



# Entropy-minimizing dynamical transport on Riemannian manifolds

Gabriele Bocchi<sup>1</sup> · Alessio Porretta<sup>1</sup>

Received: 9 January 2024 / Accepted: 20 December 2024 / Published online: 9 January 2025  
© The Author(s) 2024

## Abstract

Given a smooth Riemannian manifold  $(M, g)$ , compact and without boundary, we analyze the dynamical optimal mass transport problem where the cost is given by the sum of the kinetic energy and the relative entropy with respect to a reference volume measure  $e^{-V} dx$ . Under the only assumption that the prescribed marginals lie in  $L^1(M)$ , and a lower bound on the Ricci curvature, we characterize the minimal curves as unique weak solutions of the optimality system coupling the continuity equation with a backward Hamilton–Jacobi equation (with source given by the logarithm of the density). We give evidence that the entropic cost enhances diffusive effects in the evolution of the optimal densities, proving  $L^1 \rightarrow L^\infty$  regularization in time for any initial-terminal data, and smoothness of the solutions whenever the marginals are positive and smooth. We use displacement convexity arguments (in the Eulerian approach) and gradient bounds from quasilinear elliptic equations. We also prove the convergence of optimal curves towards the classical Wasserstein geodesics, as the entropic term is multiplied by a vanishing parameter, showing that this kind of functionals can be used to build a smoothing approximation of the standard optimal transport related to the Wasserstein distance.

**Mathematics Subject Classification** 35F20 · 49J20

## 1 Introduction

Let  $(M, g)$  be a smooth, connected  $d$ -dimensional Riemannian manifold, assumed to be compact and without boundary, endowed with a metric tensor  $g = (g_{ij})$  and a volume form  $dx$ . We denote by  $\mathcal{P}(M)$  the set of probability measures on  $M$ . Given a time horizon  $T > 0$

---

Communicated by A. Mondino.

Gabriele Bocchi's research is supported by the Department of Mathematics, University of Rome Tor Vergata. Alessio Porretta's research is supported by the Department of Mathematics, University of Rome Tor Vergata.

---

✉ Alessio Porretta  
porretta@mat.uniroma2.it

Gabriele Bocchi  
bocchi@mat.uniroma2.it

<sup>1</sup> Department of Mathematics, University of Rome Tor Vergata, Via della Ricerca Scientifica 1, 00133 Rome, Italy

and two fixed (initial and terminal) measures  $m_0, m_1 \in \mathcal{P}(M)$ , we analyze in this note the optimal transport problem

$$\min \{ \mathcal{F}_\varepsilon(m, v) \}, \quad \text{among all } (m, v) : \begin{cases} \partial_t m - \operatorname{div}_g(vm) = 0 \\ m(0) = m_0, \quad m(T) = m_1 \end{cases} \quad (1.1)$$

where  $\mathcal{F}_\varepsilon(m, v) := \int_0^T \int_M \frac{1}{2} |v|^2 \, dm \, dt + \varepsilon \int_0^T \mathcal{H}(m(t); v) \, dt,$

with  $v := e^{-V(x)} dx$  and where

$$\mathcal{H}(m; v) = \int_M \log \left( \frac{dm}{dv} \right) dm = \int_M m(\log m + V)$$

denotes the relative entropy of  $m$  with respect to a reference measure  $e^{-V} dx$ , given for some Lipschitz continuous vector field  $V$ . Here and throughout the whole paper, we identify any measure  $\mu$  which is absolutely continuous with respect to  $dx$  with the corresponding density function  $d\mu/dx$ , still denoted as  $\mu$ .

In (1.1),  $m(t)$  is an (absolutely continuous) arc joining  $m_0$  and  $m_1$  with velocity  $v$ ,  $|\cdot|$  is the length of vector fields and  $\operatorname{div}_g(\cdot)$  the intrinsic divergence operator on the manifold  $M$ .

The functional (1.1) can be seen as a perturbation of the kinetic energy functional used in the dynamical version of mass transportation [2]. The additional term in (1.1) prevents concentration effects by penalizing through the relative entropy and is supposed to enhance some form of dissipation along optimal curves. This is only one, yet very natural, among possibly different entropic regularizations of the classical optimal transport energy. In this respect, it follows a stream of research which has been very intensive in recent times, where other kind of regularizations of the Wasserstein distance were suggested (see [13, 17, 28, 30]).

The evolution of optimal transport densities with additional costs that consider the effects of congestion has been exploited so far in several directions, see e.g. [3, 5], and especially [25], where some  $L^1 - L^\infty$  regularization in the time evolution of the optimal curves was proved using variational techniques. Similar problems were addressed in [7–10, 20, 37] with a different approach based on ideas coming from mean field game theory and PDE estimates on the optimality system (state-adjoint state) associated to (1.1), which is

$$\begin{cases} -\partial_t u + \frac{1}{2} |\nabla u|^2 = \varepsilon(\log(m) + V(x)) & \text{in } (0, T) \times M \\ \partial_t m - \operatorname{div}_g(m \nabla u) = 0 & \text{in } (0, T) \times M \\ m(0) = m_0, \quad m(T) = m_1 & \text{in } M. \end{cases} \quad (1.2)$$

As firstly observed by Lions [31], (1.2) is just one instance of PDE system appearing in mean field game theory [23, 24] and some smoothness on the optimal curves of this kind of functionals can be derived from gradient estimates on the adjoint state  $u$  (the so-called Kantorovich potential, in mass transportation language). This approach, relying on the ellipticity hidden in the optimality system, was thoroughly developed in [35, 36, 39]. In particular, the case of functional (1.1), with the additional entropy term, was addressed in [39] for convex domains of  $\mathbb{R}^d$  (with no-flux condition at the boundary) assuming that the marginals  $m_0, m_1$  are positive and smooth, in which case the minima can be proved to be positive and smooth for all times. Similar results were also proved for Gaussian-like measures in the whole Euclidean space.

The goal of this paper is twofold. First of all, we give some general result on problem (1.1), under the only assumption that the marginals  $m_0, m_1 \in L^1(M)$ . In particular, the

marginals do not need to be positive (nor smooth), extending many results of [39] to the case of nonnegative initial-terminal data. Except for some special results obtained in one dimension [11], the case of compactly supported marginals had not been developed, so far.

Secondly, we analyze the problem in the setting of a Riemannian manifold, in order to get a more exhaustive comprehension of some crucial tools. In fact, in the genuine optimal transport viewpoint, it is well understood [14–16, 32–34, 38, 42–44] that the Riemannian setting is the most natural to observe the role of Ricci curvature in regularity arguments related to displacement convexity and entropy dissipation. In our results, we will require the only condition that the Ricci curvature is bounded below, and we will obtain estimates which are totally consistent with the pure mass transport problem, embedding therefore the case  $\varepsilon = 0$  in (1.1) into a family of similar problems. As an example, if  $Ric_g(M) \geq \Lambda I$ , we will prove the displacement  $\Lambda$ -convexity of the entropy along Wasserstein geodesics as the limit of  $\Lambda_\varepsilon$ -convexity along the optimal curves of (1.1).

Concerning the characterization of minimal curves of (1.1), we summarize our main results in the following statement.

**Theorem 1.1** *Let  $(M, g)$  be a smooth compact Riemannian manifold without boundary, with  $Ric_g(M)$  bounded below. Let  $m_0, m_1 \in L^1(M) \cap \mathcal{P}(M)$ , and assume that  $V \in W^{2,\infty}(M)$ ,  $\varepsilon > 0$ . Then the functional  $\mathcal{F}_\varepsilon$  in (1.1) admits a unique minimum, given by  $(m, \nabla u)$ , where  $(m, u)$  is the unique weak solution (in the sense of Definition 5.1) of the system (1.2), with  $\int_M u(T)m_1 = 0$ .*

Moreover we have:

- (i)  $m > 0$  a.e. in  $(0, T) \times M$ .
- (ii)  $u, m \in L^\infty_{loc}((0, T) \times M)$  and  $u(0) \in L^1(dm_0), u(T) \in L^1(dm_1)$ .
- (iii) if  $m_0, m_1 \in W^{1,\infty}(M)$  and are (strictly) positive, and if  $V \in C^{k,\alpha}(M)$ , then  $u \in C^{k+1,\alpha}((0, T) \times M), m \in C^{k,\alpha}((0, T) \times M)$ .

The statement of Theorem 1.1 summarizes several different results that we establish later. In fact, we will start from the case of smooth and positive marginals (see Theorem 5.4), proving the existence of smooth solutions to the system (1.2). To this purpose, we follow the strategy suggested by Lions [31], and developed in [36, 39], which consists in rewriting the system (1.2) as a quasilinear elliptic equation for  $u$  (see (3.16)) and using the continuity method, relying on gradient bounds, in order to produce a smooth solution  $u$ . Following [39], this strategy is first employed for a penalized auxiliary problem (4.14), where the  $L^\infty$ -norm of  $u$  is readily controlled. Then a compactness argument yields a smooth solution  $(m, u)$  after a suitable normalization of  $u$ .

Once we have built smooth solutions of (1.2), we will obtain all relevant estimates which remain robust for merely  $L^1$  marginals  $m_0, m_1$ . This step includes a bound on the minimal value of  $\mathcal{F}_\varepsilon$ , obtained by building suitable competitors, which in turn will yield local uniform bounds on  $u$ .

From the local bounds on  $u$  we will also obtain the local  $L^\infty$ -bound on  $m$ . Notice that this  $L^1 - L^\infty$  regularization on the density is not a straightforward extension from the Euclidean case, because the Ricci curvature is allowed to be negative. In particular, the  $L^\infty$  control on  $m$  does not follow directly from the displacement convexity inequalities as in [39].

Finally, we will derive local (in time) bounds for  $Du$  in  $L^2$  (from the HJ equation) and for the Fisher information of  $m$  (by displacement convexity estimates) that yield the suitable compactness arguments. This latter step includes a relaxation result on the system and the convergence to weak solutions, providing with the final characterization of the minima of (1.1) (see Theorem 5.10).

Many of those tools, involving estimates and stability on the optimality system (1.2), are stable as  $\varepsilon \rightarrow 0$ .

Indeed, we conclude the article by giving a result on the convergence of the optimal curves towards the Wasserstein geodesic, as  $\varepsilon \rightarrow 0$ , as well as the convergence of  $\min \mathcal{F}_\varepsilon$  towards  $\min \mathcal{F}_0$ . This convergence occurs with a rate  $O(\varepsilon)$  when the marginals have finite entropy, while we cannot prove a rate better than  $O(\varepsilon |\log \varepsilon|)$  for the general case of marginals only in  $L^1(M)$ .

**Theorem 1.2** *Under the assumptions of Theorem 1.1, let  $(m_\varepsilon, \nabla u_\varepsilon)$  be the minima of (1.1), where  $(m_\varepsilon, u_\varepsilon)$  solves (1.2), with  $\int_M u_\varepsilon(T)m_1 = 0$ , and let  $(m, \nabla u)$  be the Wasserstein geodesic between  $m_0, m_1$ . Then, as  $\varepsilon \rightarrow 0$ , we have*

$$m_\varepsilon \rightarrow m \text{ in } C^0([0, T], \mathcal{P}(M)) \text{ and weakly in } L^1((0, T) \times M),$$

$$m_\varepsilon \nabla u_\varepsilon \rightarrow m \nabla u \text{ weakly in } L^1((0, T) \times M),$$

and  $\min \mathcal{F}_\varepsilon \rightarrow \min \mathcal{F}_0$ . In particular we have

$$\min \mathcal{F}_\varepsilon = \min \mathcal{F}_0 + r_\varepsilon$$

where  $r_\varepsilon = O(\varepsilon)$  if  $m_0, m_1$  have finite entropy, otherwise  $r_\varepsilon = O(\varepsilon |\log \varepsilon|)$ .

Last but not least, we will show (see Theorem 6.4) that using a suitable approximation of  $m_0, m_1$  with sequences  $m_{0\varepsilon}, m_{1\varepsilon}$  of smooth positive functions, the Wasserstein geodesic between  $m_0, m_1$  can be approximated by smooth minimizers of  $\mathcal{F}_\varepsilon$  (related to marginals  $m_{0\varepsilon}, m_{1\varepsilon}$ ), in a way that the corresponding approximated Kantorovich potentials  $u_\varepsilon$  remain uniformly bounded in Lipschitz norm, and converge uniformly in  $M$ .

This last result shows that the purpose of smoothing the Wasserstein geodesic can be fully accomplished with the functional (1.1); in particular, this gives a general Eulerian strategy towards the proof of displacement convexity properties along the geodesics of optimal transport. As an example, we recover some results of [14, 15] with an alternative proof, avoiding the use of EVI inequalities in favor of a standard Eulerian approach which is now justified going through problems (1.1).

## 2 Notations and setting of the problem

In the following, we recall some elements of Riemannian geometry (see e.g. [41]). Throughout the paper,  $(M, g)$  denotes a smooth, compact, connected and oriented  $d$ -dimensional Riemannian manifold without boundary, with metric tensor  $g = (g_{ij})$ , inverse  $g^{-1} = (g^{ij})$  and determinant  $|g|$ . The orientation induces a unitary volume form  $dx$  (hence  $\text{Vol}(M) = 1$  by normalization). Unless otherwise specified, the integration on  $M$  is understood to be taken with respect to  $dx$ . If  $w$  and  $v$  are two vector fields on  $M$ , we denote by

$$w \cdot_g v := \sum_{ij} g_{ij}(x) w_i v_j$$

their scalar product in the tangent space  $T_x M$ . The length of a vector field is given by  $|w| = \sqrt{w \cdot_g w}$ . Correspondingly, there is a scalar product in the cotangent space  $T_x^* M$ , which is defined on differential 1-forms  $\omega$  and  $\nu$  on  $M$  as  $\omega \cdot_g \nu := \sum_{ij} g^{ij}(x) \omega_i \nu_j$ .

Let  $x_j, j = 1, \dots, d$ , be a local system of coordinates: if  $u \in C^1(M)$ , the covariant gradient of  $u$ , denoted by  $\nabla u$ , is the vector field with coordinates  $(\nabla u)_i = g^{ij}(x) u_{x_j}$ .

Therefore, given  $u, v \in C^1(M)$ ,

$$\nabla u \cdot_g \nabla v = \sum_{ij} g^{ij}(x)u_{x_i}v_{x_j}.$$

We denote the Levi–Civita connection associated to the metric  $g$  with the letter  $D$  and we will derivate covariantly vector and tensor fields on  $M$ . Recalling that, in local coordinates, the Christoffel symbols are

$$\Gamma_{ij}^k = \frac{1}{2} \sum_l \left( \frac{\partial g_{jl}}{\partial x_i} + \frac{\partial g_{li}}{\partial x_j} - \frac{\partial g_{ij}}{\partial x_l} \right) g^{lk},$$

the covariant derivative of a  $C^1$  vector field  $X = (X_j)$  along the vector field  $v = (v_i)$  is the vector field  $D_v X$  with  $k$ -th coordinate given by

$$(D_v X)_k = \sum_{ij} v_i X_j \Gamma_{ij}^k + (\nabla X_k) \cdot_g v.$$

If  $X = (X_j)$  is a  $C^1$  vector field on  $M$ , the divergence of  $X$  is defined by

$$div_g X = \frac{1}{\sqrt{|g|}} \sum_k (\sqrt{|g|} X_k)_{x_k}$$

and the Leibniz rule:  $div_g(fX) = \nabla f \cdot_g X + f div_g X$  holds for every  $f \in C^1(M)$  and any  $C^1$  vector field  $X$  on  $M$ . Furthermore, by the Stokes theorem, we have

$$\int_M div_g X = 0.$$

The Hessian  $\nabla^2 u$  of a  $C^2$  function  $u$  is the symmetric 2-tensor given by

$$(\nabla^2 u)(v, w) := (D_v \nabla u) \cdot_g w = (D_w \nabla u) \cdot_g v$$

for every vector fields  $v, w$  on  $M$ , where the last equality follows by the symmetry and the compatibility with the metric of the Levi–Civita connection. The components of the Hessian are the second covariant derivatives, given by

$$\nabla_{ij} u = u_{x_i x_j} - \sum_k \Gamma_{ij}^k u_{x_k}.$$

In particular for every  $C^2$  functions  $f, u$  and every vector field  $v$  it holds

$$v \cdot_g \nabla(\nabla f \cdot_g \nabla u) = (\nabla^2 f)(v, \nabla u) + (\nabla^2 u)(v, \nabla f).$$

As usual, we denote  $\Delta_g u = div_g(\nabla u)$  the Laplace–Beltrami operator on  $M$ . We recall the Bochner formula (see e.g. [4]):

$$\frac{1}{2} \Delta_g |\nabla f|^2 = |\nabla^2 f|^2 + \nabla(\Delta_g f) \cdot_g \nabla f + Ricc_g(\nabla f, \nabla f) \tag{2.1}$$

for every  $f \in C^3(M)$ , where  $Ricc_g$  is the curvature tensor of the metric  $g$ . Throughout all the paper, we will assume that  $Ricc_g(M)$  is bounded below, i.e.

$$Ricc_g(X, X) \geq -\lambda |X|^2 \tag{2.2}$$

for every vector field  $X$  on  $M$ , for some  $\lambda \geq 0$ .

Given  $M$ , we denote by  $\mathcal{P}(M)$  the space of probability measures on  $M$ , endowed with the Wasserstein metric. The characterization of the Wasserstein geodesics in terms of optimal mass transportation on  $M$  is well established, since  $M$  is compact, see [34, 44]. We will always identify the measures with their densities when they are absolutely continuous with respect to the unitary volume measure  $dx$ . With  $L^p(M)$  we indicate the standard Lebesgue space, for  $p \in [1, +\infty]$ , and with  $W^{k,p}(M)$  the Sobolev space of functions with  $k$  weak  $L^p$ -derivatives. We will use the Sobolev and Poincaré–Wirtinger inequalities on  $M$ , for which we refer to [22]. Finally, throughout the paper we denote  $Q_T$  the cylinder  $(0, T) \times M$ , and  $\overline{Q}_T = [0, T] \times M$ .

### 2.1 Optimal transport functional

We now make precise the sense of the minimization problem (1.1).

**Definition 2.1** Let  $m_0, m_1 \in \mathcal{P}(M)$ . A couple  $(m, v)$  is a solution of the continuity equation

$$\begin{cases} \partial_t m - \operatorname{div}_g(vm) = 0 \\ m(0) = m_0, \quad m(T) = m_1, \end{cases} \tag{2.3}$$

if  $m \in C([0, T]; \mathcal{P}(M))$  with  $m(0) = m_0$  and  $m(T) = m_1$ ,  $v(t, x)$  is a measurable vector field on  $Q_T$  such that  $\int_0^T \int_M |v|^2 dm < \infty$  and the following equality holds

$$\int_M \varphi(t) dm(t) - \int_M \varphi(s) dm(s) + \int_s^t \int_M (-\partial_t \varphi + v \cdot_g \nabla \varphi) dm = 0,$$

for every  $0 \leq s < t \leq T$  and every function  $\varphi \in C^1(\overline{Q}_T)$ .

We recall (see [1]) that weak solutions as defined above are essentially equivalent to absolutely continuous curves from  $[0, T]$  into  $\mathcal{P}(M)$  which have  $L^2$  metric derivative. We also recall that any convex, superlinear function  $F(r)$  induces a lower semicontinuous functional on the space of probability measures:

$$\Phi(m) := \begin{cases} \int_M F(m) & \text{if } m \text{ is absolutely continuous} \\ +\infty & \text{otherwise.} \end{cases}$$

Similar kind of functionals have been extensively studied, see e.g. [25] and references therein. Even if we could consider general functions  $F$ , for the sake of clarity we restrict the analysis in this paper to the specific entropic case, in which  $F(m) = m \log(m)$ , and more generally to the relative entropy with respect to a possibly inhomogeneous reference measure  $\nu = e^{-V(x)} dx$ :

$$\mathcal{H}(m; \nu) := \int_M F\left(\frac{dm}{d\nu}\right) d\nu = \int_M \log\left(\frac{dm}{d\nu}\right) dm = \int_M m(\log m + V) \tag{2.4}$$

with the convention that  $\mathcal{H}(m; \nu) = +\infty$  whenever  $m$  is not absolutely continuous with respect to  $dx$ . In what follows, we assume that  $V$  is (at least) Lipschitz continuous on  $M$ .

Thanks to Definition 2.1, the meaning of the optimal transport problem (1.1) is now clarified, to be read as

$$\begin{aligned} \min\{\mathcal{F}_\varepsilon(m, v)\}, \quad \text{among all } (m, v) : & \begin{cases} \partial_t m - \operatorname{div}_g(vm) = 0 \\ m(0) = m_0, \quad m(T) = m_1 \end{cases}, \\ \mathcal{F}_\varepsilon(m, v) := & \int_0^T \int_M \frac{1}{2} |v|^2 dm + \varepsilon \int_0^T \mathcal{H}(m; \nu), \quad \nu := e^{-V(x)} dx, \end{aligned} \tag{2.5}$$

where the transport equation is understood as in Definition 2.1.

We first establish that, for every  $m_0, m_1 \in \mathcal{P}(M)$ , there exists an arc along which the above functional is finite, so it admits a finite minimum. In addition, we can give a universal upper bound on the minimal value of  $\mathcal{F}_\varepsilon$ .

**Proposition 2.2** *Let (2.2) hold true. There exists a constant  $C(M, d, \lambda, T, \|\varepsilon V\|_\infty)$  (depending on  $M$  as well as on  $d, \lambda, T, \|\varepsilon V\|_\infty$ ) such that*

$$\min \mathcal{F}_\varepsilon \leq C \tag{2.6}$$

for every  $m_0, m_1 \in \mathcal{P}(M)$ , and every  $\varepsilon \leq 1$ .

**Proof** Consider the heat kernel  $p_t(x, y)$  associated to the volume measure  $dx$  and the curve  $\mu_0(\cdot) : [0, 1] \rightarrow \mathcal{P}(M) \cap C_+^\infty(M)$  generated by the heat semigroup  $S_t$ :

$$t \rightarrow \mu_0(t, x) = S_t(\mu_0) := \int_M p_t(x, y) d\mu_0(y).$$

It is a classical result (cfr. [21, Chapters 7, 8]) that  $\mu_0(\cdot)$  is well defined and is a smooth solution of the heat equation

$$\frac{\partial}{\partial t} \mu_0 = \Delta_g \mu_0 \quad \text{on } [0, \infty) \times M.$$

In particular we have  $\mu_0(t, \cdot) > 0$  for every  $t > 0$  by the strong maximum principle.

It follows that the velocity of such curve for  $t > 0$  is given by the vector field

$$v_0(t, x) := \frac{\nabla \mu_0}{\mu_0}.$$

By the Li–Yau inequality [26, Theorem 1.4] we know that there exists a constant  $C(d, \lambda)$  such that

$$\frac{|\nabla \mu_0|^2}{\mu_0} - 2 \frac{\partial}{\partial t} \mu_0 \leq \left(C + \frac{2d}{t}\right) \mu_0.$$

Recalling that  $\mu_0$  is a probability density for every  $t > 0$ , integrating the above inequality we get

$$\int_M |v_0|^2 \mu_0 \leq C + \frac{2d}{t}. \tag{2.7}$$

We now study the decay of the entropy along the heat flow. We recall that

$$\frac{\partial}{\partial t} \int_M \mu_0 \log(\mu_0) = \frac{\partial}{\partial t} \int_M (\mu_0 \log(\mu_0) - \mu_0) = - \int_M \frac{|\nabla \mu_0|^2}{\mu_0}.$$

By Sobolev and Poincaré–Wirtinger inequality we have (for  $2^* = \frac{2d}{d-2}$  if  $d > 2$ , or  $2^*$  any sufficiently large number if  $d = 2$ )

$$\begin{aligned} \int_M \frac{|\nabla \mu_0|^2}{\mu_0} &= 4 \int_M |\nabla \sqrt{\mu_0}|^2 \geq C_S \left( \int_M \left| \sqrt{\mu_0} - \int_M \sqrt{\mu_0} \right|^{2^*} dx \right)^{\frac{2}{2^*}} \\ &\geq c_1 \left( \int_M \sqrt{\mu_0}^{2^*} dx \right)^{\frac{2}{2^*}} - c_2 \left( \int_M \sqrt{\mu_0} \right)^2 \end{aligned}$$

$$\geq c_1 \left( \int_M \sqrt{\mu_0^{2^*}} dx \right)^{\frac{2}{2^*}} - c_2$$

where, here and below, by  $c_i$  we denote different constants only depending on  $M$ , in particular through  $d$  and  $\lambda$ . By the concavity of the log function and Jensen inequality for the probability measure  $\mu_0$

$$\begin{aligned} \log \left( \int_M \frac{1}{\mu_0} |\nabla \mu_0|^2 + c_2 \right) &\geq \frac{2}{2^*} \log \left( \int_M \sqrt{\mu_0^{2^*}} dx \right) + \log(c_1) \\ &\geq \frac{2}{2^*} \int_M \log \left( \mu_0^{\frac{2^*-2}{2}} \right) \mu_0 dx + \log(c_1) \\ &= \kappa_d \int_M \log(\mu_0) \mu_0 dx + \log(c_1) \end{aligned} \tag{2.8}$$

where  $\kappa_d = 1 - \frac{2}{2^*}$ , (in particular,  $\kappa_d = \frac{2}{d}$  if  $d > 2$ ). In other words, if  $\varphi(t) := \int_M \mu_0 \log \mu_0$ , then we deduce

$$\varphi'(t) \leq -c_1 e^{\kappa_d \varphi} + c_2$$

which implies

$$(\varphi'(t) - c_2) e^{-\kappa_d(\varphi - c_2 t)} \leq -c_1 e^{\kappa_d c_2 t} \leq -c_1.$$

Integrating in  $(t_0, t_1)$ , we get

$$-e^{-\kappa_d(\varphi(t_1) - c_2 t_1)} + e^{-\kappa_d(\varphi(t_0) - c_2 t_0)} + \kappa_d c_1 (t_1 - t_0) \leq 0.$$

In particular, letting  $t_0 \rightarrow 0$  we deduce

$$-e^{-\kappa_d(\varphi(t_1) - c_2 t_1)} + \kappa_d c_1 t_1 \leq 0.$$

Since  $t_1$  is arbitrary, this means that

$$\int_M \mu_0(t) \log \mu_0(t) = \varphi(t) \leq -\frac{1}{\kappa_d} \log(\kappa_d c_1 t) + c_2 t \tag{2.9}$$

for every  $t > 0$ .

Now, for any given  $\beta > 1$ , we consider the reparametrization  $\tilde{\mu}_0(t, \cdot) := \mu_0(t^\beta)$  for every  $t > 0$ . Its velocity field is

$$\tilde{v}_0(t, \cdot) := \beta t^{\beta-1} v_0(t^\beta, \cdot)$$

so, for any fixed  $0 < \delta_0 < \frac{T}{3}$ , by (2.7)

$$\int_0^{\delta_0} \int_M |\tilde{v}_0|^2 \tilde{\mu}_0 dt = \beta \int_0^{\delta_0^\beta} t^{1-\frac{1}{\beta}} \int_M |v_0|^2 \mu_0 dt \leq c_3 \int_0^{\delta_0^\beta} \frac{1}{t^{\frac{1}{\beta}}} dt$$

which is finite for every  $\beta > 1$ . With such a choice, if we merge this estimate with (2.9) we obtain

$$\int_0^{\delta_0} \int_M |\tilde{v}_0|^2 \tilde{\mu}_0 dt + \varepsilon \int_0^{\delta_0} \int_M \tilde{\mu}_0(t) (\log \tilde{\mu}_0(t) + V) dt \leq C_0$$

where  $C_0$  is a constant depending only on  $M, d, \lambda, T, \|\varepsilon V\|_\infty, \beta$  and  $\delta_0$ .

In a similar way, for a fixed  $T/3 < \delta_1 < T$ , we find a smooth curve of probability densities  $\tilde{\mu}_1$ , with velocity  $\tilde{v}_1$  such that

$$\int_{\delta_1}^T \int_M |\tilde{v}_1|^2 \tilde{\mu}_1 dt + \varepsilon \int_{\delta_1}^T \int_M \tilde{\mu}_1(t) (\log \tilde{\mu}_1(t) + V) dt \leq C_1$$

where  $C_1$  is a constant depending only on  $M, d, \lambda, T, \|\varepsilon V\|_\infty, \beta$  and  $\delta_1$ .

Consider now the 2-Wasserstein geodesic  $(\bar{\mu}, \bar{v})$  between  $\tilde{\mu}_0(\delta_0, \cdot)$  and  $\tilde{\mu}_1(\delta_1, \cdot)$ . We recall (see e.g. [15]) that the entropy functional is  $(-\lambda)$ -convex along the 2-Wasserstein geodesic. Hence, by (2.9),

$$\int_{\delta_0}^{\delta_1} \int_M |\bar{v}|^2 \bar{\mu} dt + \varepsilon \int_{\delta_0}^{\delta_1} \int_M \bar{\mu} (\log \bar{\mu} + V) dt \leq C$$

where the constant  $C$  depends only on  $M, d, \lambda, T, \|\varepsilon V\|_\infty, \delta_0, \delta_1$  and the Wasserstein distance  $W_2(\tilde{\mu}_0(\delta_0, \cdot), \tilde{\mu}_1(\delta_1, \cdot))$ . However, the latter is uniformly estimated in terms of the manifold, thanks to the compactness of  $M$ .

Finally, gluing the paths from  $m_0$  to  $\tilde{\mu}_0(\delta_0, \cdot)$  and from  $\tilde{\mu}_1(\delta_1, \cdot)$  to  $m_1$  with the 2-Wasserstein geodesic  $\bar{\mu}$ , we have built an admissible arc joining  $m_0$  and  $m_1$ , and with a convenient choice of  $\delta_0, \delta_1$  we estimate

$$\inf \mathcal{F}_\varepsilon(m_0, m_1) \leq C_0 + C + C_1$$

where last constants only depend on  $M, d, \lambda, T, \|\varepsilon V\|_\infty$ .

It is a classical result, after Benamou-Brenier’s trick [2] and the weak-lower semicontinuity of the entropy, that the above estimate, with the existence of an admissible curve, yields the existence of a minimizer of  $\mathcal{F}_\varepsilon$  by direct methods of Calculus of Variations. For a similar proof in Euclidean context, see e.g. [25, Proposition 2.9].  $\square$

**Remark 2.3** With a suitable choice of  $\delta_0, \delta_1$  in the above proof, it is possible to give an estimate of the dependence of the constant  $C$  in (2.6) from the time-horizon  $T$ . Alternatively, if we denote  $\min \mathcal{F}_\varepsilon^1$  the minimum in unit time  $T = 1$ , with a simple time-scaling one can estimate

$$\min \mathcal{F}_\varepsilon \leq \max \left( \frac{1}{T}, T \right) \min \mathcal{F}_\varepsilon^1 \leq \max \left( \frac{1}{T}, T \right) C(M, d, \lambda, \|\varepsilon V\|_\infty).$$

**Remark 2.4** We stress that in the above proof we can avoid the use of the Wasserstein geodesic  $(\bar{\mu}, \bar{v})$  between  $\tilde{\mu}_0(\delta_0, \cdot)$  and  $\tilde{\mu}_1(\delta_1, \cdot)$  (and consequently, avoid the use of the  $(-\lambda)$ -displacement convexity of the geodesic). In fact, since  $\tilde{\mu}_0(\delta_0, \cdot)$  and  $\tilde{\mu}_1(\delta_1, \cdot)$  are smooth and positive, we can take the smooth optimal curve of  $\mathcal{F}_\varepsilon$  joining the two measures, whose existence will be (independently) proved in Sect. 5, Theorem 5.4.

### 3 The optimality system

In this Section we discuss the structure of the optimality system satisfied by minima of functional (2.5). This is a first order PDE system which takes the following form

$$\begin{cases} -\partial_t u + \frac{1}{2} |\nabla u|^2 = \varepsilon (\log(m) + V), & t \in (0, T) \\ \partial_t m - \operatorname{div}_g(m \nabla u) = 0, & t \in (0, T). \end{cases} \tag{3.1}$$

Let us first underline that, at least formally, (3.1) is the optimality condition of (2.5), in the sense that  $(m, \nabla u)$  provides with the couple  $(m, v)$  minimizing (2.5). This is a consequence of the convexity of the functional, following the original idea of Benamou and Brenier [2]. Even if this is clear to expert readers, we provide a proof for completeness.

**Lemma 3.1** *Let  $(u, m)$  be a smooth solution of system (3.1), in the sense that  $u \in C^1([0, T] \times M)$ ,  $m \in C^0([0, T] \times M)$  with  $m(0) = m_0, m(T) = m_1$ , and  $m > 0$ . Then  $(m, \nabla u)$  is a minimum point of (2.5).*

**Proof** Let  $(\mu, v)$  be any couple which solves the continuity equation in the sense of Definition 2.1. We can assume that  $\mathcal{F}_\varepsilon(\mu, v) < \infty$  (otherwise, the inequality  $\mathcal{F}_\varepsilon(m, \nabla u) \leq \mathcal{F}_\varepsilon(\mu, v)$  is obvious), and in particular  $\mu \in L^1(Q_T)$ . Let us define the following convex and lower semicontinuous function in  $\mathbb{R}^d \times \mathbb{R}$ :

$$\Psi(p, m) = \begin{cases} \frac{|p|^2}{2m} & \text{if } m > 0, \\ 0 & \text{if } m = 0 \text{ and } p = 0, \\ +\infty & \text{otherwise.} \end{cases} \tag{3.2}$$

We set  $w = \mu v, \hat{w} = m \nabla u$ . Since  $m, u$  are smooth solutions of (3.1) (with  $m > 0$ ), then  $\partial \Psi(\hat{w}, m)$  is well defined. By convexity of  $\Psi$ , we have

$$\begin{aligned} \Psi(\hat{w}, m) - \Psi(w, \mu) &\leq \partial_p \Psi(\hat{w}, m) \cdot (\hat{w} - w) + \partial_m \Psi(\hat{w}, m)(m - \mu) \\ &= \frac{\hat{w}}{m} \cdot (\hat{w} - w) - \frac{|\hat{w}|^2}{2m^2}(m - \mu) \end{aligned}$$

Since  $\mu \in L^1$  and  $\mu|v|^2 \in L^1$ , we have  $w \in L^1$  and the above inequality is integrable on  $M$ . From the very definition of  $w, \hat{w}$  we deduce

$$\begin{aligned} \int_0^T \int_M \frac{1}{2} m |\nabla u|^2 &\leq \frac{1}{2} \int_0^T \int_M \mu |v|^2 + \frac{1}{2} \int_0^T \int_M m |\nabla u|^2 \\ &\quad - \int_0^T \int_M w \cdot \nabla u + \frac{1}{2} \int_0^T \int_M \mu |\nabla u|^2 \end{aligned} \tag{3.3}$$

Since  $u \in C^1$ , the continuity equation gives

$$\begin{aligned} \int_0^T \int_M w \cdot \nabla u &= \int_0^T \int_M \partial_t u \mu - \int_M u(T) m_1 + \int_M u(0) m_0 \\ &= -\varepsilon \int_0^T \int_M (\log m + V) \mu + \frac{1}{2} \int_0^T \int_M |\nabla u|^2 \mu \\ &\quad - \int_M u(T) m_1 + \int_M u(0) m_0. \end{aligned}$$

Hence from (3.3) we get

$$\begin{aligned} \int_0^T \int_M \frac{1}{2} m |\nabla u|^2 &\leq \frac{1}{2} \int_0^T \int_M \mu |v|^2 + \frac{1}{2} \int_0^T \int_M m |\nabla u|^2 + \int_M u(T) m_1 - \int_M u(0) m_0 \\ &\quad + \varepsilon \int_0^T \int_M (\log m + V) \mu \\ &= \frac{1}{2} \int_0^T \int_M \mu |v|^2 - \varepsilon \int_0^T \int_M m (\log(m) + V) + \varepsilon \int_0^T \int_M (\log m + V) \mu \end{aligned}$$

By convexity we obviously have  $m \log(m) - \mu \log(\mu) \leq \log(m)(m - \mu) + (m - \mu)$ , where last term disappears after integration. Then we conclude that

$$\mathcal{F}_\varepsilon(m, \nabla u) \leq \mathcal{F}_\varepsilon(\mu, v).$$

□

Since (3.1) is a Hamiltonian system, there is some invariant of motion. The proof is straightforward.

**Lemma 3.2** *Let  $(u, m)$  be a smooth solution of system (3.1). Then we have that the quantity*

$$E(m_0, m_1) := \frac{1}{2} \int_M m |\nabla u|^2 - \varepsilon \int_M m (\log(m) + V)$$

*is constant in time and it holds*

$$E(m_0, m_1) = \frac{1}{T} \mathcal{B}_\varepsilon(m_0, m_1) - \frac{2\varepsilon}{T} \int_0^T \int_M m (\log(m) + V), \tag{3.4}$$

where  $\mathcal{B}_\varepsilon(m_0, m_1) = \min \mathcal{F}_\varepsilon$ .

We observe that the quantity  $E$  only depends on the marginals  $m_0, m_1$ , and is easily estimated. In particular, by Jensen’s inequality, we have  $\int_M m (\log m + V) dx \geq -\log(v(M))$ , for the measure  $v := e^{-V} dx$ . Recalling Proposition 2.2 (and Remark 2.3), we deduce that

$$E(m_0, m_1) \leq \frac{1}{T^2 \wedge 1} C(M, d, \lambda, \|e^V\|_\infty) + 2\varepsilon \log\left(\int_M e^{-V} dx\right) \leq K \tag{3.5}$$

for some constant  $K$  which is uniform for all  $m_0, m_1 \in \mathcal{P}(M)$ ,  $V \in L^\infty(M)$  with uniform bound, and any  $\varepsilon \leq 1$ .

### 3.1 Displacement convexity estimates

In this section we study the convexity of some energy functional along the optimal curves of (2.5). This is obtained in the Eulerian approach by exploiting dissipativity properties of the solutions of system (3.1). We consider smooth solutions, which justifies the computations below; as we will see later, this is no loss of generality, since all solutions will be obtained as limit of classical ones. The following is an extension to the Riemannian setting of the results proved in [19], (or in [39] with Neumann conditions); the only new ingredient is provided by the Bochner formula (2.1).

**Proposition 3.3** *Let  $u \in C^2(\overline{Q}_T)$  and  $m \in C^1(\overline{Q}_T)$  be classical solutions to the system (3.1), where  $V \in W^{2,\infty}(M)$ .*

*Let  $U : (0, +\infty) \rightarrow \mathbb{R}$  be a  $C^1$  function such that*

$$P(r) := U'(r)r - U(r) \geq 0.$$

*Then*

$$\begin{aligned} \frac{d^2}{dt^2} \int_M U(m) &\geq \int_M \left[ P'(m)m - \left(1 - \frac{1}{d}\right)P(m) \right] (\Delta_g u)^2 \\ &\quad + \int_M P(m) Ricc_g(\nabla u, \nabla u) \\ &\quad + \int_M \varepsilon \frac{P'(m)}{m} |\nabla m|^2 + \varepsilon \int_M P'(m) \nabla m \cdot_g \nabla V \end{aligned} \tag{3.6}$$

**Proof** We begin by calculating the first derivate of the function  $t \rightarrow \int_M U(m)$ ,

$$\begin{aligned} \frac{d}{dt} \int_M U(m) &= \int_M U'(m) \partial_t m \\ &= \int_M U'(m) (m \Delta_g u + \nabla m \cdot_g \nabla u) \\ &= \int_M P(m) \Delta_g u \end{aligned}$$

recalling that  $P(r) = U'(r)r - U(r)$ . So, the second derivative takes the form

$$\begin{aligned} \frac{d^2}{dt^2} \int_M U(m) &= \int_M P'(m) \partial_t m \Delta_g u + P(m) \Delta_g (\partial_t u) \\ &= \int_M P'(m) (m \Delta_g u + \nabla m \cdot_g \nabla u) \Delta_g u \\ &\quad + \int_M P(m) \Delta_g \left( \frac{1}{2} |\nabla u|^2 - \varepsilon (\log(m) + V) \right) \\ &= \int_M [P'(m)m - P(m)] (\Delta_g u)^2 - \int_M P(m) \nabla (\Delta_g u) \cdot_g \nabla u \\ &\quad + \int_M P(m) \Delta_g \left( \frac{1}{2} |\nabla u|^2 - \varepsilon (\log(m) + V) \right). \end{aligned}$$

Now we use Böchner’s formula (2.1) to calculate  $\Delta_g (\frac{1}{2} |\nabla u|^2)$ , obtaining

$$\begin{aligned} \frac{d^2}{dt^2} \int_M U(m) &= \int_M [P'(m)m - P(m)] (\Delta_g u)^2 \\ &\quad + \int_M P(m) [tr((\nabla^2 u)^2) + Ricc_g(\nabla u, \nabla u)] \\ &\quad + \int_M \varepsilon \frac{P'(m)}{m} |\nabla m|^2 dx + \varepsilon \int_M P'(m) \nabla m \cdot_g \nabla V. \end{aligned}$$

Since, by the symmetry of  $\nabla^2 u$ , it holds  $tr((\nabla^2 u)^2) \geq \frac{1}{d} (tr(\nabla^2 u))^2 = \frac{1}{d} (\Delta_g u)^2$ , we get (3.6). □

In particular, the inequality (3.6) implies the semi-convexity of the log-entropy along the optimal curve  $m(t)$ , and the strict convexity of the relative entropy whenever  $Ricc_g + D^2 V \geq 0$ .

**Corollary 3.4** *Under the assumptions of Proposition 3.3, let  $\mathcal{H}(m(t); \nu)$  be the relative entropy defined in (2.4), for  $\nu = e^{-V} dx$ . Then we have*

$$\begin{aligned} \frac{d^2}{dt^2} \mathcal{H}(m(t); \nu) &\geq \int_M m (Ricc_g(\nabla u, \nabla u) + D^2 V(\nabla u, \nabla u)) \\ &\quad + \varepsilon \int_M |\nabla (\log m + V)|^2 m. \end{aligned} \tag{3.7}$$

Moreover, let  $\lambda \in \mathbb{R}$  satisfy (2.2), and define

$$\varphi(t) = \int_M m(t) \log m(t).$$

Then we have:

(i) there exists a constant  $\Lambda_\varepsilon$ , depending on  $M, d, \lambda, \|V\|_{W^{1,\infty}}, T, \varepsilon$ , such that

$$\varphi(t) \leq \frac{T-t}{T}\varphi(0) + \frac{t}{T}\varphi(T) + \Lambda_\varepsilon \frac{t(T-t)}{2T^2}. \tag{3.8}$$

Moreover, the constant  $\Lambda_\varepsilon$  is bounded independently of  $\varepsilon$  (for  $\varepsilon \leq 1$ ), and we have  $\Lambda_\varepsilon \xrightarrow{\varepsilon \rightarrow 0} \frac{\lambda}{T} W_2(m_0, m_1)^2$ .

(ii) there exist constants  $L = L(M, d, \lambda, \|V\|_{W^{1,\infty}}, T)$  and  $\ell_d$  (only depending on  $d$ ) such that

$$\varphi(t) \leq 2\ell_d |\log(t(T-t))| + \ell_d |\log \varepsilon| + L \quad \forall t \in (0, T). \tag{3.9}$$

**Proof** We use Proposition 3.3 with  $U(r) = r \log r - r$  (so that  $P(r) = r$ ) and we get

$$\begin{aligned} \frac{d^2}{dt^2} \int_M m \log m &= \frac{d^2}{dt^2} \int_M (m \log m - m) \geq \int_M m Ric_{\bar{c}g}(\nabla u, \nabla u) \\ &+ \int_M \varepsilon \frac{1}{m} |\nabla m|^2 + \varepsilon \int_M \nabla m \cdot_g \nabla V. \end{aligned} \tag{3.10}$$

Similarly, we compute

$$\begin{aligned} \frac{d^2}{dt^2} \int_M m V &= \frac{d}{dt} \int_M \partial_t m V = -\frac{d}{dt} \int_M m \nabla V \cdot_g \nabla u \\ &= -\int_M (\nabla V \cdot_g \nabla u) \operatorname{div}_g(m \nabla u) - \int_M m \nabla V \cdot_g \nabla \left(\frac{1}{2} |\nabla u|^2 - \varepsilon(\log(m) + V)\right) \\ &= \int_M m D^2 V(\nabla u, \nabla u) + \varepsilon \int_M m \nabla V \cdot_g \nabla(\log(m) + V). \end{aligned}$$

Adding this equality to (3.10), we obtain (3.7).

Now, if we come back to (3.10) and use the lower bound on  $Ric_g$ , we get

$$\frac{d^2}{dt^2} \int_M m \log m \geq -\lambda \int_M m |\nabla u|^2 + \frac{\varepsilon}{2} \int_M \frac{1}{m} |\nabla m|^2 - \frac{\varepsilon}{2} \int_M m |\nabla V|^2.$$

By definition of the quantity  $E(m_0, m_1)$  we obtain

$$\begin{aligned} \frac{d^2}{dt^2} \int_M m \log m &\geq -2\lambda E(m_0, m_1) - 2\lambda\varepsilon \int_M m(\log m + V) \\ &+ \frac{\varepsilon}{2} \int_M \frac{1}{m} |\nabla m|^2 - \frac{\varepsilon}{2} \int_M m |\nabla V|^2 \\ &\geq -2\lambda E(m_0, m_1) - 2\lambda\varepsilon \int_M m \log m \\ &+ \frac{\varepsilon}{2} \int_M \frac{1}{m} |\nabla m|^2 - \varepsilon c(\lambda, \|V\|_{W^{1,\infty}}). \end{aligned} \tag{3.11}$$

We estimate the Fisher information of  $m$  as in Proposition 2.2, see (2.8):

$$\int_M \frac{1}{m} |\nabla m|^2 \geq c_1 \exp\left(\kappa_d \int_M m \log m\right) - c_2,$$

where we recall that  $\kappa_d = 1 - \frac{2}{2^*}$  ( $2^* = \frac{2d}{d-2}$  if  $d > 2$ , otherwise it is any sufficiently large number if  $d = 2$ ). Therefore, if  $\varphi(t) := \int_M m \log m$ , we deduce

$$\varphi'' \geq -2\lambda E(m_0, m_1) + \varepsilon(-2\lambda\varphi + c_3 e^{\kappa_d \varphi}) - \varepsilon c(\lambda, \|V\|_{W^{1,\infty}}, M), \tag{3.12}$$

for some positive constant  $c_3$ . We note that the function  $r \rightarrow -2\lambda r + c_3 e^{\kappa_d r}$  has a finite minimum on  $[0, +\infty)$ , and that  $E(m_0, m_1)$  is bounded above by some constant  $K$  only depending on  $M, d, \lambda, T, \|V\|_\infty$ , see (3.5). Hence we have

$$\varphi'' \geq -2\lambda K - \varepsilon C(\lambda, \|V\|_{W^{1,\infty}}, M)$$

which gives the semi-convexity of the entropy along the optimal curves, with a semi-convexity constant  $\Lambda_\varepsilon$  which is bounded uniformly for  $\varepsilon \leq 1$ .

In fact, we also obtain a local bound for the entropy, independently from the initial and terminal marginals. Indeed, we deduce from (3.12) that

$$\varphi'' \geq \varepsilon c_4 e^{\kappa_d \varphi} - c_5 \quad t \in (0, T),$$

for some constant  $c_4, c_5$  depending on  $M, d, \lambda, \|V\|_{W^{1,\infty}}, T, \varepsilon$  (and uniform for  $\varepsilon \leq 1$ ). With a suitable choice of  $L$  (depending on  $c_4, c_5, T$ ), we have that the function

$$\psi(t) := -\frac{1}{\kappa_d} \log(\varepsilon t^2(T-t)^2) + L$$

is a supersolution of the same equation, i.e.  $\psi'' \leq \varepsilon c_4 e^{\kappa_d \psi} - c_5$  for  $t \in (0, T)$ . Since  $\psi$  blows up at  $t = 0, t = T$ , we conclude by comparison that  $\varphi \leq \psi$ , which gives (3.9) with  $\ell_d = \frac{1}{\kappa_d}$ .

To conclude, we observe that the semi-convexity constant  $\Lambda_\varepsilon$  can now be estimated in a more precise form. Indeed, on account of (3.4), we have

$$\begin{aligned} \Lambda_\varepsilon &= \varepsilon C + 2\lambda E(m_0, m_1) = \varepsilon C + 2\frac{\lambda}{T} \min \mathcal{F}_\varepsilon - \frac{4\lambda\varepsilon}{T} \int_0^T \int_M m(\log(m) + V) \\ &\simeq \varepsilon C(1 + |\log \varepsilon|) + 2\frac{\lambda}{T} \min \mathcal{F}_\varepsilon, \end{aligned}$$

where we used that  $\int_0^T \int_M m \log(m) = O(\log(\varepsilon))$  by (3.9). As we will prove in Sect. 6, it holds that  $\min(\mathcal{F}_\varepsilon) \rightarrow \frac{1}{2} W_2(m_0, m_1)^2$ ; hence we deduce

$$\Lambda_\varepsilon \xrightarrow{\varepsilon \rightarrow 0} \frac{\lambda}{T} W_2(m_0, m_1)^2.$$

□

### 3.2 The optimality system as an elliptic equation

System (3.1) can be recast as a single elliptic equation, in time-space variables, for  $u$ . This comes by noting that  $me^V = \exp(\frac{1}{\varepsilon}(|\nabla u|^2 - \partial_t u))$ , which can be inserted in the continuity equation, giving rise to a quasilinear elliptic equation in divergence form for  $u$ .

This is in fact a special case of a general approach suggested by P-L. Lions in his lectures at Collège de France [31], in order to handle mean-field game systems of first order, such as

$$\begin{cases} -\partial_t u + \frac{1}{2} |\nabla u|^2 = f(m) + V \\ \partial_t m - m \Delta_g u - \nabla u \cdot_g \nabla m = 0 \end{cases} \tag{3.13}$$

whenever  $f$  is an increasing function. For the reader's convenience, we derive here this equation in the Riemannian context. To this purpose, we first compute the covariant gradient

of both terms in the Hamilton–Jacobi equation:

$$f'(m)\nabla m = \nabla \left( -\partial_t u + \frac{|\nabla u|^2}{2} - V \right).$$

Taking the scalar product with  $\nabla u$  we get

$$f'(m)\nabla m \cdot_g \nabla u = \nabla \left( -\partial_t u + \frac{|\nabla u|^2}{2} - V \right) \cdot_g \nabla u.$$

Similarly, by taking the time derivative from the Hamilton–Jacobi equation, we have:

$$\begin{aligned} f'(m)\partial_t m &= \frac{\partial}{\partial t} \left( -\partial_t u + \frac{|\nabla u|^2}{2} - V \right) \\ &= -\partial_{tt} u + \nabla(\partial_t u) \cdot_g \nabla u. \end{aligned} \tag{3.14}$$

From the above equalities, merged with the continuity equation for  $m$ , we obtain

$$\begin{aligned} -\partial_{tt} u + \nabla(\partial_t u) \cdot_g \nabla u - \nabla \left( -\partial_t u + \frac{|\nabla u|^2}{2} - V \right) \cdot_g \nabla u &= f'(m) (\partial_t m - \nabla m \cdot_g \nabla u) \\ &= f'(m) m \Delta_g u \end{aligned} \tag{3.15}$$

which becomes a second-order equation in the only unknown  $u$ . Indeed, by setting  $\phi := (f)^{-1}$ , the first equation of the system reads as

$$m = \phi \left( -\partial_t u + \frac{|\nabla u|^2}{2} - V \right),$$

and so (3.15) becomes a single equation for  $u$ . In the particular case of  $f(m) = \varepsilon \log m$ , we have  $f'(m)m = \varepsilon$ , and (3.15) simplifies further; if we also scale the potential  $V$  into  $\varepsilon V$  (this is not necessary of course, but it is more consistent with the model of relative entropy), we get to the following form:

$$-\partial_{tt} u + 2\nabla u \cdot_g \nabla(\partial_t u) - (\nabla^2 u)(\nabla u, \nabla u) - \varepsilon \Delta_g u + \varepsilon \nabla u \cdot_g \nabla V = 0. \tag{3.16}$$

We can shorten such equation by writing it as a quasilinear equation in the space-time variable

$$-tr \left( \mathcal{A}(x, \nabla u) \circ \overline{\nabla}^2 u \right) + \varepsilon \nabla u \cdot_g \nabla V(x) = 0 \tag{3.17}$$

where  $\mathcal{A}(x, \eta)$  is the endomorphism of  $\mathbb{R} \times T_x M$  defined by

$$\begin{aligned} \mathbb{R} \times T_x M &\longrightarrow \mathbb{R} \times T_x M \\ [\bar{w}, w] &\longmapsto [-(\eta \cdot_g w - \bar{w}), (\eta \cdot_g w - \bar{w})\eta + \varepsilon w] \end{aligned}$$

and, for every  $C^2$  function  $f$  on  $[0, T] \times M$ , the endomorphism  $\overline{\nabla}^2 f$  is given by

$$\begin{aligned} \mathbb{R} \times T_x M &\longrightarrow \mathbb{R} \times T_x M \\ [\bar{w}, w] &\longmapsto [\bar{w}\partial_{tt} f + \nabla(\partial_t f) \cdot_g w, \bar{w}\nabla(\partial_t f) + D_w \nabla f]. \end{aligned}$$

Note that  $\mathcal{A}(x, \eta)$  is independent of  $t \in [0, T]$  and it is symmetric for every choice  $(x, \eta) \in M \times T_x M$ . The symbol  $\mathcal{A}(x, \eta)$  will denote also the bilinear form induced by such endomorphism through the product metric  $g$  of the manifold  $\mathbb{R} \times M$ . Namely,

$$\mathcal{A}(x, \eta)([\bar{w}, w], [\bar{v}, v]) := (\mathcal{A}(x, \eta)[\bar{w}, w]) \cdot_g [\bar{v}, v]$$

$$= (\eta \cdot_g w - \bar{w})(\eta \cdot_g v - \bar{v}) + \varepsilon w \cdot_g v.$$

Finally we note that such bilinear form is elliptic (though not uniformly), in fact for every  $[\bar{w}, w] \in \mathbb{R} \times T_x M$  we have

$$\mathcal{A}(x, \eta)([\bar{w}, w], [\bar{w}, w]) = (\eta \cdot_g w - \bar{w})^2 + \varepsilon |w|^2 > 0 \quad \forall [\bar{w}, w] \neq [0, 0].$$

### 4 Gradient bounds for smooth solutions

The purpose of this section is to obtain estimates for smooth solutions of the system (1.2), by exploiting the elliptic character of the quasilinear equation (3.17). As a matter of fact, the existence itself of smooth solutions of (1.2) needs to be justified, and this is why we will develop this kind of estimates for a class of penalized problems, which will serve as approximation steps towards (1.2). To be precise, we will introduce parameters  $\rho, \delta > 0$  and we will consider first the doubly penalized problem (4.4), which is an approximation of the elliptic equation (3.17). At this stage, the parameter  $\rho$  is introduced only to speed up some continuity argument for elliptic equations, in order to produce initially some solution  $u_{\rho, \delta}$  to (4.4) (this task is accomplished in ‘‘Appendix A’’). The estimates that we prove in the next Sects. 4.1 and 4.2 will allow us to let  $\rho \rightarrow 0$  and obtain a smooth solution for the  $\delta$ -penalized problem (4.13), corresponding to the system (4.14). This is now a close approximation of (1.2), with only a penalization of the marginal constraints at  $t = 0, T$ , and will be discussed in Sect. 4.3. In the end, the analysis of the penalized problems, and the uniform character with respect to  $\rho, \delta$  of the a priori estimates, will be crucial in order to deduce the existence of smooth solutions of (1.2), jointly with the regularity estimates that those solutions satisfy.

To perform the above program, we mostly follow the ideas developed in [31, 36, 39], although specifying to the case of the entropy nonlinearity allows us to simplify some argument and to give estimates in a more precise form.

As a first step, we derive from the (possibly penalized) elliptic problem the differential equations solved by some auxiliary functions of  $u$  and its derivatives.

**Lemma 4.1** *Let  $u \in C^3([0, T] \times M)$  be a solution of*

$$-tr \left( \mathcal{A}(x, \nabla u) \circ \bar{\nabla}^2 u \right) + \rho u + \varepsilon \nabla u \cdot_g \nabla V(x) = 0 \tag{4.1}$$

for some constants  $\rho, \varepsilon \geq 0$ . Then

(i) *for every  $K \in \mathbb{R}$ , the function  $h := (u + K)^2$  satisfies*

$$\begin{aligned} & -tr \left( \mathcal{A}(x, \nabla u) \circ \bar{\nabla}^2 h \right) + \varepsilon \nabla h \cdot_g \nabla V + 2\rho u (u + K) \\ & = -2\mathcal{A}(x, \nabla u) ([\partial_t u, \nabla u], [\partial_t u, \nabla u]). \end{aligned}$$

(ii) *Set  $\omega := \partial_t u$ , then it satisfies*

$$-tr \left( \mathcal{A}(x, \nabla u) \circ \bar{\nabla}^2 \omega \right) + \varepsilon \nabla \omega \cdot_g \nabla V + \rho \omega = -2|\nabla \omega|^2 + 2(\nabla^2 u)(\nabla u, \nabla \omega).$$

(iii) *Set  $\varphi := \frac{1}{2} |\nabla u|^2$ , then it satisfies*

$$\begin{aligned} & -tr \left( \mathcal{A}(x, \nabla u) \circ \bar{\nabla}^2 \varphi \right) + \varepsilon \nabla \varphi \cdot_g \nabla V + 2\rho \varphi = |\nabla \varphi|^2 - |\nabla(\partial_t u)|^2 \\ & - \varepsilon \left( |\nabla^2 u|^2 + (\nabla^2 V)(\nabla u, \nabla u) + Ricc_g(\nabla u, \nabla u) \right). \end{aligned}$$

**Proof** Equations (i) and (ii) are straightforward computations based on equation (4.1). For (i) we use the chain rule. For (ii) we derive in time equation (4.1) and we use that  $\mathcal{A}(x, \nabla u)$  is independent from time.

(iii) We now study the differential equation solved by  $\varphi := \frac{1}{2} |\nabla u|^2$ . We first observe that, if we develop the trace in (4.1), as we did in (3.16), we can rewrite the equation as

$$\partial_{tt}u + \varepsilon \Delta_g u = 2\partial_t \varphi - \nabla u \cdot_g \nabla \varphi + \varepsilon \nabla u \cdot_g \nabla V + \rho u \tag{4.2}$$

Now, using the Bochner’s formula (2.1), which means

$$\Delta_g \varphi = \frac{1}{2} \Delta_g |\nabla u|^2 = |\nabla^2 u|^2 + \nabla(\Delta_g u) \cdot_g \nabla u + Ricc_g(\nabla u, \nabla u) \tag{4.3}$$

we get

$$\begin{aligned} \partial_{tt} \varphi + \varepsilon \Delta_g \varphi &= \nabla(\partial_t u) \cdot_g \nabla(\partial_t u) + \nabla u \cdot_g \nabla(\partial_{tt} u) \\ &+ \varepsilon \left( \nabla(\Delta_g u) \cdot_g \nabla u + |\nabla^2 u|^2 + Ricc_g(\nabla u, \nabla u) \right) \\ &= \nabla(\partial_{tt} u + \varepsilon \Delta_g u) \cdot_g \nabla u + |\nabla(\partial_t u)|^2 + \varepsilon \left[ |\nabla^2 u|^2 + Ricc_g(\nabla u, \nabla u) \right]. \end{aligned}$$

Then, using (4.2), we have

$$\begin{aligned} \partial_{tt} \varphi + \varepsilon \Delta_g \varphi &= 2\nabla(\partial_t \varphi) \cdot_g \nabla u - (\nabla^2 u)(\nabla u, \nabla \varphi) - (\nabla^2 \varphi)(\nabla u, \nabla u) \\ &+ \varepsilon(\nabla^2 u)(\nabla u, \nabla V) + \varepsilon(\nabla^2 V)(\nabla u, \nabla u) + 2\rho \varphi \\ &+ |\nabla(\partial_t u)|^2 + \varepsilon \left[ |\nabla^2 u|^2 + Ricc_g(\nabla u, \nabla u) \right]. \end{aligned}$$

We note that for every vector field  $v$  on  $M$  we have  $\nabla \varphi \cdot_g v = (\nabla^2 u)(\nabla u, v)$ , so the above equality becomes

$$\begin{aligned} \partial_{tt} \varphi + \varepsilon \Delta_g \varphi &= 2\nabla(\partial_t \varphi) \cdot_g \nabla u - \nabla \varphi \cdot_g \nabla \varphi - (\nabla^2 \varphi)(\nabla u, \nabla u) + \varepsilon \nabla \varphi \cdot_g \nabla V \\ &+ \varepsilon(\nabla^2 V)(\nabla u, \nabla u) + 2\rho \varphi + |\nabla(\partial_t u)|^2 \\ &+ \varepsilon \left[ |\nabla^2 u|^2 + Ricc_g(\nabla u, \nabla u) \right]. \end{aligned}$$

Finally, if we look at the whole chain of equalities we get

$$\begin{aligned} &-tr \left( \mathcal{A}(x, \nabla u) \circ \overline{\nabla^2 \varphi} \right) + \varepsilon \nabla \varphi \cdot_g \nabla V + 2\rho \varphi = |\nabla \varphi|^2 - |\nabla(\partial_t u)|^2 \\ &- \varepsilon \left( |\nabla^2 u|^2 + (\nabla^2 V)(\nabla u, \nabla u) + Ricc_g(\nabla u, \nabla u) \right). \end{aligned}$$

□

### 4.1 Global Lipschitz bound on $u$

In this section we will prove a  $W^{1,\infty}$  bound for the solution  $u \in C^3(Q_T)$  of the system

$$\begin{cases} -tr \left( \mathcal{A}(x, \nabla u) \circ \overline{\nabla^2 u} \right) + \rho u + \varepsilon \nabla u \cdot_g \nabla V(x) = 0 & \text{in } Q_T \\ -\partial_t u + \frac{1}{2} |\nabla u|^2 = \delta u + \varepsilon(\log(m_1) + V(x)) & \text{in } t = T, x \in M \\ -\partial_t u + \frac{1}{2} |\nabla u|^2 + \delta u = \varepsilon(\log(m_0) + V(x)) & \text{in } t = 0, x \in M \end{cases} \tag{4.4}$$

where  $\rho, \delta$  are constants,  $0 \leq \rho, \delta \leq 1$ . We recall that  $Q_T = (0, T) \times M$ , and hereafter we denote

$$\Sigma_0 := \{0\} \times M \quad \Sigma_T := \{T\} \times M.$$

We notice that (4.4) is a perturbation of (3.17), containing an additional term  $\rho u$  in the interior (this will simplify a preliminary step, detailed in ‘‘Appendix’’) and an additional term  $\delta u$  on the boundary. The latter is used to control the function  $u$  in a first step. Indeed, using the maximum principle, we have

$$\delta \|u\|_\infty \leq (\|\varepsilon(\log(m_0) + V)\|_\infty + \|\varepsilon(\log(m_1) + V)\|_\infty). \tag{4.5}$$

Analogously, using Lemma 4.1 and the maximum principle, we bound the time derivative. In this way, as in [39, Lemma 3.3], we obtain the following estimate.

**Lemma 4.2** *Let  $u$  be a solution of (4.4), where  $0 \leq \rho, \delta \leq 1$ . It holds that*

$$\partial_t u(t, x) \leq \sup_{\Sigma_0 \cup \Sigma_T} (\partial_t u)_+ \quad \text{and} \quad |\partial_t u(t, x)| \leq \sup_{\Sigma_0 \cup \Sigma_T} |\partial_t u(t, x)| \quad \forall (t, x) \in Q_T \tag{4.6}$$

Finally we give a bound for the space derivative in terms of the sup-norm of  $u$ .

**Theorem 4.3** *Let  $u$  be a solution of (4.4), where  $0 \leq \rho, \delta \leq 1$ . There exists a constant  $C$ , independent from  $\rho$  and  $\delta$ , such that*

$$\|\nabla u\|_\infty \leq C(1 + \|u\|_\infty); \quad \|\partial_t u\|_\infty \leq C(1 + \|u\|_\infty^2). \tag{4.7}$$

Such constant  $C$  depends on  $\|\varepsilon \log(m_0)\|_{W^{1,\infty}}, \|\varepsilon \log(m_1)\|_{W^{1,\infty}}$  and on  $\|\varepsilon V\|_{W^{2,\infty}}$ .

**Proof** We follow P.-L. Lions’ method, as developed in [36, 39]. First of all, we replace  $u$  with the auxiliary function

$$v := u + K - C_0 \left(\frac{T-t}{T}\right), \quad \text{where } K = 2\|u\|_\infty + 1, \quad C_0 = 2K.$$

Notice that  $\nabla v = \nabla u$ , so  $v$  solves the same elliptic equation of  $u$ , up to an additional term due to the time translation. Moreover, we have  $\|v\|_\infty \leq C(1 + \|u\|_\infty)$ . We set

$$z := \frac{1}{2} |\nabla v|^2 + \frac{\gamma}{2} v^2$$

where

$$\gamma := \frac{\sigma}{(1 + \|u\|_\infty)^2}$$

for some small constant  $\sigma$  to be chosen later. The goal is to obtain an upper bound on  $z$  by means of the maximum principle. If the maximum occurs at the boundary, this means that either  $t = 0$  or  $t = T$ ; here one uses that, by construction,  $v(T) \geq 1, v(0) \leq -1$ , then reasoning exactly as in [39] (Thm 3.4, Step 1) one obtains that

$$\|\nabla v\|_\infty \leq C(1 + \|u\|_\infty)$$

for some  $C = C(\|\varepsilon \log(m_1)\|_{W^{1,\infty}}, \|\varepsilon \log(m_0)\|_{W^{1,\infty}}, \|\varepsilon \nabla V\|_\infty)$ , in case of a maximum attained at the extremal times  $t = 0, T$ .

Let us focus on a possibly interior maximum point. From Lemma 4.1, part (i) (applied to  $v$ ) and part (iii), we have

$$\begin{aligned} & -tr \left( \mathcal{A}(x, \nabla u) \circ \overline{\nabla^2 z} \right) + \varepsilon \nabla z \cdot \nabla V + 2\rho z = -\gamma \mathcal{A}(x, \nabla u) ([\partial_t v, \nabla v], [\partial_t v, \nabla v]) \\ & + |\nabla \varphi|^2 - |\nabla(\partial_t v)|^2 - \varepsilon \left( |\nabla^2 u|^2 + Ricc_g(\nabla u, \nabla u) + (\nabla^2 V)(\nabla u, \nabla u) \right) \\ & + \gamma \rho (K - C_0 \frac{T-t}{T}) v \end{aligned} \tag{4.8}$$

where  $\varphi = \frac{1}{2}|\nabla u|^2$ . By definition of the matrix  $\mathcal{A}(x, \nabla u)$ , we get

$$\mathcal{A}(x, \nabla u)([\partial_t v, \nabla v], [\partial_t v, \nabla v]) = |-\partial_t v + |\nabla v|^2|^2 + \varepsilon |\nabla v|^2$$

while the definition of  $\varphi$  implies

$$|\nabla \varphi|^2 = |\nabla z|^2 - 2\gamma v \nabla v \cdot_g \nabla z + \gamma^2 v^2 |\nabla v|^2.$$

Inserting the above equalities in (4.8) we get

$$\begin{aligned} & -tr \left( \mathcal{A}(x, \nabla u) \circ \overline{\nabla^2 z} \right) + \varepsilon \nabla z \cdot_g \nabla V + 2\rho z + \gamma |-\partial_t v + |\nabla v|^2|^2 + \gamma \varepsilon |\nabla v|^2 \\ & = |\nabla z|^2 - 2\gamma v \nabla v \cdot_g \nabla z + \gamma^2 v^2 |\nabla v|^2 - |\nabla(\partial_t v)|^2 \\ & \quad - \varepsilon \left( |\nabla^2 u|^2 + Ricc_g(\nabla u, \nabla u) + (\nabla^2 V)(\nabla u, \nabla u) \right) + \gamma \rho (K - C_0 \frac{T-t}{T}) v \end{aligned}$$

and since  $Ricc_g$  is bounded below, and  $\gamma v^2 \leq C\sigma$  by the initial choice, we can estimate the right-hand side obtaining

$$\begin{aligned} & -tr \left( \mathcal{A}(x, \nabla u) \circ \overline{\nabla^2 z} \right) + \varepsilon \nabla z \cdot_g \nabla V + 2\rho z + \gamma |-\partial_t v + |\nabla v|^2|^2 + \gamma \varepsilon |\nabla v|^2 \\ & \leq |\nabla z|^2 - 2\gamma v \nabla v \cdot_g \nabla z + C(1 + |\nabla v|^2) \end{aligned} \tag{4.9}$$

for some constant  $C$  depending on  $(Ricc_g + \nabla^2 V)_-$  and on  $\sigma$ . Let us focus on the quantity  $-\partial_t v + |\nabla v|^2$ : at an interior maximum point of  $z$ , we have

$$-\partial_t v + |\nabla v|^2 = \frac{(\max z - \partial_t u)}{Q} - \frac{C_0}{T} - \frac{\gamma}{2} v^2 + \frac{1}{2} |\nabla v|^2$$

However, thanks to Lemma 4.2, and to the boundary conditions at  $t = 0, T$ , we have

$$\begin{aligned} \partial_t u & \leq \sup_{\Sigma_0 \cup \Sigma_T} \partial_t u \\ & \leq \max z + \|\delta u\|_\infty + \frac{\gamma}{2} \|v\|_\infty^2 + \|\varepsilon V\|_\infty + \varepsilon \max(\|\log(m_0)\|_\infty, \|\log(m_1)\|_\infty) \end{aligned}$$

which implies, using  $\gamma v^2 \leq C\sigma$  and (4.5),

$$\partial_t u - \max z \leq \overline{K}$$

for a certain  $\overline{K} > 0$  depending only on  $\|\varepsilon V\|_\infty, \|\varepsilon \log(m_1)\|_\infty$  and  $\|\varepsilon \log(m_0)\|_\infty$ . Therefore, we conclude that

$$\begin{aligned} -\partial_t v + |\nabla v|^2 & \geq -\overline{K} - \frac{C_0}{T} - \frac{\gamma}{2} v^2 + \frac{1}{2} |\nabla v|^2 \\ & \geq -K + \frac{1}{2} |\nabla v|^2 \end{aligned}$$

where  $K = \overline{K} + \frac{C_0}{T} + C\sigma$ . So, either  $|\nabla v|^2 \leq 4K$  (and then  $\max z \leq 2K + C\frac{\sigma}{2}$ ) or we have  $-K + \frac{1}{2} |\nabla v|^2 \geq \frac{1}{4} |\nabla v|^2$ , and we estimate  $|-\partial_t v + |\nabla v|^2|^2 \geq \frac{1}{16} |\nabla v|^4$ . In this latter case, looking at (4.9) on a maximum point of  $z$ , where  $\nabla z = 0$ , we deduce that

$$\frac{\gamma}{16} |\nabla v|^4 \leq C(1 + |\nabla v|^2)$$

which implies  $|\nabla v|^2 \leq C(1 + \frac{1}{\gamma}) \leq C(1 + \|u\|_\infty^2)$ . Thus, we conclude in both cases with an estimate like (4.7), for the spatial gradient  $\nabla u$ . Due to Lemma 4.2, this also yields an estimate for  $\partial_t u$ , and then the full gradient  $\bar{\nabla} u$  is estimated.  $\square$

### 4.2 Local bound on the density

We next derive a local (in time) version of the gradient estimate. This is not enough to give a local Lipschitz bound for  $u$ , but provides with a local  $L^\infty$ -bound for the density  $m$ .

In the following, we denote by  $\|(Ric_g + \nabla^2 V)_-\|_\infty$  the nonnegative real number  $R_g := \inf\{r > 0 : Ric_g + \nabla^2 V \geq -r I_g\}$  (in the sense of quadratic forms).

**Proposition 4.4** *Let  $u$  be a solution of (4.4), where  $0 \leq \rho, \delta \leq 1$ .*

*Let  $0 < a < b < T$  and  $\kappa \in (0, \frac{1}{2})$ . There exists a constant  $K > 0$  (independent of  $\rho, \delta$ ) such that*

$$\varepsilon (\log(m(t)) + V) + \kappa |\nabla u|^2 \leq K \left( \frac{1}{(t-a)^2} + \frac{1}{(b-t)^2} \right) \quad \forall t \in (a, b) \quad (4.10)$$

where  $K$  depends only on  $\kappa, T, \|u\|_{L^\infty((a,b) \times M)}$  and  $\|(Ric_g + \nabla^2 V)_-\|_\infty$ .

**Proof** We follow an idea introduced in [39], where a similar result was proved in the Euclidean case.

Let us consider  $z : [a, b] \times M \rightarrow \mathbb{R}$  defined by

$$z := \theta |\nabla u|^2 - \partial_t u + \gamma \frac{u^2}{2}, \quad \text{where } \theta \in (0, 1), \gamma = \frac{1}{(1 + \|u\|_{L^\infty((a,b) \times M)})^2}.$$

By using Lemma 4.1 and denoting  $\varphi = \frac{|\nabla u|^2}{2}$ , we obtain

$$\begin{aligned} & -tr \left( \mathcal{A}(x, \nabla u) \circ \bar{\nabla}^2 z \right) + \varepsilon \nabla z \cdot_g \nabla V + \rho z \\ & \leq -2\theta \varepsilon (|\nabla^2 u|^2 + Ric_g(\nabla u, \nabla u) + (\nabla^2 V)(\nabla u, \nabla u)) + 2|\nabla(\partial_t u)|^2 \\ & \quad - 2\nabla\varphi \cdot_g \nabla(\partial_t u) + 2\theta [|\nabla\varphi|^2 - |\nabla(\partial_t u)|^2] - \gamma \mathcal{A}(x, \nabla u)(\bar{\nabla} u, \bar{\nabla} u) \end{aligned} \quad (4.11)$$

where we notice that

$$\begin{aligned} & 2|\nabla(\partial_t u)|^2 - 2\nabla\varphi \cdot_g \nabla(\partial_t u) + 2\theta [|\nabla\varphi|^2 - |\nabla(\partial_t u)|^2] \\ & = -\nabla(\partial_t u) \cdot_g (2\nabla\varphi - 2\nabla(\partial_t u)) + \theta \nabla|\nabla u|^2 \cdot_g (2\nabla\varphi - 2\nabla(\partial_t u)) - 2\theta |\nabla\varphi - \nabla(\partial_t u)|^2 \\ & = \nabla z \cdot_g (2\nabla\varphi - 2\nabla(\partial_t u)) - 2\theta |\nabla\varphi - \nabla(\partial_t u)|^2 - \gamma u \nabla u \cdot_g (2\nabla\varphi - 2\nabla(\partial_t u)) \\ & \leq \frac{1}{\theta} |\nabla z|^2 + \frac{1}{\theta} \gamma^2 u^2 |\nabla u|^2. \end{aligned}$$

By construction of  $\mathcal{A}(x, \nabla u)$ , we have

$$\mathcal{A}(x, \nabla u)(\bar{\nabla} u, \bar{\nabla} u) = |\nabla u|^2 - \partial_t u|^2 + \varepsilon |\nabla u|^2$$

so that, inserting the above estimates in (4.11), we obtain

$$\begin{aligned} & -tr \left( \mathcal{A}(x, \nabla u) \circ \bar{\nabla}^2 z \right) + \varepsilon \nabla z \cdot_g \nabla V + \rho z \\ & \leq -2\theta \varepsilon (|\nabla^2 u|^2 + Ric_g(\nabla u, \nabla u) + (\nabla^2 V)(\nabla u, \nabla u)) \\ & \quad + \frac{1}{\theta} |\nabla z|^2 + \frac{1}{\theta} \gamma^2 u^2 |\nabla u|^2 - \gamma |\nabla u|^2 - \partial_t u|^2 - \gamma \varepsilon |\nabla u|^2. \end{aligned}$$

Using  $\gamma u^2 \leq 1$  and the regularity of  $V$  we get

$$\begin{aligned}
 & -tr \left( \mathcal{A}(x, \nabla u) \circ \bar{\nabla}^2 z \right) + \varepsilon \nabla z \cdot_g \nabla V + \rho z \\
 & \leq \left( \frac{\gamma}{\theta} + \hat{C} \right) |\nabla u|^2 + \frac{1}{\theta} |\nabla z|^2 - \gamma |\nabla u|^2 - \partial_t u|^2
 \end{aligned} \tag{4.12}$$

where  $\hat{C}$  is a constant depending on the lower bound of  $\varepsilon(Ric_g + \nabla^2 V)$ . Given  $L > 0$ , let

$$\psi := L \left( \frac{1}{(t-a)^2} + \frac{1}{(b-t)^2} \right)$$

Since  $\psi$  blows up at  $t = a, b$ , we have that  $z - \psi$  admits a maximum point in  $(a, b) \times M$ . In such point we have  $\nabla z = 0$  and

$$-tr \left( \mathcal{A}(x, \nabla u) \circ \bar{\nabla}^2 z \right) \geq -tr \left( \mathcal{A}(x, \nabla u) \bar{\nabla}^2 \psi \right) = -6L \left( \frac{1}{(t-a)^4} + \frac{1}{(b-t)^4} \right).$$

If  $L_0 := \max(z - \psi) \geq \frac{1}{2}$ , then, at a maximum point of  $z - \psi$  we have

$$\begin{aligned}
 |\nabla u|^2 - \partial_t u &= (1 - \theta) |\nabla u|^2 + z - \gamma \frac{u^2}{2} \geq (1 - \theta) |\nabla u|^2 + \psi + L_0 - \frac{1}{2} \\
 &\geq (1 - \theta) |\nabla u|^2 + \psi > 0
 \end{aligned}$$

so

$$||\nabla u|^2 - \partial_t u|^2 \geq (1 - \theta)^2 |\nabla u|^4 + \psi^2.$$

Using all of these in (4.12), we get

$$-6L \left( \frac{1}{(t-a)^4} + \frac{1}{(b-t)^4} \right) \leq \left( \frac{\gamma}{\theta} + \hat{C} \right) |\nabla u|^2 - \gamma (1 - \theta)^2 |\nabla u|^4 - \gamma \psi^2$$

which gives

$$\begin{aligned}
 (\gamma L^2 - 6L) \left( \frac{1}{(t-a)^4} + \frac{1}{(b-t)^4} \right) &\leq \left( \frac{\gamma}{\theta} + \hat{C} \right) |\nabla u|^2 - \gamma (1 - \theta)^2 |\nabla u|^4 \\
 &\leq c_\theta \left( \gamma + \frac{\hat{C}^2}{\gamma} \right)
 \end{aligned}$$

for some  $c_\theta$  only depending on  $\theta$ . But the inequality above cannot hold for any too large  $L$ ; handling constants with care, this occurs for  $L = O([((b-a)^2 \vee 1] (1 + \|u\|_{L^\infty((a,b) \times M)}^2)))$ .

The conclusion is that we have  $\max(z - \psi) \leq \frac{1}{2}$ , hence

$$z \leq L \left( \frac{1}{(t-a)^2} + \frac{1}{(b-t)^2} \right) + \frac{1}{2}.$$

Using the definition of  $z$  leads to

$$\theta |\nabla u|^2 - \partial_t u \leq L \left( \frac{1}{(t-a)^2} + \frac{1}{(b-t)^2} \right) + \frac{1}{2}$$

and choosing  $\theta > \frac{1}{2}$ , from  $\partial_t u = \frac{1}{2} |\nabla u|^2 - \varepsilon(\log(m) + V)$  we obtain (4.10) with  $\kappa = \theta - \frac{1}{2} < \frac{1}{2}$ . □

### 4.3 Existence of smooth solutions for a penalized problem

We first collect all the above ingredients to show that a penalized version of the optimality system admits a classical solution. We consider, for  $\delta > 0$ , the problem

$$\begin{cases} -tr \left( \mathcal{A}(x, \nabla u) \circ \overline{\nabla}^2 u \right) + \varepsilon \nabla u \cdot_g \nabla V(x) = 0 & \text{in } Q_T \\ -\partial_t u + \frac{1}{2} |\nabla u|^2 = \delta u + \varepsilon (\log(m_1) + V(x)) & \text{in } \Sigma_T \\ -\partial_t u + \frac{1}{2} |\nabla u|^2 + \delta u = \varepsilon (\log(m_0) + V(x)) & \text{in } \Sigma_0. \end{cases} \tag{4.13}$$

which is equivalent, reasoning as in Sect. 3.2, to the system

$$\begin{cases} -\partial_t u + \frac{1}{2} |\nabla u|^2 = \varepsilon (\log(m) + V) & \text{in } Q_T \\ \partial_t m - div_g(m \nabla u) = 0 & \text{in } Q_T \\ \delta u(T) = \varepsilon \log(m(T)) - \varepsilon \log(m_1) & \text{in } \Sigma_T \\ \delta u(0) = \varepsilon \log(m_0) - \varepsilon \log(m(0)) & \text{in } \Sigma_0. \end{cases} \tag{4.14}$$

The auxiliary penalized problem (4.14) has the advantage that the  $L^\infty$ -norm of  $u$  is controlled, for  $\delta > 0$ , see (4.5). The estimates derived from the elliptic theory (see ‘‘Appendix A’’) yield a smooth solution, with  $m > 0$ , provided the marginals are positive and smooth.

**Theorem 4.5** *Assume that  $V \in W^{2,\infty}(M)$ ,  $m_0, m_1 \in C^{1,\alpha}(M)$  and  $m_0, m_1 > 0$  in  $M$ . For every  $\delta > 0$ , there exists a unique smooth solution  $(u^\delta, m^\delta)$  of (4.14), in the sense that  $u \in C^{2,\alpha}(Q_T) \cap C^{1,\alpha}(\overline{Q}_T)$ ,  $m \in C^{1,\alpha}(Q_T) \cap C^{0,\alpha}(\overline{Q}_T)$ ,  $m > 0$  and the equations are satisfied in a classical sense.*

**Proof** We first rely on Proposition 7.1 which is proved in the ‘‘Appendix’’. This gives a sequence of solutions  $u_\rho$  of problem (4.4). By maximum principle, we have  $\|u_\rho\|_\infty \leq \frac{C}{\delta}$ . By the gradient bound proved in Theorem 4.3, we have that  $\|\overline{\nabla} u_\rho\|_\infty \leq C$ . By elliptic estimates (see also Lemma 7.2 below) we deduce first that  $\|u_\rho\|_{C^{1,\alpha}(\overline{Q}_T)} \leq C$ , and then, bootstrapping Schauder estimates,  $u_\rho$  is bounded in  $C^{2,\alpha}$  on any compact subset of  $Q_T$ . Defining

$$m_\rho = \exp \left( \frac{-\partial_t u_\rho + \frac{1}{2} |\nabla u_\rho|^2}{\varepsilon} - V \right)$$

we deduce the  $C^{1,\alpha}(Q_T) \cap C^{0,\alpha}(\overline{Q}_T)$  estimates on  $m_\rho$  and, in particular,  $m_\rho$  is uniformly bounded below due to the gradient bounds for  $u_\rho$ . Passing to the limit as  $\rho \rightarrow 0$  gives the desired solution of (4.13), hence of (4.14).  $\square$

The next step will consist in letting  $\delta \rightarrow 0$  in (4.14), still assuming that the marginals  $m_0, m_1$  are positive, showing that the minima of (1.1) are smooth for positive smooth marginals. This step is left to the stability results of the next section.

## 5 Existence and regularity of optimal curves

In this section we obtain the existence and the characterization of the minima of (1.1) in terms of the optimality system (1.2), thus proving Theorem 1.1 stated in the Introduction. We first obtain the existence of smooth minima, whenever the marginals  $m_0, m_1$  are positive and smooth; this is achieved by passing to the limit as  $\delta \rightarrow 0$  in problem (4.14) and using the ‘‘elliptic’’ Lipschitz estimates of Theorem 4.3. Then we will enlarge the set of admitted

marginals  $m_0$  and  $m_1$  to merely  $L^1$ , nonnegative densities. To this purpose, we will need a relaxed definition of weak solution to the system (3.1), where merely sub-solutions of the Hamilton–Jacobi equation are taken into account. This kind of notion of weak solutions, introduced in [7] (see also [8, 9]) for first-order mean-field game systems, is by now well established also in the context of mean-field transport problems, see e.g. [20, 37]. In particular, in this latter paper a notion of trace was developed for functions which are (distributional) sub-solutions of Hamilton–Jacobi equations, e.g. when  $u$  satisfies

$$-\partial_t u + \frac{1}{2} |\nabla u|^2 \leq \alpha, \quad \alpha \in L^1_{loc}((0, T) \times \Omega) \tag{5.1}$$

in some open set  $\Omega$ . Relying on the fact that  $u$  is nondecreasing in time, up to an absolutely continuous function, one-sided traces of  $u$  were defined in the sense of limits of measurable functions (with possible range in  $[-\infty, +\infty]$ ), see [37, Prop 5.6]. In particular, if  $\alpha \in L^1(Q_T)$ , any  $u$  satisfying (5.1) admits traces at  $t = 0, t = T$ , denoted below as  $u(0), u(T)$  respectively, which are the pointwise limits of  $u(t, \cdot)$  as  $t \downarrow 0$  (respectively  $t \uparrow T$ ) in the sense of measurable functions (for a possibly well-defined precise representative of  $u$ ).

**Definition 5.1** A pair  $(u, m)$  is a weak solution of

$$\begin{cases} -\partial_t u + \frac{1}{2} |\nabla u|^2 = \varepsilon(\log(m) + V(x)) & \text{in } Q_T \\ \partial_t m - \operatorname{div}_g(m \nabla u) = 0 & \text{in } Q_T \\ m(0, \cdot) = m_0, \quad m(T, \cdot) = m_1 & \text{in } M \end{cases} \tag{5.2}$$

if  $m \in C^0([0, T]; \mathcal{P}(M)) \cap L^1(Q_T)$  with  $m(0) = m_0, m(T) = m_1$  and  $\log(m) \in L^1_{loc}((0, T); L^1(M))$ ;

$u \in L^2_{loc}((0, T); H^1(M))$  and in addition  $m |\nabla u|^2 \in L^1(Q_T), m \log m \in L^1(Q_T)$  and  $(u, m)$  satisfy

- (i)  $u$  is a weak sub-solution satisfying, in the sense of distributions,

$$-\partial_t u + \frac{1}{2} |\nabla u|^2 \leq \varepsilon(\log(m) + V(x)) \quad \text{in } Q_T$$

- (ii)  $m$  is a weak solution satisfying, in the sense of distributions, the continuity equation

$$\partial_t m - \operatorname{div}_g(m \nabla u) = 0 \quad \text{in } Q_T$$

- (iii)  $(u, m)$  satisfy the identity

$$\int_M m_0 u(0) - \int_M m_1 u(T) = \int_0^T \int_M \left( \frac{1}{2} |\nabla u|^2 m + \varepsilon m(\log m + V) \right) dt \tag{5.3}$$

where  $u(0), u(T)$  are the one-sided traces of  $u$  mentioned above.

The existence of weak solutions to (5.2) will be produced through a stability argument, which relies on several steps. Some crucial a priori estimates are given in the following lemma, which applies to solutions of both (4.14) and (5.2) (corresponding to  $\delta = 0$ ).

**Lemma 5.2** *Assume that  $(u, m)$  is a smooth solution to (4.14), for some  $\delta \in [0, 1]$ , and define  $\hat{u} := u - \int_M u(T) m_1$ . Let  $\lambda$  be given by (2.2). There exists a constant  $\hat{C} = \hat{C}(M, d, \lambda, T, \|V\|_\infty)$  such that*

$$\begin{aligned} -\frac{\hat{C}}{t} &\leq \hat{u}(t, x) \quad \forall (t, x) \in (0, T) \times M, \\ \hat{u}(t, x) &\leq \frac{\hat{C}}{T-t} \quad \forall (t, x) \in [0, T) \times M \end{aligned} \tag{5.4}$$

and

$$\int_M \hat{u}(0)m_0 \geq -\varepsilon \log \left( \int_M e^{-V} \right) \tag{5.5}$$

$$\int_0^T \int_M (m |\nabla u|^2 + \varepsilon m \log m) dt \leq \hat{C}.$$

In addition, for any  $0 < a < b < T$  there exists  $C$ , depending on  $\hat{C}$  and additionally on  $a, (T - b)$ , such that

$$\int_a^b \int_M |\nabla u|^2 dt + \varepsilon \int_a^b \int_M |\log m| dt \leq C. \tag{5.6}$$

Finally, if  $m_0, m_1 > 0$  on  $M$ , then there exists  $K$ , depending on  $\hat{C}$  and additionally on  $\min_M [\varepsilon \log(m_0 e^V)]$  and  $\min_M [\varepsilon \log(m_1 e^V)]$ , such that

$$\|\hat{u}\|_{L^\infty([0,T] \times M)} < K. \tag{5.7}$$

**Proof** By definition of  $\hat{u}$ , we have

$$\int_0^T \int_M \left( \frac{1}{2} |\nabla u|^2 m + \varepsilon m (\log m + V) \right) dt = \int_M m(0)u(0) - \int_M m(T)u(T)$$

$$= \int_M (m(0) - m_0)u(0) - \int_M (m(T) - m_1)u(T) + \int_M m_0 \hat{u}(0)$$

This implies, using the initial-terminal conditions of (4.14),

$$\int_M m_0 \hat{u}(0) = \int_0^T \int_M \left( \frac{1}{2} |\nabla u|^2 m + \varepsilon m (\log m + V) \right) dt$$

$$+ \frac{\varepsilon}{\delta} \int_M (m_0 - m(0)) [\log(m_0) - \log(m(0))]$$

$$+ \frac{\varepsilon}{\delta} \int_M (m(T) - m_1) [\log m(T) - \log(m_1)], \tag{5.8}$$

where last two terms should be treated as zero in case that  $\delta = 0$ . In particular, since the right-side is bounded below, we deduce that

$$\int_M m(0) \hat{u}(0) \geq \int_0^T \int_M \left( \frac{1}{2} |\nabla u|^2 m + \varepsilon m (\log m + V) \right) dt$$

$$\geq -\varepsilon T \log \left( \int_M e^{-V} \right) \tag{5.9}$$

where we used Jensen’s inequality in the last step. Of course,  $\hat{u}$  satisfies the same Hamilton–Jacobi equation as  $u$ , from which we derive now all the estimates.

At first, for any fixed  $x_0 \in M$ , letting  $\delta_{x_0}$  be the Dirac delta concentrated at  $x_0$ , we consider the solution  $\mu$  to the continuity equation

$$\begin{cases} \partial_t \mu - \operatorname{div}_g(\mu v) = 0 & \text{in } (s, T) \times M \\ \mu(T) = m_1, \mu(s) = \delta_{x_0} & \text{in } M \end{cases}$$

which was built in Proposition 2.2. Such a curve satisfies the estimate

$$\int_s^T \int_M \left( \frac{|v|^2}{2} \mu + \varepsilon \mu (\log \mu + V) \right) dt \leq \frac{K}{T - s}$$

with  $K$  depending only on  $M, d, \lambda, T$  and  $\|V\|_\infty$ , but independent of  $x_0$ . We multiply by  $\mu$  the equation of  $\hat{u}$  and we integrate over  $(s, T) \times M$ : we get

$$\begin{aligned} \hat{u}(x_0) &= \int_s^T \int_M v \cdot \nabla u \mu - \int_s^T \int_M \frac{1}{2} |\nabla u|^2 \mu + \varepsilon \int_s^T \int_M (\log(m) + V) \mu dt \\ &\leq \int_s^T \int_M \frac{1}{2} |v|^2 \mu + \varepsilon \int_s^T \int_M (\log(m) + V) \mu dt. \end{aligned}$$

Using that  $a \log(b) \leq a \log(a) + b$  for every  $a, b \in (0, +\infty)$  we obtain

$$\begin{aligned} \hat{u}(x_0) &\leq \int_s^T \int_M \frac{1}{2} |v|^2 \mu + \varepsilon \int_s^T \int_M \mu (\log(\mu) + V) dt + \varepsilon(T - s) \\ &\leq \frac{K}{T - s} + \varepsilon(T - s). \end{aligned}$$

So there exists a constant  $\hat{C} > 0$ , depending on  $T, \|V\|_\infty$  and  $M$ , but independent of  $\varepsilon$  and  $x_0$ , such that for every  $t \in [0, T)$

$$\hat{u}(t, x_0) \leq \frac{\hat{C}}{T - t}. \tag{5.10}$$

Reasoning in a similar way, namely using an analogous curve between  $m_0$  and  $\delta_{x_0}$ , and the bound (5.9), we conclude that there exists a constant  $\hat{C} > 0$  such that for every  $t \in (0, T]$

$$-\frac{\hat{C}}{t} \leq \hat{u}(t, x_0). \tag{5.11}$$

By the arbitrariness in the choice of  $x_0 \in M$ , we get (5.4).

Fix now  $0 < a < b < T$ , from the equation of  $\hat{u}$  we have

$$\begin{aligned} \int_M \hat{u}(a, \cdot) - \int_M \hat{u}(b, \cdot) + \int_a^b \int_M \frac{1}{2} |\nabla u|^2 dt \\ \leq -\varepsilon \int_a^b \int_M |\log(m)| dt + \varepsilon c_M (1 + \|V\|_\infty). \end{aligned}$$

Using (5.10) and (5.11), we obtain estimate (5.6). Coming back to (5.9) and using the upper bound (5.10), we also get (5.5).

Finally, to get (5.7) we use the comparison principle on the Eq. (4.13). In fact, the linear function  $\psi := \frac{\hat{C}}{T} + Bt$  is clearly a solution of the same equation, and it is a strict supersolution at  $t = T$  if  $\varepsilon(\log(m_1 e^V)) > -B$ . This is possible if  $m_1 > 0$ , choosing  $B$  sufficiently large. Since  $\psi(0) > \hat{u}(0)$  by (5.10), one can compare  $\hat{u}$  with  $\psi$  (note that  $\hat{u}$  satisfies the same condition as  $u$  at  $t = T$  up to a bounded term, and  $B$  can be chosen possibly larger to compensate). Hence  $\hat{u} \leq \frac{\hat{C}}{T} + Bt$ . Similarly we reason to get the estimate from below using the positivity of  $m_0$ , and we conclude with the global bound (5.7).  $\square$

We notice that the local bounds of  $\hat{u}$ , given by (5.4), are totally independent of  $m$ . This allows us to infer the local boundedness of  $m$  too, as a consequence of Proposition 4.4.

**Corollary 5.3** *Under the assumptions of Lemma 5.2, given any  $0 < a < b < T$  there exists a constant  $C = C(M, d, \lambda, T, \|V\|_\infty, \|(Ric_g + D^2V)_-\|_\infty, \varepsilon, a, T - b)$  such that*

$$\|m(t)\|_{L^\infty((a,b) \times M)} \leq C.$$

**Proof** By estimates (5.4), we have  $\|\hat{u}\|_{L^\infty((a,b)\times M)} \leq \hat{C} \max(a^{-1}, (T - b)^{-1})$ . Hence, from (4.10) we deduce that  $\varepsilon \log(m(t))$  is bounded above for  $t \in (a, b)$  by some constant depending on  $\hat{C}, T, \max(a^{-1}, (T - b)^{-1})$  and  $\|(Ric_g + D^2V)_-\|_\infty$ , which yields the conclusion.  $\square$

Thanks to the global bound (5.7), we can now pass to the limit as  $\delta \rightarrow 0$  in (4.14), obtaining the existence of smooth minimizers for positive smooth marginals.

**Theorem 5.4** *Assume that  $m_0, m_1 \in W^{1,\infty}(M)$  and  $m_0, m_1 > 0$  in  $M$ . Let  $V \in W^{2,\infty}(M)$ ,  $\varepsilon > 0$ . Then there exists a (unique) smooth solution  $(u, m)$  of the system (1.2) such that  $\int_M u(T)m_1 = 0$ ,  $u \in C^{2,\alpha}(Q_T) \cap C^{1,\alpha}(\bar{Q}_T)$ ,  $m \in C^{1,\alpha}(Q_T) \cap C^{0,\alpha}(\bar{Q}_T)$ ,  $\alpha \in (0, 1)$ . In addition we have  $m > 0$  in  $\bar{Q}_T$ ,  $u$  is a classical solution of the elliptic equation (3.17), and  $(m, \nabla u)$  is the unique minimum of the functional  $\mathcal{F}_\varepsilon$  in (1.1).*

*Finally, if  $V \in C^{k,\alpha}(M)$ , we have  $u \in C^{k+1,\alpha}(M)$ ,  $m \in C^{k,\alpha}(M)$ .*

**Proof** In a first step, we take  $m_0, m_1 \in C^{1,\alpha}(M)$  and positive on  $M$ . By Theorem 4.5, problem (4.14) admits a smooth solution  $(u^\delta, m^\delta)$ , and we set as before  $\hat{u}^\delta = u^\delta - \int_M u^\delta(T)m_1$ . By (5.7), we have that  $\hat{u}^\delta$  is uniformly bounded in  $Q_T$ . Then, by Theorem 4.3, we deduce that  $\hat{u}^\delta$  is uniformly bounded in Lipschitz norm (time-space). Since  $m^\delta = e^{-V} \exp\{\frac{1}{\varepsilon}(|\nabla \hat{u}^\delta|^2/2 - \partial_t \hat{u}^\delta)\}$ , the gradient bound on  $\hat{u}^\delta$  implies that  $m^\delta$  is uniformly bounded both from above and from below by a positive constant. By elliptic estimates (same as in Lemma 7.2), we have that  $\hat{u}^\delta$  is bounded in  $C^{1,\alpha}(\bar{Q}_T)$  and the bound only depends on  $\|\log(m_0)\|_{W^{1,\infty}(M)}, \|\log(m_1)\|_{W^{1,\infty}(M)}$  (and of course on  $\varepsilon, \|V\|_{W^{2,\infty}}$ ). We also have interior local bounds on  $\hat{u}^\delta$  in  $C^{2,\alpha}$ , because the elliptic equation have coefficients bounded in  $C^{0,\alpha}$ . Therefore, by compactness, we deduce that  $\hat{u}^\delta$  converges to some  $u \in C^{2,\alpha}(Q_T) \cap C^{1,\alpha}(\bar{Q}_T)$ , which is a classical solution of the elliptic equation (3.17). At the boundary, e.g. at  $t = T$ , we have

$$\begin{aligned} -\partial_t \hat{u}^\delta + \frac{1}{2} |\nabla \hat{u}^\delta|^2 &= \varepsilon(\log(m_1) + V(x)) + \delta \hat{u}^\delta + \delta \int_M u^\delta(T)m_1 \\ &= \varepsilon(\log(m_1) + V(x)) + \delta \hat{u}^\delta + \varepsilon \int_M (\log(m^\delta(T)) - \log(m_1))m_1. \end{aligned} \tag{5.12}$$

However, by (5.8) and the bound on  $\hat{u}^\delta$ , we know that

$$\int_M (\log(m^\delta(T)) - \log(m_1))(m^\delta(T) - m_1) \leq C \delta \xrightarrow{\delta \rightarrow 0} 0$$

which implies that  $m^\delta(T) \rightarrow m_1$ . In particular, last term in (5.12) vanishes as  $\delta \rightarrow 0$  by dominated convergence, since  $\log(m^\delta)$  is uniformly bounded. Since we also have  $\delta \hat{u}^\delta \rightarrow 0$  we conclude that  $u$  satisfies the boundary condition at  $t = T$

$$-\partial_t u + \frac{1}{2} |\nabla u|^2 = \varepsilon(\log(m_1) + V(x)).$$

We argue in the same way for  $t = 0$  and we conclude that the limit  $u$  satisfies the elliptic problem (4.4) with  $\rho = 0, \delta = 0$ . Furthermore, we notice that  $u$  satisfies an elliptic equation where the second-order coefficients only depend on  $\nabla u$ , and the first-order coefficients depend on  $\nabla V$ . Hence, by the interior Schauder regularity and a bootstrap argument, we have  $u \in C^{k+1,\alpha}(M)$  provided  $V \in C^{k,\alpha}(M), k \geq 1$ .

Finally, defining

$$m := e^{-V(x)} \exp\left(\frac{-\partial_t u + \frac{1}{2}|\nabla u|^2}{\varepsilon}\right),$$

we have  $m \in C^{1,\alpha}(Q_T) \cap C^{0,\alpha}(\overline{Q_T})$ ,  $m > 0$  and  $m(0) = m_0, m(T) = m_1$ . In other words,  $(u, m)$  is a smooth solution of system (1.2) (unique, with the normalization of  $u$ ), and is the minimum of the functional  $\mathcal{F}_\varepsilon$ . This is also the unique minimum, as we will prove in more generality in our next results. As a last remark, the result extends by approximation to positive marginals  $m_0, m_1 \in W^{1,\infty}(M)$  since in fact all the estimates above remain true.  $\square$

We will now extend the existence result to the case of general  $L^1$  marginals. We will need some more a priori estimates and new compactness arguments. To this purpose, we first make use of the displacement convexity estimates of Proposition 3.3 to derive a local  $L^2$ -bound on  $\nabla m$ .

**Lemma 5.5** *Let  $(u, m)$  be a smooth solution of the system (3.1). For every  $0 < a < b < T$  there exist a constant  $C = C(M, d, \lambda, T, b - a, \|V\|_{W^{1,\infty}})$  such that*

$$\int_a^b \int_M |\nabla \sqrt{m}|^2 dt \leq C \frac{|\log \varepsilon|}{\varepsilon}.$$

**Proof** First, we recall inequality (3.11) which implies, using also (3.5),

$$\frac{d^2}{dt^2} \int_M m \log m \geq -2\lambda\varepsilon \int_M m \log m + \frac{\varepsilon}{2} \int_M \frac{1}{m} |\nabla m|^2 - C$$

for some  $C$  depending on  $M, d, \lambda, T, \|V\|_{W^{1,\infty}}$ , and independent of  $\varepsilon \leq 1$ . Now we fix  $t_0 \in (0, T)$ , and  $R < R_0 := \min(t_0, T - t_0)$ ; then, for  $\tau \in (0, R)$  we let  $\xi(t)$  be a smooth cut-off function such that

$$\begin{cases} \xi(t) = 1 & \text{if } t \in (t_0 - \tau, t_0 + \tau) \\ \xi(t) = 0 & \text{if } |t - t_0| > R \\ |\xi'(t)|^2 + |\xi''(t)| \leq \alpha_\xi \end{cases}$$

for a certain  $\alpha_\xi > 0$ . Then we have

$$\begin{aligned} \frac{d^2}{dt^2} \left( \xi^2 \int_M m \log m \right) &\geq -2\lambda\varepsilon\xi^2 \int_M m \log m + \frac{\varepsilon}{2} \int_M \xi^2 \frac{|\nabla m|^2}{m} - C \xi^2 \\ &\quad + 4\xi\xi' \frac{d}{dt} \int_M m \log m + 2(\xi'^2 + \xi\xi'') \int_M m \log m. \end{aligned}$$

If we integrate in  $(0, T)$  both sides we get

$$\varepsilon \int_0^T \int_M \xi^2 |\nabla \sqrt{m}|^2 dt \leq C_{t_0, T} \left[ \int_0^T \int_M m \log m dt + C \right],$$

for a possibly different constant  $C$ . Since  $\int_0^T \int_M m \log(m)$  is estimated by (3.9), we conclude.  $\square$

We will also use the following stability result.

**Lemma 5.6** *Let  $(u_n, m_n)$  be a sequence of smooth solutions of (3.1), possibly for different parameters  $\varepsilon_n \rightarrow \varepsilon \geq 0$ . Assume that  $u_n$  satisfies (5.4) for some constant  $\hat{C}$  independent of  $n$ . Let  $(u, m)$  be such that  $u_n \rightarrow u$  weakly in  $L^2((a, b); H^1(M))$ , for any  $0 < a < b < T$ , and  $m_n \rightarrow m$  weakly in  $L^1(Q_T)$ . Then we have:*

1.  $u$  satisfies, in the sense of distributions,

$$-\partial_t u + \frac{1}{2} |\nabla u|^2 \leq \varepsilon (\log m + V) \text{ in } (0, T) \times M. \tag{5.13}$$

2. For every sequences  $m_{0n}, m_{1n}$  such that  $m_{0n} \rightarrow m_0, m_{1n} \rightarrow m_1$  strongly in  $L^1(M)$ , we have

$$\begin{aligned} \limsup_{n \rightarrow \infty} \int_M u_n(0) m_{0n} &\leq \int_M u(0) dm_0, \\ \liminf_{n \rightarrow \infty} \int_M u_n(T) m_{1n} &\geq \int_M u(T) dm_1. \end{aligned}$$

3. For every  $(\mu, v)$  which solves (2.3) and such that  $\mu(t) \in L^1(M)$  for every  $t$  and  $\mu \log \mu \in L^1(Q_T)$ , it holds

$$\begin{aligned} \int_M u(s) \mu(s) - \int_M u(t) \mu(t) &\leq \int_s^t \int_M [\mu v \cdot_g \nabla u - \frac{1}{2} |\nabla u|^2 \mu] d\tau \\ &+ \varepsilon \int_s^t \int_M (\log m + V) \mu d\tau \end{aligned} \tag{5.14}$$

for every  $0 \leq s < t \leq T$ .

**Remark 5.7** We notice that (5.13) implies

$$-\partial_t u + \frac{1}{2} |\nabla u|^2 \leq \varepsilon (m + V) \in L^1(Q_T)$$

hence  $u$  satisfies (5.1) for some  $\alpha \in L^1(Q_T)$  and admits one-sided traces up to  $t = 0$  and  $t = T$  (those traces are used in items 2,3 above).

**Proof** By definition,  $u_n$  satisfies

$$\int_0^T \int_M u_n \partial_t \varphi dt + \frac{1}{2} \int_0^T \int_M |\nabla u_n|^2 \varphi dt = \int_0^T \int_M \varepsilon (\log(m_n) + V) \varphi dt$$

for every  $\varphi \in C_c^{1,0}(Q_T)$ . In particular the integrals are restricted to some interval  $(a, b)$  containing the support of  $\varphi$ , with  $0 < a < b < T$ . Then, by simply using the weak convergence of  $\nabla u_n$  in  $L^2((a, b) \times M)$  and  $m_n$  in  $L^1((a, b) \times M)$ , and the weak lower semicontinuity for the convex functions  $|p|^2$  and  $-\log(m)$ , respectively, we conclude that  $u$  satisfies (5.13), in distributional sense.

To observe the relaxation on the initial traces, we fix  $k \in \mathbb{N}$  and consider the truncations  $u_{n,k} := \max\{u_n, -k\}$ . By a standard argument in Sobolev spaces (see e.g. Lemma 5.3 in [37]), we have that  $u_{n,k} := \max\{u_n, -k\}$  are Lipschitz functions that satisfy

$$\begin{aligned} -\partial_t u_{n,k} + \frac{1}{2} |\nabla u_{n,k}|^2 &\leq \mathbb{1}_{\{u_n \geq -k\}} \varepsilon_n (\log m_n + V) \\ &\leq \varepsilon_n (m_n + |V|). \end{aligned} \tag{5.15}$$

For every  $k \in \mathbb{N}$ , let  $\xi_k$  be the piecewise linear function such that

$$\begin{cases} \xi_k(s) = 1 & \forall s \in \left[0, \frac{\hat{C}}{k}\right] \\ \xi_k(s) = 0 & \forall s \in \left[2\frac{\hat{C}}{k}, T\right] \\ \xi'_{\delta,k}(s) = -\frac{k}{\hat{C}} & \forall s \in \left[\frac{\hat{C}}{k}, 2\frac{\hat{C}}{k}\right] \end{cases}$$

where  $\hat{C}$  is the same constant which appears in (5.4). We now fix a positive  $\varphi \in L^\infty(M)$  and we multiply (5.15) by  $\phi = \xi_k \varphi$ . Integrating by parts the time derivative we get

$$\int_M u_{n,k}(0, \cdot) \varphi - \frac{k}{\hat{C}} \int_{\frac{\hat{C}}{k}}^{2\frac{\hat{C}}{k}} \int_M u_{n,k} \varphi dt \leq C \frac{\|\varphi\|_\infty}{k}.$$

The sequence  $u_{n,k}(0, \cdot)$  is uniformly bounded in  $n$  by construction, let  $\chi_k$  be its weak-\* limit, up to subsequences, in  $L^\infty(M)$ . By (5.4), we note that  $u_{n,k} = u_n$  for  $t \geq \frac{\hat{C}}{k}$ . Then we pass to the limit for  $n \rightarrow +\infty$ , using the weak convergence of  $u_n$ , and we get

$$\int_M \chi_k \varphi - \frac{k}{\hat{C}} \int_{\frac{\hat{C}}{k}}^{2\frac{\hat{C}}{k}} \int_M u \varphi dt \leq C \frac{\|\varphi\|_\infty}{k}. \tag{5.16}$$

From Eq. (5.13), setting  $F(t, x) := \varepsilon \int_0^t (m(s, x) + V(x)) ds$ , we know that  $u + F$  is nondecreasing in time. Hence

$$\frac{k}{\hat{C}} \int_{\frac{\hat{C}}{k}}^{2\frac{\hat{C}}{k}} \int_M u \varphi dt \leq \int_M u(2\frac{\hat{C}}{k}, \cdot) \varphi - \frac{k}{\hat{C}} \int_{\frac{\hat{C}}{k}}^{2\frac{\hat{C}}{k}} \int_M \left( F(t, x) - F(2\frac{\hat{C}}{k}, x) \right) \varphi dt$$

and last term vanishes as  $k \rightarrow \infty$  because  $F$  is a primitive of an  $L^1$  function. Therefore, we have

$$\limsup_{k \rightarrow \infty} \frac{k}{\hat{C}} \int_{\frac{\hat{C}}{k}}^{2\frac{\hat{C}}{k}} \int_M u \varphi dt \leq \limsup_{k \rightarrow \infty} \int_M u(2\frac{\hat{C}}{k}, \cdot) \varphi \leq \int_M u(0) \varphi$$

where we used the pointwise convergence of  $u(t, \cdot)$  as  $t \rightarrow 0^+$  and the fact that  $u$  is bounded above by (5.4), which allows us to apply Fatou’s lemma. Finally, letting  $k \rightarrow \infty$  in (5.16), we obtain

$$\limsup_{k \rightarrow \infty} \int_M \chi_k \varphi \leq \int_M u(0) \varphi \tag{5.17}$$

for every  $\varphi \in L^\infty(M)$ . Let us define now  $T_j(f) = \min(f, j)$  the truncation operator in  $L^1(M)$ . We recall that  $m_{0n}$  converges strongly to  $m_0$  in  $L^1(M)$  and  $u(0)$  is bounded above, so

$$\begin{aligned} \limsup_n \int_M m_{0n} u_n(0) &\leq \limsup_n \int_M T_j(m_{0n}) u_n(0) \\ &\quad + \limsup_n \int_M (m_{0n} - T_j(m_{0n})) u_n(0) \\ &\leq \limsup_n \int_M T_j(m_{0n}) u_{n,k}(0) \\ &\quad + \frac{\hat{C}}{T} \limsup_n \int_M (m_{0n} - T_j(m_{0n})) \\ &\leq \int_M T_j(m_0) \chi_k dx + \frac{\hat{C}}{T} \int_M (m_0 - T_j(m_0)). \end{aligned}$$

Using (5.17) with  $\varphi = T_j(m_0)$  we obtain, letting  $k \rightarrow \infty$ ,

$$\limsup_n \int_M m_{0n} u_n(0) \leq \int_M u(0) T_j(m_0) + \frac{\hat{C}}{T} \int_M (m_0 - T_j(m_0)).$$

Letting finally  $j \rightarrow \infty$ , we get

$$\limsup_n \int_M m_{0n} u_n(0) \leq \int_M u(0) m_0.$$

Similarly we argue for  $t = T$ , using now  $u_{n,k} = \min(u_n, k)$ .

We are left to prove (5.14). To this purpose, for any  $f \in L^1(Q_T)$ , we denote by  $f_h, f_{-h}$  the time-average functions  $f_h := \frac{1}{h} \int_t^{t+h} f(x, s) ds$  and  $f_{-h} := \frac{1}{h} \int_{t-h}^t f(x, s) ds$ . Integrating (5.15) in  $(t, t + h)$  and dividing by  $h$ , we obtain, by means of Jensen inequality,

$$-\partial_t(u_{n,k})_h + \frac{1}{2} |\nabla(u_{n,k})_h|^2 \leq \frac{1}{h} \int_t^{t+h} \{ \mathbb{1}_{\{u_n \geq -k\}} \varepsilon_n (\log m_n + V) \} ds.$$

Let  $(\mu, v)$  be any solution of the continuity equation such that  $\mu(t) \in L^1(M)$  for every  $t$ , and  $\mu \log(\mu) \in L^1(Q_T)$ . By a density argument, any Lipschitz function  $\varphi$  can be used as test function in the continuity equation; hence, multiplying by  $(u_{n,k})_h$ , we get (for any  $0 \leq s < r < T - h$ )

$$\begin{aligned} & \int_M (u_{n,k})_h(s) \mu(s) - \int_M (u_{n,k})_h(r) \mu(r) \\ & \leq \int_s^r \int_M \mu v \cdot_{\mathbb{g}} \nabla(u_{n,k})_h - \frac{1}{2} |\nabla(u_{n,k})_h|^2 \mu \, d\tau \\ & \quad + \varepsilon_n \int_s^r \int_M \frac{1}{h} \int_{\tau}^{\tau+h} \{ \mathbb{1}_{\{u_n \geq -k\}} (\log m_n + V) \} \mu(\tau) \, ds \, d\tau. \end{aligned} \tag{5.18}$$

Now we let  $n \rightarrow \infty$ . Recall that  $u_n$  is bounded in  $L^2((a, b); H^1(M))$ , and, using (5.6), we have in fact  $\partial_t u_n$  bounded in  $L^1((a, b); L^1(M))$ ; using classical compactness results (see [40]), we deduce that  $u_n$  is compact in  $L^2((a, b); L^2(M))$ . Moreover,  $u_n$  is locally uniformly bounded (and bounded above up to  $t = 0$ ); since  $\mu(t) \in L^1(M)$ , we deduce that

$$\int_M (u_{n,k})_h(t) \mu(t) \xrightarrow{n \rightarrow \infty} \int_M (u_k)_h(t) \mu(t)$$

for any  $t \in [0, T)$ , where  $u_k = \max(u, -k)$ . Moreover, by weak convergence of  $u_n$  in  $L^2((a, b); H^1(M))$ , for any small  $\eta > 0$  we have

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \int_s^r \int_M \mu v \cdot_{\mathbb{g}} \nabla(u_{n,k})_h - \frac{1}{2} |\nabla(u_{n,k})_h|^2 \mu \, d\tau \\ & \leq \limsup_{n \rightarrow \infty} \int_{s+\eta}^r \int_M \mu v \cdot_{\mathbb{g}} \nabla(u_{n,k})_h - \frac{1}{2} |\nabla(u_{n,k})_h|^2 \mu \, d\tau + \int_s^{s+\eta} \int_M \frac{1}{2} \mu |v|^2 \, d\tau \\ & \leq \int_{s+\eta}^r \int_M \mu v \cdot_{\mathbb{g}} \nabla(u_k)_h - \frac{1}{2} |\nabla(u_k)_h|^2 \mu \, d\tau + \int_s^{s+\eta} \int_M \frac{1}{2} \mu |v|^2 \, d\tau, \end{aligned}$$

and letting  $\eta \rightarrow 0$  Fatou's lemma yields

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \int_s^r \int_M \mu v \cdot_{\mathbb{g}} \nabla(u_{n,k})_h - \frac{1}{2} |\nabla(u_{n,k})_h|^2 \mu \, d\tau \\ & \leq \int_s^r \int_M \mu v \cdot_{\mathbb{g}} \nabla(u_k)_h - \frac{1}{2} |\nabla(u_k)_h|^2 \mu \, d\tau. \end{aligned}$$

We are left with the last integral in (5.18). Of course, if  $\varepsilon_n \rightarrow 0$ , using  $\mu \log m_n \leq \mu \log \mu + m_n$ , we estimate

$$\begin{aligned} \varepsilon_n \int_s^r \int_M \frac{1}{h} \int_\tau^{\tau+h} \{ \mathbb{1}_{\{u_n \geq -k\}} (\log m_n + V) \} \mu(\tau) ds d\tau \\ \leq \varepsilon_n \int_s^r \int_M \frac{1}{h} \int_\tau^{\tau+h} (m_n + \mu \log \mu + V \mu) ds d\tau \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

So we suppose that  $\varepsilon_n \rightarrow \varepsilon > 0$ . In this case, from Lemma 5.5 we get that  $\sqrt{m_n}$  is bounded in  $L^2((a, b); H^1(M))$ , for any  $0 < a < b < T$ . Since  $m_n |\nabla u_n|^2$  is bounded in  $L^1(Q_T)$  and

$$(\sqrt{m_n})_t = \frac{1}{2} \operatorname{div}_g(\sqrt{m_n} \nabla u_n) + \frac{1}{2} \nabla u_n \cdot_g \nabla \sqrt{m_n}$$

we also deduce that  $(\sqrt{m_n})_t$  is bounded in  $L^2((a, b); (H^1(M))^*) + L^1((a, b) \times M)$ . By classical compactness results (see [40]), we infer that  $\sqrt{m_n}$  is strongly compact in  $L^2((a, b) \times M)$ , which means that  $m_n$  is relatively compact in the strong  $L^1$ -topology, locally in time, and converges almost everywhere, up to subsequences. In particular, still using that  $\mu \log m_n \leq \mu \log \mu + m_n$ , we can apply Fatou’s lemma in the last integral in (5.18) (notice that  $\mathbb{1}_{\{u_n \geq -k\}}$  a.e. converges to  $\mathbb{1}_{\{u \geq -k\}}$  for almost every  $k$ , which we can suppose to be the case). Finally, by (5.18) we obtain, letting  $n \rightarrow \infty$ :

$$\begin{aligned} \int_M (u_k)_h(s) \mu(s) - \int_M (u_k)_h(r) \mu(r) \\ \leq \int_s^r \int_M \mu v \cdot_g \nabla (u_k)_h - \frac{1}{2} |\nabla (u_k)_h|^2 \mu d\tau \\ + \varepsilon \int_s^r \int_M \frac{1}{h} \int_\tau^{\tau+h} \mathbb{1}_{\{u \geq -k\}} (\log m + V) \mu(\tau) ds d\tau. \end{aligned}$$

Now we let  $h \rightarrow 0$ . Once more, the last term can be handled through Fatou’s lemma; indeed, since  $\log(m) \in L^1_{loc}((0, T); L^1(M))$ , we have that

$$\frac{1}{h} \int_\tau^{\tau+h} \mathbb{1}_{\{u \geq -k\}} (\log m + V) \xrightarrow{h \rightarrow 0} \mathbb{1}_{\{u \geq -k\}} (\log m + V) \quad \text{for a.e. } \tau \in (0, T), x \in M$$

and

$$\mu(\tau) \frac{1}{h} \int_\tau^{\tau+h} \mathbb{1}_{\{u \geq -k\}} (\log m + V) \leq \mu \log \mu + \mu V + m_h$$

where last sequence is strongly convergent in  $L^1(Q_T)$ . Hence Fatou’s lemma can be applied and yields

$$\begin{aligned} \limsup_{h \rightarrow 0} \int_s^r \int_M \frac{1}{h} \int_\tau^{\tau+h} \mathbb{1}_{\{u \geq -k\}} (\log m + V) \mu(\tau) ds d\tau \\ \leq \int_s^r \int_M \mathbb{1}_{\{u \geq -k\}} (\log m + V) \mu d\tau. \end{aligned}$$

Similarly we argue for the term involving  $\nabla u_k$ , where we also use Fatou’s lemma, because  $\nabla u_k \in L^2_{loc}((0, T) \times M)$  and  $\nabla (u_k)_h \rightarrow \nabla u_k$  almost everywhere, up to extracting a suitable subsequence. Finally, using that  $u_k$  is uniformly bounded in  $(0, r + h)$  and  $u$  admits one-sided traces (in the sense of monotone limits of measurable functions, as recalled above), we have that  $(u_k)_h(t) \rightarrow u_k(t)$  (we can use here the precise representative for  $u$  at any  $t$ , otherwise we

should limit ourselves to a.e.  $t$ ). Notice that the convergence  $(u_k)_h(t) \rightarrow u_k(t)$  is pointwise but also weak- $*$   $L^\infty$ , and holds for all  $t \geq 0$ . Therefore, once  $h \rightarrow 0$  we get

$$\int_M u_k(s)\mu(s) - \int_M u_k(r)\mu(r) \leq \int_s^r \int_M [\mu v \cdot_g \nabla u_k - \frac{1}{2}|\nabla u_k|^2 \mu] d\tau + \varepsilon \int_s^r \int_M \mathbb{1}_{\{u \geq -k\}} (\log m + V) \mu d\tau.$$

Letting now  $k \rightarrow \infty$ , using Fatou’s lemma (and the monotone convergence theorem if  $s = 0$ ), we obtain

$$\int_M u(s)\mu(s) - \int_M u(r)\mu(r) \leq \int_s^r \int_M [\mu v \cdot_g \nabla u - \frac{1}{2}|\nabla u|^2 \mu] d\tau + \varepsilon \int_s^r \int_M (\log m + V) \mu d\tau.$$

With a symmetric argument, using the left time-averages  $u_{-h}$ , we also obtain the inequality

$$\int_M u(r)\mu(r) - \int_M u(t)\mu(t) \leq \int_r^t \int_M [\mu v \cdot_g \nabla u - \frac{1}{2}|\nabla u|^2 \mu] d\tau + \varepsilon \int_r^t \int_M (\log m + V) \mu d\tau$$

for every  $0 < r < t \leq T$ . Adding the last two inequalities we obtain (5.14). □

**Remark 5.8** The inequality (5.14) includes the case that  $s = 0$  or  $t = T$ . In particular, it is a byproduct of the previous proof that  $u(0) \in L^1(dm_0)$  and  $u(T) \in L^1(dm_1)$ , which would not be guaranteed a priori. This is indeed a consequence of item 2 of the statement, which implies that  $\int_M u(0)m_0$  is bounded below and  $\int_M u(T)m_1$  is bounded above. Since the other bounds are obvious from (5.4), this yields  $u(0) \in L^1(dm_0)$ ,  $u(T) \in L^1(dm_1)$ .

**Remark 5.9** We point out that inequality (5.14) remains true when  $\varepsilon = 0$  even without requiring that  $\mu \log(\mu) \in L^1(Q_T)$ . It is enough that  $\mu \in L^1(Q_T)$  in order that (5.14) holds for all  $s, t \in (0, T)$  (such that  $\mu(s), \mu(t) \in L^1(M)$ ), and even for  $s = 0, t = T$  assuming for instance that  $\mu(\cdot)$  is continuous in  $[0, T]$  in the weak  $L^1$ -topology.

In fact, we know from Proposition 4.4 that  $\varepsilon_n(\log(m_n))_+$  is locally uniformly bounded; in addition, if  $\varepsilon_n \rightarrow 0$ , using estimate (3.9) we have

$$\begin{aligned} \|\varepsilon_n(\log(m_n))_+\|_{L^1((0,T) \times M)} &\leq \varepsilon_n \int_0^T \int_M m_n(\log(m_n))_+ \\ &\leq C \varepsilon_n (1 + |\log(\varepsilon_n)|) \rightarrow 0. \end{aligned}$$

Hence  $\varepsilon_n(\log(m_n))_+$  converges to zero in  $L^1$  and weakly- $*$  in  $L^\infty((a, b) \times M)$ . This implies that the last term in (5.18) vanishes as  $n \rightarrow \infty$ , only using that  $\mu$  is in  $L^1(Q_T)$ . Thus we obtain again the inequality

$$\int_M u_k(s)\mu(s) - \int_M u_k(r)\mu(r) \leq \int_s^r \int_M [\mu v \cdot_g \nabla u_k - \frac{1}{2}|\nabla u_k|^2 \mu] d\tau$$

for any  $0 < s < r < T$ . Since  $u$  is locally bounded, here one can readily get rid of the truncation  $k$  for  $s, r \in (0, T)$ . To get the inequality up to  $t = 0$ , one can first let  $s \rightarrow 0^+$  using that  $\mu(\cdot)$  is continuous in the weak  $L^1$ -topology and  $u_k$  is uniformly bounded. This leads the first integral towards  $\int_M u_k(0)m_0$ , which gives the desired term by letting  $k \rightarrow \infty$

and using the monotone convergence theorem. Symmetrically one argue up to  $t = T$  to get (5.14) in the whole interval  $(0, T)$ .  $\square$

Now we have all the ingredients to prove our main result on the existence and characterization of minima for nonnegative marginals  $m_0, m_1$ , which are only assumed to be  $L^1(M)$ .

**Theorem 5.10** *Let  $V \in W^{2,\infty}(M)$ ,  $m_0, m_1 \in \mathcal{P}(M) \cap L^1(M)$ , and  $\varepsilon > 0$ .*

*Then there is a unique  $m$  and a unique  $u$  such that  $(u, m)$  is a weak solution of problem (5.2) with  $\int_M u(T)m_1 = 0$ . Moreover we have that  $u, m \in L^\infty_{loc}(Q_T)$ , and  $(m, \nabla u)$  is the unique minimum of the functional  $\mathcal{F}_\varepsilon$  in (1.1).*

**Proof** We first approximate  $m_0, m_1$  with positive smooth marginals. To this goal, we consider the solutions of the heat equation

$$\frac{\partial}{\partial t} \tilde{m}_i = \Delta_g \tilde{m}_i \quad \text{with } \tilde{m}_i(0, \cdot) = m_i \text{ for } i = 0, 1.$$

Thanks to the compactness of  $M$ , it is well known that such solutions are smooth on  $(0, +\infty) \times M$  and that they are curves of probability measures (cfr. [21, Chapter 7]). Furthermore we have  $\tilde{m}_i(t, \cdot) > 0$  for every  $t > 0$  by the strong parabolic maximum principle (cfr. [21, Theorem 8.11]) and  $\tilde{m}_i(t, \cdot) \rightarrow m_i$  in  $L^1(M)$  for  $t \rightarrow 0$  [21, Theorem 7.19].

For every  $n \in \mathbb{N}$  let  $m_{0n} = \tilde{m}_0(\frac{1}{n}, \cdot)$  and  $m_{1n} = \tilde{m}_1(\frac{1}{n}, \cdot)$ . By Theorem 5.4, we obtain the existence of a couple  $(u_n, m_n)$  which is a classical solution of

$$\begin{cases} -\partial_t u + \frac{1}{2} |\nabla u|^2 = \varepsilon(\log(m) + V(x)) & \text{in } Q_T \\ \partial_t m - \operatorname{div}_g(m \nabla u) = 0 & \text{in } Q_T \\ m(0, \cdot) = m_{0n}, \quad m(T, \cdot) = m_{1n} & \text{in } M \end{cases} \quad (5.19)$$

with  $\int_M u_n(T)m_{1n} = 0 \forall n \in \mathbb{N}$ .

Now we use Lemmas 5.2 and 5.5 to get estimates for  $u_n, m_n$ . In particular,  $u_n$  satisfies (5.4), hence it is locally uniformly bounded, and the same holds for  $m_n$  due to Corollary 5.3. In addition  $u_n$  is bounded in  $L^2((a, b); H^1(M))$  and  $m_n |\nabla u_n|^2$  is bounded in  $L^1(Q_T)$ . With the same compactness arguments used in Lemma 5.6, we deduce that both  $u_n$  and  $m_n$  are relatively compact in the strong  $L^1(M)$ -topology, locally in time. Therefore, we deduce that there exist functions  $u, m$  such that for a subsequence, not relabeled,

$$\begin{aligned} u_n &\rightarrow u \quad \text{weakly in } L^2((a, b); H^1(M)) \text{ and strongly in } L^p((a, b) \times (M)) \forall p > 1, \\ m_n &\rightarrow m \quad \text{strongly in } L^p((a, b) \times (M)) \forall p > 1, \text{ and a.e. in } Q_T. \end{aligned}$$

In addition, we know that  $m_n(t)$  has bounded entropy (at any time  $t$ ), so it is weakly compact in  $L^1(M)$ ; and due to the bound of  $m_n |\nabla u_n|^2$  (see (5.5)), we get that  $m_n$  is equi-continuous from  $[0, T]$  into the space of measures. By Ascoli–Arzela’s theorem,  $m_n(t) \rightarrow m(t)$  in the Wasserstein topology (uniformly in  $[0, T]$ ), and actually in  $L^1$ -weak for all  $t \in (0, T)$ , due to the bound of the entropy. By continuity, we conclude that  $m(0) = m_0$  and  $m(T) = m_1$ . Notice that, by the local strong convergence of  $m_n$ , and due to the bound (5.5), we also deduce that

$$\sqrt{m_n} \nabla u_n \rightarrow \sqrt{m} \nabla u \quad \text{weakly in } L^2((0, T) \times M), \quad (5.20)$$

and, in particular, we have that  $m_n \nabla u_n \rightarrow m \nabla u$  in the sense of distributions, and actually weakly in  $L^1((0, T) \times M)$ . Thus, we proved so far that  $m \in C^0([0, T]; \mathcal{P}(M))$  and is a weak

solution of

$$\begin{cases} \partial_t m - \operatorname{div}_g(m \nabla u) = 0 & \text{in } Q_T, \\ m(0) = m_0, m(T) = m_1. \end{cases} \tag{5.21}$$

In addition, by (5.6) and Fatou’s lemma, we have that  $\log(m) \in L^1((a, b) \times (M))$ , for any  $0 < a < b < T$ , and in particular  $m > 0$  a.e. in  $Q$ .

As for the Hamilton–Jacobi equation, we use Lemma 5.6 to deduce that

$$-\partial_t u + \frac{1}{2} |\nabla u|^2 \leq \varepsilon (\log(m) + V) \tag{5.22}$$

and we also have

$$\limsup_{n \rightarrow \infty} \int_M m_{0n} u_n(0) \leq \int u(0) dm_0. \tag{5.23}$$

In particular (since the upper bound follows from (5.4)), we deduce that  $u(0) \in L^1(dm_0)$ . Similarly we reason for  $t = T$ , obtaining

$$\int_M m_1 u(T) \leq \liminf_{n \rightarrow \infty} \int_M m_{1n} u_n(T) = 0. \tag{5.24}$$

As before, this implies, in particular, that  $u(T) \in L^1(dm_1)$ . Now we only need to conclude that  $(u, m)$  satisfies condition (iii) of Definition 5.1. To this purpose, we follow the steps of [12, 37], on account of Lemma 5.6. First we go back to (5.19), which implies, using  $\int_M u_n(T) m_{1n} = 0$ ,

$$\int_M m_{0n} \hat{u}_n(0) = \int_0^T \int_M \frac{1}{2} m_n |\nabla u_n|^2 + \varepsilon m_n (\log m_n + V) dt.$$

Using (5.20) and weak lower semicontinuity, as well as (5.23), we deduce, as  $n \rightarrow \infty$ :

$$\int_0^T \int_M \frac{1}{2} m |\nabla u|^2 + \varepsilon m (\log m + V) dt \leq \int u(0) dm_0. \tag{5.25}$$

However, applying (5.14) with  $(\mu, v) = (m, \nabla u)$ ,  $s = 0$  and  $t = T$ , we get

$$\begin{aligned} \int_M m_0 u(0) &\leq \int_0^T \int_M \frac{1}{2} |\nabla u|^2 m + \varepsilon m (\log m + V) dt + \int_M u(T) dm_1 \\ &\leq \int_0^T \int_M \frac{1}{2} |\nabla u|^2 m + \varepsilon m (\log m + V) dt \end{aligned}$$

where we used (5.24). Putting together the above information with (5.25) we obtain

$$\int_0^T \int_M \frac{1}{2} m |\nabla u|^2 + \varepsilon m (\log m + V) dt = \int_M m_0 u(0),$$

and, in between, we also get that  $\int_M m_1 u(T) = 0$ . This means that  $(u, m)$  satisfy Definition 5.1. In addition, from the bounds derived before, we have  $u, m \in L^\infty_{loc}(Q_T)$ .

We are left to prove that  $(m, \nabla u)$  is the unique minimum of  $\mathcal{F}_\varepsilon$ . To show that, let  $(\mu, v)$  be an admissible couple for the functional  $\mathcal{F}_\varepsilon$ . Without loss of generality, we can assume

that  $\mathcal{F}_\varepsilon(\mu, v) < \infty$ , in particular  $\mu \log \mu \in L^1(Q_T)$ . We use once more (5.25) together with (5.14) and we get

$$\begin{aligned} & \int_0^T \int_M \left( \frac{1}{2} m |\nabla u|^2 + \varepsilon m (\log m + V) \right) dt \\ & \leq \int_0^T \int_M [\mu v \cdot_g \nabla u - \frac{1}{2} |\nabla u|^2 \mu] dt + \varepsilon \int_0^T \int_M (\log m + V) \mu dt \quad (5.26) \\ & \leq \int_0^T \int_M \frac{1}{2} |v|^2 \mu dt + \varepsilon \int_0^T \int_M (\log m + V) \mu dt. \end{aligned}$$

By the strict convexity of  $r \mapsto r \log r - r$ , for every  $a \geq 0$  and  $b > 0$  we have

$$a \log(a) - a \geq b \log(b) - b + \log(b)(a - b)$$

where  $a \log a$  is extended as 0 for  $a = 0$ . Furthermore the equality holds if and only if  $a = b$ . This is equivalent to

$$(\log(b) - \log(a))a \leq b - a \quad (5.27)$$

with equality if and only if  $a = b$ . Applying this inequality with  $a = \mu$  and  $b = m$  (which is positive a.e.), we obtain from (5.26)

$$\begin{aligned} \int_0^T \int_M \left( \frac{1}{2} m |\nabla u|^2 + \varepsilon m (\log m + V) \right) dt & \leq \mathcal{F}_\varepsilon(\mu, v) + \varepsilon \int_0^T \int_M (m - \mu) dt \\ & = \mathcal{F}_\varepsilon(\mu, v) \end{aligned}$$

and the equality holds if and only if  $\mu = m$ . This concludes the proof that  $(m, \nabla u)$  is the unique minimum of  $\mathcal{F}_\varepsilon$ .

In fact, the solution we found is also the unique weak solution of the system (5.2) (up to addition of a constant to  $u$ ). Indeed, the uniqueness of  $(m, u)$  can be proved similarly as in [12, Thm 1.16]. Compared to this latter result, we observe that, being  $m > 0$  almost everywhere, there is no loss of information here due to the set where  $m$  vanishes.  $\square$

### 6 Convergence to Wasserstein geodesic

We now briefly analyze the limit  $\varepsilon \rightarrow 0$  to show that the minimal curves of  $\mathcal{F}_\varepsilon$  converge to the geodesics of the classical mass transport problem

$$\min \mathcal{F}_0(m, v) := \int_0^T \int_M \frac{1}{2} |v|^2 dm, \quad (m, v) : \begin{cases} \partial_t m - \operatorname{div}_g(vm) = 0 \\ m(0) = m_0 \\ m(T) = m_1. \end{cases} \quad (6.1)$$

We first consider the easier case where the marginals  $m_0, m_1$  have finite entropy. This assumption implies that  $\mathcal{F}_\varepsilon$  converges to  $\mathcal{F}_0$  with a rate of order  $\varepsilon$ .

**Theorem 6.1** *Let  $V \in W^{2,\infty}(M)$ ,  $m_0, m_1 \in \mathcal{P}(M) \cap L^1(M)$  and such that  $\mathcal{H}(m_0; v), \mathcal{H}(m_1; v) < \infty$ . For  $\varepsilon \in (0, 1)$ , let  $(m_\varepsilon, u_\varepsilon)$  be the unique solution of (5.2) given by Theorem 5.10, and  $(m_\varepsilon, \nabla u_\varepsilon)$  the unique minima of  $\mathcal{F}_\varepsilon$ .*

*As  $\varepsilon \rightarrow 0$ , we have:*

$$\begin{aligned} m_\varepsilon & \rightarrow m \text{ in } C^0([0, T], \mathcal{P}(M)) \text{ and weakly in } L^1(Q_T), \\ m_\varepsilon \nabla u_\varepsilon & \rightarrow m \nabla u \text{ weakly in } L^1(Q_T), \end{aligned} \quad (6.2)$$

where  $m$  is the Wasserstein geodesic between  $m_0, m_1$ , and  $(m, \nabla u)$  is a minimum of  $\mathcal{F}_0$ .

Moreover, we have  $(\min \mathcal{F}_0) = \lim_{\varepsilon \rightarrow 0} (\min \mathcal{F}_\varepsilon)$ , and in particular, for some  $K > 0$ ,

$$|\min \mathcal{F}_\varepsilon - \min \mathcal{F}_0| \leq K \varepsilon. \tag{6.3}$$

**Proof** For every  $\varepsilon > 0$  we can apply Theorem 5.10 and define the couple  $(u_\varepsilon, m_\varepsilon)$  which is the unique weak solution to the problem

$$\begin{cases} -\partial_t u + \frac{1}{2} |\nabla u|^2 = \varepsilon (\log(m) + V(x)) & \text{in } Q_T \\ \partial_t m - \operatorname{div}_g(m \nabla u) = 0 & \text{in } Q_T \\ m(0) = m_0, \quad m(T) = m_1 & \text{in } M \end{cases}$$

with  $\int_M u_\varepsilon(T) m_1 = 0$ .

By Lemma 5.2, we have that  $u_\varepsilon$  is bounded in  $L^2((a, b), H^1(M))$  and in  $L^\infty((a, b) \times M)$  for every  $0 < a < b < T$ , so it is weakly relatively compact in  $L^2((a, b); H^1(M))$ .

For any sequence extracted out of  $u_\varepsilon$ , by a diagonal process we can select (and fix) a function  $u \in L^2_{loc}((0, T); H^1(M)) \cap L^\infty_{loc}((0, T) \times M)$  and a subsequence (that we will not rename) such that  $u_\varepsilon$  converges weakly to  $u$  in  $L^2((a, b); H^1(M))$  for every  $0 < a < b < T$ .

Furthermore, as a consequence of estimate (5.5), we have  $d_W(m_\varepsilon(t), m_\varepsilon(s)) \leq C\sqrt{t-s}$  for any  $t > s$  and some  $C > 0$ , where  $d_W$  is the Wasserstein distance. Hence, by Ascoli–Arzela’s theorem, there exists  $m \in C([0, T]; \mathcal{P}(M))$  such that, up to a subsequence,  $m_\varepsilon(t) \rightarrow m(t)$  in the Wasserstein topology, uniformly in time. Since  $m_0, m_1$  have finite entropy, by estimate (3.8), we have that

$$\int_M m_\varepsilon(t) \log(m_\varepsilon(t)) \text{ is uniformly bounded in } (0, T). \tag{6.4}$$

We deduce that  $m_\varepsilon$  is weakly relatively compact in  $L^1(Q_T)$  and then, by Cauchy–Schwarz inequality and (5.5),  $m_\varepsilon \nabla u_\varepsilon$  is also weakly relatively compact in  $L^1(Q_T)$ . In particular, there exists  $m \in C([0, T]; \mathcal{P}(M))$ , and  $w \in L^1(Q_T)$ , such that

$$\begin{aligned} m_\varepsilon &\rightarrow m \text{ in } C([0, T]; \mathcal{P}(M)) \text{ and weakly in } L^1(Q_T), \\ m_\varepsilon \nabla u_\varepsilon &\rightarrow w \text{ weakly in } L^1(Q_T). \end{aligned}$$

We now identify  $w$  as  $m \nabla u$ . To this goal, we first use the semi-continuity for the function  $\Psi$  defined in (3.2), and we get

$$\int_0^T \int_M \frac{|w|^2}{2m} dt \leq \liminf_{\varepsilon \rightarrow 0} \int_0^T \int_M \frac{1}{2} m_\varepsilon |\nabla u_\varepsilon|^2 dt = \int_M u_\varepsilon(0) m_0 + O(\varepsilon) \tag{6.5}$$

where we used (6.4). By Lemma 5.6 we deduce

$$\begin{aligned} \int_0^T \int_M \frac{|w|^2}{2m} dt &\leq \liminf_{\varepsilon \rightarrow 0} \int_0^T \int_M \frac{1}{2} m_\varepsilon |\nabla u_\varepsilon|^2 dt \\ &\leq \limsup_{\varepsilon \rightarrow 0} \int_M u_\varepsilon(0) m_0 \leq \int_M u(0) m_0. \end{aligned} \tag{6.6}$$

Notice that this inequality also implies that  $w = 0$  a.e. in the set  $\{m = 0\}$ . Setting  $v := \frac{w}{m} \mathbb{1}_{\{m>0\}}$  we deduce that  $(m, v)$  is a solution to (2.3). We also get from Lemma 5.6 a similar inequality at  $t = T$ , namely

$$\int_M u(T) m_1 \leq \liminf_{\varepsilon \rightarrow 0} \int_M u_\varepsilon(T) m_1 = 0.$$

We insert this information into (5.14) (where  $\varepsilon = 0, s = 0, t = T$ ) and we get

$$\int_M u(0)m_0 \leq \int_0^T \int_M \left[ \frac{w}{m} \cdot_{\mathbb{g}} \nabla u - \frac{|\nabla u|^2}{2} \right] m \leq \int_0^T \int_M \frac{|w|^2}{2m} dt.$$

Combining this with (6.6) we conclude that  $w = m \nabla u$  and that

$$\int_M u(0)m_0 = \int_0^T \int_M \frac{1}{2} m |\nabla u|^2 dt,$$

and then

$$\int_M u_\varepsilon(0) m_0 \rightarrow \int_M u(0)m_0,$$

and

$$\int_0^T \int_M \frac{1}{2} m_\varepsilon |\nabla u_\varepsilon|^2 dt \rightarrow \int_0^T \int_M \frac{1}{2} m |\nabla u|^2 dt.$$

We now show that  $(m, \nabla u)$  is a minimum of  $\mathcal{F}_0$ . To this goal, we recall (see [14, 34]) that the minimum of  $\mathcal{F}_0$  is attained at a unique geodesic  $\mu^*$  such that  $\mu^*(t) \in L^1(M)$  for every  $t$  and  $\mu^*(\cdot)$  is continuous in the weak- $L^1$  topology. On account of Remark 5.9, we can use inequality (5.14) with  $\varepsilon = 0$  and  $\mu = \mu^*$ , which yields:

$$\begin{aligned} \int_0^T \int_M \frac{1}{2} m |\nabla u|^2 dt &= \int_M u(0)m_0 \leq \int_0^T \int_M [\mu^* v \cdot_{\mathbb{g}} \nabla u - \frac{1}{2} |\nabla u|^2 \mu^*] dt \\ &\leq \int_0^T \int_M \frac{1}{2} |v|^2 \mu^* dt = \min \mathcal{F}_0 \end{aligned}$$

Hence  $(m, \nabla u)$  is the minimum point of  $\mathcal{F}_0$  and coincides with the unique geodesic between  $m_0, m_1$  (notice that the previous inequality implies  $v = \nabla u$ ). Finally, we observe that, still using (5.14) for  $(m_\varepsilon, u_\varepsilon)$  and  $(m, u)$ , we have

$$\begin{aligned} \min \mathcal{F}_\varepsilon &= \int_M u_\varepsilon(0) m_0 \\ &\leq \int_0^T \int_M \frac{1}{2} m |\nabla u|^2 dt + \varepsilon \int_0^T \int_M m (\log(m_\varepsilon) + V) dt \\ &\leq \min \mathcal{F}_0 + \varepsilon \int_0^T \int_M m \log(m) + C \varepsilon = \min \mathcal{F}_0 + O(\varepsilon) \end{aligned}$$

due to the fact that  $m$  has finite entropy. Since the opposite inequality is obviously true, we conclude with (6.3). □

As a byproduct of the previous result, we have proved that whenever  $m_0, m_1$  have finite entropy, then the Wasserstein geodesic connecting  $m_0, m_1$  has finite entropy for all times. In fact, using Corollary 3.4 we also have a quantitative estimate of the convexity of the log-entropy functional along the geodesic. In this way, we recover a result proved by Daneri and Savaré [15] with a different and independent approach. We point out that we can avoid the use of this convexity property of the geodesic in all our estimates and stability results (see also Remark 2.4), so that this is just deduced from the limit  $\varepsilon \rightarrow 0$  of the convexity of the entropy along the optimal curves of  $\mathcal{F}_\varepsilon$ .

**Corollary 6.2** *In the assumptions of Theorem 6.1, let  $\Lambda \in \mathbb{R}$  be such that*

$$\text{Ric}_g + D^2V \geq \Lambda I_g$$

*in the sense that  $(\text{Ric}_g + D^2V)(X, X) \geq \Lambda |X|^2$  for every vector field  $X$  on  $M$ .*

*Then the relative entropy functional  $\mathcal{H}(m; \nu)$  is  $\Lambda$ -convex along the 2-Wasserstein geodesics. In other words, if  $m : [0, 1] \rightarrow \mathcal{P}(M)$  is the geodesic between  $m_0$  and  $m_1$ , it holds*

$$\mathcal{H}(m(t); \nu) \leq t\mathcal{H}(m_1; \nu) + (1 - t)\mathcal{H}(m_0; \nu) - \frac{\Lambda}{2}t(1 - t)W_2^2(m_0, m_1) \tag{6.7}$$

*for every  $t \in [0, 1]$ .*

**Proof** As in Theorem 6.1, let  $(u_\varepsilon, m_\varepsilon)$  be the sequence of solutions of (1.2), with  $\int_M u_\varepsilon(T)m_1 = 0$ . At first, let us suppose that  $m_0, m_1$  are smooth and positive, so that  $u_\varepsilon, m_\varepsilon$  are smooth solutions (Theorem 5.4). By (3.7) in Corollary 3.4, we have

$$\begin{aligned} \frac{d^2}{dt^2}\mathcal{H}(m_\varepsilon(t); \nu) &\geq \Lambda \int_M m_\varepsilon |\nabla u_\varepsilon|^2 dx \\ &= 2\Lambda\mathcal{B}_\varepsilon(m_0, m_1) - 2\Lambda\varepsilon\mathcal{H}(m_\varepsilon(t); \nu) \end{aligned}$$

where we used (3.4) (with  $T = 1$ ) and, we recall,  $\mathcal{B}_\varepsilon(m_0, m_1) = \min(\mathcal{F}_\varepsilon)$ . Since  $m_0, m_1$  have finite entropy, we already know from (3.8) that  $\mathcal{H}(m_\varepsilon(t); \nu)$  is bounded independently of  $\varepsilon$ , for every  $t \in (0, T)$ . Hence we get, for some  $C > 0$ :

$$\frac{d^2}{dt^2}\mathcal{H}(m_\varepsilon(t); \nu) \geq \Lambda_\varepsilon := 2\Lambda\mathcal{B}_\varepsilon(m_0, m_1) - C\varepsilon$$

which implies

$$\mathcal{H}(m_\varepsilon(t); \nu) \leq t\mathcal{H}(m_1; \nu) + (1 - t)\mathcal{H}(m_0; \nu) - \frac{\Lambda_\varepsilon}{2}t(1 - t) \quad \forall t \in ([0, 1]). \tag{6.8}$$

Note that  $\Lambda_\varepsilon$  is stable by approximation of  $m_0, m_1$  with smooth positive densities, due to Theorem 5.10, so the above inequality holds for any  $m_0, m_1$  with finite entropy. Finally, we let  $\varepsilon \rightarrow 0$ ; by Theorem 6.1 we know that  $m_\varepsilon(t)$  weakly converges in  $L^1(M)$  towards  $m(t)$ , where  $m$  is the Wasserstein geodesic between  $m_0, m_1$ . Thus we can use the weak lower semicontinuity of the entropy for the left-hand side. We also know that  $\mathcal{B}_\varepsilon(m_0, m_1) = \min(\mathcal{F}_\varepsilon)$  converges towards  $\min(\mathcal{F}_0) = \frac{1}{2}W_2^2(m_0, m_1)$ . Hence  $\Lambda_\varepsilon \rightarrow \Lambda W_2^2(m_0, m_1)$  and from (6.8) we deduce (6.7).  $\square$

We conclude by extending the convergence result of Theorem 6.1 to marginals which only belong to  $L^1(M)$ , without having necessarily finite entropy. It is known that, if  $m_0, m_1 \in L^1(M)$ , then the Wasserstein geodesic belongs to  $L^1(M)$  for all times  $t$ , see e.g. [14, 34]. This is also a byproduct of our next result, since we will prove that (6.2) still holds for merely  $L^1$  marginals. To get the necessary equi-integrability for this goal, we will use an idea suggested to us by G. Savaré, based on displacement convexity and the following lemma essentially contained in [44].

**Lemma 6.3** *Let  $m_0 \in L^1(M)$ . Then there exists a function  $U : [0, \infty) \rightarrow \mathbb{R}^+$  such that:*

- (i)  $U \in C^2(0, \infty)$ , is convex and satisfies  $\frac{U(r)}{r} \xrightarrow{r \rightarrow \infty} +\infty$ .
- (ii)  $P'(r)r - (1 - \frac{1}{d})P(r) \geq 0$ , and  $0 \leq P(r) \leq Kr$  for every  $r > 0$ , where  $P(r) = U'(r)r - U(r)$ ,  $K > 0$ .

(iii)  $U(m_0) \in L^1(M)$ .

Even if the above Lemma mostly follows from [44, Proposition 17.7] combined with De la Vallée–Poussin lemma, we will give a proof in the “Appendix”, for the reader’s convenience. Standing on Lemma 6.3, we will show that  $\min \mathcal{F}_\varepsilon(m_0, m_1) \rightarrow \min \mathcal{F}_0(m_0, m_1)$ , although now the rate of convergence appears to be of order  $|\varepsilon \log \varepsilon|$ . We will also prove a further property here, namely that, up to approximating  $m_0, m_1$  with suitable smooth sequences  $m_{0\varepsilon}, m_{1\varepsilon}$ , we can build a minimizing curve of  $\mathcal{F}_\varepsilon$  which is a smooth approximation of the Wasserstein geodesic, with the adjoint states uniformly converging to the Kantorovich potentials.

**Theorem 6.4** *Let  $V \in W^{2,\infty}(M)$ , and  $m_0, m_1 \in \mathcal{P}(M) \cap L^1(M)$ . For  $\varepsilon \in (0, 1)$ , let  $(m_\varepsilon, u_\varepsilon)$  be the unique solution of (5.2) given by Theorem 5.10, and  $m$  be the Wasserstein geodesic between  $m_0, m_1$ , with  $(m, \nabla u)$  the minimum of  $\mathcal{F}_0$ . Then we have that (6.2) holds true and  $\min \mathcal{F}_0 = \lim_{\varepsilon \rightarrow 0} \min \mathcal{F}_\varepsilon$ .*

*In particular, we have*

$$\min \mathcal{F}_0 - c_0\varepsilon \leq \min \mathcal{F}_\varepsilon \leq \min \mathcal{F}_0 + c_1 \varepsilon |\log \varepsilon| \tag{6.9}$$

for some  $c_0, c_1 > 0$ .

*In addition, there exist sequences  $m_{0\varepsilon}, m_{1\varepsilon}$ , converging respectively to  $m_0, m_1$  in  $L^1(M)$ , such that, if  $(u_\varepsilon, m_\varepsilon)$  denotes the solution of (5.2) corresponding to marginals  $m_{0\varepsilon}, m_{1\varepsilon}$ , and with  $\int_M u_\varepsilon(T) dm_1 = 0$ , then we have:*

- (i)  $(m_\varepsilon, u_\varepsilon)$  is smooth in  $(0, T) \times M$ .
- (ii)  $u_\varepsilon$  is bounded in  $W^{1,\infty}(Q_T)$  and converges uniformly to a Lipschitz continuous solution  $u$  of the Hamilton–Jacobi equation  $\partial_t u = |\nabla u|^2/2$  in  $Q_T$ .
- (iii)  $m_\varepsilon \rightarrow m$  in  $C^0([0, T], \mathcal{P}(M))$  where  $m$  is the Wasserstein geodesic connecting  $m_0, m_1$ , with  $\nabla u$  being the metric velocity of the geodesic and  $u(0), u(T)$  the Kantorovich optimal potentials.

**Proof** Let  $U$  be the function given by Lemma 6.3 (built replacing  $m_0$  with  $\max(m_0, m_1)$ ). By the properties of  $U$ , we are allowed to use Proposition 3.3 and, in addition, the first term in the right side of (3.6) is nonnegative. Hence we have

$$\begin{aligned} \frac{d^2}{dt^2} \int_M U(m_\varepsilon) &\geq \int_M P(m_\varepsilon) Ricc_g(\nabla u_\varepsilon, \nabla u_\varepsilon) + \int_M \varepsilon \frac{P'(m_\varepsilon)}{m_\varepsilon} |\nabla m_\varepsilon|^2 \\ &\quad + \varepsilon \int_M P'(m_\varepsilon) \nabla m_\varepsilon \cdot_g \nabla V \\ &\geq \int_M P(m_\varepsilon) Ricc_g(\nabla u_\varepsilon, \nabla u_\varepsilon) - \varepsilon \int_M P(m_\varepsilon) \Delta_g V. \end{aligned}$$

Using (2.2) and that  $P(r) \leq Kr$  by Lemma 6.3, we obtain

$$\frac{d^2}{dt^2} \int_M U(m_\varepsilon) \geq -\lambda K \int_M m_\varepsilon |\nabla u_\varepsilon|^2 - \varepsilon K \|\Delta_g V\|_\infty.$$

Setting  $\varphi_\varepsilon(t) = \int_M U(m_\varepsilon)(t)$ , we deduce that  $\varphi_\varepsilon$  satisfies, for some constant  $c > 0$ ,

$$\begin{cases} -\varphi_\varepsilon''(t) \leq \lambda K f_\varepsilon(t) + c\varepsilon & t \in (0, T) \\ \varphi_\varepsilon(0) = \int_M U(m_0), \varphi_\varepsilon(T) = \int_M U(m_1) \end{cases} ,$$

where  $f_\varepsilon := \int_M m_\varepsilon |\nabla u_\varepsilon|^2$  is bounded in  $L^1(0, T)$  by Lemma 5.2. By the (compact) embedding of  $H^1(0, T)$  into  $C^0([0, T])$ , we deduce that  $\varphi_\varepsilon$  is uniformly bounded and actually it is relatively compact in the uniform topology. Since  $U$  is superlinear, this means that  $m_\varepsilon(t)$  is weakly compact in  $L^1(M)$  and weakly converges to  $m(t)$ , for every  $t \in [0, T]$ . In addition,  $t \mapsto m(t)$  is continuous in  $L^1(M)$  endowed with the weak topology. With this in hand, we also have that  $m_\varepsilon \nabla u_\varepsilon$  is weakly compact in  $L^1(Q_T)$ . Moreover, from Lemma 5.2, we know that  $u_\varepsilon$  weakly converges to some  $u \in L^2_{loc}((0, T); H^1(M)) \cap L^\infty_{loc}((0, T) \times M)$ , exactly as in Theorem 6.1. In order to identify the limit of  $m_\varepsilon \nabla u_\varepsilon$  as  $m \nabla u$ , we can proceed as before, using (5.14) on account of Remark 5.9. Thus we obtain the same conclusion (6.2) as before. However, the rate of convergence (6.3) does not follow in this case since we can only estimate

$$\int_0^T \int_M m_\varepsilon \log(m_\varepsilon) dt = O(\varepsilon |\log \varepsilon|)$$

from (3.9). This yields (6.9).

Finally, we observe that we can build a smooth approximation of the Wasserstein geodesic if we approximate  $m_0, m_1$  with the heat semigroup, namely  $m_{0\varepsilon} = S_\varepsilon(m_0), m_{1\varepsilon} = S_\varepsilon(m_1)$ , where  $S_t$  is the heat semigroup as in Proposition 2.2. By using Li-Yau estimates on the heat kernel, in the improved form given e.g. in [27, Thm 1.8] for Riemannian manifolds with Ricci curvature bounded below, we have that there exists a constant  $C$ , only depending on  $M, d$ , such that

$$\varepsilon \left( \frac{|\nabla S_\varepsilon(m_0)|}{S_\varepsilon(m_0)} + |\log(S_\varepsilon(m_0))| \right) \leq C.$$

This means that  $\varepsilon \log(m_{0\varepsilon})$  is bounded in  $W^{1,\infty}(M)$ , and so is for  $\varepsilon \log(m_{1\varepsilon})$ . From (5.7) in Lemma 5.2 we deduce that  $u_\varepsilon$  is uniformly bounded, and then from Theorem 4.3  $u_\varepsilon$  is bounded in Lipschitz norm. Moreover, at fixed  $\varepsilon$ ,  $(u_\varepsilon, m_\varepsilon)$  are smooth according to Theorem 5.4. Finally, by Ascoli–Arzelà’s theorem,  $u_\varepsilon$  is relatively compact in  $C^0(\overline{Q_T})$  and converges uniformly towards its limit  $u$ . It is a standard result that  $u$  is a viscosity solution of the Hamilton Jacobi equation  $\partial_t u = \frac{1}{2} |\nabla u|^2$ .  $\square$

### 7 Appendix A: Existence of smooth solutions

Here we show the existence of solutions to the differential system

$$\begin{cases} -tr \left( A(x, \nabla u) \circ \overline{\nabla}^2 u \right) + \rho u + \varepsilon \nabla u \cdot_g \nabla V(x) = 0 & \text{in } Q \\ -\partial_t u + \frac{1}{2} |\nabla u|^2 = \delta u + \varepsilon (\log(m_1) + V(x)) & \text{in } \Sigma_T \\ -\partial_t u + \frac{1}{2} |\nabla u|^2 + \delta u = \varepsilon (\log(m_0) + V(x)) & \text{in } \Sigma_0 \end{cases} \tag{7.1}$$

and we recall (see (3.16)) that the expanded form of the first equation is

$$-\partial_{tt} u + 2\nabla(\partial_t u) \cdot_g \nabla u - \varepsilon \Delta_g u - (\nabla^2 u)(\nabla u, \nabla u) + \rho u + \varepsilon \nabla u \cdot_g \nabla V = 0. \tag{7.2}$$

**Proposition 7.1** *Let  $\rho, \delta, \varepsilon > 0$ . Assume that  $V \in W^{2,\infty}(M)$  and  $m_0, m_1 \in \mathcal{P}(M) \cap C^{1,\alpha}(M)$  with  $m_0, m_1 > 0$  in  $M$ .*

*Then there exists a classical solution  $u \in C^{2,\alpha}(\overline{Q_T})$  of the quasilinear elliptic problem (7.1).*

We will prove such result by means of the method of continuity. For convenience of notation, we set  $\varepsilon = 1$  in what follows. We consider the differential operators  $F : C^{2,\alpha}(\overline{Q}_T) \rightarrow C^\alpha(\overline{Q}_T)$  and  $G : C^{2,\alpha}(\overline{Q}_T) \rightarrow C^{1,\alpha}(\Sigma_0 \cup \Sigma_T)$  defined by

$$\begin{aligned}
 F[u] &:= -tr \left( \mathcal{A}(x, \nabla u) \circ \overline{\nabla}^2 u \right) + \rho u + \nabla u \cdot_g \nabla V(x) \\
 G[u] &:= \begin{cases} -\partial_t u + \frac{1}{2} |\nabla u|^2 - \delta u - \log(m_1) - V(x) & \text{in } \Sigma_T \\ -\partial_t u + \frac{1}{2} |\nabla u|^2 + \delta u - \log(m_0) - V(x) & \text{in } \Sigma_0. \end{cases}
 \end{aligned}$$

Finally we define the operator

$$\begin{aligned}
 P : C^{2,\alpha}(\overline{Q}_T) &\longrightarrow C^\alpha(\overline{Q}_T) \times C^{1,\alpha}(\Sigma_0 \cup \Sigma_T) \\
 u &\longmapsto (F[u], G[u])
 \end{aligned}$$

and the set

$$E := \{u \in C^{2,\alpha}(\overline{Q}_T) \mid \exists \tau \in [0, 1] \text{ s.t. } P[u] = (1 - \tau)P[0]\}$$

We note that  $0 \in E$  and that if a function  $u \in C^{2,\alpha}(\overline{Q}_T)$  satisfies  $P[u] = 0$  then it is a solution for the elliptic problem (7.1).

**Lemma 7.2** *The set  $E$  is bounded in  $C^{1,\alpha}(\overline{Q}_T)$ .*

**Proof** In the expanded form, given  $\tau \in [0, 1]$ , the problem  $P[u] = (1 - \tau)P[0]$  is

$$\begin{cases} -tr \left( \mathcal{A}(x, \nabla u) \circ \overline{\nabla}^2 u \right) + \rho u + \nabla u \cdot_g \nabla V(x) = 0 & \text{in } Q_T \\ -\partial_t u + \frac{1}{2} |\nabla u|^2 = \delta u + \tau (\log(m_1) + V(x)) & \text{in } \Sigma_T \\ -\partial_t u + \frac{1}{2} |\nabla u|^2 + \delta u = \tau (\log(m_0) + V(x)) & \text{in } \Sigma_0 \end{cases} \tag{7.3}$$

So by Theorem 4.3, (4.5), and  $\tau \leq 1$  we get

$$\begin{aligned}
 \|u\|_{W^{1,\infty}} &\leq C(\delta, \|\tau \log(m_0)\|_{W^{1,\infty}}, \|\tau \log(m_1)\|_{W^{1,\infty}}, \|\tau V\|_{W^{2,\infty}}) \\
 &\leq C(\delta, \|\log(m_0)\|_{W^{1,\infty}}, \|\log(m_1)\|_{W^{1,\infty}}, \|V\|_{W^{2,\infty}}).
 \end{aligned}$$

Moreover, since  $V \in W^{2,\infty}$ , we get that the coefficients of the elliptic problem  $P[u] = (1 - \tau)P[0]$  are  $C^{0,\alpha}$  in the interior and  $C^{1,\alpha}$  on the boundary with respect to  $(t, x)$ , independently of  $\tau$ .

Fixed any local system of coordinates on  $M$ , we recall that the second covariant derivatives of  $\psi$  are

$$\nabla_{ij} \psi = \frac{\partial^2}{\partial x_i \partial x_j} \psi - \sum_k \Gamma_{ij}^k \frac{\partial}{\partial x_k} \psi$$

where the  $\Gamma_{ij}^k$  are the Christoffel symbols.

Hence, if we localize the differential system (7.3), we get a differential problem on  $\mathbb{R}^d$  which differs from (7.3) only in the first-order terms (because of the Christoffel symbols, which are smooth), so an elliptic problem with Hölder coefficients in their arguments. Therefore, we can apply the classical results on Schauder estimates (e.g. [18], and [29, Lemma 2.4] for the boundary estimates) in every local chart of a finite atlas ( $M$  is compact). We obtain a global  $C^{1,\alpha}(\overline{Q}_T)$  estimate on  $u$ , independent of  $\tau$ . □

We now observe that, thanks to Lemma 17.29 in [18], we have that  $E$  is closed in  $C^{2,\alpha}(\overline{Q_T})$  and that the set

$$S := \{\tau \in [0, 1] \mid \exists u \in C^{2,\alpha}(\overline{Q_T}) \text{ s.t. } P[u] = (1 - \tau)P[0]\}$$

is closed. Note that  $S$  is not empty because  $0 \in S$ . We want to prove that  $S$  is also open, so that it will coincide with  $[0, 1]$ . To this purpose, we prove that for each  $u \in E$  the linear differential system induced by the Gateaux derivative of  $P$  has one and only one solution.

**Lemma 7.3** *Given  $u \in E$ ,  $\phi \in C^{0,\alpha}(\overline{Q_T})$ ,  $\zeta_1, \zeta_0 \in C^{1,\alpha}(M)$  there exists one and only one solution  $\psi \in C^{2,\alpha}(\overline{Q_T})$  of the linear problem*

$$\begin{cases} -\partial_{tt}\psi - \Delta_g\psi - (\nabla^2\psi)(\nabla u, \nabla u) + 2\nabla(\partial_t\psi) \cdot_g \nabla u \\ \quad + [2\nabla(\partial_t u) - \nabla(|\nabla u|^2) + \nabla V] \cdot_g \nabla\psi + \rho\psi = \phi & \text{in } Q_T \\ -\partial_t\psi + 2\nabla u \cdot_g \nabla\psi - \delta\psi = \zeta_1 & \text{on } \Sigma_T \\ -\partial_t\psi + 2\nabla u \cdot_g \nabla\psi + \delta\psi = \zeta_0 & \text{on } \Sigma_0. \end{cases} \tag{7.4}$$

**Proof** The uniqueness of the solution follows from the maximum principle applied to the homogeneous system ( $\zeta_1 = \zeta_0 = \phi = 0$ ), thanks to the linearity of the problem.

We now prove the existence of a solution. Let  $L : C^{2,\alpha}(\overline{Q_T}) \rightarrow C^{0,\alpha}(\overline{Q_T}) \times C^{1,\alpha}(M)^2$  defined by the left side of system (7.4), in other words a function  $\psi \in C^{2,\alpha}$  solves (7.4) if and only if  $L[\psi] = (\phi, \zeta_1, \zeta_0)$ .

Let  $L_0 : C^{2,\alpha}(\overline{Q_T}) \rightarrow C^{0,\alpha}(\overline{Q_T}) \times C^{1,\alpha}(M)^2$  be the operator given by

$$L_0[\psi] = (-\partial_{tt}\psi - \Delta_g\psi + \rho\psi, -\delta\psi, \delta\psi)$$

and, for every  $\tau \in [0, 1]$ ,

$$\begin{aligned} L_\tau : C^{2,\alpha}(Q_T) &\longrightarrow C^\alpha(Q_T) \times C^{1,\alpha}(M)^2 \\ \psi &\longmapsto \tau L[\psi] + (1 - \tau)L_0[\psi]. \end{aligned}$$

so the expanded form of the problem  $L_\tau[\psi] = (\phi, \zeta_1, \zeta_0)$  is

$$\begin{cases} -\partial_{tt}\psi - \Delta_g\psi - \tau(\nabla^2\psi)(\nabla u, \nabla u) + 2\tau\nabla(\partial_t\psi) \cdot_g \nabla u \\ \quad + \tau[2\nabla(\partial_t u) - \nabla(|\nabla u|^2) + \nabla V] \cdot_g \nabla\psi + \rho\psi = \phi & \text{in } Q_T \\ -\tau\partial_t\psi + 2\tau\nabla u \cdot_g \nabla\psi - \delta\psi = \zeta_1 & \text{on } \Sigma_T \\ -\tau\partial_t\psi + 2\tau\nabla u \cdot_g \nabla\psi + \delta\psi = \zeta_0 & \text{on } \Sigma_0. \end{cases} \tag{7.5}$$

As in Lemma 7.2, we observe that if we write system (7.5) in the coordinates of any local chart, we obtain a uniformly elliptic problem. Hence, on every differential problem induced by the localizations of a finite atlas ( $M$  is compact), we can, by ellipticity, apply Theorem 6.17 or Theorem 6.30 of [18] (the former for  $t \in (0, T)$ , the latter for  $t \in \{0, T\}$ ). We obtain the global bound

$$\|\psi\|_{C^{2,\alpha}(\overline{Q_T})} \leq C(\|\psi\|_{C^\alpha(\overline{Q_T})} + \|\phi\|_{C^\alpha(\overline{Q_T})} + \|\zeta_1\|_{C^{1,\alpha}(M)} + \|\zeta_0\|_{C^{1,\alpha}(M)})$$

where  $C$  is independent from  $\tau$ . We conclude that there exists a constant  $C > 0$ , depending only on  $\rho, \delta$  and  $M$ , such that

$$\|\psi\|_{C^{2,\alpha}(\overline{Q_T})} \leq C\|L_\tau[\psi]\|_{C^{0,\alpha}(\overline{Q_T}) \times C^{1,\alpha}(M)^2} \quad \forall \tau \in [0, 1], \quad \forall \psi \in C^{2,\alpha}(Q_T).$$

Thanks to the bound above, we can apply the method of continuity for linear elliptic differential problems and we find that, for each  $\tau \in [0, 1]$ , the operator  $L_\tau$  is surjective if and

only if  $L_0$  is surjective. This means that the system (7.4) has solution if and only if there is a solution to the system

$$\begin{cases} -\partial_{tt}\psi - \Delta_g \psi + \rho\psi = \phi & \text{in } Q_T \\ -\delta\psi = \zeta_1 & \text{on } \Sigma_T \\ \delta\psi = \zeta_0 & \text{on } \Sigma_0. \end{cases}$$

This is a standard Poisson problem, which admits a unique solution since the differential operator  $-\frac{\partial^2}{\partial t^2} - \Delta_g + \rho$  is Fredholm of index zero and injective (see Proposition 1.8, Chapter 5 of [41]). Moreover, by elliptic regularity, the solution belongs to  $C^{2,\alpha}(\overline{Q}_T)$ .  $\square$

### 8 Appendix B: Proof of Lemma 6.3

In this section we provide the proof of Lemma 6.3, which we restate here for the reader’s convenience.

**Lemma 8.1** *Let  $m_0 \in L^1(M)$ . Then there exists a function  $U : [0, \infty) \rightarrow \mathbb{R}^+$  such that:*

- (i)  $U \in C^2(0, \infty)$ , is convex and satisfies  $\frac{U(r)}{r} \xrightarrow{r \rightarrow \infty} +\infty$ .
- (ii)  $P'(r)r - (1 - \frac{1}{d})P(r) \geq 0$ , and  $0 \leq P(r) \leq Kr$  for every  $r > 0$ , and for some  $K > 0$ , where  $P(r) = U'(r)r - U(r)$ .
- (iii)  $U(m_0) \in L^1(M)$ .

**Proof** The proof consists of two steps, which rely on two general facts: first of all, by the classical De la Vallée-Poussin criterion, any  $m_0 \in L^1(M)$  is integrable for some superlinear function  $\Psi$ . Secondly, by [44, Proposition 17.7], given a superlinear function  $\Psi$  one can build a new function  $U$  which stays below and satisfies the conditions needed for the displacement convexity. However, for our purposes, it will be needed to detail the possible construction of  $\Psi$  as well as Villani’s construction of  $U$ , in order to show that  $P$  satisfies a suitable sublinear growth. This is why we detail the proof, for the reader’s convenience.

**Step 1** Here we follow and slightly modify, to our purposes, the proof of de la Vallée-Poussin’s lemma proposed in [6]. If we set  $f(\lambda) := \text{meas}(\{m_0 > \lambda\})$ , then  $f \in L^1(0, \infty)$  and we can define the sequence

$$a_n := \inf\{t : \int_t^\infty f(\lambda)d\lambda < \frac{1}{n^2}\}. \tag{8.1}$$

We possibly increase  $a_n$  by defining the sequence  $x_n$  as

$$\begin{cases} x_1 = 1 \\ x_{n+1} = \max(2x_n, a_{n+1} + 1) \end{cases}$$

Now we define a piecewise linear function (which will play the role of  $\Psi'$ ) as follows:

$$\begin{cases} \Phi(x) = x & x \in (0, 1) \\ \Phi(x) = \Phi(x_n) + d_{n+1}(x - x_n) & \text{for } x \in (x_n, x_{n+1}) \ n \geq 1 \end{cases}$$

where  $d_n$  is defined as

$$\begin{cases} d_1 = 1, \\ d_{n+1} = \min\left(d_n, \frac{1}{x_{n+1}}, \frac{n+1-\Phi(x_n)}{x_{n+1}-x_n}\right) \end{cases}$$

We notice that  $\Phi$  is increasing and concave, because  $\{d_n\}$  is positive and decreasing. In addition, since  $d_{n+1} \leq \frac{1}{x_{n+1}} \leq \frac{1}{x}$  for  $x \in (x_n, x_{n+1})$ , we deduce that  $\Phi'(x) \leq \frac{1}{x}$  for  $x \geq 1$ . Moreover, by definition of  $d_{n+1}$ , we have

$$\Phi(x) \leq \Phi(x_n) + \frac{n + 1 - \Phi(x_n)}{x_{n+1} - x_n}(x - x_n) \leq n + 1 \quad \text{for } x \in (x_n, x_{n+1}). \tag{8.2}$$

This implies that

$$\int_1^\infty \Phi(\lambda) f(\lambda) d\lambda \leq \sup_k \sum_{n=1}^k (n + 1) \int_{x_n}^{x_{n+1}} f(\lambda) d\lambda$$

where the last sum is finite, since reordering summation and using  $x_n > a_n$  defined by (8.1), we have

$$\begin{aligned} \sum_{n=1}^k (n + 1) \int_{x_n}^{x_{n+1}} f(\lambda) d\lambda &= 2 \int_{x_1}^\infty f(\lambda) d\lambda - (k + 1) \int_{x_{k+1}}^\infty f(\lambda) d\lambda + \sum_{n=2}^k \int_{x_n}^\infty f(\lambda) d\lambda \\ &\leq 2\|f\|_{L^1} + \sum_{n=2}^k \frac{1}{n^2} \leq C. \end{aligned}$$

Hence  $\Phi f \in L^1(1, \infty)$ . Finally, we check that  $\Phi(r) \xrightarrow{r \rightarrow \infty} \infty$ ; for that, being  $\Phi$  increasing, it is enough to show that  $\Phi(x_n) \rightarrow \infty$ . But we have

$$\Phi(x_{n+1}) = 1 + \sum_{k=1}^n (\Phi(x_{k+1}) - \Phi(x_k)) = 1 + \sum_{k=1}^n d_{k+1}(x_{k+1} - x_k) \tag{8.3}$$

If  $d_{k+1} = \frac{1}{x_{k+1}}$ , then  $d_{k+1}(x_{k+1} - x_k) = 1 - \frac{x_k}{x_{k+1}} \geq \frac{1}{2}$  by definition of  $x_k$ . Similarly, if  $d_{k+1} = \frac{k+1-\Phi(x_k)}{x_{k+1}-x_k}$ , then  $d_{k+1}(x_{k+1} - x_k) = k + 1 - \Phi(x_k) \geq 1$  (see (8.2)). Therefore, by definition of  $d_k$ , we conclude that  $\sum_{k=1}^n d_{k+1}(x_{k+1} - x_k)$  do not converge if  $d_{k+1} \neq d_k$  for infinitely many indexes; but this means that the only case left out is that  $d_k$  is definitively constant after some  $k_0$ , and again this implies  $\sum_{k=1}^n d_{k+1}(x_{k+1} - x_k) \geq d_{k_0}(x_{n+1} - x_{k_0}) \rightarrow \infty$ . We conclude that the sum in (8.3) diverges, hence  $\Phi(x_n) \rightarrow \infty$ .

Therefore, we proved that  $\Phi$  is a concave, positive, piecewise linear function, which is increasing and unbounded, and satisfying  $\Phi'(x) \leq \frac{1}{x}$  for  $x \geq 1$ ; in addition,  $\Phi(\lambda) f(\lambda) \in L^1(0, \infty)$ . Now, one can smoothen  $\Phi$  by convolution with some function  $\zeta \in C^\infty$  which is supported in  $(-\frac{1}{2}, \frac{1}{2})$  and with unit mass. Namely, we define

$$\Psi'(r) := \int_{\mathbb{R}} \Phi(r - s)\zeta(s)ds, \quad \Psi(r) = \int_0^r \Psi'(s)ds.$$

We notice that  $\Psi' \geq 0$ , is increasing and differentiable, and satisfies  $\Psi''(r) \leq \frac{\zeta}{r}$ . Moreover, we have  $\Psi'(r) \leq \Phi(r + \frac{1}{2}) \leq \Phi(r) + \Phi(\frac{1}{2})$  (because  $\Phi$  is concave and sublinear). Hence we also have  $\Psi'(\lambda) f(\lambda) \in L^1(0, \infty)$ . This implies  $\Psi(m_0) \in L^1(M)$ , since

$$\int_M \Psi(m_0) = \int_0^\infty \Psi'(\lambda) f(\lambda) d\lambda < \infty.$$

Notice also that  $\Psi$  is superlinear, because  $\Psi'(r) \rightarrow \infty$  as  $r \rightarrow \infty$ , by the very same property of  $\Phi$ .

**Step 2** Given  $\Psi$  built in the previous step, which is a superlinear function, now we apply [44, Proposition 17.7] which provides with a function  $U \leq \Psi$  such that  $U(0) = 0$ ,  $U$  is

convex and superlinear, belongs to  $C^2(0, \infty)$  and satisfies  $P'(r)r - (1 - \frac{1}{d})P(r) \geq 0$ , where  $P(r) = U'(r)r - U(r)$ . All those properties are proved in [44, Proposition 17.7]. In addition, we also wish to show that  $P(r) \leq Cr$ , and to this purpose we recall the construction of  $U$  in Villani’s book. Given  $\Psi(r)$ , let us define  $u(r) = r^d\Psi(r^{-d})$ , and let  $\tilde{u}(r)$  be the lower convex envelope of  $u$  (the sup of linear functions lying below  $u$ ). The function  $U$  is then defined as

$$U(r) = r\tilde{u}(r^{-1/d}).$$

Since  $\tilde{u} \leq u$ , one obviously have  $U(r) \leq \Psi(r)$  by definition. Now, let us define

$$a_r := \sup_{0 < \xi \leq r^{-1/d}} \left[ \frac{u((2r)^{-1/d}) - u(\xi)}{\xi} \right]$$

which is a finite number because  $u(\xi) \xrightarrow{\xi \rightarrow 0} \infty$ . Since  $u(\xi) \geq u((2r)^{-1/d}) - a_r\xi$  for  $\xi \leq r^{-1/d}$ , by definition of the convex envelope the same inequality is satisfied replacing  $u$  with  $\tilde{u}$ ; when  $\xi = r^{-1/d}$ , this yields

$$u((2r)^{-1/d}) - \tilde{u}(r^{-1/d}) \leq a_r r^{-1/d}. \tag{8.4}$$

We estimate now  $a_r$ ; in fact, we have

$$a_r = \sup_{\lambda \in [1, \infty)} \left[ \frac{u((2r)^{-1/d}) - u((\lambda r)^{-1/d})}{(\lambda r)^{-1/d}} \right]$$

where we can write, by definition of  $u$ ,

$$\begin{aligned} \frac{u((2r)^{-1/d}) - u((\lambda r)^{-1/d})}{(\lambda r)^{-1/d}} &= (\lambda r)^{1/d} \left[ \frac{\Psi(2r)}{2r} - \frac{\Psi(\lambda r)}{\lambda r} \right] \\ &= (\lambda r)^{1/d} \int_{\lambda}^2 \left( \frac{\Psi(s)}{s} \right)' r dt. \end{aligned}$$

Now, by the construction in Step 1, we know that  $0 \leq \Psi''(s) \leq \frac{C}{s}$  for some constant  $C > 0$ . This implies that  $\frac{\Psi(s)}{s}$  is increasing and we have

$$\left( \frac{\Psi(s)}{s} \right)' = \frac{\Psi'(s)s - \Psi(s)}{s^2} \leq \frac{C}{s}$$

by Lagrange theorem. Therefore, for any  $\lambda \geq 1$  we estimate

$$\int_{\lambda}^2 \left( \frac{\Psi(s)}{s} \right)' r dt \leq C (\log 2) \mathbb{1}_{\{\lambda \in [1, 2]\}}.$$

Coming back to  $a_r$  we deduce that

$$a_r \leq C (\log 2)(2r)^{1/d}$$

and therefore, from (8.4), we conclude that

$$u((2r)^{-1/d}) - \tilde{u}(r^{-1/d}) \leq \tilde{C} \tag{8.5}$$

for some constant  $\tilde{C} > 0$ . Now we can estimate  $P(r)$ ; using the convexity of  $U$  and its definition, we have

$$\begin{aligned} P(r) &= U'(r)r - U(r) \\ &\leq U(2r) - 2U(r) = 2r\tilde{u}((2r)^{-1/d}) - 2r\tilde{u}(r^{-1/d}) \end{aligned}$$

$$\leq 2r(u((2r)^{-1/d}) - \tilde{u}(r^{-1/d}))$$

where we used that  $\tilde{u} \leq u$ . Using (8.5) we conclude that  $P(r) \leq 2\tilde{C}r$ .  $\square$

**Acknowledgements** The authors are members of Gnampa research group of Indam. We acknowledge the support of the Excellence Project MatMod@TOV of the Department of Mathematics of the University of Rome Tor Vergata and Italian (EU Next Gen) PRIN project 2022W58BJ5 (*PDEs and optimal control methods in mean field games, population dynamics and multi-agent models*, CUP E53D23005910006). We wish to warmly thank Giuseppe Savaré for fruitful discussions on the topics of the article.

**Funding** Open access funding provided by Università degli Studi di Roma Tor Vergata within the CRUI-CARE Agreement.

**Data availability** There are no associated data concerning this manuscript.

## Declarations

**Conflict of interest** We declare no Conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows in Metric Spaces and in the Space of Probability Measures. Lectures in Mathematics ETH Zürich, 2nd edn. Birkhäuser Verlag, Basel (2008)
2. Benamou, J.-D., Brenier, Y.: A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numer. Math.* **84**, 375–393 (2000)
3. Benamou, J.-D., Carlier, G., Santambrogio, F.: Variational Mean Field Games. *Active Particles*, vol. 1, pp. 141–171. Birkhäuser, Cham (2017)
4. Berger, M., Gauduchon, P., Mazet, E.: Le spectre d'une variété riemannienne. *Lecture Notes in Mathematics*. Springer (1971)
5. Buttazzo, G., Jimenez, C., Oudet, E.: An optimization problem for mass transportation with congested dynamics. *SIAM J. Control Optim.* **48**(3), 1961–1976 (2009)
6. Canizo, J.A.: The lemma of de la Vallée-Poussin, see <https://canizo.org/tex/vallee-poussin.pdf>
7. Cardaliaguet, P.: Weak solutions for first order mean field games with local coupling. In: *Analysis and Geometry in Control Theory and Its Applications*, pp. 111–158. Springer (2015)
8. Cardaliaguet, P., Graber, P.J.: Mean field games systems of first order. *ESAIM Control Optim. Calc. Var.* **21**, 690–722 (2015)
9. Cardaliaguet, P., Graber, P.J., Porretta, A., Tonon, D.: Second order mean field games with degenerate diffusion and local coupling. *NoDEA Nonlinear Differ. Equ. Appl.* **22**, 1287–1317 (2015)
10. Cardaliaguet, P., Mészáros, A.R., Santambrogio, F.: First order mean field games with density constraints: pressure equals price. *SIAM J. Control Optim.* **54**, 2672–2709 (2016)
11. Cardaliaguet, P., Munoz, S., Porretta, A.: Free boundary regularity and support propagation in mean field games and optimal transport. *J. Math. Pures Appl.* (9) **190**, 103599 (2024)
12. Cardaliaguet, P., Porretta, A.: An introduction to Mean Field Game theory. In: *Mean Field Games. Lecture Notes in Mathematics*, vol. 2281, pp. 1–158 (CIME Found. Subser.). Springer, Cham (2020)
13. Clason, C., Lorenz, D.A., Mahler, H., Wirth, B.: Entropic regularization of continuous optimal transport problems. *J. Math. Anal. Appl.* **494**, 124–432 (2021)

14. Cordero-Erausquin, D., McCann, R.J., Schmuckenschläger, M.: A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Invent. Math.* **146**, 219–257 (2001)
15. Daneri, S., Savaré, G.: Eulerian calculus for the displacement convexity in the Wasserstein distance. *SIAM J. Math. Anal.* **40**, 1104–1122 (2008)
16. Figalli, A., Villani, C.: Strong displacement convexity on Riemannian manifolds. *Math. Z.* **257**, 251–259 (2007)
17. Gigli, N., Tamanini, L.: Benamou–Brenier and duality formulas for the entropic cost on  $RCD^*(K, N)$  spaces. *Probab. Theory Relat. Fields* **176**, 1–34 (2020)
18. Gilbarg, D., Trudinger, N.: *Elliptic Partial Differential Equations of Second Order*, 2nd edn. Springer, Berlin (1983)
19. Gomes, D., Seneci, T.: Displacement convexity for first-order mean-field games. *Minimax Theory Appl.* **3**, 261–284 (2018)
20. Graber, P.J., Mészáros, A.R., Silva, F., Tonon, D.: The planning problem in mean field games as regularized mass transport. *Calc. Var. Partial Differ. Equ.* **58**, 1–28 (2019)
21. Grigoryan, A.: *Heat Kernel and Analysis on Manifolds*. AMA/IP Studies in Advances Mathematics. AMS, Providence (2009)
22. Hebey, E.: *Sobolev Spaces on Riemannian Manifolds*. Springer (2008)
23. Lasry, J.-M., Lions, P.-L.: Jeux à champ moyen. II-Horizon fini et contrôle optimal. *Comptes Rendus Mathématique* **343**, 679–684 (2006)
24. Lasry, J.-M., Lions, P.-L.: Mean field games. *Jpn. J. Math.* **2**, 229–260 (2007)
25. Lavenant, H., Santambrogio, F.: Optimal density evolution with congestion:  $L^\infty$  bounds via flow interchange techniques and applications to variational Mean Field Games. *Comm. PDE* **43**, 1761–1802 (2018)
26. Li, P., Yau, S.T.: On the parabolic kernel of the Schrödinger operator. *Acta Math.* **156**, 153–201 (1986)
27. Li, L., Zhang, Z.L.: On Li–Yau heat kernel estimate. *Acta Math. Sin. (Eng. Ser.)* **37**, 1205–1218 (2021)
28. Leonard, C.: A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.* **34**, 1533–1574 (2014)
29. Lieberman, G.: The nonlinear oblique derivative problem for quasilinear elliptic equations. *Nonlinear Anal. TMA* **8**, 49–65 (1984)
30. Liero, M., Mielke, A., Savaré, G.: Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures. *Invent. Math.* **211**, 969–1117 (2018)
31. Lions, P.-L.: Cours at Collège de France (2009–2010). <https://www.college-de-france.fr/site/pierre-louis-lions/course-2009-2010.htm>
32. Lott, J., Villani, C.: Ricci curvature for metric-measure spaces via optimal transport. *Ann. Math.* **169**(2), 903–991 (2009)
33. McCann, R.J.: A convexity principle for interacting gases. *Adv. Math.* **128**, 153–179 (1997)
34. McCann, R.J.: Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.* **11**, 589–608 (2001)
35. Mimikos-Stamatopoulos, N., Munoz, S.: Regularity and long time behavior of one-dimensional first-order mean field games and the planning problem. *Siam J Math Anal* **56**(1), 43–78 (2024)
36. Munoz, S.: Classical and weak solutions to local first order mean field games through elliptic regularity. *Ann. I. H. Poincaré Anal. Nonlinéaire* **39**, 1–39 (2022)
37. Orrieri, C., Porretta, A., Savaré, G.: A variational approach to the mean field planning problem. *J. Funct. Anal.* **277**, 1868–1957 (2019)
38. Otto, F., Villani, C.: Generalization of an Inequality by Talagrand and Links with the Logarithmic Sobolev Inequality. *J. Funct. Anal.* **277**, 361–400 (2019)
39. Porretta, A.: Regularizing effects of the entropy functional in optimal transport and planning problems. *J. Funct. Anal.* **284**, 109–759 (2023)
40. Simon, J.: Compact sets in  $L^p(0, T; B)$ . *Ann. Mat. Pura Appl.* **146**(4), 65–96 (1987)
41. Taylor, M.E.: *Partial Differential Equations I: Basic Theory*. Applied Mathematical Sciences. Springer New York (2010)
42. von Renesse, M.-K., Sturm, K.-T.: Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Commun. Pure Appl. Math.* **58**, 923–940 (2005)
43. Villani, C.: *Topics in Optimal transportation*. Graduate Studies in Mathematics, vol. 58. American Mathematical Society, Providence (2003)
44. Villani, C.: *Optimal transport: Old and New*. Grundlehren der mathematischen Wissenschaften, vol. 338, Springer-Verlag and Heidelberg GmbH & Co. KG (2016)