Università degli Studi di Pavia

Joint Ph.D. Program in Mathematics Milano Bicocca-Pavia-INdAM

Doctoral Thesis

Optimal transport: entropic regularizations, geometry and diffusion PDEs

Author:
Nicolò De Ponti

Supervisor:
Prof. Giuseppe Savaré

UNIVERSITÀ DEGLI STUDI DI PAVIA

Abstract

Dipartimento di Matematica "F. Casorati"

Optimal transport: entropic regularizations, geometry and diffusion PDEs

by Nicolò De Ponti

The thesis is divided in three main parts:

In the first part we introduce the class of optimal Entropy-Transport problems, a recent generalization of optimal transport problems where also creation and destruction of mass is taken into account. We focus in particular on the metric properties of these problems, computed in terms of an entropy function F and a cost function. Starting from the power-like entropy $F(s) = (s^p - p(s-1) - 1)/(p(p-1))$ and a suitable cost depending on a metric d on a space X, our main result ensures that for every $p > 1$ the related Entropy-Transport cost induces a distance on the space of finite measures over X. Inspired by previous work of Gromov and Sturm, we then use these Entropy-Transport metrics to construct new complete and separable distances on the family of metric measure spaces with finite mass.
We also study in detail the pure entropic setting, that can be recovered as a particular case when the transport is forbidden. In this situation, corresponding to the classical theory of Csiszár F-divergences, we analyse some structural properties of these entropic functionals and we highlight the important role played by the class of Matusita's divergences.

The second part is devoted to the study of bounds involving Cheeger's isoperimetric constant h and the first eigenvalue λ_1 of the Laplacian.
A celebrated lower bound of λ_1 in terms of h, $\lambda_1 \geq h^2/4$, was proved by Cheeger in 1970 for smooth Riemannian manifolds. An upper bound on λ_1 in terms of h was established by Buser in 1982 (with dimensional constants) and improved (to a dimension-free estimate) by Ledoux in 2004 for smooth Riemannian manifolds with Ricci curvature bounded below.
The goal of this part is two fold. First: by adapting the approach of Ledoux via heat semigroup techniques, we sharpen the inequalities obtaining a dimension-free sharp Buser inequality for spaces with (Bakry-Émery weighted) Ricci curvature bounded below by $K \in \mathbb{R}$ (the inequality is sharp for $K > 0$ as equality is obtained on the Gaussian space). Second: all of our results hold in the higher generality of (possibly non-smooth) metric measure spaces with Ricci curvature bounded below in synthetic sense, the so-called $\mathsf{RCD}(K, \infty)$ spaces.

In the third part, given a complete, connected Riemannian manifold $\mathbb{M}^n$ with Ricci curvature bounded from below, we discuss the stability of the solutions of a porous medium-type equation with respect to the 2-Wasserstein distance. We produce (sharp) stability estimates under negative curvature bounds, which to some extent generalize well-known results by Sturm and Otto-Westdickenberg.
The strategy of the proof mainly relies on a quantitative L^1–L^∞ smoothing property of the equation considered, combined with the Hamiltonian approach developed by Ambrosio, Mondino and Savaré in a metric-measure setting.

Acknowledgements

My first thank must be addressed to Prof. Giuseppe Savaré, for the precious support during my Ph.D. studies. I am indebted with him for many fruitful discussions, for his valuable suggestions and for his remarkable generosity.
I wish to acknowledge Prof. Andrea Mondino, who accepted me as a visitor at University of Warwick and provided me the great opportunity to work with him. His contagious enthusiasm for the research is very inspiring for me.
I am deeply grateful to Matteo Muratori and Carlo Orrieri for the stimulating discussions and the nice time we had together.
I warmly thank Fabio Cavalletti who kindly invited me at SISSA where we had interesting mathematical conversations.
I thank the Ph.D. and postdoc students in Pavia for the sharing of three years of studies and for all the fun we had.
Last but not the least, I would like to thank my parents for their love and support.

Contents

List of Symbols

$\Gamma_0(\mathbb{R}_+)$	class of admissible entropy functions, see Section 2.1
$\hat{F}$	perspective function induced by F
$r \wedge t, r \vee t$	minimum and maximum between the numbers r and t
$\mathfrak{M}_p(r,t)$	p-power mean between r and t
d	(pseudo-)metric on a space X
$\mathfrak{C}(X)$	cone over the space X
$\mathcal{C}(X), \mathcal{C}_b(X)$	continuous (and bounded) real-valued functions from X
$\mathrm{Lip}(X), \mathrm{Lip}_{bs}(X)$	Lipschitz real-valued functions (with bounded support) from X
$\mathcal{C}([0,1];X)$	continuous curves from $[0,1]$ with values in the metric space X
$\mathrm{Lip}([0,1];X)$	Lipschitz curves from $[0,1]$ with values in the metric space X
$\mathscr{P}(X), \mathscr{P}_2(X)$	Borel probability measures (with finite quadratic moment) on X
$\mathscr{M}(X), \mathscr{M}_2^M(X)$	nonnegative Borel measures (with mass M and finite quadratic moment) on X
$\mathcal{W}_p$	p-Kantorovich-Wasserstein distance
D_F	F-divergence, see (3.1.1)
H_F	marginal perspective function, see Section 3.2.5
$\boldsymbol{c}$	cost function in Entropy-Transport problems
ET	Entropy-Transport cost
$\mathbf{D}_{\mathrm{ET}}$	Sturm-Entropy-Transport distance, see Definition 5.3.2
$\mathsf{RCD}(K,\infty)$	class of metric measure spaces with Ricci bounded from below by the constant K, see definition 2.6.3
$\lvert\nabla f\rvert$	slope of the Lipschitz function f
Δf	Laplacian of the function f
λ_1, λ_0	first eigenvalues of the Laplacian, as defined in (6.1.1), (6.1.2)
$h(X)$	Cheeger constant of the metric measure space X
$\mathbb{M}^n$	complete, connected, n-dimensional Riemannian manifold
$\mathcal{V}$	Riemannian volume measure on $\mathbb{M}^n$
Ric_x	Ricci curvature at $x \in \mathbb{M}^n$
T_x	tangent space at $x \in \mathbb{M}^n$
$\exp_x v$	exponential map at $x \in \mathbb{M}^n$ along $v \in T_x\mathbb{M}^n$
$\mathbb{H}^n_K$	n-dimensional hyperbolic space with sectional curvature $-K$
$\Gamma, \Gamma_2, \boldsymbol{\Gamma}_2$	classical, iterated and nonlocal *carré du champ* operator, respectively, see Section 2.6.1
$\mathcal{E}_\rho[f]$	weighted Dirichlet energy (or Hamiltonian functional), see (2.6.21)
$\mathcal{E}^*_\rho[\ell]$	dual of the Hamiltonian functional, see (2.6.22)
$Q_s\varphi$	Hopf-Lax semigroup starting from φ, see (2.5.12)

Chapter 1

Introduction

The theory of optimal transport started with a problem proposed by Gaspard Monge in 1781 in a famous report submitted to the *Académie of Sciences* [Mon81]: assume you have a certain amount of goods that have to be moved from a given initial position to a given final position. Where do you send each unit of material in order to minimize the cost of the transport? In a modern mathematical language, the problem can be stated as follows: let $\mu_1, \mu_2 \in \mathscr{P}(X)$ be two probability measures over a space X and let $\boldsymbol{c} : X \times X \to [0, \infty]$ be a cost function. Find a map $f : X \to X$ pushing the first measure onto the second, i.e. $\mu_2(A) = \mu_1(f^{-1}(A))$ for every measurable set A, and minimizing the quantity

$$\int_{X\times X} \boldsymbol{c}(x, f(x))\mathrm{d}\mu_1(x).$$

Monge developed a fine analysis of the problem and in particular of the geometric properties of the solutions, but the question of the existence of a minimizer was only addressed in the 1940s by Kantorovich [Kan42; Kan48]: instead of looking to maps, he proposed the study of the following minimization problem

$$\mathrm{T}(\mu_1, \mu_2) := \inf_{\boldsymbol{\gamma}} \int_{X\times X} \boldsymbol{c}(x_1, x_2)\mathrm{d}\boldsymbol{\gamma}(x_1, x_2), \tag{1.0.1}$$

where the infimum is taken with respect to every measure $\boldsymbol{\gamma} \in \mathscr{P}(X \times X)$ such that $\pi^i_\sharp\boldsymbol{\gamma} = \mu_i$, where π^i denotes the projection map $\pi^i : X \times X \to X$, $\pi^i(x_1, x_2) = x_i$, $i = 1, 2$. In other terms, $\boldsymbol{\gamma}$ is a so-called *transport plan*, i.e. a probability measure on the product space $X \times X$ such that

$$\boldsymbol{\gamma}(A\times X) = \mu_1(A), \quad \boldsymbol{\gamma}(X\times B) = \mu_2(B) \quad \text{for every measurable set } A, B \subset X. \tag{1.0.2}$$

Under this formulation, Kantorovich not only proved that the infimum can be replaced by a minimum for a general class of cost functions, but also stated and proved a fundamental duality result.

Despite these important contributions, only in the 1990s, following a seminal paper by Brenier [Bre91], optimal transport was rediscovered under a different light by mathematicians and since then it has gained a lot of attention in different areas, such as probability, applied mathematics, geometry, partial differential equations. We refer to the monographs [Vil03; AGS08; Vil09; San15] for a complete overview of the theory.

One of the most important feature of optimal transport is that it can be used to construct relevant distances on the space of probabilities. Indeed, let us consider a metric space X (say complete and separable), and a cost function induced by the metric d. A typical choice is $\boldsymbol{c} = \mathsf{d}^2$: in this situation it is well-known that the square root of the optimal transport cost T is a distance on the space of probability measures (with finite second moment) over X, called 2-*Wasserstein* distance $\mathcal{W}_2$. It inherits many properties of the base space X: it is a complete and separable distance that quantifies the spatial shift between the supports of two measures. Moreover, $\mathcal{W}_2$ metrizes the weak convergence in duality with continuous and bounded functions. All these properties make the distances coming from optimal transport theory very useful in several applications. In the present thesis we focus on three different aspects:

- One of the major restrictions of the classical optimal transport problem is the fact that it requires the two input measures to have the same total mass. In the first part of the thesis, we see how an entropic regularization allows to overcome this issue and we develop a study of the distances coming from optimal *Entropy-Transport* problems. In doing so, we also give new insights on the theory of Csiszár F-divergences.
- The second part of the thesis is devoted to the study of famous bounds involving the first positive eigenvalue of the Laplacian and the Cheeger constant. By means of heat semigroup techniques, we sharpen and generalize these bounds to the class of $\mathsf{RCD}(K,\infty)$ spaces. Roughly speaking, an $\mathsf{RCD}(K,\infty)$ space is a (possibly infinite-dimensional, possibly non-smooth) metric measure space with Ricci curvature bounded from below in a synthetic sense. Here, the 2-Wasserstein distance enters already in the definition of this notion of Ricci curvature bounds and its geometric properties play a crucial role in the theory.
- In the last part, we study the stability with respect to the 2-Wasserstein distance of the solutions of the porous medium equation on Riemannian manifolds with Ricci curvature bounded from below. We produce a sharp control of the $\mathcal{W}_2$-distance between the solutions of the equation in terms of the $\mathcal{W}_2$-distance of the corresponding initial data, generalizing on possibly negatively curved spaces some celebrated results by Sturm and Otto-Westdickenberg.

We now discuss in detail the three parts of the thesis:

1.1 First part

Let $F:[0,+\infty)\to[0,+\infty)$ be a superlinear, convex function such that $F(1)=0$. A central role in the following chapters is played by the F*-divergences*, i.e. functionals $D_F:\mathscr{M}(X)\times\mathscr{M}(X)\to[0,+\infty]$ of the form

$$D_F(\gamma||\mu):=\begin{cases}\int_X F\big(\frac{\mathrm{d}\gamma}{\mathrm{d}\mu}\big)\mathrm{d}\mu & \text{if } \gamma\ll\mu,\\ +\infty & \text{otherwise.}\end{cases} \tag{1.1.1}$$

Here $\mathscr{M}(X)$ denotes the set of finite, nonnegative, Borel measures over a Polish space (X,d), i.e. a complete and separable metric space.

A classical example of F-divergence is given by the possible choice $F=U_1(s):=s\ln(s)-s+1$, corresponding to the celebrated Kullback-Leibler divergence (also called relative entropy) [KL51], a functional introduced by Kullback and Leibler in 1951 intimately related to the famous Shannon's entropy (see [Sha48; Lin91]). We notice that the presence of the linear part in the function U_1 is natural in dealing with measures with possibly different total mass (see Remark 3.1.2).

Since their introduction by Csiszár [Csi63], and independently Ali and Silvey [AS66], F-divergences have become a fundamental tool in many areas of mathematics and engineering, including probability, statistics, information theory, signal processing and machine learning. They can be interpreted as a sort of "distance function" on the set of finite measures, even if they do not generally fulfill the symmetric property and the triangle inequality. We refer to Liese and Vajda [LV06; Vaj89] and references therein for a systematic presentation of these functionals, where also their applicability in statistical test is discussed.

Recently, F-divergences have been considered by Liero, Mielke, Savaré [LMS18a] as penalizing functionals in the formulation of optimal *Entropy-Transport* problems, a generalization of optimal transport problems obtained by relaxing the marginal constraints. In contrast with the classical transport setting, the theory that has been developed allows the description of phenomena where the conservation of mass may not hold, and for this reason in the literature it is also referred to as "unbalanced optimal transport". The full theory presented in [LMS18a] (see also [LMS16]) follows the same line of development of classical optimal transport theory and includes: a primal formulation of the problem; a dual formulation where also optimality conditions are discussed; the study of a particular case of optimal Entropy-Transport problem that generates the so-called "Hellinger-Kantorovich" distance HK, a complete and separable distance on $\mathscr{M}(X)$ that admits a dynamic characterization corresponding to the classical Benamou-Brenier formulation of the 2-Wasserstein distance (see [BB00]). This dynamic approach was also considered in two other parallel work: in [KMV16] and [Chi+18] the authors derived the Hellinger-Kantorovich distance by studying Lagrangian action in duality with the *non-conservative continuity equation*

$$\partial_t \rho + \nabla \cdot \omega = \xi. \tag{1.1.2}$$

In the present thesis we focus instead on the static formulation of optimal Entropy-Transport problems, which works as follows: starting from a cost function $\boldsymbol{c} : X \times X \to [0, +\infty]$, we define the Transport functional as

$$\mathcal{T}(\boldsymbol{\gamma}) := \int_{X\times X} \boldsymbol{c}(x_1, x_2) \mathrm{d}\boldsymbol{\gamma}(x_1, x_2), \qquad \boldsymbol{\gamma} \in \mathscr{M}(X \times X), \tag{1.1.3}$$

and the Entropy-Transport functional as

$$\mathcal{ET}(\boldsymbol{\gamma}||\mu_1, \mu_2) := D_F(\gamma_1||\mu_1) + D_F(\gamma_2||\mu_2) + \mathcal{T}(\boldsymbol{\gamma}), \tag{1.1.4}$$

where $\gamma_i := (\pi^i)_\sharp \boldsymbol{\gamma}$ are the marginals of the measure $\boldsymbol{\gamma}$ obtained as the push-forward through the projection maps $\pi^i(x_1, x_2) = x_i$, $i = 1, 2$. The *Entropy-Transport problem* between the measures μ_1 and μ_2 is the minimization problem

$$\mathsf{ET}(\mu_1, \mu_2) := \inf_{\boldsymbol{\gamma} \in \mathscr{M}(X \times X)} \mathcal{ET}(\boldsymbol{\gamma}||\mu_1, \mu_2). \tag{1.1.5}$$

The aim of the first part of the thesis is to investigate the metric properties of the function $\mathsf{ET} : \mathscr{M}(X) \times \mathscr{M}(X) \to [0, +\infty]$ with respect to different possible choices of the function F and the cost $\boldsymbol{c}$. As a result of our study, we will not only produce new Entropy-Transport distances on $\mathscr{M}(X)$ beside the above-mentioned Hellinger-Kantorovich distance (and the related Gaussian Hellinger-Kantorovich distance), but we also give new insights on the metric properties of the F-divergences. Moreover, by taking advantage of a construction due to Gromov and Sturm, we show how these distances on the space of finite measures can be lift to distances between metric measure spaces.

The starting point of our analysis is the *marginal perspective cost*

$$H(x_1, r; x_2, t) := \inf_{\theta > 0} F\Big(\frac{\theta}{r}\Big) r + F\Big(\frac{\theta}{t}\Big) t + \theta \boldsymbol{c}(x_1, x_2) = \mathsf{ET}(r\delta_{x_1}, t\delta_{x_2}), \tag{1.1.6}$$

a function corresponding to the solution of the Entropy-Transport problem between two Dirac masses $\mu_1 = r\delta_{x_1}$, $\mu_2 = t\delta_{x_2}$. The function H allows an "homogeneous formulation" of the Entropy-Transport problem (see Proposition 3.2.8) which is crucial in order to show that ET is a (power) of a distance on $\mathscr{M}(X)$ if the function H is a (power) of a distance on the cone space over X (see Corollary 4.2.4). The latter is the space $\mathfrak{C}(X) = Y/\sim$, where $Y = X \times [0, +\infty)$ and

$$(x_1, r) \sim (x_2, t) \iff r = t = 0 \text{ or } r = t,\, x_1 = x_2\,.$$

When the starting entropy $F(s)$ has a strict minimum at $s = 1$, and the cost c is a symmetric function such that $c(x_1, x_2) = 0$ if and only if $x_1 = x_2$, we show that the induced marginal perspective cost H is symmetric, nonnegative and $H(x_1, r; x_2, t) = 0$ if and only if $(x_1, r) \sim (x_2, t)$.

The question regarding the validity of the triangle inequality is much more challenging: first of all, in the presence of a non-trivial cost function $\boldsymbol{c}$, an explicit computation of the induced marginal perspective cost is often unavailable. In addition, the triangle inequality is known to be difficult to prove, even in the easier pure entropic case that has been previously considered in the literature (see below). In this part of the thesis, we address the question for an important class of Entropy-Transport problems, the ones generated by the choices $\boldsymbol{c}(x_1, x_2) = f(\mathsf{d}(x_1, x_2))$ for a suitable function $f : [0, +\infty) \to [0, +\infty]$, and $F = U_p$, where the latter corresponds to the class of power-like entropies defined by

$$U_1(s) := s\ln(s) - s + 1, \quad U_p(s) := \frac{1}{p(p-1)}(s^p - p(s-1) - 1) \ \text{ if } p > 1.$$

In this situation, the induced marginal perspective cost H takes the form

$$H = H_p(x_1, r; x_2, t) := \frac{2}{p}\left[\mathfrak{M}_1(r,t) - \mathfrak{M}_{1-p}(r,t)\left(1 + (1-p)\frac{\boldsymbol{c}(x_1,x_2)}{2}\right)_+^{\frac{p}{p-1}}\right] \quad p > 1,$$
$$H = H_1(x_1, r; x_2, t) := 2\Big[\mathfrak{M}_1(r,t) - \mathfrak{M}_0(r,t)e^{-\boldsymbol{c}(x_1,x_2)/2}\Big],$$

where the expressions are written in the terms of the power means

$$\mathfrak{M}_p(r,t) := \left(\frac{r^p + t^p}{2}\right)^{\frac{1}{p}},\ p \neq 0, \qquad \mathfrak{M}_0(r,t) := \sqrt{rt}.$$

In Theorem 4.2.8 and Theorem 4.2.14, which are our main results, we prove that the square root of the induced marginal perspective cost H_p is a distance on $\mathfrak{C}(X)$ for every $p > 1$, where $\boldsymbol{c}$ is given by one of the two following cost functions:

1. $\boldsymbol{c} = \boldsymbol{c}_p(x_1, x_2) := \frac{2}{p-1}\left[1 - \left(\cos(\mathsf{d}(x_1,x_2) \wedge \pi/2)\right)^{\frac{p-1}{p}}\right]$,
2. $\boldsymbol{c}(x_1, x_2) = \mathsf{d}(x_1, x_2)$.

The same result is obtained also for the cost $\boldsymbol{c}(x_1, x_2) = \mathsf{d}^2(x_1, x_2)$ and $1 < p \leq 3$.

Thus, we provide an entire class of Entropy-Transport distances on the space $\mathscr{M}(X)$, besides the Gaussian Hellinger-Kantorovich distance and the related Hellinger-Kantorovich distance studied in [LMS18a], that correspond to the case

$$p = 1, \quad \boldsymbol{c}(x_1, x_2) = \mathsf{d}^2(x_1, x_2)$$

and

$$p = 1, \quad \boldsymbol{c} = \boldsymbol{c}_{\mathrm{HK}}(x_1, x_2) := \begin{cases} -\log\left(\cos^2(\mathsf{d}(x_1, x_2)\right) & \text{if } \mathsf{d}(x_1, x_2) < \pi/2, \\ +\infty & \text{otherwise,} \end{cases}$$

respectively. Notice here that the choice of the "exotic" cost $\boldsymbol{c}_p$ defined above is motivated by its counterpart $\boldsymbol{c}_{\mathrm{HK}}$; in particular, it holds

$$\lim_{p \to 1} \boldsymbol{c}_p(x_1, x_2) = \boldsymbol{c}_{\mathrm{HK}}(x_1, x_2).$$

The class of distances we studied includes, for $p = 2$, a transport variant of the Vincze-Le Cam distance [Vin81; Cam86].

In contrast with the case $p = 1$, where Liero, Mielke and Savaré were able to take advantage of the expression of the induced marginal perspective cost, closely connected with the "natural" metric $\mathsf{d}_{\mathfrak{C}}$ of the cone space

$$\mathsf{d}_{\mathfrak{C}}^2\big((x_1, r), (x_2, t)\big) := r^2 + t^2 - 2rt\cos(\mathsf{d}(x_1, x_2) \wedge \pi),$$

the proof of Theorem 4.2.8 is based on a careful case by case inspection, where we adapt different results already present in the literature ([Ost96; ES03; Kou14]) as well as employ new techniques in order to compare the function H_p with the "model case" $p = 1$. Once that is done, the proof of Theorem 4.2.14 follows by some explicit computations, taking advantage of a well-known technical lemma (Lemma 4.1.2).

If $F = U_p$, we also show that for every cost function $\boldsymbol{c}$ it holds

$$H_p \le H_1 \le pH_p \qquad p \ge 1. \tag{1.1.7}$$

The explicit bounds (1.1.7) allow us to prove that all the Entropy-Transport distances previously considered are complete and separable metrics on $\mathscr{M}(X)$ inducing the weak topology.

Another important part of the work is devoted to the study of the problem (1.1.5) and the property of the function (1.1.6) in the *pure entropic case*, which correspond to the choice

$$\boldsymbol{c}(x_1, x_2) = \begin{cases} 0 & \text{if } x_1 = x_2, \\ +\infty & \text{otherwise.} \end{cases} \tag{1.1.8}$$

In this situation, the structure of the problem does not allow the spatial movement of the mass and we obtain

$$\mathrm{ET}(\mu_1, \mu_2) = \int_X H_F\Big(\frac{\mathrm{d}\mu_1}{\mathrm{d}\lambda}, \frac{\mathrm{d}\mu_2}{\mathrm{d}\lambda}\Big)\mathrm{d}\lambda, \tag{1.1.9}$$

where $\lambda \in \mathscr{M}(X)$ is any dominating measure of μ_1, μ_2, i.e. μ_1 and μ_2 are absolutely continuous with respect to λ. Here H_F is defined by

$$H_F(r, t) := \inf_{\theta > 0} F\Big(\frac{\theta}{r}\Big)r + F\Big(\frac{\theta}{t}\Big)t, \tag{1.1.10}$$

and we notice the close connection with the marginal perspective cost (1.1.6).

We refer to this setting as "pure entropic" since the expression (1.1.9) is an equivalent formulation of the divergence induced by the function $f(s) := H_F(s, 1)$ (see Lemma 3.1.1). In particular, the minimizing procedure (1.1.10) corresponds to the construction

$$D_f(\mu_1||\mu_2) := \inf_{\gamma\in\mathscr{M}(X)} D_F(\gamma||\mu_1) + D_F(\gamma||\mu_2)$$

at the level of divergences, and it can be seen as a simple variational way to generate a new *symmetric* divergence starting from a given convex function F.

The metric properties of F-divergences, and in particular the validity of the triangle inequality, have been investigated by many authors like Csiszár, Endres, Kafka, Osterreicher, Schindelin, Vincze ([Csi67; ES03; KOV91; Ost96; OV03]), to cite only a few. However, to the author's knowledge, the construction of the function H_F defined in (1.1.10) has never been considered in the literature and we will show that it exhibits interesting structural properties.

In contrast with the spatial inhomogeneous case, the explicit expression of the function H_F is easier to obtain and gives rise to well-known statistical functionals, which include the Hellinger distance [Hel09], the Jensen-Shannon divergence [Lin91] and more generally a class of Arimoto-type divergences that has been studied by Osterreicher and Vajda [Ost96; OV03] (see Example 4). One can also obtain other important functionals, including the symmetric Kullback-Leibler divergence [KL51] (see Example 5) and the class of Matusita's divergences (see Example 3).

Regarding the power-like entropies U_p, we prove that the induced function H_{U_p} is the square of a distance on $\mathbb{R}_+$ for every $p \geq 1$. This completes a study started by Osterreicher and Vajda [Ost96; OV03], who considered the class of functions corresponding to the case $p < 1$ (see Example 4 and Proposition 4.1.3).

Our analysis does not limit to superlinear entropy functions. This is particularly important because, by directly studying the minimizing procedure (1.1.10), we show that the only distances between F-divergences are provided by the family of the total variation divergences related to the function $F(s) = c|s-1|$, $c \in (0,+\infty)$. Some general results are proved also for the more difficult case of the F-divergences that induce a distance on $\mathscr{M}(X)$ of the form

$$(\gamma,\mu) \mapsto D_F(\gamma||\mu)^a \quad \text{for a power } a \in (0,1).$$

Here, we will emphasize the central role played by the class of Matusita's divergences $F(s) = |s^a - 1|^{\frac{1}{a}}$ [Mat64].

In the last chapter of the first part of the thesis we show an application of our study on the Entropy-Transport distances.

From the seminal work of Gromov (see [Gro99, Chapter $3\frac{1}{2}$]), it is well understood the importance of studying "geometric distances", as well as the induced notions of convergence, on the family of metric measure spaces. The theoretical applications range from concentration of measure [Shi16], Riemannian geometry [Fuk87] and, more generally, the geometry of metric measure spaces with a synthetic notion of curvature bounds [Stu06], stability of the heat flow [GMS15]. The flexibility of the representation of various "objects" as metric measure spaces makes these distances also useful for real-world applications, where the great advancement in data acquisition and storage imposes new challenges in the analysis and comparison of large datasets. In this regard, we mention the work on object matching [MS05; Mém11] and computer vision [SS13].

As we have previously mentioned, the quadratic Wasserstein distance provides a natural distance on the space of probability measures. It is thus remarkable that $\mathcal{W}_2$ can be used to construct a distance between metric measure spaces, as shown by Sturm [Stu06]: given two metric measure spaces $(X_1, \mathsf{d}_1, \mu_1)$, $(X_2, \mathsf{d}_2, \mu_2)$, the Sturm's

D-distance is defined as

$$\mathbf{D}\big((X_1,\mathsf{d}_1,\mu_1),(X_2,\mathsf{d}_2,\mu_2)\big) := \inf \mathcal{W}_2(\psi^1_\sharp\mu_1,\psi^2_\sharp\mu_2), \tag{1.1.11}$$

where the infimum is taken over all metric spaces $(\hat{X},\hat{\mathsf{d}})$ with isometric embeddings $\psi_1 : \mathsf{supp}(\mu_1) \to \hat{X}$ and $\psi_2 : \mathsf{supp}(\mu_2) \to \hat{X}$.

The function **D** turns out to be a complete and separable distance between equivalence classes of normalized metric measure spaces with finite variance, i.e. spaces where the reference measure is a probability measure with finite second moment, and it has been proved to be useful in discussing the stability of curvature bounds, just to mention an example.

However, in many applications the fact that the **D**-distance requires the input spaces to have equal mass is more a drawback of the theory rather than a feature. For instance, in image comparison this constraint does not allow to recognize differences in color intensity. It is thus interesting to replicate the construction of Sturm by replacing the 2-Wasserstein distance used in the right hand side of (1.1.11) with the Hellinger-Kantorovich distance or, more generally, with a *regular* Entropy-Transport distance (see Definition 5.2.1). In Chapter 5 we show how this construction enables to obtain a class of complete and separable distances between equivalence classes of metric measure spaces with finite total mass.

Part of the material contained in Chapters 3 and 4 can be found in the article [De 19].

The content of Chapter 5 is based on a joint collaboration with Andrea Mondino and Giuseppe Savaré and will be part of a future manuscript.

1.2 Second part

Let $(X,\mathsf{d},\mathfrak{m})$ be a metric measure space, i.e. a complete and separable metric space (X,d) endowed with a non-negative Borel measure $\mathfrak{m}$ finite on bounded sets. We denote by $\mathsf{Lip}_{bs}(X)$ the space of real-valued Lipschitz functions with bounded support. Given $f \in \mathsf{Lip}_{bs}(X)$ its slope $|\nabla f|(x)$ at $x \in X$ is defined by

$$|\nabla f|(x) := \limsup_{y\to x} \frac{|f(y)-f(x)|}{\mathsf{d}(y,x)} \tag{1.2.1}$$

with the convention $|\nabla f|(x) = 0$ if x is an isolated point. The first non-trivial eigenvalue of the Laplacian is characterized as follows:

- If $\mathfrak{m}(X) < \infty$, the non-zero constant functions are in $L^2(X,\mathfrak{m})$ and are eigenfunctions of the Laplacian with eigenvalue 0. In this case, the first non-trivial eigenvalue is given by

$$\lambda_1 = \inf\left\{ \frac{\int_X |\nabla f|^2 \mathrm{d}\mathfrak{m}}{\int_X |f|^2 \mathrm{d}\mathfrak{m}} :\ 0 \not\equiv f \in \mathsf{Lip}_{bs}(X),\ \int_X f\,\mathrm{d}\mathfrak{m} = 0 \right\}. \tag{1.2.2}$$

- When $\mathfrak{m}(X) = \infty$, 0 may not be an eigenvalue of the Laplacian and the first eigenvalue is characterized by

$$\lambda_0 = \inf\left\{ \frac{\int_X |\nabla f|^2 \mathrm{d}\mathfrak{m}}{\int_X |f|^2 \mathrm{d}\mathfrak{m}} :\ \ 0 \not\equiv f \in \mathsf{Lip}_{bs}(X) \right\}. \tag{1.2.3}$$

Note that λ_0 may be zero (for instance if $\mathfrak{m}(X) < \infty$ or if $(X, \mathsf{d}, \mathfrak{m})$ is the Euclidean space $\mathbb{R}^d$ with the Lebesgue measure) but there are examples when $\lambda_0 > 0$: for instance in the Hyperbolic plane $\lambda_0 = 1/4$ and more generally on an n-dimensional simply-connected Riemannian manifold with sectional curvatures bounded above by $k < 0$ it holds $\lambda_0 \geq (n-1)^2|k|/4$ (see [McK70]).

Given a Borel subset $A \subset X$ with $\mathfrak{m}(A) < \infty$, the *perimeter* $\mathrm{Per}(A)$ is defined as follows (see for instance [M M03]):

$$\mathrm{Per}(A) := \inf\left\{ \liminf_{n\to\infty} \int_X |\nabla f_n| \mathrm{d}\mathfrak{m} : f_n \in \mathsf{Lip}_{bs}(X), f_n \to \chi_A \text{ in } L^1(X, \mathfrak{m}) \right\}.$$

In 1970, Cheeger [Che70] introduced an isoperimetric constant, now known as *Cheeger constant*, to bound from below the first eigenvalue of the Laplacian. The *Cheeger constant* of the metric measure space $(X, \mathsf{d}, \mathfrak{m})$ is defined by

$$h(X) := \begin{cases} \inf\left\{ \frac{\mathrm{Per}(A)}{\mathfrak{m}(A)} : A \subset X \text{ Borel subset with } \mathfrak{m}(A) \leq \mathfrak{m}(X)/2 \right\} & \text{if } \mathfrak{m}(X) < \infty \\ \inf\left\{ \frac{\mathrm{Per}(A)}{\mathfrak{m}(A)} : A \subset X \text{ Borel subset with } \mathfrak{m}(A) < \infty \right\} & \text{if } \mathfrak{m}(X) = \infty. \end{cases} \tag{1.2.4}$$

The lower bound obtained in [Che70] for compact Riemannian manifolds, now known as *Cheeger inequality*, reads as

$$\lambda_1 \geq \frac{1}{4} h(X)^2. \tag{1.2.5}$$

As proved by Buser [Bus78], the constant 1/4 in (1.2.5) is optimal in the following sense: for any $h > 0$ and $\varepsilon > 0$, there exists a closed (i.e. compact without boundary) two-dimensional Riemannian manifold (M, g) with $h(M) = h$ and such that $\lambda_1 \leq \frac{1}{4} h(M)^2 + \varepsilon$.

The paper [Che70] is in the framework of smooth Riemannian manifolds; however, the stream of arguments (with some care) extends to general metric measure spaces. For the reader's convenience, we give a self-contained proof of (1.2.5) for m.m.s. in the Appendix (see Theorem 6.3.2).

Cheeger's inequality (1.2.5) revealed to be extremely useful in proving lower bounds on the first eigenvalue of the Laplacian in terms of the isoperimetric constant h. It was thus an important discovery by Buser [Bus82] that also an upper bound for λ_1 in terms of h holds, where the inequality explicitly depends on the lower bound on the Ricci curvature of the smooth Riemannian manifold. More precisely, Buser [Bus82] proved that for any compact Riemannian manifold of dimension n and $\mathrm{Ric} \geq K$, $K \leq 0$ it holds

$$\lambda_1 \leq 2\sqrt{-(n-1)K}h + 10h^2. \tag{1.2.6}$$

Note that the constant here is dimension-dependent. For a complete connected Riemannian manifold with $\mathrm{Ric} \geq K$, $K \leq 0$, Ledoux [Led04] remarkably showed that the constant can be chosen to be independent of the dimension:

$$\lambda_1 \leq \max\{6\sqrt{-K}h, 36h^2\}. \tag{1.2.7}$$

The goal of the present work is twofold:

1. The main results (namely Theorem 1.2.1 and Corollary 1.2.2) improve the constants in both the Buser-type inequalities (1.2.6)-(1.2.7) in a way that now the inequality is sharp for $K > 0$ (as equality is attained on the Gaussian space).

2. The inequalities are established in the higher generality of (possibly non-smooth) metric measure spaces satisfying Ricci curvature lower bounds in synthetic sense,

the so-called $\mathsf{RCD}(K,\infty)$ spaces.

For the precise definition of $\mathsf{RCD}(K,\infty)$ space, we refer the reader to Section 2.6. Here let us just recall that the $\mathsf{RCD}(K,\infty)$ condition was introduced by Ambrosio-Gigli-Savaré [AGS14b] (see also [Amb+15]) as a refinement of the $\mathsf{CD}(K,\infty)$ condition due to Lott-Villani [LV09] and Sturm [Stu06]. Roughly, a $\mathsf{CD}(K,\infty)$ space is a (possibly infinite-dimensional, possibly non-smooth) metric measure space with Ricci curvature bounded from below by K, in a synthetic sense. While the $\mathsf{CD}(K,\infty)$ condition allows Finsler structures, the main point of RCD is to reinforce the axiomatization (by asking linearity of the heat flow) in order to rule out Finsler structures and thus isolate the "possibly non-smooth Riemannian structures with Ricci curvature bounded below". It is out of the scopes of this introduction to survey the long list of achievements and results proved for CD and RCD spaces (to this aim, see the Bourbaki seminar [Vil19] and the recent ICM-Proceeding [Amb18]). Let us just mention that a key property of both CD and RCD is the stability under measured Gromov-Hausdorff convergence (or more generally $\mathbf{D}$-convergence of Sturm [Stu06; AGS14b], or even more generally pointed measured Gromov convergence [GMS15]) of metric measure spaces. In particular pointed measured Gromov-Hausdorff limits of Riemannian manifolds with Ricci bounded below, the so-called *Ricci limits*, are examples of (possibly non-smooth) RCD spaces. Let us also recall that weighted Riemannian manifolds with Bakry-Émery Ricci tensor bounded below are also examples of RCD spaces; for instance the Gaussian space $(\mathbb{R}^d, |\cdot|, (2\pi)^{-d/2}e^{-|x|^2/2}\mathrm{d}\mathcal{L}^d(x))$, $1 \le d \in \mathbb{N}$, satisfies $\mathsf{RCD}(1,\infty)$. It is also worth recalling that if $(X,\mathsf{d},\mathfrak{m})$ is an $\mathsf{RCD}(K,\infty)$ space for some $K>0$, then $\mathfrak{m}(X)<\infty$; since scaling the measure by a constant does not affect the synthetic Ricci curvature lower bounds, when $K>0$, without loss of generality one can then assume $\mathfrak{m}(X)=1$.

In order to state our main result, it is convenient to set

$$J_K(t) := \begin{cases} \sqrt{\frac{2}{\pi K}}\arctan\left(\sqrt{e^{2Kt}-1}\right) & \text{if } K>0, \\ \frac{2}{\sqrt{\pi}}\sqrt{t} & \text{if } K=0, \\ \sqrt{-\frac{2}{\pi K}}\operatorname{arctanh}\left(\sqrt{1-e^{2Kt}}\right) & \text{if } K<0. \end{cases} \qquad \forall t>0 \tag{1.2.8}$$

The aim of the second part of the thesis is to prove the following theorem.

Theorem 1.2.1 (Sharp implicit Buser-type inequality for $\mathsf{RCD}(K,\infty)$ spaces)**.** *Let $(X,\mathsf{d},\mathfrak{m})$ be an $\mathsf{RCD}(K,\infty)$ space, for some $K\in\mathbb{R}$.*

- *In case $\mathfrak{m}(X)=1$, then*
$$h(X) \ge \sup_{t>0}\frac{1-e^{-\lambda_1 t}}{J_K(t)}. \tag{1.2.9}$$
The inequality is sharp for $K>0$, as equality is achieved for the Gaussian space $(\mathbb{R}^d, |\cdot|, (2\pi)^{-d/2}e^{-|x|^2/2}\mathrm{d}\mathcal{L}^d(x))$, $1\le d\in\mathbb{N}$.
- *In case $\mathfrak{m}(X)=\infty$, then*
$$h(X) \ge 2\sup_{t>0}\frac{1-e^{-\lambda_0 t}}{J_K(t)}. \tag{1.2.10}$$

Using the expression (1.2.8) of J_K, in the next corollary we obtain more explicit bounds.

Corollary 1.2.2 (Explicit Buser inequality for $\mathsf{RCD}(K,\infty)$ spaces)**.** *Let* $(X,\mathsf{d},\mathfrak{m})$ *be an* $\mathsf{RCD}(K,\infty)$ *space, for some* $K\in\mathbb{R}$.

- *Case* $K>0$. *If* $\frac{K}{\lambda_1}\geq c>0$, *then*

$$\lambda_1 \leq \frac{\pi}{2c}h(X)^2. \tag{1.2.11}$$

The estimate is sharp, as equality is attained on the Gaussian space $(\mathbb{R}^d,|\cdot|,(2\pi)^{-d/2}e^{-|x|^2/2}\mathrm{d}\mathcal{L}^d(x))$, $1\leq d\in\mathbb{N}$, *for which* $K=1,\lambda_1=1,h(X)=(2/\pi)^{1/2}$.

- *Case* $K=0$, $\mathfrak{m}(X)=1$. *It holds*

$$\lambda_1 \ \leq\ \frac{4}{\pi}h(X)^2\inf_{T>0}\frac{T}{(1-e^{-T})^2}\ <\ \pi h(X)^2. \tag{1.2.12}$$

In case $\mathfrak{m}(X)=\infty$, *the estimate* (1.2.12) *holds replacing* λ_1 *with* λ_0 *and* $h(X)$ *with* $h(X)/2$.

- *Case* $K<0$, $\mathfrak{m}(X)=1$. *It holds*

$$\begin{aligned}\lambda_1 &\leq \max\left\{\sqrt{-K}\frac{\sqrt{2}\log\left(e+\sqrt{e^2-1}\right)}{\sqrt{\pi}(1-\frac{1}{e})}h(X),\frac{2\Big(\log\left(e+\sqrt{e^2-1}\right)\Big)^2}{\pi\Big(1-\frac{1}{e}\Big)^2}h(X)^2\right\}\\ &< \max\left\{\frac{21}{10}\sqrt{-K}h(X),\frac{22}{5}h(X)^2\right\}.\end{aligned} \tag{1.2.13}$$

In case $\mathfrak{m}(X)=\infty$, *the estimate* (1.2.13) *holds replacing* λ_1 *with* λ_0 *and* $h(X)$ *with* $h(X)/2$.

Comparison with previous results in the literature

Theorem 1.2.1 and Corollary 1.2.2 improve the known results about Buser-type inequalities in several aspects. First of all the best results obtained before this work are the aforementioned estimates (1.2.6)-(1.2.7) due to Buser [Bus82] and Ledoux [Led04] for smooth complete Riemannian manifolds satisfying $\mathrm{Ric}\geq K$, $K\leq 0$. Let us stress that the constants in Corollary 1.2.2 improve the ones in both (1.2.6)-(1.2.7) and are dimension-free as well. In addition, the improvements of the present work are:

- In case $K>0$, the inequalities (1.2.9) and (1.2.11) are sharp (as equality is attained on the Gaussian space).
- The results hold in the higher generality of (possibly non-smooth) $\mathsf{RCD}(K,\infty)$ spaces.

The proof of Theorem 1.2.1 is inspired by the semi-group approach of Ledoux [Led94; Led04], but it improves upon by using Proposition 6.2.1 in place of:

- A dimension-dependent Li-Yau inequality, in [Led94].
- A weaker version of Proposition 6.2.1 (see [Led04, Lemma 5.1]) analyzed only in case $K\leq 0$, in [Led04].

Theorem 1.2.1 and Corollary 1.2.2 are also the first *upper bounds* in the literature of RCD spaces for the first eigenvalue of the Laplacian. On the other hand, *lower bounds* on the first eigenvalue of the Laplacian have been throughly analyzed in both CD and RCD spaces: the sharp Lichnerowitz spectral gap $\lambda_1 \geq KN/(N-1)$ was proved under the (non-branching) $\mathsf{CD}(K,N)$ condition by Lott-Villani [LV07], under the $\mathsf{RCD}^*(K,N)$ condition by Erbar-Kuwada-Sturm [EKS15], and generalized by Cavalletti and Mondino [CM17b] to a sharp spectral gap for the p-Laplacian for essentially non-branching $\mathsf{CD}^*(K,N)$ spaces involving also an upper bound on the diameter (together with rigidity and almost rigidity statements). Jiang-Zhang [JZ16] independently showed, for $p=2$, that the improved version under an upper diameter bound holds for $\mathsf{RCD}^*(K,N)$. The rigidity of the Lichnerowitz spectral gap for $\mathsf{RCD}^*(K,N)$ spaces, $K>0$, $N \in (1,\infty)$, known as Obata's Theorem was first proved by Ketterer [Ket15]. The rigidity in the Lichnerowitz spectral gap for $\mathsf{RCD}(K,\infty)$ spaces, $K>0$, was recently proved by Gigli-Ketterer-Kuwada-Ohta [Gig+]. Local Poincaré inequalities in the framework of $\mathsf{CD}(K,N)$ and $\mathsf{CD}(K,\infty)$ spaces were proved by Rajala [Raj12]. Finally various lower bounds, together with rigidity and almost rigidity statements for the *Dirichlet first eigenvalue* of the Laplacian, have been proved by Mondino-Semola [MS18] in the framework of CD and RCD spaces. Lower bounds on Cheeger's isoperimetric constant have been obtained for (essentially non-branching) $\mathsf{CD}^*(K,N)$ spaces by Cavalletti-Mondino [CM17a; CM17b; CM18] and for $\mathsf{RCD}(K,\infty)$ spaces ($K>0$) by Ambrosio-Mondino [AM16].

The material in Chapter 6 is based on a joint collaboration with Andrea Mondino. It has been submitted to a scientific journal for publication, the preprint can be found on Arxiv [DM19].

1.3 Third part

In this part we investigate the Cauchy problem for the following *porous medium-type* equation:

$$\begin{cases} \partial_t \rho = \Delta P(\rho) & \text{in } \mathbb{M}^n \times \mathbb{R}^+ , \\ \rho(\cdot,0) = \mu_0 \geq 0 & \text{in } \mathbb{M}^n \times \{0\} , \end{cases} \tag{1.3.1}$$

where μ_0 is a suitable finite, nonnegative Borel measure and P is a nonlinearity whose model case corresponds to $P(\rho)=\rho^m$ with $m>1$, namely the *porous medium equation* (PME for short). Here $\mathbb{M}^n$ is a smooth, complete, connected, n-dimensional ($n \geq 2$) Riemannian manifold endowed with the standard Riemannian distance d and the Riemannian volume measure $\mathcal{V}$. We denote by Δ the Laplace-Beltrami operator on $\mathbb{M}^n$, which hereafter for simplicity will mostly be referred to as the "Laplacian". The initial datum μ_0 is assumed to belong to $\mathscr{M}_2^M(\mathbb{M}^n)$, namely the space of finite, nonnegative Borel measures on $\mathbb{M}^n$ having mass M and finite second moment, that is

$$\mu_0(\mathbb{M}^n) = M \qquad \text{and} \qquad \int_{\mathbb{M}^n} \mathsf{d}(x,o)^2 \, \mathrm{d}\mu_0(x) < \infty$$

for some (hence all) $o \in \mathbb{M}^n$. As is well known, one can make $\mathscr{M}_2^M(\mathbb{M}^n)$ a complete metric space by endowing it with the 2-Wasserstein distance, which we will denote by $\mathcal{W}_2$ (see Subsection 2.5.1 for more details).

Our main focus is on a stability property of the evolution (1.3.1) with respect to $\mathcal{W}_2$, when $\mathbb{M}^n$ is possibly noncompact (with infinite volume) and its Ricci curvature is merely bounded from below. This is strongly motivated by the results obtained by Sturm [Stu05] and Otto-Westdickenberg [OW05] under the nonnegativity assumption of the Ricci curvature, which we recall below. We point out that by "stability" we mean the possibility to control the $\mathcal{W}_2$-distance between two solutions of (1.3.1), along the flow, in terms of the $\mathcal{W}_2$-distance of the corresponding initial data. We will refer to this property as "contraction" when the $\mathcal{W}_2$-distance of the initial data cannot be increased by the flow.

To attack the problem we have at our disposal at least two different points of view. On the one hand, one can profit from the recent developments in the theory of nonlinear diffusion equations in non-Euclidean setting, where the connection with the geometry of the underlying structure is taken into account. On the other hand, the theory of optimal transportation can be employed to lift the problem to the space of measures endowed with the Wasserstein distance. The results obtained herein actually take advantage of the combination of techniques borrowed from both the two approaches.

For what concerns the analysis of nonlinear diffusion equations on Riemannian manifolds, we mention the following recent contributions. In [BGV08] the authors consider well-posedness and finite-time extinction phenomena for the fast-diffusion equation (i.e. (1.3.1) with $P(\rho) = \rho^m$ for $m \in (0,1)$) on Cartan-Hadamard manifolds, namely simply connected, complete Riemannian manifolds with nonpositive sectional curvature, for sufficiently integrable initial data. In the same geometric setting, in [GMP18b] the porous medium equation is investigated when initial data are finite Borel measures, by means of potential techniques. Still in the Cartan-Hadamard setting and for porous medium equation, in [GMP18a] the authors study well-posedness and blow-up phenomena for initial data possibly growing at infinity. The asymptotic behaviour for large times is addressed in [GMV17; GMV19], complementing some results previously obtained in [Váz15] in the hyperbolic space $\mathbb{H}^n$.

With regards to the theory of optimal transport, after the seminal work of Otto et al. [JKO98; Ott01] a lot of interest has been drawn in the description of certain PDEs as gradient flows in the space of probability measures endowed with the quadratic Wasserstein distance. In fact, when associated with a convex structure, such a formulation turns out to be extremely useful to obtain existence and stability results for a large class of PDEs. To that purpose, a very general theory of gradient flows of geodesically-convex functionals in metric spaces has rigorously been developed by Ambrosio, Gigli and Savaré: we refer to the monograph [AGS08] for a comprehensive treatment of this topic.

Let us first briefly comment on the analysis of the heat equation (at first in $\mathbb{R}^n$), for which the picture is by now quite clear. By setting

$$\mathsf{Ent}(\mu) := \begin{cases} \int_{\mathbb{R}^n} \rho \log \rho \, \mathrm{d}x & \text{if } \mathrm{d}\mu = \rho(x)\mathrm{d}x \,, \\ +\infty & \text{elsewhere} \,, \end{cases}$$

that is the so-called *relative entropy*, and denoting by $\mathscr{P}_2(\mathbb{R}^n)$ the space of probability measures with finite second moment, the following holds: for every initial datum $\mu_0 \in \mathscr{P}_2(\mathbb{R}^n)$ there exists a unique gradient flow of Ent in $(\mathscr{P}_2(\mathbb{R}^n), \mathcal{W}_2)$ in the sense of Evolution Variational Inequalities (EVI), whose trajectories coincide with the corresponding solution of the heat equation. The Wasserstein contraction property of the solutions is then a consequence of the *displacement convexity* of Ent in $(\mathscr{P}_2(\mathbb{R}^n), \mathcal{W}_2)$. This result was further extended to the Riemannian setting in [RS05], see also [Vil09; Erb10], upon taking into account the Ricci curvature of the manifold $\mathbb{M}^n$: it is shown that the bound $\mathrm{Ric} \geq \lambda$ ($\lambda \in \mathbb{R}$) is equivalent to both the λ-convexity of the relative entropy and the following stability property of the generated gradient flow:

$$\mathcal{W}_2(\rho(t), \hat{\rho}(t)) \leq e^{-\lambda t} \, \mathcal{W}_2(\mu_0, \hat{\mu}_0) \qquad \forall t \geq 0 \,,$$

where the densities ρ and $\hat{\rho}$ represent the solutions of the heat equation on $\mathbb{M}^n$ starting from μ_0 and $\hat{\mu}_0$, respectively. An equivalence of this form is still missing in the context of nonlinear diffusion, where only partial results can be found in the literature.

As concerns the classical porous medium equation, a gradient flow interpretation was firstly treated in [Ott01]. Then, numerous results have subsequently been obtained in the Euclidean setting even for more general PDEs. For instance, in [CMV06] the authors quantify the Wasserstein contraction for diffusion equations that may also exhibit a nonlocal structure. In the one-dimensional case, contraction estimates for granular-media models are obtained in [LT04], by exploiting the explicit formulation of the Wasserstein distance. Regularizing effects and decay estimates for porous medium evolutions (with a nonlocal pressure) can be obtained by means of the minimizing movement approximation scheme in $(\mathscr{P}_2(\mathbb{R}^n), \mathcal{W}_2)$, as is shown in [LMS18b]. Finally, we refer to [BC14] for a simple proof of the equivalence between the contraction of the flow and the convexity condition, in which the gradient-flow structure of the problem is in fact not exploited. A related argument (coming from the probabilistic coupling method) can also be found in the recent manuscript [FP19].

As already mentioned above, for nonlinear diffusions the passage from the Euclidean to the Riemannian setting is not straightforward. The first contribution in this direction was given by Sturm in [Stu05], where the equivalence between the geodesic convexity of the free energy and the curvature-dimension conditions is shown. In this setting, stability estimates for the PME on *nonnegatively* curved manifolds are still a consequence of the geodesic convexity of the free energy, thus complementing, when $\mathrm{Ric} \geq 0$, the results of [RS05] in the linear case. More precisely, the gradient-flow structure of the PME on $\mathscr{P}_2(\mathbb{M}^n)$ can be derived by introducing the free energy

$$\mathsf{E}(\mu) := \begin{cases} \int_{\mathbb{M}^n} U(\rho)\,\mathrm{d}\mathcal{V} & \text{if } \mathrm{d}\mu = \rho\,\mathrm{d}\mathcal{V}\,, \\ +\infty & \text{elsewhere}\,, \end{cases} \tag{1.3.2}$$

where U is linked to the nonlinearity of the equation through the relation $P(\rho) = \rho U'(\rho) - U(\rho)$. When $\mathbb{M}^n$ satisfies $\mathrm{Ric}_x \geq 0$ for every $x \in \mathbb{M}^n$, it is shown that under the additional assumption $\rho U'(\rho) \geq (1 - 1/n)\, U(\rho)$, the following contraction property holds along the flow:

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{W}_2(\rho(t), \hat{\rho}(t)) \leq 0 \qquad \forall t \geq 0\,.$$

Furthermore, the conditions on U and Ricci turn out to be also necessary for the contraction to hold, and they are equivalent to the displacement convexity of the functional E:

$$\mathsf{E}(\mu^s) \leq (1-s)\mathsf{E}\big(\mu^0\big) + s\mathsf{E}\big(\mu^1\big)$$

for every 2-Wasserstein geodesic $\{\mu^s\}_{0 \leq s \leq 1}$ in the space $(\mathscr{P}_2(\mathbb{M}^n), \mathcal{W}_2)$.

Let us recall that the above argument was subsequently revisited in the compact setting by Otto and Westdickenberg in [OW05] through the so-called *Eulerian* calculus. Recent developments have also been obtained in [OT11] in the context of weighted Riemannian and Finsler manifolds.

Our main goal is to obtain stability estimates for the porous medium-type evolution (1.3.1) without imposing the nonnegativity of the Ricci curvature. To that purpose, we need to introduce some key hypotheses both on the manifold and on the form of the nonlinearity we consider. First of all, we assume that $\mathbb{M}^n$ $(n \geq 3)$ supports the following *Sobolev-type* inequality:

$$\|f\|_{L^{2^\star}(\mathbb{M}^n)} \le C_S \left(\|\nabla f\|_{L^2(\mathbb{M}^n)} + \|f\|_{L^2(\mathbb{M}^n)} \right) \quad \forall f \in W^{1,2}(\mathbb{M}^n) \,, \qquad 2^\star := \frac{2n}{n-2} \,, \tag{1.3.3}$$

and has Ricci curvature bounded from below, that is

$$\mathrm{Ric}_x \ge -K \qquad \forall x \in \mathbb{M}^n \tag{1.3.4}$$

for some constant $K \ge 0$, in the sense of quadratic forms. Note that (1.3.3) is guaranteed on any complete, n-dimensional ($n \ge 3$) Riemannian manifold satisfying (1.3.4) along with the *noncollapse* condition, see Section 7.2. The 2-dimensional case can also be dealt with by means of minor modifications: we refer to Remark 7.2.8. For what concerns the nonlinearity, we assume P to be a $\mathcal{C}^1([0,+\infty))$, strictly increasing function satisfying $P(0)=0$ and the two-sided bound

$$c_0\, m\, \rho^{m-1} \le P'(\rho) \le c_1\, m\, \rho^{m-1} \qquad \forall \rho \ge 0 \,, \tag{1.3.5}$$

for some $c_1 \ge c_0 > 0$ and $m > 1$. In fact the requirement $m > 1$ corresponds to the so-called *slow diffusion* regime. Furthermore, it will also be crucial to ask that P complies with the *McCann* condition

$$\rho\, P'(\rho) - \left(1 - \tfrac{1}{n}\right) P(\rho) \ge 0 \qquad \forall \rho \ge 0 \,. \tag{1.3.6}$$

Let us observe that the *pure* porous medium nonlinearity, namely $P(\rho) = \rho^m$, obviously complies with (1.3.5) and (1.3.6).

In our main result, that is Theorem 7.2.4, we show that under the above conditions problem (1.3.1) admits a unique solution in an appropriate weak sense (see again Section 7.2 for more details). Moreover, for any pair of initial data $\mu_0, \hat\mu_0 \in \mathscr{M}_2^M(\mathbb{M}^n)$, the corresponding solutions $\mu(t) = \rho(t)\mathcal{V}$ and $\hat\mu(t) = \hat\rho(t)\mathcal{V}$ have a (bounded) density for every $t > 0$ and satisfy the following stability estimate with respect to the 2-Wasserstein distance:

$$\mathcal{W}_2(\rho(t), \hat\rho(t)) \le \exp\left\{ K\, c_1\, \mathfrak{C}_m \left[\left(tM^{m-1}\right)^{\frac{2}{2+n(m-1)}} \vee \left(tM^{m-1}\right) \right] \right\} \mathcal{W}_2(\mu_0, \hat\mu_0) \quad \forall t > 0 \,, \tag{1.3.7}$$

where a semi-explicit form of the constant $\mathfrak{C}_m > 0$ is also given. Estimate (1.3.7) seems to be new in the context of diffusion equations on manifolds, due to the presence of a nonlinear time power in the exponent. Moreover, in Theorem 7.2.5 we exhibit a nontrivial example that shows that our estimate is indeed optimal (for small times). Precisely, in the n-dimensional hyperbolic space $\mathbb{H}^n_K$ of constant sectional curvature $-K$ (thus of Ricci curvature $-(n-1)K$), given two close enough points $x, y \in \mathbb{H}^n_K$ and the associated Dirac measures $\mu_0 = M\delta_x$, $\hat\mu_0 = M\delta_y$, there holds

$$\mathcal{W}_2(\rho(t), \hat\rho(t)) \ge \left[1 + K\,\kappa \left(tM^{m-1}\right)^{\frac{2}{2+n(m-1)}} \right] \mathcal{W}_2(\mu_0, \hat\mu_0) \qquad \forall t \in (0, \bar t) \,, \tag{1.3.8}$$

for a suitable constant $\kappa = \kappa(n,m) > 0$ and a sufficiently small time $\bar t > 0$. As a consequence, we can deduce that the PME *is not a gradient flow* with respect to $\mathcal{W}_2$ on $\mathbb{H}^n_K$, or more generally on negatively-curved manifolds. We refer to Remark 7.2.6 for further details.

Strategy

The strategy we adopt has its roots in the so called *Eulerian* approach employed in [OW05; DS08] and subsequently in [AMS19]. Instead of relying on existence and smoothness of the optimal transport map, the main insight of the Eulerian approach is to directly work in the subspace of smooth densities and to take advantage of the Benamou-Brenier formulation of the Wasserstein distance. The basic idea is to link the contraction property of the Wasserstein distance to the monotonicity of the associated Lagrangian. Moreover, as is discussed in greater detail in [AMS19], the contraction of the distance under the action of the flow is also equivalent to the monotonicity of the associated Hamiltonian functional (in the sense of Fenchel duality). Such equivalence turns out to be more convenient in the context of porous medium flows; we give here a flavor of the strategy, referring to Section 7.4 for a more complete discussion. Let us start by writing the 2-Wasserstein distance as an action functional of the following form:

$$\frac{1}{2}\mathcal{W}_2(\rho_0,\hat\rho_0) = \inf\left\{\int_0^1 \mathcal{L}\big(\rho^s, \tfrac{\mathrm{d}}{\mathrm{d}s}\rho^s\big)\,\mathrm{d}s :\ s\mapsto\rho^s \ \text{ with } \rho^0=\rho_0\,,\ \rho^1=\hat\rho_0\right\},$$

where

$$\mathcal{L}(\rho,w) = \frac{1}{2}\int_{\mathbb{M}^n} |\nabla\phi|^2\rho\,\mathrm{d}\mathcal{V}\,, \qquad -\mathrm{div}(\rho\nabla\phi) = w \quad \text{in } \mathbb{M}^n\,.$$

Rather than looking directly at the Lagrangian $\mathcal{L}$, we consider the Hamiltonian functional

$$\mathcal{E}_\rho[\phi] := \frac{1}{2}\int_{\mathbb{M}^n} |\nabla\phi|^2\rho\,\mathrm{d}\mathcal{V}\,.$$

If $\rho\equiv\rho(t)$ is a solution of (1.3.1) and $\phi\equiv\phi(t)$ is the solution of the corresponding *linearized* backward flow given by $\frac{\mathrm{d}}{\mathrm{d}t}\phi = -P'(\rho)\Delta\phi$, it is not difficult to check that, at least formally, there holds (see [AMS19, Example 2.4])

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}_{\rho(t)}[\phi(t)] = \int_{\mathbb{M}^n} P(\rho(t))\,\Gamma_2(\phi(t))\,\mathrm{d}\mathcal{V} + \int_{\mathbb{M}^n}\left[\rho(t)P'(\rho(t)) - P(\rho(t))\right](\Delta\phi(t))^2\,\mathrm{d}\mathcal{V}\,,$$

where Γ_2 is the iterated *carré du champ* operator, whose definition is provided in Subsection 2.6.1. By exploiting (1.3.6) and the Bakry-Émery formulation $\mathrm{BE}(0,n)$ of the curvature bound $\mathrm{Ric}\ge 0$ (we refer again to Subsection 2.6.1), one can deduce the monotonicity of the Hamiltonian along the flow, namely $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}_{\rho(t)}[\phi(t)]\ge 0$, which is a key step in order to prove the 2-Wasserstein contraction property (see [AMS19, Proposition 2.1] in a simplified framework).

However, in the present setting we are dealing with the more general case in which the Ricci curvature is merely bounded from below. As a consequence, by employing the Bakry-Émery formulation $\mathrm{BE}(-K,n)$, a priori we only have

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}_{\rho(t)}[\phi(t)] \ge -K\int_{\mathbb{M}^n}|\nabla\phi(t)|^2\,P(\rho(t))\,\mathrm{d}\mathcal{V}\,.$$

In order to compare $\rho(t)$ with $P(\rho(t))$, and therefore to close the above differential inequality, the crucial idea is now to exploit a *quantitative* $L^1(\mathbb{M}^n)$–$L^\infty(\mathbb{M}^n)$ smoothing estimate for weak energy solutions of (1.3.1), see Proposition 7.3.3. To that purpose, it is necessary to first understand problem (1.3.1) for more regular initial data, namely

$$\begin{cases}\partial_t\rho = \Delta P(\rho) & \text{in } \mathbb{M}^n\times\mathbb{R}^+\,,\\ \rho(\cdot,0)=\rho_0\ge 0 & \text{on } \mathbb{M}^n\times\{0\}\,,\end{cases} \tag{1.3.9}$$

where $\rho_0 \in L^1(\mathbb{M}^n) \cap L^\infty(\mathbb{M}^n)$; in fact, it will also be essential to deal with a nondegenerate regularization of the equation, which will be addressed in detail in Sections 7.3 and 7.4. We point out that smoothing effects are a very important and well-established tool in the theory of a large class of nonlinear diffusion equations: we refer the reader e.g. to the monograph [Váz06]. This way we are able to integrate the differential inequality to get the estimate

$$\mathcal{E}_{\rho(t)}[\phi(t)] \geq \exp\{-K\, C(t,m,n)\}\, \mathcal{E}_{\rho_0}[\phi(0)]\,, \tag{1.3.10}$$

where an explicit computation of $C(t,m,n) > 0$ is available (see Lemma 7.4.3). The final step consists of exploiting the dual formulation of the Wasserstein distance for suitable regular curves, and we refer to Subsection 7.4.1 for a precise description of the strategy that allows one to pass from (1.3.10) to the stability estimate (1.3.7).

As for the optimality, we choose $\mathbb{M}^n$ as the hyperbolic space $\mathbb{H}^n_K$ of constant sectional curvature $-K$. The key ingredient to derive (1.3.8) is a delicate estimate on the Wasserstein distance between suitable radially-symmetric densities centered about two different (sufficiently close) points. To that purpose, we take advantage of a result originally proved by Ollivier [Oll09] in the simpler case of uniform densities, combined with the behaviour for small times of Barenblatt solutions of the PME in $\mathbb{H}^n_K$, obtained in [Váz15]. All the rigorous computations are carried out in Subsection 7.4.4.

Let us point out that the extension of the present results to a metric-measure setting appears not to be straightforward, mainly due to the PDE techniques we employ in Section 7.3. Indeed, the proof of the L^1–L^∞ smoothing estimate, which is a crucial ingredient of our strategy, is not directly applicable. The point is that we take advantage of a uniformly *parabolic* regularization of problem (1.3.1) in smooth domains, whose analogue in the metric-measure framework is in principle not available (see Remark 7.3.8). Another key tool we use, in order to show that solutions starting from data in $\mathscr{M}_2^M(\mathbb{M}^n)$ stay in $\mathscr{M}_2^M(\mathbb{M}^n)$, is the so-called *compact-support* property, that we establish again by pure PDE methods (see Proposition 7.3.4). The counterpart of this result in metric-measure spaces should be investigated by a different approach.

The material in Chapter 7 is the fruit of a joint collaboration with Matteo Muratori and Carlo Orrieri. It has been submitted to a scientific journal for publication, the preprint can be found on arXiv [DMO19].

Chapter 2

Preliminaries

The aim of the chapter is to fix the notation and recall some preliminary results, which will be useful in the sequel. We present some basic facts about convex analysis, power means, metric spaces, measure theory and optimal transport. The last part of the chapter is devoted to a brief introduction to the more recent theory of synthetic Ricci curvature bounds on metric measure spaces.

2.1 Convex analysis

A function $F : [0, +\infty) \to [0, +\infty]$ belongs to the class $\Gamma_0(\mathbb{R}_+)$ of the *admissible entropy functions* if F is convex, lower semicontinuous and $F(1) = 0$. The domain of the function F is the set

$$\mathrm{D}(F) := \big\{ s \in [0, +\infty) : F(s) < +\infty \big\}. \tag{2.1.1}$$

Let $F \in \Gamma_0(\mathbb{R}_+)$, the *recession function* $\mathrm{rec}(F)$ and the *recession constant* F'_∞ are defined by

$$\mathrm{rec}(F)(r) := \lim_{\alpha \to +\infty} \frac{F(1 + \alpha r)}{\alpha}, \qquad F'_\infty := \mathrm{rec}(F)(1). \tag{2.1.2}$$

We define the *right derivative* F'_0 at 0, and the *asymptotic affine coefficient* $\mathrm{aff}F_\infty$ as

$$F'_0 := \begin{cases} -\infty & \text{if} \quad F(0) = +\infty, \\ \lim_{s \downarrow 0} \frac{F(s) - F(0)}{s} & \text{otherwise}, \end{cases} \tag{2.1.3}$$

$$\mathrm{aff}F_\infty := \begin{cases} +\infty & \text{if} \quad F'_\infty = +\infty, \\ \lim_{s \to \infty} \big(F'_\infty s - F(s) \big) & \text{otherwise}. \end{cases} \tag{2.1.4}$$

Note that the definitions are well posed thanks to the convexity of F.

The *Legendre conjugate function* $F^* : \mathbb{R} :\to (-\infty, +\infty]$ is defined by

$$F^*(\phi) := \sup_{s \geq 0} \{ s\phi - F(s) \}. \tag{2.1.5}$$

F^* is the conjugate of the convex function $\tilde{F} : \mathbb{R} \to [0, +\infty]$ obtained by extending F to $+\infty$ for negative arguments. It is convex and lower semicontinuous. We also introduce the closed and convex subset $\mathfrak{F} \subset \mathbb{R}^2$ defined by

$$\mathfrak{F} := \{ (\phi, \psi) \in \mathbb{R}^2 : \psi \leq -F^*(\phi) \}. \tag{2.1.6}$$

Concerning the behavior of F^*, we have the following Lemma (see [LMS18a, Section 2.3]):

Lemma 2.1.1. *The function F^* is an increasing homeomorphism between (F_0', F_∞') and $(-F(0), \mathrm{aff} F_\infty)$ with $F^*(0) = 0$.*

The *reverse entropy function* $R : [0, +\infty) \to [0, +\infty]$ induced by $F \in \Gamma_0(\mathbb{R}_+)$ is defined by

$$R(s) := \begin{cases} F(\frac{1}{s})s & \text{if } s > 0, \\ F_\infty' & \text{if } s = 0. \end{cases} \tag{2.1.7}$$

In particular, R is convex, lower semicontinuous and the map $F \mapsto R$ is an involution of $\Gamma_0(\mathbb{R}_+)$. We also have

$$R(1) = 0, \quad R(0) = F_\infty', \quad R_\infty' = F(0), \quad R_0' = -\mathrm{aff} F_\infty, \quad \mathrm{aff} R_\infty = -F_0'. \tag{2.1.8}$$

The Legendre conjugates of F and R are related by

$$\psi \le -F^*(\phi) \iff \phi \le -R^*(\psi). \tag{2.1.9}$$

The *perspective function* induced by $F \in \Gamma_0(\mathbb{R}_+)$ is the function $\hat{F} : [0, +\infty) \times [0, +\infty) \to [0, +\infty]$, given by

$$\hat{F}(r, t) := \begin{cases} F\big(\frac{r}{t}\big)t & \text{if } t > 0, \\ \mathrm{rec}(F)(r) & \text{if } t = 0. \end{cases} \tag{2.1.10}$$

$\hat{F}$ is jointly convex, lower semicontinuous, positively 1-homogeneous in the sense that $\hat{F}(\lambda r, \lambda t) = \lambda \hat{F}(r, t)$ for every $\lambda \ge 0$, and $\hat{F}(1, 1) = 0$. An immediate consequence of the definitions (2.1.7) and (2.1.10) is that $\hat{F}(r, t) = \hat{R}(t, r)$. In particular, if F is equal to its reverse entropy R then $\hat{F}$ is a symmetric function, i.e.

$$\hat{F}(r, t) = \hat{F}(t, r) \qquad \text{for every } r, t \in [0, +\infty).$$

Accordingly, we denote by $\Gamma_0^s(\mathbb{R}_+)$ the set of functions $F \in \Gamma_0(\mathbb{R}_+)$ such that F is equals to its reverse entropy.

2.2 Power means

In this section we study the power means (also called generalized means), a family of functions that includes the well-known arithmetic, geometric and harmonic means. The property of these functions will be useful in Chapter 4.

In what follows r, t will denote two nonnegative real numbers and p a real parameter, which we suppose for the present not to be 0. The *p-power mean* between r and t is given by

$$\mathfrak{M}_p(r, t) := \Big(\frac{r^p + t^p}{2}\Big)^{\frac{1}{p}}, \tag{2.2.1}$$

except when $p < 0$ and r or t is zero. In this case $\mathfrak{M}_p$ is equal to zero:

$$\mathfrak{M}_p(r, t) := 0 \quad \text{if } p < 0,\ r = 0 \text{ or } t = 0). \tag{2.2.2}$$

If $p = 0$ we put

$$\mathfrak{M}_0(r, t) := \sqrt{rt} \tag{2.2.3}$$

so that $\lim_{p \to 0} \mathfrak{M}_p(r, t) = \mathfrak{M}_0(r, t)$.

It is easy to see that $\mathfrak{M}_p(r,r) = r$ for every $p \in \mathbb{R}$ and every $r \geq 0$. The function $\mathfrak{M}_p$ is positively 1-homogeneous and symmetric. Moreover, it is not difficult to prove that $M_p(r,s) \leq M_p(r,t)$ for every p, r and $s \leq t$.

$\mathfrak{M}_1$ is the well-known *arithmetic* mean, $\mathfrak{M}_0$ is the *geometric* mean and $\mathfrak{M}_{-1}$ is called *harmonic* mean.

The main result regarding the power means is the following (see [Bul03] for a proof).

Proposition 2.2.1. *If $p_1 < p_2$ then*

$$\mathfrak{M}_{p_1}(r,t) \leq \mathfrak{M}_{p_2}(r,t)$$

with the case of equality given by $r = t$, or $p_2 \leq 0$ and $r \wedge t = 0$.

In particular,

$$r \wedge t = \lim_{p \to -\infty} \mathfrak{M}_p(r,t) \leq \mathfrak{M}_p(r,t) \leq \lim_{p \to +\infty} \mathfrak{M}_p(r,t) = r \vee t, \tag{2.2.4}$$

for every $p \in \mathbb{R}$, $r,t \in [0,+\infty)$.

2.3 Metric spaces and functional spaces

A function $\mathsf{d} : X \times X \to [0,+\infty]$ is a *pseudo-metric* on the set X if for every $x,y,z \in X$,

- $\mathsf{d}(x,x) = 0$;
- d is symmetric, i.e. $\mathsf{d}(x,y) = \mathsf{d}(y,x)$;
- d satisfies the triangle inequality, i.e. $\mathsf{d}(x,z) \leq \mathsf{d}(x,y) + \mathsf{d}(y,z)$.

We call d a *metric* if it is a finite valued pseudo-metric such that $\mathsf{d}(x,y) = 0$ implies $x = y$. A *pseudo-metric space* (resp. *metric space*) is a couple (X,d), where d is a pseudo-metric (resp. metric) on the set X.

On a pseudo-metric space we consider the topology induced by the open balls $B_r(x) := \{y \in X : \mathsf{d}(x,y) < r\}$.

A *Polish space* is a separable, completely metrizable, topological space.

An *isometry* between two pseudo-metric spaces (X_1,d_1), (X_2,d_2) is a map $\psi : X_1 \to X_2$ such that $\mathsf{d}_1(x,y) = \mathsf{d}_2(\psi(x),\psi(y))$ for every $x,y \in X_1$.

Let (X,d) be a metric space, a *curve* is a continuous function $\gamma : I \to X$, where $I \subset \mathbb{R}$ is an interval. If X is complete, we say that the curve $\gamma : [0,1] \to X$ belongs to the class of *absolutely continuous curves of order* 2, denoted by $\mathrm{AC}^2([0,1];(X,\mathsf{d}))$, if there exists a square integrable function $w \in L^2(0,1)$ such that

$$\mathsf{d}(\gamma(s),\gamma(t)) \leq \int_s^t w(r)\,\mathrm{d}r \quad \text{for every } 0 \leq s \leq t \leq 1. \tag{2.3.1}$$

If $\gamma \in \mathrm{AC}^2([0,1];(X,\mathsf{d}))$, its *metric velocity*, defined as

$$|\dot{\gamma}|(r) := \lim_{h \to 0} \frac{\mathsf{d}(\gamma(r+h),\gamma(r))}{|h|},$$

exists for $\mathcal{L}^1$-a.e. $r \in (0,1)$. Moreover, $|\dot{\gamma}| \in L^2(0,1)$ and provides the minimal function w, up to $\mathcal{L}^1$-negligible sets, such that (2.3.1) holds (see e.g. [AGS08, Chapter 1]). Here with $\mathcal{L}^1$ we are denoting the 1-dimensional Lebesgue measure.

A (constant speed) *geodesic* is a curve $\gamma : [0,1] \to X$ such that

$$\mathsf{d}^2(\gamma(0),\gamma(1)) = \int_0^1 |\dot\gamma|^2(r)\,\mathrm{d}r,$$

or, equivalently,

$$\mathsf{d}(\gamma(s),\gamma(t)) = \mathsf{d}\left(\gamma(0),\gamma(1)\right)(t-s) \quad \text{for every } 0 \le s \le t \le 1.$$

In particular, a geodesic is a Lipschitz curve. We say that the space (X,d) is a *geodesic space* if every couple of points $(x,y) \in X \times X$ can be joined by a geodesic, i.e. there exists a geodesic γ such that $\gamma(0) = x, \gamma(1) = y$.

A well-known fact is that a complete metric space is a geodesic space if and only if for every pair of points $x, y \in X$ there exists $z \in X$ such that $\mathsf{d}(x,z) = \mathsf{d}(z,y) = \frac{1}{2}\mathsf{d}(x,y)$. The point z is called mid-point between x and y.

Let $\{(X_\alpha,\mathsf{d}_\alpha) \,|\, \alpha \in A\}$ be an indexed family of metric spaces, we define its *disjoint union* as

$$\bigsqcup_\alpha X_\alpha := \bigcup \Big\{X_\alpha \times \{\alpha\} | \alpha \in A\Big\}.$$

A pseudo-metric $\hat{\mathsf{d}}$ on $\bigsqcup_\alpha X_\alpha$ is called a *metric coupling* between $\{\mathsf{d}_\alpha\}$ if

$$\hat{\mathsf{d}}((x,\alpha),(y,\alpha)) = \mathsf{d}_\alpha(x,y) \quad \text{for every } x,y \in X_\alpha.$$

In this situation, the inclusion map

$$\iota_\alpha : X_\alpha \to \bigsqcup_\alpha X_\alpha, \quad \iota_\alpha(x) := (x,\alpha),$$

is an isometry with image $X_\alpha \times \{\alpha\}$. For this reason, we will often identify the space X_α with $X_\alpha \times \{\alpha\}$.

Starting from a metric space (X,d), we define the *cone over* X as the space $\mathfrak{C}(X) := Y/\sim$, where $Y := X \times [0,+\infty)$ and

$$(x_1,r_1) \sim (x_2,r_2) \iff r_1 = r_2 = 0 \text{ or } r_1 = r_2, x_1 = x_2.$$

We denote by $\mathfrak{y} = [\boldsymbol{y}] = [x,r]$ the points of $\mathfrak{C}(X)$, where the vertex $\mathfrak{o} := [x,0]$ plays a distinguished role.

Unless otherwise stated, we endowed $\mathfrak{C}(X)$ with the standard metric $\mathsf{d}_{\mathfrak{C}(X)}$, defined as

$$\mathsf{d}^2_{\mathfrak{C}(X)}\big([x_1,r_1],[x_2,r_2]\big) := r_1^2 + r_2^2 - 2r_1r_2\cos(\mathsf{d}(x_1,x_2) \wedge \pi). \tag{2.3.2}$$

The space $(\mathfrak{C}(X),\mathsf{d}_{\mathfrak{C}(X)})$ is complete and separable if (X,d) is complete and separable and it is a geodesic space if and only if (X,d) is a geodesic space at distances less than π, the latter means that if $x,y \in X$ are two points such that $\mathsf{d}(x,y) < \pi$ then there exists a geodesic in X connecting them. A proof of these results can be found in [BBI01, Chapter 3.6].

Let X, Y be two topological spaces, we denote by $\mathcal{C}(X,Y)$ the set of continuous functions $f : X \to Y$. When $Y = \mathbb{R}$ endowed with the Euclidean topology, we simply put $\mathcal{C}(X) := \mathcal{C}(X,\mathbb{R})$.

The *support* of $f \in \mathcal{C}(X)$ is the closure (with respect the topology on X) of the set of points $x \in X$ such that $f(x) \neq 0$. We denote by $\mathcal{C}_b(X)$ (resp. $\mathcal{C}_{bs}(X)$) the subspace of $\mathcal{C}(X)$ consisting of continuous and bounded (resp. continuous with bounded support) functions. We endow these spaces with the supremum norm. We also use

the notation $f \in \mathcal{C}_c(X)$ when the support of f is compact.
If (X, d) is a metric space, the space of real Lipschitz functions is denoted by $\mathsf{Lip}(X)$. We also consider the spaces $\mathsf{Lip}_b(X)$ and $\mathsf{Lip}_{bs}(X)$ with obvious meaning of the symbols. On these spaces, we consider the norm

$$\|f\|_{\mathsf{Lip}(X)} := \|f\|_\infty + c_f,$$

where c_f is the *Lipschitz constant* of the function f, i.e.

$$c_f := \sup\left\{\frac{|f(x) - f(y)|}{\mathsf{d}(x,y)} : x, y \in X,\, x \neq y\right\}.$$

Given $f \in \mathsf{Lip}(X)$ its slope $|\nabla f|(x)$ at $x \in X$ is defined by

$$|\nabla f|(x) := \limsup_{y \to x} \frac{|f(y) - f(x)|}{\mathsf{d}(y,x)}, \tag{2.3.3}$$

with the convention $|\nabla f|(x) = 0$ if x is an isolated point.

When $(X, \|\cdot\|)$ is a Banach space, we denote by X' the dual space consisting of the linear functionals $f : X \to \mathbb{R}$ with finite dual norm, where the latter is defined as

$$\|f\|_{X'} := \sup\left\{\frac{|f(x)|}{\|x\|} : x \in X,\, x \neq 0\right\}.$$

The standard duality pairing between X' and X is denoted by ${}_{X'}\langle\cdot,\cdot\rangle_X$.

The notion of *strong* convergence on X is defined by

$$x_n \xrightarrow[(X,\|\cdot\|)]{} x \iff \|x_n - x\| \to 0,$$

while the notion of *weak* convergence is defined by

$$x_n \xrightharpoonup[(X,\|\cdot\|)]{} x \iff f(x_n) \to f(x) \quad \text{for every } f \in X'. \tag{2.3.4}$$

When the space is clear by the context, we simply denote by $x_n \to x$ the strong convergence and by $x_n \rightharpoonup x$ the weak convergence.

2.4 Measure theory

Let X be a topological space. On X we always consider the σ-algebra $\mathcal{B}(X)$ of the Borel sets. We denote by $\mathscr{M}(X)$ the set of finite, nonnegative, Borel measures on X. We define $\mathscr{M}^M(X)$ as the space of measures $\mu \in \mathscr{M}(X)$ such that $\mu(X) = M$. When $M = 1$, we put $\mathscr{P}(X) := \mathscr{M}^1(X)$ and call it the space of *probability* measures.

The *support* $\mathsf{supp}(\mu)$ of $\mu \in \mathscr{M}(X)$ is defined as

$$\mathsf{supp}(\mu) := \{x \in X : \mu(U) > 0 \text{ for every neighbourhood } U \text{ of } x\}. \tag{2.4.1}$$

If μ is a measure on X, we denote by $L^p(X,\mu)$, $p \in [1,\infty]$, the space of Lebesgue p-integrable functions. When the measure μ is clear by the context, we simply write $L^p(X)$. We endow $L^p(X,\mu)$, $p < \infty$, with the standard norm

$$\|f\|_{L^p(X,\mu)} := \left(\int_X |f|^p \,\mathrm{d}\mu\right)^{1/p}.$$

On $L^\infty(X,\mu)$ we consider the *essential supremum* norm, so that $(L^p(X,\mu), \|\cdot\|_{L^p(X,\mu)})$ is a Banach space for every $1 \le p \le \infty$. For abbreviation, we also use the notation $\|f\|_p = \|f\|_{L^p} := \|f\|_{L^p(X,\mu)}$.

When $I \subset \mathbb{R}$ is an interval (endowed with the Lebesgue measure) and $(X, \|\cdot\|)$ is a Banach space, we denote by $L^p(I;X)$ the *Bochner* space (see [Mál+96, Section 1.2.6]), $1 \le p \le \infty$.

When X is regular enough, the notations $W^{1,p}(X)$ and $W^{1,p}(I,X)$ stand for the *(Bochner)-Sobolev* space of p-integrable functions with p-integrable weak derivative.

We say that a sequence of measures $(\mu_n) \in \mathscr{M}(X)$ *weakly* (or *narrowly*) converges to a measure $\mu \in \mathscr{M}(X)$ if

$$\int_X f \,\mathrm{d}\mu_n \to \int_X f \,\mathrm{d}\mu \quad \text{for any } f \in \mathcal{C}_b(X).$$

If X is a Polish space, it is well known that this notion of convergence can be induced by a distance on $\mathscr{M}(X)$ (see [AGS08, Remark 5.1.1] or the rest of the thesis), so that the family of all converging sequences is sufficient to characterize this topology.

A subset $\mathscr{K} \subset \mathscr{M}(X)$ is *bounded* if $\sup_{\mu\in\mathscr{K}} \mu(X) < \infty$ and it is *equally tight* if

$$\forall \epsilon > 0 \ \ \exists K_\epsilon \subset X \text{ compact}: \ \forall \mu \in \mathscr{K}, \quad \mu(X \setminus K_\epsilon) \le \epsilon. \tag{2.4.2}$$

Compactness properties with respect the weak topology on $\mathscr{M}(X)$ are guaranteed by the famous Prokhorov's Theorem [Sch73, Th. 3, p.379]:

Theorem 2.4.1. *Let X be a Polish space. A subset $\mathscr{K} \subset \mathscr{M}(X)$ is bounded and equally tight if and only if it is relatively compact with respect to the weak topology.*

Suppose $\mu \in \mathscr{M}(X)$, let X, Y be two Polish spaces and let $T : X \to Y$ be a Borel map. We define the *push-forward* measure $T_\sharp\mu \in \mathscr{M}(Y)$ as

$$T_\sharp\mu(B) := \mu(T^{-1}(B)) \quad \text{for every } B \in \mathcal{B}(Y). \tag{2.4.3}$$

More generally, it holds

$$\int_X f(T(x))\mathrm{d}\mu(x) = \int_Y f(y)\mathrm{d}T_\sharp\mu(y), \tag{2.4.4}$$

for every $T_\sharp\mu$-integrable Borel function $f : Y \to \mathbb{R}$.

We say that a measure $\gamma \in \mathscr{M}(X)$ is *absolutely continuous* with respect to a measure $\mu \in \mathscr{M}(X)$, and we write $\gamma \ll \mu$, if $\gamma(A) = 0$ for any set $A \in \mathcal{B}(X)$ such that $\mu(A) = 0$. In this situation it is well known that γ admits a *Radon-Nykodym derivative* with respect to μ, i.e. there exists a measurable function $f : X \to [0,\infty)$ such that

$$\gamma(B) = \int_B f \mathrm{d}\mu, \qquad \text{for any } B \in \mathcal{B}(X).$$

The function f is denoted by $\dfrac{\mathrm{d}\gamma}{\mathrm{d}\mu}$.

The measure γ is singular with respect to $\mu^\perp \in \mathscr{M}(X)$ if there exists a Borel subset $A \subset X$ such that $\mu(A) = \gamma^\perp(X \setminus A) = 0$.

More generally, the following version of the Lebesgue's decomposition Theorem holds (see [LMS18a, Lemma 2.3]).

Lemma 2.4.2. *For every $\gamma, \mu \in \mathscr{M}(X)$ with $\gamma(X) + \mu(X) > 0$, there exist Borel functions $\sigma, \rho : X \to [0,\infty)$ and a Borel partition (A, A_γ, A_μ) of X that satisfy the* *following:*

$$A = \{x \in X : \sigma(x) > 0\} = \{x \in X : \rho(x) > 0\}, \quad \sigma \cdot \rho \equiv 1 \text{ in } A,$$
$$\gamma = \sigma\mu + \gamma^\perp, \quad \sigma \in L^1_+(X,\mu), \quad \gamma^\perp(X \setminus A_\gamma) = \mu(A_\gamma) = 0,$$
$$\mu = \rho\gamma + \mu^\perp, \quad \rho \in L^1_+(X,\gamma), \quad \mu^\perp(X \setminus A_\mu) = \mu(A_\mu) = 0.$$

2.5 Optimal Transport and Wasserstein distance

Let X be a Polish space and let $\boldsymbol{c} : X \times X \to [0, +\infty]$ be a lower semicontinuous cost function. We define the *optimal transport cost* between the measures $\mu^0, \mu^1 \in \mathscr{M}(X)$ as

$$\mathrm{T}(\mu^0, \mu^1) := \inf_{\boldsymbol{\pi}} \int_{X\times X} \boldsymbol{c}(x_1, x_2) \, \mathrm{d}\boldsymbol{\pi}(x_1, x_2) \,, \tag{2.5.1}$$

where the infimum is taken among the set of *transport plans* $\boldsymbol{\pi}$ between μ^0 and μ^1. The latter is the set of measures $\boldsymbol{\pi} \in \mathscr{M}(X \times X)$ such that $\boldsymbol{\pi}(A \times X) = \mu^0(A)$ and $\boldsymbol{\pi}(X \times B) = \mu^1(B)$ for every Borel sets $A, B \in \mathcal{B}(X)$. We notice that $\mathrm{T}(\mu^0, \mu^1) = +\infty$ whenever $\mu^0(X) \neq \mu^1(X)$, and the value $\mathrm{T}(\mu^0, \mu^1)$ can be equal to $+\infty$ even if $\mu^0, \mu^1 \in \mathscr{M}^M(X)$.

A well-known fact (see e.g. [Vil09, Theorem 4.1]) is that there exist an optimal plan $\tilde{\boldsymbol{\pi}}$ such that

$$\mathrm{T}(\mu^0, \mu^1) = \int_{X\times X} \boldsymbol{c}(x_1, x_2) \, \mathrm{d}\tilde{\boldsymbol{\pi}}(x_1, x_2) \,. \tag{2.5.2}$$

In the above setting, it also holds the celebrated *Kantorovich duality* (see e.g. [Vil09, Theorem 5.10]):

Theorem 2.5.1. *For every $\mu^0, \mu^1 \in \mathscr{M}(X)$ we have*

$$\mathrm{T}(\mu^0, \mu^1) = \sup_{\substack{(\varphi,\psi)\in \mathcal{C}_b(X)\times\mathcal{C}_b(X);\\ \psi-\varphi\le c}} \left\{ \int_X \psi \, \mathrm{d}\mu^1 - \int_X \varphi \, \mathrm{d}\mu^0 \right\}, \tag{2.5.3}$$

where $\psi - \varphi \le c$ stands for

$$\psi(x_1) - \varphi(x_2) \le c(x_1, x_2) \quad \forall\, (x_1, x_2) \in X \times X.$$

2.5.1 The Wasserstein space

Let (X, d) be a Polish space. We say that the measure $\mu \in \mathscr{M}^M(X)$ has finite *p-moment*, $p \ge 1$, and we write $\mu \in \mathscr{M}^M_p(X)$, if there exists a point $o \in X$ such that

$$\int_X \mathsf{d}(x, o)^p \, \mathrm{d}\mu(x) < +\infty \,. \tag{2.5.4}$$

We define the *p-Wasserstein cost* between two measures $\mu^0, \mu^1 \in \mathscr{M}(X)$ as

$$\mathcal{W}_p^p(\mu^0, \mu^1) := \inf_{\boldsymbol{\pi}} \int_{X\times X} \mathsf{d}(x_1, x_2)^p \, \mathrm{d}\boldsymbol{\pi}(x_1, x_2) \,, \tag{2.5.5}$$

where the infimum is taken among all the transport plans $\boldsymbol{\pi}$ between μ^0 and μ^1.

An elementary fact is that

$$\mu^0 \in \mathscr{M}^M_p(X) \quad \text{and} \quad \mathcal{W}_p(\mu^0, \mu^1) < +\infty \quad \text{implies} \quad \mu^1 \in \mathscr{M}^M_p(X) \,. \tag{2.5.6}$$

We are mainly interested in the cases $p = 1$ and $p = 2$. Regarding the 1-Wasserstein distance, we will only use these two well-known facts:

$$\mathcal{W}_1(\mu^0, \mu^1) \le \mathcal{W}_2(\mu^0, \mu^1), \quad \text{for every } \mu^0, \mu^1 \in \mathscr{P}_2(X), \tag{2.5.7}$$

$$\mathcal{W}_1(\mu^0, \mu^1) = \sup\left\{ \int_X f \mathrm{d}\mu_1 - \int_X f \mathrm{d}\mu_0 : f : X \to \mathbb{R},\ f \text{ 1-Lipschitz} \right\}, \tag{2.5.8}$$

where the first property follows by an application of the Holder's inequality while the second property is a direct consequence of Theorem 2.5.1.

When $p = 2$, it can be shown that for every $M \in (0, \infty)$ the space $(\mathscr{M}_2^M(X), \mathcal{W}_2)$ is a metric space, called the *2-Wasserstein* (or simply Wasserstein) *space* of mass M over X. It inherits many geometrical properties of the ambient space X: in particular, it is complete and separable. It is also geodesic if and only if (X, d) is geodesic (for a proof of these facts see e.g. [Vil09, Chapter 6]).

To our purposes, it is convenient to recall a useful characterization of convergence in the 2-Wasserstein space, for the proof of which we refer to [AGS08, Proposition 7.1.5].

Proposition 2.5.2. *Let $\mu \in \mathscr{M}_2^M(X)$ and $\{\mu_j\}_{j\in\mathbb{N}} \subset \mathscr{M}_2^M(X)$. Then*

$$\lim_{j\to\infty} \mathcal{W}_2(\mu_j, \mu) = 0$$

if and only if $\mu_j \rightharpoonup \mu$ weakly and $\{\mu_j\}_{j\in\mathbb{N}}$ has equi-integrable second moments, where the latter means that there exists a point $o \in X$ such that

$$\lim_{k\to\infty} \limsup_{j\to\infty} \int_{X\setminus B_k(o)} \mathsf{d}^2(x, o)\mathrm{d}\mu_j(x) = 0.$$

The dual characterization of the Wasserstein distance takes the following form: for any $\mu^0, \mu^1 \in \mathscr{M}^M(X)$ there holds

$$\frac{1}{2}\mathcal{W}_2^2(\mu^0, \mu^1) = \sup_{\substack{\varphi,\psi\in \mathcal{C}_b(X): \\ \psi(x_1)\le\varphi(x_2)+\frac{1}{2}\mathsf{d}(x_1,x_2)^2\ \forall x_1,x_2\in X\times X}} \left\{ \int_X \psi \,\mathrm{d}\mu^1 - \int_X \varphi \,\mathrm{d}\mu^0 \right\}. \tag{2.5.9}$$

From (2.5.9) it is clear that, for any fixed φ, the best possible choice of ψ is given by

$$\psi(x) = Q_1\varphi(x) := \inf_{y\in X} \varphi(y) + \frac{1}{2}\mathsf{d}(x, y)^2 \qquad \forall x \in X. \tag{2.5.10}$$

Using (2.5.10) and by means of a standard cut-off and regularization argument, it is not difficult to show that the supremum in (2.5.9) can actually be taken over the space $\mathsf{Lip}_{bs}(X)$ of Lipschitz functions with bounded support. In particular it can be shown:

Proposition 2.5.3. *Let $\mu^0, \mu^1 \in \mathscr{M}^M(X)$. Then*

$$\frac{1}{2}\mathcal{W}_2^2(\mu^0, \mu^1) = \sup_{\varphi\in \mathsf{Lip}_{bs}(X)} \left\{ \int_X Q_1\varphi \,\mathrm{d}\mu^1 - \int_X \varphi \,\mathrm{d}\mu^0 \right\}. \tag{2.5.11}$$

If (X, d) is more regular, say a geodesic space, the function $Q_1\varphi$ can be seen as an endpoint of the *Hopf-Lax* evolution semigroup starting from φ. We recall that the latter is given by the family of maps $Q_s : \mathsf{Lip}_{bs}(X) \to \mathsf{Lip}_{bs}(X)$, $s \geq 0$, defined as

$$Q_s\varphi(x) := \inf_{y\in X} \varphi(y) + \frac{\mathsf{d}(x,y)^2}{2s} \quad \forall s > 0\,, \qquad Q_0\varphi(x) := \varphi(x) \qquad \forall x \in X\,. \tag{2.5.12}$$

It is readily seen that $Q_s\varphi$ satisfies

$$\inf_X \varphi \leq Q_s\varphi(x) \leq \varphi(x) \qquad \forall s \geq 0\,,\ \forall x \in X\,. \tag{2.5.13}$$

More importantly, it can be shown (see [AGS14a, Theorem 3.6]) that $(s,x) \mapsto Q_s\varphi(x)$ is the Lipschitz solution of the Hopf-Lax (or Hamilton-Jacobi) problem

$$\begin{cases} \frac{\partial}{\partial s} Q_s\varphi(x) = -\frac{1}{2}\,|\nabla Q_s\varphi|^2(x) & \text{for a.e. } (x,s) \in X \times \mathbb{R}^+\,, \\ Q_0\varphi = \varphi\,. \end{cases} \tag{2.5.14}$$

2.6 Metric measure spaces and Ricci curvature

Definition 2.6.1. *Let (X, d) be a Polish space endowed with a reference measure $\mathfrak{m}$ over the Borel σ-algebra $\mathcal{B}(X)$. Let us suppose that $\mathfrak{m}$ is non-negative and finite on bounded sets. The triple $(X, \mathsf{d}, \mathfrak{m})$ is called* metric measure space*, m.m.s for short.*

In this section, we suppose that $\mathsf{supp}(\mathfrak{m}) = X$ and $\mathfrak{m}$ satisfies an exponential growth condition: namely, there exist $x_0 \in X$, $M > 0$ and $c \geq 0$ such that

$$\mathfrak{m}(B_r(x_0)) \leq M \exp(cr^2) \quad \text{for every } r > 0.$$

Possibly enlarging $\mathcal{B}(X)$ and extending $\mathfrak{m}$, we also assume that $\mathcal{B}(X)$ is $\mathfrak{m}$-complete.

The *classical relative entropy functional* $\mathsf{Ent}_\mathfrak{m} : \mathcal{P}_2(X) \to \mathbb{R} \cup \{+\infty\}$ is defined as

$$\mathsf{Ent}_\mathfrak{m}(\mu) := \begin{cases} \int \rho \log \rho \, \mathrm{d}\mathfrak{m} & \text{if } \mu = \rho\mathfrak{m}, \\ +\infty & \text{otherwise.} \end{cases} \tag{2.6.1}$$

In the sequel we use the notation

$$\mathrm{D}(\mathsf{Ent}_\mathfrak{m}) := \{\mu \in \mathcal{P}_2(X) \,:\, \mathsf{Ent}_\mathfrak{m}(\mu) \in \mathbb{R}\}$$

for the domain of the relative entropy.

We now define the $\mathsf{CD}(K, \infty)$ condition, coming from the seminal works of Lott-Villani [LV09] and Sturm [Stu06].

Definition 2.6.2 ($\mathsf{CD}(K,\infty)$ condition)**.** *Let $K \in \mathbb{R}$. We say that $(X, \mathsf{d}, \mathfrak{m})$ is a $\mathsf{CD}(K,\infty)$ space provided that for any $\mu^0, \mu^1 \in \mathrm{D}(\mathsf{Ent}_\mathfrak{m})$ there exists a $\mathcal{W}_2$-geodesic (μ_t) such that $\mu_0 = \mu^0$, $\mu_1 = \mu^1$ and*

$$\mathsf{Ent}_\mathfrak{m}(\mu_t) \leq (1-t)\mathsf{Ent}_\mathfrak{m}(\mu_0) + t\mathsf{Ent}_\mathfrak{m}(\mu_1) - \frac{K}{2} t(1-t)\mathcal{W}_2^2(\mu_0, \mu_1). \tag{2.6.2}$$

The *Cheeger energy* (introduced in [Che99] and further studied in [AGS08]) is defined as the L^2-lower semicontinuous envelope of the functional $f \mapsto \frac{1}{2}\int_X |\nabla f|^2 \mathrm{d}\mathfrak{m}$, i.e.:

$$\mathsf{Ch}_\mathfrak{m}(f) := \inf\left\{ \liminf_{n\to\infty} \frac{1}{2}\int_X |\nabla f_n|^2 \mathrm{d}\mathfrak{m} : f_n \in \mathsf{Lip}_{bs}(X), f_n \to f \text{ in } L^2(X,\mathfrak{m}) \right\}. \tag{2.6.3}$$

If $\mathsf{Ch}_{\mathfrak{m}}(f) < \infty$, it was proved in [Che99; AGS08] that the set

$$\mathrm{G}(f) := \{g \in L^2(X, \mathfrak{m}) \,:\, \exists \{f_n\}_n \subset \mathsf{Lip}_{bs}(X),\, f_n \to f, |\nabla f_n| \rightharpoonup h \leq g \text{ in } L^2(X, \mathfrak{m})\}$$

is closed and convex, therefore it admits a unique element of minimal norm called *minimal weak upper gradient* and denoted by $|Df|_w$. The Cheeger energy can be then represented by integration as

$$\mathsf{Ch}_{\mathfrak{m}}(f) = \frac{1}{2} \int_X |Df|_w^2 \, \mathrm{d}\mathfrak{m}.$$

We recall that the minimal weak upper gradient satisfies the following locality property (see e.g. [AGS14b, equation (2.18)]):

$$|Df|_w = 0 \quad \mathfrak{m}\text{-a.e. on the set } \{f = 0\}. \tag{2.6.4}$$

One can show that $\mathsf{Ch}_{\mathfrak{m}}$ is a 2-homogeneous, lower semicontinuous, convex functional on $L^2(X, \mathfrak{m})$ whose proper domain

$$\mathbb{V} := \{f \in L^2(X, \mathfrak{m}) \,:\, \mathsf{Ch}_{\mathfrak{m}}(f) < \infty\}$$

is a dense linear subspace of $L^2(X, \mathfrak{m})$. It then admits an L^2 gradient flow which is a continuous semi-group of contractions $(H_t)_{t\geq 0}$ in $L^2(X, \mathfrak{m})$, whose continuous trajectories $t \mapsto H_t f$, for $f \in L^2(X, \mathfrak{m})$, are locally Lipschitz curves from $(0, \infty)$ with values into $L^2(X, \mathfrak{m})$ that satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t} H_t f \in -\partial \mathsf{Ch}_{\mathfrak{m}}(H_t f) \quad \text{for a.e. } t \in (0, \infty). \tag{2.6.5}$$

Here ∂ denotes the subdifferential of convex analysis, namely for every $f \in \mathbb{V}$ we have $\ell \in \partial \mathsf{Ch}_{\mathfrak{m}}(f)$ if and only if

$$\int_X \ell(g - f) \mathrm{d}\mathfrak{m} \leq \mathsf{Ch}_{\mathfrak{m}}(g) - \mathsf{Ch}_{\mathfrak{m}}(f), \quad \text{for every } g \in L^2(X, \mathfrak{m}). \tag{2.6.6}$$

We now define the $\mathsf{RCD}(K, \infty)$ condition, introduced and throughly analyzed in [AGS14b] (see also [Amb+15] for the present simplified axiomatization and the extension to the σ-finite case).

Definition 2.6.3 ($\mathsf{RCD}(K, \infty)$ condition)**.** *Let $K \in \mathbb{R}$. We say that the metric measure space $(X, \mathsf{d}, \mathfrak{m})$ is $\mathsf{RCD}(K, \infty)$ if it satisfies the $\mathsf{CD}(K, \infty)$ condition and moreover the Cheeger energy $\mathsf{Ch}_{\mathfrak{m}}$ is quadratic, i.e. it satisfies the parallelogram identity*

$$\mathsf{Ch}_{\mathfrak{m}}(f + g) + \mathsf{Ch}_{\mathfrak{m}}(f - g) = 2\mathsf{Ch}_{\mathfrak{m}}(f) + 2\mathsf{Ch}_{\mathfrak{m}}(g), \quad \forall f, g \in \mathbb{V}. \tag{2.6.7}$$

If $(X, \mathsf{d}, \mathfrak{m})$ is an $\mathsf{RCD}(K, \infty)$ space, then the Cheeger energy induces the Dirichlet form $\mathcal{E}(f) := 2\mathsf{Ch}_{\mathfrak{m}}(f)$ which is strongly local, symmetric and admits the Carré du Champ

$$\Gamma(f) := |Df|_w^2, \quad \forall f \in \mathbb{V}.$$

The space $\mathbb{V}$ endowed with the norm $\|f\|_{\mathbb{V}}^2 := \|f\|_{L^2}^2 + \mathcal{E}(f)$ is Hilbert. Moreover, the sub-differential $\partial \mathsf{Ch}_m$ is single-valued and coincides with the linear generator $-\Delta$ of the heat flow semi-group $(H_t)_{t\geq 0}$ defined above. In other terms, the semigroup can be equivalently characterized by the fact that for any $f \in L^2(X, \mathfrak{m})$ the curve

$t \mapsto H_t f \in L^2(X, \mathfrak{m})$ is locally Lipschitz from $(0,\infty)$ to $L^2(X,\mathfrak{m})$ and satisfies

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} H_t f = \Delta H_t f & \text{for } \mathcal{L}^1\text{-a.e } t \in (0,\infty), \\ \lim_{t\to 0} H_t f = f, \end{cases} \tag{2.6.8}$$

where the limit is in the strong $L^2(X,\mathfrak{m})$-topology.

The semigroup H_t extends uniquely to a strongly continuous semigroup of linear contractions in $L^p(X,\mathfrak{m})$, $p \in [1,\infty)$, for which we retain the same notation. Regarding the case $p = \infty$, it was proved in [AGS14b, Theorem 6.1] that there exists a version of the semigroup such that $H_t f(x)$ belongs to $\mathcal{C} \cap L^\infty((0,\infty) \times X)$ whenever $f \in L^\infty(X,\mathfrak{m})$. We will implicitly refer to this version of $H_t f$ when f is essentially bounded. Moreover, for any $f \in L^2 \cap L^\infty(X,\mathfrak{m})$ and for every $t > 0$ we have $H_t f \in \mathbb{V} \cap \mathsf{Lip}(X)$ with the explicit bound (see [AGS14b, Theorem 6.5] for a proof and Proposition 6.2.1 of the present thesis for an improved inequality)

$$\| |DH_t f|_w \|_\infty \le \sqrt{\frac{K}{e^{2Kt} - 1}} \, \|f\|_\infty \, . \tag{2.6.9}$$

Two crucial properties of the heat flow are the preservation of mass and the maximum principle (see [AGS08]):

$$\int_X H_t f \,\mathrm{d}\mathfrak{m} = \int_X f \,\mathrm{d}\mathfrak{m}, \quad \text{for any } f \in L^1(X,\mathfrak{m}), \tag{2.6.10}$$

$$0 \le H_t f \le C, \quad \text{for any } 0 \le f \le C \ \mathfrak{m}\text{-a.e.}, \ C > 0. \tag{2.6.11}$$

A result of Savaré [Sav14, Corollary 3.5] ensures that, in the $\mathsf{RCD}(K,\infty)$ setting, for every $f \in \mathbb{V}$ and $\alpha \in [1/2, 1]$ we have

$$|DH_t f|_w^{2\alpha} \le e^{-2\alpha K t} H_t\big(|Df|_w^{2\alpha}\big), \quad \mathfrak{m}\text{-a.e. } . \tag{2.6.12}$$

In particular, we will make use of the case $\alpha = 1/2$

$$|DH_t f|_w \le e^{-Kt} H_t(|Df|_w), \quad \mathfrak{m}\text{-a.e. } . \tag{2.6.13}$$

2.6.1 Riemannian manifold and the Bakry-Émery curvature condition

A typical example of metric measure space is a complete, connected, n-dimensional Riemannian manifold $(\mathbb{M}^n, g)$. It is indeed well known that the metric tensor g induces a Riemannian distance d and a Riemannian volume measure $\mathcal{V}$ (see e.g. [BGL14, Appendix C] and references therein). Moreover, the metric space $(\mathbb{M}^n, \mathsf{d})$ is complete, separable, locally compact and also geodesic (thus proper). On a Riemannian manifold, we denote by $T_x\mathbb{M}^n$ the tangent space at the point $x \in \mathbb{M}^n$ and we put $\mathbb{H} := L^2(\mathbb{M}^n, \mathcal{V})$.

In this situation, the Cheeger energy corresponds to the standard Dirichlet energy

$$\mathsf{Ch}_{\mathcal{V}}(f) = \frac{1}{2}\mathcal{E}(f) = \frac{1}{2}\int_{\mathbb{M}^n} \Gamma(f)\,\mathrm{d}\mathcal{V} = \frac{1}{2}\int_{\mathbb{M}^n} |\nabla f|^2\,\mathrm{d}\mathcal{V}, \tag{2.6.14}$$

where ∇ stands for the (distributional) gradient and

$$\Gamma(f,g) := \langle \nabla f, \nabla g \rangle \, , \qquad \Gamma(f) := |\nabla f|^2. \tag{2.6.15}$$

The space $\mathbb{V}$ previously defined is the usual Sobolev space $W^{1,2}(\mathbb{M}^n)$ and the $L^2(\mathbb{M}^n)$ gradient flow of the energy $\mathsf{Ch}_{\mathcal{V}}$ produces the classical heat equation

$$\frac{\mathrm{d}}{\mathrm{d}t}u = \Delta u.$$

Here $\Delta = \Delta_g$ is the Laplace-Beltrami operator. It is useful to recall that, thanks to [Str83, Theorem 2.4], the operator $(-\Delta)$ defined in $C_c^\infty(\mathbb{M}^n)$ is essentially self-adjoint on *any* complete Riemannian manifold, i.e. $\mathbb{D} := \{f \in \mathbb{V} : \Delta f \in \mathbb{H}\}$ coincides with the closure of $C_c^\infty(\mathbb{M}^n)$ with respect to the norm

$$\|u\|_{\mathbb{D}} := \left(\|u\|_{\mathbb{V}}^2 + \|\Delta u\|_{\mathbb{H}}^2\right)^{1/2}.$$

The *iterated carré du champ* is defined by

$$\Gamma_2(f,g) := \frac{1}{2}\big(\Delta\left(\Gamma(f,g)\right) - \Gamma(f,\Delta g) - \Gamma(g,\Delta f)\big). \tag{2.6.16}$$

When the Ricci curvature of $\mathbb{M}^n$ is uniformly bounded from below by a constant K, i.e. there exists $K \in \mathbb{R}$ such that

$$\mathrm{Ric}_x(v,v) \geq K|v|^2 \qquad \forall x \in \mathbb{M}^n \text{ and } v \in T_x\mathbb{M}^n\,, \tag{2.6.17}$$

by applying the Bochner-Lichnerowicz formula it follows that (for all sufficiently regular function f)

$$\Gamma_2(f) \geq K\Gamma(f) + \frac{1}{n}(\Delta f)^2\,, \tag{2.6.18}$$

which goes under the name of *Bakry-Émery curvature-dimension condition* $\mathrm{BE}(K,n)$. It is possible to show that in fact the converse implication is also true: if a Riemannian manifold $\mathbb{M}^n$ satisfies the condition $\mathrm{BE}(K,N)$, then $n \leq N$ and $\mathrm{Ric} \geq K$, see [BGL14, Subsection 1.16 and Sections C.5, C.6] for further details.

It is convenient to define a suitable "integral" version of the Γ_2 operator (see [AGS15]), in the following form:

$$\begin{aligned}\mathbf{\Gamma}_2[f;\rho] &:= \int_{\mathbb{M}^n}\left[\frac{1}{2}\Gamma(f)\,\Delta\rho - \Gamma(f,\Delta f)\,\rho\right]\mathrm{d}\mathcal{V}\\ &= \int_{\mathbb{M}^n}\left[\frac{1}{2}\Gamma(f)\,\Delta\rho + \Gamma(f,\rho)\,\Delta f + (\Delta f)^2\,\rho\right]\mathrm{d}\mathcal{V} \qquad \forall (f,\rho)\in\mathbb{D}_\infty\,,\end{aligned} \tag{2.6.19}$$

where $\mathbb{D}_\infty$ stands for the algebra of functions defined as $\mathbb{D}_\infty := \mathbb{D} \cap L^\infty(\mathbb{M}^n)$. Note that, formally, (2.6.19) is obtained upon choosing $g = f$ in (2.6.16), multiplying by ρ and integrating by parts. The introduction of the multilinear form $\mathbf{\Gamma}_2$ provides a weak version of the Bakry-Émery condition: for every $(f,\rho) \in \mathbb{D}_\infty$ with $\rho \geq 0$ there holds

$$\mathbf{\Gamma}_2[f;\rho] \geq K\int_{\mathbb{M}^n}\Gamma(f)\,\rho\,\mathrm{d}\mathcal{V} + \frac{1}{n}\int_{\mathbb{M}^n}(\Delta f)^2\,\rho\,\mathrm{d}\mathcal{V}\,. \tag{2.6.20}$$

On a Riemannian manifold, the two formulations (2.6.18) and (2.6.20) turn out to be equivalent, and we will refer to both of them as $\mathrm{BE}(K,n)$. For a proof of such equivalence see e.g. [AGS15, Subsection 2.2].

We also consider a weighted version of the Dirichlet energy (2.6.14). More precisely, given $\rho \in L^\infty(\mathbb{M}^n)$ with $\rho \geq 0$, we set $\mathcal{E}_\rho : \mathbb{V} \to [0, +\infty)$ as

$$\mathcal{E}_\rho[f] := \int_{\mathbb{M}^n} \Gamma(f)\, \rho \, \mathrm{d}\mathcal{V} . \tag{2.6.21}$$

The associated dual weighted Dirichlet energy $\mathcal{E}^*_\rho : \mathbb{V}' \to [0, +\infty]$ is defined as

$$\frac{1}{2}\mathcal{E}^*_\rho[\ell] := \sup_{f\in\mathbb{V}} {}_{\mathbb{V}'}\langle \ell, f\rangle_{\mathbb{V}} - \frac{1}{2}\mathcal{E}_\rho[f] , \tag{2.6.22}$$

where $\mathbb{V}'$ is the dual of $\mathbb{V}$.

In Chapter 7 it will be crucial to connect any two given measures $\mu^0, \mu^1 \in \mathscr{M}_2^M(\mathbb{M}^n)$ through a curve in the 2-Wasserstein space that satisfies some additional regularity properties, according to the following definition.

Definition 2.6.4. *Let $\mu \equiv \mu^s$, $s \in [0,1]$, be a curve with values in $\mathscr{M}_2^M(\mathbb{M}^n)$. We say that μ is a* regular curve *if $\mu^s = \rho^s\mathcal{V}$ and the following hold:*

(i) There exists a constant $R > 0$ such that $\|\rho^s\|_{L^\infty(\mathbb{M}^n)} \leq R$ for every $s \in [0,1]$;

(ii) $\mu \in \mathsf{Lip}\big([0,1]; \big(\mathscr{M}_2^M(\mathbb{M}^n), \mathcal{W}_2\big)\big)$;

(iii) $\sqrt{\rho^s} \in \mathbb{V}$ and there exists a constant E such that

$$\int_{\mathbb{M}^n} \Gamma\big(\sqrt{\rho^s}\big)\, \mathrm{d}\mathcal{V} \leq E \qquad \forall s \in [0,1] .$$

Remark 2.6.5. If $\mu = \rho\mathcal{V}$ is a regular curve, in particular $\rho^s \in \mathbb{V}$ for every $s \in [0,1]$. Moreover, thanks to [AMS19, Lemma 8.1], condition (ii) ensures that $\rho \in \mathsf{Lip}([0,1]; \mathbb{V}')$.

The following density result, whose proof is contained in [AMS19, Lemma 12.2], allows one to approximate Wasserstein geodesics by means of regular curves.

Lemma 2.6.6. *Let $\mathbb{M}^n$ be a complete, connected Riemannian manifold with Ricci curvature bounded from below by a constant and $\mu^0, \mu^1 \in \mathscr{M}_2^M(\mathbb{M}^n)$. Then there exist a geodesic $\{\mu^s\}_{s\in[0,1]}$ connecting μ^0 and μ^1 and a sequence of regular curves $\{\mu_j^s\}_{j\in\mathbb{N}, s\in[0,1]} \subset \mathscr{M}_2^M(\mathbb{M}^n)$ such that*

$$\lim_{j\to\infty} \mathcal{W}_2\left(\mu_j^s, \mu^s\right) = 0 \qquad \forall s \in [0,1] \tag{2.6.23}$$

and

$$\limsup_{j\to\infty} \int_0^1 \left|\dot{\mu}_j^s\right|^2 \mathrm{d}s \leq \mathcal{W}_2^2\big(\mu^0, \mu^1\big) . \tag{2.6.24}$$

Furthermore, if $\mu^0 = \rho^0\mathcal{V}$ and $\mu^1 = \rho^1\mathcal{V}$ with ρ^0, ρ^1 $\mathcal{V}$-essentialy bounded with bounded support , then $\mu^s = \rho^s\mathcal{V}$ with ρ^s uniformly (w.r.t. s) bounded and bounded supported, and in addition to (2.6.23)–(2.6.24) also the following hold:

$$\lim_{j\to\infty} \left\|\rho_j^s - \rho^s\right\|_{L^p(\mathbb{M}^n)} = 0 \qquad \forall p \in [1,\infty) , \quad \forall s \in [0,1] , \tag{2.6.25}$$

$$\limsup_{j\to\infty} \sup_{s\in[0,1]} \left\|\rho_j^s\right\|_{L^\infty(\mathbb{M}^n)} < \infty . \tag{2.6.26}$$

To conclude, given a regular curve $\mu^s = \rho^s \mathrm{d}\mathcal{V}$ in the sense of Definition 2.6.4, by combining [AMS19, Theorem 6.6, formula (6.11)] and [AMS19, Theorem 8.2, formula (8.7)] we can deduce that the following key identity holds:

$$\int_0^1 |\dot{\mu}^s|^2 \,\mathrm{d}s = \int_0^1 \mathcal{E}^*_{\rho^s}\left[\tfrac{\mathrm{d}}{\mathrm{d}s}\rho^s\right] \mathrm{d}s\,, \tag{2.6.27}$$

where $\mathcal{E}^*_\rho$ is the dual weighted Dirichlet energy introduced in (2.6.22). Note that the r.h.s. of (2.6.27) does make sense, in view of Remark 2.6.5.

Part I

Optimal Entropy-Transport and distances

Chapter 3

Optimal Entropy-Transport problems

In this chapter we give a brief introduction to the theory of optimal *Entropy-Transport* problems, a generalization of optimal transport introduced by Liero, Mielke and Savaré [LMS18a] (see also [Chi+18], [KMV16]). We focus on the static formulation of these problems, that are constructed by relaxing the marginal constraints typical of optimal transport with the introduction of an entropic penalization. Motivated by our interest in the metric properties of the theory, we limit our exposition to problems involving a Polish space and the same penalizing functional for the two marginals. A more general formulation, where one considers two (possibly) different Hausdorff topological spaces and two (possibly) different entropic penalizations, can be found in [LMS18a].

3.1 Entropy functional

Let X be a Polish space and $F \in \Gamma_0(\mathbb{R}_+)$ be an admissible entropy function as defined in Section 2.1. The F-*divergence* (also called *Csiszár divergence* or *relative entropy*) is the functional $D_F : \mathscr{M}(X) \times \mathscr{M}(X) \to [0, +\infty]$ defined by

$$D_F(\gamma||\mu) := \int_X F(\sigma)\mathrm{d}\mu + F'_\infty \gamma^\perp(X), \qquad \gamma = \sigma\mu + \gamma^\perp, \tag{3.1.1}$$

where $\gamma = \sigma\mu + \gamma^\perp$ is the Lebesgue's decomposition of the measure γ with respect to μ that follows from Lemma 2.4.2. Note that, if F is superlinear, i.e. $F'_\infty = +\infty$, $D_F(\gamma||\mu) = +\infty$ if γ has a singular part with respect to μ.
The presence of the additional term $F'_\infty \gamma^\perp(X)$ in the definition (3.1.1), togheter with the convexity of F, is crucial for the lower semicontinuity property of these functionals (see the discussion in [San15, Section 7.1.2]).

Recalling the definition of $\hat{F}$ (2.1.10), we define the perspective divergence $\hat{D}_F : \mathscr{M}(X) \times \mathscr{M}(X) \to [0, +\infty]$ as

$$\hat{D}_F(\gamma||\mu) := \int_X \hat{F}\Big(\frac{\mathrm{d}\gamma}{\mathrm{d}\lambda}, \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\Big)\mathrm{d}\lambda, \tag{3.1.2}$$

where $\lambda \in \mathscr{M}(X)$ is any dominating measure of γ and μ, i.e. $\gamma \ll \lambda$, $\mu \ll \lambda$. It is easy to see that such a measure λ always exists (take for instance $\lambda = \gamma + \mu$) and $\hat{D}_F$ does not depend on λ since $\hat{F}$ is positively 1-homogeneous.

Lemma 3.1.1. *For every $\gamma, \mu \in \mathscr{M}(X)$ we have*

$$D_F(\gamma||\mu) = \hat{D}_F(\gamma||\mu).$$

Proof. Let $\gamma = \sigma\mu + \gamma^\perp$ be the Lebesgue's decomposition and define $\lambda := \mu + \gamma^\perp$. We observe that λ dominates μ and γ. Since μ is singular with respect to $\gamma^\perp$, there exist A, B Borel subsets such that

$$A \cup B = X, \qquad A \cap B = \emptyset, \qquad \mu(A) = 0, \qquad \gamma^\perp(B) = 0.$$

Put $\gamma := \rho\lambda$, $\mu := \tau\lambda$. The densities ρ, τ satisfies

$$\rho(x) = \begin{cases} 1 & \text{if } x \in A, \\ \sigma(x) & \text{if } x \in B, \end{cases} \qquad \tau(x) = \begin{cases} 0 & \text{if } x \in A, \\ 1 & \text{if } x \in B. \end{cases} \tag{3.1.3}$$

Thus,

$$\begin{aligned} \hat{D}_F(\gamma||\mu) &= \int_X \hat{F}(\rho,\tau)\mathrm{d}\lambda = \int_B \hat{F}(\rho,\tau)\mathrm{d}\lambda + \int_A \hat{F}(\rho,\tau)\mathrm{d}\lambda \\ &= \int_B \hat{F}(\sigma,1)\mathrm{d}\lambda + \int_A \hat{F}(1,0)\mathrm{d}\lambda = \int_B F(\sigma)\mathrm{d}\lambda + \int_A F'_\infty \mathrm{d}\lambda \\ &= \int_X F(\sigma)\mathrm{d}\mu + F'_\infty \gamma^\perp(X) = D_F(\gamma||\mu). \end{aligned} \tag{3.1.4}$$

□

Some comments are in order.

Remark 3.1.2. *We have defined the F-divergence only for functions F in the class $\Gamma_0(\mathbb{R}_+)$. In principle, one could work with convex and lower semicontinuous functions $F : [0,+\infty) \to \mathbb{R}$ such that $F(1) = 0$. In this case, an easy application of Jensen's inequality ensures that D_F is nonnegative between* probability *measures. However, if there exists a point $q \in [0,\infty]$ such that $F(q) < 0$, then D_F is not nonnegative between measures with different total mass. To see this, let us consider $\gamma := r\delta_x$ and $\mu := t\delta_x$ such that $r/t = q$. It is apparent that $D_F(\gamma||\mu) = \hat{F}(r,t)t = F(q)t < 0$.*

Let us finally mention that, given a convex and lower semicontinuous function $F : [0,+\infty) \to \mathbb{R}$ such that $F(1) = 0$, the function $\tilde{F}(s) := F(s) - c(s-1)$, where c is a subderivative of the function F at $s = 1$ (which exists since F is convex), is in the class $\Gamma_0(\mathbb{R}_+)$ and we have

$$D_{\tilde{F}}(\gamma||\mu) = D_F(\gamma||\mu) - c(\mu(X) - \gamma(X)).$$

In particular $D_{\tilde{F}}$ and D_F coincide for measures with the same total mass.

Remark 3.1.3. *Starting from a function $F \in \Gamma_0(\mathbb{R}_+)$, we have seen that we can construct the perspective function $\hat{F}$ thanks to (2.1.10), the F-divergence prescribed in (3.1.1) and the perspective divergence defined by (3.1.2). Moreover, Lemma 3.1.1 tell us that the F-divergence and the induced perspective divergence coincide. If we start instead with a lower semicontinuous, jointly convex and positively 1-homogeneous function $H : [0,+\infty) \times [0,+\infty) \to [0,\infty]$ such that $H(1,1) = 0$, we can define the $\mathcal{H}$-divergence $\mathcal{H} : \mathscr{M}(X) \times \mathscr{M}(X) \to [0,+\infty]$ by the formula*

$$\mathcal{H}(\gamma||\mu) := \int_X H\Big(\frac{\mathrm{d}\gamma}{\mathrm{d}\lambda}, \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\Big)\mathrm{d}\lambda, \tag{3.1.5}$$

where λ is any dominating measure of γ and μ; H also induces a function $f \in \Gamma_0(\mathbb{R}_+)$ simply by taking $f(s) := H(s,1)$. Thus, in studying Csiszár divergences, we have two *different but equivalent points of view and, depending on the circumstances, we may choose to work with functions of* 1 *or* 2 *variables.*

We now collect some useful properties of the relative entropies. For the proof see [LMS18a, Section 2.4].

Proposition 3.1.4. *For every $\gamma, \mu \in \mathscr{M}(X)$ and for every $F \in \Gamma_0(\mathbb{R}_+)$ we have the following duality result:*

$$D_F(\gamma||\mu) = \sup\Big\{ \int_X \psi \,\mathrm{d}\mu + \int_X \phi \,\mathrm{d}\gamma : \ \phi, \psi \in \mathcal{C}_b(X), \ (\phi(x), \psi(x)) \in \mathfrak{F} \Big\} \tag{3.1.6}$$

$$= \sup\Big\{ \int_X \psi \,\mathrm{d}\mu - \int_X R^*(\psi) \,\mathrm{d}\gamma : \ \psi, R^*(\psi) \in \mathcal{C}_b(X) \Big\}, \tag{3.1.7}$$

where $\mathfrak{F}$ is the set associated to F via the definition (2.1.6) and R^* is the Legendre conjugate of the reverse entropy R.

Proposition 3.1.5. *The functional D_F is jointly convex and lower semicontinuous in $\mathscr{M}(X) \times \mathscr{M}(X)$. More generally, if $F \in \Gamma_0(\mathbb{R}_+)$ is the pointwise limit of an increasing sequence $(F_n) \subset \Gamma_0(\mathbb{R}_+)$ and $\gamma, \mu \in \mathscr{M}(X)$ are the weak limit of a sequence $(\gamma_n, \mu_n) \subset \mathscr{M}(X) \times \mathscr{M}(X)$ we have*

$$\liminf D_{F_n}(\gamma_n||\mu_n) \geq D_F(\gamma||\mu).$$

Proposition 3.1.6. *If $\mathcal{K} \subset \mathscr{M}(X)$ is bounded and $F'_\infty > 0$ then the set*

$$\mathbf{K}_C := \{\gamma \in \mathscr{M}(X) : D_F(\gamma||\mu) \leq C, \ \textit{for some } \mu \in \mathcal{K}\}$$

is bounded for every $C \geq 0$. Moreover, if $\mathcal{K}$ is also equally tight and F is superlinear, then $\mathbf{K}_C$ is equally tight for every $C \geq 0$.

3.2 Entropy-Transport problem

3.2.1 Primal formulation of Entropy-Transport problem

Definition 3.2.1. *Let us fix the setting:*

- *X is a given Polish space and we denote with the bold character $\boldsymbol{X}$ the product space $X \times X$. We denote by $\pi^i : \boldsymbol{X} \to X$ the projection map $\pi^i(x_1, x_2) = x_i$, $i = 1, 2$.*
- *$F \in \Gamma_0(\mathbb{R}_+)$ is a given admissible entropy function and $\boldsymbol{c}$ is a proper, lower semicontinuous cost function $\boldsymbol{c} : \boldsymbol{X} \to [0, +\infty]$.*
- *$\mu_i \in \mathscr{M}(X)$, $i = 1, 2$, are two measures that satisfy the compatibility condition*

  $$J := \big(m_1 \mathrm{D}(F)\big) \cap \big(m_2 \mathrm{D}(F)\big) \neq \emptyset,$$

  where $m_i := \mu_i(X)$.

We will refer to these assumptions as the basic setting.

We say that the basic setting is *coercive* if at least one of the following conditions is satisfied:

$$F \text{ is superlinear;} \tag{3.2.1}$$

$$F'_\infty + \inf \boldsymbol{c} > 0 \text{ and } \boldsymbol{c} \text{ has compact sublevels.} \tag{3.2.2}$$

Let $\boldsymbol{\gamma} \in \mathscr{M}(\boldsymbol{X})$. We denote by $\gamma_i := (\pi^i)_\sharp \boldsymbol{\gamma}$ the marginals of $\boldsymbol{\gamma}$.

We are now ready to define the primal formulation of the Entropy-Transport problem.

Definition 3.2.2. *Let us consider the basic setting defined in 3.2.1. The Entropy-Transport functional between the measures $\mu_1, \mu_2 \in \mathscr{M}(X)$ is the functional*

$$\begin{aligned} &\mathcal{ET}(\,\cdot\,||\mu_1,\mu_2) : \boldsymbol{X} \to [0,+\infty], \\ &\mathcal{ET}(\boldsymbol{\gamma}||\mu_1,\mu_2) := D_F(\gamma_1||\mu_1) + D_F(\gamma_2||\mu_2) + \int_{\boldsymbol{X}} \boldsymbol{c}(x_1,x_2)\mathrm{d}\boldsymbol{\gamma}(x_1,x_2). \end{aligned} \tag{3.2.3}$$

We define the Entropy-Transport problem between μ_1 and μ_2 as the minimization problem

$$\mathrm{ET}(\mu_1,\mu_2) := \inf_{\boldsymbol{\gamma}\in\mathscr{M}(\boldsymbol{X})} \mathcal{ET}(\boldsymbol{\gamma}||\mu_1,\mu_2). \tag{3.2.4}$$

To highlight the role of the entropy function F and the cost function $\boldsymbol{c}$, we also say that ET *is the cost of the Entropy-Transport problem induced by $(F,\boldsymbol{c})$.*

We denote by $\mathsf{OPT}_{\mathrm{ET}}(\mu_1,\mu_2)$ the set of minimizers of (3.2.4).

We say that the problem is *feasible* if the functional $\mathcal{ET}$ is not identically $+\infty$. A sufficient condition for feasibility, at least in the nondegenerate case $m_1 m_2 \neq 0$, is the existence of two functions $g_i \in L^1(X,\mu_i)$, $i = 1,2$, such that

$$\boldsymbol{c}(x_1,x_2) \leq g(x_1) + g(x_2).$$

In fact, in this case one can easily check that the plan

$$\tilde{\boldsymbol{\gamma}} := \frac{\theta}{m_1 m_2}\mu_1 \otimes \mu_2$$

provides the estimate

$$\mathrm{ET}(\mu_1,\mu_2) \leq \mathcal{ET}(\tilde{\boldsymbol{\gamma}}||\mu_1,\mu_2) \leq \sum_{i=1}^{2} F\Big(\frac{\theta}{m_i}\Big) m_i + \theta \sum_{i=1}^{2} m_i^{-1} \|g_i\|_{L^1(X,\mu_i)},$$

for every $\theta \in [0,+\infty)$.

The first general result in the theory of Entropy-Transport problems is the following Theorem (see [LMS18a, Theorem 3.3]).

Theorem 3.2.3. *Let us assume that, in the basic setting, the problem is feasible and that at least one of the coercivity conditions (3.2.1), (3.2.2) holds. Then the set $\mathsf{OPT}_{\mathrm{ET}}(\mu_1,\mu_2) \subset \mathscr{M}(\boldsymbol{X})$ is compact, convex and not empty.*

3.2.2 Dual formulation

As in the case of classical transport problems, also Entropy-Transport problems admit a dual formulation. Here we assume the basic setting defined in the previous section, and we recall that R^* denotes the Legendre transform of the inverse entropy function R induced by F.

Definition 3.2.4. *We define the* dual constraint set *as*

$$\boldsymbol{\Psi} := \{\boldsymbol{\psi} = (\psi_1,\psi_2) \in \mathcal{C}_b(X) \times \mathcal{C}_b(X) : R^*(\psi_1) + R^*(\psi_2) \leq \boldsymbol{c}\}, \tag{3.2.5}$$

where $R^*(\psi_1) + R^*(\psi_2) \le \boldsymbol{c}$ *stands for*

$$R^*(\psi_1(x_1)) + R^*(\psi_2(x_2)) \le \boldsymbol{c}(x_1, x_2) \quad \textit{for every } (x_1, x_2) \in \boldsymbol{X}.$$

The dual functional *is defined as*

$$\mathcal{D}(\boldsymbol{\psi}||\mu_1, \mu_2) = \int_X \psi_1 \, \mathrm{d}\mu_1 + \int_X \psi_2 \, \mathrm{d}\mu_2. \tag{3.2.6}$$

The next Theorem shows the connection between the primal formulation and the dual formulation. For a proof we refer to [LMS18a, Theorem 4.11].

Theorem 3.2.5. *Let us assume the basic setting defined in 3.2.1 and at least one of the coercivity conditions* (3.2.1), (3.2.2). *We have*

$$\mathrm{ET}(\mu_1, \mu_2) = \sup_{\boldsymbol{\psi} \in \boldsymbol{\Psi}} \mathcal{D}(\boldsymbol{\psi}||\mu_1, \mu_2). \tag{3.2.7}$$

A direct consequence of the previous Theorem is the following:

Corollary 3.2.6. *The functional* ET : $\mathscr{M}(X) \times \mathscr{M}(X) \to [0, +\infty]$ *is convex and positively* 1*-homogeneous, thus subadditive.*

3.2.3 Marginal perspective cost

In this section we introduce the marginal perspective cost. As we will see, this function arises by studying Entropy-Transport problems between Dirac masses.

First of all, given a number $c \in [0, +\infty)$ and an admissible entropy function F, we define the function $H_c : [0, +\infty) \times [0, +\infty) \to [0, +\infty]$ as the lower semicontinuous envelope of the function

$$\tilde{H}_c(r_1, r_2) := \inf_{\theta > 0} \Big(R\big(\frac{r_1}{\theta}\big) + R\big(\frac{r_2}{\theta}\big) + c \Big) \theta, \tag{3.2.8}$$

where R is the reverse entropy function of F. The function $\tilde{H}_c$ can be also computed as

$$\tilde{H}_c(r_1, r_2) = \inf_{\theta > 0} F\Big(\frac{\theta}{r_1}\Big) r_1 + F\Big(\frac{\theta}{r_2}\Big) r_2 + \theta c, \quad \text{if } r_1 \wedge r_2 > 0, \tag{3.2.9}$$

or in terms of the perspective function as

$$\tilde{H}_c(r_1, r_2) = \inf_{\theta > 0} \hat{F}(\theta, r_1) + \hat{F}(\theta, r_2) + \theta c \tag{3.2.10}$$

$$= \inf_{\theta > 0} \hat{R}(r_1, \theta) + \hat{R}(r_2, \theta) + \theta c. \tag{3.2.11}$$

If $c = +\infty$ we set

$$H_\infty(r_1, r_2) := F(0) r_1 + F(0) r_2. \tag{3.2.12}$$

The following Lemma, proved in [LMS18a, Lemma 5.3], gives a dual characterization of H_c.

Lemma 3.2.7. *For every* $c \in [0, +\infty]$ *the function* H_c *can be represented as*

$$H_c(r_1, r_2) = \sup\{r_1 \psi_1 + r_2 \psi_2 : \psi_i \in \mathrm{D}(R^*), \ R^*(\psi_1) + R^*(\psi_2) \le c\}. \tag{3.2.13}$$

In particular H_c is lower semicontinuous, convex, symmetric and positively 1-homogeneous with respect to (r_1, r_2), increasing and concave with respect to c. Moreover, H_c coincides with $\tilde{H}_c$ in the interior of its domain.

If X is a Polish spaces and $\boldsymbol{c}(x_1, x_2)$ is a cost function $\boldsymbol{c} : \boldsymbol{X} \to [0, +\infty]$, we define the induced *marginal perspective cost* as the function

$$H : X \times [0, +\infty) \times X \times [0, +\infty) \to [0, +\infty]$$

such that

$$H(x_1, r; x_2, t) := H_{\boldsymbol{c}(x_1,x_2)}(r, t). \tag{3.2.14}$$

At least in the superlinear case, it is a direct consequence of the primal formulation of optimal Entropy-Transport problems that for every $x_i \in X$ and $r_i > 0$, $i = 1, 2$, it holds

$$H(x_1, r; x_2, t) = \mathrm{ET}(r\delta_{x_1}, t\delta_{x_2}). \tag{3.2.15}$$

3.2.4 The homogeneous formulation

In this section, we derive a different formulation of the optimal Entropy-Transport problem. It is related to the marginal perspective cost previously defined. In the next chapter, we will take advantage of this formulation in order to connect the metric properties of the cost ET with the corresponding properties of the marginal perspective cost H on the cone over X. We tacitly assume the basic setting defined in 3.2.1 and at least one of the coercivity conditions (3.2.1), (3.2.2).

We put $Y := X \times [0, +\infty)$ and we endowed this space with the product topology. The points of Y are denoted by $y = (x, r)$, and we set $\boldsymbol{Y} := Y \times Y$. The maps π^{x_i}, $i = 1, 2$, define the projections from $\boldsymbol{Y}$ to X, with obvious meaning.

Let $p > 0$, we say that a plan $\boldsymbol{\alpha} \in \mathscr{M}(\boldsymbol{Y})$ lies in $\mathscr{M}_p(\boldsymbol{Y})$ if

$$\int_{\boldsymbol{Y}} (r_1^p + r_2^p) \mathrm{d}\boldsymbol{\alpha} < +\infty. \tag{3.2.16}$$

Let $\boldsymbol{\alpha} \in \mathscr{M}_p(\boldsymbol{Y})$, then the measure $r_i^p \boldsymbol{\alpha} \in \mathscr{M}(\boldsymbol{Y})$ and we can define the p-*homogeneous marginal* $h_i^p(\boldsymbol{\alpha})$ of the measure $\boldsymbol{\alpha}$ as

$$h_i^p(\boldsymbol{\alpha}) := \pi^{x_i}_\sharp(r_i^p \boldsymbol{\alpha}) \in \mathscr{M}(X). \tag{3.2.17}$$

Given $\mu_1, \mu_2 \in \mathscr{M}(X)$, we define the

$$\mathcal{H}^p(\mu_1, \mu_2) := \{\boldsymbol{\alpha} \in \mathscr{M}_p(\boldsymbol{Y}) : h_i^p(\boldsymbol{\alpha}) = \mu_i, \, i = 1, 2\}. \tag{3.2.18}$$

Then we have ([LMS18a, Theorem 5.8])

Proposition 3.2.8. *Let H be the marginal perspective cost induced by $(F, \boldsymbol{c})$. For every $\mu_1, \mu_2 \in \mathscr{M}(X)$ it holds*

$$\mathrm{ET}(\mu_1, \mu_2) = \min_{\boldsymbol{\alpha} \in \mathcal{H}^p(\mu_1,\mu_2)} \int_{\boldsymbol{Y}} H(x_1, r_1^p; x_2, r_2^p) \mathrm{d}\boldsymbol{\alpha}. \tag{3.2.19}$$

An important feature of the homogeneous marginals is that they are invariant with respect to dilations: let $\vartheta : \boldsymbol{Y} \to (0, \infty)$ be a map in $L^p(\boldsymbol{Y}, \boldsymbol{\alpha})$ and define the product map $\mathrm{prd}_\vartheta(\boldsymbol{y}) := (x_1, r_1/\vartheta(\boldsymbol{y}); x_2, r_2/\vartheta(\boldsymbol{y}))$ and the dilation measure as

$\mathrm{dil}_{\vartheta,p}(\boldsymbol{\alpha}) := (\mathrm{prd}_\vartheta)_\sharp(\vartheta^p\boldsymbol{\alpha}) \in \mathscr{M}_p(\boldsymbol{Y})$, i.e. the measure that satisfies

$$\int_{\boldsymbol{Y}} \varphi(\boldsymbol{y})\mathrm{d}\left(\mathrm{dil}_{\vartheta,2}(\boldsymbol{\alpha})\right) = \int_{\boldsymbol{Y}} \varphi\left(x_1, \frac{r_1}{\vartheta(\boldsymbol{y})}; x_2, \frac{r_2}{\vartheta(\boldsymbol{y})}\right)\vartheta^p(\boldsymbol{y})\mathrm{d}\boldsymbol{\alpha}(\boldsymbol{y}), \quad \forall\varphi \in \mathcal{C}_b(\boldsymbol{Y}). \tag{3.2.20}$$

Then, using (3.2.17), it is not difficult to show that

$$h_i^p\left(\mathrm{dil}_{\vartheta,p}(\boldsymbol{\alpha})\right) = h_i^p(\boldsymbol{\alpha}) \qquad \forall \boldsymbol{\alpha} \in \mathscr{M}_p(\boldsymbol{Y}). \tag{3.2.21}$$

3.2.5 Pure entropy problems: the marginal perspective function

In this section we study an important class of Entropy-Transport problems: the pure entropy problems.

Let us consider the following possible choices for the entropy function and the cost function:

$$F \in \Gamma_0(\mathbb{R}_+),\ F'_\infty = +\infty \quad \text{and} \quad \boldsymbol{c}(x_1,x_2) := \begin{cases} 0 & \text{if } x_1 = x_2, \\ +\infty & \text{otherwise.} \end{cases} \tag{3.2.22}$$

In this situation, the optimal plan for the induced Entropy-Transport problem between μ_1 and μ_2 defined in 3.2.2 is concentrated on the diagonal of $X \times X$, so that the marginals γ_1, γ_2 must coincide: we denote them by γ. By the superlinearity of the entropy, we also have that $\gamma \ll \mu_i$, $i = 1, 2$. Everything can thus be expressed in terms of a dominating measure λ, and by exploiting the explicit formulation of the problem we obtain

$$\mathrm{ET}(\mu_1,\mu_2) = \int_X H_0\Big(\frac{\mathrm{d}\mu_1}{\mathrm{d}\lambda}, \frac{\mathrm{d}\mu_2}{\mathrm{d}\lambda}\Big)\mathrm{d}\lambda, \tag{3.2.23}$$

where H_0 is the function defined in Subsection 3.2.3. In order to emphasize the role of the admissible entropy function F in the construction of H_0, we put $H_F := H_0$ so that $H_F : [0,+\infty) \times [0,+\infty) \to [0,+\infty]$ is the lower semicontinuous envelope of the function

$$\tilde{H}_F(r,t) := \inf_{\theta>0}\left(R\big(\frac{r}{\theta}\big) + R\big(\frac{t}{\theta}\big)\right)\theta. \tag{3.2.24}$$

An equivalent definition of $\tilde{H}_F$ can be given in term of the induced perspective functions $\hat{F}$ or $\hat{R}$ by:

$$\tilde{H}_F(r,t) = \inf_{\theta>0} \hat{F}(\theta,r) + \hat{F}(\theta,t) = \inf_{\theta>0} \hat{R}(r,\theta) + \hat{R}(t,\theta). \tag{3.2.25}$$

We call H_F the *marginal perspective function* induced by F.

Recalling the perspective formulation of the entropy functional given in (3.1.2), we recognize that the Entropy-Transport cost (3.2.23) corresponds to the f-divergence induced by the admissible entropy function $f(s) := H_F(s,1)$. It is thus justified to speak of pure entropy problems.

In general, an F-divergence is not a symmetric functional (take for instance $F(s) := s\log(s) - s + 1$), so that the construction developed above can be seen as a natural procedure to replace the relative entropy induced by F with a new *symmetric* Csiszár divergence.

We notice that infimum in the definition of $\tilde{H}_F$ is a minimum and it occurs in the interval $[r,t]$ (without loss of generality we are assuming $r \le t$): to see this, it is enough to note that the function $\theta \mapsto \hat{F}(\theta, r) + \hat{F}(\theta, t)$ is lower semicontinuous and it is decreasing in $[0, r]$ and increasing in $[t, +\infty)$.

We conclude the chapter by considering different examples of admissible entropy functions and by computing the expressions of the corresponding marginal perspective functions, at least in the case $r \wedge t > 0$. In this regard, it is useful to recall that the function $\tilde{H}_F$ coincides with the marginal perspective function H_F in the interior of its domain (see Lemma 3.2.7).

Example 1. *(Indicator functions) The indicator function of the closed interval with endpoints a and b, $0 \le a \le 1 \le b \le +\infty$, is defined by*

$$I_{[a,b]}(s) = \begin{cases} 0 & \text{if } \; s \in [a,b], \\ +\infty & \text{if } \; s \notin [a,b]. \end{cases} \tag{3.2.26}$$

When $F = I_{[a,b]}$ one obtains

$$H_{I_{[a,b]}}(r,t) = \begin{cases} 0 & \text{if } \frac{a}{b} \le \frac{r}{t} \le \frac{b}{a}, \\ +\infty & \text{otherwise}, \end{cases} \tag{3.2.27}$$

where $\frac{b}{a} = +\infty$ if $a = 0$ and $\frac{a}{b} = 0$ if $b = +\infty$.

Example 2. *(χ^α divergences) Given a parameter $\alpha \ge 1$, the χ^α divergence is induced by the function*

$$\chi^\alpha(s) = |s-1|^\alpha. \tag{3.2.28}$$

$\chi^1 = |s-1|$ is the famous total variation entropy.

The entropy function $F = \chi^\alpha$ gives rise to the marginal perspective function

$$H_{\chi^\alpha}(r,t) = \frac{|r-t|^\alpha}{(r+t)^{\alpha-1}}. \tag{3.2.29}$$

We can recognize the expression of the so-called Puri-Vincze divergence.

Example 3. *(Matusita's divergences) For $0 < a \le 1$ the Matusita's divergence is given by the function $M_a(s) = |s^a - 1|^{\frac{1}{a}}$. Clearly $\chi^1 = M_1$.*

When $F = M_a$ it is easy to see that

$$H_{M_a}(r,t) = 2^{1-\frac{1}{a}} |r^a - t^a|^{\frac{1}{a}}. \tag{3.2.30}$$

It is interesting to note that except for the constant factor $2^{1-\frac{1}{a}}$, the Matusita function M_a remains invariant after the minimizing procedure (3.2.24). We will come back to this point in section 4.1.2.

Example 4. *(Power like entropies) Let p be any real number. We call power-like entropy of order p the function $U_p : [0, +\infty) \to [0, +\infty]$ characterized by*

$$U_p \in \mathcal{C}^\infty(0,+\infty),\ U_p(1) = U_p'(1) = 0,\ U_p''(s) = s^{p-2},\ U_p(0) := \lim_{s \downarrow 0} U_p(s). \tag{3.2.31}$$

The function U_p can be computed explicitly and one gets:

$$\begin{cases} U_p(s) = \frac{1}{p(p-1)}(s^p - p(s-1) - 1) & \text{if } \; p \ne 0, 1, \\ U_1(s) = s\ln(s) - s + 1, \\ U_0(s) = s - 1 - \ln(s), \end{cases} \tag{3.2.32}$$

with $U_p(0) = 1/p$ *for* $p > 0$ *and* $U_p(0) = +\infty$ *for* $p \leq 0$. *This family of functions, also called Dichotomy Class, was introduced by Liese and Vajda [LV87],[Vaj89].*

Given $F = U_p$, *we obtain the following expression:*

$$\begin{cases} H_{U_p}(r,t) = \frac{1}{p}\Big[r + t - 2^{\frac{p}{p-1}}(r^{1-p} + t^{1-p})^{\frac{1}{1-p}}\Big] & p \neq 0, 1, \\ H_{U_1}(r,t) = r + t - 2\sqrt{rt}, \\ H_{U_0}(r,t) = r\ln r + t\ln t - (r+t)\ln\left(\frac{r+t}{2}\right). \end{cases} \tag{3.2.33}$$

We can recognize some well-known statistical functionals: for example in the logarithmic entropy case $p = 1$ *it appears the Hellinger distance*

$$H_{U_1}(r,t) = (\sqrt{r} - \sqrt{t})^2. \tag{3.2.34}$$

We have already noticed that the same function is obtained starting from the entropy $U_{\frac{1}{2}}(s) = 2(\sqrt{s} - 1)^2 = 2M_{\frac{1}{2}}$.

For $p = 0$ *we have the Jensen-Shannon divergence, a squared distance between measures derived from the Kullback-Leibler divergence ([ES03]).*

The quadratic entropy $U_2(s) = \frac{1}{2}(s-1)^2$ *gives rise to the triangular discrimination*

$$H_{U_2}(r,t) = \frac{1}{2}H_{\chi^2}(r,t) = \frac{1}{2}\frac{(r-t)^2}{(r+t)}. \tag{3.2.35}$$

Example 5. *(Power-logarithmic entropies) Given a real number* $p \geq 1$, *we call power-logarithmic entropy of order* p *the function* $V_p : [0, +\infty) \to [0, +\infty]$ *defined as*

$$V_p(s) := s^p - p\ln(s) - 1, \quad s > 0, \tag{3.2.36}$$

and $V_p(0) = +\infty$. *It is easy to see that* $V_p \in \mathcal{C}^\infty(0, +\infty)$ *and* $V_p(0) = \lim_{s\downarrow 0} V_p(s)$.

Starting from the power-logarithmic entropy of order p *one gets:*

$$H_{V_p}(r,t) = (r+t)\ln\left[\frac{rt(r^{p-1} + t^{p-1})}{r+t}\right] - p\big(r\ln(t) + t\ln(r)\big). \tag{3.2.37}$$

As expected, $H_{V_1} = H_{U_0}$ *since* $V_1 = U_0$. *When* $p = 2$, *one obtains the symmetric Kullback-Leibler divergence [KL51]:*

$$H_{V_2}(r,t) = (r-t)\ln\left(\frac{r}{t}\right). \tag{3.2.38}$$

Example 6. *(Double power entropies) Given two parameters* p, q *such that* $p \geq 1$, $0 < q \leq 1$ *and* $p \neq q$, *or* $p < 0$, $q \geq 1$, *the double power entropy of order* p, q *is given by*

$$W_{p,q}(s) := qs^p - ps^q + p - q, \quad s > 0. \tag{3.2.39}$$

$W_{p,q}$ *is a strictly convex function,* $W_{p,q} \in \mathcal{C}^\infty(0, +\infty)$, *and it is extended in* 0 *by continuity so that* $W_{p,q}(0) = p - q$ *when* p, q *are positive,* $W_{p,q}(0) = +\infty$ *when* $p < 0$.

A direct computation shows that:

$$H_{W_{p,q}}(r,t) = (q-p)rt\left[\frac{(r^{q-1} + t^{q-1})^p}{(r^{p-1} + t^{p-1})^q}\right]^{\frac{1}{p-q}} - (q-p)(r+t). \tag{3.2.40}$$

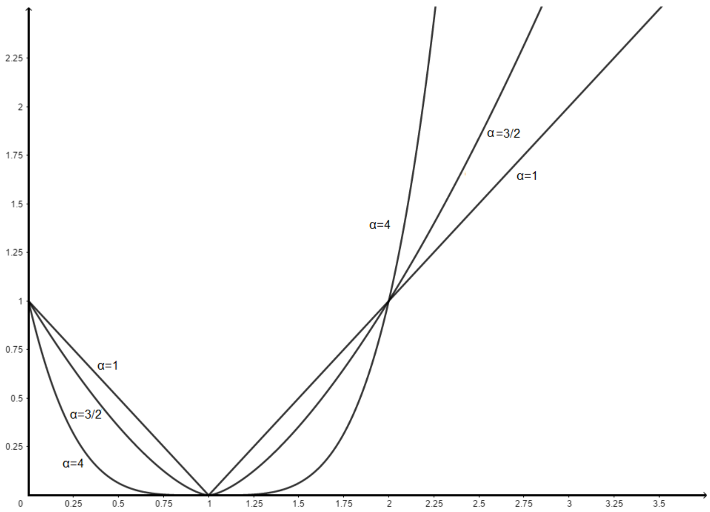

FIGURE 3.1: χ^α divergences

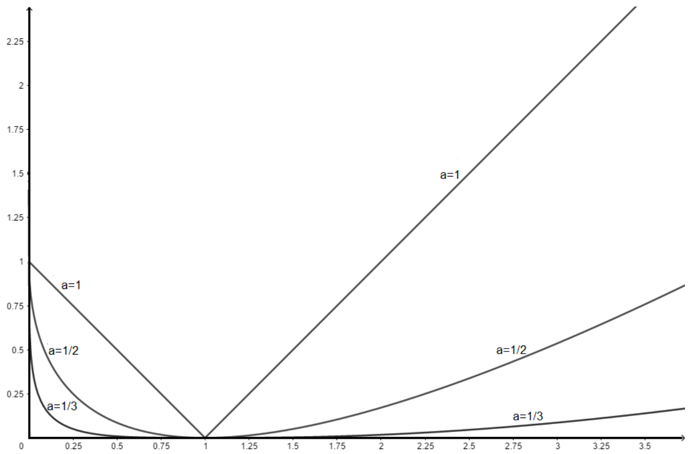

FIGURE 3.2: Matusita's divergences

For example, when $p = 3/2, q = 1/2$ one gets

$$H_{W_{\frac{3}{2},\frac{1}{2}}}(r,t) = r + t - (rt)^{\frac{1}{4}}(\sqrt{r} + \sqrt{t}). \tag{3.2.41}$$

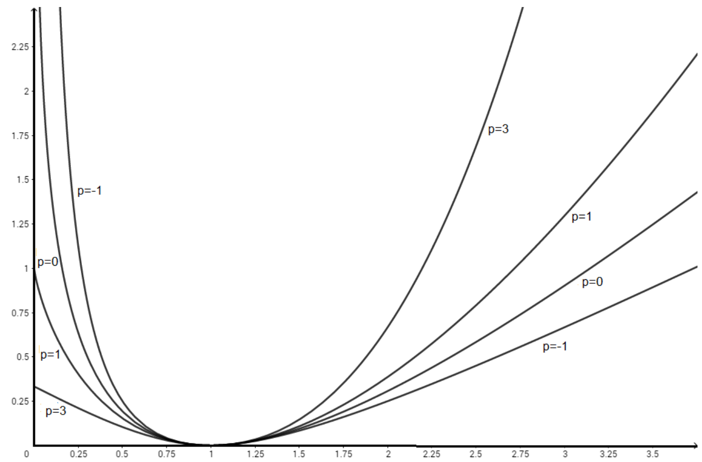

FIGURE 3.3: Power-like entropies

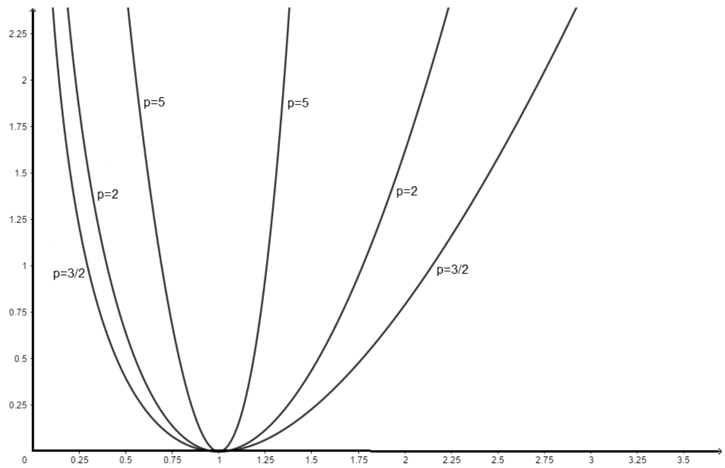

FIGURE 3.4: Power logarithmic entropies

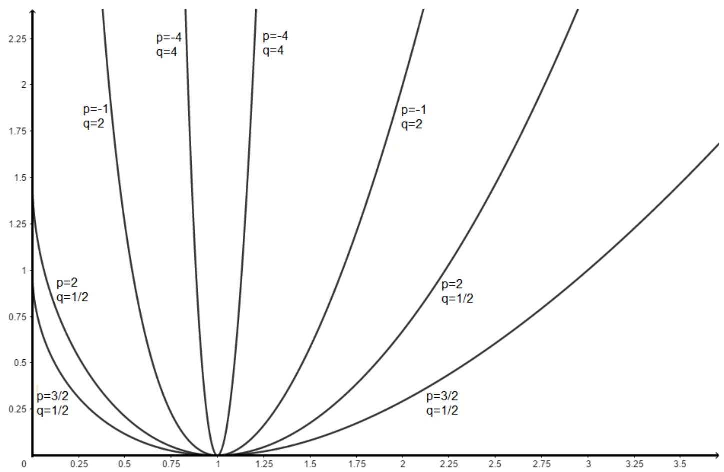

FIGURE 3.5: Double-power entropies

Chapter 4

Metric properties of Entropy-Transport problems

The aim of this chapter is studying the metric properties of Entropy-Transport problems. In our main result (Theorem 4.2.14), we produce a new class of Entropy-Transport distances. A crucial feature of these metrics is that, while maintaining some of the properties of classical transport distances, they can compare measures with possibly different total mass and they may thus be useful in all the situations when normalization is not feasible. The study performed gives also new insights on the classical notion of Csiszár F-divergence, that corresponds to the pure entropic setting of the theory.

4.1 Metric properties of homogeneous divergences

4.1.1 F-divergences and triangle inequality

In Section 3.2.5 we have seen that we can associate to a function $F \in \Gamma_0(\mathbb{R}_+)$ the marginal perspective function H_F.

By Lemma 3.2.7 it follows that H_F is nonnegative and symmetric. Moreover, if $F(s)$ has a strict minimum at $s = 1$, one can show that $H_F(r,t) = 0$ if and only if $r = t$ (see below Proposition 4.2.6 for a more general statement, where also a cost function is considered).

We begin in this section the discussion regarding the last property that H_F has to fulfill in order to be a metric on $[0, +\infty)$: the triangle inequality.

Since we will prove that the total variation is the only divergence that is also a metric, is natural to discuss when the power H_F^a is a distance on $[0, +\infty)$, for $a \in (0, 1]$.

The study of the metric properties of the function H_F in order to produce entropy distances on $\mathscr{M}(X)$ is justified by the following Proposition.

Proposition 4.1.1. *Let $F \in \Gamma_0(\mathbb{R}_+)$. The functional D_F^a is a distance on $\mathscr{M}(X)$ if and only if $\hat{F}^a$ is a distance on $[0, +\infty)$, $a \in (0, 1]$.*

Proof. We recall that, thanks to Lemma 3.1.1, $D_F(\mu_1||\mu_2) = \hat{D}_F(\mu_1||\mu_2)$.

Let us suppose that D_F^a is a distance. Then $\hat{F}^a$ is a distance since $D_F^a(r\delta_x||t\delta_x) = \hat{F}^a(r,t)$ for every $r, t \in [0, +\infty)$, where δ_x denotes the Dirac measure at $x \in X$.

For the converse implication, let $\hat{F}^a$ be a distance on the nonnegative real numbers. It is apparent that the (power) of the F-divergence is nonnegative, symmetric and $D_F^a(\mu_1||\mu_2) = 0$ if $\mu_1 = \mu_2$. Moreover, since $\hat{F}^a$ is a distance, it follows that $F(s) = 0$ if and only if $s = 1$, which also implies $F'_\infty > 0$. Thus, if $D_F^a(\mu_1||\mu_2) = 0$ we easily see using the definition of F-divergence that $\mu_1 = \mu_2$. Let $\mu_1, \mu_2, \mu_3 \in \mathscr{M}(X)$ and consider $\lambda := \mu_1 + \mu_2 + \mu_3$ so that we can write $\mu_i = \tau_i \lambda$, $i = 1, 2, 3$. Then

$$D_F^a(\mu_1||\mu_3) = \Big(\int_X \hat{F}(\tau_1,\tau_3)\mathrm{d}\lambda\Big)^a \le \Big(\int_X \big(\hat{F}^a(\tau_1,\tau_2) + \hat{F}^a(\tau_2,\tau_3)\big)^{1/a}\mathrm{d}\lambda\Big)^a$$
$$\le \Big(\int_X \big(\hat{F}^a(\tau_1,\tau_2)\big)^{1/a}\mathrm{d}\lambda\Big)^a + \Big(\int_X \big(\hat{F}^a(\tau_2,\tau_3)\big)^{1/a}\mathrm{d}\lambda\Big)^a = D_F^a(\mu_1||\mu_2) + D_F^a(\mu_2||\mu_3), \tag{4.1.1}$$

where we have used the triangle inequality for $\hat{F}^a$ and the Minkowski inequality. □

We recall this simple Lemma:

Lemma 4.1.2. *Let (X, d) be a metric space and $f : [0, +\infty) \to [0, +\infty)$ be a concave function such that $f(r) = 0$ if and only if $r = 0$. Then $(X, f(\mathsf{d}))$ is a metric space inducing the same topology.*

Proof. $f(\mathsf{d}(x_1, x_2)) \ge 0$ and $f(\mathsf{d}(x_1, x_2)) = 0$ if and only if $\mathsf{d}(x_1, x_2) = 0$ which implies $x_1 = x_2$. It is clear that

$$(x_1, x_2) \mapsto f(\mathsf{d}(x_1, x_2))$$

is a symmetric function. Since f is concave and $f(r) \ge 0$, f is also increasing and subadditive, thus

$$f\big(\mathsf{d}(x_1, x_3)\big) \le f\big(\mathsf{d}(x_1, x_2) + \mathsf{d}(x_2, x_3)\big) \le f\big(\mathsf{d}(x_1, x_2)\big) + f\big(\mathsf{d}(x_2, x_3)\big).$$

The function f is continuous because it is concave and finite valued, so that $f(\mathsf{d})$ is topological equivalent to d. □

An immediate consequence of the previous Lemma is that if H_F^a is a metric, then H_F^b is a metric for every $b \in (0, a]$.

The convexity of the function H_F implies that

$$H_F(r,t) \ge H_F(s,t) \text{ and } H_F(r,s) \le H_F(r,t) \text{ for every } 0 \le r \le s \le t. \tag{4.1.2}$$

Using the symmetry, the 1-homogeneity of the function H_F and the property (4.1.2), it follows that H_F^a satisfies the triangle inequality if and only if

$$H_F^a(u,1) \le H_F^a(u,v) + H_F^a(v,1) = v^a H_F^a\Big(\frac{u}{v},1\Big) + H_F^a(v,1), \quad \text{for every } 0 \le u < v < 1. \tag{4.1.3}$$

A last useful remark is that

$$\lim_{u\downarrow 0} H_F(u,1) < +\infty \tag{4.1.4}$$

is a necessary condition for the existence of a power a such that H_F^a is a metric.

Regarding the examples of marginal perspective function discussed in Section 3.2.5, it was proved by Kafka, Osterreicher and Vincze [KOV91] that $H_{\chi^\alpha}^a$ is a metric when $a = 1/\alpha$.

About the Matusita's divergences, it is apparent that $H_{M_a}^a$ is a distance.

When $p > 1$, $\lim_{u\downarrow 0} H_{V_p}(u,1) = +\infty$ so that, except for the case $p = 1$, the power-logarithmic entropy is not a metric for every power a.

We now turn the attention to the function $H_p := H_{U_p}$. It has the following expression

$$\begin{cases} H_p(r,t) = \frac{2}{p}\Big[\mathfrak{M}_1(r,t) - \mathfrak{M}_{1-p}(r,t)\Big], & \text{if } p \neq 0, \\ H_0(r,t) = r\ln r + t\ln t - (r+t)\ln\left(\frac{r+t}{2}\right), \end{cases} \tag{4.1.5}$$

that is also valid when $rt = 0$ with the convention $0\ln(0) = 0$. Here $\mathfrak{M}_q(r,t)$ denotes the q-power mean between r and t, as defined in Section 2.2

As we have already notice, H_p is the square of a metric on $[0,+\infty)$ for $p = 0$, $p = \frac{1}{2}$ and $p = 1$. We investigate now the same question for every real number p. This problem was already considered by Osterreicher and Vajda in the case $p < 1$ [Ost96], [OV03]. Following the same approach we prove:

Proposition 4.1.3. *The induced marginal perspective function H_p is the square of a metric on the nonnegative real numbers for any $p \in (-\infty, \frac{1}{2}] \cup [1,+\infty)$. $\sqrt{H_p}$ does not satisfy the triangle inequality if $p \in (\frac{1}{2}, 1)$.*

The proof of the previous Proposition is based on the following Lemma. It is the first example in the chapter of a fact that will be recurrent: the central role of the class of Matusita's divergences in the study of the metric properties of the marginal perspective function.

Lemma 4.1.4. *Let $a \in (0,1]$ and $F \in \Gamma_0^s(\mathbb{R}_+)$. If*

$$h(u) := \frac{(1-u^a)^{\frac{1}{a}}}{F(u)}$$

is decreasing in $[0,1)$, then $\hat{F}^a$ satisfies the triangle inequality.

Proof. Due to the monotonicity of the function $x \mapsto x^a$, one has that

$$h^a(u) = \frac{1-u^a}{F^a(u)}$$

is decreasing in $[0,1)$, so that $h^a(u) \geq h^a(v)$ and $h^a(u) \geq h^a(\frac{u}{v})$ if $0 \leq u < v < 1$. It follows that

$$\hat{F}^a(u,1) = \frac{1-u^a}{h^a(u)} = \frac{1-v^a}{h^a(u)} + \frac{v^a - u^a}{h^a(u)} \tag{4.1.6}$$

$$\leq \frac{1-v^a}{h^a(v)} + \frac{v^a\Big(1-\left(\frac{u}{v}\right)^a\Big)}{h^a(\frac{u}{v})} = \hat{F}^a(u,v) + \hat{F}^a(v,1), \tag{4.1.7}$$

which is sufficient to prove the triangle inequality since $\hat{F}$ is convex, symmetric and positively 1-homogeneous. □

Proof of Proposition 4.1.3. Using now Lemma 4.1.4, it remains to show that the function

$$h_p(u) := \frac{(1-\sqrt{u})^2}{f_p(u)}$$

is decreasing in $(0,1)$, where we have set $f_p(u) := H_p(u,1)$. The derivative of the function h_p is the following:

$$h_p'(u) = -\frac{2}{p}\left(\frac{1}{\sqrt{u}} - 1\right)\frac{1}{f_p^2(u)}\phi_p(u), \tag{4.1.8}$$

where

$$\phi_p(u) := 2^{-1}(u^{\frac{1}{2}} + 1) - 2^{-\frac{1}{1-p}}(u^{1-p} + 1)^{\frac{1}{1-p}-1}(u^{\frac{1}{2}-p} + 1). \tag{4.1.9}$$

Note that $\phi_p(1) = 0$ and $\psi_p(u) := \sqrt{u}\phi_p'(u)$ satisfies:

$$\psi_p(u) = \frac{1}{4} - 2^{-\frac{1}{1-p}}(u^{1-p} + 1)^{\frac{1}{1-p}-2}u^{-p}\Big(\frac{1 + u^{1-p}}{2} - p(1 - \sqrt{u})\Big). \tag{4.1.10}$$

The function ψ_p is such that $\psi_p(1) = 0$ and

$$\psi_p'(u) = 2^{-\frac{1}{1-p}}p(\frac{1}{2} - p)(u^{1-p} + 1)^{\frac{1}{1-p}-3}u^{-p-1}(1 - \sqrt{u})(1 - u^{1-p}). \tag{4.1.11}$$

Now let us suppose $p > 1$: we have to prove that ϕ_p is positive in $(0, 1)$. This is implied by $\psi_p(u) < 0$ in $(0, 1)$ which is true because $\psi_p'(u)$ is positive in $(0, 1)$. Similar considerations can be applied to the case $p < 0$ and $p \in (0, \frac{1}{2})$.
For $p \in (\frac{1}{2}, 1)$ one gets $\psi_p'(u) < 0$ in $(0, 1)$ so ψ_p is positive in $(0, 1)$. This implies that ϕ_p is negative and so h_p is increasing in $(0, 1)$. As a consequence, an analysis of the proof of Lemma 4.1.4 shows that the triangle inequality is reversed for these values of p. □

Remark 4.1.5. *When $p \in (\frac{1}{2}, 1)$, Osterreicher and Vajda ([OV03]) proved that H_p^{1-p} is a metric.*

4.1.2 Marginal perspective function and structural properties

In Section 3.2.5 we have shown that the construction of the marginal perspective function naturally produces a symmetric divergence. In this section we show that this is not the only feature of the minimization procedure (3.2.24): iterating this process we will highlight the important role played by the class of Matusita's divergences.

We have already seen how to produce a map $T_1 : \Gamma_0(\mathbb{R}_+) \to \Gamma_0^s(\mathbb{R}_+)$ (we recall that $\Gamma_0^s(\mathbb{R}_+)$ denotes the set of *symmetric* admissible entropy functions, as defined in Section 2.1): let $F \in \Gamma_0(\mathbb{R}_+)$, we set $T_1(F)(s) := H_F(1, s)$, where H_F is the lower semicontinuous envelope of the function $\tilde{H}_F$ defined by (3.2.24). We also denote by $T_a : \Gamma_0(\mathbb{R}_+) \to \Gamma_0^s(\mathbb{R}_+)$ the map $T_a(F) := 2^{\frac{1}{a}-1}T_1(F)$, where $a \in (0, 1]$.

It is clear that the two trivial entropies

$$F(s) \equiv 0 \qquad \text{and} \qquad F(s) = I_{\{1\}} = \begin{cases} 0 & \text{if } s = 1, \\ +\infty & \text{otherwise,} \end{cases} \tag{4.1.12}$$

are fixed points of the map T_a for any $a \in (0, 1]$. Another important property that follows immediately from the definition is that

$$F_1 \geq F_2 \implies T_a(F_1) \geq T_a(F_2). \tag{4.1.13}$$

Due to the difference between the case $a = 1$ and the case $0 < a < 1$, we have divided the analysis of the behaviour of the map T_a. Nevertheless, the strategy behind the proofs is in common: we show that, under suitable conditions, the sequence $\{T_a^{(n)}(F)\}$ is monotone and the limit is a fixed point of the map T_a. We then prove that $T_a(F) = F$ implies $F(s) = c|s^a - 1|^{\frac{1}{a}}$, where $c \in [0, +\infty]$ (in the case $c = +\infty$ we mean that $c|s^a - 1|^{\frac{1}{a}} = I_{\{1\}}(s)$).

We start with a simple Lemma that provides a crucial monotonicity property.

Lemma 4.1.6. *Let $F \in \Gamma_0^s(\mathbb{R}_+)$ and $a \in (0,1]$, if $\hat{F}^a$ satisfies the triangle inequality then $T_a(F) \geq F$.*

Proof. For any $s, t \in \mathbb{R}_+$ the convexity of the function $x \mapsto x^{\frac{1}{a}}$ yields

$$2^{\frac{1}{a}-1}\hat{F}(1,s) + 2^{\frac{1}{a}-1}\hat{F}(s,t) = \frac{1}{2}\big(2\hat{F}^a(1,s)\big)^{\frac{1}{a}} + \frac{1}{2}\big(2\hat{F}^a(s,t)\big)^{\frac{1}{a}}$$
$$\geq \big(\hat{F}^a(1,s) + \hat{F}^a(s,t)\big)^{\frac{1}{a}} \geq \hat{F}(1,t) = F(t).$$

The result follows by taking the infimum of the left hand side with respect to s. □

Lemma 4.1.7. *Given a function $F \in \Gamma_0^s(\mathbb{R}_+)$, the sequence $\{T_1^{(n)}(F)\}$ is decreasing. If we put*

$$\tilde{F}^\infty(s) := \lim_{n\to\infty} T_1^{(n)}(F)(s) \qquad \text{for any } s \in [0,+\infty)$$

and we denote by F^∞ the lower semicontinuous envelope of $\tilde{F}^\infty$, then F^∞ is a fixed point of the map T_1.

Proof. If $F = I_{\{1\}}$ the Lemma trivially holds, so let us suppose that $F \neq I_{\{1\}}$. Since the map $\theta \mapsto \hat{F}(1,\theta) + \hat{F}(\theta, s)$ is equals to $F(1,s)$ when $\theta = 1$ or $\theta = t$, it follows that $T_1(F)(s) \leq F(s)$ for any s. Thus, the sequence $T_1^{(n)}(F)(s)$ is decreasing and it has a limit $\tilde{F}^\infty(s)$ that is clearly convex, nonnegative and satisfies $\tilde{F}^\infty(1) = 0$. The domain of F must contain an interval of the form $[1/b, b]$ for some $b > 1$. By the convexity of the function F, there exists a constant $c \in (0,+\infty)$ such that

$$F \leq G := \begin{cases} c|s-1| & \text{if } \ s \in [1/b, b], \\ +\infty & \text{otherwise.} \end{cases}$$

In particular $T_1^{(n)}(F)(s) \leq T_1^{(n)}(G)(s)$ for any s. An easy computation shows that $\mathrm{D}(\tilde{G}^\infty) = (0,+\infty)$ (see below, Lemma 4.1.15), which implies $(0,+\infty) \subset \mathrm{D}(\tilde{F}^\infty)$.

Let F^∞ be the lower semicontinuous envelope of the function $\tilde{F}^\infty$. It is immediate that $F^\infty \in \Gamma_0^s(\mathbb{R}_+)$. We have to show that F^∞ is a fixed point of T_1: with the same reasoning as above, one gets $T_1(F^\infty) \leq F^\infty$; in order to prove the reverse inequality we notice that for any $1 \leq s \leq t$ and any $n \in \mathbb{N}$ it holds

$$T_1^{(n)}(F)(s) + sT_1^{(n)}(F)\left(\frac{t}{s}\right) \geq T_1^{(n+1)}(F)(t).$$

The result follows by taking the limit with respect to n, doing the lower semicontinuous envelope and then minimizing with respect to s. □

Theorem 4.1.8. *Let $F \in \Gamma_0^s(\mathbb{R}_+)$ be a fixed point of T_1. Then $F(s) = c|s-1|$ for a certain $c \in [0,+\infty]$. In particular, an induced marginal perspective function H_F is a metric on $\mathbb{R}_+$ if and only if $H_F = cM_1$, $c \in (0,+\infty)$.*

Proof. It is clear that the function cM_1 is a fixed point of T_1 for any $c \in [0,+\infty]$. We show now that they are the only fixed points: since $\theta \mapsto \hat{F}(1,\theta) + \hat{F}(\theta, s)$ is a convex function that has the same value when $\theta = 1$ and $\theta = s$, $T_1(F) = F$ implies that

$$F(\theta) + \theta F\left(\frac{s}{\theta}\right) = F(s) \qquad \text{for every } 1 \leq \theta \leq s. \tag{4.1.14}$$

If $F \neq I_{\{1\}}$, we can also assume that F is finite valued. To see this, let us assume by contradiction that $s > 1$ is such that $F(s) = +\infty$ and $F(\sqrt{s}) < +\infty$. Equation (4.1.14) fails with the choice $\theta = \sqrt{s}$.

Take now $s > 2$ and $\theta = 2$, we have

$$F(2) + 2F\left(\frac{s}{2}\right) = F(s).$$

If we choose instead $s > 2$ and $\theta = s/2$ we obtain

$$F\left(\frac{s}{2}\right) + \frac{s}{2}F(2) = F(s).$$

By taking the difference of the two obtained equations, one gets $F\left(\frac{s}{2}\right) = F(2)\left(\frac{s}{2} - 1\right)$ for any $s > 2$ and we can conclude that $F(s) = c|s-1|$, $c \in [0, +\infty]$, since $F \in \Gamma_0^s(\mathbb{R}_+)$.

To conclude the proof, let us suppose that H_F is a metric. We put $f(s) := H_F(s, 1)$ so that $f \in \Gamma_0^s(\mathbb{R}_+)$. Lemma 4.1.6 implies that $T_1(F) \geq F$; Lemma 4.1.7 provides the converse inequality. In particular f is a fixed point of T_1, and the only fixed points that induces a metric on $\mathbb{R}_+$ are the functions of the form cM_1 with $c \in (0, +\infty)$. □

In order to deal with the case $0 < a < 1$ we need some preliminary results and some additional assumptions. We start by proving that every metric of the form $\hat{F}^a$, $a \in (0, 1]$, is a complete metric.

Lemma 4.1.9. *Let $F \in \Gamma_0^s(\mathbb{R}_+)$ and let us suppose that $D := \hat{F}^a$ is a metric for a number $a \in (0, 1]$. Then there exists $c > 0$ such that*

$$F(s) > c|s^a - 1|^{\frac{1}{a}}$$

and $\hat{F}^a$ is a complete metric.

Proof. For any $0 \leq u < v < 1$ we rewrite the distance between u and 1 as

$$D(u, 1) = \frac{g(u)}{g(v)} D(v, 1) + \frac{g(u)}{g(\frac{u}{v})} D(u, v) \tag{4.1.15}$$

where $g(u) := \dfrac{F^a(u)}{1 - u^a}$. Since the triangle inequality holds, at least one of the numbers $\frac{g(u)}{g(v)}$ and $\frac{g(u)}{g(\frac{u}{v})}$ is less or equal than 1. Choosing $u := v^2$, it follows $g(v^2) \leq g(v)$ for any $v < 1$. By contradiction let us suppose it does not exists a positive constant c such that $F(s) > c|s^a - 1|^{\frac{1}{a}}$, so that there exists a sequence $v_n \in (0, 1)$ such that $g(v_n) \to 0$. Then, we can find a $\bar{v} \in (0, 1)$ such that $D(0, 1) = g(0) > g(\bar{v})$. On the other hand, the sequence w_n defined by $w_0 := \bar{v}$, $w_n := w_{n-1}^2$ converges to 0, and by continuity of the function g it follows $g(w_n) \to g(0)$. Since $g(0) > g(w_0)$ and $n \mapsto g(w_n)$ is decreasing, we get a contradiction.

Now it is easy to show that the metric D is complete: since $\hat{F}^a$ is a metric, $\hat{F}$ is symmetric and $D(0, 1) = F^a(0) := c_2 < +\infty$. From the convexity of the function F it follows $F^a(s) \leq c_2|s - 1|^a$ so that

$$c_1 M_a^a \leq D \leq c_2 M_1^a.$$

The result follows using the fact that M_a^a and M_1^a are two complete metrics inducing the same convergence. □

We now ready to prove the analogous of Theorem 4.1.8 in the case $0 < a < 1$, under an additional assumption.

Theorem 4.1.10. *Let $F \in \Gamma_0^s(\mathbb{R}_+)$ and let us suppose that $\hat{F}^a$ is a distance and $T_a(F) = F$, where $a \in (0,1)$. Then $F(s) = c|s^a - 1|^{\frac{1}{a}}$ for a constant $c \in (0, +\infty)$.*

Proof. We can assume that F is finite valued since $\hat{F}^a$ is a metric. The fixed point property $T_a(F) = F$ implies that for every $s > 1$ it exists θ, $1 \leq \theta \leq s$, such that

$$2^{\frac{1}{a}-1}\hat{F}(1,\theta) + 2^{\frac{1}{a}-1}\hat{F}(\theta, s) = \hat{F}(1,s), \tag{4.1.16}$$

Since $\hat{F}^a$ is a metric and the function $f(x) = x^a$ is concave we can deduce

$$\begin{aligned} &\hat{F}^a(1,s) \leq \hat{F}^a(1,\theta) + \hat{F}^a(\theta,s) = \frac{[2^{\frac{1}{a}}\hat{F}(1,\theta)]^a + [2^{\frac{1}{a}}\hat{F}(\theta,s)]^a}{2} \\ &\leq \Big(2^{\frac{1}{a}-1}\hat{F}(1,\theta) + 2^{\frac{1}{a}-1}\hat{F}(\theta,s)\Big)^a. \end{aligned} \tag{4.1.17}$$

Equation (4.1.16) implies the equality in the inequality (4.1.17), so that we obtain $\hat{F}(1,\theta) = \hat{F}(\theta, s)$.

Since $\hat{F}$ is symmetric, positively 1-homogogeneous and $\hat{F}^a$ is a complete metric, Lemma 4.1.9 implies that $(\mathbb{R}_+, \hat{F}^a)$ is a one dimensional geodesic space, so it must be isometric to $(\mathbb{R}_+, |\cdot|)$ (for a reference see [BBI01], chapter 2). In particular there exists $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ increasing and continuous such that we can write $\hat{F}^a(r,t) = |\phi(t) - \phi(r)|$. Using again the 1-homogeneity of the function $\hat{F}$, it follows $\hat{F}^a(r,t) = r^a\hat{F}^a(1, \frac{t}{r})$ for $r > 0$, so that

$$\phi(t) - \phi(r) = r^a\Big(\phi(\frac{t}{r}) - \phi(1)\Big), \quad t \geq r. \tag{4.1.18}$$

Evaluating equation (4.1.18) for $t = 2r$ we get

$$\phi(2r) - \phi(r) = r^a(\phi(2) - \phi(1)), \quad r \geq 1, \tag{4.1.19}$$

whereas the choice $r = 2$ yields

$$\phi(t) - \phi(2) = 2^a\Big(\phi(\frac{t}{2}) - \phi(1)\Big), \quad t \geq 2. \tag{4.1.20}$$

Now consider the previous equation with $t = 2r$, it follows

$$\phi(2r) - \phi(r) = \phi(2) + 2^a(\phi(r) - \phi(1)) - \phi(r), \quad r \geq 1. \tag{4.1.21}$$

Using now the identities (4.1.19) and (4.1.21), it follows

$$r^a(\phi(2) - \phi(1)) = \phi(2) + 2^a(\phi(r) - \phi(1)) - \phi(r) \;\text{ for any } r \geq 1,$$

and we can compute $\phi(r)$ as

$$\phi(r) = \frac{\phi(2) - \phi(1)}{2^a - 1}(r^a - 1) + \phi(1),$$

so that $F^a(r) = (r^a - 1)\dfrac{F^a(2)}{2^a - 1}$ for every $r \geq 1$, which prove the theorem. □

Remark 4.1.11. *We do not know if the assumption that $\hat{F}^a$ is a metric can be removed in order to obtain the same characterization as in Theorem 4.1.8. The difficulty is that the value of the function $\theta \mapsto 2^{\frac{1}{a}-1}\hat{F}(1,\theta) + 2^{\frac{1}{a}-1}\hat{F}(\theta,s)$ at $\theta = 1$ and $\theta = s$ is strictly greater that $\hat{F}(1,s)$, if $a < 1$.*

In order to obtain that also in the case $0 < a < 1$ the limit function is a fixed point of the map T_a, we need the following Lemma. It is an easy consequence of general results in the theory of Γ-convergence (see e.g. [Mas93]), but we give a direct proof in our simplified setting.

Lemma 4.1.12. *Let X be a compact space and let $f_n : X \to [0, +\infty]$ be a sequence of lower semicontinuous functions such that $f_n(x) \le f_{n+1}(x)$ for every $n \in \mathbb{N}$ and every $x \in X$. Then*

$$\lim_{n\to\infty} \min_{x\in X} f_n(x) = \min_{x\in X} f_\infty(x),$$

where we put $f_\infty(x) := \lim_{n\to\infty} f_n(x)$.

Proof. The functions f_n and f_∞ are lower semicontinuous over a compact set so that they have a minimum. Since $f_n(x) \le f_\infty(x)$ for every $x \in X$ it is clear that

$$\lim_{n\to\infty} \min_{x\in X} f_n(x) \le \min_{x\in X} f_\infty(x).$$

Let us suppose now $a < \min_{x\in X} f_\infty(x)$, so that for every $x \in X$ $a < f_\infty(x)$. Since $\lim_n f_n(x) = f_\infty(x)$, there exists $n = n(x)$ such that $a < f_n(x)$. It follows that the family $\{a < f_n\}_{n\in\mathbb{N}}$ is an open cover of X. Let $n_1, ..., n_j$ be a finite collection of indexes such that

$$X \subset \{a < f_{n_1}\} \cup ... \cup \{a < f_{n_j}\}.$$

Let $N := \max\{n_1, ..., n_j\}$, so that $X \subset \{a < f_N\}$ since f_n are increasing. This implies that $a < f_n(x)$ for every $x \in X$ so that $a < \lim_{n\to\infty} \min_{x\in X} f_n(x)$. Since a is an arbitrary number less than $\min_{x\in X} f_\infty(x)$, the Lemma follows. □

We can now state the Theorem about the convergence of the iterations of the map T^a.

Theorem 4.1.13. *Let $a \in (0, 1)$. Given a function $F \in \Gamma_0^s(\mathbb{R}_+)$, if $\hat{F}^a$ is a metric then the sequence $\{T_a^{(n)}(F)\}$ converges pointwise to a fixed point of the map T_a. In particular, if the limit function F^∞ is such that $(\hat{F^\infty})^a$ is a metric, then $F^\infty(s) = c|s^a - 1|^{\frac{1}{a}}$ where $c \in (0, +\infty)$.*

Proof. Lemma 4.1.6 implies that $T_a(F) \ge F$. By the monotonicity property (4.1.13) the sequence $T_a^{(n)}(F)$ is increasing so it converges pointwise to a function F^∞. It is clear that $F^\infty \in \Gamma_0^s(\mathbb{R}_+)$ (recall that the monotone increasing limit of a sequence of lower semicontinuous functions is lower semicontinuous). Since $\hat{F}^a$ is a metric, F is finite finite valued, as well as $T_a^{(n)}(F)$. We want to show that F^∞ is a fixed point of T_a:

$$\begin{aligned}
T_a(F^\infty)(s) &= sc^-\Big(2^{\frac{1}{a}-1} \inf_{\theta>0} \big(F^\infty(\theta) + \theta F^\infty(\frac{s}{\theta})\big)\Big) \\
&= sc^-\Big(2^{\frac{1}{a}-1} \lim_{n\to\infty} \inf_{\theta>0} \big(T_a^{(n)}(F)(\theta) + \theta T_a^{(n)}(F)(\frac{s}{\theta})\big)\Big) \\
&= sc^-\Big(\lim_{n\to\infty} T_a^{(n+1)}(F)(s)\Big) = F^\infty(s),
\end{aligned}$$

where we have denoted by $sc^-(f)$ the lower semicontinuous envelope of the function f and we have used Lemma 4.1.12 applied to $f_n(\theta) := T_a^{(n)}(F)(\theta) + \theta T_a^{(n)}(F)(\frac{s}{\theta})$ and $X := [1, s]$. The conclusion follows from Theorem 4.1.10. □

Remark 4.1.14. *It is not difficult to show that F^∞ can be equal to $I_{\{1\}}$. Take for instance $F(s) = |s - 1|$ and consider the sequence $T_a^{(n)}(F)$ with $a \in (0, 1)$.*

In the final part of this section we want to study the connection between the behaviour of the function F in a neighborhood of 1 and the limit function F^∞. we start with two lemmas:

Lemma 4.1.15. *Let $a \in (0,1]$, $b > 1$, $c \in (0,+\infty)$ and $\bar{F} \in \Gamma_0^s(\mathbb{R}_+)$ be the function defined by*

$$\bar{F}(s) := \begin{cases} c|s^a - 1|^{\frac{1}{a}} & s \in [1/b, b], \\ +\infty & \text{otherwise.} \end{cases}$$

Then

$$\lim_{n\to\infty} T_a^{(n)}(\bar{F})(s) = c|s^a - 1|^{\frac{1}{a}} \qquad s \in (0,+\infty).$$

Proof. It is sufficient to consider the case $s > 1$; it holds

$$T_a(\bar{F})(s) = 2^{\frac{1}{a}-1} \inf_{\theta\in[1,s]} \bar{F}(\theta) + \theta\bar{F}\left(\frac{s}{\theta}\right). \tag{4.1.22}$$

When $b^2 < s$ it is clear that $T_a(\bar{F})(s) = +\infty$. Moreover, we notice that in the case

$$\mathfrak{M}_a(1,s) \leq b, \text{ and } \frac{s}{\mathfrak{M}_a(1,s)} \leq b, \tag{4.1.23}$$

the expression (4.1.22) is minimized by $\theta = \mathfrak{M}_a(1,s)$, so that $T_a(\bar{F})(s) = c|s^a - 1|^{\frac{1}{a}}$ for such an s. Using now the bound given by Proposition 2.2.1, we deduce that the inequalities (4.1.23) are certainly satisfied when $1 \leq s \leq 2b - 1$. The theorem is now an easy consequence of the fact that the sequence $b_0 := b$, $b_{n+1} := 2b_n - 1$ is strictly increasing and it diverges to $+\infty$. □

Lemma 4.1.16. *Let $a \in (0,1]$, $b > 1$, $c \in (0,+\infty)$ and $\underline{F} \in \Gamma_0^s(\mathbb{R}_+)$ be the function defined by $\underline{F}(s) := c|s^a - 1|^{\frac{1}{a}}$ when $s \in [1/b, b]$ and extended linearly outside the interval in such a way that the left derivative of $\underline{F}$ at b is the slope of the linear extension in $[b,+\infty)$. Then $\lim_{n\to\infty} T_a^{(n)}(\underline{F})(s) = c|s^a - 1|^{\frac{1}{a}}$.*

Proof. The lemma follows if we prove that

$$\underline{F}^a(t) \leq \underline{F}^a(s) + s\underline{F}^a\left(\frac{t}{s}\right) \tag{4.1.24}$$

for every $1 \leq s \leq t$. Indeed (4.1.24) implies that H^a is a distance, so that, by Theorem 4.1.13, $T_a^{(n)}(\underline{F})$ must converge to a function F^∞ that is a fixed point of T_a. Since $T_a^{(n)}(\underline{F})(s) = c|s^a - 1|^{\frac{1}{a}}$ for every n and every $s \in [\frac{1}{b}, b]$, it holds $F^\infty(s) = c|s^a - 1|^{\frac{1}{a}}$ for every $s \in [\frac{1}{b}, b]$ and this implies that $F^\infty(s) = c|s^a - 1|^{\frac{1}{a}}$ for every s. Indeed, let us suppose by contradiction there exists $s_0 > 1$ such that $F^\infty(s_0) \neq c|(s_0)^a - 1|^{\frac{1}{a}}$ and consider the constant $k \neq c$ such that $F^\infty(s_0) = k|(s_0)^a - 1|^{\frac{1}{a}}$. Since F^∞ and $k|(s_0)^a - 1|^{\frac{1}{a}}$ are fixed points of T_a and they coincide in s_0, it must exists another number s_1, $1 < s_1 < s_0$, where they coincide. Iterating the argument it is easy to show that F^∞ and $k|(s_0)^a - 1|^{\frac{1}{a}}$ have to coincide on a sequence of numbers that converges to 1 but this is absurd since $F^\infty(s) = c|s^a - 1|^{\frac{1}{a}}$ for every $s \in [\frac{1}{b}, b]$ and the functions $c|s^a - 1|^{\frac{1}{a}}$ and $k|s^a - 1|^{\frac{1}{a}}$ coincide only at $s = 1$.

It remains to show that (4.1.24) holds. We use Lemma 4.1.4: we have to prove that the function

$$s \mapsto \frac{|s^a - 1|^{\frac{1}{a}}}{\underline{F}(s)}$$

is increasing in $(1, +\infty)$: this is obvious in the interval $(1, b]$; consider now two numbers r, t such that $b < r < t$. We define $s \mapsto l_r(s)$ to be the affine function that coincide with $\underline{F}$ at b and such that $l_r(r) = c|r^a - 1|^{\frac{1}{a}}$), and we notice that the convexity of the function $s \mapsto c|s^a - 1|^{\frac{1}{a}}$ implies that the slope of l_r is greater or equal than the positive slope of the function $\underline{F}$ in $(b, +\infty)$. Using again the convexity of the function $c|s^a - 1|^{\frac{1}{a}}$ and the trivial fact that the quotient

$$s \mapsto \frac{l_r(s)}{\underline{F}(s)}$$

is increasing in $(b, +\infty)$, we conclude because

$$\frac{|t^a - 1|^{\frac{1}{a}}}{\underline{F}(t)} \geq \frac{l_r(t)}{\underline{F}(t)} \geq \frac{l_r(r)}{\underline{F}(r)} = \frac{c|r^a - 1|^{\frac{1}{a}}}{\underline{F}(r)}. \tag{4.1.25}$$

□

Theorem 4.1.17. *Let $F \in \Gamma_0^s(\mathbb{R}_+)$ be a function such that*

$$\lim_{s \to 1} \frac{F(s)}{c|s^a - 1|^{\frac{1}{a}}} = 1. \tag{4.1.26}$$

Then

$$\lim_{n \to +\infty} T_a^{(n)}(F)(s) = c|s^a - 1|^{\frac{1}{a}} \qquad s \in (0, +\infty). \tag{4.1.27}$$

Proof. For every $\epsilon > 0$ there exists a $b > 1$ such that

$$(1 - \epsilon)c|s^a - 1|^{\frac{1}{a}} \leq F(s) \leq (1 + \epsilon)c|s^a - 1|^{\frac{1}{a}}, \quad s \in [1/b, b],$$

so that

$$(1 - \epsilon)\underline{F} \leq F \leq (1 + \epsilon)\bar{F},$$

where $\underline{F}, \bar{F}$ are defined in Lemma 4.1.15 and 4.1.16. Take now an arbitrary $s \in (0, +\infty)$, from the monotonicity property (4.1.13) it follows

$$(1 - \epsilon)T_a^{(n)}(\underline{F}) \leq T_a^{(n)}(F) \leq (1 + \epsilon)T_a^{(n)}(\bar{F}),$$

so that by Lemma 4.1.15 and Lemma 4.1.16 one gets

$$(1 - \epsilon)c|s^a - 1|^{\frac{1}{a}} \leq \liminf_{n \to \infty} T_a^{(n)}(F)(s) \leq \limsup_{n \to \infty} T_a^{(n)}(F)(s) \leq (1 + \epsilon)c|s^a - 1|^{\frac{1}{a}}.$$

Since ϵ is arbitrary, there exists the limit of $T_a^{(n)}(F)(s)$ and it is equal to $c|s^a - 1|^{\frac{1}{a}}$. □

4.2 Spatially inhomogeneous F-divergences

4.2.1 Formulation on the cone

Let assume the basic setting introduced in 3.2.1 and let us recall that we can associate to the entropy function F and to the cost $\boldsymbol{c}$ the marginal perspective cost H, whose construction is developed in Section 3.2.3.

We denote by $\mathfrak{C}(X)$ the cone over X defined in Section 2.3 with standard metric $\mathsf{d}_{\mathfrak{C}(X)}$, and by $\mathfrak{y} = [\boldsymbol{y}] = [x, r]$ the points of $\mathfrak{C}(X)$, while the vertex is denoted by $\mathfrak{o}$. We also put $\mathfrak{C}_{\mathfrak{o}} := \mathfrak{C}(X) \setminus \{\mathfrak{o}\}$. Instead of the usual quotient topology, on the cone we consider the following weaker topology $\tau_{\mathfrak{C}}$: neighborhoods of points in the set $\mathfrak{C}_{\mathfrak{o}}$ coincide with neighborhoods in $Y := X \times [0, +\infty)$; a system of open neighborhoods of the vertex $\mathfrak{o}$ is given by the sets

$$\{[x, r] : 0 \leq r < \epsilon\}, \quad \epsilon > 0. \tag{4.2.1}$$

It is not difficult to show that $\mathsf{d}_{\mathfrak{C}(X)} : \mathfrak{C}(X) \times \mathfrak{C}(X) \to [0, +\infty)$ is lower semicontinuous with respect to the product topology induced by $\tau_{\mathfrak{C}}$.

We define the canonical projection $\mathfrak{p} : Y \to \mathfrak{C}(X)$ as $\mathfrak{p}(x, r) = [x, r]$, which is continuous. It has a right inverse y: fixed a point $\bar{x} \in X$ we define

$$\mathsf{r} : \mathfrak{C}(X) \to [0, +\infty), \qquad \mathsf{r}[x, r] := r, \tag{4.2.2}$$

$$\mathsf{x} : \mathfrak{C}(X) \to X, \qquad \mathsf{x}[x, r] := \begin{cases} x \text{ if } r > 0, \\ \bar{x} \text{ if } r = 0, \end{cases} \tag{4.2.3}$$

$$\mathsf{y} : \mathfrak{C}(X) \to Y, \qquad \mathsf{y} := (\mathsf{x}, \mathsf{r}). \tag{4.2.4}$$

We notice that, using the map y, any measure $\nu \in \mathscr{M}(\mathfrak{C}(X))$ can be lifted to a measure $\bar{\nu} \in \mathscr{M}(Y)$ such that $\mathfrak{p}_\sharp \bar{\nu} = \nu$. Indeed, it is sufficient to define $\bar{\nu} := \mathsf{y}_\sharp \nu$.

Now, let us consider the product spaces

$$\boldsymbol{\mathfrak{C}}(X) := \mathfrak{C}(X) \times \mathfrak{C}(X), \quad \boldsymbol{Y} := Y \times Y.$$

We denote by $\boldsymbol{\mathfrak{y}} = (\mathfrak{y}_1, \mathfrak{y}_2) = ([x_1, r_1], [x_2, r_2])$ and $\boldsymbol{y} = (y_1, y_2)$ the points of these spaces, respectively. We put $\mathsf{r}_i(\boldsymbol{\mathfrak{y}}) := \mathsf{r}(\mathfrak{y}_i) = r_i$ and $\mathsf{x}_i(\boldsymbol{\mathfrak{y}}) := \mathsf{x}(\mathfrak{y}_i)$, while $\pi^i : \boldsymbol{\mathfrak{C}}(X) \to \mathfrak{C}(X)$ denotes the projection on the i-coordinate, $i = 1, 2$. The lift map from $\boldsymbol{\mathfrak{C}}(X)$ to $\boldsymbol{Y}$ is denoted by $\mathbf{y} := \mathsf{y} \otimes \mathsf{y}$.

We say that a plan $\boldsymbol{\alpha} \in \mathscr{M}(\boldsymbol{\mathfrak{C}}(X))$ lies in $\mathscr{M}_2(\boldsymbol{\mathfrak{C}}(X))$ if

$$\int_{\boldsymbol{\mathfrak{C}}(X)} (\mathsf{r}_1^2 + \mathsf{r}_2^2) \mathrm{d}\boldsymbol{\alpha} < +\infty. \tag{4.2.5}$$

We write $\boldsymbol{\alpha} \in \mathscr{P}_2(\boldsymbol{\mathfrak{C}}(X))$ if $\boldsymbol{\alpha} \in \mathscr{M}_2(\boldsymbol{\mathfrak{C}}(X))$ and $\boldsymbol{\alpha}$ is a probability measure.

The homogeneous marginals of the plan $\boldsymbol{\alpha}$ are defined as

$$\mathfrak{h}_i^2(\boldsymbol{\alpha}) := (\mathsf{x}_i)_\sharp(\mathsf{r}_i^2 \boldsymbol{\alpha}) = h_i^2(\bar{\boldsymbol{\alpha}}) \in \mathscr{M}(X),$$

where h_i was defined in (3.2.17) and $\bar{\boldsymbol{\alpha}} = \mathbf{y}_\sharp \boldsymbol{\alpha} \in \mathscr{M}(\boldsymbol{Y})$. The definition is well posed (i.e. it does not depend on the choice of $\bar{x}$) since the measure $\mathsf{r}_i^2 \boldsymbol{\alpha}$ does not charge $(\pi^i)^{-1}(\mathfrak{o})$.

On the cone we introduce the operation

$$\mathfrak{C}(X) \ni \mathfrak{y} \cdot \lambda := \begin{cases} \mathfrak{o} & \text{if } \mathfrak{y} = \mathfrak{o}, \\ [x, \lambda r] & \text{if } \mathfrak{y} = [x, r],\, r > 0 \end{cases} \tag{4.2.6}$$

and, given a Borel map $\vartheta : \boldsymbol{\mathfrak{C}}(X) \to (0, \infty)$ in $L^2(\boldsymbol{\mathfrak{C}}(X), \boldsymbol{\alpha})$, we define the product map $(\mathrm{prd}_\vartheta(\boldsymbol{\mathfrak{y}}))_i := \mathfrak{y}_i \cdot (\vartheta(\boldsymbol{\mathfrak{y}}))^{-1}$ and the dilation map as $\mathrm{dil}_{\vartheta,2}(\boldsymbol{\alpha}) := (\mathrm{prd}_\vartheta)_\sharp(\vartheta^2 \boldsymbol{\alpha}) \in \mathscr{M}_2(\boldsymbol{\mathfrak{C}}(X))$. Then, it is not difficult to prove (see [LMS18a, Equation 7.18]) that

$$\mathfrak{h}_i^2(\mathrm{dil}_{\vartheta,2}(\boldsymbol{\alpha})) = \mathfrak{h}_i^2(\boldsymbol{\alpha}). \tag{4.2.7}$$

Given $\mu_1, \mu_2 \in \mathscr{M}(X)$, we define the

$$\mathfrak{H}^2(\mu_1, \mu_2) := \{\boldsymbol{\alpha} \in \mathscr{M}_2(\boldsymbol{\mathfrak{C}}(X)) : \mathfrak{h}_i^2(\boldsymbol{\alpha}) = \mu_i,\, i = 1, 2\}. \tag{4.2.8}$$

Theorem 4.2.1. *Let X be a Polish space and let $\mathfrak{D}$ be a distance on the cone $\mathfrak{C}(X)$, lower semicontinuous with respect to the product topology induced by $\tau_{\mathfrak{C}}$ and positively 1-homogeneous in the sense that*

$$\mathfrak{D}([x_1,\lambda r];[x_2,\lambda t]) = \lambda\mathfrak{D}([x_1,r];[x_2,t]) \quad \text{for every } \lambda, r, t \in [0,+\infty),\ x_1,x_2 \in X. \tag{4.2.9}$$

Define the cone-distance cost $\mathfrak{CDC} : \mathscr{M}(X) \times \mathscr{M}(X) \to [0,+\infty)$ *induced by* $\mathfrak{D}$ *as*

$$\mathfrak{CDC}^2(\mu_1,\mu_2) := \min_{\boldsymbol{\alpha}\in\mathfrak{H}^2(\mu_1,\mu_2)} \int_{\mathfrak{C}(X)} \mathfrak{D}^2(\mathfrak{y}_1,\mathfrak{y}_2)\mathrm{d}\boldsymbol{\alpha}. \tag{4.2.10}$$

Then $\mathfrak{CDC}$ is well defined (i.e. the right hand side of 4.2.10 is indeed a minimum) and it is a distance on $\mathscr{M}(X)$.

Proof. We only give a sketch of the proof: by the 1-homogeneity of $\mathfrak{D}$ the right hand side of 4.2.10 is equal to (see [LMS18a, after Remark 7.5])

$$\min_{\boldsymbol{\alpha}\in C} \int_{\boldsymbol{\mathfrak{C}}(X)} \mathfrak{D}^2(\mathfrak{y}_1,\mathfrak{y}_2)\mathrm{d}\boldsymbol{\alpha}, \tag{4.2.11}$$

where

$$C := \left\{\boldsymbol{\alpha} \in \mathscr{P}(\boldsymbol{\mathfrak{C}}(X)) : \mathfrak{h}_i^2(\boldsymbol{\alpha}) = \mu_i,\ \boldsymbol{\alpha}\left(\boldsymbol{\mathfrak{C}}(X) \setminus \boldsymbol{\mathfrak{C}}(X)[R] = 0\right)\right\}. \tag{4.2.12}$$

Here $\boldsymbol{\mathfrak{C}}(X)[R]$ is the closed set defined as

$$\boldsymbol{\mathfrak{C}}(X)[R] = \mathfrak{C}(X)[R] \times \mathfrak{C}(X)[R], \qquad \mathfrak{C}(X)[R] := \{[x,r] \in \mathfrak{C}(X), r \le R\}, \tag{4.2.13}$$

and R satisfies $R^2 = \mu_1(X) + \mu_2(X)$. Using now the lower semicontinuity of $\mathfrak{D}$, the existence of the minimum of (4.2.11) (thus, of (4.2.10)) follows by the direct method of the calculus of variations since C is weakly closed and equally tight (see [LMS18a, Proof of Theorem 7.6]).

The fact that $\mathfrak{CDC}(\mu_1,\mu_2)$ is symmetric and equal to zero if and only if $\mu_1 = \mu_2$ is obvious. To prove the triangle inequality we use a version of the so-called *gluing lemma* for homogeneous marginals. We follow again [LMS18a] where the case of the Hellinger-Kantorovich distance is discussed but the proof goes through also in our situation. In particular, we can apply [LMS18a, Lemma 7.10], [LMS18a, Lemma 7.11] (only equation (7.38) is needed here), [LMS18a, Corollary 7.13] and [LMS18a, Corollary 7.14] with $\mathsf{d}_{\mathfrak{C}(X)}$ replaced by $\mathfrak{D}$ and HK replaced by $\mathfrak{CDC}$ and the result follows. □

Remark 4.2.2. *Any plan $\boldsymbol{\alpha} \in \mathscr{M}(\boldsymbol{\mathfrak{C}}(X))$ can be lifted through the map $\mathbf{y}$ to a measure $\bar{\boldsymbol{\alpha}} \in \mathscr{M}(\boldsymbol{Y})$ such that $\mathfrak{p}_\sharp\bar{\boldsymbol{\alpha}} = \boldsymbol{\alpha}$. As a consequence,*

$$\mathfrak{CDC}^2(\mu_1,\mu_2) = \min\left\{\int_{\boldsymbol{Y}} \mathfrak{D}^2(y_1,y_2)\mathrm{d}\bar{\boldsymbol{\alpha}} : \bar{\boldsymbol{\alpha}} \in \mathscr{M}(\boldsymbol{Y}),\ h_i^2(\bar{\boldsymbol{\alpha}}) = \mu_i\right\}. \tag{4.2.14}$$

Notice that, with a slight abuse of notation, we are using the same notation for the cone distance $\mathfrak{D}$ and the function $\mathfrak{D}$ defined on $\boldsymbol{Y}$ through $\mathfrak{D}(y_1,y_2) := \mathfrak{D}(\mathfrak{p}(y_1),\mathfrak{p}(y_2))$.

The cone-distance cost can also be interpreted in terms of the Wasserstein distance on the cone $\mathfrak{C}(X)$.

Corollary 4.2.3. *For every $\mu_1,\mu_2 \in \mathscr{M}(X)$ we have*

$$\mathfrak{CDC}(\mu_1,\mu_2) = \min\{\mathcal{W}_{2,\mathfrak{D}}(\alpha_1,\alpha_1) : \alpha_i \in \mathscr{P}_2(\mathfrak{C}(X)),\ \mathfrak{h}^2(\alpha_i) = \mu_i\}, \tag{4.2.15}$$

where $\mathcal{W}_{2,\mathfrak{D}}$ is the 2-Wasserstein distance on $\mathfrak{C}(X)$ induced by $\mathfrak{D}$ and $\mathfrak{h}^2 : \mathscr{P}_2(\mathfrak{C}(X)) \to \mathscr{M}(X)$ is the 2-homogeneous marginal.

Proof. The proof is a straightforward generalization of the proof of the result [LMS18a, Corollary 7.7]. □

The following Corollary is crucial in order to link the metric properties of the Entropy-Transport cost with the corresponding properties of the marginal perspective cost.

Corollary 4.2.4. *In the basic setting defined in 3.2.1, let H be the marginal perspective cost and ET be the Entropy-Transport cost induced by $(F, \boldsymbol{c})$. Let us suppose that the function H is a well defined function on $\mathfrak{C}(X)$, lower semicontinuous with respect to the product topology induced by $\tau_{\mathfrak{C}}$ and it is a square of a distance on $\mathfrak{C}(X)$. Then ET is the square of a distance on $\mathscr{M}(X)$.*

Proof. The formulation of the Entropy-Transport cost ET given by Proposition 3.2.8 corresponds to the right hand side of (4.2.14) with

$$\mathfrak{D}^2([x_1, r_1], [x_2, r_2]) := H(x_1, r_1^2; x_2, r_2^2). \tag{4.2.16}$$

By the assumptions on H, $\mathfrak{D}$ is lower with respect to the product topology induced by $\tau_{\mathfrak{C}}$ and positively 1-homogeneous since H has the same property (Lemma 3.2.7). The result follows by Theorem 4.2.1. □

Remark 4.2.5. *In this section, we have decided to work with the 2-homogeneous marginals in order to make even more transparent the connection between the Hellinger-Kantorovich distance and the distance $\mathsf{d}_{\mathfrak{C}(X)}$ (see below and [LMS18a, Chapter 7]), however everything can be stated for p-homogeneous marginals as discussed in section 3.2.4. In particular, one can also generalize Corollary 4.2.4 in order to prove that if H^a is a distance on the cone, then ET^a is a distance on $\mathscr{M}(X)$, $a \in (0, 1]$. We do not need this general version in the thesis.*

We conclude the section by showing some properties of the marginal perspective cost.

Proposition 4.2.6. *Let $F(s)$ be an admissible entropy function with a strict minimum at $s = 1$ and let $\boldsymbol{c} : X \times X \to [0, \infty]$ be a symmetric function such that $\boldsymbol{c}(x_1, x_2) = 0$ if and only if $x_1 = x_2$. Then the induced marginal perspective cost H is nonnegative, symmetric in the sense that $H(x_1, r; x_2, t) = H(x_2, t; x_1, r)$ for every $x_1, x_2 \in X$, $r, t \in \mathbb{R}_+$, and $H(x_1, r; x_2, t) = 0$ if and only if $(x_1, r) \sim (x_2, t)$.*

Proof. It is clear that $H \geq 0$. When $r = t = 0$ it follows from the dual representation (3.2.13) that $H(x_1, r; x_2, t) = 0$. If $(x_1, r) \sim (x_2, t)$ and $r = t > 0$ then $\boldsymbol{c}(x_1, x_2) = 0$ and the fact that the marginal perspective cost is null follows from the possible choice $\theta = r$ in the expression (3.2.8). Since $\boldsymbol{c}$ is symmetric it is also apparent that

$$H(x_1, r; x_2, t) = H(x_2, t; x_1, r).$$

It remains to prove that $H = 0$ implies $(x_1, r) \sim (x_2, t)$. Lemma 2.1.1 and equation (2.1.8) tell us that R^* is an increasing homeomorphism between $(-\mathrm{aff} F_\infty, F(0))$ and $(-F'_\infty, -F'_0)$ with $R^*(0) = 0$. Since $F(s)$ is a convex function with a strict minimum at $s = 1$, it holds $\mathrm{aff} F_\infty > 0$, $F(0) > 0$, $F'_\infty > 0, F'_0 < 0$. In particular, there exists a positive number $k > 0$ such that the function R^* is finite, continuous and strictly increasing in $(-k, k)$. Hence, it follows again from the representation (3.2.13) that

$H(x_1, r; x_2, t) = 0$ and $\boldsymbol{c}(x_1, x_2) > 0$ implies $r = t = 0$. Moreover, when $\boldsymbol{c}(x_1, x_2) = 0$ we must have $r = t$: suppose by contradiction that $0 = r < t$ (the other case is similar), in the equation (3.2.13) we find $-k < \psi_1 < 0 < \psi_2 < k$ such that $R^*(\psi_1) + R^*(\psi_2) \leq 0$, contradicting the fact $H = 0$. Finally, when $H(x_1, r; x_2, t) = 0$, $\boldsymbol{c}(x_1, x_2) = 0$ and r, t are positive we can prove that $r = t$ using the fact that $\tilde{H}_0 = 0$ implies $r = t$ because, using now the expression (3.2.9), we know that for every natural n there exists θ_n such that

$$0 \leq F\Big(\frac{\theta_n}{r}\Big) r + F\Big(\frac{\theta_n}{t}\Big) t < \frac{1}{n}.$$

In particular, for n large enough, $\theta_n \in [K_1, K_2]$ for some constants $0 < K_1 < 1 < K_2$, and by extracting a subsequence θ_{n_j} it follows that $\theta_{n_j} \to \bar{\theta}$. The lower semicontinuity of F forces $\frac{\bar{\theta}}{r} = \frac{\bar{\theta}}{t} = 1$ so that $r = t$. □

4.2.2 Triangle inequality in the Entropy-Transport case

In this section we take advantage of the Corollary 4.2.4 in order to produce new distances in the space of measures coming from Entropy-Transport problems. We consider a Polish space X with a metric d and the class of power-like entropies $F = U_p$, $p \in \mathbb{R}$. The cost function $\boldsymbol{c} : X \times X \to [0, +\infty]$ will be a suitable function of the given metric d to be specified later, i.e. $\boldsymbol{c}(x_1, x_2) := \ell(\mathsf{d}(x_1, x_2))$ for a certain $\ell : [0, +\infty) \to [0, +\infty]$.

With these choices, we denote by H_p the induced marginal perspective cost. If $p \neq 0, 1$ it holds:

$$H_p(x_1, r; x_2, t) = \frac{2}{p}\Bigg[\mathfrak{M}_1(r, t) - \mathfrak{M}_{1-p}(r, t)\Big(1 + (1-p)\frac{\boldsymbol{c}(x_1, x_2)}{2}\Big)_+^{\frac{p}{p-1}}\Bigg]. \tag{4.2.17}$$

When $p = 1$ or $p = 0$ one gets:

$$H_1(x_1, r; x_2, t) = 2\Big[\mathfrak{M}_1(r, t) - \mathfrak{M}_0(r, t) e^{-\boldsymbol{c}(x_1, x_2)/2}\Big], \tag{4.2.18}$$

$$H_0(x_1, r; x_2, t) = r \ln r + t \ln t - (r + t) \ln\left(\frac{r + t}{2 + \boldsymbol{c}(x_1, x_2)}\right), \tag{4.2.19}$$

with standard meaning when $\boldsymbol{c} = +\infty$.

For a general cost function $\boldsymbol{c}$, if $p < 1$, we notice that $H_p(x_1, 0; x_2, t)$ depends on x_1 since $\mathfrak{M}_{1-p}(0, t) > 0$. In particular, H_p is not well defined on the cone $\mathfrak{C}(X)$ if $p < 1$. As we will see, when $p \geq 1$ the situation is more interesting.

We start by proving explicit bounds of H_p in terms of H_1, which immediately yield the corresponding bounds at the level of the induced Entropy-Transport cost.

Proposition 4.2.7. *Let H_p and H_1 be the functions defined in (4.2.17) and (4.2.18). Then for any $p > 1$, $x_1, x_2 \in X$ and $r, t \in [0, +\infty)$ it holds*

$$H_p(x_1, r; x_2, t) \leq H_1(x_1, r; x_2, t) \leq p H_p(x_1, r; x_2, t). \tag{4.2.20}$$

Proof. In order to prove the left inequality, we have to show that for any $p > 1$, $c \in [0, +\infty]$ and $r, t \in [0, +\infty)$ it holds

$$\mathfrak{M}_1(r, t) - \mathfrak{M}_{1-p}(r, t)\Big(1 + (1-p)c\Big)_+^{\frac{p}{p-1}} \leq p\mathfrak{M}_1(r, t) - p\mathfrak{M}_0(r, t) e^{-c}. \tag{4.2.21}$$

It is sufficient to study the case $c = 0$. Indeed the function

$$f(c) := \mathfrak{M}_{1-p}(r,t)\Big(1 + (1-p)c\Big)_+^{\frac{p}{p-1}} - p\mathfrak{M}_0(r,t)e^{-c} \tag{4.2.22}$$

is increasing in $[0, +\infty]$. This is clear when $c > \frac{1}{p-1}$, when $c \in [0, \frac{1}{p-1}]$ we compute the derivative of the function so that

$$f'(c) = p\mathfrak{M}_0(r,t)e^{-c} - p\mathfrak{M}_{1-p}(r,t)\Big(1 + (1-p)c\Big)^{\frac{1}{p-1}}$$

and this is nonnegative since $\mathfrak{M}_0 \geq \mathfrak{M}_{1-p}(r,t)$ and

$$e^{-c} \geq \Big(1 + (1-p)c\Big)^{\frac{1}{p-1}}. \tag{4.2.23}$$

Since $p - 1 > 0$, the last inequality easily follows from the well known $e^x \geq 1 + x$, $x \in \mathbb{R}$. So, we have to prove that

$$\mathfrak{M}_1(r,t) - \mathfrak{M}_{1-p}(r,t) \leq p\mathfrak{M}_1(r,t) - p\mathfrak{M}_0(r,t),$$

which can be rewritten (when $r \neq t$, otherwise it is obvious) as

$$1 - p \leq \frac{\mathfrak{M}_{1-p}(r,t) - \mathfrak{M}_0(r,t)}{\mathfrak{M}_1(r,t) - \mathfrak{M}_0(r,t)}$$

and the result follows by the bounds proved in [Kou14].

The right inequality in (4.2.20) is easier to obtain since it is equivalent to

$$0 \leq \mathfrak{M}_0(r,t)e^{-c} - \mathfrak{M}_{1-p}(r,t)\Big(1 + (1-p)c\Big)_+^{\frac{p}{p-1}},$$

and one can conclude using that $\mathfrak{M}_0(r,t) \geq \mathfrak{M}_{1-p}(r,t)$ and $e^{-c} \geq \Big(1 + (1-p)c\Big)_+^{\frac{p}{p-1}}$, which follows from the inequality (4.2.23) proved above and the monotonicity of the function x^p when $x \in [0,1]$ and $p > 1$. □

We prove now a crucial result in order to deduce that H_p satisfies the triangle inequality for $p \geq 1$.

Theorem 4.2.8. *Let* d *be a metric on a Polish space* X *and let* $h : [0, +\infty) \to [0, 1]$ *be a decreasing function such that* $h(0) = 1$. *Let us suppose that*

$$\bar{H}_1(x_1, r; x_2, t) := \mathfrak{M}_1(r,t) - \mathfrak{M}_0(r,t)h\left(\mathsf{d}(x_1, x_2)\right) \tag{4.2.24}$$

is the square of a distance on the cone $\mathfrak{C}(X)$. *Then*

$$\bar{H}_p(x_1, r; x_2, t) := \mathfrak{M}_1(r,t) - \mathfrak{M}_{1-p}(r,t)h\left(\mathsf{d}(x_1, x_2)\right) \tag{4.2.25}$$

is the square of a distance on the cone $\mathfrak{C}(X)$ *for every* $p > 1$.

Proof. Recalling the properties of the power means we have seen in section 2.2, it is apparent that $\bar{H}_p$ is nonnegative, symmetric and $\bar{H}_p(x_1, r; x_2, t) = 0$ if and only if $(x_1, r) \sim (x_2, t)$.

We have to prove that for every $p > 1$, for every metric d on X and for every $r, s, t \in [0, +\infty)$, $x_1, x_2, x_3 \in X$ it holds:

$$\sqrt{\bar{H}_p(x_1,r;x_3,t)} \leq \sqrt{\bar{H}_p(x_1,r;x_2,s)} + \sqrt{\bar{H}_p(x_2,s;x_3,t)}. \tag{4.2.26}$$

Since the function

$$d \mapsto \sqrt{\mathfrak{M}_1(r,t) - \mathfrak{M}_{1-p}(r,t)h(d)} \tag{4.2.27}$$

is increasing in $[0,+\infty)$ we can assume

$$\mathsf{d}(x_1,x_3) = \mathsf{d}(x_1,x_2) + \mathsf{d}(x_2,x_3).$$

Without loss of generality, we can also assume $r \leq t$ and we have to deal with three cases:

- $t < s$,
- $r < s \leq t$, (4.2.28)
- $s \leq r$.

Step 1. *Case $t < s$*

Lemma 4.2.9. *For any fixed r, t, x_1, x_2, x_3, the function*

$$s \mapsto \sqrt{\bar{H}_p(x_1,r;x_2,s)} + \sqrt{\bar{H}_p(x_2,s;x_3,t)} \tag{4.2.29}$$

is increasing in $[t,+\infty)$.

Proof. The result follows if we prove that for any fixed x_1, x_2 the function

$$f_p(u) = \bar{H}_p(x_1,1;x_2,u)$$

is increasing in $[1,+\infty)$. This is easy to prove since

$$f_p'(u) = \frac{1}{2} - \frac{u^{-p}}{2}\Big(\frac{1+u^{1-p}}{2}\Big)^{\frac{p}{1-p}} h(\mathsf{d}(x_1,x_2)) \geq \frac{1}{2} - \frac{u^{-p}}{2}\Big(\frac{1+u^{1-p}}{2}\Big)^{\frac{p}{1-p}} > 0, \tag{4.2.30}$$

where the last inequality holds because it is equivalent to the following

$$\mathfrak{M}_{1-p}(1,u) < u.$$

□

Thus, it is sufficient to prove the triangle inequality when $s \leq t$.

Step 2. *Case $r < s \leq t$*

We start with a useful lemma:

Lemma 4.2.10. *Let A, B, C three nonnegative numbers. Then*

$$\sqrt{C} \leq \sqrt{A} + \sqrt{B} \tag{4.2.31}$$

if and only if for every $\alpha, \beta \in (0,1)$ such that $\alpha + \beta = 1$ we have

$$C \leq \frac{A}{\alpha} + \frac{B}{\beta}. \tag{4.2.32}$$

Proof. Let us suppose (4.2.31). Then

$$C \le \left(\alpha\frac{\sqrt{A}}{\alpha} + \beta\frac{\sqrt{B}}{\beta}\right)^2 \le \frac{A}{\alpha} + \frac{B}{\beta}$$

where we have used the Jensen inequality for the convex function $f(x) = x^2$. In order to show that (4.2.32) $\Rightarrow$ (4.2.31) we notice that if $A = 0$ or $B = 0$ the result is clearly true, otherwise we choose α, β such that $\frac{\sqrt{A}}{\alpha} = \frac{\sqrt{B}}{\beta}$. Thus

$$(\sqrt{A} + \sqrt{B})^2 = \left(\alpha\frac{\sqrt{A}}{\alpha} + \beta\frac{\sqrt{B}}{\beta}\right)^2 = \frac{A}{\alpha} + \frac{B}{\beta} \ge C.$$

□

In order to simplify the notation, from now on we put $\mathsf{d}(x_1, x_3) = d_{13}$, $\mathsf{d}(x_1, x_2) = d_{12}$, $\mathsf{d}(x_2, x_3) = d_{23}$. Then, we can use Lemma 4.2.10 and the triangle inequality in the case $p = 1$ in order to derive a new inequality. Given $\alpha, \beta \in (0, 1)$ such that $\alpha + \beta = 1$, one gets:

$$\begin{aligned}
\bar{H}_p(x_1, r; x_3, t) &= \bar{H}_1(x_1, r; x_3, t) + \Big[\mathfrak{M}_0(r,t) - \mathfrak{M}_{1-p}(r,t)\Big]h(d_{13}) \\
&\le \frac{\bar{H}_1(x_1, r; x_2, s)}{\alpha} + \frac{\bar{H}_1(x_2, s; x_3, t)}{\beta} + \Big[\mathfrak{M}_0(r,t) - \mathfrak{M}_{1-p}(r,t)\Big]h(d_{13}) \\
&\le \frac{\bar{H}_p(x_1, r; x_2, s)}{\alpha} - \frac{\Big[\mathfrak{M}_0(r,s) - \mathfrak{M}_{1-p}(r,s)\Big]h(d_{12})}{\alpha} + \frac{\bar{H}_p(x_2, s; x_3, t)}{\beta} \\
&\quad - \frac{\Big[\mathfrak{M}_0(s,t) - \mathfrak{M}_{1-p}(s,t)\Big]h(d_{23})}{\beta} + \Big[\mathfrak{M}_0(r,t) - \mathfrak{M}_{1-p}(r,t)\Big]h(d_{13}) \\
&\le \frac{\bar{H}_p(x_1, r; x_2, s)}{\alpha} + \frac{\bar{H}_p(x_2, s; x_3, t)}{\beta},
\end{aligned} \tag{4.2.33}$$

where the last inequality in (4.2.33) is valid if and only if (using again Lemma 4.2.10):

$$\begin{aligned}
&\sqrt{\Big[\mathfrak{M}_0(r,t) - \mathfrak{M}_{1-p}(r,t)\Big]h(d_{13})} \\
\le &\sqrt{\Big[\mathfrak{M}_0(r,s) - \mathfrak{M}_{1-p}(r,s)\Big]h(d_{12})} + \sqrt{\Big[\mathfrak{M}_0(s,t) - \mathfrak{M}_{1-p}(s,t)\Big]h(d_{23})}.
\end{aligned} \tag{4.2.34}$$

We notice that $h(d_{13}) \le h(d_{12}) \wedge h(d_{23})$ since h is decreasing. Thus, it is enough to prove (4.2.34) in the case $d_{13} = d_{12} = d_{23} = 0$. Now, we adapt the strategy used in the proof of [ES03, Lemma 2] we put $u := \frac{r}{s} \in (0, 1)$, $\beta u := \frac{t}{s} \in (1, +\infty)$, so that β is a real number greater than 1. Thus, $\frac{1}{\beta} < u < 1$ and, denoted by $F(s)$ the function

$$F(s) = \sqrt{\Big[\mathfrak{M}_0(r,s) - \mathfrak{M}_{1-p}(r,s)\Big]} + \sqrt{\Big[\mathfrak{M}_0(s,t) - \mathfrak{M}_{1-p}(s,t)\Big]}, \tag{4.2.35}$$

it follows

$$4\sqrt{s}\frac{d}{ds}F(s) = g_p(u) + g_p(\beta u), \tag{4.2.36}$$

where

$$g_p(u) := \frac{\mathfrak{M}_0(u,1) - \frac{2}{u^{1-p}+1}\mathfrak{M}_{1-p}(u,1)}{\sqrt{\mathfrak{M}_0(u,1) - \mathfrak{M}_{1-p}(u,1)}}. \tag{4.2.37}$$

Lemma 4.2.11. *The function*

$$u \mapsto g_p(u) + g_p(\beta u)$$

is increasing in $(\frac{1}{\beta}, 1)$ *with only one zero inside the interval, so that* F *is minimized when* $s = r$ *or* $s = t$ *and the inequality* (4.2.34) *holds.*

Proof. Since g_p is continuous in $(0,1)$ and $(1,+\infty)$, it is enough to show that g_p is increasing in $(0,1)$ and $(1,+\infty)$, and

$$\lim_{u\to 1^-} g_p(u) = \sqrt{2(p-1)}, \quad \lim_{u\to 1^+} g_p(u) = -\sqrt{2(p-1)}.$$

The limits are easy to compute expanding the function near $u = 1$. When $u \in (0,1)\cup(1,+\infty)$ it follows:

$$g_p'(u) = \frac{(p-\frac{1}{2})u^{-p}\Big(\frac{u^{1-p}+1}{2}\Big)^{\frac{2p}{1-p}} - pu^{-p+\frac{1}{2}}\Big(\frac{u^{1-p}+1}{2}\Big)^{\frac{2p-1}{1-p}} + \frac{1}{2}}{2\big[\mathfrak{M}_0(u,1) - \mathfrak{M}_{1-p}(u,1)\big]^{\frac{3}{2}}}. \tag{4.2.38}$$

The proof is complete if we show that

$$(p-\frac{1}{2})u^{-p}\Big(\frac{u^{1-p}+1}{2}\Big)^{\frac{2p}{1-p}} - pu^{-p+\frac{1}{2}}\Big(\frac{u^{1-p}+1}{2}\Big)^{\frac{2p-1}{1-p}} + \frac{1}{2} > 0$$

for any $p > 1$ and any positive u. We put $v = \frac{u^{1-p}+1}{2}$, so that we have to prove

$$(p-\frac{1}{2})\Big(\frac{2v-1}{v^2}\Big)^{\frac{-p}{1-p}} - p\Big(\frac{2v-1}{v^2}\Big)^{\frac{1-2p}{2(1-p)}} + \frac{1}{2} > 0$$

for any $p > 1$ and $v \in (\frac{1}{2}, +\infty)$. Finally we put $w = \Big(\frac{2v-1}{v^2}\Big)^{\frac{1}{p-1}} \in (0,1)$ and we prove that

$$z_p(w) := (p-\frac{1}{2})w^p - pw^{p-\frac{1}{2}} + \frac{1}{2} > 0,$$

for any $p > 1$ and $w \in (0,1)$. To prove the last inequality, we notice that $z_p(1) = 0$ and z_p is a decreasing function because

$$z_p'(w) = p(p-\frac{1}{2})w^{p-\frac{3}{2}}(\sqrt{w}-1) < 0.$$

□

Step 3. *Case* $s \le r$

The strategy is to use again Lemma 4.2.10 and the triangle inequality for the case $p = 1$, but we have to derive a different inequality with respect to the previous step.

Lemma 4.2.12. *We denote with* $\theta_p : [0,+\infty)\times[0,+\infty) \to [0,+\infty)$ *the function*

$$\theta_p(r,t) := \frac{\mathfrak{M}_{1-p}(r,t)}{\mathfrak{M}_0(r,t)}.$$

Then $\theta_p(s,t) \le \theta_p(r,t)$.

Proof. It is sufficient to prove that $\theta_p(u,1)$ is increasing in $(0,1)$. This is easy to prove, indeed

$$\sqrt{u}\frac{d}{du}\theta_p(u,1) = \theta_p(u,1)\Big(\frac{u^{1-p}}{u^{1-p}+1} - \frac{1}{2}\Big) \ge 0.$$

□

Let α, β be any two numbers in $(0,1)$ such that $\alpha+\beta = 1$. Let us suppose, at first, $\theta_p(s,r) \le \theta_p(r,t)$. Then

$$\begin{aligned}\bar{H}_p(x_1,r;x_3,t) &= \bar{H}_1(x_1,r;x_3,t)\theta_p(r,t) + \mathfrak{M}_1(r,t)\big(1-\theta_p(r,t)\big)\\ &\le \frac{\bar{H}_1(x_1,r;x_2,s)}{\alpha}\theta_p(r,t) + \frac{\mathfrak{M}_1(r,s)}{\alpha}\big(1-\theta_p(r,t)\big)\\ &+ \frac{\bar{H}_1(x_2,s;x_3,t)}{\beta}\theta_p(r,t) + \frac{\mathfrak{M}_1(s,t)}{\beta}\big(1-\theta_p(r,t)\big)\\ &\le \frac{\bar{H}_p(x_1,r;x_2,s)}{\alpha} + \frac{\bar{H}_p(x_2,s;x_3,t)}{\beta}, \end{aligned} \tag{4.2.39}$$

where the first inequality of (4.2.39) follows by the triangle inequalities satisfied by $\sqrt{\bar{H}_1}$ and by $\sqrt{\mathfrak{M}_1}$ (where the latter is straightforward to prove), while the second inequality follows because

$$\theta_p(s,r) \le \theta_p(r,t), \quad \theta_p(s,t) \le \theta_p(r,t) \quad \text{and} \quad \bar{H}_1 \le \mathfrak{M}_1.$$

It remains to investigate the case $\theta_p(s,r) > \theta_p(r,t)$. Let us suppose

$$\sqrt{\mathfrak{M}_1(r,t)\big(1-\theta_p(r,t)\big)} \le \sqrt{\mathfrak{M}_1(s,r)\big(1-\theta_p(s,r)\big)} + \sqrt{\mathfrak{M}_1(s,t)\big(1-\theta_p(r,t)\big)}. \tag{4.2.40}$$

Then

$$\begin{aligned}\bar{H}_p(x_1,r;x_3,t) &= \bar{H}_1(x_1,r;x_3,t)\theta_p(r,t) + \mathfrak{M}_1(r,t)\big(1-\theta_p(r,t)\big)\\ &\le \frac{\bar{H}_1(x_1,r;x_2,s)}{\alpha}\theta_p(r,t) + \frac{\mathfrak{M}_1(s,r)}{\alpha}\big(1-\theta_p(s,r)\big)\\ &+ \frac{\bar{H}_1(x_2,s;x_3,t)}{\beta}\theta_p(r,t) + \frac{\mathfrak{M}_1(s,t)}{\beta}\big(1-\theta_p(r,t)\big)\\ &\le \frac{\bar{H}_1(x_1,r;x_2,s)}{\alpha}\theta_p(s,r) + \frac{\mathfrak{M}_1(s,r)}{\alpha}\big(1-\theta_p(s,r)\big)\\ &+ \frac{\bar{H}_1(x_2,s;x_3,t)}{\beta}\theta_p(r,t) + \frac{\mathfrak{M}_1(s,t)}{\beta}\big(1-\theta_p(r,t)\big)\\ &\le \frac{\bar{H}_p(x_1,r;x_2,s)}{\alpha} + \frac{\bar{H}_p(x_2,s;x_3,t)}{\beta}, \end{aligned} \tag{4.2.41}$$

where in the first inequality we use (4.2.40), in the second we use the hypothesis $\theta_p(s,r) > \theta_p(r,t)$, in the third we reason as in the second step of the inequality (4.2.39) in order to replace $\theta_p(r,t)$ with $\theta_p(s,t)$.

Finally, the proof is complete if we prove the inequality (4.2.40). Since the case $r = s$ is trivial, we put $u := \frac{s}{r} < 1$, $v := \frac{t}{r} > 1$, so that we can rewrite the inequality (4.2.40) in the following equivalent way

$$(1+\sqrt{u})^2 \frac{\mathfrak{M}_0(u,1) - \mathfrak{M}_{-1}(u,1)}{\mathfrak{M}_0(u,1) - \mathfrak{M}_{1-p}(u,1)} \le \frac{\sqrt{v}\big(\sqrt{u+v}+\sqrt{1+v}\big)^2}{\mathfrak{M}_0(1,v) - \mathfrak{M}_{1-p}(1,v)}. \tag{4.2.42}$$

Now we use the estimate

$$\big(\sqrt{u+v}+\sqrt{1+v}\big)^2 \ge 1+4v,$$

so that it is sufficient to prove that for any $u \in (0,1)$ and any $v \in (1,+\infty)$

$$(1+\sqrt{u})^2 \frac{\mathfrak{M}_0(u,1) - \mathfrak{M}_{-1}(u,1)}{\mathfrak{M}_0(u,1) - \mathfrak{M}_{1-p}(u,1)} \le \frac{\sqrt{v}(1+4v)}{\mathfrak{M}_0(1,v) - \mathfrak{M}_{1-p}(1,v)}. \tag{4.2.43}$$

It is easy to see that the last inequality is true at least if $p \ge \frac{3}{2}$. For example, one can bound the left hand side with

$$l(u) := (1+\sqrt{u})^2 \frac{\mathfrak{M}_0(u,1) - \mathfrak{M}_{-1}(u,1)}{\mathfrak{M}_0(u,1) - \mathfrak{M}_{-\frac{1}{2}}(u,1)},$$

and the right hand side with

$$r(u) := \frac{\sqrt{v}(1+4v)}{\sqrt{v}-1}.$$

Then, standard computations show that:

$$\sup_{u\in(0,1)} l(u) < \inf_{u\in(1,+\infty)} r(u).$$

If $1 < p < \frac{3}{2}$ one needs precise bounds that we have found in [Kou14]. The supremum of the left hand side of (4.2.43) is $\frac{4}{p-1}$. For the right hand side of (4.2.43) one has:

$$\begin{aligned}\frac{\sqrt{v}(1+4v)}{\mathfrak{M}_0(v,1) - \mathfrak{M}_{1-p}(v,1)} &= \frac{\mathfrak{M}_{p-1}(1,v)(1+4v)}{\mathfrak{M}_{p-1}(1,v) - \mathfrak{M}_0(1,v)} \\ &\ge \frac{\sqrt{v}(1+4v)}{\mathfrak{M}_{p-1}(1,v) - \mathfrak{M}_0(1,v)} \ge 4\frac{\mathfrak{M}_1(1,v) - \mathfrak{M}_0(1,v)}{\mathfrak{M}_{p-1}(1,v) - \mathfrak{M}_0(1,v)},\end{aligned}$$

and again using the results in [Kou14] it is proved that the sharp lower bound for the last expression is $\frac{4}{p-1}$.

□

Let us now define the two Entropy-Transport distances studied in [LMS18a, Section 7].

Definition 4.2.13. *We define the Hellinger-Kantorovich distance* HK *as the square root of the Entropy-Transport cost induced by the entropy function* U_1 *and the cost*

$$\boldsymbol{c}_{\mathsf{HK}}(x_1,x_2) = \begin{cases} -\log\big(\cos^2(\mathsf{d}(x_1,x_2))\big) & \text{if } \mathsf{d}(x_1,x_2) < \frac{\pi}{2}, \\ +\infty & \text{otherwise,} \end{cases} \tag{4.2.44}$$

which produce the marginal perspective cost

$$H(x_1, r; x_2, t) := r + t - 2\sqrt{rt}\cos(\mathsf{d}(x_1, x_2) \wedge \pi/2). \tag{4.2.45}$$

We define the Gaussian Hellinger-Kantorovich distance GHK *as the square root of the Entropy-Transport cost induced by the entropy function* U_1 *and the cost* $\boldsymbol{c}(x_1, x_2) = \mathsf{d}^2(x_1, x_2)$*, which produce the marginal perspective cost*

$$H(x_1, r; x_2, t) := r + t - 2\sqrt{rt}\exp\left(-\mathsf{d}^2(x_1, x_2)/2\right). \tag{4.2.46}$$

The name Hellinger-Kantorovich is suggested by the fact that HK can be understood as an inf-convolution (see [LMS18a, Remark 8.19]) between the Kantorovich-Wasserstein distance $\mathcal{W}_2$ and the Hellinger distance He defined by

$$\mathsf{He}^2(\gamma, \mu) := D_{M_{1/2}}(\gamma || \mu), \tag{4.2.47}$$

where $M_{1/2}$ is the Matusita's divergence introduced in Example 3.

By taking advantage of Corollary 4.2.4, it is straightforward to prove that HK and GHK are indeed distances since the expressions (4.2.45) and (4.2.46) are strictly related to the metric $\mathsf{d}_{\mathfrak{C}}$ defined in (2.3.2).

In the next Theorem we discuss some possible choices for the cost function $\boldsymbol{c}$ and we produce a new class of metrics in the cone space $\mathfrak{C}(X)$ (thus, also in $\mathscr{M}(X)$ thanks to Corollary 4.2.4).

Theorem 4.2.14. *Let* (X, d) *be a Polish space,* $F = U_p$ *and consider one of the two following cost functions:*

1. $\boldsymbol{c}_p(x_1, x_2) := \frac{2}{p-1}\left[1 - \left(\cos(\mathsf{d}(x_1, x_2) \wedge \pi/2)\right)^{\frac{p-1}{p}}\right]$.
2. $\boldsymbol{c}(x_1, x_2) = \mathsf{d}(x_1, x_2)$.

Then, the induced marginal perspective cost is the square of a distance on $\mathfrak{C}(X)$ *for every* $p > 1$.

Moreover, if

3. $\boldsymbol{c}(x_1, x_2) = \mathsf{d}^2(x_1, x_2)$

the induced marginal perspective cost is the square of a distance on $\mathfrak{C}(X)$ *for every* $1 < p \le 3$.

In particular, for every possible choice of entropy and cost functions mentioned in the theorem the induced Entropy-Transport cost ET *is the square of a distance on* $\mathscr{M}(X)$.

Proof. **Case 1.** With the choice $F = U_p$ and $\boldsymbol{c}_p$ defined above, we obtain

$$H(x_1, r; x_2, t) = \frac{2}{p}\left[\mathfrak{M}_1(r, t) - \mathfrak{M}_{1-p}(r, t)\cos(\mathsf{d}(x_1, x_2) \wedge \pi/2)\right].$$

The assertion follows from Theorem 4.2.8 applied with $h(d) := \cos(d \wedge \pi/2)$ and the fact that

$$\sqrt{\mathfrak{M}_1(r, t) - \mathfrak{M}_0(r, t)\cos(d \wedge \pi/2)}$$

is a metric on $\mathfrak{C}(X)$.

Case 2. In this situation we obtain

$$H(x_1, r; x_2, t) = \frac{2}{p}\left[\mathfrak{M}_1(r, t) - \mathfrak{M}_{1-p}(r, t)\left(1 + (1-p)\frac{\mathsf{d}(x_1, x_2)}{2}\right)_+^{\frac{p}{p-1}}\right].$$

Again we want to apply Theorem 4.2.8 with

$$h(d) := \left(1 + (1-p)\frac{\mathsf{d}(x_1, x_2)}{2}\right)_+^{\frac{p}{p-1}}.$$

It is clear that $h(0) = 1$ and h is nonnegative and decreasing since $1 - p < 0$ and $p/(p-1) > 0$. It remains to show that

$$\mathfrak{M}_1(r,t) - \mathfrak{M}_0(r,t)\left(1 + (1-p)\frac{\mathsf{d}(x_1, x_2)}{2}\right)_+^{\frac{p}{p-1}} \tag{4.2.48}$$

is the square of a metric on $\mathfrak{C}(X)$. We already know, as a consequence of Lemma 3.2.7, that the map

$$d \mapsto \left(1 + (1-p)\frac{d}{2}\right)_+^{\frac{p}{p-1}}$$

is convex and decreasing with values in $[0, 1]$. Since $x \mapsto \arccos(x)$ is a concave function in $[0, 1]$, it follows that

$$l_p(d) := \arccos\left[\left(1 + (1-p)\frac{d}{2}\right)_+^{\frac{p}{p-1}}\right]$$

is concave and $l_p(d) = 0$ if and only if $d = 0$. Now we conclude by taking advantage of Lemma 4.1.2, since (4.2.45) is the square of a metric on $\mathfrak{C}(X)$ for every metric d on X and

$$\begin{aligned}\mathfrak{M}_1(r,t) - \mathfrak{M}_0(r,t)\cos(l_p(\mathsf{d}(x_1,x_2)) \wedge \pi/2)&\\ = \mathfrak{M}_1(r,t) - \mathfrak{M}_0(r,t)\left(1 + (1-p)\frac{\mathsf{d}(x_1, x_2)}{2}\right)_+^{\frac{p}{p-1}}&.\end{aligned}$$

Case 3. We reason as in the proof of point 2). Now, we have to show that the function

$$f_p(d) = \arccos\left[(1 - (p-1)\frac{d^2}{2})_+^{\frac{p}{p-1}}\right]$$

is concave and $f_p(d) = 0$ if and only if $d = 0$. The second statement is obvious, for the first one we cannot reason as before so we proceed with explicit computations. We notice that it is enough to prove that the function is concave when $d \in \left(0, \sqrt{\frac{2}{p-1}}\right)$. Let us compute the second derivative: we put

$$g_p(d) = \left(1 + (1-p)\frac{d^2}{2}\right)^{\frac{p}{p-1}},$$

so that

$$f_p(d) = \arccos(g_p(d)),$$

$$g_p'(d) = \frac{-pdg_p(d)}{\left(1 + (1-p)\frac{d^2}{2}\right)},$$

$$g_p''(d) = \frac{p\left((p+1)\frac{d^2}{2} - 1\right)g_p(d)}{\left(1 + (1-p)\frac{d^2}{2}\right)^2}.$$

Thus

$$
\begin{aligned}
f_p''(d) &= -\frac{(1-g_p(d)^2)g_p''(d)+g_p(d)g_p'(d)^2}{(1-g_p(d)^2)^{\frac{3}{2}}}\\
&= -\frac{p\Big((p+1)\frac{d^2}{2}-1\Big)g_p(d)\Big(1-g_p(d)^2\Big)+p^2d^2g_p(d)^3}{\Big(1-g_p(d)^2\Big)^{\frac{3}{2}}\Big(1+(1-p)\frac{d^2}{2}\Big)^2} \qquad (4.2.49)\\
&= -\frac{pg_p(d)\Big[(p+1)\frac{d^2}{2}-1+g_p(d)^2\big((p-1)\frac{d^2}{2}+1\big)\Big]}{\Big(1-g_p(d)^2\Big)^{\frac{3}{2}}\Big(1+(1-p)\frac{d^2}{2}\Big)^2}
\end{aligned}
$$

and f_p is concave if and only if

$$
(p+1)\frac{d^2}{2}-1+\Big(1+(1-p)\frac{d^2}{2}\Big)^{\frac{2p}{p-1}}\big((p-1)\frac{d^2}{2}+1\big)\geq 0 \tag{4.2.50}
$$

for every $p\in(1,3]$ and $d\in\Big(0,\sqrt{\frac{2}{p-1}}\Big)$.

We put $z:=1-(p-1)\frac{d^2}{2}\in(0,1)$ and (4.2.50) follows if we prove that

$$
h_p(z):=\frac{p+1}{p-1}(1-z)-1+z^{\frac{2p}{p-1}}(2-z)\geq 0 \tag{4.2.51}
$$

for every $z\in(0,1)$ and every $p\in(1,3]$. We have

$$
h_p'(z)=-\frac{(3p-1)z^{\frac{2p}{p-1}}-4pz^{\frac{p+1}{p-1}}+p+1}{p-1},
$$

$$
h_p''(z)=-\frac{2p(3p-1)z^{\frac{p+1}{p-1}}-4p(p+1)z^{\frac{2}{p-1}}}{(p-1)^2},
$$

and it is straightforward to deduce that in the interval $(0,1)$, for every $p\in(1,3]$, the function h_p is convex, so that h_p' is increasing and thus nonpositive since $h_p'(1)=0$. This implies that h_p is decreasing but $h_p(1)=0$ so inequality (4.2.51) is satisfied.

The last assertion is a consequence of Corollary 4.2.4 if we prove that the marginal perspective costs in consideration are lower semicontinuous with respect to the product topology induced by $\tau_{\mathfrak{C}(X)}$. To see this we notice that if $\boldsymbol{c}(x_1,x_2)=\mathsf{d}^2(x_1,x_2)$ then the induced marginal perspective cost is metrically equivalent to the marginal perspective cost (4.2.46) thanks to Proposition 4.2.7. If $\boldsymbol{c}(x_1,x_2)=\mathsf{d}(x_1,x_2)$, we observe that if (X,d) is a metric space then also $(X,\sqrt{\mathsf{d}})$ is a metric space inducing the same topology thanks to Lemma 4.1.2, so we can conclude using the previous point. The same argument can be applied also to the case

$$
\boldsymbol{c}_p(x_1,x_2)=\frac{2}{p-1}\left[1-\big(\cos(\mathsf{d}(x_1,x_2)\wedge\pi/2)\big)^{\frac{p-1}{p}}\right]
$$

since the function

$$
g(d):=\frac{2}{p-1}\left[1-\big(\cos(d\wedge\pi/2)\big)^{\frac{p-1}{p}}\right]
$$

is convex, strictly increasing (it is sufficient to notice that the cosine function is concave strictly decreasing in $[0,\pi/2]$ and $(p-1)/p\in(0,1)$) and $g(0)=0$, so that its inverse function is concave, strictly increasing and $g^{-1}(0)=0$. □

Corollary 4.2.15. *Every Entropy-Transport metric considered in Theorem 4.2.14 is a complete and separable distance on $\mathscr{M}(X)$ inducing the weak topology.*

Proof. In [LMS18a, Theorem 7.25] it is proved that, for every metric d on X, the Gaussian Hellinger-Kantorovich distance is a complete and separable metric on the space $\mathscr{M}(X)$ inducing the weak topology. From the proof of Theorem 4.2.14 it follows that every Entropy-Transport metric in consideration is topological equivalent to the Gaussian Hellinger-Kantorovich distance and thus the conclusion follows. □

Chapter 5

Sturm-Entropy-Transport distances

In this chapter we adapt a construction due to Sturm [Stu06], who introduced the **D**-distance (see Definition 5.3.1 below) on the family of isomorphism classes of metric measure spaces (X, d, μ) such that $\mu \in \mathscr{P}_2(X)$. Instead of the 2-Wasserstein distance considered by Sturm, we use here the Entropy-Transport distances studied in the previous chapter and we define new complete and separable distances between metric measure spaces without requiring neither the normalization of the mass nor the finiteness of the second moment.

5.1 Metric measure spaces and couplings

Let us consider two metric measure spaces $(X_1, \mathsf{d}_1, \mu_1)$ and $(X_2, \mathsf{d}_2, \mu_2)$ in the sense of definition 2.6.1. We say that $(X_1, \mathsf{d}_1, \mu_1)$ and $(X_2, \mathsf{d}_2, \mu_2)$ are *isomorphic* if there exists an isometry $\psi : \mathsf{supp}(\mu_1) \to \mathsf{supp}(\mu_2)$ such that $\psi_\sharp \mu_1 = \mu_2$. It is apparent that the relation of being isomorphic is an equivalence relation on the family of metric measure spaces and that $\mu_1(X_1) = \mu_2(X_2)$ is a necessary condition in order to be isomorphic.

We denote by $\mathbf{X}$ the family of all isomorphism classes of metric measure spaces with finite measure, namely a class $[(X, \mathsf{d}, \mu)] \in \mathbf{X}$ if (X, d, μ) is a metric measure space and $\mu \in \mathscr{M}(X)$. From now on, with a slight abuse of notation, we will identify a metric measure space with its class.

Let $(X_1, \mathsf{d}_1, \mu_1)$ and $(X_2, \mathsf{d}_2, \mu_2)$ be two metric measure spaces and let $X_1 \sqcup X_2$ be their disjoint union. We say that a pseudo-metric $\hat{\mathsf{d}}$ on $X_1 \sqcup X_2$ is a *metric coupling* between d_1 and d_2 if $\hat{\mathsf{d}}(x, y) = \mathsf{d}_1(x, y)$ for every $x, y \in X_1$ and $\hat{\mathsf{d}}(x, y) = \mathsf{d}_2(x, y)$ for every $x, y \in X_2$.

It is not difficult to show that we can always consider a finite valued coupling $\hat{\mathsf{d}}$ between two metrics d_1 and d_2: to construct it, fix two points $\bar{x}_1 \in X_1$, $\bar{x}_2 \in X_2$, a number $c \in \mathbb{R}_+$, and define $\hat{\mathsf{d}}$ as:

$$\hat{\mathsf{d}}(x, y) := \begin{cases} \mathsf{d}_1(x, y) & \text{if } x, y \in X_1\,, \\ \mathsf{d}_2(x, y) & \text{if } x, y \in X_2\,, \\ \mathsf{d}_1(x, \bar{x}_1) + c + \mathsf{d}_2(\bar{x}_2, y) & \text{if } x \in X_1\,, y \in X_2, \\ \mathsf{d}_1(y, \bar{x}_1) + c + \mathsf{d}_2(\bar{x}_2, x) & \text{if } y \in X_1\,, x \in X_2. \end{cases} \tag{5.1.1}$$

Moreover, from any finite value coupling $\hat{\mathsf{d}}$ between d_1 and d_2 and any $\epsilon > 0$, we can obtain a complete, separable metric $\hat{\mathsf{d}}_\epsilon$ which is again a coupling of d_1 and d_2 in the following way:

$$\hat{\mathsf{d}}_\epsilon := \begin{cases} \hat{\mathsf{d}} & \text{on } (X_1 \times X_1) \sqcup (X_2 \times X_2), \\ \hat{\mathsf{d}} + \epsilon & \text{on } (X_1 \times X_2) \sqcup (X_2 \times X_1). \end{cases} \tag{5.1.2}$$

We say that a measure $\boldsymbol{\gamma} \in \mathscr{M}(X_1 \times X_2)$ is a *measure coupling* between μ_1 and μ_2 if

$$\boldsymbol{\gamma}(A \times X_2) = \mu_1(A) \quad \text{and} \quad \boldsymbol{\gamma}(X_1 \times B) = \mu_2(B), \tag{5.1.3}$$

for every Borel sets $A \subset X_1$ and $B \subset X_2$.

We keep the usual notation γ_i for the (standard) marginals of the measure $\boldsymbol{\gamma} \in \mathscr{M}(X_1 \times X_2)$ and π^i for the projection maps $\pi^i : X_1 \times X_2 \to X_i$, $i = 1, 2$.

5.2 Regular Entropy-Transport distances

In Section 4.2.2 we have constructed and studied a class of Entropy-Transport distances. We see now how we can use some of these metrics to define new distances between metric measure spaces with finite measure.

Let us introduce the class of *regular* Entropy-Transport distances.

Definition 5.2.1. *Let (X, d) be a Polish space. We say that D_{ET} is a regular Entropy-Transport distance if D_{ET} is one of the following possible distances on $\mathscr{M}(X)$:*

1. *The Hellinger-Kantorovich distance HK induced by the entropy U_1 defined in the Example 4 and the cost $\boldsymbol{c}_{\mathrm{HK}}$ defined in (4.2.44).*
2. *The Entropy-Transport distances induced by the entropy U_p defined in the Example 4, $1 \le p \le 3$, and the cost $\boldsymbol{c} = \mathsf{d}^2$.*

We will refer to the Entropy-Transport distances considered in the second point of the previous definition as *power-like-Wasserstein distance of order p*. We notice that this class includes, for $p = 1$, the Gaussian Hellinger-Kantorovich distance GHK.

Every regular Entropy-Transport distance is induced by a *superlinear* entropy function F, and by a cost function of the form $\boldsymbol{c} = \ell(\mathsf{d})$, where $\ell : [0, \infty) \to [0, \infty]$ is

$$\ell(d) = \ell_{\mathrm{HK}}(d) := \begin{cases} -\log\left(\cos^2(d)\right) & \text{if } d < \frac{\pi}{2}, \\ +\infty & \text{otherwise}, \end{cases} \quad \text{or} \quad \ell(d) = d^2. \tag{5.2.1}$$

In particular, the possible functions ℓ in consideration are continuous and convex, possibly attaining the value $+\infty$. Moreover, $\ell''_{\mathrm{HK}}(d) = 2\exp(\ell_{\mathrm{HK}}(d))$ and $\ell_{\mathrm{HK}}(0) = \ell'_{\mathrm{HK}}(0) = 0$ so that

$$\ell_{\mathrm{HK}}(d) \ge d^2 \qquad \text{for every } d \in [0, \infty). \tag{5.2.2}$$

We recall that, thanks to Corollary 4.2.15, every regular Entropy-Transport distance is a complete and separable metric on $\mathscr{M}(X)$ inducing the weak convergence.

The following estimates will be useful later on.

Proposition 5.2.2. *Let D_{ET} be a regular Entropy-Transport distance on a Polish space (X, d). Then*

$$\mathsf{D}_{\mathrm{ET}}(\mu_1, \mu_2) \le \mathsf{HK}(\mu_1, \mu_2) \le \min\left(\mathsf{He}(\mu_1, \mu_2), \mathcal{W}_2(\mu_1, \mu_2)\right) \quad \textit{for every } \mu_1, \mu_2 \in \mathscr{M}(X), \tag{5.2.3}$$

where He is the Hellinger distance defined in (4.2.47).

Proof. The left inequality is a consequence of $\mathsf{GHK} \le \mathsf{HK}$ (which follows directly from 5.2.2) and of Proposition 4.2.7. The right inequality is proved in [LMS18a, Section 7.7]. □

5.3 Sturm-Entropy-Transport distances

Let us recall the definition of the $\mathbf{D}$-distance given by Sturm [Stu06, Section 3.1].

Definition 5.3.1. *Let $(X_1, \mathsf{d}_1, \mu_1)$ and $(X_2, \mathsf{d}_2, \mu_2)$ be two metric measure spaces, the Sturm $\mathbf{D}$-distance is defined as*

$$\mathbf{D}\big((X_1, \mathsf{d}_1, \mu_1), (X_2, \mathsf{d}_2, \mu_2)\big) := \inf \mathcal{W}_2(\psi^1_\sharp \mu_1, \psi^2_\sharp \mu_2) \tag{5.3.1}$$

where the infimum is taken over all metric spaces $(\hat{X}, \hat{\mathsf{d}})$ with isometric embeddings $\psi^1 : \mathsf{supp}(\mu_1) \to \hat{X}$ and $\psi^2 : \mathsf{supp}(\mu_2) \to \hat{X}$.

Actually, $\mathbf{D}$ is a metric (see [Stu06, Theorem 3.6]) only on the family $\tilde{\mathbf{X}}(M)$ of all isomorphism classes of metric measure spaces with the same total mass M and *finite variance*, i.e. $[(X, \mathsf{d}, \mu)] \in \tilde{\mathbf{X}}(M)$ if (X, d, μ) is a metric measure space with $\mu(X) = M$ and

$$\mathrm{Var}(X, \mathsf{d}, \mu) := \inf \int_{X'} \big(\mathsf{d}'(z, x)\big)^2 \,\mathrm{d}\mu'(x) < +\infty, \tag{5.3.2}$$

where the infimum is taken over all the metric measure spaces (X', d', μ') isomorphic to (X, d, μ) and over all $z \in X'$.

We are now ready to define in a similar way the *Sturm-Entropy-Transport distance* induced by D_{ET}.

Definition 5.3.2. *Let D_{ET} be a regular Entropy-Transport distance. Let $(X_1, \mathsf{d}_1, \mu_1)$ and $(X_2, \mathsf{d}_2, \mu_2)$ be two metric measure spaces, we define the Sturm-Entropy-Transport $\mathbf{D}_{\mathrm{ET}}$-distance induced by D_{ET} as*

$$\mathbf{D}_{\mathrm{ET}}\big((X_1, \mathsf{d}_1, \mu_1), (X_2, \mathsf{d}_2, \mu_2)\big) := \inf \mathsf{D}_{\mathrm{ET}}(\psi^1_\sharp \mu_1, \psi^2_\sharp \mu_2) \tag{5.3.3}$$

where the infimum is taken over all metric spaces $(\hat{X}, \hat{\mathsf{d}})$ with isometric embeddings $\psi^1 : \mathsf{supp}(\mu_1) \to \hat{X}$ and $\psi^2 : \mathsf{supp}(\mu_2) \to \hat{X}$.

It is not difficult to prove that the definition is well-posed. Indeed, let us suppose $(X_i', \mathsf{d}_i', \mu_i')$ is isomorphic to $(X_i, \mathsf{d}_i, \mu_i)$ through the map φ^i, $i = 1, 2$. Then, for every metric space $\hat{X}$ and every isometric embedding $\psi^i : \mathsf{supp}(\mu_i) \to \hat{X}$, $i = 1, 2$, we have that

$$\mathsf{D}_{\mathrm{ET}}\left((\psi^1 \circ \varphi^1)_\sharp \mu_1, (\psi^2 \circ \varphi^2)_\sharp \mu_2\right) = \mathsf{D}_{\mathrm{ET}}(\psi^1_\sharp \mu_1, \psi^2_\sharp \mu_2).$$

Remark 5.3.3. *In the Definition 5.3.2 we can suppose without loss of generality that $\hat{X} = X_1 \sqcup X_2$, $\psi^1 = \iota_1$, $\psi^2 = \iota_2$ be respectively the inclusion of X_1 and X_2 in $X_1 \sqcup X_2$ and the infimum is taken over all coupling $\hat{\mathsf{d}}$ of d_1 and d_2. In this case we will identify μ_k with $(\iota_k)_\sharp \mu_k$, $k = 1, 2$.*

By taking advantage of the primal formulation of the Entropy-Transport problem introduced in Definition 3.2.2, we can give a more explicit formulation of the distance $\mathbf{D}_{\mathrm{ET}}$.

Proposition 5.3.4. *Let D_{ET} be a regular Entropy-Transport distance induced by the entropy F and the cost $\boldsymbol{c} = \ell(\mathsf{d})$. For every $(X_1, \mathsf{d}_1, \mu_1)$, $(X_2, \mathsf{d}_2, \mu_2)$ metric measure spaces we have*

$$\begin{aligned} &\mathbf{D}^2_{\mathrm{ET}}\left((X_1, \mathsf{d}_1, \mu_1), (X_2, \mathsf{d}_2, \mu_2)\right) = \\ &= \inf \left\{ \sum_{i=1}^{2} D_F(\gamma_i || \mu_i) + \int_{X_1 \times X_2} \ell(\hat{\mathsf{d}}(x_1, x_2)) \mathrm{d}\boldsymbol{\gamma} : \begin{array}{l} \boldsymbol{\gamma} \in \mathscr{M}(X_1 \times X_2), \\ \hat{\mathsf{d}} \text{ coupling of } \mathsf{d}_1 \text{ and } \mathsf{d}_2. \end{array} \right\} \end{aligned} \tag{5.3.4}$$

Proof. This follows directly from the definition of $\mathbf{D}_{\mathrm{ET}}$, recalling that every regular Entropy-Transport distance is the square root of the corresponding Entropy-Transport cost, Remark 5.3.3 and the primal formulation of Entropy-Transport problems. □

In the next Lemma we collect some basic properties of the function $\mathbf{D}_{\mathrm{ET}}$.

Lemma 5.3.5. *Let D_{ET} be a regular Entropy-Transport. Then*

(i) For any $M \geq 0$ it holds

$$\mathbf{D}_{\mathrm{ET}}((X_1, \mathsf{d}_1, M\mu_1), (X_2, \mathsf{d}_2, M\mu_2)) = \sqrt{M}\mathbf{D}_{\mathrm{ET}}((X_1, \mathsf{d}_1, \mu_1), (X_2, \mathsf{d}_2, \mu_2)). \tag{5.3.5}$$

(ii) If $(X_1, \mathsf{d}_1) = (X_2, \mathsf{d}_2)$ then

$$\mathbf{D}_{\mathrm{ET}}((X_1, \mathsf{d}_1, \mu_1), (X_2, \mathsf{d}_2, \mu_2)) \leq \mathsf{D}_{\mathrm{ET}}(\mu_1, \mu_2). \tag{5.3.6}$$

(iii) The following inequality holds

$$\mathbf{D}_{\mathrm{ET}}((X_1, \mathsf{d}_1, \mu_1), (X_2, \mathsf{d}_2, \mu_2)) \leq \mathbf{D}((X_1, \mathsf{d}_1, \mu_1), (X_2, \mathsf{d}_2, \mu_2)). \tag{5.3.7}$$

(iv) The set

$$\mathbf{X}_* := \Big\{(X, \mathsf{d}, \mu) \in \mathbf{X},\ \mathsf{supp}(\mu) = \{x_1, ..., x_N\},\ N \in \mathbb{N},\ \mu = M\sum_{n=1}^{N} \delta_{x_n},\ M \in \mathbb{R}_+\Big\} \tag{5.3.8}$$

is dense in $(\mathbf{X}, \mathbf{D}_{\mathrm{ET}})$.

(v) If

$$\mu = M\sum_{n=1}^{N} \delta_{x_n} \text{ and } \mu' = M\sum_{n=1}^{N} \delta_{x'_n}, \tag{5.3.9}$$

then

$$\mathbf{D}_{\mathrm{ET}}((X, \mathsf{d}, \mu), (X', \mathsf{d}', \mu')) \leq \sqrt{MN} \sup_{i,j} |\mathsf{d}_{ij} - \mathsf{d}'_{ij}|, \tag{5.3.10}$$

where we put $\mathsf{d}_{ij} := \mathsf{d}(x_i, x_j)$ and $\mathsf{d}'_{ij} := \mathsf{d}'(x'_i, x'_j)$.

(vi) For any $M \geq 0$,

$$\mathbf{D}_{\mathrm{ET}}((X, \mathsf{d}, \mu), (X, \mathsf{d}, M\mu)) \leq |\sqrt{M} - 1|\sqrt{\mu(X)}. \tag{5.3.11}$$

Proof. (i) This is a direct consequence of the 1-homogeneity of the functional ET (Corollary 3.2.6) and the fact that $\psi_\sharp(M\mu) = M\psi_\sharp\mu$.

(ii) The result follows from the definition of $\mathbf{D}_{\mathrm{ET}}$, since $(\hat{X}, \hat{\mathsf{d}}) = (X_1, \mathsf{d}_1)$ with $\psi_1 = \psi_2 = \mathrm{Id}$ is an admissible competitor for the infimum.

(iii) This follows again by the definition of the distances $\mathbf{D}_{\mathrm{ET}}$ and $\mathbf{D}$ and the bound $\mathsf{D}_{\mathrm{ET}} \leq \mathcal{W}_2$.

(iv) The result is a consequence of the second point of this Lemma, the fact that D_{ET} metrizes the weak convergence and the density in $\mathscr{M}(X)$ (with respect to the weak convergence) of the measures of the form $\mu = M\sum_{n=1}^{N} \delta_{x_i}$.

(v) This follows from the point (iii) of this Lemma and the point (iv) of [Stu06, Lemma 3.5].

(vi) It is sufficient to use the definition of $\mathbf{D}_{\mathrm{ET}}$ and the bound $\mathsf{D}_{\mathrm{ET}} \leq \mathsf{He}$.

□

The next Lemma shows the existence of the optimal couplings.

Lemma 5.3.6. *Let D_{ET} be a regular Entropy-Transport distance induced by the entropy F and the cost $\boldsymbol{c} = \ell(\mathsf{d})$. Let $(X_1, \mathsf{d}_1, \mu_1)$ and $(X_2, \mathsf{d}_2, \mu_2)$ be two metric measure spaces. Then there exist a measure $\boldsymbol{\gamma} \in \mathscr{M}(X_1 \times X_2)$ between μ_1 and μ_2 and a metric coupling $\hat{\mathsf{d}}$ between d_1 and d_2 such that*

$$\mathbf{D}^2_{\mathrm{ET}}((X_1, \mathsf{d}_1, \mu_1), (X_2, \mathsf{d}_2, \mu_2)) = \sum_{i=1}^{2} D_F(\gamma_i || \mu_i) + \int_{X_1 \times X_2} \ell\big(\hat{\mathsf{d}}(x, y)\big) \mathrm{d}\boldsymbol{\gamma}. \tag{5.3.12}$$

Proof. **Step 1**: tightness of the plans.
By Proposition 5.3.4 there exists a sequence $\boldsymbol{\gamma}_n \in \mathscr{M}(X_1 \times X_2)$ and $\hat{\mathsf{d}}_n$ couplings of $\mathsf{d}_1, \mathsf{d}_2$ such that

$$\sum_{i=1}^{2} D_F((\gamma_n)_i || \mu_i) + \int_{X_1 \times X_2} \ell\big(\hat{\mathsf{d}}_n(x, y)\big) \mathrm{d}\boldsymbol{\gamma}_n < \mathbf{D}^2_{\mathrm{ET}}((X_1, \mathsf{d}_1, \mu_1), (X_2, \mathsf{d}_2, \mu_2)) + \frac{1}{n}. \tag{5.3.13}$$

Since the entropy functionals from the fixed measures μ_1 and μ_2 are bounded, we can apply Theorems 2.4.1 and Proposition 3.1.6 in order to obtain the existence of subsequences (from now on we will not relabel them) such that $(\gamma_n)_i$ converges weakly to some $\gamma^i \in \mathscr{M}(X_i)$, $i = 1, 2$. Since $(\gamma_n)_i$ are marginals of the measure $\boldsymbol{\gamma}_n$, the tightness of $(\gamma_n)_i$ implies the tightness of $\boldsymbol{\gamma}_n$, so that the sequence $\boldsymbol{\gamma}_n \in \mathscr{M}(X_1 \times X_2)$ is converging to some $\boldsymbol{\gamma}$. Moreover, by the continuity of the operator $\pi^i_\sharp$ with respect the weak topology, the marginals of $\boldsymbol{\gamma}$ coincide with γ^i, $i = 1, 2$.

Step 2: pre-compactness of the couplings.
Regarding the sequence $\hat{\mathsf{d}}_n$, by the triangle inequality we have that

$$|\hat{\mathsf{d}}_n(x_1, y_1) - \hat{\mathsf{d}}_n(x_2, y_2)| \leq |\mathsf{d}_1(x_1, x_2) + \mathsf{d}_2(y_1, y_2)|.$$

In particular, $\hat{\mathsf{d}}_n$ is uniformly 1-Lipschitz with respect to the complete and separable metric $\mathsf{d}_1 + \mathsf{d}_2$ on $X_1 \times X_2$. We claim it is also uniformly bounded in a point. To see this, take $(\bar{x}, \bar{y}) \in \mathsf{supp}(\gamma)$; since $\ell(d) \geq d^2$ and by taking advantage of Holder inequality there exist some positive constants C, c, ϵ, δ such that for all n sufficiently large

$$\begin{aligned} C > \int_{X_1 \times X_2} \hat{\mathsf{d}}_n(x, y) \mathrm{d}\boldsymbol{\gamma}_n(x, y) &\geq \int_{B_\epsilon(\bar{x}) \times B_\epsilon(\bar{y})} [\hat{\mathsf{d}}_n(\bar{x}, \bar{y}) - 2\epsilon] \mathrm{d}\boldsymbol{\gamma}_n(x, y) \\ &\geq [\hat{\mathsf{d}}_n(\bar{x}, \bar{y}) - 2\epsilon][\boldsymbol{\gamma}(B_\epsilon(\bar{x}) \times B_\epsilon(\bar{y})) - \delta] \geq c[\hat{\mathsf{d}}_n(\bar{x}, \bar{y}) - 2\epsilon]. \end{aligned} \tag{5.3.14}$$

We can thus apply Ascoli-Arzelà's theorem to infer the existence of a limit function $\mathsf{d} : X_1 \times X_2 \to [0, \infty)$ such that d_n converges (up to subsequence) pointwise to d and the convergence is uniform on compact sets. We can extend d to $(X_1 \sqcup X_2) \times (X_1 \sqcup X_2)$ in order to get a limit coupling, that we denote in the same way.

Step 3: passing to the limit.
Next, we pass to the limit in the following expression

$$\sum_{i=1}^{2} D_F((\gamma_n)_i||\mu_i) + \int_{X_1\times X_2} \ell\big(\hat{\mathsf{d}}_n(x,y)\big)\mathrm{d}\boldsymbol{\gamma}_n.$$

By Proposition 3.1.5, the entropy is jointly lower semicontinuous and thus

$$\liminf_n D_F((\gamma_n)_i||\mu_i) \geq D_F(\gamma_i||\mu_i).$$

So, it is sufficient to prove that

$$\liminf_n \int_{X_1\times X_2} \ell\big(\hat{\mathsf{d}}_n(x,y)\big)\mathrm{d}\boldsymbol{\gamma}_n \geq \int_{X_1\times X_2} \ell\big(\hat{\mathsf{d}}(x,y)\big)\mathrm{d}\boldsymbol{\gamma}. \tag{5.3.15}$$

Using the equi-tightness of $\{\boldsymbol{\gamma}_k\}$ we can find a sequence of compact sets $K_{1,n} \subset X_1$ and $K_{2,n} \subset X_2$ such that

$$\boldsymbol{\gamma}_k\big(X_1 \times X_2 \setminus (K_{1,n} \times K_{2,n})\big) \leq \frac{1}{n}$$

for every k. We define $\ell_m(r) := \min(\ell(r), m)$, so that the sequence of functions $(x,y) \mapsto \ell_m(\mathsf{d}_n(x,y))$ converges uniformly on compact subsets of $X_1 \times X_2$, as $n \to \infty$. Possibly by taking another subsequence via a diagonal argument, we can infer that $\|\ell_m(\mathsf{d}) - \ell_m(\mathsf{d}_n)\|_{\infty;n} \to 0$ when $n \to \infty$, where we denote by $\|\cdot\|_{\infty;n}$ the supremum norm in the set $K_{1,n} \times K_{2,n}$. Let M be a positive constant such that $\gamma_n(X_1 \times X_2) \leq M$ for every n. We can bound the integral on the left hand side of (5.3.15) in the following way:

$$\begin{aligned}\int_{X_1\times X_2} \ell(\hat{\mathsf{d}}_n)\mathrm{d}\gamma_n &\geq \int_{X_1\times X_2} \ell_m(\hat{\mathsf{d}}_n)\mathrm{d}\gamma_n \geq \int_{K_n^1\times K_n^2} \ell_m(\hat{\mathsf{d}}_n)\mathrm{d}\gamma_n \\ &\geq \int_{K_n^1\times K_n^2} \ell_m(\hat{\mathsf{d}})\mathrm{d}\gamma_n - M\|\ell_m(\hat{\mathsf{d}}) - \ell_m(\hat{\mathsf{d}}_n)\|_{\infty;n} \\ &\geq \int_{X_1\times X_2} \ell_m(\hat{\mathsf{d}})\mathrm{d}\gamma_n - M\|\ell_m(\hat{\mathsf{d}}) - \ell_m(\hat{\mathsf{d}}_n)\|_{\infty;n} - m/n \end{aligned} \tag{5.3.16}$$

Now we can pass to the limit with respect to n using the weak convergence of $\{\boldsymbol{\gamma}_n\}$, and we obtain

$$\liminf \int_{X_1\times X_2} \ell(\hat{\mathsf{d}}_n)\mathrm{d}\boldsymbol{\gamma}_n \geq \int_{X_1\times X_2} \ell_m(\hat{\mathsf{d}})\mathrm{d}\boldsymbol{\gamma}$$

and then we conclude using the Beppo Levi's monotone convergence theorem with respect to m. □

Remark 5.3.7. *It is clear that the optimal coupling $\hat{\mathsf{d}}$ whose existence is proven in the previous Theorem is in general only a pseudo-metric and not a metric on $X_1 \sqcup X_2$. To see this, it is sufficient to consider two isomorphic metric measure spaces $(X_1, \mathsf{d}_1, \mu_1)$, $(X_2, \mathsf{d}_2, \mu_2)$. If we denote by $\psi : X_1 \to X_2$ the isometry between (X_1, d_1) and (X_2, d_2), the optimal coupling $\hat{\mathsf{d}}$ satisfies $\hat{\mathsf{d}}(x_1, \psi(x_1)) = 0$ for μ_1-a.e x_1.*

The next theorem is the main result of the chapter.

Theorem 5.3.8. *Let D_{ET} be a regular Entropy-Transport distance induced by the entropy F and the cost $\boldsymbol{c} = \ell(\mathsf{d})$. Then $(\mathbf{X}, \mathbf{D}_{\mathrm{ET}})$ is a complete, separable metric space.*

Proof. **Step 1**: $\mathbf{D}_{\mathrm{ET}}$ defines a metric.
It is clear that $\mathbf{D}_{\mathrm{ET}}$ is symmetric, finite valued, nonnegative and

$$\mathbf{D}_{\mathrm{ET}}\big((X_1,\mathsf{d}_1,\mu_1),(X_2,\mathsf{d}_2,\mu_2)\big)=0 \text{ if } (X_1,\mathsf{d}_1,\mu_1)=(X_2,\mathsf{d}_2,\mu_2). \tag{5.3.17}$$

We claim that $\mathbf{D}_{\mathrm{ET}}\big((X_1,\mathsf{d}_1,\mu_1),(X_2,\mathsf{d}_2,\mu_2)\big)=0$ implies that the metric measure spaces (X_1,d_1,μ_1) and (X_2,d_2,μ_2) are isomorphic. By Lemma 5.3.6 there exist a measure $\boldsymbol{\gamma}\in\mathscr{M}(X_1\times X_2)$ and a metric coupling $\hat{\mathsf{d}}$ such that

$$0=\sum_{i=1}^{2}D_F\big(\gamma_i||\mu_i\big)+\int_{X_1\times X_2}\ell\big(\hat{\mathsf{d}}(x,y)\big)\mathrm{d}\boldsymbol{\gamma}.$$

All the terms are nonnegative, so that $\gamma_i=\mu_i$ and $\hat{\mathsf{d}}(x,y)=0$ for γ-a.e (x,y) and thus (using triangle inequality and that $\hat{\mathsf{d}}$ is a pseudo-metric coupling between d_1 and d_2)

$$\hat{\mathsf{d}}(x,y)=0 \quad \text{for all } (x,y)\in\mathsf{supp}(\boldsymbol{\gamma}). \tag{5.3.18}$$

Using that d_1 and d_2 are metrics, it follows that for every $x_1\in\mathsf{supp}(\mu_1)$ there exists a unique $x_2\in\mathsf{supp}(\mu_2)$ such that $(x_1,x_2)\in\mathsf{supp}(\boldsymbol{\gamma})$. Switching the role of X_1 and X_2 in the argument above, we obtain the existence of a bijection $\psi:\mathsf{supp}(\mu_1)\to\mathsf{supp}(\mu_2)$ such that $\boldsymbol{\gamma}=(\mathrm{Id},\psi)_\sharp\mu_1$ and (in virtue of (5.3.18))

$$\hat{\mathsf{d}}(x,\psi(x))=0 \quad \text{for all } x\in\mathsf{supp}(\mu_1). \tag{5.3.19}$$

Let $x,y\in\mathsf{supp}(\mu_1)$, from (5.3.19) and the triangle inequality it follows

$$\begin{aligned}
\mathsf{d}_1(x,y)&=\hat{\mathsf{d}}(x,y)\le\hat{\mathsf{d}}(x,\psi(x))+\hat{\mathsf{d}}(\psi(x),\psi(y))+\hat{\mathsf{d}}(y,\psi(y))=\mathsf{d}_2(\psi(x),\psi(y)),\\
\mathsf{d}_2(\psi(x),\psi(y))&=\hat{\mathsf{d}}(\psi(x),\psi(y))\le\hat{\mathsf{d}}(x,\psi(x))+\hat{\mathsf{d}}(x,y)+\hat{\mathsf{d}}(y,\psi(y))=\mathsf{d}_1(x,y),
\end{aligned}$$

which implies that $\psi:\mathsf{supp}(\mu_1)\to\mathsf{supp}(\mu_2)$ is an isometry.
Hence (X_1,d_1,μ_1) and (X_2,d_2,μ_2) are isomorphic, as claimed.

Regarding the triangle inequality, let (X_i,d_i,μ_i), $i=1,2,3$, be three metric measure spaces. From the definition of $\mathbf{D}_{\mathrm{ET}}$, for every $\epsilon>0$ we find a metric coupling d_{12} between d_1 and d_2, and a metric coupling d_{23} between d_2 and d_3 such that

$$\mathbf{D}_{\mathrm{ET}}\big((X_1,\mathsf{d}_1,\mu_1),(X_2,\mathsf{d}_2,\mu_2)\big)\ge(\mathsf{D}_{\mathrm{ET}})_{\mathsf{d}_{12}}(\mu_1,\mu_2)-\epsilon, \tag{5.3.20}$$
$$\mathbf{D}_{\mathrm{ET}}\big((X_2,\mathsf{d}_2,\mu_2),(X_3,\mathsf{d}_3,\mu_3)\big)\ge(\mathsf{D}_{\mathrm{ET}})_{\mathsf{d}_{23}}(\mu_2,\mu_3)-\epsilon, \tag{5.3.21}$$

where we have denoted by $(\mathsf{D}_{\mathrm{ET}})_{\mathsf{d}}$ the Entropy-Transport distance induced by the metric d. Set $X:=X_1\sqcup X_2\sqcup X_3$ and define a metric d on X in the following way

$$\mathsf{d}(x,y):=\begin{cases}\mathsf{d}_{12}(x,y) & \text{if } x,y\in X_1\sqcup X_2,\\ \mathsf{d}_{23}(x,y) & \text{if } x,y\in X_2\sqcup X_3,\\ \inf_{z\in X_2}[\mathsf{d}_{12}(x,z)+\mathsf{d}_{23}(z,y)] & \text{if } x\in X_1 \text{ and } y\in X_3,\\ \inf_{z\in X_2}[\mathsf{d}_{23}(x,z)+\mathsf{d}_{12}(z,y)] & \text{if } x\in X_3 \text{ and } y\in X_1.\end{cases} \tag{5.3.22}$$

We notice that d coincide with d_i when restricted to X_i. By applying the point (ii) of Lemma 5.3.5 and the triangle inequality of D_{ET} we obtain

$$\begin{aligned}\mathbf{D}_{\mathrm{ET}}\big((X_1,\mathsf{d}_1,\mu_1),(X_3,\mathsf{d}_3,\mu_3)\big) &\le (\mathsf{D}_{\mathrm{ET}})_{\mathsf{d}}(\mu_1,\mu_3) \le (\mathsf{D}_{\mathrm{ET}})_{\mathsf{d}}(\mu_1,\mu_2)+(\mathsf{D}_{\mathrm{ET}})_{\mathsf{d}}(\mu_2,\mu_3)\\ &= (\mathsf{D}_{\mathrm{ET}})_{\mathsf{d}_{12}}(\mu_1,\mu_2)+(\mathsf{D}_{\mathrm{ET}})_{\mathsf{d}_{23}}(\mu_2,\mu_3)\\ &\le \mathbf{D}_{\mathrm{ET}}\big((X_1,\mathsf{d}_1,\mu_1),(X_2,\mathsf{d}_2,\mu_2)\big)+\mathbf{D}_{\mathrm{ET}}\big((X_2,\mathsf{d}_2,\mu_2),(X_3,\mathsf{d}_3,\mu_3)\big)+2\epsilon. \end{aligned} \tag{5.3.23}$$

and the conclusion follows since $\epsilon > 0$ is arbitrary.

Step 2: Separability of $\mathbf{D}_{\mathrm{ET}}$.

Thanks to (iv) of Lemma 5.3.5 it is enough to show that the set $\mathbf{X}_*$, defined in (5.3.8), is separable. To this aim, we notice that $\mathbf{X}_*$ can be written as $\bigsqcup_{n\in\mathbb{N}}\tilde{\mathcal{K}}_n$ where

$$\tilde{\mathcal{K}}_n := \{(X,\mathsf{d},\mu)\in\mathbf{X}_* : \mathsf{supp}(\mu) \text{ has } n \text{ points}\}.$$

Using the points (v), (vi) of Lemma 5.3.5, each $\tilde{\mathcal{K}}_n$ can be identified with the set of all $(D,M)=(D_{ij},M)\in\mathbb{R}_+^{n\times n}\times\mathbb{R}_+$ such that

$$D_{ij}=D_{ji},\quad D_{ij}=0 \iff i=j,\quad D_{ij}\le D_{ik}+D_{kj}, \tag{5.3.24}$$

that is separable as a subset of the Euclidean space.

Step 3: Completeness of $\mathbf{D}_{\mathrm{ET}}$.

In order to prove completeness, let $\{(X_n,\mathsf{d}_n,\mu_n)\}_{n\in\mathbb{N}}$ be a Cauchy sequence in the space $(\mathbf{X},\mathbf{D}_{\mathrm{ET}})$. By a standard result, it is enough to prove that there exists a converging subsequence. Let us consider a subsequence such that

$$\mathbf{D}^2_{\mathrm{ET}}\big((X_{n_k},\mathsf{d}_{n_k},\mu_{n_k}),(X_{n_{k+1}},\mathsf{d}_{n_{k+1}},\mu_{n_{k+1}})\big) < 2^{-(k+1)}. \tag{5.3.25}$$

Since $\ell(d)\ge d^2$, we can find a measure $\boldsymbol{\gamma}_{k+1}\in\mathscr{M}(X_{n_k}\times X_{n_{k+1}})$ between μ_{n_k} and $\mu_{n_{k+1}}$, and a metric coupling $\hat{\mathsf{d}}_{k+1}$ between $\mathsf{d}_{X_{n_k}}$ and $\mathsf{d}_{X_{n_{k+1}}}$ such that

$$\int_{X_{n_k}} F(\sigma_{n_k})\mathrm{d}\mu_{n_k} + \int_{X_{n_{k+1}}} F(\sigma_{n_{k+1}})\mathrm{d}\mu_{n_{k+1}} + \int_{X_{n_k}\times X_{n_{k+1}}} \hat{\mathsf{d}}^2_{k+1}(x,y)\mathrm{d}\boldsymbol{\gamma}_{k+1} < 2^{-k}, \tag{5.3.26}$$

where σ_{n_k} (resp. $\sigma_{n_{k+1}}$) is the Radon-Nykodim derivative of the first (resp. second) marginal of γ_{k+1} with respect to μ_{n_k} (resp. $\mu_{n_{k+1}}$). Without loss of generality (recall the construction (5.1.2)), we can also assume that $(X_{n_k}\sqcup X_{n_{k+1}},\hat{\mathsf{d}}_{k+1})$ is a complete and separable metric space. Now we want to define a sequence $\big\{(X'_k,\mathsf{d}'_k)\big\}_{k=1}^{\infty}$ of metric spaces such that $X_{n_k}\subset X'_k$ and $X'_k\subset X'_{k+1}$. We proceed in the following way: we put

$$(X'_1,\mathsf{d}'_1) := (X_{n_1},\mathsf{d}_{X_{n_1}}), \tag{5.3.27}$$

$$X'_{k+1} := X'_k\sqcup X_{n_{k+1}}/\sim, \tag{5.3.28}$$

where $x\sim y$ if $\mathsf{d}'_{k+1}(x,y)=0$ and the latter is defined as

$$\mathsf{d}'_{k+1}(x,y) := \begin{cases} \mathsf{d}'_k(x,y) & \textit{if } \; x,y\in X'_k,\\ \hat{\mathsf{d}}_{k+1}(x,y) & \textit{if } \; x,y\in X_{n_k}\sqcup X_{n_{k+1}},\\ \inf_{z\in X_{n_k}} \mathsf{d}'_k(x,z)+\hat{\mathsf{d}}_{k+1}(z,y) & \textit{if } \; x\in X'_k,\ y\in X_{n_{k+1}},\\ \inf_{z\in X_{n_k}} \mathsf{d}'_k(y,z)+\hat{\mathsf{d}}_{k+1}(z,x) & \textit{if } \; y\in X'_k,\ x\in X_{n_{k+1}}. \end{cases} \tag{5.3.29}$$

From the definition of d'_k, it is clear that we can endow the space $X' := \bigcup_{k=1}^{\infty} X'_k$ with a limit metric d'. Now we consider the completion (X, d) of (X', d') and we notice that $(X_{n_k}, \mathsf{d}_{X_{n_k}})$ is isometrically embedded in this space for every k. Using the embedding, we can also define a measure $\bar{\mu}_{n_k}$ as the push-forward of the measure μ_{n_k}. Combining (5.3.26), it follows

$$\begin{aligned}&(\mathsf{D}_{\mathrm{ET}})^2_{\mathsf{d}}(\bar{\mu}_{n_k}, \bar{\mu}_{n_{k+1}}) \leq \\ &\int_{X_{n_k}} F(\sigma_{n_k}) \mathrm{d}\mu_{n_k} + \int_{X_{n_{k+1}}} F(\sigma_{n_{k+1}}) \mathrm{d}\mu_{n_{k+1}} + \int_{X_{n_k} \times X_{n_{k+1}}} \hat{\mathsf{d}}^2_{k+1}(x, y) \mathrm{d}\boldsymbol{\gamma}_{k+1} < 2^{-k},\end{aligned} \tag{5.3.30}$$

where $(\mathsf{D}_{\mathrm{ET}})_{\mathsf{d}}$ is the power-like-Wasserstein distance computed in the space (X, d). In particular, (5.3.30) implies that $(\bar{\mu}_{n_k})_{k \in \mathbb{N}}$ is a Cauchy sequence in $(\mathscr{M}(X), (\mathsf{D}_{\mathrm{ET}})_{\mathsf{d}})$. Since the latter is complete, $\bar{\mu}_{n_k}$ weakly converges to some measure $\mu \in \mathscr{M}(X)$.

Using again that $(X_{n_k}, \mathsf{d}_{X_{n_k}})$ is isometrically embedded in (X, d) and the point (ii) of the Lemma 5.3.5, we can conlude that

$$\mathbf{D}^2_{\mathrm{ET}}\big((X_{n_k}, \mathsf{d}_{n_k}, \mu_{n_k}), (X, \mathsf{d}, \mu)\big) \leq (\mathsf{D}_{\mathrm{ET}})^2_{\mathsf{d}}(\bar{\mu}_{n_k}, \mu) \to 0. \tag{5.3.31}$$

□

Part II

Sharp Cheeger-Buser type inequalities in $\mathsf{RCD}(K,\infty)$ spaces

Chapter 6

Sharp Cheeger-Buser type inequalities in $\mathsf{RCD}(K,\infty)$ spaces

In this chapter we sharpen and generalise bounds involving Cheeger isoperimetric constant and the first eigenvalue of the Laplacian in the class of $\mathsf{RCD}(K,\infty)$ spaces. The proof of our main result (Theorem 6.1.1) is based on the semigroup approach implemented by Ledoux in [Led94; Led04], but it improves upon mainly by taking advantage of the sharp inequality (6.2.5), which is interesting by its own.

6.1 The main result

Let $(X,\mathsf{d},\mathfrak{m})$ be a metric measure space. We suppose that there exist $x_0 \in X$, $M > 0$ and $c \geq 0$ such that

$$\mathfrak{m}(B_r(x_0)) \leq M\exp(cr^2) \quad \text{for every } r \geq 0.$$

Possibly enlarging the σ-algebra $\mathcal{B}(X)$ and extending $\mathfrak{m}$, we also assume that $\mathcal{B}(X)$ is $\mathfrak{m}$-complete.

We define the first non-trivial eigenvalue as follows:

- If $\mathfrak{m}(X) < \infty$, the non-zero constant functions are in $L^2(X,\mathfrak{m})$ and are eigenfunctions of the Laplacian with eigenvalue 0. In this case, the first non-trivial eigenvalue is given by

$$\lambda_1 = \inf\left\{\frac{\int_X |\nabla f|^2 \mathrm{d}\mathfrak{m}}{\int_X |f|^2 \mathrm{d}\mathfrak{m}} :\ 0 \not\equiv f \in \mathsf{Lip}_{bs}(X),\ \int_X f\,\mathrm{d}\mathfrak{m} = 0\right\}. \tag{6.1.1}$$

- When $\mathfrak{m}(X) = \infty$, 0 may not be an eigenvalue of the Laplacian and the first eigenvalue is characterized by

$$\lambda_0 = \inf\left\{\frac{\int_X |\nabla f|^2 \mathrm{d}\mathfrak{m}}{\int_X |f|^2 \mathrm{d}\mathfrak{m}} :\ 0 \not\equiv f \in \mathsf{Lip}_{bs}(X)\right\}. \tag{6.1.2}$$

Let $A \subset X$ be a Borel set with $\mathfrak{m}(A) < \infty$, the *perimeter* $\mathrm{Per}(A)$ is defined as follows:

$$\mathrm{Per}(A) := \inf\left\{\liminf_{n\to\infty} \int_X |\nabla f_n| \mathrm{d}\mathfrak{m} : f_n \in \mathsf{Lip}_{bs}(X), f_n \to \chi_A \text{ in } L^1(X,\mathfrak{m})\right\}.$$

The *Cheeger constant* of the metric measure space $(X, \mathsf{d}, \mathfrak{m})$ is defined by

$$h(X) := \begin{cases} \inf\left\{\frac{\mathrm{Per}(A)}{\mathfrak{m}(A)} : A \subset X \text{ Borel subset with } \mathfrak{m}(A) \le \mathfrak{m}(X)/2\right\} & \text{if } \mathfrak{m}(X) < \infty \\ \inf\left\{\frac{\mathrm{Per}(A)}{\mathfrak{m}(A)} : A \subset X \text{ Borel subset with } \mathfrak{m}(A) < \infty\right\} & \text{if } \mathfrak{m}(X) = \infty. \end{cases} \tag{6.1.3}$$

We set

$$J_K(t) = \begin{cases} \sqrt{\frac{2}{\pi K}}\arctan\left(\sqrt{e^{2Kt}-1}\right) & \text{if } K > 0, \\ \frac{2}{\sqrt{\pi}}\sqrt{t} & \text{if } K = 0, \qquad \forall t > 0 \\ \sqrt{-\frac{2}{\pi K}}\operatorname{arctanh}\left(\sqrt{1-e^{2Kt}}\right) & \text{if } K < 0. \end{cases} \tag{6.1.4}$$

The aim of the chapter is to prove the following theorem.

Theorem 6.1.1 (Sharp implicit Buser-type inequality for $\mathsf{RCD}(K,\infty)$ spaces)**.** *Let $(X, \mathsf{d}, \mathfrak{m})$ be an $\mathsf{RCD}(K,\infty)$ space, for some $K \in \mathbb{R}$.*

- *In case $\mathfrak{m}(X) = 1$, then*
$$h(X) \ge \sup_{t>0} \frac{1-e^{-\lambda_1 t}}{J_K(t)}. \tag{6.1.5}$$
The inequality is sharp for $K > 0$, as equality is achieved for the Gaussian space $(\mathbb{R}^d, |\cdot|, (2\pi)^{-d/2}e^{-|x|^2/2}\mathrm{d}\mathcal{L}^d(x))$, $1 \le d \in \mathbb{N}$.

- *In case $\mathfrak{m}(X) = \infty$, then*
$$h(X) \ge 2\sup_{t>0} \frac{1-e^{-\lambda_0 t}}{J_K(t)}. \tag{6.1.6}$$

Using the expression (6.1.4) of J_K, in the next corollary we obtain more explicit bounds.

Corollary 6.1.2 (Explicit Buser inequality for $\mathsf{RCD}(K,\infty)$ spaces)**.** *Let $(X, \mathsf{d}, \mathfrak{m})$ be an $\mathsf{RCD}(K,\infty)$ space, for some $K \in \mathbb{R}$.*

- *Case $K > 0$. If $\frac{K}{\lambda_1} \ge c > 0$, then*
$$\lambda_1 \le \frac{\pi}{2c}h(X)^2. \tag{6.1.7}$$
The estimate is sharp, as equality is attained on the Gaussian space $(\mathbb{R}^d, |\cdot|, (2\pi)^{-d/2}e^{-|x|^2/2}\mathrm{d}\mathcal{L}^d(x))$, $1 \le d \in \mathbb{N}$, for which $K = 1, \lambda_1 = 1, h(X) = (2/\pi)^{1/2}$.

- *Case $K = 0$, $\mathfrak{m}(X) = 1$. It holds*
$$\lambda_1 \le \frac{4}{\pi}h(X)^2 \inf_{T>0}\frac{T}{(1-e^{-T})^2} < \pi h(X)^2. \tag{6.1.8}$$
In case $\mathfrak{m}(X) = \infty$, the estimate (6.1.8) holds replacing λ_1 with λ_0 and $h(X)$ with $h(X)/2$.

- *Case $K<0$, $\mathfrak{m}(X)=1$. It holds*

$$\lambda_1 \le \max\left\{\sqrt{-K}\frac{\sqrt{2}\log\left(e+\sqrt{e^2-1}\right)}{\sqrt{\pi}(1-\frac{1}{e})}h(X), \frac{2\Big(\log\left(e+\sqrt{e^2-1}\right)\Big)^2}{\pi\Big(1-\frac{1}{e}\Big)^2}h(X)^2\right\}$$
$$< \max\left\{\frac{21}{10}\sqrt{-K}h(X), \frac{22}{5}h(X)^2\right\}. \tag{6.1.9}$$

In case $\mathfrak{m}(X)=\infty$, the estimate (6.1.9) holds replacing λ_1 with λ_0 and $h(X)$ with $h(X)/2$.

Remark 6.1.3. *Only if the m.m.s $(X,\mathsf{d},\mathfrak{m})$ admits a compact embedding of $\mathbb{V}$ in $L^2(X,\mathfrak{m})$, we have the right to speak of the* first positive *eigenvalue of $-\Delta$. Indeed, by a standard result of spectral theory (see [DS63] for a general reference and [GMS15] for some results in the* RCD *setting), the above-mentioned assumption is equivalent to ask that $-\Delta$ has discrete spectrum consisting of an increasing sequence of non-negative eigenvalues $\{\lambda_n\}_{n=0}^{\infty}$ such that $\lim_{n\to\infty}\lambda_n \to +\infty$. Anyway, we point out that the proof goes through also in the presence of a non-discrete spectrum.*

6.2 The Proof

We denote by $I:[0,1]\to[0,\frac{1}{\sqrt{2\pi}}]$ the Gaussian isoperimetric function defined by $I:=\varphi\circ\Phi^{-1}$ where

$$\Phi(x) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-u^2/2}\,\mathrm{d}u, \quad x\in\mathbb{R},$$

and $\varphi=\Phi'$. The function I is concave, continuous, $I(0)=I(1):=0$ and $0\le I(x)\le I(\frac{1}{2})=\frac{1}{\sqrt{2\pi}}$, for all $x\in[0,1]$. Moreover, $I\in\mathcal{C}^\infty((0,1))$, it satisfies the identity

$$I(x)I''(x)=-1, \quad \text{for every } x\in(0,1). \tag{6.2.1}$$

and (see [BL96])

$$\lim_{x\to 0}\frac{I(x)}{x\sqrt{2\log\frac{1}{x}}}=1. \tag{6.2.2}$$

Given $K\in\mathbb{R}$, we define the function $j_K:(0,\infty)\to(0,\infty)$ as

$$j_K(t) := \begin{cases} \frac{K}{e^{2Kt}-1} & \text{if } K\neq 0,\\ \frac{1}{2t} & \text{if } K=0. \end{cases} \tag{6.2.3}$$

Notice that j_K is increasing as a function of K.
The next proposition was proved in the smooth setting by Bakry, Gentil and Ledoux (see [BL96], [BGL15] and [BGL14, Proposition 8.6.1]).

Proposition 6.2.1 (Bakry-Gentil-Ledoux Inequality in $\mathsf{RCD}(K,\infty)$ spaces)**.** *Consider an $\mathsf{RCD}(K,\infty)$ space $(X,\mathsf{d},\mathfrak{m})$, for some $K\in\mathbb{R}$. Then for every function $f\in L^2(X,\mathfrak{m})$, $f:X\to[0,1]$ it holds*

$$|D(H_tf)|_w^2 \le j_K(t)\Big(\big[I(H_tf)\big]^2-\big[H_t(I(f))\big]^2\Big), \quad \mathfrak{m}\text{-a.e., for every } t>0. \tag{6.2.4}$$

In particular, for every $f \in L^2 \cap L^\infty(X,\mathfrak{m})$, *it holds*

$$\| |D(H_t f)|_w \|_\infty \leq \sqrt{\frac{2}{\pi}} \sqrt{j_K(t)}\, \|f\|_\infty\,, \quad \mathfrak{m}\text{-a.e., for every } t>0. \tag{6.2.5}$$

Proof. Given $\varepsilon > 0$, $\eta > 2\varepsilon$ and $\delta > 0$ sufficiently small, consider $f \in L^2(X,\mathfrak{m})$ with values in $[0, 1-\eta]$. We define

$$\phi_\varepsilon(x) := I(x+\varepsilon) - I(\varepsilon), \tag{6.2.6}$$

$$\Psi_\varepsilon(s) := \Big[H_s(\phi_\varepsilon(H_{t-s}f))\Big]^2, \quad \text{for every } s \in (0,t). \tag{6.2.7}$$

We notice that $\phi_\varepsilon(0) = 0$ and $\phi_\varepsilon(x) \geq 0$ for every $x \in [0, 1-\eta]$. Moreover, using the property (2.6.11), ϕ_ε is Lipschitz in the range of $H_{t-s}f$. Since $t \mapsto H_t f$ is a locally Lipschitz map with values in $L^p(X,\mathfrak{m})$ for $1 < p < \infty$ ([Ste70, Theorem 1, Section III]), we have that Ψ_ε is a locally Lipschitz map with values in $L^1(X,\mathfrak{m})$. Let $\psi \in L^1 \cap L^\infty(X,\mathfrak{m})$ be a non-negative function. By the chain rule for locally Lipschitz maps, the fundamental theorem of calculus for the Bochner integral and the properties of the semigroup H_t we have that for any $\varepsilon > 0$ and $0 < \delta < t$ it holds

$$\begin{aligned}
&\int_X \left(\Big[H_\delta(\phi_\varepsilon(H_{t-\delta}f))\Big]^2 - \Big[H_{t-\delta}(\phi_\varepsilon(H_\delta f))\Big]^2\right)\psi\,\mathrm{d}\mathfrak{m} \\
&\qquad = \int_\delta^{t-\delta} \left(-\frac{d}{ds}\int_X \Big[H_s(\phi_\varepsilon(H_{t-s}f))\Big]^2 \psi\,\mathrm{d}\mathfrak{m}\right)\mathrm{d}s \\
&= -2\int_\delta^{t-\delta}\left(\int_X H_s\big(\phi_\varepsilon(H_{t-s}f)\big)H_s\big(\Delta\phi_\varepsilon(H_{t-s}f) - \phi_\varepsilon'(H_{t-s}f)\Delta H_{t-s}f\big)\psi\,\mathrm{d}\mathfrak{m}\right)\mathrm{d}s \\
&\quad = 2\int_\delta^{t-\delta}\left(\int_X H_s\big(\phi_\varepsilon(H_{t-s}f)\big)H_s\big(-\phi_\varepsilon''(H_{t-s}f)|DH_{t-s}f|_w^2\big)\psi\,\mathrm{d}\mathfrak{m}\right)\mathrm{d}s.
\end{aligned} \tag{6.2.8}$$

Applying the Cauchy-Schwarz inequality

$$H_s(X)H_s(Y) \geq \big[H_s\big(\sqrt{XY}\big)\big]^2,$$

and the identity $I(x)I''(x) = -1$, for all $x \in (0,1)$, we get that the right-hand side of (6.2.8) is bounded below by

$$2\int_\delta^{t-\delta}\left(\int_X \left[H_s\left(\sqrt{\left(1 - \frac{I(\varepsilon)}{I(H_{t-s}f+\varepsilon)}\right)|DH_{t-s}f|_w^2}\right)\right]^2 \psi\,\mathrm{d}\mathfrak{m}\right)\mathrm{d}s. \tag{6.2.9}$$

Noticing that

$$\int_X \left[H_s\left(\sqrt{\left(1 - \frac{I(\varepsilon)}{I(H_{t-s}f+\varepsilon)}\right)|DH_{t-s}f|_w^2}\right)\right]^2 \psi\,\mathrm{d}\mathfrak{m} \leq \int_X \Big[H_s\big(|DH_{t-s}f|_w\big)\Big]^2\psi\,\mathrm{d}\mathfrak{m}$$

and that, for any fixed $\delta > 0$,

$$\int_\delta^{t-\delta}\left(\int_X \Big[H_s\big(|DH_{t-s}f|_w\big)\Big]^2\psi\,\mathrm{d}\mathfrak{m}\right)\mathrm{d}s < \infty$$

thanks to the bound (2.6.9), we can pass to the limit as $\varepsilon \to 0$ in (6.2.9) using the Dominated Convergence Theorem.

Since I is continuous, $I(0)=0$ and $I(x)>0$ for every $x\in(0,1)$, using the locality property (2.6.4), the Dominated Convergence Theorem yields

$$\begin{aligned}&\int_X\left(\Big[H_\delta(I(H_{t-\delta}f))\Big]^2-\Big[H_{t-\delta}(I(H_\delta f))\Big]^2\right)\psi\,\mathrm{d}\mathfrak{m}\\&\geq 2\int_\delta^{t-\delta}\left(\int_X\Big[H_s\big(|DH_{t-s}f|_w\big)\Big]^2\psi\,\mathrm{d}\mathfrak{m}\right)\mathrm{d}s,\end{aligned}\tag{6.2.10}$$

for every $\delta\in(0,t)$. Now, we can bound the right-hand side of (6.2.10) using the inequality (2.6.13) in order to obtain

$$2\int_\delta^{t-\delta}\left(\int_X\Big[H_s\big(|DH_{t-s}f|_w\big)\Big]^2\psi\,\mathrm{d}\mathfrak{m}\right)\mathrm{d}s\geq 2\int_X\left(\int_\delta^{t-\delta}e^{2Ks}ds\right)|DH_tf|_w^2\,\psi\,\mathrm{d}\mathfrak{m}.\tag{6.2.11}$$

From (6.2.2) it follows that for every $0<a<1$ there exists $C=C(a)>0$ and $\bar{x}=\bar{x}(a)\in(0,1)$ such that $I(x)\leq Cx^a$ for all $x\in(0,\bar{x})$. In particular, if $g\in L^2(X,\mathfrak{m})$, $g:X\to[0,1-\eta]$, then $I(g)\in L^p(X,\mathfrak{m})$ for every $p>2$. We now apply this argument for $p=4$, so that we can take advantage of the continuity of I and the continuity of the semigroup and pass to the limit as $\delta\downarrow 0$. We obtain

$$\int_X\left(\Big[I(H_tf)\Big]^2-\Big[H_t(I(f))\Big]^2\right)\psi\,\mathrm{d}\mathfrak{m}\geq\frac{1}{j_K(t)}\int_X|DH_tf|_w^2\,\psi\,\mathrm{d}\mathfrak{m},\tag{6.2.12}$$

for every $\eta>0$ sufficiently small, every $f\in L^2(X,\mathfrak{m})$, $f:X\to[0,1-\eta]$.
Now, for $f\in L^2(X,\mathfrak{m})$, $f:X\to[0,1]$, consider the truncation $f_\eta:=\min\{f,1-\eta\}$. Applying (6.2.12) to f_η, we have

$$\int_X\left(\Big[I(H_tf_\eta)\Big]^2-\Big[H_t(I(f_\eta))\Big]^2\right)\psi\,\mathrm{d}\mathfrak{m}\geq\frac{1}{j_K(t)}\int_X|DH_tf_\eta|_w^2\,\psi\,\mathrm{d}\mathfrak{m}.\tag{6.2.13}$$

From $f_\eta\to f$ in $L^2\cap L^\infty(X,\mathfrak{m})$ as $\eta\downarrow 0$, we get that $H_tf_\eta\to H_tf$ in $\mathbb{V}$ for every $t>0$; we can then pass to the limit as $\eta\downarrow 0$ in (6.2.13) and obtain

$$\int_X\left(\Big[I(H_tf)\Big]^2-\Big[H_t(I(f))\Big]^2\right)\psi\,\mathrm{d}\mathfrak{m}\geq\frac{1}{j_K(t)}\int_X|DH_tf|_w^2\,\psi\,\mathrm{d}\mathfrak{m}.$$

Since $\psi\in L^1\cap L^\infty(X,\mathfrak{m})$, $\psi\geq 0$ is arbitrary, the desired estimate (6.2.4) follows. Recalling that $0\leq I\leq\frac{1}{\sqrt{2\pi}}$, the inequality (6.2.4) yields

$$|D(H_tf)|_w\leq\sqrt{\frac{j_K(t)}{2\pi}},\quad\mathfrak{m}\text{-a.e.},\quad\text{for every } t>0,\tag{6.2.14}$$

for any $f\in L^2(X,\mathfrak{m})$, $f:X\to[0,1]$. For any $f\in L^2\cap L^\infty(X,\mathfrak{m})$, write $f=f^+-f^-$ with $f^+=\max\{f,0\}$, $f^-=\max\{-f,0\}$. Applying (6.2.14) to $f^+/\|f\|_\infty, f^-/\|f\|_\infty$ and summing up we obtain

$$\||DH_tf|_w\|_\infty\leq\big\||DH_tf^+|_w\big\|_\infty+\big\||DH_tf^-|_w\big\|_\infty\leq\sqrt{\frac{2}{\pi}}\sqrt{j_K(t)}\,\|f\|_\infty,\quad\forall t>0.$$

□

We next recall the definition of the first non-trivial eigenvalue of the laplacian $-\Delta$. First of all, if $\mathfrak{m}(X)<\infty$, the non-zero constant functions are in $L^2(X,\mathfrak{m})$ and are eigenfunctions of $-\Delta$ with eigenvalue 0. In this case, the first non-trivial eigenvalue is given by λ_1

$$\lambda_1 = \inf\left\{\frac{\int_X |Df|_w^2 \mathrm{d}\mathfrak{m}}{\int_X |f|^2 \mathrm{d}\mathfrak{m}} :\ 0 \not\equiv f \in \mathbb{V}, \int_X f \mathrm{d}\mathfrak{m} = 0\right\}. \tag{6.2.15}$$

When $\mathfrak{m}(X) = \infty$, 0 may not be an eigenvalue of $-\Delta$ and the first eigenvalue is characterized by

$$\lambda_0 = \inf\left\{\frac{\int_X |Df|_w^2 \mathrm{d}\mathfrak{m}}{\int_X |f|^2 \mathrm{d}\mathfrak{m}} :\ \ 0 \not\equiv f \in \mathbb{V}\right\}. \tag{6.2.16}$$

Observe that, by the very definition of Cheeger energy (2.6.3), the definition (6.1.1) of λ_1 (resp. (6.1.2) of λ_0) given at the beginning of the chapter in terms of slope of Lipschitz functions, is equivalent to (6.2.15) (resp. (6.2.16)).

It is also convenient to set

$$J_K(t) := \sqrt{\frac{2}{\pi}} \int_0^t \sqrt{j_K(s)}\, \mathrm{d}s, \tag{6.2.17}$$

where j_K was defined in (6.2.3).

Proof of Theorem 6.1.1. **Step 1**: Proof of (6.1.5), the case $\mathfrak{m}(X) = 1$.
First of all, we claim that for any $f \in L^2(X,\mathfrak{m})$ with zero mean it holds

$$\|H_t f\|_2 \le e^{-\lambda_1 t} \|f\|_2 . \tag{6.2.18}$$

To prove (6.2.18) let $0 \not\equiv f \in L^2(X,\mathfrak{m})$ such that $0 = \int_X f \mathrm{d}\mathfrak{m} = \int_X H_t f \mathrm{d}\mathfrak{m}$. Then

$$2\lambda_1 \int_X |H_t f|^2 \mathrm{d}\mathfrak{m} \le 2\int_X |D(H_t f)|_w^2 \mathrm{d}\mathfrak{m} = -2\int_X H_t f \Delta(H_t f) \mathrm{d}\mathfrak{m} = -\frac{d}{dt}\int_X |H_t f|^2 \mathrm{d}\mathfrak{m}, \tag{6.2.19}$$

and the Gronwall's inequality yields (6.2.18).
Next we claim that, by duality, the bound (6.2.5) implies

$$\|f - H_t f\|_1 \le J_K(t)\, \||Df|_w\|_1 , \quad \text{for all } f \in \mathsf{Lip}_{bs}(X), \tag{6.2.20}$$

where $J_K(t)$ was defined in (6.2.17).
To prove (6.2.20) we take a function g, $\|g\|_\infty \le 1$, and observe that

$$\begin{aligned}\int_X g(f - H_t f)\mathrm{d}\mathfrak{m} = -\int_0^t \Big(\int_X g\Delta H_s f \mathrm{d}\mathfrak{m}\Big)\mathrm{d}s = \int_0^t \Big(\int_X DH_s g \cdot Df \mathrm{d}\mathfrak{m}\Big)\mathrm{d}s \\ \le \||Df|_w\|_1 \int_0^t \||D(H_s g)|_w\|_\infty \,\mathrm{d}s.\end{aligned}$$

Since g is arbitrary, the claimed (6.2.20) follows from the last estimate combined with (6.2.5).
We now combine the above claims in order to conclude the proof. Let $A \subset X$ be a Borel subset and let $f_n \in \mathsf{Lip}_{bs}(X)$, $f_n \to \chi_A$ in $L^1(X,\mathfrak{m})$, be a recovery sequence for the perimeter of the set A, i.e.:

$$\mathrm{Per}(A) = \lim_{n\to\infty} \int_X |\nabla f_n| \,\mathrm{d}\mathfrak{m} \ge \limsup_{n\to\infty} \int_X |Df_n|_w \,\mathrm{d}\mathfrak{m}.$$

Inequality (6.2.20) passes to the limit since H_t is continuous in $L^1(X,\mathfrak{m})$ [AGS08, Theorem 4.16] and we can write

$$\begin{aligned} J_K(t)\mathrm{Per}(A) &\geq \|\chi_A - H_t(\chi_A)\|_1 = \int_A [1-H_t(\chi_A)]\mathrm{d}\mathfrak{m} + \int_{A^c} H_t(\chi_A)\mathrm{d}\mathfrak{m} \\ &= 2\Big(\mathfrak{m}(A) - \int_A H_t(\chi_A)\mathrm{d}\mathfrak{m}\Big) = 2\Big(\mathfrak{m}(A) - \int_X \chi_A H_{t/2}(H_{t/2}(\chi_A))\mathrm{d}\mathfrak{m}\Big) \\ &= 2\Big(\mathfrak{m}(A) - \int_X H_{t/2}(\chi_A)H_{t/2}(\chi_A)\mathrm{d}\mathfrak{m}\Big) = 2\big(\mathfrak{m}(A) - \big\|H_{t/2}(\chi_A)\big\|_2^2\big), \end{aligned} \quad (6.2.21)$$

where we used properties (2.6.10), (2.6.11), together with the semigroup property and the self-adjointness of the semigroup. We observe that $\int_X H_{t/2}(\chi_A - \mathfrak{m}(A))\, d\mathfrak{m} = 0$ thanks to (2.6.10) and the fact that $H_t\mathbf{1} = \mathbf{1}$ when $\mathfrak{m}(X) = 1$, where we have denoted with $\mathbf{1}$ the constant function equals to 1. We can thus apply (6.2.18) in order to bound $\big\|H_{t/2}(\chi_A)\big\|_2^2$ in the following way

$$\big\|H_{t/2}(\chi_A)\big\|_2^2 = \mathfrak{m}(A)^2 + \big\|H_{t/2}(\chi_A - \mathfrak{m}(A))\big\|_2^2 \leq \mathfrak{m}(A)^2 + e^{-\lambda_1 t}\,\|\chi_A - \mathfrak{m}(A)\|_2^2\,. \quad (6.2.22)$$

A direct computation gives $\|\chi_A - \mathfrak{m}(A)\|_2^2 = \mathfrak{m}(A)(1-\mathfrak{m}(A))$, so that the combination of (6.2.21) and (6.2.22) yields

$$J_K(t)\mathrm{Per}(A) \geq 2\mathfrak{m}(A)(1-\mathfrak{m}(A))(1-e^{-\lambda_1 t}), \qquad \text{for every } t>0. \quad (6.2.23)$$

Recalling that in the definition of the Cheeger constant $h(X)$ one considers only Borel subsets $A \subset X$ with $\mathfrak{m}(A) \leq 1/2$, the last inequality (6.2.23) gives (6.1.5).

Step 2: Proof of (6.1.6), the case $\mathfrak{m}(X) = \infty$.
Arguing as in (6.2.19) using Gronwall Lemma, for any $f \in L^2(X,\mathfrak{m})$ it holds

$$\|H_t f\|_2 \leq e^{-\lambda_0 t}\,\|f\|_2\,. \quad (6.2.24)$$

Note that in order to establish (6.2.21), the finiteness of $\mathfrak{m}(X)$ played no role. Now we can directly use (6.2.24) to bound the right-hand side of the equation (6.2.21) in order to achieve

$$\frac{\mathrm{Per}(A)}{\mathfrak{m}(A)} \geq 2\sup_{t>0}\Big\{\frac{1-e^{-\lambda_0 t}}{J_K(t)}\Big\},$$

for any Borel subset $A \subset X$ with $\mathfrak{m}(A) < \infty$. The estimate (6.1.6) follows. □

Remark 6.2.2. *It was proved in [GMS15] that an* $\mathsf{RCD}(K,\infty)$ *space, with* $K > 0$ *(or with finite diameter) has discrete spectrum (as the Sobolev imbedding* $\mathbb{V}$ *into* L^2 *is compact). Even in case of infinite measure the embedding of* $\mathbb{V}$ *in* L^2 *may be compact. An example is given by* $\mathbb{R}$ *with the Euclidean distance* $\mathsf{d}(x,y) = |x-y|$ *and the measure* $\mathfrak{m} := \frac{1}{\sqrt{2\pi}}e^{x^2/2}\mathrm{d}\mathcal{L}^1$*. It is a* $\mathsf{RCD}(-1,\infty)$ *space and a result of Wang [Wan02] ensures that the spectrum is discrete.*

6.2.1 From the implicit to explicit bounds. Proof of Corollary 6.1.2

In this section we show how to derive explicit bounds for λ_1 (resp. λ_0) in term of the Cheeger constant $h(X)$, starting from (6.1.5) (resp. (6.1.6)). We also show that (6.1.5) is sharp, since equality is achieved on the Gaussian space.
First of all, the expression of the function J_K defined in (6.2.17) can be explicitly computed as:

$$J_K(t) = \begin{cases} \sqrt{\frac{2}{\pi K}}\arctan\left(\sqrt{e^{2Kt}-1}\right) & \text{if } K>0, \\ \frac{2}{\sqrt{\pi}}\sqrt{t} & \text{if } K=0, \\ \sqrt{-\frac{2}{\pi K}}\operatorname{arctanh}\left(\sqrt{1-e^{2Kt}}\right) & \text{if } K<0. \end{cases} \qquad \forall t>0 \tag{6.2.25}$$

Case $K = 0$

When $K = 0$, the estimate (6.1.5) combined with (6.1.4) gives

$$h(X) \geq \frac{\sqrt{\pi}}{2}\sup_{t>0}\frac{1-e^{-\lambda_1 t}}{\sqrt{t}} = \frac{\sqrt{\pi\lambda_1}}{2}\sup_{T>0}\frac{1-e^{-T}}{\sqrt{T}}, \tag{6.2.26}$$

where we set $T = \lambda_1 t$ in the last identity.

Let $W_{-1} : [-1/e, 0) \to (-\infty, -1]$ be the lower branch of the Lambert function, i.e. the inverse of the function $x \mapsto xe^x$ in the interval $(-\infty, -1]$. An easy computation yields

$$M := \sup_{T>0}\frac{1-e^{-T}}{\sqrt{T}} = \frac{\sqrt{-4W_{-1}\Big(-\frac{1}{2\sqrt{e}}\Big)-2}}{2W_{-1}\Big(-\frac{1}{2\sqrt{e}}\Big)}, \quad \text{achieved at } T = -W_{-1}\Big(-\frac{1}{2\sqrt{e}}\Big)-\frac{1}{2}. \tag{6.2.27}$$

A good lower estimate of M is given by $2/\pi$. Using this bound, we obtain

$$\lambda_1 < \pi h^2(X).$$

Case $K > 0$

We start with the following

Lemma 6.2.3. *Let $f_1 : (0,\infty) \to (0,\infty)$ be defined as*

$$f_1(x) := \frac{\sqrt{x}}{\arctan\left(\sqrt{e^{Tx}-1}\right)}, \tag{6.2.28}$$

where $T > 0$ is a fixed number. Then f_1 is an increasing function and $f_1(x) \geq \frac{1}{\sqrt{T}}$.

Proof. The function f_1 is differentiable and the derivative of f_1 is non-negative if and only if

$$\sqrt{e^{Tx}-1}\arctan\left(\sqrt{e^{Tx}-1}\right) - Tx \geq 0, \quad x > 0.$$

We put $y := \sqrt{e^{Tx}-1}$ so that we have to prove

$$y\arctan(y) - \log(y^2+1) \geq 0, \quad y > 0. \tag{6.2.29}$$

Called $g_1(y)$ the function $g_1(y) := y\arctan(y) - \log(y^2+1)$, we have that $g_1(0) = 0$ and

$$g_1'(y) = \arctan(y) - \frac{y}{1+y^2} \geq 0,$$

so that the inequality (6.2.29) is proved and f_1 is increasing for any $T > 0$. The proof is finished since

$$\lim_{x\downarrow 0} f_1(x) = \frac{1}{\sqrt{T}}.$$

□

Rewriting the estimate (6.1.5) using (6.1.4) in case $K>0$, we obtain

$$\sqrt{\frac{2}{\pi}}h(X) \geq \sqrt{K}\sup_{t>0}\frac{1-e^{-\lambda_1 t}}{\arctan\left(\sqrt{e^{2Kt}-1}\right)}$$
$$= \sqrt{\lambda_1}\sup_{T>0}\frac{\sqrt{\frac{K}{\lambda_1}}}{\arctan\left(\sqrt{e^{2\frac{K}{\lambda_1}T}-1}\right)}\left(1-e^{-T}\right). \tag{6.2.30}$$

Thanks to the Lemma 6.2.3 it is clear that we can always obtain the same lower bound of the case $K=0$ (as expected), but this can be improved as soon as we have a positive lower bound of the quotient K/λ_1. Indeed, let us suppose $K/\lambda_1 \geq c > 0$. Then, observing that

$$\sup_{T>0}\frac{1-e^{-T}}{\arctan(\sqrt{e^{2cT}-1})} \geq \lim_{T\to+\infty}\frac{1-e^{-T}}{\arctan(\sqrt{e^{2cT}-1})} = \frac{2}{\pi},$$

from (6.2.30), we obtain

$$\sqrt{\frac{2}{c\pi}}h(X) \geq \sqrt{\lambda_1}\sup_{T>0}\frac{1-e^{-T}}{\arctan(\sqrt{e^{2cT}-1})} \geq \frac{2}{\pi}\sqrt{\lambda_1}. \tag{6.2.31}$$

When $X=\mathbb{R}^d$ endowed with the Euclidean distance $\mathsf{d}(x,y)=|x-y|$ and the Gaussian measure $(2\pi)^{-d/2}e^{-|x|^2/2}d\mathcal{L}^d$, $1\leq d\in\mathbb{N}$, we have that $h(X)=\sqrt{\frac{2}{\pi}}$, $K=1$ and $\lambda_1=1$ (see [BGL14, Section 4.1]). Thus, we can take $c=1$ and the equality in (6.2.31) is achieved, making sharp the lower bound.

Case $K<0$

We begin by noticing that

$$J_K(t)=\sqrt{-\frac{2}{\pi K}}\operatorname{arctanh}\left(\sqrt{1-e^{2Kt}}\right)=\sqrt{-\frac{2}{\pi K}}\log\left(e^{-Kt}+\sqrt{e^{-2Kt}-1}\right). \tag{6.2.32}$$

The following lemma holds:

Lemma 6.2.4. *Let $f_2:(0,\infty)\to(0,\infty)$ be defined as*

$$f_2(x):=\frac{\sqrt{x}}{\log\left(e^{Tx}+\sqrt{e^{2Tx}-1}\right)}, \tag{6.2.33}$$

where $T>0$ is a fixed number. Then f_2 is a decreasing function.

Proof. A direct computation shows that the derivative of f_2 is non-positive if and only if

$$\sqrt{e^{2Tx}-1}\,\log\left(e^{Tx}+\sqrt{e^{2Tx}-1}\right)\leq 2Txe^{Tx}, \quad \text{for all } x>0,$$

which is equivalent to

$$\sqrt{1-e^{-2Tx}}\,\log\left(1+\sqrt{1-e^{-2Tx}}\right)\leq\left(2-\sqrt{1-e^{-2Tx}}\right)Tx, \quad \text{for all } x>0. \tag{6.2.34}$$

We put $y := \sqrt{1-e^{-2Tx}}$, and we write (6.2.34) as

$$y\log(1+y) + \frac{1}{2}(2-y)\log(1-y^2) \le 0, \quad \text{for all } 0<y<1,$$

which in turn is equivalent to

$$\left(1+\frac{y}{2}\right)\log(1+y) + \left(1-\frac{y}{2}\right)\log(1-y) \le 0, \quad \text{for all } 0<y<1. \tag{6.2.35}$$

Now define $g_2 : (0,1) \to \mathbb{R}$ as $g_2(y) := (1+\frac{y}{2})\log(1+y) + (1-\frac{y}{2})\log(1-y)$ and observe that g_2 is concave with $g_2(0)=0$, $g_2'(0)=0$. Thus g_2 is non-positive on $(0,1)$ and the inequality (6.2.35) is proved. □

The combination of (6.1.5), (6.1.4) and (6.2.32) implies that if $(X,\mathsf{d},\mathfrak{m})$ is an $\mathsf{RCD}(K,\infty)$ space with $K<0$ and $\mathfrak{m}(X)=1$ then

$$h(X) \ge \sqrt{-\frac{\pi K}{2}} \sup_{t>0} \frac{1-e^{-\lambda_1 t}}{\log\left(e^{-Kt}+\sqrt{e^{-2Kt}-1}\right)}. \tag{6.2.36}$$

We make two different choices:

- When $\lambda_1 \le -K$, we choose $t=-\frac{1}{K}$ in (6.2.36) so that

$$h(X) \ge \sqrt{-\frac{\pi K}{2}} \frac{1-e^{\frac{\lambda_1}{K}}}{\log\left(e+\sqrt{e^2-1}\right)} \ge \lambda_1 \sqrt{-\frac{\pi}{2K}} \frac{1-\frac{1}{e}}{\log\left(e+\sqrt{e^2-1}\right)}, \tag{6.2.37}$$

  where we used the inequality

$$1-e^{-x} \ge \left(1-\frac{1}{e}\right)x, \quad \text{for all } 0\le x\le 1.$$

- When $\lambda_1 > -K$, we choose $t=\frac{1}{\lambda_1}$ in (6.2.36) so that

$$h(X) \ge \sqrt{\frac{\pi}{2}}\sqrt{\lambda_1}\left(1-\frac{1}{e}\right) \frac{\sqrt{-\frac{K}{\lambda_1}}}{\log\left(e^{-\frac{K}{\lambda_1}}+\sqrt{e^{-2\frac{K}{\lambda_1}}-1}\right)}.$$

  Applying now Lemma 6.2.4, we obtain

$$\lambda_1 \le \frac{2\Big(\log\left(e+\sqrt{e^2-1}\right)\Big)^2}{\pi\left(1-\frac{1}{e}\right)^2} h(X)^2. \tag{6.2.38}$$

The combination of (6.2.37) and (6.2.38) gives that, if $(X,\mathsf{d},\mathfrak{m})$ is an $\mathsf{RCD}(K,\infty)$ space with $K<0$ and $\mathfrak{m}(X)=1$

$$\begin{aligned}\lambda_1 &\le \max\left\{\sqrt{-K}\frac{\sqrt{2}\log\left(e+\sqrt{e^2-1}\right)}{\sqrt{\pi}(1-\frac{1}{e})}h(X), \frac{2\Big(\log\left(e+\sqrt{e^2-1}\right)\Big)^2}{\pi\left(1-\frac{1}{e}\right)^2}h(X)^2\right\}\\ &< \max\left\{\frac{21}{10}\sqrt{-K}h(X), \frac{22}{5}h(X)^2\right\}.\end{aligned} \tag{6.2.39}$$

In case $(X,\mathsf{d},\mathfrak{m})$ is an $\mathsf{RCD}(K,\infty)$ space with $K<0$ and $\mathfrak{m}(X)=\infty$ then, using (6.1.6) instead of (6.1.5), the estimates (6.2.36) and (6.2.39) hold with λ_1 replaced by λ_0 and $h(X)$ replaced by $h(X)/2$. Thus, in case $\mathfrak{m}(X)=\infty$, we obtain:

$$\lambda_0 \le \max\left\{\sqrt{-K}\frac{\log\left(e+\sqrt{e^2-1}\right)}{\sqrt{2\pi}(1-\frac{1}{e})}h(X), \frac{\left(\log\left(e+\sqrt{e^2-1}\right)\right)^2}{2\pi\left(1-\frac{1}{e}\right)^2}h(X)^2\right\}$$
$$< \max\left\{\frac{21}{20}\sqrt{-K}h(X), \frac{11}{10}h(X)^2\right\}. \quad (6.2.40)$$

□

Remark 6.2.5. *Another bound, similar to the one obtained in the case $K>0$, can be achieved in the presence of a lower bound for K/λ_1, if $\mathfrak{m}(X)=1$ (resp. a lower bound for K/λ_0, if $\mathfrak{m}(X)=\infty$). To see this, let us suppose $K/\lambda_1\ge -c$, $c>0$ (resp. $K/\lambda_0\ge -c$). Then, using (6.1.5) (resp. (6.1.6)), (6.1.4) and Lemma 6.2.4, we have that (resp. the left-hand side can be improved to $h(X)/\sqrt{2\pi}$)*

$$\sqrt{\frac{2}{\pi}}h(X) \ge \sqrt{\lambda_1}\sup_{T>0}\frac{\sqrt{-\frac{K}{\lambda_1}}}{\log\left(e^{-\frac{K}{\lambda_1}T}+\sqrt{e^{-2\frac{K}{\lambda_1}T}-1}\right)}\left(1-e^{-T}\right)$$
$$\ge \sqrt{c\lambda_1}\sup_{T>0}\frac{1-e^{-T}}{\log\left(e^{cT}+\sqrt{e^{2cT}-1}\right)}. \quad (6.2.41)$$

6.3 Appendix A: Cheeger's inequality in general metric measure spaces

The Buser-type inequalities of Theorem 6.1.1 and Corollary 6.1.2 give an upper bound on λ_1 (resp. on λ_0, in case $\mathfrak{m}(X)=\infty$) in terms of the Cheeger constant $h(X)$. It is natural to ask if also a reverse inequality holds, namely if it possible to give a lower bound on λ_1 (resp. on λ_0, in case $\mathfrak{m}(X)=\infty$) in terms of $h(X)$. The answer is affirmative in the higher generality of metric measure spaces with a non-negative locally bounded measure *without curvature conditions*, see Theorem 6.3.2 below. This generalizes to the metric measure setting a celebrated result by Cheeger [Che70], known as Cheeger's inequality. In contrast to the previous section, here we do not assume the separability of the space.

A key tool in the proof of Cheeger's inequality is the co-area formula; more precisely, in the arguments it is enough to have an inequality in the co-area formula. For the reader's convenience, we give below the statement and a self-contained proof.

Proposition 6.3.1 (Coarea inequality)**.** *Let (X,d) be a complete metric space and let $\mathfrak{m}$ be a non-negative Borel measure finite on bounded subsets.*
Let $u\in\mathsf{Lip}_{bs}(X)$, $u:X\to[0,\infty)$ and set $M=\sup_X u$. Then for $\mathcal{L}^1$-a.e. $t>0$ the set $\{u>t\}$ has finite perimeter and

$$\int_0^M \mathrm{Per}(\{u>t\})\,\mathrm{d}t \le \int_X |\nabla u|\,\mathrm{d}\mathfrak{m}. \quad (6.3.1)$$

Proof. The proof is quite standard, but since we did not find it in the literature stated at this level of generality (tipically one assumes some extra condition like measure

doubling and gets a stronger statement, namely equality in the co-area formula; see for instance [M M03]) we add it for the reader's convenience.
Let $E_t := \{u > t\}$ and set $V(t) := \int_{E_t} |\nabla u| \,\mathrm{d}\mathfrak{m}$. The function $t \mapsto V(t)$ is non-increasing and bounded, thus differentiable for $\mathcal{L}^1$-a.e. $t > 0$.
Since $\int_X u \,\mathrm{d}\mathfrak{m} < \infty$, we also have that $\mathfrak{m}(\{u = t\}) = 0$ for $\mathcal{L}^1$-a.e. $t > 0$.

Fix $t > 0$ a differentiability point for V for which $\mathfrak{m}(\{u = t\}) = 0$, and define $\psi : (0,t) \times (0,\infty) \to [0,1]$ as

$$\psi(h,s) := \begin{cases} 0 & \text{for } s \leq t-h \\ \frac{1}{h}(s-t)+1 & \text{for } t-h < s \leq t \\ 1 & \text{for } s > t. \end{cases} \tag{6.3.2}$$

For $h > 0$ define $u_h(x) = \psi(h, u(x))$ and observe that the sequence $(u_h)_h \subset \mathsf{Lip}_{bs}(X)$. We first claim that

$$u_h \to \chi_{E_t} \quad \text{in } L^1(X,\mathfrak{m}) \quad \text{as } h \downarrow 0. \tag{6.3.3}$$

Indeed

$$\begin{aligned} \int_X |u_h - \chi_{E_t}| \,\mathrm{d}\mathfrak{m} &= \int_{\{t-h<u\leq t\}} \psi(h,u) \,\mathrm{d}\mathfrak{m} \\ &\leq \mathfrak{m}\left(\{t-h < u \leq t\}\right) \to \mathfrak{m}(\{u=t\}) = 0 \qquad \text{as } h \downarrow 0, \end{aligned}$$

by Dominated Convergence Theorem, since by assumption u has bounded support, $\mathfrak{m}$ is finite on bounded sets and $\chi_{\{t-h<u\leq t\}} \to \chi_{\{u=t\}}$ pointwise as $h \downarrow 0$.
In order to prove that E_t is a set of finite perimeter it is then sufficient to show that $\limsup_{h\downarrow 0} \int_X |\nabla u_h| \,\mathrm{d}\mathfrak{m} < \infty$. To this aim observe that

$$\int_X |\nabla u_h| \,\mathrm{d}\mathfrak{m} = \frac{1}{h} \int_{\{t-h<u\leq t\}} |\nabla u| \,\mathrm{d}\mathfrak{m} = \frac{V(t-h) - V(t)}{h}.$$

Since by assumption $t > 0$ is a differentiability point for V, we obtain that E_t is a finite perimeter set satisfying

$$\mathrm{Per}(E_t) \leq \lim_{h\downarrow 0} \int_X |\nabla u_h| \,\mathrm{d}\mathfrak{m} = -V'(t). \tag{6.3.4}$$

Using that (6.3.4) holds for $\mathcal{L}^1$-a.e. $t > 0$ and that V is non-increasing, we get

$$\int_0^M \mathrm{Per}(E_t) \,\mathrm{d}t \leq -\int_0^M V'(t) \,\mathrm{d}t \leq V(0) - V(M) = \int_X |\nabla u| \,\mathrm{d}\mathfrak{m}. \tag{6.3.5}$$

□

Theorem 6.3.2 (Cheeger's Inequality in metric measure spaces)**.** *Let (X, d) be a complete metric space and let $\mathfrak{m}$ be a non-negative Borel measure finite on bounded subsets.*

1. *If $\mathfrak{m}(X) < \infty$ then*
$$\lambda_1 \geq \frac{1}{4} h(X)^2. \tag{6.3.6}$$

2. *If $\mathfrak{m}(X) = \infty$ then*
$$\lambda_0 \geq \frac{1}{4} h(X)^2. \tag{6.3.7}$$

As proved by Buser [Bus78], the constant 1/4 in (6.3.6) is optimal in the following sense: for any $h > 0$ and $\varepsilon > 0$, there exists a closed (i.e. compact without boundary) two-dimensional Riemannian manifold (M, g) with $h(M) = h$ and such that $\lambda_1 \leq \frac{1}{4}h(M)^2 + \varepsilon$.

Proof. We give a proof of (6.3.6), the arguments for showing (6.3.7) being analogous (and even simpler).

By the very definition of λ_1 as in (6.2.15), for every $\varepsilon > 0$ there exists $f \in \mathsf{Lip}_{bs}(X)$ with $\int_X f \,\mathrm{d}\mathfrak{m} = 0$, $f \not\equiv 0$ such that

$$\lambda_1 \geq \frac{\int_X |\nabla f|^2 \,\mathrm{d}\mathfrak{m}}{\int_X f^2 \,\mathrm{d}\mathfrak{m}} - \varepsilon. \tag{6.3.8}$$

Let m be any median of the function f and set $f^+ := \max\{f - m, 0\}$, $f^- := -\min\{f - m, 0\}$. Applying the co-area inequality (6.3.1) to $u = (f^+)^2$ (respectively $(f^-)^2$) and recalling the definition of Cheeger's constant $h(X)$ as in (6.1.3), we obtain

$$\begin{aligned} &\int_X |\nabla(f^+)^2| \,\mathrm{d}\mathfrak{m} + \int_X |\nabla(f^-)^2| \,\mathrm{d}\mathfrak{m} \\ &\geq \int_0^{\sup\{(f^+)^2\}} \mathrm{Per}(\{(f^+)^2 > t\}) \,\mathrm{d}t + \int_0^{\sup\{(f^-)^2\}} \mathrm{Per}(\{(f^-)^2 > t\}) \,\mathrm{d}t \\ &\geq h(X) \int_0^{\sup\{(f^+)^2\}} \mathfrak{m}(\{(f^+)^2 > t\}) \,\mathrm{d}t + h(X) \int_0^{\sup\{(f^-)^2\}} \mathfrak{m}(\{(f^-)^2 > t\}) \,\mathrm{d}t \\ &= h(X) \int_X (f^+)^2 \mathrm{d}\mathfrak{m} + h(X) \int_X (f^-)^2 \mathrm{d}\mathfrak{m} = h(X) \int_X |f - m|^2 \,\mathrm{d}\mathfrak{m}. \end{aligned} \tag{6.3.9}$$

Since

$$|\nabla g^2| \leq 2|g|\,|\nabla g|,$$

and

$$|\nabla f^+| \leq |\nabla f|, \quad |\nabla f^-| \leq |\nabla f|,$$

we can apply the Cauchy-Schwarz inequality and get

$$2\Big(\int_X |\nabla f|^2 \,\mathrm{d}\mathfrak{m}\Big)^{\frac{1}{2}} \Big(\int_X |f - m|^2 \,\mathrm{d}\mathfrak{m}\Big)^{\frac{1}{2}} \geq \int_X |\nabla(f^+)^2| \,\mathrm{d}\mathfrak{m} + \int_X |\nabla(f^-)^2| \,\mathrm{d}\mathfrak{m}, \tag{6.3.10}$$

where we have used that $|f^+| + |f^-| = |f - m|$. It follows from (6.3.9) and (6.3.10) that for every median m of f it holds

$$\frac{\int_X |\nabla f|^2 \,\mathrm{d}\mathfrak{m}}{\int_X |f - m|^2 \,\mathrm{d}\mathfrak{m}} \geq \frac{h(X)^2}{4}. \tag{6.3.11}$$

Finally, since $\int_X f \mathrm{d}\mathfrak{m} = 0$ and the mean minimises $\mathbb{R} \ni c \mapsto \int_X |f - c|^2 \mathrm{d}\mathfrak{m}$, we have

$$\frac{\int_X |\nabla f|^2 \,\mathrm{d}\mathfrak{m}}{\int_X |f|^2 \,\mathrm{d}\mathfrak{m}} \geq \frac{\int_X |\nabla f|^2 \,\mathrm{d}\mathfrak{m}}{\int_X |f - m|^2 \,\mathrm{d}\mathfrak{m}}$$

and we can conclude thanks to (6.3.8) and the fact that $\varepsilon > 0$ is arbitrary. □

Part III

Wasserstein stability of porous medium-type equations on manifolds with Ricci curvature bounded below

Chapter 7

Wasserstein stability of porous medium equations on manifolds

The last part of the thesis is devoted to the study of Wasserstein-stability properties of porous medium-type equations on Riemannian manifolds with Ricci curvature bounded from below.
Stability properties of the porous medium equation on *nonnegatively* curved manifolds are known since the work of Sturm [Stu05] and are a consequence of the displacement convexity of the associated energy functional on the Wasserstein space. In order to generalise the result on (noncollapsed) manifolds with merely $\mathrm{Ric} \geq -K$, we take advantage of a quantitative L^1–L^∞ smoothing estimate satisfied by the equation, combined with a clever Hamiltonian approach developed by Ambrosio, Mondino and Savaré [AMS19].

7.1 Notations

Throughout, we will deal with a smooth, complete and connected Riemannian manifold $(\mathbb{M}^n, \mathfrak{g})$. In the sequel, for simplicity, we will omit the explicit dependence of the geometric quantities on the metric $\mathfrak{g}$. On $\mathbb{M}^n$ we always consider the associated Riemannian distance d and the Riemannian volume measure $\mathcal{V}$. The former, with some abuse of notation, will also be used to denote distance between sets. The symbol $T_x\mathbb{M}^n$ will stand for the tangent space at $x \in \mathbb{M}^n$, endowed with a scalar product $\langle \cdot, \cdot \rangle$ that induces the Riemannian norm $|\cdot|$.

If the measures $\mu, \nu \in \mathscr{M}(\mathbb{M}^n)$ have densities w.r.t. $\mathcal{V}$, say ρ_μ and ρ_ν, we will often write $\mathcal{W}_2(\rho_\mu, \rho_\nu)$ in place of $\mathcal{W}_2(\mu, \nu)$.

For simplicity's sake, in the following we use the notations $\mathbb{H}$, $\mathbb{V}$ and $\mathbb{D}$ for the Hilbert spaces

$$\mathbb{H} := L^2(\mathbb{M}^n), \qquad \mathbb{V} := W^{1,2}(\mathbb{M}^n), \qquad \mathbb{D} := \{f \in \mathbb{V} : \Delta f \in \mathbb{H}\}, \tag{7.1.1}$$

with associated norms $\|f\|_{\mathbb{V}}^2 := \|f\|_{\mathbb{H}}^2 + \|\nabla f\|_{\mathbb{H}}^2$ and $\|f\|_{\mathbb{D}}^2 := \|f\|_{\mathbb{V}}^2 + \|\Delta f\|_{\mathbb{H}}^2$.

Given $T > 0$ and two Hilbert spaces X and Y continuously embedded in a Banach space U, we introduce the space of time-dependent functions

$$W^{1,2}((0,T); X, Y) := \left\{ u \in W^{1,2}((0,T); U) : \, u \in L^2((0,T); X), \, \tfrac{\mathrm{d}u}{\mathrm{d}t} \in L^2((0,T); Y) \right\},$$

with associated norm

$$\|u\|_{W^{1,2}((0,T);X,Y)}^2 := \|u\|_{L^2((0,T);X)}^2 + \left\| \tfrac{\mathrm{d}u}{\mathrm{d}t} \right\|_{L^2((0,T);Y)}^2 .$$

Let $T > 0$. For any function $F \in \mathcal{C}^1(\mathbb{R})$ with $F(0) = 0$, such that $0 < \lambda \le F'(r) \le \lambda^{-1}$ for every $r \in \mathbb{R}$, for some $\lambda > 0$, in agreement with [AMS19, Section 3.3] we introduce the set

$$\mathcal{ND}_F(0,T) := \left\{ u \in W^{1,2}((0,T);\mathbb{H}) \cap \mathcal{C}^1([0,T];\mathbb{V}') : \, F(u) \in L^2((0,T);\mathbb{D}) \right\}.$$

As a general rule, we will use superscripts to denote the parameter of curves that are related to geodesics in the Wasserstein space over $(\mathbb{M}^n, \mathsf{d})$ and subscripts to denote the index or parameter of an approximation. Since subscripts are also typically used to refer to initial data of a Cauchy problem as in (1.3.1) or (1.3.9), we will try to avoid ambiguity as much as possible.

Finally, when referring to a function $\rho : D \subseteq \mathbb{M}^n \times \mathbb{R}^+ \to \mathbb{R}$ (or to a measure) evaluated at some time t *as a whole*, we will adopt the notation $\rho(t)$ (or $\mu(t)$). As for its time derivative, we will write $\frac{\partial \rho}{\partial t}$ whenever it can be understood as a classical partial derivative; we will write $\frac{\mathrm{d}\rho}{\mathrm{d}t}$ instead if it must be interpreted as the time derivative of ρ as a curve in a suitable Banach space. The notation $\dot{\rho}$ will mostly be used for *metric velocity*.

7.2 Statement of the main results

We consider the following nonlinear diffusion equation:

$$\begin{cases} \partial_t \rho = \Delta P(\rho) & \text{in } \mathbb{M}^n \times \mathbb{R}^+ , \\ \rho(\cdot, 0) = \mu_0 \ge 0 & \text{in } \mathbb{M}^n \times \{0\} , \end{cases} \tag{7.2.1}$$

where $\mu_0 \in \mathscr{M}_2^M(\mathbb{M}^n)$ and $\rho \mapsto P(\rho)$ is a suitable $\mathcal{C}^1([0,+\infty))$ function of *porous medium type*. We require that $\mathbb{M}^n$ and P satisfy a precise set of hypotheses.

Hypotheses 7.2.1 (Manifold)**.** *We assume throughout that $\mathbb{M}^n$ ($n \ge 3$) is a smooth, complete and connected Riemannian manifold. Moreover, it will comply with either one or more of the following conditions:*

- *The Ricci curvature is uniformly bounded from below, i.e. there exists $K \ge 0$ such that*
$$\mathrm{Ric}_x(v,v) \ge -K|v|^2 \qquad \forall x \in \mathbb{M}^n \text{ and } v \in T_x\mathbb{M}^n ; \tag{H1}$$
- *For some $C_S > 0$ there holds the Sobolev-type inequality*
$$\|f\|_{L^{2^\star}(\mathbb{M}^n)} \le C_S \left(\|\nabla f\|_{L^2(\mathbb{M}^n)} + \|f\|_{L^2(\mathbb{M}^n)} \right) \quad \forall f \in W^{1,2}(\mathbb{M}^n) , \tag{H2}$$
with $2^\star := 2n/(n-2)$.

A result originally due to Varopoulos [Var89] asserts that (H2) does hold on any complete, n-dimensional ($n \ge 3$) Riemannian manifold satisfying (H1) along with the *noncollapse* condition

$$\inf_{x \in \mathbb{M}^n} \mathcal{V}(B_1(x)) > 0 , \tag{7.2.2}$$

where $B_1(x) := \{ y \in \mathbb{M}^n : \mathsf{d}(x,y) < 1 \}$. We refer in particular to [Heb99, Theorem 3.2] (in fact B_1 could be replaced by B_r for any $r > 0$). Condition (7.2.2) is also necessary for (H2) to hold, see [Heb99, Lemma 2.2]. Note that (H1) and (7.2.2) are for free on any *compact* Riemannian manifold, a simple subcase of the frameworks we will work within. On the other hand, if $\mathbb{M}^n$ is noncompact and has finite volume, or more in general has an end with finite volume, then (7.2.2) (and therefore (H2)) necessarily fails.

As concerns the nonlinearity P appearing in (7.2.1), we introduce the following set of hypotheses. We write them separately in order to be able to single out the specific assumption(s) needed for each result we will prove.

Hypotheses 7.2.2 (Nonlinearity). *We assume throughout that $P \in \mathcal{C}^1([0,+\infty))$. Moreover, it will comply with either one or more of the following conditions:*

$$P(0) = 0 \text{ and the map } \rho \mapsto P(\rho) \text{ is strictly increasing}; \tag{H3}$$

there exist $c_1 \geq c_0 > 0$ and $m > 1$ such that

$$c_0\, m\, \rho^{m-1} \leq P'(\rho) \leq c_1\, m\, \rho^{m-1} \qquad \forall \rho \geq 0\,, \tag{H4}$$

$$\rho\, P'(\rho) - \left(1 - \tfrac{1}{n}\right) P(\rho) \geq 0 \qquad \forall \rho \geq 0\,. \tag{H5}$$

It is straightforward to check that (H5) is implied by (H4) provided $c_1 \leq c_0\, m\, \frac{n}{n-1}$.

Let us firstly notice that the choice $P(\rho) = \rho^m$ for some $m > 1$ (corresponding to the PME) obviously implies (H3), (H4) and (H5). We point out that condition (H4) is essential to establish the smoothing effect (see (7.2.5)) and the compact-support property (see Proposition 7.3.4), while (H5) is a key tool to develop the Hamiltonian approach in its abstract formulation (we refer to Lemma 7.4.2).

We start by providing a good notion of weak solution of (7.2.1) for initial data in $\mathscr{M}_2^M(\mathbb{M}^n)$ and for a general nonlinearity P, which is inspired by the (wide) existing literature, see Section 7.3.

Definition 7.2.3 (Weak Wasserstein solutions). *Let P comply with assumption* (H3). *Given $\mu_0 \in \mathscr{M}_2^M(\mathbb{M}^n)$, we say that a nonnegative measurable function ρ is a* Wasserstein solution *of* (7.2.1) *if, for every $T > \tau > 0$, there hold*

$$\rho, P(\rho) \in L^2(\mathbb{M}^n \times (\tau, T))\,, \quad \nabla P(\rho) \in L^2(\mathbb{M}^n \times (\tau, T))\,, \tag{7.2.3}$$

$$\int_0^T \int_{\mathbb{M}^n} \rho\, \partial_t \eta \, \mathrm{d}\mathcal{V} \mathrm{d}t = \int_0^T \int_{\mathbb{M}^n} \langle \nabla P(\rho)\,, \nabla \eta \rangle \, \mathrm{d}\mathcal{V} \mathrm{d}t \tag{7.2.4}$$

for every $\eta \in W_c^{1,2}((0,T); L^2(\mathbb{M}^n))$ with $\nabla \eta \in L^2((0,T); L^2(\mathbb{M}^n))$, and

$$\mu \in \mathcal{C}\big([0,T]; \big(\mathscr{M}_2^M(\mathbb{M}^n), \mathcal{W}_2\big)\big) \text{ with } \mu(0) = \mu_0\,,$$

where $\mu(t) = \rho(t)\mathcal{V}$ for $t > 0$.

We are now in position to state our main results, which will be proved in Section 7.4.

Theorem 7.2.4 (Wasserstein stability). *Let $\mathbb{M}^n$ $(n \geq 3)$ comply with assumptions* (H1) *and* (H2).

Let moreover P comply with assumptions (H3), (H4) *and* (H5). *Let $\mu_0 \in \mathscr{M}_2^M(\mathbb{M}^n)$. Then there exists a* unique *Wasserstein solution ρ of* (7.2.1), *which satisfies the smoothing estimate*

$$\|\rho(t)\|_{L^\infty(\mathbb{M}^n)} \leq C \left(t^{-\frac{n}{2+n(m-1)}}\, M^{\frac{2}{2+n(m-1)}} + M \right) \qquad \forall t > 0\,, \tag{7.2.5}$$

where $C \geq 1$ is a constant depending only on C_S, n, c_0 and independent of m ranging in a bounded subset of $(1,+\infty)$. Furthermore, if $\hat{\rho}$ is the Wasserstein solution of (7.2.1) corresponding to another initial datum $\hat{\mu}_0 \in \mathscr{M}_2^M(\mathbb{M}^n)$, the stability estimate

$$\mathcal{W}_2(\rho(t),\hat{\rho}(t)) \leq \exp\left\{K\,c_1\,\mathfrak{C}_m\left[\left(tM^{m-1}\right)^{\frac{2}{2+n(m-1)}} \vee \left(tM^{m-1}\right)\right]\right\}\mathcal{W}_2(\mu_0,\hat{\mu}_0) \quad \forall t>0 \tag{7.2.6}$$

holds, where $\mathfrak{C}_m := C^{m-1}\,2^{m-2}\,[2+n(m-1)]$.

In fact (7.2.6) is sharp, as $t \downarrow 0$, in the hyperbolic space $\mathbb{H}^n_K$ of sectional curvature $-K$, i.e. of Ricci curvature $-(n-1)K$.

Theorem 7.2.5 (Optimality). *Estimate (7.2.6) is* optimal *in $\mathbb{M}^n = \mathbb{H}^n_K$, for $P(\rho) = \rho^m$, with the choices $\mu_0 = M\delta_x$ and $\hat{\mu}_0 = M\delta_y$, provided the points $x,y \in \mathbb{H}^n_K$ are close enough. More precisely, upon setting $\delta := \mathsf{d}(x,y) > 0$, there exist constants $\kappa = \kappa(n,m) > 0$, $\overline{\delta} = \overline{\delta}(n,K,m) > 0$ and $\overline{t} = \overline{t}(\delta,n,K,m,M) > 0$ such that if $\delta \in (0,\overline{\delta})$ then*

$$\mathcal{W}_2(\rho(t),\hat{\rho}(t)) \geq \left[1 + K\,\kappa\left(tM^{m-1}\right)^{\frac{2}{2+n(m-1)}}\right]\mathcal{W}_2(\mu_0,\hat{\mu}_0) \qquad \forall t \in (0,\overline{t})\,. \tag{7.2.7}$$

The proof of Theorem 7.2.5 will be provided in Subsection 7.4.4. Some comments regarding both Theorem 7.2.4 and Theorem 7.2.5 are now in order.

Remark 7.2.6 (The PME, the heat equation and gradient flows)**.** As mentioned above, the explicit choice $P(\rho) = \rho^m$ corresponds to the well-known *porous medium equation* (PME). In this case estimate (7.2.6) holds with $c_1 = 1$. In particular, if we let $m \downarrow 1$, thanks to the fact that $\mathfrak{C}_m \to 1$ we recover exactly the following stability estimate for the heat equation:

$$\mathcal{W}_2(\rho(t),\hat{\rho}(t)) \leq e^{K\,t}\,\mathcal{W}_2(\mu_0,\hat{\mu}_0) \qquad \forall t > 0\,. \tag{7.2.8}$$

We recall that the Ricci bound (H1) is *equivalent* to the $(-K)$-gradient flow formulation of the heat equation with respect to the relative entropy in $(\mathscr{P}_2(\mathbb{M}^n),\mathcal{W}_2)$, from which (7.2.8) follows: we refer to [RS05, Theorem 1.1 and Corollary 1.4] for more details. We stress that, as a byproduct of Theorem 7.2.5, we can deduce that in general on negatively-curved manifolds the porous medium equation *cannot* be seen as the gradient flow of some λ-convex functional with respect to the 2-Wasserstein distance, at least in the sense of Evolution Variational Inequalities (see [AGS08]). Indeed, if it was, then the estimate

$$\mathcal{W}_2(\rho(t),\hat{\rho}(t)) \leq e^{\lambda\,t}\,\mathcal{W}_2(\mu_0,\hat{\mu}_0) \qquad \forall t > 0$$

would hold for some $\lambda \in \mathbb{R}$, thus contradicting (7.2.7). On the other hand, it is known that the PME *can* indeed be seen as the gradient flow of the free energy (1.3.2) in the case where the Ricci curvature is nonnegative (we refer to [Stu05] and [Ott01; OW05]), so that (7.2.6) holds with $K = 0$.

Remark 7.2.7 (The Cartan-Hadamard case)**.** If, in place of (H2), the manifold $\mathbb{M}^n$ supports a *Euclidean* Sobolev inequality, namely

$$\|f\|_{L^{2^\star}(\mathbb{M}^n)} \leq C_S\,\|\nabla f\|_{L^2(\mathbb{M}^n)} \qquad \forall f \in \mathcal{C}^1_c(\mathbb{M}^n)\,, \tag{7.2.9}$$

then it is not difficult to deduce that (7.2.6) turns into a better estimate:

$$\mathcal{W}_2(\rho(t),\hat{\rho}(t)) \leq \exp\left\{K\,c_1\,\mathfrak{C}_m\left(tM^{m-1}\right)^{\frac{2}{2+n(m-1)}}\right\}\mathcal{W}_2(\mu_0,\hat{\mu}_0) \qquad \forall t > 0\,.$$

This is a simple consequence of our method of proof, since in that case the smoothing effect (7.2.5) holds with no additional M term in the right-hand side, which causes the linear term to appear in the exponent of (7.2.6). We recall that (7.2.9) does hold, for instance, on any *Cartan-Hadamard* manifold, that is a complete, simply connected Riemannian manifold with everywhere nonpositive sectional curvature (see [GMP18b] and references therein).

Remark 7.2.8 (The 2-dimensional case)**.** The results of Theorem 7.2.4 can also be extended to the dimension $n = 2$. In that case, the Sobolev inequality should be replaced by the Gagliardo-Nirenberg inequality

$$\|f\|_{L^r(\mathbb{M}^2)} \le C_{GN} \left(\|\nabla f\|_{L^2(\mathbb{M}^2)} + \|f\|_{L^2(\mathbb{M}^2)} \right)^{\frac{r-s}{r}} \|f\|_{L^s(\mathbb{M}^2)}^{\frac{s}{r}} \quad \forall f \in W^{1,2} \cap L^s(\mathbb{M}^2) \, , \tag{7.2.10}$$

for some $r > s > 0$ and $C_{GN} > 0$. We recall that, by [Bak+95, Theorem 3.3], the validity of (7.2.10) for *some* $r > s > 0$ yields the validity of the same inequality for *every* $r > s > 0$. In particular, this allows us to reproduce the proof of Proposition 7.3.3 also for $n = 2$, starting from (7.2.10) in place of (7.3.33). The rest of the results we need in order to establish Theorem 7.2.4 also hold for $n = 2$. Note that, again, inequality (7.2.10) is satisfied (e.g. with $r > 2$ and $s = r - 2$) on any 2-dimensional Riemannian manifold complying with (H1) and (7.2.2): this is a simple consequence of [Heb99, Lemma 2.1 and Theorem 3.2]. As concerns the optimality result contained in Theorem 7.2.5, we just observe that its proof follows with no modifications in the case $n = 2$ as well (see Subsection 7.4.4).

7.3 Fundamental properties of porous medium-type equations on manifolds

This section is devoted to the study of (7.2.1) for more regular initial data, that is the problem

$$\begin{cases} \partial_t \rho = \Delta P(\rho) & \text{in } \mathbb{M}^n \times \mathbb{R}^+ \, , \\ \rho(\cdot, 0) = \rho_0 \ge 0 & \text{on } \mathbb{M}^n \times \{0\} \, , \end{cases} \tag{7.3.1}$$

with $\rho_0 \in L^1(\mathbb{M}^n) \cap L^\infty(\mathbb{M}^n)$. To begin with, we will introduce the notion of *weak energy* solution and then discuss some crucial related properties. In particular, we will focus on the smoothing effect and on a bound on the support of such solutions (when the initial data are compactly supported). Inspired by [AMS19], for a restricted class of nonlinearities we will also give an alternative (variational) notion of solution and consequently prove the equivalence with the weak-energy one. Finally, with regards to the Hamiltonian strategy mentioned in the Introduction, we will discuss well-posedness results for the *forward linearized* equation associated with (7.3.1) and for the related *backward adjoint* equation.

For convenience, in the following we make the additional (implicit) assumption that $\mathbb{M}^n$ is *noncompact* and with *infinite volume*, as well as in Subsection 7.4.2. Note that for our purposes there is no point in considering noncompact manifolds with *finite* volume, since most of our results require the validity of the Sobolev inequality (H2), which does not hold on such manifolds.

7.3.1 Weak energy solutions

The concept of *weak energy* solution of (7.3.1) has been proved to be well suited for porous medium-type equations: see e.g. [Váz07, Subsections 5.3.2 and 11.2.1], [FM17, Section 3], [GMP13, Subsections 3.1 and 3.2] or [GMP18b, Section 2]. Here we mostly take inspiration from [FM17, Section 3]: there the framework is purely Euclidean, but the basic definitions and properties are straightforwardly adaptable to the Riemannian setting.

Even if in Subsection 7.1 we introduced the more synthetic notations (7.1.1), here we keep the standard notations typically used in the PDE framework.

Definition 7.3.1 (Weak energy solutions). *Let P comply with assumption* (H3). *Given a nonnegative $\rho_0 \in L^1(\mathbb{M}^n) \cap L^\infty(\mathbb{M}^n)$, we say that a nonnegative measurable function ρ is a* weak energy solution *of* (7.3.1) *if, for every $T > 0$, there hold*

$$\rho, P(\rho) \in L^2(\mathbb{M}^n \times (0,T)) , \qquad \nabla P(\rho) \in L^2(\mathbb{M}^n \times (0,T))$$

and

$$\int_0^T \int_{\mathbb{M}^n} \rho\, \partial_t \eta \, \mathrm{d}\mathcal{V}\mathrm{d}t = - \int_{\mathbb{M}^n} \rho_0(x)\, \eta(x,0)\, \mathrm{d}\mathcal{V}(x) + \int_0^T \int_{\mathbb{M}^n} \langle \nabla P(\rho) \, , \nabla \eta \rangle \, \mathrm{d}\mathcal{V}\mathrm{d}t \quad (7.3.2)$$

for every $\eta \in W^{1,2}((0,T); L^2(\mathbb{M}^n))$ with $\nabla\eta \in L^2((0,T); L^2(\mathbb{M}^n))$ such that $\eta(T) = 0$.

Existence and uniqueness of weak energy solutions, at least for the class of data $L^1(\mathbb{M}^n) \cap L^\infty(\mathbb{M}^n)$, is by now a well-established issue (see e.g. the references quoted above). Nevertheless, since it will be very useful to our later purposes, we recall here the approximation procedure that allows one to construct such solutions: the essential idea is to approximate the possibly degenerate nonlinearity $P \in \mathcal{C}^1([0,+\infty))$ by means of suitable *nondegenerate* nonlinearities. More precisely, for every $\varepsilon > 0$ we define $P_\varepsilon \in \mathcal{C}^1([0,+\infty))$ by

$$(P_\varepsilon)'(\rho) := \begin{cases} P'(\rho) + \varepsilon & \text{if } \rho \in \left[0, \frac{1}{\varepsilon}\right], \\ P'\left(\frac{1}{\varepsilon}\right) + \varepsilon & \text{if } \rho > \frac{1}{\varepsilon}, \end{cases} \qquad P_\varepsilon(0) = 0 \, . \qquad (7.3.3)$$

It is apparent that P_ε satisfies

$$P_\varepsilon(\rho) \le P(\rho) + \varepsilon\rho \qquad \forall \rho \ge 0 \qquad (7.3.4)$$

and

$$\varepsilon \le (P_\varepsilon)'(\rho) \le \max_{\rho \in [0, 1/\varepsilon]} P'(\rho) + \varepsilon \qquad \forall \rho \ge 0 \, ;$$

in particular, P_ε is also strictly increasing. Moreover, by construction,

$$(P_\varepsilon)'(\rho) \ge P'(\rho) \qquad \forall \rho \in \left[0, \tfrac{1}{\varepsilon}\right] , \qquad (7.3.5)$$

and

$$\rho\, (P_\varepsilon)'(\rho) - \left(1 - \tfrac{1}{n}\right) P_\varepsilon(\rho) \ge 0 \qquad \forall \rho \ge 0 \qquad (7.3.6)$$

provided P satisfies the same inequality, i.e. (H5). Note that if P complies with the left-hand bound in (H4) so does P_ε' in the interval $[0, 1/\varepsilon]$, thanks to (7.3.5). Such a bound is crucial to establish the *smoothing effect*, which is a key ingredient to our strategy (see Proposition 7.3.3 below). Accordingly, we thus address the following approximate version of (7.3.1):

$$\begin{cases} \partial_t \rho_\varepsilon = \Delta P_\varepsilon(\rho_\varepsilon) & \text{in } \mathbb{M}^n \times \mathbb{R}^+ , \\ \rho_\varepsilon(\cdot, 0) = \rho_0 & \text{on } \mathbb{M}^n \times \{0\} . \end{cases} \qquad (7.3.7)$$

Problem (7.3.7) can be interpreted both from the viewpoint of *linear* and *nonlinear* theory, in the sense that P_ε is a nonlinear function but it is "uniformly elliptic", hence one expects that the solutions of (7.3.7) enjoy, to some extent, properties similar to those satisfied by the solutions of the *heat equation* (we refer to Propositions 7.3.6 and 7.3.7 below). We will mainly take advantage of the linear interpretation in Section 7.4, in agreement with the approach of [AMS19]. The nonlinear interpretation is exploited in the present section.

Proposition 7.3.2 (Existence, uniqueness, properties of weak energy solutions). *Let P comply with* (H3). *Given a nonnegative $\rho_0 \in L^1(\mathbb{M}^n) \cap L^\infty(\mathbb{M}^n)$, there exists a* unique *weak energy solution ρ of* (7.3.1)*, which enjoys the following additional properties:*

- L^1-continuity: *$\{\rho(t)\}_{t\geq 0}$ is a continuous curve with values in $L^1(\mathbb{M}^n)$;*
- Energy inequality: *ρ satisfies*
$$\int_0^T \int_{\mathbb{M}^n} |\nabla P(\rho)|^2 \, \mathrm{d}\mathcal{V}\mathrm{d}t + \int_{\mathbb{M}^n} \Psi(\rho(x,T)) \, \mathrm{d}\mathcal{V}(x) \leq \int_{\mathbb{M}^n} \Psi(\rho_0) \, \mathrm{d}\mathcal{V} \qquad \forall T > 0 \, , \tag{7.3.8}$$
where $\Psi(\rho) := \int_0^\rho P(r) \, \mathrm{d}r$;
- Nonexpansivity of the L^p norms: *for every $p \in [1, \infty]$ there holds*
$$\|\rho(t)\|_{L^p(\mathbb{M}^n)} \leq \|\rho_0\|_{L^p(\mathbb{M}^n)} \qquad \forall t > 0 \, ; \tag{7.3.9}$$
- Mass conservation: *if in addition $\mathbb{M}^n$ satisfies* (H1) *then*
$$\int_{\mathbb{M}^n} \rho(x,t) \, \mathrm{d}\mathcal{V}(x) = \int_{\mathbb{M}^n} \rho_0 \, \mathrm{d}\mathcal{V} \qquad \forall t > 0 \, ; \tag{7.3.10}$$
- Approximation: *if $\varepsilon > 0$ and ρ_ε is the weak energy solution of* (7.3.7)*, where $P_\varepsilon(\rho)$ is defined in* (7.3.3)*, then*
$$\lim_{\varepsilon \downarrow 0} \|\rho_\varepsilon(t) - \rho(t)\|_{L^1_{\mathrm{loc}}(\mathbb{M}^n)} = 0 \qquad \forall t > 0 \, ; \tag{7.3.11}$$
if in addition (H1) *is satisfied, then*
$$\lim_{\varepsilon \downarrow 0} \|\rho_\varepsilon(t) - \rho(t)\|_{L^1(\mathbb{M}^n)} = 0 \qquad \forall t > 0 \, ; \tag{7.3.12}$$
- L^1-contraction: *if $\hat\rho$ is the weak energy solution corresponding to another nonnegative initial datum $\hat\rho_0 \in L^1(\mathbb{M}^n) \cap L^\infty(\mathbb{M}^n)$, then*
$$\|\rho(t) - \hat\rho(t)\|_{L^1(\mathbb{M}^n)} \leq \|\rho_0 - \hat\rho_0\|_{L^1(\mathbb{M}^n)} \qquad \forall t > 0 \, . \tag{7.3.13}$$

Proof. We start by recalling that uniqueness of weak energy solutions follows from a standard trick due to Oleĭnik: given $T > 0$, one plugs the (admissible) test function
$$\eta(x,t) = \int_t^T [P(\rho(x,s)) - P(\hat\rho(x,s))] \, \mathrm{d}s \, , \qquad (x,t) \in \mathbb{M}^n \times [0,T] \, ,$$

into the weak formulation satisfied by the difference between ρ and $\hat{\rho}$ (the latter being two possibly different solutions corresponding to the same initial datum), thus obtaining

$$\begin{aligned}&\int_0^T\int_{\mathbb{M}^n}(\rho-\hat{\rho})\,(P(\rho)-P(\hat{\rho}))\,\mathrm{d}\mathcal{V}\mathrm{d}t\\&=\int_0^T\int_{\mathbb{M}^n}\left\langle\nabla\left[P(\rho(x,t))-P(\hat{\rho}(x,t))\right],\int_t^T\nabla\left[P(\rho(x,s))-P(\hat{\rho}(x,s))\right]\mathrm{d}s\right\rangle\mathrm{d}\mathcal{V}(x)\mathrm{d}t\,.\end{aligned}\tag{7.3.14}$$

A simple time integration in (7.3.14) yields

$$\begin{aligned}&\int_0^T\int_{\mathbb{M}^n}(\rho-\hat{\rho})\,(P(\rho)-P(\hat{\rho}))\,\mathrm{d}\mathcal{V}\mathrm{d}t\\&\qquad+\frac{1}{2}\int_{\mathbb{M}^n}\left|\int_0^T\nabla\left[P(\hat{\rho}(x,s))-P(\rho(x,s))\right]\mathrm{d}s\right|^2\mathrm{d}\mathcal{V}(x)=0\,,\end{aligned}\tag{7.3.15}$$

which ensures that $\rho=\hat{\rho}$ given the strict monotonicity of $\rho\mapsto P(\rho)$ and the arbitrariness of T. Note that here we have only used the validity of (7.3.2) for functions $\eta\in W^{1,2}((0,T);W^{1,2}(\mathbb{M}^n))$. Furthermore, the fact that $\rho_0\in L^1(\mathbb{M}^n)\cap L^\infty(\mathbb{M}^n)$ is unimportant. These observations will be useful in the proof of Proposition 7.3.6 below.

As concerns the construction of a weak energy solution, we will not provide a complete proof since the procedure is quite standard: see e.g. [Váz07, Theorem 5.7 and Lemma 5.8] or [GMP13, Theorems 3.4 and 3.7] in Euclidean or weighted-Euclidean contexts. The basic idea consists first of solving problem (7.3.1) in a sequence D_k of bounded regular domains that form an exhaustion for $\mathbb{M}^n$ (see the proof of Lemma 7.3.4 below for more details on such a sequence), with homogeneous Dirichlet boundary conditions on ∂D_k. In order to do this, it is convenient to make a further approximation by replacing P with P_ε: let us denote by $\rho_{\varepsilon,k}$ the corresponding solutions, which are therefore regular enough (up to approximating also the initial datum ρ_0 and approximating further P_ε in case P' is merely continuous – we skip this passages). A first key estimate is provided by the energy inequality itself, which is obtained upon multiplying the differential equation by $P_\varepsilon(\rho_{\varepsilon,k})$ and integrating by parts:

$$\int_0^T\int_{D_k}|\nabla P_\varepsilon(\rho_{\varepsilon,k})|^2\,\mathrm{d}\mathcal{V}\mathrm{d}t+\int_{D_k}\Psi_\varepsilon(\rho_{\varepsilon,k}(x,T))\,\mathrm{d}\mathcal{V}(x)=\int_{D_k}\Psi_\varepsilon(\rho_0)\,\mathrm{d}\mathcal{V}\qquad\forall T>0\,,\tag{7.3.16}$$

where $\Psi_\varepsilon(\rho):=\int_0^\rho P_\varepsilon(r)\,\mathrm{d}r$. Note that for the moment the energy inequality is in fact an identity. Another crucial estimate involves time derivatives and is obtained by multiplying the differential equation by $\zeta\,P_\varepsilon'(\rho_{\varepsilon,k})\,\partial_t\rho_{\varepsilon,k}$ and again integrating by parts, where $\zeta\in\mathcal{C}_c^\infty((0,+\infty))$ is any cut-off function that depends only on time and satisfies $0\le\zeta\le 1$; this yields

$$\int_0^T\int_{D_k}\zeta\,|\partial_t\Upsilon_\varepsilon(\rho_{\varepsilon,k})|^2\,\mathrm{d}\mathcal{V}\mathrm{d}t=\frac{1}{2}\int_0^T\int_{D_k}\zeta'\,|\nabla P_\varepsilon(\rho_{\varepsilon,k})|^2\,\mathrm{d}\mathcal{V}\mathrm{d}t\qquad\forall T>0\,,\tag{7.3.17}$$

where $\Upsilon_\varepsilon(\rho):=\int_0^\rho\sqrt{P_\varepsilon'(r)}\,\mathrm{d}r$. Finally, by using $\rho_{\varepsilon,k}$ itself as a test function we obtain

$$\int_0^T\int_{D_k}|\nabla\Upsilon_\varepsilon(\rho_{\varepsilon,k})|^2\,\mathrm{d}\mathcal{V}\mathrm{d}t+\frac{1}{2}\int_{D_k}\rho_{\varepsilon,k}(x,T)^2\,\mathrm{d}\mathcal{V}(x)=\frac{1}{2}\int_{D_k}\rho_0^2\,\mathrm{d}\mathcal{V}\qquad\forall T>0\,;\tag{7.3.18}$$

a similar computation ensures that in fact all $L^p(D_k)$ norms do not increase:

$$\|\rho_{\varepsilon,k}(t)\|_{L^p(D_k)} \leq \|\rho_0\|_{L^p(D_k)} \qquad \forall t>0\,, \quad \forall p\in[1,\infty]\,. \tag{7.3.19}$$

If $\hat{\rho}_{\varepsilon,k}$ is another (approximate) solution corresponding to a different nonnegative $\hat{\rho}_0 \in L^1(\mathbb{M}^n)\cap L^\infty(\mathbb{M}^n)$, the L^1-contraction property simply follows upon multiplying the differential equation satisfied by $(\rho_{\varepsilon,k}-\hat{\rho}_{\varepsilon,k})$ formally by the test function $\operatorname{sign}(\rho_{\varepsilon,k}-\hat{\rho}_{\varepsilon,k})$ and integrating: this leads to

$$\|\rho_{\varepsilon,k}(t)-\hat{\rho}_{\varepsilon,k}(t)\|_{L^1(D_k)} \leq \|\rho_0-\hat{\rho}_0\|_{L^1(D_k)} \qquad \forall t>0\,.$$

Actually, to be more rigorous, the sign function should further be approximated by regular nondecreasing functions, see [Váz07, Proposition 3.5]. We are now ready to pass to the limit into the weak formulation satisfied by each $\rho_{\varepsilon,k}$, which reads

$$\int_0^T\int_{D_k}\rho_{\varepsilon,k}\,\partial_t\eta\,\mathrm{d}\mathcal{V}\mathrm{d}t = -\int_{D_k}\rho_0(x)\,\eta(x,0)\,\mathrm{d}\mathcal{V}(x) + \int_0^T\int_{D_k}\langle\nabla P_\varepsilon(\rho_{\varepsilon,k})\,,\nabla\eta\rangle\,\mathrm{d}\mathcal{V}\mathrm{d}t \tag{7.3.20}$$

for every $T>0$ and every $\eta\in W^{1,2}((0,T);L^2(D_k))\cap L^2((0,T);W^{1,2}_0(D_k))$ such that $\eta(T)=0$. Indeed, the energy estimate (7.3.16) ensures that $\{\nabla P_\varepsilon(\rho_{\varepsilon,k})\}_{\varepsilon>0}$ weakly converges (up to subsequences) as $\varepsilon\downarrow 0$ to some vector field $\vec{w}$ in $L^2(D_k\times(0,T))$, whereas (7.3.19) yields weak convergence of $\{\rho_{\varepsilon,k}\}_{\varepsilon>0}$ for instance in $L^2(D_k\times(0,T))$ to some limit function ρ_k, still up to subsequences. On the other hand, estimates (7.3.16)–(7.3.18) guarantee that $\{\Upsilon_\varepsilon(\rho_{\varepsilon,k})\}_{\varepsilon>0}$ is locally bounded in $W^{1,2}(D_k\times(0,T))$; in particular it admits a subsequence that converges pointwise almost everywhere. Since Υ_ε, $\Upsilon_\varepsilon^{-1}$ and P_ε are continuous, monotone increasing functions converging pointwise (and therefore locally uniformly) as $\varepsilon\downarrow 0$ to their continuous limits $\Upsilon(\rho):=\int_0^\rho\sqrt{P'(r)}\,\mathrm{d}r$, Υ^{-1} and P, respectively, we can assert that also $\{\rho_{\varepsilon,k}\}_{\varepsilon>0}$ and $\{P_\varepsilon(\rho_{\varepsilon,k})\}_{\varepsilon>0}$ converge pointwise, up to subsequences. This is the key to guarantee the identification $\vec{w}=\nabla P(\rho_k)$, so that by letting $\varepsilon\downarrow 0$ in (7.3.20) we end up with

$$\int_0^T\int_{D_k}\rho_k\,\partial_t\eta\,\mathrm{d}\mathcal{V}\mathrm{d}t = -\int_{D_k}\rho_0(x)\,\eta(x,0)\,\mathrm{d}\mathcal{V}(x) + \int_0^T\int_{D_k}\langle\nabla P(\rho_k)\,,\nabla\eta\rangle\,\mathrm{d}\mathcal{V}\mathrm{d}t\,,$$

which is valid for every $T>0$ and the same type of test functions η as in (7.3.20). Note that all the above estimates pass to the limit as $\varepsilon\downarrow 0$ e.g. by lower semicontinuity, yielding

$$\int_0^T\int_{D_k}|\nabla P(\rho_k)|^2\,\mathrm{d}\mathcal{V}\mathrm{d}t + \int_{D_k}\Psi(\rho_k(x,T))\,\mathrm{d}\mathcal{V}(x) \leq \int_{D_k}\Psi(\rho_0)\,\mathrm{d}\mathcal{V} \qquad \forall T>0\,, \tag{7.3.21}$$

$$\int_0^T\int_{D_k}\zeta\,|\partial_t\Upsilon(\rho_k)|^2\,\mathrm{d}\mathcal{V}\mathrm{d}t \leq \frac{\max_{\mathbb{R}^+}|\zeta'|}{2}\int_{D_k}\Psi(\rho_0)\,\mathrm{d}\mathcal{V} \qquad \forall T>0\,, \tag{7.3.22}$$

$$\int_0^T\int_{D_k}|\nabla\Upsilon(\rho_k)|^2\,\mathrm{d}\mathcal{V}\mathrm{d}t + \frac{1}{2}\int_{D_k}\rho_k(x,T)^2\,\mathrm{d}\mathcal{V}(x) \leq \frac{1}{2}\int_{D_k}\rho_0^2\,\mathrm{d}\mathcal{V} \qquad \forall T>0\,, \tag{7.3.23}$$

$$\|\rho_k(t)\|_{L^p(D_k)} \leq \|\rho_0\|_{L^p(D_k)} \qquad \forall t>0\,, \quad \forall p\in[1,\infty]\,, \tag{7.3.24}$$

$$\|\rho_k(t)-\hat{\rho}_k(t)\|_{L^1(D_k)} \leq \|\rho_0-\hat{\rho}_0\|_{L^1(D_k)} \qquad \forall t>0\,. \tag{7.3.25}$$

At this point we are allowed to let $k\to\infty$, so that D_k will eventually become the whole manifold $\mathbb{M}^n$. By exploiting estimates (7.3.21)–(7.3.25) and reasoning similarly to the previous step, we can easily deduce that $\{\rho_k\}_{k\in\mathbb{N}}$ (extended to zero in $\mathbb{M}^n\setminus D_k$) suitably converges as $k\to\infty$ to the energy solution ρ of (7.3.1), which therefore satisfies (7.3.8), (7.3.9) and (7.3.13) (upon repeating the same procedure starting from $\hat\rho_0$), along with

$$\int_0^T\int_{\mathbb{M}^n}\zeta\,|\partial_t\Upsilon(\rho)|^2\,\mathrm{d}\mathcal{V}\mathrm{d}t\le\frac{\max_{\mathbb{R}^+}|\zeta'|}{2}\int_{\mathbb{M}^n}\Psi(\rho_0)\,\mathrm{d}\mathcal{V}\qquad\forall T>0\,,\tag{7.3.26}$$

$$\int_0^T\int_{\mathbb{M}^n}|\nabla\Upsilon(\rho)|^2\,\mathrm{d}\mathcal{V}\mathrm{d}t+\frac12\int_{\mathbb{M}^n}\rho(x,T)^2\,\mathrm{d}\mathcal{V}(x)\le\frac12\int_{\mathbb{M}^n}\rho_0^2\,\mathrm{d}\mathcal{V}\qquad\forall T>0\,.\tag{7.3.27}$$

Note that since $P\in\mathcal{C}^1([0,+\infty))$ and $\rho_0\in L^1(\mathbb{M}^n)\cap L^\infty(\mathbb{M}^n)$ the r.h.s. of (7.3.26) is surely finite. We are thus left with proving L^1-continuity, mass conservation and (7.3.11)–(7.3.12).

In order to establish the mass-conservation property, we take advantage of a recent result contained in [BS18], which ensures that under (H1) for every $R\ge1$ there exist positive constants C,γ independent of R and a nonnegative function $\phi_R\in\mathcal{C}_c^\infty(\mathbb{M}^n)$ such that $\phi_R=1$ in $B_R(o)$, $\operatorname{supp}\phi_R\subset B_{\gamma R}(o)$ (let $o\in\mathbb{M}^n$ be a fixed pole), $\phi_R\le1$ and $|\Delta\phi_R|\le C/R$. See in particular [BS18, Corollary 2.3]. So let us plug into (7.3.2) the test function $\eta(x,t)=\phi_R(x)\xi(t)$, where $\xi\in C_c^\infty([0,T))$ with $\xi(0)=1$; we obtain

$$\int_0^T\int_{\mathbb{M}^n}\rho\,\phi_R\,\xi'\,\mathrm{d}\mathcal{V}\mathrm{d}t=-\int_{\mathbb{M}^n}\rho_0\,\phi_R\,\mathrm{d}\mathcal{V}+\int_0^T\int_{\mathbb{M}^n}\xi\,\langle\nabla P(\rho)\,,\nabla\phi_R\rangle\,\mathrm{d}\mathcal{V}\mathrm{d}t\,.\tag{7.3.28}$$

If we suitably let $\xi\to\chi_{[0,T]}$ and we integrate by parts the second term in the r.h.s. of (7.3.28), we end up with

$$\int_{\mathbb{M}^n}\rho(x,T)\,\phi_R(x)\,\mathrm{d}\mathcal{V}(x)\mathrm{d}t=\int_{\mathbb{M}^n}\rho_0\,\phi_R\,\mathrm{d}\mathcal{V}+\int_0^T\int_{\mathbb{M}^n}P(\rho)\,\Delta\phi_R\,\mathrm{d}\mathcal{V}\mathrm{d}t\,.$$

By letting $R\to\infty$, exploiting the integrability properties of ρ (note that $P(\rho)\in L^1(\mathbb{M}^n\times(0,T))$) along with the above estimate on $\Delta\phi_R$ and the arbitrariness of T, we deduce (7.3.10).

As concerns L^1-continuity, as a first step we point out that it could be proved by means of an alternative construction of weak energy solutions that takes advantage of time-discretization and the Crandall-Liggett Theorem: see e.g. [FM17, Remark 3.7]. More comments on such a construction will be made in Remark 7.3.8 at the end of this section. However, in the present framework it can be obtained in a more direct fashion, at least under (H1). Indeed, if we let $\zeta\to\chi_{[0,T]}$ in (7.3.17), upon a passage to the limit as $\varepsilon\downarrow0$ and $k\to\infty$ we infer that

$$\int_0^T\int_{\mathbb{M}^n}|\partial_t\Upsilon(\rho)|^2\,\mathrm{d}\mathcal{V}\mathrm{d}t+\frac12\int_{\mathbb{M}^n}|\nabla P(\rho(x,T))|^2\,\mathrm{d}\mathcal{V}(x)\le\frac12\int_{\mathbb{M}^n}|\nabla P(\rho_0)|^2\,\mathrm{d}\mathcal{V}\ \ \forall T>0\,.$$

This in particular ensures that, at least for initial data $\rho_0\in\mathcal{C}_c^1(\mathbb{M}^n)$, the curve $t\mapsto\Upsilon(\rho(t))$ is in $W^{1,2}\big((0,T);L^2(\mathbb{M}^n)\big)$, which further guarantees that $\rho(t)\to\rho_0$ as $t\downarrow0$ in $L^1_{\mathrm{loc}}(\mathbb{M}^n)$ (recall the uniform boundedness of ρ); on the other hand, the just proved mass-conservation property implies $\|\rho(t)\|_{L^1(\mathbb{M}^n)}=\|\rho_0\|_{L^1(\mathbb{M}^n)}$ for all $t>0$, so that the convergence does occur in $L^1(\mathbb{M}^n)$. By virtue of the contraction estimate (7.3.13), the L^1-continuity of $t\mapsto\rho(t)$ at $t=0$ yields the L^1-continuity at any other time, so that in fact $\rho\in\mathcal{C}([0,+\infty);L^1(\mathbb{M}^n))$. This holds provided $\rho_0\in\mathcal{C}_c^1(\mathbb{M}^n)$: for a general initial datum $\rho_0\in L^1(\mathbb{M}^n)\cap L^\infty(\mathbb{M}^n)$, if we take a sequence $\{\rho_{j,0}\}_{j\in\mathbb{N}}\subset\mathcal{C}_c^1(\mathbb{M}^n)$ such that $\rho_{j,0}\to\rho_0$ in $L^1(\mathbb{M}^n)$, with $\rho_{j,0}\ge0$, still the contraction estimate (7.3.13) ensures that the corresponding sequence of energy solutions $\{\rho_j\}_{j\in\mathbb{N}}$ converges to ρ in $L^\infty(\mathbb{R}^+;L^1(\mathbb{M}^n))$, hence also $t\mapsto\rho(t)$ belongs to $\mathcal{C}([0,+\infty);L^1(\mathbb{M}^n))$ (so that a posteriori we have the right to write all the above estimates for *every* rather than *almost every* t or T).

Let us finally establish the approximation properties (7.3.11)–(7.3.12). Given $\varepsilon>0$, if ρ_ε is the weak energy solution of (7.3.7) then it satisfies (7.3.8) (with $P\equiv P_\varepsilon$ and $\Psi\equiv\Psi_\varepsilon$), (7.3.9), (7.3.10) and (7.3.26)–(7.3.27) (with $\Upsilon\equiv\Upsilon_\varepsilon$ and $\Psi\equiv\Psi_\varepsilon$): by proceeding as in the first part of the proof, one can easily infer that $\{\rho_\varepsilon\}_{\varepsilon>0}$ converges pointwise almost everywhere in $\mathbb{M}^n\times\mathbb{R}^+$ as $\varepsilon\downarrow 0$ to ρ, up to subsequences. This implies convergence in $L^1_{\mathrm{loc}}(\mathbb{M}^n)$ for a.e. $t\in\mathbb{R}^+$, given the uniform boundedness of $\{\rho_\varepsilon\}_{\varepsilon>0}$. In order to show that such convergence occurs at *every* t, note that by (7.3.26) the family $\{\Upsilon_\varepsilon(\rho_\varepsilon)\}_{\varepsilon>0}$ is equicontinuous with values in $L^2(\mathbb{M}^n)$, at least for times bounded away from zero:

$$\|\Upsilon_\varepsilon(\rho_\varepsilon(t))-\Upsilon_\varepsilon(\rho_\varepsilon(s))\|_{L^2(\mathbb{M}^n)}\le\sqrt{t-s}\,\|\partial_t\Upsilon_\varepsilon(\rho_\varepsilon)\|_{L^2(\mathbb{M}^n\times(s,t))}\qquad\forall t>s>0\,;$$

by the Ascoli-Arzelà theorem we then deduce that $\{\Upsilon_\varepsilon(\rho_\varepsilon(t))\}_{\varepsilon>0}$ converges locally in $L^2(\mathbb{M}^n)$ to $\Upsilon(\rho(t))$ for every $t>0$, whence the convergence of $\{\rho_\varepsilon(t)\}_{\varepsilon>0}$ in $L^1_{\mathrm{loc}}(\mathbb{M}^n)$, thanks to the just recalled uniform boundedness of $\{\rho_\varepsilon\}_{\varepsilon>0}$. Finally, the global convergence under (H1) is again a consequence of mass conservation. □

As mentioned above, a fundamental ingredient to the strategy of proof of Theorem 7.2.4 (see Section 7.4) is the smoothing effect, namely a quantitative $L^1(\mathbb{M}^n)$–$L^\infty(\mathbb{M}^n)$ regularization property of the nonlinear evolution that depends only on the L^1 norm of the initial datum. To this end we need to ask some crucial extra assumptions: the validity of the Sobolev-type inequality (H2) and a bound from below on the degeneracy of P given by the left-hand side of (H4). The proof is largely inspired from [FM17, Section 4], where a Moser-type iteration is exploited (see also references quoted therein); nevertheless, here we are also interested in keeping track of the dependence of the multiplying constants on m as $m\downarrow 1$.

Proposition 7.3.3 (Smoothing effect). *Let $\mathbb{M}^n$ ($n\ge 3$) comply with* (H2). *Let P comply with* (H3) *and the left-hand inequality in* (H4). *Let $\varepsilon>0$ and*

$\rho_0\in L^1(\mathbb{M}^n)\cap L^\infty(\mathbb{M}^n)$ be nonnegative. Then the weak energy solution ρ_ε of (7.3.7)*, where P_ε is defined by* (7.3.3)*, satisfies the* smoothing estimate

$$\|\rho_\varepsilon(t)\|_{L^\infty(\mathbb{M}^n)}\le C\left(t^{-\frac{n}{2+n(m-1)}}\,\|\rho_0\|_{L^1(\mathbb{M}^n)}^{\frac{2}{2+n(m-1)}}+\|\rho_0\|_{L^1(\mathbb{M}^n)}\right)\qquad\forall t>0\qquad(7.3.29)$$

provided

$$\|\rho_0\|_{L^\infty(\mathbb{M}^n)}\le\frac{1}{\varepsilon}\,,\qquad(7.3.30)$$

where $C\ge 1$ is a constant depending only on c_0, C_S, n and independent of m ranging in a bounded subset of $(1,+\infty)$. As a consequence, if ρ is the weak energy solution of (7.3.1) *starting from the same initial datum, there holds*

$$\|\rho(t)\|_{L^\infty(\mathbb{M}^n)}\le C\left(t^{-\frac{n}{2+n(m-1)}}\,\|\rho_0\|_{L^1(\mathbb{M}^n)}^{\frac{2}{2+n(m-1)}}+\|\rho_0\|_{L^1(\mathbb{M}^n)}\right)\qquad\forall t>0\,.\qquad(7.3.31)$$

Proof. Given $t>0$, we consider the sequence of time steps $t_j:=(1-2^{-j})t$, for all $j\in\mathbb{N}$, so that $t_0=0$ and $t_\infty=t$. Associated with $\{t_j\}_{j\in\mathbb{N}}$, we take an increasing sequence of exponents $\{p_j\}_{j\in\mathbb{N}}$ to be defined later, such that $p_0\ge 2$ and $p_\infty=\infty$. Throughout, we will work with the approximate solutions $\{\rho_{\varepsilon,k}\}_{\varepsilon>0,k\in\mathbb{N}}$ defined in the proof of Proposition 7.3.2, so that the computations we will perform below are justified. The key starting point consists of multiplying the differential equation in (7.3.7) by the (p_j-1)-th power of $\rho_{\varepsilon,k}$, integrating by parts in $D_k \times [t_j, t_{j+1}]$, using (1.3.5) (only the bound from below) and (7.3.5) along with (7.3.19) and (7.3.30), so as to obtain

$$\begin{aligned}
&\frac{4\,c_0\,m\,p_j\,(p_j-1)}{(m+p_j-1)^2}\int_{t_j}^{t_{j+1}}\int_{D_k}\left|\nabla\left(\rho_{\varepsilon,k}^{(m+p_j-1)/2}\right)\right|^2 \mathrm{d}\mathcal{V}\mathrm{d}t \\
&\le p_j\,(p_j-1)\int_{t_j}^{t_{j+1}}\int_{D_k}\rho_{\varepsilon,k}^{p_j-2}\,P_\varepsilon'(\rho_{\varepsilon,k})\,|\nabla\rho_{\varepsilon,k}|^2\,\mathrm{d}\mathcal{V}\mathrm{d}t \\
&= \|\rho_{\varepsilon,k}(t_j)\|_{p_j}^{p_j} - \|\rho_{\varepsilon,k}(t_{j+1})\|_{p_j}^{p_j} \le \|\rho_{\varepsilon,k}(t_j)\|_{p_j}^{p_j}\,.
\end{aligned} \tag{7.3.32}$$

For readability's sake, we set $\|\cdot\|_{L^p(D_k)} = \|\cdot\|_p$. Before proceeding further, it is convenient to recall (see [Bak+95, Theorem 3.1]) that the Sobolev-type inequality (H2) can equivalently be rewritten in a "Gagliardo-Nirenberg" form as

$$\begin{gathered}
\|f\|_{L^r(\mathbb{M}^n)} \le \widetilde{C}_S\left(\|\nabla f\|_{L^2(\mathbb{M}^n)} + \|f\|_{L^2(\mathbb{M}^n)}\right)^{\vartheta(s,r,n)}\|f\|_{L^s(\mathbb{M}^n)}^{1-\vartheta(s,r,n)} \quad \forall f\in W^{1,2}\cap L^s(\mathbb{M}^n) \\
\text{for every } 0<s<r\le 2^\star\,, \quad \text{where } \vartheta=\vartheta(s,r,N) := \frac{2n\,(r-s)}{r\,[2n-s(n-2)]}\in(0,1)
\end{gathered} \tag{7.3.33}$$

and $\widetilde{C}_S$ is another positive constant that can be taken independent of r, s. Taking advantage of Young's inequality, it is not difficult to show that (7.3.33) implies

$$\begin{gathered}
\|f\|_{L^r(\mathbb{M}^n)} \le \widetilde{C}_S\left(\|\nabla f\|_{L^2(\mathbb{M}^n)} + \|f\|_{L^s(\mathbb{M}^n)}\right)^{\vartheta(s,r,n)}\|f\|_{L^s(\mathbb{M}^n)}^{1-\vartheta(s,r,n)} \quad \forall f\in W^{1,2}\cap L^s(\mathbb{M}^n) \\
\text{for every } 0<s<r\le 2^\star \text{ with } s\le 2\,,
\end{gathered} \tag{7.3.34}$$

for a possibly different positive constant $\widetilde{C}_S$ as above that we do not relabel. We are now in position to handle the l.h.s. of (7.3.32) by applying (7.3.34) to the function

$$f = \rho_{\varepsilon,k}^{(m+p_j-1)/2}(t)\,,$$

which yields (we can suppose that the solution is not identically zero)

$$\begin{aligned}
&\frac{2\,c_0\,m\,p_j\,(p_j-1)}{\widetilde{C}_S^{\frac{2}{\vartheta}}\,(m+p_j-1)^2}\int_{t_j}^{t_{j+1}}\frac{\|\rho_{\varepsilon,k}(t)\|_{r(m+p_j-1)/2}^{(m+p_j-1)/\vartheta}}{\|\rho_{\varepsilon,k}(t)\|_{s(m+p_j-1)/2}^{(1-\vartheta)(m+p_j-1)/\vartheta}}\,\mathrm{d}t \\
&\le \|\rho_{\varepsilon,k}(t_j)\|_{p_j}^{p_j} + \frac{4\,c_0\,m\,p_j\,(p_j-1)}{(m+p_j-1)^2}\int_{t_j}^{t_{j+1}}\|\rho_{\varepsilon,k}(t)\|_{s(m+p_j-1)/2}^{m+p_j-1}\,\mathrm{d}t\,.
\end{aligned} \tag{7.3.35}$$

Upon making the (feasible) choices

$$s = \frac{2p_j}{m+p_j-1}\,, \qquad r = 2+\frac{2s}{n} = 2\,\frac{(n+2)p_j+n(m-1)}{n(m+p_j-1)}\,,$$

recalling the recursive definition of $\{t_j\}_{j\in\mathbb{N}}$ and using (7.3.19), from (7.3.35) we can infer that

$$\begin{aligned}&\frac{c_0\, m\, p_j\,(p_j-1)\,t}{\widetilde{C}_S^{\frac{2}{\vartheta}}\, 2^j (m+p_j-1)^2}\,\frac{\|\rho_{\varepsilon,k}(t_{j+1})\|_{p_{j+1}}^{p_{j+1}}}{\|\rho_{\varepsilon,k}(t_j)\|_{p_j}^{2p_j/n}}\\ &\le \|\rho_{\varepsilon,k}(t_j)\|_{p_j}^{p_j}+\frac{2\,c_0\,m\,p_j\,(p_j-1)\,t}{2^j(m+p_j-1)^2}\,\|\rho_{\varepsilon,k}(t_j)\|_{p_j}^{m+p_j-1},\end{aligned} \tag{7.3.36}$$

where p_j is also defined recursively by

$$p_{j+1}=\frac{n+2}{n}\,p_j+m-1 \quad\Longrightarrow\quad p_j=\left[p_0+\frac{n(m-1)}{2}\right]\left(\frac{n+2}{n}\right)^j-\frac{n(m-1)}{2} \quad \forall j\in\mathbb{N}. \tag{7.3.37}$$

From here on, we will denote by H a generic positive constant that depends only on $c_0, \widetilde{C}_S, n, p_0$ and is independent of m ranging in a bounded subset of $(1,+\infty)$, which may vary from line to line. Hence estimate (7.3.36) can be rewritten as

$$\|\rho_{\varepsilon,k}(t_{j+1})\|_{p_{j+1}}^{p_{j+1}}\le H\left(\frac{2^j}{t}\,\|\rho_{\varepsilon,k}(t_j)\|_{p_j}^{\frac{n+2}{n}\,p_j}+\|\rho_{\varepsilon,k}(t_j)\|_{p_j}^{\frac{n+2}{n}\,p_j+m-1}\right). \tag{7.3.38}$$

By combining (7.3.19), the monotonicity of $\{p_j\}_{j\in\mathbb{N}}$, interpolation and Young's inequalities, we easily obtain:

$$\|\rho_{\varepsilon,k}(t_j)\|_{p_j}\le\|\rho_0\|_\infty+\|\rho_0\|_{p_0},$$

whence from (7.3.38) there follows

$$\|\rho_{\varepsilon,k}(t_{j+1})\|_{p_{j+1}}\le H^{\frac{j+1}{p_{j+1}}}\left[t^{-1}+\left(\|\rho_0\|_\infty+\|\rho_0\|_{p_0}\right)^{m-1}\right]^{\frac{1}{p_{j+1}}}\|\rho_{\varepsilon,k}(t_j)\|_{p_j}^{\frac{n+2}{n}\frac{p_j}{p_{j+1}}}. \tag{7.3.39}$$

Iterating (7.3.39) and exploiting again (7.3.19) (in the l.h.s. of (7.3.39)) yields

$$\begin{aligned}&\|\rho_{\varepsilon,k}(t)\|_{p_{j+1}}\\ &\le H^{\frac{\sum_{h=1}^{j+1}h\left(\frac{n+2}{n}\right)^{j+1-h}}{p_{j+1}}}\left[t^{-1}+\left(\|\rho_0\|_\infty+\|\rho_0\|_{p_0}\right)^{m-1}\right]^{\frac{\sum_{h=0}^{j}\left(\frac{n+2}{n}\right)^h}{p_{j+1}}}\|\rho_0\|_{p_0}^{\left(\frac{n+2}{n}\right)^{j+1}\frac{p_0}{p_{j+1}}};\end{aligned}$$

by letting $j\to\infty$, recalling (7.3.37), we thus end up with

$$\|\rho_{\varepsilon,k}(t)\|_\infty\le H\left[t^{-1}+\left(\|\rho_0\|_\infty+\|\rho_0\|_{p_0}\right)^{m-1}\right]^{\frac{n}{2p_0+n(m-1)}}\|\rho_0\|_{p_0}^{\frac{2p_0}{2p_0+n(m-1)}},$$

whence

$$\|\rho_{\varepsilon,k}(t)\|_\infty\le H\left[t^{-\frac{n}{2p_0+n(m-1)}}+\left(\|\rho_0\|_\infty+\|\rho_0\|_{p_0}\right)^{\frac{n(m-1)}{2p_0+n(m-1)}}\right]\|\rho_0\|_{p_0}^{\frac{2p_0}{2p_0+n(m-1)}}. \tag{7.3.40}$$

At this point we need to take advantage of the following version of Young's inequality:

$$A^\theta\,B^{1-\theta}\le\epsilon\,\theta\,A+\epsilon^{-\frac{\theta}{1-\theta}}\,(1-\theta)\,B \qquad \forall A,B,\epsilon>0, \quad \forall\theta\in(0,1).$$

Upon choosing

$$A = \|\rho_0\|_\infty + \|\rho_0\|_{p_0}\,, \quad B = \|\rho_0\|_{p_0}\,, \quad \theta = \frac{n(m-1)}{2p_0 + n(m-1)}\,, \quad \epsilon = \left(H\,\theta\, 2^{1+\frac{\theta}{m-1}} \right)^{-1},$$

from (7.3.40) we infer that

$$\|\rho_{\varepsilon,k}(t)\|_\infty \le \frac{\|\rho_0\|_\infty}{2^{1+\frac{\theta}{m-1}}} + H\, t^{-\frac{\theta}{m-1}} \, \|\rho_0\|_{p_0}^{1-\theta} + \left[2^{-1-\frac{\theta}{m-1}} + H \left(H\,\theta\, 2^{1+\frac{\theta}{m-1}} \right)^{\frac{\theta}{1-\theta}} \right] \|\rho_0\|_{p_0}\,; \tag{7.3.41}$$

since θ stays bounded away from 1 and $\theta/(m-1)$ stays bounded as m ranges in a bounded subset of $(1,+\infty)$, we can equivalently rewrite (7.3.41) as

$$\|\rho_{\varepsilon,k}(t)\|_\infty \le \frac{\|\rho_0\|_\infty}{2^{1+\frac{\theta}{m-1}}} + H\, t^{-\frac{\theta}{m-1}} \, \|\rho_0\|_{p_0}^{1-\theta} + H\,\|\rho_0\|_{p_0} \qquad \forall t>0\,. \tag{7.3.42}$$

In order to remove the dependence of the r.h.s. of (7.3.42) on $\|\rho_0\|_\infty$, we can use a time-shift argument, namely for each $j \in \mathbb{N}$ we consider (7.3.42) evaluated at $t \equiv t/2^j$ with time origin shifted from 0 to $t/2^{j+1}$ (we implicitly rely on the uniqueness of energy solutions). This, along with (7.3.19), ensures that

$$\left\|\rho_{\varepsilon,k}\!\left(t/2^j\right)\right\|_\infty \le \frac{\left\|\rho_{\varepsilon,k}\!\left(t/2^{j+1}\right)\right\|_\infty}{2^{1+\frac{\theta}{m-1}}} + 2^{\frac{\theta(j+1)}{m-1}}\, H\, t^{-\frac{\theta}{m-1}} \, \|\rho_0\|_{p_0}^{1-\theta} + H\,\|\rho_0\|_{p_0} \qquad \forall j \in \mathbb{N}\,. \tag{7.3.43}$$

By iterating (7.3.43) from $j=0$ to $j=J\in\mathbb{N}$, we obtain:

$$\begin{aligned} &\|\rho_{\varepsilon,k}(t)\|_\infty \\ &\le \frac{\|\rho_0\|_\infty}{2^{\left(1+\frac{\theta}{m-1}\right)(J+1)}} + 2^{\frac{\theta}{m-1}} H\, t^{-\frac{\theta}{m-1}} \, \|\rho_0\|_{p_0}^{1-\theta} \sum_{j=0}^{J} 2^{-j} + H\,\|\rho_0\|_{p_0} \sum_{j=0}^{J} 2^{-\left(1+\frac{\theta}{m-1}\right)j}\,, \end{aligned}$$

so that taking limits as $J\to\infty$ yields

$$\|\rho_{\varepsilon,k}(t)\|_\infty \le H \left(t^{-\frac{\theta}{m-1}} \, \|\rho_0\|_{p_0}^{1-\theta} + \|\rho_0\|_{p_0} \right) \qquad \forall t>0\,. \tag{7.3.44}$$

We finally need to extend the just proved estimate to the case $p_0=1$, the one we are primarily interested in. Given any $p_0 \ge 2$ as above (fixed), let us plug the interpolation inequality

$$\|\rho_0\|_{p_0} \le \|\rho_0\|_\infty^{1-\frac{1}{p_0}} \, \|\rho_0\|_1^{\frac{1}{p_0}}$$

into (7.3.44):

$$\begin{aligned} &\|\rho_{\varepsilon,k}(t)\|_\infty \\ &\le C\,\|\rho_0\|_\infty^{\frac{2(p_0-1)}{2p_0+n(m-1)}} \left(t^{-\frac{n}{2p_0+n(m-1)}} \, \|\rho_0\|_1^{\frac{2}{2p_0+n(m-1)}} + \|\rho_0\|_\infty^{\frac{n(m-1)(p_0-1)}{p_0[2p_0+n(m-1)]}} \, \|\rho_0\|_1^{\frac{1}{p_0}} \right), \end{aligned} \tag{7.3.45}$$

for every $t>0$, where C stands for a generic positive constant as in the statement. By exploiting again a time-shift argument, it is readily seen that (7.3.45) entails, for all $j \in \mathbb{N}$,

$$\left\| \rho_{\varepsilon,k}(t/2^j) \right\|_\infty \le 2^{\frac{n(j+1)}{2p_0+n(m-1)}} \, C \left\| \rho_{\varepsilon,k}(t/2^{j+1}) \right\|_\infty^{\frac{2(p_0-1)}{2p_0+n(m-1)}} \times \left(t^{-\frac{n}{2p_0+n(m-1)}} \, \|\rho_0\|_1^{\frac{2}{2p_0+n(m-1)}} + \|\rho_0\|_\infty^{\frac{n(m-1)(p_0-1)}{p_0[2p_0+n(m-1)]}} \, \|\rho_0\|_1^{\frac{1}{p_0}} \right). \tag{7.3.46}$$

Since

$$\frac{2(p_0-1)}{2p_0+n(m-1)} \le 1 - \frac{1}{p_0} \, ,$$

a straightforward iteration of (7.3.46) ensures that

$$\begin{aligned} \|\rho_{\varepsilon,k}(t)\|_\infty &\le C \left(t^{-\frac{n}{2p_0+n(m-1)}} \, \|\rho_0\|_1^{\frac{2}{2p_0+n(m-1)}} + \|\rho_0\|_\infty^{\frac{n(m-1)(p_0-1)}{p_0[2p_0+n(m-1)]}} \, \|\rho_0\|_1^{\frac{1}{p_0}} \right)^{\frac{2p_0+n(m-1)}{2+n(m-1)}} , \\ &\le C \left(t^{-\frac{n}{2+n(m-1)}} \, \|\rho_0\|_1^{\frac{2}{2+n(m-1)}} + \|\rho_0\|_\infty^{\frac{n(m-1)(p_0-1)}{p_0[2+n(m-1)]}} \, \|\rho_0\|_1^{\frac{2p_0+n(m-1)}{p_0[2+n(m-1)]}} \right). \end{aligned} \tag{7.3.47}$$

By applying a Young-type inequality similar to the one that led us to (7.3.41), from (7.3.47) we easily deduce that

$$\begin{aligned} \|\rho_{\varepsilon,k}(t)\|_\infty \le{} & \frac{\|\rho_0\|_\infty}{2^{1+\frac{n}{2+n(m-1)}}} + C\, t^{-\frac{n}{2+n(m-1)}} \, \|\rho_0\|_1^{\frac{2}{2+n(m-1)}} \\ & + C \left(C \, \frac{n(m-1)(p_0-1)}{p_0[2+n(m-1)]} \, 2^{1+\frac{n}{2+n(m-1)}} \right)^{\frac{n(m-1)(p_0-1)}{2p_0+n(m-1)}} \|\rho_0\|_1 \\ \le{} & \frac{\|\rho_0\|_\infty}{2^{1+\frac{n}{2+n(m-1)}}} + C\, t^{-\frac{n}{2+n(m-1)}} \, \|\rho_0\|_1^{\frac{2}{2+n(m-1)}} + C \, \|\rho_0\|_1 \qquad \forall t > 0 \, . \end{aligned} \tag{7.3.48}$$

Estimate (7.3.48) is completely analogous to (7.3.42), so that by reasoning in the same fashion we end up with

$$\|\rho_{\varepsilon,k}(t)\|_{L^\infty(D_k)} \le C \left(t^{-\frac{n}{2+n(m-1)}} \, \|\rho_0\|_{L^1(D_k)}^{\frac{2}{2+n(m-1)}} + \|\rho_0\|_{L^1(D_k)} \right) \qquad \forall t > 0 \, . \tag{7.3.49}$$

Recalling the convergence results encompassed by Proposition 7.3.2, the smoothing effect (7.3.29) follows by letting $k \to \infty$ in (7.3.49), whereas (7.3.31) follows by letting $\varepsilon \downarrow 0$ in (7.3.29). □

In the next proposition we will see that every solution starting from bounded and compactly-supported datum stays with compact support, at least for short times. It is a consequence of the power degeneracy of P induced by assumption (H4) (here we need both sides), hence it is a purely *nonlinear* effect. We stress that this property will be crucial in order to show two essential facts: solutions starting from data in $\mathscr{M}_2^M(\mathbb{M}^n)$ belong to $\mathscr{M}_2^M(\mathbb{M}^n)$ for all times and they form a *continuous* curve with values in $(\mathscr{M}_2^M(\mathbb{M}^n), \mathcal{W}_2)$.

Proposition 7.3.4 (Compactness of the support)**.** *Let P comply with* (H3) *and* (H4)*. Let $\rho_0 \in L^1(\mathbb{M}^n) \cap L^\infty(\mathbb{M}^n)$ be nonnegative with* compact support*. Then there exist $t_1 > 0$ and a compact set $B \subset \mathbb{M}^n$, depending on ρ_0, m, c_0, c_1 and $\mathbb{M}^n$, such that the weak energy solution ρ to* (7.3.1) *satisfies*

$$\operatorname{supp} \rho(t) \subset B \qquad \forall t \in [0, t_1] \, . \tag{7.3.50}$$

Proof. Since $\mathbb{M}^n$ is a smooth, complete, connected and noncompact Riemannian manifold, it is well known that it admits a regular exhaustion, namely a sequence of open sets $D_k \subset \mathbb{M}^n$ such that $\overline{D}_k$ is a smooth, compact manifold with boundary (for all $k \in \mathbb{N}$) and there hold

$$\overline{D}_k \Subset D_{k+1} \qquad \text{and} \qquad \bigcup_{k=1}^{\infty} D_k = \mathbb{M}^n \, .$$

In particular, ∂D_k is a smooth $(n-1)$-dimensional, compact, orientable submanifold of $\mathbb{M}^n$, with a natural orientation given by the outward-pointing normal field w.r.t. D_k. For such a construction we refer e.g. to [Lee13, Proposition 2.28, Theorem 6.10, Propositions 15.24 and 15.33]. Given $\epsilon > 0$, let us define the set of all points inside D_k whose distance from ∂D_k is smaller than ϵ, that is

$$D_k^\epsilon := \{ x \in D_k : \ \mathsf{d}(x, \partial D_k) < \epsilon \} \, .$$

Since ∂D_k enjoys the above recalled regularity properties, if ϵ is sufficiently small then each $x \in D_k^\epsilon$ admits a unique projection $\pi(x)$ onto ∂D_k. Hence every such point is uniquely identified by the pair $\Pi(x) := (\pi(x), \delta(x))$, where $\delta(x)$ is the geodesic distance from x to $\pi(x)$ (or equivalently to ∂D_k). Moreover, the map Π is a diffeomorphism between D_k^ϵ and $\partial D_k \times (0, \epsilon)$, so that one can use $\delta = \delta(x)$ and $\pi = \pi(x)$ as coordinates that span the whole D_k^ϵ (see e.g. [Foo84]). It is not difficult to check that δ being a geodesic coordinate, the Laplacian of a regular function ϕ (defined on D_k^ϵ) that depends only on δ reads

$$\Delta \phi(\pi, \delta) = \phi''(\delta) + \mathsf{m}(\pi, \delta) \, \phi'(\delta) \qquad \forall (\pi, \delta) \in \partial D_k \times (0, \epsilon) \, , \tag{7.3.51}$$

where $\mathsf{m}(\pi, \delta)$ is also regular (in fact it is the Laplacian of the distance function itself).

Taking advantage of such framework, first of all we pick k so large that $\operatorname{supp} \rho_0 \subset D_{k-1}$ and $\epsilon > 0$ so small that, alongside with the unique-projection property, there holds $D_{k-1} \cap D_k^\epsilon = \emptyset$. Then we define

$$\Sigma_\epsilon := \Pi^{-1}(\partial D_k \times \{\epsilon\}) \, ,$$

namely the set of points inside D_k whose distance to ∂D_k is equal to ϵ, which describes a smooth submanifold having analogous properties to ∂D_k (note that, since one has the right to choose ϵ arbitrarily small, Π can smoothly be extended up to $\partial D_k \times \{\epsilon\}$). We also define Ω_ϵ to be the regular domain enclosed by Σ_ϵ. Now let us consider the Cauchy-Dirichlet problem

$$\begin{cases} \partial_t u = \Delta P(u) & \text{in } \mathbb{M}^n \setminus \Omega_\epsilon \times (0, t_1) \, , \\ u = \|\rho_0\|_\infty & \text{on } \Sigma_\epsilon \times (0, t_1) \, , \\ u = 0 & \text{on } \mathbb{M}^n \setminus \Omega_\epsilon \times \{0\} \, , \end{cases} \tag{7.3.52}$$

where $t_1 > 0$ is a small enough time to be chosen later. Since $\rho \le \|\rho_0\|_\infty$ in $\mathbb{M}^n \times \mathbb{R}^+$ and $\operatorname{supp} \rho_0 \subset \Omega_\epsilon$, it is apparent that ρ is a *subsolution* of (7.3.52). Our aim is to construct a *supersolution* which depends spatially only on δ and has compact support for all $t \in [0, t_1]$. The candidate profile is modeled after Euclidean planar *traveling waves* for the porous medium equation, see [Váz07, Section 4.3]. That is, we consider the following function:

$$\overline{u}(\delta,t) := P^{-1}\left(\left[C_1\left(C_2 t + \delta - \frac{\epsilon}{2}\right)_+\right]^{\frac{m}{m-1}}\right) \qquad \forall (\delta,t)\in(0,\epsilon]\times[0,t_1]\,, \tag{7.3.53}$$

where C_1 and C_2 are positive constants to be selected. In view of the assumptions on P, it is not difficult to deduce the following inequalities:

$$\left(\frac{v}{c_1}\right)^{\frac{1}{m}} \le P^{-1}(v) \le \left(\frac{v}{c_0}\right)^{\frac{1}{m}} \qquad \forall v\ge 0\,, \tag{7.3.54}$$

$$\left[P^{-1}\right]'(v) \ge \frac{c_0^{1-\frac{1}{m}}}{c_1\, m}\, v^{\frac{1}{m}-1} \qquad \forall v>0\,. \tag{7.3.55}$$

Clearly $\overline{u}(\delta,0)\ge 0$ and, thanks to (7.3.54),

$$\overline{u}(\epsilon,t) \ge c_1^{-\frac{1}{m}}\left[C_1\,\frac{\epsilon}{2}\right]^{\frac{1}{m-1}} \qquad \forall t\ge 0\,;$$

hence a first requirement to make sure that $\overline{u}$ complies with the boundary condition in (7.3.52) is

$$C_1 \ge \frac{2}{\epsilon}\, c_1^{\frac{m-1}{m}}\, \|\rho_0\|_\infty^{m-1}\,. \tag{7.3.56}$$

Let us now compute the derivatives of $\overline{u}$ and $P(\overline{u})$ we need:

$$\begin{aligned}\partial_t\overline{u}(\delta,t) &= C_2\, C_1^{\frac{m}{m-1}}\,\frac{m}{m-1}\left(C_2 t+\delta-\frac{\epsilon}{2}\right)_+^{\frac{1}{m-1}}\left[P^{-1}\right]'\left(\left[C_1\left(C_2 t+\delta-\frac{\epsilon}{2}\right)_+\right]^{\frac{m}{m-1}}\right)\\ &\overset{(7.3.55)}{\ge} C_2\, C_1^{\frac{1}{m-1}}\,\frac{c_0^{\frac{m-1}{m}}}{(m-1)c_1}\left(C_2 t+\delta-\frac{\epsilon}{2}\right)_+^{\frac{2-m}{m-1}}\,,\end{aligned} \tag{7.3.57}$$

$$\partial_\delta(P(\overline{u}))(\delta,t) = C_1^{\frac{m}{m-1}}\,\frac{m}{m-1}\left(C_2 t+\delta-\frac{\epsilon}{2}\right)_+^{\frac{1}{m-1}}\,, \tag{7.3.58}$$

$$\partial_{\delta\delta}(P(\overline{u}))(\delta,t) = C_1^{\frac{m}{m-1}}\,\frac{m}{(m-1)^2}\left(C_2 t+\delta-\frac{\epsilon}{2}\right)_+^{\frac{2-m}{m-1}}\,. \tag{7.3.59}$$

We pick t_1 in such a way that the distance of the support of $\overline{u}$ from ∂D_k is not smaller than $\epsilon/4$ for all $t\in[0,t_1]$, namely

$$t_1 = \frac{\epsilon}{4C_2}\,. \tag{7.3.60}$$

Let σ denote the maximum of $\mathsf{m}(\pi,\delta)$ in the region $E_\epsilon := \partial D_k\times[\epsilon/4,\epsilon]$. Because $\overline{u}$ is nondecreasing in δ and (7.3.60) ensures that the support of $\overline{u}$ lies in E_ϵ, in order to guarantee that the latter is a (weak) supersolution of the differential equation in (7.3.52) it suffices to ask that (recalling (7.3.51))

$$\partial_t\overline{u}(\delta,t) \ge \partial_{\delta\delta}P(\overline{u})(\delta,t) + \sigma\,\partial_\delta P(\overline{u})(\delta,t) \qquad \forall(\delta,t)\in[\epsilon/4,\epsilon]\times[0,t_1]\,. \tag{7.3.61}$$

Thanks to (7.3.57)–(7.3.59), after some simplifications we find that (7.3.61) holds if

$$C_2 \, \frac{c_0^{\frac{m-1}{m}}}{(m-1)c_1} \geq C_1 \, \frac{m}{(m-1)^2} \left[1 + (m-1)\sigma \left(C_2 t + \delta - \tfrac{\epsilon}{2} \right)_+ \right] \quad \text{for every} \quad (\delta, t) \in [\epsilon/4, \epsilon] \times [0, t_1] \, ,$$

the latter inequality being in turn implied by

$$C_2 \geq C_1 \, \frac{c_1 \, m}{(m-1) c_0^{\frac{m-1}{m}}} \left[1 + \frac{3(m-1)\sigma\epsilon}{4} \right] . \tag{7.3.62}$$

Hence by choosing C_1 as in (7.3.56), C_2 as in (7.3.62) and finally t_1 as in (7.3.60), we infer that (7.3.53) is indeed a supersolution of (7.3.52) (obviously extended in $\mathbb{M}^n \setminus D_k$). By comparison we can therefore assert that $\rho \leq \overline{u}$ in $\mathbb{M}^n \setminus \Omega_\epsilon \times [0, t_1]$, which yields (7.3.50) with $B = \overline{D}_k$.

As concerns the comparison principle we have just applied, let us point out that in order to justify it rigorously one should know a priori that ρ is also a strong solution, namely that it has an $L^1(\mathbb{M}^n)$ time derivative: see [Váz07, Section 8.2], we refer in particular to the analogue of [Váz07, Lemma 8.11] in our framework. On the other hand $\overline{u}$ is a strong supersolution by construction. To circumvent this issue, it is enough (for instance) to exploit the fact that ρ can always be seen as the limit of solutions ρ_j to homogeneous Dirichlet problems set up on each D_j (recall the proof of Proposition 7.3.2). Since every ρ_j is a strong solution in D_j (see e.g. [Váz07, Corollary 8.3] in the Euclidean setting) and $\overline{u}$ clearly satisfies homogeneous Dirichlet boundary conditions on ∂D_j for j large enough, one obtains $\rho_j \leq \overline{u}$ in $D_j \setminus \Omega_\epsilon \times [0, t_1]$ for every $j \in \mathbb{N}$ by proceeding as above, and then lets $j \to \infty$. □

7.3.2 Variational solutions, linearized and adjoint equation

For the purposes of proving Theorem 7.2.4, we first introduce a suitable (variational) notion of solution of the approximate problem (7.3.7) and we show its equivalence with the notion of weak energy solution discussed in the previous subsection. Hereafter we identify $\mathbb{H}\mathbb{H}$ with its dual $\mathbb{H}'$ and consider the following Hilbert triple:

$$\mathbb{V} \hookrightarrow \mathbb{H} \equiv \mathbb{H}' \hookrightarrow \mathbb{V}' .$$

Problem (7.3.7) reads

$$\frac{\mathrm{d}}{\mathrm{d}t} \rho = \Delta(P_\varepsilon(\rho)) \, , \qquad \rho(0) = \rho_0 \, , \tag{7.3.63}$$

where ρ is seen as a curve with values in $\mathbb{H}$ and, accordingly, Δ is the realization of the (self-adjoint) Laplace-Beltrami operator in $\mathbb{H}$. In agreement with the notations of Subsection 7.1, for every $\varepsilon > 0$ and $T > 0$ we recall the definition of the set $\mathcal{ND}(0, T)$ associated with P_ε:

$$\mathcal{ND}_{P_\varepsilon}(0, T) := \left\{ u \in W^{1,2}((0,T); \mathbb{H}) \cap \mathcal{C}^1([0,T]; \mathbb{V}') : \; u \geq 0 \, , \; P_\varepsilon(u) \in L^2((0,T); \mathbb{D}) \right\} .$$

Note that the nonlinearity P_ε falls within the class of functions considered in [AMS19, Subsection 3.3], in the more general framework of Dirichlet forms.

Definition 7.3.5 (Strong variational solutions)**.** *Let P comply with* (H3) *and P_ε ($\varepsilon > 0$) be defined by* (7.3.3)*. Let $\rho_0 \in \mathbb{H}$, with $\rho_0 \geq 0$, and $T > 0$. We say that a curve $\rho \in W^{1,2}((0,T); \mathbb{V}, \mathbb{V}')$, with $\rho \geq 0$, is a* strong variational solution *of* (7.3.63) *in the time interval* $(0,T)$ *if there holds*

$$- {}_{\mathbb{V}'}\big\langle \tfrac{\mathrm{d}}{\mathrm{d}t}\rho(t), \eta \big\rangle_{\mathbb{V}} = \int_{\mathbb{M}^n} \langle \nabla P_\varepsilon(\rho(t)) \, , \nabla \eta \rangle \, \mathrm{d}\mathcal{V} \qquad \textit{for a.e. } t \in (0,T) \, , \quad \forall \eta \in \mathbb{V} \, , \tag{7.3.64}$$

and $\lim_{t\downarrow 0} \rho(t) = \rho_0$ *in* $\mathbb{H}$.

We point out that Definition 7.3.5 does make sense since $\rho \in W^{1,2}((0,T);\mathbb{V},\mathbb{V}')$ implies $\rho \in \mathcal{C}([0,T];\mathbb{H})$, see [AMS19, formula (3.28)] (this is indeed a rather general fact).

The following well-posedness result is established by [AMS19, Theorem 3.4].

Proposition 7.3.6 (Existence of strong variational solutions)**.** *Let* P *comply with* (H3) *and* P_ε *(*$\varepsilon > 0$*) be defined by* (7.3.3)*. Let* $\rho_0 \in \mathbb{H}$*, with* $\rho_0 \geq 0$*, and* $T > 0$*. Then there exists a* unique *strong variational solution of* (7.3.63)*, in the sense of Definition 7.3.5. If in addition* $\rho_0 \in \mathbb{V}$ *then* $\rho \in \mathcal{ND}_{P_\varepsilon}(0,T)$*.*

Weak energy solutions and strong variational solutions in fact coincide.

Proposition 7.3.7 (Equivalent notions of solution)**.** *Let* P *comply with* (H3) *and* P_ε *(*$\varepsilon > 0$*) be defined by* (7.3.3)*. Let* $T > 0$*. Then for any nonnegative* $\rho_0 \in L^1(\mathbb{M}^n) \cap L^\infty(\mathbb{M}^n)$ *the weak energy solution of* (7.3.7) *(provided by Proposition 7.3.2) and the strong variational solution of* (7.3.63) *(provided by Proposition 7.3.6) are equal, up to* $t = T > 0$*.*

Proof. Let us denote by $\hat\rho$ the solution constructed in Proposition 7.3.6. Thanks to the integrability properties of $\hat\rho$ and the $\mathcal{C}^1$ regularity of the map $\rho \mapsto P_\varepsilon(\rho)$, we know that $\hat\rho \in L^2(\mathbb{M}^n \times (0,T))$, which is equivalent to $P_\varepsilon(\hat\rho) \in L^2(\mathbb{M}^n \times (0,T))$, and $\nabla\hat\rho \in L^2(\mathbb{M}^n \times (0,T))$, which is equivalent to $\nabla P_\varepsilon(\hat\rho) \in L^2(\mathbb{M}^n \times (0,T))$. By (7.3.64), for any curve $\eta \in W^{1,2}((0,T);W^{1,2}(\mathbb{M}^n))$ with $\eta(T) = 0$ there holds

$$- {}_{\mathbb{V}'}\big\langle \tfrac{\mathrm{d}}{\mathrm{d}t}\hat\rho(t), \eta(t) \big\rangle_{\mathbb{V}} = \int_{\mathbb{M}^n} \langle \nabla P_\varepsilon(\hat\rho(t)) \, , \nabla \eta(t) \rangle \, \mathrm{d}\mathcal{V} \qquad \text{for a.e. } t \in (0,T) \, ; \tag{7.3.65}$$

since both $\hat\rho$ and η are continuous curves with values in $L^2(\mathbb{M}^n)$, integrating (7.3.65) between $t = 0$ and $t = T$ yields

$$\int_0^T \int_{\mathbb{M}^n} \hat\rho \, \partial_t \eta \, \mathrm{d}\mathcal{V}\mathrm{d}t + \int_{\mathbb{M}^n} \rho_0(x) \, \eta(x,0) \, \mathrm{d}\mathcal{V}(x) = \int_0^T \int_{\mathbb{M}^n} \langle \nabla P_\varepsilon(\hat\rho) \, , \nabla\eta \rangle \, \mathrm{d}\mathcal{V}\mathrm{d}t \, ,$$

which shows that $\hat\rho$ is also a weak energy solution of (7.3.7) starting from ρ_0 and therefore it coincides with the one provided by Proposition 7.3.2, up to the observations made in the first part of the corresponding proof. □

Remark 7.3.8 (On possibly different constructions of weak energy solutions)**.** In Subsection 7.3.1 we used a well-established approach to prove existence of weak energy solutions of (7.3.1), which consists in the first place of solving *evolution* problems associated with nondegenerate nonlinearities on regular domains. As shown above, this technique is suitable to prove several key estimates, especially the smoothing effect of Proposition 7.3.3. Nevertheless, we mention that there exists at least another fruitful method, which relies first on solving a *discretized* version (in time) of problem (7.3.1) by means of the Crandall-Liggett Theorem (see [Váz07, Chapter 10] in the Euclidean context). This is precisely the technique employed in [AMS19, Section 3.3] to construct solutions of (7.3.1) in the general setting considered therein; the advantage of such an approach is that it also works in nonsmooth frameworks (like metric-measure spaces). However, in that case the proof of the smoothing effect is less trivial and should be investigated further (one can no longer differentiate L^p norms along the flow), for instance by taking advantage of the abstract tools developed in [CH16], which a priori work upon assuming the validity of the stronger *Euclidean* Sobolev inequality (7.2.9).

To implement the Hamiltonian approach described in the Introduction, it is necessary to study the linearization of (7.3.63) along with its formal adjoint. More precisely, in the variational setting $\mathbb{V} \hookrightarrow \mathbb{H} \hookrightarrow \mathbb{V}'$ described above, we can consider the *forward linearized* equation

$$\frac{\mathrm{d}}{\mathrm{d}t} w = \Delta\big[P_\varepsilon'(\rho)\, w\big]\,, \qquad w(0) = w_0\,, \tag{7.3.66}$$

and the *backward adjoint* equation

$$\frac{\mathrm{d}}{\mathrm{d}t} \phi = -P_\varepsilon'(\rho)\, \Delta\phi\,, \qquad \phi(T) = \phi_T\,. \tag{7.3.67}$$

Following [AMS19, Theorem 4.5], we begin with rephrasing in our setting a well-posedness result for (7.3.66). Hereafter we denote by $\mathbb{D}'$ the dual of $\mathbb{D}$, recalling that $\mathbb{H} \hookrightarrow \mathbb{V}' \hookrightarrow \mathbb{D}'$ with continuous and dense inclusions.

Theorem 7.3.9 (Forward linearized equation)**.** *Let P comply with* (H3) *and P_ε ($\varepsilon > 0$) be defined by* (7.3.3)*. Let $T > 0$. For every nonnegative $\rho \in L^2((0,T);\mathbb{H})$ and for every $w_0 \in \mathbb{V}'$, there exists a unique weak solution $w \in W^{1,2}((0,T);\mathbb{H},\mathbb{D}')$ of* (7.3.66)*, in the sense that it satisfies*

$${}_{\mathbb{V}'}\langle w(r), \theta(r)\rangle_{\mathbb{V}} - \int_0^r \int_{\mathbb{M}^n} \big[\partial_t\theta(t) + P_\varepsilon'(\rho(t))\, \Delta\theta(t)\big]\, w(t)\, \mathrm{d}\mathcal{V}\mathrm{d}t = {}_{\mathbb{V}'}\langle w_0, \theta(0)\rangle_{\mathbb{V}} \tag{7.3.68}$$

for every $r \in [0,T]$ and every $\theta \in W^{1,2}((0,T);\mathbb{D},\mathbb{H})$.

As concerns (7.3.67) we have the following result, whose proof can be found in [AMS19, Theorem 4.1].

Theorem 7.3.10 (Backward adjoint equation)**.** *Let P comply with* (H3) *and P_ε ($\varepsilon > 0$) be defined by* (7.3.3)*. Let $T > 0$. For every nonnegative $\rho \in L^2((0,T);\mathbb{H})$ and for every $\phi_T \in \mathbb{V}$, there exists a unique strong solution $\phi \in W^{1,2}((0,T);\mathbb{D},\mathbb{H})$ of* (7.3.67)*. Moreover, if $\phi_T \in L^\infty(\mathbb{M}^n) \cap \mathbb{V}$ then $\|\phi(t)\|_{L^\infty(\mathbb{M}^n)} \le \|\phi_T\|_{L^\infty(\mathbb{M}^n)}$ for every $t \in [0,T]$.*

7.4 Proof of the main results

This section is entirely devoted to proving Theorems 7.2.4 and 7.2.5. After a brief introduction to the strategy of proof of Theorem 7.2.4 (Subsection 7.4.1), we will first treat the noncompact case (Subsection 7.4.2) and then shortly address the compact case (Subsection 7.4.3). Finally, in Subsection 7.4.4 we will show that our estimate is optimal for small times, namely Theorem 7.2.5.

7.4.1 Outline of the strategy

The idea is to prove the stability estimate (7.2.6) for a suitable approximation of problem (7.2.1), passing to the limit in the approximation scheme only at the very end. Let us briefly sketch the main steps of the proof.

1. We firstly consider the "elliptic" nonlinearity P_ε as in (7.3.3) and introduce a regular initial density ρ_0 belonging to $L^\infty_c(\mathbb{M}^n) \cap \mathbb{V}$. We denote by ρ, ϕ and w the solutions of the approximated problems (7.3.7), (7.3.66) and (7.3.67), respectively (for the moment for simplicity we omit the subscript ε).

2. We estimate the derivative $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}_{\rho(t)}[\phi(t)]$ of the Hamiltonian functional defined in (2.6.21). Here it is essential to exploit the lower bound on the Ricci curvature in the Bakry-Émery form (2.6.20), which allows us to deduce that (Lemma 7.4.2)
$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\,\mathcal{E}_{\rho(t)}[\phi(t)] \ge -K \int_{\mathbb{M}^n} \Gamma(\phi(t))\, P_\varepsilon(\rho(t))\,\mathrm{d}\mathcal{V}\,.$$
We then use the smoothing effect provided by Proposition 7.3.3 to integrate the above differential inequality; this yields the estimate
$$\mathcal{E}_{\rho(t)}[\phi(t)] \ge \exp(-K\,C(t,m,n))\,\mathcal{E}_{\rho_0}[\phi(0)]\,,$$
where an explicit computation of $C(t,m,n) > 0$ is given in Lemma 7.4.3.

3. We take a pair of initial data $\rho_0^0, \rho_0^1 \in L^\infty_c(\mathbb{M}^n) \cap \mathbb{V}$ and connect them by a *regular* curve $\{\rho_0^s\}_{s\in[0,1]}$ (in the sense of Definition 2.6.4). For any ρ_0^s, hereafter $t \mapsto \rho^s(t)$ will stand for the corresponding solution of (7.3.7) and ϕ^s for a solution of (7.3.67) with $\rho \equiv \rho^s$. We then denote by $(s,x) \mapsto Q_s\varphi(x)$ the (Lipschitz) solution of the Hopf-Lax problem (2.5.14) starting from an arbitrary $\varphi \in \mathsf{Lip}_c(\mathbb{M}^n)$ and by $w^s(t) \equiv t \mapsto \frac{\mathrm{d}}{\mathrm{d}s}\rho^s(t)$ the solution of the linearized equation (7.3.66). For every $t > 0$ we compute the Wasserstein distance $\mathcal{W}_2(\rho^0(t), \rho^1(t))$ in the (Kantorovich) formulation recalled by Proposition 2.5.3 in terms of the Hamiltonian. The "duality" relation between ϕ^s and w^s (Lemma 7.4.4) guarantees that
$$\begin{aligned}&\int_{\mathbb{M}^n} Q_1\varphi\,\rho^1(t)\,\mathrm{d}\mathcal{V} - \int_{\mathbb{M}^n} \varphi\,\rho^0(t)\,\mathrm{d}\mathcal{V}\\ &= \int_0^1 \left(-\frac{1}{2}\,\mathcal{E}_{\rho^s(t)}[\phi^s(t)] + {}_{\mathbb{V}'}\langle w^s(0), \phi^s(0)\rangle_{\mathbb{V}}\right)\mathrm{d}s\,,\end{aligned}$$
where the final datum of ϕ^s is given at time $T \equiv t$ by $\phi^s(t) = Q_s\varphi$.

4. By exploiting the regularity of the curve $s \mapsto \rho_0^s\mathcal{V} =: \mu^s$, we can take advantage of the key identity
$$\int_0^1 |\dot{\mu}^s|^2\,\mathrm{d}s = \int_0^1 \mathcal{E}^*_{\rho_0^s}\left[\tfrac{\mathrm{d}}{\mathrm{d}s}\rho_0^s\right]\mathrm{d}s\,.$$
By combining the latter with the estimate obtained in Step 2 and recalling the definition (2.6.22) of the (Fenchel) *dual* Hamiltonian $\mathcal{E}^*_\rho$, we can deduce that
$$\int_{\mathbb{M}^n} Q_1\varphi\,\rho^1(t)\,\mathrm{d}\mathcal{V} - \int_{\mathbb{M}^n} \varphi\,\rho^0(t)\,\mathrm{d}\mathcal{V} \le \frac{1}{2}\exp\{K\,C(t,m,n)\} \int_0^1 |\dot{\mu}^s|^2\,\mathrm{d}s\,;$$
this is the content of Lemma 7.4.5.

5. We use Lemma 2.6.6, which ensures that the right-hand side can be made arbitrarily close to the squared Wasserstein distance between ρ_0^0 and ρ_0^1 (this in fact implies a further approximation of the initial data). As a consequence, we end up with
$$\int_{\mathbb{M}^n} Q_1\varphi\,\rho^1_\varepsilon(t)\,\mathrm{d}\mathcal{V} - \int_{\mathbb{M}^n} \varphi\,\rho^0_\varepsilon(t)\,\mathrm{d}\mathcal{V} \le \frac{1}{2}\exp\{K\,C(t,m,n)\}\,\mathcal{W}_2^2(\rho_0^0, \rho_0^1)\,,$$

where we have reintroduced the dependence on ε in view of the last passage to the limit.

6. By virtue of (7.3.12), we are allowed to first pass to the limit as $\varepsilon \downarrow 0$ and then take the supremum over all $\varphi \in \mathsf{Lip}_c(\mathbb{M}^n)$, which yields

$$\mathcal{W}_2\big(\rho^0(t), \rho^1(t)\big) \le \exp\{K\, C(t,m,n)\}\, \mathcal{W}_2\big(\rho_0^0, \rho_0^1\big) .$$

7. We exploit Proposition 7.3.4 in order to show that such solutions do belong to $\mathscr{M}_2^M(\mathbb{M}^n)$ for all times; here we apply inductively the stability estimate itself in the form $\mathcal{W}_2(\rho(t), \rho(t+\tau))$, for small $\tau > 0$, along with (2.5.6). Then, upon approximating the initial data, we show that the stability estimate extends to the whole class $\mathscr{M}_2^M(\mathbb{M}^n)$.

8. As a final step, we prove that the solutions constructed above are indeed weak Wasserstein solutions, in the sense of Definition 7.2.3. This basically follows from the smoothing effect (7.2.5) and the energy inequality (7.3.8). Uniqueness of Wasserstein solutions is also a direct consequence of the uniqueness result for weak energy solutions, together with their regularity properties.

7.4.2 The noncompact case

Throughout this whole subsection we will assume again that $\mathbb{M}^n$ is in addition noncompact and with infinite volume, hence we will carry out the proof of Theorem 7.2.4 in this case only. We will then discuss in Subsection 7.4.3 the (simple) modifications required to deal with compact manifolds.

Let ρ be a weak energy solution of (7.3.7) and let ϕ be a strong variational solution of the associated *backward adjoint* problem, according to Theorem 7.3.10. Upon recalling (2.6.21), we define the *Hamiltonian* functional as

$$\mathcal{E}_{\rho(t)}[\phi(t)] := \int_{\mathbb{M}^n} \Gamma(\phi(t))\, \rho(t)\, \mathrm{d}\mathcal{V} .$$

Following [AMS19], we firstly connect the time derivative of the Hamiltonian with the *carré du champ* operators defined in (2.6.15) and (2.6.19) (see [AMS19, Theorem 11.1 and Lemma 11.2] for a detailed proof).

Lemma 7.4.1. *Let P comply with (H3), P_ε ($\varepsilon > 0$) be defined by (7.3.3) and $T > 0$. Let $\rho \in \mathcal{ND}_{P_\varepsilon}(0,T)$ be a bounded solution of (7.3.7), provided by Proposition 7.3.6. Let $\phi \in W^{1,2}((0,T); \mathbb{D}, \mathbb{H})$ be a bounded strong solution of (7.3.67), provided by Theorem 7.3.10. Then the map $t \mapsto \mathcal{E}_{\rho(t)}[\phi(t)]$ is absolutely continuous in $[0,T]$ and satisfies the identity*

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\, \mathcal{E}_{\rho(t)}[\phi(t)] = \mathbf{\Gamma}_2[\phi(t); P_\varepsilon(\rho(t))] + \int_{\mathbb{M}^n} R(\rho(t))\, (\Delta\phi(t))^2\, \mathrm{d}\mathcal{V} \qquad a.e \text{ in } (0,T) ,$$

where

$$R(\rho) := \rho\, (P_\varepsilon)'(\rho) - P_\varepsilon(\rho) \qquad \forall \rho \ge 0 .$$

By requiring the additional assumption (H5) on the nonlinearity, we are able to exploit the curvature bound (H1) in the Bakry-Émery form (2.6.20).

Lemma 7.4.2. *Let the hypotheses of Lemma 7.4.1 hold. Assume in addition that* $\mathbb{M}^n$ *($n \ge 3$) complies with* (H1) *and* P *complies with* (H5). *Then*

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\,\mathcal{E}_{\rho(t)}[\phi(t)] \ge -K \int_{\mathbb{M}^n} \Gamma(\phi(t))\,P_\varepsilon(\rho(t))\,\mathrm{d}\mathcal{V} \qquad a.e \text{ in } (0,T)\,. \tag{7.4.1}$$

Proof. By combining Lemma 7.4.1 and the Bakry-Émery condition (2.6.20) with $f \equiv \phi(t)$ and $\rho \equiv P_\varepsilon(\rho(t))$, we deduce that

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\,\mathcal{E}_{\rho(t)}[\phi(t)] \ge &-K \int_{\mathbb{M}^n} \Gamma(\phi(t))\,P_\varepsilon(\rho(t))\,\mathrm{d}\mathcal{V} \\ &+ \int_{\mathbb{M}^n} \left[\rho(t)\,(P_\varepsilon)'(\rho(t)) - \left(1-\tfrac{1}{n}\right)P_\varepsilon(\rho(t))\right](\Delta\phi(t))^2\,\mathrm{d}\mathcal{V}\,.\end{aligned}$$

The conclusion follows upon taking advantage of (7.3.6). □

If $K > 0$ in general it is not clear how to bound the r.h.s. of (7.4.1) in terms of the Hamiltonian itself. Nevertheless, if P complies with (H4) and the Sobolev-type inequality (H2) holds, the smoothing effect provided by Proposition 7.3.3 allows us to do so.

Lemma 7.4.3. *Let* $\mathbb{M}^n$ *($n \ge 3$) comply with* (H1) *and* (H2). *Let* P *comply with* (H3), (H4) *and* (H5). *Let* $T > 0$ *and* $\rho_\varepsilon \in \mathcal{ND}_{P_\varepsilon}(0,T)$ *be the (weak energy) solution of* (7.3.7) *corresponding to some nonnegative* $\rho_0 \in L^1(\mathbb{M}^n)\cap L^\infty(\mathbb{M}^n)\cap W^{1,2}(\mathbb{M}^n)$ *with* $\|\rho_0\|_{L^1(\mathbb{M}^n)} =: M$ *(recall Proposition 7.3.7), where* P_ε *($\varepsilon > 0$) is defined by* (7.3.3). *Let* $\phi \in W^{1,2}((0,T);\mathbb{D},\mathbb{H})$ *be a bounded solution of* (7.3.67) *provided by Theorem 7.3.10. Suppose that* ε *is so small that*

$$\|\rho_0\|_{L^\infty(\mathbb{M}^n)} \le \frac{1}{\varepsilon}\,.$$

Then

$$\begin{aligned}&\mathcal{E}_{\rho_\varepsilon(t)}[\phi(t)] \ge \\ &\exp\left\{-2K\,c_1\,\mathfrak{C}_m\left[\left(tM^{m-1}\right)^{\frac{2}{2+n(m-1)}} \vee \left(tM^{m-1}\right) + \frac{\varepsilon}{c_1\mathfrak{C}_m}t\right]\right\}\mathcal{E}_{\rho_0}[\phi(0)] \quad \forall t \ge 0\,,\end{aligned} \tag{7.4.2}$$

where $C > 0$ *is the same constant appearing in* (7.3.29) *and*

$$\mathfrak{C}_m := C^{m-1}\,2^{m-2}\,[2+n(m-1)]\,. \tag{7.4.3}$$

Proof. By combining inequalities (7.3.4) and (7.4.1), we obtain:

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\,\mathcal{E}_{\rho_\varepsilon(t)}[\phi(t)] \ge -K \int_{\mathbb{M}^n} \Gamma(\phi(t))\,[P(\rho_\varepsilon(t)) + \varepsilon\rho_\varepsilon(t)]\,\mathrm{d}\mathcal{V}\,. \tag{7.4.4}$$

Thanks to Proposition 7.3.3, we know that

$$\begin{aligned}\|\rho_\varepsilon(t)\|_{L^\infty(\mathbb{M}^n)}^{m-1} &\le C^{m-1}\left(t^{-\frac{n}{2+n(m-1)}}\,\|\rho_0\|_{L^1(\mathbb{M}^n)}^{\frac{2}{2+n(m-1)}} + \|\rho_0\|_{L^1(\mathbb{M}^n)}\right)^{m-1} \\ &= C^{m-1}M^{m-1}\,g_m\!\left(t\,M^{m-1}\right) \qquad \forall t > 0\,,\end{aligned} \tag{7.4.5}$$

where

$$g_m(s) := \left(s^{-\frac{n}{2+n(m-1)}} + 1\right)^{m-1} \qquad \forall s > 0\,.$$

It is apparent that

$$g_m(s) \le \begin{cases} 2^{m-1}\, s^{-\frac{n(m-1)}{2+n(m-1)}} & \text{if } s \in (0,1)\,, \\ 2^{m-1} & \text{if } s \ge 1\,. \end{cases} \tag{7.4.6}$$

If we plug (7.4.5) in (7.4.4) and recall that $P(\rho)/\rho \le c_1\, \rho^{m-1}$, we find:

$$\begin{aligned} \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\, \mathcal{E}_{\rho_\varepsilon(t)}[\phi(t)] &\ge -\, K \int_{\mathbb{M}^n} \Gamma(\phi(t))\, \rho_\varepsilon(t) \left[c_1\, \rho_\varepsilon(t)^{m-1} + \varepsilon\right] \mathrm{d}\mathcal{V} \\ &\ge -\, K \left[c_1\, C^{m-1} M^{m-1}\, g_m\!\left(t\, M^{m-1}\right) + \varepsilon\right] \mathcal{E}_{\rho_\varepsilon(t)}[\phi(t)]\,; \end{aligned} \tag{7.4.7}$$

by integrating (7.4.7) we therefore obtain

$$\mathcal{E}_{\rho_\varepsilon(t)}[\phi(t)] \ge \exp\left\{ -2K \left(c_1\, C^{m-1} \int_0^{t\, M^{m-1}} g_m(s)\, \mathrm{d}s + \varepsilon t \right) \right\} \mathcal{E}_{\rho_0}[\phi(0)] \qquad \forall t \ge 0\,. \tag{7.4.8}$$

In order to suitably simplify (7.4.8), by exploiting (7.4.6) we easily infer that

$$\int_0^\tau g_m(s)\, \mathrm{d}s \le \begin{cases} 2^{m-1}\, \frac{2+n(m-1)}{2}\, \tau^{\frac{2}{2+n(m-1)}} & \text{if } \tau \in (0,1)\,, \\ 2^{m-1} \left[\tau + \frac{n(m-1)}{2}\right] & \text{if } \tau \ge 1\,, \end{cases}$$

which implies

$$\int_0^\tau g_m(s)\, \mathrm{d}s \le 2^{m-2}\, [2 + n(m-1)] \left(\tau^{\frac{2}{2+n(m-1)}} \vee \tau \right) \qquad \forall \tau > 0\,,$$

whence (7.4.2). □

In the following, we will connect any two (sufficiently regular) initial data ρ_0^0 and ρ_0^1 with a *regular* curve $\{\rho_0^s\}_{s\in[0,1]}$ (in the sense of Definition 2.6.4) and consider the corresponding solution $t \mapsto \rho_\varepsilon^s(t)$ of (7.3.7) with initial datum ρ_0^s, that is

$$\begin{cases} \partial_t \rho_\varepsilon^s = \Delta P_\varepsilon(\rho_\varepsilon^s) & \text{in } \mathbb{M}^n \times \mathbb{R}^+\,, \\ \rho_\varepsilon^s(0) = \rho_0^s & \text{on } \mathbb{M}^n \times \{0\}\,. \end{cases} \tag{7.4.9}$$

Reasoning as in [AMS19], we will exploit the lower bound on the Hamiltonian ensured by Lemma 7.4.3 in order to prove the stability estimate (7.2.6). We start by studying the quantity

$$s \mapsto \int_{\mathbb{M}^n} Q_s\varphi\, \rho_\varepsilon^s(t)\, \mathrm{d}\mathcal{V}\,,$$

where $\varphi \in \mathsf{Lip}_c(\mathbb{M}^n)$ is arbitrary but fixed and $[0,1] \times \mathbb{M}^n \ni (s,x) \mapsto Q_s\varphi(x)$ is the (Lipschitz and compactly-supported) solution of the Hopf-Lax problem (2.5.14). To this aim, for (almost) every $s \in (0,1)$ we also introduce the solution w^s of the linearized equation (7.3.66) starting from $\frac{\mathrm{d}}{\mathrm{d}s}\rho_0^s$:

$$\begin{cases} \partial_t w^s = \Delta[P_\varepsilon'(\rho_\varepsilon^s)\, w^s] & \text{in } \mathbb{M}^n \times \mathbb{R}^+\,, \\ w^s(0) = \frac{\mathrm{d}}{\mathrm{d}s}\rho_0^s & \text{on } \mathbb{M}^n \times \{0\}\,. \end{cases} \tag{7.4.10}$$

Thanks to Theorem 7.3.9 and Remark 2.6.5, if $\{\rho_0^s\}_{s\in[0,1]}$ is a regular curve we can guarantee that (7.4.10) admits a weak solution, at least for almost every $s \in (0,1)$. Moreover, [AMS19, Theorem 4.6] ensures that $w^s(t) = \frac{\mathrm{d}}{\mathrm{d}s}\rho_\varepsilon^s(t)$.

Lemma 7.4.4. *Let P comply with (H3) and P_ε ($\varepsilon > 0$) be defined by (7.3.3). Given a regular curve $\{\rho_0^s\}_{s\in[0,1]}$ and $T > 0$, let $\rho_\varepsilon^s \in \mathcal{ND}_{P_\varepsilon}(0,T)$ be the corresponding (weak energy) solution of (7.4.9). Then, for every $\varphi \in \mathsf{Lip}_c(\mathbb{M}^n)$ and every $t \in (0,T)$, the map $s \mapsto \int_{\mathbb{M}^n} Q_s\varphi\, \rho_\varepsilon^s(t)\, \mathrm{d}\mathcal{V}$ is Lipschitz continuous in $[0,1]$ and satisfies*

$$\frac{\mathrm{d}}{\mathrm{d}s} \int_{\mathbb{M}^n} Q_s\varphi\, \rho_\varepsilon^s(t)\, \mathrm{d}\mathcal{V} = -\frac{1}{2} \int_{\mathbb{M}^n} \Gamma(Q_s\varphi)\, \rho_\varepsilon^s(t)\, \mathrm{d}\mathcal{V} + {}_{\mathbb{V}'}\langle w^s(t), Q_s\varphi\rangle_{\mathbb{V}} \quad \text{for a.e. } s \in (0,1), \tag{7.4.11}$$

where $(s,x) \mapsto Q_s\varphi(x)$ is the (Lipschitz and compactly-supported) solution of the Hopf-Lax problem (2.5.14) and $w^s(t) = \frac{\mathrm{d}}{\mathrm{d}s}\rho_\varepsilon^s(t)$ is the weak solution of (7.4.10) provided by Theorem 7.3.9.

Moreover, if we denote by $r : (0,t) \mapsto \phi^s(r)$ the solution of the backward adjoint problem (7.3.67) corresponding to $\rho \equiv \rho_\varepsilon^s$ with final condition $\phi^s(t) = Q_s\varphi$, provided by Theorem 7.3.10, the following identities hold:

$$\begin{aligned} &{}_{\mathbb{V}'}\langle w^s(t), Q_s\varphi\rangle_{\mathbb{V}} = {}_{\mathbb{V}'}\langle w^s(t), \phi^s(t)\rangle_{\mathbb{V}} \\ &= {}_{\mathbb{V}'}\langle w^s(0), \phi^s(0)\rangle_{\mathbb{V}} = {}_{\mathbb{V}'}\left\langle \tfrac{\mathrm{d}}{\mathrm{d}s}\rho_0^s, \phi^s(0)\right\rangle_{\mathbb{V}} \quad \text{for a.e. } s \in (0,1)\,. \end{aligned} \tag{7.4.12}$$

Proof. The Lipschitz-continuity of the map $s \mapsto \int_{\mathbb{M}^n} Q_s\varphi\, \rho_\varepsilon^s(t)\, \mathrm{d}\mathcal{V}$ follows since the function $(s,x) \mapsto Q_s\varphi(x)$ is Lipschitz continuous (plus the boundedness of its support) and the Lipschitz-continuity of the curve $s \mapsto \rho_0^s$ with values in $\mathbb{V}'$ (recall Remark 2.6.5) along with the fact that the semigroup generated by (7.4.9) turns out to be also a contraction with respect to $\|\cdot\|_{\mathbb{V}'}$. For more details we refer the reader to [AMS19, Proof of Theorem 12.5]. Once we have observed this, identity (7.4.11) is a direct consequence of (2.5.14) and the equality $w^s(t) = \frac{\mathrm{d}}{\mathrm{d}s}\rho_\varepsilon^s(t)$ (for a.e. $s \in (0,1)$ independently of t), which can rigorously be proved by proceeding as in [AMS19, Theorem 4.6].

As concerns (7.4.12), it is enough to observe that it is nothing but formula (7.3.68) with $\rho \equiv \rho_\varepsilon^s$, $w \equiv w^s$ and $\theta \equiv \phi^s$ (actually with r and t interchanged). □

Lemma 7.4.5. *Let $\mathbb{M}^n$ ($n \geq 3$) comply with assumptions (H1) and (H2). Let moreover P comply with assumptions (H3), (H4), (H5) and P_ε ($\varepsilon > 0$) be defined by (7.3.3). Let ρ_ε^0 and ρ_ε^1 be any two (weak energy) solutions of (7.3.7) corresponding to the initial data ρ_0^0 and ρ_0^1, respectively, both nonnegative, belonging to $L_c^\infty(\mathbb{M}^n) \cap W^{1,2}(\mathbb{M}^n)$ and having the same mass $M > 0$. Suppose that $\{\rho_0^s\}_{s\in[0,1]}$ is any regular curve (in the sense of Definition 2.6.4) connecting ρ_0^0 with ρ_0^1, which satisfies*

$$\|\rho_0^s\|_{L^\infty(\mathbb{M}^n)} \leq \frac{1}{\varepsilon} \qquad \forall s \in [0,1]\,. \tag{7.4.13}$$

Then for every $\varphi \in \mathsf{Lip}_c(\mathbb{M}^n)$ there holds

$$\begin{aligned} &\int_{\mathbb{M}^n} Q_1\varphi\, \rho_\varepsilon^1(t)\, \mathrm{d}\mathcal{V} - \int_{\mathbb{M}^n} \varphi\, \rho_\varepsilon^0(t)\, \mathrm{d}\mathcal{V} \\ &\leq \frac{1}{2} \exp\left\{ 2K\, c_1\, \mathfrak{C}_m \left[\left(tM^{m-1}\right)^{\frac{2}{2+n(m-1)}} \vee \left(tM^{m-1}\right) + \frac{\varepsilon}{c_1 \mathfrak{C}_m} t \right] \right\} \int_0^1 |\dot{\mu}^s|^2\, \mathrm{d}s\,, \end{aligned} \tag{7.4.14}$$

where $\mu^s := \rho_0^s\, \mathcal{V}$ and $\{Q_s\varphi\}_{s\in[0,1]}$ is the (Lipschitz and compactly-supported) solution of the Hopf-Lax problem (2.5.14) and the constant $\mathfrak{C}_m$ is defined in (7.4.3).

Proof. We follow the line of proof of [AMS19, Theorem 12.5], keeping the same notations as in Lemma 7.4.4. By combining (7.4.11) and (7.4.12), we obtain:

$$\begin{aligned}&\int_{\mathbb{M}^n} Q_1\varphi\, \rho_\varepsilon^1(t)\,\mathrm{d}\mathcal{V} - \int_{\mathbb{M}^n} \varphi\, \rho_\varepsilon^0(t)\,\mathrm{d}\mathcal{V}\\ &= \int_0^1 \left(-\frac{1}{2} \int_{\mathbb{M}^n} \Gamma(\phi^s(t))\, \rho_\varepsilon^s(t)\,\mathrm{d}\mathcal{V} + {}_{\mathbb{V}'}\langle \tfrac{\mathrm{d}}{\mathrm{d}s}\rho_0^s, \phi^s(0)\rangle_{\mathbb{V}} \right)\mathrm{d}s\\ &= \int_0^1 \left(-\frac{1}{2}\,\mathcal{E}_{\rho_\varepsilon^s(t)}[\phi^s(t)] + {}_{\mathbb{V}'}\langle \tfrac{\mathrm{d}}{\mathrm{d}s}\rho_0^s, \phi^s(0)\rangle_{\mathbb{V}} \right)\mathrm{d}s\,.\end{aligned}$$

Now we can apply, at every $s \in [0,1]$, estimate (7.4.2) from Lemma 7.4.3 with $\rho_\varepsilon(t) \equiv \rho_\varepsilon^s(t)$ and $\phi(t) \equiv \phi^s(t)$, under assumption (7.4.13). This yields, upon recalling (2.6.22),

$$\begin{aligned}&\int_{\mathbb{M}^n} Q_1\varphi\, \rho_\varepsilon^1(t)\,\mathrm{d}\mathcal{V} - \int_{\mathbb{M}^n} \varphi\, \rho_\varepsilon^0(t)\,\mathrm{d}\mathcal{V} \le\\ &\int_0^1 \left({}_{\mathbb{V}'}\langle \tfrac{\mathrm{d}}{\mathrm{d}s}\rho_0^s, \phi^s(0)\rangle_{\mathbb{V}} - \frac{1}{2}\, e^{-2K\,c_1\,\mathfrak{C}_m\left[\left(tM^{m-1}\right)^{\frac{2}{2+n(m-1)}} \vee \left(tM^{m-1}\right) + \frac{\varepsilon}{c_1\mathfrak{C}_m}t\right]} \mathcal{E}_{\rho_0^s}[\phi^s(0)] \right)\mathrm{d}s\\ &= e^{2K\,c_1\,\mathfrak{C}_m\left[\left(tM^{m-1}\right)^{\frac{2}{2+n(m-1)}} \vee \left(tM^{m-1}\right) + \frac{\varepsilon}{c_1\mathfrak{C}_m}t\right]} \int_0^1 \left({}_{\mathbb{V}'}\langle \tfrac{\mathrm{d}}{\mathrm{d}s}\rho_0^s, \psi^{s,t}\rangle_{\mathbb{V}} - \frac{1}{2}\,\mathcal{E}_{\rho_0^s}\left[\psi^{s,t}\right] \right)\mathrm{d}s\\ &\le e^{2K\,c_1\,\mathfrak{C}_m\left[\left(tM^{m-1}\right)^{\frac{2}{2+n(m-1)}} \vee \left(tM^{m-1}\right) + \frac{\varepsilon}{c_1\mathfrak{C}_m}t\right]} \int_0^1 \frac{1}{2}\,\mathcal{E}^*_{\rho_0^s}\left[\tfrac{\mathrm{d}}{\mathrm{d}s}\rho_0^s\right]\mathrm{d}s\,,\end{aligned} \tag{7.4.15}$$

where we have set

$$\psi^{s,t} := e^{2K\,c_1\,\mathfrak{C}_m\left[\left(tM^{m-1}\right)^{\frac{2}{2+n(m-1)}} \vee \left(tM^{m-1}\right) + \frac{\varepsilon}{c_1\mathfrak{C}_m}t\right]}\,\phi^s(0)\,.$$

Estimate (7.4.14) thus follows from (7.4.15) in view of (2.6.27). □

In order to prove Theorem 7.2.4, we need first to approximate the geodesic connecting μ_0 and $\hat{\mu}_0$ in $(\mathscr{M}_2^M(\mathbb{M}^n), \mathcal{W}_2)$ by regular curves, let $\varepsilon \to 0$ in (7.3.7) and finally pass to the limit in the approximation of the measures μ_0 and $\hat{\mu}_0$ by bounded and compactly supported densities as in Lemma 7.4.5.

Proof of Theorem 7.2.4 (noncompact case). To begin with, we suppose that $\mu_0 = \rho_0\mathcal{V}$ and $\hat{\mu}_0 = \hat{\rho}_0\mathcal{V}$, where ρ_0 and $\hat{\rho}_0$ are initial data complying with the assumptions of Lemma 7.4.5: we will remove this hypothesis only at the very end of the proof. By virtue of Lemma 2.6.6, we know that there exists a sequence of regular curves $\{\rho_j^s\}_{j\in\mathbb{N}, s\in[0,1]}$ satisfying (2.6.23)–(2.6.26) (let $\rho^1 = \rho_0$ and $\rho^0 = \hat{\rho}_0$ according to the corresponding notations). Given $\varepsilon > 0$, if we denote by $t \mapsto (\rho_j^s)_\varepsilon(t)$ each weak energy solution of (7.3.7) starting from $\rho_0 \equiv \rho_j^s$, then by Lemma 7.4.5 we know that

$$\begin{aligned}&\int_{\mathbb{M}^n} Q_1\varphi\,(\rho_j^1)_\varepsilon(t)\,\mathrm{d}\mathcal{V} - \int_{\mathbb{M}^n} \varphi\,(\rho_j^0)_\varepsilon(t)\,\mathrm{d}\mathcal{V}\\ &\le \frac{1}{2}\exp\left\{ 2K\,c_1\,\mathfrak{C}_m\left[\left(tM^{m-1}\right)^{\frac{2}{2+n(m-1)}} \vee \left(tM^{m-1}\right) + \frac{\varepsilon}{c_1\mathfrak{C}_m}t \right] \right\} \int_0^1 \left|\dot{\mu}_j^s\right|^2 \mathrm{d}s\end{aligned} \tag{7.4.16}$$

for every $\varphi \in \mathsf{Lip}_c(\mathbb{M}^n)$, provided

$$\left\|\rho_j^s\right\|_{L^\infty(\mathbb{M}^n)} \le \frac{1}{\varepsilon} \qquad \forall s \in [0,1]\,. \tag{7.4.17}$$

Let us pass to the limit in (7.4.16) as $j \to \infty$. In the sequel, we denote by ρ_ε and $\hat\rho_\varepsilon$ the weak energy solutions of (7.3.7) starting from ρ_0 and $\hat\rho_0$, respectively. Thanks to (2.6.24), (2.6.25) (with $p = 1$) and the L^1-contraction property (7.3.13) of weak energy solutions, which guarantees that $(\rho_j^0)_\varepsilon(t) \to \rho_\varepsilon(t)$ and $(\rho_j^1)_\varepsilon(t) \to \hat\rho_\varepsilon(t)$ in $L^1(\mathbb{M}^n)$, we deduce that

$$\begin{aligned}&\int_{\mathbb{M}^n} Q_1\varphi\,\rho_\varepsilon(t)\,\mathrm{d}\mathcal{V} - \int_{\mathbb{M}^n} \varphi\,\hat\rho_\varepsilon(t)\,\mathrm{d}\mathcal{V}\\ &\le \frac{1}{2}\exp\left\{2K\,c_1\,\mathfrak{C}_m\left[\left(tM^{m-1}\right)^{\frac{2}{2+n(m-1)}} \vee \left(tM^{m-1}\right) + \frac{\varepsilon}{c_1\mathfrak{C}_m}t\right]\right\}\mathcal{W}_2^2(\rho_0,\hat\rho_0)\end{aligned} \tag{7.4.18}$$

upon requiring

$$\limsup_{j\to\infty}\,\sup_{s\in[0,1]}\left\|\rho_j^s\right\|_{L^\infty(\mathbb{M}^n)} \le \frac{1}{2\varepsilon}$$

in view of (7.4.17), which holds for ε small enough thanks to (2.6.26). We are now in position to let $\varepsilon \downarrow 0$. The r.h.s. of (7.4.18) is clearly stable as $\varepsilon \downarrow 0$. In order to pass to the limit in the l.h.s. we need to exploit Proposition 7.3.2: in particular, formula (7.3.12) ensures that $\{\rho_\varepsilon(t)\}_{\varepsilon>0}$ and $\{\hat\rho_\varepsilon(t)\}_{\varepsilon>0}$ converge in $L^1(\mathbb{M}^n)$ to $\rho(t)$ and $\hat\rho(t)$, respectively, so that (7.4.18) yields

$$\begin{aligned}&\int_{\mathbb{M}^n} Q_1\varphi\,\rho(t)\,\mathrm{d}\mathcal{V} - \int_{\mathbb{M}^n} \varphi\,\hat\rho(t)\,\mathrm{d}\mathcal{V}\\ &\le \frac{1}{2}\exp\left\{2K\,c_1\,\mathfrak{C}_m\left[\left(tM^{m-1}\right)^{\frac{2}{2+n(m-1)}} \vee \left(tM^{m-1}\right)\right]\right\}\mathcal{W}_2^2(\rho_0,\hat\rho_0)\,.\end{aligned} \tag{7.4.19}$$

If we take the supremum of the l.h.s. of (7.4.19) over all $\varphi \in \mathsf{Lip}_c(\mathbb{M}^n)$, then by virtue of Proposition 2.5.3 we obtain

$$\mathcal{W}_2(\rho(t),\hat\rho(t)) \le \exp\left\{K\,c_1\,\mathfrak{C}_m\left[\left(tM^{m-1}\right)^{\frac{2}{2+n(m-1)}} \vee \left(tM^{m-1}\right)\right]\right\}\mathcal{W}_2(\rho_0,\hat\rho_0) \quad \forall t>0\,, \tag{7.4.20}$$

namely (7.2.6) restricted to initial data $\rho_0,\hat\rho_0 \in L^\infty_c(\mathbb{M}^n)\cap W^{1,2}(\mathbb{M}^n)$. It is apparent that estimate (7.4.20) remains true in the wider class $\rho_0,\hat\rho_0 \in L^1(\mathbb{M}^n)\cap L^\infty(\mathbb{M}^n)\cap\mathscr{M}_2^M$: indeed by local regularization and a standard truncation argument, one can pick sequences of nonnegative initial data of mass M belonging to $L^\infty_c(\mathbb{M}^n)\cap W^{1,2}(\mathbb{M}^n)$ which converge to ρ_0 and $\hat\rho_0$, respectively, both in $L^1(\mathbb{M}^n)$ and in $(\mathscr{M}_2^M(\mathbb{M}^n),\mathcal{W}_2)$ (recall Proposition 2.5.2). Thanks again to (7.3.13), i.e. the stability of solutions in $L^1(\mathbb{M}^n)$, this suffices to pass to the limit in (7.4.19) and hence in (7.4.20).

We still have to prove that $\rho(t) \in \mathscr{M}_2^M(\mathbb{M}^n)$ for all $t > 0$, since the mass-conservation property (7.3.10) only ensures that $\rho(t) \in \mathscr{M}^M(\mathbb{M}^n)$. To this aim, we take advantage of Proposition 7.3.4: from the latter we know that if $\rho_0 \in L^\infty_c(\mathbb{M}^n)$ then the weak energy solution $\rho(t)$ of (7.3.1) stays (uniformly) bounded with (uniform) compact support in a suitable time interval $[0,t_1]$, so that in particular $\rho(t) \in \mathscr{M}_2^M(\mathbb{M}^n)$ for all $t \in [0,t_1]$. Let $\tau \in (0,t_1]$. Since $\{\rho(t+\tau)\}_{t\ge0}$ is the weak energy solution of (7.3.1) starting from $\rho(\tau) \in L^1(\mathbb{M}^n)\cap L^\infty(\mathbb{M}^n)\cap\mathscr{M}_2^M(\mathbb{M}^n)$, estimate (7.4.20) applied to $\hat\rho(t) = \rho(t+\tau)$ guarantees that

$$\begin{aligned}&\mathcal{W}_2(\rho(t),\rho(t+\tau))\\ &\le \exp\left\{K\,c_1\,\mathfrak{C}_m\left[\left(tM^{m-1}\right)^{\frac{2}{2+n(m-1)}} \vee \left(tM^{m-1}\right)\right]\right\}\mathcal{W}_2(\rho_0,\rho(\tau)) < \infty \quad \forall t>0\,,\end{aligned} \tag{7.4.21}$$

whence $\rho(t) \in \mathscr{M}_2^M(\mathbb{M}^n)$ also for all $t \in (t_1, 2t_1]$ upon recalling (2.5.6). It is then clear how one can set up an induction procedure to establish that in fact $\rho(t) \in \mathscr{M}_2^M(\mathbb{M}^n)$

for *all* $t>0$. Furthermore, $\rho \in \mathcal{C}([0,+\infty); \mathscr{M}_2^M(\mathbb{M}^n))$. Indeed, the just mentioned property of compactness of the support for short times and the L^1-continuity ensured by Proposition 7.3.2 easily imply, along with Proposition 2.5.2, that

$$\lim_{t\downarrow 0} \mathcal{W}_2(\rho(t),\rho_0) = 0 \, . \tag{7.4.22}$$

Hence by combining (7.4.21) (understood for all $t,\tau>0$) and (7.4.22), we deduce that for every $t_0>0$ there holds

$$\begin{aligned} &\lim_{t\to t_0} \mathcal{W}_2(\rho(t),\rho(t_0)) \\ &\le \exp\left\{ K\, c_1\, \mathfrak{C}_m \left[\left(t_0 M^{m-1}\right)^{\frac{2}{2+n(m-1)}} \vee \left(t_0 M^{m-1}\right) \right] \right\} \lim_{t\to t_0} \mathcal{W}_2(\rho(|t-t_0|),\rho_0) = 0 \, . \end{aligned}$$

We have therefore shown the validity of Theorem 7.2.4 under the additional assumptions $\mu_0=\rho_0\mathcal{V}$ and $\hat\mu_0 = \hat\rho_0 \mathcal{V}$ with $\rho_0,\hat\rho_0 \in L^\infty_c(\mathbb{M}^n)\cap W^{1,2}(\mathbb{M}^n)$. In order to be able to deal with general initial data as in the statement, first of all we take a sequence of nonnegative functions $(\rho_{j,0},\hat\rho_{j,0}) \in [L^\infty_c(\mathbb{M}^n)\cap W^{1,2}(\mathbb{M}^n)]^2$ of mass M such that

$$\lim_{j\to\infty} \rho_{j,0} = \mu_0 \qquad \text{and} \qquad \lim_{j\to\infty} \hat\rho_{j,0} = \hat\mu_0 \qquad \text{in } \left(\mathscr{M}_2^M(\mathbb{M}^n), \mathcal{W}_2\right), \tag{7.4.23}$$

which exists as a consequence of Definition 2.6.4, Remark 2.6.5 and Lemma 2.6.6 (only applied at the endpoints $s=0,1$): the additional property of the compactness of the support can be obtained again by a straightforward truncation argument. Estimate (7.4.20) applied to the corresponding sequences of solutions, which we denote by $\{(\rho_j,\hat\rho_j)\}_{j\in\mathbb{N}}$, yields

$$\begin{aligned} \mathcal{W}_2(\rho_j(t),\rho_i(t)) &\le \exp\left\{ K\, c_1\, \mathfrak{C}_m \left[\left(t M^{m-1}\right)^{\frac{2}{2+n(m-1)}} \vee \left(t M^{m-1}\right) \right] \right\} \mathcal{W}_2(\rho_{j,0},\rho_{i,0}) \, , \\ \mathcal{W}_2(\hat\rho_j(t),\hat\rho_i(t)) &\le \exp\left\{ K\, c_1\, \mathfrak{C}_m \left[\left(t M^{m-1}\right)^{\frac{2}{2+n(m-1)}} \vee \left(t M^{m-1}\right) \right] \right\} \mathcal{W}_2(\hat\rho_{j,0},\hat\rho_{i,0}) \, , \end{aligned} \tag{7.4.24}$$

for every $t>0$ and $i,j\in\mathbb{N}$, whereas the smoothing effect (7.3.31) ensures that

$$\|\rho_j(t)\|_{L^\infty(\mathbb{M}^n)} \vee \|\hat\rho_j(t)\|_{L^\infty(\mathbb{M}^n)} \le C\left(t^{-\frac{n}{2+n(m-1)}} M^{\frac{2}{2+n(m-1)}} + M \right) \quad \forall t>0\, , \ \forall j \in \mathbb{N} \, . \tag{7.4.25}$$

From (7.4.23) and (7.4.24) we infer that $\{\rho_j\}_{j\in\mathbb{N}}$ and $\{\hat\rho_j\}_{j\in\mathbb{N}}$ are Cauchy sequences in the space $\mathcal{C}([0,T];(\mathscr{M}_2^M(\mathbb{M}^n),\mathcal{W}_2))$ for every $T>0$, hence they converge to two corresponding curves ρ and $\hat\rho$, respectively, both in $\mathcal{C}([0,T];(\mathscr{M}_2^M(\mathbb{M}^n),\mathcal{W}_2))$ for all $T>0$. By construction estimates (7.4.25) and (7.4.20) (applied to $\rho\equiv\rho_j$ and $\hat\rho\equiv\hat\rho_j$) are preserved at the limit, ensuring the validity of (7.2.5)–(7.2.6). We are thus left with proving that ρ and $\hat\rho$ are indeed Wasserstein solutions of (7.2.1) in the sense of Definition 7.2.3, i.e. they comply with (7.2.3) and (7.2.4). Of course it is enough to show it for ρ only. Since the latter satisfies (7.2.5) and $\|\rho(t)\|_{L^1(\mathbb{M}^n)} = M$ for all $t>0$, the first property in (7.2.3) is trivially fulfilled. In order to establish the second one and (7.2.4), we take advantage of the energy estimate (7.3.8) applied to each $\rho\equiv\rho_j$ (with time origin shifted from 0 to $\tau\in(0,T)$) combined with (H4) and (7.4.25), which yield

$$\begin{aligned}&\int_\tau^T \int_{\mathbb{M}^n} |\nabla P(\rho_j)|^2 \, \mathrm{d}\mathcal{V}\mathrm{d}t + \int_{\mathbb{M}^n} \Psi(\rho_j(x,T)) \, \mathrm{d}\mathcal{V}(x) \\ \le & \frac{c_1}{m+1} \int_{\mathbb{M}^n} \rho_j(x,\tau)^{m+1} \, \mathrm{d}\mathcal{V}(x) \le \frac{c_1 \, C^m M}{m+1} \left(\tau^{-\frac{n}{2+n(m-1)}} M^{\frac{2}{2+n(m-1)}} + M \right)^m . \end{aligned} \tag{7.4.26}$$

Starting from (7.4.26), using in a similar way the analogues of (7.3.26)–(7.3.27) with $\rho \equiv \rho_j$ and the time origin shifted from 0 to τ, one can reason as in the proof of Proposition 7.3.2 to deduce that $\{\rho_j\}_{j\in\mathbb{N}}$ converges to ρ and $\{\nabla P(\rho_j)\}_{j\in\mathbb{N}}$ converges to $\nabla P(\rho)$ weakly in $L^2(\mathbb{M}^n \times (\tau,T))$ as $j \to \infty$, whence the validity of (7.2.4) upon passing to the limit in the weak formulation satisfied by every ρ_j.

Finally, the uniqueness of Wasserstein solutions is a simple consequence of the continuity in $(\mathscr{M}_2^M(\mathbb{M}^n), \mathcal{W}_2)$ down to $t=0$ and the uniqueness of weak energy solutions (Proposition 7.3.2). Indeed, if ρ and $\hat\rho$ are two Wasserstein solutions starting from the same initial datum, they can be seen as weak energy solutions starting from the initial data $\rho(\tau)$ and $\hat\rho(\tau)$, respectively, for every $\tau>0$. In particular, there holds

$$\begin{aligned}&\mathcal{W}_2(\rho(t),\hat\rho(t)) \\ \le & \exp\left\{ K \, c_1 \, \mathfrak{C}_m \left[\left(t M^{m-1} \right)^{\frac{2}{2+n(m-1)}} \vee \left(t M^{m-1} \right) \right] \right\} \mathcal{W}_2(\rho(\tau),\hat\rho(\tau)) \quad \forall t > \tau > 0 \, , \end{aligned} \tag{7.4.27}$$

whence $\mathcal{W}_2(\rho(t),\hat\rho(t)) = 0$ upon letting $\tau \downarrow 0$ in (7.4.27). □

7.4.3 The compact case

If $\mathbb{M}^n$ is a *compact* manifold, the construction of the Wasserstein solutions of (7.2.1) performed in Subsection 7.3.1 is in fact easier with respect to the one performed in the noncompact case. Indeed, in the proofs of Propositions 7.3.2 and 7.3.3, there is no need to fill $\mathbb{M}^n$ with a regular exhaustion $\{D_k\}_{k\in\mathbb{N}}$: it is enough to solve the approximate problems (i.e. the ones associated with the nonlinearity P_ε) directly on the compact manifold, where integrations by parts are always justified. Moreover, mass conservation is plain because space-constant functions are admissible test functions in the weak formulation (7.3.2). The compact-support property established in Proposition 7.3.4 is clearly for free.

As concerns the variational framework considered in Subsection 7.3.2, some less trivial modifications have to be implemented. That is, one defines the space

$$\mathbb{V}'_{\mathcal{E}} := \left\{ \ell \in \mathbb{V}' : \left| {}_{\mathbb{V}'}\langle \ell, f \rangle_{\mathbb{V}} \right| \le C \sqrt{\mathcal{E}(f)} \ \text{ for every } f \in \mathbb{V} \, , \ \text{for some } C>0 \right\}$$

endowed with the norm

$$\|\ell\|_{\mathbb{V}'_{\mathcal{E}}} := \sup_{f\in\mathbb{V}: \ \mathcal{E}(f)\neq 0} \frac{\left| {}_{\mathbb{V}'}\langle \ell, f \rangle_{\mathbb{V}} \right|}{\sqrt{\mathcal{E}(f)}} \, ,$$

and the space

$$\mathbb{D}'_{\mathcal{E}} := \left\{ \ell \in \mathbb{D}' : \left| {}_{\mathbb{D}'}\langle \ell, f \rangle_{\mathbb{D}} \right| \le C \, \|\Delta f\|_{\mathbb{H}} \ \text{ for every } f \in \mathbb{D} \, , \ \text{for some } C>0 \right\} ,$$

endowed with the norm

$$\|\ell\|_{\mathbb{D}'_{\mathcal{E}}} := \sup_{f\in\mathbb{D}: \ \Delta f\neq 0} \frac{\left| {}_{\mathbb{D}'}\langle \ell, f \rangle_{\mathbb{D}} \right|}{\|\Delta f\|_{\mathbb{H}}} \, .$$

Upon replacing $\mathbb{V}'$ with $\mathbb{V}'_{\mathcal{E}}$ and $\mathbb{D}'$ with $\mathbb{D}'_{\mathcal{E}}$, respectively, the results stated in Subsections 7.3.2 and 7.4.2 continue to hold. Here we refer again to the machinery developed in [AMS19].

We point out that, in view of the standard Dirichlet form we have dealt with, the only reason why $\mathbb{V}'_{\mathcal{E}}$ and $\mathbb{D}'_{\mathcal{E}}$ do not coincide with $\mathbb{V}'$ and $\mathbb{D}'$, respectively, is that in the compact case the kernel of the Dirichlet energy functional $\mathcal{E} : \mathbb{H} \to [0, +\infty]$ coincides with the set of constant functions, hence is nontrivial. In fact $\mathbb{V}'_{\mathcal{E}}$ and $\mathbb{D}'_{\mathcal{E}}$ turn out to be identified as those elements of $\mathbb{V}'$ and $\mathbb{D}'$, respectively, that vanish on constant functions. On the contrary, in the noncompact case there holds

$$\mathcal{E}(f) = 0 \quad \text{and} \quad f \in \mathbb{H} \quad \Longrightarrow \quad f = 0$$

provided $\mathcal{V}(\mathbb{M}^n) = \infty$, which is always true if (H2) is satisfied.

7.4.4 Optimality for small times

In what follows, even if the discussion could in principle be made more general, we will restrict ourselves to $\mathbb{M}^n = \mathbb{H}^n_K$, that is the n-dimensional hyperbolic space of constant sectional curvature $\mathrm{Sec} = -K$. The key starting point to show optimality is the next delicate result, inspired by [Oll09, Proposition 6].

Lemma 7.4.6. *Let $K > 0$, $x \in \mathbb{H}^n_K$ and v be a unit tangent vector of $T_x \mathbb{H}^n_K$. Let $r, \delta > 0$. Denote by $v^\perp \subset T_x \mathbb{H}^n_K$ the orthogonal subspace to v and set $\mathrm{E} := \exp_x v^\perp \subset \mathbb{H}^n_K$. Let $w \in v^\perp$ be another unit tangent vector. Consider the point $y := \exp_x \delta v$ and set $w' := I^x_y(w)$, where $I^x_y : T_x \mathbb{H}^n_K \to T_y \mathbb{H}^n_K$ stands for the parallel-transport map along the geodesic $t \mapsto \exp_x tv$. Then*

$$\mathsf{d}\big(\exp_y rw', \mathrm{E}\big) = \delta \left(1 + \frac{K}{2}\, r^2 + O\big(r^3 + \delta r^2\big) \right) \qquad \text{as } (r, \delta) \to 0\,. \tag{7.4.28}$$

More in general, if $u' \in T_y \mathbb{H}^n_K$ is a unit tangent vector, then

$$\begin{aligned} \mathsf{d}\big(\exp_y ru', \mathrm{E}\big) = {} & \delta \left(1 + \frac{K}{2}\, r^2 \sin^2\!\alpha(u', I^x_y(v)) + O(r^3 + \delta r^2) \right) \\ & + r \cos\alpha(u', I^x_y(v)) + O(r^3) \qquad \text{as } (r, \delta) \to 0\,, \end{aligned} \tag{7.4.29}$$

where $\alpha(\cdot,\cdot) \in [0, \pi]$ denotes the angle between unit vectors in $T_y \mathbb{H}^n_K$. In all the above identities, the remainder terms $O(\cdot)$ can be considered independent of the chosen tangent vectors.

Proof. The expansion of formula (7.4.28) is exactly what is proved in [Oll09, Section 8]. Consider now a general unit tangent vector $u' \in T_y \mathbb{H}^n_K$. Let us denote by $P_v(ru')$ and $P_{v^\perp}(ru')$ the projections, in the tangent space $T_y \mathbb{H}^n_K$, of the vector ru' on the subspace generated by $I^x_y(v)$ and on its orthogonal subspace $I^x_y(v^\perp)$, respectively. Clearly, we have:

$$\big|P_v(ru')\big| = \big|r \cos\alpha(u', I^x_y(v))\big|\,, \qquad \big|P_{v^\perp}(ru')\big| = \big|r \sin\alpha(u', I^x_y(v))\big| \tag{7.4.30}$$

and

$$P_v(ru') \perp P_{v^\perp}(ru')\,, \qquad P_v(ru') + P_{v^\perp}(ru') = ru'\,. \tag{7.4.31}$$

In agreement with [Gav07], we put

$$\exp_y(P_v(ru')\,, P_{v^\perp}(ru')) := \exp_{\exp_y P_v(ru')} \left[I^y_{\exp_y P_v(ru')}\big(P_{v^\perp}(ru')\big) \right].$$

Thanks to (7.4.30) and (7.4.31), we can apply (7.4.28) with y replaced by $\exp_y P_v(ru')$ and rw' replaced by the vector $I^y_{\exp_y P_v(ru')}(P_{v^\perp}(ru'))$ (so that δ is replaced by $\delta + r\cos\alpha(u', I^x_y(v))$ and r replaced by $|r\sin\alpha(u', I^x_y(v))|$), which yields

$$\begin{aligned} &\mathsf{d}\big(\exp_y(P_v(ru'), P_{v^\perp}(ru')), \mathrm{E}\big) \\ &= \delta\left(1 + \frac{K}{2}\, r^2 \sin^2\!\alpha(u', I^x_y(v)) + O(r^3 + \delta r^2)\right) + r\cos\alpha(u', I^x_y(v)) + O(r^3)\,. \end{aligned} \tag{7.4.32}$$

In order to establish (7.4.29), first of all we take advantage of the triangle inequality, so as to obtain

$$\begin{aligned} &\big|\mathsf{d}(\exp_y ru', \mathrm{E}) - \mathsf{d}\big(\exp_y(P_v(ru')\,, P_{v^\perp}(ru')), \mathrm{E}\big)\big| \\ \leq&\,\mathsf{d}\big(\exp_y ru', \exp_y(P_v(ru')\,, P_{v^\perp}(ru'))\big)\,. \end{aligned} \tag{7.4.33}$$

Still in agreement with [Gav07], we denote by $h_y(P_v(ru')\,, P_{v^\perp}(ru'))$ the unique vector of $T_y\,\mathbb{H}^n_K$ such that

$$\exp_y\big(h_y(P_v(ru')\,, P_{v^\perp}(ru'))\big) = \exp_y(P_v(ru')\,, P_{v^\perp}(ru'))\,;$$

on the other hand, by virtue of [Gav07, formula (3)] there holds

$$\big|h_y(P_v(ru')\,, P_{v^\perp}(ru')) - ru'\big| = O(r^3)\,,$$

so that

$$\mathsf{d}\big(\exp_y ru', \exp_y(P_v(ru')\,, P_{v^\perp}(ru'))\big) = O(r^3) \tag{7.4.34}$$

upon recalling the well-known fact that the Riemannian distance locally can be replaced by the Euclidean distance up an error of order $O(r^3)$ (see e.g. [Vil09, formula (14.1)]). Estimate (7.4.29) then follows from (7.4.32), (7.4.33) and (7.4.34). □

Taking advantage of Lemma 7.4.6, we are able to prove a lower bound for the Wasserstein distance between suitable radially-symmetric probability densities in $\mathbb{H}^n_K$.

Lemma 7.4.7. *Let $K > 0$ and $\{\rho^\epsilon\}_{\epsilon\in(0,1)}$ be a family of (continuous) radially-symmetric probability densities in $\mathbb{H}^n_K$, i.e. each $\rho^\epsilon : [0, +\infty) \mapsto [0, +\infty)$ satisfies*

$$\frac{\left|\mathbb{S}^{n-1}\right|}{K^{\frac{n-1}{2}}} \int_0^{+\infty} \rho^\epsilon(r)\sinh\!\left(\sqrt{K}r\right)^{n-1} \mathrm{d}r = 1 \qquad \forall \epsilon \in (0,1)\,. \tag{7.4.35}$$

Suppose in addition that there exist some $\theta \in (0,1)$ and constants $C_1, C_2 > 0$ (independent of ϵ) such that

$$\frac{C_1}{\epsilon^n}\,\chi_{[0,\theta\epsilon]}(r) \leq \rho^\epsilon(r) \leq \frac{C_2}{\epsilon^n}\,\chi_{[0,\epsilon]}(r) \qquad \forall \epsilon \in (0,1)\,, \quad \forall r \geq 0\,. \tag{7.4.36}$$

Let $x, y \in \mathbb{H}^n_K$ with $\mathsf{d}(x,y) =: \delta > 0$ and consider the probability measures μ^ϵ_x and μ^ϵ_y obtained by centering ρ^ϵ at x and y, respectively. That is, put $\mu^\epsilon_x := \rho^\epsilon(\mathsf{d}(\cdot, x))\mathcal{V} \in \mathscr{P}(\mathbb{H}^n_K)$ and $\mu^\epsilon_y := \rho^\epsilon(\mathsf{d}(\cdot, y))\mathcal{V} \in \mathscr{P}(\mathbb{H}^n_K)$.

Then there exist constants $\overline{\delta} = \overline{\delta}(n, K, C_1, C_2, \theta) > 0$ and $\kappa = \kappa(n, C_1, C_2, \theta) > 0$ such that, if $\delta \in (0, \overline{\delta})$,

$$\mathcal{W}_2(\mu^\epsilon_x, \mu^\epsilon_y) \geq \delta\left(1 + \kappa\, K\, \epsilon^2\right) \qquad \forall \epsilon \in (0, \overline{\epsilon})\,, \tag{7.4.37}$$

where $\overline{\epsilon} = \overline{\epsilon}(\delta, n, K, C_1, C_2, \theta) \in (0,1)$.

Proof. For simplicity we assume $K = 1$ and set $\mathbb{H}^n := \mathbb{H}^n_1$, since the modifications in order to deal with a general $K > 0$ are inessential. So, let $v \in T_x\mathbb{H}^n$ be the unit vector such that $\exp_x \delta v = y$. Let $i : \mathbb{R}^n \to T_x\mathbb{H}^n$ be an isometric isomorphism that preserves orientation. As in Lemma 7.4.6, we denote by I^x_y the parallel-transport map between $T_x\mathbb{H}^n$ and $T_y\mathbb{H}^n$ along the geodesic $t \mapsto \exp_x tv$. We then define the maps $\varphi_x : \mathbb{R}^n \to \mathbb{H}^n$ and $\varphi_y : \mathbb{R}^n \to \mathbb{H}^n$ as follows:

$$\varphi_x := \exp_x \circ\, i\,, \qquad \varphi_y := \exp_y \circ I^x_y \circ i\,.$$

First of all, we normalize ρ^ϵ in such a way that it is a probability measure on $\mathbb{R}^n$, namely we set

$$\rho^\epsilon_E(r) := h(\epsilon)\,\rho^\epsilon(r) \qquad \forall r \ge 0$$

with

$$h(\epsilon) := \frac{1}{1 - \int_0^\epsilon \rho^\epsilon(r)\left(\sinh(r)^{n-1} - r^{n-1}\right) \mathrm{d}r} = 1 + O(\epsilon^2)\,, \tag{7.4.38}$$

where we used (7.4.35) and (7.4.36). Hence we put $\mu^\epsilon_E := \rho^\epsilon_E(|\cdot|)\mathscr{L}^n$, the symbol $\mathscr{L}^n$ standing for the Lebesgue measure on $\mathbb{R}^n$. Now we push forward the probability measure μ^ϵ_E on $\mathbb{H}^n$ by means of the maps φ_x and φ_y:

$$\hat{\mu}^\epsilon_x := (\varphi_x)_\sharp\, \mu^\epsilon_E\,, \qquad \hat{\mu}^\epsilon_y := (\varphi_y)_\sharp\, \mu^\epsilon_E\,. \tag{7.4.39}$$

It is possible to show that $\hat{\mu}^\epsilon_x$ and $\hat{\mu}^\epsilon_y$ are absolutely continuous w.r.t. to μ^ϵ_x and μ^ϵ_y, respectively, in a quantitative way; more precisely, there exist bounded functions $f^\epsilon_x : \mathbb{H}^n \to \mathbb{R}$ and $f^\epsilon_y : \mathbb{H}^n \to \mathbb{R}$ such that

$$\mathrm{d}\mu^\epsilon_x = \left(1 + \epsilon^2 f^\epsilon_x\right) \mathrm{d}\hat{\mu}^\epsilon_x\,, \qquad \mathrm{d}\mu^\epsilon_y = \left(1 + \epsilon^2 f^\epsilon_y\right) \mathrm{d}\hat{\mu}^\epsilon_y \tag{7.4.40}$$

and

$$\int_{\mathbb{H}^n} f^\epsilon_x\, \mathrm{d}\hat{\mu}^\epsilon_x = \int_{\mathbb{H}^n} f^\epsilon_y\, \mathrm{d}\hat{\mu}^\epsilon_y = 0\,. \tag{7.4.41}$$

Indeed, by construction φ_x and φ_y preserve radial lengths and angles. As a consequence, both $\hat{\mu}^\epsilon_x$ and $\hat{\mu}^\epsilon_y$ are represented on $\mathbb{H}^n$ by the same radial density $\hat{\rho}^\epsilon$ via the relation

$$\hat{\rho}^\epsilon(r)\, \sinh(r)^{n-1} = \rho^\epsilon_E(r)\, r^{n-1} = h(\epsilon)\, \rho^\epsilon(r)\, r^{n-1} \qquad \forall r \in (0, \epsilon)\,,$$

whence

$$\begin{aligned} \rho^\epsilon(r) &= \frac{\sinh(r)^{n-1}}{h(\epsilon)\, r^{n-1}}\, \hat{\rho}^\epsilon(r) = \left(1 + \epsilon^2\, \frac{\sinh(r)^{n-1} - h(\epsilon)\, r^{n-1}}{\epsilon^2\, h(\epsilon)\, r^{n-1}}\right) \hat{\rho}^\epsilon(r) \\ &=: \left(1 + \epsilon^2 f^\epsilon(r)\right) \hat{\rho}^\epsilon(r) \qquad \forall r \in (0, \epsilon) \end{aligned}$$

and therefore (7.4.40) holds with $f^\epsilon_x(\cdot) = f^\epsilon(\mathsf{d}(\cdot, x))$ and $f^\epsilon_y(\cdot) = f^\epsilon(\mathsf{d}(\cdot, y))$. Note that, in view of (7.4.38) and a standard Taylor expansion of $\sinh(r)$, the function f^ϵ is uniformly bounded by a constant that depends only on n and C_2. On the other hand, identity (7.4.41) just follows by the fact that μ^ϵ_x, $\hat{\mu}^\epsilon_x$, μ^ϵ_y, $\hat{\mu}^\epsilon_y$ are all probability measures.

Let E_0 and E_1 be the two disjoint, open, connected components in $\mathbb{H}^n$ separated by E, the latter set being defined as in Lemma 7.4.6. Assume for convenience that E_1 contains the point y. In order to prove (7.4.37), as in [Oll09, Section 8] we choose the following 1-Lipschitz function $g : \mathbb{H}^n \to \mathbb{R}$:

$$g(z) := \begin{cases} \quad \mathsf{d}(z, \mathrm{E}) & \text{if } z \in \mathrm{E}_1 \,, \\ - \quad \mathsf{d}(z, \mathrm{E}) & \text{otherwise} \,. \end{cases}$$

Upon recalling the duality formula (2.5.8) along with (2.5.7) and (7.4.40), we obtain:

$$\begin{aligned} \mathcal{W}_2(\mu_x^\epsilon, \mu_y^\epsilon) \geq & \mathcal{W}_1(\mu_x^\epsilon, \mu_y^\epsilon) \\ & \geq \int_{\mathbb{H}^n} g(z) \left(1 + \epsilon^2 f_y^\epsilon(z)\right) \mathrm{d}\hat{\mu}_y^\epsilon(z) - \int_{\mathbb{H}^n} g(z) \left(1 + \epsilon^2 f_x^\epsilon(z)\right) \mathrm{d}\hat{\mu}_x^\epsilon(z) \,. \end{aligned} \tag{7.4.42}$$

Since μ_x^ϵ is represented by a radially-symmetric density about x and $\mathbb{H}^n$ also has a radially-symmetric structure (about any point), by the definition of g it is not difficult to check that in fact

$$\int_{\mathbb{H}^n} g(z) \, \mathrm{d}\mu_x^\epsilon(z) = \int_{\mathbb{H}^n} g(z) \left(1 + \epsilon^2 f_x^\epsilon(z)\right) \mathrm{d}\hat{\mu}_x^\epsilon(z) = 0 \,, \tag{7.4.43}$$

therefore we can focus on the first integral. By virtue of (7.4.39), we have:

$$\int_{\mathbb{H}^n} g(z) \left(1 + \epsilon^2 f_y^\epsilon(z)\right) \mathrm{d}\hat{\mu}_y^\epsilon(z) = \int_{\mathbb{R}^n} g(\varphi_y(q)) \left(1 + \epsilon^2 f_y^\epsilon(\varphi_y(q))\right) \mathrm{d}\mu_E^\epsilon(q) \,; \tag{7.4.44}$$

on the other hand, thanks to (7.4.29) and the fact that μ_E^ϵ is supported in the Euclidean ball B_ϵ centered at the origin, we can write

$$\begin{aligned} & \int_{\mathbb{R}^n} g(\varphi_y(q)) \left(1 + \epsilon^2 f_y^\epsilon(\varphi_y(q))\right) \mathrm{d}\mu_E^\epsilon(q) \\ = & \int_{B_\epsilon} \left[\delta \left(1 + \frac{|q|^2 - (q \cdot p_v)^2}{2} + O(|q|^3 + \delta |q|^2) \right) + O(|q|^3) \right] \left(1 + \epsilon^2 f^\epsilon(|q|)\right) \rho_E^\epsilon(|q|) \, \mathrm{d}q \,, \end{aligned} \tag{7.4.45}$$

where $p_v := i^{-1}(v)$ and we have not considered the term $q \cdot p_v$ in the expansion since, by simmetry, it vanishes when integrated against any radial density. Hence, thanks to (7.4.36) (still the right-hand inequality) and (7.4.41), from (7.4.45) we can infer that

$$\begin{aligned} & \int_{\mathbb{R}^n} g(\varphi_y(q)) \left(1 + \epsilon^2 f_y^\epsilon(\varphi_y(q))\right) \mathrm{d}\mu_E^\epsilon(q) \\ = & \delta \left[1 + \frac{n-1}{2n} \int_{B_\epsilon} |q|^2 \rho_E^\epsilon(|q|) \, \mathrm{d}q + O\left(\epsilon^3 + \delta \epsilon^2\right) \right] + O(\epsilon^3) \,. \end{aligned}$$

In view of the left-hand inequality in (7.4.36), there exists a constant $\kappa > 0$ as in the statement such that

$$\int_{\mathbb{R}^n} g(\varphi_y(q)) \left(1 + \epsilon^2 f_y^\epsilon(\varphi_y(q))\right) \mathrm{d}\mu_E^\epsilon(q) \geq \delta \left[1 + 3 \, \kappa \, \epsilon^2 + O\left(\epsilon^3 + \delta \epsilon^2\right) \right] + O(\epsilon^3) \,. \tag{7.4.46}$$

Upon collecting (7.4.42), (7.4.43), (7.4.44) and (7.4.46), the thesis follows by choosing $\overline{\delta}$ so small that $\left|O(\delta \epsilon^2)\right| \leq \kappa \, \epsilon^2$ for all $\delta \in (0, \overline{\delta})$ and $\overline{\epsilon}$ so small that $\left|\delta O(\epsilon^3)\right| + \left|O(\epsilon^3)\right| \leq \kappa \, \delta \, \epsilon^2$ for all $\epsilon \in (0, \overline{\epsilon})$ and all $\delta \in (0, \overline{\delta})$. □

Proof of Theorem 7.2.5. Let $M = 1$. Thanks to [Váz15, Theorem 1.1], we know that $\rho(\cdot, t)$ and $\hat{\rho}(\cdot, t)$ are represented by the same radial density centered at x and y, respectively. That is, $\rho(\cdot, t) = \tilde{\rho}(\mathsf{d}(\cdot, x), t)$ and $\hat{\rho}(\cdot, t) = \tilde{\rho}(\mathsf{d}(\cdot, y), t)$ for a suitable continuous, bounded, radially-nonincreasing family of densities $(r,t) : \mathbb{R}^+ \times \mathbb{R}^+ \mapsto \tilde{\rho}(r,t)$. First of all we observe that, since $\mathbb{H}^n_K$ is a Cartan-Hadamard manifold, $\tilde{\rho}(r,t)$ lies below the *Euclidean* Barenblatt solution $\tilde{\rho}_E(r,t)$, see [GMP18b, Remark 2.12] and [Váz15, Introduction]. This means that there exist constants $D = D(n,m) > 0$ and $k = k(n,m) > 0$ such that

$$\tilde{\rho}(r,t) \leq t^{-\frac{n}{2+n(m-1)}} \left(D - k\, r^2\, t^{-\frac{2}{2+n(m-1)}} \right)_+^{m-1} =: \tilde{\rho}_E(r,t) \qquad \forall (r,t) \in \mathbb{R}^+ \times \mathbb{R}^+ . \tag{7.4.47}$$

In particular,

$$\tilde{\rho}(r,t) \leq \frac{D^{m-1}}{t^{\frac{n}{2+n(m-1)}}}\, \chi_{[0,A(t)]}(r) \quad \forall (r,t) \in \mathbb{R}^+ \times \mathbb{R}^+ , \qquad A(t) := \sqrt{\tfrac{D}{k}}\, t^{\frac{1}{2+n(m-1)}} . \tag{7.4.48}$$

Now let

$$I(t) := \inf_{r \in \left[0, \frac{A(t)}{2}\right]} \tilde{\rho}(r,t) \qquad \forall t > 0 .$$

By mass conservation, (7.4.47) and the fact that $\tilde{\rho}(\cdot,t)$ is nonincreasing, we can deduce the following:

$$\begin{aligned}
&\frac{1}{|\mathbb{S}^{n-1}|} = \\
&K^{-\frac{n-1}{2}} \int_0^{\frac{A(t)}{2}} \tilde{\rho}(r,t) \sinh\!\left(\sqrt{K} r\right)^{n-1} \mathrm{d}r + K^{-\frac{n-1}{2}} \int_{\frac{A(t)}{2}}^{A(t)} \tilde{\rho}(r,t) \sinh\!\left(\sqrt{K} r\right)^{n-1} \mathrm{d}r \leq \\
&K^{-\frac{n-1}{2}} \int_0^{\frac{A(t)}{2}} \tilde{\rho}_E(r,t) \sinh\!\left(\sqrt{K} r\right)^{n-1} \mathrm{d}r + K^{-\frac{n-1}{2}}\, I(t) \int_{\frac{A(t)}{2}}^{A(t)} \sinh\!\left(\sqrt{K} r\right)^{n-1} \mathrm{d}r = \\
&\left[\frac{\lambda}{|\mathbb{S}^{n-1}|} + I(t)\, C\, t^{\frac{n}{2+n(m-1)}} \right] \left[1 + O\!\left(t^{\frac{2}{2+n(m-1)}} \right) \right] ,
\end{aligned} \tag{7.4.49}$$

where

$$\lambda := \left| \mathbb{S}^{n-1} \right| \int_0^{\frac{1}{2}\sqrt{\frac{D}{k}}} \tilde{\rho}_E(r,1)\, r^{n-1}\, \mathrm{d}r < 1 , \qquad C := \int_{\frac{1}{2}\sqrt{\frac{D}{k}}}^{\sqrt{\frac{D}{k}}} r^{n-1}\, \mathrm{d}r > 0 .$$

Note that in the last passage we have exploited the scaling properties of $\tilde{\rho}_E$. From (7.4.49) and the definition of $I(t)$, it is therefore apparent that there exist constants $D_1 = D_1(n,m) > 0$ and $t_1 = t_1(n,K,m) > 0$ such that

$$\tilde{\rho}(r,t) \geq \frac{D_1}{t^{\frac{n}{2+n(m-1)}}}\, \chi_{\left[0, \frac{A(t)}{2}\right]}(r) \qquad \forall (r,t) \in \mathbb{R}^+ \times (0, t_1) . \tag{7.4.50}$$

Hence, in order to estimate $\mathcal{W}_2(\rho(t), \hat{\rho}(t))$ from below, we are in position to apply Lemma 7.4.7. Indeed, if we set $\epsilon \equiv A(t)$ and $\rho^\epsilon \equiv \tilde{\rho}(\cdot,t)$, then by virtue of (7.4.48) and (7.4.50) we can claim that (7.4.36) is satisfied with $\theta = 1/2$ and suitable positive constants C_1, C_2 depending only on n and m, provided $\epsilon < A(t_1)$ (condition (7.4.36) is required to hold for $\epsilon \in (0,1)$ only for convenience). Estimate (7.2.7) for $M = 1$ is just (7.4.37), upon exploiting the above relation between t and ϵ, along with the trivial identity $\mathcal{W}_2(\delta_x, \delta_y) = \mathsf{d}(x,y)$.

In order to deal with a general mass $M > 0$, it is enough to notice that $M\rho(tM^{m-1})$ and $M\hat{\rho}(tM^{m-1})$ are still solutions of (7.2.1) starting from $M\delta_x$ and $M\delta_y$, respectively (recall that $\mathcal{W}_2^2$ is proportional to the mass). □

Bibliography

[AGS08] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Second. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2008, pp. x+334.

[AGS14a] L. Ambrosio, N. Gigli, and G. Savaré. "Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below". In: *Invent. Math.* 195.2 (2014), pp. 289–391.

[AGS14b] L. Ambrosio, N. Gigli, and G. Savaré. "Metric measure spaces with Riemannian Ricci curvature bounded from below". In: *Duke Math. J.* 163.7 (2014), pp. 1405–1490.

[AGS15] L. Ambrosio, N. Gigli, and G. Savaré. "Bakry-Émery curvature-dimension condition and Riemannian Ricci curvature bounds". In: *Ann. Probab.* 43.1 (2015), pp. 339–404.

[AM16] L. Ambrosio and A. Mondino. "Gaussian-type isoperimetric inequalities in $\mathsf{RCD}(K,\infty)$ probability spaces for positive K". In: *Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl.* 27.4 (2016), pp. 497–514.

[Amb+15] L. Ambrosio et al. "Riemannian Ricci curvature lower bounds in metric measure spaces with σ-finite measure". In: *Trans. Amer. Math. Soc.* 367.7 (2015), pp. 4661–4701.

[Amb18] L. Ambrosio. "Calculus, heat flow and curvature-dimension bounds in metric measure spaces". In: *Proceedings of the ICM 2018* 1 (2018), pp. 301–340.

[AMS19] L. Ambrosio, A. Mondino, and G. Savaré. "Nonlinear diffusion equations and curvature conditions in metric measure spaces". In: *Mem. Amer. Math. Soc (In Press)* (2019). URL: https://arxiv.org/abs/1509.07273.

[AS66] S.M. Ali and S.D. Silvey. "A general class of coefficients of divergence of one distribution from another". In: *J. Roy. Stat. Soc. Ser. B* 28 (1966), pp. 131–142.

[Bak+95] D. Bakry et al. "Sobolev inequalities in disguise". In: *Indiana Univ. Math. J.* 44.4 (1995), pp. 1033–1074.

[BB00] J-D Benamou and Y. Brenier. "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem". In: *Numer. Math.* 84.3 (2000), pp. 375–393.

[BBI01] D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry*. Vol. 33. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2001, pp. xiv+415.

[BC14] F. Bolley and J.A Carrillo. "Nonlinear diffusion: geodesic convexity is equivalent to Wasserstein contraction". In: *Comm. Partial Differential Equations* 39.10 (2014), pp. 1860–1869.

[BGL14] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*. Vol. 348. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer, Cham, 2014, pp. xx+552.

[BGL15] D. Bakry, I. Gentil, and M. Ledoux. "On Harnack inequalities and optimal transportation". In: *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)* 14.3 (2015), pp. 705–727.

[BGV08] M. Bonforte, G. Grillo, and J.L. Vázquez. "Fast diffusion flow on manifolds of nonpositive curvature". In: *J. Evol. Equ.* 8.1 (2008), pp. 99–128.

[BL96] D. Bakry and M. Ledoux. "Lévy-Gromov's isoperimetric inequality for an infinite-dimensional diffusion generator". In: *Invent. Math.* 123.2 (1996), pp. 259–281.

[Bre91] Y. Brenier. "Polar factorization and monotone rearrangement of vector-valued functions". In: *Comm. Pure Appl. Math.* 44.4 (1991), pp. 375–417.

[BS18] D. Bianchi and A.G. Setti. "Laplacian cut-offs, porous and fast diffusion on manifolds and other applications". In: *Calc. Var. Partial Differential Equations* 57.1 (2018), Art. 4, 33.

[Bul03] P.S. Bullen. *Handbook of Means and Their Inequalities*. 3300 AA Dordrecht, the Netherlands: Kluwer Academic Publishers, 2003.

[Bus78] P. Buser. "Über eine Ungleichung von Cheeger". In: *Math. Z.* 158.3 (1978), pp. 245–252.

[Bus82] P. Buser. "A note on the isoperimetric constant". In: *Ann. Sci. École Norm. Sup. (4)* 15.2 (1982), pp. 213–230.

[Cam86] L. Le Cam. *Asymptotic Methods in Statistical Decision Theory*. Springer, 1986.

[CH16] T. Coulhon and D. Hauer. "Regularisation effects of nonlinear semigroups". In: *arXiv e-prints*, arXiv:1604.08737 (Apr. 2016). arXiv: `1604.08737 [math.AP]`.

[Che70] J. Cheeger. "A lower bound for the smallest eigenvalue of the Laplacian". In: *Problems in analysis (Papers dedicated to Salomon Bochner, 1969)*. Princeton University Press, 1970, pp. 195–199.

[Che99] J. Cheeger. "Differentiability of Lipschitz functions on metric measure spaces". In: *Geom. Funct. Anal.* 9.3 (1999), pp. 428–517.

[Chi+18] L. Chizat et al. "Unbalanced optimal transport: dynamic and Kantorovich formulations". In: *J. Funct. Anal.* 274.11 (2018), pp. 3090–3123.

[CM17a] F. Cavalletti and A. Mondino. "Sharp and rigid isoperimetric inequalities in metric-measure spaces with lower Ricci curvature bounds". In: *Invent. Math.* 208.3 (2017), pp. 803–849.

[CM17b] F. Cavalletti and A. Mondino. "Sharp geometric and functional inequalities in metric measure spaces with lower Ricci curvature bounds". In: *Geom. Topol.* 21.1 (2017), pp. 603–645.

[CM18] F. Cavalletti and A. Mondino. "Isoperimetric inequalities for finite perimeter sets under lower Ricci curvature bounds". In: *Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl.* 29.3 (2018), pp. 413–430.

[CMV06] J.A. Carrillo, R.J. McCann, and C. Villani. “Contractions in the 2-Wasserstein length space and thermalization of granular media”. In: *Arch. Ration. Mech. Anal.* 179.2 (2006), pp. 217–263.

[Csi63] I. Csiszar. “Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizitat von Markoffschen Ketten”. In: *Magyar. Tud. Akad. Mat. Kutato Int. Kozl* 8 (1963), pp. 85–108.

[Csi67] I. Csiszar. “Information-type measures of difference of probability distributions and indirect observation”. In: *Studia Scientiarum Mathematicarum Hungarica* 2 (1967), pp. 229–318.

[De 19] N. De Ponti. “Metric properties of homogeneous and spatially inhomogeneous F-divergences”. In: *IEEE Transactions on Information Theory* (2019). early access. DOI: 10.1109/TIT.2019.2937485.

[DM19] N. De Ponti and A. Mondino. “Sharp Cheeger-Buser type inequalities in $\mathsf{RCD}(K,\infty)$ spaces”. In: *arXiv e-prints*, arXiv:1902.03835 (Feb. 2019). arXiv: 1902.03835 [math.FA].

[DMO19] N. De Ponti, M. Muratori, and C. Orrieri. “Wasserstein stability of porous medium-type equations on manifolds with Ricci curvature bounded below”. In: *arXiv e-prints*, arXiv:1908.03147 (Aug. 2019). arXiv: 1908.03147 [math.AP].

[DS08] S. Daneri and G. Savaré. “Eulerian calculus for the displacement convexity in the Wasserstein distance”. In: *SIAM J. Math. Anal.* 40.3 (2008), pp. 1104–1122.

[DS63] N. Dunford and J.T. Schwartz. *Linear operators. Part II: Spectral theory. Self adjoint operators in Hilbert space.* With the assistance of William G. Bade and Robert G. Bartle. Interscience Publishers John Wiley & Sons New York-London, 1963, ix+pp. 859–1923+7.

[EKS15] M. Erbar, K. Kuwada, and K-T. Sturm. “On the equivalence of the entropic curvature-dimension condition and Bochner’s inequality on metric measure spaces”. In: *Invent. Math.* 201.3 (2015), pp. 993–1071.

[Erb10] M. Erbar. “The heat equation on manifolds as a gradient flow in the Wasserstein space”. In: *Ann. Inst. Henri Poincaré Probab. Stat.* 46.1 (2010), pp. 1–23.

[ES03] D.M. Endres and J.E. Schindelin. “A new metric for probability distributions”. In: *IEEE Transactions on Information Theory* 7 (7 2003).

[FM17] A.R. Fotache and M. Muratori. “Smoothing effects for the filtration equation with different powers”. In: *J. Differential Equations* 263.6 (2017), pp. 3291–3326.

[Foo84] R.L. Foote. “Regularity of the distance function”. In: *Proc. Amer. Math. Soc.* 92.1 (1984), pp. 153–155.

[FP19] N. Fournier and B. Perthame. “Monge-Kantorovich distance for PDEs: the coupling method”. In: *arXiv e-prints*, arXiv:1903.11349 (Mar. 2019). arXiv: 1903.11349 [math.AP].

[Fuk87] K. Fukaya. “Collapsing of Riemannian manifolds and eigenvalues of Laplace operator”. In: *Invent. Math.* 87.3 (1987), pp. 517–547.

[Gav07] A.V. Gavrilov. “The double exponential mapping and covariant differentiation”. In: *Sibirsk. Mat. Zh.* 48.1 (2007), pp. 68–74.

[Gig+] N. Gigli et al. “Rigidity for the spectral gap on $\mathsf{RCD}(K,\infty)$-spaces”. In: *to appear in American Journ. Math* (). URL: https://arxiv.org/abs/1709.04017.

[GMP13] G. Grillo, M. Muratori, and M. Porzio. “Porous media equations with two weights: smoothing and decay properties of energy solutions via Poincaré inequalities”. In: *Discrete Contin. Dyn. Syst.* 33.8 (2013), pp. 3599–3640.

[GMP18a] G. Grillo, M. Muratori, and F. Punzo. “The porous medium equation with large initial data on negatively curved Riemannian manifolds”. In: *J. Math. Pures Appl. (9)* 113 (2018), pp. 195–226.

[GMP18b] G. Grillo, M. Muratori, and F. Punzo. “The porous medium equation with measure data on negatively curved Riemannian manifolds”. In: *J. Eur. Math. Soc. (JEMS)* 20.11 (2018), pp. 2769–2812.

[GMS15] N. Gigli, A. Mondino, and G. Savaré. “Convergence of pointed non-compact metric measure spaces and stability of Ricci curvature bounds and heat flows”. In: *Proc. Lond. Math. Soc. (3)* 111.5 (2015), pp. 1071–1129.

[GMV17] G. Grillo, M. Muratori, and J.L. Vázquez. “The porous medium equation on Riemannian manifolds with negative curvature. The large-time behaviour”. In: *Adv. Math.* 314 (2017), pp. 328–377.

[GMV19] G. Grillo, M. Muratori, and J.L. Vázquez. “The porous medium equation on Riemannian manifolds with negative curvature: the superquadratic case”. In: *Math. Ann.* 373.1-2 (2019), pp. 119–153.

[Gro99] M. Gromov. *Metric structures for Riemannian and non-Riemannian spaces*. Vol. 152. Progress in Mathematics. Based on the 1981 French original [MR0682063 (85e:53051)], With appendices by M. Katz, P. Pansu and S. Semmes, Translated from the French by Sean Michael Bates. Birkhäuser Boston, Inc., Boston, 1999, pp. xx+585.

[Heb99] E. Hebey. *Nonlinear analysis on manifolds: Sobolev spaces and inequalities*. Vol. 5. Courant Lecture Notes in Mathematics. New York University, Courant Institute of Mathematical Sciences, New York, 1999, pp. x+309.

[Hel09] E. Hellinger. “Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen”. In: *J. Reine Angew. Math* 136 (1909).

[JKO98] R. Jordan, D. Kinderlehrer, and F. Otto. “The variational formulation of the Fokker-Planck equation”. In: *SIAM J. Math. Anal.* 29.1 (1998), pp. 1–17.

[JZ16] Y. Jiang and H-C. Zhang. “Sharp spectral gaps on metric measure spaces”. In: *Calc. Var. Partial Differential Equations* 55.1 (2016), Art. 14, 14.

[Kan42] L. Kantorovitch. “On the translocation of masses”. In: *C. R. (Doklady) Acad. Sci. URSS (N.S.)* 37 (1942), pp. 199–201.

[Kan48] L. Kantorovitch. “On a problem of Monge”. In: *Uspehi Matem. Nauk (N.S.)* 3 (1948).

[Ket15] C. Ketterer. “Obata’s rigidity theorem for metric measure spaces”. In: *Anal. Geom. Metr. Spaces* 3.1 (2015), pp. 278–295.

[KL51] S. Kullback and R.A. Leibler. “On Information and Sufficiency”. In: *Annals of Mathematical Statistics* 22 (1951).

[KMV16] S. Kondratyev, L. Monsaingeon, and D. Vorotnikov. "A new optimal transport distance on the space of finite Radon measures". In: *Adv. Differential Equations* 21.11-12 (2016), pp. 1117–1164.

[Kou14] O. Kouba. "Bounds for the ratios of differences of power means in two arguments". In: *Mathematical Inequalities and Applications* 17.3 (2014).

[KOV91] P. Kafka, F. Osterreicher, and I. Vincze. "On powers of f-divergences defining a distance". In: *Studia Sci. Math. Hungar.* 26 (1991).

[Led04] M. Ledoux. "Spectral gap, logarithmic Sobolev constant, and geometric bounds". In: *Surveys in differential geometry. Vol. IX.* Vol. 9. Surv. Differ. Geom. Int. Press, Somerville, MA, 2004, pp. 219–240.

[Led94] M. Ledoux. "A simple analytic proof of an inequality by P. Buser". In: *Proc. Amer. Math. Soc.* 121.3 (1994), pp. 951–959.

[Lee13] J.M. Lee. *Introduction to smooth manifolds.* Second. Vol. 218. Graduate Texts in Mathematics. Springer, New York, 2013, pp. xvi+708.

[Lin91] J. Lin. "Divergence measures based on the Shannon entropy". In: *IEEE Transactions on Information Theory* 37 (1991), pp. 145–151.

[LMS16] M. Liero, A. Mielke, and G. Savaré. "Optimal transport in competition with reaction: the Hellinger-Kantorovich distance and geodesic curves". In: *SIAM J. Math. Anal.* 48.4 (2016), pp. 2869–2911.

[LMS18a] M. Liero, A. Mielke, and G. Savaré. "Optimal entropy-transport problems and a new Hellinger-Kantorovich distance between positive measures". In: *Invent. Math.* 211.3 (2018), pp. 969–1117.

[LMS18b] S. Lisini, E. Mainini, and A. Segatti. "A gradient flow approach to the porous medium equation with fractional pressure". In: *Arch. Ration. Mech. Anal.* 227.2 (2018), pp. 567–606.

[LT04] H. Li and G. Toscani. "Long-time asymptotics of kinetic models of granular flows". In: *Arch. Ration. Mech. Anal.* 172.3 (2004), pp. 407–428.

[LV06] F. Liese and I. Vajda. "On divergences and informations in statistics and information theory". In: *IEEE Transactions on Information Theory* 52 (2006), pp. 4394–4412.

[LV07] J. Lott and C. Villani. "Weak curvature conditions and functional inequalities". In: *J. Funct. Anal.* 245.1 (2007), pp. 311–333.

[LV09] J. Lott and C. Villani. "Ricci curvature for metric-measure spaces via optimal transport". In: *Ann. of Math. (2)* 169.3 (2009), pp. 903–991.

[LV87] F. Liese and I. Vajda. *Convex Statistical Distances.* Teubner, 1987.

[M M03] Jr. M. Miranda. "Functions of bounded variation on "good" metric spaces". In: *J. Math. Pures Appl. (9)* 82.8 (2003), pp. 975–1004.

[Mál+96] J. Málek et al. *Weak and measure-valued solutions to evolutionary PDEs.* Vol. 13. Applied Mathematics and Mathematical Computation. Chapman & Hall, London, 1996, pp. xii+317.

[Mas93] G. Dal Maso. *An introduction to* Γ*-convergence.* Vol. 8. Progress in Nonlinear Differential Equations and their Applications. Birkhäuser Boston, 1993, pp. xiv+340.

[Mat64] K. Matusita. "Distances and decision rules". In: *Annals of the Institute of Statistical Mathematics* 16 (1964), pp. 305–320.

[McK70] H.P. McKean. "An upper bound to the spectrum of Δ on a manifold of negative curvature". In: *J. Differential Geometry* 4 (1970), pp. 359–366.

[Mém11] F. Mémoli. "Gromov-Wasserstein distances and the metric approach to object matching". In: *Found. Comput. Math.* 11.4 (2011), pp. 417–487.

[Mon81] G. Monge. "Mémoire sur la théorie des déblais et des remblais". In: *Histoire de l'Académie Royale des Sciences de Paris* (1781), pp. 666–704.

[MS05] F. Mémoli and G. Sapiro. "A theoretical and computational framework for isometry invariant recognition of point cloud data". In: *Found. Comput. Math.* 5.3 (2005), pp. 313–347.

[MS18] A. Mondino and D. Semola. "Polya-Szego inequality and Dirichlet p-spectral gap for non-smooth spaces with Ricci curvature bounded below". In: *arXiv e-prints*, arXiv:1807.04453 (July 2018). arXiv: 1807.04453 [math.FA].

[Oll09] Y. Ollivier. "Ricci curvature of Markov chains on metric spaces". In: *J. Funct. Anal.* 256.3 (2009), pp. 810–864.

[Ost96] F. Osterreicher. "On a class of perimeter-type distances of probability distributions". In: *Kybernetika* 32 (4 1996), pp. 389–393.

[OT11] S-i. Ohta and A. Takatsu. "Displacement convexity of generalized relative entropies". In: *Adv. Math.* 228.3 (2011), pp. 1742–1787.

[Ott01] F. Otto. "The geometry of dissipative evolution equations: the porous medium equation". In: *Comm. Partial Differential Equations* 26.1-2 (2001), pp. 101–174.

[OV03] F. Osterreicher and I. Vajda. "A new metric divergences on probability spaces and its applicability in statistics". In: *Annals of the Institute of Statistical Mathematics* 55 (3 2003), pp. 639–653.

[OW05] F. Otto and M. Westdickenberg. "Eulerian calculus for the contraction in the Wasserstein distance". In: *SIAM J. Math. Anal.* 37.4 (2005), pp. 1227–1255.

[Raj12] T. Rajala. "Local Poincaré inequalities from stable curvature conditions on metric spaces". In: *Calc. Var. Partial Differential Equations* 44.3-4 (2012), pp. 477–494.

[RS05] M.K. von Renesse and K-T. Sturm. "Transport inequalities, gradient estimates, entropy, and Ricci curvature". In: *Comm. Pure Appl. Math.* 58.7 (2005), pp. 923–940.

[San15] F. Santambrogio. *Optimal transport for applied mathematicians.* Vol. 87. Progress in Nonlinear Differential Equations and their Applications. Calculus of variations, PDEs, and modeling. Birkhäuser/Springer, Cham, 2015, pp. xxvii+353.

[Sav14] G. Savaré. "Self-improvement of the Bakry-Émery condition and Wasserstein contraction of the heat flow in $\mathsf{RCD}(K,\infty)$ metric measure spaces". In: *Discrete Contin. Dyn. Syst.* 34.4 (2014), pp. 1641–1661.

[Sch73] L. Schwartz. *Radon measures on arbitrary topological spaces and cylindrical measures.* Tata Institute of Fundamental Research Studies in Mathematics, No. 6. Published for the Tata Institute of Fundamental Research, Bombay by Oxford University Press, London, 1973, pp. xii+393.

[Sha48] C. E. Shannon. "A mathematical theory of communication". In: *Bell System Tech. J.* 27 (1948), pp. 379–423, 623–656.

[Shi16] T. Shioya. *Metric measure geometry.* Vol. 25. IRMA Lectures in Mathematics and Theoretical Physics. Gromov's theory of convergence and concentration of metrics and measures. EMS Publishing House, Zürich, 2016, pp. xi+182.

[SS13] B. Schmitzer and C. Schnörr. "Modelling convex shape priors and matching based on the Gromov-Wasserstein distance". In: *J. Math. Imaging Vision* 46.1 (2013), pp. 143–159.

[Ste70] E.M. Stein. *Topics in harmonic analysis related to the Littlewood-Paley theory.* Annals of Mathematics Studies, No. 63. Princeton University Press, Princeton, N.J., 1970, pp. viii+146.

[Str83] R.S. Strichartz. "Analysis of the Laplacian on the complete Riemannian manifold". In: *J. Funct. Anal.* 52.1 (1983), pp. 48–79.

[Stu05] K-T. Sturm. "Convex functionals of probability measures and nonlinear diffusions on manifolds". In: *J. Math. Pures Appl. (9)* 84.2 (2005), pp. 149–168.

[Stu06] K-T. Sturm. "On the geometry of metric measure spaces. I". In: *Acta Math.* 196.1 (2006), pp. 65–131.

[Vaj89] I. Vajda. *Theory of statistical inference and information.* Springer Netherlands, 1989.

[Var89] N.T. Varopoulos. "Small time Gaussian estimates of heat diffusion kernels. I. The semigroup technique". In: *Bull. Sci. Math.* 113.3 (1989), pp. 253–277.

[Váz06] J.L. Vázquez. *Smoothing and decay estimates for nonlinear diffusion equations.* Vol. 33. Oxford Lecture Series in Mathematics and its Applications. Equations of porous medium type. Oxford University Press, Oxford, 2006, pp. xiv+234.

[Váz07] J.L. Vázquez. *The porous medium equation.* Oxford Mathematical Monographs. Mathematical theory. The Clarendon Press, Oxford University Press, Oxford, 2007, pp. xxii+624.

[Váz15] J.L. Vázquez. "Fundamental solution and long time behavior of the porous medium equation in hyperbolic space". In: *J. Math. Pures Appl. (9)* 104.3 (2015), pp. 454–484.

[Vil03] C. Villani. *Topics in optimal transportation.* Vol. 58. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2003, pp. xvi+370.

[Vil09] C. Villani. *Optimal transport.* Vol. 338. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Old and new. Springer-Verlag, Berlin, 2009, pp. xxii+973.

[Vil19] C. Villani. "Inégalités isopérimétriques dans les espaces métriques mesurés [d'après F. Cavalletti & A. Mondino]". In: 407. Séminaire Bourbaki. Vol. 2016/2017. Exposés 1120–1135. 2019. URL: http://www.bourbaki.ens.fr/TEXTES/1127.pdf.

[Vin81] I. Vincze. "On the Concept and Measure of Information Contained in an Observation". In: *Contributions to Probability.* Academic Press, 1981, pp. 207–214.

[Wan02] F-Y. Wang. “Functional inequalities and spectrum estimates: the infinite measure case”. In: *J. Funct. Anal.* 194.2 (2002), pp. 288–310.