Università di Pisa

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
Corso di Laurea Magistrale in Matematica

Tesi di Laurea Magistrale

Riemannian-like structures on the set of probability measures

A comparison between Euclidean and discrete spaces

Candidate:
Roberto Daluiso

Advisor:
Prof. Luigi Ambrosio

17th September 2012

Contents

Introduction **v**

Notation **vii**

1 Optimal transport **1**
1.1 Statement of the problem . . . 1
1.2 Duality and Brenier's theorem . . . 3
1.3 The Wasserstein distance . . . 10
1.3.1 Absolute continuity in metric spaces . . . 14
1.3.2 Dynamical couplings . . . 17

2 Differential structure of $\mathscr{P}_2(\mathbb{R}^n)$ **23**
2.1 Velocity fields and AC curves . . . 23
2.1.1 The continuity equation . . . 25
2.1.2 Benamou-Brenier's theorem and tangent fields . . . 30
2.2 Differentials in $\mathscr{P}_2(\mathbb{R}^n)$ . . . 38
2.2.1 Potential energy . . . 44
2.2.2 Internal energy . . . 46
2.3 Gradient flows . . . 55

3 A new distance on discrete spaces **63**
3.1 The new distance . . . 63
3.1.1 Setting and continuity equation . . . 63
3.1.2 The 2-point space . . . 66
3.1.3 The n-point space . . . 71
3.2 Riemannian structure . . . 86
3.3 Differentiable functionals and gradient flows . . . 96

4 Conclusions **99**

Bibliography **101**

Acknowledgements **103**

Introduction

Since the late nineties, the theory of optimal transport has been applied to the study of parabolic PDEs (see [11], [14]). The idea is that, if an equation is mass preserving and positivity preserving, then for every $\rho_0 \geq 0$ such that $\int_{\mathbb{R}^n} \rho_0 = 1$, one can interpret ρ_t as the density of a probability measure. It turned out that for some important equations, like the Fokker-Planck equation, the resulting "flow of probability measures" can be thought as the "gradient flow" of a suitable functional. The prototype of this kind of correspondences is the identification of the heat flow in $\mathbb{R}^n$ with the gradient flow of entropy at the level of probability measures: in some sense, heat diffusion tends to decrease entropy "as fast as possible".

To speak properly of gradient flows, however, we need a "Riemannian-like" structure on (a subset of) the space of probability measures $\mathscr{P}(\mathbb{R}^n)$. This is possible in the framework of optimal transport, which provides first of all the "right" distance W_2 between probability measures. Moreover, Benmaou-Brenier's theorem gives an interpretation of $W_2^2(\mu,\nu)$ as the minimum of $\int_0^1 \|v_t\|^2_{L^2(\mu_t)}\,dt$ among the curves joining μ to ν, where v_t is a "tangent" velocity field.

This formulation shares much with the classical definition of Riemannian distance, except for the fact that we are not able to put a differentiable atlas on the space of probability measures, and so we will have to define the "tangent" velocity to a curve (μ_t) as the solution v_t of the continuity equation

$$\frac{d}{dt}\mu_t + \nabla \cdot (v_t \mu_t) = 0 \tag{CE}$$

such that $\|v_t\|_{L^2(\mu_t)}$ is minimal for a.e. t.

As we noted above, the structure which we will put on the space of probability measures is not Riemannian in the proper sense of the term; but it behaves as if it were, allowing for instance a definition of (sub)differential of a functional which retains, in suitable senses, all the good properties of a gradient.

The theory we are speaking about is remarkable for elegance and deepness, but exploits crucially the structure of the underlying space $\mathbb{R}^n$. In particular, it appears to lose sense when $\mathbb{R}^n$ is replaced by a non-geodesic space: most evidently, when the underlying space is discrete. Nevertheless, in 2011, Jan Maas's paper [12] has reproduced on finite spaces many of the results true in the Euclidean setting. In the absence of a differentiable structure on the base space X, Maas uses an irreducible Markov kernel to build the structure of $\mathscr{P}(X)$. The Markov kernel provides both a natural reference measure (the invariant probability measure) with respect to which entropy may be calculated, and a classical definition of heat flow: hence, one can wonder if the heat flow is still the gradient flow of entropy in some sense.

The answer is positive if one defines a new distance $\mathcal{W} \neq W_2$ via a discrete analogous of Benamou-Brenier's formula. In order to do this, the crucial point is to find a proper discrete counterpart of (CE): while the concepts of discrete gradient of a vector and discrete divergence of a matrix arise quite naturally, it is not immediately clear what should play the role of the term $v_t\mu_t$. An explicit computation when X has only two points will suggest the matrix

$$\big(\theta(\rho(x), \rho(y))V_t(x, y)\big)_{x,y\in X},$$

where $\theta(\rho(x), \rho(y))$ may intuitively represent the amount of mass effectively affected by the velocity $V_t(x, y)$. If we want the heat flow to coincide with the gradient flow of entropy, the right θ turns out to be the so-called logarithmic mean, but it is worth developing the theory for quite general functions θ. This time, the resulting structure on the relevant sets of probability measures will be truly Riemannian.

The purpose of this thesis is to to illustrate in a parallel and fairly complete way the two theories. In our intention, the set of Chapters 1 and 2 is a self-contained introduction to the theory of optimal transport and "Otto calculus", complete in its proofs but not indulging in the maximum generality (especially in Chapter 2); while the set of Chapters 2 and 3 is a comparison between the Euclidean and discrete theories introduced above. The arrival point of both chapters is the correspondence between heat flow and gradient flow of entropy.

We tried to follow a parallel order of exposition as far as we could: though a unified treatment seems presently not possible, the correspondence is conceptually transparent, and sometimes even allowed the reproduction of the proofs with minor modifications. We think that one theory can enlighten the other: till now, the Euclidean theory has been the model for the discrete one, but one can imagine that in the future new problems might be firstly approached in the simpler discrete context.

Notation

The meaning of each symbol used in the text, unless otherwise stated, is the one specified in the following list.

LHS, RHS	left hand side, right hand side (of an equation).
l.s.c.	lower semicontinuous.
I_E	indicator function of the set E.
S_n	permutations of $\{1, \dots, n\}$.
$C_0(\mathbb{R}^n)$	continuous functions on $\mathbb{R}^n$ infinitesimal at infinity.
c	lower semicontinuous nonnegative cost function, see Problem 1.1.1.
X, Y	in Chapter 1, Polish spaces; in Chapter 3, finite sets.
d	distance function.
$\mathscr{P}(X)$	probability measures on X (for Chapter 3, see also equation (3.1.1)).
$C_b(X)$	continuous bounded functions on X.
$\rightharpoonup$	narrow convergence of probability measures (i.e. in duality with $C_b(X)$).
$\mu \llcorner E$	measure $\mu(\cdot \cap E)$.
p_x, p_y	canonical projections of $X \times Y$.
$\Gamma(\mu, \nu)$	transport plans from μ to ν, see Definition 1.1.5.
$\Gamma_o(\mu, \nu)$	optimal plans from μ to ν, see Problem 1.1.6.
$\mathrm{supp}(\pi)$	support of the measure π: $\mathrm{supp}(\pi) := \{x : \pi(B_r(x)) > 0 \ \forall r > 0\}$.
ϕ^c	c-transform of the function ϕ, see Definition 1.2.5.
$\pi_y \otimes \nu$	measure on $X \times Y$ s.t. $\int_{X \times Y} f \ d(\pi_y \otimes \nu) = \int_Y \nu(dy) \int_X f(x, y) \pi_y(dx)$.
$\mathrm{Lip}(\phi)$	Lipschitz constant of the function ϕ.
$\mathscr{L}^n$	Lebesgue measure on $\mathbb{R}^n$.
$\mathscr{P}_p(X)$	$\{\mu \in \mathscr{P}(X) : \int d(x, \bar{x})^p d\mu(x) < \infty$ for some $\bar{x} \in X\}$.
$\mathscr{P}^a(\mathbb{R}^n)$	$\{\mu \in \mathscr{P}(\mathbb{R}^n) : \mu \ll \mathscr{L}^n\}$.
$\mathscr{P}^a_p(\mathbb{R}^n)$	$\mathscr{P}^a(\mathbb{R}^n) \cap \mathscr{P}_p(\mathbb{R}^n)$.
$\mathrm{Int}(A)$	interior of the set A.
$\mathrm{Dom}(f)$	set where the function f is defined and finite.
W_p	p-Wassertein distance, see Definition 1.3.1.
$AC(I; X)$	absolutely continuous functions $I \to X$, see Definition 1.3.12.
$\lvert f' \rvert$	metric derivative, see Proposition 1.3.13.
$\mathrm{len}(f)$	lenght of the curve f, see Definition 1.3.15.
$A_p(\gamma)$	p-action of the AC curve γ, see Remark 1.3.18. ($p = 2$ if omitted.)
$\mathrm{Geo}(X)$	constant speed geodesics $(0, 1) \to X$, see page 15.
e_t	evaluation map $\gamma \mapsto \gamma(t)$.
$\mathrm{OptGeo}(\mu, \nu)$	optimal geodesic plans from μ to ν, see Remark 1.3.22.
v_t	in Chapter 2, velocity field satisfying Hypothesis 2.1.8.

$X_t(x)$	maximal solution of the characteristic equation, see Theorem 2.1.6.
$X_t(x,s)$	maximal solution starting from x at time s, see Remark 2.1.7.
$\operatorname{Tan}_\mu \mathscr{P}_2(\mathbb{R}^n)$	tangent space to $\mathscr{P}_2(\mathbb{R}^n)$ at μ, see Definition 2.1.19.
$\lvert\nabla E\rvert$	slope, see Definition 2.2.1.
T_μ^ν	unique optimal map from $\mu \in \mathscr{P}^a(\mathbb{R}^n)$ to $\nu \in \mathscr{P}(\mathbb{R}^n)$.
$\nabla^W E$	Wasserstein differential of E, see Definition 2.2.6.
$\partial^W E$	Wasserstein subdifferential of E, see Definition 2.2.8.
$\partial^o E$	"minimal" subdifferential, see Remark 2.2.16.
$\mathcal{V}, V$	potential energy and potential: for Chapter 2, see Definition 2.2.18; for Chapter 3, see Definition 3.3.1.
$\mathcal{U}, U$	internal energy and the function defining it: for Chapter 2, see Definition 2.2.23; for Chapter 3, see Definition 3.3.3.
L_U	Legendre transform of U, see Remark 2.2.29.
K	in Chapter 3, irreducible Markov kernel on X. (Reversible after Hypothesis 3.1.18.)
π	invariant probability vector of K.
$\nabla\psi, \nabla\cdot\Psi$	in Chapter 3, discrete gradient and divergence: see Definition 3.1.1.
$\Delta\psi$	in Chapter 3, discrete Laplacian $\nabla\cdot\nabla\psi$.
$\langle\phi,\psi\rangle_\pi, \lVert\phi\rVert_\pi$	see Definition 3.1.1.
$\langle\Phi,\Psi\rangle_K, \lVert\Phi\rVert_K$	see Definition 3.1.1.
$\langle\Phi,\Psi\rangle_{\hat\rho}, \lVert\Phi\rVert_{\hat\rho}$	see page 70.
$A \bullet B$	componentwise multiplication of two matrices.
$\hat\rho, \theta$	see Hypothesis 3.1.2 (and then Hypothesis 3.1.32).
ρ^β	see equation (3.1.2).
$\mathcal{W}_{p,q}$	distance on the 2-point space defined in Theorem 3.1.14.
$\mathcal{W}$	new distance on $\mathscr{P}(X)$ if X is finite, see Definition 3.1.16.
d_g	graph distance associated to K, see Definition 3.1.20.
$d_{TV}(\rho,\sigma)$	total variation distance, i.e. $\lVert\rho-\sigma\rVert_{L^1(\pi)}$.
$\lVert\theta\rVert'_\infty$	see page 73.
C_d	doubling constant of θ, see Hypothesis 3.1.32.
C_θ	finiteness parameter, see equation (3.1.18).
$x \sim_{\boldsymbol\rho} y$	see Definition 3.1.37.
$A(\rho), B(\rho)$	matrices defined in equations (3.2.1) and (3.2.2).
Π	diagonal matrix with diagonal entries $\pi(x)$.
$\mathscr{P}_\sigma(X)$	$\{\rho\in\mathscr{P}(X) : \mathcal{W}(\rho,\sigma)<\infty\}$.
$\mathscr{P}'_\sigma(X)$	see page 87.
$\mathscr{P}^b_\sigma(X)$	$\big\{\rho\in\mathscr{P}'_\sigma(X) : \rho(x)\ge b\ \forall x\in\operatorname{supp}(\sigma)\big\}$.
$\mathscr{P}_*(X)$	$\{\rho\in\mathscr{P}(X) : \rho(x)>0\ \forall x\in X\}$.
B_ρ	restriction of $B(\rho)$ to $\operatorname{Ran} A(\rho)$.
T_ρ	"tangent space" to $\mathscr{P}'_\sigma(X)$ at ρ, see Theorem 3.2.7.
$\mathcal{I}_\rho$	see Theorem 3.2.7.
$D_t\rho$	tangent velocity field to $(\rho_t)\subseteq\mathscr{P}'_\sigma(X)$, see Theorem 3.2.7.
Proj_H	orthogonal projection on the subspace $H\subseteq\mathbb{R}^X$.

Chapter 1

Optimal transport

In this chapter we briefly review the basic concepts of the theory of optimal transport. For the sake of completeness, we will also give a sketch of the proofs; but we will omit many details and verifications, which may be performed by the interested reader. A rich reference book on the topic is [17].

1.1 Statement of the problem

Our starting point is the following variational problem.

Problem 1.1.1 (Monge, 1781, generalized)**.** Given two distributions of mass on two measurable spaces X, Y, represented by two probability measures μ and ν, and a measurable function $c : X \times Y \to [0, \infty]$ representing the "cost of transport", find a **transport map** $T : X \to Y$, i.e. a map such that $T_{\#}\mu = \nu$, realizing

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

Hypothesis 1.1.2. From now on, unless otherwise specified, X, Y will be Polish spaces (i.e. metric, separable and complete), endowed with their Borel σ-algebra; c will be lower semicontinuous (one of the main reasons is that we want Theorem 1.1.10 below to hold); $\mathscr{P}(X)$ will denote the space of Borel probability measures on X, endowed with the topology of narrow convergence (i.e. the topology obtained via duality with $C_b(X)$). Convergence in this topology will be denoted by $\rightharpoonup$.

Remark 1.1.3 (Ill-posedness). The problem in general is ill-posed, for various reasons:

- Sometimes no admissible T exists: take $a, b \in X$ such that $a \neq b$, $\mu = \delta_a$, $\nu = \frac{1}{2}(\delta_a + \delta_b)$.
- The infimum might not be attained:

Example 1.1.4 (inf not attained)**.** On $X = Y = \mathbb{R}^2$ with $c(x, y) = |x - y|^2$, take

$$\mu = \mathcal{H}^1 \llcorner (\{0\} \times [0, 1]), \quad \nu = \frac{1}{2}\mathcal{H}^1 \llcorner (\{-1\} \times [0, 1]) + \frac{1}{2}\mathcal{H}^1 \llcorner (\{1\} \times [0, 1]),$$

and observe that $|T(x) - x| \geq 1$ μ-a.s., and having a.s. equality is impossible. However it is quite easy to construct transport maps whose cost is arbitrarily close to 1.

So, we choose to extend the problem: we look for a "transport plan" possibly sending part of the mass at the same point to different points. For a precise formulation, let us denote by p_x, p_y the canonical projections of $X \times Y$.

Definition 1.1.5 (Transport plan)**.** $\pi \in \mathscr{P}(X \times Y)$ is called a **transport plan** from μ to ν if $(p_x)_\# \pi = \mu$ and $(p_y)_\# \pi = \nu$. The set of such measures π will be denoted by $\Gamma(\mu, \nu)$. (Note that $\mu \times \nu \in \Gamma(\mu, \nu)$, which therefore is nonempty.)

$\pi(A \times B)$ represents the amount of mass from A sent into B.

Problem 1.1.6 (Kantorovich)**.** Given $\mu \in \mathscr{P}(X), \nu \in \mathscr{P}(Y)$, and a cost function $c : X \times Y \to [0, \infty]$, find a transport plan realizing the minimum **cost**:

$$\inf_{\pi \in \Gamma(\mu,\nu)} \int_{X \times Y} c(x, y) d\pi(x, y). \tag{K}$$

Such a plan is said to be **optimal**. The set of optimal plans will be denoted by $\Gamma_o(\mu, \nu)$.

Remark 1.1.7. Transport maps can be seen as special transport plans, in the sense that for every transport map T we can define the **induced plan** $(Id \times T)_\# \mu \in \Gamma(\mu, \nu)$ which has (Kantorovich) cost exactly equal to the (Monge) cost of T.

More precisely:

Proposition 1.1.8 (Plans induced by maps)**.** *$\pi \in \Gamma(\mu, \nu)$ is induced by a Borel transport map if and only if it is concentrated on a π-measurable graph Γ.*

Proof. For the nontrivial implication, let Γ_0, Γ_1 be Borel subsets of $X \times Y$ such that $\Gamma_0 \subseteq \Gamma \subseteq \Gamma_1$ and $\pi(\Gamma_1 \setminus \Gamma_0) = 0$. Using Ulam's lemma take compact sets $K_n \uparrow \Gamma_0$ π-a.s.; each K_n is the graph of a continuous function f_n (in fact, $p_x(K_n)$ is compact and hence measurable; the continuity follows from the compactness of K_n). We conclude noting that for any ϕ Borel and bounded

$$\int_{X \times Y} \phi(x, y) d\pi(x, y) = \int_{X \times Y} \phi(x, T(x)) d\pi(x, y) = \int_X \phi(x, y) d\left[(Id \times T)_\# \mu\right](x, y).$$

$\square$

In particular, inf(K)$\leq$ inf(M). The inequality may be strict:

Example 1.1.9. (inf(K)$<$ inf(M)) Take the measures of Example 1.1.4, but with the cost equal to 1 if the two points' second coordinates coincide, 2 otherwise. It is easy to verify that the set on which x and $T(x)$ have the same second coordinate must be μ-negligible to have $T_\# \mu = \nu$.

Theorem 1.1.10 (Existence of optimal plans)**.** *$\Gamma_o(\mu, \nu)$ is nonempty, convex and compact (with respect to the narrow topology).*

Note. Sequential compactness or compactness are in fact the same concept, because if X is metric, then $\mathscr{P}(X)$ with narrow topology is metrizable: see Remark 1.3.7 below.

Proof. Convexity is evident. Consider $\Gamma(\mu, \nu)$ with the narrow topology. We use Prokhorov's theorem to see that it is compact. It is obviously closed. It is tight: if H, K are compact sets such that $\mu(X \setminus H) < \varepsilon$ and $\nu(Y \setminus K) < \varepsilon$ (they exist by Ulam's lemma), then every admissible plan π satisfies $\pi((X \times Y) \setminus (H \times K)) < 2\,\varepsilon$. We conclude observing that $\pi \mapsto \int c \, d\pi$ is lower semicontinuous: take $c_n \uparrow c$ continuous bounded functions and note that our map is the supremum of the continuous maps $\pi \mapsto \int c_n \, d\pi$. □

Remark 1.1.11. If we have $\mu_h \rightharpoonup \mu, \nu_h \rightharpoonup \nu$ and $\pi_h \in \Gamma(\mu_h, \nu_h)$, then (π_h) has narrow limit points (which obviously belong to $\Gamma(\mu, \nu)$). In fact the argument showing compactness of $\Gamma(\mu, \nu)$ in the previous proof, can be used to prove that the tightness of the sequences $(\mu_h), (\nu_h)$ implies the tightness of (π_h). For a discussion of the optimality of the limit, see Theorem 1.2.21 below.

1.2 Duality and Brenier's theorem

A property intuitively desirable for the support of a candidate optimal plan is the following one:

Definition 1.2.1 (c-monotonicity)**.** A set $\Gamma \subseteq X \times Y$ is said to be **c-monotone** if whenever $(x_i, y_i) \in \Gamma$ for $i = 1, \ldots, n$, if $\sigma \in S_n$ is any permutation, it holds that $\sum_{i=1}^n c(x_i, y_i) \leq \sum_{i=1}^n c(x_i, y_{\sigma(i)})$. Since every permutation is represantable as a product of cycles, it is sufficient that the inequality is true for cyclic permutations.

Lemma 1.2.2 (c-monotonicity of $\mathrm{supp}(\pi)$ for c continuous)**.** *If c is continuous and $\pi \in \Gamma_o(\mu, \nu)$, then* $\mathrm{supp}(\pi)$ *is c-monotone. (See also Remark 1.2.15.)*

Proof. By contradiction

$$\sum_{i=1}^n c(x_i, y_i) > \sum_{i=1}^n c(x_i, y_{\sigma(i)}) \quad \text{for some } \sigma \in S_n, \ (x_i, y_i)_{i=1,\ldots,n} \subseteq \mathrm{supp}(\pi).$$

By continuity $\sum_{i=1}^n c(u_i, v_i) > \sum_{i=1}^n c(u_i, v_{\sigma(i)})$ for $(u_i, v_i) \in U_i \times V_i$ neighbourhoods of (x_i, y_i). By definition of support we have that $\pi_i := \frac{1}{\pi(U_i \times V_i)} \pi \llcorner (U_i \times V_i)$ is well defined. Now take $L_i = (z_i, w_i)$ random variables on a common probability space with law π_i with respect to the probability measure $\mathbb{P}$, and put

$$\eta := \alpha \Big[\sum (z_i, w_{\sigma(i)})_\# \, \mathbb{P} - \sum (z_i, w_i)_\# \, \mathbb{P} \Big]$$

with $0 < \alpha < \min_i(\pi(U_i \times V_i))$ so that $\pi + \eta$ is a positive measure. Note that $(p_x)_\# \eta = 0, (p_y)_\# \eta = 0$ and $\int c \, d\nu < 0$, hence $\pi + \eta$ contradicts the minimality of π. □

Example 1.2.3. Suppose that $X = Y = \mathbb{R}^n$ and $c(x,y) = \frac{1}{2}|x-y|^2$, and take an optimal plan induced by a map T. We now know that $\mathrm{supp}((Id \times T)_\# \mu)$ is c-monotone: this reduces to the classical concept of **monotonicity of a map**, i.e. the existence of a μ-null set N such that $\langle T(x) - T(y), x - y \rangle \geq 0$ for every $x, y \in X \setminus N$.

The fundamental Theorem below shows that there is a "dual problem" which is equivalent to Kantorovich's problem (K):

Theorem 1.2.4 (Kantorovich's duality)**.** *Consider the problem*

$$\sup\left\{ \int \phi \, d\mu + \int \psi \, d\nu : \phi \in L^1(\mu), \psi \in L^1(\nu), \phi(x) + \psi(y) \leq c(x,y) \ \forall x, y \right\}. \quad \text{(D)}$$

Then $\sup \text{(D)} = \min \text{(K)}$.

For the proof of this Theorem we need the following concepts:

Definition 1.2.5. The **c-transform** of a function $\phi : X \to \bar{\mathbb{R}}$ is the function $\phi^c : Y \to \bar{\mathbb{R}}$ given by $\phi^c(y) := \inf_x (c(x,y) - \phi(x))$ (let us decide in this case that $+\infty - \infty = +\infty$). It is the greatest ψ such that $\phi(x) + \psi(y) \leq c(x,y) \ \forall x, y$.

A function ψ is **c-concave** if there exists a ϕ such that $\psi = \phi^c$. It is easy to see that a function of the form $\psi(y) = \inf_{i \in I}(c(x_i, y) + \alpha_i)$ is c-concave.

A function ψ is **c-convex** if $-\psi$ is c-concave.

Remark 1.2.6. Using the fact that $\phi^{cc} \geq \phi$ we get immediately that $\phi^{ccc} = \phi^c$. In particular ϕ is c-concave if and only if $\phi^{cc} = \phi$.

Remark 1.2.7. One inequality in the theorem is trivial: for all $\phi \in L^1(\mu), \psi \in L^1(\nu)$ with $\phi(x) + \psi(y) \leq c(x,y) \ \forall x, y$, and for all $\pi \in \Gamma(\mu, \nu)$, it holds

$$\int \phi \, d\mu + \int \psi \, d\nu = \int (\phi(x) + \psi(y)) \, d\pi(x,y) \leq \int c \, d\pi. \tag{1.2.1}$$

Note that the chain of inequalities makes sense even if we only know that ϕ and ψ are upper semi-integrable.

Remark 1.2.8. Suppose now in addition that for some ϕ measurable and $> -\infty$ a.s., it holds that $\phi(x) + \phi^c(y) = c(x,y)$ π-a.s.; then we can prove that ϕ^c is measurable. To this aim, disintegrate π as $\pi_y \otimes \nu$, so that π_y is concentrated on the set $\{(x,y) : \phi(x) + \phi^c(y) = c(x,y)\}$ for ν-a.e. y: then it suffices to observe that $\phi^c(y) = \int (c(x,y) - \phi(x)) d\pi_y(x)$ for ν-a.e. y.

We note that if ϕ and ϕ^c are upper semi-integrable, then we can put $\psi = \phi^c$ in equation (1.2.1), and moreover all the inequalities become equalities: so, a posteriori, $\phi \in L^1(\mu)$ and $\phi^c \in L^1(\nu)$, and (ϕ, ϕ^c) realizes the equality in the Kantorovich's duality. We also deduce that π is optimal with finite cost.

Note that without loss of generality we can suppose that ϕ is c-concave, because our assumption implies that $\phi = \phi^{cc}$ π-a.s..

With this in mind, we give the following definition:

Definition 1.2.9. Given $\pi \in \mathscr{P}(X \times X)$ with first marginal μ, a function $\phi \in L^1(\mu)$ will be called a **Kantorovich potential** if it is c-concave and $\phi(x) + \phi^c(y) = c(x,y)$ π-a.s..

Here is a general sufficient condition for the existence of a Kantorovich potential:

Theorem 1.2.10 (When c-monotonicity implies optimality). *Let c be a finite cost. Let $\pi \in \Gamma(\mu, \nu)$ be concentrated on a Borel c-monotone set.*

1. *If* $\min(\mathrm{K}) < \infty$*, then π is optimal and Kantorovich's duality holds.*
2. *If $\mu\Big\{x : \int c(x,y)d\nu(y) < \infty\Big\} > 0$ and $\nu\Big\{x : \int c(x,y)d\mu(x) < \infty\Big\} > 0$, then there exists a measurable c-concave Kantorovich potential ϕ, such that $\phi(x) + \phi^c(y) = c(x,y)$ π-a.s.; as a consequence, π is optimal with finite cost and the equality in Kantorovich's duality is achieved.*

For a converse, see Corollary 1.2.14.

Remark 1.2.11. The conditions of part 2. are satisfied if $\int c \; d(\mu \times \nu) < \infty$; for instance, if $c(x,y) \le a(x) + b(y)$ where $a \in L^1(\mu), b \in L^1(\nu)$.

Proof. Let Γ be a c-monotone set on which π is concentrated; combining Lusin's theorem and Ulam's lemma we can suppose the existence of compact sets $\Gamma_k \uparrow \Gamma$ such that $c_{|\Gamma_k}$ is continuous.

Fix $(x_0, y_0) \in \Gamma_1$ and take

$$\phi(x) := \inf\left\{ c(x, y_p) - \sum_{i=0}^{p} c(x_i, y_i) + \sum_{i=1}^{p} c(x_i, y_{i-1}) : p \in \mathbb{N}, (x_i, y_i)_{i=1,\dots,p} \subseteq \Gamma \right\}.$$

We claim that $\phi(x) + \phi^c(y) = c(x,y)$ on Γ (and so π-a.s.): the nontrivial inequality is

$$\phi(x') \le \phi(x) + c(x', y) - c(x,y) \quad \forall x' \in X, \forall (x,y) \in \Gamma \tag{1.2.2}$$

which is evident by definition of ϕ. Obviously, $\phi(x_0) = 0$; then, using (1.2.2) with $x' = x_0$, we see that $\phi > -\infty$ on $p_x(\Gamma)$ (σ-compact and hence measurable), therefore μ-a.s..

We claim that ϕ is measurable. In fact take $c_l \uparrow c$ continuous functions, and for every p, m, l put

$$\phi_{p,m,l}(x) := \inf\left\{ c_l(x, y_p) - \sum_{i=0}^{p} c(x_i, y_i) + \sum_{i=1}^{p} c_l(x_i, y_{i-1}) : (x_i, y_i)_{i=1,\dots,p} \subseteq \Gamma_m \right\},$$

which is u.s.c. because it is an infimum of continuous functions. It is easily proven that if $f_l \uparrow f$ are defined on a compact set and f_l are continuous then $\min f_l \to \min f$: hence $\phi_{p,m,l} \to \phi_{p,m}$, where $\phi_{p,m}$ is defined as $\phi_{p,m,l}$ but with c_l replaced by c. The conclusion follows noting that $\phi = \lim_{p\to\infty} \lim_{m\to\infty} \phi_{p,m}$.

We now show that if $(x,y) \in \Gamma$ then $\phi^c(y) = c(x,y) - \phi(x)$: we have to prove that $c(x', y) - \phi(x') \ge c(x,y) - \phi(x)$ for every $x' \in X$, i.e. that

$$\phi(x') \le \inf\left\{ c(x, y_p) - \sum_{i=0}^{p} c(x_i, y_i) + \sum_{i=1}^{p} c(x_i, y_{i-1}) : p \in \mathbb{N}, (x_i, y_i)_{i=1,\dots,p} \subseteq \Gamma \right\} + c(x', y) - c(x,y)$$

which is immediate by definition of ϕ.

Therefore, as a consequence of Remark 1.2.8, $\psi := \phi^c$ is measurable.

We consider the functions $\phi_n := (\phi \wedge n) \vee (-n)$, $\psi_n := (\psi \wedge n) \vee (-n)$; note that $\phi_n(x) + \psi_n(y) \leq c(x, y)$. Hence $\int (\phi_n(x) + \psi_n(y)) d\pi \leq \sup(\mathrm{D})$ by definition.

We observe that $\pi\{\phi(x) + \psi(y) = c(x, y) \geq 0\} = 1$. We claim that on this set $\phi_n(x) + \psi_n(y) \uparrow \phi(x) + \psi(y)$: in fact until n is such that ϕ or ψ is less or equal than $(-n)$, then the other one is greater or equal than n and so $\phi_n(x) + \psi_n(y) = 0$; while for greater values of n, $\phi_n(x) + \psi_n(y) = \phi(x) \wedge n + \psi(y) \wedge n$ which evidently increases to $\phi(x) + \psi(y)$. From this monotonicity we also get that $\phi_n(x) + \psi_n(y) \geq \phi_0(x) + \psi_0(y) = 0$. To sum up, we are in the position to apply the monotone convergence theorem: we get $\int c\, d\pi = \lim_n (\int \phi_n d\mu + \int \psi_n d\nu) \leq \sup(\mathrm{D})$, so that π satisfies the nontrivial inequality in Kantorovich's duality, and hence is optimal.

With the hypotheses of 2., instead, we can prove that ψ^+ itself is ν-integrable: in fact we can choose $\bar{x} \in X$ such that $\int c(\bar{x}, y) d\nu(y) < \infty$ and $\phi(\bar{x}) > -\infty$, so $\psi \leq c(\bar{x}, \cdot) - \phi(\bar{x})$ concludes. By the same argument, ϕ is μ-upper semi-integrable. Now Remark 1.2.8 gives the conclusion. □

Example 1.2.12. When $c(x, y) = d(x, y)$, it is straightforward to prove that ϕ is c-concave if and only if it is 1-Lipschitz; for such a ϕ we have $\phi^c = -\phi$. We can apply part 2. of the previous Theorem if $\int d(\bar{x}, x) d\mu(x) < \infty$ for any $\bar{x} \in X$ (in particular, if X is bounded): in this case we conclude that $\min(\mathrm{K}) = \max_{\mathrm{Lip}(\phi) \leq 1} \int \phi\, d(\mu - \nu)$.

Example 1.2.13. When $c(x, y) = \frac{1}{2}|x - y|^2$ in $\mathbb{R}^n$, the definition of c-concavity of a function ϕ reduces to requiring that $\phi(x) - \frac{|x|^2}{2}$ is an infimum of affine functions, that is to say, that $\phi(x) - \frac{|x|^2}{2}$ is concave and upper semicontinuous.

Proof of Kantorovich's duality. For $c \in C_b$ it is an immediate consequence of the Theorem above, because we already know that $\mathrm{supp}(\pi)$ is c-monotone for every π optimal.

In general take continuous and bounded cost functions $c_n \uparrow c$ and π_n optimal plans relative to them; by Remark 1.1.11, without loss of generality $\pi_n \rightharpoonup \pi$ for a certain π. We know that

$$\int c_n\, d\pi_n = \sup\left\{\int \phi d\mu + \int \psi d\nu : \phi(x) + \psi(y) \leq c_n(x, y)\right\} \leq$$
$$\leq \sup\left\{\int \phi d\mu + \int \psi d\nu : \phi(x) + \psi(y) \leq c(x, y)\right\}.$$

Therefore we would conclude if we were able to prove that $\int c\, d\pi \leq \liminf_n \int c_n d\pi_n$. Fix $l \in \mathbb{N}$ and a subsequence $n(k)$ which realizes $\liminf_n \int c_n d\pi_n$.

$$\int c_l d\pi = \lim_{k\to\infty} \int c_l d\pi_{n(k)} \overset{c_l \leq c_{n(k)} \text{ for } k \gg 1}{\leq} \liminf_{k\to\infty} \int c_{n(k)} d\pi_{n(k)} = \liminf_{n\to\infty} \int c_n d\pi_n.$$

Letting $l \to \infty$, the claim is proved. □

Corollary 1.2.14 (c-monotonicity vs. optimality). *Every optimal plan is concentrated on a c-monotone set. In particular, in the hypotheses of part* 1. *or part* 2. *of Theorem 1.2.10, optimality is equivalent to concentration on a c-monotone set.*

Proof. For π optimal plan, let $\phi_n \in L^1(\mu), \psi_n \in L^1(\nu)$ be such that

$$\phi_n(x) + \psi_n(y) \leq c(x, y)\ \forall x, y, \quad \text{and} \int \phi_n d\mu + \int \psi_n d\nu \uparrow \int c\ d\pi;$$

put $g_n(x, y) := c(x, y) - \phi(x) - \psi(y) \geq 0$, and note that we are saying that $g_n \xrightarrow{L^1(\pi)} 0$; possibly passing to a subsequence we can assume that they converge π-a.s. too. Call Γ a set on which $g_n \to 0$ pointwise and such that $\pi(\Gamma) = 1$. We conclude noting that Γ is c-monotone: in fact if $(x_i, y_i)_{i=1,\dots,n} \subseteq \Gamma$, and $\sigma \in S_n$, we have

$$\sum_{i=1}^n c(x_i, y_{\sigma(i)}) \geq \sum_{i=1}^n \Big(\phi_k(x_i) + \psi_k(y_{\sigma(i)})\Big) = \sum_{i=1}^n \Big(\phi_k(x_i) + \psi_k(y_i)\Big) \xrightarrow{k\to\infty} \sum_{i=1}^n c(x_i, y_i).$$

□

Remark 1.2.15. If c is continuous, then the closure of a c-monotone set is c-monotone: as a consequence, $\mathrm{supp}(\pi)$ is c-monotone for every $\pi \in \Gamma_o(\mu, \nu)$ (as we already knew by Lemma 1.2.2). We can even find a c-monotone closed set on which *all* the optimal plans are concentrated. First of all, if we find any set with this property, we conclude taking the closure. To complete the argument, consider the set $\bigcup_{\pi\in\Gamma_o(\mu,\nu)} \mathrm{supp}(\pi)$. By definition of c-monotonicity, it is sufficient to prove that $\bigcup_{i=1}^k \mathrm{supp}(\pi_i)$ is c-monotone for every $\pi_1, \dots, \pi_k \in \Gamma_o(\mu, \nu)$; this is true because that finite union coincides with $\mathrm{supp}(\frac{1}{k}\sum_{i=1}^k \pi_i)$, support of an optimal plan.

Remark 1.2.16. We cannot state in full generality that c-monotonicity of the support (or of any π-full set) is equivalent to optimality: take $X = Y = [0, 1]$, $\mu = \nu = \mathscr{L}^1$, $\alpha \in [0, 1] \setminus \mathbb{Q}$, and put $c(x, x) = 1$; $c(x, x + \alpha \pmod 1) = 2$; $c(x, y) = +\infty$ in every other case. Obviously $T(x) := x + \alpha \pmod 1$ is not optimal; but it is not difficult to show that the support of $(Id \times T)_\#\mu$ is c-monotone.

Remark 1.2.17 (On uniqueness of optimal plans). In general the optimal plan is not unique; there can even be different *maps* solving Kantorovich's problem, as shown by the following "book shifting" example. Let $X = Y = \mathbb{R}$, $c(x, y) = |x - y|$, $\mu = \frac{1}{n} I_{[0,n]} \mathscr{L}^1$, $\nu = \frac{1}{n} I_{[1,n+1]} \mathscr{L}^1$. Then Example 1.2.12, applied to the 1-Lipschitz function $-t$, gives that the minimum of (K) is at least 1; which is attained by the two different maps $T_1(x) := x + 1$ and $T_2(x) := x + nI_{[0,1]}(x)$.

On the other hand, if every optimal plan is induced by a map, then there is uniqueness. In fact if $\pi = (Id \times T)_\#\mu$ and $\pi' = (Id \times T')_\#\mu$ are both optimal, then so is $\pi'' = \frac{1}{2}(\pi + \pi')$ which is of the form $\mu \otimes \frac{1}{2}(\delta_{T(x)} + \delta_{T'(x)})$: by hypothesis it is induced by a map, so $\frac{1}{2}(\delta_{T(x)} + \delta_{T'(x)})$ is a delta μ-a.s.: therefore $T = T'$ μ-a.s..

Remark 1.2.18 (Invertibility of optimal maps). Suppose in addition that also the optimal transport from ν to μ is induced by a map (uniqueness is obvious by symmetry of (K)). Then the unique element of $\Gamma_o(\mu, \nu)$ is concentrated both on $\{(x, T(x)) : x \in E\}$ and on $\{(S(y), y) : y \in F\}$, where $\mu(E) = 1$ and $\nu(F) = 1$. From this, one easily deduces that $S \circ T = Id$ μ-a.s. and $T \circ S = Id$ ν-a.s..

We now study an important case in which we know that the optimal plan is unique and induced by a transport.

Notation. For $p \in [1, \infty)$, we will use the notations:
$\mathscr{P}_p(X) := \{\mu \in \mathscr{P}(X) : \int d(x, \bar{x})^p d\mu(x) < \infty$ for some, and hence all, $\bar{x} \in X\}$.
$\mathscr{P}^a(\mathbb{R}^n) := \{\mu \in \mathscr{P}(\mathbb{R}^n) : \mu \ll \mathscr{L}^n\}$.
$\mathscr{P}^a_p(\mathbb{R}^n) := \mathscr{P}^a(\mathbb{R}^n) \cap \mathscr{P}_p(\mathbb{R}^n)$.

Theorem 1.2.19 (Brenier). *Given $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^n)$, $\mu \ll \mathscr{L}^n$, $c(x, y) := \frac{1}{2}|x - y|^2$, then:*

1. *The optimal plan is unique and induced by a map T;*

2. *$T = \nabla f$ for some $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ convex l.s.c. such that μ is concentrated on* $\mathrm{Int}(\mathrm{Dom}(f))$.

Conversely, if $\mu \in \mathscr{P}^a_2(\mathbb{R}^n)$, f is convex l.s.c., μ is concentrated on $\mathrm{Dom}(f)$, *and $T := \nabla f$ is an element of $L^2(\mu)$, then T induces the unique optimal plan from μ to $T_\# \mu$.*

Remark 1.2.20. A convex l.s.c. function f is locally Lipschitz (and hence differentiable at a.e. point by Rademacher's theorem) in the interior of its domain. First of all, it is locally bounded: take any closed hypercube in $\mathrm{Int}(\mathrm{Dom}(f))$ and observe that the values in it are bounded from below by lower semicontinuity; and from above by the values at the vertices. Then, using the monotonicity of the difference quotients of a convex function $\mathbb{R} \to \mathbb{R}$, it is easy to prove that, calling $\mathrm{Osc}(f, B_R(x)) := \sup_{y,z \in B_R(x)} |f(y) - f(z)|$, it holds

$$\mathrm{Lip}_{B_r(x)}(f) \le \frac{\mathrm{Osc}(f, B_R(x))}{R - r} \quad \text{whenever } B_r(x) \subset B_R(x) \subset \mathrm{Int}(\mathrm{Dom}(f)).$$

Proof (Brenier). Denote by π an optimal plan. Note that $c(x, y) \le |x|^2 + |y|^2$ sum of an $L^1(\mu)$ and an $L^1(\nu)$ function, so there exists a c-concave Kantorovich potential ϕ by Remark 1.2.11; we also know that, with this cost, c-concavity means that $f(x) := \frac{|x|^2}{2} - \phi(x)$ is convex lower semicontinuous. We observe that $\partial(\mathrm{Dom}(f))$ is always $\mathscr{L}^n$ negligible (since the boundary of a convex set is always locally a Lipschitz graph), and f is differentiable a.e. in $\mathrm{Int}(\mathrm{Dom}(f))$ by the previous Remark; from $\mu \ll \mathscr{L}^n$ we deduce that μ is concentrated on a set on which f is finite and differentiable.

To sum up, without loss of generality π is concentrated on Γ c-monotone such that f (and hence ϕ) is differentiable on $p_x(\Gamma)$ and $\phi(x) + \phi^c(y) = c(x, y)$ for every $(x, y) \in \Gamma$. The latter statement means that for every $(x, y) \in \Gamma$, $\frac{1}{2}|x' - y|^2 - \phi(x')$ attains its minimum in $x' = x$; by the differentiability assumption we get that $x - y - \nabla\phi(x) = 0$, i.e. $y = \nabla f(x)$.

Conversely, $T \in L^2$ means that $\nu := T_\# \mu$ is an element of $\mathscr{P}_2(\mathbb{R}^n)$, so that, as above, optimality is equivalent to concentration on a c-monotone set. This c-monotonicity is easily proved using that the gradient of a convex function is monotone, i.e. $\langle \nabla f(x) - \nabla f(y), x - y \rangle \ge 0 \ \forall x, y$. $\square$

Note. It is not really difficult to adapt the proof of the first part to a cost of the form $h(x-y)$ with h strictly convex (typically, $h(x,y) = |x-y|^p$ with $p > 1$); roughly speaking, one replaces the gradient of $|\cdot|^2$ with the subdifferential of h. For the details, see [3]. However, we do not have the nice characterisation of the potentials obtained in Example 1.2.13, so part 2 of the above theorem is typical of the quadratic cost.

On the other hand, McCann generalized both parts of the theorem replacing $\mathbb{R}^n$ with a Riemannian manifold and $\mathscr{L}^n$ with the Riemannian volume measure: see [13].

Theorem 1.2.21 (Stability of optimal plans)**.**

1. *Let c_h be finite continuous costs uniformly converging to c. Suppose that $\mu_h \rightharpoonup \mu$ in $\mathscr{P}(X)$, $\nu_h \rightharpoonup \nu$ in $\mathscr{P}(Y)$, $\pi_h \in \Gamma_o(\mu_h, \nu_h)$ with respect to the cost c_h.*

   If $\pi_h \rightharpoonup \pi$ with $\liminf_h \int c_h d\pi_h < \infty$, then $\pi \in \Gamma_o(\mu,\nu)$ with respect to c, with finite cost $\int c\, d\pi \leq \liminf_h \int c_h d\pi_h$. (Recall that (π_h) always has limit points by Remark 1.1.11.)

2. *In the setting of the previous point, take $Y = \mathbb{R}^n$, $\mu_h \equiv \mu$. Suppose that π_h, π are induced by maps T_h, T (true for example if $X = \mathbb{R}^n$ and $\mu \ll \mathscr{L}^n$). Assume that there is a compact set K such that $\mathrm{supp}(\nu_h) \subseteq K$ for every h. Then $T_h \xrightarrow{L^p(\mu)} T$ for every $p \in [1,\infty)$.*

Note. For a setting in which the hypothesis of existence of K can be dropped, see Proposition 1.3.9.

Proof. $\boxed{1.}$ Obviously $\pi \in \Gamma(\mu,\nu)$. Fix $\varepsilon > 0$. Since c is continuous and nonnegative, then $\pi \mapsto \int c\, d\pi$ is l.s.c.: so, for k large we have

$$\int c\, d\pi \leq \varepsilon + \int c\, d\pi_k \leq 2\,\varepsilon + \int c_k\, d\pi_k.$$

Letting $k \to \infty$ along an appropriate subsequence, we get the desired inequality. In particular $\int c\, d\pi < \infty$, so that to conclude it suffices to prove that $\mathrm{supp}(\pi)$ is c-monotone.

To this aim, take $(x_i, y_i)_{i=1,\dots,n} \subseteq \mathrm{supp}(\pi)$; a simple fact about narrow convergence tells that for $i = 1, \dots, n$ there are $\left(x_i^{(h)}, y_i^{(h)}\right) \in \mathrm{supp}(\pi_h)$ converging to (x_i, y_i) as $h \to \infty$. Hence

$$\sum_{i=1}^n \left[c\left(x_i, y_i\right) - c\left(x_i, y_{\sigma(i)}\right)\right] \overset{h\gg1}{\leq} \varepsilon + \sum_{i=1}^n \left[c\left(x_i^{(h)}, y_i^{(h)}\right) - c\left(x_i^{(h)}, y_{\sigma(i)}^{(h)}\right)\right] \leq$$
$$\overset{h\gg1}{\leq} 2\,\varepsilon + \sum_{i=1}^n \left[c_h\left(x_i^{(h)}, y_i^{(h)}\right) - c_h\left(x_i^{(h)}, y_{\sigma(i)}^{(h)}\right)\right] \leq 2\,\varepsilon$$

because we know that $\mathrm{supp}(\pi_h)$ is c_h-monotone. Letting $\varepsilon \to 0$ the c-monotonicity follows.

$\boxed{2.}$ We prove that if $|T_h| \le M < \infty$ and $(Id \times T_h)_\# \mu \rightharpoonup (Id \times T)_\# \mu$, then $T_h \to T$ in $L^p(\mu)$ for every $p \in [1, \infty)$; thanks to the boundedness of T_h we only have to prove convergence in μ-measure.

To this aim, fix $\varepsilon > 0$ and find by Lusin's theorem a continuous function $\tilde{T}$ such that $\tilde{T} \le M$ and $\mu\{T \ne \tilde{T}\} < \frac{\varepsilon}{2M}$. Putting $\phi(x, y) := 2M \wedge |y - \tilde{T}(x)|$, this means that $\int \phi \, d(Id \times T)_\# \mu \le \varepsilon$ and so $\lim_h \int \phi \, d(Id \times T_h)_\# \mu \le \varepsilon$. This implies

$$\limsup_{h \to \infty} \int_{\{T = \tilde{T}\}} (|T_h(x) - T(x)| \wedge 2M) \; d\mu(x) \le \varepsilon;$$

but the argument of the lim sup is $\int_{\{T=\tilde{T}\}} |T_h(x) - T(x)| \; d\mu(x)$, from which we deduce that $\limsup_h \mu\{T = \tilde{T}, |T_h - T| > \sqrt{\varepsilon}\} \le \sqrt{\varepsilon}$. We conclude that

$$\limsup_{h \to \infty} \mu\{|T_h - T| > \sqrt{\varepsilon}\} \le \varepsilon + \sqrt{\varepsilon}.$$

□

1.3 The Wasserstein distance

The theory of optimal transport permits to give to $\mathscr{P}_p(X)$ a natural structure of metric space.

Definition 1.3.1. For $p \in [1, \infty)$, consider the cost $c(x, y) := d(x, y)^p$. For $\mu, \nu \in \mathscr{P}_p(X)$, take an optimal plan π_o. Then the **p-Wasserstein distance** between the two measures is $W_p(\mu, \nu) := (\int d(x, y)^p d\pi_o)^{1/p}$.

Remark 1.3.2 (Comparison). Evidently $W_p \le W_q$ if $p \le q$; if X is bounded one immediately gets that the two distances are equivalent.

Proposition 1.3.3. *W_p is a distance on $\mathscr{P}_p(X)$.*

Proof. The triangular inequality can be proved taking a "3-plan" such that the marginal on the first two coordinates is optimal from μ_1 and μ_2 and the marginal on the last two coordinates is optimal between μ_2 and μ_3 (such a probability measure can be built by disintegration with respect to the second coordinate).

Using the triangle inequality with a Dirac mass as intermediate measure, we see that W_p is always finite by the very definition of $\mathscr{P}_p(X)$. All the other properties of a distance are obvious. □

Proposition 1.3.4 (Separability)**.** *If X is any separable metric space, then also $(\mathscr{P}_p(X), W_p)$ is separable.*

Proof. Let $(x_n)_{n \ge 1}$ be a dense sequence in X; we claim that

$$\mathcal{D} := \left\{ \sum_{i=1}^N a_i \delta_{x_i} : N \in \mathbb{N}, \; (a_i) \in \mathbb{Q}^N, \; a_i \ge 0 \; \forall i, \; \sum_{i=1}^N a_i = 1 \right\}$$

is dense in $\mathscr{P}_p(X)$. By an obvious verification, it is surely dense in $\tilde{\mathcal{D}}$ defined as $\mathcal{D}$ but allowing $(a_i) \in \mathbb{R}^N$: so we only have to approximate a generic $\mu \in \mathscr{P}(X)$ with elements of $\tilde{\mathcal{D}}$.

We note that, for every $\varepsilon > 0$ fixed, the balls $B_i := B(x_i, \varepsilon)$ cover X: therefore there exists N such that

$$\int_{(\cup_{i=1}^N B_i)^c} d(x, x_0)^p d\mu \le \varepsilon \,.$$

Then, if we define $T(x)$ to be equal to x_0 if $x \notin \cup_{i=1}^N B_i$, and otherwise equal to the minimal i such that $x \in B_i$, we have that T is measurable, $T_\# \mu \in \tilde{\mathcal{D}}$, and

$$W_p^p(\mu, T_\# \mu) \le \int_{(\cup_{i=1}^N B_i)^c} d(x, x_0)^p d\mu + \varepsilon^p \le \varepsilon + \varepsilon^p \,.$$

□

Proposition 1.3.5 (Completeness). *If X is any complete metric space, then also $(\mathscr{P}_p(X), W_p)$ is complete.*

Proof. Take a sequence in $\mathscr{P}_p(X)$ such that $\sum_{i=1}^\infty W_p(\mu_i, \mu_{i+1}) < \infty$. Put $\Lambda_1 := \mu_1$. Generalizing inductively the construction sketched in the proof of the triangular inequality, we can build for every $h \ge 2$ a probability Λ_h on X^h such that its marginal on the first $h-1$ coordinates is Λ_{h-1} and its marginal on the last two coordinates is optimal between μ_{h-1} and μ_h. By Kolmogorov's extension theorem we have a probability Λ on $X^{\mathbb{N}}$ whose marginal on the first h coordinates if Λ_h. Denote by $p_n : X^{\mathbb{N}} \to X$ the n-th canonical projection: then one easily sees that $\|d(p_h, p_{h+1})\|_{L^p(\Lambda)} = W_p(\mu_h, \mu_{h+1})$ summable by hypothesis: hence $(p_h)_{h \ge 1}$ has a limit in $L^p(\Lambda)$, which we call p_∞. To conclude, put $\mu_\infty := (p_\infty)_\# \Lambda$: it is immediate to see that $W_p(\mu_n, \mu_\infty) \le \|d(p_n, p_\infty)\|_{L^p(\Lambda)} \to 0$. □

Theorem 1.3.6 (Characterisation of convergence). *The following conditions are equivalent:*

1. $\mu_n \xrightarrow{W_p} \mu$;

2. $\mu_n \rightharpoonup \mu$ *and* $\int d(x, x_0)^p d\mu_n \to \int d(x, x_0)^p d\mu$ *for one/every* $x_0 \in X$ *(convergence of moments).*

In particular, if X is bounded, then the second condition is redundant and W_p metrizes the narrow convergence (hence if X is compact then $(\mathscr{P}_p(X), W_p)$ is compact as well).

Remark 1.3.7 (Narrow convergence is metrizable). One can always put on X the bounded metric $\tilde{d}(x, y) := d(x, y) \wedge 1$, which has the same convergent and Cauchy sequences of d. So the concept of narrow convergence is the same, and (X, d) is complete (resp. separable) if and only if $(X, \tilde{d})$ is complete (resp. separable). We conclude that even for unbounded X, there exists a distance on $\mathscr{P}(X)$ which metrizes the narrow convergence, which is complete (resp. separable) if X is complete (resp. separable).

Proof. $\boxed{1 \Rightarrow 2}$ The convergence of moments is simply $W_2(\mu_n, \delta_{x_0})^2 \to W_2(\mu, \delta_{x_0})^2$. For ϕ bounded *and Lipschitz*, denoting by π_n an element of $\Gamma_o(\mu_n, \mu)$, we have

$$\left| \int \phi \; d(\mu_n - \mu) \right| = \left| \int (\phi(x) - \phi(y)) \; d\pi_n(x,y) \right| \leq \operatorname{Lip}(\phi) \int d(x,y) d\pi_n(x,y) \leq$$

$$\leq \operatorname{Lip}(\phi) \left(\int d(x,y)^p d\pi_n(x,y) \right)^{1/p} = \operatorname{Lip}(\phi) W_p(\mu_n, \mu) \to 0.$$

For general $\phi \in C_b$, take ϕ_n, ϕ^n Lipschitz functions such that $\phi_n \uparrow \phi, \phi^n \downarrow \phi$ and $-\|\phi\|_\infty \leq \phi_n \leq \phi \leq \phi^n \leq \|\phi\|_\infty$. The conclusion follows noting that

$$\int \phi d\mu = \sup_n \int \phi_n d\mu = \sup_n \lim_m \int \phi_n d\mu_m \leq \liminf_m \sup_n \int \phi_n d\mu_m = \liminf_m \int \phi d\mu_m,$$
$$\int \phi d\mu = \inf_n \int \phi^n d\mu = \inf_n \lim_m \int \phi^n d\mu_m \geq \limsup_m \inf_n \int \phi^n d\mu_m = \limsup_m \int \phi d\mu_m.$$

$\boxed{2 \Rightarrow 1,\ X \text{ compact}}$ For the remarked equivalence of the various Wasserstein metrics in this case, it is sufficient to prove that $W_1(\mu_n, \mu) \to 0$. We know that $W_1(\mu_n, \mu) = \max_{\operatorname{Lip}(\phi) \leq 1} \int \phi \; d(\mu_n - \mu)$ (see Example 1.2.12); of course we can restrict the supremum to $E := \{\phi : \operatorname{Lip}(\phi) \leq 1 \text{ and } \phi(x_0) = 0\}$, which is compact by Ascoli-Arzelà's theorem.

Define on E the functional $L_n : \phi \mapsto \int \phi \; d(\mu - \mu_n)$. L_n are equicontinuous and equibounded (easy) and their pointwise limit is 0, therefore they converge uniformly to 0, which is what we need.

$\boxed{2 \Rightarrow 1,\ X \text{ general}}$ Fix x_0 such that 2. holds. We only need to prove that $\{\mu_n\}_n$ is relatively compact in $\mathscr{P}_p(X)$, because then by the proven implication all the limit points must be μ. We note that $\sigma_n := d(\cdot, x_0)^p \mu_n \rightharpoonup d(\cdot, x_0)^p \mu =: \sigma$: in fact the measure of the whole space converges by hypothesis, and the measure of an open set A satisfies $\sigma(A) \leq \liminf_n \sigma_n(A)$ as easily proven taking continuous functions $\phi_m \uparrow I_A d(\cdot, x_0)^p$ (they exist by lower semicontinuity). Then by Prokhorov's theorem for every $\varepsilon > 0$ there is K_ε compact such that $\sigma_n(X \setminus K_\varepsilon) < \varepsilon$ for all n. Put $\mu_{n,\varepsilon} := I_{K_\varepsilon} \mu_n + (1 - \mu_n(K_\varepsilon)) \delta_{x_0}$, relatively compact (with ε fixed) thanks to the case "X compact". With a diagonal argument we find a subsequence such that $\mu_{n(l),1/k}$ W_p-converges for $l \to \infty$ for every $k \in \mathbb{N}_0$.

Thanks to the completeness of $\mathscr{P}_p(X)$ we must only show that $\mu_{n(l)}$ is a Cauchy sequence. But by the triangular inequality

$$\limsup_{l,l' \to \infty} W_p(\mu_{n(l)}, \mu_{n(l')}) \leq 2 \limsup_{l \to \infty} W_p(\mu_{n(l)}, \mu_{n(l),1/k}).$$

Finally, we estimate the latter $\limsup$ using the plan from μ_n to $\mu_{n,1/k}$ that keeps the mass in $K_{1/k}$ fixed and moves the rest to x_0: we conclude

$$\text{RHS} \leq 2 \limsup_{l \to \infty} \int_{X \setminus K_{1/k}} d(x, x_0)^p \; d\mu_{n(l)} = 2 \limsup_{l \to \infty} \sigma_{n(l)}(X \setminus K_{1/k}) \leq \frac{2}{k}.$$

□

Remark 1.3.8 (Two other characterisations). Suppose that $(\mu_n) \subseteq \mathscr{P}_p(X)$ converges narrowly to $\mu \in \mathscr{P}_p(X)$. Then $\mu_n \xrightarrow{W_p} \mu$ if and only if one of the following holds (for one/every x_0):

a. $\int d(x,x_0)^p d\mu_n(x) \to \int d(x,x_0)^p d\mu(x)$;

b. $\lim\limits_{R\to\infty} \limsup\limits_{n\to\infty} \int_{X\setminus B_R(x_0)} d(x,x_0)^p d\mu_n = 0$;

c. $d(\cdot,x_0)^p \mu_n \rightharpoonup d(\cdot,x_0)^p \mu$.

In fact a. is known by the Theorem above; a.⇒c. was seen in its proof; c.⇒b. because every narrowly convergent sequence is tight; and b.⇒a. taking $\phi_R \in C_b$ equal to $d(\cdot,x_0)^p$ in $B_R(x_0)$ and to 0 out of $B_{R+1}(x_0)$, so that $\int \phi_R d\mu_n \to \int \phi_R d\mu$ gives the thesis for $R \to \infty$.

Using *b.*, with the same argument, we get even more. Namely, take a continuous function with **p-growth**, i.e. such that $|f(x)| \le C(1 + d(x,x_0)^p)$. Then $\int f \; d\mu_n \to \int f \; d\mu$.

With this concept of convergence, we can improve the result about stability of optimal plans stated in Theorem 1.2.21.

Proposition 1.3.9 (Stability of optimal maps)**.** *Let $\mu,\nu \in \mathscr{P}_p(\mathbb{R}^d)$ be such that there is an unique optimal plan from μ to ν, induced by a transport map T. Suppose also that $\nu_n \to \nu$ and that T_n is an optimal transport map from μ to ν_n. Then $T_n \to T$ in $L^p(\mu)$.*

Proof. We already know that $(Id \times T_n)_\# \mu \rightharpoonup (Id \times T)_\# \mu$. We note that $\nu_n \to \nu$ gives immediately the convergence of the p-moments of $(Id \times T_n)_\# \mu$ to the p-moment of $(Id \times T)_\# \mu$, so that $(Id \times T_n)_\# \mu \to (Id \times T)_\# \mu$ in $\mathscr{P}_p(\mathbb{R}^d)$.

The convergence of moments of (ν_n) yields also that $\|T_n\|_{L^p(\mu)} \to \|T\|_{L^p(\mu)}$: in particular, these norms are bounded, so that every subsequence of (T_n) has weak limit points in $L^p(\mu)$ (thought as the dual of $L^{p'}(\mu)$; recall that $p > 1$).

We now observe that for for every $\zeta \in C_b(\mathbb{R}^d)$, and calling $(e_i)_{i=1,\dots,d}$ the canonical basis of $\mathbb{R}^d$, the function $f : (\mathbb{R}^d)^2 \to \mathbb{R}$ defined by $f(x,y) := \zeta(x)\langle y, e_i\rangle$ has p-growth, so that applying the previous Remark to $(Id \times T_n)_\# \mu \to (Id \times T)_\# \mu$ we get

$$\int_{\mathbb{R}^d} \zeta(x)\langle T_n(x), e_i\rangle \; d\mu(x) \to \int_{\mathbb{R}^d} \zeta(x)\langle T(x), e_i\rangle \; d\mu(x)$$

so that any weak limit point of (T_n) is forced to be T. To sum up, $T_n \rightharpoonup T$ in $L^p(\mu)$ with converging norms, hence $T_n \to T$ in $L^p(\mu)$-norm. □

Remark 1.3.10 (Lower semicontinuity of W_p). Even when X has infinite diameter, W_p is lower semicontinuous with respect to narrow convergence: this is an immediate consequence of the stability of optimal plans (Theorem 1.2.21).

Remark 1.3.11. Unlike compactness, local compactness does not pass from X to $\mathscr{P}_p(X)$: in fact $\mathscr{P}_p(\mathbb{R}^n)$ is not locally compact. To see this, take $\mu_n := (1-\varepsilon_n)\delta_0 + \varepsilon_n \, \delta_{x_n}$ with $\varepsilon_n \to 0$ so that $\mu_n \rightharpoonup \delta_0$, but such that $\varepsilon_n \, |x_n|^2 = c > 0$ constant so that $W_p(\mu_n, \delta_0)$ does not go to zero.

1.3.1 Absolute continuity in metric spaces

In order to give an useful reformulation of Kantorovich's problem, we need to develop a theory of absolutely continuous curves with values in a general metric space.

Definition 1.3.12 (AC in metric spaces)**.** For $I \subseteq \mathbb{R}$ interval, $f : I \to X$ is called an **absolutely continuous curve** if there exists $g \in L^1(I)$ such that

$$d(f(s), f(t)) \leq \int_s^t g(r)\, dr \qquad \forall [s,t] \subseteq I. \tag{1.3.1}$$

We write $f \in AC(I; X)$.

Proposition 1.3.13 (Existence of the metric derivative)**.** *Let $f \in AC(I; X)$. Then for a.e. $t \in I$ there exists the limit*

$$\lim_{h \to 0} \frac{d(f(t+h), f(t))}{|h|} =: |f'|(t)$$

called ***metric derivative*** *of f. Moreover, $|f'|$ is the smallest g (up to negligible sets) satisfying* (1.3.1).

Proof. Without loss of generality I is bounded, so by uniform continuity $f(I)$ is precompact; take in it a dense sequence (x_n) and put $\phi_n(t) := d(f(t), x_n)$, which is in $AC(I; \mathbb{R})$ and so satisfies $\phi_n(t) - \phi_n(s) = \int_s^t \phi'(r)\, dr$. We note that $|\phi_n(t) - \phi_n(s)| \leq d(f(s), f(t))$ for all $s, t \in I$, so $|\phi_n'| \leq g$ for every g satisfying (1.3.1). Hence $|f'|(t) := \sup_n |\phi_n'(t)|$ is smaller than g. Note that if t_0 is a differentiability point of ϕ_n for every n, then

$$\liminf_{h \to 0} \frac{d(f(t_0), f(t_0+h))}{|h|} \geq \liminf_{h \to 0} \frac{|\phi_n(t_0) - \phi_n(t_0+h)|}{|h|} = |\phi_n'(t)| :$$

the supremum on n gives one inequality of the thesis.

For the converse, $|\phi_n(s) - \phi_n(t)| \leq \int_s^t |\phi_n'(r)| dr \leq \int_s^t |f'|(r)\, dr$, and by density the left hand side is arbitrarily close to $d(f(s), f(t))$. As a consequence,

$$\frac{d(f(t), f(t+h))}{|h|} \leq \frac{1}{|h|} \left| \int_t^{t+h} |f'|(r)\, dr \right| :$$

if t is a Lebesgue point of $|f'|$, for $h \to 0$ we get the desired inequality. □

Remark 1.3.14. As in the real case, it can be proven that a curve is AC if and only if $\forall\, \varepsilon > 0\ \exists \delta > 0$ such that for every collection $\{(a_i, b_i)\}_{i=1,\dots,n}$ of disjoint subintervals of I satisfying $\sum_i (b_i - a_i) < \delta$, it holds $\sum_i d(f(a_i), f(b_i)) < \varepsilon$.

Definition 1.3.15. The **length** of $f \in AC((a,b); X)$ is $\mathrm{len}(f) := \int_a^b |f'|(s)\, ds$.

Remark 1.3.16. Length is invariant by reparametrisation, and $d(f(a), f(b)) \leq \mathrm{len}(f)$. If for every pair of points we have $d(x,y) = \inf_{f(0)=x, f(1)=y} \mathrm{len}(f)$, we say that X is a **length space**; if in addition the infimum is attained, that X is **geodesic**, and every minimizing f is called a **minimizing geodesic**. From now on, we will omit the word "minimizing" when it is clear that we are not speaking of geodesics in the classical differential sense.

Lemma 1.3.17 (Reparametrisation by arclength)**.** *Given $f \in AC((a,b);X)$ with $\mathrm{len}(f) = L$, there exist a 1-Lipschitz curve $\tilde{f} : [0,L] \to X$ and an AC nondecreasing reparametrisation σ which maps $[a,b]$ onto $[0,L]$, such that $|\tilde{f}'| = 1$ a.e. and $f = \tilde{f} \circ \sigma$.*

Proof. Put $\sigma(t) := \int_a^t |f'|(r)\ dr$, $\tau(s) := \min\{t \in [a,b] : \sigma(t) = s\}$, so that $\sigma(\tau(s)) = s$ and $\tau(\sigma(t)) \le t$. Note that $f(\tau \circ \sigma(t)) = f(t)$ since

$$d(f(t), f(\tau \circ \sigma(t))) \le \int_{\tau\circ\sigma(t))}^{t} |f'|(r)\ dr = \sigma(t) - \sigma(\tau \circ \sigma(t)) = 0.$$

Take $\tilde{f} := f \circ \tau$: it is 1-Lipschitz (and so $|f'|$ exists a.e. and is at most 1) because

$$d(f \circ \tau(s_1), f \circ \tau(s_2)) \le \int_{\tau(s_1)}^{\tau(s_2)} |f'|(r)\ dr = \sigma(\tau(s_2)) - \sigma(\tau(s_1)) = s_2 - s_1.$$

On the other hand, it is easily seen that for a composition of AC curves the chain rule holds, so that $|f'|(t) = |\tilde{f}'|(\sigma(t))\sigma'(t)$ a.e.: hence with a change of variables

$$\int_0^L |\tilde{f}'|(s)\ ds = \int_a^b |\tilde{f}'|(\sigma(t))\sigma'(t)\ dt = \int_a^b |f'|(t)\ dt = L$$

which gives that $|\tilde{f}'| = 1$ a.e.. □

Remark 1.3.18. For $p \in (1,\infty)$ fixed and $\gamma \in AC_p((0,1);X)$, we note that the **action** $A_p(\gamma) := \int_0^1 |\gamma'|(s)^p\ ds$ is greater or equal than $\mathrm{len}(\gamma)^p$ by Hölder's inequality, with equality if and only if $|\gamma'|$ is constant: thanks to the Lemma above,

$$\inf_{\substack{\gamma\in AC((0,1);X)\\ \gamma(0)=a,\gamma(1)=b}} \mathrm{len}(\gamma) = \inf_{\substack{\gamma\in AC((0,1);X)\\ \gamma(0)=a,\gamma(1)=b}} A_p(\gamma)^{1/p}.$$

If a geodesic exists, we can reparametrise it by arclength obtaining a **constant speed geodesic**. Note that the following are equivalent:

1. γ is a constant speed geodesic;
2. $d(\gamma(s),\gamma(t)) = |t-s| d(\gamma(0),\gamma(1))\ \forall s,t \in [0,1]$;
3. $d(\gamma(s),\gamma(t)) \le |t-s| d(\gamma(0),\gamma(1))\ \forall s,t \in [0,1]$.

1. $\Rightarrow$ 2. because for every $s < t$ and every geodesic $d(\gamma(s),\gamma(t)) = \int_s^t |\gamma'|(r)\ dr$: otherwise we would have $d(\gamma(0),\gamma(1)) > \int_0^1 |\gamma'|(r)\ dr$;

2. $\Rightarrow$ 1.: just divide by $|t-s|$ and let $s \to t$;

2. $\Leftrightarrow$ 3. because if for some $s < t$ we had the strict inequality, summing it to the weak inequality in $[0,s]$ and in $[t,1]$ we would get the contradictory $d(\gamma(0),\gamma(1)) < (s + (t-s) + (1-t))d(\gamma(0),\gamma(1))$.

Notation.

$\mathrm{Geo}(X) := \{\gamma \in AC((0,1);X) : \gamma \text{ is a constant speed geodesic}\}$.

$A(\gamma) := A_2(\gamma)$ for brevity.

$e_t : AC((0,1);X) \to X$ evaluation map $\gamma \mapsto \gamma(t)$.

Note. For the sake of clarity, in the sequel we will restrict to the case $p = 2$. The results of this section are quite easily generalized to $p > 1$ generic; but the generalisation of the results of the next chapter is much subtler, see [3].

Proposition 1.3.19 (Formula for the action)**.** *If $\gamma : [0,1] \to X$ is a continuous curve, then the two quantities*

$$A_1 := \sup_{n\geq 1} n \sum_{i=1}^{n} d^2\Big(\gamma\Big(\frac{i-1}{n}\Big), \gamma\Big(\frac{i}{n}\Big)\Big), \quad A_2 := \limsup_{n\to\infty} n \sum_{i=1}^{n} d^2\Big(\gamma\Big(\frac{i-1}{n}\Big), \gamma\Big(\frac{i}{n}\Big)\Big)$$

coincide; they are finite if and only if $\gamma \in AC_2((0,1);X)$, and in this case they are equal to $A(\gamma)$. As a corollary, A is lower semicontinuous with respect to uniform convergence.

The proof of the Proposition requires the following Lemma.

Lemma 1.3.20 (Semicontinuity of the L^p relative norm)**.** *Let $p \in (1,\infty]$, $\mu_n \rightharpoonup \mu$ probability measures on X, $\nu_n = f_n\mu_n$ finite measures, $\sup_n \|f_n\|_{L^p(\mu_n)} < \infty$. Then (ν_n) has narrow limit points, and each limit point is of the form $f\mu$ where $\|f\|_{L^p(\mu)} \leq \limsup_n \|f_n\|_{L^p(\mu_n)}$.*

Proof. For $g \in C_b(X)$ we have that

$$|\langle \nu_n, g\rangle| = \left|\int g f_n \, d\mu_n\right| \leq \|f_n\|_{L^p(\mu_n)} \|g\|_{L^{p'}(\mu_n)} \tag{1.3.2}$$

Since RHS $\leq \|f_n\|_{L^p(\mu_n)} \|g\|_\infty$, we have that ν_n are equibounded measures and so have limit points.

Now it is sufficient to prove the statement when $\nu_n \rightharpoonup \nu$. We observe that $\|g\|_{L^{p'}(\mu_n)} \to \|g\|_{L^{p'}(\mu)}$, so that letting $n \to \infty$ in (1.3.2) we get

$$|\langle \nu, g\rangle| \leq \liminf_n \|f_n\|_{L^p(\mu_n)} \|g\|_{L^{p'}(\mu)} \quad \forall g \in C_b(X),$$

i.e. $\langle \nu, \cdot\rangle$ is the restriction to $C_b(X)$ of a linear continuous functional on $L^{p'}(\mu)$ with norm less or equal than $\liminf_n \|f_n\|_{L^p(\mu_n)}$: hence it can be represented by a function $f \in L^p(\mu)$ with $\|f\|_{L^p(\mu)} \leq \liminf_n \|f_n\|_{L^p(\mu_n)}$, and the conclusion follows. □

Proof of the Proposition. If $\gamma \in AC_2((0,1);X)$ then

$$d\Big(\gamma\Big(\frac{i-1}{n}\Big), \gamma\Big(\frac{i}{n}\Big)\Big) \leq \int_{\frac{i-1}{n}}^{\frac{i}{n}} |\gamma'|(s)\, ds \overset{\text{Hölder}}{\leq} \frac{1}{\sqrt{n}} \left(\int_{\frac{i-1}{n}}^{\frac{i}{n}} |\gamma'|^2(s)\, ds\right)^{\frac{1}{2}},$$

so that, squaring and summing on i, we get $A_1 \leq A(\gamma)$.

Conversely, suppose that $A_2 < \infty$. Put

$$\mu_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{i/n} \rightharpoonup I_{[0,1]}\mathscr{L}^1, \quad \nu_n := \sum_{i=1}^{n} d\Big(\gamma\Big(\frac{i-1}{n}\Big), \gamma\Big(\frac{i}{n}\Big)\Big)\delta_{i/n}.$$

Note that $\int_0^1 \left|\frac{d\nu_n}{d\mu_n}\right|^2 d\mu_n = n\sum_{i=1}^n d^2\big(\gamma(\frac{i-1}{n}), \gamma(\frac{i}{n})\big)$. So we can apply the Lemma, obtaining a subsequence $\nu_{n(k)} \rightharpoonup f\mathscr{L}^1$ with $\int_0^1 f^2(t)\, dt \leq A_2$. We conclude if we prove that $|\gamma'| \le f$, i.e. $d(\gamma(s), \gamma(t)) \le \int_s^t f(r)\ dr$ for all $s < t$. To this purpose, take $i_{n(k)}, j_{n(k)} \in \mathbb{N}$ such that $\frac{i_{n(k)}}{n(k)} \downarrow s, \frac{j_{n(k)}}{n(k)} \uparrow t$, so that

$$d(\gamma(s),\gamma(t)) = \lim_{k\to\infty} d\left(\gamma\left(\frac{i_{n(k)}}{n(k)}\right), \gamma\left(\frac{j_{n(k)}}{n(k)}\right)\right) \overset{\text{Def } \nu_n}{\le}$$
$$\le \limsup_{k\to\infty} \nu_{n(k)}\left(\left[\frac{i_{n(k)}}{n(k)}, \frac{j_{n(k)}}{n(k)}\right]\right) \le \nu_{n(k)}([s,t]) \overset{\nu_n \rightharpoonup \nu}{\le} \int_s^t f(r)\ dr.$$

□

1.3.2 Dynamical couplings

In this section we switch point of view: instead of prescribing only the amount of mass sent from a set A to a set B, we take care also of the path along which the mass is moved.

Note. We consider on $AC_2([0,1];X)$ the measurable structure inherited from the space $C([0,1];X)$.

Problem 1.3.21 (Minimal action)**.** Given $\mu, \nu \in \mathscr{P}_2(X)$, a **dynamical coupling** of μ and ν is $\Lambda \in \mathscr{P}(AC_2([0,1];X))$ such that $(e_0)_\#\Lambda = \mu$ and $(e_1)_\#\Lambda = \nu$. In this class of Λ, we seek

$$\inf_\Lambda \int_{AC_2([0,1];X)} A(\gamma)\ d\Lambda(\gamma). \tag{MA}$$

Remark 1.3.22 (inf (MA) = min (K) if the space is geodesic)*.* Note that

$$\int_{AC_2} A\ d\Lambda \ge \int_{AC_2} d^2(\gamma(0),\gamma(1)) d\Lambda(\gamma) = \int_{X^2} d^2(x,y)\ d(e_0,e_1)_\#\Lambda(x,y) \ge W_2^2(\mu,\nu)$$

so that inf (MA) $\ge$ min (K). Moreover, the inequalities above are equalities if and only if Λ is concentrated on $\mathrm{Geo}(X)$ and $(e_0,e_1)_\#\Lambda \in \Gamma_o(\mu,\nu)$. By a standard measurable selection argument we get that if X is geodesic, then there is at least an **optimal geodesic plan** Λ realizing the equality min (MA) = min (K): indeed, it suffices to choose measurably for every couple $(x,y) \in X \times X$ a constant speed geodesic $\Phi(x,y)$ from x to y, and to put $\Lambda := \Phi_\#\pi$ where π is any optimal plan from μ to ν. We will denote by $\mathrm{OptGeo}(\mu,\nu)$ the set of all optimal geodesic plans from μ to ν.

More generally, one sees that in a length space inf (MA) = min (K).

Theorem 1.3.23 ($\mathrm{Geo}(\mathscr{P}_2(X)) \leftrightarrow$ optimal geodesic plans)**.** *Let X be a geodesic space. Then $(\mu_t)_{t\in[0,1]} \subseteq \mathscr{P}_2(X)$ is a constant speed geodesic if and only if it is of the form $\mu_t = (e_t)_\#\Lambda$ for some $\Lambda \in \mathrm{OptGeo}(\mu_0,\mu_1)$. In particular, $\mathscr{P}_2(X)$ is geodesic too.*

Proof of the Theorem, "if" part (which does not use that X is geodesic). Thanks to the previous Remark,

$$W_2^2\big((e_t)_\#\Lambda,(e_s)_\#\Lambda\big)\le\int_{\mathrm{Geo}(X)}d^2\big(\gamma(s),\gamma(t)\big)d\Lambda(\gamma)=$$
$$=|t-s|^2\int_{\mathrm{Geo}(X)}d^2\big(\gamma(0),\gamma(1)\big)d\Lambda(\gamma)=|t-s|^2W_2^2\big((e_0)_\#\Lambda,(e_1)_\#\Lambda\big)$$

which implies that $\big((e_t)_\#\Lambda\big)_t\in\mathrm{Geo}(\mathscr{P}_2(X))$ (see Remark 1.3.18). $\square$

For the converse implication we need two lemmas:

Lemma 1.3.24 (Ascoli-Arzelà in AC_2)**.** *Let $K\subseteq AC_2([0,1];X)$ satisfy*

1. $\sup_{\gamma\in K}A(\gamma)<\infty$;
2. *For every $n\ge 1$ there are compact sets $X_{1,n},\dots,X_{n,n}\subseteq X$ such that $\gamma\big(\frac{i}{n}\big)\in X_{i,n}$ for every $\gamma\in K$ and for every $i=1,\dots,n$.*

Then K is relatively compact with respect to uniform convergence.

Proof. Given any sequence $(\gamma_h)_{h\in\mathbb{N}}\subseteq K$, using 2. and a diagonal argument we can find a subsequence $h(k)$ such that $\big(\gamma_{h(k)}\big(\frac{i}{n}\big)\big)_{k\in\mathbb{N}}$ converges $\forall n\ge1,\forall i=1,\dots,n$. But 1. implies that K is equi-$\frac{1}{2}$-Hölder, which gives uniform convergence on $[0,1]$ by a standard argument. $\square$

Lemma 1.3.25 (tightness in $\mathscr{P}(AC_2)$)**.** *Let $\mathscr{F}\subseteq\mathscr{P}(AC_2([0,1];X))$ satisfy*

1. $\sup_{\Lambda\in\mathscr{F}}\int A(\gamma)d\Lambda(\gamma)<\infty$;
2. $\{(e_t)_\#\Lambda\}_{\Lambda\in\mathscr{F}}$ *is tight in $\mathscr{P}(X)$ for every $t\in[0,1]$.*

Then $\mathscr{F}$ is tight.

Proof. Fix $\varepsilon>0$, put $K_1:=\{\gamma\in AC_2:A(\gamma)\le M\}$ for some $M>0$. Thanks to 1., we have that

$$\Lambda(AC_2\setminus K_1)\le\frac{1}{M}\int A(\gamma)d\Lambda(\gamma)<\frac{\varepsilon}{2}$$

for an appropriate $M\gg1$ independent of $\Lambda\in\mathscr{F}$.

In addition, using 2., for every $n>0$ and every $i=0,\dots,n$ there is a compact set $K_{i,n}\subseteq X$ such that

$$(e_{i/n})_\#\Lambda\big(X\setminus K_{i,n}\big)\le2^{-i-1}2^{-n-1}\,\varepsilon\quad\forall\Lambda\in\mathscr{F}$$

or, in other terms, $\Lambda\big\{\gamma:\gamma\big(\frac{i}{n}\big)\notin K_{i,n}\big\}\le2^{-i-1}2^{-n-1}\,\varepsilon$ for all $\Lambda\in\mathscr{F}$.

We conclude defining the closed set

$$K:=K_1\cap\bigcap_{n=1}^{\infty}\bigcap_{i=1}^{n}\Big\{\gamma:\gamma\Big(\frac{i}{n}\Big)\in K_{i,n}\Big\}$$

which is compact by the previous Lemma, and observing that

$$\Lambda(AC_2\setminus K)\le\Lambda(AC_2\setminus K_1)+\sum_{n=1}^{\infty}\sum_{i=1}^{n}\Lambda\Big\{\gamma:\gamma\Big(\frac{i}{n}\Big)\notin K_{i,n}\Big\}\le\varepsilon\,.$$

$\square$

Proof of the Theorem, "only if" part.

For every $n \geq 2$, build, as in the proof of Theorem 1.3.5, $\theta_n \in \mathscr{P}(X^{n+1})$ such that $(p_{i-1}, p_i)_\# \theta_n \in \Gamma_o(\mu_{(i-1)/n}, \mu_{i/n})$ for $i = 1, \dots, n$. With a measurable selection argument, find a measurable map $X^{n+1} \to AC_2([0,1]; X)$ such that $(x_0, \dots, x_n)$ is mapped to a curve which on each interval $\left[\frac{i-1}{n}, \frac{i}{n}\right]$ is a constant speed geodesic from x_{i-1} to x_i. Call $\Lambda_n \in \mathscr{P}(AC_2([0,1]; X))$ the law of this map under θ_n.

We prove that $(\Lambda_n) \subseteq \mathscr{P}(AC_2)$ is tight using the previous lemma. As for condition 1.,

$$
\begin{aligned}
\int_{AC_2} A(\gamma) d\Lambda_n(\gamma) &= \int_{AC_2} \sum_{i=1}^{n} n \cdot d^2\Big(\gamma\Big(\frac{i-1}{n}\Big), \gamma\Big(\frac{i}{n}\Big)\Big)\ d\Lambda_n(\gamma) \overset{\text{Def } \Lambda_n}{=} \\
&= \sum_{i=1}^{n} n \cdot W_2^2(\mu_{(i-1)/n}, \mu_{i/n})) \quad \overset{(\mu_t) \in \text{Geo}}{=} \quad W_2^2(\mu_0, \mu_1). \quad (1.3.3)
\end{aligned}
$$

To prove that also 2. holds, we fix t and prove a little more, namely that $\{(e_t)_\# \Lambda_n\}_n$ is precompact in $\mathscr{P}_2(X)$. Note first that the family $\left\{\left(e_{\lfloor nt \rfloor/n}\right)_\# \Lambda_n\right\}_n$ is a subfamily of $\{\mu_t\}_{t \in [0,1]}$, which is compact as continuous image of $[0,1]$. Now, as observed many times above,

$$
W_2^2\Big(\left(e_{\lfloor nt \rfloor/n}\right)_\# \Lambda_n, (e_t)_\# \Lambda_n\Big) \leq \int_{AC_2} \Big(t - \frac{\lfloor nt \rfloor}{n}\Big) \int_{\lfloor nt \rfloor/n}^{t} |\gamma'|^2(s)\ ds\ d\Lambda_n(\gamma)
$$

which can be estimated

$$
\leq \frac{1}{n} \int_{AC_2} \int_{\frac{\lfloor nt \rfloor}{n}}^{\frac{\lceil nt \rceil}{n}} |\gamma'|^2(s)\ ds\ d\Lambda_n(\gamma) = W_2^2\left(\mu_{\lfloor nt \rfloor/n}, \mu_{\lceil nt \rceil/n}\right) \to 0
$$

by uniform continuity of $t \mapsto \mu_t$. To sum up, for every $\varepsilon > 0$ the sequence $\left((e_t)_\# \Lambda_n\right)_n$ is (definitively) in an ε-neighbourhood of a compact set, hence it is precompact.

Finally, take a subsequence $\Lambda_{n(k)} \rightharpoonup \Lambda$. By the lower semicontinuity of the action, (1.3.3) implies $\int A(\gamma) d\Lambda(\gamma) \leq W_2^2(\mu_0, \mu_1)$, hence Λ is optimal. Consider now the equality $\left(e_{\lfloor nt \rfloor/n}\right)_\# \Lambda_n = \mu_{\lfloor nt \rfloor/n}$, evaluate it at $n = n(k)$ and let $k \to \infty$. RHS $\to \mu_t$ obviously.

As for the left hand side, we first note that we can equivalently compute the limit of $(e_t)_\# \Lambda_{n(k)}$ since we have shown above that $W_2\Big(\left(e_{\lfloor nt \rfloor/n}\right)_\# \Lambda_n, (e_t)_\# \Lambda_n\Big) \to 0$. But the narrow limit of $(e_t)_\# \Lambda_{n(k)}$ is $(e_t)_\# \Lambda$, since $(e_t)_\#$ is continuous with respect to narrow convergence. In conclusion, passing to the limit as $k \to \infty$ we get $(e_t)_\# \Lambda = \mu_t$. □

Remark 1.3.26 (Non uniqueness of the geodesic). Unfortunately, even in the simplest case $X = \mathbb{R}^n$, there may be more than one geodesic between two elements of $\mathscr{P}_2(X)$. As an example, take

$$
X = \mathbb{R}^2, \quad \mu_0 = \frac{1}{2}\delta_{(-1,1)} + \frac{1}{2}\delta_{(1,-1)}, \quad \mu_1 = \frac{1}{2}\delta_{(1,1)} + \frac{1}{2}\delta_{(-1,-1)}
$$

which can be connected by either a "horizontal" or a "vertical" movement.

Definition 1.3.27 (Non-branching)**.** A metric space X is **non-branching** if for all $t \in (0,1)$ the map $(e_0, e_t) : \mathrm{Geo}(X) \to X^2$ is injective.

Remark 1.3.28. It means that two minimizing geodesics can meet again only at the final point. From this, it is easy to see that X is non-branching if and only if $(e_0, e_{\bar{t}})$ is injective for *some* $\bar{t} \in (0,1)$.

Theorem 1.3.29. *Let X be non-branching. Then:*

1. *$\mathscr{P}_2(X)$ is non-branching;*
2. *If $(\mu_s)_{s\in[0,1]} \in \mathrm{Geo}(\mathscr{P}_2(X))$, then for every $t > 0$ the optimal geodesic plan from μ_t to μ_1 is unique;*
3. *For every $(\mu_s)_{s\in[0,1]} \in \mathrm{Geo}(\mathscr{P}_2(X))$, the optimal geodesic plan Λ* such that *$(e_t)_\#\Lambda = \mu_t$ for every $t \in [0,1]$ is unique.*

Proof. Denote by $\mathrm{Geo}_I(X)$ the set of constant speed geodesic defined on the interval I, so that $\mathrm{Geo}(X) = \mathrm{Geo}_{[0,1]}(X)$. Call $res_I : \mathrm{Geo}(X) \to \mathrm{Geo}_I(X)$ the restriction map; note that $res_{[0,t]}$ is injective by hypothesis and so it has a left inverse ext_t.

Take any $\Lambda_1 \in \mathscr{P}(\mathrm{Geo}_{[0,t]}(X))$ optimal geodesic plan from μ_0 to μ_t rescaled in the time interval $[0,t]$, and $\Lambda_2 \in \mathscr{P}(\mathrm{Geo}_{[t,1]}(X))$ optimal geodesic plan from μ_t and μ_1 rescaled in $[t,1]$. We want to build a "concatenated" plan. To this aim, disintegrate both Λ_i with respect to (e_t), finding $\Lambda_i = \Lambda_i^{(x)} \otimes \mu_t$ where $\Lambda_i^{(x)}$ is concentrated on the curves γ such that $\gamma(t) = x$. Denote by $\Lambda^{(x)} \in \mathscr{P}(AC_2([0,1];X))$ the law under $\Lambda_1^{(x)} \times \Lambda_2^{(x)}$ of the concatenation of curves: under it, $\gamma(t) = x$ a.s. too. Finally, put $\Lambda := \Lambda^{(x)} \otimes \mu_t$, which clearly satisfies $(res_{[0,t]})_\#\Lambda = \Lambda_1, (res_{[t,1]})_\#\Lambda = \Lambda_2$, and in particular is a dynamical coupling of μ_0 and μ_1. We claim that Λ is an optimal geodesic plan.

Note that

$$\left(\int d^2(\gamma(0),\gamma(1))d\Lambda(\gamma)\right)^{\frac{1}{2}} \leq \left(\int \left[d(\gamma(0),\gamma(t)) + d(\gamma(t),\gamma(1))\right]^2 d\Lambda(\gamma)\right)^{\frac{1}{2}}$$
$$\overset{\text{Minkowski}}{\leq} \left(\int d^2(\gamma(0),\gamma(t))d\Lambda(\gamma)\right)^{\frac{1}{2}} + \left(\int d^2(\gamma(t),\gamma(1))d\Lambda(\gamma)\right)^{\frac{1}{2}}.$$

But in the last expression, the two integrals can be equivalently computed with respect to the measures Λ_1 and Λ_2 respectively, which are optimal (rescaled), yielding

$$\mathrm{RHS} = W_2(\mu_0,\mu_t) + W_2(\mu_t,\mu_1) = W_2(\mu_0,\mu_1).$$

Therefore $(e_0,e_1)_\#\Lambda \in \Gamma_o(\mu_0,\mu_1)$; moreover, a posteriori, all the inequalities are equalities, so that $d(\gamma(0),\gamma(1)) = d(\gamma(0),\gamma(t)) + d(\gamma(t),\gamma(1))$ for Λ-a.e. γ; since Λ-a.e. γ is a geodesic both in $[0,t]$ and in $[t,1]$, we conclude that $\gamma \in \mathrm{Geo}(X)$ Λ-a.s., which concludes by the characterisation of optimal geodesic plans.

By the non-branching hypothesis, $\gamma \in \mathrm{Geo}(X)$ implies that $\gamma = ext_t(\gamma_{|[0,t]})$, hence $\Lambda = (ext_t)_\#\Lambda_1$, and in particular Λ_2 (which was taken any optimal geodesic plan from μ_t to μ_1) is uniquely determined: 2. is proved.

A fortiori, the geodesic from μ_t to μ_1 is determined too. By symmetry of the problem, also the geodesic from μ_t and μ_0 is determined, and 1. follows.

To prove 3., take $\Lambda \in \mathscr{P}(\mathrm{Geo}(X))$ optimal such that $(e_t)_\# \Lambda = \mu_t$ for every $t \in [0,1]$. Put $\Lambda_t := (res_{[t,1]})_\# \Lambda$: it is optimal, since it is concentrated on geodesics and

$$\int d^2(\gamma(t), \gamma(1)) d\Lambda_t(\gamma) = (1-t)^2 \int d^2(\gamma(0), \gamma(1)) d\Lambda(\gamma) = W_2^2(\mu_t, \mu_1)$$

Point 2. implies that Λ_t is determined $\forall t > 0$. We claim that Λ is determined by $(\Lambda_t)_{t>0}$; it is sufficient that the measure of every element of a countable basis of the topology is determined, but this is true because

$$\Big\{\gamma : \|\gamma - \gamma_i\|_\infty < a_i,\ i = 1, \ldots, n\Big\} = \bigcap_n \Big\{\gamma : \big\|(\gamma - \gamma_i)_{|[1/n,1]}\big\|_\infty < a_i, i = 1, \ldots, n\Big\}$$

where the right hand side can be computed using the measures $\Lambda_{1/n}$: the conclusion follows. □

Chapter 2

Differential structure of $\mathscr{P}_2(\mathbb{R}^n)$

In this chapter, we fix the space $\mathbb{R}^n$ and the cost $c(x,y) := \frac{1}{2}|x-y|^2$, and show that there is a natural concept of velocity field tangent to a curve in $\mathscr{P}_2(\mathbb{R}^n)$. This will suggest the construction of a Riemannian-like structure in $\mathscr{P}_2(\mathbb{R}^n)$; with respect to this structure, many important functionals are (sub)differentiable, and their gradient flow can be identified with the solution of some well known parabolic PDEs.

2.1 Velocity fields and AC curves

Let $\mu_0, \mu_1 \in \mathscr{P}_2(\mathbb{R}^n)$, $\Lambda \in \mathrm{OptGeo}(\mu_0,\mu_1)$. We know that $(e_0,e_1)_\#\Lambda =: \gamma$ is an element of $\Gamma_o(\mu_0,\mu_1)$; since Λ is concentrated on $\mathrm{Geo}(\mathbb{R}^n)$, and the unique constant speed geodesic between two points $x,y \in \mathbb{R}^n$ is $t \mapsto g_t(x,y) := (1-t)x + ty$, we conclude that $\Lambda = g_\#\gamma$, and so $(e_t)_\#\Lambda = (g_t)_\#\gamma$. To sum up, the generic constant speed geodesic between μ_0 and μ_1 is $\mu_t := ((1-t)p_x + tp_y)_\#\gamma$ with $\gamma \in \Gamma_o(\mu_0,\mu_1)$ and $p_x, p_y : (\mathbb{R}^n)^2 \to \mathbb{R}^n$ canonical projections.

Remark 2.1.1. With the same argument, if in addition $\mu_0 \ll \mathscr{L}^n$, and T is the optimal transport map from μ_0 to μ_1 (given by Brenier's theorem), then $\mu_t = ((1-t)Id + tT)_\#\mu_0$. Put $T_t := (1-t)Id + tT$. Since (a suitable version of) T is the gradient of a convex function, then (for this choice) $\langle Tx - Ty, x-y\rangle \geq 0$ and so $\langle T_t x - T_t y, x-y\rangle \geq (1-t)|x-y|^2$: as a consequence, T_t is injective with Lipschitz inverse. In particular $\mu_t \ll \mathscr{L}^n$: in fact

$$\mathscr{L}^n(B) = 0 \Rightarrow \mathscr{L}^n(T_t^{-1}(B)) = 0 \Rightarrow \mu_0(T_t^{-1}(B)) = 0 \Rightarrow \mu_t(B) = 0.$$

For every $(\mu_t) \in \mathrm{Geo}(\mathscr{P}_2(\mathbb{R}^n))$, we can give a description of the evolution of μ_t in terms of a "velocity field":

Proposition 2.1.2. *Let $(\mu_t)_{t\in[0,1]}$ be a constant speed geodesic in $\mathscr{P}_2(\mathbb{R}^n)$, and let $\Lambda \in \mathrm{OptGeo}(\mu_0,\mu_1)$ be such that $\mu_t = (e_t)_\#\Lambda$. Put $\gamma := (e_0,e_1)_\#\Lambda$. Then:*

1. *μ_t satisfies in $(0,1)\times\mathbb{R}^n$ the so called "continuity equation"*

$$\frac{d}{dt}\mu_t + \nabla\cdot(v_t\mu_t) = 0 \quad \textit{(in distributional sense: see the Definition below)}$$

for a velocity field $v_t : \mathbb{R}^n \to \mathbb{R}^n$ defined implicitly by $v_t\mu_t = (g_t)_\#(p_y - p_x)\gamma$. Moreover,

$$\int_{\mathbb{R}^n} |v_t|^2 \, d\mu_t \le W_2^2(\mu_0, \mu_1) \qquad \forall t \in [0,1]. \tag{2.1.1}$$

2. *Suppose in addition that $\mu_0 \ll \mathscr{L}^n$; call ϕ_1 a Kantorovich potential for the couple (μ_0, μ_1) with respect to the cost $c(x,y) = \frac{1}{2}|x-y|^2$ (see the proof of Brenier's theorem). Then $\phi_t := t\phi_1$ is a Kantorovich potential for (μ_0, μ_t); moreover, putting $\psi_t := \phi_t^c$, the v_t defined in the first part of the Theorem is equal to $\dfrac{\nabla\psi_t}{t}$ for every $t \in (0,1)$.*

Note. Benamou-Brenier's theorem (Theorem 2.1.15 below) will imply that (2.1.1) is in fact an equality for a.e. t; while Proposition 2.1.18 will tell us that even the v_t of part 1. is "almost a gradient".

Proof. $\boxed{1.}$ First of all, v_t is well defined, because $(p_y - p_x)\gamma \ll \gamma$ implies that $(g_t)_\#(p_y - p_x)\gamma \ll (g_t)_\#\gamma = \mu_t$. By this argument we also easily get that

$$\int |v_t|^2 \, d\mu_t \le \int |y-x|^2 \, d\gamma(x,y) = W_2^2(\mu_0, \mu_1).$$

To prove that (CE) is satisfied, just observe that for every $\phi \in C_c^\infty((0,1) \times \mathbb{R}^n)$ it holds

$$\begin{aligned} \int_0^1 \int_{\mathbb{R}^n} \partial_t\phi(t,x) \, d\mu_t(x)dt &= \int_0^1 \int_{\mathbb{R}^n} \partial_t\phi(t, g_t(x,y)) \, d\gamma(x,y)dt = \\ &= \int_0^1 \int_{\mathbb{R}^n} \left\{ \frac{d}{dt}\big[\phi(g_t(x,y),t)\big] - \big\langle \nabla_x\phi(g_t(x,y),t), y-x \big\rangle \right\} d\gamma(x,y)dt \end{aligned}$$

where the integral of the total derivative is zero thanks to the compactness of the support, so the last expression is

$$= -\int_0^1 \int_{\mathbb{R}^n} \nabla_x\phi(z,t) \cdot v_t(z) d\mu_t(z)dt$$

as we wanted.

$\boxed{2.}$ We still denote by T the optimal transport map from μ_0 to μ_1, and by T_t the function $(1-t)Id + tT$.

Fix $t \in (0,1)$. Recall that $T = \nabla f$ with f convex; in the proof of Brenier's theorem we also saw that $\nabla f(x) = x - \nabla\phi_1(x)$. Since T_t is the gradient of a convex function (it inherits this property from T), then it is optimal thanks to Brenier's theorem, and again $T_t(x) = x - \nabla\phi_t(x)$: this implies $\nabla\phi_t = t\nabla\phi_1$, from which the first assertion.

We have seen that $\mu_t \ll \mathscr{L}^n$, which implies that $S_t(y) = y - \nabla\psi_t(y)$ is the optimal transport map from μ_t to μ_0; it satisfies $T_t \circ S_t = Id$ μ_0-a.s. (Remark 1.2.18).

To conclude, take any ξ bounded and Borel, and observe that

$$\int \xi v_t d\mu_t = \int \xi((1-t)x+ty)(y-x)[d(Id\times T)_{\#}\mu_0](x,y) =$$
$$= \int \xi((1-t)x+tT(x))(T(x)-x)d\mu_0(x) = \int \xi(T_t(x))\frac{T_t(x)-x}{t}d(S_t)_{\#}\mu_t(x) =$$
$$= \int \xi(y)\frac{y-S_t(y)}{t}d\mu_t(y) = \int \xi\frac{\nabla\psi_t}{t}d\mu_t.$$

□

Remark 2.1.3 (Eulerian description of geodesics). By definition of c-transform, the ψ_t defined in the second part of the previous Theorem satisfies

$$\frac{\psi_t(y)}{t} = \inf_x \left(\frac{|x-y|^2}{2t} - \phi_1(x)\right).$$

The right hand side is the so-called **Hopf-Lax formula** for the solution of the **Hamilton-Jacobi equation**

$$\begin{cases}\partial_t u_t + \frac{1}{2}|\nabla u_t|^2 = 0\\ u_0 = -\phi_1.\end{cases}$$

(A complete treatment of this kind of PDE can be found in [9].)

To sum up, we have proven that also $\mathscr{P}_2^a(\mathbb{R}^n)$ is a geodesic space, and that its constant speed geodesics satisfy the equations

$$\begin{cases}\frac{d}{dt}\mu_t + \nabla\cdot(v_t\mu_t) = 0\\ v_t = \nabla\psi(t,\cdot)\\ \partial_t\psi + \frac{1}{2}|\nabla\psi|^2 = 0.\end{cases} \tag{2.1.2}$$

This fact is sometimes referred to as the **Eulerian description of geodesics**. For a formal variational argument suggesting it, see Remark 2.1.24 below.

In order to continue the analysis of the curves in $\mathscr{P}_2(\mathbb{R}^n)$, we have to study the solutions of the continuity equation in some detail.

2.1.1 The continuity equation

Definition 2.1.4. Given $(\mu_t)_{t\in[t_0,T]} \subseteq \mathscr{P}(\mathbb{R}^n)$, and $(x,t)\mapsto v_t(x)\in\mathbb{R}^n$ a Borel "velocity field" defined for $(t,x)\in[t_0,T]\times\mathbb{R}^n$, we say that $(\mu_t,v_t)_{t\in[t_0,T]}$ satisfies the **continuity equation** in $[t_0,T]$ if $\int_{t_0}^T\int_{\mathbb{R}^n}|v_t|\,d\mu_t\,dt<\infty$, and

$$\frac{d}{dt}\mu_t + \nabla\cdot(v_t\mu_t) = 0 \tag{CE}$$

holds in the sense of distributions in $(t_0,T)\times\mathbb{R}^n$; i.e., if it holds

$$\int_{t_0}^T\int_{\mathbb{R}^n}\big(\partial_t\phi(x,t)+\langle v_t(x),\nabla_x\phi(x,t)\rangle\big)\,d\mu_t\,dt = 0 \qquad \forall\phi\in C_c^\infty((t_0,T)\times\mathbb{R}^n).$$

(Then by density it is true also for $\phi\in C^1$.)
If $t\mapsto\mu_t$ is narrowly continuous, we will call (μ_t) a **continuous solution** of (CE).

Remark 2.1.5 (Time reparametrisation). Of course by a translation we can always assume that $t_0 = 0$. Moreover, if $\tau : [t_0', T'] \to [t_0, T]$ is a C^1 diffeomorphism (a "time reparametrisation"), it is easy to see that (μ_t, v_t) satisfies (CE) in $[t_0, T]$ if and only if $(\mu_{\tau(t)}, \dot\tau(t) v_{\tau(t)})_t$ satisfies (CE) in $[t_0', T']$: just change variables in the distributional version of the equation.

Natural solutions of (CE) are described via the so called **characteristic equation**, which is the ordinary differential equation $u'(t) = v_t(u(t))$. We need a slight extension of the classical Cauchy-Lipschitz theorem:

Theorem 2.1.6 ("Cauchy-Lipschitz"). *Consider the Cauchy problem*

$$\begin{cases} u'(t) = v_t(u(t)) \\ u(0) = x \end{cases}$$

and call ***solution*** *of the problem in $[0,T]$ any $u \in AC_{loc}[0,T)$ such that $u(0) = x$ and the first equality is satisfied for a.e. $t \in (0,T)$.*

1. *Let $v_t : \mathbb{R}^n \to \mathbb{R}^n$ be Borel for every $t \in [0,T]$ and satisfy*

$$S := \int_0^T \left(\sup_{\mathbb{R}^n} |v_t| + \mathrm{Lip}(v_t) \right) \, dt < \infty.$$

Then for every initial condition $x \in \mathbb{R}^n$, the solution of the Cauchy problem in $[0,T]$ exists unique. Moreover, if $X_t(x)$ is this solution, then $\mathrm{Lip}(X_t) \le e^S$ for every $t \in [0,T]$.

2. *Let v_t satisfy only*

$$\int_0^T \left(\sup_{B} |v_t| + \mathrm{Lip}(v_t, B) \right) \, dt < \infty \quad \forall B \subset\subset \mathbb{R}^n .$$

Then for every initial condition $x \in \mathbb{R}^n$, the Cauchy problem above has a solution for t in a right neighbourhood of 0.

Moreover, for every x there exists a maximal solution $X_t(x)$, defined on a relatively open subinterval of $[0,T]$, such that every other solution $u(t)$ is extended by $X_t(x)$. We denote by $\tau(x)$ the supremum of the domain of the maximal solution.

Finally, if $X_t(x)$ is bounded for $t \in (0, \tau(x))$, then $X_t(x)$ is defined ***globally****, i.e. for every $t \in [0,T]$.*

Proof. Identical to the classical case. For the first part, prove that the map $F : C([0,\bar t]; \mathbb{R}^n) \to C([0,\bar t]; \mathbb{R}^n)$ given by $Fu(t) := x + \int_0^t v_s(u(s)) ds$ is well defined, and a contraction if $\bar t$ satisfies $\int_0^{\bar t} \mathrm{Lip}(v_s) \, ds < 1$. Conclude that the solution in $[0,\bar t]$ exists unique, and by a finite number of iterations of this argument get global existence and uniqueness. The bound on the Lipschitz constant is an immediate application of Gronwall's lemma.

For the second part, take $\tilde v_t(x)$ equal to $v_t(x)$ in $B_R(x)$ and to $v_t\left(\frac{x}{|x|} R\right)$ otherwise, and consider the ODE with $\tilde v$ in the place of v. To this apply the first part. Observe that the solution to this ODE starting from x is also a solution of the original Cauchy problem for t in a right neighbourhood of 0: hence a solution exists. On the other hand, two solutions which coincide for some t coincide in a neighbourhood of that t (because locally they are solutions for a suitable $\tilde{v}$ to which the first part applies): uniqueness follows. The existence of a maximal solution is a straightforward construction; this solution is defined on a relatively open set since otherwise it could be extended.

Finally, if a solution u defined on $[0,\tilde{t})$ is bounded, then there is a sequence $t_n \uparrow \tilde{t}$ such that $u(t_n)$ converges to a certain y. The hypothesis on the supremum of v easily implies that $u(t) \to y$ as $t \uparrow \tilde{t}$. To the limit, one sees that $y = u(0) + \int_0^{\tilde{t}} v_s(u(s))\, ds$, so that putting $u(\tilde{t}) = y$ we have a solution on the closed set $[0,\tilde{t}]$; if $\tilde{t} < T$ it can be extended, from which the thesis follows. □

Remark 2.1.7 (Internal initial conditions). More generally, without changing the hypotheses, given $s \in (0,T)$ one has existence and uniqueness (resp. global and local) of a solution such that $u(s) = x$. The maximal solution with this "initial" condition will be denoted by $X_t(x,s)$. In case 1., the bound on the Lipschitz constant remains true as well.

Hypothesis 2.1.8. From now on, v_t will at least satisfy the hypotheses of part 2. of the previous Theorem, and X_t will denote the maximal solution of the characteristic equation.

Proposition 2.1.9 (An explicit solution)**.** *Let $\mu_0 \in \mathscr{P}(\mathbb{R}^n)$, $\bar{t} > 0$, and suppose that $X_t(x)$ is defined in $[0,\bar{t}]$ for μ_0-a.e. x. For $t \in [0,\bar{t}]$, put $\mu_t := (X_t)_\# \mu_0$, and suppose to know a priori that $\int_0^T \int_{\mathbb{R}^n} |v_t|\, d\mu_t\, dt < \infty$. Then (μ_t) is a continuous solution of* (CE) *in* $[0,\bar{t}]$.

Proof. μ_t is continuous: if $\zeta \in C_b(\mathbb{R}^n)$ then

$$\lim_{s\to t} \int \zeta d\mu_s = \lim_{s\to t} \int \zeta(X_s(x))\, d\mu_0(x) \overset{\text{Lebesgue}}{=} \int \zeta(X_t(x))\, d\mu_0(x) = \int \zeta d\mu_t.$$

Take $\phi \in C_c^\infty(\mathbb{R}^n \times (0,\bar{t}))$. We note that $t \to \phi_x(t) := \phi(X_t(x),t)$ is absolutely continuous with derivative $\dot{\phi}_x(t) = \partial_t \phi(X_t(x),t) + \langle \nabla\phi(X_t(x),t), v_t(X_t(x))\rangle$, hence

$$0 \overset{\text{compact supp}}{=} \int_{\mathbb{R}^n} (\phi_x(\bar{t}) - \phi_x(0))\, d\mu_0(x) = \int_{\mathbb{R}^n} \left(\int_0^{\bar{t}} \dot{\phi}_x(t) dt \right) d\mu_0(x)$$

and if we could apply Fubini's theorem we would get:

$$= \int_0^{\bar{t}} \int_{\mathbb{R}^n} (\partial_t \phi + \langle \nabla\phi, v\rangle)(X_t(x),t)\ d\mu_0(x) dt = \int_0^{\bar{t}} \int_{\mathbb{R}^n} (\partial_t\phi + \langle\nabla\phi, v\rangle)(x,t)\ d\mu_t(x) dt$$

hence this expression would be equal to 0, as desired. Fubini can be applied because

$$\int_0^{\bar{t}} \int_{\mathbb{R}^n} |\partial_t\phi + \langle\nabla\phi, v\rangle|(x,t)\ d\mu_t(x)dt \le \mathrm{Lip}(\phi)T + \int_0^T \int_{\mathbb{R}^n} \mathrm{Lip}(\phi)|v_t(x)|\ d\mu_t(x) dt < \infty.$$

□

Lemma 2.1.10 (Test functions nonzero at the extremes)**.** *Let $(\mu_t)_{t\in[0,T]}$ be a continuous solution of* (CE)*. Then for every $\phi \in C_c^1(\mathbb{R}^n \times [0,T])$ and every $[t_1,t_2] \subseteq [0,T]$ it holds*

$$\int_{\mathbb{R}^n} \phi(x,t_2)\ d\mu_{t_2} - \int_{\mathbb{R}^n} \phi(x,t_1)\ d\mu_{t_1} = \int_{t_1}^{t_2} \int_{\mathbb{R}^n} (\partial_t \phi + \langle \nabla_x \phi, v\rangle)(x,t)\ d\mu_t(x)dt.$$

Proof. Take $\eta_\varepsilon \in C_c^\infty(0,T)$ with values in $[0,1]$, converging pointwise to $I_{(t_1,t_2)}$ with $\eta_\varepsilon' dt \rightharpoonup \delta_{t_2} - \delta_{t_1}$, use $\eta_\varepsilon(t)\phi(x,t)$ as a test function for (CE), and let $\varepsilon \to 0$. □

Corollary 2.1.11. *If (μ_t) is a continuous solution of* (CE) *in $[0,t]$ and in $[t,T]$, then it is a solution in $[0,T]$.*

Proof. Take any test function for (CE) in $[0,T]$; apply the Lemma in $[0,t]$ and in $[t,T]$; and sum the two equalities. □

Remark 2.1.12 (Transport equation). The quantity $\partial_t\phi + \langle \nabla_x \phi, v\rangle$, at least with smoothness hypotheses, can be made into any function ψ. Precisely, the classical (backward) **transport equation**

$$\begin{cases} \partial_t \phi(x,t) + \langle \nabla_x \phi(x,t), v_t(x)\rangle = \psi(x,t) & \forall t \in (0,T)\ \forall x \in \mathbb{R}^n \\ \phi(x,T) = \phi_T(x) \quad \forall x \in \mathbb{R}^n \end{cases}$$

has existence and uniqueness of solutions at least when ϕ_T is C^1, ψ is C_b, and $X_t(x)$ is globally defined and satisfies $\frac{d}{dt}X_t(x) = v_t(X_t(x))$ for *all* t, x.

In fact, if ϕ is a solution, then $\frac{d}{dt}\phi(X_t(x),t) = \psi(X_t(x),t)$, or integrating, $\phi(X_T(x),T) - \phi(X_t(x),t) = \int_t^T \psi(X_s(x),s)\ ds$. This, replacing x with $X_0(x,t)$, becomes

$$\phi(x,t) = \phi_T(X_T(x,t)) - \int_t^T \psi(X_s(x,t),s)\ ds. \tag{2.1.3}$$

Conversely, taking this as the definition of ϕ, it is straightforward to verify that ϕ satisfies the transport equation.

Theorem 2.1.13 (Comparison and uniqueness)**.** *Let $(\sigma_t)_{t\in[0,T]}$ be a continuous family of signed measures such that*

$$\int_0^T \int_{\mathbb{R}^n} |v_t|\ d|\sigma_t|\ dt < \infty$$

and $\int_0^T |\sigma_t|(B)\ dt < \infty$ for every $B \subset\subset \mathbb{R}^n$. Suppose that (σ_t, v_t) satisfies (CE)*.*

Then $\sigma_0 \le 0$ implies $\sigma_t \le 0$ $\forall t$. As a consequence, given v_t, there is at most one continuous family of probability measures starting from a given μ_0 and solving (CE) *in $[0,T]$.*

Proof. It is sufficient to prove that $\int \psi\ d\sigma_t\ dt \le 0$ for every $\psi \in C_c^\infty(\mathbb{R}^n \times (0,T))$ nonnegative (for homogeneity, we can also suppose $\psi \le 1$). In fact, in this case, for $\psi(x,t) = f(x)g(t)$, we would get

$$\int_0^T \left(\int_{\mathbb{R}^n} f(x)\ d\sigma_t(x) \right) g(t)dt \le 0 \qquad \forall f, g \ge 0$$

where the quantity in parentheses is continuous in t: for arbitrariness of g, this quantity would be nonpositive for every $f \geq 0$, and we would conclude.

The idea to prove the claim is to apply Lemma 2.1.10 to a function ϕ such that $(\partial_t \phi + \langle \nabla \phi, v \rangle)(x,t) = \psi$ and $\phi(\cdot, T) \equiv 0$, which would give immediately the conclusion. By the previous Remark we know that such a ϕ exists if v_t is regular, so rigorously we need an approximation argument.

To understand what kind of approximation we need, we suppose that we have functions $v^k \in C_c^\infty(\mathbb{R}^n \times [0,T])$ which tend to v in some sense to be established later; we call ϕ^k the solution of the transport equation

$$\begin{cases} \partial_t \phi^k(x,t) + \langle \nabla_x \phi^k(x,t), v_t^k(x) \rangle = \psi(x,t) \quad \forall t \in (0,T)\ \forall x \in \mathbb{R}^n \\ \phi^k(x,T) = 0 \quad \forall x \in \mathbb{R}^n \end{cases}$$

which is C^∞ and has values in $[-T, 0]$ by the previous Remark.

We want to mimic the computation we would do in the regular case; but since we cannot expect good convergence on the whole $\mathbb{R}^n$, we fix $R > 0$ and introduce a cut-off function ξ_R equal to 1 on $B_R(0)$ and to 0 out of $B_{2R}(0)$, such that $|\nabla \xi_R| \leq \frac{2}{R}$. Now we compute

$$0 \geq -\int \phi^k \xi_R d\sigma_0 \overset{\text{Lemma 2.1.10}}{=} \int_0^T \int_{\mathbb{R}^n} \left\{ \xi_R \partial_t \phi^k + \langle v_t, \xi_R \nabla \phi^k + \phi^k \nabla \xi_R \rangle \right\} d\sigma_t\ dt =$$

$$= \int_0^T \int_{\mathbb{R}^n} \xi_R \psi\ d\sigma_t\ dt + \int_0^T \int_{\mathbb{R}^n} \xi_R \langle v_t - v_t^k, \nabla \phi^k \rangle d\sigma_t\ dt + \int_0^T \int_{\mathbb{R}^n} \phi^k \langle v_t, \nabla \xi_R \rangle d\sigma_t\ dt.$$

The first integral is the term we wanted; the third is bounded by

$$\int_0^T \int_{\mathbb{R}^n} T|v_t| \frac{2}{R}\ d|\sigma_t|\ dt \xrightarrow{R \to \infty} 0;$$

so, if we had that

$$\int_0^T \int_{\mathbb{R}^n} \xi_R \langle v_t - v_t^k, \nabla \phi^k \rangle d\sigma_t\ dt \xrightarrow{k \to \infty} 0, \tag{2.1.4}$$

letting first $k \to \infty$ and then $R \to \infty$ we would get $\int_0^T \int_{\mathbb{R}^n} \psi\ d\sigma_t\ dt \leq 0$ as desired.

It remains to find an approximating sequence v^k such that (2.1.4) holds.

Firstly, we note that (2.1.4) does not depend on the values of $v_t(x)$ for x out of $B_{2R}(0)$, so we can suppose that $v_t(x)$ is zero out of a large ball and that

$$\int_0^T \left(\sup_{\mathbb{R}^n} |v_t| + \mathrm{Lip}(v_t) \right)\ dt < \infty.$$

Secondly, we observe that the two following properties would be sufficient:

a. $v^k \to v$ in $L^1(B_{2R}(0) \times [0,T], \sigma_t \otimes dt)$;

b. $\nabla \phi^k$ is equibounded in $B_{2R}(0)$.

Moreover, by the explicit representation (2.1.3) of ϕ^k, b. is true if $\left(\nabla X_s^k\right)_k$ are equibounded; equivalently, if $\mathrm{Lip}(X_s^k(t,\cdot))$ is bounded by a constant independent of k, t. We obtain this from the "Cauchy-Lipschitz" theorem above if it holds:

b'. $\displaystyle\int_0^T \left(\sup_{\mathbb{R}^n} |v_t^k| + \mathrm{Lip}(v_t^k) \right) dt \le C < \infty.$

Eventually, the existence of $v^k \in C_c^\infty(\mathbb{R}^n \times [0,T])$ satisfying a. and b'. follows by a standard mollification argument. ☐

Theorem 2.1.14 (Characterisation of solutions). *Let $(\mu_t)_{t\in[0,T]}$ be continuous family of probability measures on $\mathbb{R}^n$ which solves* (CE)*. Then for μ_0-a.e. initial condition $x \in \mathbb{R}^n$, the maximal solution $X_t(x)$ to the associated ODE is defined globally, and $\mu_t = (X_t)_\#\mu_0$.*

Proof. Call $\tau(x)$ the maximum $t \in [0,T]$ such that the maximal solution is defined in $[0,\tau(x))$. We want to prove that the solution is μ_0-a.s. bounded in $(0,\tau(x))$, so that it is globally defined in $[0,T]$ thanks to Theorem 2.1.6. Since we know that in this case $(X_t)_\#\mu_0$ solves (CE) (Proposition 2.1.9), the uniqueness of the solution will then give the desired result.

For $s \in [0,T]$, put $E_s := \{x \in \mathbb{R}^n : \tau(x) > s\}$ and consider $\nu_t^{(s)} := (X_t)_\#(I_{E_s}\mu_0)$, which satisfies (CE) in $[0,s]$ with the velocity field v_t thanks to Proposition 2.1.9. The comparison theorem gives that $\nu_t^{(s)} \le \mu_t$ for every $t \in [0,s]$. Hence

$$\int_{\mathbb{R}^n} \sup_{t\in(0,\tau(x))} |X_t(x) - x| \, d\mu_0(x) \le \int_{\mathbb{R}^n} \int_0^{\tau(x)} |v_t|(X_t(x)) \, dt \, d\mu_0(x) \overset{\text{Fubini}}{=}$$

$$= \int_0^T \int_{E_t} |v_t|(X_t(x)) \, d\mu_0(x) \, dt = \int_0^T \int_{\mathbb{R}^n} |v_t| \, d\nu_t^{(t)} \, dt \le \int_0^T \int_{\mathbb{R}^n} |v_t| \, d\mu_t \, dt < \infty$$

by hypothesis. The conclusion follows. ☐

2.1.2 Benamou-Brenier's theorem and tangent fields

The following theorem gives a "time dependent" interpretation of the W_2 distance.

Theorem 2.1.15 (Benamou-Brenier). *For every $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^n)$, it holds:*

$$W_2^2(\mu,\nu) = \min\left\{ \int_0^1 \int_{\mathbb{R}^n} |v_t|^2 \, d\mu_t : \frac{d}{dt}\mu_t + \nabla\cdot(v_t\mu_t) = 0, \mu_0 = \mu, \mu_1 = \nu \right\}. \quad \text{(BB)}$$

Proof. At the very beginning of the chapter we found a solution $(\mu_t, v_t)_t$ of (CE) with $\mu_0 = \mu, \mu_1 = \nu$ generic elements of $\mathscr{P}_2(\mathbb{R}^n)$, such that $\int |v_t|^2 \, d\mu_t \le W_2^2(\mu_0,\mu_1)$: hence we only need to prove the inequality $\le$.

Take any admissible couple (μ_t, v_t); we can assume that $\int_0^1 |v_t|^2 d\mu_t \, dt < \infty$, since otherwise the inequality is trivial. We want to use the characterisation of the solutions of (CE) in terms of the associated ODE; however, we do not know that v_t satisfies Hypothesis 2.1.8, hence we will use a regularisation argument.

Precisely, take a *strictly positive* convolution kernel $\rho \in C^\infty(\mathbb{R}^n)$; for instance, $\rho(x) := (2\pi)^{-n/2} \exp\left(-\frac{|x|^2}{2}\right)$. Put $\rho_\varepsilon(x) := \varepsilon^{-n}\rho\left(\frac{x}{\varepsilon}\right)$, $\mu_t^\varepsilon := \mu_t * \rho_\varepsilon$. Acting by convolution on the continuity equation for (μ_t, v_t), we get that

$$\frac{d}{dt}\mu_t^\varepsilon + \nabla\cdot((v_t\mu_t) * \rho_\varepsilon) = 0$$

which is of the form $\frac{d}{dt}\mu_t^\varepsilon + \nabla \cdot (v_t^\varepsilon \mu_t^\varepsilon)$ if we can define v_t^ε implicitly by the relation $v_t^\varepsilon \mu_t^\varepsilon = (v_t\mu_t) * \rho_\varepsilon$. This can be done because both $(v_t\mu_t) * \rho_\varepsilon$ and μ_t^ε are absolutely continuous with respect to $\mathscr{L}^n$, so that identifying them with their densities (which we will do in the sequel of the proof), it is sufficient to put

$$v_t^\varepsilon = \frac{(v_t\mu_t) * \rho_\varepsilon}{\mu_t^\varepsilon}$$

where the denominator is nonzero because $\mu_t \neq 0$ and ρ is strictly positive.

$\boxed{\text{Step 1}}$ We prove that

$$\int_{\mathbb{R}^n} |v_t^\varepsilon|^2 \, d\mu_t^\varepsilon \leq \int_{\mathbb{R}^n} |v_t|^2 \, d\mu_t \tag{2.1.5}$$

To do this, we compute

$$|v_t^\varepsilon(x)|^2 \mu_t^\varepsilon(x) \ = \left| \frac{\int v_t(y)\rho_\varepsilon(x-y)d\mu_t(y)}{\int \rho_\varepsilon(x-y)d\mu_t(y)} \right|^2 \int \rho_\varepsilon(x-y)d\mu_t(y).$$

Now we apply Jensen's inequality to the function $t \mapsto t^2$ and the probability measure obtained normalizing $\rho_\varepsilon(x-\cdot)\mu_t$: we get that the latter expression is

$$\leq \int |v_t(y)|^2 \rho_\varepsilon(x-y)d\mu_t(y).$$

Integration in dx gives the desired inequality.

$\boxed{\text{Step 2}}$ Fix $\varepsilon > 0$. We claim that $\mu_t^\varepsilon = (X_t^\varepsilon)_\# \mu_0^\varepsilon$ where $X_t^\varepsilon(x)$ is the maximal solution of $u'(t) = v_t^\varepsilon(u(t))$ with $u(0) = x$. We want to apply Theorem 2.1.14. Note that by Jensen $\int_0^1 \int |v_t| \, d\mu_t \, dt \leq (\int_0^1 \int |v_t|^2 \, d\mu_t \, dt)^{1/2} < \infty$, so we only have to verify Hypothesis 2.1.8.

In order to do this, we prove that $\sup_{x\in B} |v_t^\varepsilon(x)|$ and $\sup_{x\in B} |\nabla v_t^\varepsilon(x)|$ (which estimates the Lipschitz constant at least when B is a ball) are summable functions of t for every $B \subset\subset \mathbb{R}^n$. After expanding the derivative of the quotient that defines v_t^ε, this follows from the remarks:

- $\sup_{\mathbb{R}^n} |(v_t\mu_t) * \rho_\varepsilon| \leq \|v_t\mu_t\|_1 \|\rho_\varepsilon\|_\infty$ is summable in t because we made the initial assumption $\int_0^1 \|v_t\mu_t\|_1 \, dt < \infty$;
- $\sup_{\mathbb{R}^n} |\nabla((v_t\mu_t) * \rho_\varepsilon)| = \sup_{\mathbb{R}^n} |(v_t\mu_t) * \nabla\rho_\varepsilon| \leq \|v_t\mu_t\|_1 \|\nabla\rho_\varepsilon\|_\infty$ is summable in t for the same reason;
- For every $B \subset\subset \mathbb{R}^n$ there exist $c, C \in \mathbb{R}$ such that $0 < c < \mu_t * \rho_\varepsilon < C < \infty$, because $\mu_t * \rho_\varepsilon$ is a continuous strictly positive function of (t,x) in the compact set $[0,1] \times \bar{B}$;
- $\sup_{\mathbb{R}^n} |\nabla(\mu_t * \rho_\varepsilon)| = \sup_{\mathbb{R}^n} |\mu_t * \nabla\rho_\varepsilon| \leq \|\mu_t\|_1 \|\nabla\rho_\varepsilon\|_\infty = \|\nabla\rho_\varepsilon\|_\infty$ which is a constant.

Conclusion By lower semicontinuity, $W_2^2(\mu,\nu) \le \liminf_{\varepsilon\to 0} W_2^2(\mu_0^\varepsilon,\mu_1^\varepsilon)$. Therefore, if we could show that $W_2^2(\mu_0^\varepsilon,\mu_1^\varepsilon) \le \int_0^1\int_{\mathbb{R}^n}|v_t^\varepsilon|^2\, d\mu_t^\varepsilon\, dt$, then the thesis would follow using (2.1.5). But

$$W_2^2(\mu_0^\varepsilon,\mu_1^\varepsilon) \le \int_{\mathbb{R}^n}|X_1^\varepsilon(x)-x|^2 d\mu_0^\varepsilon(x) = \int_{\mathbb{R}^n}\left|\int_0^1 v_t^\varepsilon(X_t^\varepsilon(x))\, dt\right|^2 d\mu_0^\varepsilon(x) \overset{\text{Jensen}}{\le}$$
$$\le \int_{\mathbb{R}^n}\int_0^1 |v_t^\varepsilon|^2(X_t^\varepsilon(x))\, dt\, d\mu_0^\varepsilon(x) = \int_0^1\int_{\mathbb{R}^n}|v_t^\varepsilon|^2(y)\, d\mu_t^\varepsilon(y)\, dt$$

and the proof is complete.

□

The next theorem gives more insight in the meaning of Benamou-Brenier's theorem:

Theorem 2.1.16 (Description of $AC_2(\mathscr{P}_2(\mathbb{R}^n))$)**.**

1. *Let $(\mu_t,v_t)_{t\in[0,1]}$ be a solution of* (CE) *with $\mu_t\in\mathscr{P}_2(\mathbb{R}^n)$ for every t, and suppose that $\int_0^1|v_t|^2 d\mu_t dt<\infty$. Then $(\mu_t)_{t\in[0,1]}$ is an AC_2 curve in $\mathscr{P}_2(\mathbb{R}^n)$ and $|\mu_t'|\le\|v_t\|_{L^2(\mu_t)}$ for a.e. $t\in[0,1]$.*
2. *Conversely, if $(\mu_t)_{t\in[0,1]}$ is an AC_2 curve in $\mathscr{P}_2(\mathbb{R}^n)$, then it satisfies* (CE) *for some v_t such that $\|v_t\|_{L^2(\mu_t)}=|\mu_t'|$ for a.e. $t\in[0,1]$ (the least possible value, according to part 1.). Moreover, two such v_t coincide for a.e. t.*

Proof.
1. Take (a_i,b_i) disjoint subintervals of $(0,1)$ with $\sum_i(b_i-a_i)<\varepsilon$. Then

$$\sum_i W_2(\mu_{a_i},\mu_{b_i}) \overset{\text{(BB)}}{\le} \sum_i\sqrt{b_i-a_i}\sqrt{\int_{a_i}^{b_i}|v_t|^2 d\mu_t dt} \le \sum_i \frac{1}{2}\left[(b_i-a_i)+\int_{a_i}^{b_i}|v_t|^2 d\mu_t dt\right]$$

which goes to 0 as $\varepsilon\to 0$: the absolute continuity is proven.

To show that $|\mu_t'|\le\|v_t\|_{L^2(\mu_t)}$ for a.e. t, we just note that

$$\frac{W_2^2(\mu_t,\mu_{t+h})}{h^2} \overset{\text{(BB)}}{\le} \frac{1}{h}\int_t^{t+h}\|v_s\|^2_{L^2(\mu_s)}\, ds$$

which for $h\to 0$ gives the desired inequality if t is a point of metric differentiability for $t\mapsto\mu_t$, and a Lebesgue point for $t\mapsto\|v_t\|^2_{L^2(\mu_t)}$.

2. For every $m\in\mathbb{N}$, we approximate μ_t with a curve μ_t^m such that $\mu_t(\frac{i}{m})=\mu_t^m(\frac{i}{m})$ for $i=0,\dots,m$, and μ_t^m is a constant speed geodesic on each interval of the form $[\frac{i}{m},\frac{i+1}{m}]$. For constant speed geodesics we already know that there is a velocity field satisfying our thesis (Proposition 2.1.2). Gluing these velocity fields, we get a v_t^m such that (μ_t^m,v_t^m) solves (CE) (in each interval, hence in $[0,1]$ by Corollary 2.1.11). We also get that for every $t\in[\frac{i}{m},\frac{i+1}{m}]$ it holds

$$\int|v_t^m|^2\, d\mu_t^m \le m^2 W_2^2(\mu_{i/m},\mu_{(i+1)/m}) \le m^2\left(\int_{\frac{i}{m}}^{\frac{i+1}{m}}|\mu_s'|ds\right)^2 \le m\int_{\frac{i}{m}}^{\frac{i+1}{m}}|\mu_s'|^2 ds.$$

Integrating with respect to t in $[\frac{i}{m}, \frac{i+1}{m}]$ and summing on i one concludes

$$\int_0^1 \int_{\mathbb{R}^n} |v_t^m|^2 \, d\mu_t^m dt \le \int_0^1 |\mu_t'|^2 dt \tag{2.1.6}$$

which is finite by hypothesis.

We note that $\mu_t^m \to \mu_t$ in $\mathscr{P}_2(\mathbb{R}^n)$, hence narrowly; from this, using the definition of narrow convergence and Lebesgue's theorem, one immediately gets that $\mu_t^m \otimes dt \rightharpoonup \mu_t \otimes dt$. The inequality (2.1.6) and the semicontinuity of the L^2 relative norm (Lemma 1.3.20) imply the existence of a subsequence $(m(k))$ and of a function v_t such that $v_t^{m(k)} \mu_t^{m(k)} \otimes dt \rightharpoonup v_t \mu_t \otimes dt$ and

$$\int_0^1 \int_{\mathbb{R}^n} |v_t|^2 \, d\mu_t dt \le \int_0^1 |\mu_t'|^2 dt. \tag{2.1.7}$$

Now (μ_t, v_t) is a solution of (CE): just write the definition of the fact that (μ_t^m, v_t^m) solves (CE) in distributional sense, and let $m \to \infty$ along the subsequence $(m(k))$. As a consequence, the first part of the Theorem tells us that $|\mu_t'| \le \|v_t\|_{L^2(\mu_t)}$ for a.e. $t \in [0,1]$, which combined with (2.1.7) concludes the proof of existence.

To prove uniqueness, take v_t, w_t velocity fields such that (μ_t) satisfies (CE) and $\|v_t\|_{L^2(\mu_t)} = \|w_t\|_{L^2(\mu_t)} = |\mu_t'|$ for a.e. t. Then defining $u_t := \frac{1}{2}(v_t + w_t)$, part 1. yields $|\mu_t'| \le \|u_t\|_{L^2(\mu_t)}$ for a.e. t; which implies that $v_t = w_t$ for a.e. t by strict convexity of the L^2 norm. $\square$

Definition 2.1.17. The velocity field v_t whose existence and uniqueness was proven in the second part of the previous Theorem, will be said **tangent** to the curve (μ_t). (See also Remark 2.1.20.)

Proposition 2.1.18 (Characterisation of the tangent field)**.** *Let $(\mu_t)_{t\in[0,1]}$ be an AC_2 curve in $\mathscr{P}_2(\mathbb{R}^n)$, and suppose that (μ_t, v_t) satisfy* (CE)*. Then v_t is tangent to (μ_t) if and only if, for a.e. t, v_t is in the $L^2(\mu_t;\mathbb{R}^n)$-closure of $\{\nabla\phi : \phi \in C_c^\infty(\mathbb{R}^n)\}$.*

Proof. Note that $(v_t + w_t, \mu_t)$ satisfies (CE) if and only if for a.e. t, $\nabla \cdot (w_t \mu_t) = 0$, which by the very definition of divergence means $w_t \in \{\nabla\phi : \phi \in C_c^\infty(\mathbb{R}^n)\}^\perp$ in the Hilbert space $L^2(\mu_t)$. The element of smallest norm among these $v_t + w_t$ is the projection of v_t onto $\{\nabla\phi : \phi \in C_c^\infty(\mathbb{R}^n)\}^{\perp\perp}$, and the conclusion follows. $\square$

With this Proposition in mind, we give the following Definition:

Definition 2.1.19. The **tangent space** to $\mathscr{P}_2(\mathbb{R}^n)$ at a point $\mu \in \mathscr{P}_2(\mathbb{R}^n)$ is the set

$$\mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^n) := \overline{\{\nabla\phi : \phi \in C_c^\infty(\mathbb{R}^n)\}}^{L^2(\mu;\mathbb{R}^n)}.$$

Remark 2.1.20 (Tangent field to generic AC curves)*.* If (μ_t) is any absolutely continuous curve in $\mathscr{P}_2(\mathbb{R}^n)$, we can always reparametrise it in such a way that it becomes an AC_2 curve, and since the continuity equation is invariant under reparametrisation, we find by Theorem 2.1.16 that there exist $v_t \in L^2(\mu_t;\mathbb{R}^n)$ such that (μ_t, v_t) satisfies (CE). By projection (as in Proposition 2.1.18), we see that we can take $v_t \in \mathrm{Tan}_{\mu_t} \mathscr{P}_2(\mathbb{R}^n)$ for a.e. t, and that this determines one precise velocity field (v_t) (up to a negligible set of t). Extending Definition 2.1.17, we will say that this v_t is the **tangent** velocity field to (μ_t).

There are more "constructive" results which indicate v_t as the "tangent" vector to the curve:

Theorem 2.1.21 (Characterisation of v_t)**.** *Let $(\mu_t) \subseteq \mathscr{P}_2(\mathbb{R}^n)$ be an AC curve and v_t be its tangent velocity field. Then for a.e. t, denoting by γ_h any element of $\Gamma_o(\mu_t, \mu_{t+h})$, it holds*

$$\left(p_x, \frac{1}{h}(p_y - p_x)\right)_{\#} \gamma_h \xrightarrow{h\to 0} (Id \times v_t)_{\#}\mu_t \quad in\ \mathscr{P}_2((\mathbb{R}^n)^2). \tag{2.1.8}$$

Moreover, the above relation implies

$$\lim_{h\to 0} \frac{W_2\left(\mu_{t+h}, (Id + hv_t)_{\#}\, \mu_t\right)}{|h|} = 0. \tag{2.1.9}$$

Finally, for a.e. t such that $\mu_t \ll \mathscr{L}^n$, if T_t^{t+h} denotes the optimal map from μ_t to μ_{t+h}, then

$$\frac{1}{h}(T_t^{t+h} - Id) \xrightarrow{h\to 0} v_t \quad in\ L^2(\mu_t).$$

Proof. For simplicity of notation, we denote the measures $\left(p_x, \frac{1}{h}(p_y - p_x)\right)_{\#} \gamma_h$ as ν_h. We call $\nu_0 := \lim_{i\to\infty} \nu_{h(i)}$ any limit point of (ν_h) for $h \to 0$ in the sense of duality with $C_0((\mathbb{R}^n)^2)$. (A priori, the mass of ν_0 might be less than 1.) The first marginal of ν_h is μ_t for every h, from which one easily proves that the first marginal of ν_0 is a constant multiple of μ_t: so we can disintegrate ν_0 as $\mu_t \otimes \nu_{0x}$. Our first goal is to show that $\nu_{0x} = \delta_{v_t(x)}$ for a.e. t, which will give $\nu_0 = (Id \times v_t)_{\#}\mu_t$.

Fix $\phi \in C_c^\infty(\mathbb{R}^n)$. On the one hand

$$\begin{aligned}\int \phi\ d\mu_{t+h} - \int \phi\ d\mu_t &= \int (\phi(y) - \phi(x)) d\gamma_h(x, y) = \\ &= \int (\phi(x + hy) - \phi(x)) d\nu_h = h \int \langle \nabla\phi(x), y\rangle d\nu_h(x, y) + o(h)\end{aligned}$$

where we used that the error in the Taylor formula has an uniform estimate because ϕ is C_c^∞.

On the other hand

$$\frac{1}{h}\left(\int \phi\ d\mu_{t+h} - \int \phi\ d\mu_t\right) \xrightarrow{h\to 0} \int_{\mathbb{R}^n} \langle \nabla\phi, v_t\rangle d\mu_t$$

for a.e. t (see Lemma 2.1.10).

Putting $\tilde{v}_t(x) := \int y\ d\nu_{0,x}(y)$, we have proven that for a.e. t it holds

$$\int \langle \nabla\phi, v_t\rangle d\mu_t = \int \langle \nabla\phi, \tilde{v}_t\rangle d\mu_t;$$

using this equation for ϕ varying in a countable dense subset of $C_c^\infty(\mathbb{R}^n)$ we conclude that $\nabla\cdot((\tilde v_t - v_t)\mu_t) = 0$ for a.e. t, so that $(\mu_t, \tilde v_t)$ solves (CE) too. But

$$\begin{aligned}\|\tilde v_t\|^2_{L^2(\mu_t)} &= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} y\, d\nu_{0x}(y)\right)^2 d\mu_t(x) \le \int_{\mathbb{R}^n}\nu_{0x}(\mathbb{R}^n)\int_{\mathbb{R}^n}|y|^2 d\nu_{0x}(y)\, d\mu_t(x) \le \\ &\le \int_{(\mathbb{R}^n)^2}|y|^2 d\nu_0(x,y) \le \liminf_{i\to\infty}\int_{(\mathbb{R}^n)^2}|y|^2 d\nu_{h(i)}(x,y) = \\ &= \liminf_{i\to\infty}\int_{(\mathbb{R}^n)^2}\frac{|y-x|^2}{h(i)^2}d\gamma_{h(i)}(x,y) = \liminf_{i\to\infty}\frac{W_2^2(\mu_t,\mu_{t+h(i)})}{h(i)^2} = |\mu'_t|^2\end{aligned}$$

for a.e. t; we know that the converse inequality holds for a.e. t for every solution of (CE), so that we have equality and $\tilde v_t = v_t$ for a.e. t (by definition of v_t).

Note that the chain of (in)equalities holds also replacing $h(i)$ with the subsequence of $h(i)$ realising $\limsup\limits_{i\to\infty}\int_{(\mathbb{R}^n)^2}|y|^2 d\nu_{h(i)}(x,y)$: from which we deduce that

$$\int_{(\mathbb{R}^n)^2}|y|^2 d\nu_0(x,y) = \lim_{i\to\infty}\int_{(\mathbb{R}^n)^2}|y|^2 d\nu_{h(i)}(x,y) = |\mu'_t|^2 \quad \text{for a.e. } t.$$

If we combine this with the fact that the first marginal of ν_h is constant, we easily deduce that the sequence $\nu_{h(i)}$ is tight, and so ν_0 is a probability measure and the convergence is narrow. As a consequence, $(p_x)_\#\nu_0 = \mu_t$.

Now both $\int_{(\mathbb{R}^n)^2}|x|^2 d\nu_{h(i)}(x,y)$ and $\int_{(\mathbb{R}^n)^2}|x|^2 d\nu_0(x,y)$ are equal to the constant finite quantity $\int_{\mathbb{R}^n}|x|^2 d\mu_t(x)$, so we have shown that the second moments of $\nu_{h(i)}$ converge to the second moment of ν_0. By the basic criterion of convergence in $\mathscr{P}_2$, we conclude that $\nu_{h(i)}\to\nu_0$ in W_2 distance.

We observe that to have equality in all the inequalities above, it is necessary that for μ_t-a.e. x, ν_{0x} is a Dirac mass. In this case, the definition of $\tilde v_t$ forces this Dirac mass to be in $\tilde v_t = v_t$, and the first part of the Theorem is proved.

As for (2.1.9), we disintegrate γ_h as $\gamma_{h,x}\otimes\mu_t$ and consider the probability measure on $(\mathbb{R}^n)^3$ given by $\left(\gamma_{h,x}\times\delta_{x+hv_t(x)}\right)\otimes\mu_t$: since two of its marginals are respectively μ_{t+h} and $(Id+hv_t)_\#\mu_t$, then we can estimate

$$\begin{aligned}\frac{W_2^2\left(\mu_{t+h},(Id+hv_t)_\#\,\mu_t\right)}{h^2} &\le \int_{(\mathbb{R}^n)^2}\frac{1}{h^2}|x+hv_t(x)-y|^2 d\gamma_h(x,y) = \\ &= \int_{(\mathbb{R}^n)^2}|v_t(x)-y|^2 d\nu_h(x,y)\end{aligned}$$

by definition of ν_h. We claim that (2.1.8) implies that this quantity is infinitesimal. In fact, for every $\varepsilon>0$ fixed, we can take by density $v^\varepsilon\in C_c(\mathbb{R}^n)$ such that $\int|v_t-v^\varepsilon|^2 d\mu_t<\varepsilon$, and estimate

$$\int_{(\mathbb{R}^n)^2}|v_t(x)-y|^2 d\nu_h(x,y) \le 2\int|v_t-v^\varepsilon|^2 d\mu_t + 2\int_{(\mathbb{R}^n)^2}|v^\varepsilon(x)-y|^2 d\nu_h(x,y)$$

where $|v^\varepsilon(x)-y|^2$ has 2-growth (see Remark 1.3.8), and so for $h\gg 1$ the second integral is less then

$$2\int_{(\mathbb{R}^n)^2}|v^\varepsilon(x)-y|^2 d\nu_0(x,y)+\varepsilon = 2\int|v_t-v^\varepsilon|^2 d\mu_t+\varepsilon\,.$$

To sum up, we have proven that $\frac{W_2^2(\mu_{t+h},(Id+hv_t)_\#\mu_t)}{h^2} \leq 5\,\varepsilon$ for $h \gg 1$, which is (2.1.9) for arbitrariness of ε.

Finally, in the case in which $\mu_t \ll \mathscr{L}^n$, (2.1.8) reduces to

$$\left(Id \times \frac{1}{h}\left(T_t^{t+h} - Id\right)\right)_\# \mu_t \to (Id \times v_t)_\# \mu_t.$$

A direct consequence of the definitions is that if $(Id \times f_h)_\#\mu \rightharpoonup (Id \times f)_\#\mu$, then $f_h\mu \rightharpoonup f\mu$: therefore $\frac{1}{h}(T_t^{t+h} - Id)\mu_t \rightharpoonup v_t\mu_t$. But

$$\left\|\frac{1}{h}(T_t^{t+h} - Id)\right\|_{L^2(\mu_t)} = \frac{1}{h}W_2(\mu_t, \mu_{t+h}) \to |\mu_t'| = \|v_t\|_{L^2(\mu_t)} \quad \text{for a.e. } t.$$

In particular the $L^2(\mu_t)$ norms of $\frac{1}{h}(T_t^{t+h} - Id)$ are bounded, hence the sequence of this functions is weakly sequentially compact in $L^2(\mu_t)$; but any $L^2(\mu_t)$-weak limit for $h \to 0$ is forced to be v_t, from which we conclude that $\frac{1}{h}(T_t^{t+h} - Id)$ converges to v_t weakly in $L^2(\mu_t)$. Since the norms converge, the convergence is also in $L^2(\mu_t)$-norm. □

Proposition 2.1.22. *For $\mu \in \mathscr{P}_2^a(\mathbb{R}^n), \nu \in \mathscr{P}_2(\mathbb{R}^n)$, let T_μ^ν be the unique optimal transport map from μ to ν. Then $(T_\mu^\nu - Id) \in \mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^n)$.*

Proof. Suppose first that ν has compact support. Brenier's theorem tells us that T_μ^ν is of the form $\nabla\phi$ with ϕ convex and μ concentrated on $\mathrm{Int}(\mathrm{Dom}(\phi))$. It is straightforward to prove that $\nabla\phi$ is continuous on its domain: simply observe that the graph of the subdifferential of a convex function is closed. Hence by truncation and mollification there exist $\phi_k \in C_c^\infty(\mathbb{R}^n)$ such that $(\nabla\phi_k)_{k\in\mathbb{N}}$ is bounded and converges a.e. to $\nabla\phi$ on $\mathrm{Int}(\mathrm{Dom}(\phi))$; since μ is concentrated on this set and $\mu \ll \mathscr{L}^n$, they converge also μ-a.e.. By dominated convergence we conclude that $\nabla\phi_k \to \nabla\phi$ in $L^2(\mu)$.

For general ν, take ν_k compact supported and such that $\nu_k \to \nu$, and use the stability of optimal maps (Proposition 1.3.9). □

Remark 2.1.23 (Tangent space to $\mu \in \mathscr{P}_2^a(\mathbb{R}^n)$). The proposition shows that

$$\overline{\{\lambda(T_\mu^\nu - Id) : \lambda \in \mathbb{R}, \nu \in \mathscr{P}_2(\mathbb{R}^n)\}}^{L^2(\mu;\mathbb{R}^n)} \subseteq \mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^n).$$

This is actually an equality: in fact, for any $\phi \in C_c^\infty(\mathbb{R}^n)$, we can take $\lambda \gg 1$ such that $\frac{1}{2}|x|^2 + \lambda^{-1}\phi(x)$ is convex, from which $Id + \lambda^{-1}\nabla\phi$ is an optimal trasport map T_μ^ν by Brenier's theorem.

Remark 2.1.24 (Hamilton-Jacobi equation, variational argument). With a formal variational argument, we can obtain in a new way the equations of geodesics (2.1.2).

In fact, let $(\mu_t)_{t\in[0,1]}$ be a constant speed geodesic, with $\mu_0, \mu_1 \ll \mathscr{L}^n$. From part 2. of Proposition 2.1.2 we know that the tangent velocity field is of the form $\nabla\psi_t$, and that $\mu_t \ll \mathbb{R}^n$ for every t; put $\rho_t := \frac{d\mu_t}{d\mathscr{L}^n}$.

Now take $(\sigma_t)_{t\in[0,1]}$ any AC_2 curve in $\mathscr{P}_2^a(\mathbb{R}^n)$ from μ_0 to μ_1, and w_t the corresponding tangent field. Put $s_t := \frac{d\sigma_t}{d\mathscr{L}^n}$. For $\varepsilon \in \mathbb{R}$, define

$$\mu_t^\varepsilon := (1-\varepsilon)\mu_t + \varepsilon\,\sigma_t = ((1-\varepsilon)\rho_t + \varepsilon\, s_t)\,\mathscr{L}^n =: \rho_t^\varepsilon\,\mathscr{L}^n$$

and suppose that this is an element of $\mathscr{P}_2(\mathbb{R}^n)$ for every $t \in [0,1]$ and every ε sufficiently small. (This is not obvious for $\varepsilon < 0$; but it is true for instance if s_t is bounded and $s_t \neq 0$ only on a set where ρ_t is bounded from below by a positive constant.) It clearly holds

$$\frac{d}{dt}\mu_t^\varepsilon + \nabla \cdot ((1-\varepsilon)\mu_t v_t + \varepsilon\, \sigma_t w_t) = 0$$

which is of the form (CE) if v_t^ε is defined by

$$v_t^\varepsilon := \frac{(1-\varepsilon)\rho_t v_t + \varepsilon\, s_t w_t}{(1-\varepsilon)\rho_t + \varepsilon\, s_t}.$$

where we use the convention $\frac{0}{0} = 0$. Now

$$\int_0^1 \int_{\mathbb{R}^n} |v_t^\varepsilon|^2(x)\rho_t^\varepsilon(x)dx\ dt = \int_0^1 \int_{\mathbb{R}^n} \frac{|(1-\varepsilon)\rho_t v_t + \varepsilon\, s_t w_t|^2}{(1-\varepsilon)\rho_t + \varepsilon\, s_t} d\mathscr{L}^n\, dt$$

attains a local minimum at $\varepsilon = 0$; hopefully, we can differentiate under the integral sign and impose the result to be zero. The derivative of the integrand with respect to ε is

$$2\frac{(1-\varepsilon)\rho_t v_t + \varepsilon\, s_t w_t}{(1-\varepsilon)\rho_t + \varepsilon\, s_t} \cdot [s_t w_t - \rho_t v_t] - \left[\frac{(1-\varepsilon)\rho_t v_t + \varepsilon\, s_t w_t}{(1-\varepsilon)\rho_t + \varepsilon\, s_t}\right]^2 [s_t - \rho_t]$$

so that in $\varepsilon = 0$ we obtain the condition

$$0 = \int_0^1 \int_{\mathbb{R}^n} 2v_t \cdot [s_t w_t - \rho_t v_t] d\mathscr{L}^n\, dt + \int_0^1 \int_{\mathbb{R}^n} |v_t|^2 [\rho_t - s_t] d\mathscr{L}^n\, dt.$$

Now note that $v_t = \nabla\psi_t$: hence, using that $\nabla\cdot(\sigma_t w_t - \mu_t v_t) = \dot{\mu}_t - \dot{\sigma}_t$ (and neglecting issues of regularity of ψ), we can rewrite the first addend as

$$-2\int_0^1 \int_{\mathbb{R}^n} \partial_t \psi_t d(\sigma_t - \mu_t) dt.$$

To sum up, the condition is

$$\int_0^1 \int_{\mathbb{R}^n} \left(2\partial_t\psi_t + |\nabla\psi_t|^2\right) d(\mu_t - \sigma_t)dt = 0 \tag{2.1.10}$$

whenever $(\sigma_t)_{t\in[0,1]}$ is an AC_2 curve from μ_0 to μ_1 (satisfying some technical assumptions). We cannot expect that this implies that the integrand is 0: the only constraint on ψ_t in the argument above is the value of its gradient, so we can for instance add to ψ_t a constant function $f(t)$. However, we can hope that, with a suitable additive perturbation to ψ_t, we can find a potential which solves the Hamilton-Jacobi equation. In the setting of Chapter 3, the discrete counterpart of this statement will be given a formal justification: see Remark 3.1.42.

2.2 Differentials in $\mathscr{P}_2(\mathbb{R}^n)$

We define a couple of concepts, meaningful in generic metric spaces, which will be useful in the sequel.

Definition 2.2.1 (Slope)**.** Let (X, d) be a metric space. The **slope** of a function $E : X \to \mathbb{R} \cup \{+\infty\}$ in a point $x \in X$ such that $E(x) < \infty$ is the quantity

$$|\nabla E|(x) := \limsup_{y \to x} \frac{(E(x) - E(y))^+}{d(x, y)}.$$

By definition, it is the smallest $s \in [0, \infty]$ such that $E(y) \geq E(x) - s \cdot d(x, y) + o(d(x, y))$. We will use the notations

$$\mathrm{Dom}(E) =: \{x \in X : E(x) < \infty\}, \quad \mathrm{Dom}(|\nabla E|) := \{x \in \mathrm{Dom}(E) : |\nabla E|(x) < \infty\}.$$

Definition 2.2.2 (λ-convexity)**.** Let (X, d) be a geodesic metric space, $\lambda \in \mathbb{R}$. A function $E : X \to \mathbb{R} \cup \{+\infty\}$ is **geodesically $\boldsymbol{\lambda}$-convex** (resp. **$\boldsymbol{\lambda}$-concave**) if for every $x, y \in X$ there *exists* a constant speed geodesic $\gamma : [0, 1] \to X$ from x to y such that

$$E(\gamma(t)) \leq (1 - t)E(\gamma(0)) + tE(\gamma(1)) - \frac{\lambda}{2} t(1 - t)d^2(\gamma(0), \gamma(1)) \qquad (\text{resp. } \geq).$$

When there is no possibility of confusion (for instance, when there is no linear structure on X), we will simply say that E is **$\boldsymbol{\lambda}$-convex** (resp. **$\boldsymbol{\lambda}$-concave**).

Example 2.2.3 (Hilbert spaces)**.** If X is a Hilbert space, a direct verification shows that E is λ-convex if and only if $x \mapsto E(x) - \frac{\lambda}{2}|x|^2$ is convex.

Moreover, given any functional E, there is a classical object denoted with $\nabla E(x)$: precisely, the element of minimal norm in the subdifferential of E in x (when this subdifferential is nonempty). The nontrivial fact is that, at least if E is λ-convex, the notation introduced above is coherent with this, i.e. that $|\nabla E(x)|$ is the slope of E in x.

One inequality does not need convexity:

$$E(y) \geq E(x) + \langle \nabla E(x), y - x \rangle + o(|x - y|) \geq E(x) - |\nabla E(x)||y - x| + o(|x - y|)$$

implies that $|\nabla E(x)|$ is greater or equal than the slope.

For the converse inequality, we firstly note that $E(y) \geq E(x) - s|x - y| + o(|x - y|)$ is equivalent to

$$E(y) \geq E(x) - s|x - y| + \frac{\lambda}{2}|x - y|^2 \qquad \forall y$$

(the proof is simple; the same ideas will be used in the proofs of Lemma 2.2.4 and Proposition 2.2.10). Then, in $H \times \mathbb{R}$, the convex sets

$$A := \left\{(y, t) : E(y) - E(x) - \frac{\lambda}{2}|x - y|^2 < t\right\}, \quad B := \{(y, t) : t < -s|x - y|\}$$

can be separated by Hahn-Banach's theorem, and from this it is not difficult to conclude.

Much of what seen in this example, in a suitable sense, will be valid replacing H with $\mathscr{P}_2(\mathbb{R}^n)$.

Lemma 2.2.4 (Global formula for the slope)**.** *If E is λ-convex and $x \in \mathrm{Dom}(E)$, then*

$$|\nabla E|(x) = \sup_{y \neq x} \left[\frac{E(x) - E(y)}{d(x,y)} + \frac{\lambda}{2} d(x,y) \right]^+ .$$

Proof. The inequality $\leq$ is obvious. Conversely, given any $y \in X$, we apply the definition of λ-convexity to find a geodesic γ from x to y such that

$$E(\gamma(t)) - E(x) \leq t(E(y) - E(x)) - \frac{\lambda}{2} t(1-t) d(x,y).$$

Now it is sufficient to divide this relation by the equality $d(x, \gamma(t)) = t \cdot d(x,y)$ and let $t \downarrow 0$. □

Corollary 2.2.5. *If E is λ-convex and lower semicontinuous, then $|\nabla E|$ is lower semicontinuous on* $\mathrm{Dom}(E)$ *(as it is a supremum of lower semicontinuous functionals).*

Notation. For $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^n)$, $\mu \ll \mathscr{L}^n$, we will denote by T_μ^ν the unique optimal transport map from μ to ν.

In analogy with the classical concepts from differential calculus, we give the following definitions:

Definition 2.2.6. For $\mu \in \mathrm{Dom}(|\nabla E|) \cap \mathscr{P}_2^a(\mathbb{R}^n)$, we will say that $\xi \in \mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^n)$ is the **Wasserstein differential of E**, and write $\xi = \nabla^W E(\mu)$, if

$$E(\nu) - E(\mu) = \int_{\mathbb{R}^n} \langle \xi(x), T_\mu^\nu(x) - x \rangle \, d\mu(x) + o(W_2(\mu,\nu)).$$

Remark 2.2.7 (Uniqueness of the differential)*.* The Wasserstein differential is unique: in fact for every $\nu \in \mathscr{P}_2(\mathbb{R}^n)$, take the unique constant speed geodesic from μ to ν, namely $\mu_t := (Id + t(T_\mu^\nu - Id))_\# \mu$ (see the beginning of the chapter); observe that $W_2(\mu, \mu_t) = t W_2(\mu, \nu)$ and $T_\mu^{\mu_t} = Id + t(T_\mu^\nu - Id)$, and conclude from the definition that the value of $\int_{\mathbb{R}^n} \langle \xi(x), T_\mu^\nu(x) - x \rangle \, d\mu(x)$ is uniquely determined. For arbitrariness of ν, and using the characterisation of the tangent space seen in Remark 2.1.23, ξ is determined.

Definition 2.2.8. For $\mu \in \mathrm{Dom}(|\nabla E|) \cap \mathscr{P}_2^a(\mathbb{R}^n)$, we will say that $\xi \in L^2(\mu; \mathbb{R}^n)$ is in the **Wasserstein subdifferential of E**, and write $\xi \in \partial^W E(\mu)$, if

$$\liminf_{\nu \to \mu} \frac{E(\nu) - E(\mu) - \int_{\mathbb{R}^n} \langle \xi(x), T_\mu^\nu(x) - x \rangle \, d\mu(x)}{W_2(\mu,\nu)} \geq 0.$$

Remark 2.2.9. Once we have a $\xi \in L^2(\mu; \mathbb{R}^n)$ satisfying the inequality of the definition above, we can find an element of the subdifferential belonging to $\mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^n)$ by projection, thanks to the fact that $(T_\mu^\nu - Id) \in \mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^n)$ for every ν by Proposition 2.1.22. Note that the projected subdifferential has a strictly smaller L^2 norm, unless ξ was already tangent.

In the λ-convex case, we have the global characterisation:

Proposition 2.2.10 (λ-convex case). *Suppose that E is λ-convex, and consider $\mu \in \mathrm{Dom}(|\nabla E|) \cap \mathscr{P}_2^a(\mathbb{R}^n)$. Then $\xi \in L^2(\mu;\mathbb{R}^n)$ is in $\partial^W E(\mu)$ if and only if*

$$E(\nu) - E(\mu) \geq \int \langle \xi(x), T_\mu^\nu(x) - x\rangle \, d\mu(x) + \frac{\lambda}{2} W_2^2(\mu,\nu) \quad \forall \nu \in \mathrm{Dom}(E). \tag{2.2.1}$$

Proof. For the nontrivial implication, take the unique constant speed geodesic from μ to ν, namely $\mu_t := (Id + t(T_\mu^\nu - Id))_\# \mu$ (see the beginning of the chapter). The hypothesis gives

$$E(\mu_t) - E(\mu) \leq -tE(\mu) + tE(\nu) - \frac{\lambda}{2} t(1-t) W_2^2(\mu,\nu) \tag{2.2.2}$$

Since $W_2(\mu,\mu_t) = tW_2(\mu,\nu)$ and $T_\mu^{\mu_t} = Id + t(T_\mu^\nu - Id)$, the definition of subdifferential gives

$$\liminf_{t\to 0+} \frac{E(\mu_t) - E(\mu)}{t} \geq \liminf_{t\to 0+} \frac{1}{t} \int \langle \xi(x), T_\mu^{\mu_t}(x) - x\rangle \, d\mu(x) = \int \langle \xi(x), T_\mu^\nu(x) - x\rangle \, d\mu(x).$$

Dividing (2.2.2) by t, taking the $\liminf$ for $t \to 0+$, and using the relation just obtained, gives the desired inequality. □

Proposition 2.2.11 (Closure). *Suppose that E is λ-convex and lower semicontinuous, and let $\mu_k \to \mu$ in $\mathscr{P}_2^a(\mathbb{R}^n)$. Assume that $\xi_k \in \partial^W E(\mu_k), \xi \in L^1(\mu)$ satisfy $\xi_k \mu_k \to \xi\mu$ in the sense of distributions, and that $\sup_k \int |\xi_k|^2 d\mu_k < \infty$. Then $\xi \in \partial^W E(\mu)$.*

Proof. For every ν fixed and every μ_k, we rewrite (2.2.1) using the probability measure $P_k := (Id \times \xi_k \times T_{\mu_k}^\nu)_\# \mu_k \in \mathscr{P}((\mathbb{R}^n)^3)$, as

$$E(\nu) \geq E(\mu_k) + \int_{(\mathbb{R}^n)^3} \langle x_2, x_3 - x_1\rangle \, dP_k(x_1,x_2,x_3) + \frac{\lambda}{2} W_2^2(\mu_k,\nu). \tag{2.2.3}$$

We note that the second moments $\int_{(\mathbb{R}^n)^3} \left(|x_1|^2 + |x_2|^2 + |x_3|^2\right) dP_k(x_1,x_2,x_3)$ are equibounded by hypothesis, and so the sequence (P_k) is tight: by Prohorov's theorem, there exists a subsequence $P_{k(h)}$ narrowly converging to some probability P. Note that $|\langle x_2, x_3 - x_1\rangle| \leq C\left(|x_1|^2 + |x_2|^2 + |x_3|^2\right)$ whose integrals in dP_k are equibounded, so a standard truncation argument gives

$$\int_{(\mathbb{R}^n)^3} \langle x_2, x_3 - x_1\rangle \, dP_{k(h)}(x_1,x_2,x_3) \xrightarrow{h\to\infty} \int_{(\mathbb{R}^n)^3} \langle x_2, x_3 - x_1\rangle \, dP(x_1,x_2,x_3).$$

The marginals of $P_{k(h)}$ obviously converge: in particular $(p_{1,3})_\# P$ is optimal from μ to ν (Theorem 1.2.21), i.e. $(p_{1,3})_\# P = (Id \times T_\mu^\nu)_\# \mu$. So, calling $\gamma := (p_{1,2})_\# P$ and disintegrating P with respect to x_2, it is immediate to prove that

$$\int f(x_1,x_2,x_3) \, dP(x_1,x_2,x_3) = \int f\left(x_1,x_2,T_\mu^\nu(x_1)\right) d\gamma(x_1,x_2)$$

whenever any of the two makes sense.

Using all the observations above, if we let $k \to \infty$ in (2.2.3) along the subsequence $k(h)$, and recall the semicontinuity of E, we get

$$E(\nu) \geq E(\mu) + \int_{(\mathbb{R}^n)^2} \langle x_2, T_\mu^\nu(x_1) - x_1 \rangle \; d\gamma(x_1, x_2) + \frac{\lambda}{2} W_2^2(\mu, \nu).$$

If we disintegrate γ as $\gamma^{(x_1)} \otimes \mu$, and put $\bar{\gamma}(x_1) := \int x_2 \; d\gamma^{(x_1)}(x_2)$ (well defined: $x_2 \in L^2(\gamma)$), we therefore find

$$E(\nu) \geq E(\mu) + \int_{(\mathbb{R}^n)^2} \langle \bar{\gamma}(x_1), T_\mu^\nu(x_1) - x_1 \rangle \; d\mu(x_1) + \frac{\lambda}{2} W_2^2(\mu, \nu)$$

and our last task is to prove that $\bar{\gamma} = \xi$ μ-a.s.. Since $\bar{\gamma}, \xi \in L^2(\mu; \mathbb{R}^n)$, then it is sufficient to verify that for every test function $F \in C_c^\infty(\mathbb{R}^n; \mathbb{R}^n)$ we have $\int \langle F, \bar{\gamma} \rangle d\mu = \int \langle F, \xi \rangle d\mu$. This is true because

$$\begin{aligned} &\int \langle F(x_1), \bar{\gamma}(x_1) \rangle d\mu(x_1) = \int \langle F(x_1), x_2 \rangle d\gamma(x_1, x_2) = \\ &= \lim_{h \to \infty} \int \langle F(x_1), x_2 \rangle dP_{k(h)}(x_1, x_2, x_3) = \lim_{h \to \infty} \int \langle F, \xi_{k(h)} \rangle d\mu_{k(h)} = \int \langle F, \xi \rangle d\mu \end{aligned}$$

by the assumed distributional convergence.

□

Remark 2.2.12. (Subdifferentiability vs. finite slope) If $\xi \in \partial^W E(\mu)$, then the slope of E clearly satisfies $|\nabla E|(\mu) \leq \|\xi\|_{L^2}$. We wonder if a converse holds, namely if the finiteness of the slope is equivalent to $\partial^W E(\mu) \neq \varnothing$. The main tool to find elements of the subdifferential is to solve a variational problem (the idea is: "in a minimum point, zero is in the subdifferential"). Hence, we will often need the hypothesis:

Hypothesis 2.2.13.

1. There exists $\tau^* > 0$ such that for every $\tau \in (0, \tau^*)$ and every $\mu \in \mathscr{P}_2(\mathbb{R}^n)$, the functional $\Phi_E(\tau, \mu; \cdot) : \mathscr{P}_2(\mathbb{R}^n) \to (-\infty, +\infty]$ defined by

$$\Phi_E(\tau, \mu; \nu) := \frac{1}{2\tau} W_2^2(\mu, \nu) + E(\nu)$$

   admits at least a minimum point $\mu_\tau \in \mathrm{Dom}(E)$.

2. $\mathrm{Dom}(|\nabla E|) \subseteq \mathscr{P}_2^a(\mathbb{R}^n)$.

Proposition 2.2.14. *Let $\mu_\tau \in \mathrm{Dom}(E)$ be a minimum point of $\Phi_E(\tau, \mu; \cdot)$. Then $\mu_\tau \in \mathrm{Dom}(|\nabla E|)$. If moreover $\mu_\tau \in \mathscr{P}_2^a(\mathbb{R}^n)$ (for instance, if the second part of Hypothesis 2.2.13 holds), then $\frac{1}{\tau}(T_{\mu_\tau}^\mu - Id) \in \partial^W E(\mu_\tau)$.*

Proof. The minimality reads

$$E(\nu) - E(\mu_\tau) \geq \frac{1}{2\tau} \left(W_2^2(\mu_\tau, \mu) - W_2^2(\nu, \mu) \right) \qquad \forall \nu \in \mathscr{P}_2(\mathbb{R}^n) \tag{2.2.4}$$

and since

$$\begin{aligned}\left|W_2^2(\mu_\tau,\mu)-W_2^2(\nu,\mu)\right| &= |W_2(\mu_\tau,\mu)-W_2(\nu,\mu)|\cdot|W_2(\mu_\tau,\mu)+W_2(\nu,\mu)| \le \\ &\le W_2(\mu_\tau,\nu)\cdot(2W_2(\mu_\tau,\mu)+W_2(\mu_\tau,\nu)) = sW_2(\mu_\tau,\nu)+o(W_2(\mu_\tau,\nu))\end{aligned}$$

we see that $\mu_\tau \in \mathrm{Dom}(|\nabla E|)$.

If $\mu_\tau \in \mathscr{P}_2^a(\mathbb{R}^n)$ then we have

$$W_2^2(\mu_\tau,\mu) = \int |T_{\mu_\tau}^\mu(x)-x|^2\, d\mu_\tau(x), \quad W_2^2(\nu,\mu) \le \int |T_{\mu_\tau}^\nu(x)-T_{\mu_\tau}^\mu(x)|^2\, d\mu_\tau(x).$$

We insert these relations in inequality (2.2.4), and use the elementary identity $\frac{1}{2}|a|^2-\frac{1}{2}|b|^2 = \langle a, a-b\rangle - \frac{1}{2}|a-b|^2$, to get

$$\begin{aligned}E(\nu)-E(\mu_\tau) &\ge \frac{1}{2\tau}\int \left(|T_{\mu_\tau}^\mu(x)-x|^2-|T_{\mu_\tau}^\nu(x)-T_{\mu_\tau}^\mu(x)|^2\right) d\mu_\tau(x) = \\ &= \int \left\langle \frac{1}{\tau}\left(T_{\mu_\tau}^\mu(x)-x\right), T_{\mu_\tau}^\nu(x)-x\right\rangle d\mu_\tau(x) - \frac{1}{2\tau}\int \left|T_{\mu_\tau}^\nu(x)-x\right|^2\, d\mu_\tau(x);\end{aligned}$$

recognising that the second integral is $W_2^2(\mu_\tau,\nu)$, we conclude as desired that $\frac{1}{\tau}\left(T_{\mu_\tau}^\mu - Id\right) \in \partial^W E(\mu_\tau)$. □

With the help of our additional hypothesis, we can now answer positively to our question about finiteness of slope:

Proposition 2.2.15 (Finiteness of slope)**.** *Let E be λ-convex and lower semicontinuous, and suppose that Hypothesis 2.2.13 holds. Let $\mu \in \mathrm{Dom}(E)$. Then $\mu \in \mathrm{Dom}(|\nabla E|)$ if and only if $\partial^W E(\mu) \neq \varnothing$.*

Proof. If $\xi \in \partial^W E(\mu)$, then clearly $|\nabla E|(\mu) \le \|\xi\|_{L^2(\mu;\mathbb{R}^n)}$. Conversely, suppose that $|\nabla E|(\mu) < \infty$, and consider the minimizer μ_τ whose existence is granted by Hypothesis 2.2.13. Putting $\xi_\tau := \frac{1}{\tau}\left(T_{\mu_\tau}^\mu - Id\right)$, we now know that $\xi_\tau \in \partial^W E(\mu_\tau)$; moreover, by definition, $\int |\xi_\tau|^2 d\mu_\tau = \frac{W_2^2(\mu,\mu_\tau)}{\tau^2}$.

To estimate this quantity, we note that

$$-|\nabla E|(\mu)W_2(\mu,\mu_\tau) + \frac{\lambda}{2}W_2^2(\mu,\mu_\tau) \le E(\mu_\tau)-E(\mu) \le -\frac{1}{2\tau}W_2^2(\mu,\mu_\tau)$$

and so

$$\frac{W_2(\mu,\mu_\tau)}{\tau} \le \frac{2|\nabla E|(\mu)}{1+\tau\lambda} \quad \text{(in particular } \mu_\tau \to \mu\text{)}.$$

Hence we have

$$\limsup_{\tau\to 0} \int |\xi_\tau|^2 d\mu_\tau \le 4|\nabla E|(\mu)^2 < \infty.$$

Thanks to the semicontinuity of the relative L^2 norm (Lemma 1.3.20), there exists a sequence $\tau_k \downarrow 0$ such that $\xi_{\tau_k}\mu_{\tau_k}$ narrowly converges to some measure of the form $\xi\mu$. The "closure" of the subdifferential (Proposition 2.2.11) yields the conclusion. □

Remark 2.2.16. One can actually say more: in the same hypotheses and when any of the two equivalent conditions in the theorem holds, if we call $\partial^o E(\mu)$ the element of minimal L^2 norm in $\partial^W E(\mu)$ (it exists unique since $\partial^W E(\mu)$ is convex and closed), then $|\nabla E|(\mu) = \|\partial^o E(\mu)\|_{L^2(\mu;\mathbb{R}^n)}$, in perfect analogy with the Hilbert case. However, we will never use this fact: we therefore skip the proof, to avoid an inappropriate digression into the purely metric theory of slope. The interested reader may refer to [3]. Anyway, for the example of greatest interest for this thesis, i.e. internal energy, the result is directly proven in Theorem 2.2.32 below. Moreover, we will obtain incidentally that this equality is true at least a.e. along the trajectory of gradient flows: see Remark 2.3.10.

Finally, we see another situation in which our differential behaves like the classical gradient in ordinary calculus:

Proposition 2.2.17 (Chain rule). *Let $(\mu_t)_{t\in(a,b)}$ be an AC curve contained in* $\mathrm{Dom}(E) \subseteq \mathscr{P}_2(\mathbb{R}^n)$*, with tangent velocity field v_t. Consider a point $\bar{t} \in (a,b)$ such that*

1. *$\mu_{\bar{t}} \ll \mathscr{L}^n$, and $\partial^W E(\mu_{\bar{t}})$ is nonempty: we call $\xi_{\bar{t}}$ any of its elements;*
2. *$E \circ \mu : (a,b) \to \mathbb{R}$ is differentiable in $\bar{t}$;*
3. *$\lim_{h\to 0} \frac{1}{h}\left(T_{\mu_{\bar{t}}}^{\mu_{\bar{t}+h}} - Id\right) = v_{\bar{t}}$ in $L^2(\mu_{\bar{t}})$ (this holds for a.e. $t \in (a,b)$ thanks to Theorem 2.1.21);*
4. *$\lim_{h\to 0} \frac{W_2(\mu_{\bar{t}},\mu_{\bar{t}+h})}{h} = O(h)$ (this also holds for a.e. t: it is true in every point of metric differentiability of μ).*

Then $\displaystyle \left.\frac{d}{dt}\right|_{t=\bar{t}} [E\circ\mu] = \int \langle \xi_{\bar{t}}, v_{\bar{t}}\rangle \, d\mu_{\bar{t}}.$

Proof. Using the hypothesis $\left(T_{\mu_{\bar{t}}}^{\mu_{\bar{t}+h}} - Id\right) = h\cdot v_{\bar{t}} + o(h)$ in $L^2(\mu_{\bar{t}})$, we get from the definition of $\partial^W E(\mu_{\bar{t}})$ that

$$E(\mu_{\bar{t}+h}) - E(\mu_{\bar{t}}) \geq h \int \langle \xi_{\bar{t}}, v_{\bar{t}}\rangle \, d\mu_{\bar{t}} + o(h) + o(W_2(\mu_{\bar{t}},\mu_{\bar{t}+h}))$$

where the last term is $o(h)$ too by hypothesis. If one considers separately the cases $h > 0$ and $h < 0$, and in both cases divides the above inequality by h, the result is

$$\left.\frac{d^+}{dt}\right|_{t=\bar{t}} [E\circ\mu] \geq \int \langle \xi_{\bar{t}}, v_{\bar{t}}\rangle \, d\mu_{\bar{t}}, \qquad \left.\frac{d^-}{dt}\right|_{t=\bar{t}} [E\circ\mu] \leq \int \langle \xi_{\bar{t}}, v_{\bar{t}}\rangle \, d\mu_{\bar{t}},$$

and since the right and left derivatives of $E\circ\mu$ (exist and) coincide by hypothesis, the conclusion follows. □

Our next intention is to study two special kind of functionals: potential energy and internal energy.

2.2.1 Potential energy

Definition 2.2.18. Given a lower semicontinuous “potential” $V : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ bounded from below, the **potential energy** functional is $\mathcal{V}(\mu) := \int V d\mu$.

Remark 2.2.19 (Semicontinuity). The hypotheses on V obviously guarantee that the functional is well defined and lower semicontinuous on $\mathscr{P}_2(\mathbb{R}^n)$.

Proposition 2.2.20 (λ-convexity)**.** *$\mathcal{V}$ is λ-convex if and only if V is λ-convex. In this case, it is λ-convex along* every *geodesic.*

Proof. One implication follows considering Dirac masses. For the converse, consider a generic geodesic $\mu_t := ((1-t)p_x + tp_y)_\# \gamma$, where $\gamma \in \Gamma_o(\mu_0, \mu_1)$, and note

$$\mathcal{V}(\mu_t) = \int_{\mathbb{R}^n} V d\mu_t = \int_{(\mathbb{R}^n)^2} V((1-t)x + ty) d\gamma(x, y)$$

which, applying λ-convexity of V, gives immediately the result. □

Remark 2.2.21 (Candidate differential). Suppose that V is C^1. Then

$$\begin{aligned} \mathcal{V}(\nu) - \mathcal{V}(\mu) &= \int_{\mathbb{R}^n} \left[V(T_\mu^\nu(x)) - V(x) \right] d\mu(x) = \\ &= \int_{\mathbb{R}^n} \langle \nabla V(x), T_\mu^\nu(x) - x \rangle \, d\mu(x) + \int_{\mathbb{R}^n} \omega_x(T_\mu^\nu(x)) |T_\mu^\nu(x) - x| d\mu(x) \end{aligned} \tag{2.2.5}$$

where $\omega_x(y) \to 0$ as $y \to x$, so we can hope that under suitable hypotheses the second integral becomes small as $\nu \to \mu$. But if the second integral were $o(W_2(\mu, \nu))$, we would be very close to the statement that $\nabla^W \mathcal{V} \equiv \nabla V$. However, to hope that such a relation is true, of course we need that $\nabla V \in L^2(\mu; \mathbb{R}^n)$. The most natural (and simplest) condition which ensures this is $|\nabla V(x)| \leq C(1 + |x|)$. This is sufficient:

Theorem 2.2.22 (Differentiability)**.** *If V is C^1 and there exists $C > 0$ such that $|\nabla V(x)| \leq C(1 + |x|)$, then $\mathcal{V}$ is Wasserstein-differentiable at every $\mu \in \mathscr{P}_2^a(\mathbb{R}^n)$ with differential ∇V.*

Proof. First of all, we prove that the second integral in (2.2.5) is $o(W_2(\mu, \nu))$. We note that this integral is less or equal than

$$\left(\int \omega_x(T_\mu^\nu(x))^2 d\mu(x) \right)^{\frac{1}{2}} \left(\int |T_\mu^\nu(x) - x|^2 d\mu(x) \right)^{\frac{1}{2}}$$

where the second factor is precisely $W_2(\mu, \nu)$, so it is sufficient to prove that the first factor is infinitesimal as $\nu \to \mu$.

Let ν_n be any sequence tending to μ in $\mathscr{P}_2(\mathbb{R}^n)$, and call $T_n := T_\mu^{\nu_n}$. We know by the stability of optimal maps (Proposition 1.3.9) that $T_n \to Id$ in $L^2(\mu)$. In particular, every subsequence has a sub-subsequence which converges μ-a.s.; if we could prove that the limit is zero along these subsequences, we would conclude by a standard argument, so let us suppose that the whole sequence T_n converges μ-a.s.. Thanks to the elementary Lagrange theorem, we see that $\omega_x(\cdot)$ is continuous, and that

$$|\omega_x(y)| \leq \sup_{B_{|x|+|y|}} |\nabla V| \leq C(1 + |x| + |y|).$$

With this in mind, we split

$$\int \omega_x(T_n(x))^2 d\mu(x) \leq \int_{|x|+|T_n(x)|\leq R} \omega_x(T_n(x))^2 d\mu(x) +$$
$$+ \int_{|x|>\frac{R}{2}} \tilde{C}(1+|x|^2+|T_n(x)|^2) d\mu(x) + \int_{|T_n(x)|>\frac{R}{2}} \tilde{C}(1+|x|^2+|T_n(x)|^2) d\mu(x) =$$
$$=: I + II + III. \quad (2.2.6)$$

With R fixed, I is infinitesimal by dominated convergence. II is less or equal than a constant times

$$\int_{|x|>\frac{R}{2}} (1+2|x|^2) d\mu(x) + \int_{\mathbb{R}^n} |T_n(x)-x|^2 d\mu(x)$$

where the first addend is infinitesimal for $R \to \infty$ (independently of n), and the second addend for $n \to \infty$ (independently on R). Finally, III is less or equal than a constant times

$$\int_{|T_n(x)|>\frac{R}{2}} (1+2|x|^2) d\mu(x) + \int_{\mathbb{R}^n} |T_n(x)-x|^2 d\mu(x);$$

the second addend is infinitesimal as above. As for the first,

$$\mu\left\{|T_n| > \frac{R}{2}\right\} \leq \mu\left\{x : |x| > \frac{R}{2} - 1\right\} + \mu\left\{x : |T_n(x)-x| \geq 1\right\}$$

and the convergence of T_n to Id in μ-measure implies that the second contribution is infinitesimal for $n \to \infty$. Since $(1+|x|^2) \in L^1(\mu)$, then the first addend is infinitesimal too (by absolute continuity of the integral). To sum up, choosing a sufficiently large R and then a sufficiently large n, we can make (2.2.6) as small as we wish.

To conclude, we need only that ∇V is an element of $\mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^n)$. To prove this, we find by mollification $\phi_\varepsilon \in C_c^\infty(\mathbb{R}^n)$ such that $\|\phi_\varepsilon - V\|_{C^1(B_{2R})} < \varepsilon$. We call ζ_R a cutoff function which is 1 in $\bar{B}_R$, 0 out of B_{2R}, and such that $0 \leq \zeta_R \leq 1$, $|\nabla \zeta_R| \leq \frac{2}{R}$. Now $\int_{\mathbb{R}^n} |\nabla V - \nabla(\phi_\varepsilon \zeta_R)|^2 d\mu$ is less or equal than

$$\int_{\bar{B}_R} |\nabla V - \nabla \phi_\varepsilon|^2 d\mu + c \int_{\bar{B}_{2R}\setminus \bar{B}_R} \left[|\nabla V|^2 + |\phi_\varepsilon|^2 \frac{4}{R^2} + |\nabla \phi_\varepsilon|^2\right] d\mu + \int_{\bar{B}_{2R}^c} |\nabla V|^2 d\mu.$$

The first integral is less then ε; the second is less than

$$c' \int_{\bar{B}_{2R}\setminus \bar{B}_R} \left[\frac{V^2}{R^2} + |\nabla V|^2 + \varepsilon^2\right] d\mu \leq$$
$$\leq c'' \int_{\bar{B}_{2R}\setminus \bar{B}_R} \left[\frac{V(0)^2}{R^2} + \frac{C^2(1+|x|)^2|x|^2}{R^2} + C^2(1+|x|)^2 + \varepsilon^2\right] d\mu(x)$$

which is arbitrarily small if $\varepsilon \ll 1, R \gg 1$. Finally, the third integral is infinitesimal as $R \to \infty$, since $\int |\nabla V|^2 d\mu < \infty$ by the hypothesis on the growth of ∇V.

□

2.2.2 Internal energy

Definition 2.2.23. Let $U : [0,\infty) \to \mathbb{R} \cup \{+\infty\}$ be a convex lower semicontinuous function such that

$$U(0) = 0, \qquad \liminf_{s\to 0+} \frac{U(s)}{s^\alpha} > -\infty \quad \text{for some } \alpha > \frac{n}{n+2}. \tag{2.2.7}$$

Then the **internal energy** is the functional

$$\mathcal{U}(\mu) := \begin{cases} \int_{\mathbb{R}^n} U(\rho(x))dx, & \text{if } \mu = \rho \cdot \mathscr{L}^n \\ +\infty, & \text{otherwise.} \end{cases}$$

Remark 2.2.24. Condition (2.2.7) is made in such a way that the functional is well defined (and $> -\infty$). In fact, by convexity we have that the graph of U lies above a straight line, and so evidently there exist $s_0, c_1 > 0$ such that $U^-(s) \le c_1 s$ for every $s > s_0$; while (2.2.7) gives $U^-(s) \le c_2 s^\alpha$ for every $s \in [0, s_0]$. If $\alpha > 1$, this is also $\le c_1' s$, from which $U^-(\rho) \in L^1$; otherwise, the same conclusion follows observing

$$\int \rho(x)^\alpha dx \overset{\text{Hölder, } \frac{1}{p}=\alpha}{\le} \left(\int \rho(x)(1+|x|^2)dx\right)^\alpha \left(\int \frac{1}{(1+|x|^2)^{\alpha p'}}dx\right)^{1-\alpha} < \infty \tag{2.2.8}$$

because $\rho\mathscr{L}^n \in \mathscr{P}_2(\mathbb{R}^n)$ and $2\alpha p' > n$ (the latter follows by (2.2.7)).

Theorem 2.2.25 (Semicontinuity)**.** *If* $\lim_{s\to+\infty} \frac{U(s)}{s} = +\infty$*, then* $\mathcal{U}$ *is lower semicontinuous in* $\mathscr{P}_2(\mathbb{R}^n)$*.*

Proof. For every $R > 0$ and $\rho \in L^1(B_R)$, we define

$$\Phi_R(\rho) := \int_{B_R} U(\rho)\, d\mathscr{L}^n .$$

As a functional on $L^1(B_R)$, Φ_R is obviously convex. It is also lower semicontinuous: in fact, U is the supremum of a family of affine functions U_i, and so $\Phi_R = \sup \Phi_R^i$ is a supremum of continuous functionals.

Let us take $\mu_k \to \mu$ in $\mathscr{P}_2(\mathbb{R}^n)$. Our aim is to prove that if $\mathcal{U}(\mu_k) \le C < \infty$, then $\mathcal{U}(\mu) \le \liminf \mathcal{U}(\mu_k)$. Thanks to the finiteness of $\mathcal{U}(\mu_k)$ we can write $\mu_k := \rho_k \mathscr{L}^n$.

We observe that, since the second moment of μ_k is bounded independently of k, then (2.2.8) gives an uniform estimate on $\int U^-(\rho_k)\, d\mathscr{L}^n$, which combined with $\mathcal{U}(\mu_k) \le C$ yields that for any $R > 0$ the quantity $\Phi_R(\mu_k)$ is bounded above independently of k. The hypothesis on the growth of U is exactly the one which enables the use of Dunford-Pettis's compactness theorem: so, we can suppose that (ρ_k) converges weakly in $L^1(B_R)$ to some function $\rho^{(R)}$. But $\rho_k \mathscr{L}^n \rightharpoonup \mu$: hence μ has a density $\rho^{(R)}$ on each ball B_R, and by arbitrariness of R we conclude that $\mu = \rho\mathscr{L}^n$.

By the semicontinuity in $L^1(B_R)$, we have

$$\begin{aligned} &\int_{B_R} U(\rho)\, d\mathscr{L}^n \le \liminf_{k\to\infty} \int_{B_R} U(\rho_k)\, d\mathscr{L}^n \le \\ &\quad \le \liminf_{k\to\infty} \left[\int_{\mathbb{R}^n} U(\rho_k)\, d\mathscr{L}^n - \int_{B_R^c} U^+(\rho_k)\, d\mathscr{L}^n\right] + \limsup_{k\to\infty} \int_{B_R^c} U^-(\rho_k)\, d\mathscr{L}^n. \end{aligned} \tag{2.2.9}$$

We note that, arguing as in Remark 2.2.24, the quantity $\int_{B_R^c} U^-(\rho_k)\, d\mathscr{L}^n$ can be estimated by

$$c_1 \int_{B_R^c} \rho_k(x)dx + I_{[0,1]}(\alpha)\cdot c_2 \left(\int_{B_R^c} \rho_k(x)(1+|x|^2)dx\right)^{\alpha} \left(\int_{B_R^c} \frac{1}{(1+|x|^2)^{\alpha p'}}dx\right)^{1-\alpha}$$

which is infinitesimal for $R\to\infty$ uniformly in k, since:

- $\int_{B_R^c} \rho_k(x)dx = \mu_k(B_R^c)$ which tends to zero uniformly in k thanks to the tightness of the narrowly converging sequence μ_k;
- $\int_{B_R^c} \rho_k(x)(1+|x|^2)dx$ is uniformly bounded since so are the second moments of the measures μ_k;
- $\int_{B_R^c} \frac{1}{(1+|x|^2)^{\alpha p'}}dx \to 0$ as $R\to\infty$ since the integrand is an L^1 function.

Therefore, if we let $R\to\infty$ in (2.2.9), we get

$$\begin{aligned}\int_{\mathbb{R}^n} U(\rho)\, d\mathscr{L}^n &\le \liminf_{R\to\infty}\liminf_{k\to\infty}\left[\int_{\mathbb{R}^n} U(\rho_k)\, d\mathscr{L}^n - \int_{B_R^c} U^+(\rho_k)\, d\mathscr{L}^n\right] = \\ &= \sup_R \liminf_{k\to\infty}\left[\int_{\mathbb{R}^n} U(\rho_k)\, d\mathscr{L}^n - \int_{B_R^c} U^+(\rho_k)\, d\mathscr{L}^n\right] \le \\ &\le \liminf_{k\to\infty}\sup_R\left[\int_{\mathbb{R}^n} U(\rho_k)\, d\mathscr{L}^n - \int_{B_R^c} U^+(\rho_k)\, d\mathscr{L}^n\right] = \liminf_{k\to\infty}\int_{\mathbb{R}^n} U(\rho_k)\, d\mathscr{L}^n\end{aligned}$$

and semicontinuity is proven. □

Theorem 2.2.26 (Convexity). *Suppose that U satisfies **McCann's condition**:*

$$s\mapsto s^n U(s^{-n}) \quad \textit{is convex and nonincreasing for } s\in(0,\infty). \tag{MC}$$

Then $\mathcal{U}$ is convex along every *geodesic of $\mathscr{P}_2(\mathbb{R}^n)$.*

Remark 2.2.27 ((MC) implies convexity). McCann's condition is stronger than convexity of U. Firstly, the required monotonicity can be equivalently stated saying that $z\mapsto z^{-1}U(z)$ is nondecreasing: recalling that $U(0)=0$, this gives the convexity inequality when one of the two extremes is 0. Secondly, we observe that $\Phi(z) := zU(z^{-1})$ is a convex function of z, as composition of a convex nonincreasing map with the concave map $z\mapsto z^{1/n}$. Then, for every $a,b>0$ and every $t\in[0,1]$, we call $s\in[0,1]$ the real number such that $(1-s)a^{-1}+sb^{-1} = [(1-t)a+tb]^{-1}$; i.e. $s = tb[(1-t)a+tb]^{-1}$. We have

$$\begin{aligned}U((1-t)a+tb) &= \Phi\left([(1-t)a+tb]^{-1}\right)[(1-t)a+tb] = \frac{\Phi\left((1-s)a^{-1}+sb^{-1}\right)}{(1-s)a^{-1}+sb^{-1}} \le \\ &\le \frac{(1-s)\Phi\left(a^{-1}\right)+s\Phi\left(b^{-1}\right)}{(1-s)a^{-1}+sb^{-1}} = \frac{(1-s)a^{-1}U(a)+sb^{-1}U(b)}{(1-s)a^{-1}+sb^{-1}}\end{aligned}$$

which, substituting $s = tb[(1-t)a+tb]^{-1}$, after a little algebra turns out to be equal to $(1-t)U(a)+tU(b)$.

To prove the theorem, we first establish an "explicit" formula for $\mathcal{U}$ along a geodesic:

Proposition 2.2.28. *Let $\mu_0, \mu_1 \in \mathscr{P}_2^a(\mathbb{R}^n)$, $\mu_0 = \rho_0 \mathscr{L}^n$. Call T the optimal transport map from μ_0 to μ_1, and $T_t := (1-t)Id + tT$, so that the constant speed geodesic between μ_0 and μ_1 is $\mu_t = (T_t)_\# \mu_0$. Recall that we can put $T_1 = \nabla f$ where f is convex. Then there exists a μ_0-full set Σ such that for every $t \in [0,1]$ the following hold:*

1. *$\mu_t \ll \mathscr{L}^n$, and T_t is injective on Σ;*
2. *On Σ, besides being injective, T_t is differentiable with ∇T_t symmetric and $JT_t := \det \nabla T_t > 0$. Moreover, it holds $\dfrac{d\mu_t}{d\mathscr{L}^n} = \left[\left(\dfrac{\rho_0}{JT_t}\right) \circ T_t^{-1}\right] \cdot I_{T_t(\Sigma)}$;*
3. *$\mathcal{U}(\mu_t) = \displaystyle\int_\Sigma U\left(\frac{\rho_0}{JT_t}\right) JT_t \; dt$.*

Proof of the Proposition.
$\boxed{1.}$ In Remark 2.1.1 we already proved that, for $t \in [0,1)$, $\mu_t \ll \mathscr{L}^n$ and T_t is injective on the set of differentiability points of f, which is a μ_0-full set. The (essential) injectivity of T_1 is a consequence of the (essential) invertibility of the optimal map T_1 (see Remark 1.2.18).
$\boxed{2.}$ Aleksandrov's theorem (see [10]) tells us that $T_1 = \nabla f$ is a.e. differentiable on its domain (hence, μ_0-a.s. in $\mathbb{R}^n$), with $\nabla T_1 = \nabla^2 f$ a symmetric positive semidefinite matrix. Note that $JT_1 > 0$ μ_0-a.s.: in fact $\mu_0\{JT_1 = 0\} \le \mu_1(T_1\{JT_1 = 0\})$ which is zero since $\mathscr{L}^n(T_1\{JT_1 = 0\}) = 0$ by the area formula.

Hence, $\nabla T_t = t\nabla T_1 + (1-t)Id$ exists and is simmetric positive definite on the same μ_0-full set. To sum up, there exists a μ_0-full set Σ on which T_t is injective and differentiable with $JT_t > 0$. But then, by the change of variables formula, for every $\psi \ge 0$ it holds

$$\int_{\mathbb{R}^n} \psi \; d\mu_t = \int_\Sigma (\psi \circ T_t(x))\rho_0(x)dx = \int_\Sigma \psi(T_t(x)) \frac{\rho_0(x)}{JT_t(x)} JT_t(x)dy = \\ = \int_{T_t(\Sigma)} \psi(y) \frac{\rho_0}{JT_t} \circ T_t^{-1}(y)dy$$

from which the conclusion follows.
$\boxed{3.}$ By the previous point, and changing variables once again,

$$\mathcal{U}(\mu_t) = \int_{T_t(\Sigma)} U\left(\frac{\rho_0}{JT_t} \circ T_t^{-1}(y)\right) dy \overset{y:=T_t(x)}{=} \int_\Sigma U\left(\frac{\rho_0(x)}{JT_t(x)}\right) JT_t(x) \; dx.$$

$\square$

Proof of Theorem 2.2.26. The convexity inequality is trivial unless $\mu_0, \mu_1 \ll \mathscr{L}^n$. In view of the Proposition, it is sufficient to prove the convexity of $t \mapsto U\left(\frac{\rho_0(x)}{JT_t(x)}\right) JT_t(x)$ for every x fixed.

Note that, as a function of t, this is the composition of the affine function $t \mapsto tT - (1-t)Id$, the map $A \mapsto (\det A)^{1/n}$, and the convex nonincreasing map

$s \mapsto s^n U(s^{-n}\rho_0(x))$. Using again that $a \circ b$ is convex if a is convex nonincreasing and b is concave, it is sufficient to prove that $A \mapsto (\det A)^{1/n}$ is concave at least for

$$A \in \mathrm{Sym}_+^{n\times n} := \{\text{symmetric positive definite } n \times n \text{ matrices}\}.$$

First of all, let us prove that

$$[\det((1-t)A + tId)]^{1/n} \geq (1-t)[\det(A)]^{1/n} + t. \qquad \forall A \in \mathrm{Sym}_+^{n\times n}. \tag{2.2.10}$$

With a change of basis, we can suppose that A is diagonal with diagonal elements $\lambda_i > 0$. The thesis becomes

$$(1-t)\prod_{i=1}^n \frac{\lambda_i^{1/n}}{[(1-t)\lambda_i + t]^{1/n}} + t\prod_{i=1}^n \frac{1}{[(1-t)\lambda_i + t]^{1/n}} \leq 1.$$

But estimating the two geometric means which appear at the left hand side with the corresponding arithmetic means, we get precisely that

$$\mathrm{LHS} \leq \frac{1}{n}\sum_{i=1}^n \left[\frac{\lambda_i(1-t)}{(1-t)\lambda_i + t} + \frac{t}{(1-t)\lambda_i + t}\right] = 1.$$

Finally, for every $A, B \in \mathrm{Sym}_+^{n\times n}$, we have

$$\det((1-t)A + tB) = \det(B)\det((1-t)AB^{-1} + tId),$$

so (2.2.10) implies the concavity inequality. □

Remark 2.2.29 (Candidate subdifferential). Let us perform a heuristic formal computation: if $(\rho_t \mathscr{L}^n, v_t)$ satisfies the continuity equation (CE), then hopefully

$$\frac{d}{dt}\mathcal{U}(\rho_t \mathscr{L}^n) = \int_{\mathbb{R}^n} U'(\rho_t(x))\frac{d}{dt}\rho_t(x)dx = \int_{\mathbb{R}^n} \langle \nabla[U' \circ \rho_t], v_t\rangle \rho_t d\mathscr{L}^n .$$

Hence, the chain rule suggests that that $\partial^W \mathcal{U}(\rho_0 \mathscr{L}^n)$ contains the element $\nabla[U' \circ \rho_0]$.

Note that, defining the **Legendre transform** of U by $L_U(z) := zU'(z) - U(z)$, our candidate subdifferential may be written as $\frac{\nabla[L_U \circ \rho]}{\rho}$. (We will conventionally put $L_U(0) = 0$ and $\frac{0}{0} = 0$.) To make the above intuition into a formal theorem, we will need some technical hypotheses: one of them is the following.

Definition 2.2.30. The potential U satisfies the **doubling condition** if there exists $C > 0$ such that $U(2z) \leq C(1 + U(z))$ for every $z \geq 0$.

Remark 2.2.31. Since U is convex and $U(0) = 0$, then $U(tz) \leq tU(z)$ for every $t \in [0, 1]$; combining this with the doubling condition, we immediately obtain that for every R there exists M such that $U^+(rz) \leq M(1 + U^+(z))$ for every $r \in [0, R]$ and every z.

Moreover, writing $U(x + y) = U(2\frac{x+y}{2})$ and using first the doubling condition and then convexity, we see that $U(x + y) \leq C'(1 + U(x) + U(y))$ for some other constant C'.

Theorem 2.2.32 (Subdifferentiability)**.** *Suppose that the potential U is $C^1(0,\infty)$ and satisfies the doubling condition just defined. Let $\mu := \rho\mathscr{L}^n$ be such that $\mathcal{U}(\mu) < \infty$. Consider the two conditions:*

1. *The slope $|\nabla\mathcal{U}|(\mu)$ is finite;*
2. *$[L_U \circ \rho] \in W^{1,1}_{\rm loc}(\mathbb{R}^n)$ and $\frac{\nabla[L_U\circ\rho]}{\rho} \in L^2(\mu;\mathbb{R}^n)$ (with the convention $\frac{0}{0} = 0$).*

Then $1 \Rightarrow 2$. *Moreover, if in addition U satisfies McCann's condition* (MC)*, then* 1. *and* 2. *are equivalent, and they imply* $\partial^W\mathcal{U}(\mu) \cap \mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^n) = \left\{\dfrac{\nabla[L_U \circ \rho]}{\rho}\right\}$.

Proof of $1 \Rightarrow 2$. Take any $\psi \in C^\infty_c(\mathbb{R}^n;\mathbb{R}^n)$, call $T_\varepsilon := Id + \varepsilon\,\psi$ and consider $\mu_\varepsilon := (T_\varepsilon)_\#\mu$. Note that $JT_\varepsilon > 0$ everywhere if ε is sufficiently small, so arguing as in Proposition 2.2.28 we get that $\mathcal{U}(\mu_\varepsilon) = \int_{\rho>0} U\left(\frac{\rho}{JT_\varepsilon}\right) JT_\varepsilon$. By direct computation, the derivative of the integrand with respect to ε is

$$I(\varepsilon, x) := -L_U\left(\frac{\rho}{JT_\varepsilon}\right)\frac{d}{d\,\varepsilon}JT_\varepsilon.$$

We would like to pass the derivative in $\varepsilon = 0$ under the integral sign; this is correct because:

Claim. $I(\varepsilon, x)$ is controlled by an integrable function of x, independently of ε in a neighbourhood of 0.

This is a tedious verification; however, since it is the point where the doubling hypothesis plays its role, we will perform it. But before, let us see how the claim is used to conclude the proof.

Passing the derivative under the integral sign, and since by elementary calculus $\frac{d}{d\,\varepsilon}\big|_{\varepsilon=0} JT_\varepsilon = \mathrm{div}(\psi)$, we get

$$\mathcal{U}(\mu_\varepsilon) - \mathcal{U}(\mu) = -\,\varepsilon\int L_U(\rho)\mathrm{div}(\psi)\;dx + o(\varepsilon).$$

But the left hand side, by definition of slope, is greater or equal than

$$-|\nabla\mathcal{U}|(\mu)W_2(\mu,\mu_\varepsilon)+o(W_2(\mu,\mu_\varepsilon)) \overset{\text{definition of } W_2}{\geq} -|\nabla\mathcal{U}|(\mu)\,\varepsilon\left(\int|\psi|^2\rho\;dx\right)^{1/2}+o(\varepsilon),$$

so we can conclude that

$$\int L_U(\rho)\mathrm{div}(\psi)\;dx \leq |\nabla\mathcal{U}|(\mu)\left(\int|\psi|^2\rho\;dx\right)^{1/2};$$

since this is true also for $-\psi$, we deduce that

$$\left|\int L_U(\rho)\mathrm{div}(\psi)\;dx\right| \leq |\nabla\mathcal{U}|(\mu)\left(\int|\psi|^2\rho\;dx\right)^{1/2}.$$

This relation means that $\psi \mapsto \int L_U(\rho)\mathrm{div}(\psi)dx$ is the restriction to $C^\infty_c(\mathbb{R}^n;\mathbb{R}^n)$ of a continuous functional on $L^2(\mu;\mathbb{R}^n)$, with norm less or equal than $|\nabla\mathcal{U}|(\mu)$.

So, by Riesz's theorem, we conclude that there exists $g \in L^2(\mu;\mathbb{R}^n)$ such that $\|g\|_{L^2(\mu;\mathbb{R}^n)} \le |\nabla\mathcal{U}|(\mu)$ and

$$\int L_U(\rho)\mathrm{div}(\psi)\ dx = \int g\cdot\psi\ d\mu = \int (g\rho)\cdot\psi\ dx$$

for every $\psi \in C_c^\infty(\mathbb{R}^n;\mathbb{R}^n)$; i.e. $g\rho$ is the distributional derivative of $L_U\circ\rho$. The fact that $g \in L^2(\mu;\mathbb{R}^n)$ translates exactly into $\frac{\nabla[L_U\circ\rho]}{\rho} \in L^2(\mu;\mathbb{R}^n)$. We emphasise for future use that

$$\left\|\frac{\nabla[L_U\circ\rho]}{\rho}\right\|_{L^2(\mu;\mathbb{R}^n)} \le |\nabla\mathcal{U}|(\mu). \tag{2.2.11}$$

Moreover, since $L^1(\mu) \subseteq L^2(\mu)$, we have that $\int \left|\frac{\nabla[L_U\circ\rho]}{\rho}\right| \rho\ dx < \infty$, and therefore $\nabla[L_U\circ\rho] \in L^1(\mathscr{L}^n)$.

Finally, we prove that $[L_U\circ\rho] \in L^1_{\mathrm{loc}}(\mathscr{L}^n)$. In fact, by convexity

$$U(z) = U(z) - U(0) \le U'(z)z \le U(2z) - U(z) \le C(1+U(z)) - U(z),$$

from which $0 \le L_U(z) \le C'(1+|U(z)|)$, and since $U\circ\rho \in L^1(\mathscr{L}^n)$ by hypothesis, then the conclusion follows. □

Proof of the Claim. Firstly, we note that for every ε, $I(\varepsilon,x) = 0$ if x is outside the compact set $K := \mathrm{supp}(\psi)$. Secondly, we can exploit the just proved inequalities $0 \le L_U(z) \le C'(1+|U(z)|)$ to infer

$$|I(\varepsilon,\cdot)| \le C'\left[1 + U^+\left(\frac{\rho}{JT_\varepsilon}\right) + U^-\left(\frac{\rho}{JT_\varepsilon}\right)\right]\cdot\left|\frac{d}{d\varepsilon}JT_\varepsilon\right|.$$

Let us choose $\varepsilon_0 > 0$ such that $\frac{1}{2} \le JT_\varepsilon(x) \le \frac{3}{2}$ for (x,ε) in the compact set $K\times[-\varepsilon_0,\varepsilon_0]$; then the desired estimate follows, because:

- $U^+\left(\frac{\rho}{JT_\varepsilon}\right) \le M(1+U^+(\rho))$ by Remark 2.2.31, and $U^+(\rho)$ is integrable because $\mathcal{U}(\mu) < \infty$;

- As we saw in Remark 2.2.24,

$$U^-\left(\frac{\rho}{JT_\varepsilon}\right) \le c_1\left(\frac{\rho}{JT_\varepsilon}\right) + I_{[0,1]}(\alpha)\ c_2\left(\frac{\rho}{JT_\varepsilon}\right)^\alpha \le \tilde{C}\left(I_{[0,1]}(\alpha)\rho^\alpha + \rho\right)$$

  which is integrable;

- $\frac{d}{d\varepsilon}JT_\varepsilon(x)$ is equibounded for (x,ε) in the compact set $K\times[-\varepsilon_0,\varepsilon_0]$.

□

For the remaining part of the proof, we need an approximation lemma.

Lemma 2.2.33. *Suppose that U satisfies the doubling condition, and consider a probability measure $\nu \in \mathscr{P}_2^a(\mathbb{R}^n)$ such that $\mathcal{U}(\nu) < \infty$. Then there exist compactly supported $\nu_k \in \mathscr{P}_2^a(\mathbb{R}^n)$ W_2-converging to ν and such that $\mathcal{U}(\nu_k) \to \mathcal{U}(\nu)$.*

Proof. If $\nu = \rho \mathscr{L}^n$, we take $\nu_k := \frac{1}{z_k}\rho I_{B_k}\mathscr{L}^n$ where $z_k \uparrow 1$ are normalizing constants. The convergence $\nu_k \to \nu$ in $\mathscr{P}_2$ is immediate using the basic criterion (Theorem 1.3.6). Moreover, the densities converge in L^1, and so the argument in the proof of semicontinuity (Theorem (2.2.25), from equation (2.2.9)) can be repeated literally to get $\mathcal{U}(\nu) \le \liminf \mathcal{U}(\nu_k)$. It remains only to prove that $\limsup \mathcal{U}(\nu_k) \le \mathcal{U}(\nu)$.

We note that, since $U(0) = 0$, from the definition of convexity one deduces easily that U^+ is nondecreasing. Fix any $\beta \in (1, 2)$: since $\frac{1}{\beta z_k} < 1$ if $k \gg 1$, then

$$\limsup_{k\to\infty} \mathcal{U}(\nu_k) = \limsup_{k\to\infty} \int_{B_k} U\left(\frac{\rho}{z_k}\right) d\mathscr{L}^n \overset{\text{convexity}}{\le} \limsup_{k\to\infty} \frac{1}{\beta z_k} \int_{B_k} U(\beta\rho) d\mathscr{L}^n .$$

If we knew that $U^+(2\rho)$ is integrable, then thanks to $U^+(\beta\rho) \le U^+(2\rho)$ we could apply Fatou's lemma and get

$$\limsup_{k\to\infty} \mathcal{U}(\nu_k) \le \frac{1}{\beta} \int_{\mathbb{R}^n} U(\beta\rho) d\mathscr{L}^n$$

and another application of Fatou's lemma, this time for $\beta \downarrow 1$, would yield the desired inequality.

Hence, we must only show that $U^+(2\rho)$ is integrable. If $U^+ \equiv 0$, this is trivial. Otherwise, take $z_0 > 0$ such that $U^+(z_0) > 0$: then $U^+(z) \ge U^+(z_0)$ for every $z \ge z_0$, so the doubling condition implies

$$U^+(2z) \le C(1 + U^+(z)) \le C\left(\frac{U^+(z)}{U^+(z_0)} + U^+(z)\right) \le \bar{C}U^+(z) \qquad \forall z \ge z_0,$$

and since $U(z) \le \dfrac{z}{z_0}U(z_0)$ for every $z \le z_0$, we conclude

$$\int_{\mathbb{R}^n} U^+(2\rho) d\mathscr{L}^n \le \int_{\mathbb{R}^n} \tilde{C}(\rho + U^+(\rho)) d\mathscr{L}^n < \infty.$$

□

Proof of Theorem 2.2.32, second part. Suppose that (MC) holds. Defining

$$w := \frac{\nabla[L_U \circ \rho]}{\rho},$$

we firstly want to prove that

$$\mathcal{U}(\nu) - \mathcal{U}(\mu) \ge \int \langle w, T_\mu^\nu - Id\rangle d\mu \qquad \forall \nu \in \mathscr{P}_2(\mathbb{R}^n) \tag{2.2.12}$$

and of course we can suppose that $\mathcal{U}(\nu) < \infty$. Thanks to the Lemma, we immediately reduce to the case in which $\mathrm{supp}(\nu)$ is compact. We write T instead of T_μ^ν for brevity, and put $T_t := (1-t)Id + tT$, $\nu_t := (T_t)_{\#}\mu$ as usual.

In the proof of Theorem 2.2.26 we saw that $\mathcal{U}(\nu_t) = \int_\Sigma U\left(\frac{\rho}{JT_t}\right) JT_t \, d\mathscr{L}^n$ where Σ is a μ-full set and that the integrand is a convex function of t. Therefore the difference quotients $\frac{1}{t}\left[U\left(\frac{\rho}{JT_t}\right) JT_t - U(\rho)\right]$ decrease as $t \downarrow 0$ to the value

$$\left.\frac{d}{dt}\right|_{t=0} \left[U\left(\frac{\rho}{JT_t}\right) JT_t\right] = -L_U(\rho) \left.\frac{d}{dt}\right|_{t=0} JT_t = -L_U(\rho) \cdot \mathrm{div}(T - Id)$$

where div denotes pointwise divergence. Therefore

$$\mathcal{U}(\nu) - \mathcal{U}(\mu) = \int \left[U\left(\frac{\rho}{JT_1}\right) JT_1 - U(\rho) \right] d\mathscr{L}^n \geq \int L_U(\rho) \cdot \mathrm{div}(Id - T)\ d\mathscr{L}^n\,. \tag{2.2.13}$$

If we could integrate by parts, we would immediately obtain (2.2.12). However we cannot, both because the divergence is not distributional, and because the supports are not compact.

As for the first problem, we recall that by Brenier's theorem $T = \nabla f$ with f convex, so it is not difficult to see that the distributional derivative DT is positive semidefinite; therefore the distributional divergence $\mathrm{Div}(T)$ is nonnegative, and so it is greater or equal than its absolutely continuous part $\mathrm{div}(T)$. We will use this in a moment.

Now we try to solve the issue of the support choosing compactly supported cutoff functions $\eta_k \in W^{1,\infty}(\mathbb{R}^n)$ such that $0 \leq \eta_k \uparrow 1$ for $k \to \infty$. Recall that $L_U \circ \rho \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^n)$: therefore $\eta_k \cdot (L_U \circ \rho) \in W^{1,1}(\mathbb{R}^n)$ with

$$\nabla[\eta_k \cdot (L_U \circ \rho)] = (\nabla \eta_k) \cdot (L_U \circ \rho) + \eta_k \nabla[L_U \circ \rho].$$

Note that for $g \in C^\infty_c(\mathbb{R}^n)$ nonnegative,

$$\int g \cdot \mathrm{div}(Id)\ d\mathscr{L}^n = -\int \langle \nabla g, Id \rangle\ d\mathscr{L}^n,$$

$$-\int g \cdot \mathrm{div}(T)\ d\mathscr{L}^n \geq -\int g \cdot \mathrm{Div}(T)\ d\mathscr{L}^n = \int \langle \nabla g, T \rangle\ d\mathscr{L}^n\,.$$

Summing the two relations above,

$$\int g \cdot \mathrm{div}(Id - T)\ d\mathscr{L}^n \geq \int \langle \nabla g, T - Id \rangle\ d\mathscr{L}^n\,. \tag{2.2.14}$$

We claim that if $g_h \to g$ a.e. and in $W^{1,1}$, $g_h \geq 0$, and the supports of g_h, g are all contained in a common bounded subset B of $\mathrm{supp}(\mu)$, then this inequality goes to the limit. In fact:

- $\int g_h \cdot \mathrm{div}(Id)\ d\mathscr{L}^n = n \int g_h\ d\mathscr{L}^n$ obviously goes to the limit, and this limit is finite;

- $\int g \cdot \mathrm{div}(T)\ d\mathscr{L}^n \leq \liminf_h \int g_h \cdot \mathrm{div}(T)\ d\mathscr{L}^n$ by Fatou's lemma (recall that $g_h, \mathrm{div}(T) \geq 0$);

- The right hand side goes to the limit because $\nabla g_h \xrightarrow{L^1} \nabla g$, and $T - Id$ is bounded on B thanks to compactness of the support of ν.

Therefore we can put $g = L_U \circ \rho$ in (2.2.14): the result is

$$\begin{aligned}\int \eta_k L_U(\rho) \cdot \mathrm{div}(Id - T)\ d\mathscr{L}^n &\geq \\ &\geq \int \eta_k \langle \nabla[L_U \circ \rho], T - Id \rangle\ d\mathscr{L}^n + \int L_U(\rho) \langle \nabla \eta_k, T - Id \rangle\ d\mathscr{L}^n.\end{aligned} \tag{2.2.15}$$

Letting $k \to \infty$, the left hand side tends to $\int L_U(\rho) \cdot \mathrm{div}(Id - T)$, as we see using twice the monotone convergence theorem (recall that $L_U \geq 0$). The first integral at the right hand side converges to $\int \langle \nabla[L_U \circ \rho], T - Id \rangle = \int \langle w, T - Id \rangle \, d\mu$ by dominated convergence: in fact $\langle \nabla[L_U \circ \rho], T - Id \rangle \in L^1(\mathscr{L}^n)$, since this is equivalent to $\langle w, T - Id \rangle \in L^1(\mu)$ which is true by Cauchy-Schwarz's inequality because $T, Id \in L^2(\mu; \mathbb{R}^n)$ (as $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^n)$) and $w \in L^2(\mu, \mathbb{R}^n)$ by hypothesis.

So, if we could neglect the second integral at the right hand side of (2.2.15), we would get exactly the desired (2.2.12). Let us choose $\eta_k(x) := \phi_k(|x|)$ where $\phi_k : [0, \infty) \to \mathbb{R}$ is the only continuous function equal to 1 on $[0, k-1]$, affine on $[k-1, k]$, and equal to 0 on $[k, \infty)$. The function η_k is concave on the ball B_k, and so for $x \in B_k$ we have

$$\langle \nabla \eta_k(x), T(x) - x \rangle \geq \eta_k(T(x)) - \eta_k(x).$$

But the support of ν is compact: hence, for k sufficiently large, $\eta_k(T(x)) = 1$ for μ-a.e. x. So we have proven $\langle \nabla \eta_k(x), T(x) - x \rangle \geq 0$ for μ-a.e. x (in B_k, but also outside since there the inequality is trivial). Therefore, the last integral in (2.2.15) is nonnegative, as we wanted.

To resume, till now we have proven (2.2.12), i.e. that $w \in \partial^W \mathcal{U}(\mu)$. This obviously yields the finiteness of the slope; more precisely, if $\tilde{w}$ is any element of $\partial^W \mathcal{U}(\mu)$, then $|\nabla \mathcal{U}|(\mu) \leq \|\tilde{w}\|_{L^2(\mu;\mathbb{R}^n)}$. But the finiteness of the slope allows the use of the first part of the theorem; and during its proof we obtained that $\|w\|_{L^2(\mu;\mathbb{R}^n)} \leq |\nabla \mathcal{U}|(\mu)$ (equation (2.2.11)). Therefore, the L^2 norm of w is minimal in $\partial^W \mathcal{U}(\mu)$, and so by Remark 2.2.9 we conclude that $w \in \mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^n)$.

Eventually, suppose that $w' \in \mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^n) \cap \partial^W \mathcal{U}(\mu)$. Take any $\phi \in C_c^\infty(\mathbb{R}^n)$, and consider $T_t := Id + t\nabla\phi$: for t small this is the gradient of a convex function, and as such it is optimal by Brenier's theorem. So, at least for small values of t, we still have the representation formula

$$\frac{1}{t}\left[\mathcal{U}(\nu_t) - \mathcal{U}(\mu)\right] = \int \frac{1}{t}\left[U\left(\frac{\rho}{JT_t}\right) JT_t - U(\rho)\right] d\mathscr{L}^n$$

in which the left hand side is by hypothesis greater or equal than $\int \langle w', \nabla\phi \rangle \, d\mu + o(1)$, while the right hand side decreasingly converges to

$$-\int L_U(\rho) \cdot \mathrm{div}(\nabla\phi) \, d\mathscr{L}^n = \int \langle \nabla[L_U \circ \rho], \nabla\phi \rangle \, d\mathscr{L}^n = \int \langle w, \nabla\phi \rangle \, d\mu.$$

In conclusion,

$$\int \langle w' - w, \nabla\phi \rangle \, d\mu \leq 0 \qquad \forall \phi \in C_c^\infty(\mathbb{R}^n);$$

using this inequality for both ϕ and $-\phi$ we see that in fact

$$\int \langle w' - w, \nabla\phi \rangle \, d\mu = 0 \qquad \forall \phi \in C_c^\infty(\mathbb{R}^n),$$

i.e. $w' - w \perp \{\nabla\phi : \phi \in C_c^\infty(\mathbb{R}^n)\}$ in $L^2(\mu; \mathbb{R}^n)$. But $w - w' \in \mathrm{Tan}_\mu \mathscr{P}_2(\mathbb{R}^n)$, and recalling the definition of this space this forces $w - w' = 0$. □

Example 2.2.34 (Entropy)**.** All the above results can be applied to the potential $U(z) := z \log z$. The resulting functional is called **entropy**, and is therefore geodesically convex and lower semicontinuous in $\mathscr{P}_2(\mathbb{R}^n)$.

2.3 Gradient flows

We want to adapt to our context the following concept from classical analysis:

Definition 2.3.1. Let H be a Hilbert space, and $F : H \to (-\infty, +\infty]$ be a functional; denote by ∂F the subdifferential of F. Then $u \in AC_{\mathrm{loc}}((0,\infty); H)$ is a **gradient flow** of F if $u'(t) \in -\partial F(u(t))$ for a.e. t (in particular, $u(t) \in \mathrm{Dom}(F)$ for a.e. t). If $\lim_{t\to 0} u(t) = \bar{x}$, then we will say that the gradient flow **starts** from $\bar{x}$.

Example 2.3.2 (Heat flow)**.** Take $H := L^2(\mathbb{R}^n)$ and put

$$F(u) := \begin{cases} \frac{1}{2}\int_{\mathbb{R}^n} |\nabla u|^2 d\mathscr{L}^n, & \text{if } u \in W^{1,2}(\mathbb{R}^n) \\ +\infty, & \text{otherwise,} \end{cases}$$

which is obviously convex. In the sequel, Δ will denote the distributional Laplacian. We claim that $\partial F(u) \neq \varnothing$ if and only if $\Delta u \in L^2(\mathbb{R}^n)$, and that in this case $\partial F(u) = \{-\Delta u\}$: so, the gradient flow equation for F reduces to the **heat equation** $\frac{\partial u}{\partial t}(t,x) = \Delta_x u(t,x)$.

In fact, if $\Delta u \in L^2(\mathbb{R}^n)$, then for every $\phi \in W^{1,2}(\mathbb{R}^n)$ we have

$$F(u+\phi) = F(u) + \int \langle \nabla u, \nabla \phi \rangle d\mathscr{L}^n + \int |\nabla \phi|^2 d\,\mathscr{L}^n \geq F(u) + \langle -\Delta u, \phi \rangle_{L^2};$$

for $\phi \notin W^{1,2}(\mathbb{R}^n)$ the inequality is trivially true, so we conclude $-\Delta u \in \partial F(u)$. Conversely, if $\xi \in \partial F(u)$, then by convexity $F(u + \varepsilon\,\phi) - F(u) \geq \varepsilon \langle \xi, \phi \rangle_{L^2}$, which for $\phi \in W^{1,2}(\mathbb{R}^n)$ means

$$\frac{1}{2}\,\varepsilon^2 \int |\nabla \phi|^2 d\mathscr{L}^n + \varepsilon \int \langle \nabla u, \nabla \phi \rangle d\mathscr{L}^n \geq \varepsilon \int \phi \xi \; d\mathscr{L}^n \,.$$

For $\varepsilon \downarrow 0$ we get

$$\int \langle \nabla u, \nabla \phi \rangle d\mathscr{L}^n \geq \int \phi \xi \; d\mathscr{L}^n \qquad \forall \phi \in W^{1,2}(\mathbb{R}^n)$$

and since this is true both for ϕ and for $-\phi$, in fact there is equality. From this, by definition of distributional Laplacian, we get exactly $\Delta u = -\xi$.

One of the main goals of this section is to show that, in a suitable sense, the heat flow can be seen as a gradient flow also in $\mathscr{P}_2(\mathbb{R}^n)$ (of course, the functional will be different).

Since we have a concept of subdifferential also in $\mathscr{P}_2(\mathbb{R}^n)$, it is natural to define:

Definition 2.3.3. Let and $E : \mathscr{P}_2(\mathbb{R}^n) \to (-\infty, +\infty]$ be a functional such that $\mathrm{Dom}(|\nabla E|) \subseteq \mathscr{P}_2^a(\mathbb{R}^n)$. Then $(\mu_t) \in AC_{\mathrm{loc}}((0,\infty); \mathscr{P}_2(\mathbb{R}^n))$ is a **gradient flow** of E if its tangent velocity field v_t satisfies

$$v_t \in -\partial^W E(\mu_t) \qquad \text{for a.e. } t$$

(in particular, $\mu_t \in \mathscr{P}_2^a(\mathbb{R}^n)$ for a.e. t). If $\lim_{t\to 0} u(t) = \bar{\mu}$ in $\mathscr{P}_2(\mathbb{R}^n)$, then we will say that the gradient flow **starts** from $\bar{\mu}$.

Another approach would be to look for a purely metric generalisation of the concept of gradient flow. There is a variety of possible definitions: the detailed study of them, also in connection with the differential one given above, is the subject of the monograph [3]. We will only introduce one of the strongest definitions, and prove its equivalence with the differential one in our setting: our motivation for this digression is that some of the properties we are going to deal with are more transparently proved forgetting the rich additional structure we endowed $\mathscr{P}_2(\mathbb{R}^n)$ with.

The basic remark is that in a Hilbert space H, $u \in AC_{\mathrm{loc}}((0,\infty); H)$ is a gradient flow of the λ-convex functional F if and only if

$$F(y) - F(u(t)) \geq -\langle u'(t), y - u(t)\rangle + \frac{\lambda}{2}|y - u(t)|^2 \qquad \forall y \in H \text{ for a.e. } t.$$

But the right hand side can be written as $\frac{1}{2}\frac{d}{dt}|y-u(t)|^2 + \frac{\lambda}{2}|y-u(t)|^2$, involving only the distance and not the scalar product. This suggests the following definition.

Definition 2.3.4. Let (X, d) be a metric space and $F : X \to (-\infty, +\infty]$ be a functional. Then $u \in AC_{\mathrm{loc}}((0,\infty); X)$ is an **EVI($\boldsymbol{\lambda}$)-gradient flow** of F if

$$\frac{1}{2}\frac{d}{dt}d^2(u(t), y) + F(u(t)) + \frac{\lambda}{2}d^2(u(t), y) \leq F(y) \qquad \forall y \in X \text{ for a.e. } t. \tag{EVI}$$

(EVI stands for "evolution variational inequality".)

The equivalence of the two formulations, in the Hilbert setting, is an immediate consequence of the expression for the derivative $\frac{d}{dt}d^2(u(t), y)$. Analogously, the equivalence in $\mathscr{P}_2(\mathbb{R}^n)$ relies on the following:

Theorem 2.3.5 (Derivative of W_2)**.** *Let (μ_t) be an AC curve in $\mathscr{P}_2(\mathbb{R}^n)$ and call (v_t) its tangent velocity field. Then for a.e. t it holds*

$$\frac{d}{dt}W_2^2(\mu_t, \sigma) = 2\int_{(\mathbb{R}^n)^2} \langle x - y, v_t(x)\rangle d\gamma(x, y) \qquad \forall \sigma \in \mathscr{P}_2(\mathbb{R}^n)\ \forall \gamma \in \Gamma_o(\mu_t, \sigma).$$

Proof. We will prove that the equality is true (for a certain σ and every γ) if (2.1.9) holds in t and t is a differentiability point of $t \mapsto W_2(\mu_t, \sigma)$. Note that for a.e. t, this is true for all $\sigma \in \mathscr{P}_2(\mathbb{R}^n)$: in fact, if $(\sigma_n)_{n\in\mathbb{N}} \subseteq \mathscr{P}_2(\mathbb{R}^n)$ is any sequence, then $t \mapsto W^2(\mu_t, \sigma_n)$ is differentiable for every n, for every t out of null-set N; and from this, if (σ_n) is dense (recall that $\mathscr{P}_2(\mathbb{R}^n)$ is separable by Proposition 1.3.4), it is easy to prove that

$$\left(\frac{W_2(\mu_{t+h}, \sigma) - W_2(\mu_t, \sigma)}{h}\right)_h$$

is Cauchy for $h \to 0$ for every $\sigma \in \mathscr{P}_2(\mathbb{R}^n)$ and every $t \notin N$.

Property (2.1.9) implies that $W_2(\mu_{t+h}, \sigma) - W_2((Id + hv_t)_\# \mu_t, \sigma) = o(h)$, and so the limit which defines $\frac{d}{dt}W_2^2(\mu_t, \sigma)$ can be computed as

$$\lim_{h\to 0} \frac{W_2^2((Id + hv_t)_\# \mu_t, \sigma) - W_2^2(\mu_t, \sigma)}{h}.$$

Denoting as usual with $p_x, p_y : (\mathbb{R}^n)^2 \to \mathbb{R}^n$ the canonical projections, we can estimate $W_2^2((Id + hv_t)_\# \mu_t, \sigma)$ using the plan $\eta := (p_x + hv_t \circ p_x, p_y)_\# \gamma$: we obtain

$$W_2^2((Id+hv_t)_\#\mu_t,\sigma) \leq \int_{(\mathbb{R}^n)^2} |x-y|^2 d\eta(x,y) = \int_{(\mathbb{R}^n)^2} |x+hv_t(x)-y|^2 d\gamma(x,y) =$$
$$= \int_{(\mathbb{R}^n)^2} |x-y|^2 d\gamma(x,y) + 2h \int_{(\mathbb{R}^n)^2} \langle x-y, v_t(x)\rangle d\gamma(x,y) + o(h) =$$
$$= W_2^2(\mu_t,\sigma) + 2h \int_{(\mathbb{R}^n)^2} \langle x-y, v_t(x)\rangle d\gamma(x,y) + o(h).$$

Considering separately the cases $h > 0$ and $h < 0$, we divide the above inequality by h to get

$$\lim_{h\to 0+} \frac{W_2^2((Id+hv_t)_\#\mu_t,\sigma) - W_2^2(\mu_t,\sigma)}{h} \leq 2\int_{(\mathbb{R}^n)^2} \langle x-y, v_t(x)\rangle d\gamma(x,y),$$
$$\lim_{h\to 0-} \frac{W_2^2((Id+hv_t)_\#\mu_t,\sigma) - W_2^2(\mu_t,\sigma)}{h} \geq 2\int_{(\mathbb{R}^n)^2} \langle x-y, v_t(x)\rangle d\gamma(x,y).$$

But we saw that these limits coincide with $\frac{d}{dt} W_2^2(\mu_t, \sigma)$: the proof is complete. □

Corollary 2.3.6 (Equivalence of gradient flow formulations)**.** *Consider a λ-convex functional $E : \mathscr{P}_2(\mathbb{R}^n) \to (-\infty, +\infty]$ such that* $\mathrm{Dom}(|\nabla E|) \subseteq \mathscr{P}_2^a(\mathbb{R}^n)$*. Then u is a gradient flow of E in the "differential" sense if and only if it is an EVI(λ)-gradient flow of E with respect to W_2.*

Proof. Firstly, we prove that if u is an EVI-gradient flow, then $u(t) \in \mathrm{Dom}(|\nabla E|)$ for almost every $t > 0$ (in fact, it is true in general metric spaces, but the proof is more involved: see Theorem 2.3.9). From the theorem above, we can rewrite EVI(λ) as

$$E(u(t)) - E(y) \leq -\int_{(\mathbb{R}^n)^2} \langle x-y, v_t(x)\rangle d\gamma(x,y) - \frac{\lambda}{2} W_2^2(u(t),\sigma)$$
$$\forall \sigma \in \mathscr{P}_2(\mathbb{R}^n)\ \forall \gamma \in \Gamma_o(u(t),\sigma), \text{ for a.e. } t, \quad (2.3.1)$$

where v_t is the tangent field to u. The integral part, by Cauchy-Schwarz's inequality, is less or equal than $W_2(u(t), \sigma) \, \|v_t\|_{L^2}$, from which we immediately deduce $|\nabla E|(u(t)) \leq \|v_t\|_{L^2} < \infty$.

Now in expression (2.3.1), which is equivalent to EVI(λ), we can substitute $\gamma := \left(Id \times T^\sigma_{u(t)}\right)_\# u(t)$; the result, recalling the characterisation of $\partial^W E$ when E is λ-convex (Proposition 2.2.10), is precisely $v_t \in -\partial^W E(u(t))$.

□

Theorem 2.3.7 (Uniqueness and contractivity)**.** *Let $u, v \in AC_{\mathrm{loc}}((0,\infty); \mathbb{R}^n)$ be EVI(λ)-gradient flows of a functional F. Then $d(u(t), v(t)) \leq e^{-\lambda(t-s)} d(u(s), v(s))$ for every $t \geq s > 0$. As a consequence, two gradient flows with the same starting point coincide for every t.*

To prove this, we want to estimate the derivative of $d^2(u(t), v(t))$ knowing bounds on $\frac{\partial}{\partial s}[d^2(u(s), v(t))]$ and $\frac{\partial}{\partial t}[d^2(u(s), v(t))]$. Therefore, the following lemma is useful.

Lemma 2.3.8. *Let $f(s,t) : (a,b)^2 \to \mathbb{R}$ be a map, and suppose that there exists $w \in AC_{\mathrm{loc}}((a,b))$ such that*

$$|f(s,t)-f(s',t)| \le |w(s)-w(s')|,\ |f(s,t)-f(s,t')| \le |w(t)-w(t')| \quad \forall s,s',t,t' \in (a,b).$$

Call $\phi(t) := f(t,t)$. Then $\phi \in AC_{\mathrm{loc}}((0,1))$ and

$$\frac{d\phi}{dt} \le \limsup_{h\to 0+} \frac{f(t,t)-f(t-h,t)}{h} + \limsup_{h\to 0+} \frac{f(t,t+h)-f(t,t)}{h} \quad \text{a.e. in } (a,b).$$

Proof. Translating and rescaling, we can suppose $(a,b) = (0,1)$.

From $|\phi(t)-\phi(s)| \le 2|w(t)-w(s)|$ we see that ϕ is locally absolutely continuous. Therefore, it is sufficient to estimate its distributional derivative: let us take any $\zeta \in C_c^\infty((0,1))$, $\zeta \ge 0$, and compute

$$\int_0^1 \phi'(t)\zeta(t)dt = -\int_0^1 \phi(t)\zeta'(t)dt = -\lim_{h\to 0+}\int_0^1 \phi(t)\frac{\zeta(t+h)-\zeta(t)}{h}dt =$$
$$= \lim_{h\to 0+}\int_0^1 \frac{\phi(t)-\phi(t-h)}{h}\zeta(t)dt \le \limsup_{h\to 0+}\int_0^1 \frac{f(t,t)-f(t-h,t)}{h}\zeta(t)dt +$$
$$+ \limsup_{h\to 0+}\int_0^1 \frac{f(t-h,t)-f(t-h,t-h)}{h}\zeta(t)dt =$$
$$= \limsup_{h\to 0+}\int_0^1 \frac{f(t,t)-f(t-h,t)}{h}\zeta(t)dt + \limsup_{h\to 0+}\int_0^1 \frac{f(t,t+h)-f(t,t)}{h}\zeta(t+h)dt.$$

We want to pass the lim sup inside the integrals; a classical extension of Fatou's lemma allows this if we know that the integrands are uniformly integrable. But

$$h^{-1}|f(t-h,t)-f(t-h,t-h)|\cdot|\zeta(t)| \le h^{-1}|w(t)-w(t-h)|\cdot C\cdot I_{\operatorname{supp}\zeta}(t)$$

which is uniformly integrable since $h^{-1}|w(t)-w(t-h)| \to |w'(t)|$ in $L^1_{\mathrm{loc}}((0,1))$: in fact the difference quotients of any $g \in AC((a,b))$ always converge in L^1, because they converge a.e. and are uniformly integrable.
(The latter can be seen noting that $h^{-1}[g(x)-g(x-h)] = (g' * h^{-1}I_{[0,h]})(x)$ and so $\int_E h^{-1}|g(x)-g(x-h)|dx \le \int_E |g'(x)|dx \int_0^h h^{-1}dx \to 0$ as $|E| \to 0$.)

The same argument applies to the second integrand. □

Proof of the Theorem. We can apply the lemma to $f(s,t) := d^2(u(s), v(t))$ on every $(a,b) \subset\subset (0,\infty)$. (EVI) tells us that for a.e. t

$$\lim_{h\to 0+}\frac{d^2(u(t),v(t))-d^2(u(t-h),v(t))}{h} \le -\lambda d^2(u(t),v(t)) + 2F(v(t)) - 2F(u(t)),$$
$$\lim_{h\to 0+}\frac{d^2(u(t),v(t+h))-d^2(u(t),v(t))}{h} \le -\lambda d^2(u(t),v(t)) + 2F(u(t)) - 2F(v(t)).$$

So, the lemma yields that $d^2(u(t),v(t))$ is locally absolutely continuous and satisfies $\frac{d}{dt}[d^2(u(t),v(t))] \le -2\lambda d^2(u(t),v(t))$, i.e. $\frac{d}{dt}[e^{2\lambda t}d^2(u(t),v(t))] \le 0$: this means that $[e^{\lambda t}d(u(t),v(t))]^2$ is nonincreasing, and so $e^{\lambda t}d(u(t),v(t))$ is nonincreasing as well, which gives the conclusion. □

The next theorem shows that an EVI-gradient flow u of F is such that $F \circ u$ decreases "as fast as possible". In a Hilbert setting and if F is λ-convex, this will allow to write the gradient flow "equation" (which in the definition is an inclusion) as a true equality, precisely the familiar $u' = -\nabla F(u)$. Finally, this classical result will suggest an analogous statement in the $\mathscr{P}_2(\mathbb{R}^n)$ setting.

Theorem 2.3.9. *Let $u : (0, \infty) \to X$ be an EVI(λ)-gradient flow of F, and suppose that F is lower semicontinuous. Then:*

1. *u and $F \circ u$ are locally Lipschitz and satisfy*
$$\frac{d}{dt}(F \circ u)(t) = -|u'|(t)^2 = -|\nabla F|(u(t))^2 \quad \text{for a.e. } t;$$
2. *If (X, d) is a Hilbert space and F is λ-convex, then $u'(t) = -(\nabla F)(u(t))$ for a.e. t (for the notation, see Example 2.2.3);*
3. *If $(X, d) = (\mathscr{P}_2(\mathbb{R}^n), W_2)$ and F is λ-convex, calling v_t the tangent velocity field to $u(t)$ and $\partial^o F(u(t))$ the element of minimal norm in $\partial^W F(u(t))$ (see Remark 2.2.16), then $v_t = -\partial^o F(u(t))$ for a.e. t.*

Proof.
$\boxed{1.}$ Since $t \mapsto u(t+h)$ is a gradient flow as well, then the contractivity estimates yield $d(u(t), u(t+h)) \le e^{-\lambda(t-s)} d(u(s), u(s+h))$ for every $t \ge s > 0$, and so $|u'|(t) \ge e^{-\lambda(t-s)} |u'|(s)$ for a.e. $t \ge s$ if s is a point of metric differentiability of u: hence $|u'| \in L^\infty((\varepsilon, M))$ whenever $0 < \varepsilon < M < \infty$, and so u is locally Lipschitz. For a.e. t, we have that for all $y \in X$
$$F(u(t)) - F(y) \le -\frac{1}{2}\frac{d}{dt} d^2(u(t), y) - \frac{\lambda}{2} d^2(u(t), y) \le d(u(t), y)|u'|(t) - \frac{\lambda}{2} d^2(u(t), y), \tag{2.3.2}$$
as is readily verified using the definition of $|u'|(t)$ as a limit. Dividing by $d(u(t), y)$ and letting $y \to u(t)$, we deduce that
$$|\nabla F|(u(t)) \le |u'|(t) \qquad \text{for a.e. } t. \tag{2.3.3}$$

If we put $y := u(s)$ in (2.3.2), instead, and we recall that u is locally Lipschitz, we get that for $s \in [\varepsilon, M]$ it holds
$$F(u(t)) - F(u(s)) \le d(u(t), u(s))|u'|(t) - \frac{\lambda}{2} d^2(u(t), u(s)) \le (L + C) d(u(t), u(s))$$
($d^2(u(t), u(s)) \le C$ for $t, s \in [\varepsilon, M]$ by continuity), and so
$$F(u(t)) - F(u(s)) \le \tilde{L}\, d(u(t), u(s)) \qquad \forall s \in [\varepsilon, M], \text{ for a.e. } t \in [\varepsilon, M].$$
By lower semicontinuity, the inequality is true for *all* $t \in [\varepsilon, M]$. Now the roles of s and t are symmetric, so we can exchange them to get also the inequality $F(u(s)) - F(u(t)) \le \tilde{L}\, d(u(t), u(s))$: the local Lipschitz property of $F \circ u$ is proven.

Let t be any point of differentiability for $F \circ u$ and of metric differentiability for u. We write (EVI) in integral form: for all $h > 0$ and $y \in X$, it holds
$$\int_t^{t+h} \left[F(u(r)) - F(y) + \frac{\lambda}{2} d^2(u(r), y) \right] dr \le \frac{1}{2} d^2(u(t), y) - \frac{1}{2} d^2(u(t+h), y).$$

For $y := u(t)$, we can write $F(u(r)) = F(u(t)) + (F \circ u)'(t)(r-t) + o((r-t)^2)$ and $d(u(r), u(t)) = (r-t)|u'|(t) + o((r-t)^2)$: the above inequality reduces to

$$\frac{h^2}{2}(F \circ u)'(t) + o(h^2) \leq -\frac{1}{2}|u'|(t)^2 h^2 + o(h^2)$$

from which $(F \circ u)'(t) \leq -|u'|(t)^2$.

But from the definitions $(F \circ u)'(t) \geq -|\nabla F|(u(t))|u'|(t)$ whenever the two members are defined: combining this with the inequality just proven, we deduce that $-|\nabla F|(u(t))|u'(t)| \leq -|u'|(t)^2$ and therefore $|\nabla F|(u(t)) \geq |u'|(t)$ for a.e. t. Since in (2.3.3) we showed the converse, we conclude that $|\nabla F|(u(t)) = |u'|(t)$ for a.e. t. In light of this, the proven inequalities

$$-|\nabla F|(u(t))|u'|(t) \leq (F \circ u)'(t) \leq -|u'|(t)^2 \qquad \text{for a.e. } t$$

yield the desired chain of equalities.

$\boxed{2.}$ For λ-convex functionals, we have the equivalent differential formulation $u'(t) \in -\partial F(u(t))$ for a.e. t. But we now know that $|u'(t)| = |\nabla F|(u(t))$ which is the least possible value for the norm of an element in the subdifferential, and so by definition of ∇F we have that $u'(t) = -\nabla F(u(t))$ for a.e. t.

$\boxed{3.}$ Exactly as above, we can use the equivalent differential formulation to get $v_t \in -\partial^W F(u(t))$ for a.e. t. But $\|v_t\|_{L^2(u(t))} = |u'|(t) = |\nabla F|(u(t))$ which is the least possible value for the norm of an element in the subdifferential: we conclude that $u'(t) = -\partial^o F(u(t))$ for a.e. t. □

Remark 2.3.10. The proof of point 3. shows incidentally, a.e along the trajectory of u, the cited general equality $|\nabla F| = \|\partial^o F\|_{L^2}$.

Here is the fundamental example of gradient flow in $\mathscr{P}_2(\mathbb{R}^n)$.

Theorem 2.3.11 (Heat flow as gradient flow of entropy)**.** *Let $\mu_0 \in \mathscr{P}_2(\mathbb{R}^n)$ be of the form $\rho_0 \mathscr{L}^n$. Then there exists an unique gradient flow starting from μ_0 of the entropy functional*

$$\mathrm{Ent}(\mu) := \begin{cases} \int_{\mathbb{R}^n} \rho(x) \log \rho(x) dx, & \text{if } \mu = \rho \cdot \mathscr{L}^n \\ +\infty, & \text{otherwise,} \end{cases}$$

and it is given by $\mu_t = u_t \mathscr{L}^n$, where u_t is the solution of the heat equation $\frac{d}{dt}u_t = \Delta u_t$ "starting" from ρ_0 (see the following Remark).

Remark 2.3.12 (Representation formula for the heat equation). We recall the elementary fact that for every $u_0 \in L^1(\mathbb{R}^n)$, there exists a unique (classical) solution in $(0, \infty)$ of the heat equation "starting" from u_0, i.e. such that u_t converges to u_0 in the sense of distributions as $t \downarrow 0$. Moreover , u_t is given by the convolution formula

$$u_t(x) = \frac{1}{(4\pi t)^{n/2}} \int_{\mathbb{R}^n} \rho_0(y) e^{-\frac{|x-y|^2}{4t}} dy \tag{2.3.4}$$

(as one might directly verify).

Proof of the Theorem. We apply the characterisation of the subdifferential of the internal energy (Theorem 2.2.32) with potential $U(z) = z \log z$: since $L_U(z) = z$, we get that μ_t is a gradient flow of the entropy functional if and only if $\mu_t = \rho_t \mathscr{L}^n$ where

1. $\rho_t \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^n)$ and $|\nabla \rho_t|^2 \rho_t \in L^1(\mathbb{R}^n)$ for a.e. t;
2. $\dfrac{d}{dt}(\rho_t \mathscr{L}^n) - \operatorname{div}\left[\dfrac{\nabla \rho_t}{\rho_t}(\rho_t \mathscr{L}^n)\right] = 0$ (in distributional sense).

(The second equation is very close to $\frac{d}{dt}\rho_t = \Delta \rho_t$.) Elementary results about the heat equation in $\mathbb{R}^n$ show that the solution $(u_t)_{t\geq 0}$ starting from ρ_0 is such that for every $t > 0$ it holds

$$u_t \geq 0, \qquad \int u_t(x)dx = 1, \quad u_t \in C^\infty(\mathbb{R}^n), \quad u_t, |\nabla u_t| \in L^\infty(\mathbb{R}^n):$$

these are all immediate consequences of the representation formula (2.3.4).

Therefore $\mu_t := u_t \mathscr{L}^n$ satisfies conditions 1. and 2. above: to show that it is a gradient flow of Ent starting from μ_0, it remains to prove that $(\mu_t)_{t\in(0,\infty)}$ is in $AC_{\mathrm{loc}}((0,\infty); \mathscr{P}_2(\mathbb{R}^n))$ and that $\mu_t \to \mu_0$ in $\mathscr{P}_2(\mathbb{R}^n)$ as $t \downarrow 0$.

The latter is true because for $t > 0$ the second moment of μ_t is

$$\int_{\mathbb{R}^n} |x|^2 u_t(x) \overset{\text{Fubini}}{=} \int_{\mathbb{R}^n} \rho_0(y) \frac{1}{(4\pi t)^{n/2}} \int_{\mathbb{R}^n} |x|^2 e^{-\frac{|x-y|^2}{4t}} dx\, dy = \int_{\mathbb{R}^n} \rho_0(y)(2nt+|y|^2)dy$$

(the second equality is evident if is seen as the calculation of the second moment of a Gaussian random variable). With the usual argument, we deduce from this that the family $(\mu_t)_{t>0}$ is tight, and therefore has narrow limit points for $t \downarrow 0$; but every narrow limit point is a distributional limit point, and therefore must coincide with $\rho_0 \mathscr{L}^n$. To sum up, $\mu_t \rightharpoonup \mu_0$ as $t \downarrow 0$ with converging second moments: i.e., $\mu_t \to \mu_0$ in $\mathscr{P}_2(\mathbb{R}^n)$.

Eventually, we have to prove the local absolute continuity of $t \mapsto \mu_t \in \mathscr{P}_2(\mathbb{R}^n)$. From the characterisation of $AC_2(\mathscr{P}_2(\mathbb{R}^n))$ (Theorem 2.1.16), it is sufficient to see that for every $(a,b) \subset\subset (0,\infty)$ it holds

$$\int_a^b \left\| \frac{\nabla u_t}{u_t} \right\|_{L^2(u_t \mathscr{L}^n)} dt < \infty.$$

But the left hand side is $\int_a^b \int_{\mathbb{R}^n} |\nabla u_t(x)|^2 u_t(x)\ dx\ dt$, which is finite since $|\nabla u_t(x)|$ is equibounded (for example, by the representation formula). $\square$

Remark 2.3.13. The identifiability between the heat flow and the gradient flow of entropy seems to be a very deep relationship: in Chapter 3 we will see that it holds, in a suitable sense, also in a discrete setting (Theorem 3.3.7). In a different direction, the above result is even true in generic metric measure spaces, perhaps the most general setting in which one may think about both gradient flows and entropy: this was shown in [4] ([5] is a more readable simplified account). The development of the theory needed to state and prove this result is far beyond the scope of this thesis; it shares little with the "differential" proof given above, which therefore retains its independent interest.

In a similar way, the solution of some other parabolic PDEs can be identified with a gradient flow in $\mathscr{P}_2(\mathbb{R}^n)$. Without any pretensions to rigorousness, we mention the two most classical examples, which motivated two fundamental papers in this field (respectively [11] and [14]).

Example 2.3.14 (Fokker-Planck)**.** We consider on $\mathscr{P}_2(\mathbb{R}^n)$ a functional of the form $E := \mathrm{Ent} + \mathcal{V}$, obtained adding to entropy a potential energy $\mathcal{V}(\mu) := \int V d\mu$. We know that (under suitable hypotheses) $\partial^W E(\rho \mathscr{L}^n) = \frac{\nabla\rho}{\rho} + \nabla V$: arguing as at the beginning of the above proof, we see that the gradient flow equation (under the identification of measures and densities) is formally equivalent to

$$\frac{d}{dt}\rho_t - \mathrm{div}(\nabla\rho_t + \rho_t V) = 0$$

which is the **Fokker-Planck equation**.

Example 2.3.15 (Porous medium)**.** We consider the internal energy with potential $U(z) := \frac{1}{m-1} z^m$, where $m \neq 1$ is a real number. Straightforward computations show that U is convex and satisfies McCann's condition for every $m \geq 1 - \frac{1}{n}$, $m \neq 1$ (note that for $m \to 1$, $U(z)$ tends to $z \log z$, which defines the entropy functional...). Therefore, we can apply the characterisation of the subdifferential of the the internal energy: since $L_U(z) = z^m$, the formal equivalent of the gradient flow equation becomes $\frac{d}{dt}\rho_t - \mathrm{div}\nabla((\rho_t)^m) = 0$, i.e.

$$\frac{d}{dt}\rho_t = \Delta[(\rho_t)^m]$$

which is the **porous medium equation**.

Chapter 3

A new distance on discrete spaces

Following [12], we define a new "transport" distance on a finite set endowed with a Markov irreducible kernel, in such a way that the gradient flow of the entropy is still identifiable with what in this context is called heat flow. The definition turns out to be a discrete analogous of Benamou-Brenier's formula. We will exploit this analogy to "translate" the continuous theory in this setting; we will try to stress the similarities, following a parallel order of exposition as far as this is possible.

3.1 The new distance

3.1.1 Setting and continuity equation

Our space X will be a finite set.

K will be an irreducible Markov kernel on X: given $x, y \in X$, $K(x, y)$ will be the probability of the transition from x to y.

By irreducibility, there is an unique probability measure on X invariant for K, which we will denote by π: this means that $\pi^T K = \pi^T$.

We will say that K is **reversible** if the **detailed balance** equations hold: $\pi(x)K(x, y) = \pi(y)K(y, x)$ for every $x, y \in X$. This means that, with starting distribution π, the mass going from x to y is equal to the mass going in the opposite direction.

Since by irreducibility $\pi(x) > 0$ for every $x \in X$, every measure on X has a density with respect to π. We will always identify the measure with its density: in particular we will think

$$\mathscr{P}(X) \equiv \left\{ \rho : X \to [0, \infty) : \sum_{x \in X} \rho(x)\pi(x) = 1 \right\}. \tag{3.1.1}$$

We need discrete versions of the concepts of gradient and divergence:

Definition 3.1.1. Given $\psi : X \to \mathbb{R}$, its **gradient** is the function $\nabla\psi : X \times X \to \mathbb{R}$ defined by $\nabla\psi(x, y) := \psi(x) - \psi(y)$.

For $\phi, \psi : X \to \mathbb{R}$ and $\Phi, \Psi : X \times X \to \mathbb{R}$, we will use the notations:

$$\langle \phi, \psi \rangle_\pi := \sum_{x \in X} \phi(x)\psi(x)\pi(x), \qquad \langle \Phi, \Psi \rangle_K := \frac{1}{2} \sum_{x,y \in X} \Phi(x,y)\Psi(x,y)K(x,y)\pi(x).$$

(Which, apart from the factor $\frac{1}{2}$ which is inserted for future convenience, are the natural "scalar products" on $\mathbb{R}^X$ and $\mathbb{R}^{X \times X}$, given the Markov structure.)

In the usual way, we define $\|\phi\|_\pi := \sqrt{\langle \phi, \phi \rangle_\pi}$ and $\|\Phi\|_K := \sqrt{\langle \Phi, \Phi \rangle_K}$.

Given $\Psi : X \times X \to \mathbb{R}$, its **divergence** $(\nabla \cdot \Psi) : X \to \mathbb{R}$ is defined in such a way that $\langle \nabla \psi, \Psi \rangle_K = -\langle \psi, \nabla \cdot \Psi \rangle_\pi$. Direct computations lead to

$$(\nabla \cdot \Psi)(x) = \frac{1}{2} \sum_{y \in X} \left[-K(x,y)\Psi(x,y) + K(y,x)\frac{\pi(y)}{\pi(x)}\Psi(y,x) \right]$$

which if K is reversible reduces to $\frac{1}{2} \sum_{y \in X} K(x,y)[\Psi(y,x) - \Psi(x,y)]$.

Notation. Given two matrices A, B with the same dimensions, $A \bullet B$ will denote the matrix obtained by componentwise multiplication: $(A \bullet B)_{ij} := A_{ij} B_{ij}$.

We are tempted to use this discrete divergence to define a formal analogous of the continuity equation. The role of the velocity field, in this context, will be played by a matrix $V_t \in \mathbb{R}^{X \times X}$, defined for $t \in [0,T]$. The naivest generalisation of (CE) would be $\dot{\rho}_t + \nabla \cdot ((\rho_t(x)V_t(x,y))_{x,y}) = 0$. However, we will see in the next paragraph that some more freedom is desirable: so, we substitute $(\rho_t(x)V_t(x,y))_{x,y}$ with $\hat{\rho}_t \bullet V_t$, where $\hat{\rho}_t \in \mathbb{R}^{X \times X}$ is a nonnegative symmetric matrix depending only on ρ_t; intuitively, $\hat{\rho}_t(x,y)$ may represent the amount of mass effectively affected by the velocity $V_t(x,y)$. Reasonable hypotheses on $\hat{\rho}$ are:

Hypothesis 3.1.2. For every $\rho \in \mathscr{P}(X)$, we will put $\hat{\rho}(x,y) = \theta(\rho(x), \rho(y))$, where $\theta : [0,\infty) \times [0,\infty) \to [0,\infty)$ is a fixed continuous function such that

1. $\theta(r,s) = \theta(s,r)$ (symmetry);
2. $\theta(r,s) \geq 0$ for every r, s, with equality if and only if either $r = 0$ or $s = 0$;
3. $\theta(r, \cdot)$ and $\theta(\cdot, s)$ are nondecreasing functions for every r, s.

In fact, when we will study the Riemannian structure of $\mathscr{P}(X)$, it will be useful to assume that θ is at least $C^1((0,\infty)^2)$.

Definition 3.1.3. Given $(\rho_t) \in AC([0,T]; \mathbb{R}^X)$ and $(V_t)_{t \in [0,T]} \subseteq \mathbb{R}^{X \times X}$ Borel, we say that $(\rho_t, V_t)_t$ solves the **continuity equation** if

$$\dot{\rho}_t + \nabla \cdot (\hat{\rho}_t \bullet V_t) = 0 \qquad \text{for a.e. } t. \tag{CE'}$$

Remark 3.1.4 (Time reparametrisation). Suppose that $(\rho_t, V_t)_{t \in [0,T]}$ solves (CE'), and let $\tau : [0,T'] \to [0,T]$ be an increasing AC reparametrisation. Then it is staightforward to see that $s \mapsto \rho_{\tau(s)}$ is AC, and that $(\rho_{\tau(s)}, V_{\tau(s)}\tau'(s))_{s \in [0,T']}$ solves (CE') as well.

Remark 3.1.5 (Total mass is preserved). The total mass $\sum_{x\in X}\pi(x)\rho_t(x)$ of a solution of the continuity equation has zero derivative (immediate verification), and hence is constant. However, in general, some of the components of the solution ρ_t might become negative.

Remark 3.1.6 (On existence and uniqueness of solutions). Given V_t and θ, the continuity equation for ρ_t is a system of ordinary differential equations. Thanks to "generalised Cauchy-Lipschitz theorem" (Theorem 2.1.6), given an initial condition ρ_0, we have local existence and uniqueness of the solution ρ_t at least if

$$\int_0^T |V_t|(x,y)dt < \infty \qquad \forall x, y \in X.$$

If the maximal solution is always nonnegative, then it is also bounded by the previous Remark, from which we deduce global existence in $[0,T]$. (For a case in which this is known a priori, see Remark 3.1.41 below.)

Finally, in the setting of reversible Markov chains, there is a natural concept of heat flow. The analytic way to justify its definition is to note the following:

Definition 3.1.7. Let K be a reversible Markov kernel. Given $f : X \to \mathbb{R}$, its **Laplacian** is the function $\Delta f : X \to \mathbb{R}$ defined by $\Delta f = \nabla\cdot(\nabla f)$. An immediate calculation gives that the operator Δ is simply left multiplication by the matrix $K - I$: with the usual identification of matrices and operators, we write $\Delta = K - I$.

We will say that a C^1 function $(f_t)_{t\in[0,T]}$ with values in $\mathbb{R}^X$ satisfies the **heat equation** if $\partial_t f_t = \Delta f_t$ for every t. The flow of this ordinary differential equation, namely

$$(t, f_0) \mapsto H(t)f_0 := e^{t\Delta} f_0,$$

will be called the **heat flow**; and $(H(t))_t$ will be called the **heat semigroup**.

Remark 3.1.8 (Probabilistic interpretation). There is also a probabilistic meaning of the name "heat flow". Recall that the **transition semigroup** of a continuous-time Markov kernel $(N_t(x,\cdot))_{t\geq 0, x\in Y} \subseteq \mathscr{P}(Y)$ on a measurable space Y is set of functions $(H(t))_t$ mapping every ϕ bounded and Borel to $(H(t)\phi)(x) := \int_Y \phi(y)N_t(x,dy)$. The transition semigroup of the n-dimensional Brownian kernel is therefore

$$(H(t)\phi)(x) = \frac{1}{(2\pi t)^{n/2}}\int_{\mathbb{R}^n} \phi(y)e^{-\frac{|y-x|^2}{2t}}dy$$

which is the solution of $\partial_t f = \frac{\Delta}{2} f$ with initial condition ϕ. In conclusion, the transition semigroup of the n-dimensional Brownian kernel (up to a rescaling of time) is the heat semigroup on $\mathbb{R}^n$.

The natural concept of "random walk" in our discrete setting is the continuous time Markov process associated to the discrete Markov kernel K: that is to say, the piecewise constant Markov process which, starting from $x \in X$, waits a random exponentially distributed time, and then jumps to y with probability $K(x,y)$. A classical result says that its transition semigroup is given by $H(t) = e^{(K-I)t}$, from which the definition above is coherent with this interpretation of the heat semigroup. The book [7] is entirely devoted to this topic.

3.1.2 The 2-point space

Note. Unlike in most of the other paragraphs of this thesis, here we prefer to introduce the hypotheses one by one where we need them, and to perform the computations *before* the statement of the result. Perhaps this makes the exposition less ordered; but we desire to show a sensible way in which the results can be *guessed*. Otherwise, the definition of the new distance, and the fact that "it works so well", would look incomprehensible and rather magical. However, the relevant hypotheses and the main results are summarized at the end, in Theorem 3.1.14.

We consider the simplest case: $X = \{a, b\}$. Call $K(a,b) = p, K(b,a) = q$: by irreducibility $p, q > 0$. It is immediate to verify that the invariant probability is given by $\pi(a) = \frac{q}{p+q}, \pi(b) = \frac{p}{p+q}$. Obviously, the Markov kernel is reversible.

The probability measures on X are of the form $\frac{1}{2}(1-\beta)\delta_a + \frac{1}{2}(1+\beta)\delta_b$ where $\beta \in [-1,1]$ is a convenient parameter. The corresponding density ρ^β is

$$\rho^\beta(a) = \frac{p+q}{q}\frac{1-\beta}{2}, \qquad \rho^\beta(b) = \frac{p+q}{p}\frac{1+\beta}{2}. \tag{3.1.2}$$

As for the heat flow, a straightforward calculation gives that

$$H(t) = \frac{1}{p+q}\left(\begin{bmatrix} q & p \\ q & p \end{bmatrix} + e^{-(p+q)t}\begin{bmatrix} p & -p \\ -q & q \end{bmatrix}\right).$$

In terms of the parameter β, a direct computation shows that $H(t)\rho^\beta := \rho^{\beta_t}$ where

$$\beta_t = \frac{p-q}{p+q}(1 - e^{-(p+q)t}) + \beta e^{-(p+q)t}$$

which is the solution of the Cauchy problem

$$\begin{cases} \dot\beta_t = p(1-\beta_t) - q(1+\beta_t) \\ \beta_0 = \beta. \end{cases} \tag{3.1.3}$$

Remark 3.1.9 (Inadequacy of W_2). If our aim is to study absolutely continuous curves and gradient flows in $\mathscr{P}(X)$, then W_2 is not the right distance. In fact $W_2(\rho^\alpha, \rho^\beta) = d(a,b)\sqrt{|\beta - \alpha|}$, and a classical exercise tells us that in $\mathbb{R}$ with the distance $\sqrt{|x-y|}$, the only AC curves are constant.

We wonder whether there exists a distance on $\mathscr{P}(X)$ which, as in the continuous case, gives to $\mathscr{P}(X)$ a "Riemannian" structure such that the heat flow at the level of densities is the gradient flow at the level of measures of the entropy (relative to π)

$$\mathcal{H}(\rho) = \sum_{x\in X} \rho(x)\log(\rho(x))\pi(x) =: \sum_{x\in X} U(\rho(x))\pi(x)$$

where we defined for brevity $U(s) := s\log s$.

We expect $\mathscr{P}(X)$ to have dimension 1, because it is parametrised by $\beta \in [-1,1]$. Therefore, we proceed heuristically and look for a distance given by an isometry $J : \rho^\beta \mapsto \phi(\beta) \in \mathbb{R}$. Any metric concept of gradient flow is obviously invariant by isometry; but on $\mathbb{R}$ and for C^1 functionals, any formulation is implied by the classical differential one. So, it is sufficient to find ϕ such that $\frac{d}{dt}\phi(\beta_t) = -\tilde{\mathcal{H}}'(\phi(\beta_t))$ where $\tilde{\mathcal{H}} := \mathcal{H} \circ J^{-1}$.

Now it is a matter of calculus. First of all

$$\tilde{\mathcal{H}}(\phi(\alpha)) = \mathcal{H}(\rho^\alpha) = \frac{q}{p+q}U(\rho^\alpha(a)) + \frac{p}{p+q}U(\rho^\alpha(b))$$

and if ϕ is differentiable we deduce

$$\tilde{\mathcal{H}}'(\phi(\alpha))\phi'(\alpha) = \frac{q}{p+q}U'(\rho^\alpha(a))\frac{d}{d\alpha}\rho^\alpha(a) + \frac{p}{p+q}U'(\rho^\alpha(b))\frac{d}{d\alpha}\rho^\alpha(b) = \frac{U'(\rho^\alpha(b)) - U'(\rho^\alpha(a))}{2}.$$

If $\phi' > 0$, we can divide by $\phi'(\alpha)$, so that what we want is that that the flow of the equation

$$\frac{d}{dt}\phi(\beta_t) = \frac{-U'(\rho^{\beta_t}(b)) + U'(\rho^{\beta_t}(a))}{2\phi'(\beta_t)}$$

is the heat flow. Rewriting the equation in the the form

$$\dot{\beta}_t = \frac{U'(\rho^{\beta_t}(a)) - U'(\rho^{\beta_t}(b))}{2\phi'(\beta_t)^2}$$

and recalling (3.1.3), it is sufficient to impose

$$\frac{U'(\rho^\beta(a)) - U'(\rho^\beta(b))}{2\phi'(\beta)^2} = p(1-\beta) - q(1+\beta)$$

or equivalently

$$\phi'(\beta)^2 = \frac{U'(\rho^\beta(a)) - U'(\rho^\beta(b))}{2[p(1-\beta) - q(1+\beta)]} = \frac{p+q}{4pq}\frac{U'(\rho^\beta(a)) - U'(\rho^\beta(b))}{\rho^\beta(a) - \rho^\beta(b)}.$$

To be precise, we have a problem if $\rho^\beta(a) = \rho^\beta(b)$, which is not a serious issue since this happens only for one value $\bar{\beta}$, where RHS can be extended by continuity with value $U''(\rho^{\bar{\beta}}(a))$. With this convention, and noting that $U'' > 0$, we have that the right hand side is strictly positive, so we can conclude that

$$\phi(\beta) := \frac{1}{2}\sqrt{\frac{1}{p} + \frac{1}{q}}\int_0^\beta \sqrt{\frac{U'(\rho^r(a)) - U'(\rho^r(b))}{\rho^r(a) - \rho^r(b)}}\,dr =: \frac{1}{2}\sqrt{\frac{1}{p} + \frac{1}{q}}\int_0^\beta \frac{1}{\sqrt{h(r)}}dr,$$

if finite, is a solution to our problem. For $\beta \in (-1, 1)$, finiteness is obvious; if we allow β to take the values ± 1, then ϕ may take the values $\pm\infty$. (In fact this will not be the case for $U(x) = x \log x$, but in a moment we will wish to choose U differently.)

Hence, with this definition of ϕ, and for $\alpha, \beta \in [-1, 1]$, we define the extended distance ("extended" because it can take the value ∞):

$$\mathcal{W}(\rho^\alpha, \rho^\beta) := |\phi(\alpha) - \phi(\beta)| = \frac{1}{2}\sqrt{\frac{1}{p} + \frac{1}{q}}\left|\int_\alpha^\beta \frac{1}{\sqrt{h(r)}}dr\right|, \tag{3.1.4}$$

$$h(r) := \frac{\rho^r(a) - \rho^r(b)}{U'(\rho^r(a)) - U'(\rho^r(b))}. \tag{3.1.5}$$

In fact, definition (3.1.4) makes sense for every function $h : (-1, 1) \to (0, \infty)$, no matter whether it is defined by the relation (3.1.5) or not.

If we denote by $\mathscr{P}_*(X) := \{\rho^\beta : \beta \in (-1, 1)\}$ the space of (probability measures with) strictly positive densities, we have just proven:

Proposition 3.1.10. *$(\mathscr{P}_*(X), \mathcal{W})$ is a one-dimensional Riemannian manifold on which the gradient flow of entropy is the heat flow.*

Remark 3.1.11 (More general functionals). We never used that $U(x) = x \log x$: everything (even in the sequel) works replacing $\mathcal{H}$ with an "internal energy" functional

$$\mathcal{U}(\rho) := \sum_{x \in X} U(\rho(x))\pi(x)$$

where $U \in C^2((0, \infty))$ is such that $U'' > 0$.

We can obtain a different characterisation of $\mathcal{W}$ using the following classical observation:

Lemma 3.1.12. *Let $h : (-1, 1) \to (0, \infty)$ be continuous. Then for $-1 \leq \alpha \leq \beta \leq 1$, it holds*

$$\left(\int_\alpha^\beta \frac{1}{\sqrt{h(r)}} dr\right)^2 = \min_{\gamma(0)=\alpha, \gamma(1)=\beta} \int_0^1 \frac{\dot{\gamma}(t)^2}{h(\gamma(t))} dt \qquad \textit{(possibly infinite)}$$

where γ runs over all AC functions $[0, 1] \to [-1, 1]$; the minimum is attained by a C^1 function.

Proof. Firstly, by Jensen's inequality, for every admissible γ

$$\int_0^1 \frac{\dot{\gamma}(t)^2}{h(\gamma(t))} dt \geq \left(\int_0^1 \frac{\dot{\gamma}(t)}{\sqrt{h(\gamma(t))}} dt\right)^2$$

which with a change of variables gives one inequality of our thesis.

The converse inequality is nontrivial only if $\int_\alpha^\beta \frac{1}{\sqrt{h(r)}} dr < \infty$. Take now any γ admissible, C^1 and strictly increasing (for instance, the constant speed parametrisation of the segment $[\alpha, \beta]$). For $t \in [0, 1]$, put

$$\sigma(t) := \frac{\int_0^t \frac{\dot{\gamma}(r)}{\sqrt{h(\gamma(r))}} dr}{\int_0^1 \frac{\dot{\gamma}(r)}{\sqrt{h(\gamma(r))}} dr}$$

where the denominator is equal to $\int_\alpha^\beta \frac{1}{\sqrt{h(r)}} dr$, which is finite by hypothesis. Observe that $\sigma \in C^1((0, 1)) \cap C([0, 1])$ has strictly positive derivative in $(0, 1)$, hence it has a $C^1((0, 1)) \cap C([0, 1])$ inverse τ satisfying

$$\dot{\tau}(s) = \left(\int_\alpha^\beta \frac{1}{\sqrt{h(r)}} dr\right) \left(\frac{\dot{\gamma}(\tau(s))}{\sqrt{h(\gamma \circ \tau(s))}}\right)^{-1}.$$

To conclude, we define $\tilde{\gamma} := \gamma \circ \tau$ and note that

$$\int_0^1 \frac{\dot{\tilde{\gamma}}(s)^2}{h(\tilde{\gamma}(s))} ds = \int_0^1 \frac{\dot{\gamma}(\tau(s))^2 \dot{\tau}(s)^2}{h(\gamma \circ \tau(s))} dt = \left(\int_\alpha^\beta \frac{1}{\sqrt{h(r)}} dr \right)^2$$

as desired.

$\square$

Remark 3.1.13 (Choice of the velocity field). In the two-point space, the continuity equation reduces to

$$\begin{cases} \dot{\rho}_t(a) = \frac{1}{2} \hat{\rho}_t(a,b)\ p\ [V_t(b,a) - V_t(a,b)] \\ \dot{\rho}_t(b) = \frac{1}{2} \hat{\rho}_t(a,b)\ q\ [V_t(a,b) - V_t(b,a)]. \end{cases}$$

Given any AC curve $(\rho_t) \subseteq \mathscr{P}(X)$, it is immediate to find a V_t satisfying these equations. First of all, the value of V_t on $\{t : \rho_t \notin \mathscr{P}_*(X)\}$ is irrelevant, because $\dot{\rho}_t = 0$ for almost every t in this set. Elsewhere, it is necessary and sufficient that

$$V_t(b,a) - V_t(a,b) = \frac{2\dot{\rho}_t(a)}{\hat{\rho}_t(a,b)\ p} \tag{3.1.6}$$

because then the other equation is automatically satisfied, since

$$\pi(a)\rho_t(a) + \pi(b)\rho_t(b) = 1.$$

We recall that in the continuous case the "right" choice of the velocity field was (almost) a gradient. Here there is no difficulty to find a suitable V_t of the form $V_t = \nabla\psi_t$: in fact the only restriction on V_t is (3.1.6), which becomes

$$\psi_t(b) - \psi_t(a) = \frac{\dot{\rho}_t(a)}{\hat{\rho}_t(a,b)\ p}. \tag{3.1.7}$$

In light of the Lemma, we have

$$\mathcal{W}(\rho^\alpha, \rho^\beta)^2 = \min_{\gamma(0)=\alpha, \gamma(1)=\beta} \frac{p+q}{4pq} \int_0^1 \frac{\dot{\gamma}(t)^2}{h(\gamma(t))} dt.$$

Putting $\rho_t := \rho^{\gamma(t)}$, this can be rewritten in terms of the two quantities

$$\dot{\rho}_t(a) = -\frac{p+q}{2q}\dot{\gamma}(t), \qquad \dot{\rho}_t(b) = \frac{p+q}{2p}\dot{\gamma}(t).$$

which in turn can be expressed using the continuity equation. Using both the identities to give the same relevance both to a and to b, we get

$$\begin{aligned} \mathcal{W}(\rho^\alpha, \rho^\beta)^2 &= \min_{\rho_0=\rho^\alpha, \rho_1=\rho^\beta} \frac{1}{2} \int_0^1 \frac{1}{h(\gamma(t))} \left\{ \frac{q}{p(p+q)} [\dot{\rho}_t(a)]^2 + \frac{p}{q(p+q)} [\dot{\rho}_t(b)]^2 \right\} dt = \\ &= \min_{\rho_0=\rho^\alpha, \rho_1=\rho^\beta} \frac{1}{2} \int_0^1 \left\{ \frac{\hat{\rho}_t(a,b)^2}{h(\gamma(t))} [\nabla\psi_t(a,b)]^2 K(a,b)\pi(a) + \right. \\ &\qquad\qquad \left. + \frac{\hat{\rho}_t(a,b)^2}{h(\gamma(t))} [\nabla\psi_t(a,b)]^2 K(b,a)\pi(b) \right\} dt. \end{aligned}$$

(Recall that by (3.1.7), (ρ_t) determines $\nabla\psi_t(a,b)$ for the values of t where $\hat{\rho}_t(a,b) \neq 0$, which are the ones relevant to the value of the integral above.)

The computation above works for every h, and whatever θ we are using in (CE'). The result becomes simpler if we require that $\hat{\rho}_t(a,b) = h(\gamma(t))$. There are two conceptually different situations in which this can be done:

1. If we have h of the form (3.1.5) (because we are trying to represent the heat flow, and so we are in the setting of Proposition 3.1.10), then we impose

$$\theta(r,s) := \frac{s-r}{U'(s)-U'(r)} \qquad \forall r,s>0, s\neq r$$

   which can be extended to $s = r$ defining $\theta(s,s) = \frac{1}{U''(s)}$. We complete the definition of θ with the condition $\theta(\cdot,0) = \theta(0,\cdot) = 0$; for general U this θ is not continuous, in the sequel we assume that it is. By direct verification, the standard case $U(x) = x\log x$ satisfies this assumption (see Example 3.1.15 below).

2. Conversely, if we are given θ, we can realize the condition $\hat{\rho}_t(a,b) = h(\gamma(t))$ simply choosing $h(r) := \theta(\rho^r(a), \rho^r(b))$.

In both cases, the final result reads as follows:

$$\begin{aligned}\mathcal{W}(\rho_0,\rho_1)^2 = \min_{(\rho_t,\nabla\psi_t)_t \text{ solves (CE')}} \int_0^1 \frac{1}{2}\big\{&[\nabla\psi_t(a,b)]^2\hat{\rho}_t(a,b)K(a,b)\pi(a)+\\ &+[\nabla\psi_t(b,a)]^2\hat{\rho}_t(b,a)K(b,a)\pi(b)\big\}\,dt. \quad (3.1.8)\end{aligned}$$

We recognize in the integrand an expression very similar to $\|\nabla\psi_t\|_K^2$, except for the presence of the weight $\hat{\rho}_t$. Hence, even if X has more than two points, we introduce the notation:

Notation. For $\Phi,\Psi : X\times X \to \mathbb{R}$ and $\rho \in \mathscr{P}(X)$, we define

$$\langle\Phi,\Psi\rangle_{\hat{\rho}} := \frac{1}{2}\sum_{x,y\in X}\Phi(x,y)\Psi(x,y)\hat{\rho}(x,y)K(x,y)\pi(x), \qquad \|\Phi\|_{\hat{\rho}} := \sqrt{\langle\Phi,\Phi\rangle_{\hat{\rho}}}.$$

Now we can write the result (3.1.8) in a very suggestive way, formally similar to Benamou-Brenier's formula (BB) (rewritten using the fact that the optimal velocity field is in the closure of gradients):

$$\mathcal{W}(\rho_0,\rho_1)^2 := \inf\left\{\int_0^1 \|\nabla\psi_t\|_{\hat{\rho}_t}^2 : (\rho_t,\nabla\psi_t)_t \text{ solves (CE')}\right\}.$$

To avoid confusion, we summarize the main achievements of this paragraph.

Theorem 3.1.14 (Summary)**.** *Suppose that X has only two points a,b, and put $p := K(a,b)$, $q := K(b,a)$. Call ρ^β the density of $\frac{1}{2}(1-\beta)\delta_a + \frac{1}{2}(1-\beta)\delta_b$ with respect to π. Then:*

1. *For θ as in Hypothesis 3.1.2, the two following formulas define the same extended distance on $\mathscr{P}(X)$:*

$$\mathcal{W}(\rho^\alpha,\rho^\beta) := \frac{1}{2}\sqrt{\frac{1}{p}+\frac{1}{q}}\left|\int_\alpha^\beta \frac{1}{\sqrt{\theta(\rho^r(a),\rho^r(b))}}dr\right|,$$
$$\mathcal{W}(\rho_0,\rho_1)^2 := \inf\left\{\int_0^1 \|\nabla\psi_t\|^2_{\hat\rho_t} : (\rho_t,\nabla\psi_t)_t \text{ solves (CE')}\right\},$$

and the infimum is realised by a C^1 curve (ρ_t).

In particular, from the first equation, $\mathcal{W}$ is a true distance at least on the space $\mathscr{P}_*(X) := \{\rho^\beta : \beta \in (-1,1)\}$, and $(\mathscr{P}_*(X),\mathcal{W})$ is isometric to an open interval of $\mathbb{R}$.

2. *Let $U \in C^2((0,\infty))$ be such that $U'' > 0$, and put*

$$\theta(r,s) := \begin{cases} 0, & \text{if } rs = 0\\ \dfrac{s-r}{U'(s)-U'(r)}, & \text{if } r,s>0, s\neq r\\ [U''(s)]^{-1}, & \text{if } r = s. \end{cases}$$

Suppose that this θ is continuous also in the boundary points of $[0,\infty)^2$, so that the first part of the theorem applies. Then the gradient flow starting from ρ_0 of the functional $\mathcal{U}(\rho) := U(\rho(a))\pi(a)+U(\rho(b))\pi(b)$ on the space $(\mathscr{P}_*(X),\mathcal{W})$ is given by the heat flow $\rho_t = e^{(K-I)t}\rho_0$.

This part applies to $U(x) := x\log x$, in which case $\mathcal{U}$ coincides with the entropy functional $\mathcal{H}$.

Notation. Sometimes we will emphasise the dependence of the distance $\mathcal{W}$ on the parameters p,q denoting it with the symbol $\mathcal{W}_{p,q}$.

Example 3.1.15 (Classical case)**.** If $f(x) = x\log x$, then a calculation gives that $\theta(r,s) = \int_0^1 s^{1-t}r^t\,dt$, which is the so-called **logarithmic mean**. If in addition $p = q$, then the explicit definition of $\mathcal{W}$ reduces to

$$\mathcal{W}(\rho^\alpha,\rho^\beta) := \frac{1}{\sqrt{2p}}\left|\int_\alpha^\beta \sqrt{\frac{\operatorname{arctanh} r}{r}}dr\right|. \tag{3.1.9}$$

By elementary calculus, one can now verify that $\mathcal{W}(\rho^\alpha,\rho^\beta)$ is finite even when α,β take the values ±1.

3.1.3 The n-point space

The second definition of $\mathcal{W}$ makes sense even if $|X| > 2$:

Definition 3.1.16. For $\rho_0,\rho_1 \in \mathbb{R}^X$ (and in particular if $\rho_0,\rho_1 \in \mathscr{P}(X)$), we put

$$\mathcal{W}(\rho_0,\rho_1) := \sqrt{\inf\left\{\int_0^1 \|\nabla\psi_t\|^2_{\hat\rho_t} : (\rho_t,\nabla\psi_t)_{t\in[0,1]} \text{ solves (CE')}\right\}}.$$

Remark 3.1.17 (Reduction to K reversible). By direct inspection, one sees that both (CE') and $\|\nabla\psi_t\|^2_{\hat\rho_t}$ are the same if we substitute K with the irreducible Markov kernel $\tilde K$ defined by $\tilde K(x,y) := \frac{1}{2}\left(K(x,y) + \frac{\pi(y)}{\pi(x)}K(y,x)\right)$. The invariant π is the same, but $\tilde K$ is reversible: so, in order to study the properties of $\mathcal{W}$, the following assumption is not really restrictive:

Hypothesis 3.1.18. From now on, K will be a *reversible* irreducible Markov kernel.

Remark 3.1.19 (Associated graph). A reversible Markov kernel can be naturally associated to a weighted unoriented graph on X: it is sufficient to draw the edge between x and y if and only if $K(x,y) > 0$, and to give it the weight $\pi(x)K(x,y)$. Note that the sum of the weights of the edges touching x (i.e. the "weight of x") is $\pi(x)$.

Conversely, given an unoriented weighted graph such that the sum of the weights of its points is equal to 1, we can recover a (unique) Markov kernel inducing this graph: we have just seen that $\pi(x)$ is determined, hence $K(x,y)$ is determined too by

$$K(x,y) = \frac{(\text{weight of the edge } x \leftrightarrow y)}{\pi(x)}.$$

Definition 3.1.20. For $x, y \in X$, $d_g(x,y)$ will be the **graph distance** between x and y on the graph associated to K, i.e. the minimum number of edges of a path connecting x to y.

Theorem 3.1.21. *$\mathcal{W}$ is an extended distance on $\mathscr{P}(X)$.*

Symmetry is obvious. The following lemma immediately implies the triangle inequality:

Lemma 3.1.22. *For every $\eta, \sigma \in \mathscr{P}(X)$ and $T > 0$, it holds*

$$\mathcal{W}(\rho,\sigma) = \inf\left\{\int_0^T \|\nabla\psi_t\|_{\hat\rho_t} : (\rho_t, \nabla\psi_t)_{t\in[0,T]} \text{ solves (CE')}, \rho_0 = \eta, \rho_1 = \sigma\right\}.$$

Proof. The infimum in the thesis is evidently invariant by time rescaling, hence we can suppose $T = 1$. Then $\geq$ is an application of Jensen's inequality.

The converse is nontrivial only if the infimum on the right hand side if a finite number C. In this case, fix $\varepsilon > 0$ and pick $(\rho_t, \nabla\psi_t)_{t\in[0,1]}$ solving (CE') and such that

$$\int_0^1 \|\nabla\psi_t\|_{\hat\rho_t} < C + \varepsilon\,.$$

We know that (CE') is preserved by AC strictly increasing reparametrisations (Remark 3.1.4). The idea is to use a reparametrisation τ such that $\tau'(s)\left\|\nabla\psi_{\tau(s)}\right\|_{\hat\rho_{\tau(s)}}$ becomes "approximately constant", so to have "almost equality" in Jensen's inequality. Precisely, for $s \in [0,1]$ put

$$\sigma(s) := \frac{1}{c_\varepsilon}\int_0^s \sqrt{\|\nabla\psi_r\|^2_{\hat\rho_r} + \varepsilon^2}dr$$

where c_ε is chosen to have $\sigma_\varepsilon(1) = 1$, i.e.

$$c_\varepsilon := \int_0^1 \sqrt{\|\nabla\psi_r\|_{\hat\rho_r}^2 + \varepsilon^2} dr \le \int_0^1 \left(\|\nabla\psi_r\|_{\hat\rho_r} + \varepsilon\right) dr < C + 2\,\varepsilon\,.$$

σ is AC with $\sigma' > \varepsilon$, so it has an AC strictly increasing inverse τ such that for a.e. $s \in [0,1]$

$$\tau'(s) = c_\varepsilon \left(\sqrt{\left\|\nabla\psi_{\tau(s)}\right\|_{\hat\rho_{\tau(s)}}^2 + \varepsilon^2}\right)^{-1}.$$

To conclude, put $\tilde\rho_s := \rho_{\tau(s)}$ and $\tilde\psi_s := \psi_{\tau(s)}\tau'(s)$: it is an admissible couple in the infimum defining $\mathcal{W}$, therefore

$$\mathcal{W}(\eta,\sigma)^2 \le \int_0^1 \left\|\nabla\tilde\psi_s\right\|_{\widehat{\tilde\rho}_s}^2 ds = \int_0^1 \left\|\nabla\psi_{\tau(s)}\right\|_{\hat\rho_{\tau(s)}}^2 \tau'(s)^2 ds =$$

$$= c_\varepsilon^2 \int_0^1 \frac{\left\|\nabla\psi_{\tau(s)}\right\|_{\hat\rho_{\tau(s)}}^2}{\left\|\nabla\psi_{\tau(s)}\right\|_{\hat\rho_{\tau(s)}}^2 + \varepsilon^2} ds \le c_\varepsilon^2 \le (C + 2\,\varepsilon)^2.$$

Letting $\varepsilon \to 0$, we get the desired inequality. □

To conclude that $\mathcal{W}$ is an extended distance, we need only that $\mathcal{W}(\rho,\sigma) > 0$ for every $\rho \ne \sigma$. We can actually say more:

Definition 3.1.23. For $\rho,\sigma \in \mathscr{P}(X)$, the **total variation distance** between ρ and σ is

$$d_{TV}(\rho,\sigma) := \sum_{x\in X} |\rho(x) - \sigma(x)|\pi(x) = \|\rho - \sigma\|_{L^1(\pi)}\,.$$

Notation. $\|\theta\|'_\infty := \sup\left\{\theta(s,t) : 0 \le s,t \le \left(\min_{x\in X}\pi(x)\right)^{-1}\right\}.$

Lemma 3.1.24 (Lower bound)**.** *Let $\rho_0,\rho_1 \in \mathscr{P}(X)$. Then*

$$\frac{1}{\sqrt{2}} d_{TV}(\rho_0,\rho_1) \le \sqrt{2} W_1(\rho_0,\rho_1) \le \sqrt{\|\theta\|'_\infty}\,\mathcal{W}(\rho_0,\rho_1)$$

where W_1 is the Wasserstein distance with respect to d_g.

Proof. The first inequality follows observing that $\frac{1}{2}d_{TV} = W_1$ when the distance on X is $\tilde d(x,y) := 1$ for every $x \ne y$; but $\tilde d \le d_g$ and the inequality is proven.

To prove the remaining inequality, we can suppose that $\mathcal{W}(\rho_0,\rho_1) < \infty$. We fix $\varepsilon > 0$ and take a $(\rho_t, \nabla\psi_t)_{t\in[0,1]}$ solving (CE') and such that

$$\int_0^1 \|\nabla\psi_t\|_{\hat\rho_t}^2\, dt < \mathcal{W}(\rho_0,\rho_1)^2 + \varepsilon\,.$$

We want to use Kantorovich's duality: so, for any ϕ 1-Lipschitz for d_g, we compute

$$|\langle\phi, \rho_0 - \rho_1\rangle_\pi| = \left|\int_0^1 \langle\phi, \dot\rho_t\rangle_\pi dt\right| \overset{\text{(CE')}}{=} \left|\int_0^1 \langle\nabla\phi, \hat\rho_t \bullet \nabla\psi_t\rangle_K dt\right|. \tag{3.1.10}$$

If we apply Cauchy-Schwarz's inequality to the symmetric bilinear positive semidefinite form

$$(f,g) \mapsto \int_0^1 \langle \nabla f_t, \hat{\rho}_t \bullet \nabla g_t \rangle_K dt$$

we infer that

$$(3.1.10) \leq \left(\int_0^1 \langle \nabla \phi, \hat{\rho}_t \bullet \nabla \phi \rangle_K dt \right)^{1/2} \left(\int_0^1 \langle \nabla \psi_t, \hat{\rho}_t \bullet \nabla \psi_t \rangle_K dt \right)^{1/2}.$$

The second integral is less or equal than $\mathcal{W}(\rho_0,\rho_1)^2+\varepsilon$; it remains to estimate the first integral, which is

$$\int_0^1 \frac{1}{2} \sum_{x,y\in X} (\phi(x)-\phi(y))^2 \hat{\rho}_t(x,y) K(x,y)\pi(x)\ dt. \tag{3.1.11}$$

But in (3.1.11), the only nonzero addends are such that $K(x,y)\neq 0$ and $x\neq y$, so that $d_g(x,y)=1$. Hence for ϕ 1-Lipschitz

$$(3.1.11) \leq \frac{1}{2} \int_0^1 \sum_{x,y\in X} \theta(\rho_t(x),\rho_t(y)) K(x,y)\pi(x) \leq \frac{1}{2} \|\theta\|'_\infty . \tag{3.1.12}$$

Letting $\varepsilon \to 0$ the conclusion follows.

□

Remark 3.1.25 (Simpler bound). If we are not interested in the bound involving W_1 (for example, because we only want to prove that $\mathcal{W}(\eta,\sigma)>0$ if $\eta\neq\sigma$), then we can avoid using Kantorovich's duality (which may be desirable, because the theory exposed in this chapter is independent from optimal transport, except for inspiration): in fact, the computations leading to (3.1.11) are valid even if we only know that $\phi \in L^\infty(X)$, but then

$$(3.1.11) \leq 2 \|\phi\|^2_{L^\infty(X)} \|\theta\|'_\infty \sum_{x,y\in X} K(x,y)\pi(x) = 2 \|\phi\|^2_{L^\infty(X)} \|\theta\|'_\infty .$$

Letting $\varepsilon \to 0$, we have shown that

$$|\langle \phi, \rho_0-\rho_1\rangle_\pi| \leq \sqrt{2\|\theta\|'_\infty} \mathcal{W}(\rho_0,\rho_1) \|\phi\|_{L^\infty(X)}$$

which by duality of L^1 and L^∞ easily yields the bound involving d_{TV}.

Remark 3.1.26 (Sharper bound). Suppose that θ is concave and 1-homogeneous. Then

$$\theta(s,t) = \frac{1}{2}(\theta(s,t)+\theta(t,s)) \leq \theta\left(\frac{s+t}{2},\frac{s+t}{2}\right) = \frac{s+t}{2}\theta(1,1).$$

Inserting this inequality in (3.1.12) (instead of using simply $\theta(\dots)\leq\|\theta\|'_\infty$), one obtains

$$\sqrt{2}W_1(\rho_0,\rho_1) \leq \sqrt{\theta(1,1)}\mathcal{W}(\rho_0,\rho_1).$$

which is slightly sharper, and independent of K.

This applies if θ is the logarithmic mean, for which $\theta(1,1)=1$.

As in the continuous case, we can drop the request that the velocity field is a gradient:

Proposition 3.1.27 (General velocity fields)**.** *For every $\rho_0, \rho_1 \in \mathscr{P}(X)$, it holds*

$$\mathcal{W}(\rho_0, \rho_1) = \sqrt{\inf \left\{ \int_0^1 \|V_t\|^2_{\hat{\rho}_t} : (\rho_t, V_t)_{t\in[0,1]} \textit{ solves } (\text{CE'}) \right\}}. \tag{3.1.13}$$

Proof. If $(\rho_t, V_t)_t$ is a solution of (CE') and W_t is another velocity field, then $(\rho_t, W_t)_t$ solves (CE') as well if and only if $\nabla \cdot ((V_t - W_t) \bullet \hat{\rho}_t) = 0$. In some sense this is still equivalent to "orthogonality to the gradients", since

$$\langle \Psi, \nabla\phi \rangle_{\hat{\rho}} = -\langle \nabla \cdot (\Psi \bullet \hat{\rho}), \phi \rangle_\pi. \tag{3.1.14}$$

To make the argument precise, we note that $\|V_t\|^2_{\hat{\rho}_t}$ and (CE') depend only on the values $V_t(x, y)$ in the couples (x, y) such that $K(x, y)\hat{\rho}_t(x, y) > 0$. Hence the only thing that matters is the equivalence class of V_t in the space

$$H_{\rho_t} := \frac{\mathbb{R}^{X\times X}}{\sim}$$

where $\Phi \sim \Psi$ if they coincide on $\{(x, y) : K(x, y)\hat{\rho}_t(x, y) > 0\}$.

On H_ρ, we consider $\langle \cdot, \cdot \rangle_{\hat{\rho}}$, which thanks to the identification $\sim$ is a nondegenerate positive scalar product: so (3.1.14) means that

$$\Psi \mapsto \nabla \cdot (\Psi \bullet \hat{\rho}) \qquad (\text{well defined in } H_\rho)$$

is minus the adjoint to the "gradient" operator $\nabla : L^2(\pi) \to H_\rho$. Hence its kernel is $\text{Ran}(\nabla)^\perp$.

To sum up, given that $(\rho_t, V_t)_t$ solves (CE'), then $(\rho_t, W_t)_t$ is another solution if and only if for a.e. t it holds

$$W_t = V_t + \Psi_t \qquad \text{in } H_{\rho_t}$$

where Ψ_t is any element orthogonal to the gradients in H_{ρ_t}. Among these W_t, the one of least norm is the orthogonal projection of V_t onto $\{\nabla\psi : \psi \in \mathbb{R}^X\}$, from which the conclusion follows.

□

Theorem 3.1.28 (The infimium is attained)**.** *Suppose that ρ_0, ρ_1 are such that $\mathcal{W}(\rho_0, \rho_1) < \infty$. Then the infimum which defines $\mathcal{W}$ is attained (and a fortiori also the infimum in* (3.1.13)*).*

Proof. Take (ρ^n, V^n) any minimizing sequence in (3.1.13). Firstly, we note that

$$\dot{\rho}_t(x)^2 = \frac{1}{4}\left(\sum_{y\in X}(V_t(x, y) - V_t(y, x))\hat{\rho}_t(x, y)K(x, y)\pi(x)\right)^2 \overset{\text{Jensen}}{\leq}$$
$$\leq \frac{1}{4}\sum_{y\in X}(V_t(x, y) - V_t(y, x))^2\hat{\rho}_t(x, y)^2K(x, y)\pi(x) \leq C < \infty$$

thanks to the fact that (ρ^n, V^n) is minimizing and $\hat{\rho}_t$ is equibounded. Therefore, (ρ^n) is a bounded sequence in $W^{1,2}(0,1)$ and so, possibly passing to a subsequence, we can suppose that $\rho^n \rightharpoonup \rho^\infty$ in $W^{1,2}(0,1)$ and, by compactness of the embedding of $W^{1,2}(0,1)$ in $C([0,1])$, also that $\rho^n \to \rho^\infty$ uniformly. In particular, $\rho_0^\infty = \rho_0$ and $\rho_1^\infty = \rho_1$.

Now we can use a form of semicontinuity of the L^2 relative norm (in the spirit of Lemma 1.3.20). Precisely, consider on $[0,1]$ the measures

$$\mu^n_{x,y}(dt) := \hat{\rho}^n_t(x,y)K(x,y)\pi(x)dt, \qquad \mu^\infty_{x,y}(dt) := \hat{\rho}^\infty_t(x,y)K(x,y)\pi(x)dt.$$

Evidently, the total masses are bounded and $\mu^n \rightharpoonup \mu^\infty$ as vector-valued measures. Put $\nu^n_{x,y}(dt) := V^n_t(x,y)\mu^n_{x,y}dt$.

If μ is any measure with values in $\mathbb{R}^{X\times X}$ and nonnegative components, then $L^2(\mu;\mathbb{R}^{X\times X})$ with the scalar product $\langle f,g\rangle_{L^2(\mu)} := \sum_{x,y}\langle f_{x,y}, g_{x,y}\rangle_{L^2(\mu_{x,y})}$ is Hilbert. For $g \in C_b([0,1];\mathbb{R}^{X\times X})$, we have

$$|\langle \nu_n, g\rangle| = \left|\sum_{x,y}\int_{[0,1]} g_{x,y}V^n(x,y)\ d\mu^n_{x,y}\right| \overset{\text{Cauchy-Schwarz}}{\leq} \|V^n\|_{L^2(\mu^n)}\|g\|_{L^2(\mu^n)} \tag{3.1.15}$$

Since $\|g\|_{L^2(\mu^n)} \leq C\|g\|_\infty$, we infer that ν^n are equibounded vector-valued measures, and so we can suppose that (ν^n) converges weakly (i.e. in duality with $C_0\left([0,1];\mathbb{R}^{X\times X}\right)$) to a certain ν. Letting $n\to\infty$ in (3.1.15), we get that $g\mapsto |\langle\nu,g\rangle|$ is the restriction to $C_0\left([0,1];\mathbb{R}^{X\times X}\right)$ of a continuous linear functional on $L^2(\mu^\infty)$ with norm less or equal than $\lim \|V^n\|_{L^2(\mu^n)} = \sqrt{2}\mathcal{W}(\rho_0,\rho_1)$. Hence, it can be represented as $\langle V^\infty,\cdot\rangle_{L^2(\mu)}$ where $\|V^\infty\|_{L^2(\mu^\infty)} \leq \sqrt{2}\mathcal{W}(\rho_0,\rho_1)$.

In other terms, there exists a function $V^\infty_t(x,y)$ such that

$$V^n_t(x,y)\hat{\rho}^n_t(x,y)K(x,y)\pi(x)dt \rightharpoonup V^\infty_t(x,y)\hat{\rho}^\infty_t(x,y)K(x,y)\pi(x)dt \quad \forall x,y\in X \tag{3.1.16}$$

and

$$\frac{1}{2}\int_0^1 \sum_{x,y\in X} V^\infty_t(x,y)^2\hat{\rho}^\infty_t(x,y)K(x,y)\pi(x)dt \leq \mathcal{W}(\rho_0,\rho_1)^2.$$

Finally, we prove that $(\rho^\infty_t, V^\infty_t)$ solves (CE'). We already know that ρ^∞ has a distributional derivative in $L^2(0,1)$; hence we can let $n\to\infty$ in the continuity equation for (ρ^n_t, V^n_t), and using (3.1.16) the conclusion follows.

The result of the argument above is that the infimum in (3.1.13) is attained; but then, thanks to the construction used in the proof of Proposition 3.1.27, there is also a minimizing couple of the form $(\rho_t, \nabla\psi_t)_t$.

□

Theorem 3.1.29 (Existence of geodesics). *For every $\rho_0,\rho_1\in\mathscr{P}(X)$ such that $\mathcal{W}(\rho_0,\rho_1)<\infty$, if (ρ_t,V_t) is a minimizing couple in* (3.1.13)*, then*

$$\|V_t\|_{\hat{\rho}_t} = \mathcal{W}(\rho_0,\rho_1) \qquad \textit{for a.e. } t\in[0,1]$$

and ρ_t is a constant speed geodesic in $\mathscr{P}(X)$. So, in light of the previous theorem, a constant speed geodesic from ρ_0 to ρ_1 exists. In particular, if $\mathcal{W}$ is always finite, then $\mathscr{P}(X)$ is a geodesic space.

Proof. Let (ρ_t, V_t) be a minimizing couple in (3.1.13). Then we have

$$\mathcal{W}(\rho_0,\rho_1) \overset{\text{Lemma 3.1.22}}{\leq} \int_0^1 \|V_t\|_{\hat{\rho}_t}\,dt \overset{\text{Jensen}}{\leq} \sqrt{\int_0^1 \|V_t\|_{\hat{\rho}_t}^2\,dt} = \mathcal{W}(\rho_0,\rho_1).$$

As a consequence, equality is achieved in Jensen's inequality: hence the integrand is a.e. constant, precisely $\|V_t\|_{\hat{\rho}_t} = \mathcal{W}(\rho_0,\rho_1)$ for a.e. t. Using the definition we deduce that $\mathcal{W}(\rho_s,\rho_t) \leq (t-s)\mathcal{W}(\rho_0,\rho_1)$ for every $[s,t] \subseteq [0,1]$, so (ρ_t) is a constant speed geodesic. □

A complete description of $AC_2(\mathscr{P}(X))$ is possible, in perfect analogy with Theorem 2.1.16.

Theorem 3.1.30 (Description of $AC_2(\mathscr{P}(X))$)**.**

1. *Let $(\rho_t, V_t)_{t\in[0,1]}$ be a solution of* (CE') *with $\int_0^1 \|V_t\|_{\hat{\rho}_t}^2 < \infty$. Then $(\rho_t)_{t\in[0,1]}$ is an AC_2 curve in $\mathscr{P}(X)$ and, denoting by $|\rho_t'|_{\mathscr{P}(X)}$ its* metric *derivative, it holds*
$$|\rho_t'|_{\mathscr{P}(X)} \leq \|V_t\|_{\hat{\rho}_t} \qquad \textit{for a.e. } t\in[0,1].$$
2. *Conversely, if $(\rho_t)_{t\in[0,1]}$ is an AC_2 curve in $\mathscr{P}(X)$, then it satisfies* (CE') *for some V_t such that $\|V_t\|_{\hat{\rho}_t} = |\rho_t'|_{\mathscr{P}(X)}$ for a.e. $t\in[0,1]$ (the least possible value, according to part 1.). It is also possible to take V_t of the form $\nabla\psi_t$.*

Proof. The proof of part 1. is absolutely identical to its continuous counterpart.

Also for part 2. we can exploit the idea of the old proof: namely, for $m\in\mathbb{N}$, we use Theorem 3.1.29 to build $(\rho_t^m, V_t^m)_{t\in[0,1]}$ solving (CE'), such that ρ_t^m coincides with ρ_t in $t=\frac{i}{m}$ for $i=0,\dots,m$, and

$$\|V_t^m\|_{\hat{\rho}_t^m} = m\mathcal{W}(\rho_{(i-1)/m},\rho_{i/m}) \qquad \text{for every } i=1,\dots,m \text{ and a.e. } t\in\left[\frac{i-1}{m},\frac{i}{m}\right].$$

In particular, for a.e. $t\in\left[\frac{i-1}{m},\frac{i}{m}\right]$, it holds

$$\|V_t^m\|_{\hat{\rho}_t^m}^2 = m^2\mathcal{W}(\rho_{(i-1)/m},\rho_{i/m})^2 \leq m\int_{(i-1)/m}^{i/m} |\rho_s'|_{\mathscr{P}(X)}^2 ds.$$

Integrating in dt for $t\in\left[\frac{i-1}{m},\frac{i}{m}\right]$ and summing over i, we deduce

$$\int_0^1 \|V_t^m\|_{\hat{\rho}_t^m}^2\,dt \leq \int_0^1 |\rho_s'|_{\mathscr{P}(X)}^2 ds < \infty. \tag{3.1.17}$$

We observe that $\rho_t^m \to \rho_t$ for $m\to\infty$ uniformly: in fact, for $t\in[0,1]$, we have

$$|\rho_t^m - \rho_t| \leq \left|\rho_t - \rho_{\lfloor mt\rfloor/m}\right| + \left|\rho_t^m - \rho_{\lfloor mt\rfloor/m}^m\right|$$

where the first addend goes to 0 for uniform continuity of ρ_t, and the second is equal to

$$\left(t-\frac{\lfloor mt\rfloor}{m}\right)\mathcal{W}\left(\rho_{\lfloor mt\rfloor/m},\rho_{\lceil mt\rceil/m}\right) \xrightarrow{m\to\infty} 0.$$

This uniform convergence and inequality (3.1.17) allow the repetition of the semi-continuity argument used in the proof of Theorem 3.1.28. The result is that there exists a velocity field V_t such that $(\rho_t, V_t)_{t\in[0,1]}$ satisfies the continuity equation and

$$\int_0^1 \|V_t\|^2_{\hat{\rho}_t}\, dt \le \int_0^1 |\rho'_t|^2_{\mathscr{P}(X)} dt.$$

But part 1. tells us that pointwise we have the converse inequality $\|V_t\|_{\hat{\rho}_t} \ge |\rho'_t|_{\mathscr{P}(X)}$ for a.e. t: the conclusion follows. We know that we can replace V_t with a velocity field of the form $\nabla\psi_t$ without increasing the norm, so the proof is complete. □

Remark 3.1.31. Every V_t of the form $\nabla\psi_t$ and solving (CE') satisfies part 2.: in fact, among the velocity fields that solve (CE'), the gradients have minimum norm (see the proof of Proposition 3.1.27). However, this time the minimizing gradient velocity field may not be unique. We will anyway find a canonical way to choose a "tangent" field: see Remark 3.2.8.

Before we continue our comparison with the continuous case, it is worth studying in detail on which subspaces of $\mathscr{P}(X)$ the extended distance $\mathcal{W}$ is a finite distance. We will need the following assumption:

Hypothesis 3.1.32. The function θ, besides satisfying Hypothesis 3.1.2, has the **doubling property**: for every $T > 0$ there exists a constant $C_d = C_d(T) > 0$ such that $\theta(2r, 2s) \le 2C_d\, \theta(r, s)$ for every $s, t \in [0, T]$.

Remark 3.1.33. Any 1-homogeneous θ (like the logarithmic mean) satisfies this assumption with $C_d = 1$.

Remark 3.1.34. Combining the doubling property and the monotonicity of θ, we have that for every $M > 0$ there exists a C_M such that $\theta(\alpha r, \alpha s) \le C_M \theta(r, s)$ for every $\alpha \in [0, M]$.

Notation.

$$C_\theta := \int_0^1 \frac{1}{\sqrt{\theta(1-r, 1+r)}} dr. \tag{3.1.18}$$

Remark 3.1.35 (Meaning of C_θ). $\sqrt{\frac{2}{p}} C_\theta$ is the distance between a Dirac mass and the uniform distribution on the two point space with $p = q > 0$ (Theorem 3.1.14). Hence $C_\theta < \infty$ if and only if $\mathcal{W}$ is finite on the space of probability measures on this 2-point space. We know that this is true if θ is the logarithmic mean.

We can extend our considerations to generic positive measures (with a common fixed total mass), because finiteness of $\mathcal{W}$ is scaling invariant:

Lemma 3.1.36 (Positive measures)**.** *Let $m > 0$. Then there exist c, C positive constants such that*

$$c\, \mathcal{W}(\rho_0, \rho_1) \le \mathcal{W}(m\rho_0, m\rho_1) \le C\, \mathcal{W}(\rho_0, \rho_1).$$

Moreover, if θ is 1-homogeneous, then $\mathcal{W}(m\rho_0, m\rho_1) = \sqrt{m}\, \mathcal{W}(\rho_0, \rho_1)$.

Proof. It is sufficient to prove that for every $m > 0$ and every $\rho_0, \rho_1 \in \mathbb{R}^X$ it holds $\mathcal{W}(m\rho_0, m\rho_1) \leq C_m\, \mathcal{W}(\rho_0, \rho_1)$, with $C_m = \sqrt{m}$ in the 1-homogeneous case. In fact, using this inequality with $\frac{1}{m}$ in the place of m would then conclude the proof of the lemma.

Suppose that (ρ_t, V_t) satisfies $\partial_t \rho_t + \nabla \cdot (\hat{\rho}_t \bullet V_t) = 0$. This can be written as $\partial_t(m\rho_t) + \nabla \cdot (\widehat{m\rho_t} \bullet \tilde{V}_t)$ if we call

$$\tilde{V}_t(x,y) := \frac{m\hat{\rho}_t(x,y)}{\widehat{m\rho_t}(x,y)} V_t(x,y) \qquad \left(\text{convention: } \frac{0}{0} = 0\right).$$

Now we observe that

$$\sum_{x,y\in X} \tilde{V}_t(x,y)^2 \widehat{m\rho_t}(x,y) K(x,y)\pi(x) = \sum_{x,y\in X} V_t(x,y)^2 \hat{\rho}_t(x,y) K(x,y)\pi(x) \frac{m^2 \hat{\rho}_t(x,y)}{\widehat{m\rho_t}(x,y)}.$$

If θ is 1-homogeneous, we immediately deduce that the fraction at the right hand side is equal to m; otherwise, we can estimate it from above with a constant by Remark 3.1.34. In both cases, for arbitrariness of (ρ_t, V_t) the conclusion easily follows. $\square$

Definition 3.1.37. For $\rho \in \mathscr{P}(X)$ and $x, y \in X$, we will write $\mathbf{x} \sim_{\boldsymbol{\rho}} \mathbf{y}$ if either $x = y$ or there exists a path in the graph associated to K linking x with y and involving only points where $\rho > 0$ (in particular $\rho(x), \rho(y) > 0$). This defines an equivalence relation on X; the points of $X \setminus \mathrm{supp}(\rho)$ are isolated, while the equivalence classes in $\mathrm{supp}(\rho)$ will be called the **connected components of the support of $\boldsymbol{\rho}$**.

Theorem 3.1.38 (Characterisation of finiteness)**.**

1. *If $C_\theta < \infty$, then $\mathcal{W}(\rho_0, \rho_1) < \infty$ for every $\rho_0, \rho_1 \in \mathscr{P}(X)$.*
2. *If $C_\theta = \infty$, then $\mathcal{W}(\rho_0, \rho_1) < \infty$ if and only if ρ_0 and ρ_1 have the same support and each connected component of it has the same total mass for the two measures.*

For the proof, we need two lemmas which compare $\mathcal{W}$ on X with the explicitly known distances $\mathcal{W}_{p,q}$ on the 2-point space.

Lemma 3.1.39 (Comparison I)**.** *Let $a, b \in X$ be such that $K(a,b) > 0$, and suppose that $\rho_0, \rho_1 \in \mathscr{P}(X)$ coincide on $X \setminus \{a,b\}$; put $p := K(a,b)\pi(a)$. Consider on the 2-point space $Y = \{a,b\}$ the Markov kernel $\bar{K}$ such that $\bar{K}(a,b) = \bar{K}(b,a) = p$, and denote by $\mathcal{W}_{p,p}$ the induced distance between measures on Y; call $\bar{\rho}_0, \bar{\rho}_1$ the (densities of the) restrictions to Y of the measures on X whose densities are ρ_0, ρ_1. Then*

$$\mathcal{W}(\rho_0, \rho_1) \leq \sqrt{C_d}\, \mathcal{W}_{p,p}(\bar{\rho}_0, \bar{\rho}_1).$$

Proof. Let $(\bar{\rho}_t)_{t\in[0,1]} \subseteq \mathscr{P}(Y)$, $\big(\bar{\psi}_t\big)_{t\in[0,1]} \subseteq \mathbb{R}^Y$ be such that $(\bar{\rho}_t, \nabla\bar{\psi}_t)_t$ satisfies (CE'), i.e.

$$\begin{cases} \dot{\bar{\rho}}_t(a) + \big(\bar{\psi}_t(b) - \bar{\psi}_t(a)\big)\, p\, \hat{\bar{\rho}}_t(a,b) = 0 \\ \dot{\bar{\rho}}_t(b) + \big(\bar{\psi}_t(a) - \bar{\psi}_t(b)\big)\, p\, \hat{\bar{\rho}}_t(a,b) = 0. \end{cases} \tag{3.1.19}$$

Then we can extend the measure (whose density w.r.t. $\bar{\pi}$ is) $\bar{\rho}_t$ to a probability measure (whose density w.r.t. π is) $\rho_t \in \mathscr{P}(X)$, simply imposing $\rho_t(x) = \rho_0(x)$ for every $x \in X \setminus \{a, b\}$. ($\rho_t$ is a probability measure since ρ_0 is and (CE') preserves total mass.) Since the invariant probability vector for $\bar{K}$ is $\bar{\pi}(a) = \bar{\pi}(b) = \frac{1}{2}$, we are saying $2\pi(a)\rho_t(a) = \bar{\rho}_t(a)$ and $2\pi(b)\rho_t(b) = \bar{\rho}_t(b)$.

If we write the system (3.1.19) in terms of ρ_t, we can give to it the form of a continuity equation for the curve (ρ_t) using the velocity field

$$\begin{aligned} V_t(a,b) &:= -V_t(b,a) := \frac{\hat{\bar{\rho}}_t(a,b)}{2\hat{\rho}_t(a,b)} \left(\bar{\psi}_t(a) - \bar{\psi}_t(b)\right), \qquad \left(\frac{0}{0} := 0\right) \\ V_t(x,y) &:= 0 \qquad \text{if } (x,y) \notin \{(a,b),(b,a)\}. \end{aligned}$$

Hence

$$\begin{aligned} \mathcal{W}(\rho_0,\rho_1)^2 &\le \frac{1}{2}\int_0^1 [V_t(a,b)^2\hat{\rho}_t(a,b)K(a,b)\pi(a) + V_t(b,a)^2\hat{\rho}_t(b,a)K(b,a)\pi(b)]dt = \\ &= \frac{1}{4}\int_0^1 \frac{\hat{\bar{\rho}}_t(a,b)^2}{\hat{\rho}_t(a,b)} \left(\bar{\psi}(b) - \bar{\psi}(a)\right)^2 p\ dt. \quad (3.1.20) \end{aligned}$$

The monotonicity and doubling properties of θ imply

$$\hat{\bar{\rho}}_t(a,b) = \theta\left(2\pi(a)\rho_t(a), 2\pi(b)\rho_t(b)\right) \le \theta\left(2\rho_t(a), 2\rho_t(b)\right) \le 2C_d\ \hat{\rho}_t(a,b).$$

Using this to eliminate $\hat{\rho}_t$ in (3.1.20), and since $p = 2\bar{\pi}(a)\bar{K}(a,b) = 2\bar{\pi}(b)\bar{K}(b,a)$, we obtain that $\mathcal{W}(\rho_0,\rho_1)^2$ is less or equal than

$$\frac{1}{2}C_d \int_0^1 \nabla\bar{\psi}_t(a,b)^2\hat{\bar{\rho}}_t(a,b)[\bar{\pi}(a)\bar{K}(a,b) + \bar{\pi}(b)\bar{K}(b,a)]dt = C_d \int_0^1 \left\|\nabla\bar{\psi}_t\right\|^2_{\hat{\bar{\rho}}_t} dt.$$

The infimum over all the admissible choices of $(\bar{\rho}, \bar{\psi})$ gives the conclusion.

□

Lemma 3.1.40 (Comparison II). *Consider on the 2-point space $Y = \{a,b\}$ the Markov kernel such that $\bar{K}(a,b) = \bar{K}(b,a) = 1$, and call $\mathcal{W}_{1,1}$ the induced distance between measures on Y. Fix $o \in X$; for every $\rho \in \mathscr{P}(X)$ call $\underline{\rho}$ the element of $\mathscr{P}(Y)$ giving mass $\pi(o)\rho(o)$ to the point a.*

Then there exists a constant c depending only on K and θ, such that

$$\mathcal{W}(\rho_0,\rho_1) \ge c\ \mathcal{W}_{1,1}\left(\underline{\rho_0}, \underline{\rho_1}\right).$$

Moreover, if θ is concave and 1-homogeneous (like the logarithmic mean), then the inequality is true with $c = 1$.

Proof. We preliminarily note that for every $\rho \in \mathscr{P}(X)$ it holds

$$\underline{\rho}(a) = 2\pi(o)\rho(o), \qquad \underline{\rho}(b) = 2\sum_{x\ne o}\pi(x)\rho(x).$$

Let $(\rho_t)_{t\in[0,1]}$ be any curve in $\mathscr{P}(X)$ from ρ_0 to ρ_1, and suppose that $(\rho_t, \nabla\psi_t)_t$ satisfies the continuity equation. Then

$$\dot{\rho}_t(o) + \sum_{x\ne o}\left(\psi_t(x) - \psi_t(o)\right)\hat{\rho}_t(o,x)K(o,x) = 0.$$

We multiply this equation by $2\pi(o)$ to get

$$\underline{\dot{\rho}_t}(a) + 2\sum_{x\neq o} (\psi_t(x) - \psi_t(o))\, \hat{\rho}_t(o,x)K(o,x)\pi(o) = 0.$$

We would like to write this in the form

$$\underline{\dot{\rho}_t}(a) + \big(\underline{\psi_t}(b) - \underline{\psi_t}(a)\big)\, \underline{\hat{\rho}_t}(a,b) = 0$$

so that $(\underline{\rho_t}, \nabla\underline{\psi_t})$ solves (CE') (recall that the equation for $\underline{\dot{\rho}_t}(b)$ is then automatically satisfied because total mass is constant). To this aim, we put (with the convention $\frac{0}{0} = 0$):

$$\begin{aligned}
\underline{\psi_t}(a) &:= \frac{2\sum_{x\neq o} \hat{\rho}_t(o,x)K(o,x)\pi(o)}{\underline{\hat{\rho}_t}(a,b)}\psi_t(o), \\
\underline{\psi_t}(b) &:= \frac{2\sum_{x\neq o} \hat{\rho}_t(o,x)K(o,x)\pi(o)\psi_t(x)}{\underline{\hat{\rho}_t}(a,b)}.
\end{aligned}$$

For simplicity of notation, we will write S_t for the sum $\sum_{x\neq o} \hat{\rho}_t(o,x)K(o,x)\pi(o)$.

We have to estimate

$$\|\nabla\psi_t\|^2_{\hat{\rho}_t} = \frac{1}{2}\sum_{x,y\in X} (\psi_t(x) - \psi_t(y))^2\, \hat{\rho}_t(x,y)K(x,y)\pi(x).$$

Firstly, we forget the addends with $x \neq o$ and $y \neq o$, getting

$$\|\nabla\psi_t\|^2_{\hat{\rho}_t} \geq \sum_{x\neq o} (\psi_t(x) - \psi_t(o))^2\, \hat{\rho}_t(o,x)K(o,x)\pi(o).$$

(We used that K is reversible.) Expanding the square, this is equal to

$$\sum_{x\neq o} \psi_t(x)^2\hat{\rho}_t(o,x)K(o,x)\pi(o) - 2\psi_t(o)\sum_{x\neq o} \psi_t(x)\hat{\rho}_t(o,x)K(o,x)\pi(o) + S_t\psi_t(o)^2. \tag{3.1.21}$$

We want to express this in terms of $\underline{\rho_t}$, $\underline{\psi_t}$.

To begin, let us suppose that $S_t > 0$. Jensen's inequality gives

$$\frac{1}{S_t}\sum_{x\neq o} \psi_t(x)^2\hat{\rho}_t(o,x)K(o,x)\pi(o) \geq \left(\frac{1}{S_t}\sum_{x\neq o} \psi_t(x)\hat{\rho}_t(o,x)K(o,x)\pi(o)\right)^2$$

Inserting this in (3.1.21) and using the definition of $\underline{\psi_t}$, we get

$$\begin{aligned}
(3.1.21) \geq \frac{1}{4S_t}\underline{\psi_t}(b)^2\underline{\hat{\rho}_t}(a,b)^2 - \underline{\psi_t}(b)\underline{\hat{\rho}_t}(a,b)\frac{\underline{\psi_t}(a)\underline{\hat{\rho}_t}(a,b)}{2S_t} + S_t\frac{\underline{\psi_t}(a)^2\underline{\hat{\rho}_t}(a,b)^2}{4S_t^2} &= \\
= \frac{\underline{\hat{\rho}_t}(a,b)^2}{4S_t}\big(\underline{\psi_t}(b) - \underline{\psi_t}(a)\big)^2&.
\end{aligned}$$

If we could prove that for some $c > 0$ it holds

$$\frac{\underline{\hat{\rho}_t}(a,b)}{2S_t} \geq c^2 \qquad \text{(with } c = 1 \text{ if } \theta \text{ is concave 1-homogeneous),} \tag{3.1.22}$$

then we would conclude that

$$\|\nabla\psi_t\|^2_{\hat{\rho}_t} \geq c^2\left(\underline{\psi_t}(b) - \underline{\psi_t}(a)\right)^2 \frac{1}{2}\underline{\hat{\rho}_t}(a,b) = c^2\left\|\nabla\underline{\psi_t}\right\|^2_{\underline{\hat{\rho}_t}}$$

which is true by direct inspection also if $S_t = 0$; integration on t would then give the thesis for arbitrariness of ρ_t and ψ_t.

It remains only to prove (3.1.22). We once remarked that the doubling property implies that for a suitable constant C,

$$\begin{aligned}
2S_t = 2\pi(o)\sum_{x\neq o} K(o,x)\theta(\rho_t(o),\rho_t(x)) &\overset{\text{Remark 3.1.34}}{\leq} \\
&\leq C\sum_{x\neq o}\theta\Big(2\pi(o)K(o,x)\rho_t(o), 2\pi(o)K(o,x)\rho_t(x)\Big) = \\
&= C\sum_{x\neq o}\theta\Big(2\pi(o)K(o,x)\rho_t(o), 2\pi(x)K(x,o)\rho_t(x)\Big).
\end{aligned}$$

Using monotonicity of θ and the fact that $K(o,x), K(x,o) \leq 1$, the right hand side is less or equal than

$$C\sum_{x\neq o}\theta(2\pi(o)\rho_t(o), 2\pi(x)\rho_t(x)) \overset{\text{monotonicity}}{\leq} C\sum_{x\neq o}\theta\left(2\pi(o)\rho_t(o), 2\sum_{y\neq o}\pi(y)\rho_t(y)\right)$$

which is equal to $C|X|\underline{\hat{\rho}_t}(a,b)$, and the desired inequality is proven.

Finally, with the additional hypothesis that θ is concave 1-homogeneous, we can perform a better estimate. In fact,

$$\underline{\hat{\rho}_t}(a,b) = \theta\left(2\rho_t(o)\pi(o), 2\sum_{y\neq o}\pi(y)\rho_t(y)\right) = 2\pi(o)\theta\left(\rho_t(o), \sum_{y\neq o}\rho_t(y)\frac{\pi(y)}{\pi(o)}\right).$$

Noting that $\pi(y) = \sum_{x\in X}\pi(x)K(x,y) \geq \pi(o)K(o,y)$ and using twice the monotonicity property we get that the last expression is greater or equal than

$$2\pi(o)\theta\left(\rho_t(o), \sum_{y\neq o}\rho_t(y)K(o,y)\right) \geq 2\pi(o)\theta\left(\sum_{y\neq o}\rho_t(o)K(o,y), \sum_{y\neq o}\rho_t(y)K(o,y)\right)$$

which by concavity and 1-homogeneity is greater or equal than

$$2\pi(o)\sum_{y\neq o}K(o,y)\theta\left(\rho_t(o),\rho_t(y)\right) = 2S_t.$$

□

Proof of Theorem 3.1.38 (Characterisation of finiteness).

$\boxed{1.}$ If $C_\theta < \infty$, thanks to Lemma 3.1.39, if $\eta, \sigma \in \mathscr{P}(X)$ differ only in a couple of points a, b such that $K(a,b) > 0$, then $\mathcal{W}(\eta, \sigma) < \infty$. Now given any $\rho_0, \rho_1 \in \mathscr{P}(X)$ distinct, we claim that there exists $\tilde{\rho}_0 \in \mathscr{P}(X)$ at a finite distance from ρ_0 and such that

$$\#\{x : \tilde{\rho}_0(x) \neq \rho_1(x)\} < \#\{x : \rho_0(x) \neq \rho_1(x)\}.$$

Iterating this argument a finite number of times, the conclusion would follow.

Let us prove the claim. Since the total mass is 1, there are $\bar{x}, \bar{y} \in X$ such that $\rho_0(\bar{x}) > \rho_1(\bar{x})$ and $\rho_0(\bar{y}) < \rho_1(\bar{y})$. Call m the exceeding mass which ρ_0 has in $\bar{x}$. For irreducibility of K, there exists a chain $\bar{x} = x_0, x_1, \ldots, x_N = \bar{y}$ such that $K(x_{i-1}, x_i) > 0$ for every $i = 1, \ldots, N$. We recursively build intermediate measures $(\rho^{(i)})_{i=0,\ldots,N}$ as follows:

- $\rho^{(0)} = \rho_0$;
- For $i = 0, \ldots, N-1$, $\rho^{(i+1)}$ differs from $\rho^{(i)}$ only in the fact that a mass m has been "moved" from x_i to x_{i+1}.

By construction, $\mathcal{W}(\rho^{(i-1)}, \rho^{(i)}) < \infty$, $\rho^{(0)} = \rho_0$ and $\rho^{(N)}$ differs from ρ_0 only in the points $\bar{x}, \bar{y}$; moreover, $\rho^{(N)}(\bar{x}) = \rho_1(\bar{x})$. Putting $\tilde{\rho}_0 := \rho^{(N)}$, the claim is proved.

$\boxed{2.}$ If ρ_0, ρ_1 have the same support and give the same mass to each connected component of the support, then we can still build $\tilde{\rho}_0$ at a finite distance from ρ_0 and differing from ρ_1 in a smaller number of points. In fact, we can repeat the construction above, with the only additional attention to take $\bar{x}, \bar{y}$ in the same connected component of the support (here we use that the total mass of this component is equal for the two measures), and to chose the x_i in such a way that $\rho_0(x_i) > 0$ for every i (possible because $x \sim_{\rho_0} y$). We note that at each step, the application of Lemma 3.1.39 gives an upper bound of $\mathcal{W}(\rho^{(i-1)}, \rho^{(i)})$ in terms of the distance between two measures with strictly positive densities on the 2-point space: this distance is always finite, and the conclusion follows.

Conversely, suppose that $\mathcal{W}(\rho_0, \rho_1) < \infty$. If by contradiction the supports were different, there would be $o \in X$ such that $\rho_0(o) > 0$ and $\rho_1(o) = 0$ or vice versa, so that an application of Lemma 3.1.40 would give $\mathcal{W}(\rho_0, \rho_1) = \infty$, which is not the case. Therefore, $\mathrm{supp}(\rho_0) = \mathrm{supp}(\rho_1)$.

Take now any $(\rho_t, \nabla\psi_t)_{t\in[0,1]}$ satisfying (CE') and such that $\int_0^1 \|\nabla\psi_t\|_{\hat{\rho}_t}^2 \, dt < \infty$. Then by definition $\mathcal{W}(\rho_0, \rho_t) < \infty$, so the previous argument gives $\mathrm{supp}(\rho_0) = \mathrm{supp}(\rho_t)$. But then $K(x,y)\hat{\rho}_t(x,y) = 0$ unless $x \sim_{\rho_0} y$, so the continuity equation reduces to

$$\dot{\rho}_t(x) + \sum_{y \sim_{\rho_0} x} (\psi_t(y) - \psi_t(x)) \, K(x,y)\hat{\rho}_t(x,y) = 0 \qquad \forall x \in X. \tag{3.1.23}$$

For every $z \in X$ fixed, consider the identities above for $x \sim_{\rho_0} z$, multiply each of them by $\pi(x)$ and sum over x: the result (using reversibility of K) is $\sum_{x \sim_{\rho_0} z} \dot{\rho}_t(x)\pi(x) = 0$, i.e. the total mass of the connected component of z is preserved.

□

Remark 3.1.41 (On global existence of solutions to (CE')). Assume that $C_\theta = \infty$, and suppose that $(V_t)_{t\in[0,T]} \subseteq \mathbb{R}^{X\times X}$ is such that $\sum_{x,y}\int_0^T V_t(x,y)^2 dt < \infty$. Then we claim that there exists a globally defined solution $(\rho_t)_{t\in[0,1]}$ of (CE') for every initial condition $\rho_0 \in \mathscr{P}_*(X)$. By Remark 3.1.6, it is sufficient to show that if ρ_t is a solution defined for $t \in [0,\tau)$, then $\rho_t > 0$ for every $t \in [0,\tau)$. If this was false, there would be a first value $\bar t$ such that $\rho_{\bar t}$ has some component equal to 0; evidently $\rho_{\bar t} \geq 0$, so it is an element of $\mathscr{P}(X) \setminus \mathscr{P}_*(X)$. But the hypothesis on V_t implies $\mathcal{W}(\rho_0, \rho_{\bar t}) < \infty$, and this contradicts the previous Theorem.

Remark 3.1.42 (Eulerian description of geodesics, formal proof). Mimicking Remark 2.1.24, we now formally derive an analogous of the "Eulerian" equations of geodesics (2.1.2). The result can be proven rigorously in a different way, as we will explain in Theorem 3.2.14.

Let $(\rho_t)_{t\in[0,1]}$ be a constant speed geodesic in $\mathscr{P}(X)$, and consider an "optimal" velocity field $\nabla\psi_t$ realising the infimum in the definition of $\mathcal{W}$. Consider any other curve $(\sigma_t)_{t\in[0,1]}$ from ρ_0 to ρ_1, and any $(W_t)_{t\in[0,1]} \subseteq \mathbb{R}^{X\times X}$ such that $(\sigma_t, W_t)_t$ solves (CE').

For $\varepsilon \in \mathbb{R}$, put $\rho_t^\varepsilon := (1-\varepsilon)\rho_t + \varepsilon\,\sigma_t$. This defines a curve in $\mathscr{P}(X)$ at least if $\rho_t(x) > c > 0$ for every t, x and ε is small: let us assume that this is true. With the usual trick, we find that a velocity field such that ρ_t^ε solves the continuity equation is

$$V_t^\varepsilon(x,y) := \frac{(1-\varepsilon)\nabla\psi_t(x,y)\hat\rho_t(x,y) + \varepsilon\, W_t(x,y)\hat\sigma_t(x,y)}{\hat\rho_t^\varepsilon(x,y)}.$$

So the following quantity has a local minimum in $\varepsilon = 0$:

$$\int_0^1 \|V_t^\varepsilon\|^2_{\hat\rho_t^\varepsilon}\,dt =$$
$$\int_0^1 \sum_{x,y} \left[(1-\varepsilon)\nabla\psi_t(x,y)\hat\rho_t(x,y) + \varepsilon\, W_t(x,y)\sigma(x,y)\right]^2 \hat\rho_t^\varepsilon(x,y)^{-1} K(x,y)\pi(x)dt.$$

Supposing that $\theta \in C^1((0,\infty)^2)$, and writing for brevity $\partial_1\hat\rho(x,y)$ for $\partial_1\theta(\rho(x),\rho(y))$, the derivative of the integrand (as a function of ε) in $\varepsilon = 0$ is (straightforward computation):

$$2\sum_{x,y}\Big\{\nabla\psi_t(x,y)\left[W_t(x,y)\hat\sigma_t(x,y) - \nabla\psi_t(x,y)\hat\rho_t(x,y)\right] +$$
$$-\nabla\psi_t(x,y)^2\partial_1\hat\rho_t(x,y)\left(\sigma_t(x)-\rho_t(x)\right)\Big\}K(x,y)\pi(x).$$

Now we note that

$$\sum_{x,y}\nabla\psi_t(x,y)W_t(x,y)\hat\sigma_t(x,y)K(x,y)\pi(x) = 2\langle\nabla\psi_t, W_t \bullet \hat\sigma_t\rangle_K =$$
$$= -2\langle\psi_t, \nabla\cdot(W_t\bullet\hat\sigma_t)\rangle_\pi = 2\langle\psi_t, \dot\sigma_t\rangle_\pi$$

and for the same reasons

$$\sum_{x,y}\nabla\psi_t(x,y)^2\hat\rho_t(x,y)K(x,y)\pi(x) = 2\langle\psi_t,\dot\rho_t\rangle_\pi$$

Hence, if we can differentiate under the integral sign and impose the result to be 0, the condition is

$$0 = \int_0^1 \sum_x \Big\{ 2\psi_t(x) \left(\dot\sigma_t(x) - \dot\rho_t(x)\right) \pi(x) + \\ - \sum_y \nabla\psi_t(x,y)^2 \partial_1\hat\rho_t(x,y) \left(\sigma_t(x) - \rho_t(x)\right) K(x,y)\pi(x) \Big\} \, dt.$$

Let us suppose that ψ_t is a C^1 function of t. (In fact it is not restrictive; we do not give any proof, because it will become clear after the Riemannian structure of $\mathscr{P}_*(X)$ is described: see the next section.) Integration by parts transforms this equation into

$$0 = \int_0^1 \sum_x \left[\dot\psi_t(x) + \frac{1}{2} \sum_y \nabla\psi_t(x,y)^2 \partial_1\hat\rho_t(x,y) K(x,y) \right] \left(\rho_t(x) - \sigma_t(x)\right) \pi(x) dt \tag{3.1.24}$$

for every (σ_t) curve in $\mathscr{P}(X)$ from ρ_0 to ρ_1 solving the continuity equation for some velocity field.

As in the continuous case, we cannot expect that this implies that the integrand is zero, because the couple $(\rho_t, \nabla\psi_t)$ is unchanged if we add to ψ_t a constant function $f(t)$. However, we claim that if a function $(\phi_t)_{t\in[0,1]} \subseteq \mathbb{R}^X$ satisfies

$$0 = \int_0^1 \sum_{x\in X} \phi_t(x) \left(\rho_t(x) - \sigma_t(x)\right) \pi(x) dt \tag{3.1.25}$$

for every (σ_t) as above, then ϕ_t *must* be constant in space (up to a negligible set of times). This would immediately imply that, with an additive perturbation, we can find a potential ψ_t satisfying the following "Hamilton-Jacobi"-like condition:

$$\dot\psi_t(x) + \frac{1}{2} \sum_{y\in X} \nabla\psi_t(x,y)^2 \partial_1\hat\rho_t(x,y) K(x,y) = 0 \qquad \forall x \in X. \tag{3.1.26}$$

To prove the claim, we choose any $y, z \in X$, and any $\eta \in C_c^\infty((0,1))$, and consider $\sigma_t := \mu_t + \varepsilon(\frac{\delta_y}{\pi(y)} - \frac{\delta_z}{\pi(z)})\eta(t)$, which is in $\mathscr{P}(X)$ if ε is small. Since σ_t is a.e. differentiable, then one can see that it satisfies the continuity equation for some velocity field (see for instance Theorem 3.2.7 below): hence it an admissible variation. But for this choice of σ_t, equation (3.1.25) becomes

$$\int_0^1 \left[\phi_t(y) - \phi_t(z)\right] \eta(t) dt = 0,$$

which for arbitrariness of $\eta \in C_c^\infty((0,1))$ yields $\phi_t(y) = \phi_t(z)$ for a.e. t as desired.

Finally, we note that in the case $C_\theta = \infty$, even if we do not know that $\rho_0 > 0$, the above argument can be adapted. In fact, the manipulations leading to (3.1.24) still work if we impose additionally that $\mathcal{W}(\sigma_t, \rho_0) < \infty$ for every t: by our characterisation of finiteness of $\mathcal{W}$, this means that the connected components of the supports of σ_t and ρ_0 are the same and have the same total mass. As for the claim, the variations σ_t used in its proof are now admissible if and only if $y \sim_{\rho_0} z$. Hence the conclusion is somewhat weaker: ϕ_t is constant *on the connected components of* $\operatorname{supp}(\rho_0)$. This is not a real issue, since if we add to ψ_t a function which is constant on the components of $\operatorname{supp}(\rho_0)$, then the quantities involved in the definition of $\mathcal{W}$ do not vary. Hence, we can still conclude that, for a suitable choice of the "optimal" potential ψ_t, equation (3.1.26) holds.

The system of ordinary differential equations

$$\begin{cases} \partial_t \rho_t + \nabla \cdot (\nabla \psi_t \bullet \hat{\rho}_t) = 0 \\ \partial_t \psi_t + \frac{1}{2} \sum_{y \in X} \nabla \psi_t(\cdot, y)^2 \partial_1 \hat{\rho}_t(\cdot, y) K(\cdot, y) = 0 \end{cases} \tag{3.1.27}$$

strongly resembles the "Eulerian" description of geodesics in $\mathscr{P}_2(\mathbb{R}^n)$ given by the system (2.1.2), except for the fact that now the equation for ψ_t depends also on ρ_t. It actually characterises constant speed geodesics: see Theorem 3.2.14.

3.2 Riemannian structure

Our aim is to show that under suitable hypotheses, $\mathcal{W}$ is a "Riemannian" distance. For easier manipulation, we write the quadratic quantity $\|\nabla \psi_t\|^2_{\hat{\rho}_t}$ in matrix notation as $\langle A(\rho_t)\psi_t, \psi_t \rangle$; it turns out that $A(\rho) \in \mathbb{R}^{X \times X}$ is given by

$$A_{x,y}(\rho) := \begin{cases} \sum_{z \neq x} K(x,z)\hat{\rho}(x,z)\pi(x), & \text{if } x = y \\ -K(x,y)\hat{\rho}(x,y)\pi(x), & \text{if } x \neq y. \end{cases} \tag{3.2.1}$$

Similarly, we can write the continuity equation as $\dot{\rho}_t = B(\rho_t)\psi_t$ where

$$B_{x,y}(\rho) := \begin{cases} \sum_{z \neq x} K(x,z)\hat{\rho}(x,z), & \text{if } x = y \\ -K(x,y)\hat{\rho}(x,y), & \text{if } x \neq y; \end{cases} \tag{3.2.2}$$

note that $A = \Pi B$ if Π is the diagonal matrix with diagonal entries $\pi(x)$. We collect for future use some linear algebra remarks.

Remark 3.2.1. Since $A(\rho)$ is symmetric positive semidefinite, then

$$\begin{aligned} \operatorname{Ker} A(\rho) &= \left\{ \psi \in \mathbb{R}^X : \langle A(\rho)\psi, \psi \rangle = 0 \right\} \\ &= \left\{ \psi \in \mathbb{R}^X : \psi(x) = \psi(y) \text{ whenever } x \sim_\rho y \right\} \end{aligned}$$

by definition of $\langle A(\rho)\psi, \psi \rangle$. Moreover, by symmetry of $A(\rho)$, we have

$$\operatorname{Ran} A(\rho) = (\operatorname{Ker} A(\rho))^\perp = \left\{ \psi \in \mathbb{R}^X : \sum_{z \sim_\rho x} \psi(z) = 0 \ \forall x \in X \right\}.$$

From the relation $A = \Pi B$, we also deduce that

$$\begin{aligned} \operatorname{Ker} B(\rho) &= \left\{ \psi \in \mathbb{R}^X : \psi(x) = \psi(y) \text{ whenever } x \sim_\rho y \right\}, \\ \operatorname{Ran} B(\rho) &= \left\{ \psi \in \mathbb{R}^X : \sum_{z \sim_\rho x} \psi(z)\pi(z) = 0 \ \forall x \in X \right\}. \end{aligned}$$

Notation. For $\sigma \in \mathscr{P}(X)$ and $b > 0$, we will write:

$$\begin{aligned}
\mathscr{P}_\sigma(X) :=& \{\rho \in \mathscr{P}(X) : \mathcal{W}(\rho, \sigma) < \infty\}; \\
\mathscr{P}'_\sigma(X) :=& \left\{ \rho \in \mathscr{P}(X) : \sum_{y \sim_\rho x} \rho(y)\pi(y) = \sum_{y \sim_\sigma x} \sigma(y)\pi(y) \ \forall x \in X \right\}; \\
\mathscr{P}^b_\sigma(X) :=& \left\{ \rho \in \mathscr{P}'_\sigma(X) : \rho(x) \geq b \ \forall x \in \operatorname{supp}(\sigma) \right\}; \\
\mathscr{P}_*(X) :=& \mathscr{P}'_{\mathbf{1}}(X) = \{\rho \in \mathscr{P}(X) : \rho(x) > 0 \ \forall x \in X\}.
\end{aligned}$$

(Note that every $\rho \in \mathscr{P}'_\sigma(X)$ satisfies $\operatorname{supp}(\rho) = \operatorname{supp}(\sigma)$.)

Thanks to the characterisation of finiteness of $\mathcal{W}$, we have that $\mathscr{P}_\sigma(X)$ is equal to $\mathscr{P}(X)$ if $C_\theta < \infty$, and to $\mathscr{P}'_\sigma(X)$ otherwise.

Lemma 3.2.2. *For every $\sigma \in \mathscr{P}(X)$ and $b > 0$ there exist constants $0 < c < C < \infty$ such that*

$$c \|\psi\| \leq \|B(\rho)\psi\| \leq C \|\psi\| \qquad \forall \psi \in \operatorname{Ran} A(\sigma) \ \forall \rho \in \mathscr{P}^b_\sigma(X).$$

Proof. The second inequality is immediate since the entries of $B(\rho)$ are bounded.

As for the first one, we note that by symmetry $A(\rho)$ restricts to an isomorphism on its range, and so $B(\rho) = \Pi^{-1}A(\rho)$ restricts to an isomorphism B_ρ of $\operatorname{Ran} A(\rho)$ onto $\operatorname{Ran} B(\rho)$. But $\operatorname{Ran} A(\rho) = \operatorname{Ran} A(\sigma)$ and $\operatorname{Ran} B(\rho) = \operatorname{Ran} B(\sigma)$ by Remark 3.2.1: hence, $\rho \mapsto B_\rho^{-1}$ is a well defined continuous mapping defined on the compact set $\mathscr{P}^b_\sigma(X)$ with values in the linear isomorphisms of $\operatorname{Ran} B(\sigma)$ onto $\operatorname{Ran} A(\sigma)$, and the conclusion follows. □

Corollary 3.2.3 (Partial converse of Lemma 3.1.24)**.** *For every $\sigma \in \mathscr{P}(X)$ and $b > 0$, there exists a constant $C > 0$ such that for every $\rho_0, \rho_1 \in \mathscr{P}^b_\sigma(X)$, setting $\rho_t := (1-t)\rho_0 + t\rho_1$ and for a suitable $\nabla\psi_t$ satisfying* (CE')*, it holds*

$$\mathcal{W}(\rho_0, \rho_1) \leq \left(\int_0^1 \|\nabla\psi_t\|^2 \, dt \right)^{1/2} \leq C d_{TV}(\rho_0, \rho_1).$$

Proof. Note that $\rho_t \in \mathscr{P}^b_\sigma(X)$ satisfies $\dot\rho_t = \rho_0 - \rho_1 \in \operatorname{Ran} B(\sigma)$ by the characterisation of the range of $B(\sigma)$ (Remark 3.2.1). Hence, with the notations of the proof of Lemma 3.2.2, we can take $\psi_t := B_{\rho_t}^{-1}\dot\rho_t \in \operatorname{Ran} A(\sigma)$, so that the continuity equation $\dot\rho_t = B(\rho_t)\psi_t$ is satisfied and $\|\psi_t\| \leq C \|\rho_0 - \rho_1\| \leq C' d_{TV}(\rho_0, \rho_1)$ by the same Lemma 3.2.2. In conclusion,

$$\mathcal{W}(\rho_0, \rho_1)^2 \leq \int_0^1 \langle A(\rho_t)\psi_t, \psi_t \rangle dt \leq \bar{C} d_{TV}(\rho_0, \rho_1)^2$$

where we also used the obvious fact that the matrices $A(\rho)$ are equibounded. □

Remark 3.2.4 ($\mathscr{P}'_\sigma(X)$ as a smooth manifold)*.* $\mathscr{P}'_\sigma(X)$, as a subset of $\mathbb{R}^X$, is simply the open subset $\{\psi \in \mathbb{R}^X : \psi(x) > 0 \ \forall x \in \operatorname{supp}(\sigma)\}$ of the affine subspace

$$\left\{ \psi \in \mathbb{R}^X : \psi(z) = 0 \ \forall z \notin \operatorname{supp}(\sigma), \ \sum_{y \sim_\sigma x} \psi(y)\pi(y) = \sum_{y \sim_\sigma x} \sigma(y)\pi(y) \ \forall x \in X \right\}.$$

Hence, it has a natural structure of smooth manifold of dimension $|\operatorname{supp}(\sigma)| - n(\sigma)$, where $n(\sigma)$ is the number of independent conditions of the form

$$\sum_{y \sim_\sigma x} \psi(y)\pi(y) = \sum_{y \sim_\sigma x} \sigma(y)\pi(y)$$

as x runs in X; i.e., $n(\sigma)$ is the number of connected components in $\operatorname{supp}(\sigma)$.

This structure is compatible with the topology induced by $\mathcal{W}$:

Proposition 3.2.5 (convergence in $\mathscr{P}_\sigma(X)$)**.** *Let $\sigma \in \mathscr{P}(X)$ and $\rho_\alpha, \rho \in \mathscr{P}_\sigma(X)$. Then $\rho_\alpha \xrightarrow{\mathcal{W}} \rho$ if and only if $\rho_\alpha \xrightarrow{d_{TV}} \rho$.*

Proof. We already know that d_{TV} is bounded above by a multiple of $\mathcal{W}$ (Lemma 3.1.24), so one implication is trivial. Conversely, suppose that $(\rho_\alpha) \subseteq \mathscr{P}_\sigma(X)$ converges to $\rho \in \mathscr{P}_\sigma(X)$ in total variation.

If $C_\theta < \infty$, for each α we can estimate $\mathcal{W}(\rho_\alpha, \rho)$ using the procedure used at page 83 in the first part of the proof of Theorem 3.1.38. In fact, the measures $\rho^{(i)}$ defined there satisfy

$$\mathcal{W}\left(\rho^{(i-1)}, \rho^{(i)}\right) \overset{\text{Lemma 3.1.39}}{\leq} \mathcal{W}_{p^{(i)},p^{(i)}}\left(\rho^{\beta^{(i)}}, \rho^{\tilde\beta^{(i)}}\right) = \frac{1}{\sqrt{2p^{(i)}}}\left|\int_{\beta^{(i)}}^{\tilde\beta^{(i)}} \frac{1}{\sqrt{\theta(\rho^r(a), \rho^r(b))}} dr\right|.$$

In this estimate, the number of intermediate measures $\rho^{(i)}$ is at most $|X|+1$ independently of α; $p^{(i)}$ can take only a finite number of values (see Lemma 3.1.39), which are strictly positive; and $\beta^{(i)} - \tilde\beta^{(i)}$ is infinitesimal in α thanks to the hypothesis $\rho_\alpha \xrightarrow{d_{TV}} \rho$. The absolute continuity of the integral gives the conclusion.

If $C_\theta = \infty$, then the hypothesis $\rho_\alpha, \rho \in \mathscr{P}_\sigma(X) = \mathscr{P}'_\sigma(X)$ enables the use of the procedure from the *second* part of the cited proof of Theorem 3.1.38. Now we can repeat the argument above literally, with the only additional remark that $\beta^{(i)}$ and $\tilde\beta^{(i)}$ definitively belong to some subinterval $(\varepsilon, 1-\varepsilon) \subset\subset (0,1)$, where the absolute continuity of the integral holds even without the hypothesis $C_\theta < \infty$. □

Proposition 3.2.6 (Completeness)**.** *The metric space $(\mathscr{P}_\sigma(X), \mathcal{W})$ is complete for every $\sigma \in \mathscr{P}(X)$.*

Proof. Let $(\rho_n) \subseteq \mathscr{P}_\sigma(X)$ be Cauchy with respect to $\mathcal{W}$.

If $C_\theta < \infty$ then $\mathscr{P}_\sigma(X) = \mathscr{P}(X)$; (ρ_n) is Cauchy also for d_{TV} (Lemma 3.1.24), and therefore converges in total variation: the conclusion comes from the previous proposition.

If $C_\theta = \infty$, from the $\mathcal{W}$-boundedness of the set $\{\rho_n : n \in \mathbb{N}\}$, by comparison with the two point space (Lemma 3.1.40), we get that there exists $b > 0$ such that $\rho_n \in \mathscr{P}^b_\sigma(X)$ for every n. But on $\mathscr{P}^b_\sigma(X)$, the distances d_{TV} and $\mathcal{W}$ are strongly equivalent by Lemma 3.1.24 and Corollary 3.2.3: since $\left(\mathscr{P}^b_\sigma(X), d_{TV}\right)$ is complete, the conclusion follows. □

We saw that $\mathscr{P}'_\sigma(X) = \sigma + \operatorname{Ran} B(\sigma)$: so, the tangent space can be identified with $\operatorname{Ran} B(\sigma)$. But there is also a way of thinking of the "tangent space" as a space of gradients via the continuity equation, in analogy with the definition of $\operatorname{Tan}_\mu \mathscr{P}_2(\mathbb{R}^n)$ from Chapter 2:

Theorem 3.2.7. *Fix $\sigma \in \mathscr{P}(X)$. Then:*

1. *For every $\rho \in \mathscr{P}'_\sigma(X)$ and $\phi \in \operatorname{Ran} B(\sigma)$, there exists a unique $\psi \in \operatorname{Ran} A(\sigma)$ such that $\phi = B(\rho)\psi$; the map $\mathcal{I}_\rho$ defined by*

$$\mathcal{I}_\rho \phi := \nabla\psi$$

   is a linear isomorphism of $\operatorname{Ran} B(\sigma)$ *onto*

$$T_\rho := \{\nabla\psi : \psi \in \operatorname{Ran} A(\sigma)\}.$$

2. *If $\rho : [0,1] \to \mathscr{P}'_\sigma(X)$ is differentiable at a point $t \in [0,1]$, then*

$$D_t\rho := \mathcal{I}_{\rho_t}\dot{\rho}_t$$

   is the unique element $\nabla\psi_t \in T_{\rho_t}$ satisfying $\dot{\rho}_t + \nabla\cdot(\hat{\rho}_t \bullet \nabla\psi_t) = 0$. In particular, every a.e. differentiable curve satisfies (CE') *for a suitable gradient velocity field.*

Proof.
$\boxed{1.}$ We already observed in the proof of Lemma 3.2.2 that $B(\rho)$ is a linear isomorphism of $\operatorname{Ran} A(\sigma)$ onto $\operatorname{Ran} B(\sigma)$: it remains to prove that $\psi \mapsto \nabla\psi$ is injective on $\operatorname{Ran} A(\sigma)$ (hence an isomorphism with its image T_ρ). This is true because if $\nabla\phi = 0$, then $0 = \|\nabla\psi\|^2_{\hat{\rho}} = \langle A(\rho)\psi, \psi\rangle$; $A(\rho)$ is symmetric positive semidefinite, therefore $A(\rho)\psi = 0$, and since $A(\rho)$ is an isomorphism on its image we deduce $\psi = 0$.
$\boxed{2.}$ By definition of $B(\rho)$, the equation $\dot{\rho}_t + \nabla\cdot(\hat{\rho}_t \bullet \nabla\psi_t) = 0$ can be rewritten as $\dot{\rho}_t = B(\rho_t)\psi_t$, so part 1. yields the conclusion.

□

Remark 3.2.8 (Tangent field). Let $(\rho_t) \subseteq \mathscr{P}'_\sigma(X)$ be any absolutely continuous curve with respect to $\mathcal{W}$. There exists an AC_2 reparametrisation of the curve, so Theorem 3.1.30 implies that ρ solves (CE') for some velocity field.

Moreover, if $(\rho_t) \subseteq \mathscr{P}'_\sigma(X)$ is any a.e. differentiable curve, then the Theorem above provides a canonical choice $D_t\rho$ of a velocity field $\nabla\psi_t$ satisfying (CE') (up to a negligible set of times). This **tangent velocity field** is characterized by the fact that for a.e. t, $\nabla\psi_t$ belongs to the **tangent space** T_{ρ_t} to $\mathscr{P}'_\sigma(X)$ in ρ_t. Note that the tangent field realizes

$$\inf_{(\nabla\psi_t):(\rho_t,\nabla\psi_t)_t \text{ solves (CE')}} \int_0^1 \|\nabla\psi_t\|^2_{\hat{\rho}_t}\, dt$$

(see Remark 3.1.31).

In the case of $\mathscr{P}_*(X)$, the tangent space is truly the "space of gradients" as in the continuous setting: $T_\rho = \{\nabla\psi : \psi \in \mathbb{R}^X\}$. In fact, when we defined $A(\sigma)$, we noted that the condition $\psi \in \operatorname{Ran} A(\sigma)$ is equivalent to $\sum_x \psi(x) = 0$, which is readily accomplished by addition of the same constant to all components, thus not modifying $\nabla\psi$.

Now that we can identify T_ρ with the tangent space, $\langle\cdot,\cdot\rangle_{\hat{\rho}}$ can be seen as a Riemannian metric (which is C^r if $\theta \in C^r((0,\infty)^2)$): the scalar product is nondegenerate since, as we noticed in the previous proof, $\|\nabla\psi\|^2_{\hat{\rho}}$ with $\nabla\psi \in T_\rho$ implies $\nabla\psi = 0$.

Theorem 3.2.9 ($\mathcal{W}$ as a Riemannian distance)**.**

1. *If $C_\theta = \infty$ and $\sigma \in \mathscr{P}(X)$, then the Riemannian metric $\langle \cdot, \cdot \rangle_{\hat\rho}$ induces on $\mathscr{P}_\sigma(X)$ the distance $\mathcal{W}$.*

2. *If $C_\theta < \infty$ and θ is concave (like the logarithmic mean), then the Riemannian metric $\langle \cdot, \cdot \rangle_{\hat\rho}$ induces on $\mathscr{P}_*(X)$ the distance $\mathcal{W}$.*

Proof.
$\boxed{1.}$ Let $\rho_0, \rho_1 \in \mathscr{P}_\sigma(X)$. Note that in the infimum that defines $\mathcal{W}$ one can obviously restrict to the curves (ρ_t) such that the quantity to be minimized is finite. Supposing that $C_\theta = \infty$, this implies that $(\rho_t) \subseteq \mathscr{P}'_\sigma(X)$: hence, by the previous Remark, an optimal choice of $\nabla\psi_t$ is the "tangent" one. To sum up,

$$\mathcal{W}(\rho_0, \rho_1)^2 = \inf_{(\rho_t) \subseteq \mathscr{P}_\sigma(X) \text{ a.e. differentiable}} \int_0^1 \|D_t\rho\|^2_{\hat\rho_t}\, dt$$

which is precisely the squared Riemannian distance.
$\boxed{2.}$ Let $\rho_0, \rho_1 \in \mathscr{P}_*(X)$. The argument of part 1. can be repeated literally if we can show that, in the definition of $\mathcal{W}$, we can restrict to the curves (ρ_t) entirely contained in $\mathscr{P}_*(X)$. Let us prove this.

Fix $\varepsilon > 0$, and find $(\rho_t, \Psi_t)_{t\in[0,1]}$ solving (CE') such that

$$\int_0^1 \|\Psi_t\|^2_{\hat\rho_t}\, dt < \mathcal{W}(\rho_0, \rho_1)^2 + \varepsilon\,.$$

Set $\rho^\varepsilon_t := (1-\varepsilon)\rho_t + \varepsilon$, which solves (CE') with velocity field

$$\Psi^\varepsilon_t(x, y) := (1-\varepsilon)\frac{\hat\rho_t(x, y)}{\hat\rho^\varepsilon_t(x, y)}\Psi_t(x, y).$$

The function $\frac{1}{\theta}$ is convex as a composition of the convex nonincreasing $x \mapsto x^{-1}$ with the concave θ. Hence, $(x, s, t) \mapsto \frac{x^2}{\theta(s,t)}$ is convex too, from which

$$2\,\|\Psi^\varepsilon_t\|^2_{\hat\rho^\varepsilon_t} = \sum_{x,y\in X} \frac{\Psi^\varepsilon_t(x,y)^2}{\theta\left(\rho^\varepsilon_t(x), \rho^\varepsilon_t(y)\right)}\hat\rho^\varepsilon_t(x,y)^2 K(x,y)\pi(x) \overset{\text{convexity}}{\leq}$$

$$\leq \sum_{x,y\in X} (1-\varepsilon)\frac{\left[\frac{\hat\rho_t(x,y)}{\hat\rho^\varepsilon_t(x,y)}\Psi_t(x,y)\right]^2}{\theta\left(\rho_t(x), \rho_t(y)\right)}\hat\rho^\varepsilon_t(x,y)^2 K(x,y)\pi(x) = (1-\varepsilon)2\,\|\Psi_t\|^2_{\hat\rho_t}\,.$$

Therefore,

$$\int_0^1 \|\Psi^\varepsilon_t\|^2_{\hat\rho^\varepsilon_t}\, dt \leq (1-\varepsilon)\int_0^1 \|\Psi_t\|^2_{\hat\rho_t}\, dt < (1-\varepsilon)\mathcal{W}(\rho_0, \rho_1)^2 + \varepsilon\,.$$

Unfortunately, $\rho^\varepsilon_0 \neq \rho_0$ and $\rho^\varepsilon_1 \neq \rho_1$. However, by Corollary 3.2.3 they are "near": precisely, for $i = 0, 1$, if we put

$$\rho^{i,\varepsilon}_t := (1-t)\rho_i + t\rho^\varepsilon_i,$$

then there exists a velocity field $(\Psi_t^{i,\varepsilon})_t$ such that $(\rho_t^{i,\varepsilon}, \Psi_t^{i,\varepsilon})_t$ solves (CE') and

$$\int_0^1 \left\|\Psi_t^{i,\varepsilon}\right\|^2_{\hat\rho_t^{i,\varepsilon}} dt \le C^2 d_{TV}(\rho_i, \rho_i^\varepsilon)^2 \le C' \varepsilon^2$$

for some constants C, C' independent of ε.

Now we simply rescale $\left(\rho^{0,\varepsilon}, \Psi^{0,\varepsilon}\right)$ to be defined in the time interval $[0,\varepsilon]$, $(\rho^\varepsilon, \Psi^\varepsilon)$ to be defined in $[\varepsilon, 1-\varepsilon]$, and $\left(\rho^{1,\varepsilon}, \Psi^{1,\varepsilon}\right)$ to be defined in $[1-\varepsilon, 1]$: gluing the three resulting curves together, we get $(\bar\rho_t^\varepsilon, \bar\Psi_t^\varepsilon)_{t\in[0,1]} \subseteq \mathscr{P}_*(X)$ solving (CE') and connecting ρ_0 to ρ_1, such that

$$\begin{aligned}\int_0^1 \|\bar\Psi_t^\varepsilon\|^2_{\hat\rho_t^\varepsilon} dt &\le \frac{1}{\varepsilon}\int_0^1 \left\|\Psi_t^{0,\varepsilon}\right\|^2_{\hat\rho_t^{0,\varepsilon}} dt + \frac{1}{1-2\varepsilon}\int_0^1 \|\Psi_t^\varepsilon\|^2_{\hat\rho_t^\varepsilon} dt + \frac{1}{\varepsilon}\int_0^1 \left\|\Psi_t^{1,\varepsilon}\right\|^2_{\hat\rho_t^{1,\varepsilon}} dt \\ &\le 2C'\varepsilon + \frac{(1-\varepsilon)W(\rho_0,\rho_1)^2 + \varepsilon}{1-2\varepsilon}.\end{aligned}$$

Since ε is arbitrary, and recalling as usual that we can replace $\bar\Psi_t^\varepsilon$ with a gradient without increasing its norm, the conclusion follows. ☐

With almost the same proof, we can show the following:

Theorem 3.2.10 (restriction to $\rho \in C^\infty$)**.** *Suppose that $\theta \in C^\infty((0,\infty)^2)$ is concave. Then in the infimum that defines $\mathcal{W}$, one can restrict to $(\rho_t) \in C^\infty((0,1);\mathbb{R}^X)$.*

Note. In the original article [12], the distance $\mathcal{W}$ was defined requiring (ρ_t) to be piecewise C^1; in light of the Theorem, this makes no difference at least in the most relevant case of the logarithmic mean.

Proof. Given any $(\rho_t, V_t)_{t\in[0,1]}$ solving the continuity equation, we define a solution of (CE') in $[-\varepsilon, 1+\varepsilon]$ as follows:

$$(\sigma_t^\varepsilon, W_t^\varepsilon) := \begin{cases} (\rho_0, 0)\,, & \text{if } t \in [-\varepsilon, \varepsilon] \\ \left(\rho_{\frac{t-\varepsilon}{1-2\varepsilon}}, \dfrac{1}{1-2\varepsilon} V_{\frac{t-\varepsilon}{1-2\varepsilon}}\right), & \text{if } t \in [\varepsilon, 1-\varepsilon] \\ (\rho_1, 0)\,, & \text{if } t \in [1-\varepsilon, 1+\varepsilon]. \end{cases}$$

It evidently satisfies

$$\int_0^1 \|W_t^\varepsilon\|^2_{\hat\sigma_t^\varepsilon} dt = \frac{1}{1-2\varepsilon}\int_0^1 \|V_t\|^2_{\hat\rho_t} dt. \tag{3.2.3}$$

We now mollify in time: $\rho^\varepsilon := \sigma^\varepsilon * j_\varepsilon$ is an admissible $C^\infty((0,1);\mathbb{R}^X)$ curve between ρ_0 and ρ_1, which satisfies (CE') with velocity field

$$V^\varepsilon(x,y) := \frac{[W^\varepsilon(x,y)\hat\sigma^\varepsilon(x,y)] * j_\varepsilon}{\hat\rho^\varepsilon(x,y)}.$$

Again by convexity of $(x,s,t) \mapsto \frac{x^2}{\theta(s,t)}$, we get

$$\begin{aligned} 2\|V_t^\varepsilon\|^2_{\hat\rho_t^\varepsilon} &= \sum_{x,y\in X} \frac{V_t^\varepsilon(x,y)^2 \hat\rho_t^\varepsilon(x,y)^2}{\theta\left(\rho_t^\varepsilon(x), \rho_t^\varepsilon(y)\right)} K(x,y)\pi(x) \overset{\text{convexity}}{\le} \\ &\le \sum_{x,y\in X}\left[\frac{W^\varepsilon(x,y)^2\hat\sigma^\varepsilon(x,y)^2}{\theta\left(\sigma^\varepsilon(x),\sigma^\varepsilon(y)\right)} * j_\varepsilon\right](t)\ K(x,y)\pi(x). \end{aligned}$$

Integrating in dt,

$$\int_0^1 \|V_t^\varepsilon\|^2_{\hat\rho_t^\varepsilon}\, dt \le \int_0^1 \|W_t^\varepsilon\|^2_{\hat\sigma_t^\varepsilon}\, dt \overset{(3.2.3)}{\le} \frac{1}{1-2\,\varepsilon}\int_0^1 \|V_t\|^2_{\hat\rho_t}\, dt.$$

Finally, we can replace V^ε with a gradient field with the usual projection argument. □

Remark 3.2.11. If in addition one knows that $\rho_0, \rho_1 \in \mathscr{P}_*(X)$, then the above proof in fact shows that we can also require ψ to be C^∞. This is not surprising if one recalls the Riemannian structure of $\mathscr{P}_*(X)$.

Example 3.2.12 ($\mathcal{W}$ not Riemannian on $\mathscr{P}'_\sigma(X)$)**.** When $C_\theta < \infty$, we cannot expect that $\mathcal{W}$ is induced by the Riemannian metric on the smooth manifold $\mathscr{P}'_\sigma(X)$ for general σ, as the present example shows.

Let θ be such that $C_\theta < \infty$. Consider on the three-point space $X := \{a, b, c\}$ a reversible irreducible Markov kernel K in which the roles of a and b are symmetric and $K(x, y) > 0$ if $x \neq y$. Call ρ_0 the probability measure giving mass $\frac{1}{3}$ to a and $\frac{2}{3}$ to b, ρ_1 the probability measure giving mass $\frac{2}{3}$ to a and $\frac{1}{3}$ to b. Take $\sigma := \rho_0$, and call d_σ the distance induced on $\mathscr{P}'_\sigma(X)$ by the Riemannian metric. By definition,

$$d_\sigma(\rho_0, \rho_1) \ge \left(\inf_{\substack{(\rho_t, \nabla\psi_t) \text{ solves (CE')} \\ (\rho_t) \subseteq \mathscr{P}'_\sigma(X)}} \int_0^1 \|\nabla\psi_t\|^2_{\hat\rho_t}\, dt \right)^{1/2}.$$

In our situation, the argument used for the comparison Lemma 3.1.39 can be in some sense reversed:

Claim. Put $p := K(a, b)\pi(a)$. Consider on the 2-point space $Y = \{a, b\}$ the Markov kernel $\bar K$ such that $\bar K(a, b) = \bar K(b, a) = p$, and denote by $\mathcal{W}_{p,p}$ the induced distance between measures on Y; call $\bar\rho_0, \bar\rho_1$ the (densities of the) restrictions to Y of the probability measures on X whose densities are ρ_0, ρ_1. Then there exists a constant $C > 0$, depending only on θ and π, such that

$$\inf_{\substack{(\rho_t, \nabla\psi_t) \text{ solves (CE')} \\ (\rho_t) \subseteq \mathscr{P}'_\sigma(X)}} \int_0^1 \|\nabla\psi_t\|^2_{\hat\rho_t}\, dt \ge C^2 \mathcal{W}_{p,p}(\bar\rho_0, \bar\rho_1)^2.$$

We postpone the proof, which is a modification of the proof of Lemma 3.1.39. Using the claim, we get

$$d_\sigma(\rho_0, \rho_1) \ge C\mathcal{W}_{p,p}(\bar\rho_0, \bar\rho_1).$$

By the explicit formula (3.1.9) for $\mathcal{W}_{p,p}$, this becomes

$$d_\sigma(\rho_0, \rho_1) \ge C \frac{1}{\sqrt{2K(a, b)\pi(a)}} \int_{-1/3}^{1/3} \sqrt{\frac{\operatorname{arctanh} r}{r}}\, dr. \tag{3.2.4}$$

On the other hand, if we call ρ_2 the probability measure giving mass $\frac{1}{3}$ to each point of X, then

$$\mathcal{W}(\rho_0, \rho_1) \le \mathcal{W}(\rho_0, \rho_2) + \mathcal{W}(\rho_2, \rho_1) \overset{\text{symmetry}}{=} 2\mathcal{W}(\rho_1, \rho_2).$$

By comparison with the two point space (Lemma 3.1.39),

$$\mathcal{W}(\rho_1, \rho_2) \le \sqrt{C_d}\mathcal{W}_{p',p'}(\bar\rho_1, \bar\rho_2),$$

where $p' = K(a,c)\pi(a)$, and $\bar\rho_1, \bar\rho_2$ represent measures on $\{a, c\}$ of total mass $\frac{2}{3}$; the first one is concentrated on a, while the second one gives equal mass to the two points. Using the scaling property in Lemma 3.1.36 and then the explicit formula (3.1.9), we have

$$\mathcal{W}_{p',p'}(\bar\rho_1, \bar\rho_2) = \sqrt{\frac{2}{3}} \frac{1}{\sqrt{2K(a,c)\pi(a)}} \int_0^1 \sqrt{\frac{\operatorname{arctanh} r}{r}} dr.$$

To sum up, we have proved the two inequalities

$$d_\sigma(\rho_0, \rho_1) \ge \frac{C}{\sqrt{2K(a,b)\pi(a)}} \int_{-1/3}^{1/3} \sqrt{\frac{\operatorname{arctanh} r}{r}} dr,$$
$$\mathcal{W}(\rho_0, \rho_1) \le 2\sqrt{C_d}\sqrt{\frac{2}{3}} \frac{1}{\sqrt{2K(a,c)\pi(a)}} \int_0^1 \sqrt{\frac{\operatorname{arctanh} r}{r}} dr.$$

Hence, if we choose K in such a way that

$$\frac{C}{\sqrt{2K(a,b)\pi(a)}} \int_{-1/3}^{1/3} \sqrt{\frac{\operatorname{arctanh} r}{r}} dr > 2\sqrt{C_d}\sqrt{\frac{2}{3}} \frac{1}{\sqrt{2K(a,c)\pi(a)}} \int_0^1 \sqrt{\frac{\operatorname{arctanh} r}{r}} dr,$$

then $d_\sigma(\rho_0, \rho_1) > \mathcal{W}(\rho_0, \rho_1)$.

Choose π in any way: for instance, $\pi := (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. The above condition becomes that $K(a,b)$ is smaller than a constant multiple of $K(a,c)$. A kernel K with the chosen π and satisfying this property can be easily built, for instance via the associated weighted graph (see Remark 3.1.19).

Note that the found condition is about relative smallness of $K(a,b)$, which is intuitive: if the probability of the direct transition from a to b is too small, then it is less expensive to pass through c.

Proof of the Claim. Take any $(\rho_t, \nabla\psi_t)$ solving (CE') and such that $\rho_t(x) = 0$ for every $x \notin \{a, b\}$. Define a potential $\bar\psi_t : \{a, b\} \to \mathbb{R}$ satisfying

$$\nabla\bar\psi_t(a,b) = \frac{2\hat\rho_t(a,b)}{\hat{\bar\rho}_t(a,b)} \nabla\psi_t(a,b), \qquad \left(\frac{0}{0} := 0\right).$$

Then $(\bar\rho_t, \nabla\bar\psi_t)$ solves (CE') (immediate verification). We observe that $\int_0^1 \|\nabla\psi_t\|^2_{\hat\rho_t}\,dt$ equals

$$\frac{1}{2}\int_0^1 [\nabla\psi_t(a,b)^2 \hat\rho_t(a,b) K(a,b)\pi(a) + \nabla\psi_t(b,a)^2 \hat\rho_t(b,a) K(b,a)\pi(b)]dt =$$
$$= \frac{1}{4}\int_0^1 \frac{\hat{\bar\rho}_t(a,b)^2}{\hat\rho_t(a,b)} \left(\bar\psi_t(b) - \bar\psi_t(a)\right)^2 p\ dt. \quad (3.2.5)$$

The monotonicity and doubling properties of θ imply that, if k is a positive integer such that $2\pi(a), 2\pi(b) \ge 2^{-k}$, then

$$\hat{\bar\rho}_t(a,b) = \theta\left(2\pi(a)\rho_t(a), 2\pi(b)\rho_t(b)\right) \ge \theta\left(2^{-k}\rho_t(a), 2^{-k}\rho_t(b)\right) \ge (2C_d)^{-k}\ \hat\rho_t(a,b).$$

Using this to eliminate $\hat\rho_t$ in (3.2.5), and noting that $p = 2\bar\pi(a)\bar K(a,b) = 2\bar\pi(b)\bar K(b,a)$, we obtain

$$\int_0^1 \|\nabla\psi_t\|^2_{\hat\rho_t}\,dt \ge \frac{1}{4}(2C_d)^{-k}\int_0^1 \nabla\bar\psi_t(a,b)^2\hat{\bar\rho}_t(a,b)[\bar\pi(a)\bar K(a,b)+\bar\pi(b)\bar K(b,a)]dt =$$
$$= \frac{1}{2}(2C_d)^{-k}\int_0^1 \|\nabla\bar\psi_t\|^2_{\hat{\bar\rho}_t}\,dt \ge C^2\mathcal{W}_{p,p}(\bar\rho_0,\bar\rho_1)^2.$$

□

Remark 3.2.13 ("Differential" geodesics). If either $C_\theta = \infty$, or $\sigma = (1,\dots,1)$ and θ is concave, then we know that $\mathscr{P}_\sigma(X)$ is a Riemannian manifold, so we have the classical differential concept of geodesic. "Differential" geodesics always exist locally; in the case $C_\theta = \infty$ we have also global existence by Hopf-Rinow's theorem, since the space $\mathscr{P}_\sigma(X)$ is complete. From the same theorem, we recover the existence of minimizing geodesics between any couple of points.

For classical geodesics, we can obtain rigorously the "Hamilton-Jacobi"-like equation in (3.1.27):

Theorem 3.2.14 (Equations of geodesics). *Assume that $\theta \in C^1((0,\infty)^2)$. Consider a curve (ρ_t) in $\mathscr{P}_\sigma(X)$, where either $C_\theta = \infty$, or $\sigma = (1,\dots,1)$ and θ is concave. Suppose that $\dot\rho_t$ exists for a.e. t, and let $\psi_t \in \operatorname{Ran} A(\sigma)$ be characterized by $\nabla\psi_t = D_t\rho$. Then (ρ_t) is a (classical) constant speed geodesic with tangent field $(\nabla\psi_t)$ if and only if*

$$\begin{cases}\partial_t\rho_t + \nabla\cdot(\nabla\psi_t \bullet \hat\rho_t) = 0\\ \partial_t\psi_t + \frac{1}{2}\operatorname{Proj}_{\operatorname{Ran} A(\sigma)}\sum_{z\in X}\nabla\psi_t(\cdot,z)^2\partial_1\hat\rho_t(\cdot,z)K(\cdot,z) = 0,\end{cases} \tag{3.2.6}$$

where Proj *denotes orthogonal projection and*

$$\partial_1\hat\rho(x,z) := \begin{cases}\partial_1\theta\ (\rho(x),\rho(z)), & \text{if } \rho(x)\rho(z) > 0\\ 0, & \text{otherwise.}\end{cases}$$

The proof of the theorem uses the **cogeodesic flow**, whose definition is here recalled to fix the notations. The idea is to write the Euler-Lagrange equations of geodesics in Hamiltonian form.

Let (M,g) be a Riemannian manifold, and choose local coordinates y^α. We will use the Einstein summation convention. For every $x \in M$ and $x' \in T_xM$, write $x' := x'^\beta\partial_\beta$; put

$$p(x,x') := g_{\alpha\beta}(x)x'^\beta dy^\alpha \in T^*_x(M).$$

$p(x,x')$ can be intrinsically defined as the 1-form $p(x,x') := g(x)[x',\cdot]$. The inverse transformation maps $x \in M$, $p := p_\beta dy^\beta \in T^*_xM$ into

$$x'(x,p) := g^{\alpha\beta}(x)p_\beta\partial_\alpha \in T_xM.$$

The classical result is that $\rho(t)$ is a constant speed geodesic if and only if, putting $p(t) := p(\rho(t),\dot\rho(t)) := p_\alpha(t)d_{\rho(t)}y^\alpha$, it holds

$$\begin{cases}\dot\rho^\alpha = g^{\alpha\beta}(\rho)p_\beta\\ \dot p_\alpha = -\frac{1}{2}\frac{\partial g^{\beta\gamma}}{\partial y^\alpha}(\rho)p_\beta p_\gamma,\end{cases}$$

where ρ^α are the coordinates of ρ. We note that the first equation simply states that $p(t) = p(\rho(t), \dot\rho(t))$, while the second one can be compactly written using the concept of differential of a function, as:

$$\dot p_\alpha(t)dy^\alpha = -d_{\rho(t)}\left[\frac{1}{2}g_{\beta\gamma}(\cdot)x'^\beta(\cdot, p(t)), x'^\gamma(\cdot, p(t))\right]. \tag{3.2.7}$$

Proof of the Theorem. Since both geodesics and solutions to the system of ODE locally exist and are unique (with suitable initial conditions), it is sufficient to prove that if (ρ_t) is a constant speed geodesic and $(\nabla\psi_t)$ is its tangent field with $\psi_t \in \operatorname{Ran} A(\sigma)$, then (ρ_t, ψ_t) satisfies the system; the first equation is the continuity equation, so we must only verify the second one.

We want to write the second equation of the cogeodesic flow in our context; since our manifold is an open set in an affine subspace $\sigma + \operatorname{Ran} B(\sigma)$ of $\mathbb{R}^X$, we can choose as coordinates an appropriate subset $\{y^\alpha\}_{\alpha\in J}$ of the canonical coordinate functions $\left\{y^\beta\right\}_{\beta\in X}$ of $\mathbb{R}^X$.

Let us denote by B_ρ the restriction of $B(\rho)$ to $\operatorname{Ran} A(\rho)$. By definition of the metric, for all $\rho \in \mathscr{P}_\sigma(X)$ and for every couple of tangent vectors $\rho', \tilde\rho' \in \operatorname{Ran} B(\sigma)$,

$$g_\rho[\rho', \tilde\rho'] = \langle B_\rho^{-1}\rho', A(\rho)B_\rho^{-1}\tilde\rho'\rangle;$$

since $A(\rho) = \Pi B(\rho)$ and $B(\rho)B_\rho^{-1} = Id$, the right hand side equals

$$\langle B_\rho^{-1}\rho', \Pi\tilde\rho'\rangle = \langle \Pi B_\rho^{-1}\rho', \tilde\rho'\rangle.$$

This means that

$$p(\rho, \rho') = \sum_{\beta\in X}\left(\Pi B_\rho^{-1}\rho'\right)(\beta)dy^\beta = \sum_{\beta\in X}\left(B_\rho^{-1}\rho'\right)(\beta)\pi(\beta)dy^\beta$$

as 1-forms *on* $\mathscr{P}_\sigma(X)$. But $\psi_t = B_{\rho_t}^{-1}\dot\rho_t$: hence

$$\sum_{\alpha\in J} p_\alpha(t)dy^\alpha := p(\rho_t, \rho_t') = \sum_{\beta\in X}\psi_t(\beta)\pi(\beta)dy^\beta \qquad \text{on } \mathscr{P}_\sigma(X). \tag{3.2.8}$$

In other terms, $\Pi\psi_t$ is the projection on $\operatorname{Ran} B(\sigma)$ of

$$v^\alpha(t) := \begin{cases} 0, & \text{if } \alpha \notin J \\ p_\alpha(t), & \text{otherwise.}\end{cases}$$

For future use, we restate this as

$$\psi_t = \operatorname{Proj}_{\operatorname{Ran} A(\sigma)}\left[\Pi^{-1}v(t)\right]. \tag{3.2.9}$$

Since on our manifold every y^β is a constant plus a linear combination with constant coefficients of the functions $\{y^\alpha\}_{\alpha\in J}$, differentiation of (3.2.8) with respect to t yields

$$\sum_{\alpha\in J}\dot p_\alpha(t)dy^\alpha = \sum_{\beta\in X}\dot\psi_t(\beta)\pi(\beta)dy^\beta \qquad \text{on } \mathscr{P}_\sigma(X). \tag{3.2.10}$$

It remains to compute the right hand side of (3.2.7). In our context, the expression in square brackets is exactly

$$\frac{1}{4}\sum_{\alpha,z\in X}(\psi_t(\alpha)-\psi_t(z))^2\theta(\rho(\alpha),\rho(z))K(\alpha,z)\pi(\alpha),$$

where ψ_t has to be thought as a function ρ and $p(t)$. By (3.2.9), we see that in fact it is a function of the only $p(t)$ and does not depend on ρ. Thanks to this observation, the right hand side of the (3.2.7) becomes

$$-\frac{1}{2}\left\{\sum_{\alpha,z\in X}(\psi_t(\alpha)-\psi_t(z))^2\,\partial_1\theta\left(\rho_t(\alpha),\rho_t(z)\right)K(\alpha,z)\pi(\alpha)dy^\alpha\right\}\quad \text{on } \mathscr{P}_\sigma(X).$$

Combining this with (3.2.10), we can restate (3.2.7) as the equality of the projections on $\operatorname{Ran} B(\sigma)$ of the two vectors $\Pi\dot{\psi}_t$ and

$$-\Pi\cdot\frac{1}{2}\sum_{z\in X}\nabla\psi_t(\cdot,z)^2\partial_1\hat{\rho}_t(\cdot,z)K(\cdot,z);$$

the former already belongs to $\operatorname{Ran} B(\sigma)$, so the projection operator can be omitted. Applying Π^{-1} to this equality, the conclusion follows. □

Remark 3.2.15 (Projections and gradients). If $\psi\in\mathbb{R}^X$ and $x\sim_\sigma z$, then the value of $\nabla\psi(x,z)$ is unaltered if we substitute ψ with $\operatorname{Proj}_{\operatorname{Ran} A(\sigma)}\psi$. In fact, an orthogonal basis of $\operatorname{Ran} A(\sigma)^\perp$ is given by the elements of $\mathbb{R}^X$ corresponding to indicator functions of connected components of $\operatorname{supp}(\sigma)$. Using this basis, it is easy to verify that

$$\left[\operatorname{Proj}_{\operatorname{Ran} A(\sigma)}\psi\right](x)=\psi(x)-c_x,$$

where c_x depends only on the connected component of x.

Thanks to this observation, to compute the trajectory of a geodesic, one might solve the simplified system

$$\begin{cases}\partial_t\rho_t+\nabla\cdot(\nabla\psi_t\bullet\hat{\rho}_t)=0\\ \partial_t\psi_t+\frac{1}{2}\sum_{z\in X}\nabla\psi_t(\cdot,z)^2\partial_1\hat{\rho}_t(\cdot,z)K(\cdot,z)=0:\end{cases}$$

in fact, our remark implies that then $\left(\rho_t,\operatorname{Proj}_{\operatorname{Ran} A(\sigma)}\psi_t\right)$ solves (3.2.6). Though this is the form in which the Theorem was stated in the original article [12], we preferred equation (3.2.6), because Maas's procedure does not produce the "canonical" tangent field ψ_t: note for instance that the last equation implies that each component of ψ_t has nonpositive derivative, while we know from $\psi_t\in\operatorname{Ran} A(\sigma)$ that the tangent field satisfies $\sum_{x\in X}\psi_t(x)\equiv 0$.

3.3 Differentiable functionals and gradient flows

We now compute the (classical) gradient of the potential and internal energy functionals on $\mathscr{P}_*(X)$, with respect to our metric. This will yield the identification of the heat flow with the gradient flow of entropy.

Definition 3.3.1. Let $V : X \to \mathbb{R}$ be any function. Then the **potential energy** functional $\mathcal{V} : \mathscr{P}(X) \to \mathbb{R}$ is defined by

$$\mathcal{V}(\rho) := \sum_{x\in X} V(x)\rho(x)\pi(x).$$

Proposition 3.3.2 (gradient of $\mathcal{V}$)**.** *The potential energy functional is differentiable on $\mathscr{P}_*(X)$, and* $\operatorname{grad}\mathcal{V} \equiv \nabla V$.

Proof. Since the differentiable structure of $\mathscr{P}_*(X)$ is inherited by $\mathbb{R}^X$, differentiability is clear. Moreover, if $(\rho_t) \subseteq \mathscr{P}_*(X)$ is a C^1 curve with tangent field $\nabla\psi_t$, where $\psi_t \in \operatorname{Ran} A(\rho_t)$, then we have

$$\begin{aligned}\frac{d}{dt}\mathcal{V}(\rho_t) = \sum_{x\in X} V(x)\dot{\rho}_t(x)\pi(x) = \langle V, \dot{\rho}_t\rangle_\pi = -\langle V, \nabla\cdot(\hat{\rho}_t \bullet \nabla\psi_t)\rangle_\pi = \\ = \langle \nabla V, \hat{\rho}_t \bullet \nabla\psi_t\rangle_K = \langle \nabla V, \nabla\psi_t\rangle_{\hat{\rho}_t}.\end{aligned}$$

□

Definition 3.3.3. Let $U \in C^1((0,\infty))$. Then the **internal energy** functional $\mathcal{U} : \mathscr{P}(X) \to \mathbb{R}$ is defined by

$$\mathcal{U}(\rho) := \sum_{x\in X} U(\rho(x))\pi(x).$$

Proposition 3.3.4 (gradient of $\mathcal{U}$)**.** *The potential energy functional is differentiable on $\mathscr{P}_*(X)$, and*

$$\operatorname{grad}\mathcal{U}(\rho) = \nabla(U' \circ \rho) \qquad \forall\rho \in \mathscr{P}_*(X).$$

Proof. Again, differentiability is obvious. Moreover, if $(\rho_t) \subseteq \mathscr{P}_*(X)$ is a C^1 curve with tangent field $\nabla\psi_t$, where $\psi_t \in \operatorname{Ran} A(\rho_t)$, then we have

$$\begin{aligned}\frac{d}{dt}\mathcal{U}(\rho_t) = \sum_{x\in X} U'(\rho_t)\dot{\rho}_t(x)\pi(x) = \langle U'\circ\rho_t, \dot{\rho}_t\rangle_\pi = -\langle U'\circ\rho_t, \nabla\cdot(\hat{\rho}_t \bullet \nabla\psi_t)\rangle_\pi = \\ = \langle \nabla(U'\circ\rho_t), \hat{\rho}_t \bullet \nabla\psi_t\rangle_K = \langle \nabla(U'\circ\rho_t), \nabla\psi_t\rangle_{\hat{\rho}_t}.\end{aligned}$$

□

Remark 3.3.5 (gradients in general $\mathscr{P}_\sigma(X)$)*.* The situation is a little more involved if we compute the gradients in $\mathscr{P}_\sigma(X)$, since it might happen that $\nabla V, \nabla(U'\circ\rho) \notin T_\rho$, in which case they must be projected to obtain the true gradient in $\mathscr{P}_\sigma(X)$.

Since $\mathscr{P}_*(X)$ is Riemannian, and a submanifold of $\mathbb{R}^X$, we can use the simple classical concept of gradient flow:

Definition 3.3.6. $\rho : [0,\infty) \to \mathscr{P}(X)$ is a gradient flow of the differentiable functional $\mathcal{F} : \mathscr{P}_*(X) \to \mathbb{R}$ if and only if it is continuous in total variation, and

$$\rho_t \in \mathscr{P}_*(X), \quad D_t\rho = -\operatorname{grad}\mathcal{F}(\rho_t) \qquad \forall t > 0.$$

Theorem 3.3.7 (Heat flow as gradient flow of entropy)**.** *Let $U \in C^2((0,\infty))$ be such that $U'' > 0$, and suppose that θ is of the form*

$$\theta(r,s) := \begin{cases} 0, & \text{if } rs = 0 \\ \dfrac{s-r}{U'(s)-U'(r)}, & \text{if } r,s > 0, r \neq s \\ [U''(s)]^{-1}, & \text{if } r = s > 0. \end{cases}$$

Assume either that $C_\theta = \infty$ or that θ is concave.

Then the gradient flow of the internal energy functional $\mathcal{U}$ starting from any $\rho_0 \in \mathscr{P}(X)$ is the heat flow $\rho_t = e^{(K-I)t}\rho_0$.

The theorem applies to $U(x) := x \log x$; in this case, θ is the logarithmic mean and $\mathcal{U}$ coincides with the entropy functional $\mathcal{H}$.

Proof. Of course, (ρ_t) is continuous in total variation, and $C^\infty((0,\infty))$. Note that $e^{t(K-I)} = e^{Kt}e^{-t}$, so all its entries are positive by definition of exponential and irreducibility of K. As a consequence, $\rho_t \in \mathscr{P}_*(X)$ for every $t > 0$.

Moreover,

$$\hat{\rho}(x,y) = \frac{\rho(x)-\rho(y)}{U'(\rho(x)) - U'(\rho(y))}$$

(extended by continuity where $\rho(x) = \rho(y)$ or $\rho(x)\rho(y) = 0$): therefore

$$\Delta\rho = \nabla \cdot (\nabla\rho) = \nabla \cdot \left(\hat{\rho} \bullet \nabla(U' \circ \rho)\right).$$

Since $\rho_t = \Delta\rho_t$ by definition of heat flow, then the above equality yields

$$\dot{\rho}_t - \nabla \cdot \left(\hat{\rho}_t \bullet \nabla(U' \circ \rho_t)\right) = 0$$

which is what we wanted, in light of Proposition 3.3.4.

The second part of the theorem is an obvious verification. $\square$

Chapter 4

Conclusions

The similarities between the two theories developed in Chapters 2 and 3 are evident. Since one of the main purposes of this thesis is to facilitate the comparison, we put here a "conversion table" to help the reader find which is the discrete counterpart of any Euclidean result, and vice versa.

Result	**Euclidean version**	**Discrete version**
Benamou-Brenier's formula	Theorem 2.1.15	Definition 3.1.16
restriction to gradients	Proposition 2.1.18	Proposition 3.1.27
inf is attained	Proposition 2.1.2	Theorem 3.1.28
(Quasi-)arclength reparametrisation	Lemma 1.3.17	Lemma 3.1.22
Existence of geodesics	Theorem 1.3.23	Theorem 3.1.29
Description of AC_2	Theorem 2.1.16	Theorem 3.1.30
Tangent field to a curve	Remark 2.1.20	Remark 3.2.8
Eulerian description of geodesics	Remark 2.1.3	Theorem 3.2.14
formal variational proof	Remark 2.1.24	Remark 3.1.42
Completeness	Proposition 1.3.5	Proposition 3.2.6
Differentiability of potential energy	Theorem 2.2.22	Proposition 3.3.2
(Sub)differentiability of internal energy	Theorem 2.2.32	Proposition 3.3.4
Heat flow as gradient flow of entropy	Theorem 2.3.11	Theorem 3.3.7

Unfortunately, the way in which the two theories are presently developed looks incompatible with a unified treatment. Nevertheless, such a tight conceptual connection may be a fertile source of inspiration both for the Euclidean and for the discrete setting.

Bibliography

[1] L. Ambrosio. Lecture notes on optimal transport problems. In P. Colli and J. F. Rodrigues, editors, *Mathematical aspects of evolving interfaces*, Lecture notes in mathematics, pages 1–52. Springer-Verlag, Berlin Heidelberg, 2003.

[2] L. Ambrosio and N. Gigli. A user's guide to optimal transport. CVGMT preprint server - http://cvgmt.sns.it/paper/195/, 2011.

[3] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics ETH Zurich. Birkhäuser Verlag, Basel, second edition, 2008.

[4] L. Ambrosio, N. Gigli, and G. Savaré. Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below. arXiv:1106.2090v3 [math.MG], September 2011.

[5] L. Ambrosio, N. Gigli, and G. Savaré. Heat flow and calculus on metric measure spaces with Ricci curvature bounded below - the compact case. arXiv:1205.3288v1 [math.AP], May 2012.

[6] L. Ambrosio, N. Gigli, and G. Savaré. Metric measure spaces with Riemannian Ricci curvature bounded from below. arXiv:1106.2090v3 [math.MG], January 2012.

[7] W. J. Anderson. *Continuous-time Markov chains: an applications-oriented approach.* Springer-Verlag, New York, 1991.

[8] M. Erbar and J. Maas. Ricci curvature of finite Markov chains via convexity of the entropy. arXiv 1111.2687v1 [Math.MG], November 2011.

[9] L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics.* American Mathematical Society, 2nd edition, 2010.

[10] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions.* Studies in advanced mathematics. CRC press, Boca Raton, FL, 1992.

[11] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, January 1998.

[12] J. Maas. Gradient flows of the entropy for finite Markov chains. *Journal of Functional Analysis*, 261(8):2250–2292, 2011.

[13] R. J. McCann. Polar factorization of maps on Riemannian manifolds. *Geometric and Functional Analysis*, 11(3):589–608, 2001.

[14] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Communications in Partial Differential Equations*, 26:101–174, 2001.

[15] D. W. Stroock. *Probability theory: an analytic view.* Cambridge University Press, New York, 2nd edition, 2011.

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

[17] C. Villani. *Optimal transport: old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften.* Springer-Verlag, Berlin, 2009.

Acknowledgements

I would like to thank Prof. Luigi Ambrosio for the patience, willingness and competence with which he guided me throughout the preparation of this thesis, read the manuscript and suggested corrections.

I also thank Prof. Jan Maas for kindly answering my questions about his paper.

This text is only the top of the iceberg of my university studies, so I desire to express my gratitude also to those who, though not directly involved in the writing of this thesis, contributed crucially to the success of this adventure.

In particular, special thanks go to the Carozza family, whose support went beyond the economical side. Thanks to Anna Subitoni and all Comunità Nuova, who teach all of us how helping the neighbour can make a better world, also with their smaller-size initiatives, like the Edelweiss project I was so lucky to be chosen for.

Thanks to Eleonora Bardelli, Maria Colombo and Denis Nardin, because my understanding of (and enthusiasm about) mathematics would have been much poorer without their help. I regret wasting too often the possibility of such a brilliant and pleasant company, not only in study, because of my perhaps too busy lifestyle. With them, I thank all those whose friendship made life in Pisa in these five years so great.

Finally, thanks to my family, who is always near, also when I am geographically far.