

A formula for the time derivative of the entropic cost and applications

Giovanni Conforti ^{*} Luca Tamanini [†]

December 23, 2020

Abstract

In the recent years the Schrödinger problem has gained a lot of attention because of the connection, in the small-noise regime, with the Monge-Kantorovich optimal transport problem. Its optimal value, the *entropic cost* \mathcal{C}_T , is here deeply investigated. In this paper we study the regularity of \mathcal{C}_T with respect to the parameter T under a curvature condition and explicitly compute its first and second derivative. As applications:

- we determine the large-time limit of \mathcal{C}_T and provide sharp exponential convergence rates; we obtain this result not only for the classical Schrödinger problem but also for the recently introduced Mean Field Schrödinger problem [3];
- we improve the Taylor expansion of $T \mapsto T\mathcal{C}_T$ around $T = 0$ from the first to the second order.

Contents

1	Introduction and statement of the main results	2
1.1	The Schrödinger problem	2
1.2	First and second derivative of the entropic cost	6
1.3	Long-time behavior of entropic cost and energy	7
1.4	Short-time behavior of the entropic cost	9
1.5	Long time behavior of the mean field entropic cost	9
2	Preliminaries	11
3	Proof of the main results	17
3.1	A Γ -convergence proof of Theorem 1.2	17
3.2	Proof of Theorem 1.1 and Theorem 1.6	19
3.3	Proof of Theorem 1.2 and Corollary 1.3	26
3.4	Proof of Theorem 1.4	30
3.5	Proof of Theorem 1.7	34

^{*}Département de Mathématiques Appliquées, École Polytechnique, Route de Saclay, 91128 Palaiseau Cedex, France. email: giovanni.conforti@polytechnique.edu

[†]CEREMADE (UMR CNRS 7534), Université Paris Dauphine PSL, Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, France and INRIA-Paris, MOKAPLAN, 2 Rue Simone Iff, 75012, Paris, France. email: tamanini@ceremade.dauphine.fr

1 Introduction and statement of the main results

The entropic transportation cost is the optimal value in a probabilistic version of the Monge-Kantorovich optimal transport problem, the Schrödinger problem, whose study has already shown to have far reaching consequences in various fields, ranging from statistical machine learning to functional inequalities. The goal of the present article is to advance in the study of the entropic cost as a function of the time (regularization) parameter T (see Definition 1.2 below). Following Mikami's contribution [26] linking the Schrödinger problem to optimal transport, several results have been obtained in the last years concerning the behavior of the entropic cost in the short-time (small noise) limit $T \rightarrow 0$. In particular, as a byproduct of the research line originated in [1] another step forward was made with the computation of the first derivative of the rescaled entropic cost at $T = 0$, see also [12], the recent works [29, 10] and references therein. However, very few results beyond the short-time limit have been obtained. In particular, very little is known about the long-time regime $T \rightarrow \infty$, where the entropic cost is expected to converge to the sum of the marginal entropies. This lack of knowledge was one of the main motivations for our work and in this respect, our contribution includes:

- A formula for the first and second derivative of the (rescaled) entropic cost for a general value of T in terms of the so-called “energy” (defined at (1.6) below).
- A rigorous identification of the large-time limit of the entropic cost as the sum of the marginal entropies.
- Sharp exponential convergence rates under a curvature condition in the long-time regime. We obtain this result not only for the classical Schrödinger problem but also for the recently introduced Mean Field Schrödinger problem [3].

We also establish some results in the short-time limit. Their interest resides in the fact that they allow to obtain a clean expression of the second derivative of the rescaled entropic cost that, to the best of our knowledge, was not known before. However, the regularity assumptions we impose on the marginals are not the weakest ones. For this part, our contribution can be resumed as follows:

- A formula for the second derivative of the rescaled cost around $T = 0$ that yields the local convexity of the cost in the time variable.
- A non-asymptotic sharp quantitative bound for the convergence to the Wasserstein distance depending only on the integral of the Fisher information functional along the corresponding Wasserstein geodesic.

1.1 The Schrödinger problem

We consider a triplet (M, d_g, \mathbf{m}) , where M is a smooth, connected and complete Riemannian manifold without boundary and with metric tensor g , d_g is the induced distance and \mathbf{m} is given by $\mathbf{m}(dx) = \exp(-2U(x))\text{vol}(dx)$, where U is a C^2 potential satisfying the Bakry-Émery $\text{CD}(\kappa, \infty)$ condition, namely

$$U \in C^2(M; \mathbb{R}), \quad \text{Ric}_g + \text{Hess}(2U) \geq \kappa g \quad \text{for some } \kappa \in \mathbb{R}. \quad (\text{H1})$$

The measure \mathbf{m} is invariant for the SDE

$$dX_t = -\nabla U(X_t)dt + dB_t \quad (1.1)$$

(here B_t denotes a standard Brownian motion) whose joint law at time 0 and T will be denoted by $R_{0,T}$. Within this framework and given $\mu, \nu \in \mathcal{P}(M)$ (as usual, for a measurable space (E, \mathcal{E}) we denote $\mathcal{P}(E)$ the set of probability measures over (E, \mathcal{E})), the entropic transportation cost $\mathcal{C}_T(\mu, \nu)$ is defined as the optimal value in the corresponding Schrödinger problem, namely

$$\mathcal{C}_T(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \mathcal{H}(\pi | R_{0,T}), \quad (1.2)$$

where $\Pi(\mu, \nu) \subset \mathcal{P}(M \times M)$ is the set of couplings of μ and ν and \mathcal{H} is the relative entropy functional defined for two probability measures p, q on the same (arbitrary) measurable space as

$$\mathcal{H}(p|q) = \begin{cases} \int \log\left(\frac{dp}{dq}\right) dp, & \text{if } p \ll q \\ +\infty, & \text{otherwise.} \end{cases} \quad (1.3)$$

When q is not a probability, the precise definition is postponed to Section 2.

The physical meaning of the variational problem (1.2) is described in the seminal papers [30, 31], where E. Schrödinger addressed the problem of finding the most likely evolution of a system of independent random particles driven by (1.1) conditionally on the observation of their initial and final configuration. With this picture in mind, the short- and long-time behavior of the entropic cost sounds perfectly natural.

Assumptions on \mathbf{m} , μ and ν . We state here the additional assumptions we will make either on the reference measure \mathbf{m} or on the marginals μ, ν . Concerning the former, we will often assume that

$$\mathbf{m} \text{ is a probability measure.} \quad (\text{H2})$$

Notice that when $\kappa > 0$, (H2) automatically holds and in addition $\mathbf{m} \in \mathcal{P}_2(M)$ (see for instance [32, Theorem 4.26]). For the latter, we shall suppose that either

$$\mu, \nu \in \mathcal{P}_2(M) \text{ and } \mathcal{H}(\mu | \mathbf{m}), \mathcal{H}(\nu | \mathbf{m}) < +\infty, \quad (\text{H3})$$

where $\mathcal{P}_2(M) \subset \mathcal{P}(M)$ denotes the space of probability measures over M with finite second moment, or, more frequently, that

$$\mu = \rho \mathbf{m}, \nu = \sigma \mathbf{m} \text{ and } \rho, \sigma \in L^\infty(\mathbf{m}) \text{ with compact support.} \quad (\text{H4})$$

Notice that (H4) is stronger than (H3).

The Benamou-Brenier formulation. The fluid-dynamical formulation of the entropic cost [7, 15, 19, 34] asserts that, at least under (H4), we have

$$\begin{aligned} \mathcal{C}_T(\mu, \nu) &= \frac{1}{2} \left(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m}) \right) + \inf_{(\bar{\rho}, \bar{v})} \int \int_0^T \left(\frac{1}{2} |\bar{v}_t|^2 + \frac{1}{8} |\nabla \log \bar{\rho}_t|^2 \right) \rho_t dt d\mathbf{m} \\ &= \frac{1}{2} \left(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m}) \right) + \int \int_0^T \left(\frac{1}{2} |\bar{v}_t^T|^2 + \frac{1}{8} |\nabla \log \bar{\rho}_t^T|^2 \right) \bar{\rho}_t^T dt d\mathbf{m} \end{aligned} \quad (1.4)$$

where $(\bar{\rho}^T, \bar{v}^T)$ is the unique optimal curve and the infimum runs over all weak solutions $(\bar{\rho}, \bar{v})$ of the continuity equation

$$\partial_t(\bar{\rho}_t \mathbf{m}) + \operatorname{div}_{\mathbf{m}}(\bar{v}_t \bar{\rho}_t \mathbf{m}) = 0, \quad (1.5)$$

satisfying the marginal constraints $\rho_0 \mathbf{m} = \mu$ and $\rho_T \mathbf{m} = \nu$, namely among all $(\bar{\rho}_t) \subset L^\infty(\mathbf{m})$ with $\bar{\rho}_t \mathbf{m} \in \mathcal{P}(M)$ and all Borel vector fields (\bar{v}_t) such that:

- (i) $t \mapsto \bar{\rho}_t \mathbf{m}$ is weakly continuous and there exists $C > 0$ such that $\bar{\rho}_t \leq C$ for all $t \in [0, T]$;
- (ii) the map $t \mapsto \int |\bar{v}_t|^2 \bar{\rho}_t \mathbf{m}$ is Borel and belongs to $L^1(0, T)$;
- (iii) for any $f \in C_c^\infty(M)$ the map $[0, T] \ni t \mapsto \int f \bar{\rho}_t \mathbf{m}$ is absolutely continuous and it holds

$$\frac{d}{dt} \int f \bar{\rho}_t \mathbf{m} = \int df(\bar{v}_t) \bar{\rho}_t \mathbf{m} \quad \text{a.e. } t.$$

In the above, we denoted by $\operatorname{div}_{\mathbf{m}}$ the divergence w.r.t. \mathbf{m} and by df the differential of f .

From a physical point of view, (1.4) states that the trajectory $(\bar{\rho}_t^T)_{t \in [0, T]}$, also called *entropic interpolation*, is the one minimizing a functional consisting of two terms: the former is purely kinetic, while the latter is given by the Fisher information functional. Hence $(\bar{\rho}_t^T)_{t \in [0, T]}$ is given by the balance between deterministic and chaotic behavior.

The energy. The quantity

$$\frac{1}{2} \int |\bar{v}_t^T|^2 \bar{\rho}_t^T \mathbf{m} - \frac{1}{8} \int |\nabla \log \bar{\rho}_t^T|^2 \bar{\rho}_t^T \mathbf{m} \quad (1.6)$$

is conserved along the optimal flow $(\bar{\rho}_t^T)_{t \in [0, T]}$, i.e. its value does not depend on t , see [9, Corollary 1.1] and [17, Lemma 3.2] (although in these papers the setting is more restrictive, the argument therein carries over verbatim to the present setting). We call this constant the *energy* and denote it $\mathcal{E}_T(\mu, \nu)$. In order to justify the term “energy”, we remind that the formal Riemannian calculus on $\mathcal{P}_2(M)$ introduced by Otto [27] allows to draw an analogy between the minimization (1.4) and the much simpler problem (see [16]) on \mathbb{R}^d

$$\inf_{\substack{x: [0, T] \rightarrow \mathbb{R}^d \\ x_0 = x, x_T = y}} \int_0^T \frac{1}{2} |\dot{x}_t|^2 + \frac{1}{8} |\nabla f|^2(x_t) dt, \quad (1.7)$$

where the role of the function f in (1.7) is played by the relative entropy $\mathcal{H}(\cdot | \mathbf{m})$ in (1.4). For (1.7) it is well known that along any critical curve the total energy (given by the sum of the kinetic and potential components)

$$\frac{1}{2} |\dot{x}_t|^2 - \frac{1}{8} |\nabla f|^2(x_t)$$

is conserved. The analogous of this simple fact for the problem (1.4) is the conservation of (1.6) along optimal flows.

Short-time behavior of the entropic cost. The fact that the rescaled entropic cost converges to the squared Wasserstein distance of order two in the short-time limit, i.e.

$$\lim_{T \rightarrow 0} T \mathcal{C}_T(\mu, \nu) = \frac{1}{2} W_2^2(\mu, \nu), \quad (1.8)$$

has generated a surge of interest around the Schrödinger problem (henceforth SP), since SP is more regular and numerically easier to solve than the Monge-Kantorovich problem. There exist nowadays several proofs of (1.8), see for instance [23, 26] for Γ -convergence results. A proof that is valid under the hypotheses (H1) and (H4) can be found in [18, Remark 5.11]. In [1], (see also [11, 12, 29]) a further fundamental step was taken that consists in computing the first order term in the expansion of $T \mathcal{C}_T(\mu, \nu)$ around $T = 0$. Referring to the above mentioned articles for precise statements, we have the following expansion¹

$$T \mathcal{C}_T(\mu, \nu) = \frac{1}{2} W_2^2(\mu, \nu) + \frac{T}{2} (\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m})) + o(T). \quad (1.9)$$

In this work we establish at Theorem 1.6 a non-asymptotic bound for $T \mathcal{C}_T(\mu, \nu) - W_2^2(\mu, \nu)/2$ which is sharp in the limit $T \rightarrow 0$ and we compute the second order term in (1.9), thus getting

$$T \mathcal{C}_T(\mu, \nu) = \frac{1}{2} W_2^2(\mu, \nu) + \frac{T}{2} (\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m})) + \frac{T^2}{8} \iint_0^1 |\nabla \log \rho_t^0|^2 \rho_t^0 dt d\mathbf{m} + o(T^2), \quad (1.10)$$

where $(\rho_t^0)_{t \in [0,1]}$ denotes the (unique) Wasserstein geodesic between μ and ν . Let us remark that (1.10) tells that the rescaled cost is convex around $T = 0$ and using functional inequalities such as the HWI inequality [28] one can also estimate from below its second derivative under the Bakry-Émery condition (H1). It is an interesting question to obtain general versions of (1.10) in terms of Γ -convergence. In addition, it is worth noticing that the energy, once properly rescaled, also converges to the squared Wasserstein distance, namely

$$\lim_{T \rightarrow 0} T^2 \mathcal{E}_T(\mu, \nu) = \frac{1}{2} W_2^2(\mu, \nu).$$

To see, at least formally, why this is expected to be true, one can integrate (1.6) in time, use (1.4) and the energy conservation to get

$$T \mathcal{E}_T(\mu, \nu) = \mathcal{C}_T(\mu, \nu) - \frac{1}{2} (\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m})) - \frac{1}{4} \iint_0^T |\nabla \log \bar{\rho}_t^T|^2 \bar{\rho}_t^T dt d\mathbf{m}$$

Multiplying by T , letting $T \rightarrow 0$ and using (1.8) we can then handle the first term on the right-hand side as well as $\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m})$. The fact that the last integral converges to 0 has been proved in [17, Lemma 3.3] and the same argument can be adapted verbatim to the present setting (see [34]); yet it is a non-trivial fact, as the entropic interpolation $(\bar{\rho}_t^T)_{t \in [0,T]}$ converges to the Wasserstein geodesic $(\rho_t^0)_{t \in [0,T]}$ between μ and ν , whose Fisher information needs not to be defined.

¹In the above mentioned references, one often finds $\mathcal{H}(\mu | \mathbf{m}) - \mathcal{H}(\nu | \mathbf{m})$ instead of $\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m})$ in the first order term. This is due to a slightly different choice of reference measure $R_{0,T}$. Since we prefer the reference measure to be reversible, we gain an extra $\mathcal{H}(\mu | \mathbf{m})$ term in the Taylor expansion.

Long-time behavior of the entropic cost. Although its asymptotic regime has been the object of recent studies in connection with the ergodic behavior of Schrödinger bridges [3] and the limiting behavior of Sinkhorn divergence [13], very few results are available. In this article we prove that under mild assumptions we have

$$\lim_{T \rightarrow +\infty} \mathcal{C}_T(\mu, \nu) = \mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m}), \quad (1.11)$$

The intuition behind (1.11) is that $R_{0,T}$ converges towards $\mathbf{m} \otimes \mathbf{m}$ and therefore the optimal coupling in SP converges towards the independent coupling of μ and ν . We remark that in a different although related context [13], the convergence of Sinkhorn divergences towards MMD divergences in the limit when the regularization parameter goes to $+\infty$ shares many analogies with (1.11).

Our second main result deals with the approximation error in (1.11) and states that it is exponentially small (see Theorem 1.4 for the rigorous statement)

$$|\mathcal{C}_T(\mu, \nu) - \mathcal{H}(\mu | \mathbf{m}) - \mathcal{H}(\nu | \mathbf{m})| \lesssim \exp\left(-\frac{\kappa}{2}T\right) \quad (1.12)$$

provided the Bakry-Émery condition (H1) is satisfied. The proof of (1.12) is based on Theorem 1.1 asserting that the time derivative of $\mathcal{C}_T(\mu, \nu)$ is precisely $-\mathcal{E}_T(\mu, \nu)$ and on two functional inequalities. The first one is a version of the Talagrand inequality (obtained in [9] and also called entropic Talagrand inequality) and the second one is a functional inequality relating $|\mathcal{E}_T(\mu, \nu)|$ with $\mathcal{C}_T(\mu, \nu)$, that we call “energy-transport” inequality (cf. Lemma 3.8). A similar inequality has been proved very recently in [3]; there it has been used to obtain the so-called *turnpike property* for mean field Schrödinger bridges. It is worth noticing that estimates such as (1.12) do not seem to follow from classical functional inequalities such as Talagrand and Log-Sobolev, whereas they can be proven using the new family of functional inequalities involving the entropic cost $\mathcal{C}_T(\mu, \nu)$. The final contribution of the article is to derive a bound similar to (1.12) for the Mean Field Schrödinger problem introduced in [3]. Since we could not establish a generalization of the differentiation formula at Theorem 1.1 to the mean field setup, the proof of this estimate follows a different scheme, but is still based on a class of functional inequalities derived in [3] that are the mean field versions of the entropic Talagrand and of the energy-transport inequalities mentioned above.

Organization of the paper. The document is structured as follows: in the remainder of Section 1 we state and comment the main results, whose proofs are contained in Section 3. In Section 2 we collect, for reader’s sake, all relevant results and bibliographical references on SP. Finally, in Appendix A we prove the sharpness of a functional inequality introduced by the first-named author in [9], which plays an important role in this paper.

1.2 First and second derivative of the entropic cost

The regularity of the entropic cost w.r.t. to the time variable T has never been investigated to the best of our knowledge, so that the following is the first result of such a kind and it plays a pivotal role in the study of both the long- and short-time behavior of $\mathcal{C}_T(\mu, \nu)$.

Theorem 1.1. *Given (H1) and (H4) the following hold:*

(i) the map $T \mapsto \mathcal{C}_T(\mu, \nu)$ is $C^1((0, \infty))$, twice differentiable a.e. and the first derivative is given by

$$\frac{d}{dT} \mathcal{C}_T(\mu, \nu) = -\mathcal{E}_T(\mu, \nu), \quad \forall T > 0;$$

(ii) the map $T \mapsto T\mathcal{C}_T(\mu, \nu)$ is $C^1((0, \infty))$ and twice differentiable a.e. The first derivative is given for all $T > 0$ by

$$\frac{d}{dT} (T\mathcal{C}_T(\mu, \nu)) = \mathcal{C}_T(\mu, \nu) - T\mathcal{E}_T(\mu, \nu)$$

or equivalently by

$$\frac{d}{dT} (T\mathcal{C}_T(\mu, \nu)) = \frac{1}{2} \mathcal{H}(\mu | \mathbf{m}) + \frac{1}{2} \mathcal{H}(\nu | \mathbf{m}) + \frac{1}{4} \iint_0^T |\nabla \log \bar{\rho}_t^T|^2 \bar{\rho}_t^T dt d\mathbf{m},$$

where $(\bar{\rho}^T, \bar{v}^T)$ is the optimal solution in (1.4). The second derivative writes as

$$\frac{d^2}{dT^2} (T\mathcal{C}_T(\mu, \nu)) = \frac{1}{4} \frac{d}{dT} \iint_0^T |\nabla \log \bar{\rho}_t^T|^2 \bar{\rho}_t^T dt d\mathbf{m} = -2\mathcal{E}_T(\mu, \nu) - T \frac{d}{dT} \mathcal{E}_T(\mu, \nu)$$

for a.e. $T > 0$.

As implicitly stated above, by Theorem 1.1 we see that $T \mapsto \mathcal{E}_T(\mu, \nu)$ is continuous on $(0, \infty)$ and differentiable a.e.

1.3 Long-time behavior of entropic cost and energy

As a first step we show that

Theorem 1.2 (Long-time behavior of entropic cost and energy). *Under (H1) with $\kappa \geq 0$ and (H2), for any $\mu, \nu \in \mathcal{P}(M)$ satisfying (H3) it holds*

$$\lim_{T \rightarrow \infty} \mathcal{C}_T(\mu, \nu) = \mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m}).$$

If $\mu, \nu \in \mathcal{P}(M)$ satisfy (H4), then it also holds

$$\lim_{T \rightarrow \infty} \mathcal{E}_T(\mu, \nu) = 0.$$

As concerns the long-time behavior of $\mathcal{C}_T(\mu, \nu)$, we provide two different proofs:

- the former is direct and relies on a Γ -convergence argument (see Section 3.1);
- the latter is more technical and requires slightly stronger assumptions on μ and ν , but as an advantage it allows us to determine the long-time behavior of the so-called “ (f, g) -decomposition” of the optimal coupling in SP (see Section 2 for its definition, Section 3.3 for the proof).

As an application of this result we provide a new proof of the logarithmic Sobolev inequality based on entropic interpolations, in the same spirit of the recent paper [17], where an “entropic” proof of the HWI inequality is established.

Corollary 1.3 (Log-Sobolev inequality). *Under (H1) with $\kappa > 0$, for any $\mu = \rho \mathbf{m} \in \mathcal{P}(M)$ it holds*

$$\mathcal{H}(\mu | \mathbf{m}) \leq \frac{1}{2\kappa} \int |\nabla \log \rho|^2 d\mu, \quad (1.13)$$

where the right-hand side is set equal to $+\infty$ if $\log \rho$ is not locally Sobolev.

As a further step, in the following result we improve Theorem 1.2 by providing sharp rates of convergence for both $\mathcal{C}_T(\mu, \nu)$ and $\mathcal{E}_T(\mu, \nu)$. The key message is that, under a positive curvature condition, the approximation error is asymptotically smaller than $\exp(-\kappa T/2)$, up to constant factors depending on $\mathcal{H}(\mu | \mathbf{m})$ and $\mathcal{H}(\nu | \mathbf{m})$, and the rate $\exp(-\kappa T/2)$ is sharp. As already pointed out, recall that if (H1) holds with $\kappa > 0$, then (H2) also holds.

Theorem 1.4 (Long-time behavior of entropic cost and energy - rate of convergence). *Let us assume that (H1) with $\kappa > 0$ and (H4) are satisfied. Then for all $T > 0$ it holds*

$$|\mathcal{C}_T(\mu, \nu) - \mathcal{H}(\mu | \mathbf{m}) - \mathcal{H}(\nu | \mathbf{m})| \leq \frac{2}{\exp(\kappa T/2) - 1} \left(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m}) \right) \quad (1.14)$$

and

$$|\mathcal{E}_T(\mu, \nu)| \leq \frac{\kappa \exp(-\kappa T/2)}{(1 - \exp(-\kappa T/2))^2} \left(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m}) \right)^2. \quad (1.15)$$

Furthermore, the convergence rate $\exp(-\kappa T/2)$ in (1.14) and (1.15) is sharp in the following sense: for all μ, ν as in (H4) it holds

$$\lim_{T \rightarrow \infty} \frac{1}{T} \log |\mathcal{C}_T(\mu, \nu) - \mathcal{H}(\mu | \mathbf{m}) - \mathcal{H}(\nu | \mathbf{m})| \leq -\frac{\kappa}{2}, \quad \lim_{T \rightarrow \infty} \frac{1}{T} \log |\mathcal{E}_T(\mu, \nu)| \leq -\frac{\kappa}{2}$$

and there exists a triplet $(M', \mathbf{d}'_g, \mathbf{m}')$ satisfying (H1) with $\kappa > 0$ such that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \log |\mathcal{C}_T(\mu, \nu) - \mathcal{H}(\mu | \mathbf{m}) - \mathcal{H}(\nu | \mathbf{m})| \leq -\alpha\kappa, \quad \lim_{T \rightarrow \infty} \frac{1}{T} \log |\mathcal{E}_T(\mu, \nu)| \leq -\alpha\kappa$$

holds for all μ, ν satisfying (H4) if and only if $\alpha \leq 1/2$.

In other words, it may be possible to improve the constant factor in (1.14), but the convergence rate $\exp(-\kappa T/2)$ is asymptotically sharp.

Remark 1.5. Actually, in Section 3.4 we shall prove the following (stronger) bounds

$$|\mathcal{C}_T(\mu, \nu) - \mathcal{H}(\mu | \mathbf{m}) - \mathcal{H}(\nu | \mathbf{m})| \leq 2 \frac{\sqrt{\mathcal{H}(\mu | \mathbf{m})\mathcal{H}(\nu | \mathbf{m}) + \delta \exp(-\kappa T/2)}}{\exp(\kappa T/2) - 1}, \quad (1.16a)$$

$$|\mathcal{E}_T(\mu, \nu)| \leq \frac{\kappa \exp(-\kappa T/2)}{(1 - \exp(-\kappa T/2))^2} \left(\mathcal{H}(\mu | \mathbf{m})\mathcal{H}(\nu | \mathbf{m}) + \delta \exp(-\kappa T/2) \right), \quad (1.16b)$$

where

$$\delta = \mathcal{H}(\mu | \mathbf{m})^2 + \exp(-\kappa T/2)\mathcal{H}(\mu | \mathbf{m})\mathcal{H}(\nu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m})^2.$$

However we prefer the more compact, although slightly less precise, formulation given above. ■

1.4 Short-time behavior of the entropic cost

We provide a non-asymptotic bound for the difference $T\mathcal{C}_T(\mu, \nu) - W_2^2(\mu, \nu)/2$ under the assumption that the integral of the Fisher information along the displacement interpolation is finite and we also prove that this bound is sharp, as it coincides with the Taylor expansion of $T \mapsto T\mathcal{C}_T(\mu, \nu)$ around $T = 0$.² This result is of particular interest as the second order term in the expansion of the rescaled cost has not been computed before (to the best of our knowledge) and its form is general enough to formulate more general version of (1.18) below, for instance in terms of Γ -convergence, that deserve to be the object of future work.

Theorem 1.6 (Short-time behavior of the entropic cost). *Assume that (H1) and (H4) hold. Then for all $T > 0$ we have:*

$$0 \leq T\mathcal{C}_T(\mu, \nu) - \frac{1}{2}W_2^2(\mu, \nu) \leq \frac{T}{2}(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m})) + \frac{T^2}{8} \iint_0^1 |\nabla \log \rho_t^0|^2 \rho_t^0 dt d\mathbf{m}, \quad (1.17)$$

where $(\rho_t^0 \mathbf{m})_{t \in [0,1]}$ denotes the Wasserstein geodesic between μ and ν and $\int |\nabla \log \rho_t^0|^2 \rho_t^0 d\mathbf{m} := +\infty$ whenever ρ_t^0 is not Sobolev.

If in addition the Bakry-Émery $\text{CD}(\kappa, N)$ condition

$$\text{Ric}_{U,N} := \text{Ric}_g - (N - n) \frac{\text{Hess}(\exp(-\frac{2}{N-n}U))}{\exp(-\frac{2}{N-n}U)} \geq \kappa g$$

holds for some $N \geq n = \dim(M)$ and $\iint_0^1 |\nabla \log \rho_t^0|^2 \rho_t^0 dt d\mathbf{m} < \infty$, then $T \mapsto T\mathcal{C}_T(\mu, \nu)$ is twice differentiable at $T = 0$ and the following expansion holds

$$T\mathcal{C}_T(\mu, \nu) = \frac{1}{2}W_2^2(\mu, \nu) + \frac{T}{2}(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m})) + \frac{T^2}{8} \iint_0^1 |\nabla \log \rho_t^0|^2 \rho_t^0 dt d\mathbf{m} + o(T^2). \quad (1.18)$$

Note, for instance, that the $\text{CD}(\kappa, N)$ condition is satisfied when \mathbf{m} is the volume measure.

1.5 Long time behavior of the mean field entropic cost

In this section we move from the Riemannian to the Euclidean setting and turn our attention to the Mean Field Schrödinger problem (henceforth MFSP) recently introduced in [3]. Intuitively, this problem models the well-known thought experiment proposed by Schrödinger [30, 31] in the case when the underlying particles are not independent but interact through a pair potential W . An important difference with the classical case, is that MFSP cannot be reduced to a static problem. Therefore it is formulated for probability measures on the space of continuous paths $C([0, T]; \mathbb{R}^d)$, that we equip with the Borel σ -algebra generated by the norm $\|\cdot\|_\infty$. As usual, we also consider the canonical filtration on $C([0, T]; \mathbb{R}^d)$ and denote $(X_t)_{t \in [0, T]}$ the canonical process. MFSP is then stated as

$$\inf \left\{ \mathcal{H}(P | \Gamma(P)) : P \in \mathcal{P}_1(C([0, T]; \mathbb{R}^d)), P_0 = \mu, P_T = \nu \right\} \quad (\text{MFSP})$$

²added in proof: With respect to a previous version of the paper, we have been able to remove quite demanding regularity assumptions. A similar but less general result (valid only in the Euclidean setting) has recently been obtained in [8] by other means, hence our proof is independent.

where $\mathcal{P}_1(C([0, T]; \mathbb{R}^d))$ denotes the space of probability measures over $C([0, T]; \mathbb{R}^d)$ with finite first moment w.r.t. $\|\cdot\|_\infty$ and, for any $P \in \mathcal{P}_1(C([0, T]; \mathbb{R}^d))$, P_0 (resp. P_T) denotes its marginal law at time 0 (resp. T), whereas $\Gamma(P)$ is the law of the controlled SDE

$$\begin{cases} dZ_t &= -\nabla W * P_t(Z_t)dt + dB_t, \\ Z_0 &\sim \mu. \end{cases}$$

In the above we denoted by $*$ the usual convolution operator. The interaction potential W satisfies the assumptions

$$W \text{ is of class } C^2(\mathbb{R}^d; \mathbb{R}) \text{ and symmetric, i.e. } W(x) = W(-x) \text{ for all } x \in \mathbb{R}^d, \quad (\text{H}'1)$$

$$\sup_{z, v \in \mathbb{R}^d, |v|=1} v \cdot \nabla^2 W(z) \cdot v < +\infty.$$

The role of the functional $\mathcal{H}(\cdot | \mathbf{m})$ is now taken by the functional $\mathcal{F} : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$, which is obtained by shifting by a constant the functional $\tilde{\mathcal{F}}$ defined by

$$\tilde{\mathcal{F}}(\sigma) = \begin{cases} \int_{\mathbb{R}^d} \log\left(\frac{d\sigma}{d\lambda}\right) d\sigma + \int_{\mathbb{R}^d} W * \frac{d\sigma}{d\lambda} d\sigma, & \text{if } \sigma \ll \lambda \\ +\infty, & \text{otherwise} \end{cases} \quad (1.19)$$

where λ denotes the Lebesgue measure on \mathbb{R}^d , in such a way that

$$\inf_{\sigma \in \mathcal{P}_2(\mathbb{R}^d) : \int x \sigma(dx) = \int x \mu(dx)} \mathcal{F}(\sigma) = 0.$$

In [3] the mean field entropic cost is defined as the optimal value in (MFSP). However, if $W = 0$ this definition does not give back the usual entropic cost \mathcal{C}_T and the reason is simply the following: in the large deviations formulation of (1.2) particles are sampled with initial distribution the invariant measure \mathbf{m} , whereas in MFSP particles are sampled according to μ . For this reason, and to strengthen the analogy with (1.2), we prefer to define the mean field entropic cost \mathcal{C}_T^{mf} as

$$\mathcal{C}_T^{mf}(\mu, \nu) := \mathcal{F}(\mu) + \mathcal{H}(P^* | \Gamma(P^*)), \quad P^* \text{ optimal in (MFSP)}. \quad (1.20)$$

With this definition we recover $\mathcal{C}_T(\mu, \nu)$ defined in (1.2) when $W = 0$.

After this premise, the following plays the same role of assumption (H1)

$$\exists \kappa > 0 \quad \text{s.t.} \quad \forall z \in \mathbb{R}^d, \quad 2\nabla^2 W(z) \geq \kappa \mathbb{I}_{d \times d}, \quad (\text{H}'2)$$

while regarding the marginal constraints μ and ν , we assume that they have finite second moment, they belong to the domain of \mathcal{F} and have the same barycenter, i.e.

$$\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d), \quad \mathcal{F}(\mu), \mathcal{F}(\nu) < +\infty, \quad \int_{\mathbb{R}^d} x \mu(dx) = \int_{\mathbb{R}^d} x \nu(dx). \quad (\text{H}'3)$$

The proof we gave of Theorem 1.1 is hard to replicate for MFSP essentially because uniqueness of optimizers is currently not known. However, the long-time behavior of the mean field entropic cost can still be studied and exponential rate of convergence can be derived as well.

Theorem 1.7. *Assume (H'1)-(H'3) and that $\mathcal{F}(\mu), \mathcal{F}(\nu) < +\infty$. Then:*

(i) it holds

$$\lim_{T \rightarrow \infty} \mathcal{C}_T^{mf}(\mu, \nu) = \mathcal{F}(\mu) + \mathcal{F}(\nu);$$

(ii) the rate of convergence is at least $\exp(-\kappa T)$, i.e. there exists a decreasing function $B(\cdot)$ such that

$$|\mathcal{C}_T^{mf}(\mu, \nu) - \mathcal{F}(\mu) - \mathcal{F}(\nu)| \leq B(\kappa)(\mathcal{F}(\mu) + \mathcal{F}(\nu)) \exp(-\kappa T/2) \quad (1.21)$$

uniformly in $T \geq 1$.

Remark 1.8. One could be more precise in the above statement and get that

$$\begin{aligned} (\mathcal{F}(\mu) + \mathcal{F}(\nu)) \frac{\sinh(\kappa T)}{\sinh(\kappa T/2)} + 2 \frac{\exp(\kappa T/2)}{(\exp(\kappa T/2) - 1)^2} \left(\mathcal{F}(\mu)\mathcal{F}(\nu) + \delta \exp(-\kappa T/2) \right)^{1/2} \\ \leq \mathcal{C}_T^{mf}(\mu, \nu) - \mathcal{F}(\mu) + \mathcal{F}(\nu) \leq \frac{1}{\exp(\kappa T/2) - 1} (\mathcal{F}(\mu) + \mathcal{F}(\nu)) \end{aligned}$$

where $\delta = \mathcal{F}(\mu)^2 + \mathcal{F}(\nu)^2 + \exp(-\kappa T/2)\mathcal{F}(\mu)\mathcal{F}(\nu)$. For sake of clarity, we prefer the more compact, although slightly less precise, formulation given above. \blacksquare

Remark 1.9. It is important to stress that in [3] the interaction potential W satisfies $\nabla^2 W(z) \geq \kappa \mathbb{I}_{d \times d}$ instead of (H'2). We prefer to work with the latter condition rather than with the former because in such a way the convergence rate (1.21) is consistent with (1.14). As a flip side, when addressed to [3] the reader has to keep in mind this difference. \blacksquare

2 Preliminaries

In this section we collect some useful results concerning Markov semigroups and Schrödinger problem, either already present in the literature or extended to our framework.

Markov semigroups and heat kernels. It is well known that there is a Markov semigroup P_t associated to the SDE (1.1), which has $L = \Delta/2 - \nabla U \cdot \nabla$ as generator and $\mathbf{m}(dx) = \exp(-2U(x))\text{vol}(dx)$ as invariant measure. We address the reader to [5] for a comprehensive discussion on the topic. However, it is worth stressing that in [5] a different convention is adopted and the generator of P_t is $2L = \Delta - 2\nabla U \cdot \nabla$; as a consequence, there exists a discrepancy in the factors appearing in the estimates (2.1), (2.3), (2.4), (2.5), (2.7), (2.8) below and the corresponding estimates in [5].

With this remark in mind, let us present all the useful information about P_t . First of all, it enjoys the following standard a priori estimate

$$\|LP_t f\|_{L^2(\mathbf{m})} \leq \frac{2}{t^2} \|f\|_{L^2(\mathbf{m})}, \quad \forall t > 0, \forall f \in L^2(\mathbf{m}), \quad (2.1)$$

which can be obtained by differentiating $t \mapsto \|\nabla P_t f\|_{L^2(\mathbf{m})}^2$. The semigroup is also ergodic. This means that:

(i) if $\mathbf{m}(M) = 1$, then for all $f \in L^2(\mathbf{m})$ it holds

$$\lim_{t \rightarrow \infty} P_t f = \int f d\mathbf{m}, \quad \text{in } L^2(\mathbf{m}), \quad (2.2)$$

(ii) if $\mathfrak{m}(M) = \infty$, then $\mathbf{P}_t f \rightarrow 0$ in $L^2(\mathfrak{m})$ for all $f \in L^2(\mathfrak{m})$.

The curvature assumption (H1) then yields many important consequences, the first of which is the Bakry-Émery commutation estimate

$$|\nabla \mathbf{P}_t u|^2 \leq e^{-\kappa t} \mathbf{P}_t (|\nabla u|^2), \quad \forall u \in C_c^\infty(M), \forall t \geq 0. \quad (2.3)$$

For its proof as well as for all the regularizing properties of \mathbf{P}_t that will be used throughout the paper, we address once more the reader to [5]. Secondly, \mathbf{P}_t enjoys an L^∞ -Lipschitz regularization (see [4]), namely for all $u \in L^\infty(\mathfrak{m})$ it holds

$$\|\nabla \mathbf{P}_t u\|_{L^\infty(\mathfrak{m})} \leq \frac{1}{\sqrt{I_\kappa(t)}} \|u\|_{L^\infty(\mathfrak{m})} \quad \text{with} \quad I_\kappa(t) := \frac{\exp(\kappa t) - 1}{\kappa}, \quad \forall t > 0. \quad (2.4)$$

Then let us recall that under (H1) Hamilton's gradient estimate is satisfied (see [22]): for any positive function $u \in L^p \cap L^\infty(\mathfrak{m})$ for some $p \in [1, \infty)$ it holds

$$t |\nabla \log \mathbf{P}_t u|^2 \leq 2(1 + \kappa^- t) \log \left(\frac{\|u\|_{L^\infty(\mathfrak{m})}}{\mathbf{P}_t u} \right), \quad \forall t > 0 \quad (2.5)$$

pointwise, where $\kappa^- := \max\{0, -\kappa\}$. Within our framework it is also well known (see for instance [21]) that there exists a unique kernel (transition probability) $\mathfrak{p}_t(x, y)$ representing \mathbf{P}_t in the following sense:

$$\mathbf{P}_t u(x) = \int u(y) \mathfrak{p}_t(x, y) \mathfrak{m}(dy), \quad \forall u \in L^\infty(\mathfrak{m}). \quad (2.6)$$

The function \mathfrak{p}_t can also be seen as the density of $R_{0,T}$ (the joint law at time 0 and T of the solution to (1.1)) w.r.t. $\mathfrak{m} \otimes \mathfrak{m}$; it is a smooth function on $(0, \infty) \times M \times M$ and the second-named author recently proved in [33] that in great generality (and in particular under the assumption (H1)) the following upper Gaussian estimate for the kernel holds

$$\mathfrak{p}_t(x, y) \leq \frac{1}{\sqrt{\mathfrak{m}(B_{\sqrt{t/2}}(x)) \mathfrak{m}(B_{\sqrt{t/2}}(y))}} \exp \left(C_\varepsilon (1 + C_\kappa t) - \frac{\mathfrak{d}^2(x, y)}{(2 + \varepsilon)t} \right) \quad (2.7)$$

for all $t > 0$, $x, y \in M$ and $\varepsilon > 0$, where C_κ can be chosen equal to 0 if $\kappa \geq 0$ in (H1). As concerns Gaussian lower bounds, if (H2) holds then by [35, Corollary 1.3] the following is satisfied

$$\mathfrak{p}_t(x, y) \geq \exp \left(- \frac{\kappa \mathfrak{d}^2(x, y)}{2(e^{\kappa t/2} - 1)} \right) \quad (2.8)$$

for all $t > 0$ and $x, y \in M$.

The relative entropy functional. Let us first recall the definition of the relative entropy functional in the case of a reference measure with possibly infinite mass (see [24] for more details). Given a σ -finite measure q on M , there exists a measurable function $W : M \rightarrow [0, \infty)$ such that

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

Introducing the probability measure $q_W := z_W^{-1} e^{-W} q$, for any $p \in \mathcal{P}(M)$ such that $\int W dp < +\infty$ the Boltzmann-Shannon entropy is defined as

$$\mathcal{H}(p|q) := \mathcal{H}(p|q_W) - \int W dp - \log z_W$$

where $\mathcal{H}(p|q_W)$ is in turn defined as in (1.3). Notice that Jensen's inequality and the fact that $q_W \in \mathcal{P}(M)$ grant that $\mathcal{H}(p|q_W)$ is well defined and non-negative, in particular the definition makes sense. The definition is also meaningful, because if $\int W' dp < +\infty$ for another function W' such that $z_{W'} < +\infty$, then

$$\mathcal{H}(p|q_W) - \int W dp - \log z_W = \mathcal{H}(p|q_{W'}) - \int W' dp - \log z_{W'}.$$

Hence $\mathcal{H}(p|q)$ is well defined for all $p \in \mathcal{P}(M)$ such that $\int W dp < +\infty$ for some non-negative measurable function W with $z_W < +\infty$. In particular, it is well known that under (H1) W can be chosen as $Cd^2(\bar{x}, \cdot)$ for any $\bar{x} \in M$ and $C > 0$ sufficiently large (see for instance [32, Theorem 4.26]).

The (f, g) -decomposition. It is important to stress that solving (1.2) is equivalent to finding non-negative Borel functions f^T, g^T , also called decomposition, such that

$$\mu = f^T \mathbf{P}_T g^T \mathbf{m}, \quad \nu = g^T \mathbf{P}_T f^T \mathbf{m}. \quad (2.9)$$

This pair of equations is known as Schrödinger system and its solvability holds under very mild assumptions (see [25]), in particular under:

- (H4), as a consequence of [18, Proposition 2.1], the smoothness and positivity of the heat kernel and the boundedness of $\text{supp}(\mu), \text{supp}(\nu)$;
- (H1), (H2) and (H3), because of [25, Proposition 2.5] and (2.8).

Still under (H1)-(H4), the couple (f^T, g^T) solving (2.9) is unique up to the trivial transformation $(f^T, g^T) \mapsto (cf^T, g^T/c)$ with $c > 0$, as proven for instance in [18, Proposition 2.1]: this fact will play an important role in several proofs (e.g. the extension of (2.20) to our setting and Lemma 3.4). A good feature of f^T, g^T is that they inherit the regularity (smoothness and integrability) of ρ, σ , the densities of μ, ν respectively. More precisely,

- (a) $f^T, g^T \in L^\infty(\mathbf{m})$ since so are ρ, σ ;
- (b) if $\rho \in C^k(M)$ for some $k \in \mathbb{N} \cup \{\infty\}$ (resp. σ), then f^T (resp. g^T) also belongs to $C^k(M)$.

A proof of this property can be found in [17, Proposition 2.7].

About point (a) a more quantitative statement is actually possible. In [18, Proposition 2.1] the second-named author proved that for all μ, ν as in (H4), the following integral bounds hold for the decomposition (f^T, g^T)

$$\|f^T\|_{L^\infty(\mathbf{m})} \|g^T\|_{L^1(\mathbf{m})} \leq \frac{\|\rho\|_{L^\infty(\mathbf{m})}}{c_T}, \quad \|g^T\|_{L^\infty(\mathbf{m})} \|f^T\|_{L^1(\mathbf{m})} \leq \frac{\|\sigma\|_{L^\infty(\mathbf{m})}}{c_T},$$

where c_T is a suitable positive constant. If we normalize g^T in such a way that

$$\|g^T\|_{L^1(\mathbf{m})} = 1, \quad \forall T > 0, \quad (2.10)$$

and from now on this choice will always be done, then the bounds above become

$$\|f^T\|_{L^\infty(\mathfrak{m})} \leq \frac{\|\rho\|_{L^\infty(\mathfrak{m})}}{c_T}, \quad \|g^T\|_{L^\infty(\mathfrak{m})}\|f^T\|_{L^1(\mathfrak{m})} \leq \frac{\|\sigma\|_{L^\infty(\mathfrak{m})}}{c_T}$$

Aim of the next lemma is to improve this result by showing that the same kind of bounds holds when T ranges in a compact subset of $(0, \infty)$ or even closed half-line if we also assume (H2).

Lemma 2.1. *Given (H1) and μ, ν as in (H4), the following hold:*

(i) *for all $0 < T_0 < T_1 < \infty$ there exists $c_{T_0, T_1} > 0$ such that*

$$\|f^T\|_{L^\infty(\mathfrak{m})} \leq \frac{\|\rho\|_{L^\infty(\mathfrak{m})}}{c_{T_0, T_1}}, \quad \|g^T\|_{L^\infty(\mathfrak{m})}\|f^T\|_{L^1(\mathfrak{m})} \leq \frac{\|\sigma\|_{L^\infty(\mathfrak{m})}}{c_{T_0, T_1}}, \quad \forall T \in [T_0, T_1]. \quad (2.11)$$

(ii) *if (H2) is satisfied, then for all $T_0 > 0$ there exists $c_{T_0} > 0$ such that*

$$\|f^T\|_{L^\infty(\mathfrak{m})} \leq \frac{\|\rho\|_{L^\infty(\mathfrak{m})}}{c_{T_0}}, \quad \|g^T\|_{L^\infty(\mathfrak{m})}\|f^T\|_{L^1(\mathfrak{m})} \leq \frac{\|\sigma\|_{L^\infty(\mathfrak{m})}}{c_{T_0}}, \quad \forall T \geq T_0. \quad (2.12)$$

Proof. The first equation in the Schrödinger system (2.9) and the representation formula (2.6) entail

$$f^T(x) \int g^T(y) \mathfrak{p}_T(x, y) d\mathfrak{m}(y) = \rho(x), \quad \text{for } \mathfrak{m}\text{-a.e. } x.$$

As $\mathfrak{p}_T(x, y)$ is smooth in $T \in (0, \infty)$ and $x, y \in M$ and f^T, g^T are uniformly compactly supported (since they are supported in $\text{supp}(\mu)$ and $\text{supp}(\nu)$ respectively) we deduce that there exists $c_{T_0, T_1} > 0$ such that $\mathfrak{p}_T(x, y) \geq c_{T_0, T_1}$ in $\text{supp}(\mu) \times \text{supp}(\nu)$ for all $T \in [T_0, T_1]$, whence

$$f^T = f^T \|g^T\|_{L^1(\mathfrak{m})} \leq \frac{\|\rho\|_{L^\infty(\mathfrak{m})}}{c_{T_0, T_1}}, \quad \mathfrak{m}\text{-a.e. in } \text{supp}(\mu)$$

provided $T \in [T_0, T_1]$, and this proves the first inequality in (2.11). For the second one it is sufficient to swap the roles of f^T and g^T . If we further assume (H2), then the Gaussian lower bound (2.8) holds. Hence for all $T_0 > 0$ there exists $c_{T_0} > 0$ such that $\mathfrak{p}_T(x, y) \geq c_{T_0}$ in $\text{supp}(\mu) \times \text{supp}(\nu)$ for all $T \geq T_0$, whence by the same argument as above the first inequality in (2.12) follows. By interchanging the roles of f^T and g^T , the same conclusion follows for g^T . \square

The Benamou-Brenier formulation. The equivalence between (2.9) and (1.2) is extremely fruitful, as it enables to fully describe the optimal pair $(\bar{\rho}^T, \bar{v}^T)$ in the fluid-dynamical description (1.4) of the entropic cost. Indeed,

$$\bar{\rho}_t^T = \mathbf{P}_t f^T \mathbf{P}_{T-t} g^T, \quad \bar{v}_t^T = \frac{1}{2} \nabla \log \mathbf{P}_{T-t} g^T - \frac{1}{2} \nabla \log \mathbf{P}_t f^T, \quad t \in [0, T].$$

Actually, (1.4) and the identities above are nothing but a reparametrization of the Benamou-Brenier-like formula for the entropic cost established in [19, Theorem 4.2] and [34], which reads as

$$T \mathcal{E}_T(\mu, \nu) = \frac{T}{2} \left(\mathcal{H}(\mu | \mathfrak{m}) + \mathcal{H}(\nu | \mathfrak{m}) \right) + \inf_{(\rho, v)} \int \int_0^1 \left(\frac{1}{2} |v_t|^2 + \frac{T^2}{8} |\nabla \log \rho_t|^2 \right) \rho_t dt d\mathfrak{m}, \quad (2.13)$$

and of

$$\rho_t^T := \mathbf{P}_{Tt} f^T \mathbf{P}_{T(1-t)} g^T, \quad v_t^T := \frac{T}{2} \nabla \log \mathbf{P}_{T(1-t)} g^T - \frac{T}{2} \nabla \log \mathbf{P}_{Tt} f^T, \quad t \in [0, 1].$$

In other words

$$\rho_t^T = \tilde{\rho}_{Tt}^T \quad \text{and} \quad v_t^T = T \tilde{v}_{Tt}^T. \quad (2.14)$$

In (2.13) the infimum runs over all weak solutions of the continuity equation (1.5) satisfying the marginal constraints $\rho_0 \mathbf{m} = \mu$ and $\rho_1 \mathbf{m} = \nu$, while (ρ^T, v^T) defined above is the unique optimal density-velocity couple for (2.13). It is also useful to see the second term in the right-hand side of (2.13) as an action functional, whose arguments are the curves (ρ_t) and (v_t) , i.e.

$$\mathcal{A}_T(\rho, v) := \int \int_0^1 \left(\frac{1}{2} |v_t|^2 + \frac{T^2}{8} |\nabla \log \rho_t|^2 \right) \rho_t dt d\mathbf{m}.$$

When $T = 0$, \mathcal{A}_T is a purely kinetic energy and by the well-known Benamou-Brenier formulation of the optimal transport problem [6],

$$\frac{1}{2} W_2^2(\mu, \nu) = \inf_{(\rho, v)} \mathcal{A}_0(\rho, v), \quad (2.15)$$

where the infimum is taken over the same set as in (2.13). In this case optimal couples shall be denoted by (ρ^0, v^0) . Let us also recall that

$$\begin{aligned} \mathcal{E}_T(\mu, \nu) &= \mathcal{H}(\mu | \mathbf{m}) + \frac{T}{2} \int \int_0^1 |\nabla \log \mathbf{P}_{T(1-t)} g^T|^2 \rho_t^T dt d\mathbf{m} \\ &= \mathcal{H}(\nu | \mathbf{m}) + \frac{T}{2} \int \int_0^1 |\nabla \log \mathbf{P}_{Tt} f^T|^2 \rho_t^T dt d\mathbf{m}, \end{aligned} \quad (2.16)$$

which trivially implies

$$\int \int_0^1 |\nabla \log \mathbf{P}_{Tt} f^T|^2 \rho_t^T dt d\mathbf{m}, \int \int_0^1 |\nabla \log \mathbf{P}_{T(1-t)} g^T|^2 \rho_t^T dt d\mathbf{m} < \infty \quad (2.17)$$

Finally, the expression of the conserved quantity $\mathcal{E}_T(\mu, \nu)$ in terms of the rescaled optimal couple (ρ^T, v^T) reads as

$$\begin{aligned} \mathcal{E}_T(\mu, \nu) &= \frac{1}{2T^2} \int |v_t^T|^2 \rho_t^T d\mathbf{m} - \frac{1}{8} \int |\nabla \log \rho_t^T|^2 \rho_t^T d\mathbf{m} \\ &= -\frac{1}{2} \int \nabla \log \mathbf{P}_{T(1-t)} g^T \cdot \nabla \log \mathbf{P}_{Tt} f^T \rho_t^T d\mathbf{m}. \end{aligned} \quad (2.18)$$

Dual formulation. For future reference, it is also worth mentioning the fact that, for all μ, ν satisfying (H4), the (rescaled) entropic cost admits the following dual representation (see [19, 34])

$$T \mathcal{C}_T(\mu, \nu) = T \mathcal{H}(\mu | \mathbf{m}) + \sup_{\phi \in C_b(M)} \left\{ \int Q_1^T \phi d\mu - \int \phi d\nu \right\}, \quad (2.19)$$

where $Q_1^T \phi := -T \log \mathbf{P}_T(\exp(-\phi/T))$.

Geometric information. As regards the relationship between Schrödinger problem and lower Ricci bounds, let us recall that in [9] the first-named author showed that (H1) implies the following distorted κ -convexity inequality of the entropy along entropic interpolations:

$$\begin{aligned} \mathcal{H}(\mu_t^T | \mathbf{m}) &\leq \frac{1 - \exp(-\kappa T(1-t))}{1 - \exp(-\kappa T)} \mathcal{H}(\mu | \mathbf{m}) + \frac{1 - \exp(-\kappa T t)}{1 - \exp(-\kappa T)} \mathcal{H}(\nu | \mathbf{m}) \\ &\quad - \frac{\cosh(\kappa T/2) - \cosh(\kappa T(t-1/2))}{\sinh(\kappa T/2)} \mathcal{C}_T(\mu, \nu), \end{aligned} \quad (2.20)$$

for all $t \in [0, 1]$, where $\mu_t^T := \rho_t^T \mathbf{m}$. Truth to be told, the result in [9] was stated in the framework of compact Riemannian manifolds satisfying (H1) with $\kappa > 0$ and endowed with the volume measure, but the proof can be adapted to our setting when (H2) holds, by relying on [17, Lemma 3.7]. Indeed, if we set

$$\begin{aligned} \rho_t^{T,\delta} &:= c_\delta (\mathbf{P}_{Tt} f^T + \delta) (\mathbf{P}_{T(1-t)} g^T + \delta), & \mu_t^{T,\delta} &:= \rho_t^{T,\delta} \mathbf{m}, \\ v_t^{T,\delta} &:= \frac{T}{2} \nabla \log(\mathbf{P}_{T(1-t)} g^T + \delta) - \frac{T}{2} \nabla \log(\mathbf{P}_{Tt} f^T + \delta), \end{aligned} \quad (2.21)$$

where c_δ is a normalization constant so that $\rho_t^{T,\delta}$ is still a probability density, then the identities

$$\begin{aligned} \partial_t \log(\mathbf{P}_{Tt} f^T + \delta) &= \frac{T}{2} |\nabla \log(\mathbf{P}_{Tt} f^T + \delta)|^2 + T \mathbf{L} \log(\mathbf{P}_{Tt} f^T + \delta) \\ \partial_t \log(\mathbf{P}_{T(1-t)} g^T + \delta) &= \frac{T}{2} |\nabla \log(\mathbf{P}_{T(1-t)} g^T + \delta)|^2 + T \mathbf{L} \log(\mathbf{P}_{T(1-t)} g^T + \delta) \\ \partial_t \rho_t^{T,\delta} + \operatorname{div}_{\mathbf{m}}(v_t^{T,\delta} \rho_t^{T,\delta}) &= 0 \end{aligned}$$

hold both in the classical sense and as strong $W^{1,2}$ -limits, where $\mathbf{L} := \Delta/2 - \nabla U \cdot \nabla$ is the generator of (\mathbf{P}_t) . Together with (2.4) and (2.5), this is sufficient to follow the lines of [9, Lemma 3.6, Lemma 3.7] and [17, Lemma 3.7] and deduce the following

Lemma 2.2. *Under (H1), given μ, ν as in (H4) and with the same notations as in (2.21), for all $\delta > 0$ and $t \in [0, 1]$ define*

$$h_f(t) := \int \log(\mathbf{P}_{Tt} f^T + \delta) d\mu_t^{T,\delta}, \quad h_b(t) := \int \log(\mathbf{P}_{T(1-t)} g^T + \delta) d\mu_t^{T,\delta}. \quad (2.22)$$

Then $h_f \in C([0, 1]) \cap C^2((0, 1))$, $h_b \in C([0, 1]) \cap C^2((0, 1))$,

$$h'_f(t) = -\frac{T}{2} \int |\nabla \log(\mathbf{P}_{Tt} f^T + \delta)|^2 d\mu_t^{T,\delta}, \quad h'_b(t) = \frac{T}{2} \int |\nabla \log(\mathbf{P}_{T(1-t)} g^T + \delta)|^2 d\mu_t^{T,\delta} \quad (2.23)$$

and

$$h''_f(t) \geq -\kappa T h'_f(t), \quad h''_b(t) \geq \kappa T h'_b(t).$$

By [9, Lemma 4.1] and following the proof of Theorem 1.4 therein, we obtain

$$\begin{aligned} \mathcal{H}(\mu_t^{T,\delta} | \mathbf{m}) &\leq \frac{1 - \exp(-\kappa T(1-t))}{1 - \exp(-\kappa T)} \mathcal{H}(\mu_0^{T,\delta} | \mathbf{m}) + \frac{1 - \exp(-\kappa T t)}{1 - \exp(-\kappa T)} \mathcal{H}(\mu_1^{T,\delta} | \mathbf{m}) \\ &\quad - \frac{\cosh(\kappa T/2) - \cosh(\kappa T(t-1/2))}{\sinh(\kappa T/2)} \mathcal{C}_T(\mu_0^{T,\delta}, \mu_1^{T,\delta}). \end{aligned}$$

It is now sufficient to pass to the limit as $\delta \downarrow 0$: $\mathcal{H}(\mu_t^{T,\delta} | \mathbf{m})$ converges to $\mathcal{H}(\mu_t^T | \mathbf{m})$, $\mathcal{H}(\mu_0^{T,\delta} | \mathbf{m})$ to $\mathcal{H}(\mu | \mathbf{m})$ and $\mathcal{H}(\mu_1^{T,\delta} | \mathbf{m})$ to $\mathcal{H}(\nu | \mathbf{m})$ by dominated convergence theorem; as concerns $\mathcal{C}_T(\mu_0^{T,\delta}, \mu_1^{T,\delta})$, this converges to $\mathcal{C}_T(\mu, \nu)$ for the following reason: $\tilde{f}^T := \sqrt{c_\delta}(f^T + \delta)$ and $\tilde{g}^T := \sqrt{c_\delta}(g^T + \delta)$ solve (2.9) with $\mu_0^{T,\delta}, \mu_1^{T,\delta}$ as marginal constraints, so that

$$\mathcal{C}_T(\mu_0^{T,\delta}, \mu_1^{T,\delta}) = \mathcal{H}(\tilde{f}^T \otimes \tilde{g}^T R_{0,T} | R_{0,T})$$

and again by dominated convergence the right-hand side above converges to

$$\mathcal{H}(f^T \otimes g^T R_{0,T} | R_{0,T}) = \mathcal{C}_T(\mu, \nu).$$

Let us stress that the fact that \tilde{f}^T and \tilde{g}^T solve (2.9) with $\mu_0^{T,\delta}, \mu_1^{T,\delta}$ as marginal constraints and the equivalence between SP and (2.9) immediately imply that $(\rho^{T,\delta}, v^{T,\delta})$ is optimal in (2.13).

3 Proof of the main results

3.1 A Γ -convergence proof of Theorem 1.2

A rather straightforward proof of the long-time behavior of $\mathcal{C}_T(\mu, \nu)$ can be obtained by a Γ -convergence argument. Its essence is contained in the following

Lemma 3.1. *Under assumption (H1) with $\kappa \geq 0$ and (H2), assume that $\mu, \nu \in \mathcal{P}_2(M)$ with $\mathcal{H}(\mu | \mathbf{m}), \mathcal{H}(\nu | \mathbf{m}) < \infty$. Let $(T_n)_{n \in \mathbb{N}} \subset \mathbb{R}_{\geq 0}$ be a sequence converging to $+\infty$ and, for any n , consider the functional $\mathcal{H}(\cdot | R_{0,T_n})$ defined on $\Pi(\mu, \nu)$ endowed with the weak topology. Then*

$$\Gamma - \lim_{n \rightarrow \infty} \mathcal{H}(\cdot | R_{0,T_n}) = \mathcal{H}(\cdot | \mathbf{m} \otimes \mathbf{m})$$

Proof. As a first step, we claim that for any lower bounded lower semicontinuous function ϕ on $M \times M$ it holds

$$\liminf_{n \rightarrow \infty} \int_{M \times M} \phi dR_{0,T_n} \geq \int_{M \times M} \phi d(\mathbf{m} \otimes \mathbf{m}). \quad (3.1)$$

Since $R_{0,T_n}(M \times M) = 1$, the claim is trivial for constant functions; hence, without loss of generality we can assume that $\phi \geq 0$ and, under this further assumption, the Gaussian lower bound (2.8) together with Fatou's lemma yields (3.1). By Portmanteau theorem this is equivalent to say that

$$R_{0,T_n} \rightharpoonup \mathbf{m} \otimes \mathbf{m} \quad (3.2)$$

in duality with $C_b(M \times M)$.

After this premise, let $\pi \in \Pi(\mu, \nu)$, $(\pi_n) \subset \Pi(\mu, \nu)$ and assume that $\pi_n \rightharpoonup \pi$. The lower semicontinuity of the relative entropy w.r.t. to both its arguments together with (3.2) gives that

$$\liminf_{n \rightarrow \infty} \mathcal{H}(\pi_n | R_{0,T_n}) \geq \mathcal{H}(\pi | \mathbf{m} \otimes \mathbf{m}),$$

thus the Γ -liminf inequality. To complete the proof, let $\pi \in \Pi(\mu, \nu)$ and find a sequence $(\pi^n) \subset \Pi(\mu, \nu)$ such that

$$\limsup_{n \rightarrow \infty} \mathcal{H}(\pi^n | R_{0,T_n}) \leq \mathcal{H}(\pi | \mathbf{m} \otimes \mathbf{m}). \quad (3.3)$$

To this aim it is not restrictive to assume that $\mathcal{H}(\pi | \mathbf{m} \otimes \mathbf{m}) < \infty$ and it is also easy to see that $\pi^n \equiv \pi$ satisfies (3.3). Indeed, the Gaussian lower bound (2.8) and the fact that (H1) holds with $\kappa \geq 0$ imply that

$$-\log \mathbf{p}_{T_n}(x, y) \leq \frac{1}{2T_n} \mathbf{d}^2(x, y)$$

for all $x, y \in M$ and for all $n \in \mathbb{N}$. In particular, since $\mathbf{d}^2(x, y) \leq 2\mathbf{d}^2(x, z) + 2\mathbf{d}^2(y, z)$ for all $z \in M$ and $\mu, \nu \in \mathcal{P}_2(M)$ we have that $\int_{M \times M} -\log \mathbf{p}_{T_n}(x, y) \pi(\mathrm{d}x \mathrm{d}y)$ is well defined for all n and

$$-\int_{M \times M} \log \mathbf{p}_{T_n}(x, y) \pi(\mathrm{d}x \mathrm{d}y) \leq \frac{C}{2T_n} \quad (3.4)$$

for some constant C independent of n . Next, observe that by definition of $\mathbf{p}_{T_n}(\cdot, \cdot)$ and π^n we have

$$\mathcal{H}(\pi^n | R_{0, T_n}) = \mathcal{H}(\pi | \mathbf{m} \otimes \mathbf{m}) - \int \log \mathbf{p}_{T_n}(x, y) \pi(\mathrm{d}x, \mathrm{d}y), \quad (3.5)$$

where the sum on the right-hand side is well defined because $0 \leq \mathcal{H}(\pi | \mathbf{m} \otimes \mathbf{m}) < \infty$ by assumption and the fact that $\mathbf{m} \otimes \mathbf{m}$ is a probability, and a fortiori $\int \log \mathbf{p}_{T_n} \mathrm{d}\pi \in [-\infty, +\infty)$, since $\mathcal{H}(\pi^n | R_{0, T_n}) \geq 0$. Plugging (3.4) into (3.5) and letting $n \rightarrow \infty$ yields (3.3) and thus the Γ -limsup inequality. \square

Let us point out that in the previous lemma and thus in Theorem 1.2, (H2) is required only when $\kappa = 0$. As regards the long-time behavior of $\mathcal{E}_T(\mu, \nu)$, we need to determine in which way $\mathcal{E}_T(\mu, \nu)$ is controlled in terms of $\mathcal{C}_T(\mu, \nu)$.

Lemma 3.2. *Under (H1) and (H4) it holds*

$$T|\mathcal{E}_T(\mu, \nu)| \leq \mathcal{C}_T(\mu, \nu) - \frac{1}{2} \left(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m}) \right). \quad (3.6)$$

Proof. Let us start observing that

$$\begin{aligned} T\mathcal{E}_T(\mu, \nu) &= T \int_0^1 \mathcal{E}_T(\mu, \nu) \mathrm{d}t = \iint_0^1 \left(\frac{1}{2T} |v_t^T|^2 - \frac{T}{8} |\nabla \log \rho_t^T|^2 \right) \rho_t^T \mathrm{d}t \mathrm{d}\mathbf{m} \\ &= \mathcal{C}_T(\mu, \nu) - \frac{1}{2} \left(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m}) \right) - \frac{T}{4} \iint_0^1 |\nabla \log \rho_t^T|^2 \rho_t^T \mathrm{d}t \mathrm{d}\mathbf{m}, \end{aligned}$$

whence trivially the desired upper bound for $T\mathcal{E}_T(\mu, \nu)$. On the other hand, Young's inequality and (2.16) yield

$$\begin{aligned} \frac{T}{4} \iint_0^1 |\nabla \log \rho_t^T|^2 \rho_t^T \mathrm{d}t \mathrm{d}\mathbf{m} &\leq \frac{T}{2} \iint_0^1 \left(|\nabla \log \mathbf{P}_{Tt} f^T|^2 + |\nabla \log \mathbf{P}_{T(1-t)} g^T|^2 \right) \rho_t^T \mathrm{d}t \mathrm{d}\mathbf{m} \\ &= 2\mathcal{C}_T(\mu, \nu) - \mathcal{H}(\mu | \mathbf{m}) - \mathcal{H}(\nu | \mathbf{m}), \end{aligned}$$

so that plugging this inequality into the previous identity gives also the lower bound for $T\mathcal{E}_T(\mu, \nu)$. \square

We are now in the position to prove Theorem 1.2.

Proof. It is easily seen that the unique optimal coupling in

$$\inf_{\pi \in \Pi(\mu, \nu)} \mathcal{H}(\pi | \mathbf{m} \otimes \mathbf{m}) \quad (3.7)$$

is $\mu \otimes \nu$ and that the optimal value is $\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m})$. Consider a sequence $(T_n)_{n \in \mathbb{N}}$ such that $T_n \rightarrow \infty$ and let π_n be the optimal coupling in the corresponding Schrödinger problem, i.e. $\mathcal{H}(\pi_n | R_{0, T_n}) = \mathcal{E}_{T_n}(\mu, \nu)$. Since $\Pi(\mu, \nu)$ is compact for the weak topology, π^n has an accumulation point π^* . The basic results of Γ -convergence ensure that π^* is optimal for (3.7), whence $\pi^* = \mu \otimes \nu$ by uniqueness. Using again the basic properties of Γ -convergent sequences we finally obtain the convergence of the optimal values, namely

$$\lim_{n \rightarrow \infty} \mathcal{E}_{T_n}(\mu, \nu) = \mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m}).$$

Dividing by T (3.6) and letting $T \rightarrow \infty$, the long-time behavior of $\mathcal{E}_T(\mu, \nu)$ is established as well. \square

3.2 Proof of Theorem 1.1 and Theorem 1.6

The proof of Theorem 1.1 requires the preparatory Lemmas 3.3, 3.4 and 3.5. In the first one we show that the Fisher information of the entropic interpolation is non-increasing as a function of T .

Lemma 3.3. *Under the assumptions of Theorem 1.1, let (ρ^T, v^T) be optimal for the formulation (2.13) and (ρ^0, v^0) for (2.15). Then the function*

$$[0, \infty) \ni T \mapsto \int_0^1 \int_0^1 |\nabla \log \rho_t^T|^2 \rho_t^T dt d\mathbf{m}$$

is non-increasing, where $\int |\nabla \log \rho_t^0|^2 \rho_t^0 d\mathbf{m} := +\infty$ whenever $\rho_t^0 \notin W_{loc}^{1,2}(M)$, the space of locally Sobolev functions on M .

Proof. Fix $0 \leq T_1 < T_2$. Summing the inequalities

$$\mathcal{A}_{T_1}(\rho^{T_1}, v^{T_1}) \leq \mathcal{A}_{T_1}(\rho^{T_2}, v^{T_2}), \quad \mathcal{A}_{T_2}(\rho^{T_2}, v^{T_2}) \leq \mathcal{A}_{T_2}(\rho^{T_1}, v^{T_1})$$

and dividing by $(T_2^2 - T_1^2)/8$ we obtain

$$\int_0^1 \int_0^1 |\nabla \log \rho_t^{T_2}|^2 \rho_t^{T_2} dt d\mathbf{m} \leq \int_0^1 \int_0^1 |\nabla \log \rho_t^{T_1}|^2 \rho_t^{T_1} dt d\mathbf{m},$$

which is the desired conclusion. \square

In the second lemma we prove the continuity in T of the functions f^T and g^T given by (2.9) with respect to the $L^p(\mathbf{m})$ norm, for any $p \in [1, \infty)$.

Lemma 3.4. *Under the assumptions of Theorem 1.1, the functions $T \mapsto f^T$ and $T \mapsto g^T$ are continuous from $(0, \infty)$ to $L^p(\mathbf{m})$, for any $p \in [1, \infty)$.*

As a byproduct, the functions $T \mapsto \mathbf{P}_T f^T$ and $T \mapsto \mathbf{P}_T g^T$ are continuous from $(0, \infty)$ to $L^p(\mathbf{m})$, $p \in [1, \infty)$, as well.

Proof. Let $T_0 > 0$, $0 < \delta < T_0$ and denote by C_δ the positive constant provided by Lemma 2.1-(i) on the interval $[T_0 - \delta, T_0 + \delta]$. From the first bound in (2.11) we immediately deduce that $(f^T)_{T \in [T_0 - \delta, T_0 + \delta]}$ is bounded in $L^\infty(\mathbf{m})$, hence in $L^p(\mathbf{m})$ for all $p \in [1, \infty]$, because all the functions f^T are supported in $\text{supp}(\mu)$ and this has finite mass, because bounded. To prove the same property for $(g^T)_{T \in [T_0 - \delta, T_0 + \delta]}$ requires a more technical argument.

From (2.7) with $\varepsilon = 1$ we first observe that

$$\begin{aligned} 1 &= \int f^T(x)g^T(y)\mathbf{p}_T(x,y)\mathbf{m}(dx)\mathbf{m}(dy) \\ &\leq \int f^T(x)g^T(y)\frac{C(\kappa, T_0, \delta)}{\sqrt{\mathbf{m}(B_{\sqrt{T_0-\delta}}(x))\mathbf{m}(B_{\sqrt{T_0-\delta}}(y))}}\exp\left(-\frac{d^2(x,y)}{5(T_0+\delta)}\right)\mathbf{m}(dx)\mathbf{m}(dy) \\ &\leq \int f^T(x)g^T(y)\frac{C(\kappa, T_0, \delta)}{\sqrt{\mathbf{m}(B_{\sqrt{T_0-\delta}}(x))\mathbf{m}(B_{\sqrt{T_0-\delta}}(y))}}\mathbf{m}(dx)\mathbf{m}(dy), \quad \forall T \in [T_0 - \delta, T_0 + \delta] \end{aligned}$$

and since $\text{supp}(f^T) \subset \text{supp}(\mu)$, $\text{supp}(g^T) \subset \text{supp}(\nu)$ for all $T > 0$, μ, ν have bounded support and $x \mapsto \mathbf{m}(B_t(x))$ is a continuous function,

$$\inf_{x \in \text{supp}(f^T)} \mathbf{m}(B_{\sqrt{T_0-\delta}}(x)), \quad \inf_{y \in \text{supp}(g^T)} \mathbf{m}(B_{\sqrt{T_0-\delta}}(y)) \geq C_\delta > 0$$

with C_δ independent of T . Thus if we combine this inequality with the previous one and recall the chosen normalization (2.10) we obtain $1 \leq C'_\delta \|f^T\|_{L^1(\mathbf{m})}$ for all $T \in [T_0 - \delta, T_0 + \delta]$ for some $C'_\delta > 0$ (depending on T_0 and κ too), i.e.

$$\|f^T\|_{L^1(\mathbf{m})} \geq \frac{1}{C'_\delta}, \quad \forall T \in [T_0 - \delta, T_0 + \delta].$$

Plugging this inequality into the second bound in (2.11) yields

$$\|g^T\|_{L^\infty(\mathbf{m})} \leq \frac{C'_\delta}{C_\delta} \|\sigma\|_{L^\infty(\mathbf{m})}, \quad \forall T \in [T_0 - \delta, T_0 + \delta]. \quad (3.8)$$

Hence also $(g^T)_{T \in [T_0 - \delta, T_0 + \delta]}$ is bounded in $L^p(\mathbf{m})$ for all $p \in [1, \infty]$, because all the functions g^T are supported in $\text{supp}(\nu)$ and this has finite mass.

After this premise, we are now in the position to prove the continuity of $T \mapsto f^T$ and $T \mapsto g^T$. Indeed, if $|h| < \delta$ then by Banach-Alaoglu theorem there exist a (not relabeled) subsequence and functions $\tilde{f}, \tilde{g} \in L^1 \cap L^\infty(\mathbf{m})$ such that $f^{T_0+h} \xrightarrow{*} \tilde{f}$ and $g^{T_0+h} \xrightarrow{*} \tilde{g}$ in $L^\infty(\mathbf{m})$ as $h \rightarrow 0$. We claim that

$$\lim_{h \rightarrow 0} \mathbf{P}_{T_0+h} f^{T_0+h} = \mathbf{P}_{T_0} \tilde{f} \quad \text{and} \quad \lim_{h \rightarrow 0} \mathbf{P}_{T_0+h} g^{T_0+h} = \mathbf{P}_{T_0} \tilde{g}, \quad (3.9)$$

where both limits have to be understood in $L^p(\mathbf{m})$ for $p \in [1, \infty)$.

To this aim, write

$$\|\mathbf{P}_{T_0+h} f^{T_0+h} - \mathbf{P}_{T_0} \tilde{f}\|_{L^p(\mathbf{m})} \leq \|\mathbf{P}_{T_0+h} f^{T_0+h} - \mathbf{P}_{T_0+h} \tilde{f}\|_{L^p(\mathbf{m})} + \|\mathbf{P}_{T_0+h} \tilde{f} - \mathbf{P}_{T_0} \tilde{f}\|_{L^p(\mathbf{m})}$$

and observe that the second term on the right-hand side trivially vanishes as $h \rightarrow 0$. As regards the first one,

$$\begin{aligned} \|\mathbf{P}_{T_0+h} f^{T_0+h} - \mathbf{P}_{T_0+h} \tilde{f}\|_{L^p(\mathbf{m})} &= \|\mathbf{P}_{\delta+h}(\mathbf{P}_{T_0-\delta} f^{T_0+h} - \mathbf{P}_{T_0-\delta} \tilde{f})\|_{L^p(\mathbf{m})} \\ &\leq \|\mathbf{P}_{T_0-\delta} f^{T_0+h} - \mathbf{P}_{T_0-\delta} \tilde{f}\|_{L^p(\mathbf{m})}, \end{aligned}$$

because $\mathbf{P}_T : L^p(\mathfrak{m}) \rightarrow L^p(\mathfrak{m})$ is a contraction for all $p \in [1, \infty]$. Secondly, by (2.6), the fact that $f^{T_0+h} \xrightarrow{*} \tilde{f}$ in $L^\infty(\mathfrak{m})$ and $\mathbf{p}_t(x, \cdot) \in L^1(\mathfrak{m})$ for all $x \in M$, we deduce that

$$\lim_{h \rightarrow 0} \mathbf{P}_{T_0-\delta} f^{T_0+h}(x) = \mathbf{P}_{T_0-\delta} \tilde{f}(x), \quad \forall x \in M. \quad (3.10)$$

Since, as already remarked, $(f^T)_{T \in [T_0-\delta, T_0+\delta]}$ is bounded in $L^\infty(\mathfrak{m})$ and all the functions f^T are supported in $\text{supp}(\mu)$, there exists $M > 0$ sufficiently large such that $0 \leq f^T \leq M \mathbf{1}_{\text{supp}(\mu)}$ for all $T \in [T_0 - \delta, T_0 + \delta]$, whence

$$0 \leq \mathbf{P}_{T_0-\delta} f^T \leq M \mathbf{P}_{T_0-\delta}(\mathbf{1}_{\text{supp}(\mu)}), \quad \forall T \in [T_0 - \delta, T_0 + \delta]$$

by the maximum principle. As the right-hand side above belongs to $L^1 \cap L^\infty(\mathfrak{m})$ and does not depend on T , by (3.10) and the dominated convergence theorem we thus infer that

$$\lim_{h \rightarrow 0} \|\mathbf{P}_{T_0-\delta} f^{T_0+h} - \mathbf{P}_{T_0-\delta} \tilde{f}\|_{L^p(\mathfrak{m})} = 0,$$

whence, combining this fact with the previous steps, the validity of the first limit in (3.9). The argument for the second limit is completely analogous.

As a consequence, up to extract a further (not relabeled) subsequence, we have that the limits in (3.9) hold \mathfrak{m} -a.e. Therefore, if we look at the Schrödinger system (2.9) at time $T_0 + h$, which reads as

$$\rho = f^{T_0+h} \mathbf{P}_{T_0+h} g^{T_0+h}, \quad \sigma = g^{T_0+h} \mathbf{P}_{T_0+h} f^{T_0+h},$$

and rewrite it as

$$f^{T_0+h} = \frac{\rho}{\mathbf{P}_{T_0+h} g^{T_0+h}}, \quad g^{T_0+h} = \frac{\sigma}{\mathbf{P}_{T_0+h} f^{T_0+h}},$$

which is possible because $\mathbf{P}_{T_0+h} f^{T_0+h}, \mathbf{P}_{T_0+h} g^{T_0+h} > 0$, we deduce that

$$\lim_{h \rightarrow 0} f^{T_0+h} = \frac{\rho}{\mathbf{P}_{T_0} \tilde{g}} \quad \text{and} \quad \lim_{h \rightarrow 0} g^{T_0+h} = \frac{\sigma}{\mathbf{P}_{T_0} \tilde{f}},$$

both limits being understood \mathfrak{m} -a.e. This is compatible with $f^{T_0+h} \xrightarrow{*} \tilde{f}$ and $g^{T_0+h} \xrightarrow{*} \tilde{g}$ in $L^\infty(\mathfrak{m})$ as $h \rightarrow 0$ if and only if

$$\tilde{f} = \frac{\rho}{\mathbf{P}_{T_0} \tilde{g}} \quad \text{and} \quad \tilde{g} = \frac{\sigma}{\mathbf{P}_{T_0} \tilde{f}},$$

namely if and only if

$$\rho = \tilde{f} \mathbf{P}_{T_0} \tilde{g}, \quad \sigma = \tilde{g} \mathbf{P}_{T_0+h} \tilde{f}.$$

This means that $\tilde{f} = f^{T_0}$ and $\tilde{g} = g^{T_0}$ by uniqueness of the (f, g) -decomposition and by the normalization (2.10). Since this is true for any convergent subsequence obtained via Banach-Alaoglu theorem, we deduce that the whole sequences (f^{T_0+h}) and (g^{T_0+h}) converge in $L^p(\mathfrak{m})$ to f^{T_0} and g^{T_0} respectively, whence the desired conclusion.

It is then sufficient to combine the fact that $\tilde{f} = f^{T_0}$ and $\tilde{g} = g^{T_0}$ with (3.9) to get also the continuity of $T \mapsto \mathbf{P}_T f^T$ and $T \mapsto \mathbf{P}_T g^T$. \square

Finally, in the next lemma we show that the cost $\mathcal{C}_T(\mu, \nu)$, the energy $\mathcal{E}_T(\mu, \nu)$ and the integrated-in-time Fisher information $\int \int_0^1 |\nabla \log \rho_t^T|^2 d\mu_t^T$ are continuous as a function of T (the regularity of $\mathcal{C}_T(\mu, \nu)$ will be greatly improved in Theorem 1.1, of course).

Lemma 3.5. *Under the assumptions of Theorem 1.1, the functions $T \mapsto \mathcal{E}_T(\mu, \nu)$, $T \mapsto \mathcal{E}_T(\mu, \nu)$ and $T \mapsto \int \int_0^1 |\nabla \log \rho_t^T|^2 d\mu_t^T$ are continuous on $(0, \infty)$.*

Proof. For $T \mapsto \mathcal{E}_T(\mu, \nu)$, by (2.9) we observe that the cost can be rewritten as

$$\mathcal{E}_T(\mu, \nu) = \int (f^T \log f^T) \mathbf{P}_T g^T d\mathbf{m} + \int (g^T \log g^T) \mathbf{P}_T f^T d\mathbf{m}$$

and by Lemma 3.4, the locally uniform L^∞ -bounds provided by Lemma 2.1 and by dominated convergence it is easy to see that the right-hand side above continuously depends on T .

As regards $T \mapsto \mathcal{E}_T(\mu, \nu)$, let us start observing that the second identity in (2.18) can be equivalently rewritten as

$$\mathcal{E}_T(\mu, \nu) = -\frac{1}{2} \int \nabla \mathbf{P}_{Tt} f^T \cdot \nabla \mathbf{P}_{T(1-t)} g^T d\mathbf{m},$$

so fix $t \in (0, 1)$ and note that, by integration by parts first and Cauchy-Schwarz inequality then, we obtain

$$\begin{aligned} |\mathcal{E}_{T+h}(\mu, \nu) - \mathcal{E}_T(\mu, \nu)| &\leq \frac{1}{2} \left| \int \nabla \mathbf{P}_{(T+h)t} f^{T+h} \cdot \nabla (\mathbf{P}_{(T+h)(1-t)} g^{T+h} - \mathbf{P}_{T(1-t)} g^T) d\mathbf{m} \right| \\ &\quad + \frac{1}{2} \left| \int \nabla (\mathbf{P}_{(T+h)t} f^{T+h} - \mathbf{P}_{Tt} f^T) \cdot \nabla \mathbf{P}_{T(1-t)} g^T d\mathbf{m} \right| \\ &\leq \frac{1}{2} \left| \int \mathbf{L} \mathbf{P}_{(T+h)t} f^{T+h} (\mathbf{P}_{(T+h)(1-t)} g^{T+h} - \mathbf{P}_{T(1-t)} g^T) d\mathbf{m} \right| \\ &\quad + \frac{1}{2} \left| \int (\mathbf{P}_{(T+h)t} f^{T+h} - \mathbf{P}_{Tt} f^T) \mathbf{L} \mathbf{P}_{T(1-t)} g^T d\mathbf{m} \right| \\ &\leq \frac{1}{2} \|\mathbf{L} \mathbf{P}_{(T+h)t} f^{T+h}\|_{L^2(\mathbf{m})} \|\mathbf{P}_{(T+h)(1-t)} g^{T+h} - \mathbf{P}_{T(1-t)} g^T\|_{L^2(\mathbf{m})} \\ &\quad + \frac{1}{2} \|\mathbf{P}_{(T+h)t} f^{T+h} - \mathbf{P}_{Tt} f^T\|_{L^2(\mathbf{m})} \|\mathbf{L} \mathbf{P}_{T(1-t)} g^T\|_{L^2(\mathbf{m})}. \end{aligned}$$

The second summand after the last inequality clearly vanishes as $h \rightarrow 0$, because of Lemma 3.4. As concerns the first one, argue as in the proof of Lemma 3.4 and notice that for h sufficiently small, say $|h| < \delta$ with $0 \leq \delta \leq T$, all the functions f^{T+h} are uniformly bounded in $L^\infty(\mathbf{m})$ and supported in $\text{supp}(\mu)$, hence also uniformly bounded in $L^2(\mathbf{m})$. By the a priori estimate (2.1), this is sufficient to conclude that there exists $M > 0$ sufficiently large such that $\|\mathbf{L} \mathbf{P}_{(T+h)t} f^{T+h}\|_{L^2(\mathbf{m})} \leq M$ for all $|h| < \delta$. Hence, by Lemma 3.4, also the first summand after the last inequality converges to 0 in the limit, whence the continuity of $T \mapsto \mathcal{E}_T(\mu, \nu)$.

Finally, the continuity of $T \mapsto \int \int_0^1 |\nabla \log \rho_t^T|^2 d\mu_t^T$ is a pure matter of algebraic manipulation, since by (2.13) and (2.18)

$$\int \int_0^1 |\nabla \log \rho_t^T|^2 d\mu_t^T = \frac{4}{T} \mathcal{E}_T(\mu, \nu) - 4\mathcal{E}_T(\mu, \nu) - \frac{2}{T} \left(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m}) \right)$$

and by the previous discussion the right-hand side is continuous on $(0, \infty)$. \square

Let us now prove Theorem 1.1.

Proof of Theorem 1.1. Consider the map $T \mapsto \mathcal{A}_T(\rho^T, v^T)$, let $h > 0$, write

$$\begin{aligned} \frac{\mathcal{A}_{T+h}(\rho^{T+h}, v^{T+h}) - \mathcal{A}_T(\rho^T, v^T)}{h} &= \frac{\mathcal{A}_{T+h}(\rho^{T+h}, v^{T+h}) - \mathcal{A}_{T+h}(\rho^T, v^T)}{h} \\ &\quad + \frac{\mathcal{A}_{T+h}(\rho^T, v^T) - \mathcal{A}_T(\rho^T, v^T)}{h} \end{aligned} \quad (3.11)$$

and observe that the first term on the right-hand side is non-positive by optimality of (ρ^{T+h}, v^{T+h}) for \mathcal{A}_{T+h} . For the second one

$$\mathcal{A}_{T+h}(\rho^T, v^T) - \mathcal{A}_T(\rho^T, v^T) = \left(\frac{Th}{4} + \frac{h^2}{8} \right) \iint_0^1 |\nabla \log \rho_t^T|^2 \rho_t^T dt dm,$$

so that

$$\lim_{h \downarrow 0} \frac{\mathcal{A}_{T+h}(\rho^T, v^T) - \mathcal{A}_T(\rho^T, v^T)}{h} = \frac{T}{4} \iint_0^1 |\nabla \log \rho_t^T|^2 \rho_t^T dt dm = (\partial_T \mathcal{A}_T)(\rho^T, v^T), \quad (3.12)$$

where $\partial_T \mathcal{A}_T$ denotes the partial derivative w.r.t. T of $(T, \rho, v) \mapsto \mathcal{A}_T(\rho, v)$. Combining these two facts we deduce

$$\limsup_{h \downarrow 0} \frac{\mathcal{A}_{T+h}(\rho^{T+h}, v^{T+h}) - \mathcal{A}_T(\rho^T, v^T)}{h} \leq (\partial_T \mathcal{A}_T)(\rho^T, v^T). \quad (3.13)$$

On the other hand we can also write

$$\begin{aligned} \frac{\mathcal{A}_{T+h}(\rho^{T+h}, v^{T+h}) - \mathcal{A}_T(\rho^T, v^T)}{h} &= \frac{\mathcal{A}_{T+h}(\rho^{T+h}, v^{T+h}) - \mathcal{A}_T(\rho^{T+h}, v^{T+h})}{h} \\ &\quad + \frac{\mathcal{A}_T(\rho^{T+h}, v^{T+h}) - \mathcal{A}_T(\rho^T, v^T)}{h} \end{aligned} \quad (3.14)$$

and remark that now the second term on the right-hand side is non-negative by optimality of (ρ^T, v^T) for \mathcal{A}_T . As regards the first one,

$$\mathcal{A}_{T+h}(\rho^{T+h}, v^{T+h}) - \mathcal{A}_T(\rho^{T+h}, v^{T+h}) = \left(\frac{Th}{4} + \frac{h^2}{8} \right) \iint_0^1 |\nabla \log \rho_t^{T+h}|^2 \rho_t^{T+h} dt dm$$

and by Lemma 3.5 we infer that

$$\begin{aligned} \lim_{h \downarrow 0} \frac{\mathcal{A}_{T+h}(\rho^{T+h}, v^{T+h}) - \mathcal{A}_T(\rho^{T+h}, v^{T+h})}{h} &\geq \frac{T}{4} \iint_0^1 |\nabla \log \rho_t^T|^2 \rho_t^T dt dm \\ &= (\partial_T \mathcal{A}_T)(\rho^T, v^T). \end{aligned} \quad (3.15)$$

This yields

$$\liminf_{h \downarrow 0} \frac{\mathcal{A}_{T+h}(\rho^{T+h}, v^{T+h}) - \mathcal{A}_T(\rho^T, v^T)}{h} \geq (\partial_T \mathcal{A}_T)(\rho^T, v^T),$$

which together with (3.13) implies that $T \mapsto \mathcal{A}_T(\rho^T, v^T)$ is everywhere right differentiable on $(0, \infty)$. Left differentiability follows by an analogous argument. Indeed if $h < 0$, then the first term on the right-hand side in (3.11) is non-negative and (3.12) still holds true with $h \uparrow 0$ instead of $h \downarrow 0$, whence

$$\liminf_{h \uparrow 0} \frac{\mathcal{A}_{T+h}(\rho^{T+h}, v^{T+h}) - \mathcal{A}_T(\rho^T, v^T)}{h} \geq (\partial_T \mathcal{A}_T)(\rho^T, v^T).$$

Applying the same considerations to (3.14) and (3.15) gives

$$\limsup_{h \uparrow 0} \frac{\mathcal{A}_{T+h}(\rho^{T+h}, v^{T+h}) - \mathcal{A}_T(\rho^T, v^T)}{h} \leq (\partial_T \mathcal{A}_T)(\rho^T, v^T)$$

and combining this inequality with the previous one entails the desired left differentiability. Therefore $T \mapsto \mathcal{A}_T(\rho^T, v^T)$ is everywhere differentiable on $(0, \infty)$, a fortiori continuous therein, and

$$\frac{d}{dT} \mathcal{A}_T(\rho^T, v^T) = (\partial_T \mathcal{A}_T)(\rho^T, v^T).$$

Therefore $T \mapsto T\mathcal{C}_T(\mu, \nu)$ is everywhere differentiable (hence continuous) on $(0, \infty)$ as well and, by (2.13) and the very definition of $\mathcal{C}_T(\mu, \nu)$ and $\mathcal{E}_T(\mu, \nu)$,

$$\begin{aligned} \frac{d}{dT}(T\mathcal{C}_T(\mu, \nu)) &= \frac{1}{2} \left(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m}) \right) + (\partial_T \mathcal{A}_T)(\rho^T, v^T) \\ &= \frac{1}{2} \left(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m}) \right) + \frac{T}{4} \iint_0^1 |\nabla \log \rho_t^T|^2 \rho_t^T d\mathbf{m} \\ &= \mathcal{C}_T(\mu, \nu) - T\mathcal{E}_T(\mu, \nu), \end{aligned}$$

which proves both formulas for the first derivative of $T \mapsto T\mathcal{C}_T(\mu, \nu)$, thanks to (2.14). Now notice that the right-hand side above is continuous on $(0, \infty)$: $T \mapsto \mathcal{C}_T(\mu, \nu)$ is continuous on $(0, \infty)$ since so is $T \mapsto T\mathcal{E}_T(\mu, \nu)$, while $T \mapsto T\mathcal{E}_T(\mu, \nu)$ is continuous by Lemma 3.5. Hence $T \mapsto T\mathcal{C}_T(\mu, \nu)$ belongs to $C^1((0, \infty))$. The fact that $T \mapsto T\mathcal{C}_T(\mu, \nu)$ is also twice differentiable a.e. follows from the fact that

$$\frac{d}{dT}(T\mathcal{C}_T(\mu, \nu)) = \frac{1}{2} \left(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m}) \right) + \frac{T}{4} \iint_0^1 |\nabla \log \rho_t^T|^2 \rho_t^T d\mathbf{m} \quad (3.16)$$

and the last term on the right-hand side is the product of a linear function and a monotone (non-increasing) one, thanks to Lemma 3.3.

This concludes the proof of (ii). Claim (i) is a straightforward consequence as well as the formula for the second derivative of $T \mapsto T\mathcal{C}_T(\mu, \nu)$. \square

From Theorem 1.1, the proof of Theorem 1.6 follows.

Proof of Theorem 1.6. As regards (1.17), we only prove the upper bound, the lower bound being an immediate consequence of the Benamou-Brenier formulation (2.13). Since $T \mapsto T\mathcal{C}_T(\mu, \nu)$ belongs to $C^1((0, \infty))$ we can use the fundamental theorem of calculus that gives for all $\varepsilon > 0$

$$\begin{aligned} T\mathcal{C}_T(\mu, \nu) - \varepsilon\mathcal{C}_\varepsilon(\mu, \nu) &= \frac{\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m})}{2} (T - \varepsilon) + \int_\varepsilon^T \frac{S}{4} \left(\iint_0^1 |\nabla \log \rho_t^S|^2 \rho_t^S dt d\mathbf{m} \right) dS \\ &\leq \frac{\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m})}{2} (T - \varepsilon) + \left(\frac{T^2}{8} - \frac{\varepsilon^2}{8} \right) \iint_0^1 |\nabla \log \rho_t^0|^2 \rho_t^0 dt d\mathbf{m}, \end{aligned}$$

where the inequality is motivated by Lemma 3.3. The conclusion follows letting $\varepsilon \rightarrow 0$ and using the convergence of the rescaled entropic cost towards $W_2^2(\mu, \nu)/2$.

In order to prove (1.18), assume $\iint_0^1 |\nabla \log \rho_t^0|^2 \rho_t^0 dt d\mathbf{m} < \infty$ and accept, for the moment, that the following three auxiliary facts hold: (a) $T \mapsto \iint_0^1 |\nabla \log \rho_t^T|^2 \rho_t^T dt d\mathbf{m}$ is continuous at

$T = 0$; (b) the first derivative of $T \mapsto T\mathcal{C}_T(\mu, \nu)$ is continuous up to $T = 0$; (c) $T \mapsto T\mathcal{C}_T(\mu, \nu)$ is twice differentiable at $T = 0$. Then (c) implies that $T \mapsto T\mathcal{C}_T(\mu, \nu)$ has a Taylor expansion of the form

$$T\mathcal{C}_T(\mu, \nu) = \frac{1}{2}W_2^2(\mu, \nu) + AT + B\frac{T^2}{2} + o(T^2).$$

The coefficients can be identified thanks to Theorem 1.1. Indeed

$$A = \frac{d}{dT}\Big|_{T=0^+} (T\mathcal{C}_T(\mu, \nu)) = \lim_{T' \downarrow 0} \frac{d}{dT}\Big|_{T=T'} (T\mathcal{C}_T(\mu, \nu)) = \frac{1}{2}(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m})),$$

the second identity being a consequence of (b), the third one following from (3.16) and the finiteness of $\iint_0^1 |\nabla \log \rho_t^0|^2 \rho_t^0 dt d\mathbf{m}$. As regards B , this is given by

$$\begin{aligned} B &= \frac{d^2}{dT^2}\Big|_{T=0^+} (T\mathcal{C}_T(\mu, \nu)) = \lim_{T' \downarrow 0} \frac{1}{T'} \left(\frac{d}{dT}\Big|_{T=T'} (T\mathcal{C}_T(\mu, \nu)) - \frac{d}{dT}\Big|_{T=0^+} (T\mathcal{C}_T(\mu, \nu)) \right) \\ &= \lim_{T' \downarrow 0} \frac{1}{4} \iint_0^1 |\nabla \log \rho_t^{T'}|^2 \rho_t^{T'} dt d\mathbf{m} = \frac{1}{4} \iint_0^1 |\nabla \log \rho_t^0|^2 \rho_t^0 dt d\mathbf{m} \end{aligned}$$

thanks to (a). Therefore it only remains to show (a), (b), and (c); we shall start with the first one. On the one hand, by Lemma 3.3

$$\limsup_{T \downarrow 0} \iint_0^1 |\nabla \log \rho_t^T|^2 \rho_t^T dt d\mathbf{m} \leq \iint_0^1 |\nabla \log \rho_t^0|^2 \rho_t^0 dt d\mathbf{m}. \quad (3.17)$$

On the other hand, as $\text{CD}(\kappa, N)$ holds, we can and shall rely on powerful facts established in [18]: by Propositions 5.1 and 5.4 therein we know that

$$\rho_t^T \xrightarrow{*} \rho_t^0 \quad \text{in } L^\infty(\mathbf{m}), \quad \forall t \in [0, 1] \quad (3.18)$$

and by [18, Lemma 4.2] we also know that for all $\bar{x} \in M$ there exist $C', C'', r > 0$ such that

$$\rho_t^T(x) \leq C' \exp\left(-\frac{C'' d_g^2(x, \bar{x})}{T}\right), \quad \forall x \in M \setminus B_r(\bar{x}), t \in [0, 1], T \in (0, 1). \quad (3.19)$$

With this said, let $B \subset M$ be any Borel set and let us prove that

$$(\rho_t^T \mathbf{m})(B) \rightarrow (\rho_t^0 \mathbf{m})(B) \quad \text{as } T \downarrow 0, \quad \forall t \in [0, 1]. \quad (3.20)$$

To this end, since μ, ν are compactly supported, there exists a bounded set containing the supports of ρ_t^0 for all $t \in [0, 1]$, so that up to choose a larger r we can assume that $\text{supp}(\rho_t^0) \subset B_r(\bar{x})$ for all $t \in [0, 1]$. By (3.18) and the trivial fact that $\mathbf{1}_{B \cap B_r(\bar{x})} \in L^1(\mathbf{m})$ this implies that $(\rho_t^T \mathbf{m})(B \cap B_r(\bar{x})) \rightarrow (\rho_t^0 \mathbf{m})(B \cap B_r(\bar{x}))$ as $T \downarrow 0$ for all $t \in [0, 1]$. By (3.19) and the fact that under the $\text{CD}(\kappa, N)$ condition the function in the right-hand side of (3.19) is integrable and converges monotonically to 0 as $T \downarrow 0$, it also follows

$$(\rho_t^T \mathbf{m})(B \setminus B_r(\bar{x})) \rightarrow 0 = (\rho_t^0 \mathbf{m})(B \setminus B_r(\bar{x})) \quad \text{as } T \downarrow 0,$$

whence (3.20). By [2, Theorem 7.6] this implies

$$\liminf_{T \downarrow 0} \int |\nabla \log \rho_t^T|^2 \rho_t^T d\mathbf{m} \geq \int |\nabla \log \rho_t^0|^2 \rho_t^0 dt d\mathbf{m}, \quad \forall t \in [0, 1]$$

whence by Fatou's lemma

$$\liminf_{T \downarrow 0} \iint_0^1 |\nabla \log \rho_t^T|^2 \rho_t^T dt d\mathbf{m} \geq \iint_0^1 |\nabla \log \rho_t^0|^2 \rho_t^0 dt d\mathbf{m}.$$

Combining this inequality with (3.17) yields the desired continuity at $T = 0$ and thus (a).

As regards (b), by (1.17) and the finiteness of $\iint_0^1 |\nabla \log \rho_t^0|^2 \rho_t^0 dt d\mathbf{m}$ it immediately follows that

$$\limsup_{T \downarrow 0} \left\{ \mathcal{C}_T(\mu, \nu) - \frac{1}{2T} W_2^2(\mu, \nu) \right\} \leq \frac{1}{2} \left(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m}) \right). \quad (3.21)$$

On the other hand, by following the argument of Theorem 1.1 and recalling (2.15) we see that

$$\begin{aligned} T\mathcal{C}_T(\mu, \nu) - \frac{1}{2} W_2^2(\mu, \nu) &= \frac{T}{2} \left(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m}) \right) + \mathcal{A}_T(\rho^T, v^T) - \mathcal{A}_0(\rho^0, v^0) \\ &\geq \frac{T}{2} \left(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m}) \right) + \mathcal{A}_T(\rho^T, v^T) - \mathcal{A}_0(\rho^T, v^T) \\ &= \frac{T}{2} \left(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m}) \right) + \frac{T^2}{8} \iint_0^1 |\nabla \log \rho_t^T|^2 \rho_t^T dt d\mathbf{m}, \end{aligned}$$

where (v_t^0) is any choice of gradient of intermediate Kantorovich potentials associated with the geodesic $(\rho_t^0 \mathbf{m})$, and by (a) this implies

$$\liminf_{T \downarrow 0} \left\{ \mathcal{C}_T(\mu, \nu) - \frac{1}{2T} W_2^2(\mu, \nu) \right\} \geq \frac{1}{2} \left(\mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m}) \right).$$

Combining this inequality with (3.21) yields the differentiability of $T \mapsto T\mathcal{C}_T(\mu, \nu)$ at $T = 0$ and by (3.16) we see that the derivative is continuous there. Finally, by (2.13), the computations above for determining the coefficients A and B , and the properties (a) and (b) it is not difficult to see that (c) holds too. \square

3.3 Proof of Theorem 1.2 and Corollary 1.3

In this section we provide the reader with a second proof of Theorem 1.2, whose crucial ingredient is a long-time analogue of Lemma 3.4. This is precisely the content of the next result. With respect to the proof given in Section 3.1, here the marginals μ, ν satisfy (H4), hence a condition stronger than (H3); as a consequence we gain information on the long-time behavior of f^T and g^T separately, determining their limits as $T \rightarrow \infty$. Let us point out again that (H2) below is required only when $\kappa = 0$.

Lemma 3.6. *Under the assumptions (H1) with $\kappa \geq 0$, (H2) and (H4), for the functions f^T, g^T given by (2.9) it holds*

$$\lim_{T \rightarrow \infty} f^T = \rho, \quad \lim_{T \rightarrow \infty} g^T = \sigma \quad (3.22)$$

where both limits are in $L^p(\mathbf{m})$ for any $p \in [1, \infty)$.

As a byproduct

$$\lim_{T \rightarrow \infty} \mathbf{P}_T f^T = \lim_{T \rightarrow \infty} \mathbf{P}_T g^T = 1, \quad (3.23)$$

where both limits are in $L^p(\mathbf{m})$ for any $p \in [1, \infty)$.

Proof. Let $T_0 > 0$ and denote by C the positive constant provided by Lemma 2.1-(ii) on the interval $[T_0, \infty)$. From the first bound in (2.12) we immediately deduce that $(f^T)_{T \geq T_0}$ is bounded in $L^\infty(\mathfrak{m})$, hence in $L^p(\mathfrak{m})$ for all $p \in [1, \infty]$, as already argued in the proof of Lemma 3.4. From (2.7) with $C_\kappa = 0$ (since $\kappa \geq 0$), (2.10), the fact that μ, ν have bounded support and arguing as in Lemma 3.4 we also have

$$1 = \int f^T(x)g^T(y)\mathfrak{p}_T(x, y)d\mathfrak{m}(x)d\mathfrak{m}(y) \leq C'\|f^T\|_{L^1(\mathfrak{m})}, \quad \forall T \geq T_0$$

for some $C' > 0$, whence

$$\|f^T\|_{L^1(\mathfrak{m})} \geq \frac{1}{C'}, \quad \forall T \geq T_0. \quad (3.24)$$

Plugging this inequality into the second bound in (2.12) yields that also $(g^T)_{T \geq T_0}$ is bounded in $L^\infty(\mathfrak{m})$, thus in $L^p(\mathfrak{m})$ for all $p \in [1, \infty]$. As a direct consequence of this and of the maximum, we deduce that also $(\mathfrak{P}_T f^T)_{T \geq T_0}$ and $(\mathfrak{P}_T g^T)_{T \geq T_0}$ are bounded in $L^\infty(\mathfrak{m})$.

We are now in the position to prove (3.22). Indeed, by Banach-Alaoglu theorem there exist a (not relabeled) subsequence and functions $\tilde{f}, \tilde{g} \in L^\infty(\mathfrak{m})$ such that $f^T \xrightarrow{*} \tilde{f}$ and $g^T \xrightarrow{*} \tilde{g}$ in $L^\infty(\mathfrak{m})$ as $T \rightarrow \infty$. We claim that

$$\lim_{T \rightarrow \infty} \mathfrak{P}_T f^T = \int \tilde{f} d\mathfrak{m} \quad \text{and} \quad \lim_{T \rightarrow \infty} \mathfrak{P}_T g^T = \int \tilde{g} d\mathfrak{m}, \quad (3.25)$$

where both limits have to be understood in $L^p(\mathfrak{m})$ for $p \in [1, \infty)$. To this aim observe that

$$\begin{aligned} \|\mathfrak{P}_T f^T - \int \tilde{f} d\mathfrak{m}\|_{L^p(\mathfrak{m})} &\leq \|\mathfrak{P}_T f^T - \mathfrak{P}_T \tilde{f}\|_{L^p(\mathfrak{m})} + \|\mathfrak{P}_T \tilde{f} - \int \tilde{f} d\mathfrak{m}\|_{L^p(\mathfrak{m})} \\ &\leq \|\mathfrak{P}_{T_0} f^T - \mathfrak{P}_{T_0} \tilde{f}\|_{L^p(\mathfrak{m})} + \|\mathfrak{P}_T \tilde{f} - \int \tilde{f} d\mathfrak{m}\|_{L^p(\mathfrak{m})}, \end{aligned} \quad (3.26)$$

where the second inequality is motivated by the fact that $\mathfrak{P}_T = \mathfrak{P}_{T-T_0} \circ \mathfrak{P}_{T_0}$ and \mathfrak{P}_{T-T_0} is a contraction in $L^p(\mathfrak{m})$. Therefore, arguing as in the proof of Lemma 3.4 we can see that the first term on the right-hand side above vanishes as $T \rightarrow \infty$. On the other hand, the ergodicity of \mathfrak{P}_T and (H2) entail that

$$\lim_{T \rightarrow \infty} \mathfrak{P}_T \tilde{f} = \int \tilde{f} d\mathfrak{m}, \quad \text{in } L^2(\mathfrak{m}), \quad (3.27)$$

hence in $L^p(\mathfrak{m})$ for all $p \in [1, 2]$, while for $p > 2$ it is sufficient to observe that

$$\begin{aligned} \|\mathfrak{P}_T \tilde{f} - \int \tilde{f} d\mathfrak{m}\|_{L^p(\mathfrak{m})}^p &\leq \|\mathfrak{P}_T \tilde{f} - \int \tilde{f} d\mathfrak{m}\|_{L^\infty(\mathfrak{m})}^{p-2} \|\mathfrak{P}_T \tilde{f} - \int \tilde{f} d\mathfrak{m}\|_{L^2(\mathfrak{m})}^2 \\ &\leq (\|\tilde{f}\|_{L^\infty(\mathfrak{m})} + \|\tilde{f}\|_{L^1(\mathfrak{m})})^{p-2} \|\mathfrak{P}_T \tilde{f} - \int \tilde{f} d\mathfrak{m}\|_{L^2(\mathfrak{m})}^2 \end{aligned}$$

and use (3.27). Therefore, also the second term on the second line on the right-hand side in (3.26) converges to 0 as $T \rightarrow \infty$ and this proves the first identity in (3.25). An analogous argument leads to the second one.

As a consequence, up to extract a further (not relabeled) subsequence, the limits in (3.25) hold \mathfrak{m} -a.e. and if we pass to the limit as $T \rightarrow \infty$ in the Schrödinger system (2.9) we get

$$\rho = \tilde{f} \int \tilde{g} d\mathfrak{m}, \quad \sigma = \tilde{g} \int \tilde{f} d\mathfrak{m}$$

and because of (2.10) it must hold $\|\tilde{g}\|_{L^1(\mathfrak{m})} = 1$, whence $\tilde{f} = \rho$ and this in turn implies $\tilde{g} = \sigma$.

It is now sufficient to combine this information with (3.25) to get (3.23). \square

We are now in the position to determine the long-time behavior of the entropic cost.

Proof of Theorem 1.2. Let us first observe that for all $T_0 > 0$ there exists $C > 0$ such that for all $T \geq T_0$ it holds

$$C^{-1} \leq \mathbf{P}_T f^T \leq C, \quad \mathfrak{m}\text{-a.e. in } \text{supp}(\nu), \quad (3.28a)$$

$$C^{-1} \leq \mathbf{P}_T g^T \leq C, \quad \mathfrak{m}\text{-a.e. in } \text{supp}(\mu). \quad (3.28b)$$

Indeed, both upper bounds have been proven in Lemma 3.6, while the lower ones can be deduced relying on (2.8) and the boundedness of $\text{supp}(\mu)$, $\text{supp}(\nu)$. More precisely, for (3.28a) there exists $C > 0$ such that

$$\mathbf{P}_T g^T(x) = \int g^T(y) \mathbf{p}_T(x, y) \, d\mathfrak{m} \geq C \|g^T\|_{L^1(\mathfrak{m})} = C_\delta, \quad \forall x \in \text{supp}(\mu), \forall T \geq T_0,$$

the last identity being motivated by (2.10). For the same reason and taking (3.24) into account

$$\mathbf{P}_T f^T(x) = \int f^T(y) \mathbf{p}_T(x, y) \, d\mathfrak{m} \geq C \|f^T\|_{L^1(\mathfrak{m})} \geq \frac{C}{C'}, \quad \forall x \in \text{supp}(\nu), \forall T \geq T_0,$$

which proves also (3.28b). With this said, by the very definition of the entropic cost and its equivalence with (2.9) we have

$$\mathcal{E}_T(\mu, \nu) = \int \log f^T \, d\mu + \int \log g^T \, d\nu = \mathcal{H}(\mu | \mathfrak{m}) + \mathcal{H}(\nu | \mathfrak{m}) - \int \log \mathbf{P}_T f^T \, d\nu - \int \log \mathbf{P}_T g^T \, d\mu$$

and by (3.28a), (3.28b), (3.23) and the dominated convergence theorem we can pass to the limit as $T \rightarrow \infty$ in the identity above, thus getting

$$\lim_{T \rightarrow \infty} \mathcal{E}_T(\mu, \nu) = \mathcal{H}(\mu | \mathfrak{m}) + \mathcal{H}(\nu | \mathfrak{m}).$$

\square

As concerns Corollary 1.3, let us recall an approximation result, whose proof can be found in [17, Lemma 3.1].

Lemma 3.7. *Let $\mu \in \mathcal{P}(M)$ with $\mathcal{H}(\mu | \mathfrak{m}) < \infty$ and*

$$\int |\nabla \log \left(\frac{d\mu}{d\mathfrak{m}} \right)|^2 d\mu < \infty.$$

Then there exists $(\mu_n) \subset \mathcal{P}(M)$ with $\mu_n = \rho_n \mathfrak{m}$, $\rho_n \in C_c^\infty(M)$ such that $\mathcal{H}(\mu_n | \mathfrak{m}) \rightarrow \mathcal{H}(\mu | \mathfrak{m})$ and

$$\int |\nabla \log \left(\frac{d\mu_n}{d\mathfrak{m}} \right)|^2 d\mu_n \rightarrow \int |\nabla \log \left(\frac{d\mu}{d\mathfrak{m}} \right)|^2 d\mu$$

as $n \rightarrow \infty$.

We can now provide an “entropic” proof of the logarithmic Sobolev inequality.

Proof of Corollary 1.3. First of all, we can assume that the right-hand side in (1.13) is finite, otherwise the statement is trivially true; by Lemma 3.7 we can also assume that μ has compact support, is absolutely continuous w.r.t. \mathbf{m} and its density ρ is smooth. Hence, if we take any probability measure $\nu = \sigma \mathbf{m}$ with compact support and smooth density, (H4) is satisfied. In addition, f^T and g^T are also compactly supported and smooth, as they inherit the regularity of ρ and σ respectively.

After this premise, let us show that $t \mapsto \mathcal{H}(\mu_t^T | \mathbf{m})$ is differentiable at $t = 0$ and

$$\left. \frac{d}{dt} \mathcal{H}(\mu_t^T | \mathbf{m}) \right|_{t=0} = \int \nabla \log \rho \cdot v_0^T d\mu. \quad (3.29)$$

To this aim, let us recall that from [17, Lemma 3.7] for every $\delta > 0$ the map $t \mapsto \mathcal{H}(\mu_t^{T,\delta} | \mathbf{m})$ belongs to $C([0, 1]) \cap C^2((0, 1))$ and it holds

$$\frac{d}{dt} \mathcal{H}(\mu_t^{T,\delta} | \mathbf{m}) = \int \nabla \log \rho_t^{T,\delta} \cdot v_t^{T,\delta} d\mu_t^{T,\delta}, \quad (3.30)$$

where $\rho_t^{T,\delta}$, $\mu_t^{T,\delta}$ and $v_t^{T,\delta}$ are defined as in (2.21). If we integrate (3.30) on $[0, t]$ with $t \leq 1$ we obtain

$$\mathcal{H}(\mu_t^{T,\delta} | \mathbf{m}) - \mathcal{H}(\mu_0^{T,\delta} | \mathbf{m}) = \int_0^t \int \nabla \log \rho_s^{T,\delta} \cdot v_s^{T,\delta} d\mu_s^{T,\delta} ds \quad (3.31)$$

and by the dominated convergence theorem it is easy to see that the left-hand side converges to $\mathcal{H}(\mu_t^T | \mathbf{m}) - \mathcal{H}(\mu | \mathbf{m})$ as $\delta \downarrow 0$. As regards the right-hand one, to prove that

$$\lim_{\delta \downarrow 0} \int_0^t \int \nabla \log \rho_s^{T,\delta} \cdot v_s^{T,\delta} d\mu_s^{T,\delta} ds = \int_0^t \int \nabla \log \rho_s^T \cdot v_s^T d\mu_s^T ds$$

we borrow an argument used in [17, Lemma 3.11] and we report it here for reader's sake. By the very definition of $v_t^{T,\delta}$ and since

$$T \log \rho_t^{T,\delta} = T \log(\mathbb{P}_{Tt} f^T + \delta) + T \log(\mathbb{P}_{T(1-t)} g^T + \delta),$$

the desired conclusion is achieved if we are able to prove that

$$\lim_{\delta \downarrow 0} \int \int_0^t |\nabla \log(\mathbb{P}_{Ts} f^T + \delta)|^2 \rho_s^{T,\delta} ds d\mathbf{m} = \int \int_0^t |\nabla \log \mathbb{P}_{Ts} f^T|^2 \rho_s^T ds d\mathbf{m}, \quad (3.32a)$$

$$\lim_{\delta \downarrow 0} \int \int_0^t |\nabla \log(\mathbb{P}_{T(1-s)} g^T + \delta)|^2 \rho_s^{T,\delta} ds d\mathbf{m} = \int \int_0^t |\nabla \log \mathbb{P}_{T(1-s)} g^T|^2 \rho_s^T ds d\mathbf{m}. \quad (3.32b)$$

To this aim, notice that $\rho_t^{T,\delta} = \rho_t^T + \delta \mathbb{P}_{Tt} f^T + \delta \mathbb{P}_{T(1-t)} g^T + \delta^2$, whence using either $\mathbb{P}_{Tt} f^T + \delta \geq \mathbb{P}_{Tt} f^T$ or $\mathbb{P}_{Tt} f^T + \delta \geq \delta$ it is easy to infer that

$$\begin{aligned} |\nabla \log(\mathbb{P}_{Tt} f^T + \delta)|^2 \rho_t^T &= T^2 \frac{|\nabla \mathbb{P}_{Tt} f^T|^2}{(\mathbb{P}_{Tt} f^T + \delta)^2} \rho_t^T \leq T^2 \frac{|\nabla \mathbb{P}_{Tt} f^T|^2}{(\mathbb{P}_{Tt} f^T)^2} \rho_t^T = |\nabla \log \mathbb{P}_{Tt} f^T|^2 \rho_t^T, \\ \delta |\nabla \log(\mathbb{P}_{Tt} f^T + \delta)|^2 \mathbb{P}_{Tt} f^T &= \delta T^2 \frac{|\nabla \mathbb{P}_{Tt} f^T|^2}{(\mathbb{P}_{Tt} f^T + \delta)^2} \mathbb{P}_{Tt} f^T \leq T^2 |\nabla \mathbb{P}_{Tt} f^T|^2, \\ \delta |\nabla \log(\mathbb{P}_{Tt} f^T + \delta)|^2 \mathbb{P}_{T(1-t)} g^T &= \delta T^2 \frac{|\nabla \mathbb{P}_{Tt} f^T|^2}{(\mathbb{P}_{Tt} f^T + \delta)^2} \mathbb{P}_{T(1-t)} g^T \leq T^2 \frac{|\nabla \mathbb{P}_{Tt} f^T|^2}{\mathbb{P}_{Tt} f^T} \mathbb{P}_{T(1-t)} g^T \\ &= |\nabla \log \mathbb{P}_{Tt} f^T|^2 \rho_t^T, \\ \delta^2 |\nabla \log \mathbb{P}_{Tt} f^T|^2 &= \delta^2 T^2 \frac{|\nabla \mathbb{P}_{Tt} f^T|^2}{(\mathbb{P}_{Tt} f^T + \delta)^2} \leq T^2 |\nabla \mathbb{P}_{Tt} f^T|^2. \end{aligned}$$

All the right-hand sides above are integrable on $[0, 1] \times M$ (either by (2.17) or by the Bakry-Émery contraction estimate (2.3) together with $f^T \in C_c^\infty(M)$), thus by dominated convergence (3.32a) follows. An analogous argument holds for (3.32b). Therefore we can pass to the limit as $\delta \downarrow 0$ in (3.31) and get

$$\mathcal{H}(\mu_t^T | \mathbf{m}) - \mathcal{H}(\mu | \mathbf{m}) = \int_0^t \int \nabla \log \rho_s^T \cdot v_s^T d\mu_s^T ds,$$

whence (3.29). This allows us to differentiate (2.20) at $t = 0$ and obtain

$$e^{-\kappa T} \mathcal{H}(\mu | \mathbf{m}) \leq -\frac{1 - e^{-\kappa T}}{\kappa T} \int \nabla \log \rho \cdot v_0^T d\mu + \mathcal{H}(\nu | \mathbf{m}) - (1 - e^{-\kappa T}) \mathcal{C}_T(\mu, \nu). \quad (3.33)$$

We claim that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int \nabla \log \rho \cdot v_0^T d\mu = -\frac{1}{2} \int |\nabla \log \rho|^2 d\mu \quad (3.34)$$

Indeed,

$$v_0^T = \frac{T}{2} \nabla \log \mathbf{P}_T g^T - \frac{T}{2} \nabla \log f^T = T \nabla \log \mathbf{P}_T g^T - \frac{T}{2} \nabla \log \rho$$

from the very definition of v_0^T , so that

$$\begin{aligned} \frac{1}{T} \int \nabla \log \rho \cdot v_0^T d\mu &= \int \nabla \rho \cdot \nabla \log \mathbf{P}_T g^T d\mathbf{m} - \frac{1}{2} \int |\nabla \log \rho|^2 d\mathbf{m} \\ &= - \int (\mathbf{L}\rho) \log \mathbf{P}_T g^T d\mathbf{m} - \frac{1}{2} \int |\nabla \log \rho|^2 d\mathbf{m}, \end{aligned}$$

where $\mathbf{L} = \Delta/2 - \nabla U \cdot \nabla$ is the generator of \mathbf{P}_t , self-adjoint w.r.t. \mathbf{m} . By (3.28b), (3.23) and the boundedness of $\text{supp}(\mu)$ we infer that the first integral on the right-hand side vanishes as $T \rightarrow \infty$ by dominated convergence, whence (3.34).

From (3.33), (3.34) and Theorem 1.2 the logarithmic Sobolev inequality (1.13) follows. \square

3.4 Proof of Theorem 1.4

The proof of Theorem 1.4 relies on Theorem 1.1 and two other ingredients: the entropic Talagrand inequality put forward in [9] and an “energy-transport” inequality relating $\mathcal{E}_T(\mu, \nu)$ and $\mathcal{C}_T(\mu, \nu)$. The former states that if (H1) holds with $\kappa > 0$, then for all μ, ν as in (H3) and for all $t \in (0, 1)$

$$\mathcal{E}_T(\mu, \nu) \leq \frac{1}{1 - \exp(-\kappa T t)} \mathcal{H}(\mu | \mathbf{m}) + \frac{1}{1 - \exp(-\kappa T (1 - t))} \mathcal{H}(\nu | \mathbf{m}). \quad (3.35)$$

Truth to be told, as for (2.20) also the entropic Talagrand inequality (3.35) was stated in [9] in the framework of compact Riemannian manifolds satisfying (H1) with $\kappa > 0$ and endowed with the volume measure, but since (3.35) is deduced from (2.20) and the latter has been generalized to the present framework in Section 2, there is no problem in applying it.

The “energy-transport” inequality relating $\mathcal{E}_T(\mu, \nu)$ and $\mathcal{C}_T(\mu, \nu)$ is expressed in the following

Lemma 3.8 (Energy-Transport inequality). *Assume that (H1) and (H4) hold. Then for all $T > 0$ it holds*

$$|\mathcal{E}_T(\mu, \nu)| \leq \frac{\kappa}{\exp(\kappa T/2) - 1} \sqrt{(\mathcal{C}_T(\mu, \nu) - \mathcal{H}(\mu | \mathbf{m}))(\mathcal{C}_T(\mu, \nu) - \mathcal{H}(\nu | \mathbf{m}))} \quad (3.36)$$

if $\kappa \neq 0$ and

$$|\mathcal{E}_T(\mu, \nu)| \leq \frac{1}{T} \sqrt{(\mathcal{C}_T(\mu, \nu) - \mathcal{H}(\mu | \mathbf{m}))(\mathcal{C}_T(\mu, \nu) - \mathcal{H}(\nu | \mathbf{m}))} \quad (3.37)$$

if $\kappa = 0$.

Proof. From the second identity in (2.18) we have

$$|\mathcal{E}_T(\mu, \nu)| = \int_0^1 |\mathcal{E}_T(\mu, \nu)| dt \leq \frac{1}{2} \iint_0^1 |\nabla \log P_{T(1-t)} g^T| |\nabla \log P_{Tt} f^T| \rho_t^T dt d\mathbf{m}, \quad (3.38)$$

so that by Cauchy-Schwarz inequality and by (2.16), (3.37) follows.

As regards (3.36), using the same notations as in (2.21), let us first point out that, being $\rho_t^{T,\delta}$ and $\nu_t^{T,\delta}$ optimal for SP with marginals $\mu_0^{T,\delta}$ and $\mu_1^{T,\delta}$, we have

$$\mathcal{E}_T(\mu_0^{T,\delta}, \mu_1^{T,\delta}) = -\frac{1}{2} \int \nabla \log(P_{T(1-t)} g^T + \delta) \cdot \nabla \log(P_{Tt} f^T + \delta) d\mu_t^{T,\delta}, \quad (3.39)$$

so that if we define

$$\Phi(t) := \int |\nabla \log(P_{Tt} f^T + \delta)|^2 d\mu_t^{T,\delta}, \quad \Psi(t) := \int |\nabla \log(P_{T(1-t)} g^T + \delta)|^2 d\mu_t^{T,\delta}$$

for $\delta > 0$, then by (3.39) evaluated in $t = 1/2$ and by Cauchy-Schwarz inequality

$$|\mathcal{E}_T(\mu_0^{T,\delta}, \mu_1^{T,\delta})| \leq \frac{1}{2} \sqrt{\Phi(1/2)\Psi(1/2)}. \quad (3.40)$$

In order to estimate the right-hand side above, observe that by (2.23)

$$\Phi(t) = -\frac{2}{T} h'_f(t), \quad \Psi(t) = \frac{2}{T} h'_b(t)$$

with h_f, h_b defined as in (2.22), so that by Lemma 2.2

$$\Phi'(t) \leq -\kappa T \Phi(t), \quad \Psi'(t) \geq \kappa T \Psi(t).$$

Hence, by Grönwall's lemma

$$\Phi(t) \geq \exp(\kappa T(1/2 - t)) \Phi(1/2), \quad \forall t \in [0, 1/2], \quad (3.41a)$$

$$\Psi(t) \geq \exp(\kappa T(t - 1/2)) \Psi(1/2), \quad \forall t \in [1/2, 1]. \quad (3.41b)$$

On the one hand, by (2.16) and (3.41b)

$$\begin{aligned} \mathcal{E}_T(\mu_0^{T,\delta}, \mu_1^{T,\delta}) - \mathcal{H}(\mu_0^{T,\delta} | \mathbf{m}) &\geq \frac{T}{2} \int_{1/2}^1 \Psi(t) dt \geq \frac{T}{2} \Psi(1/2) \int_{1/2}^1 \exp(\kappa T(t - 1/2)) dt \\ &= \frac{1}{2\kappa} \Psi(1/2) (\exp(\kappa T/2) - 1), \end{aligned}$$

while on the other hand, by (3.41a) and a completely analogous argument

$$\mathcal{E}_T(\mu_0^{T,\delta}, \mu_1^{T,\delta}) - \mathcal{H}(\mu_1^{T,\delta} | \mathbf{m}) \geq \frac{1}{2\kappa} \Phi(1/2)(\exp(\kappa T/2) - 1).$$

Combining these inequalities with (3.40), we obtain

$$|\mathcal{E}_T(\mu_0^{T,\delta}, \mu_1^{T,\delta})| \leq \frac{\kappa}{\exp(\kappa T/2) - 1} \sqrt{(\mathcal{E}_T(\mu_0^{T,\delta}, \mu_1^{T,\delta}) - \mathcal{H}(\mu_0^{T,\delta} | \mathbf{m}))(\mathcal{E}_T(\mu_0^{T,\delta}, \mu_1^{T,\delta}) - \mathcal{H}(\mu_1^{T,\delta} | \mathbf{m}))}$$

and it is now sufficient to pass to the limit as $\delta \downarrow 0$ to get the conclusion. The fact that $\mathcal{H}(\mu_0^{T,\delta} | \mathbf{m}) \rightarrow \mathcal{H}(\mu | \mathbf{m})$ and $\mathcal{H}(\mu_1^{T,\delta} | \mathbf{m}) \rightarrow \mathcal{H}(\nu | \mathbf{m})$ is motivated by dominated convergence; the same is true for $\mathcal{E}_T(\mu_0^{T,\delta}, \mu_1^{T,\delta}) \rightarrow \mathcal{E}_T(\mu, \nu)$ taking (3.39) into account. \square

Remark 3.9. Let us point out that the bound (3.37) is better than the one obtained from (3.36) via a Taylor expansion around $\kappa = 0$, so that (3.36) is not sharp and the reader might wonder whether (3.36) could be improved by replacing $\exp(\kappa T/2)$ with $\exp(\kappa T)$. A posteriori, this is not possible because if (3.36) held with $\exp(\alpha \kappa T)$ instead of $\exp(\kappa T/2)$ for some $\alpha > 1/2$, then this would contradict the sharpness of (1.14) as stated in Theorem 1.4. \blacksquare

We are now in the position to prove Theorem 1.4.

Proof of Theorem 1.4. From Theorem 1.1 and Theorem 1.2 we have that

$$|\mathcal{E}_T(\mu, \nu) - \mathcal{H}(\mu | \mathbf{m}) - \mathcal{H}(\nu | \mathbf{m})| = \left| \int_T^\infty \mathcal{E}_S(\mu, \nu) \, dS \right| \leq \int_T^\infty |\mathcal{E}_S(\mu, \nu)| \, dS \quad (3.42)$$

and by Lemma 3.8 the right-hand side above is controlled as follows

$$\int_T^\infty |\mathcal{E}_S(\mu, \nu)| \, dS \leq \kappa \int_T^\infty \frac{\sqrt{\mathcal{E}_S(\mu, \nu) - \mathcal{H}(\mu | \mathbf{m})} \sqrt{\mathcal{E}_S(\mu, \nu) - \mathcal{H}(\nu | \mathbf{m})}}{\exp(\kappa S/2) - 1} \, dS. \quad (3.43)$$

By the entropic Talagrand inequality (3.35) for $t = 1/2$ and by (H1) with $\kappa > 0$, which implies $\mathcal{H}(\cdot | \mathbf{m}) \geq 0$, it holds

$$\begin{aligned} \mathcal{E}_S(\mu, \nu) - \mathcal{H}(\mu | \mathbf{m}) &\leq \exp(\kappa S/2) \frac{\exp(-\kappa S/2) \mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m})}{\exp(\kappa S/2) - 1} \\ &\leq \exp(\kappa S/2) \frac{\exp(-\kappa T/2) \mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m})}{\exp(\kappa S/2) - 1} \end{aligned}$$

for all $S \geq T$ and an analogous inequality holds with μ and ν swapped. Therefore the right-hand side in (3.43) is bounded from above by

$$\kappa \sqrt{\exp(-\kappa T/2) \mathcal{H}(\mu | \mathbf{m}) + \mathcal{H}(\nu | \mathbf{m})} \sqrt{\exp(-\kappa T/2) \mathcal{H}(\nu | \mathbf{m}) + \mathcal{H}(\mu | \mathbf{m})} \int_T^\infty \frac{\exp(\kappa S/2)}{(\exp(\kappa S/2) - 1)^2} \, dS.$$

The bound (1.16a) follows by calculating explicitly the integral term and this, in turn, trivially implies (1.14).

For the sharpness of (1.14), we shall prove that there exists a triplet $(M', \mathbf{d}'_g, \mathbf{m}')$ satisfying (H1) with $\kappa > 0$ and $\mu, \nu \in \mathcal{P}(M')$ satisfying (H4) such that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \log |\mathcal{E}_T(\mu, \nu) - \mathcal{H}(\mu | \mathbf{m}) - \mathcal{H}(\nu | \mathbf{m})| \geq -\frac{\kappa}{2}. \quad (3.44)$$

To this aim, consider $M' = \mathbb{R}$, d'_g the Euclidean distance and choose $U(x) := \kappa x^2/4$, so that $\text{Hess}(2U) = \kappa$, \mathbf{m}' is the Gaussian measure

$$\mathbf{m}'(dx) = \sqrt{\frac{\kappa}{2\pi}} \exp\left(-\frac{\kappa}{2}x^2\right)dx \quad (3.45)$$

and the stochastic process associated to the SDE (1.1) is the stationary Ornstein-Uhlenbeck process, whence by well-known results (see e.g. [5]) its transition probabilities w.r.t. \mathbf{m}' admit the following explicit representation

$$\mathbf{p}_t(x, y) = \frac{1}{\sqrt{1 - \exp(-\kappa t)}} \exp\left(-\frac{\kappa(x^2 - 2\exp(\kappa t/2)xy + y^2)}{2(\exp(\kappa t) - 1)}\right). \quad (3.46)$$

With this said, choose $\mu, \nu \in \mathcal{P}(\mathbb{R})$ with Lipschitz densities and compact supports with the further condition that

$$\int x \mu(dx) = -\int y \nu(dy) = 1 \quad \text{and} \quad \int x^2 \mu(dx) = \int y^2 \nu(dy) = \sigma^2$$

for some $\sigma > 0$. From the very definition of the entropic cost (1.2) and the identity $R_{0T}(dxdy) = \mathbf{p}_T(x, y)\mathbf{m}(dx)\mathbf{m}(dy)$ we have

$$\mathcal{C}_T(\mu, \nu) = \inf_{\gamma \in \Pi(\mu, \nu)} \left\{ \mathcal{H}(\gamma | \mathbf{m}' \otimes \mathbf{m}') - \int \log \mathbf{p}_T d\gamma \right\}$$

and since $\int x^2 \mu(dx) = \int y^2 \nu(dy) = \sigma^2$ by construction, if we take also (3.46) into account, then the above expression can be rewritten as

$$\begin{aligned} \mathcal{C}_T(\mu, \nu) &= \inf_{\gamma \in \Pi(\mu, \nu)} \left\{ \mathcal{H}(\gamma | \mathbf{m}' \otimes \mathbf{m}') - \frac{\kappa \exp(-\kappa T/2)}{2(1 - \exp(-\kappa T))} \int xy \gamma(dxdy) \right\} \\ &\quad + \underbrace{\frac{\kappa \sigma^2}{\exp(\kappa T) - 1} + \frac{1}{2} \log(1 - \exp(-\kappa T))}_{:=\beta(T)}. \end{aligned}$$

Now observe that there exists a unique optimal $\gamma^T \in \Pi(\mu, \nu)$ in the minimization problem above, as this is nothing but a rephrasing of (1.2), and $\mathcal{H}(\gamma^T | \mathbf{m}' \otimes \mathbf{m}') \geq \mathcal{H}(\mu | \mathbf{m}') + \mathcal{H}(\nu | \mathbf{m}')$, because $\mu \otimes \nu$ is the unique minimizer in (3.7). Moreover, by Lemma 3.1 and Theorem 1.2 we know that $\gamma^T \rightarrow \mu \otimes \nu$ as $T \rightarrow \infty$ and, together with the fact that μ, ν have compact support (whence a fortiori γ^T in $\mathbb{R} \times \mathbb{R}$ too), this implies that

$$\lim_{T \rightarrow \infty} \int xy \gamma^T(dxdy) = \int x \mu(dx) \int y \nu(dy),$$

so that, thanks to the fact that $\int x \mu(dx) = 1$ and $\int y \nu(dy) = -1$, it holds $\int xy \gamma^T(dxdy) \leq -1/2$ for T large enough. Finally, note that there exists $C > 0$ such that $|\beta(T)| \leq C \exp(-\kappa T)$ for all T sufficiently large. Combining all these ingredients, we deduce that

$$\mathcal{C}_T(\mu, \nu) - \mathcal{H}(\mu | \mathbf{m}') - \mathcal{H}(\nu | \mathbf{m}') \geq \frac{\kappa \exp(-\kappa T/2)}{4(1 - \exp(-\kappa T))} - C \exp(-\kappa T)$$

and it is now sufficient to remark that the right-hand side is asymptotically positive, so that for T large enough we are allowed to apply the logarithm to both sides of the inequality above, whence (3.44).

As regards (1.15), it is sufficient to estimate the right-hand side in (3.36) by following the same reasoning as above. To prove that also (1.15) is sharp, it is easy to see that if there exists $\alpha > 1/2$ such that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \log |\mathcal{E}_T(\mu, \nu)| \leq -\alpha\kappa$$

holds for all μ, ν satisfying (H4), then for any such μ, ν there exist $C, T_0 > 0$ sufficiently large such that

$$|\mathcal{E}_T(\mu, \nu)| \leq C \exp(-\alpha\kappa T), \quad \forall T \geq T_0$$

and if we plug this inequality into (3.42), instead of Lemma 3.8, then we get $|\mathcal{E}_T(\mu, \nu) - \mathcal{H}(\mu | \mathbf{m}) - \mathcal{H}(\nu | \mathbf{m})| \leq C \exp(-\alpha\kappa T)$, whence

$$\lim_{T \rightarrow \infty} \frac{1}{T} \log |\mathcal{E}_T(\mu, \nu) - \mathcal{H}(\mu | \mathbf{m}) - \mathcal{H}(\nu | \mathbf{m})| \leq -\alpha\kappa$$

and this clearly contradicts the sharpness of (1.14). \square

3.5 Proof of Theorem 1.7

The proof relies on the results of [3] and we shall borrow some notation from the above mentioned work. In order to keep the size of the present work under control, we prefer not to reintroduce all the notation, but rather to give precise references to the place where this is done in [3]. In particular, Theorem 1.3 therein ensures that to any optimizer P for (MFSP) there is an associated stochastic process $(\Psi_t(X_t))_{t \in [0, T]}$ such that

$$\Psi_t(X_t) - \int_0^t \tilde{\mathbb{E}}_{\hat{P}} \left[\nabla^2 W(X_s - \tilde{X}_s) \cdot (\Psi_s(X_s) - \Psi_s(\tilde{X}_s)) \right] ds \quad (3.47)$$

is a continuous martingale under P on $[0, T[$, where $(\tilde{X}_t)_{t \in [0, T]}$ is an independent copy of $(X_t)_{t \in [0, T]}$ and $\tilde{\mathbb{E}}_{\hat{P}}$ the corresponding expectation. In [3, Lemma 4.5] it is proven that if P is an optimizer for (MFSP), then its *time reversal* \hat{P} , i.e. the law of $(X_{T-t})_{t \in [0, T]}$, is optimal for

$$\inf \left\{ \mathcal{H}(Q | \Gamma(Q)) : Q \in \mathcal{P}_1(\Omega), Q_0 = \nu, Q_T = \mu \right\}. \quad (3.48)$$

The optimality of \hat{P} implies the existence of an associated process $\hat{\Psi}$ as in (3.47), explicitly described by the relation

$$\hat{\Psi}_t(X_t) = -\Psi_{T-t}(X_t) + \nabla \log P_{T-t}(X_t) + 2\nabla W * P_{T-t}(X_t), \quad \hat{P}\text{-a.s.} \quad (3.49)$$

for all $t \in [0, T]$, see [3, Eq. 92] and also [14].

After this premise, the first ingredient needed for the proof of Theorem 1.7 is the following

Lemma 3.10. *Under the same assumptions as in Theorem 1.7, let P be optimal for (MFSP). Then*

$$\mathcal{E}_T^{mf}(\mu, \nu) = \mathcal{F}(\mu) + \mathcal{F}(\nu) + \frac{1}{2} \mathbb{E}_P \left[\int_0^{T/2} |\Psi_t(X_t)|^2 dt \right] + \frac{1}{2} \mathbb{E}_{\hat{P}} \left[\int_0^{T/2} |\hat{\Psi}_t(X_t)|^2 dt \right] - \mathcal{F}(P_{T/2})$$

Proof. Using the time-reversal relation (3.49) we deduce that

$$\begin{aligned} \frac{1}{2} \mathbb{E}_{\hat{\mathbb{P}}} \left[\int_0^{T/2} |\hat{\Psi}_t(X_t)|^2 dt \right] &= \frac{1}{2} \mathbb{E}_{\hat{\mathbb{P}}} \left[\int_0^{T/2} |\Psi_{T-t}(X_t)|^2 dt \right] \\ &\quad - \mathbb{E}_{\hat{\mathbb{P}}} \left[\int_0^{T/2} \left(\Psi_{T-t} - \frac{1}{2} \chi_{T-t} \right) \cdot \chi_{T-t}(X_t) dt \right] \end{aligned}$$

where

$$\chi_t := \nabla \log P_t + 2 \nabla W * P_t.$$

Using the change of variable $s = T - t$ and the very definition of time reversal, we rewrite the right-hand side as

$$\frac{1}{2} \mathbb{E}_{\mathbb{P}} \left[\int_{T/2}^T |\Psi_s(X_s)|^2 ds \right] - \mathbb{E}_{\mathbb{P}} \left[\int_{T/2}^T \left(\Psi_s - \frac{1}{2} \chi_s \right) \cdot \chi_s(X_s) ds \right]$$

and using [3, Lemma 4.4 (ii)] we find that the second member in the above expression is worth $-\mathcal{F}(\nu) + \mathcal{F}(P_{T/2})$. Because of the mean field control formulation [3, Lemma 1.1 and Theorem 1.3] of MFSP we have

$$\mathcal{C}_T^{mf}(\mu, \nu) = \mathcal{F}(\mu) + \frac{1}{2} \mathbb{E}_{\mathbb{P}} \left[\int_0^T |\Psi_t(X_t)|^2 dt \right],$$

so that the conclusion is easily obtained. \square

As for the proof of Theorem 1.4, also in this case we rely on an entropic Talagrand inequality. In the mean field setting (and taking into account Remark 1.9) this reads as follows

$$\mathcal{C}_T^{mf}(\mu, \nu) \leq \frac{1}{1 - \exp(-\kappa T t)} \mathcal{F}(\mu) + \frac{1}{1 - \exp(-\kappa T (1 - t))} \mathcal{F}(\nu) \quad (3.50)$$

for all $t \in (0, 1)$ and μ, ν satisfying (H'3), see [3, Corollary 1.3].

We are now in position to prove Theorem 1.7.

Proof. From the previous lemma we get that

$$-\mathcal{F}(P_{T/2}) \leq \mathcal{C}_T^{mf}(\mu, \nu) - \mathcal{F}(\mu) - \mathcal{F}(\nu) \leq \frac{1}{2} \mathbb{E}_{\mathbb{P}} \left[\int_0^{T/2} |\Psi_t(X_t)|^2 dt \right] + \frac{1}{2} \mathbb{E}_{\hat{\mathbb{P}}} \left[\int_0^{T/2} |\hat{\Psi}_t(X_t)|^2 dt \right]. \quad (3.51)$$

Now observe that, thanks to the fact that $\hat{\mathbb{P}}$ is optimal for (3.48), both terms on the right-hand side in (3.51) can be bounded using first [3, Lemma 4.3 Eq. 73] and then the entropic Talagrand inequality (3.50) for $t = 1/2$, keeping in mind Remark 1.9 and that the mean field entropic cost defined in this paper is related to the one defined in [3] via the identity (1.20). We thus obtain, after some simple calculations

$$\frac{1}{2} \mathbb{E}_{\mathbb{P}} \left[\int_0^{T/2} |\Psi_t(X_t)|^2 dt \right] + \frac{1}{2} \mathbb{E}_{\hat{\mathbb{P}}} \left[\int_0^{T/2} |\hat{\Psi}_t(X_t)|^2 dt \right] \leq \frac{1}{\exp(\kappa T/2) - 1} (\mathcal{F}(\mu) + \mathcal{F}(\nu)),$$

which gives the desired upper bound for $\mathcal{C}_T^{mf}(\mu, \nu) - \mathcal{F}(\mu) - \mathcal{F}(\nu)$. For the lower bound, the conclusion directly follows from [3, Theorem 1.4] and Remark 1.9, which allow to estimate $\mathcal{F}(P_{T/2})$. Putting together the results obtained for the upper and lower bound we conclude. \square

A On the sharpness of the entropic Talagrand inequality

Under (H1) with $\kappa > 0$ it holds $\mathbf{m} \in \mathcal{P}_2(M)$, so that \mathbf{m} satisfies (H3) and it is then licit to choose $\nu = \mathbf{m}$ in (3.35), which gives rise to the following version of the entropic Talagrand inequality, closer to the classical Talagrand inequality known in optimal transport:

$$\mathcal{C}_T(\mu, \mathbf{m}) \leq \frac{1}{1 - \exp(-\kappa T)} \mathcal{H}(\mu | \mathbf{m}) \quad (\text{A.1})$$

for all μ as in (H3). In [9] neither the sharpness of (A.1) nor the one of (3.35) was investigated. Aim of this appendix is to remedy this lack.

Theorem A.1. *Assume that (H1) holds with $\kappa > 0$. Then the entropic Talagrand inequality (A.1) is sharp, since there exists a triplet $(M', \mathbf{d}'_g, \mathbf{m}')$ satisfying (H1) with $\kappa > 0$ such that*

$$\mathcal{C}_T(\mu, \mathbf{m}) \leq C \mathcal{H}(\mu | \mathbf{m}), \quad \forall \mu \in \mathcal{P}(M) \text{ satisfying (H3)} \quad (\text{A.2})$$

does not hold for any

$$C < \frac{1}{1 - \exp(-\kappa T)}.$$

As a byproduct, also the entropic Talagrand inequality (3.35) is sharp in the following sense: there exists a triplet $(M', \mathbf{d}'_g, \mathbf{m}')$ satisfying (H1) with $\kappa > 0$ such that

$$\mathcal{C}_T(\mu, \nu) \leq \frac{1}{1 - \exp(-\alpha T t)} \mathcal{H}(\mu | \mathbf{m}) + \frac{1}{1 - \exp(-\alpha T (1 - t))} \mathcal{H}(\nu | \mathbf{m}),$$

$$\forall t \in (0, 1), \forall \mu, \nu \in \mathcal{P}(M) \text{ satisfying (H3)}$$

does not hold for any $\alpha > \kappa$.

Proof. As a first step, we shall prove that (A.2) is equivalent to

$$\int \exp\left(-\frac{1}{TC}\phi\right) d\mathbf{m} \leq \exp\left(-\frac{1}{TC} \int \mathbf{Q}_1^T \phi d\mathbf{m}\right), \quad \forall \phi \in C_b(M), \quad (\text{A.3})$$

where $\mathbf{Q}_1^T \phi := -T \log \mathbf{P}_{Tt}(\exp(-\phi/T))$. To this aim, notice that by (2.19) and the symmetry of the entropic cost, i.e. $\mathcal{C}_T(\mu, \mathbf{m}) = \mathcal{C}_T(\mathbf{m}, \mu)$, (A.2) is equivalent to

$$\int \mathbf{Q}_1^T \phi d\mathbf{m} - \int \phi d\mu - TC \mathcal{H}(\mu | \mathbf{m}) \leq 0 \quad (\text{A.4})$$

for all $\mu \in \mathcal{P}(M)$ satisfying (H3) and for all $\phi \in C_b(M)$. Then observe that

$$\log \int \exp(\psi) d\mathbf{m} = \sup \left\{ \int \psi d\mu - \mathcal{H}(\mu | \mathbf{m}) : \mu \in \mathcal{P}(M), \int \psi^+ d\mu < \infty \right\},$$

where $\psi^+ := \max\{\psi, 0\}$, the identity being a byproduct of the well-known variational representation of the entropy (see [20, 24]). Hence by taking the supremum over μ in (A.4), we get the equivalent formulation

$$\int \mathbf{Q}_1^T \phi d\mathbf{m} + TC \log \int \exp\left(-\frac{1}{TC}\phi\right) d\mathbf{m} \leq 0, \quad \forall \phi \in C_b(M),$$

which is clearly equivalent to (A.3) by algebraic manipulations.

Relying on this characterization, we are now in the position to prove the sharpness of (A.1). As in the proof of Theorem 1.4, consider the real line endowed with the Euclidean distance and choose $U(x) := \kappa x^2/4$, so that $\text{Hess}(2U) = \kappa$, \mathbf{m} is the Gaussian measure defined in (3.45) and the stochastic process associated to the SDE (1.1) is the stationary Ornstein-Uhlenbeck process, whence the explicit representation for its transition probabilities $\mathbf{p}_t(x, y)$ described in (3.46). With this said, assume that (A.2) holds for some constant $C > 0$, hence (A.3) as well. By standard approximation arguments, in (A.3) $C_b(\mathbb{R})$ can be replaced by continuous functions with at most linear growth at infinity, so that

$$\int \exp\left(-\frac{1}{TC}\phi\right) d\mathbf{m} \leq \exp\left(-\frac{1}{TC} \int \mathbf{Q}_1^T \phi d\mathbf{m}\right) \quad (\text{A.5})$$

holds for $\phi(x) := \alpha x$. On the one hand

$$\int \exp\left(-\frac{1}{TC}\phi\right) d\mathbf{m} = \exp\left(\frac{\alpha^2}{2\kappa T^2 C^2}\right),$$

while on the other hand

$$\mathbf{Q}_1^T \phi(x) = -T \log \int \exp\left(-\frac{\alpha}{T}y\right) \mathbf{p}_T(x, y) \mathbf{m}(dy) = x\alpha \exp\left(-\frac{\kappa T}{2}\right) - \alpha^2 \frac{1 - \exp(-\kappa T)}{2\kappa T},$$

so that

$$\exp\left(-\frac{1}{TC} \int \mathbf{Q}_1^T \phi d\mathbf{m}\right) = \exp\left(\alpha^2 \frac{1 - \exp(-\kappa T)}{2\kappa T^2 C}\right).$$

Combining these identities with (A.5) we obtain

$$\frac{\alpha^2}{2\kappa T^2 C^2} \leq \alpha^2 \frac{1 - \exp(-\kappa T)}{2\kappa T^2 C}$$

and this inequality is satisfied if and only if $C \geq (1 - \exp(-\kappa T))^{-1}$, as claimed.

The sharpness of (3.35) immediately follows from the one of (A.1). Indeed, if there exists $\alpha > \kappa$ such that

$$\mathcal{E}_T(\mu, \nu) \leq \frac{1}{1 - \exp(-\alpha T t)} \mathcal{H}(\mu | \mathbf{m}) + \frac{1}{1 - \exp(-\alpha T(1-t))} \mathcal{H}(\nu | \mathbf{m})$$

holds for all $\mu, \nu \in \mathcal{P}(M)$ satisfying (H3), then by choosing $\nu = \mathbf{m}$ and letting $t \rightarrow 1$ we get a contradiction. \square

Acknowledgements. The second author gratefully acknowledges support by the European Union through the ERC-AdG ‘‘RicciBounds’’ for Prof. K. T. Sturm.

References

- [1] S. ADAMS, N. DIRR, M. A. PELETIER, AND J. ZIMMER, *From a Large-Deviations Principle to the Wasserstein Gradient Flow: A New Micro-Macro Passage*, Communications in Mathematical Physics, 307 (2011), p. 791.
- [2] L. AMBROSIO, N. GIGLI, AND G. SAVARÉ, *Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below*, Invent. Math., 195 (2014), pp. 289–391.

- [3] J. BACKHOFF, G. CONFORTI, I. GENTIL, AND C. LÉONARD, *The mean field Schrödinger problem: ergodic behavior, entropy estimates and functional inequalities*, Probability Theory and Related Fields, (2020), pp. 1–56.
- [4] D. BAKRY, *Functional inequalities for Markov semigroups*, in Probability measures on groups: recent directions and trends, Tata Inst. Fund. Res., Mumbai, 2006, pp. 91–147.
- [5] D. BAKRY, I. GENTIL, AND M. LEDOUX, *Analysis and geometry of Markov diffusion operators*, vol. 348 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer, Cham, 2014.
- [6] 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.
- [7] Y. CHEN, T. T. GEORGIU, 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.
- [8] L. CHIZAT, P. ROUSSILLON, F. LÉGER, F.-X. VIALARD, AND G. PEYRÉ, *Faster Wasserstein Distance Estimation with the Sinkhorn Divergence*, Preprint, arXiv:2006.08172, (2020).
- [9] G. CONFORTI, *A second order equation for Schrödinger bridges with applications to the hot gas experiment and entropic transportation cost*, Probability Theory and Related Fields, 174 (2019), pp. 1–47.
- [10] S. DI MARINO AND A. GEROLIN, *An optimal transport approach for the schrödinger bridge problem and convergence of sinkhorn algorithm*, arXiv preprint arXiv:1911.06850, (2019).
- [11] M. H. DUONG, V. LASCHOS, AND M. RENGER, *Wasserstein gradient flows from large deviations of many-particle limits*, ESAIM: Control, Optimisation and Calculus of Variations, 19 (2013), pp. 1166–1188.
- [12] M. ERBAR, J. MAAS, AND M. RENGER, *From large deviations to wasserstein gradient flows in multiple dimensions*, Electron. Commun. Probab., 20 (2015), p. 12 pp.
- [13] J. FEYDY, T. SÉJOURNÉ, F.-X. VIALARD, S.-I. AMARI, A. TROUVÉ, AND G. PEYRÉ, *Interpolating between optimal transport and mmd using sinkhorn divergences*, Preprint, arXiv:1810.08278, (2018).
- [14] H. FÖLLMER, *Time reversal on wiener space*, in Stochastic Processes Mathematics and Physics, Springer, 1986, pp. 119–129.
- [15] 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.
- [16] ———, *Dynamical aspects of generalized schrödinger problem via Otto calculus - A heuristic point of view*, Preprint, arXiv:1806.01553, (2018).
- [17] I. GENTIL, C. LÉONARD, L. RIPANI, AND L. TAMANINI, *An entropic interpolation proof of the HWI inequality*, Stochastic Processes and their Applications, 130 (2020), pp. 907–923.
- [18] N. GIGLI AND L. TAMANINI, *Second order differentiation formula on $RCD^*(K, N)$ spaces*. Accepted at JEMS, arXiv:1802.02463, (2018).
- [19] N. GIGLI AND L. TAMANINI, *Benamou-Brenier and duality formulas for the entropic cost on $RCD^*(K, N)$ spaces*, Probability Theory and Related Fields, (2019), pp. 1–34.
- [20] N. GOZLAN AND C. LÉONARD, *Transport inequalities. A survey*, Markov Processes and Related Fields, 16 (2010), pp. 635–736.
- [21] A. GRIGOR'YAN, *Heat kernel and analysis on manifolds*, vol. 47, American Mathematical Soc., 2009.

- [22] B. L. KOTSCHWAR, *Hamilton's gradient estimate for the heat kernel on complete manifolds*, Proc. Amer. Math. Soc., 135 (2007), pp. 3013–3019.
- [23] C. LÉONARD, *From the Schrödinger problem to the Monge-Kantorovich problem*, J. Funct. Anal., 262 (2012), pp. 1879–1920.
- [24] ———, *Some properties of path measures*, Séminaire de Probabilités XLVI, (2014), pp. 207–230.
- [25] ———, *A survey of the Schrödinger problem and some of its connections with optimal transport*, Discrete Contin. Dyn. Syst., 34 (2014), pp. 1533–1574.
- [26] 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.
- [27] F. OTTO, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations, 26 (2001), pp. 101–174.
- [28] 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.
- [29] S. PAL, *On the difference between entropic cost and the optimal transport cost*, Preprint, arXiv:1905.12206, (2019).
- [30] E. SCHRÖDINGER, *Über die Umkehrung der Naturgesetze*, Verlag der Akademie der Wissenschaften in Kommission bei Walter De Gruyter, 1931.
- [31] E. SCHRÖDINGER, *Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique*, Ann. Inst. H. Poincaré, 2 (1932), pp. 269–310.
- [32] K.-T. STURM, *On the geometry of metric measure spaces. I*, Acta Math., 196 (2006), pp. 65–131.
- [33] L. TAMANINI, *From Harnack inequality to heat kernel estimates on metric measure spaces and applications*, Preprint, arXiv:1907.07163, (2019).
- [34] ———, *The Schrödinger problem on RCD spaces*, work in progress, (2020).
- [35] F.-Y. WANG, *Equivalent semigroup properties for the curvature-dimension condition*, Bulletin des sciences mathématiques, 135 (2011), pp. 803–815.