. SCUOLA INTERNAZIONALE SUPERIORE DI STUDI AVANZATI

SISSA Digital Library

Benamou–Brenier and duality formulas for the entropic cost on RCD∗(K, N) spaces

Original
Benamou–Brenier and duality formulas for the entropic cost on RCD∗(K, N) spaces / Gigli, N., Tamanini, L.. - In: PROBABILITY THEORY AND RELATED FIELDS. - ISSN 0178-8051. - 176:1-2(2020), pp. 1-34. [10.1007/s00440-019-00909-1]

Availability:
This version is available at: 20.500.11767/111348 since: 2025-01-15T14:15:46Z

Publisher:

Published
DOI:10.1007/s00440-019-00909-1

Terms of use:

Testo definito dall'ateneo relativo alle clausole di concessione d'uso

Publisher copyright
Springer

This version is available for education and non-commercial purposes.

note finali coverpage

(Article begins on next page)

09 September 2026

Benamou-Brenier and duality formulas for the entropic cost on $\mathsf{RCD}^*(K,N)$ spaces

Nicola Gigli * Luca Tamanini †

March 20, 2019

Abstract

In this paper we prove that, within the framework of $\mathsf{RCD}^*(K,N)$ spaces with $N<\infty$, the entropic cost (i.e. the minimal value of the Schrödinger problem) admits:

- a threefold dynamical variational representation, in the spirit of the Benamou-Brenier formula for the Wasserstein distance;
- a Hamilton-Jacobi-Bellman dual representation, in line with Bobkov-Gentil-Ledoux and Otto-Villani results on the duality between Hamilton-Jacobi and continuity equation for optimal transport;
- a Kantorovich-type duality formula, where the Hopf-Lax semigroup is replaced by a suitable 'entropic' counterpart.

We thus provide a complete and unifying picture of the equivalent variational representations of the Schrödinger problem as well as a perfect parallelism with the analogous formulas for the Wasserstein distance. Riemannian manifolds with Ricci curvature bounded from below are a relevant class of $\mathsf{RCD}^*(K,N)$ spaces and our results are new even in this setting.

Dedicated to the memory of Prof. Kazumasa Kuwada

Contents

1 Introduction **2**
1.1 Optimal transport 2
1.2 Schrödinger problem 4
1.3 Our result 6

2 Preliminaries **9**
2.1 Analysis and optimal transport in RCD spaces 9
2.2 The Schrödinger problem in RCD spaces 11

3 Definitions and auxiliary results **13**

*SISSA, Trieste. email: ngigli@sissa.it
†Institut für Angewandte Mathematik, Universität Bonn. email: tamanini@iam.uni-bonn.de

4 Dynamical and dual representations of the entropic cost **16**
4.1 Benamou-Brenier formulas for $\mathscr{I}_\varepsilon$ 16
4.2 Dual representations for $\mathscr{I}_\varepsilon$ 24

A Proofs of Lemmas 3.3, 3.4, 3.5 **26**

1 Introduction

1.1 Optimal transport

It is well-known that the optimal transport problem with quadratic cost admits two equivalent formulations: the Kantorovich dual one [22] and the Benamou-Brenier dynamical one [5]. Given two compactly supported probability measures $\mu_0 = \rho_0 \mathcal{L}^d, \mu_1 = \rho_1 \mathcal{L}^d$ in $\mathbb{R}^d$, the former tells us that the squared Wasserstein distance between μ_0 and μ_1 can be represented as

$$\frac{1}{2} W_2^2(\mu_0, \mu_1) = \sup_{\phi \in C_b(\mathbb{R}^d)} \int \phi \, \mathrm{d}\mu_0 + \int \phi^c \, \mathrm{d}\mu_1 \tag{1.1}$$

with ϕ^c the c-conjugate of ϕ, namely

$$\phi^c(x) := \inf_{y \in \mathbb{R}^d} \frac{|x-y|^2}{2} - \phi(y).$$

On the other hand, J.-D. Benamou and Y. Brenier observed that the optimal transport problem admits a fluid-dynamics interpretation in the following sense:

$$W_2^2(\mu_0, \mu_1) = \inf_{(\rho, v)} \iint_0^1 |v(x,t)|^2 \rho(x,t) \, \mathrm{d}t \mathrm{d}x \tag{1.2}$$

where the infimum runs over all couples (ρ, v) solving the continuity equation

$$\partial_t \rho + \mathrm{div}(\rho v) = 0, \qquad \text{in } \mathbb{R}^d \times (0,1)$$

with marginal constraints $\rho(\cdot, 0) = \rho_0$ and $\rho(\cdot, 1) = \rho_1$ in $\mathbb{R}^d$.

Although the two results may appear completely different in spirit at a first glance, as a matter of fact they are deeply linked, since it is possible to pass directly from the one to the other, gaining a further variational representation of the Wasserstein distance. This relies on the duality between Hamilton-Jacobi and continuity equation, hidden in (1.2) and already noticed by Benamou and Brenier. Indeed, the Hamilton-Jacobi equation

$$\partial_t \phi + \frac{1}{2} |\nabla \phi|^2 = 0, \qquad \text{in } \mathbb{R}^d \times (0,1)$$

already arises in the optimality conditions for v in (1.2), which read as $v = \nabla \phi$ with ϕ solving the PDE above. More generally, if (ϕ_t) is a subsolution to the Hamilton-Jacobi equation and (ρ, v) is a solution to the continuity equation, then

$$\int \phi(\cdot, 1) \, \mathrm{d}\mu_1 - \int \phi(\cdot, 0) \, \mathrm{d}\mu_0 \leq \frac{1}{2} \iint_0^1 |v(x,t)|^2 \rho(x,t) \mathrm{d}t \mathrm{d}x$$

and, as first noticed by F. Otto and C. Villani in [33] and by S. Bobkov, I. Gentil and M. Ledoux in [6], if we saturate the left-hand side with the supremum and the right-hand one with the infimum, then the inequality turns out to be an equality, which in particular yields

$$\frac{1}{2}W_2^2(\mu_0,\mu_1) = \sup \int \phi(\cdot,1)\,\mathrm{d}\mu_1 - \int \phi(\cdot,0)\,\mathrm{d}\mu_0 \tag{1.3}$$

where the supremum runs over all subsolutions to the Hamilton-Jacobi equation with initial condition $\phi(\cdot,0)=\phi_0$ for all possible $\phi_0 \in C_b(\mathbb{R}^d)$. Now it is sufficient to recall that maximal subsolutions to the Hamilton-Jacobi equation are obtained via the Hopf-Lax semigroup Q_t, defined for any $f:\mathbb{R}^d \to \mathbb{R}\cup\{+\infty\}$ and $t>0$ as

$$Q_t f(x) := \inf_{y\in\mathbb{R}^d} \frac{|x-y|^2}{2t} + f(y).$$

This allows us to restate (1.3) as

$$\frac{1}{2}W_2^2(\mu_0,\mu_1) = \sup_{\phi\in C_b(\mathbb{R}^d)} \int Q_1\phi\,\mathrm{d}\mu_1 - \int \phi\,\mathrm{d}\mu_0, \tag{1.4}$$

which is equivalent to (1.1) by the very definition of the Hopf-Lax semigroup, thus closing the loop.

The great interest in the study of metric spaces led subsequently the focus on the possible generalization of these results. For what concerns the 'static' quantities, this comes quite for free: (1.4) holds in great generality and actually can be stated for more general cost functions. Similarly, (1.3) holds in arbitrary metric spaces provided one defines $\phi(t,\cdot)$ via the Hopf-Lax formula. The link of such formula with the Hamilton-Jacobi equation (with $\operatorname{lip}\phi$ in place of $|\nabla\phi|$) has been first studied in [2] and [20]. Things are technically more complicated for what concerns (1.2), because it is not clear how to formulate the continuity equation on a metric space. Yet, a natural notion of solution to such equation can be given on metric *measure* spaces $(\mathrm{X},\mathsf{d},\mathfrak{m})$: adopting the language developed by the first author in [14] and recalled in Section 2.1, we first say that a curve $(\mu_t)\subset\mathscr{P}_2(\mathrm{X})$ is a solution of the continuity equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mu_t + \operatorname{div}(X_t\mu_t) = 0,$$

provided the following holds. The X_t's are vector fields in $L^0(T\mathrm{X})$ (in the smooth setting this is the class of Borel vector fields identified up to equality a.e.) possibly defined only for a.e. $t\in[0,1]$, the map $t\mapsto\mu_t$ is weakly continuous, $t\mapsto\int|X_t|^2\mathrm{d}\mu_t$ belongs to $L^1(0,1)$, $\mu_t\le C\mathfrak{m}$ for all $t\in[0,1]$ for some $C>0$ and for any $f\in W^{1,2}(\mathrm{X})$ the map $[0,1]\ni t\mapsto\int f\,\mathrm{d}\mu_t$ is absolutely continuous with

$$\frac{\mathrm{d}}{\mathrm{d}t}\int f\,\mathrm{d}\mu_t = \int \mathrm{d}f(X_t)\,\mathrm{d}\mu_t \qquad \text{a.e. } t.$$

With this premise, if $(\mathrm{X},\mathsf{d},\mathfrak{m})$ is infinitesimally Hilbertian and $\mu,\nu\in\mathscr{P}_2(\mathrm{X})$ are such that there exists a W_2-geodesic (μ_t) connecting them such that $\mu_t\le C\mathfrak{m}$ for all $t\in[0,1]$ for some $C>0$ (the statement in [16] is slightly more general), then

$$W_2^2(\mu,\nu) = \min \int_0^1 \int |X_t|^2 \mathrm{d}\mu_t\mathrm{d}t,$$

where the minimum is taken among all solutions (μ_t,X_t) of the continuity equation such that $\mu_0=\mu$ and $\mu_1=\nu$.

1.2 Schrödinger problem

Formally similar to optimal transport but with completely different motivation and interpretation, the Schrödinger problem is an optimization and interpolation problem too. While optimal transport was originally formulated by G. Monge for engeneering purposes, such as resource allocation, on the contrary Schrödinger problem is physical in nature, as in trying to explain the wave-particle duality via a classical mechanics example, E. Schrödinger landed in a maximal likelihood problem. In both cases two probability measures μ_0, μ_1 are assigned as data, but while in optimal transport they are seen as initial and final configurations of resources whose transportation cost has to be minimized among all possible couplings $\boldsymbol{\gamma}$ between μ_0 and μ_1, in Schrödinger problem μ_0 and μ_1 represent initial and final probability distributions of diffusive particles and one looks for the most likely evolution from μ_0 to μ_1, thus suggesting a strong link with large deviations theory. Let us briefly describe what this means in the Euclidean setting.

Given two probability measures $\mu_0 = \rho_0 \mathcal{L}^d$, $\mu_1 = \rho_1 \mathcal{L}^d$ on $\mathbb{R}^d$, one looks for a coupling between them that takes into account the fact that the particles are driven by a diffusion process. As shown in [12] (see also [26] for a detailed explanation), this amounts to solve the following minimization problem

$$\inf_{\boldsymbol{\gamma} \in \mathrm{Adm}(\mu_0, \mu_1)} H(\boldsymbol{\gamma} \,|\, \mathsf{R}) \tag{1.5}$$

where $\mathrm{Adm}(\mu_0, \mu_1)$ denotes the family of all couplings between μ_0 and μ_1, $H(\cdot \,|\, \cdot)$ is the Boltzmann-Shannon entropy and $\mathrm{d}\mathsf{R}(x, y) = \mathsf{r}_{1/2}(x, y)\mathcal{L}^{2d}$, $\mathsf{r}_{1/2}$ being the heat kernel at time $t = 1/2$ (the time choice plays no special role, but is convenient in computations). The fact that the Boltzmann-Shannon entropy appears in (1.5) is a hint of the very probabilistic nature of the Schrödinger problem, which can be intuitively described as follows. Given n independent Brownian particles, whose trajectories are described by stochastic processes $Y^1, ..., Y^n$ with values in $C([0, 1], \mathbb{R}^d)$, consider the empirical measure

$$Z^n := \frac{1}{n} \sum_{i=1}^{n} \delta_{Y^i}$$

and assume that its initial (resp. final) marginal Z_0^n (resp. Z_1^n) converges to μ_0 (resp. μ_1) as $n \to \infty$. By the law of large numbers we expect Z_1^n to converge to the law at $t = 1$ of the Brownian motion with initial distribution μ_0; thus if μ_1 does not match with it, then an unlikely (but still possible, as n is finite) event occurred. The aim is to find the evolution of the particle system conditionally on this rare event. The answer relies on large deviations theory and more precisely on Sanov's theorem, whence it is possible to deduce the informal statement

$$\lim_{n \to \infty} \frac{1}{n} \log \mathrm{Prob}(Z_1^n \sim \mu_1 \,|\, Z_0^n \sim \mu_0) = - \inf_{\mathcal{P}} \Big\{ H(\mathcal{P} \,|\, \mathcal{R}) - H(\mu_0 \,|\, \mathcal{L}^d) \Big\},$$

where $Z_i^n \sim \mu_1$ should be intended as 'the law of Z_i^n is close to μ_1', $\mathcal{R}$ is the law of the reversible Brownian motion with initial distribution $\mathcal{L}^d$ (notice that $\mathcal{R}$ has infinite mass) and the infimum runs over all probability measures $\mathcal{P}$ on $C([0, 1], \mathbb{R}^d)$ matching with the marginal constraints μ_0, μ_1. Following [12], it is then easy to see that the right-hand side above is nothing but a restatement of (1.5).

After this premise, let us discuss the solvability of (1.5). It turns out that in great generality this problem admits a unique solution $\boldsymbol{\gamma}$ and the structure of the minimizer is very rigid: indeed $\boldsymbol{\gamma} = f \otimes g\, \mathsf{R}$ for some Borel functions $f, g : \mathbb{R}^d \to [0, \infty)$, where $f \otimes g(x, y) := f(x)g(y)$. As a consequence

$$\rho_0 = f\, \mathsf{h}_{1/2} g \qquad\qquad \rho_1 = g\, \mathsf{h}_{1/2} f, \tag{1.6}$$

where $\mathsf{h}_t f$ is the heat flow starting at f evaluated at time t. This suggests us to interpolate from ρ_0 to ρ_1 by defining

$$\rho_t := \mathsf{h}_{t/2} f\, \mathsf{h}_{(1-t)/2} g.$$

This is called entropic interpolation, in analogy with displacement one. Introducing the Schrödinger potentials φ_t, ψ_t (in connection with Kantorovich ones) as

$$\varphi_t := \log \mathsf{h}_{t/2} f \qquad\qquad \psi_t := \log \mathsf{h}_{(1-t)/2} g, \tag{1.7}$$

the parallelism between optimal transport and Schrödinger problem can be fully appreciated. Indeed, by direct computation it is not difficult to see that $(\varphi_t), (\psi_t)$ solve the Hamilton-Jacobi-Bellman equations

$$\partial_t \varphi_t = \frac{1}{2}|\nabla \varphi_t|^2 + \frac{1}{2}\Delta \varphi_t \qquad\qquad -\partial_t \psi_t = \frac{1}{2}|\nabla \psi_t|^2 + \frac{1}{2}\Delta \psi_t, \tag{1.8}$$

and they are linked to (ρ_t) via the Fokker-Planck equations

$$-\partial_t \rho_t + \mathrm{div}(\nabla \varphi_t\, \rho_t) = \frac{1}{2}\Delta \rho_t \qquad\qquad \partial_t \rho_t + \mathrm{div}(\nabla \psi_t\, \rho_t) = \frac{1}{2}\Delta \rho_t.$$

Thus, denoting by $\mathscr{I}(\mu_0, \mu_1)$ the minimal value of (1.5), the entropic analogue of the Benamou-Brenier formula can be proved to be

$$\begin{aligned} \mathscr{I}(\mu_0, \mu_1) &= H(\mu_0 \,|\, \mathcal{L}^d) + \inf_{(\nu^+, v^+)} \int_0^1 \int \frac{|v_t^+|^2}{2} \mathrm{d}\nu_t^+ \mathrm{d}t \\ &= H(\mu_1 \,|\, \mathcal{L}^d) + \inf_{(\nu^-, v^-)} \int_0^1 \int \frac{|v_t^-|^2}{2} \mathrm{d}\nu_t^- \mathrm{d}t \end{aligned} \tag{1.9}$$

where the infimum is taken among all suitable weak solutions of the forward Fokker-Planck equation in the first case and of the backward one in the second case, with marginal constraints $\nu_0^{\pm} = \mu_0$ and $\nu_1^{\pm} = \mu_1$. If we also introduce the functions $\vartheta_t := \frac{\psi_t - \varphi_t}{2}$ it is not hard to check that it holds

$$\partial_t \rho_t + \mathrm{div}(\nabla \vartheta_t\, \rho_t) = 0, \tag{1.10}$$

and a third Benamou-Brenier formula for $\mathscr{I}(\mu_0, \mu_1)$ is available, namely

$$\mathscr{I}(\mu_0, \mu_1) = \frac{1}{2}\Big(H(\mu_0 \,|\, \mathcal{L}^d) + H(\mu_1 \,|\, \mathcal{L}^d)\Big) + \inf_{(\eta, v)} \left\{ \iint_0^1 \Big(\frac{|v_t|^2}{2} + \frac{1}{8}|\nabla \log \eta_t|^2\Big)\eta_t \mathrm{d}t \mathrm{d}\mathfrak{m} \right\} \tag{1.11}$$

where the infimum now runs over all suitable weak solutions of the continuity equation with marginal constraints $\eta_0 \mathcal{L}^d = \mu_0$ and $\eta_1 \mathcal{L}^d = \mu_1$. These formulas have been proved, in the Euclidean setting, in [26], [11] and then extended to a slightly more general setting in [13], moving from closely related results contained in [32], [12], [31] and the subsequent literature. A heuristic discussion can be found also in [23].

As concerns the entropic analogue of Kantorovich duality, the natural guess is then to replace solutions of the Hamilton-Jacobi equation with those of the Hamilton-Jacobi-Bellman equation in (1.3) and thus to substitute the Hopf-Lax formula with a suitable semigroup providing us with solutions of the latter PDE. This is given by

$$\tilde{Q}_t\phi(x) := \log\left(\mathsf{h}_t e^{\phi}\right), \qquad \forall \phi \in C_b(\mathbb{R}^d) \tag{1.12}$$

and thus (1.3) becomes

$$\mathscr{I}(\mu_0,\mu_1) = H(\mu_0 \,|\, \mathcal{L}^d) + \sup\left\{ \int \phi_1 \,\mathrm{d}\mu_1 - \int \phi_0 \,\mathrm{d}\mu_0 \right\}, \tag{1.13}$$

as shown in [28], where the supremum is taken among all supersolutions to the backward Hamilton-Jacobi-Bellman equation with final condition $\phi(\cdot,1) = \phi_1$ for all possible $\phi_1 \in C_b^\infty(\mathbb{R}^d)$, while (1.4) turns into

$$\mathscr{I}(\mu_0,\mu_1) = H(\mu_0 \,|\, \mathcal{L}^d) + \sup_{\phi \in C_b(\mathbb{R}^d)} \left\{ \int \phi \,\mathrm{d}\mu_1 - \int \tilde{Q}_1\phi \,\mathrm{d}\mu_0 \right\}, \tag{1.14}$$

as proved in [13]. Both (1.13) and (1.14) are forward representations and thus admit backward counterparts. Truth to be told, in [28] and in the subsequent work [29] the Schrödinger problem is not explicitly mentioned; nonetheless, a direct link between (1.9) and (1.14) is established.

1.3 Our result

As for optimal transport, also for the Schrödinger problem it is reasonable to investigate what can be said in the curved and possibly non-smooth setting. In fact, the construction of entropic interpolation and Schrödinger potentials can be done in great generality, as only a heat kernel is needed. In this direction, in the recent work [19] the authors brought the Schrödinger problem to finite-dimensional $\mathsf{RCD}^*(K,N)$ spaces, obtaining new (even in the Euclidean setting) estimated for the various objects appearing in the entropic interpolation (we shall recall some of them in Section 2.2).

Still, the generalization of (1.9), (1.11), (1.13) and (1.14) to the non-smooth setting has not been achieved yet, not even on smooth Riemannian manifolds (except for a partial result obtained for (1.14) in [21], where Kantorovich duality for general transport costs is established in the metric setting and applies to the Schrödinger problem in the case the space is assumed to be compact). As concerns the Benamou-Brenier formulas for the entropic cost, this is essentially due to the fact that in [28], [13] and [11] a more or less probabilistic approach is always adopted: either via stochastic control techniques or (as it is in [13]) by strongly relying on Girsanov's theorem. On the contrary, we propose here a purely analytic proof which fits to the RCD framework, thus extending the previous results and including, in particular, the relevant case of Riemannian manifolds with Ricci curvature bounded from below: we recall indeed that a Riemannian manifold (M,g) satisfies the $\mathsf{RCD}^*(K,N)$-condition if and only if $\mathrm{Ric}_g \geq K$ and $\dim(M) \leq N$ (in particular, compact, connected and smooth manifolds are covered by our analysis). As a further advantage, with slight modifications the same argument allows us to obtain the Hamilton-Jacobi-Bellman duality (1.13) and, as a direct corollary, the Kantorovich-type duality formula (1.14) for the entropic cost, that were also missing in the Riemannian setting. This will be achieved as follows:

- (1.11) and (1.9) are proved in Theorems 4.2 and 4.4 respectively;

- (1.13) is established in Theorem 4.6;

- (1.14) is shown to hold in Theorem 4.7.

We thus provide a complete and unifying picture of the equivalent variational representations of the Schrödinger problem as well as a perfect parallelism with the analogous formulas for the Wasserstein distance. Our approach is based on a suitable (but very natural) variant of Kuwada's lemma, which has been extremely successful in dealing with the same kind of dualism for the optimal transport problem (in particular in connection with the heat flow on non-smooth spaces, see [17] and [2]). Let us informally outline our argument for proving the Benamou-Brenier formula (1.11), neglecting all the smoothness and compactness issues.

Let (η_t, v_t) be a solution of the continuity equation with $\eta_0\mathfrak{m} = \mu_0$, $\eta_1\mathfrak{m} = \mu_1$, and $(\varphi_t), (\psi_t)$ arbitrary supersolution and subsolution of the first and second in (1.8), respectively. Put $\vartheta_t := (\psi_t - \varphi_t)/2$. Then differentiate the map $t \mapsto \int \vartheta_t \eta_t \, \mathrm{d}\mathfrak{m}$ to obtain

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big(\int \vartheta_s \eta_s \, \mathrm{d}\mathfrak{m}\Big)_{|s=t} = \int (\frac{\mathrm{d}}{\mathrm{d}s}\vartheta_s|_{s=t})\eta_t \, \mathrm{d}\mathfrak{m} + \frac{\mathrm{d}}{\mathrm{d}s}\Big(\int \vartheta_t \eta_s \, \mathrm{d}\mathfrak{m}\Big)_{|s=t}$$

for a.e. $t \in [0,1]$. For the first term on the right-hand side, the definition of ϑ_t and the choice of φ_t, ψ_t yield

$$\int (\frac{\mathrm{d}}{\mathrm{d}s}\vartheta_s|_{s=t})\eta_t \mathrm{d}\mathfrak{m} \leq \int \Big(- \frac{|\nabla\psi_t|^2}{4} - \frac{|\nabla\varphi_t|^2}{4} + \frac{1}{4}\langle \nabla(\psi_t + \varphi_t), \nabla \log \eta_t\rangle\Big)\eta_t \mathrm{d}\mathfrak{m},$$

with equality if φ, ψ are solutions of the respective equations, and by Young's inequality

$$\langle \nabla(\psi_t + \varphi_t), \nabla \log \eta_t\rangle \leq \frac{1}{2}|\nabla(\psi_t + \varphi_t)|^2 + \frac{1}{2}|\nabla \log \eta_t|^2$$

with equality if and only if $\nabla(\psi_t + \varphi_t) = \nabla \log \eta_t$ $\mathfrak{m}$-a.e. On the other hand, the fact that (η_t, v_t) is a solution of the continuity equation implies that

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big(\int \vartheta_t \eta_s \mathrm{d}\mathfrak{m}\Big)_{|s=t} = \frac{1}{2}\int \langle \nabla(\psi_t - \varphi_t), v_t\rangle \eta_t \mathrm{d}\mathfrak{m}$$

and by Young's inequality

$$\langle \nabla(\psi_t - \varphi_t), v_t\rangle \leq \frac{1}{4}|\nabla(\psi_t - \varphi_t)|^2 + |v_t|^2$$

with equality if and only if $v_t = \nabla(\psi_t - \varphi_t)/2$ $\mathfrak{m}$-a.e. Plugging these observations together and integrating over $[0,1]$ we deduce that

$$\sup_{\psi,\varphi} \tfrac{1}{2}\int (\psi_1 - \varphi_1)\, \mathrm{d}\mu_1 - \tfrac{1}{2}\int (\psi_0 - \varphi_0)\, \mathrm{d}\mu_0 \leq \inf_{(\eta,v)} \iint_0^1 \Big(\frac{|v_t|^2}{2} + \frac{1}{8}|\nabla \log \eta_t|^2\Big)\nu_t \mathrm{d}t\mathrm{d}\mathfrak{m}, \quad (1.15)$$

where the sup and inf are taken among all (sub/super) solutions of the stated equations. Now we make a specific choice for these functions: we pick φ, ψ to be the Schrödinger potentials associated to μ_0, μ_1, the densities η_t to be their entropic interpolation and $v_t := \nabla\vartheta_t$. We observe that:

i) these choices are admissible in (1.15) (recall the identities (1.8) and (1.10)),

ii) with these choices, we have equalities in all the above inequalites,

iii) by the explicit structure of the minimum for the Schrödinger problem, the relations (1.6), (1.7) and little algebraic manipulation we see that the left hand side of (1.15) equals to

$$\mathscr{I}(\mu_0,\mu_1)-\tfrac{1}{2}H(\mu_0\,|\,\mathfrak{m})-\tfrac{1}{2}H(\mu_1\,|\,\mathfrak{m}).$$

These informations taken together give (1.11) and the same line of thoughts gives also (1.9), (1.13) and (1.14). In fact, at least in the compact case and under appropriate smoothness assumptions on μ_0,μ_1 we can also ensure that equality is achieved only with such choices of φ,ψ,η,v, see Remark 4.5; except for (1.11), the general case remains elusive because to justify our computations we have to perform some cut-off argument which makes hard to keep track of the equality condition.

If we replace $\mathsf{r}_{1/2}$ by $\mathsf{r}_{\varepsilon/2}$ in (1.5), denote by $\mathscr{I}_\varepsilon(\mu_0,\mu_1)$ the minimal value of the associated problem and rescale properly Fokker-Planck and Hamilton-Jacobi-Bellman equations, this can be summarized as follows

	Optimal transport $W_2^2(\mu_0,\mu_1)/2$	Schrödinger problem $\varepsilon\mathscr{I}_\varepsilon(\mu_0,\mu_1)$
Primal problem Static version	$\inf_{\boldsymbol{\gamma}\in\mathrm{Adm}(\mu_0,\mu_1)}\int\frac{\lvert x-y\rvert^2}{2}\mathrm{d}\boldsymbol{\gamma}(x,y)$	$\inf_{\boldsymbol{\gamma}\in\mathrm{Adm}(\mu_0,\mu_1)}\varepsilon H(\boldsymbol{\gamma}\,\lvert\,\mathsf{R})$
Primal problem Dynamical version	$\inf_{CE}\iint_0^1\frac{\lvert v_t\rvert^2}{2}\rho_t\,\mathrm{d}t\mathrm{d}\mathcal{L}^d$	$\varepsilon H_0+\inf_{fFP}\int_0^1\int\frac{\lvert v_t^+\rvert^2}{2}\mathrm{d}\nu_t^+\mathrm{d}t$ $\varepsilon H_1+\inf_{bFP}\int_0^1\int\frac{\lvert v_t^-\rvert^2}{2}\mathrm{d}\nu_t^-\mathrm{d}t$ $\frac{\varepsilon}{2}\big(H_0+H_1\big)+\inf_{CE}\iint_0^1\Big(\frac{\lvert v_t\rvert^2}{2}$ $+\frac{\varepsilon^2}{8}\lvert\nabla\log\eta_t\rvert^2\Big)\eta_t\mathrm{d}t\mathrm{d}\mathcal{L}^d$
Dual problem Static version	$\sup_{\phi\in C_b(\mathbb{R}^d)}\int Q_1\phi\,\mathrm{d}\mu_1-\int\phi\,\mathrm{d}\mu_0$	$\varepsilon H_0+\sup_{\phi\in C_b(\mathbb{R}^d)}\int\phi\,\mathrm{d}\mu_1-\int\tilde{Q}_1\phi\,\mathrm{d}\mu_0$ $\varepsilon H_1+\sup_{\phi\in C_b(\mathbb{R}^d)}\int\phi\,\mathrm{d}\mu_0-\int\tilde{Q}_1\phi\,\mathrm{d}\mu_1$
Dual problem Dynamical version	$\sup_{HJ}\int\phi(\cdot,1)\,\mathrm{d}\mu_1-\int\phi(\cdot,0)\,\mathrm{d}\mu_0$	$\varepsilon H_0+\sup_{bHJB}\int\phi(\cdot,1)\,\mathrm{d}\mu_1-\int\phi(\cdot,0)\,\mathrm{d}\mu_0$ $\varepsilon H_1+\sup_{fHJB}\int\tilde{\phi}(\cdot,0)\,\mathrm{d}\mu_0-\int\tilde{\phi}(\cdot,1)\,\mathrm{d}\mu_1$

where $H_0:=H(\mu_0\,|\,\mathcal{L}^d)$, $H_1:=H(\mu_1\,|\,\mathcal{L}^d)$ and CE, FP, HJ, HJB, f and b are short-hand notations for continuity equation, Fokker-Planck, Hamilton-Jacobi, Hamilton-Jacobi-Bellman, forward and backward respectively.

The choice of replacing $\mathsf{r}_{1/2}$ with $\mathsf{r}_{\varepsilon/2}$ in (1.5) is motivated by the fact that optimal transport and Schrödinger problem are intertwined by an even stronger link, as one can guess

from letting $\varepsilon \downarrow 0$ in the right-hand column above. Indeed, the Monge-Kantorovich probem can be seen as the zero-noise limit of rescaled Schrödinger problems. The basic idea is that if the heat kernel admits the asymptotic expansion $\varepsilon \log \mathsf{r}_\varepsilon(x,y) \sim -\frac{\mathsf{d}^2(x,y)}{2}$ (in the sense of Large Deviations), then the rescaled entropy functionals $\varepsilon H(\cdot \,|\, \mathsf{R}_\varepsilon)$ converge to $\frac{1}{2}\int \mathsf{d}^2(x,y)\,\mathrm{d}\cdot$ (in the sense of Γ-convergence). This has been obtained by Mikami in [27] for the quadratic cost on $\mathbb{R}^d$, later on by Mikami-Thieullen [29] for more general cost functions and finally by Léonard [24] for Polish spaces and general diffusion processes (we refer to [26] for a deeper discussion of this topic, historical remarks and much more). For this reason and to highlight the rescaling factor, throughout the paper we shall always make ε explicit.

2 Preliminaries

2.1 Analysis and optimal transport in RCD spaces

By $C([0,1],(\mathrm{X},\mathsf{d}))$, or simply $C([0,1],\mathrm{X})$, we denote the space of continuous curves with values on the metric space (X,d). For the notion of **absolutely continuous curve** in a metric space and of **metric speed** see for instance Section 1.1 in [1]. The collection of absolutely continuous curves on $[0,1]$ is denoted $AC([0,1],(\mathrm{X},\mathsf{d}))$, or simply by $AC([0,1],\mathrm{X})$.

By $\mathscr{P}(\mathrm{X})$ we denote the space of Borel probability measures on (X,d) and by $\mathscr{P}_2(\mathrm{X}) \subset \mathscr{P}(\mathrm{X})$ the subclass of those with finite second moment.

Let $(\mathrm{X},\mathsf{d},\mathfrak{m})$ be a complete and separable metric measure space endowed with a Borel non-negative measure which is finite on bounded sets.

For the definition of the **Sobolev space** $W^{1,2}(\mathrm{X})$ and of **minimal weak upper gradient** $|Df|$ see [7] (and the works [2], [34] for alternative - but equivalent - definitions of Sobolev functions). The local counterpart of $W^{1,2}(\mathrm{X})$ is introduced as follows: $L^2_{loc}(\mathrm{X})$ is defined as the space of functions $f \in L^0(\mathrm{X})$ such that for all $\Omega \subset\subset \mathrm{X}$, i.e. for all open set $\Omega \subset \mathrm{X}$ with compact closure, there exists a function $g \in L^2(\mathrm{X})$ such that $f = g$ $\mathfrak{m}$-a.e. in Ω and the local Sobolev space $W^{1,2}_{loc}(\mathrm{X})$ is then defined as

$$W^{1,2}_{loc}(\mathrm{X}) := \{f \in L^0(\mathrm{X}) \,:\, \forall \Omega \subset\subset \mathrm{X}\ \exists g \in W^{1,2}(\mathrm{X}) \text{ s.t. } f = g \ \mathfrak{m}\text{-a.e. in } \Omega\}. \tag{2.1}$$

The local minimal weak upper gradient of a function $f \in W^{1,2}_{loc}(\mathrm{X})$ is denoted by $|Df|$, omitting the locality feature, and defined for all $\Omega \subset\subset \mathrm{X}$ as $|Df| := |Dg|$ $\mathfrak{m}$-a.e. in Ω, where g is as in (2.1). The definition does depend neither on Ω nor on the choice of g associated to it by locality of the minimal weak upper gradient.

If $W^{1,2}(\mathrm{X})$ is Hilbert, which from now on we shall always assume, then $(\mathrm{X},\mathsf{d},\mathfrak{m})$ is said **infinitesimally Hilbertian** (see [15]). The language of L^0-normed modules (see [14]) allows to introduce the **differential** as a well-defined linear map d from $W^{1,2}_{loc}(\mathrm{X})$ with values in $L^0(T^*\mathrm{X})$, the family of (measurable) 1-forms. The dual of $L^0(T^*\mathrm{X})$ as L^0-normed module is denoted by $L^0(T\mathrm{X})$, it is canonically isomorphic to $L^0(T^*\mathrm{X})$ and its elements are called vector fields; the isomorphism sends the differential $\mathrm{d}f$ to the gradient ∇f.

After $W^{1,2}_{loc}(\mathrm{X})$ we can also introduce

$$D(\mathrm{div}_{loc}) := \{v \in L^0(T\mathrm{X}) \,:\, \forall \Omega \subset\subset \mathrm{X}\ \exists w \in D(\mathrm{div}) \text{ s.t. } v = w \ \mathfrak{m}\text{-a.e. in } \Omega\}$$
$$D(\Delta_{loc}) := \{f \in L^0(\mathrm{X}) \,:\, \forall \Omega \subset\subset \mathrm{X}\ \exists g \in D(\Delta) \text{ s.t. } f = g \ \mathfrak{m}\text{-a.e. in } \Omega\}$$

so that the notions of **divergence** and **Laplacian** can be extended by locality to locally integrable vector fields and functions respectively.

As regards the properties of $\mathrm{d}, \mathrm{div}, \Delta$, the differential satisfies the following calculus rules which we shall use extensively without further notice:

$$\begin{aligned} |\mathrm{d}f| &= |Df| \quad \mathfrak{m}\text{-a.e.} && \forall f \in S^2(\mathrm{X}) \\ \mathrm{d}f &= \mathrm{d}g \qquad \mathfrak{m}\text{-a.e. on } \{f = g\} && \forall f, g \in S^2(\mathrm{X}) \\ \mathrm{d}(\varphi \circ f) &= \varphi' \circ f \, \mathrm{d}f && \forall f \in S^2(\mathrm{X}),\ \varphi : \mathbb{R} \to \mathbb{R} \text{ Lipschitz} \\ \mathrm{d}(fg) &= g \, \mathrm{d}f + f \, \mathrm{d}g && \forall f, g \in L^\infty \cap S^2(\mathrm{X}) \end{aligned}$$

where it is part of the properties the fact that $\varphi \circ f, fg \in S^2(\mathrm{X})$ for φ, f, g as above. For the divergence, the formula

$$\mathrm{div}(fv) = \mathrm{d}f(v) + f\mathrm{div}(v) \qquad \forall f \in W^{1,2}(\mathrm{X}),\ v \in D(\mathrm{div}),\ \text{such that } |f|, |v| \in L^\infty(\mathrm{X})$$

holds, where it is intended in particular that $fv \in D(\mathrm{div})$ for f, v as above, and for the Laplacian

$$\begin{aligned} \Delta(\varphi \circ f) &= \varphi'' \circ f |\mathrm{d}f|^2 + \varphi' \circ f \Delta f \\ \Delta(fg) &= g\Delta f + f\Delta g + 2\langle \nabla f, \nabla g \rangle \end{aligned}$$

where in the first equality we assume that $f \in D(\Delta), \varphi \in C^2(\mathbb{R})$ are such that $f, |\mathrm{d}f| \in L^\infty(\mathrm{X})$ and $\varphi', \varphi'' \in L^\infty(\mathbb{R})$ and in the second that $f, g \in D(\Delta) \cap L^\infty(\mathrm{X})$ and $|\mathrm{d}f|, |\mathrm{d}g| \in L^\infty(\mathrm{X})$ and it is part of the claims that $\varphi \circ f, fg$ are in $D(\Delta)$. On $W^{1,2}_{loc}(\mathrm{X})$ as well as on $D(\mathrm{div}_{loc})$ and $D(\Delta_{loc})$ the same calculus rules hold with slight adaptations (see for instance [19]).

The Laplacian Δ is the infinitesimal generator of a 1-parameter semigroup (h_t) called **heat flow** (see [2]). For such a flow it holds

$$u \in L^2(\mathrm{X}) \qquad \Rightarrow \qquad (\mathsf{h}_t u) \in C([0,\infty), L^2(\mathrm{X})) \cap AC_{loc}((0,\infty), W^{1,2}(\mathrm{X})) \tag{2.2}$$

and for any $u \in L^2(\mathrm{X})$ the curve $t \mapsto \mathsf{h}_t u$ is the only solution of

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathsf{h}_t u = \Delta \mathsf{h}_t u \qquad \mathsf{h}_t u \to u \text{ as } t \downarrow 0.$$

If moreover $(\mathrm{X}, \mathsf{d}, \mathfrak{m})$ is an $\mathsf{RCD}(K, \infty)$ space (see [3]), the following a priori estimates hold true for every $u \in L^2(\mathrm{X})$ and $t > 0$:

$$\| |\nabla \mathsf{h}_t u| \|^2_{L^2(\mathrm{X})} \le \frac{1}{2t} \|u\|^2_{L^2(\mathrm{X})} \qquad \|\Delta \mathsf{h}_t u\|^2_{L^2(\mathrm{X})} \le \frac{1}{2t^2} \|u\|^2_{L^2(\mathrm{X})} \tag{2.3}$$

and the **Bakry-Émery contraction estimate** (see [3]) is satisfied:

$$|\mathrm{d}\mathsf{h}_t f|^2 \le e^{-2Kt} \mathsf{h}_t(|\mathrm{d}f|^2) \quad \mathfrak{m}\text{-a.e.} \qquad \forall f \in W^{1,2}(\mathrm{X}),\ t \ge 0. \tag{2.4}$$

Furthermore if $u \in L^\infty(\mathrm{X})$, then $\mathsf{h}_t u$ is Lipschitz on $\mathrm{supp}(\mathfrak{m})$ for all $t > 0$ and

$$\sqrt{2 \int_0^t e^{2Ks} \mathrm{d}s} \, \mathrm{Lip}(\mathsf{h}_t u) \le \|u\|_{L^\infty(\mathrm{X})}. \tag{2.5}$$

Still within the RCD framework, there exists the **heat kernel**, namely a function

$$(0,\infty) \times \mathrm{X}^2 \ni (t, x, y) \quad \mapsto \quad \mathsf{r}_t[x](y) = \mathsf{r}_t[y](x) \in (0,\infty) \tag{2.6}$$

such that $\mathsf{h}_t f(x) = \int f(y)\mathsf{r}_t[x](y)\,\mathrm{d}\mathfrak{m}(y)$ for all $t>0$ and for every $f \in L^2(\mathrm{X})$. For every $x \in \mathrm{X}$ and $t>0$, $\mathsf{r}_t[x]$ is a probability density; thus the heat flow can be extended to $L^1(\mathrm{X})$, is **mass preserving** and satisfies the **maximum principle**, i.e.

$$f \le c \quad \mathfrak{m}-a.e. \qquad\qquad \Rightarrow \qquad\qquad \mathsf{h}_t f \le c \quad \mathfrak{m}\text{-a.e.},\ \forall t>0.$$

On finite-dimensional $\mathsf{RCD}^*(K,N)$ spaces (introduced in [15]), a well-known consequence of lower Ricci curvature bounds (see e.g. [8], [9], [10]) is the existence of '**good cut-off functions**', typically intended as cut-offs with bounded Laplacian; for our purposes the following result will be sufficient:

Lemma 2.1. *Let $(\mathrm{X},\mathsf{d},\mathfrak{m})$ be an $\mathsf{RCD}^*(K,N)$ space with $K\in\mathbb{R}$ and $N\in[1,\infty)$. Then for all $R>0$ and $x\in\mathrm{X}$ there exists a function $\chi_R:\mathrm{X}\to\mathbb{R}$ satisfying:*

(i) $0\le\chi_R\le 1$, $\chi_R\equiv 1$ on $B_R(x)$ and $\mathrm{supp}(\chi_R)\subset B_{R+1}(x)$;

(ii) $\chi_R\in D(\Delta)\cap L^\infty(\mathrm{X})$, $|\nabla\chi_R|\in L^\infty(\mathrm{X})$, $\Delta\chi_R\in W^{1,2}\cap L^\infty(\mathrm{X})$.

Moreover, there exist constants $C,C'>0$ depending on K,N only such that

$$\||\nabla\chi_R|\|_{L^\infty(\mathrm{X})}\le C, \qquad\qquad \|\Delta\chi_R\|_{L^\infty(\mathrm{X})}\le C'. \tag{2.7}$$

The proof can be obtained following verbatim the arguments given in Lemma 3.1 of [30] (inspired by [4], see also [18] for an alternative approach): there the authors are interested in cut-off functions such that $\chi\equiv 1$ on $B_R(x)$ and $\mathrm{supp}(\chi)\subset B_{2R}(x)$: for this reason they fix $R>0$ and then claim that for all $x\in\mathrm{X}$ and $0<r<R$ there exists a cut-off function χ satisfying (i), (ii) and (2.7) with C,C' also depending on R. However, as far as one is concerned with cut-off functions χ where the distance between $\{\chi=0\}$ and $\{\chi=1\}$ is always equal to 1, the proof of [30] in the case $R=1$ applies and does not affect (2.7).

We conclude recalling that on $\mathsf{RCD}^*(K,N)$ spaces with $N\in[1,\infty)$ the reference measure $\mathfrak{m}$ satisfies the following volume growth condition: for all $x\in\mathrm{X}$ there exists a constant $C>0$ depending on x,K,N such that

$$\mathfrak{m}(B_r(x))\le Ce^{Cr}, \qquad \forall r>0. \tag{2.8}$$

For this reason we shall consider the weighted $L^2(\mathrm{X},e^{-V}\mathfrak{m})$ and $W^{1,2}(\mathrm{X},e^{-V}\mathfrak{m})$ spaces, where $V:=M\mathsf{d}^2(\cdot,\bar x)$. Indeed $e^{-V}\mathfrak{m}$ has finite mass for every $M>0$. For $L^2(\mathrm{X},e^{-V}\mathfrak{m})$ no comments are required. The weighted Sobolev space is defined as

$$W^{1,2}(\mathrm{X},e^{-V}\mathfrak{m}):=\{f\in W^{1,2}_{loc}(\mathrm{X})\,:\,f,|Df|\in L^2(\mathrm{X},e^{-V}\mathfrak{m})\}$$

where $|Df|$ is the local minimal weak upper gradient already introduced. Since V is locally bounded, $W^{1,2}(\mathrm{X},e^{-V}\mathfrak{m})$ turns out to coincide with the Sobolev space built over the metric measure space $(\mathrm{X},\mathsf{d},e^{-V}\mathfrak{m})$, thus motivating the choice of the notation.

2.2 The Schrödinger problem in RCD spaces

Let us first recall the definition of the relative entropy functional in the case of a reference measure with possibly infinite mass (see [25] for more details). Given a σ-finite measure ν on a Polish space (Y,τ), there exists a measurable function $W:\mathrm{Y}\to[0,\infty)$ such that

$$z_W:=\int e^{-W}\mathrm{d}\nu<+\infty.$$

Introducing the probability measure $\nu_W := z_W^{-1} e^{-W} \nu$, for any $\sigma \in \mathscr{P}(\mathrm{Y})$ such that $\int W \mathrm{d}\sigma < +\infty$ the **Boltzmann-Shannon entropy** is defined as

$$H(\sigma \,|\, \nu) := H(\sigma \,|\, \nu_W) - \int W \mathrm{d}\sigma - \log z_W \tag{2.9}$$

where $H(\sigma \,|\, \nu_W) := \int \rho \log(\rho) \, \mathrm{d}\nu_W$ if $\sigma = \rho \nu_W$ and $+\infty$ otherwise; notice that Jensen's inequality and the fact that $\tilde{\nu} \in \mathscr{P}(\mathrm{Y})$ grant that $\int \rho \log(\rho) \, \mathrm{d}\tilde{\nu}$ is well-defined and non-negative, in particular the definition makes sense.

Because of (2.8), on an $\mathsf{RCD}^*(K, N)$ space $(\mathrm{X}, \mathsf{d}, \mathfrak{m})$ with $K \in \mathbb{R}$ and $N \in [1, \infty)$ we can choose $W = \mathsf{d}^2(\cdot, \bar{x})$ in the definition above, so that $H(\cdot \,|\, \mathfrak{m})$ turns out to be well-defined on $\mathscr{P}_2(\mathrm{X})$ and W_2-lower semicontinuous. If we also introduce the following measure on X^2

$$\mathrm{d}\mathsf{R}^\varepsilon(x, y) := \mathsf{r}_\varepsilon[x](y) \, \mathrm{d}\mathfrak{m}(x) \, \mathrm{d}\mathfrak{m}(y),$$

where $\mathsf{r}_\varepsilon[x](y)$ is the heat kernel (2.6), then the choice $W : \mathrm{X}^2 \to [0, \infty)$, $W(x, x') := \mathsf{d}^2(x, \bar{x}) + \mathsf{d}^2(x', \bar{x})$ entails that, given any two probability measures $\mu_0 = \rho_0 \mathfrak{m}$, $\mu_1 = \rho_1 \mathfrak{m}$ with bounded densities and supports, $H(\cdot \,|\, \mathsf{R}^\varepsilon)$ is well-defined in $\mathrm{Adm}(\mu_0, \mu_1)$ and narrowly lower semicontinuous therein, as shown in [19].

Therefore, the minimization problem

$$\inf_{\boldsymbol{\gamma} \in \mathrm{Adm}(\mu_0, \mu_1)} H(\boldsymbol{\gamma} \,|\, \mathsf{R}^{\varepsilon/2}),$$

also known as **Schrödinger problem** (the choice of working with $\mathsf{R}^{\varepsilon/2}$ is convenient for the computations we will do later on) is meaningful. Actually, given μ_0, μ_1 as above, there exists a unique minimizer $\boldsymbol{\gamma}^\varepsilon$ and $\boldsymbol{\gamma}^\varepsilon = f^\varepsilon \otimes g^\varepsilon \mathsf{R}^{\varepsilon/2}$ for appropriate Borel functions $f, g : \mathrm{X} \to [0, \infty)$ which are $\mathfrak{m}$-a.e. unique up to the trivial transformation $(f, g) \to (cf, g/c)$ for some $c > 0$. In addition, $f^\varepsilon, g^\varepsilon$ belong to $L^\infty(\mathrm{X})$ and their supports are included in $\mathrm{supp}(\mu_0)$ and $\mathrm{supp}(\mu_1)$ respectively (cf. Proposition 2.1 and Theorem 2.2 in [19]). Thus, the **entropic cost** $\mathscr{I}_\varepsilon$ relative to $\mathsf{R}^{\varepsilon/2}$, defined as

$$\mathscr{I}_\varepsilon(\mu_0, \mu_1) := \min_{\boldsymbol{\gamma} \in \mathrm{Adm}(\mu_0, \mu_1)} H(\boldsymbol{\gamma} \,|\, \mathsf{R}^{\varepsilon/2}),$$

is finite.

Now let us fix the notations that we shall use in the sequel. For any $\varepsilon > 0$ we set $\rho_0^\varepsilon := \rho_0$, $\rho_1^\varepsilon := \rho_1$, $\mu_0^\varepsilon := \mu_0$, $\mu_1^\varepsilon := \mu_1$ and

$$\begin{cases} f_t^\varepsilon := \mathsf{h}_{\varepsilon t/2} f^\varepsilon \\ \varphi_t^\varepsilon := \varepsilon \log f_t^\varepsilon \\ \text{for } t \in (0, 1] \end{cases} \qquad \begin{cases} g_t^\varepsilon := \mathsf{h}_{\varepsilon(1-t)/2} g^\varepsilon \\ \psi_t^\varepsilon := \varepsilon \log g_t^\varepsilon \\ \text{for } t \in [0, 1) \end{cases} \qquad \begin{cases} \rho_t^\varepsilon := f_t^\varepsilon g_t^\varepsilon \\ \mu_t^\varepsilon := \rho_t^\varepsilon \mathfrak{m} \\ \vartheta_t^\varepsilon := \frac{1}{2}(\psi_t^\varepsilon - \varphi_t^\varepsilon) \\ \text{for } t \in (0, 1) \end{cases}$$

and we also define

$$\begin{aligned} \varphi_0^\varepsilon &:= \varepsilon \log(f^\varepsilon) \qquad \text{in } \mathrm{supp}(\mu_0), \\ \psi_1^\varepsilon &:= \varepsilon \log(g^\varepsilon) \qquad \text{in } \mathrm{supp}(\mu_1). \end{aligned} \tag{2.10}$$

As shown in [19] all the functions above are well defined, $\mu_t^\varepsilon \in \mathscr{P}_2(\mathrm{X})$ for every $t \in [0,1], \varepsilon > 0$

$$[0,1] \ni t \mapsto H(\mu_t^\varepsilon \,|\, \mathfrak{m}) \quad \text{is continuous} \tag{2.11}$$

and moreover for any $\varepsilon > 0$ it holds:

a) $f_t^\varepsilon, g_t^\varepsilon, \rho_t^\varepsilon$ belong to $D(\Delta)$ for all $t \in \mathcal{I}$, where $\mathcal{I}$ is the respective domain of definition (for (ρ_t^ε) we pick $\mathcal{I} = (0,1)$);

b) $\varphi_t^\varepsilon, \psi_t^\varepsilon, \vartheta_t^\varepsilon$ belong to $D(\Delta_{loc})$ for all $t \in \mathcal{I}$, where $\mathcal{I}$ is the respective domain of definition.

Secondly, $(f_t^\varepsilon), (g_t^\varepsilon), (\rho_t^\varepsilon) \in C([0,1], L^2(\mathrm{X})) \cap AC_{loc}(\mathcal{I}, W^{1,2}(\mathrm{X})) \cap L^\infty([0,1], L^\infty(\mathrm{X}))$ for any $\varepsilon > 0$ and their time derivatives are given by the following expressions for a.e. $t \in [0,1]$:

$$\frac{\mathrm{d}}{\mathrm{d}t} f_t^\varepsilon = \frac{\varepsilon}{2}\Delta f_t^\varepsilon \qquad \frac{\mathrm{d}}{\mathrm{d}t} g_t^\varepsilon = -\frac{\varepsilon}{2}\Delta g_t^\varepsilon \qquad \frac{\mathrm{d}}{\mathrm{d}t}\rho_t^\varepsilon + \mathrm{div}(\rho_t^\varepsilon \nabla \vartheta_t^\varepsilon) = 0. \tag{2.12}$$

As concerns $(\varphi_t^\varepsilon), (\psi_t^\varepsilon), (\vartheta_t^\varepsilon)$, for all $\mathcal{C} \subset \mathcal{I}$ compact and $\bar{x} \in \mathrm{X}$ there exists $M > 0$ depending on $K, N, \rho_0, \rho_1, \mathcal{C}, \bar{x}$ such that they belong to $AC(\mathcal{C}, W^{1,2}(\mathrm{X}, e^{-V}\mathfrak{m}))$ where $V = M\mathsf{d}^2(\cdot, \bar{x})$; their time derivatives are given by the following expressions for a.e. $t \in [0,1]$:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\varphi_t^\varepsilon = \frac{1}{2}|\nabla\varphi_t^\varepsilon|^2 + \frac{\varepsilon}{2}\Delta\varphi_t^\varepsilon \qquad & -\frac{\mathrm{d}}{\mathrm{d}t}\psi_t^\varepsilon = \frac{1}{2}|\nabla\psi_t^\varepsilon|^2 + \frac{\varepsilon}{2}\Delta\psi_t^\varepsilon \\ \frac{\mathrm{d}}{\mathrm{d}t}\vartheta_t^\varepsilon + \frac{|\nabla\vartheta_t^\varepsilon|^2}{2} = -\frac{\varepsilon^2}{8}\Big(2\Delta\log\rho_t^\varepsilon + |\nabla\log\rho_t^\varepsilon|^2\Big) & \end{aligned} \tag{2.13}$$

In addition, for every $\delta \in (0,1)$ and $\bar{x} \in \mathrm{X}$ there exist constants $C, C' > 0$ which depend on $K, N, \bar{x}, \rho_0, \rho_1$ and $C'' > 0$ (depending also on δ) such that

$$\rho_t^\varepsilon \le Ce^{-C'\mathsf{d}^2(\cdot,\bar{x})}, \qquad \mathfrak{m}\text{-a.e.}, \forall t \in [0,1], \tag{2.14a}$$

$$\mathrm{lip}(\varphi_t^\varepsilon) + \mathrm{lip}(\psi_{1-t}^\varepsilon) \le C''\big(1 + \mathsf{d}(\cdot,\bar{x})\big), \qquad \mathfrak{m}\text{-a.e.}, \forall t \in [\delta, 1]. \tag{2.14b}$$

As a final remark, let us recall (Lemma 4.9 in [19]) that

$$\iint_0^1 |\nabla\varphi_t^\varepsilon|^2\rho_t^\varepsilon \mathrm{d}t\mathrm{d}\mathfrak{m} < \infty \qquad \iint_0^1 |\nabla\psi_t^\varepsilon|^2\rho_t^\varepsilon \mathrm{d}t\mathrm{d}\mathfrak{m} < \infty \qquad \iint_0^1 |\nabla\vartheta_t^\varepsilon|^2\rho_t^\varepsilon \mathrm{d}t\mathrm{d}\mathfrak{m} < \infty. \tag{2.15}$$

3 Definitions and auxiliary results

We start giving the definition of 'distributional' solutions of the continuity equation and of the forward/backward Fokker-Planck equation in our setting:

Definition 3.1. *Let $(\mathrm{X}, \mathsf{d}, \mathfrak{m})$ be an infinitesimally Hilbertian metric measure space, $t \mapsto X_t \in L^0(T\mathrm{X})$ a Borel family of vector fields, possibly defined only for a.e. $t \in [0,1]$, and $c \ge 0$, $\sigma \in \{-1, 1\}$. A curve $(\mu_t) \subset \mathscr{P}_2(\mathrm{X})$ is a solution of*

$$\sigma\frac{\mathrm{d}}{\mathrm{d}t}\mu_t + \mathrm{div}(X_t\mu_t) = c\Delta\mu_t$$

if:

(i) it is weakly continuous and there exists $C > 0$ such that $\mu_t \le C\mathfrak{m}$ for all $t \in [0,1]$;

(ii) the map $t \mapsto \int |X_t|^2 \mathrm{d}\mu_t$ *is Borel and belongs to* $L^1(0,1)$*;*

(iii) for any $f \in D(\Delta)$ *the map* $[0,1] \ni t \mapsto \int f \,\mathrm{d}\mu_t$ *is absolutely continuous and it holds*

$$\sigma \frac{\mathrm{d}}{\mathrm{d}t} \int f \,\mathrm{d}\mu_t = \int \Big(\mathrm{d}f(X_t) + c\Delta f \Big) \mathrm{d}\mu_t \qquad \text{a.e. } t.$$

When

- $\sigma = 1$ *and* $c > 0$, (μ_t) *is said to be a solution of the forward Fokker-Planck equation;*
- $\sigma = -1$ *and* $c > 0$, (μ_t) *is said to be a solution of the backward Fokker-Planck equation;*
- $c = 0$, (μ_t) *is said to be a solution of the continuity equation.*

We will refer to (X_t) *as drift or velocity field.*

Let us point out that this definition of solution of the continuity equation is consistent with the one proposed in [16] and recalled in the Introduction, because if

$$\frac{\mathrm{d}}{\mathrm{d}t} \int f \,\mathrm{d}\mu_t = \int \mathrm{d}f(X_t) \,\mathrm{d}\mu_t \qquad \text{a.e. } t$$

holds for every $f \in D(\Delta)$, then it also holds for $f \in W^{1,2}(\mathrm{X})$: it is sufficient to integrate the equality on $[t_0, t_1] \subset [0,1]$ and argue by density thanks to the fact that by (ii)

$$\int_0^1 \int |X_t|^2 \mathrm{d}\mu_t \mathrm{d}t < \infty.$$

In view of Theorem 4.6 let us also provide a suitable notion of 'strong' supersolution of the forward/backward Hamilton-Jacobi-Bellman equation.

Definition 3.2. *Let* $\sigma \in \{-1, 1\}$ *and* $c, T > 0$*. A curve* $[0,T] \ni t \mapsto \phi_t \in L^0(\mathrm{X})$ *is a supersolution of the forward (resp. backward) Hamilton-Jacobi-Bellman equation provided:*

(i) there exists $C > 0$ *such that* $\|\phi_t\|_{L^\infty(\mathrm{X})} \le C$ *for all* $t \in [0,T]$*;*

(ii) $\phi_t \in D(\Delta_{loc})$ *for all* $t \in [0,T]$ *with* $(|\nabla \phi_t|), (\Delta \phi_t) \in L^\infty((0,T), L^2(\mathrm{X}))$*.*

(iii) there exist $x \in \mathrm{X}$ *and* $M > 0$ *such that* $(\phi_t) \in AC([0,T], L^2(\mathrm{X}, e^{-V}\mathfrak{m}))$*, where* $V := M\mathsf{d}^2(\cdot, x)$*, and its time derivative satisfies*

$$\sigma \frac{\mathrm{d}}{\mathrm{d}t} \phi_t \ge \frac{1}{2} |\nabla \phi_t|^2 + c\Delta \phi_t \qquad \text{for a.e. } t \in [0,T]$$

with $\sigma = 1$ *(resp.* $\sigma = -1$*).*

In this section we also collect some auxiliary results that will be used several times in the proof of the main theorems; for a fluid readability their proof is postponed to Appendix A. We start with an integrability statement (a stronger result is actually true, see [19], but this is sufficient for our purposes).

Lemma 3.3. *With the same assumptions and notation as in Section 2.2, the following holds.*

For any $\varepsilon > 0$ and $t \in \mathcal{I}$ let h_t^ε denote any of $\varphi_t^\varepsilon, \psi_t^\varepsilon, \vartheta_t^\varepsilon, \log \rho_t^\varepsilon$ and let H_t^ε denote any of the functions

$$\rho_t^\varepsilon h_t^\varepsilon, \quad \rho_t^\varepsilon |h_t^\varepsilon|^2, \quad \rho_t^\varepsilon |\nabla h_t^\varepsilon|^2,$$

where $\mathcal{I}$ is the domain of definition of h_t^ε (for $\log \rho_t^\varepsilon$ we pick $\mathcal{I} = (0,1)$). Then $H_t^\varepsilon \in L^1(\mathrm{X})$ for every $\varepsilon \in (0,1)$ and $t \in \mathcal{I}$ and, for any $\mathcal{C} \subset\subset \mathcal{I}$, $H_\cdot^\varepsilon \in L^1(\mathcal{C} \times \mathrm{X}, \mathrm{d}t \otimes \mathfrak{m})$. Moreover, $\mathcal{I} \ni t \mapsto \int H_t^\varepsilon \, \mathrm{d}\mathfrak{m}$ is continuous.

The following result is in the same spirit of the previous lemma and of the reminders of Section 2.2.

Lemma 3.4. *Let $(\mathrm{X}, \mathsf{d}, \mathfrak{m})$ be an $\mathsf{RCD}^*(K,N)$ space with $K \in \mathbb{R}$ and $N < \infty$, $u \in L^2 \cap L^\infty(\mathrm{X})$ be non-negative and $\delta > 0$. Put $\phi_t^\delta := \log(\mathsf{h}_t u + \delta)$ for all $t \geq 0$. Then:*

(i) there exists $C > 0$ such that $\|\phi_t^\delta\|_{L^\infty(\mathrm{X})} \leq C$ for all $t \geq 0$;

(ii) for all $x \in \mathrm{X}$ and $M > 0$, $(\phi_t^\delta) \in C([0,\infty), L^2(\mathrm{X}, e^{-V}\mathfrak{m})) \cap AC_{loc}((0,\infty), L^2(\mathrm{X}, e^{-V}\mathfrak{m}))$, where $V := M\mathsf{d}^2(\cdot, x)$, and its time derivative is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\phi_t^\delta = |\nabla \phi_t^\delta|^2 + \Delta \phi_t^\delta \qquad \text{for a.e. } t > 0; \tag{3.1}$$

(iii) $(|\nabla \phi_t^\delta|), (\Delta \phi_t^\delta) \in L^\infty_{loc}((0,\infty), L^2(\mathrm{X}))$;

(iv) let $(\mu_t)_{t \geq 0} \subset \mathscr{P}(\mathrm{X})$ be weakly continuous with $\mu_t \leq C\mathfrak{m}$ for some $C > 0$ independent of t, set $\eta_t := \frac{\mathrm{d}\mu_t}{\mathrm{d}\mathfrak{m}}$ and denote by H_t^δ any of the functions

$$\phi_t^\delta \eta_t, \quad |\phi_t^\delta|^2 \eta_t, \quad |\nabla \phi_t^\delta| \eta_t, \quad |\nabla \phi_t^\delta|^2 \eta_t.$$

Then $H_t^\delta \in L^1(\mathrm{X})$ for every $t, \delta > 0$ and, for any $\mathcal{C} \subset\subset (0,\infty)$, $H_\cdot^\delta \in L^1(\mathcal{C} \times \mathrm{X}, \mathrm{d}t \otimes \mathfrak{m})$.

Finally, we shall also make use of the following simple lemma valid on general metric measure spaces.

Lemma 3.5. *Let $(\mathrm{Y}, \mathsf{d}_\mathrm{Y}, \mathfrak{m}_\mathrm{Y})$ be an infinitesimally Hilbertian metric measure space endowed with a non-negative measure $\mathfrak{m}_\mathrm{Y}$ which is finite on bounded sets. Let (μ_t) be a solution of*

$$\sigma \frac{\mathrm{d}}{\mathrm{d}t}\mu_t + \mathrm{div}(X_t \mu_t) = c\Delta \mu_t$$

in the sense of Definition 3.1 and let $f \in D(\Delta)$. Then $t \mapsto \int f \, \mathrm{d}\mu_t$ is absolutely continuous and

$$\Big|\frac{\mathrm{d}}{\mathrm{d}t}\int f \, \mathrm{d}\mu_t\Big| \leq \int \big(|\mathrm{d}f||X_t| + c|\Delta f|\big)\mathrm{d}\mu_t \qquad \text{a.e. } t \in [0,1], \tag{3.2}$$

where the exceptional set can be chosen to be independent of f.

Moreover, if $(f_t) \in AC([0,1], L^2(\mathrm{Y})) \cap L^\infty([0,1], W^{1,2}(\mathrm{Y}))$ with $(\Delta f_t) \in L^\infty([0,1], L^2(\mathrm{Y}))$, then the map $t \mapsto \int f_t \, \mathrm{d}\mu_t$ is absolutely continuous and

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big(\int f_s \, \mathrm{d}\mu_s\Big)_{|s=t} = \int \big(\frac{\mathrm{d}}{\mathrm{d}s} f_{s|s=t}\big)\, \mathrm{d}\mu_t + \frac{\mathrm{d}}{\mathrm{d}s}\Big(\int f_t \, \mathrm{d}\mu_s\Big)_{|s=t} \tag{3.3}$$

for a.e. $t \in [0,1]$.

If $c = 0$, it is sufficient to assume $f \in W^{1,2}(\mathrm{Y})$ in (3.2) and $(f_t) \in AC([0,1], L^2(\mathrm{Y})) \cap L^\infty([0,1], W^{1,2}(\mathrm{Y}))$ in (3.3).

4 Dynamical and dual representations of the entropic cost

4.1 Benamou-Brenier formulas for $\mathscr{I}_\varepsilon$

In this section we extend to the RCD setting the dynamical representations (1.9) and (1.11) of the entropic cost. We start with the following threefold dynamical (but not yet variational) representation formula.

Proposition 4.1. *With the same assumptions and notations as in Section 2.2, for any $\varepsilon > 0$ the following holds:*

$$\begin{aligned}\varepsilon\mathscr{I}_\varepsilon(\mu_0,\mu_1) &= \frac{\varepsilon}{2}\Big(H(\mu_0\,|\,\mathfrak{m}) + H(\mu_1\,|\,\mathfrak{m})\Big) + \iint_0^1 \Big(\frac{|\nabla\vartheta_t^\varepsilon|^2}{2} + \frac{\varepsilon^2}{8}|\nabla\log\rho_t^\varepsilon|^2\Big)\rho_t^\varepsilon \mathrm{d}t\mathrm{d}\mathfrak{m}\\ &= \varepsilon H(\mu_0\,|\,\mathfrak{m}) + \iint_0^1 \frac{|\nabla\psi_t^\varepsilon|^2}{2}\rho_t^\varepsilon \mathrm{d}t\mathrm{d}\mathfrak{m} \qquad (4.1)\\ &= \varepsilon H(\mu_1\,|\,\mathfrak{m}) + \iint_0^1 \frac{|\nabla\varphi_t^\varepsilon|^2}{2}\rho_t^\varepsilon \mathrm{d}t\mathrm{d}\mathfrak{m}.\end{aligned}$$

proof Fix $\varepsilon > 0$ and let us prove the first identity in (4.1). To this aim fix $\bar{x} \in \mathrm{X}$, $R > 0$ and let χ_R be a cut-off function as in Lemma 2.1; recalling that $(\rho_t^\varepsilon) \in AC([0,1], L^2(\mathrm{X}))$ and for all compact set $\mathcal{C} \subset (0,1)$ there exists $M > 0$ such that $(\vartheta_t^\varepsilon) \in AC(\mathcal{C}, W^{1,2}(\mathrm{X}, e^{-V}\mathfrak{m}))$ with $V = M\mathsf{d}^2(\cdot,\bar{x})$, we see that $t \mapsto \int \chi_R\vartheta_t^\varepsilon\rho_t^\varepsilon \,\mathrm{d}\mathfrak{m}$ belongs to $AC_{loc}((0,1))$ with

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\int \chi_R\vartheta_t^\varepsilon\rho_t^\varepsilon \,\mathrm{d}\mathfrak{m} = \int \chi_R\Big(-\frac{|\nabla\vartheta_t^\varepsilon|^2}{2} - \frac{\varepsilon^2}{4}\Delta\log\rho_t^\varepsilon - \frac{\varepsilon^2}{8}|\nabla\log\rho_t^\varepsilon|^2\Big)\rho_t^\varepsilon\,\mathrm{d}\mathfrak{m}\\ - \int \chi_R\vartheta_t^\varepsilon \mathrm{div}(\rho_t^\varepsilon\nabla\vartheta_t^\varepsilon)\mathrm{d}\mathfrak{m} \qquad \text{a.e. } t \in (0,1).\end{aligned}$$

Integration by parts formula and integration in time on $[\delta, 1-\delta]$ with $\delta \in (0,1/2)$ then yield

$$\begin{aligned}\int \chi_R\vartheta_{1-\delta}^\varepsilon\rho_{1-\delta}^\varepsilon\,\mathrm{d}\mathfrak{m} - \int \chi_R\vartheta_\delta^\varepsilon\rho_\delta^\varepsilon\,\mathrm{d}\mathfrak{m} = \iint_\delta^{1-\delta} \chi_R\Big(\frac{|\nabla\vartheta_t^\varepsilon|^2}{2} + \frac{\varepsilon^2}{8}|\nabla\log\rho_t^\varepsilon|^2\Big)\rho_t^\varepsilon\,\mathrm{d}t\mathrm{d}\mathfrak{m}\\ + \iint_\delta^{1-\delta}\langle\nabla\chi_R, \nabla\rho_t^\varepsilon\rangle\,\mathrm{d}t\mathrm{d}\mathfrak{m} + \iint_\delta^{1-\delta}\vartheta_t^\varepsilon\langle\nabla\chi_R, \nabla\vartheta_t^\varepsilon\rangle\rho_t^\varepsilon\,\mathrm{d}t\mathrm{d}\mathfrak{m}.\end{aligned}$$

We claim that the limit as $R \to \infty$ can be carried under the integral signs. For the first summand on the right-hand side this is true by monotonicity, for all the other terms this follows from the dominated convergence theorem. Indeed $\vartheta_\delta^\varepsilon\rho_\delta^\varepsilon, \vartheta_{1-\delta}^\varepsilon\rho_{1-\delta}^\varepsilon \in L^1(\mathrm{X})$ by Lemma 3.3; $|\nabla\rho_t^\varepsilon|$ and $\vartheta_t^\varepsilon|\nabla\vartheta_t^\varepsilon|\rho_t^\varepsilon$ are, locally in t, uniformly bounded by an $L^1(\mathrm{X})$ function, since

$$|\nabla\rho_t^\varepsilon| = \rho_t^\varepsilon|\nabla\log\rho_t^\varepsilon|, \qquad |\vartheta_t^\varepsilon||\vartheta_t^\varepsilon\rho_t^\varepsilon| \le \frac{1}{2}\rho_t^\varepsilon\big(|\vartheta_t^\varepsilon|^2 + |\nabla\vartheta_t^\varepsilon|^2\big)$$

and because of Lemma 3.3 again; $|\chi_R|, |\nabla\chi_R|$ are uniformly bounded in $L^\infty(\mathrm{X})$ w.r.t. R and converge $\mathfrak{m}$-a.e. to $1, 0$ respectively by construction. Thus, we obtain

$$\int \vartheta_{1-\delta}^\varepsilon\rho_{1-\delta}^\varepsilon\,\mathrm{d}\mathfrak{m} - \int \vartheta_\delta^\varepsilon\rho_\delta^\varepsilon\,\mathrm{d}\mathfrak{m} = \iint_\delta^{1-\delta}\Big(\frac{|\nabla\vartheta_t^\varepsilon|^2}{2} + \frac{\varepsilon^2}{8}|\nabla\log\rho_t^\varepsilon|^2\Big)\rho_t^\varepsilon\,\mathrm{d}t\mathrm{d}\mathfrak{m}.$$

Now let $\delta \downarrow 0$: convergence of the right-hand side is trivial by monotonicity. For the left-hand side consider $t \mapsto \int \vartheta_t^\varepsilon\rho_t^\varepsilon\,\mathrm{d}\mathfrak{m}$, use the identity $\vartheta_t^\varepsilon = \psi_t^\varepsilon - \frac{\varepsilon}{2}\log\rho_t^\varepsilon$ and observe that

$$t \mapsto \int \psi_t^\varepsilon\rho_t^\varepsilon\,\mathrm{d}\mathfrak{m} \quad \text{ and } \quad t \mapsto -\frac{\varepsilon}{2}H(\mu_t^\varepsilon\,|\,\mathfrak{m})$$

are both continuous at $t=0$, the former by Lemma 3.3 and the latter by (2.11). This implies that

$$\lim_{\delta\downarrow 0}\int \vartheta^\varepsilon_\delta\rho^\varepsilon_\delta\,\mathrm{d}\mathfrak{m}=\int\psi^\varepsilon_0\rho_0\,\mathrm{d}\mathfrak{m}-\frac{\varepsilon}{2}H(\mu_0\,|\,\mathfrak{m}).$$

The same argument with the identity $\vartheta^\varepsilon_t=-\varphi^\varepsilon_t+\frac{\varepsilon}{2}\log\rho^\varepsilon_t$ allows us to handle $\int\vartheta^\varepsilon_{1-\delta}\rho^\varepsilon_{1-\delta}\,\mathrm{d}\mathfrak{m}$ too, so that

$$\begin{aligned}-\int\psi^\varepsilon_0\rho_0\,\mathrm{d}\mathfrak{m}-\int\varphi^\varepsilon_1\rho_1\,\mathrm{d}\mathfrak{m}+\frac{\varepsilon}{2}\Big(H(\mu_0\,|\,\mathfrak{m})+H(\mu_1\,|\,\mathfrak{m})\Big)\\=\iint_0^1\Big(\frac{|\nabla\vartheta^\varepsilon_t|^2}{2}+\frac{\varepsilon^2}{8}|\nabla\log\rho^\varepsilon_t|^2\Big)\rho^\varepsilon_t\,\mathrm{d}t\mathrm{d}\mathfrak{m}.\end{aligned}$$

Thanks to the identity $\varphi^\varepsilon_0+\psi^\varepsilon_0=\varepsilon\log\rho_0$ in $\mathrm{supp}(\mu_0)$ and the integrability of $\psi^\varepsilon_0\rho_0,\rho_0\log\rho_0$ we deduce that $\varphi^\varepsilon_0\rho_0\in L^1(\mathrm{X})$ too. An analogous statement holds in $t=1$ and thus the previous identity is in turn equivalent to

$$\begin{aligned}\int\varphi^\varepsilon_0\rho_0\,\mathrm{d}\mathfrak{m}+\int\psi^\varepsilon_1\rho_1\,\mathrm{d}\mathfrak{m}-\frac{\varepsilon}{2}\Big(H(\mu_0\,|\,\mathfrak{m})+H(\mu_1\,|\,\mathfrak{m})\Big)\\=\iint_0^1\Big(\frac{|\nabla\vartheta^\varepsilon_t|^2}{2}+\frac{\varepsilon^2}{8}|\nabla\log\rho^\varepsilon_t|^2\Big)\rho^\varepsilon_t\,\mathrm{d}t\mathrm{d}\mathfrak{m}\end{aligned}$$

and now it is sufficient to observe that by (2.10)

$$\varepsilon\mathscr{I}_\varepsilon(\mu_0,\mu_1)=\varepsilon H\big(f^\varepsilon\otimes g^\varepsilon\mathsf{R}^{\varepsilon/2}\,|\,\mathsf{R}^{\varepsilon/2}\big)=\int\varphi^\varepsilon_0\mathrm{d}\mu_0+\int\psi^\varepsilon_1\mathrm{d}\mu_1.\tag{4.2}$$

For the second and third identities in (4.1), the argument closely follows the one we have just presented. Indeed, it is just a matter of computation to rewrite the continuity equation solved by $(\rho^\varepsilon_t,\vartheta^\varepsilon_t)$ as forward and backward Fokker-Planck equations with velocity fields given by $\nabla\psi^\varepsilon_t$ and $\nabla\varphi^\varepsilon_t$ respectively, i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho^\varepsilon_t+\mathrm{div}(\rho^\varepsilon_t\nabla\psi^\varepsilon_t)=\frac{\varepsilon}{2}\Delta\rho^\varepsilon_t\tag{4.3a}$$

$$-\frac{\mathrm{d}}{\mathrm{d}t}\rho^\varepsilon_t+\mathrm{div}(\rho^\varepsilon_t\nabla\varphi^\varepsilon_t)=\frac{\varepsilon}{2}\Delta\rho^\varepsilon_t\tag{4.3b}$$

where the time derivatives are meant in the $L^2(\mathrm{X})$ sense as in (2.12). Therefore, arguing as above it is not difficult to see that

$$\int\varphi^\varepsilon_1\mathrm{d}\mu_1-\int\varphi^\varepsilon_0\mathrm{d}\mu_0=-\iint_0^1\frac{|\nabla\varphi^\varepsilon_t|^2}{2}\rho^\varepsilon_t\mathrm{d}t\mathrm{d}\mathfrak{m}$$

and an analogous identity holds true for ψ^ε_t. Finally using the identities $\varphi^\varepsilon_0+\psi^\varepsilon_0=\varepsilon\log\rho_0$ in $\mathrm{supp}(\mu_0)$ and $\varphi^\varepsilon_1+\psi^\varepsilon_1=\varepsilon\log\rho_1$ in $\mathrm{supp}(\mu_1)$, the entropic cost can be rewritten as

$$\varepsilon\mathscr{I}_\varepsilon(\mu_0,\mu_1)=\varepsilon H(\mu_0\,|\,\mathfrak{m})+\int\psi^\varepsilon_1\mathrm{d}\mu_1-\int\psi^\varepsilon_0\mathrm{d}\mu_0=\varepsilon H(\mu_1\,|\,\mathfrak{m})+\int\varphi^\varepsilon_0\mathrm{d}\mu_0-\int\varphi^\varepsilon_1\mathrm{d}\mu_1,$$

whence the conclusion. □

From this result we notice that the entropic cost can be expressed as the evaluation of an action functional in three different ways: as a purely kinetic energy evaluated along $(\rho_t^\varepsilon, \psi_t^\varepsilon)$ and $(\rho_t^\varepsilon, \varphi_t^\varepsilon)$ or as the sum of kinetic energy and Fisher information along $(\rho_t^\varepsilon, \vartheta_t^\varepsilon)$. As we have just seen, three different 'PDEs' are associated to these couples, so that three different minimization problems can be introduced, as already discussed in the Introduction.

The first of them reads as

$$\inf_{(\eta,v)} \iint_0^1 \Big(\frac{|v_t|^2}{2} + \frac{\varepsilon^2}{8} |\nabla \log \eta_t|^2 \Big) \eta_t \mathrm{d}t \mathrm{d}\mathfrak{m},$$

the infimum being taken among all solutions $(\eta\mathfrak{m}, v)$ of the continuity equation in the sense of Definition 3.1. In line with the smooth theory, we are now able to prove that, if we add $\frac{\varepsilon}{2}(H(\mu_0 \,|\, \mathfrak{m}) + H(\mu_1 \,|\, \mathfrak{m}))$, this infimum coincides with $\varepsilon \mathscr{I}_\varepsilon(\mu_0, \mu_1)$, thus providing a first variational representation of the entropic cost and extending (1.11) to the RCD framework. Moreover, the infimum is attained if and only if $(\eta_t, v_t) = (\rho_t^\varepsilon, \nabla \vartheta_t^\varepsilon)$. We remark that the uniqueness of the minimizers is not stated in [29], [11] and [13].

Theorem 4.2 (Benamou-Brenier formula for the entropic cost, 1st form)**.** *With the same assumptions and notations as in Section 2.2, for any $\varepsilon > 0$ the following holds:*

$$\varepsilon \mathscr{I}_\varepsilon(\mu_0, \mu_1) = \frac{\varepsilon}{2} \Big(H(\mu_0 \,|\, \mathfrak{m}) + H(\mu_1 \,|\, \mathfrak{m}) \Big) + \min_{(\eta,v)} \left\{ \iint_0^1 \Big(\frac{|v_t|^2}{2} + \frac{\varepsilon^2}{8} |\nabla \log \eta_t|^2 \Big) \eta_t \mathrm{d}t \mathrm{d}\mathfrak{m} \right\} \quad (4.4)$$

where the minimum is taken among all couples $(\eta\mathfrak{m}, v)$ solving the continuity equation in the sense of Definition 3.1 under the constraint $\eta_0 \mathfrak{m} = \mu_0$ and $\eta_1 \mathfrak{m} = \mu_1$; moreover, the minimum is attained if and only if $(\eta_t, v_t) = (\rho_t^\varepsilon, \nabla \vartheta_t^\varepsilon)$.

proof
Inequality $\geq$ in (4.4). Proposition 4.1, the third in (2.12) and the last in (2.15) imply that $(\rho_t^\varepsilon, \nabla \vartheta_t^\varepsilon)$ solves the continuity equation in the sense of Definition 3.1.
Inequality $\leq$ in (4.4). Start noticing that the assumptions on μ_0, μ_1 grant that $\mathscr{I}_\varepsilon(\mu_0, \mu_1)$ is finite. Thus, given a solution (η, v) of the continuity equation in the sense of the statement, without loss of generality we can assume that

$$\iint_0^1 \Big(\frac{|v_t|^2}{2} + \frac{\varepsilon^2}{8} |\nabla \log \eta_t|^2 \Big) \eta_t \mathrm{d}t \mathrm{d}\mathfrak{m} < +\infty. \quad (4.5)$$

Now fix $x \in \mathrm{X}$, $R > 0$ and pick a cut-off function χ_R as in Lemma 2.1; take also $\delta > 0$ and define for all $t \in [0,1]$

$$\varphi_t^{\varepsilon,\delta} := \varepsilon \log(f_t^\varepsilon + \delta) \qquad \psi_t^{\varepsilon,\delta} := \varepsilon \log(g_t^\varepsilon + \delta) \qquad \vartheta_t^{\varepsilon,\delta} := \frac{1}{2}\big(\psi_t^{\varepsilon,\delta} - \varphi_t^{\varepsilon,\delta}\big).$$

By Lemma 3.4 $(\chi_R \vartheta_t^{\varepsilon,\delta}) \in AC_{loc}((0,1), L^2(\mathrm{X})) \cap L^\infty_{loc}((0,1), W^{1,2}(\mathrm{X}))$. Thus, given $t_0, t_1 \in (0,1)$ with $t_0 < t_1$, Lemma 3.5 applies to $(\eta_t \mathfrak{m})$ and $t \mapsto \chi_R \vartheta_t^{\varepsilon,\delta}$ on $[t_0, t_1]$, whence

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big(\int \chi_R \vartheta_s^{\varepsilon,\delta} \eta_s \mathrm{d}\mathfrak{m} \Big)_{|s=t} = \int \chi_R \big(\frac{\mathrm{d}}{\mathrm{d}s} \vartheta_s^{\varepsilon,\delta}{}_{|s=t} \big) \eta_t \mathrm{d}\mathfrak{m} + \frac{\mathrm{d}}{\mathrm{d}s}\Big(\int \chi_R \vartheta_t^{\varepsilon,\delta} \eta_s \mathrm{d}\mathfrak{m} \Big)_{|s=t}$$

for a.e. $t \in [t_0, t_1]$. For the first term on the right-hand side, the fact that $\vartheta_t^{\varepsilon,\delta} = (\psi_t^{\varepsilon,\delta} - \varphi_t^{\varepsilon,\delta})/2$ and the 'PDEs' solved by $\psi_t^{\varepsilon,\delta}, \varphi_t^{\varepsilon,\delta}$ (namely (3.1) up to rescaling and change of sign) yield

$$\begin{aligned}\int \chi_R\big(\frac{\mathrm{d}}{\mathrm{d}s}\vartheta_s^{\varepsilon,\delta}|_{s=t}\big)\eta_t \mathrm{d}\mathfrak{m} &= -\int \chi_R\Big(\frac{|\nabla\psi_t^{\varepsilon,\delta}|^2}{4} + \frac{\varepsilon}{4}\Delta\psi_t^{\varepsilon,\delta} + \frac{|\nabla\varphi_t^{\varepsilon,\delta}|^2}{4} + \frac{\varepsilon}{4}\Delta\varphi_t^{\varepsilon,\delta}\Big)\eta_t \mathrm{d}\mathfrak{m}\\ &= \int \chi_R\Big(-\frac{|\nabla\psi_t^{\varepsilon,\delta}|^2}{4} - \frac{|\nabla\varphi_t^{\varepsilon,\delta}|^2}{4} + \frac{\varepsilon}{4}\langle\nabla(\psi_t^{\varepsilon,\delta} + \varphi_t^{\varepsilon,\delta}), \nabla\log\eta_t\rangle\Big)\eta_t \mathrm{d}\mathfrak{m}\\ &\quad + \frac{\varepsilon}{4}\int\langle\nabla(\psi_t^{\varepsilon,\delta} + \varphi_t^{\varepsilon,\delta}), \nabla\chi_R\rangle\,\eta_t \mathrm{d}\mathfrak{m}\end{aligned}$$

and by Young's inequality

$$\varepsilon\langle\nabla(\psi_t^{\varepsilon,\delta} + \varphi_t^{\varepsilon,\delta}), \nabla\log\eta_t\rangle \le \frac{1}{2}|\nabla(\psi_t^{\varepsilon,\delta} + \varphi_t^{\varepsilon,\delta})|^2 + \frac{\varepsilon^2}{2}|\nabla\log\eta_t|^2. \tag{4.6}$$

On the other hand, the fact that (η, v) is a solution of the continuity equation and $\chi_R\vartheta_t^{\varepsilon,\delta} \in W^{1,2}(\mathrm{X})$ imply by Lemma 3.5 that

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big(\int \chi_R\vartheta_t^{\varepsilon,\delta}\eta_s \mathrm{d}\mathfrak{m}\Big)_{|s=t} = \frac{1}{2}\int\Big(\chi_R\langle\nabla(\psi_t^{\varepsilon,\delta} - \varphi_t^{\varepsilon,\delta}), v_t\rangle + (\psi_t^{\varepsilon,\delta} - \varphi_t^{\varepsilon,\delta})\langle\nabla\chi_R, v_t\rangle\Big)\eta_t \mathrm{d}\mathfrak{m}$$

and by Young's inequality

$$\langle\nabla(\psi_t^{\varepsilon,\delta} - \varphi_t^{\varepsilon,\delta}), v_t\rangle \le \frac{1}{4}|\nabla(\psi_t^{\varepsilon,\delta} - \varphi_t^{\varepsilon,\delta})|^2 + |v_t|^2. \tag{4.7}$$

Plugging these observations together and integrating over $[t_0, t_1]$ we deduce that

$$\begin{aligned}\int \chi_R\vartheta_{t_1}^{\varepsilon,\delta}\eta_{t_1}\,\mathrm{d}\mathfrak{m} - \int \chi_R\vartheta_{t_0}^{\varepsilon,\delta}\eta_{t_0}\,\mathrm{d}\mathfrak{m} &\le \iint_{t_0}^{t_1}\chi_R\Big(\frac{|v_t|^2}{2} + \frac{\varepsilon^2}{8}|\nabla\log\eta_t|^2\Big)\eta_t \mathrm{d}t\mathrm{d}\mathfrak{m}\\ &\quad + \frac{\varepsilon}{4}\iint_{t_0}^{t_1}\langle\nabla(\psi_t^{\varepsilon,\delta} + \varphi_t^{\varepsilon,\delta}), \nabla\chi_R\rangle\,\eta_t \mathrm{d}t\mathrm{d}\mathfrak{m}\\ &\quad + \frac{1}{2}\iint_{t_0}^{t_1}(\psi_t^{\varepsilon,\delta} - \varphi_t^{\varepsilon,\delta})\langle\nabla\chi_R, v_t\rangle\,\eta_t \mathrm{d}t\mathrm{d}\mathfrak{m}.\end{aligned}$$

Now notice that

$$\lim_{R\to\infty}\iint_{t_0}^{t_1}\chi_R\Big(\frac{|v_t|^2}{2} + \frac{\varepsilon^2}{8}|\nabla\log\eta_t|^2\Big)\eta_t \mathrm{d}t\mathrm{d}\mathfrak{m} = \iint_{t_0}^{t_1}\Big(\frac{|v_t|^2}{2} + \frac{\varepsilon^2}{8}|\nabla\log\eta_t|^2\Big)\eta_t \mathrm{d}t\mathrm{d}\mathfrak{m}$$

by monotonicity. Moreover, on the left-hand side $\vartheta_t^{\varepsilon,\delta}\eta_t \in L^1(\mathrm{X})$ for $t = t_0, t_1$ by (iv) in Lemma 3.4 and the very definition of $\vartheta_t^{\varepsilon,\delta}$; on the right-hand side we know that $\||\nabla\chi_R|\|_{L^\infty(\mathrm{X})}$ is uniformly bounded w.r.t. R, $|\nabla(\psi_t^{\varepsilon,\delta} + \varphi_t^{\varepsilon,\delta})|\eta_t \in L^1([t_0, t_1]\times\mathrm{X}, \mathrm{d}t\otimes\mathfrak{m})$ by Lemma 3.4 again,

$$|(\psi_t^{\varepsilon,\delta} - \varphi_t^{\varepsilon,\delta})\langle\nabla\chi_R, v_t\rangle\,\eta_t| \le \frac{1}{2}|\nabla\chi_R|\eta_t\big(|\psi_t^{\varepsilon,\delta} - \varphi_t^{\varepsilon,\delta}|^2 + |v_t|^2\big)$$

and $|v_t|^2\eta_t \in L^1([0,1]\times\mathrm{X}, \mathrm{d}t\otimes\mathfrak{m})$ by (4.5). Thus the dominated convergence theorem applies to all the remaining terms and, since $|\nabla\chi_R| \to 0$ $\mathfrak{m}$-a.e. as $R \to \infty$, this implies

$$\int \vartheta_{t_1}^{\varepsilon,\delta}\eta_{t_1}\,\mathrm{d}\mathfrak{m} - \int \vartheta_{t_0}^{\varepsilon,\delta}\eta_{t_0}\,\mathrm{d}\mathfrak{m} \le \iint_{t_0}^{t_1}\Big(\frac{|v_t|^2}{2} + \frac{\varepsilon^2}{8}|\nabla\log\eta_t|^2\Big)\eta_t \mathrm{d}t\mathrm{d}\mathfrak{m}. \tag{4.8}$$

Passing to the limit as $t_0 \downarrow 0$, the convergence of the right-hand side is trivial by monotonicity. As regards the left-hand side, notice that

$$\left| \int \psi_{t_0}^{\varepsilon,\delta} \eta_{t_0} \, \mathrm{d}\mathfrak{m} - \int \psi_0^{\varepsilon,\delta} \, \mathrm{d}\mu_0 \right| \leq \int |\psi_{t_0}^{\varepsilon,\delta} - \psi_0^{\varepsilon,\delta}| \eta_{t_0} \, \mathrm{d}\mathfrak{m} + \left| \int \psi_0^{\varepsilon,\delta} \eta_{t_0} \, \mathrm{d}\mathfrak{m} - \int \psi_0^{\varepsilon,\delta} \, \mathrm{d}\mu_0 \right| \tag{4.9}$$

and the fact that log is Lipschitz on $[\delta, \infty)$ together with the fact that $\eta_t \leq C$ for all $t \in [0,1]$ entails

$$\int |\psi_{t_0}^{\varepsilon,\delta} - \psi_0^{\varepsilon,\delta}| \eta_{t_0} \, \mathrm{d}\mathfrak{m} \leq C' \| g_{t_0}^\varepsilon - g_0^\varepsilon \|_{L^2(\mathrm{X})},$$

so that the first term on the right-hand side of (4.9) vanishes as $t_0 \downarrow 0$, since $t \mapsto g_t^\varepsilon$ is L^2-continuous. As regards the second one, it also disappears: indeed $g_0^\varepsilon \in C_b(\mathrm{X})$ (as a consequence of $g^\varepsilon \in L^\infty(\mathrm{X})$ with compact support, (2.5) and the maximum principle for the heat equation) so that $\psi_0^{\varepsilon,\delta} \in C_b(\mathrm{X})$ too, and $\eta_{t_0}\mathfrak{m} \rightharpoonup \mu_0$ as $t_0 \downarrow 0$ by definition.

For $\varphi_t^{\varepsilon,\delta}$ we argue in an analogous way: we write

$$\left| \int \varphi_{t_0}^{\varepsilon,\delta} \eta_{t_0} \, \mathrm{d}\mathfrak{m} - \int \varphi_0^{\varepsilon,\delta} \, \mathrm{d}\mu_0 \right| \leq \int |\varphi_{t_0}^{\varepsilon,\delta} - \varphi_0^{\varepsilon,\delta}| \eta_{t_0} \, \mathrm{d}\mathfrak{m} + \left| \int \varphi_0^{\varepsilon,\delta} \eta_{t_0} \, \mathrm{d}\mathfrak{m} - \int \varphi_0^{\varepsilon,\delta} \, \mathrm{d}\mu_0 \right|$$

and get rid of the first term on the right-hand side as just done for $\psi_t^{\varepsilon,\delta}$. For the second one, let us stress that in general neither f^ε nor $\varphi_0^{\varepsilon,\delta}$ belong to $C_b(\mathrm{X})$ and thus narrow convergence can not be applied. However, f^ε has compact support, since so does ρ_0: this means that $\varphi_0^{\varepsilon,\delta}$ is constant outside a bounded set and thus for all $\alpha > 0$ we can find $h \in C_b(\mathrm{X})$ such that $\|\varphi_0^{\varepsilon,\delta} - h\|_{L^1(\mathrm{X})} < \alpha$; therefore, using again $\eta_t \leq C$, we get

$$\begin{aligned} \left| \int \varphi_0^{\varepsilon,\delta} \eta_{t_0} \, \mathrm{d}\mathfrak{m} - \int \varphi_0^{\varepsilon,\delta} \, \mathrm{d}\mu_0 \right| &\leq \int |\varphi_0^{\varepsilon,\delta} - h| \eta_{t_0} \, \mathrm{d}\mathfrak{m} + \left| \int h \eta_{t_0} \, \mathrm{d}\mathfrak{m} - \int h \, \mathrm{d}\mu_0 \right| + \int |\varphi_0^{\varepsilon,\delta} - h| \, \mathrm{d}\mu_0 \\ &\leq 2C\alpha + \left| \int h \eta_{t_0} \, \mathrm{d}\mathfrak{m} - \int h \, \mathrm{d}\mu_0 \right| \end{aligned}$$

and the arbitrariness of α together with $\eta_{t_0}\mathfrak{m} \rightharpoonup \mu_0$ as $t_0 \downarrow 0$ allows us to conclude in the same way as for $\psi_t^{\varepsilon,\delta}$. Since $\vartheta_t^{\varepsilon,\delta} := (\psi_t^{\varepsilon,\delta} - \varphi_t^{\varepsilon,\delta})/2$, this implies

$$\lim_{t_0 \downarrow 0} \int \vartheta_{t_0}^{\varepsilon,\delta} \eta_{t_0} \, \mathrm{d}\mathfrak{m} = \int \vartheta_0^{\varepsilon,\delta} \, \mathrm{d}\mu_0.$$

In a completely analogous way we can handle the case $t_1 \uparrow 1$, whence from (4.8)

$$\int \vartheta_1^{\varepsilon,\delta} \, \mathrm{d}\mu_1 - \int \vartheta_0^{\varepsilon,\delta} \, \mathrm{d}\mu_0 \leq \iint_0^1 \Big(\frac{|v_t|^2}{2} + \frac{\varepsilon^2}{8} |\nabla \log \eta_t|^2 \Big) \eta_t \mathrm{d}t \mathrm{d}\mathfrak{m}. \tag{4.10}$$

Now use the identity $\vartheta_1^{\varepsilon,\delta} = \psi_1^{\varepsilon,\delta} - \frac{\varepsilon}{2} \log((f_1^\varepsilon + \delta)(g^\varepsilon + \delta))$ and observe that by monotonicity

$$\lim_{\delta \downarrow 0} \int \psi_1^{\varepsilon,\delta} \, \mathrm{d}\mu_1 = \int \psi_1^\varepsilon \, \mathrm{d}\mu_1, \qquad \lim_{\delta \downarrow 0} \int \log((f_1^\varepsilon + \delta)(g^\varepsilon + \delta)) \mathrm{d}\mu_1 = H(\mu_1 \,|\, \mathfrak{m}).$$

Moreover both limits are finite, since the assumptions on ρ_1 grant that $H(\mu_1 \,|\, \mathfrak{m}) < \infty$, while the first one can be rewritten as $\varepsilon H(\mu_1 \,|\, \mathfrak{m}) - \int \varphi_1^\varepsilon \, \mathrm{d}\mu_1$ and $\varphi_1^\varepsilon \rho_1 \in L^1(\mathrm{X})$ by Lemma 3.3. From

the identity $\vartheta_0^{\varepsilon,\delta} = -\varphi_0^{\varepsilon,\delta} + \frac{\varepsilon}{2}\log((f^\varepsilon+\delta)(g_0^\varepsilon+\delta))$ an analogous statement holds for $\int \vartheta_0^{\varepsilon,\delta}\,\mathrm{d}\mu_0$ as $\delta \downarrow 0$ too, so that combining these remarks with (4.10) and (4.2) we get

$$\varepsilon\mathscr{I}_\varepsilon(\mu_0,\mu_1) \le \frac{\varepsilon}{2}\Big(H(\mu_0\,|\,\mathfrak{m}) + H(\mu_1\,|\,\mathfrak{m})\Big) + \iint_0^1 \Big(\frac{|v_t|^2}{2} + \frac{\varepsilon^2}{8}|\nabla\log\eta_t|^2\Big)\eta_t\mathrm{d}t\mathrm{d}\mathfrak{m},$$

whence the conclusion by taking the infimum.

Uniqueness of the minimizer. As a first step, notice that if $(\eta_t\mathfrak{m}, v_t)$ solves the continuity equation as in Definition 3.1 and we introduce the 'momentum' variable $m_t := \eta_t v_t$, then

$$\Gamma := \{(\eta, m) \,:\, (\eta\mathfrak{m}, v) \text{ solves the continuity equation, } \eta_0\mathfrak{m} = \mu_0,\ \eta_1\mathfrak{m} = \mu_1\}$$

is closed w.r.t. convex combination. Secondly, the function $\Phi : [0,\infty)\times\mathbb{R} \to [0,\infty]$ defined by

$$\Phi(x,y) := \begin{cases} \dfrac{y^2}{x} & \text{if } x > 0,\\ 0 & \text{if } x = y = 0,\\ +\infty & \text{otherwise} \end{cases}$$

is convex and lower semicontinuous. Therefore the functionals $\mathscr{K}, \mathscr{F}, \mathscr{A} : \Gamma \to [0,\infty]$ defined as

$$\mathscr{K}(\eta,m) := \frac{1}{2}\iint_0^1 \frac{|m_t|^2}{\eta_t}\,\mathrm{d}t\mathrm{d}\mathfrak{m}, \quad \mathscr{F}(\eta,m) := \frac{\varepsilon^2}{8}\iint_0^1 \frac{|\nabla\eta_t|^2}{\eta_t}\,\mathrm{d}t\mathrm{d}\mathfrak{m}, \quad \mathscr{A} := \mathscr{K} + \mathscr{F}$$

(in case the set of t's where $\eta_t \notin W^{1,2}_{loc}(\mathrm{X})$ has positive $\mathcal{L}^1$-measure, we set $\mathscr{F}(\eta,m) := +\infty$) are convex too. Thus, if $(\overline{\eta},\overline{v})$ is a minimizer for (4.4) and $\overline{m}_t := \overline{\eta}_t\overline{v}_t$, then by Proposition 4.1 it follows that $\mathscr{A}(\overline{\eta},\overline{m}) = \mathscr{A}(\rho^\varepsilon, \rho^\varepsilon\nabla\vartheta^\varepsilon)$ and so, by convexity of $\mathscr{A}$,

$$\mathscr{A}(\eta^\lambda, m^\lambda) = (1-\lambda)\mathscr{A}(\overline{\eta},\overline{m}) + \lambda\mathscr{A}(\rho^\varepsilon,\rho^\varepsilon\nabla\vartheta^\varepsilon) \qquad \forall\lambda\in(0,1),$$

where $\eta_t^\lambda := (1-\lambda)\eta_t + \lambda\rho_t^\varepsilon$ and $m_t^\lambda := (1-\lambda)m_t + \lambda\rho_t^\varepsilon\nabla\vartheta_t^\varepsilon$. As a byproduct the same identity holds with $\mathscr{F}$ instead of $\mathscr{A}$, since both $\mathscr{K}$ and $\mathscr{F}$ are convex and $\mathscr{A} = \mathscr{K} + \mathscr{F}$, whence

$$\Phi(\eta_t^\lambda, |\nabla\eta_t^\lambda|) = (1-\lambda)\Phi(\overline{\eta}_t, |\nabla\overline{\eta}_t|) + \lambda\Phi(\rho_t^\varepsilon, |\nabla\rho_t^\varepsilon|). \tag{4.11}$$

Taking into account the fact that Φ is linear only along the lines passing through the origin, we deduce that $(\overline{\eta}_t, |\nabla\overline{\eta}_t|)$ and $(\rho_t^\varepsilon, |\nabla\rho_t^\varepsilon|)$ are collinear. Since $\rho_t^\varepsilon > 0$, it is not restrictive to assume that the collinearity condition can be expressed as follows: for all $t\in(0,1)$ and $\mathfrak{m}$-a.e. $x\in\mathrm{X}$ there exists $\alpha_{x,t}\ge 0$ such that

$$(\overline{\eta}_t, |\nabla\overline{\eta}_t|) = \alpha_{x,t}(\rho_t^\varepsilon, |\nabla\rho_t^\varepsilon|). \tag{4.12}$$

Since $\alpha_{x,t} = \overline{\eta}_t/\rho_t^\varepsilon$ with $\rho_t^\varepsilon \in W^{1,2}(\mathrm{X})$ locally bounded away from 0 and $\overline{\eta}_t \in W^{1,2}_{loc}(\mathrm{X})$ for a.e. t, this implies that $\alpha_{x,t} \in W^{1,2}_{loc}(\mathrm{X})$ for a.e. t. Furthermore, on the one hand by the very definition of η_t^λ together with (4.12) we have

$$\eta_t^\lambda = \big((1-\lambda)\alpha_{x,t} + \lambda\big)\rho_t^\varepsilon$$

and on the other hand (4.11) also implies that $(\eta_t^\lambda, |\nabla\eta_t^\lambda|)$ is collinear with $(\rho_t^\varepsilon, |\nabla\rho_t^\varepsilon|)$: this implies

$$(\eta_t^\lambda, |\nabla\eta_t^\lambda|) = \big((1-\lambda)\alpha_{x,t} + \lambda\big)(\rho_t^\varepsilon, |\nabla\rho_t^\varepsilon|).$$

Considering the gradient of the identity on the first coordinate we see that

$$|\nabla\eta_t^\lambda|^2 = \big((1-\lambda)\alpha_{x,t}+\lambda\big)^2|\nabla\rho_t^\varepsilon|^2+(1-\lambda)^2(\rho_t^\varepsilon)^2|\nabla\alpha_{x,t}|^2+2(1-\lambda)\big((1-\lambda)\alpha_{x,t}+\lambda\big)\rho_t^\varepsilon\langle\nabla\rho_t^\varepsilon,\nabla\alpha_{x,t}\rangle$$

and plugging the identity on the second coordinate therein this becomes

$$|\nabla\alpha_{x,t}|^2 = -2\Big(\alpha_{x,t}+\frac{\lambda}{1-\lambda}\Big)\langle\nabla\log\rho_t^\varepsilon,\nabla\alpha_{x,t}\rangle, \qquad \forall\lambda\in(0,1). \tag{4.13}$$

Since $\alpha_{x,t}$ does not depend on λ, this can be true only if $\langle\nabla\log\rho_t^\varepsilon,\nabla\alpha_{x,t}\rangle = 0$ $\mathfrak{m}$-a.e. thus forcing also the left-hand side of (4.13) to vanish, namely $\alpha_{x,t}$ to be constant. Since $\overline{\eta}_t, \rho_t^\varepsilon$ are probability densities, we conclude that $\alpha_{x,t}\equiv 1$.

Thus $\overline{\eta}_t = \rho_t^\varepsilon$ for all $t\in[0,1]$ and now it is sufficient to use the strict convexity of $(v_t)\mapsto\iint_0^1|v_t|^2\rho_t^\varepsilon\,\mathrm{d}t\mathrm{d}\mathfrak{m}$ to conclude that $\overline{v}_t = \nabla\vartheta_t^\varepsilon$ for all $t\in[0,1]$. □

Remark 4.3 (Uniqueness of the minimizer)**.** As proved in [35], if X is assumed to be a compact $\mathsf{RCD}^*(K,N)$ space and ρ_0,ρ_1 are sufficiently regular (i.e. they belong to $\mathrm{Test}^\infty_{>0}(\mathrm{X})$ with the notation used therein), then the fact that the minimum is attained in (4.4) if and only if $(\eta_t,v_t) = (\rho_t^\varepsilon,\nabla\vartheta_t^\varepsilon)$ is straightforward. This is essentially due to the fact that no cut-off argument is needed and $(\vartheta_t^\varepsilon)\in AC([0,1],W^{1,2}(\mathrm{X}))$, so that $t\mapsto\int\vartheta_t^\varepsilon\eta_t\,\mathrm{d}\mathfrak{m}$ belongs to $AC([0,1])$, no limit as $t_0\downarrow 0, t_1\uparrow 1$ and $\delta\downarrow 0$ appears in the previous proof and by the case of equality in (4.6) and (4.7) the infimum is attained if and only if

$$\nabla(\psi_t^\varepsilon+\varphi_t^\varepsilon) = \varepsilon\nabla\log\eta_t \qquad\text{and}\qquad v_t = \frac{1}{2}\nabla(\psi_t^\varepsilon-\varphi_t^\varepsilon),$$

which completely characterizes the optimal pair (η_t,v_t). ■

The second and third dynamical variational representation of the entropic cost (and thus the non-smooth analogue of (1.9)) are the content of the next result.

Theorem 4.4 (Benamou-Brenier formula for the entropic cost, 2nd form)**.** *With the same assumptions and notations as in Section 2.2, for any $\varepsilon>0$ the following holds:*

(i)

$$\varepsilon\mathscr{I}_\varepsilon(\mu_0,\mu_1) = \varepsilon H(\mu_0\,|\,\mathfrak{m}) + \min_{(\nu^+,v^+)}\Big\{\int_0^1\int\frac{|v_t^+|^2}{2}\mathrm{d}\nu_t^+\mathrm{d}t\Big\} \tag{4.14}$$

where the minimum is taken among all couples (ν^+,v^+) solving the forward Fokker-Planck equation in the sense of Definition 3.1 with $c=\varepsilon/2$ under the constraint $\nu_0=\mu_0$ and $\nu_1=\mu_1$; the minimum is attained at $(\rho_t^\varepsilon\mathfrak{m},\nabla\psi_t^\varepsilon)$.

(ii)

$$\varepsilon\mathscr{I}_\varepsilon(\mu_0,\mu_1) = \varepsilon H(\mu_1\,|\,\mathfrak{m}) + \min_{(\nu^-,v^-)}\Big\{\int_0^1\int\frac{|v_t^-|^2}{2}\mathrm{d}\nu_t^-\mathrm{d}t\Big\} \tag{4.15}$$

where the minimum is taken among all couples (ν^-,v^-) solving the backward Fokker-Planck equation in the sense of Definition 3.1 with $c=\varepsilon/2$ under the constraint $\nu_0=\mu_0$ and $\nu_1=\mu_1$; the minimum is attained at $(\rho_t^\varepsilon\mathfrak{m},\nabla\varphi_t^\varepsilon)$.

proof It is sufficient to prove (i), as (ii) follows by swapping μ_0 and μ_1.
Inequality $\geq$ in (4.14). Proposition 4.1, (4.3a) and the second in (2.15) imply that $(\rho_t^\varepsilon, \nabla\psi_t^\varepsilon)$ solves the forward Fokker-Planck equation in the sense of Definition 3.1.
Inequality $\leq$ in (4.14). For sake of notation, let us drop the apex $+$ in (ν^+, v^+). Fix $x \in \mathrm{X}$, $R > 0$ and pick a cut-off function χ_R as in Lemma 2.1; take also $\delta > 0$ and define $\psi_t^{\varepsilon,\delta}$ as in (i). By Lemma 3.4 $(\chi_R\psi_t^{\varepsilon,\delta}) \in AC_{loc}([0,1), L^2(\mathrm{X})) \cap L^\infty_{loc}([0,1), W^{1,2}(\mathrm{X}))$ with $(\Delta(\chi_R\psi_t^{\varepsilon,\delta})) \in L^\infty_{loc}([0,1), L^2(\mathrm{X}))$. Thus, given $t_1 \in [0,1)$, Lemma 3.5 applies to (ν_t) and $t \mapsto \chi_R\psi_t^{\varepsilon,\delta}$ on $[0,t_1]$, whence

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big(\int \chi_R\psi_s^{\varepsilon,\delta}\,\mathrm{d}\nu_s\Big)_{|s=t} = \int \chi_R\big(\frac{\mathrm{d}}{\mathrm{d}s}\psi_s^{\varepsilon,\delta}{}_{|s=t}\big)\mathrm{d}\nu_t + \frac{\mathrm{d}}{\mathrm{d}s}\Big(\int \chi_R\psi_t^{\varepsilon,\delta}\mathrm{d}\nu_s\Big)_{|s=t}$$

for a.e. $t \in [0,t_1]$. For the first term on the right-hand side, the fact that $\psi_t^{\varepsilon,\delta}$ solves (3.1) (up to rescaling and change of sign) yields

$$\int \chi_R\big(\frac{\mathrm{d}}{\mathrm{d}s}\psi_s^{\varepsilon,\delta}{}_{|s=t}\big)\mathrm{d}\nu_t = -\int \chi_R\Big(\frac{|\nabla\psi_t^{\varepsilon,\delta}|^2}{2} + \frac{\varepsilon}{2}\Delta\psi_t^{\varepsilon,\delta}\Big)\mathrm{d}\nu_t.$$

On the other hand, the fact that (ν, v) is a solution of the forward Fokker-Planck equation with $c = \varepsilon/2$ and $\chi_R\psi_t^{\varepsilon,\delta} \in D(\Delta)$ imply that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\Big(\int \chi_R\psi_t^{\varepsilon,\delta}\,\mathrm{d}\nu_s\Big)_{|s=t} &= \int \Big(\chi_R\langle\nabla\psi_t^{\varepsilon,\delta}, v_t\rangle + \psi_t^{\varepsilon,\delta}\langle\nabla\chi_R, v_t\rangle\Big)\mathrm{d}\nu_t \\ &\quad + \frac{\varepsilon}{2}\int \Big(\chi_R\Delta\psi_t^{\varepsilon,\delta} + 2\langle\nabla\chi_R, \nabla\psi_t^{\varepsilon,\delta}\rangle + \psi_t^{\varepsilon,\delta}\Delta\chi_R\Big)\mathrm{d}\nu_t\end{aligned}$$

and by Young's inequality

$$\langle\nabla\psi_t^{\varepsilon,\delta}, v_t\rangle \leq \frac{1}{2}|\nabla\psi_t^{\varepsilon,\delta}|^2 + \frac{1}{2}|v_t|^2. \tag{4.16}$$

Plugging these observations together and integrating over $[0,t_1]$ we deduce that

$$\begin{aligned}\int \chi_R\psi_{t_1}^{\varepsilon,\delta}\,\mathrm{d}\nu_{t_1} - \int \chi_R\psi_0^{\varepsilon,\delta}\,\mathrm{d}\mu_0 &\leq \int_0^{t_1}\int \chi_R\frac{|v_t|^2}{2}\mathrm{d}\nu_t\mathrm{d}t + \int_0^{t_1}\int \psi_t^{\varepsilon,\delta}\langle\nabla\chi_R, v_t\rangle\mathrm{d}\nu_t\mathrm{d}t \\ &\quad + \frac{\varepsilon}{2}\int_0^{t_1}\int \Big(\psi_t^{\varepsilon,\delta}\Delta\chi_R + 2\langle\nabla\psi_t^{\varepsilon,\delta}, \nabla\chi_R\rangle\Big)\mathrm{d}\nu_t\mathrm{d}t.\end{aligned}$$

Since Lemma 2.1 ensures that $\|\Delta\chi_R\|_{L^\infty(\mathrm{X})}$ is uniformly bounded w.r.t. R and $t \mapsto \int |v_t|^2\mathrm{d}\nu_t$ belongs to $L^1(0,1)$ by Definition 3.1, the argument by dominated convergence explained in the proof of Theorem 4.2 applies and thus, passing to the limit as $R \to \infty$, we get

$$\int \psi_{t_1}^{\varepsilon,\delta}\,\mathrm{d}\nu_{t_1} - \int \psi_0^{\varepsilon,\delta}\,\mathrm{d}\mu_0 \leq \int_0^{t_1}\int \frac{|v_t|^2}{2}\mathrm{d}\nu_t\mathrm{d}t.$$

Both limits as $t_1 \uparrow 1$ and $\delta \downarrow 0$ can also be handled as in Theorem 4.2, whence

$$\int \psi_1^\varepsilon\,\mathrm{d}\mu_1 - \int \psi_0^\varepsilon\,\mathrm{d}\mu_0 \leq \int_0^1\int \frac{|v_t|^2}{2}\mathrm{d}\nu_t\mathrm{d}t,$$

and now it is sufficient to use the identity $\psi_0^\varepsilon = -\varphi_0^\varepsilon + \varepsilon\log\rho_0$ in $\mathrm{supp}(\mu_0)$ in conjunction with (4.2) to get the conclusion. □

Remark 4.5 (Uniqueness of the drift)**.** In the case X is assumed to be a compact $\mathsf{RCD}^*(K,N)$ space and ρ_0,ρ_1 are sufficiently regular, then as proved in [35] the drift of the optimal couple is uniquely determined, namely: the minimum is attained in (4.14) (resp. (4.15)) if and only if $v_t=\nabla\psi_t^\varepsilon$ (resp. $v_t=\nabla\varphi_t^\varepsilon$). As for Remark 4.3, this is essentially due to the fact that no cut-off argument is needed and $(\psi_t^\varepsilon)\in AC([0,1],W^{1,2}(\mathrm{X}))$, so that $t\mapsto\int\psi_t^\varepsilon\eta_t\,\mathrm{d}\mathfrak{m}$ belongs to $AC([0,1])$, no limit as $t_1\uparrow 1$ and $\delta\downarrow 0$ appears in the previous proof and by the case of equality in (4.16) the infimum is attained if and only if $v_t=\nabla\psi_t^\varepsilon$. ■

4.2 Dual representations for $\mathscr{I}_\varepsilon$

As already suggested by the proof of Theorem 4.4, the duality between Hamilton-Jacobi and continuity equation that appears in optimal transport is here replaced by the duality between forward (resp. backward) Hamilton-Jacobi-Bellman and backward (resp. forward) Fokker-Planck equation. This will be the content of the next result.

Theorem 4.6 (HJB duality for the entropic cost)**.** *Let* $(\mathrm{X},\mathsf{d},\mathfrak{m})$ *be an* $\mathsf{RCD}^*(K,N)$ *space with* $K\in\mathbb{R}$ *and* $N<\infty$*. Then, given any supersolution* (ϕ_t) *of the backward Hamilton-Jacobi-Bellman equation in the sense of Definition 3.2 and any solution* (ν_t,v_t) *of the forward Fokker-Planck equation in the sense of Definition 3.1 with the same parameter* c*, it holds*

$$\int\phi_1\,\mathrm{d}\nu_1-\int\phi_0\,\mathrm{d}\nu_0\le\frac12\int_0^1\int|v_t|^2\,\mathrm{d}\nu_t\mathrm{d}t. \tag{4.17}$$

Analogously, for any supersolution $(\tilde\phi_t)$ *of the forward Hamilton-Jacobi-Bellman equation in the sense of Definition 3.2 and any solution* $(\tilde\nu_t,\tilde v_t)$ *of the backward Fokker-Planck equation in the sense of Definition 3.1 with the same parameter* c*, we have*

$$\int\tilde\phi_0\,\mathrm{d}\tilde\nu_0-\int\tilde\phi_1\,\mathrm{d}\tilde\nu_1\le\frac12\int_0^1\int|\tilde v_t|^2\,\mathrm{d}\tilde\nu_t\mathrm{d}t. \tag{4.18}$$

In particular, with the same assumptions and notations as in Section 2.2, for any $\varepsilon>0$ *the following duality formula holds:*

$$\varepsilon\mathscr{I}_\varepsilon(\mu_0,\mu_1)=\varepsilon H(\mu_0\,|\,\mathfrak{m})+\sup\left\{\int\phi_1\,\mathrm{d}\mu_1-\int\phi_0\,\mathrm{d}\mu_0\right\} \tag{4.19a}$$

$$=\varepsilon H(\mu_1\,|\,\mathfrak{m})+\sup\left\{\int\tilde\phi_0\,\mathrm{d}\mu_0-\int\tilde\phi_1\,\mathrm{d}\mu_1\right\} \tag{4.19b}$$

where the supremum is taken among all supersolution of the backward (resp. forward) Hamilton-Jacobi-Bellman equation in the sense of Definition 3.2 with $c=\varepsilon/2$ *in* (4.19a) *(resp.* (4.19b)*).*

proof In order to prove (4.17) let (ϕ_t) and (ν_t,v_t) be as in the statement, fix $x\in\mathrm{X}$, $R>0$ and let χ_R be a cut-off function as in Lemma 2.1: by Definition 3.2 it follows that $(\chi_R\phi_t)\in AC([0,1],L^2(\mathrm{X}))\cap L^\infty([0,1],W^{1,2}(\mathrm{X}))$ with $(\Delta(\chi_R\phi_t))\in L^\infty([0,1],L^2(\mathrm{X}))$. Thus Lemma 3.5 applies to (ν_t) and $t\mapsto\chi_R\phi_t$ on $[0,1]$, whence

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big(\int\chi_R\phi_s\,\mathrm{d}\nu_s\Big)_{|s=t}=\int\chi_R\big(\frac{\mathrm{d}}{\mathrm{d}s}\phi_s{}_{|s=t}\big)\mathrm{d}\nu_t+\frac{\mathrm{d}}{\mathrm{d}s}\Big(\int\chi_R\phi_t\mathrm{d}\nu_s\Big)_{|s=t}$$

for a.e. $t \in [0,1]$. For the first term on the right-hand side, the fact that ϕ_t is a supersolution of the backward Hamilton-Jacobi-Bellman equation yields

$$\int \chi_R\big(\frac{\mathrm{d}}{\mathrm{d}s}\phi_s|_{s=t}\big)\mathrm{d}\nu_t \le -\int \chi_R\Big(\frac{|\nabla\phi_t|^2}{2} + c\Delta\phi_t\Big)\mathrm{d}\nu_t.$$

On the other hand, the fact that (ν_t, v_t) is a solution of the forward Fokker-Planck equation and $\chi_R\phi_t \in D(\Delta)$ imply that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\Big(\int \chi_R\phi_t\,\mathrm{d}\nu_s\Big)|_{s=t} &= \int \Big(\chi_R\langle\nabla\phi_t, v_t\rangle + \phi_t\langle\nabla\chi_R, v_t\rangle\Big)\mathrm{d}\nu_t \\ &\quad + c\int \Big(\chi_R\Delta\phi_t + 2\langle\nabla\chi_R, \nabla\phi_t\rangle + \phi_t\Delta\chi_R\Big)\mathrm{d}\nu_t\end{aligned}$$

and by Young's inequality

$$\langle\nabla\phi_t, v_t\rangle \le \frac{1}{2}|\nabla\phi_t|^2 + \frac{1}{2}|v_t|^2.$$

From these observations and integrating over $[0,1]$ we obtain

$$\begin{aligned}\int \chi_R\phi_1\,\mathrm{d}\nu_1 - \int \chi_R\phi_0\,\mathrm{d}\nu_0 &\le \int_0^1\int \chi_R\frac{|v_t|^2}{2}\mathrm{d}\nu_t\mathrm{d}t + \int_0^1\int \phi_t\langle\nabla\chi_R, v_t\rangle\mathrm{d}\nu_t\mathrm{d}t \\ &\quad + c\int_0^1\int \Big(\phi_t\Delta\chi_R + 2\langle\nabla\phi_t, \nabla\chi_R\rangle\Big)\mathrm{d}\nu_t\mathrm{d}t\end{aligned}$$

and, as in the proof of Theorems 4.2 and 4.4, we can pass the limit $R \to \infty$ under the integral sign by dominated convergence: indeed, all the integrability properties of $\varphi_t^{\varepsilon,\delta}, \psi_t^{\varepsilon,\delta}, \vartheta_t^{\varepsilon,\delta}$ that we used are still true for ϕ_t by Definition 3.2. Keeping in mind that $\chi_R \to 1$, $|\nabla\chi_R|, \Delta\chi_R \to 0$ $\mathfrak{m}$-a.e. as $R \to \infty$ and $\chi_R, |\chi_R|, \Delta\chi_R$ are uniformly bounded in $L^\infty(\mathrm{X})$ w.r.t. R, (4.17) follows.

As concerns (4.19a) the '$\ge$' inequality is a direct consequence of (4.17) and (4.14). For the opposite inequality, notice that $(\psi_t^{\varepsilon,\delta})$ as defined in the proof of Theorem 4.2 is a solution to the backward Hamilton-Jacobi-Bellman equation in the sense of Definition 3.2 only on the compact subsets of $[0,1)$. Thus let $\delta, s > 0$ and put

$$\phi_t^{\delta,s} := \varepsilon\log(\mathsf{h}_s g_t^\varepsilon + \delta) = \varepsilon\log(\mathsf{h}_{\varepsilon(1-t)/2+s}g^\varepsilon + \delta).$$

By Lemma 3.4 $(\phi_t^{\delta,s})$ is now a solution to the backward Hamilton-Jacobi-Bellman equation in the sense of Definition 3.2 on the whole $[0,1]$, whence

$$\int \phi_1^{\delta,s}\,\mathrm{d}\mu_1 - \int \phi_0^{\delta,s}\,\mathrm{d}\mu_0 \le \sup\Big\{\int \phi_1\,\mathrm{d}\mu_1 - \int \phi_0\,\mathrm{d}\mu_0\Big\}, \qquad \forall\delta, s > 0,$$

the supremum being considered as in (4.19a). The continuity of $s \mapsto \phi_t^{\delta,s}$ in $L^2(\mathrm{X}, e^{-V}\mathfrak{m})$ as $s \downarrow 0$ for all $t \in [0,1]$ (with $V = M\mathsf{d}^2(\cdot, x)$ for some $x \in \mathrm{X}$ and $M > 0$) together with the fact that μ_0, μ_1 have compact support entails

$$\lim_{s\downarrow 0}\int \phi_i^{\delta,s}\,\mathrm{d}\mu_1 = \int \psi_i^{\varepsilon,\delta}\,\mathrm{d}\mu_i, \qquad i = 0, 1$$

and arguing as in Theorem 4.2 we can pass to the limit as $\delta \downarrow 0$, thus getting

$$\int \psi_1^\varepsilon\,\mathrm{d}\mu_1 - \int \psi_0^\varepsilon\,\mathrm{d}\mu_0 \le \sup\Big\{\int \phi_1\,\mathrm{d}\mu_1 - \int \phi_0\,\mathrm{d}\mu_0\Big\}.$$

Now it is sufficient to use the identity $\psi_0^\varepsilon = -\varphi_0^\varepsilon + \varepsilon\log\rho_0$ in $\mathrm{supp}(\mu_0)$ together with (4.2) to conclude.

By reversing time and following the same strategy, (4.18) and (4.19b) also follow. □

It is not difficult to deduce from the previous result that the entropic cost admits a Kantorovich-like dual representation, the Hopf-Lax semigroup being replaced by (1.12) suitably rescaled.

Theorem 4.7 (Kantorovich duality for the entropic cost)**.** *With the same assumptions and notations as in Section 2.2, for any $\varepsilon > 0$ the following duality formula holds:*

$$\begin{aligned}\varepsilon \mathscr{I}_\varepsilon(\mu_0,\mu_1) &= \varepsilon H(\mu_0 \,|\, \mathfrak{m}) + \sup_{u\in\mathbb{V}} \left\{ \int u \,\mathrm{d}\mu_1 - \int Q_1^\varepsilon u \,\mathrm{d}\mu_0 \right\} \\ &= \varepsilon H(\mu_1 \,|\, \mathfrak{m}) + \sup_{u\in\mathbb{V}} \left\{ \int u \,\mathrm{d}\mu_0 - \int Q_1^\varepsilon u \,\mathrm{d}\mu_1 \right\}\end{aligned}$$

where $\mathbb{V} := \{u : \mathrm{X} \to \overline{\mathbb{R}} : e^{u/\varepsilon} \in L^2 \cap L^\infty(\mathrm{X})\}$, with the convention $e^{-\infty} = 0$, and

$$Q_1^\varepsilon u := \varepsilon \log\left(\mathsf{h}_{\varepsilon/2} e^{u/\varepsilon}\right).$$

proof Let us prove the second duality formula, as for the first one the argument is analogous. The '$\le$' inequality is a trivial consequence of (4.2), the identity $\varphi_1^\varepsilon + \psi_1^\varepsilon = \varepsilon \log \rho_1$ in $\mathrm{supp}(\mu_1)$ and the facts that $\varphi_0^\varepsilon \in \mathbb{V}$, $\varphi_1^\varepsilon = Q_1^\varepsilon \varphi_0^\varepsilon$. For the converse inequality, let $\delta, s > 0$ and define for all $t \in [0,1]$ and $u \in \mathbb{V}$

$$Q_t^{\varepsilon,\delta,s} u := \varepsilon \log\left(\mathsf{h}_{\varepsilon t/2+s} e^{u/\varepsilon} + \delta\right).$$

By Lemma 3.4 ($Q_t^{\varepsilon,\delta,s} u$) is a solution to the forward Hamilton-Jacobi-Bellman equation in the sense of Definition 3.2 on $[0,1]$, so that by (4.19b)

$$\varepsilon \mathscr{I}_\varepsilon(\mu_0,\mu_1) \ge \varepsilon H(\mu_1 \,|\, \mathfrak{m}) + \int Q_0^{\varepsilon,\delta,s} u \,\mathrm{d}\mu_0 - \int Q_1^{\varepsilon,\delta,s} u \,\mathrm{d}\mu_1.$$

Then let us pass to the limit as $s \downarrow 0$ as just done in the proof of Theorem 4.6 and, by monotonicity and the very definition of $\mathbb{V}$, as $\delta \downarrow 0$ too, thus getting

$$\varepsilon \mathscr{I}_\varepsilon(\mu_0,\mu_1) \ge \varepsilon H(\mu_1 \,|\, \mathfrak{m}) + \int u \,\mathrm{d}\mu_0 - \int Q_1^\varepsilon u \,\mathrm{d}\mu_1.$$

Since this is true for all $u \in \mathbb{V}$, we conclude. □

A Proofs of Lemmas 3.3, 3.4, 3.5

This appendix is devoted to the proofs of the auxiliary lemmas contained in Section 3.

proof of Lemma 3.3 Since $(\rho_t^\varepsilon) \in C([0,1], L^2(\mathrm{X}))$ and $(h_t^\varepsilon), (|\nabla h_t^\varepsilon) \in C(\mathcal{I}, L^2(\mathrm{X}, e^{-V}\mathfrak{m}))$, all the functions appearing in the statement are continuous from $\mathcal{I}$ to $L^0(\mathrm{X})$ equipped with the topology of convergence in measure on bounded sets. Therefore the continuity of $\mathcal{I} \ni t \mapsto \int H_t^\varepsilon \,\mathrm{d}\mathfrak{m}$ for these maps will follow as soon as we show that they are, locally in $t \in \mathcal{I}$, uniformly dominated by an $L^1(\mathrm{X})$ function. This will also imply all the other statements. Furthermore, it is sufficient to consider the case $h_t^\varepsilon = \varphi_t^\varepsilon$, as the estimates for ψ_t^ε can be obtained by symmetric arguments and the ones for $\vartheta_t^\varepsilon, \log \rho_t^\varepsilon$ follow from the identities $\vartheta_t^\varepsilon = \frac{\psi_t^\varepsilon - \varphi_t^\varepsilon}{2}$ and $\varepsilon \log \rho_t^\varepsilon = \varphi_t^\varepsilon + \psi_t^\varepsilon$.

From (2.14a) and (2.14b) we immediately see that for any $\bar x \in \mathrm{X}$ and $\delta > 0$ there exist constants $c_1, c_2 > 0$ depending on $K, N, \delta, \bar x, \rho_0, \rho_1$ only such that

$$\rho_t^\varepsilon |\nabla \varphi_t^\varepsilon|^2 \le c_1\big(1 + \mathsf{d}^2(\cdot, \bar x)\big) \exp\big(-c_2 \mathsf{d}^2(\cdot, \bar x)\big) \qquad \mathfrak{m}\text{-a.e.}$$

for every $t \in [\delta, 1]$ and $\varepsilon \in (0,1)$. The volume growth (2.8) then implies that the right-hand side is integrable and thus the conclusion.

For $\rho_t^\varepsilon \varphi_t^\varepsilon$ and $\rho_t^\varepsilon |h_t^\varepsilon|^2$, observe that from (2.14b) and the fact that X is a geodesic space it follows that

$$|\varphi_t^\varepsilon(x) - \varphi_t^\varepsilon(\bar x)| \le C_\delta \mathsf{d}(x, \bar x)(1 + \mathsf{d}(x, \bar x)) \le C_\delta(1 + \mathsf{d}^2(x, \bar x)) \quad \forall t \in [\delta, 1],$$

which means that φ_t^ε has quadratic growth. We then argue as before, coupling this information with (2.14a) and (2.8). □

proof of Lemma 3.4 Fix $x \in \mathrm{X}$, $M > 0$ and let V be defined as in the statement. By the maximum principle for the heat flow $\log \delta \le \phi_t^\delta \le \log(\|u\|_{L^\infty(\mathrm{X})} + \delta)$ for all $t \ge 0$, so that

$$\sup_{t \in [0,\infty)} \|\phi_t^\delta\|_{L^\infty(\mathrm{X})} < \infty \qquad \text{and} \qquad (\phi_t^\delta) \in L^\infty((0,\infty), L^2(\mathrm{X}, e^{-V}\mathfrak{m})).$$

Since log is smooth with bounded derivatives on $[\delta, +\infty)$, the chain and Leibniz rules entail that

$$|\nabla \phi_t^\delta| \le \frac{|\nabla \mathsf{h}_t u|}{\delta} \qquad |\Delta \phi_t^\delta| \le \frac{|\Delta \mathsf{h}_t u|}{\delta} + \frac{|\nabla \mathsf{h}_t u|^2}{\delta^2}$$

and, by the a priori estimates (2.3) and the Lipschitz regularization (2.5), (iii) follows. Furthermore, notice that (2.2) and the chain rule grant that $\mathfrak{m}$-a.e. (3.1) holds for a.e. t; since (iii) implies in particular that $(|\nabla \phi_t^\delta|), (\Delta \phi_t^\delta) \in L^2_{loc}((0,\infty), L^2(\mathrm{X}, e^{-V}\mathfrak{m}))$, this means that $(\phi_t^\delta) \in AC_{loc}((0,\infty), L^2(\mathrm{X}, e^{-V}\mathfrak{m}))$ and (3.1) actually holds when the left-hand side is intended as limit of the difference quotients in $L^2(\mathrm{X}, e^{-V}\mathfrak{m})$.

The continuity in $t = 0$ follows by dominated convergence from (i) and the fact that for any sequence $t_n \downarrow 0$ there exists a subsequence $t_{n_k} \downarrow 0$ such that $\mathsf{h}_{t_{n_k}} u \to u$ $\mathfrak{m}$-a.e.

Finally, given $\mathcal{C} \subset\subset (0,\infty)$, observe that from (i), (iii) and the fact that $\mu_t \le C\mathfrak{m}$ for all $t \ge 0$ we get

$$\sup_{t \in \mathcal{C}} \int |\phi_t^\delta| \eta_t \,\mathrm{d}\mathfrak{m} + \int |\phi_t^\delta|^2 \eta_t \,\mathrm{d}\mathfrak{m} + \int |\nabla \phi_t^\delta|^2 \eta_t \,\mathrm{d}\mathfrak{m} < \infty,$$

whence integrability on $\mathcal{C} \times \mathrm{X}$ by Fubini's theorem. For $|\nabla \phi_t^\delta| \eta_t$ it is sufficient to notice that $|\nabla \phi_t^\delta| \eta_t \le \frac{1}{2} |\nabla \phi_t^\delta|^2 \eta_t + \frac{1}{2} \eta_t$ and then argue as above. □

proof of Lemma 3.5 The absolute continuity of $t \mapsto \int f \,\mathrm{d}\mu_t$ and the bound (3.2) are trivial consequences of Definition 3.1. The fact that the exceptional set can be chosen independently of f follows from the separability of $W^{1,2}(\mathrm{Y})$ and standard approximation procedures, carried out, for instance, in [14].

This implies that the second derivative in the right hand side of (3.3) exists for a.e. t, so that the claim makes sense. The absolute continuity of $t \mapsto \int f_t \,\mathrm{d}\mu_t$ follows from the fact that for any $t_0, t_1 \in [0,1]$, $t_0 < t_1$ it holds

$$\begin{aligned}
\Big| \int f_{t_1} \,\mathrm{d}\mu_{t_1} - \int f_{t_0} \,\mathrm{d}\mu_{t_0} \Big| &\le \Big| \int f_{t_1} \,\mathrm{d}\mu_{t_1} - \int f_{t_1} \,\mathrm{d}\mu_{t_0} \Big| + \Big| \int \big(f_{t_1} - f_{t_0}\big) \mathrm{d}\mu_{t_0} \Big| \\
&\le \int_{t_0}^{t_1} \int \big(|\mathrm{d} f_{t_1}||X_t| + c|\Delta f_{t_1}|\big) \mathrm{d}\mu_t \mathrm{d}t + \iint_{t_0}^{t_1} \Big| \frac{\mathrm{d}}{\mathrm{d}t} f_t \Big| \,\mathrm{d}t \,\mathrm{d}\mu_{t_0}
\end{aligned}$$

and our assumptions on (f_t). Now fix a point t of differentiability for (f_t) and observe that the fact that $\frac{f_{t+h}-f_t}{h}$ strongly converges in $L^2(\mathrm{Y})$ to $\frac{\mathrm{d}}{\mathrm{d}t}f_t$ and μ_{t+h} weakly converges to μ_t as $h \to 0$ and the densities are equibounded is sufficient to get

$$\lim_{h\to 0}\int \frac{f_{t+h}-f_t}{h}\,\mathrm{d}\mu_{t+h} = \int \frac{\mathrm{d}}{\mathrm{d}t}f_t\,\mathrm{d}\mu_t = \lim_{h\to 0}\int \frac{f_{t+h}-f_t}{h}\,\mathrm{d}\mu_t.$$

Hence the conclusion comes dividing by h the trivial identity

$$\begin{aligned}\int f_{t+h}\,\mathrm{d}\mu_{t+h} - \int f_t\,\mathrm{d}\mu_t = \int f_t\,\mathrm{d}\mu_{t+h} - \int f_t\,\mathrm{d}\mu_t + \int (f_{t+h}-f_t)\mathrm{d}\mu_t + \\ + \int (f_{t+h}-f_t)\mathrm{d}\mu_{t+h} - \int (f_{t+h}-f_t)\mathrm{d}\mu_t\end{aligned}$$

and letting $h \to 0$.

The last statement is straightforward. □

References

[1] 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 ed., 2008.

[2] ——, *Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below*, Invent. Math., 195 (2014), pp. 289–391.

[3] ——, *Metric measure spaces with Riemannian Ricci curvature bounded from below*, Duke Math. J., 163 (2014), pp. 1405–1490.

[4] L. Ambrosio, A. Mondino, and G. Savaré, *On the Bakry-Émery condition, the gradient estimates and the Local-to-Global property of $RCD^*(K,N)$ metric measure spaces*, The Journal of Geometric Analysis, 26 (2014), pp. 1–33.

[5] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numerische Mathematik, 84 (2000), pp. 375–393.

[6] S. G. Bobkov, I. Gentil, and M. Ledoux, *Hypercontractivity of Hamilton-Jacobi equations*, Journal de Mathématiques Pures et Appliquées, 80 (2001), pp. 669 – 696.

[7] J. Cheeger, *Differentiability of Lipschitz functions on metric measure spaces*, Geom. Funct. Anal., 9 (1999), pp. 428–517.

[8] J. Cheeger and T. H. Colding, *On the structure of spaces with Ricci curvature bounded below. I*, J. Differential Geom., 46 (1997), pp. 406–480.

[9] ——, *On the structure of spaces with Ricci curvature bounded below. II*, J. Differential Geom., 54 (2000), pp. 13–35.

[10] ——, *On the structure of spaces with Ricci curvature bounded below. III*, J. Differential Geom., 54 (2000), pp. 37–74.

[11] Y. Chen, T. T. Georgiou, and M. Pavon, *On the Relation Between Optimal Transport and Schrödinger Bridges: A Stochastic Control Viewpoint*, Journal of Optimization Theory and Applications, 169 (2016), pp. 671–691.

[12] H. Föllmer, *Random fields and diffusion processes*, in École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87, vol. 1362 of Lecture Notes in Math., Springer, Berlin, 1988, pp. 101–203.

[13] I. Gentil, C. Léonard, and L. Ripani, *About the analogy between optimal transport and minimal entropy*, Ann. Fac. Sci. Toulouse Math. Série 6, 26 (2017), pp. 569–600.

[14] N. Gigli, *Nonsmooth differential geometry - an approach tailored for spaces with Ricci curvature bounded from below.* Accepted at Mem. Amer. Math. Soc., arXiv:1407.0809, 2014.

[15] ——, *On the differential structure of metric measure spaces and applications*, Mem. Amer. Math. Soc., 236 (2015), pp. vi+91.

[16] N. Gigli and B. Han, *The continuity equation on metric measure spaces*, Calc. Var. Partial Differential Equations, 53 (2013), pp. 149–177.

[17] N. Gigli, K. Kuwada, and S.-i. Ohta, *Heat flow on Alexandrov spaces*, Communications on Pure and Applied Mathematics, 66 (2013), pp. 307–331.

[18] N. Gigli and S. Mosconi, *The abstract Lewy-Stampacchia inequality and applications.* Preprint, arXiv:1401.4911, 2014.

[19] N. Gigli and L. Tamanini, *Second order differentiation formula on $RCD^*(K, N)$ spaces.* Accepted at JEMS, arXiv:1802.02463, (2018).

[20] N. Gozlan, C. Roberto, and P.-M. Samson, *Hamilton Jacobi equations on metric spaces and transport entropy inequalities.*, Revista matemática iberoamericana, 30 (2014), pp. 133–163.

[21] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali, *Kantorovich duality for general transport costs and applications*, Journal of Functional Analysis, 273 (2017), pp. 3327 – 3405.

[22] L. V. Kantorovich, *On the translocation of masses*, in Dokl. Akad. Nauk. USSR (NS), vol. 37, 1942, pp. 199–201.

[23] F. Léger, *A geometric perspective on regularized optimal transport*, Preprint, arXiv:1703.10243, (2017).

[24] C. Léonard, *From the Schrödinger problem to the Monge-Kantorovich problem*, J. Funct. Anal., 262 (2012), pp. 1879–1920.

[25] ——, *Some properties of path measures*, Séminaire de Probabilités XLVI, (2014), pp. 207–230.

[26] ——, *A survey of the Schrödinger problem and some of its connections with optimal transport*, Discrete Contin. Dyn. Syst., 34 (2014), pp. 1533–1574.

[27] T. Mikami, *Monge's problem with a quadratic cost by the zero-noise limit of h-path processes*, Probab. Theory Related Fields, 129 (2004), pp. 245–260.

[28] T. Mikami and M. Thieullen, *Duality theorem for the stochastic optimal control problem*, Stochastic Process. Appl., 116 (2006), pp. 1815–1835.

[29] ——, *Optimal transportation problem by stochastic optimal control*, SIAM J. Control Optim., 47 (2008), pp. 1127–1139.

[30] A. Mondino and A. Naber, *Structure Theory of Metric-Measure Spaces with Lower Ricci Curvature Bounds I.* Preprint, arXiv:1405.2222, 2014.

[31] M. Nagasawa, *Schrödinger equations and diffusion theory*, vol. 86 of Monographs in Mathematics, Birkhäuser Verlag, Basel, 1993.

[32] E. Nelson, *Dynamical theories of Brownian motion*, Princeton University Press, Princeton, N.J., 1967.

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

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

[35] L. Tamanini, *Analysis and geometry of RCD spaces via the Schrödinger problem*, PhD thesis, Université Paris Nanterre and SISSA, 2017.