# CONTRACTION AND REGULARIZING PROPERTIES OF HEAT FLOWS IN METRIC MEASURE SPACES

Giulia Luise

Department of Computer Science, University College London
Gower Street, London WC1E 6BT, UK

Giuseppe Savaré *

Dipartimento di Matematica 'Felice Casorati', University of Pavia
Via Ferrata 1, 27100 Pavia, Italy

*Dedicated to Alexander Mielke on the occasion of his 60th birthday*

(Communicated by the associate editor name)

Abstract. We illustrate some novel contraction and regularizing properties of the Heat flow in metric-measure spaces that emphasize an interplay between Hellinger-Kakutani, Kantorovich-Wasserstein and Hellinger-Kantorovich distances. Contraction properties of Hellinger-Kakutani distances and general Csiszár divergences hold in arbitrary metric-measure spaces and do not require assumptions on the linearity of the flow.

When weaker transport distances are involved, we will show that contraction and regularizing effects rely on the dual formulations of the distances and are strictly related to lower Ricci curvature bounds in the setting of $\mathrm{RCD}(K,\infty)$ metric measure spaces. As a byproduct, when $K \geq 0$ we will also find new estimates for the asymptotic decay of the solution.

## Contents



---

2010 *Mathematics Subject Classification.* Primary: 49Q20, 47D07; Secondary: 30L99.

*Key words and phrases.* Heat flows, contraction, Hellinger distance, Wasserstein distance, curvature bounds, $\mathrm{RCD}(K,\infty)$ spaces.

The second author is partially supported by PRIN2015 grant from MIUR for the project *Calculus of Variations* and by IMATI-CNR.

* Corresponding author: Giuseppe Savaré.

1. **Introduction.** The study of contraction properties of $L^p$ norms and more general convex entropy functionals with respect to the action of Markov semigroups is a very classic subject (see e.g. [9]). More recently, the role of the Kantorovich-Rubinstein-Wasserstein metric $\mathsf{W}_2$ for second order diffusion equations in the space of probability measures has been deeply investigated, starting from the pioneering contribution by F. Otto [35]. Many investigations have clarified the relations between analytic estimates depending on the structure of the generating differential operator and geometric properties of the underlying spaces, with an increasing level of generality. An incomplete list of contributions includes the contraction of a general class of evolution equations combining diffusion, interaction and drift [13], the gradient-flow structure and the geodesic convexity in Euclidean spaces [25, 35, 2], the Heat flow in Riemannian manifolds and the Ricci curvature [36, 37, 41, 17, 19, 42], Hilbert geometry [7], the duality with gradient estimates and the Alexandrov spaces [30, 21], the RCD metric measure spaces and the Bakry-Émery condition [3, 4, 5, 10, 20, 6].

In one of the most general formulations, we will deal with a metric-measure space $(X, \mathsf{d}, \mathfrak{m})$ given by a complete and separable metric space $(X, \mathsf{d})$ endowed with a Borel positive measure $\mathfrak{m}$ with full support satisfying the growth condition

$$\exists o \in X,\ \kappa \geq 0 : \quad \mathfrak{m}(\{x : \ \mathsf{d}(x, o) < r\}) \leq \mathrm{e}^{\kappa r^2}. \tag{1}$$

We introduce the Cheeger energy functional $\mathsf{Ch} : L^2(X, \mathfrak{m}) \to [0, +\infty]$

$$\mathsf{Ch}(f) := \inf \Big\{ \liminf_{n\to\infty} \frac{1}{2} \int_X |\mathrm{D} f_n|^2 \,\mathrm{d}\mathfrak{m},\ f_n \in \mathrm{Lip}_b(X),\ f_n \to f \text{ in } L^2(X, \mathfrak{m}) \Big\} \tag{2}$$

where

$$|\mathrm{D} f|(x) := \limsup_{y\to x} \frac{|f(y) - f(x)|}{\mathsf{d}(x, y)}; \quad |\mathrm{D} f|(x) := 0 \text{ if } x \text{ is isolated.} \tag{3}$$

$\mathsf{Ch}$ is a convex, 2-homogeneous and lower semicontinuous functional whose proper domain $\mathcal{D}(\mathsf{Ch}) = \{f \in L^2(X, \mathfrak{m}) : \mathsf{Ch}(f) < \infty\}$ provides one of the equivalent characterization of the metric Sobolev space $\mathrm{W}^{1,2}(X, \mathsf{d}, \mathfrak{m})$ (see also [22, 28, 39, 11, 23]). A local weak gradient $|\mathrm{D} f|_w \in L^2(X, \mathfrak{m})$ can be associated to each function $f \in \mathrm{W}^{1,2}(X, \mathsf{d}, \mathfrak{m})$ so that the Cheeger energy admits the integral representation

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

The $L^2$ subdifferential of $\mathsf{Ch}$ (whose minimal selection will be denoted by $-\Delta$) generates a continuous semigroup of order preserving contractions $(\mathsf{P}_t)_{t\geq 0}$ in $L^2(X, \mathfrak{m})$, which is canonically attached to the metric-measure structure $(X, \mathsf{d}, \mathfrak{m})$.

Even if in general the operators $\mathsf{P}_t$ are not linear, one can prove [3] that the semigroup is contractive with respect to all the $L^p$ norms, $p \in [1, +\infty]$,

$$\|\mathsf{P}_t f - \mathsf{P}_t g\|_{L^p(X,\mathfrak{m})} \leq \|f - g\|_{L^p(X,\mathfrak{m})} \quad \text{for every } f, g \in L^2 \cap L^p(X, \mathfrak{m}), \tag{4}$$

and all the integral functionals with convex integrand $\phi : \mathbb{R} \to [0, +\infty)$

$$\int_X \phi(\mathsf{P}_t f) \,\mathrm{d}\mathfrak{m} \le \int_X \phi(f) \,\mathrm{d}\mathfrak{m} \quad \text{for every } f \in L^2(X, \mathfrak{m}). \tag{5}$$

A first important result we will prove in Section 4 is the extension of (4)-(5) to arbitrary convex integral functionals on evolving pairs:

$$\int_X E(\mathsf{P}_t f, \mathsf{P}_t g) \,\mathrm{d}\mathfrak{m} \le \int_X E(f, g) \,\mathrm{d}\mathfrak{m} \quad \text{for every } f, g \in L^2(X, \mathfrak{m}), \tag{6}$$

whenever $E : \mathbb{R}^2 \to [0, +\infty]$ is a lower semicontinuous convex integrand with $E(0, 0) = 0$. As a byproduct, we obtain that the action of $(\mathsf{P}_t)_{t \ge 0}$ on nonnegative functions $f, g \in L^1(X, \mathfrak{m})$ is a contraction with respect to arbitrary Csiszár divergences (see [16, 33] and Section 2), as the Kullback-Leibler entropy functional [29] associated to $E(r, s) = r \ln(r/s) - r + s$ if $r, s > 0$, yielding (since $\mathsf{P}_t$ is mass preserving)

$$\int_{\mathsf{P}_t g > 0} \ln\Big(\mathsf{P}_t f / \mathsf{P}_t g\Big) \mathsf{P}_t f \,\mathrm{d}\mathfrak{m} \le \int_{g > 0} \ln\Big(f/g\Big) f \,\mathrm{d}\mathfrak{m},$$

or the Hellinger-Kakutani distances [24, 26]

$$\int_X |(\mathsf{P}_t f)^{1/p} - (\mathsf{P}_t g)^{1/p}|^p \,\mathrm{d}\mathfrak{m} \le \int_X |f^{1/p} - g^{1/p}|^p \,\mathrm{d}\mathfrak{m} \quad p \in [1, +\infty),$$

associated to $E(r, s) = |r^{1/p} - s^{1/p}|^p$, $r, s \ge 0$

The most relevant connections with optimal transport metrics occur when $\mathsf{Ch}$ is also a quadratic form, i.e. it satisfies the parallelogram rule

$$\mathsf{Ch}(f + g) + \mathsf{Ch}(f - g) = 2\mathsf{Ch}(f) + 2\mathsf{Ch}(g), \quad \text{for every } f, g \in \mathcal{D}(\mathsf{Ch}). \tag{7}$$

In this case $-\Delta$ is a linear positive selfadjoint operator in $L^2(X, \mathfrak{m})$ and $(\mathsf{P}_t)_{t \ge 0}$ is a linear Markov semigroup associated to a strongly local symmetric Dirichlet form $\mathcal{E}$ on $L^2(X, \mathfrak{m})$, admitting Carré du Champ $\Gamma : \mathcal{D}(\mathsf{Ch}) \times \mathcal{D}(\mathsf{Ch}) \to L^1(X, \mathfrak{m})$ which provides a bilinear extension of the weak gradient, since

$$\Gamma(f, f) = |\mathrm{D} f|_w^2 \quad \text{for every } f \in \mathrm{W}^{1,2}(X, \mathsf{d}, \mathfrak{m}).$$

If every bounded function $f \in \mathrm{W}^{1,2}(X, \mathsf{d}, \mathfrak{m})$ with $|\mathrm{D} f|_w \le 1$ $\mathfrak{m}$-a.e. admits a $\mathsf{d}$-continuous representative (still denoted by $f$) which satisfies the 1-Lipschitz condition

$$|f(y) - f(x)| \le \mathsf{d}(x, y) \quad \text{for every } x, y \in X$$

then $\Delta$ satisfies (a suitable weak formulation of) the Bakry-Émery condition $\mathrm{BE}(K, \infty)$, $K \in \mathbb{R}$,

$$\Gamma_2(f) = \frac{1}{2} \Delta \Gamma(f, f) - \Gamma(f, \Delta f) \ge K\, \Gamma(f) \tag{8}$$

if and only if $(\mathsf{P}_t)_{t \ge 0}$ admits a (unique) extension $(\mathsf{P}_t^*)_{t \ge 0}$ to the space of finite Borel measures $\mathcal{M}(X)$ and satisfies the contraction property (see [5])

$$\mathsf{W}_2(\mathsf{P}_t^* \mu_0, \mathsf{P}_t^* \mu_1) \le \mathrm{e}^{-Kt} \mathsf{W}_2(\mu_0, \mu_1) \quad \text{for every } \mu_0, \mu_1 \in \mathcal{P}_2(X); \tag{9}$$

here $\mathsf{W}_2$ denotes the 2-Kantorovich-Wasserstein distance between probability measures of $\mathcal{P}_2(X)$ with finite quadratic moments

$$\mathsf{W}_2^2(\mu_0, \mu_1) := \min \Big\{ \int_{X \times X} \mathsf{d}^2(x_0, x_1) \,\mathrm{d}\boldsymbol{\mu}(x_0, x_1) : \boldsymbol{\mu} \in \mathcal{P}(X \times X), \\ \pi^0_\sharp \boldsymbol{\mu} = \mu_0,\ \pi^1_\sharp \boldsymbol{\mu} = \mu_1 \Big\}, \quad \pi^i(x_0, x_1) := x_i,\ i = 0, 1.$$

In fact, this property is deeply related with the synthetic theory of $\mathsf{CD}(K,\infty)$ metric-measure spaces with Ricci curvature bounded from below developed by Lott-Villani [34] and Sturm [40]. The combination of the Lott-Sturm-Villani condition $\mathsf{CD}(K,\infty)$ with the quadratic property of the Cheeger energy (7) provides one of the equivalent characterizations of the so-called $\mathrm{RCD}(K,\infty)$ metric-measure space [4], which turned out to be equivalent with the Bakry-Émery functional-analytic approach we have adopted here [5].

The link between (8) and (9) becomes more apparent if we consider that (8) is in fact equivalent to the Bakry-Émery commutation estimate

$$|\mathrm{D}\mathsf{P}_t f|^2 \le \mathrm{e}^{-2Kt}\mathsf{P}_t\big(|\mathrm{D}f|^2\big)\quad\text{for every } f\in \mathrm{Lip}_b(X), \tag{10}$$

combined with the duality formula expressing the distance $\mathsf{W}_2$ in terms of regular subsolutions $\zeta\in \mathrm{C}^1([0,1];\mathrm{Lip}_b(X))$ to the Hamilton-Jacobi equation [36, 3, 1]

$$\frac12\mathsf{W}_2^2(\mu_0,\mu_1)=\sup\Big\{\int_X\zeta_1\,\mathrm{d}\mu_1-\int_X\zeta_0\,\mathrm{d}\mu_0:\partial_t\zeta_t+\frac12|\mathrm{D}\zeta_t|^2\le 0\Big\}, \tag{11}$$

thanks to the dual representation formula for $\mathsf{P}_t^*$:

$$\int_X f\,\mathrm{d}(\mathsf{P}_t^*\mu)=\int_X \mathsf{P}_t f\,\mathrm{d}\mu\quad\text{for every } f\in \mathrm{C}_b(X),\ \mu\in\mathcal{M}(X). \tag{12}$$

(10) shows in fact that $(\mathsf{P}_t)_{t\ge0}$ preserves (up to an exponential factor) subsolutions to the Hamilton-Jacobi equation (11).

In Section 5 we improve (9) in two directions. First of all, we will show that after a strictly positive time $\mathsf{P}_t$ exhibits a regularizing effect, providing a control of the stronger 2-Hellinger distance

$$\mathsf{He}_2^2(\mu_0,\mu_1):=\int_X\Big(\sqrt{\varrho_1}-\sqrt{\varrho_0}\Big)^2\,\mathrm{d}\mu,\quad \mu_i=\varrho_i\mu,$$

in terms of the weaker Wasserstein distance between the initial measures:

$$\mathsf{He}_2(\mathsf{P}_t^*\mu_0,\mathsf{P}_t^*\mu_1)\le\frac{1}{2\sqrt{R_K(t)}}\mathsf{W}_2(\mu_0,\mu_1)\quad\text{for every } \mu_0,\mu_1\in\mathcal{P}_2(X) \tag{13}$$

where

$$R_K(t):=\begin{cases}\dfrac{\mathrm{e}^{2Kt}-1}{K}&\text{if } K\neq0\\ 2t&\text{if } K=0.\end{cases} \tag{14}$$

Notice that when $\mathfrak{m}\in\mathcal{P}_2(X)$ and $K\ge0$ we obtain the asymptotic estimate

$$\mathsf{He}_2(\mathsf{P}_t^*\mu_0,\mathfrak{m})\le\frac{1}{2\sqrt{R_K(t)}}\mathsf{W}_2(\mu_0,\mathfrak{m}),$$

proving in particular Hellinger convergence of $\mathsf{P}_t\mu_0$ to $\mathfrak{m}$ as $t\to\infty$, with exponential rate if $K>0$.

A second and more refined estimate involves the recently introduced family of Hellinger-Kantorovich distances $\mathsf{HK}_\alpha$, $\alpha>0$, [15, 14, 27, 31, 32], which can be defined in terms of an Optimal Entropy–Transport problem [31, 32]

$$\mathsf{HK}_\alpha^2(\mu_0,\mu_1):=\min_{\boldsymbol{\gamma}\in\mathcal{M}(X\times X)}\mathsf{KL}(\gamma_0|\mu_0)+\mathsf{KL}(\gamma_1|\mu_1)+\int_{X\times X}\ell_\alpha(\mathsf{d}(x_0,x_1))\,\mathrm{d}\boldsymbol{\gamma},$$

where $\gamma_0,\gamma_1$ are the marginals of $\boldsymbol{\gamma}$, $\mathsf{KL}$ is the Kullback-Leibler divergence

$$\mathsf{KL}(\gamma|\mu):=\int_X\Big(\varrho\log\varrho-\varrho+1\Big)\,\mathrm{d}\mu,\quad \gamma=\varrho\mu\ll\mu,$$

and $\ell_\alpha$ is the cost function

$$\ell_\alpha(r) := \begin{cases} \log\Big(1+\tan^2\big(r/\sqrt{\alpha}\big) & \text{if } \mathsf{d}(x_0,x_1) < \sqrt{\alpha}\pi/2, \\ +\infty & \text{otherwise.} \end{cases} \tag{15}$$

It turns out that $\mathsf{HK}_\alpha$ (corresponding to $\mathsf{HK}_{\alpha,4}$ in the more general notation of [31, 32]) admits a dual dynamic representation formula [32]

$$\mathsf{HK}^2_\alpha(\mu_0,\mu_1) = \sup\Big\{\int \zeta_1\,\mathrm{d}\mu_1 - \int \zeta_0\,\mathrm{d}\mu_0 : \partial_t\zeta_t + \frac{\alpha}{4}|\mathrm{D}\zeta_t|^2 + \zeta_t^2 \le 0\Big\},$$

so that when the Bakry-Émery condition $\mathrm{BE}(0,\infty)$ holds one has [32]

$$\mathsf{HK}_\alpha(\mathsf{P}_t\mu_0,\mathsf{P}_t\mu_1) \le \mathsf{HK}_\alpha(\mu_0,\mu_1) \quad \text{for every } \mu_0,\mu_1 \in \mathcal{M}(X).$$

Actually, the stronger Hellinger distance at time $t > 0$ can be estimated in terms of the weaker Hellinger-Kantorovich one: for every $t > 0$

$$\mathsf{He}_2(\mathsf{P}^*_t\mu_0,\mathsf{P}^*_t\mu_1) \le \mathsf{HK}_{\alpha(t)}(\mu_0,\mu_1) \quad \text{where } \alpha(t) = 4R_K(t). \tag{16}$$

Differently from other well known properties, the estimates (13) and (16) cannot be deduced by a regularization effect on a single initial datum, since $\mathsf{He}_2$, $\mathsf{W}_2$ and $\mathsf{HK}_\alpha$ are not translation invariant. In this respect, the dual dynamic approach plays a crucial role.

*Plan of the paper.* The paper is organized as follows: in Section 2 we will collect a few preliminary results on Csiszár divergences, Hellinger-Kakutani, Kantorovich-Wasserstein and Hellinger-Kantorovich metrics.

Section 3 is devoted to a short review of the main tools of calculus in metric-measure spaces, which are used throughout the work. A brief description of the main properties of $\mathrm{RCD}(K,\infty)$ metric measures spaces is also presented.

The last two sections contain novel results. Section 4 is dedicated to the proof of (6) in general metric measure spaces. Section 5 discusses the regularization estimates (13) and (16).

## 2. **Distances and entropies on the space of finite measures.**

### 2.1. **Csiszár divergences/Relative entropies.**

We first recall a few basic facts on convex and 1-homogeneous functionals of positive measures.

Let $(\Omega,\mathscr{B})$ be a measurable space. We will denote the space of finite nonnegative measures on $(\Omega,\mathscr{B})$ by $\mathcal{M}(\Omega)$. If $\mu_0,\mu_1 \in \mathcal{M}(\Omega)$, we say that $\lambda \in \mathcal{M}(\Omega)$ is a *common dominating measure* if $\mu_i \ll \lambda$, $i = 0,1$. Such a $\lambda$ always exists, for instance we may take $\lambda = \mu_0 + \mu_1$. We will also often consider the Lebesgue decomposition of $\mu_0$ w.r.t. $\mu_1$ given by

$$\mu_0 = \varrho\mu_1 + \mu_0^\perp, \quad \mu^\perp \perp \mu_1, \quad \varrho := \frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}. \tag{17}$$

We consider the class of Csiszàr density functions

$$F : [0,\infty) \to [0,+\infty] \quad \text{l.s.c. and convex}, \quad F(1) = 0, \tag{18a}$$

with recession constant defined by

$$F'(\infty) := \lim_{r\to\infty}\frac{F(r)}{r} = \sup_{r>0}\frac{F(r)}{r-1},$$

and the corresponding class of homogeneous perspective functions

$$H : [0,\infty) \times [0,\infty) \to [0,+\infty] \quad \text{l.s.c., convex, and positively 1-homogeneous,}$$
$$H(\theta r, \theta s) = \theta H(r,s), \quad H(r,r) = 0 \quad \text{for every } r,s,\theta \geq 0. \tag{18b}$$

There is a one-to-one correspondence between the two classes given by the formula

$$F(r) = H(r,1), \quad H(r,s) = \begin{cases} sF(r/s) & \text{if } s > 0, \\ F'(\infty) & \text{if } s = 0 \end{cases} \tag{18c}$$

**Definition 2.1.** Let $F, H$ be as in (18a,b) and let $\mu_0, \mu_1 \in \mathcal{M}(\Omega)$ with Lebesgue decomposition $\mu_0 = \varrho\mu_1 + \mu_0^\perp$ as in (17). The *Csiszár divergence* associated with $F$ is defined as

$$\mathscr{F}(\mu_0 \mid \mu_1) := \int_\Omega F(\varrho)\,\mathrm{d}\mu_1 + F'(\infty)\mu_0^\perp(\Omega). \tag{19}$$

The $\mathscr{H}$-perspective functional is defined as

$$\mathscr{H}(\mu_0 \mid \mu_1) := \int_\Omega H(\varrho_0, \varrho_1)\,\mathrm{d}\lambda \tag{20}$$

where $\mu_i = \varrho_i\lambda \ll \lambda$, $i = 0,1$, and $\lambda$ is any common dominating measure.
If $F$ and $H$ are related by (18c) then

$$\mathscr{F}(\mu_0 \mid \mu_1) = \mathscr{H}(\mu_0 \mid \mu_1) \quad \text{for every } \mu_0, \mu_1 \in \mathcal{M}(\Omega). \tag{21}$$

Notice that (20) does not depend on the choice of the dominating measure $\lambda$, since the function $H$ is positively 1-homogeneous.

(21) can be easily checked by observing that $\lambda := \mu_1 + \mu_0^\perp$ is a dominating measure for the couple $\mu_0, \mu_1$; if $B_0, B_1$ are measurable subsets of $\Omega$ such that

$$B_0 \cap B_1 = \emptyset, \quad \Omega = B_0 \cup B_1, \quad \mu_0^\perp(B_1) = 0, \quad \mu_1(B_0) = 0,$$

we can easily calculate the densities $\varrho_0, \varrho_1$ by

$$\varrho_0(x) := \begin{cases} 1 & \text{if } x \in B_0 \\ \varrho(x) & \text{if } x \in B_1 \end{cases}, \qquad \varrho_1(x) := \begin{cases} 0 & \text{if } x \in B_0 \\ 1 & \text{if } x \in B_1 \end{cases}$$

so that

$$\begin{aligned} \int_\Omega H(\varrho_0,\varrho_1)\,\mathrm{d}\lambda &= \int_{B_0} H(\varrho_0,\varrho_1)\,\mathrm{d}\lambda + \int_{B_1} H(\varrho_0,\varrho_1)\,\mathrm{d}\lambda \\ &= \int_{B_0} H(1,0)\,\mathrm{d}\mu_0^\perp + \int_{B_1} H(\varrho,1)\,\mathrm{d}\mu_1 = F'(\infty)\mu_0^\perp(B_0) + \int_{B_1} F(\varrho)\,\mathrm{d}\mu_1 \\ &= F'(\infty)\mu_0^\perp(\Omega) + \int_\Omega F(\varrho)\,\mathrm{d}\mu_1 = \mathscr{F}(\mu_0 \mid \mu_1). \end{aligned}$$

An important class of entropy functions is provided by the power like functions which have the following explicit formulas

$$E_p(s) := \begin{cases} \frac{1}{p(p-1)}(r^p - p(r-1) - 1) & \text{if } p \neq 0,1 \\ r\log r - r + 1 & \text{if } p = 1 \\ r - 1 - \log r & \text{if } p = 0. \end{cases}$$

For $p = 1$, the entropy function $E_1(r) = r\log r - r + 1$ generates the well known Kullback-Leibler divergence, often referred to as *relative logarithmic entropy*. Notice that $E_1$ is superlinear, so that $E_1'(\infty) = +\infty$ and its corresponding perspective function is

$$H_{\mathsf{KL}}(r_0, r_1) := \begin{cases} r_0(\ln r_0 - \ln r_1) + r_1 - r_0 & \text{if } r_0, r_1 > 0, \\ r_1 & \text{if } r_0 = 0 \\ +\infty & \text{if } r_0 > 0, r_1 = 0. \end{cases} \tag{22}$$

**Definition 2.2** (Kullback-Leibler divergence (relative logarithmic entropy))**.** Let $\mu_0$ and $\mu_1$ be two finite nonnegative measures. The logarithmic entropy of $\mu_0$ with respect to $\mu_1$ is given by the Csiszàr functional associated to $E_1(r) := r \log r - (r-1)$:

$$\mathsf{KL}(\mu_0 \mid \mu_1) = \begin{cases} \displaystyle\int_\Omega \big(\varrho \log \varrho - \varrho + 1\big)\,\mathrm{d}\mu_1 & \text{if } \mu_0 = \varrho\mu_1 \\ +\infty & \text{otherwise.} \end{cases} \tag{23a}$$

$$= \int_\Omega H_{\mathsf{KL}}(\varrho_0, \varrho_1)\,\mathrm{d}\mu, \quad \mu_i = \varrho_i \mu \ll \mu, \tag{23b}$$

The functionals $\mathscr{F}, \mathscr{H}$ admit a useful dual representation. Let us denote by $\mathrm{B}_b(\Omega)$ the set of bounded Borel functions on $\Omega$ and by $F^* : \mathbb{R} \to (-\infty, +\infty]$ the Legendre conjugate function of $F$, given by

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

We introduce the closed convex subsets $\mathfrak{F}, \mathfrak{H}$ of $\mathbb{R}^2$ given by

$$\mathfrak{F} := \{(\phi, \psi) \in \mathbb{R}^2 : \ \psi \leq -F^*(\phi)\} = \{(\phi, \psi) \in \mathbb{R}^2 : \ r\phi + \psi \leq F(r) \ \ \forall r > 0\}$$

$$\mathfrak{H} := \{(\phi, \psi) \in \mathbb{R}^2 : \ r\phi + s\psi \leq H(r, s) \ \ \forall\, r, s > 0\}.$$

Since $F$ is lower semicontinuous, it can be recovered from $F^*$ and $\mathfrak{F}$ by the Fenchel-Moreau formula [32]

$$F(r) = \sup_{\phi \in \mathbb{R}} (r\phi - F^*(\phi)) = \sup_{(\phi, \psi) \in \mathfrak{F}} r\phi + \psi.$$

Similarly, we have

$$H(r, s) = \sup_{(\phi, \psi) \in \mathfrak{H}} r\phi + s\psi,$$

and $\mathfrak{F} = \mathfrak{H}$ if (18c) holds.

**Theorem 2.3.** *For every $\mu_0, \mu_1 \in \mathcal{M}(\Omega)$ we have*

$$\mathscr{F}(\mu_0 \mid \mu_1) = \sup\Big\{ \int_\Omega \phi\,\mathrm{d}\mu_0 + \int_\Omega \psi\,\mathrm{d}\mu_1 : \phi, \psi \in \mathrm{B}_b(\Omega), \ (\phi(x), \psi(x)) \in \mathfrak{F} \ \ \forall x \in \Omega \Big\},$$

$$\mathscr{H}(\mu_0 \mid \mu_1) = \sup\Big\{ \int_\Omega \phi\,\mathrm{d}\mu_0 + \int_\Omega \psi\,\mathrm{d}\mu_1 : \phi, \psi \in \mathrm{B}_b(\Omega), \ (\phi(x), \psi(x)) \in \mathfrak{H} \ \ \forall x \in \Omega \Big\}.$$

*Proof.* [32, Th. 2.7] □

2.2. **Hellinger distances.** We consider a specific example of perspective functionals $\mathscr{H}$, which gives raise to the Hellinger distances.

**Definition 2.4.** For $\mu_0, \mu_1 \in \mathcal{M}(\Omega)$ and $p \in [1, +\infty)$ the $p$-Hellinger distance is defined by

$$\mathsf{He}_p^p(\mu_0, \mu_1) := \|\varrho_0^{1/p} - \varrho_1^{1/p}\|_{L^p(\Omega, \lambda)}^p = \int_\Omega \Big|\varrho_0^{1/p} - \varrho_1^{1/p}\Big|^p\,\mathrm{d}\lambda \tag{24}$$

where $\mu_i = \varrho_i \lambda \ll \lambda$, $i = 0, 1$, and $\lambda$ is an arbitrary dominating measure.

Notice that the above definition corresponds to (20), (19) for the choices

$$H_p(r,s) := \Big|r^{1/p} - s^{1/p}\Big|^p, \quad F_p(r) = \Big|r^{1/p} - 1\Big|^p. \tag{25}$$

An immediate consequence of the above definition, choosing $\lambda = \mu_0 + \mu_1$ is the uniform bound

$$\mathsf{He}_p(\mu_0,\mu_1) \le \mu_0(X) + \mu_1(X).$$

For $p = 1$ the definition above gives the usual total variation distance, which we will still denote by $\mathsf{He}_1$. The total variation distance and the $L^p$-Hellinger distance $\mathsf{He}_p$ induce the same topology on the space $\mathcal{M}(\Omega)$ and the following relation holds.

**Theorem 2.5.** *Let $q \in (1,\infty]$ be the conjugate exponent of $p$. For every $p > 1$ and arbitrary nonnegative finite measures $\mu_0$ and $\mu_1$ in $\mathcal{M}(\Omega)$,*

$$\mathsf{He}_p^p(\mu_0,\mu_1) \le \mathsf{He}_1(\mu_0,\mu_1) \le c_p\big(\mu_0(\Omega)^{1/q} + \mu_1(\Omega)^{1/q}\big)\mathsf{He}_p(\mu_0,\mu_1), \tag{26}$$

*where $c_p := \max(p/2, 1)$.*

*Proof.* The first part of (26) follows immediately by the representation (24) and the elementary inequality

$$\Big|a^{1/p} - b^{1/p}\Big|^p \le |a-b| \quad \text{for every } a,\, b \ge 0.$$

The second inequality of (26) is a consequence of

$$|a^p - b^p| \le c_p\, |a-b|\,(a^{p-1} + b^{p-1}), \quad a, b \ge 0, \tag{27}$$

which can be easily obtained by integration (without loss of generality we can assume $a \le b$)

$$b^p - a^p = p(b-a)\int_0^1 ((1-t)a + tb)^{p-1}\,\mathrm{d}t = p(b-a)I$$

where

$$I = \int_0^1 ((1-t)a+tb)^{p-1}\,\mathrm{d}t \le \begin{cases} \int_0^1 (1-t)a^{p-1} + tb^{p-1}\,\mathrm{d}t = \frac{1}{2}(a^{p-1} + b^{p-1}) & \text{if } p \ge 2, \\ \int_0^1 \Big((1-t)^{p-1}a^{p-1} + t^{p-1}b^{p-1}\Big)\,\mathrm{d}t \le \frac{1}{p}(a^{p-1} + b^{p-1}) & \text{if } p \le 2. \end{cases}$$

(27) with the choices $a = \varrho_0^{1/p}$ and $b = \varrho_1^{1/p}$, combined with Hölder inequality, yields

$$\begin{aligned} \mathsf{He}_1(\mu_0,\mu_1) &= \int_\Omega |\varrho_0 - \varrho_1|\,\mathrm{d}\lambda \le c_p \int_\Omega \Big|\big(\varrho_0^{1/p} - \varrho_1^{1/p}\big)\big(\varrho_0^{1/q} + \varrho_1^{1/q}\big)\Big|\,\mathrm{d}\lambda \\ &\le c_p \|\varrho_0^{1/p} - \varrho_1^{1/p}\|_{L^p(\Omega,\lambda)} \|\varrho_0^{1/q} + \varrho_1^{1/q}\|_{L^q(\Omega,\lambda)} \\ &\le c_p \|\varrho_0^{1/p} - \varrho_1^{1/p}\|_{L^p(\Omega,\lambda)} \big(\|\varrho_0^{1/q}\|_{L^q(\Omega,\lambda)} + \|\varrho_1^{1/q}\|_{L^q(\Omega,\lambda)}\big) \\ &= c_p \mathsf{He}_p(\mu_0,\mu_1)(\mu_0(\Omega)^{1/q} + \mu_1(\Omega)^{1/q}). \end{aligned}$$

□

An interesting characterization of $\mathsf{He}_2$ in terms of $\mathsf{KL}$ is provided by the following property [32]:

**Proposition 2.6.** *For any two measures $\mu_0$ and $\mu_1$ in $\mathcal{M}(\Omega)$*

$$\mathsf{He}_2^2(\mu_0,\mu_1) = \min_{\mu \in \mathcal{M}(\Omega)} \mathsf{KL}(\mu,\mu_0) + \mathsf{KL}(\mu,\mu_1). \tag{28}$$

*In particular*

$$\mathsf{He}_2^2(\mu_0,\mu_1) \le \mathsf{KL}(\mu_0 \mid \mu_1). \tag{29}$$

*Proof.* Recalling (22) and (23b), (28) follows by the simple calculation

$$\min_{r\geq 0} H_{\mathsf{KL}}(r,r_0)+H_{\mathsf{KL}}(r,r_1)=r_0+r_1-2\sqrt{r_0r_1}=H_2^2(r_0,r_1),$$

attained at $r=\sqrt{r_0r_1}$. □

We now look at the Hellinger distance in its dual formulation. We focus on a 'static-dual' formulation first and then we proceed to the dynamic dual formulation in terms of subsolution of the equation $\partial\zeta_s+(p-1)\zeta_s^q=0$. This expression will play a crucial role in the contraction result of Proposition 5.1 and the regularizing estimates of Theorems 5.2 and 5.4. In the next computation we adopt the convention to write

$$x^a:=\begin{cases}x^a & \text{if } x>0,\\ 0 & \text{if } x=0,\\ -(-x)^a & \text{if } x<0,\end{cases}\qquad \text{for every } x\in\mathbb{R},\ a>0.$$

**Corollary 2.7.** *Let $p\in(1,\infty)$ and $q$ be the conjugate of $p$. The Hellinger distance admits the following dual formulation:*

$$\begin{aligned}\mathsf{He}_p^p(\mu_0,\mu_1)=\sup\Big\{\int_\Omega\psi_1\,\mathrm{d}\mu_1+\int_\Omega\psi_0\,\mathrm{d}\mu_0:\psi_0,\psi_1\in \mathrm{B}_b(\Omega)\\ \psi_0,\psi_1<1,\ (1-\psi_0^{q-1})(1-\psi_1^{q-1})\geq 1\Big\}.\end{aligned}\tag{30}$$

*Proof.* The result is a consequence of Theorem 2.3 and the computation of the convex set $\mathfrak{F}_p$ associated to the perspective function $F_p$ of (25); it is sufficient to prove that

$$\mathfrak{F}_{\mathfrak{p}}=\{(\psi_0,\psi_1)\in\mathbb{R}^2:\psi_i<1,\ (1-\psi_0^{q-1})(1-\psi_1^{q-1})\geq 1\}.\tag{31}$$

In order to show (31) we first compute the Legendre transform of $F_p$, obtaining

$$F_p^*(\psi)=\sup_{r>0} r\psi-|r^{1/p}-1|^p=\sup_{s>0}s^p\psi-|s-1|^p=\begin{cases}\dfrac{\psi}{(1-\psi^{q-1})^{p-1}} & \text{if } \psi<1,\\ +\infty & \text{if } \psi\geq 1.\end{cases}$$

Recalling that $(q-1)(p-1)=1$, the inequality $-\psi_1\geq F_p^*(\psi_0)$ for $\psi_0,\psi_1\in\mathbb{R}$ is equivalent to

$$\psi_0<1\quad\text{and}\quad -\psi_1^{q-1}(1-\psi_0^{q-1})\geq\psi_o^{q-1}=1-(1-\psi_0^{q-1}).$$

We then obtain

$$\begin{aligned}(\psi_0,\psi_1)\in\mathfrak{F}_p\quad&\Leftrightarrow\quad \psi_1\leq -F_p^*(\psi_0)\\ &\Leftrightarrow\quad \psi_0<1,\ \psi_1<1,\ (1-\psi_0^{q-1})(1-\psi_1^{q-1})\geq 1,\end{aligned}$$

which yields (31). □

The dynamic counterpart of the dual formulation is outlined in the proposition below.

**Proposition 2.8.** *Let $p\in(1,+\infty)$ and let $q$ be the conjugate of $p$. For every $\mu_0$, $\mu_1$ in $\mathcal{M}(\Omega)$,*

$$\begin{aligned}\mathsf{He}_p^p(\mu_0,\mu_1)=\sup\Big\{\int_\Omega\zeta_1\,\mathrm{d}\mu_1-\int_\Omega\zeta_0\,\mathrm{d}\mu_0:\\ \zeta\in \mathrm{C}^1([0,1],\mathrm{B}_b(\Omega)),\ \ \partial_t\zeta_t+(p-1)\zeta_t^q\leq 0\Big\}.\end{aligned}\tag{32}$$

*Proof.* First of all we manipulate the formulation (32) so that we can maximize with respect to one function only. We first observe that replacing, e.g. $\psi_i$ by $\psi_{i,\varepsilon} := \psi_i - \varepsilon$, $\varepsilon > 0$, the couple $(\psi_{0,\varepsilon}, \psi_{1,\varepsilon})$ is still admissible and

$$\sum_i \int_\Omega \psi_i \,\mathrm{d}\mu_i = \lim_{\varepsilon\downarrow 0} \sum_i \int_\Omega \psi_{i,\varepsilon} \,\mathrm{d}\mu_i,$$

so that it is not restrictive to assume $\sup \psi_i < 1$ in (30).

It is clear from the proof of Corollary 2.7 that for every choice of $\psi_0 \in \mathrm{B}_b(\Omega)$ satisfying $\sup \psi_0 < 1$ the best selection of $\psi_1$ in order to maximize $\sum_i \int_\Omega \psi_i \,\mathrm{d}\mu_i$ is given by

$$\psi_1 = -F_p^*(\psi_0) = \frac{-\psi_0}{\left(1 - \psi_0^{q-1}\right)^{p-1}}.$$

Setting $\zeta_0 := -\psi_0$ we obtain the formula

$$\mathsf{He}_p^p(\mu_0, \mu_1) = \sup_{\zeta_0 \in \mathrm{B}_b(\Omega),\ \zeta_0 > -1} \left( \int_\Omega \frac{\zeta_0}{\left(1 + \zeta_0^{q-1}\right)^{p-1}} \,\mathrm{d}\mu_1 - \int_\Omega \zeta_0 \,\mathrm{d}\mu_0 \right).$$

On the other hand we observe that the function $\zeta_1 := \frac{\zeta_0}{\left(1+\zeta_0^{q-1}\right)^{p-1}}$ corresponds to the solution at time $t = 1$ of

$$\begin{cases} \partial_t \zeta(t,x) + (p-1)\zeta^q(t,x) = 0 & \text{in } [0,1] \times \Omega, \\ \zeta(0,x) = \zeta_0(x) & \text{in } \Omega. \end{cases} \tag{33}$$

and by the comparison theorem for ordinary differential equation, any subsolution to (33) will satisfy $\zeta(1,x) \le \zeta_1(x)$. □

### 2.3. **Kantorovich-Wasserstein and Hellinger-Kantorovich distances.**

*Kantorovich-Wasserstein distance.* The standard definition of the Kantorovich - Wasserstein distance arises in a natural way in the frame of optimal transport. Here we recall the definition only and we refer to [2, 42] for further details.

We will deal with a complete and separable metric space $(X, \mathsf{d})$; we denote by $\mathscr{B}(X)$ its Borel $\sigma$-algebra and by $\mathcal{P}(X)$ the space of Borel probability measures on $X$. For $p \ge 1$ we set

$$\mathcal{P}_p(X) := \left\{ \mu \in \mathcal{P}(X) : \int_X \mathsf{d}^p(x,o) \,\mathrm{d}\mu(x) < +\infty \right\},$$

where $o$ is an arbitrary point of $X$ (the definition is independent of the choice of $o$).

If $\boldsymbol{t} : X \to Y$ is a Borel map between two metric spaces, we denote by $\boldsymbol{t}_\sharp : \mathcal{P}(X) \to \mathcal{P}(Y)$ the corresponding push-forward operation, defined by

$$\boldsymbol{t}_\sharp \mu(B) := \mu(\boldsymbol{t}^{-1}(B)) \quad \text{for every } B \in \mathscr{B}(Y).$$

In particular, when we consider the canonical cartesian projections $\pi^i : X \times X \to X$ defined by $\pi^i(x_0, x_1) := x_i$, $i = 0, 1$, and a general measure (also called transport plan) $\boldsymbol{\mu} \in \mathcal{P}(X \times X)$, the measures $\mu_i = \pi^i_\sharp \mu$ are the marginals of $\boldsymbol{\mu}$.

**Definition 2.9.** Let $p \in [1, \infty)$. For any $\mu_0, \mu_1 \in \mathcal{P}_p(X)$ the $p$-Kantorovich-Wasserstein distance is defined by

$$\mathsf{W}_p^p(\mu_0, \mu_1) := \min \left\{ \int \mathsf{d}^p(x_0, x_1) \,\mathrm{d}\boldsymbol{\mu}(x_0, x_1) : \boldsymbol{\mu} \in \mathcal{P}(X \times X),\ \pi^i_\sharp \boldsymbol{\mu} = \mu_i \right\}.$$

As we will see, a key ingredient we will extensively use in our arguments is given by the dynamic dual formulation of the Wasserstein distance, in terms of the subsolutions of the Hamilton-Jacobi equation. Such a result, which has been formulated in different form by [36, 3, 6, 1], holds if $(X,\mathsf{d})$ is a *length space*, i.e. if for every $x_0, x_1 \in X$ and every $\theta > 1/2$ there exists an approximate mid-point $x_\theta \in X$ such that

$$\max\big(\mathsf{d}(x_0,x_\theta),\mathsf{d}(x_\theta,x_1)\big) \le \theta\mathsf{d}(x_0,x_1).$$

We denote by $\mathrm{Lip}_b(X)$ the Banach space of bounded Lipschitz functions $f : X \to \mathbb{R}$ endowed with the norm

$$\|f\|_{\mathrm{Lip}_b} := \sup_{x\in X}|f| + \mathrm{Lip}(f,X), \quad \mathrm{Lip}(f,X) := \sup_{x,y\in X,\ x\neq y}\frac{|f(x)-f(y)|}{\mathsf{d}(x,y)}.$$

**Proposition 2.10.** *If $(X,\mathsf{d})$ is a length space then for every $\mu_0,\mu_1 \in \mathcal{P}_p(X)$*

$$\begin{aligned}\frac1p\mathsf{W}_p^p(\mu_0,\mu_1) = \sup\Big\{\int_X \zeta_1\,\mathrm{d}\mu_1 - \int_X \zeta_0\,\mathrm{d}\mu_0 :& \\ \zeta\in \mathrm{C}^1([0,1],\mathrm{Lip}_b(X))\ \textit{s.t.}\ \partial_t\zeta_t + \frac1q\,|\mathrm{D}\zeta_t|^q \le 0\Big\},&\end{aligned} \tag{34}$$

*where $q$ is the conjugate of $p$.*

*Proof.* Let $\mu_0,\mu_1 \in \mathcal{P}_p(X)$; since $(X,\mathsf{d})$ is a length space, also $(\mathcal{P}_p(X),\mathsf{W}_p)$ is a length space, so that for every $a > 1$ we can find a Lipschitz curve $\mu : [0,1] \to \mathcal{P}_p(X)$ such that

$$|\dot\mu_t|_{\mathsf{W}_p} := \limsup_{h\to0}\frac{\mathsf{W}_p(\mu_t,\mu_{t+h})}{|h|} \le a\mathsf{W}_p(\mu_0,\mu_1) \quad \text{for every } t\in[0,1]. \tag{35}$$

It follows that for every curve $\zeta \in \mathrm{C}^1([0,1],\mathrm{Lip}_b(X))$ the map $t \mapsto \int_X \zeta_t\,\mathrm{d}\mu_t$ is Lipschitz continuous and by [6, Lemma 6.4, Theorem 6.6]

$$\int_X \zeta_1\,\mathrm{d}\mu_1 - \int_X \zeta_2\,\mathrm{d}\mu_2 \le \int_0^1\int_X\Big(\partial_t\zeta_t + \frac1q|\mathrm{D}\zeta_t|^q(X)\Big)\,\mathrm{d}\mu_t\,\mathrm{d}t + \frac1p\int_0^1|\dot\mu_t|^p_{\mathsf{W}_p}\,\mathrm{d}t;$$

if $\zeta$ is also a subsolution to the Hamilton-Jacobi equation

$$\partial_t\zeta + \frac1q|\mathrm{D}\zeta|^q \le 0 \quad \text{in } [0,1]\times X, \tag{36}$$

then the previous inequality, the bound (35) on the metric velocity $|\dot\mu_t|_{\mathsf{W}_p}$ and the arbitrariness of $a > 1$ yield

$$\int_X \zeta_1\,\mathrm{d}\mu_1 - \int_X \zeta_2\,\mathrm{d}\mu_2 \le \frac1p\mathsf{W}_p^p(\mu_0,\mu_1) \quad \text{for every } \zeta\in\mathrm{C}^1([0,1];\mathrm{Lip}_b(X)) \text{ as in (36)}.$$

On the other hand, for every $a < 1$ we can use the Hopf-Lax semigroup

$$\mathsf{Q}_t\zeta(x) := \inf_{y\in X}\frac{1}{qt^{q-1}}\mathsf{d}^q(x,y) + \zeta(y)$$

and Kantorovich duality for the Wasserstein distance to find $\zeta_0 \in \mathrm{Lip}_b(X)$ such that

$$\int_X \mathsf{Q}_1\zeta_0\,\mathrm{d}\mu_1 - \int_X \zeta_0\,\mathrm{d}\mu_0 \ge \frac ap\mathsf{W}_p^p(\mu_0,\mu_1).$$

Using the refined estimate on the Hopf-Lax semigroup of [3] we can show that $\zeta_t := \mathsf{Q}_t\zeta_0$ is uniformly bounded in $\mathrm{Lip}_b(X)$, is Lipschitz continuous with values in $\mathrm{C}_b(X)$ and satisfies

$$\partial_t^+\zeta + \frac1q|\mathrm{D}\zeta|^q \le 0 \quad \text{in } [0,1)\times X, \quad \partial_t^+\zeta_t(x) = \lim_{h\downarrow0}h^{-1}\big(\zeta_{t+h}(x)-\zeta_t(x)\big).$$

By using a rescaling argument of [1] and the smoothing technique of the proof of [32, Theorem 8.12] we conclude. □

*Hellinger-Kantorovich distance.* After Hellinger-Kakutani and Kantorovich-Wasserstein distances, we recall the definition of a third distance between probability measures, that plays a role in the main contributions of this work.

Let $(X, \mathsf{d})$ be a separable complete metric space. The Hellinger-Kantorovich distances are defined on the space of finite nonnegative Borel measures $\mathcal{M}(X)$ and they do not require measures to have the same mass. As in the previous cases of $\mathsf{He}_p$ or $\mathsf{W}_p$, the Hellinger-Kantorovich distances admit different formulations that we summarize below. Here we focus on the family of distances $\mathsf{HK}_\alpha$ depending on a tuning parameter $\alpha > 0$; they correspond to the case $\mathsf{HK}_{\alpha,\beta}$ of [31] with the choice $\beta := 4$. In the even more specific case $\alpha = 1$, $\mathsf{HK}_1$ coincides with the distance $\mathsf{HK}$ which has been extensively studied in [32]. The general case $\alpha \neq 1$ can be reduced to the case $\alpha = 1$ by rescaling the distance $\mathsf{d}$ by a factor $\alpha^{-1/2}$.

The first formulation comes from the Logarithmic-Entropy-Transport problem, where the constraints on the marginals typical of optimal transport problems (2.9) are relaxed by the introduction of two penalizing functionals. The primal formulation of the Hellinger-Kantorovich distance is the following:

**Definition 2.11.** For any $\mu_0\, \mu_1 \in \mathcal{M}(X)$,

$$\mathsf{HK}^2_\alpha(\mu_0, \mu_1) := \min \Big\{ \sum_i \mathsf{KL}(\gamma_i|\mu_i) + \int \ell_\alpha(\mathsf{d}(x_0, x_1))\, d\boldsymbol{\gamma} : \\ \boldsymbol{\gamma} \in \mathcal{M}(X \times X),\ \pi^i_\sharp \boldsymbol{\gamma} = \gamma_i \ll \mu_i, \quad i = 0, 1 \Big\},$$

where $\ell_\alpha : [0, +\infty[ \to [0, +\infty]$ is the cost function defined by (15).

A direct comparison with (28) by restricting $\boldsymbol{\gamma}$ to plans of the form $\boldsymbol{\gamma} := \iota_\sharp \mu$ where $\mu \in \mathcal{M}(X)$ is an arbitrary measure dominating $\mu_i$ and $\iota : X \to X \times X$ is the diagonal identity map $\iota(x) := (x, x)$, immediately yields

$$\mathsf{HK}_\alpha(\mu_0, \mu_1) \le \mathsf{He}_2(\mu_0, \mu_1) \quad \text{for every } \mu_0, \mu_1 \in \mathcal{M}(X) \text{ and } \alpha > 0. \tag{37}$$

[32, Theorem 7.22] also shows that

$$\lim_{\alpha \downarrow 0} \mathsf{HK}_\alpha(\mu_0, \mu_1) = \mathsf{He}_2(\mu_0, \mu_1).$$

[32, Proposition 7.23, Theorem 7.24] provide two further useful bounds of $\mathsf{HK}_\alpha$ in terms of $\mathsf{W}_2$, when $\mu_0, \mu_1 \in \mathcal{P}_2(X)$:

$$\sqrt{\alpha}\mathsf{HK}_\alpha(\mu_0, \mu_1) \le \mathsf{W}_2(\mu_0, \mu_1), \qquad \lim_{\alpha \uparrow +\infty} \sqrt{\alpha}\mathsf{HK}_\alpha(\mu_0, \mu_1) = \mathsf{W}_2(\mu_0, \mu_1). \tag{38}$$

The $\mathsf{HK}_\alpha$ distance admits an equivalent dual formulation in terms of subsolutions to a suitable version of the Hamilton-Jacobi equation, which can be compared with (32) and (34): in fact, it is possible to show [32, Section 8.4] that

$$\mathsf{HK}^2_\alpha(\mu_0, \mu_1) = \sup \Big\{ \int_X \zeta_1 \,\mathrm{d}\mu_1 - \int_X \zeta_0 \,\mathrm{d}\mu_0 : \ \zeta \in \mathrm{C}^1([0,1], \mathrm{Lip}_b(X)) \text{ s.t.} \\ \partial_t \zeta_t + \frac{\alpha}{4} |\mathrm{D}\zeta_t|^2 + \zeta_t^2 \le 0 \text{ in } [0,1] \times X \Big\}. \tag{39}$$

3. **Metric measure spaces with curvature bounds.** This section is dedicated to a brief review of a few notions related to calculus and Sobolev spaces in metric measure spaces. We refer to [3] and [4] for a complete treatment of the topic.

3.1. **Calculus in metric measure spaces: basic notions.** Let $(X, \mathsf{d})$ be a complete and separable metric space, endowed with a Borel positive measure $\mathfrak{m}$ satisfying the growth condition (1) and $\operatorname{supp}(\mathfrak{m}) = X$. As we already mentioned in the Introduction, on this class of metric measure space it is possible to introduce an effective metric counterpart of the classic Dirichlet energy form in Euclidean spaces and of the corresponding Sobolev spaces. In the following, we will recall the basic notions only, which are strictly necessary to understand the main results of the work, by adopting the Cheeger point of view.

**Definition 3.1.** A function $G \in L^2(X, \mathfrak{m})$ is a relaxed gradient of $f \in L^2(X, \mathfrak{m})$ if there exist Borel $\mathsf{d}$-Lipschitz functions $f_n \in L^2(X, \mathfrak{m})$ such that:

a) $f_n \to f$ in $L^2(X, \mathfrak{m})$ and $|\mathrm{D} f_n|$ weakly converge to $\tilde{G}$ in $L^2(X, \mathfrak{m})$;
b) $\tilde{G} \le G$ $\mathfrak{m}$-a.e. in $X$.

We say that $G$ is the minimal relaxed gradient of $f$ if its $L^2(X, \mathfrak{m})$ norm is minimal among relaxed gradients. We shall denote by $|\mathrm{D}f|_w$ the minimal relaxed gradient.

The minimal relaxed gradient is used to give an integral formulation of the Cheeger energy (2), which can be represented as

$$\mathsf{Ch}(f) = \frac{1}{2}\int_X |\mathrm{D}f|_w^2 \,\mathrm{d}\mathfrak{m} \quad \text{if } f \text{ has a } L^2 \text{ relaxed gradient,}$$

and set equal to $+\infty$ if $f$ has no relaxed gradients. The Cheeger energy is a convex, 2-homogeneous lower semicontinuous functional on $L^2(X, \mathfrak{m})$ with dense domain $\mathcal{D}(\mathsf{Ch})$ [3, Th. 4.5]. From the lower semicontinuity of $\mathsf{Ch}$ it follows that the domain $\mathcal{D}(\mathsf{Ch})$ endowed with the norm

$$\|f\|_{\mathrm{W}^{1,2}} := \sqrt{\|f\|_2^2 + \||\mathrm{D}f|_w\|_2^2}$$

is a Banach space, which is called $\mathrm{W}^{1,2}(X, \mathsf{d}, \mathfrak{m})$. In general it is not a Hilbert space and this causes the potential non linearity of the heat flow. The following proposition summaries some useful properties of the minimal relaxed gradient, which will be helpful for our purposes.

**Proposition 3.2.** *Let $f \in L^2(X, \mathfrak{m})$. Then the following properties hold:*

*a) $|\mathrm{D}f|_w = |\mathrm{D}g|_w$ $\mathfrak{m}$-a.e. on $\{f - g = c\}$ for all constants $c \in \mathbb{R}$ and $g \in L^2(X, \mathfrak{m})$ with $\mathsf{Ch}(g) < +\infty$;*
*b) $\phi(f) \in \mathcal{D}(\mathsf{Ch})$ and $|\mathrm{D}\phi(f)|_w \le |\phi'(f)|\,|\mathrm{D}f|_w$ for any Lipschitz function $\phi$ on an interval $I$ containing the image of $f$; the inequality refines to the equality $|\mathrm{D}\phi(f)|_w = |\phi'(f)|\,|\mathrm{D}f|_w$ if in addition $\phi$ is nondecreasing;*
*c) if $f, g \in \mathcal{D}(\mathsf{Ch})$ and $\phi : \mathbb{R} \to \mathbb{R}$ is a nondecreasing contraction, then*

$$|\mathrm{D}(f + \phi(g - f))|_w^2 + |\mathrm{D}(g - \phi(g - f))|_w^2 \le |\mathrm{D}f|_w^2 + |\mathrm{D}g|_w^2 \qquad \mathfrak{m}\text{-a.e. in } X.$$

*Proof.* [3, Prop. 4.8] □

3.2. **Gradient flow of the Cheeger energy in metric-measure spaces.** The metric-measure counterpart of the Laplacian operator can be defined in terms of the element of minimal $L^2$-norm in the subdifferential $\partial\mathsf{Ch}$ of $\mathsf{Ch}$. $\partial\mathsf{Ch}$ is the multivalued operator in $L^2(X, \mathfrak{m})$ defined for all $f \in \mathcal{D}(\mathsf{Ch})$ by the following relation:

$$l \in \partial\mathsf{Ch}(f) \quad \Longleftrightarrow \quad \int_X l(g - f)\,\mathrm{d}\mathfrak{m} \le \mathsf{Ch}(g) - \mathsf{Ch}(f) \quad \forall\, g \in L^2(X, \mathfrak{m}).$$

**Definition 3.3** (Metric-measure Laplacian)**.** *The metric-measure Laplacian $\Delta f$ of $f \in L^2(X, \mathfrak{m})$ is defined for any $f$ such that $\partial\mathsf{Ch}(f) \neq \emptyset$. For those $f$, $-\Delta f$ is the element of minimal $L^2(X, \mathfrak{m})$ norm in $\partial\mathsf{Ch}(f)$.*

The domain of $\Delta$ is denoted by $\mathcal{D}(\Delta)$ and is a dense subset of $\mathcal{D}(\mathsf{Ch})$. The metric-measure heat flow can be obtained by applying the classic theory of gradient flows in Hilbert spaces [12] and it enjoys further properties which have been studied in [3]. More refined contraction properties will be proved in Section 4.

**Theorem 3.4** (Gradient flow of $\mathsf{Ch}$ in $L^2(X, \mathfrak{m})$)**.** *For any $f \in L^2(X, \mathfrak{m})$ there exists a unique locally absolutely continuous curve $(0, \infty) \ni t \to \mathsf{P}_t f \in L^2(X, \mathfrak{m})$ such that $\mathsf{P}_t f \to f$ in $L^2(X, \mathfrak{m})$ as $t \to 0$ and*

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{P}_t f \in -\partial\mathsf{Ch}(\mathsf{P}_t f) \qquad \text{for a.e. } t > 0.$$

*The following properties hold:*

*(1) The curve $t \mapsto \mathsf{P}_t f$ is locally Lipschitz, $\mathsf{P}_t f \in \mathcal{D}(\Delta)$ for any $t > 0$ and it holds*

$$\frac{\mathrm{d}^+}{\mathrm{d}t}\mathsf{P}_t f = \Delta\mathsf{P}_t f \qquad \forall t > 0.$$

*(2) The curve $t \to \mathsf{Ch}(\mathsf{P}_t f)$ is locally Lipschitz in $(0, +\infty)$, infinitesimal at $+\infty$ and continuous in $0$ if $f \in \mathcal{D}(\mathsf{Ch})$. Its right derivative is given by $-\|\Delta\mathsf{P}_t f\|^2_{L^2}$ for every $t > 0$.*

*(3) The family of maps $(\mathsf{P}_t)_{t\geq 0}$ is a strongly continuous semigroup of contractions in $L^2(X, \mathfrak{m})$ which can be extended in a unique way to a strongly continuous semigroup of contractions in every $L^p(X, \mathfrak{m})$, $1 \leq p < \infty$ (still denoted by $(\mathsf{P}_t)_{t\geq 0}$) thus satisfying*

$$\|\mathsf{P}_t f - \mathsf{P}_t g\|_{L^p(X,\mathfrak{m})} \leq \|f - g\|_{L^p(X,\mathfrak{m})} \quad \text{for every } f, g \in L^p(X, \mathfrak{m}). \tag{40}$$

3.3. $\mathrm{RCD}(K, \infty)$ **metric measure spaces.** In this subsection we briefly recall the definition and some properties of a class of metric measure spaces which generalize the notion of Riemannian manifolds with Ricci curvature bounded from below. This will be the general setting of the regularization result that we propose in Section 5, where, indeed, the bound on the curvature plays a direct role.

On a general metric measure space, the Cheeger energy is not a quadratic form and this translates into a potential lack of linearity of its $L^2$-gradient flow $(\mathsf{P}_\mathsf{t})_{t\geq 0}$. If we require the Cheeger energy to be quadratic, and hence the heat flow to be linear, we restrict the choice of the underlying metric domain to class of metric measure spaces which can be considered a nonsmooth generalization of Riemannian manifolds: among them, the so called Bakry-Émery curvature condition can be used to select the class of $\mathrm{RCD}(K, \infty)$ metric measure spaces (we refer to [3, 4] for a complete discussion and the other important equivalent characterization we mentioned in the Introduction). As in the previous section, $(X, \mathsf{d}, \mathfrak{m})$ will denote a complete and separable metric measure space satisfying the volume growth condition (1).

**Definition 3.5** (The $\mathrm{RCD}(K, \infty)$-condition)**.** $(X, \mathsf{d}, \mathfrak{m})$ is a $\mathrm{RCD}(K, \infty)$ metric measure space if the Cheeger energy is quadratic (7), every function $f \in \mathcal{D}(\mathsf{Ch}) \cap L^\infty(X, \mathfrak{m})$ with $|\mathrm{D}f|_w \leq 1$ admits a 1-Lipschitz representative (still denoted by $f$) and

$$|\mathrm{D}(\mathsf{P}_t f)|^2_w \leq \mathrm{e}^{-2Kt}\mathsf{P}_t(|\mathrm{D}f|^2_w) \qquad \mathfrak{m}\text{-a.e. in } X. \tag{41}$$

Equation (41) is one of the equivalent formulation of the celebrated Bakry-Émery condition [8], [5]. Notice that the $\mathrm{RCD}(K,\infty)$ condition implies in particular that every bounded function $f \in \mathcal{D}(\mathsf{Ch})$ with $|\mathrm{D}f|_w \in L^\infty(X,\mathfrak{m})$ has a Lipschitz continuous representative (identified with $f$) satisfying

$$\sup_X |\mathrm{D}f| = \mathrm{Lip}(f,X) \le \big\| \, |\mathrm{D}f|_w \, \big\|_{L^\infty(X,\mathfrak{m})}.$$

On $\mathrm{RCD}(K,\infty)$ spaces, an even stronger version of (41) holds true, together with crucial regularization properties which we collect in the next statement.

**Theorem 3.6.** *Let $(X,\mathsf{d},\mathfrak{m})$ be a $\mathrm{RCD}(K,\infty)$ space.*

*(1) For every $f \in L^\infty(X,\mathfrak{m})$ and $t > 0$ the function $\mathsf{P}_t f$ has a unique continuous representative $\tilde{\mathsf{P}}_t f \in \mathrm{Lip}_b(X)$ (in the following, with a slight abuse of notation, we will identify $\mathsf{P}_t f$ with $\tilde{\mathsf{P}}_t f$, whenever $f \in L^\infty(X,\mathfrak{m})$).*

*(2) For every $f \in \mathcal{D}(\mathsf{Ch})$ with $f, |\mathrm{D}f|_w \in L^\infty(X,\mathfrak{m})$ and $t > 0$*

$$|\mathrm{D}\mathsf{P}_t f| = |\mathrm{D}\mathsf{P}_t f|_w \ \mathfrak{m}\text{-a.e. in } X, \qquad |\mathrm{D}\mathsf{P}_t f| \le \mathrm{e}^{-Kt}\mathsf{P}_t |\mathrm{D}f|_w \quad \text{in } X. \tag{42}$$

*(3) For every $f \in L^\infty(X,\mathfrak{m})$ and $t > 0$*

$$R_K(t)|\mathrm{D}\mathsf{P}_t f|^2 \le \mathsf{P}_t(f^2) - (\mathsf{P}_t f)^2 \qquad \text{in } X, \tag{43}$$

*where $R_k$ has been defined in (14). In particular*

$$\sqrt{R_K(t)}\,\mathrm{Lip}(\mathsf{P}_t f, X) \le \|f\|_{L^\infty(X,\mathfrak{m})}.$$

*Proof.* Property (1) is a consequence of [5, Corollary 4.18]. The first identity of (42) is stated in [5, Theorem 3.17]; the second one is stated in [38, Corollary 4.3]. (23b) is a consequence of the above properties and the estimate of [5, Corollary 2.3(iv)]. □

## 4. **Contraction properties for the Heat flow in metric measure spaces.**

This section is devoted to some fairly general contraction properties of the heat flow in the metric-measure setting. Our main result concerns the behaviour of the functional

$$\mathscr{E}(f,g) := \int_X E(f,g)\,\mathrm{d}\mathfrak{m}, \quad f,g \in L^2(X,\mathfrak{m}) \tag{44a}$$

where

$$E : \mathbb{R}^2 \to \mathbb{R} \cup \{+\infty\} \text{ is a proper, l.s.c. and convex function.} \tag{44b}$$

Since $E$ is bounded from below by an affine map, when $\mathfrak{m}(X) < \infty$ the integral of (44a) is always well defined (possibly taking the value $+\infty$). In the general case, in order to avoid integrability issues, we will also assume that

$$E \text{ is nonnegative, } E(0,0) = 0 \quad \text{if } \mathfrak{m}(X) = +\infty. \tag{44c}$$

**Theorem 4.1.** *Let $(X,\mathsf{d},\mathfrak{m})$ be a metric measure space with the Heat semigroup $(\mathsf{P}_t)_{t\ge 0}$ generated by the Cheeger energy $\mathsf{Ch}$ in $L^2(X,\mathfrak{m})$, and let $\mathscr{E}$ be defined as in (44a,b,c). Then, for $f,g \in L^2(X,\mathfrak{m})$*

$$\mathscr{E}(\mathsf{P}_t f, \mathsf{P}_t g) \le \mathscr{E}(f,g) \quad \text{for every } t \ge 0.$$

We prove some useful lemmas first. The first one shows a generalization of part $c)$ in Proposition 3.2 and is the core of the proof of the main theorem.

**Lemma 4.2.** *Let* $(X, \mathsf{d})$ *be a metric space, let* $E : \mathbb{R}^2 \to \mathbb{R}$ *be a* $\mathrm{C}^2$ *convex function with* 1*-Lipschitz (w.r.t the Euclidean norm) gradient* $\nabla E : \mathbb{R}^2 \to \mathbb{R}^2$*, and let* $J : \mathbb{R}^2 \to \mathbb{R}^2$ *be the map* $J := Id - \nabla E$*. For every bounded Lipschitz map* $\boldsymbol{f} := (f_1, f_2) : \mathbb{R}^2 \to \mathbb{R}^2$*, the function* $\boldsymbol{g} = (g_1, g_2) := J \circ \boldsymbol{f}$ *satisfies*

$$|\mathrm{D}g_1|^2(x) + |\mathrm{D}g_2|^2(x) \le |\mathrm{D}f_1|^2(x) + |\mathrm{D}f_2|^2(x) \quad \textit{for every } x \in X. \tag{45}$$

*Proof.* Since $\nabla^2 E$ is positive definite and $J$ is 1-Lipschitz, we observe that $A := \mathrm{D}J = I - \nabla^2 E$ satisfies

$$0 \le \boldsymbol{z}^T A(\boldsymbol{w})\boldsymbol{z} \le |\boldsymbol{z}|^2 \quad \text{for every } \boldsymbol{w},\ \boldsymbol{z} \in \mathbb{R}^2. \tag{46}$$

For every $x, y \in X$, $x \neq y$, and $f : X \to \mathbb{R}$ we set

$$R(f, x, y) := \frac{|f(x) - f(y)|}{\mathsf{d}(x, y)} \quad \text{so that} \quad |\mathrm{D}f|(x) = \limsup_{y \to x} R(f, x, y).$$

Let us now fix $x \in X$; it is possible to find two sequences of points $(x_i^n)_{n \in \mathbb{N}}$, $i = 1, 2$, such that

$$\lim_{n \to +\infty} R(g_i, x, x_i^n) = |\mathrm{D}g_i|(x). \tag{47}$$

Taking a linear combination of the difference quotients $R(g_i, x, x_i^n)$ with the *positive* coefficients $v_i := |\mathrm{D}g_i|(x)$ it holds

$$\lim_{n \to +\infty} \sum_i v_i\, R(g_i, x, x_i^n) = \sum_i |\mathrm{D}g_i|^2(x).$$

Now, $g_i(x) = J_i(f_1(x), f_2(x))$ and hence

$$\lim_{n \to +\infty} v_i R(g_i, x, x_i^n) = \lim_{n \to +\infty} v_i \frac{|J_i(f_1(x), f_2(x)) - J_i(f_1(x_i^n), f_2(x_i^n))|}{\mathsf{d}(x, x_i^n)}.$$

Since $J$ is $\mathrm{C}^1$, a first order expansion at $\boldsymbol{z} = \boldsymbol{f}(x)$ with $\boldsymbol{z}_i^n := \boldsymbol{f}(x_i^n)$ and the Lipschitz character of $\boldsymbol{f}$ yield

$$\begin{aligned} J(\boldsymbol{z}_i^n) - J(\boldsymbol{z}) &= A(\boldsymbol{z})(\boldsymbol{z}_i^n - \boldsymbol{z}) + o(|\boldsymbol{z}^n - \boldsymbol{z}|) \\ &= \partial_1 J(\boldsymbol{f}(x))(f_1(x_i^n) - f_1(x)) + \partial_2 J(\boldsymbol{f}(x))(f_2(x_i^n) - f_2(x)) + o(\mathsf{d}(x_i^n, x)). \end{aligned}$$

Estimating the first component $J_1$ of $J$ along the sequence $(x_1^n)_n$ and the second component $J_2$ of $J$ along $(x_2^n)_n$ we get for $i = 1, 2$

$$\begin{aligned} &\lim_{n \to +\infty} \frac{|J_i(f_1(x), f_2(x)) - J_i(f_1(x_i^n), f_2(x_i^n))|}{\mathsf{d}(x, x_i^n)} \\ &= \lim_{n \to \infty} \Big|\partial_1 J_i(f_1(x), f_2(x)) R(f_1, x, x_i^n) + \partial_2 J_i(f_1(x), f_2(x)) R(f_2, x, x_i^n)\Big| \\ &\le |\partial_1 J_i(f_1(x), f_2(x))| \limsup_{n \to \infty} R(f_1, x, x_i^n) + |\partial_2 J_i(f_1(x), f_2(x))| \limsup_{n \to \infty} R(f_2, x, x_i^n) \\ &\le |\partial_1 J_i(f_1(x), f_2(x))||\mathrm{D}f_1|(x) + |\partial_2 J_i(f_1(x), f_2(x))||\mathrm{D}f_2|(x). \end{aligned}$$

Recalling (47), since the coefficients $v_1, v_2$ are nonnegative, we get

$$\sum_i |\mathrm{D}g_i|^2(x) \le \sum_{i,j} \hat{A}_{j,i}(f_1(x), f_2(x))|\mathrm{D}f_j|(x) v_i,$$

where for every $w \in \mathbb{R}^2$ $\hat{A}(\boldsymbol{w})$ is the symmetric matrix defined by

$$\hat{A}_{i,j}(\boldsymbol{w}) := |A_{i,j}(\boldsymbol{w})|.$$

(46) and the next elementary Lemma yield

$$\sum_i |\mathrm{D}g_i|^2(x) \le \Big(\sum_i |v_i|^2\Big)^{1/2}\Big(\sum_i |\mathrm{D}f_i|^2(x)\Big)^{1/2}$$

thus obtaining (45). □

**Lemma 4.3.** *Let $A \in \mathbb{R}^{2\times 2}$ be a symmetric matrix and let $\hat A \in \mathbb{R}^{2\times 2}$ be defined by $\hat A_{ij} := |A_{ij}|$, $i,j = 1,2$. If*

$$0 \le \boldsymbol{z}^T A \boldsymbol{z} \le |\boldsymbol{z}|^2 \quad \textit{for every } \boldsymbol{z} \in \mathbb{R}^2, \tag{48}$$

*then also $\hat A$ satisfies*

$$0 \le \boldsymbol{z}^T \hat A \boldsymbol{z} \le |\boldsymbol{z}|^2 \quad \textit{and} \quad \boldsymbol{z}^T \hat A \boldsymbol{w} \le |\boldsymbol{z}|\,|\boldsymbol{w}| \quad \textit{for every } \boldsymbol{z},\boldsymbol{w} \in \mathbb{R}^2. \tag{49}$$

*Proof.* It is easy to check that a symmetric matrix $A$ satisfies $0 \le \boldsymbol{z}^T A \boldsymbol{z} \le |\boldsymbol{z}|^2$ for every $\boldsymbol{z} \in \mathbb{R}^2$ (48) if and only if

$$0 \le A_{11} \le 1, \quad 0 \le A_{22} \le 1, \quad A_{12}^2 \le A_{11}A_{22}, \quad A_{12}^2 \le 1 + A_{11}A_{22} - A_{11} - A_{22}, \tag{50}$$

and it is clear that (50) is preserved if we replace the coefficients $A_{ij}$ by $|A_{ij}|$. The second inequality of (49) follows immediately by the first one and the Cauchy-Schwartz inequality, since

$$\boldsymbol{z}^T \hat A \boldsymbol{w} \le \big(\boldsymbol{z}^T \hat A \boldsymbol{z}\big)^{1/2}\big(\boldsymbol{w}^T \hat A \boldsymbol{w}\big)^{1/2} \le |\boldsymbol{z}|\,|\boldsymbol{w}|.$$

□

**Lemma 4.4.** *Let $E : \mathbb{R}^2 \to \mathbb{R}$ be a $\mathrm{C}^{1,1}$ convex function as in (44b) and (44c) with 1-Lipschitz (w.r.t the Euclidean norm) gradient $\nabla E : \mathbb{R}^2 \to \mathbb{R}^2$, and let $J : \mathbb{R}^2 \to \mathbb{R}^2$ be the map $J := Id - \nabla E$. For every couple bounded Lipschitz map $\boldsymbol{f} := (f_1, f_2)$ with $f_i \in \mathrm{W}^{1,2}(X,\mathsf{d},\mathfrak{m})$, the components $g_i$ of $\boldsymbol{g} := J \circ \boldsymbol{f}$ belong to $\mathrm{W}^{1,2}(X,\mathsf{d},\mathfrak{m})$ and satisfy*

$$|\mathrm{D}g_1|_w^2(x) + |\mathrm{D}g_2|_w^2(x) \le |\mathrm{D}f_1|_w^2(x) + |\mathrm{D}f_2|_w^2(x) \quad \textit{for } \mathfrak{m}\textit{-a.e. } x \in X. \tag{51}$$

*Proof.* Let us consider the case when $\mathfrak{m}(X) = +\infty$ (the case of a finite measure is simper, and it follows by obvious modifications of the arguments below): notice that (44c) yields $\nabla E(0,0) = 0$.

Let us first notice that $|J(\boldsymbol{f})| \le 2|\boldsymbol{f}|$ so that for every $f_i \in L^2(X,\mathsf{d},\mathfrak{m})$ the functions $g_i$ belong to $L^2(X,\mathsf{d},\mathfrak{m})$ as well.

Let us first prove that

$$|\mathrm{D}g_1|_w^2(x) + |\mathrm{D}g_2|_w^2(x) \le |\mathrm{D}f_1|^2(x) + |\mathrm{D}f_2|^2(x) \quad \text{for } \mathfrak{m}\text{-a.e. } x \in X, \tag{52}$$

whenever $f_1, f_2$ are bounded and Lipschitz and $E$ is of class $C^{1,1}$. To this aim, it is sufficient to regularize $\nabla E$ e.g. by convolution with a family of smooth kernels $\kappa_n : \mathbb{R}^2 \to [0,+\infty[$, $n \in \mathbb{N}$ satisfying

$$\kappa \in \mathrm{C}^\infty_c(\mathbb{R}^2), \quad \kappa \ge 0, \quad \kappa(-z) = \kappa(z), \quad \int_{\mathbb{R}^2} \kappa\,\mathrm{d}x = 1, \quad \kappa_n(z) := n^2\kappa(nz) \quad z \in \mathbb{R}^2.$$

We then set

$$\begin{aligned} E_n(x) &:= \int_{\mathbb{R}^2} \Big(E(x-z) - x\cdot\nabla E(-z)\Big)\kappa_n(z)\,\mathrm{d}z, \\ \nabla E_n(x) &= \int_{\mathbb{R}^2} \Big(\nabla E(x-z) - \nabla E(-z)\Big)\kappa_n(z)\,\mathrm{d}z, \\ J_n(x) &:= x - \nabla E_n(x), \quad \boldsymbol{g}_n := J_n \circ \boldsymbol{f}. \end{aligned} \tag{53}$$

Applying Lemma 4.2 we get

$$|\mathrm{D}g_{n,1}|^2(x) + |\mathrm{D}g_{n,2}|^2(x) \le |\mathrm{D}f_1|^2(x) + |\mathrm{D}f_2|^2(x).$$

Since

$$\begin{aligned}|J_n(x) - J(x)| &\le |\nabla E_n(x) - \nabla E(x)| \\ &\le \int_{\mathbb{R}^2} \Big(\big|\nabla E(x-z) - \nabla E(x)\big| + \big|\nabla E(0) - \nabla E(-z)\big|\Big)\kappa_n(z)\,\mathrm{d}z \\ &\le 2\int_{\mathbb{R}^2} |z|\kappa_n(z)\,\mathrm{d}z = \frac{2}{n}\int_{\mathbb{R}^2}|z|\kappa(z)\,\mathrm{d}z,\end{aligned}$$

passing to the limit as $n \to \infty$ we have $g_{n,i} \to g_i$ in $L^2(X, \mathfrak{m})$; up to the extraction of a suitable subsequence (not relabelled) we can also assume that

$$|\mathrm{D}g_{n,1}| \rightharpoonup G_1, \quad |\mathrm{D}g_{n,2}| \rightharpoonup G_2 \quad \text{weakly in } L^2(X, \mathfrak{m}) \text{ as } n \to \infty.$$

We claim that

$$G_1^2 + G_2^2 \le |\mathrm{D}f_1|^2 + |\mathrm{D}f_2|^2 \quad \mathfrak{m}\text{-a.e. in } X. \tag{54}$$

In fact, for an arbitrary measurable set $A \subset X$ we have

$$\begin{aligned}\int_A \Big(G_1^2 + G_2^2\Big)\,\mathrm{d}\mathfrak{m} &= \lim_{n\to\infty}\int_A \Big(|\mathrm{D}g_{n,1}|G_1 + |\mathrm{D}g_{n,2}|G_2\Big)\,\mathrm{d}\mathfrak{m} \\ &\le \Big(\int_A |\mathrm{D}f_1|^2 + |\mathrm{D}f_2|^2\,\mathrm{d}\mathfrak{m}\Big)^{1/2}\Big(\int_A \Big(G_1^2 + G_2^2\Big)\,\mathrm{d}\mathfrak{m}\Big)^{1/2}\end{aligned}$$

so that for every measurable set $A \subset X$

$$\int_A \Big(G_1^2 + G_2^2\Big)\,\mathrm{d}\mathfrak{m} \le \int_A \Big(|\mathrm{D}f_1|^2 + |\mathrm{D}f_2|^2\Big)\,\mathrm{d}\mathfrak{m}.$$

Since $|\mathrm{D}g_i|_w \le G_i$, (54) yields (52).

(51) then follows by (52) by a similar argument: we select optimal sequences $(f_{i,n})_n$ of bounded Lipschitz functions converging to $f_i$ in $L^2(X, \mathfrak{m})$ such that

$$|\mathrm{D}f_{i,n}| \to |\mathrm{D}f_i|_w \quad \text{in } L^2(X, \mathfrak{m}), \quad i = 1, 2,$$

and we consider the corresponding sequences $\boldsymbol{g}_n = J \circ \boldsymbol{f}_n$, converging to $\boldsymbol{g} = J \circ \boldsymbol{f}$ in $L^2(X, \mathfrak{m})$. We then pass to the limit in the inequality

$$|\mathrm{D}g_{1,n}|_w^2(x) + |\mathrm{D}g_{2,n}|_w^2(x) \le |\mathrm{D}f_{1,n}|^2(x) + |\mathrm{D}f_{2,n}|^2(x) \quad \text{for } \mathfrak{m}\text{-a.e. } x \in X.$$

□

Next lemma focuses on a useful property of the metric Laplacian which relies on the estimate that we have just proved.

**Lemma 4.5.** *If $f, g \in \mathcal{D}(\Delta)$ and $E : \mathbb{R}^2 \to \mathbb{R}$ is a $\mathrm{C}^{1,1}$ convex function satisfying (44c), then*

$$\int_X (\partial_1 E(f, g)\Delta f + \partial_2 E(f, g)\Delta g)\,\mathrm{d}\mathfrak{m} \le 0. \tag{55}$$

*Proof.* It is not restrictive to assume that $E$ is 1-Lipschitz. As we observed in the proof of Lemma 4.4, we also note that $\partial_i E(f, g)$ belongs to $L^2(X, \mathfrak{m})$, since when $\mathfrak{m}(X) = +\infty$ (44c) yields $\partial_i E(0, 0)$; therefore the integral in (55) is well defined. Recall that

$$l \in \partial\mathsf{Ch}(\varphi) \quad \Longleftrightarrow \quad \int_X l(\psi - \varphi)\,\mathrm{d}\mathfrak{m} \le \mathsf{Ch}(\psi) - \mathsf{Ch}(\varphi) \quad \text{for every } \psi \in L^2(X, \mathfrak{m})$$

and that $-\Delta\varphi \in \partial\mathsf{Ch}(\varphi)$. Hence taking in our case $\varphi = f$ and $\psi = f - \partial_1 E(f,g)$ we get

$$\int_X \partial_1 E(f,g)\Delta f \,\mathrm{d}\mathfrak{m} = \int_X -\Delta f(\psi-\varphi)\,\mathrm{d}\mathfrak{m} \le \mathsf{Ch}(\psi) - \mathsf{Ch}(\varphi) = \mathsf{Ch}(f - \partial_1 E(f,g)) - \mathsf{Ch}(f),$$

and similarly

$$\int_X \partial_2 E(f,g)\Delta g \,\mathrm{d}\mathfrak{m} \le \mathsf{Ch}(g - \partial_2 E(f,g)) - \mathsf{Ch}(g).$$

By definition of the Cheeger functional and Lemma 4.4 we obtain (55). □

With the previously developed tools we can conclude the proof of Theorem 4.1.

*Proof.* Let us set $f_t = \mathsf{P}_t f$ and $g_t = \mathsf{P}_t g$. Assume first that $E$ is $\mathrm{C}^{1,1}$ with Lipschitz gradient $\nabla E$, so that $E$ has at most quadratic growth. Recalling that $t \mapsto f_t, g_t$ are differentiable as $L^2$-valued maps, we get

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathscr{E}(f_t,g_t) &= \int_X \frac{\mathrm{d}}{\mathrm{d}t}E(f_t,g_t)\,\mathrm{d}\mathfrak{m} = \Big(\int_X \partial_1 E(f_t,g_t)\Delta_{\mathsf{d},\mathfrak{m}} f_t + \partial_2 E(f_t,g_t)\Delta_{\mathsf{d},\mathfrak{m}} g_t\Big)\,\mathrm{d}\mathfrak{m}\\ &= \int_X \Big(\partial_1 E(f_t,g_t)\Delta_{\mathsf{d},\mathfrak{m}} f_t + \partial_2 E(f_t,g_t)\Delta_{\mathsf{d},\mathfrak{m}} g_t\Big)\,\mathrm{d}\mathfrak{m} \le 0\end{aligned}$$

thanks to (55). We thus obtain

$$\mathscr{E}(\mathsf{P}_t f, \mathsf{P}_t g) \le \mathscr{E}(f,g) \quad \text{for every } t \ge 0. \tag{56}$$

In the general case, we apply (56) to the functional $\mathscr{E}_\lambda$ associated to the Yosida approximation $E_\lambda$ of $E$,

$$E_\lambda(r,s) := \inf_{(r',s')\in\mathbb{R}^2} \frac{1}{2\lambda}\Big((r'-r)^2 + (s'-s)^2\Big) + E(r',s') \quad (r,s)\in\mathbb{R}^2,\ \lambda>0. \tag{57}$$

It is well known [12] that $E_\lambda$ is convex of class $C^{1,1}$ with Lipschitz gradient $\nabla E_\lambda$; moreover, if (44c) holds, then also $E_\lambda$ is nonnegative and it is immediate to check from (57) that $E_\lambda(0,0) = 0$. Since $E_\lambda \le E$, (56) then yields

$$\mathscr{E}_\lambda(\mathsf{P}_t f, \mathsf{P}_t g) \le \mathscr{E}_\lambda(f,g) \le \mathscr{E}(f,g) \quad \text{for every } t \ge 0,\ \lambda > 0.$$

We can eventually pass to the limit as $\lambda \downarrow 0$ and applying Beppo Levi monotone convergence theorem, since $E_\lambda(r,s) \uparrow E(r,s)$ as $\lambda \downarrow 0$ for every $r, s \in \mathbb{R}^2$. □

A few particular cases follow as corollaries of the main result. The first one states the contraction in the Hellinger metric for measures which are absolutely continuous w.r.t. $\mathfrak{m}$: with a slight abuse of notation, for every $f, g \in L^1(X,\mathfrak{m})$, $f, g \ge 0$, we will set

$$\mathsf{He}_p^p(f,g) := \mathsf{He}_p(f\mathfrak{m}, g\mathfrak{m}) = \int_X \left|f^{1/p} - g^{1/p}\right|^p \mathrm{d}\mathfrak{m}.$$

**Corollary 4.6.** *For every nonnegative $f, g \in L^1(X,\mathfrak{m})$ we have*

$$\mathsf{He}_p(\mathsf{P}_t f, \mathsf{P}_t g) \le \mathsf{He}_p(f,g) \quad \textit{for every } t \ge 0. \tag{58}$$

*Proof.* It is sufficient to prove (58) for every couple of nonnegative functions $f, g \in L^1 \cap L^2(X,\mathfrak{m})$ and then argue by approximation using (40) for $p = 1$. We can then apply Theorem 4.1 with the function $E : \mathbb{R}^2 \to \mathbb{R} \cup \{+\infty\}$ given by

$$E(r,s) := \begin{cases} \left|r^{1/p} - s^{1/p}\right|^p & \text{if } r, s \ge 0,\\ +\infty & \text{otherwise.}\end{cases}$$

□

More generally, the same contraction result holds true for any Csiszàr divergence; recalling the discussion of Section 2.1 and keeping the same notation of (18a), (18b), (18c) and Definition 2.1, we first set

$$\mathscr{F}(f|g) := \mathscr{F}(f\mathfrak{m} \mid g\mathfrak{m}) = \int_X H(f,g)\,\mathrm{d}\mathfrak{m} \quad \text{for every nonnegative } f,g \in L^1(X,\mathfrak{m}).$$

**Corollary 4.7.** *Let $\mathscr{F}$ be a Csiszàr divergence as in Definition 2.1. Then, for every nonnegative $f,g \in L^1(X,\mathfrak{m})$,*

$$\mathscr{F}(\mathsf{P}_t f \mid \mathsf{P}_t g) \le \mathscr{F}(f \mid g).$$

*Proof.* Recalling (21), it is sufficient to apply Theorem 4.1 to the integral functional associated to the function $H$ of (18b), extended to $+\infty$ if $r<0$ or $s<0$.

□

## 5. Regularizing properties of the Heat flow in RCD metric measure spaces.

In the previous section we have shown contraction results involving convex functionals and metric heat flows in metric measure spaces, thus covering the case of nonlinear flows in Finsler-like geometries.

In the linear case, the Hellinger contraction (58) can also be proved by a different approach, based on the dual dynamic formulation of the Hellinger distance that we have discussed in 2.8. We first explain this technique in the simple case of a submarkovian operator $\mathsf{P}$ on the set of bounded measurable functions and we will then show how to extend this approach to prove new regularization results for the Heat semigroup in RCD metric measure spaces.

### 5.1. Hellinger contraction for submarkovian operators.

Let $(\Omega,\mathscr{B})$ be a measurable space and let $\mathsf{P} : \mathrm{B}_b(\Omega) \to \mathrm{B}_b(\Omega)$ be a linear submarkovian operator [18, Chap. IX, Sect. 1]: this means that for every bounded measurable maps $f, f_n \in \mathrm{B}_b(\Omega)$

$$0 \le f \le 1 \quad \Rightarrow \quad 0 \le \mathsf{P}f \le 1, \tag{59a}$$

$$f_n \ge 0,\ f_n \downarrow 0 \text{ as } n \to \infty \quad \Rightarrow \quad \mathsf{P}f_n \downarrow 0, \tag{59b}$$

where convergence in (59b) has to be intended pointwise everywhere. Notice that for every $x,y \in \Omega$

$$\begin{aligned}0 \le \mathsf{P}\Big((f-\mathsf{P}f(y))^2\Big)(x) &= \mathsf{P}f^2(x) - 2\mathsf{P}f(y)\mathsf{P}f(x) + (\mathsf{P}f(y))^2\mathsf{P}1\\ &\le \mathsf{P}f^2(x) - \big(\mathsf{P}f(x)\big)^2 + \big(\mathsf{P}f(x)-\mathsf{P}f(y)\big)^2\end{aligned}$$

so that choosing $x=y$ we get the Jensen's inequality

$$(\mathsf{P}f)^2 \le \mathsf{P}f^2. \tag{60}$$

We can define the adjoint operator $\mathsf{P}^*$ acting on $\mathcal{M}(\Omega)$ by the formula

$$\int_\Omega f\,\mathrm{d}\mathsf{P}^*\mu := \int_\Omega \mathsf{P}f\,\mathrm{d}\mu \quad \text{for every } f \in \mathrm{B}_b(\Omega).$$

The next result could also be derived by a more refined Jensen inequality for submarkovian operator. Here we want to highlight the role of the dual dynamic point of view.

**Proposition 5.1.** *Let* $(\Omega, \mathscr{B})$ *be a measure space and let* $\mathsf{P}$ *be a linear submarkovian operator in* $\mathrm{B}_b(\Omega)$. *Then, for any* $\mu_0, \mu_1 \in \mathcal{M}(\Omega)$

$$\mathsf{He}_2(\mathsf{P}^*\mu_0, \mathsf{P}^*\mu_1) \le \mathsf{He}_2(\mu_0, \mu_1). \tag{61}$$

*Proof.* Let us consider $(\zeta_s)_{s\in[0,1]} \in \mathrm{C}^1([0,1], \mathrm{B}_b(\Omega))$ a solution of

$$\partial_s \zeta_s + \zeta_s^2 \le 0 \quad \text{ in } \Omega \times [0,1]. \tag{62}$$

We apply the map $\mathsf{P}$ to this solution; since the linear map $\mathsf{P}$ is continuous with respect to the supremum norm in $\mathrm{B}_b(\Omega)$, $(\mathsf{P}\zeta_s)_s \in \mathrm{C}^1([0,1], \mathrm{B}_b(\Omega))$. Moreover, from (60) applied to $\zeta_s$ it follows that $s \to \mathsf{P}\zeta_s$ is also a subsolution to (62):

$$\partial_s \mathsf{P}\zeta_s + (\mathsf{P}\zeta_s)^2 \le \mathsf{P}\partial_s\zeta_s + \mathsf{P}(\zeta_s^2) = \mathsf{P}(\partial_s\zeta_s + \zeta_s^2) \le 0,$$

since $\mathsf{P}$ is positivity preserving. Then, recalling the formulation (32) of the Hellinger distance, we have

$$\int_\Omega \zeta_1 \,\mathrm{d}(\mathsf{P}^*\mu_1) - \int_\Omega \zeta_0 \,\mathrm{d}(\mathsf{P}^*\mu_0) = \int_\Omega \mathsf{P}\zeta_1 \,\mathrm{d}\mu_1 - \int_\Omega \mathsf{P}\zeta_0 \,\mathrm{d}\mu_0 \le \mathsf{He}^2(\mu_0, \mu_1).$$

Taking the supremum of the left hand side with respect to all the subsolutions of (62) and applying (32) once more, we eventually get (61). □

**Remark 1.** The same argument combined with the $p$-Jensen inequality for $\mathsf{P}$ yields

$$\mathsf{He}_p(\mathsf{P}^*\mu_0, \mathsf{P}\mu_1) \le \mathsf{He}_p(\mu_0, \mu_1),$$

for every $p \in [1, +\infty)$. The proof can also be extended to submarkovian operators in $L^1(\Omega, \mathfrak{m})$ with respect to a given reference measure $\mathfrak{m} \in \mathcal{M}(\Omega)$, obtaining in this case an Hellinger estimate for measures absolutely continuous w.r.t. $\mathfrak{m}$.

5.2. **Regularization $\mathsf{W}_p$- $\mathsf{He}_p$ for $p \in (1,2]$.** Let us now focus on the regularization estimates for the particular class of Markovian operators provided by the heat semigroup $(\mathsf{P}_t)_{t\ge0}$ in a metric measure space $(X, \mathsf{d}, \mathfrak{m})$ satisfying the $\mathrm{RCD}(K,\infty)$ condition. Since $(\mathsf{P}_t)_{t\ge0}$ maps $\mathrm{C}_b(X)$ into $\mathrm{C}_b(X)$, we can use (12) to define the adjoint heat semigroup $(\mathsf{P}_t^*)_{t\ge0}$ on arbitrary positive and finite measure of $\mathcal{M}(X)$ (see [5, Section 3.2] for more details).

**Theorem 5.2.** *Let* $(X, \mathsf{d}, \mathfrak{m})$ *be a* $\mathrm{RCD}(K,\infty)$ *metric measure space and* $p \in [1,2]$. *Then, for every* $\mu_0, \mu_1 \in \mathcal{P}_p(X)$

$$\mathsf{He}_p(\mathsf{P}_t^*\mu_0, \mathsf{P}_t^*\mu_1) \le \frac{1}{p(R_K(t))^{1/2}} \mathsf{W}_p(\mu_0, \mu_1) \qquad \text{for all } t > 0, \tag{63}$$

*where* $R_K$ *has been defined in* (14).

*Proof.* Let us set $\mathscr{C}^1(\mathrm{B}_b) := \mathrm{C}^1([0,1], \mathrm{B}_b(X))$ and $\mathscr{C}^1(\mathrm{Lip}_b) := \mathrm{C}^1([0,1], \mathrm{Lip}_b(X))$ to shorten the notation. From the dual dynamic formulations (32) and (34) (recall that a RCD space is a length space) we know

$$\mathsf{He}_p^p(\mu_0, \mu_1) = \sup\Big\{ \int_\Omega \zeta_1 \,\mathrm{d}\mu_1 - \int_\Omega \zeta_0 \,\mathrm{d}\mu_0 : \ \zeta \in \mathscr{C}^1(\mathrm{B}_b), \quad \partial_s\zeta_s + \frac{\zeta_s^q}{q-1} \le 0 \Big\}$$

and

$$\frac{1}{p}\mathsf{W}_p^p(\mu_0, \mu_1) = \sup\Big\{ \int_X \zeta_1 \,\mathrm{d}\mu_1 - \int_X \zeta_0 \,\mathrm{d}\mu_0 : \ \zeta \in \mathscr{C}^1(\mathrm{Lip}_b), \quad \partial_s\zeta_s + \frac{1}{q}\,|\mathrm{D}\zeta_s|^q \le 0 \Big\}.$$

A simple rescaling argument, replacing $\zeta$ by $\frac{p}{a}\zeta$, shows that for $a>0$

$$\frac{1}{a}\mathsf{W}_p^p(\mu_0,\mu_1)=\sup\Big\{\int_X\zeta_1\,\mathrm{d}\mu_1-\int_X\zeta_0\,\mathrm{d}\mu_0:\ \zeta\in\mathscr{C}^1(\mathrm{Lip}_b),\ \partial_s\zeta_s+\frac{a^{q-1}}{q\,p^{q-1}}\,|\mathrm{D}\zeta_s|^q\le 0\Big\}. \tag{64}$$

Now, take $\zeta\in\mathrm{C}^1([0,1],\mathrm{B}_b(X))$ such that $\partial_s\zeta_s+\frac{1}{q-1}\zeta_s^q\le 0$. We apply the order preserving semigroup $(\mathsf{P}_t)_{t\ge0}$ to $(\zeta_s)_s$ and we get

$$\partial_s\mathsf{P}_t\zeta_s+\frac{1}{q-1}\mathsf{P}_t\zeta_s^q\le 0. \tag{65}$$

The Lipschitz regularization property stated in Theorem 3.6 ensures that $\mathsf{P}_t(\zeta_s)\in\mathrm{Lip}_b(X)$ and that it satisfies the refined Bakry-Emery condition (43), where we neglect the last negative term:

$$R_K(t)\,|\mathrm{D}\mathsf{P}_t\zeta_s|^2\le\mathsf{P}_t\zeta_s^2\qquad\text{in } X. \tag{66}$$

Since $p\in(1,2]$ , the conjugate $q$ is in $[2,+\infty)$ and hence $q/2\ge1$. Taking the power $q/2$ in (66) and using Jensen's inequality we obtain

$$(R_K(t))^{\frac{q}{2}}\,|\mathrm{D}\mathsf{P}_t\zeta_s|^q\le\big(\mathsf{P}_t(\zeta_s)\big)^{q/2}\le\mathsf{P}_t(\zeta_s^q).$$

The combination of this inequality and (65) yields

$$\partial_s\mathsf{P}_t\zeta_s+\frac{p^{q-1}q(R_K(t))^{\frac{q}{2}}}{q-1}\frac{|\mathrm{D}\mathsf{P}_t\zeta_s|^q}{p^{q-1}q}=\partial_s\mathsf{P}_t\zeta_s+p^q(R_K(t))^{\frac{q}{2}}\frac{|\mathrm{D}\mathsf{P}_t\zeta_s|^q}{p^{q-1}q}\le 0$$

which shows that $\tilde\zeta_s:=\mathsf{P}_t\zeta_s$ is a subsolution of the Hamilton-Jacobi equation as in (64) with the time-and-curvature dependent weight

$$a(t):=\Big(p^q(R_K(t))^{\frac{q}{2}}\Big)^{1/q-1}=p^p(R_K(t))^{p/2}. \tag{67}$$

All these facts lead to

$$\int_X\zeta_1\,\mathrm{d}(\mathsf{P}_t^*\mu_1)-\int_X\zeta_0\,\mathrm{d}(\mathsf{P}_t^*\mu_0)=\int_X\mathsf{P}_t\zeta_1\,\mathrm{d}\mu_1-\int_X\mathsf{P}_t\zeta_0\,\mathrm{d}\mu_0\le\frac{1}{a(t)}\mathsf{W}_p^p(\mu_0,\mu_1).$$

Thus, taking the supremum over all the subsolutions to $\partial_s\zeta_s+\frac{1}{q-1}\zeta_s^q\le0$ we conclude

$$\mathsf{He}_p^p(\mathsf{P}_t^*\mu_0,\mathsf{P}_t^*\mu_1)\le\frac{1}{a(t)}\mathsf{W}_p^p(\mu_0,\mu_1)$$

where $a(t)$ as in (67), which yields (63). □

As a byproduct, when $K\ge0$, we obtain an precise decay rate for the asymptotic behaviour of $\mathsf{P}_t^*$.

**Corollary 5.3.** *Let $(X,\mathsf{d},\mathfrak{m})$ be a $\mathrm{RCD}(K,\infty)$ metric measure space with $K\ge0$ and let $\mathfrak{m}\in\mathcal{P}_p(X)$, $p\in[1,2]$. For every $\mu_0\in\mathcal{P}_p(X)$ we have*

$$\mathsf{He}_p(\mathsf{P}_t^*\mu_0,\mathfrak{m})\le\frac{1}{p(R_K(t))^{1/2}}\mathsf{W}_p(\mu_0,\mathfrak{m})\quad\textit{for every } t>0. \tag{68}$$

In the case $p=2$ and $K>0$ it is interesting to compare (68) with the well known exponential decay rates of the logarithmic entropy and of the Wasserstein distance

$$\mathsf{KL}(\mathsf{P}_t^*\mu_0|\mathfrak{m})\le\mathrm{e}^{-2Kt}\mathsf{KL}(\mu_0|\mathfrak{m}),\quad\mathsf{W}_2(\mathsf{P}_t^*\mu_0,\mathfrak{m})\le\mathrm{e}^{-Kt}\mathsf{W}_2(\mu_0,\mathfrak{m}) \tag{69}$$

which follow by the $K$-geodesic convexity of the $\mathsf{KL}$ functional in $\mathsf{CD}(K,\infty)$ spaces. In particular, recalling (29), the first estimate of (69) provides

$$\mathsf{He}_2(\mathsf{P}_t^*\mu_0|\mathfrak{m})\le\mathrm{e}^{-Kt}\mathsf{KL}(\mu_0|\mathfrak{m})$$

which exhibits the same exponential behaviour of (68); however, (68) only requires $\mu_0 \in \mathcal{P}_2(X)$.

5.3. **Regularization $\mathsf{He}_2$-$\mathsf{HK}$.** With a similar argument we prove that the Hellinger distance at time $t$ can be estimated from above by the weighted Hellinger-Kantorovich distance $\mathsf{HK}_\alpha$, in which the parameter $\alpha$ acts on the transport part of the distance with a time-dependent factor and does not affect the reaction part. Note that this embodies a natural combination of the Hellinger-Kantorovich estimate above and the Hellinger contraction that we proved in Proposition 5.1.

**Theorem 5.4.** *Let $(X, \mathsf{d}, \mathfrak{m})$ be a* $\mathrm{RCD}(K,\infty)$ *metric measure space. For every $\mu_0, \mu_1 \in \mathcal{M}(X)$*

$$\mathsf{He}_2(\mathsf{P}_t^*\mu_0, \mathsf{P}_t^*\mu_1) \le \mathsf{HK}_{\alpha(t)}(\mu_0,\mu_1), \qquad \alpha(t) = 4R_K(t) \text{ as defined in (14)}. \tag{70}$$

*Proof.* As in the previous proof, we set $\mathscr{C}^1(\mathrm{B}_b) := \mathrm{C}^1([0,1], \mathrm{B}_b(\Omega))$ and $\mathscr{C}^1(\mathrm{Lip}_b) := \mathrm{C}^1([0,1], \mathrm{Lip}_b(\Omega))$ and we recall that

$$\mathsf{He}_2^2(\mu_0,\mu_1) = \sup\Big\{ \int_X \zeta_1 \,\mathrm{d}\mu_1 - \int_X \zeta_0 \,\mathrm{d}\mu_0 : \ \zeta \in \mathscr{C}^1(\mathrm{B}_b), \quad \partial_s\zeta_s + \zeta_s^2 \le 0 \Big\}.$$

and that (39)

$$\begin{aligned}\mathsf{HK}_\alpha^2(\mu_0,\mu_1) = \sup\Big\{ \int_X \zeta_1 \,\mathrm{d}\mu_1 - \int_X \zeta_0 \,\mathrm{d}\mu_0 \ : \ \zeta \in \mathscr{C}^1(\mathrm{Lip}_b),\\ \partial_s\zeta_s(x) + \frac{\alpha}{4}|D_X\zeta_s|^2(x) + \zeta_s^2 \le 0 \Big\}\end{aligned} \tag{71}$$

We consider a solution $\zeta \in \mathscr{C}^1(\mathrm{B}_b)$ of $\partial_s\zeta_s + \zeta_s^2 \le 0$ and we apply the linear operator $\mathsf{P}_t$, $t > 0$, obtaining

$$\partial_s \mathsf{P}_t\zeta_s + \mathsf{P}_t(\zeta_s)^2 \le 0.$$

Theorem 3.6 ensures that $\mathsf{P}_t\zeta_s$ is Lipschitz and satisfies

$$R_K(t)\,|\mathrm{D}\mathsf{P}_t\zeta_s|^2 + (\mathsf{P}_t\zeta_s)^2 \le \mathsf{P}_t(\zeta_s^2)$$

so that

$$\partial_s\mathsf{P}_t\zeta_s + R_K(t)\,|\mathrm{D}\mathsf{P}_t\zeta_s|^2 + (\mathsf{P}_t\zeta_s)^2 \le 0;$$

this inequality corresponds to the subsolutions of Hamilton-Jacobi equation in (71) weighted with $\alpha = 4R_K(t) = \alpha(t)$. Therefore

$$\int_X \zeta_1 \,\mathrm{d}(\mathsf{P}_t^*\mu_1) - \int_X \zeta_0 \,\mathrm{d}(\mathsf{P}_t^*\mu_0) = \int_X \mathsf{P}_t\zeta_1 \,\mathrm{d}\mu_1 - \int_X \mathsf{P}_t\zeta_0 \,\mathrm{d}\mu_0 \le \mathsf{HK}_{\alpha(t)}^2(\mu_0,\mu_1),$$

and taking the supremum with respect to the subsolutions to $\partial_s\zeta_s + \zeta_s^2 \le 0$ we get (70). □

It is worth noticing that (70) yields the pure Hellinger contraction estimate (58) thanks to (37). Similarly, choosing $\mu_0, \mu_1 \in \mathcal{P}_2(X)$ and applying (38) one recovers (63) in the case $p = 2$.

## REFERENCES

[1] L. Ambrosio, M. Erbar, and G. Savaré. Optimal transport, Cheeger energies and contractivity of dynamic transport distances in extended spaces. *Nonlinear Anal.*, 137:77–134, 2016.
[2] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.
[3] L. Ambrosio, N. Gigli, and G. Savaré. Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below. *Invent. Math.*, 195(2):289–391, 2014.
[4] L. Ambrosio, N. Gigli, and G. Savaré. Metric measure spaces with Riemannian Ricci curvature bounded from below. *Duke Math. J.*, 163(7):1405–1490, 2014.
[5] L. Ambrosio, N. Gigli, and G. Savaré. Bakry-Émery curvature-dimension condition and Riemannian Ricci curvature bounds. *Ann. Probab.*, 43(1):339–404, 2015.
[6] L. Ambrosio, A. Mondino, and G. Savaré. Nonlinear diffusion equations and curvature conditions in metric measure spaces. *ArXiv e-prints*, 1509.07273, Sept. 2015.
[7] L. Ambrosio, G. Savaré, and L. Zambotti. Existence and stability for Fokker-Planck equations with log-concave reference measure. *Probab. Theory Relat. Fields*, 145(3-4):517–564, 2009.
[8] D. Bakry and M. Émery. Diffusions hypercontractives. *Séminaire de probabilités de Strasbourg*, 19:177–206, 1985.
[9] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.
[10] D. Bakry, I. Gentil, and M. Ledoux. On Harnack inequalities and optimal transportation. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 14(3):705–727, 2015.
[11] A. Björn and J. Björn. *Nonlinear potential theory on metric spaces*, volume 17 of *EMS Tracts in Mathematics*. European Mathematical Society (EMS), Zürich, 2011.
[12] H. Brézis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*. North-Holland Publishing Co., Amsterdam, 1973. North-Holland Mathematics Studies, No. 5. Notas de Matemática (50).
[13] J. A. Carrillo, R. J. McCann, and C. Villani. Contractions in the 2-Wasserstein length space and thermalization of granular media. *Arch. Ration. Mech. Anal.*, 179(2):217–263, 2006.
[14] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. An interpolating distance between optimal transport and Fisher-Rao metrics. *Found. Comput. Math.*, 18(1):1–44, 2018.
[15] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. Unbalanced optimal transport: dynamic and Kantorovich formulations. *J. Funct. Anal.*, 274(11):3090–3123, 2018.
[16] I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2:299–318, 1967.
[17] S. Daneri and G. Savaré. Eulerian calculus for the displacement convexity in the Wasserstein distance. *SIAM J. Math. Anal.*, 40(3):1104–1122, 2008.
[18] C. Dellacherie and P.-A. Meyer. *Probabilities and potential. C*, volume 151 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1988. Potential theory for discrete and continuous semigroups, Translated from the French by J. Norris.
[19] M. Erbar. The heat equation on manifolds as a gradient flow in the Wasserstein space. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques*, 46(1):1–23, 2010.
[20] 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. *Invent. Math.*, 201(3):993–1071, 2015.
[21] N. Gigli, K. Kuwada, and S. Ohta. Heat flow on Alexandrov spaces. *Comm. Pure Appl. Math.*, 66:307331, 2013.
[22] J. Heinonen and P. Koskela. Quasiconformal maps in metric spaces with controlled geometry. *Acta Math.*, 181(1):1–61, 1998.
[23] J. Heinonen, P. Koskela, N. Shanmugalingam, and J. T. Tyson. *Sobolev spaces on metric measure spaces*, volume 27 of *New Mathematical Monographs*. Cambridge University Press, Cambridge, 2015. An approach based on upper gradients.
[24] E. Hellinger. Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. *J. Reine Angew. Math.*, 136:210–271, 1909.
[25] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck Equation. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, 1998.

[26] S. Kakutani. On equivalence of infinite product measures. *Ann. of Math. (2)*, 49:214–224, 1948.

[27] S. Kondratyev, L. Monsaingeon, and D. Vorotnikov. A new optimal transport distance on the space of finite Radon measures. *Adv. Differential Equations*, 21(11-12):1117–1164, 2016.

[28] P. Koskela and P. MacManus. Quasiconformal mappings and Sobolev spaces. *Studia Math.*, 131(1):1–17, 1998.

[29] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Statistics*, 22:79–86, 1951.

[30] K. Kuwada. Duality on gradient estimates and Wasserstein controls. *J. Funct. Anal.*, 258(11):3758–3774, 2010.

[31] M. Liero, A. Mielke, and G. Savaré. Optimal transport in competition with reaction: the Hellinger-Kantorovich distance and geodesic curves. *SIAM J. Math. Anal.*, 48(4):2869–2911, 2016.

[32] M. Liero, A. Mielke, and G. Savaré. Optimal entropy-transport problems and a new Hellinger-Kantorovich distance between positive measures. *Invent. Math.*, 211(3):969–1117, 2018.

[33] F. Liese and I. Vajda. On divergences and informations in statistics and information theory. *IEEE Trans. Inform. Theory*, 52(10):4394–4412, 2006.

[34] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math. (2)*, 169(3):903–991, 2009.

[35] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.

[36] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.

[37] F. Otto and M. Westdickenberg. Eulerian calculus for the contraction in the Wasserstein distance. *SIAM J. Math. Anal.*, 37(4):1227–1255 (electronic), 2005.

[38] G. Savaré. Self-improvement of the Bakry-Émery condition and Wasserstein contraction of the heat flow in $\mathrm{RCD}(K,\infty)$ metric measure spaces. *Discrete Contin. Dyn. Syst.*, 34(4):1641–1661, 2014.

[39] N. Shanmugalingam. Newtonian spaces: an extension of Sobolev spaces to metric measure spaces. *Rev. Mat. Iberoamericana*, 16(2):243–279, 2000.

[40] K.-T. Sturm. On the geometry of metric measure spaces. I. *Acta Math.*, 196(1):65–131, 2006.

[41] K.-T. Sturm and M.-K. von Renesse. Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.*, 58(7):923–940, 2005.

[42] C. Villani. *Optimal transport. Old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 2009.

*E-mail address*: g.luise.16@ucl.ac.uk

*E-mail address*: giuseppe.savare@unipv.it