1] \times \mathbb{T}^d$. We have*

$$\sup_{\xi \in C([0, 1] \times \mathbb{T}^d; \mathbb{R}^d)} \int_0^1 \int \xi \cdot dm - \frac{1}{2} \int_0^1 \int |\xi|^2 d\rho(t) dt = \begin{cases} \frac{1}{2} \int_0^1 \int |v|^2 d\rho dt, & \text{if } m \ll \rho(t) \otimes dt \text{ and } m = v\rho(t) \otimes dt, \\ +\infty, & \text{otherwise.} \end{cases} \tag{13}$$

Proof. We first prove " \leq ". If m is absolutely continuous w.r.t. $\rho(t) \otimes dt$ with Radon-Nikodym derivative v , it just follows from the inequality $|v|^2/2 \geq \xi \cdot v - |\xi|^2/2$, and otherwise, there is nothing to prove.

Let us now prove " \geq ". If m is not absolutely continuous w.r.t. $\rho(t) \otimes dt$, then for all, there exists ξ continuous such that $\iint \xi \, dm - \frac{1}{2} \iint |\xi|^2 \, d\rho(t) \, dt > 0$ (approximate $v1_A$ for some well chosen $v \in \mathbb{R}^d$ and $A \subset [0, 1] \times \mathbb{T}^d$ measurable satisfying $\iint_A v \cdot dm > 0$ and $\rho(t) \otimes dt(A) = 0$). Then the result follows from taking $\lambda \xi$ for $\lambda \rightarrow +\infty$. If $m \ll \rho(t) \otimes dt$, of Radon-Nikodym derivative v , and if $\iint |v|^2 \, d\rho \, dt < +\infty$, this is simply the density of continuous functions in L^2 . Finally, if $m \ll \rho(t) \otimes dt$, of Radon-Nikodym derivative v , but $\iint |v|^2 \, d\rho \, dt = +\infty$, then one needs to use the density of continuous function in L^2 on the truncated function $v1_{|v| \leq M}$ and let M tend to $+\infty$. \square

We are now ready to prove Proposition 15.

Proof of Proposition 15. Step 1. $t \mapsto \rho(t)$ is narrowly continuous.

It suffices to show that for all $\varphi \in C_b(\mathbb{R}^d)$, the map

$$t \mapsto \int \varphi \, d\rho(t) = \int \varphi(\omega(t)) \, dP(\omega)$$

is continuous. This is a direct consequence of the dominated convergence theorem.

Step 2. m is well defined.

It is well defined because for all ξ measurable and bounded

$$\begin{aligned} \left| \int \left\{ \int_0^1 \xi(t, \omega(t)) \cdot \dot{\omega}(t) \, dt \right\} dP(\omega) \right| &\leq \int \left| \int_0^1 \xi(t, \omega(t)) \cdot \dot{\omega}(t) \, dt \right| dP(\omega) \\ &\leq \sup |\xi| \int \int_0^1 |\dot{\omega}(t)| \, dt \, dP(\omega) \\ &\leq \sup |\xi| \underbrace{\int \int_0^1 |\dot{\omega}(t)|^2 \, dt \, dP(\omega)}_{< +\infty}. \end{aligned}$$

In particular, this formula defines a continuous form on the space $C_c([0, 1] \times \mathbb{R}^d)$, that is, a Radon measure m on $[0, 1] \times \mathbb{T}^d$.

Step 3. $m \ll \rho(t) \otimes dt$ and $v := \frac{dm}{d\rho(t) \otimes dt}$ satisfies (12).

We use Lemma 17. Let $\xi \in C([0, 1] \times \mathbb{T}^d; \mathbb{R}^d)$. By definition of ρ and m , we have

$$\begin{aligned} \int_0^1 \int \xi \cdot dm - \frac{1}{2} \int_0^1 \int |\xi|^2 \, d\rho(t) \, dt &= \int \left\{ \int_0^1 \xi(t, \omega(t)) \cdot \dot{\omega}(t) \, dt \right\} dP(\omega) - \frac{1}{2} \int_0^1 \int |\xi(t, \omega(t))|^2 \, dP(\omega) \, dt \\ &= \int \left\{ \int_0^1 \left(\xi(t, \omega(t)) \cdot \dot{\omega}(t) - \frac{1}{2} |\xi(t, \omega(t))|^2 \right) dt \right\} dP(\omega) \\ &\leq \mathcal{A}(P). \end{aligned}$$

So the result directly follows from Lemma 17.

Step 4. Distributional solution.

The proof relies on the following chain rule: whenever $\omega \in H^1([0, 1]; \mathbb{R}^d)$ and $\varphi \in C_c^1((0, 1) \times \mathbb{R}^d)$, then $(t \mapsto \varphi(t, \omega(t))) \in H^1([0, 1])$, and its distributional derivative is equal to

$$\partial_t \varphi(t, \omega(t)) + \nabla \varphi(t, \omega(t)) \cdot \dot{\omega}(t)$$

for dt -almost every t (use some density argument). Therefore, for all $\varphi \in C_c^1((0, 1) \times \mathbb{R}^d)$, integrating the previous formula between 0 and 1, we find that for all $\omega \in H^1([0, 1]; \mathbb{R}^d)$,

$$\int_0^1 \{ \partial_t \varphi(t, \omega(t)) + \nabla \varphi(t, \omega(t)) \cdot \dot{\omega}(t) \} dt = 0.$$

Integrating w.r.t. P (which is valid since P only charges absolutely H^1 curves), we find

$$\iint_0^1 \{\partial_t \varphi(t, \omega(t)) + \nabla \varphi(t, \omega(t)) \cdot \dot{\omega}(t)\} dt dP(\omega) = 0.$$

Now, by definition of ρ and m , this means

$$\int_0^1 \int \partial_t \varphi d\rho dt + \int_0^1 \int \nabla \varphi \cdot dm = 0.$$

Finally, writing $m = v\rho dt$, we find the expected result. □

2.4 All solutions are induced by generalized flows

The purpose of the rest of the section is to give a kind of converse of Proposition 15: *Every solution of the continuity equation is induced by a general flow P .*

The theorem states as follows.

Theorem 18. *Let $\rho : [0, 1] \rightarrow \mathcal{P}(\mathbb{T}^d)$ be a measurable curve and $v : [0, 1] \times \mathbb{T}^d \rightarrow \mathbb{R}^d$ be a measurable vector field. If the pair (ρ, v) is a solution of the continuity equation in the sense of Definition 6, then there exists $P \in \mathcal{P}(\Omega)$ such that:*

- For all $t \in [0, 1]$, $\rho(t) = X_{t\#}P$.
- The following estimate holds:

$$\mathcal{A}(P) \leq \frac{1}{2} \int_0^1 \int |v(t, x)|^2 d\rho(t, x) dt. \tag{14}$$

Remark 19. We could go further and prove that actually, for the P built in the proof below, P -almost every ω satisfies

$$\dot{\omega}(t) = v(t, \omega(t))$$

dt -almost everywhere and distributionally. In particular, with our construction, the solution of the continuity equation associated to P given by Proposition 15 is the pair (ρ, v) we started from, and (14) is in fact an equality. This additional fact is a bit technical, and is not necessary to prove the Benamou-Brenier formula Theorem 20 below, so we decided to skip it. For a complete proof, see [1, Theorem 3.4]. Note that our construction allows to build solutions to the ODE associated with v even when v is not smooth! The price to pay is that there is no uniqueness, and that we have solution only for almost every initial condition.

Proof of Theorem 18. The proof goes through regularization. So we consider $(\tau_\varepsilon)_{\varepsilon>0}$ a regularization kernel that is positive everywhere on \mathbb{T}^d (think of the heat kernel). For all $t \in [0, 1]$ and $\varepsilon > 0$, we define

$$\rho^\varepsilon(t) := \tau_\varepsilon * \rho(t).$$

This is a continuous function in t and x , smooth w.r.t. x , uniformly w.r.t. t . For all $t \in [0, 1]$ where $v(t, \cdot) \in L^2(\rho(t))$ (this is dt -almost everywhere the case, we write $t \in \mathcal{T}$), we define

$$v^\varepsilon(t, \cdot) = \frac{\tau_\varepsilon * (v(t, \cdot)\rho(t))}{\rho^\varepsilon(t)}.$$

Observe that with this regularization, $(\rho^\varepsilon, v^\varepsilon)$ is still a solution of the continuity equation. Notice also that for all $t \in \mathcal{T}$ and all $\varepsilon > 0$, v^ε is well define (because ρ^ε is positive everywhere) and smooth. In particular, as a function of t and x , it is easy to check that v^ε satisfies (6). Let us therefore define ϕ^ε the flow given by Theorem 9, and P^ε as defined in Example 14.

By Theorem 10 and the development given in Example 14, for all $t \in [0, 1]$, $\rho^\varepsilon(t) = X_{t\#}P^\varepsilon$. On the other hand, we can derive a uniform bound for $\mathcal{A}(P^\varepsilon)$. Indeed, calling $m^\varepsilon := v^\varepsilon \rho^\varepsilon(t) \otimes dt$, we have

$$\begin{aligned} \mathcal{A}(P^\varepsilon) &= \frac{1}{2} \iint_0^1 |v^\varepsilon(t, x)|^2 dt d\rho^\varepsilon(t, x) \\ &= \sup_{\xi \in C([0,1] \times \mathbb{T}^d; \mathbb{R}^d)} \int_0^1 \int \xi \cdot dm^\varepsilon - \frac{1}{2} \int_0^1 \int |\xi|^2 d\rho^\varepsilon(t) dt \\ &= \sup_{\xi \in C([0,1] \times \mathbb{T}^d; \mathbb{R}^d)} \int_0^1 \int \xi * \tau_\varepsilon \cdot dm - \frac{1}{2} \int_0^1 \int |\xi|^2 * \tau_\varepsilon d\rho(t) dt, \end{aligned}$$

where $m := v\rho dt$. By Jensen's inequality, we have everywhere for all ξ , $|\xi|^2 * (\tau_\varepsilon) \geq |\xi * \tau_\varepsilon|^2$, so

$$\begin{aligned} \mathcal{A}(P^\varepsilon) &\leq \sup_{\xi \in C([0,1] \times \mathbb{T}^d; \mathbb{R}^d)} \int_0^1 \int \xi * \tau_\varepsilon \cdot dm - \frac{1}{2} \int_0^1 \int |\xi * \tau_\varepsilon|^2 d\rho(t) dt \\ &\leq \sup_{\xi \in C([0,1] \times \mathbb{T}^d; \mathbb{R}^d)} \int_0^1 \int \xi \cdot dm - \frac{1}{2} \int_0^1 \int |\xi|^2 d\rho(t) dt \\ &= \frac{1}{2} \int_0^1 \int |v^2(t, x)| d\rho(t, x) dt. \end{aligned}$$

Therefore, in virtue of Lemma 13 (P^ε) admits a limit point P in the narrow topology, that satisfies

$$\mathcal{A}(P) \leq \frac{1}{2} \int_0^1 \int |v(t, x)|^2 d\rho(t, x) dt.$$

The fact that for all $t \in [0, 1]$, $X_{t\#}P = \rho(t)$ is just a consequence of the fact that $X_t : \Omega \rightarrow \mathbb{T}^d$ is continuous so we can take the limit in $X_{t\#}P^\varepsilon = \rho^\varepsilon(t)$. \square

3 Benamou-Brenier formula

The Benamou-Brenier formula is the following theorem.

Theorem 20. *Let $\mu, \nu \in \mathcal{P}(\mathbb{T}^d)$. We have*

$$\frac{1}{2} W_2^2(\mu, \nu) = \inf \left\{ \frac{1}{2} \int_0^1 \int |v|^2 d\rho dt \mid (\rho, v) \text{ solution of the continuity equation with } \rho(0) = \mu \text{ and } \rho(1) = \nu \right\}.$$

Moreover, both inf are achieved.

This theorem will be seen as an immediate consequence of the following proposition which states an equivalence between three optimization problems.

Proposition 21. *Given $\mu, \nu \in \mathcal{P}(\mathbb{T}^d)$, we define*

$$\begin{aligned} \mathcal{CE}(\mu, \nu) &:= \left\{ (\rho, v) \text{ solution of the continuity equation with } \rho(0) = \mu \text{ and } \rho(1) = \nu \right\}, \\ \mathcal{GF}(\mu, \nu) &:= \left\{ P \in \mathcal{P}(\Omega) \text{ with } X_{0\#}P = \mu \text{ and } X_{1\#}P = \nu \right\}, \\ C_f &: \gamma \in \Pi(\mu, \nu) \mapsto \frac{1}{2} \int d(x, y)^2 d\gamma(x, y), \\ C_s &: (\rho, v) \in \mathcal{CE}(\mu, \nu) \mapsto \frac{1}{2} \int_0^1 \int |v|^2 d\rho(t) dt, \\ C_f &: P \in \mathcal{GF}(\mu, \nu) \mapsto \mathcal{A}(P). \end{aligned}$$

There are maps $T_{fp} : \Pi(\mu, \nu) \rightarrow \mathcal{GF}(\mu, \nu)$, $T_{pf} : \mathcal{GF}(\mu, \nu) \rightarrow \Pi(\mu, \nu)$, $T_{fc} : \mathcal{CE}(\mu, \nu) \rightarrow \mathcal{GF}(\mu, \nu)$ and $T_{cf} : \mathcal{GF}(\mu, \nu) \rightarrow \mathcal{CE}(\mu, \nu)$ such that

$$C_p \circ T_{fp} \leq C_f, \quad C_f \circ T_{fp} \leq C_p, \quad C_c \circ T_{cf} \leq C_f, \quad C_f \circ T_{fc} \leq C_c.$$

As a consequence, considering the three following problems:

$$\begin{aligned} (OT) \quad & \frac{1}{2}W_2^2(\mu, \nu) := \inf \left\{ C_p(\gamma) \mid \gamma \in \Pi(\mu, \nu) \right\}, \\ (BB) \quad & I(\mu, \nu) := \inf \left\{ C_c(\rho, \nu) \mid (\rho, \nu) \in \mathcal{CE}(\mu, \nu) \right\}, \\ (GF) \quad & J(\mu, \nu) := \inf \left\{ C_f(P) \mid P \in \mathcal{GF}(\mu, \nu) \right\}, \end{aligned}$$

the three inf coincide, are achieved, and the maps T_{fp} , T_{pf} , T_{fc} and T_{cf} send optimizers to optimizers.

Proof. For T_{cf} , it suffices to take the map described in Proposition 15. For T_{cf} , it suffices to take the map built in Theorem 18. The corresponding inequalities are already proved in those results.

We choose $T_{pf} := (X_0, X_1)$. The inequality $C_p \circ T_{pf} \leq C_f$ is a direct consequence of

$$\forall \omega \in \Omega, \quad \frac{1}{2}d(\omega(0), \omega(1))^2 \leq \frac{1}{2} \left(\int_0^1 |\dot{\omega}(t)| dt \right)^2 \leq A(\omega).$$

where $\int_0^1 |\dot{\omega}(t)| dt$ is set to $+\infty$ if $\omega \notin W^{1,1}$.

Finally, for defining T_{fp} , we choose any measurable map $\psi : \mathbb{T}^d \times \mathbb{T}^d \rightarrow \Omega$ sending any (x, y) to a geodesic between x and y , and we define for all $T_{fp} : \gamma \in \Pi(\mu, \nu) \mapsto \psi_{\#}\gamma$. This time, by definition of a geodesic, for all $(x, y) \in \mathbb{T}^d \times \mathbb{T}^d$,

$$\frac{1}{2}d(x, y)^2 = A(\psi(x, y)),$$

and $C_f \circ T_{fp} \leq C_p$ follows, actually with an equality.

The fact that the three inf coincide is an obvious consequence, the fact that solutions are sent onto solutions as well, and therefore the fact that the inf are achieved follows from the existence of a solution of (OT). \square

Remark 22. We saw in the proof that with our choices, $C_f \circ T_{fp} = C_p$. Also, we easily see that $T_{pf} \circ T_{fp} = \text{Id}$. That means that the optimization problem (OT) is a projection of the optimization problem (GF).

The problem (BB) is also a projection of (GF). Indeed, as a consequence of Remark 19, we also have $C_f \circ T_{fc} = C_c$ and $T_{cf} \circ T_{fc} = \text{Id}$.

As an exercise, the interested reader can show that (OT) is a projection of (BB).

The problem with the most information is henceforth (GF), followed by (BB), followed by (OT).

Remark 23. It is possible to prove that both I and J in the statement of Proposition 21 are achieved using the direct method of calculus of variations (that is, showing that minimizing sequences converge towards optimizers).

4 Interpretation, additional information

Let us make more precise the analogy between $\mathcal{P}(\mathbb{T}^d)$ and a Riemannian manifold. A Riemannian manifold is a manifold M equipped with a bilinear form on its tangent spaces, that is, for all $p \in M$, there is a bilinear form $g_p : T_p M \times T_p M$ (we do not talk about regularity assumptions).

Here, a priori, we see the set of vector fields in $L^2(\rho)$ as the formal tangent space of $\mathcal{P}(\mathbb{T}^d)$ at point $\rho \in \mathcal{P}(\mathbb{T}^d)$, and what plays the role of g is simply the L^2 norm, that is, for all $\rho \in \mathcal{P}(\mathbb{T}^d)$ and $v, w \in L^2(\rho)$,

$$g_\rho(v, w) := \int v \cdot w d\rho.$$

. The reason behind this idea is that at least formally, whenever we have a map $\rho \in \mathcal{P}(\mathbb{T}^d) \mapsto v(\rho) \in L^2(\rho)$, we can "integrate" the "vector field" by finding the curves ρ solving the continuity equation

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

In that sense, the continuity equation plays the role of ODEs in the classical theory.

It can be frustrating with this idea that starting just from a curve ρ , it is not clear what is the velocity of ρ at time t . Another annoying observation is that it is perfectly possible that two "vector fields" $v(\rho)$ and $w(\rho)$ induce the same solutions of the continuity equation: it suffices that for all ρ ,

$$\operatorname{div}(\rho v(\rho)) = \operatorname{div}(\rho w(\rho)).$$

These two obstructions are solved by the following theorem that we will not prove (see [5, Theorem 5.14] for some very similar result):

Theorem 24. *Let $\rho : t \in [0, 1] \mapsto \rho(t) \in \mathcal{P}(\mathbb{T}^d)$. The curve ρ satisfies the following bound¹*

$$\sup_{\substack{0=t_0 < \dots < t_n=1, \\ n \in \mathbb{N}^*}} \sum_{k=1}^n \frac{W_2^2(\rho(t_k), \rho(t_{k-1}))}{t_k - t_{k-1}}$$

if and only if there exists v such that (ρ, v) is a solution of the continuity equation in the sense of Definition 6. Moreover,

$$\sup_{\substack{0=t_0 < \dots < t_n=1, \\ n \in \mathbb{N}^*}} \sum_{k=1}^n \frac{W_2^2(\rho(t_k), \rho(t_{k-1}))}{t_k - t_{k-1}} = \inf \left\{ \int_0^1 \int |v|^2 d\rho dt \mid v \text{ s.t. } (\rho, v) \text{ solves the continuity equation} \right\},$$

the inf is achieved uniquely in $L^2(\rho(t) \otimes dt)$, and for almost every t , $v(t, \cdot)$ is in the closure of the sets of gradients

$$\left\{ \nabla \varphi \mid \varphi \in C^\infty(\mathbb{T}^d) \right\}$$

in the $L^2(\rho(t))$ topology.

In particular, this theorem shows that a posteriori, to get a better analogy with classical Riemannian manifold, it is better to see the tangent of $\mathcal{P}(\rho)$ at point ρ as the set of gradients (or more precisely, as its closure in $L^2(\rho)$). For the following, we can put aside this remark, and be confident that every sufficiently regular curves are indeed solutions of the continuity equation, so that working with solutions of the continuity equation is not restrictive.

In classical Riemannian geometry, the geodesics between points $p, p' \in M$ are the optimizers of the quantity

$$\inf_{\substack{\omega \in H^1([0,1]; M) \\ \omega(0)=p, \omega(1)=p'}} \int_0^1 |\dot{\omega}(t)|^2 dt.$$

This is exactly replaced in our context by the problem

$$\inf \left\{ \frac{1}{2} \int_0^1 \int |v|^2 d\rho dt \mid (\rho, v) \text{ solution of the continuity equation with } \rho(0) = \mu \text{ and } \rho(1) = \nu \right\},$$

whose solutions are therefore called geodesics in the Wasserstein space $(\mathcal{P}(\mathbb{T}^d), W_2)$.

On a Riemannian manifold (M, g) , there is a particular type of ODEs: the gradient flows. The idea is the following. Any smooth function $U : M \rightarrow \mathbb{R}$ admits a gradient at any point $p \in M$. This is the only vector $\nabla U(p) \in T_p M$ such that for all $v \in T_p(M)$, $DU(p)(v) = g_p(\nabla U(p), v)$, where $DU(p) : T_p M \rightarrow \mathbb{R}$ is

¹We say that $\rho \in AC^2([0, 1]; \mathcal{P}(\mathbb{T}^d))$.

the differential of U at p . In particular, the gradient of U is characterized by the fact that for all smooth curve ω and for all t ,

$$\frac{d}{dt}U(\omega(t)) = g_{\omega(t)}(\nabla U(\omega(t)), \dot{\omega}(t)).$$

The gradient flow (or steepest descent) associated to U is then the ODE

$$\dot{\omega}(t) = -\nabla U(\omega(t)).$$

It had the very useful property that if ω is a solution, then $t \mapsto U(\omega(t))$ is a nonincreasing function.

Here we can play the same game. We can for instance try to understand the "gradient" of the following functional on $\mathcal{P}(\mathbb{T}^d)$:

$$U : \rho \in \mathcal{P}(\mathbb{T}^d) = \int V d\rho + \frac{1}{2} \int W(x-y) d\rho(y) d\rho(x) + \int f\left(\frac{d\rho}{dx}(x)\right) dx,$$

where $V, W : \mathbb{T}^d \rightarrow \mathbb{R}$ and $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ are smooth, W is even, and where the last integral is set to $+\infty$ if ρ is not absolutely continuous w.r.t. dx . From now on, we do not differentiate anymore between ρ and its Radon-Nikodym derivative w.r.t. Lebesgue, so we write $f(\rho(x))$ in place of $f(\frac{d\rho}{dx}(x))$. In the applications, U is the energy associated with a population of particles of distribution ρ , evolving in a potential energy V , having pairwise interaction described by the potential W , and having internal energy described by f . Assume that ρ is a sufficiently regular curve (i.e. AC^2 and everywhere absolutely continuous w.r.t. Lebesgue) and let v be such that (ρ, v) is a solution of the continuity equation. Then a quick computation shows

$$\begin{aligned} \frac{d}{dt}U(\rho(t)) &= \int \nabla V(x) \cdot v(x) d\rho(t) + \int \nabla W(x-y)v(t, x) d\rho(t, y) d\rho(t, x) + \int \nabla(f'(\rho(x))) \cdot v(t, x) d\rho(t, x) \\ &= \int \left(\nabla V(x) + \int \nabla W(x-y) d\rho(t, y) + \nabla f'(\rho(x)) \right) \cdot v(t, x) d\rho(t, x). \end{aligned}$$

Therefore, the "gradient" of U at ρ is the vector field in $L^2(\rho)$ defined by the formula

$$\nabla V + \nabla W * \rho + \nabla f'(\rho).$$

So the curve ρ is a solution of the formal gradient flow associated with U if for all t , the "velocity" of ρ at time t is given by minus this "gradient". Translated in the vocabulary of the continuity equation, this means that ρ solves the PDE

$$\partial_t \rho = \operatorname{div} \left(\left\{ \nabla V + \nabla W * \rho + \nabla f'(\rho) \right\} \rho \right).$$

Example 25. Choose $V = W = 0$ and $f : x \mapsto x \log x$. Then, this PDE rewrites

$$\partial_t \rho = \operatorname{div} \left((\nabla \log \rho) \rho \right) = \Delta \rho.$$

The heat equation is the gradient flow of the entropy functional in the Wasserstein space.

Choose $V = W = 0$ and $f : x \mapsto x^p/(p-1)$, with $p > 1$. We find

$$\partial_t \rho = \frac{p}{p-1} \operatorname{div} \left((\nabla \rho^{p-1}) \rho \right) = \Delta(\rho^p).$$

The porous medium equation is the gradient flow of the power functional in the Wasserstein space.

5 Extensions

We made our analysis in \mathbb{T}^d , but actually a lot of things would be true in \mathbb{R}^d , or even in more general spaces. Let us make a list of contexts where the different results of this lecture are true.

- The Wasserstein distance W^p can be defined on any metric base space (X, d) on the set $\mathcal{P}_p(X)$ of probability measures ρ for which there exists $x_0 \in X$ such that

$$\int d(x_0, x)^p d\rho(x) < +\infty.$$

In that case, this condition holds replacing x_0 by any other point of X .

Whenever X is separable and complete, W_p is a distance [3, Section 7.1]. If $X = \mathbb{R}^d$ or even an infinite dimensional Hilbert space, we have the following equivalence holding for all $(\rho_n) \in \mathcal{P}_p(X)^N$ and $\rho \in \mathcal{P}_p(X)$:

$$W_p(\rho_n, \rho) \xrightarrow{n \rightarrow +\infty} 0 \quad \Leftrightarrow \quad \left(\rho_n \xrightarrow{n \rightarrow +\infty} \rho \text{ narrowly and } \int |x|^p d\rho_n(x) \xrightarrow{n \rightarrow +\infty} \int |x|^p d\rho(x) \right),$$

see [3, Remark 7.1.11].

- The theory of the continuity equation would be almost the same in \mathbb{R}^d . Definition 6 could be written in \mathbb{R}^d instead of \mathbb{T}^d . Then, Theorem 10 would still hold (we could even relax (6) by some local bounds up to adding an assumption of linear growth at infinity for v , see [1, Remark 2.4]). Proposition 15 and Theorem 18 would be the same. In the case of Theorem 18, we could even relax (4), see [1, Theorem 3.4], and some similar relaxation could be made for Proposition 15.
- The Benamou-Brenier Theorem 20 holds true with exactly the same formulation in \mathbb{R}^d using the definition of a solution of the continuity equation as in the previous point, and assuming in addition $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, see [3, Chapter 8]. In this context, Proposition 21 works exactly in the same way.
- The equivalence between the problems (OT) and (GF) works in a much wider context: \mathbb{T}^d could be replaced by any *geodesic* metric space, see [2, Theorem 2.10].

References

- [1] L. Ambrosio and G. Crippa. Continuity equations and ODE flows with non-smooth velocity. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, 144(6):1191–1244, 2014.
- [2] L. Ambrosio and N. Gigli. A user’s guide to optimal transport. In *Modelling and optimisation of flows on networks*, volume 2062 of *Lecture Notes in Math.*, pages 1–155. Springer, Heidelberg, 2013.
- [3] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Springer, 2006.
- [4] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011.
- [5] F. Santambrogio. *Optimal transport for applied mathematicians*. Birkäuser, NY, 2015.