Regularizing effects of the entropy functional in optimal transport and planning problems

Alessio Porretta*

November 18, 2022

Abstract

We analyze optimal transport problems with additional entropic cost evaluated along curves in the Wasserstein space which join two probability measures \(m_0, m_1\). The effect of the additional entropy functional results into an elliptic regularization for the (so-called) Kantorovich potentials of the dual problem. Assuming the initial and terminal measures to be positive and smooth, we prove that the optimal curve remains smooth for all time. We focus on the case that the transport problem is set on a convex bounded domain in the \(d\)-dimensional Euclidean space (with no-flux condition on the boundary), but we also mention the case of Gaussian-like measures in the whole space. The approach follows ideas introduced by P.-L. Lions in the theory of mean-field games [27]. The result provides with a smooth approximation of minimizers in optimization problems with penalizing congestion terms, which appear in mean-field control or mean-field planning problems. This allows us to exploit new estimates for this kind of problems by using displacement convexity properties in the Eulerian approach.

Contents

1 Introduction 1
2 Notations and assumptions 6
3 Quasilinear elliptic equations 8
4 A penalized optimal transport problem 15
5 Displacement convexity and density estimates 17
6 Existence results 22
   6.1 Smooth solutions ......................................................... 22
   6.2 Transport with entropy in \(\mathbb{R}^d\) ...................................... 25
   6.3 Weak solutions, extensions and further comments ................. 27
7 Appendix: existence of solutions to the elliptic problem 31

1 Introduction

The aim of this article is to analyze a regularized version of the classical mass optimal transport problem between given measures in the Euclidean space \(\mathbb{R}^d\). As is well-known, the Kantorovich formulation of this problem reads as follows: given two probability measures \(m_0, m_1\) in \(\mathbb{R}^d\), find

\[
W_2(m_0, m_1) := \min \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \, d\gamma(x, y) : \gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d), \pi_1^\ast \gamma = m_0, \, \pi_2^\ast \gamma = m_1 \right\},
\]

where \(\gamma\) is a probability measure in \(\mathbb{R}^d \times \mathbb{R}^d\), \(\pi^i : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d\), \(i = 1, 2\), stand, respectively, for the \(x\) and \(y\) projections, and where \(\pi^\ast \gamma\) denotes the push forward of the measure \(\gamma\) through the mapping \(\pi^i\).

*Dipartimento di Matematica, Università di Roma Tor Vergata. Via della Ricerca Scientifica 1, 00133 Roma, Italy. Email: porretta@mat.uniroma2.it. The author is member of GNAMPA research group of Indam.
The dynamic formulation of the problem, due to Benamou and Brenier \[3\], shows that the optimal value is realized by the energy-minimizing problem

$$ W_2(m_0, m_1) = \min \int_0^T \int_{\mathbb{R}^d} \frac{1}{2} |v|^2 dm, \quad (m,v) : \begin{cases} m_t - \text{div}(vm) = 0 \\ m(0) = m_0, m(T) = m_1 \end{cases} \tag{1.2} $$

where the minimum is meant on all couples \((m,v)\) which satisfy the continuity equation in a suitable sense. The equivalence between (1.1) and (1.2) has a clear geometrical meaning in terms of the Wasserstein space of probability measures (with finite second moments), since \(W_2\) defines a metric in this space (so-called Kantorovich-Wasserstein distance) and the minimum curve \(m(t)\) in (1.2) is the corresponding geodesic connecting \(m_0\) and \(m_1\). The optimal velocity field \(v\) is actually uniquely associated to this geodesic and the energy term in (1.2) can be interpreted as a classical kinetic energy in terms of the metric derivative of the curve \(m(t)\) in the Wasserstein space (\[1\]). It is also known that the geodesic \(m(t)\) coincides with the so-called McCann’s displacement interpolation of the optimal plan \[28\], i.e.

$$ m(t) = ((1 - t) \gamma + t \pi_0) \gamma $$

for those and many other classical results in optimal transport theory, we refer to the books \[1\], \[32\], \[35\].

In this paper we study a natural regularization of the Wasserstein geodesic and, more generally, of curves \(m(t)\) which are optimal for similar transportation costs. It is well-known that the regularity of the geodesic curve, which is minimal in (1.2), depends not only on the regularity of the marginals but also on the positivity of the density; however, the positivity set of \(m\) may shrink in time, for \(t \in (0,1)\), see e.g. \[53\], Thm 1. This issue motivated the suggestion of possible regularizations of the classical optimal transport problem, the most popular being the so-called entropic regularization which has been intensively investigated in the recent times for numerical efficiency (see \[10\] and references therein). We also refer to \[12\], \[13\], \[14\], \[26\] for other related entropic perturbations, or regularizations, of the classical problem.

What we analyze here is the modification of the classical optimal transport functional by addition of an \(\varepsilon\)-entropy along the curve \(m(t)\). In the model case, this amounts to study

$$ \min \int_0^T \int_{\mathbb{R}^d} \frac{1}{2} |v|^2 dm + \varepsilon \int_0^T \int_{\mathbb{R}^d} \log \left( \frac{dm}{dv} \right) dm, \quad (m,v) : \begin{cases} m_t - \text{div}(vm) = 0 \\ m(0) = m_0, m(T) = m_1 \end{cases} \tag{1.3} $$

where the additional entropy is computed with respect to a reference measure \(\nu\). Typical choices for \(\nu\) include the classical Lebesgue measure, in the case that the transport problem is restricted to a bounded domain \(\Omega\), or the case of Gaussian measures, for problems in the whole space.

There are several motivations for the interest in this problem. First of all, we will see that the additional entropy term produces a regularization of the geodesics of the classical optimal transport problem. Indeed, the addition of \(\varepsilon\)-entropy in the functional yields an elliptic regularization which is reminiscent of the vanishing viscosity approximation in first order problems. This elliptic regularization is observed in terms of the Kantorovich potential provided by the dual problem; exploiting this approach may enrich the viewpoints on optimal transport problems.

Secondly, this kind of regularization is natural from the intrinsic geometry of optimal transport, since it can be readily interpreted as a deformation of the Wasserstein metric which regularizes the corresponding geodesics. In particular, even if we develop here this analysis in the Euclidean setting, the extension to Riemannian manifolds will be natural. We plan to exploit this issue in future work.

As a further motivation, this approach provides with a natural setting where Eulerian calculus can be fully justified. In particular, we will prove estimates related to displacement convexity inequalities which are naturally robust in this kind of approximation.

To put things into perspective, this kind of study was originally motivated by mean-field theories of optimal control and differential games. To this purpose, it is necessary to embed the classical optimal transport problem into a larger class of dynamical optimization problems, where the cost criterion involves both a kinetic-type energy of the curve \(m(t)\) and additional terms which minimize congestion effects:

$$ \min \int_0^T \int_{\mathbb{R}^d} \frac{1}{2} |v|^2 dm + \int_0^T \int_{\mathbb{R}^d} F(m(t)) dx dt : \begin{cases} m_t - \text{div}(vm) = 0 \\ m(0) = m_0, m(T) = m_1 \end{cases} \tag{1.4} $$

In (1.4), we will suppose \(F\) to be a nondecreasing function defined on the density of the curve \(m(t)\), so that the functional will reduce concentration effects on the minimal configurations.

Minimization problems such as (1.4) have been extensively studied in the last decade, especially in connection with mean-field control theory. In particular, the introduction of mean field game theory by J.-M. Lasry and P.-L. Lions \([20, 21, 27]\) boosted the interest in this kind of problems, giving an interpretation...
of the minima in terms of Nash equilibria of large populations’ differential games. In that context, any generic agent is represented as a dynamical state \( x(t) \) and aims at minimizing an individual cost given by

\[
\min \int_0^1 \frac{|x'(t)|^2}{2} dt + \int_0^1 \int_\Omega f(m(t)) dx dt
\]

(1.5)

where \( m(t) \) is a probability density which represents the distribution law of the population. More precisely, the agents consider \( m(t) \) as an exogenous data (an anticipated guess of the collective state) which at the Nash equilibrium coincides with the density of the population driven by the optimal strategies of the individuals. It turns out that, when \( f = F' \), the couple \((u,m)\) given, respectively, by the value function and by the density of the agents, satisfy the system

\[
\begin{cases}
-u_t + \frac{|Du|^2}{2} = F'(m) & \text{in } (0,T) \times \Omega, \\
m_t - \text{div}(m Du) = 0 & \text{in } (0,T) \times \Omega,
\end{cases}
\]

(1.6)

which coincides with the state-adjoint state system of the optimization problem (1.3).

The analysis of problem (1.3), together with natural generalizations, was initiated by P.-L. Lions in his courses at Collège de France devoted to mean field games, and then pursued by many authors in different directions, mostly relying on the study of the primal-dual optimization problems and/or on the Lagrangian formulation (1.5). For a non exhaustive list of contributions, see [22, 27, 46, 47, 48, 17, 2, 22, 51]. Let us stress that in this literature the state space if often assumed to be the flat torus (periodic case) and some final cost is often prescribed rather than prescribing the final marginal \( m(T) \).

In this article, we develop one of the pioneering ideas suggested by P.-L. Lions in this context ([27, Lessons 27/11 and 4/12 2009], which relies on the reduction of system (1.6) to a single elliptic equation on \( m \). This equation is possibly degenerate in the set where \( m \) vanishes, but adding an entropic term along the curve \( m(t) \) preserves the ellipticity in a strong form. In the particular case that \( F(m) = \epsilon m (\log m - 1) \), which corresponds to the functional (1.3) with uniform Lebesgue measure \( \nu \), this equation takes the following form:

\[
-u_t + 2 Du \cdot Du_t - D^2 u Du \cdot Du - \epsilon \Delta u = 0.
\]

(1.7)

Let us recall that here \( Du \) is the optimal velocity field for the functional (1.3), and the corresponding \( m \) can be reconstructed as \( m := \exp \left( \frac{1}{\epsilon} \left( \frac{|Du|^2}{2} - u_t \right) \right) \).

In this viewpoint, it is clear that the perturbed functional (1.3) yields a \textit{vanishing viscosity} approximation of the first order case represented by optimal transport trajectories. It would actually be justified to call problem (1.3) as the \textit{vanishing entropic} approximation of optimal transport. It is to be noticed that the optimal transportation problem corresponds to equation (1.7) complemented with \textit{nonlinear Neumann boundary conditions} at time \( t = 0 \) and \( t = T \):

\[
\begin{cases}
  u_t = \frac{|Du|^2}{2} - \epsilon \log(m_0) & \text{at } t = 0, \\
  u_t = \frac{|Du|^2}{2} - \epsilon \log(m_1) & \text{at } t = T.
\end{cases}
\]

(1.8)

The purpose of this article is to analyze the \( \epsilon \)-entropic problem (1.3) from this new perspective. Our results include two main goals:

(i) to give a suitable existence and regularity result for the nonlinear elliptic problem (1.7)-(1.8) in the case that the marginals \( m_0, m_1 \) are positive and smooth. This problem is also identified as a singular limit of penalized optimal transportation problems.

(ii) to use this smooth setting to justify different type of a priori estimates, possibly independent of the positivity of \( m \) (and possibly uniform as \( \epsilon \to 0 \)). Some of those estimates directly come from displacement convexity arguments developed in what is called the Eulerian viewpoint. This will provide an extension of regularity results which were obtained for mean field game systems (1.6) with different approaches ([22]).

In order not to overlap with existing results (mostly restricted to dynamics on the flat torus), and in order to provide tools which may be useful in further applications, we develop our analysis in general convex subdomains \( \Omega \subset \mathbb{R}^d \) where a no-flux condition is prescribed on the continuity equation, which implies conservation of mass. In addition, it is convenient to add a potential energy term to the functional; this introduces some inhomogeneity in the problem and allows us to include the case where the entropy additional term is computed with respect to general reference measures \( \nu \).

Let us now be more precise on the statement of our results, starting from the very model case of quadratic cost.
Theorem 1.1 Let \( \Omega \) be a convex bounded smooth domain, and assume that \( V \in W^{2,\infty}(\Omega) \). Let \( m_0,m_1 \in \mathcal{P}(\Omega) \cap W^{1,\infty}(\Omega) \) and suppose that \( m_0,m_1 > 0 \) in \( \overline{\Omega} \). Then there exists \( u \in C^2(Q) \cap C^1(\overline{Q}) \) which is a classical solution of the problem

\[
\begin{align*}
-u_t + 2Du \cdot Du_t - D^2uDu \cdot Du - \varepsilon \Delta u + DV \cdot Du &= 0 & & \text{in } (0,T) \times \Omega, \\
Du \cdot \nu &= 0 & & \text{on } (0,T) \times \partial \Omega, \\
u_t + \varepsilon \log(m_0) - V(x) &= 0 & & \text{at } t = 0, x \in \Omega, \\
u_t + \varepsilon \log(m_1) - V(x) &= 0 & & \text{at } t = T, x \in \Omega.
\end{align*}
\]

Moreover, the function \( u \) is unique up to a constant and the function \( m \) defined as

\[
m := \exp \left( \frac{1}{\varepsilon} \left( \frac{|Du|^2}{2} - u_t - V(x) \right) \right)
\]

is the unique minimizer of the functional

\[
B(m,v) := \int_0^T \int_{\mathbb{R}^d} \frac{1}{2}|v|^2 dm + \varepsilon \int_0^T \int_{\mathbb{R}^d} \log \left( \frac{dm}{d\varrho} \right) dm,
\]

\[
\begin{align*}
&m_t - \div(vm) = 0 & & \text{in } (0,T) \times \Omega, \\
&v \cdot \nu = 0 & & \text{on } (0,T) \times \partial \Omega, \\
&m(0) = m_0, m(T) = m_1
\end{align*}
\]

where \( \varrho = e^{-V(x)} dx \). In particular, we have that \( m \in C^1(Q) \cap C^0(\overline{Q}) \) and \( m > 0 \) in \( \overline{Q} \).

The proof of Theorem 1.1 relies on gradient estimates for quasilinear elliptic equations. To this respect, we follow Lions’ approach but we also exploit some extension. In particular, even if we restrict here our analysis to positive measures, we give extra estimates which are local in time and independent of the positivity of the marginals \( m_0, m_1 \). Precisely, we prove that for \( t \in (0,T) \) there exists a constant \( K_t \) (depending only on the upper bounds of \( u, m \), and on the distance of \( t \) from \( t = 0 \), \( T \)) such that

\[
\theta |Du(t)|^2 + \varepsilon \log m(t) \leq K_t \quad \forall t \in (0,T),
\]

for some \( \theta \in (0,1) \). This gives a quantitative estimate of the gradient of \( u \) on the support of \( m \), which we hope can be useful for further generalizations to non strictly positive measures.

An extension of Theorem 1.1 is also provided for the case of optimal transport in the whole space, relying on the dissipation property of the potential \( V \), whenever the additional entropy is taken with respect to Gaussian-like measures. This will serve as a prototype for a development of similar analysis on non compact manifolds, which seems of interest in the optimal transport literature. For a more readable statement, here we rescale the potential \( V \) according to the parameter \( \varepsilon \).

Theorem 1.2 Assume that \( V \in W^{2,\infty}(\mathbb{R}^d) \) satisfies \( D^2V(x)\xi \cdot \xi \geq \gamma |\xi|^2 \) for some \( \gamma > 0 \), and for every \( x, \xi \in \mathbb{R}^d \). Let \( m_0, m_1 \in C^1(\mathbb{R}^d) \) satisfy

\[
m_0 e^V, m_1 e^V \in W^{1,\infty}(\mathbb{R}^d) \quad \text{and} \quad c_0 e^{-V(x)} \leq m_0 \leq C_0 e^{-V(x)}, \quad c_1 e^{-V(x)} \leq m_1 \leq C_1 e^{-V(x)}
\]

for some positive constants \( c_i, C_i > 0, i = 0, 1 \). Then the problem (1.12) with \( \nu = e^{-V(x)} dx \) admits a unique minimizer \( m \) given by \( m = e^{-V} \exp \left( \frac{1}{\varepsilon} \left( \frac{|Du|^2}{2} - u_t \right) \right) \), where \( u \) is a solution of

\[
\begin{align*}
-u_t + 2Du \cdot Du_t - D^2uDu \cdot Du - \varepsilon \Delta u + DV \cdot Du &= 0 & & \text{in } (0,T) \times \mathbb{R}^d, \\
\end{align*}
\]

Moreover, we have \( m \in C^1([0,T] \times \mathbb{R}^d) \) and there exist constants \( \gamma_0, \gamma_1 \) (depending on \( \varepsilon, \gamma, c_0, c_1, C_0, C_1 \)) such that

\[
\gamma_0 e^{-V(x)} \leq m(t,x) \leq \gamma_1 e^{-V(x)} \quad \forall (t,x) \in [0,T] \times \mathbb{R}^d.
\]

As we mentioned before, the results on the classical optimal transport with quadratic cost are embedded in a larger class of results provided for the system

\[
\begin{align*}
-u_t + H(Du) &= f(m) + \varepsilon \log m + V(x) & & \text{in } Q, \\
m_t - \div(m H_p(Du)) &= 0 & & \text{in } Q, \\
H_p(Du) \cdot \nu &= 0 & & \text{on } (0,T) \times \partial \Omega, \\
m(0,\cdot) &= m_0, m(T,\cdot) = m_1
\end{align*}
\]

where \( \Omega \) is a smooth bounded convex subset of \( \mathbb{R}^d \), and \( Q = (0,T) \times \Omega \). Here \( H \) is a strictly convex radial function, with power-like growth \( (H \simeq |Du|^q \text{ with } q > 1) \) and \( f \) is a nondecreasing function, satisfying quite general conditions, which are precisely given in the next Section.
When $H$ is uniformly convex and with quadratic growth, the results that we obtain for problem (1.12) include both the existence of smooth solutions (when the marginals $m_0, m_1$ are positive) and further global and local estimates independent of the positivity of $m$. In the following statement we gather results which appear, in more details, in Theorem 3.1 and Proposition 5.2.

**Theorem 1.3** Let $Q$ be a bounded smooth convex domain in $\mathbb{R}^d$ and let $V \in W^{2,\infty}(\Omega)$. Let $H = h(|p|)$ for some (strictly) convex increasing function $h \in C^2([0,\infty))$ with $h'(0) = 0$. Assume in addition that $H$ satisfies conditions (2.3)- (2.4), and let $f \in C^2(0,\infty)$ be a nondecreasing function. Assume that $m_0, m_1 \in W^{1,\infty}(\Omega)$ and $m_0, m_1 > 0$ in $\Omega$.

Then there exists a unique couple $(u, m)$ such that $u \in C^{2,\alpha}(\Omega) \cap C^{1,\alpha}(\partial \Omega)$, $m \in C^{1,\alpha}(\Omega) \cap C^{0,\alpha}(\partial \Omega)$, $\int_\Omega u(T)m_1 = 0$, $\int_\Omega m(t) = 1$ for all $t$, and (1.12) is satisfied in a classical sense. We also have that $m \geq 0$ in $Q$ and is a minimizer of the corresponding optimal transport problem, while $u$ is a solution to the elliptic problem

\[
\begin{aligned}
- \text{tr}(A(Du)) D^2 u + DV(x) \cdot H_p(Du) &= 0 & \text{in } Q, \\
- u_t + H(Du) - V(x) &= f(m_1) + \varepsilon \log(m_1) & \text{at } t = T, x \in \Omega, \\
- u_t + H(Du) - V(x) &= f(m_0) + \varepsilon \log(m_0) & \text{at } t = 0, x \in \Omega,
\end{aligned}
\]

(1.13)

where $A$ is given by (2.8).

Moreover, if $f$ satisfies

\[
\exists c_0, r_0 > 0 : \quad f'(r) \geq \frac{c_0}{r} \quad \forall r \geq r_0,
\]

then there exist constants $K_0, K_1$, independent of $\varepsilon$, such that, for $1 \leq p \leq \infty$:

\[
\|m(t)\|_{L^p(\Omega)} \leq K_0 (\|m_0\|_{L^p(\Omega)} + \|m_1\|_{L^p(\Omega)} + 1) \quad \forall t \in [0, T]
\]

(1.14)

and

\[
\|m(t)\|_{L^p(\Omega)} \leq K_1 (t^{-q} + (T - t)^{-q}) \quad \forall t \in (0, T)
\]

(1.15)

for some $q > 0$.

Let us comment on the above statement in relation to the previous literature which analyzed problem (1.12) (often called mean-field game systems).

In the first part of Theorem 1.3 we prove that the system has smooth solutions whenever an $\varepsilon$-entropy term is added; this generalizes Theorem 1.1 (in which $H(p) = \frac{|p|^2}{2}$ and $f = 0$). Previous results showing the existence of smooth solutions were obtained for the periodic case ($\Omega$ is the torus) in [30], when a final condition is prescribed on $u$ rather than for the transport problem; this latter one was previously discussed in [27], as we already said. Recent results for the one-dimensional case can also be found in [2, 29].

Let us stress that the main difference of the transport problem, compared to the case when a final condition is prescribed on $u$, lies in the estimate of the sup-norms of $m$ and $u$. Indeed, while prescribing a final cost fixes the $L^\infty$-norm of $u$, this is no longer the case for the transport problem, which is invariant by addition of constants to $u$. This leads to a different strategy in order to close the gradient estimate, because this latter one depends on the $L^\infty$-norm of $u$, see Theorem 3.1. In the present paper we circumvent this problem through the obtention of preliminary estimates on $m$ and through the choice of a suitable normalization for $u$. Both those ingredients contain novel estimates, locally and globally in time. In particular, this is where estimates (1.14) and (1.15) play a role. Such global and local (in time) $L^p$ estimates only depend on the behavior of $f(m)$ for large $m$, so they are not relying on the positivity of the marginals. Letting $\varepsilon \to 0$, those estimates are exported for so-called weak solutions of mean field game systems, as introduced in [5, 6] and studied in [7, 8, 31]. A statement in this direction can be found in Theorem 5.3. Notice that those weak solutions coincide with relaxed minima of the functionals. In particular, the estimates (1.14), (1.15) extend similar estimates proved by H. Lavenant and F. Santambrogio [22] for the case $H(p) = \frac{|p|^2}{2}$ using flow-interchange techniques. Our approach is different and relies on displacement convexity inequalities which were first exploited by D. Gomes and T. Seneci for mean-field game systems on the torus [10]. To this respect, our approach can be seen as the Eulerian route to the $L^p$ estimates of $m$, compared to the Lagrangian and gradient flow approach developed in [22]. Let us observe that the global estimates (1.14) are classically known for the plain optimal transport problem (1.14), since McCann’s work [23], as a consequence of the geodesic convexity of $m \mapsto \int_\Omega m^p$. The local estimates (1.15) are instead a new effect induced by entropic or congestion-like terms in the functional.

Finally, we will extend some of the previous results to more general Hamiltonians. In Theorem 6.1 we include the case that the Hamiltonian has $q$-growth in the gradient, although the uniform convexity is still needed to get smooth solutions. Conversely, the case that $H = |Du|^q$ corresponds to possible degeneracy
and convex. In order to build smooth solutions (say, $u$ coming from the entropy functional. Of course, we could have embedded the term sense); in particular, $W$ function to the boundary. In particular, by smoothness of $\Omega$, we have that $C$ with the space of $\vec{\nu}$ the boundary of $\Omega$ and by $\vec{\nu}$ the unit outward normal vector on $\partial \Omega$. Throughout the paper, we denote by $\Omega$ an open bounded convex subset of $\mathbb{R}^d$. In Section 2 we introduce precise assumptions and notations. In Section 3 we discuss the nonlinear elliptic problem and we derive the main gradient estimates which are needed to construct smooth solutions. Here we essentially follow the same strategy introduced by P.L.-Lions and refined in [30], which leads to gradient bounds depending on the $L^\infty$-norm of $u$ (Theorem 3.4).

In Section 4, we build solutions for a penalized approximation of the optimal transportation problem, in which the constraint on $m$ at time $t = 0, t = T$ is relaxed. This is similar to the natural relaxation occurring in control theory, when exact controllability problems are relaxed into approximate controllability problems penalizing the final target. This construction is important in our strategy, since it allows us to show how a suitable normalization of $u$ can be controlled by $L^\infty$-bounds of $m$ (see Lemma 4.1). In Section 5 we extend to our setting the displacement convexity inequalities obtained in [16] (see also [2]) and we develop new $L^p$ estimates for the density, both globally and locally in time (Proposition 5.2). This is where we recover and generalize the results of [22]. In Section 6 we deduce our main results, say Theorem 6.1 and 6.2, with specific attention to the model case of classical optimal transportation. We also prove a few extra results which show possible further developments; for example, we consider the limit as the logarithmic term goes to zero, showing convergence towards weak solutions of mean-field game systems (Theorem 6.5).

Finally, we have detailed in the Appendix the construction of solutions to the elliptic equation with nonlinear Neumann boundary conditions. Unfortunately, due to the specific origin of our elliptic problem, which is set in the cylindrical domain $(0, T) \times \Omega$, we cannot rely on classical results (mostly obtained in smooth domains or in nonsmooth domains with simpler operators). This is why we provide a self-contained proof, where we use a reflection argument to handle the Neumann condition on $\partial \Omega$. It is only in this step that we use the radial structure of the Hamiltonian. This extra condition would not be used in the periodic case, or if $x$ was taken in a compact manifold without boundary.

## 2 Notations and assumptions

Throughout the paper, we denote by $\Omega$ an open bounded convex subset of $\mathbb{R}^d$, $d \geq 1$. We denote by $\partial \Omega$ the boundary of $\Omega$ and by $\vec{\nu}$ the unit outward normal vector on $\partial \Omega$. We will assume that $\Omega$ is of class $C^3$ (i.e. $\partial \Omega$ is locally the graph of a $C^3$ function) and, for $x \in \Omega$, we denote by $d(x) := \text{dist}(x, \partial \Omega)$ the distance function to the boundary. In particular, by smoothness of $\Omega$, we have that $d(x)$ is a $C^2$ function in the set $\{x \in \Omega : 0 < d(x) < \delta_0\}$ for some $\delta_0 > 0$. In addition, we also have $D^2d \leq 0$ since $\Omega$ is convex.

Given $\Omega$, we denote by $\mathcal{P}(\Omega)$ the space of probability measures in $\Omega$, and by $L^p(\Omega)$ the standard Lebesgue space, for $p \in [1, \infty]$. The space $W^{k, \infty}(\Omega)$ is the space of functions with bounded $k$-th derivatives (in weak sense); in particular, $W^{1, \infty}(\Omega)$ coincides with the space of Lipschitz continuous functions, and $W^{2, \infty}(\Omega)$ with the space of $C^1$ functions whose first derivatives are Lipschitz continuous.

Given $T > 0$ and $t \in (0, T)$, we consider the following first order evolution system in $(0, T) \times \Omega$:

\begin{equation}
\begin{aligned}
-u_t + H(Du) &= f(m) + \varepsilon \log m + V(x) & \text{in } Q := (0, T) \times \Omega, \\
m_t - \text{div}(m H_p(Du)) &= 0 & \text{in } Q, \\
H_p(Du) \cdot \vec{\nu} &= 0 & \text{on } (0, T) \times \partial \Omega, \\
m(0, \cdot) &= m_0, \quad m(T, \cdot) = m_1 & \text{in } \Omega,
\end{aligned}
\end{equation}

(2.1)

where $u, m$ are functions of $(t, x)$ and $u_t, m_t$ denote the partial derivatives in time, $Du$ the gradient vector of $x$-derivatives, and $\text{div}(-)$ is the divergence operator in the $x$-variables ($\text{div}(F) = \sum_{i=1}^d \partial_x F_i$).

In (2.1), we assume that $V(x)$ is a Lipschitz continuous function (but it will be often required to be in $W^{2, \infty}(\Omega)$) and that $f : (0, \infty) \to \mathbb{R}$ is a nondecreasing $C^1$ function.

(2.2)

Of course, we could have embedded the term $\varepsilon \log m$ into $f$, but we decided to make it explicit the contribution coming from the entropy functional.

The function $H : \mathbb{R}^d \to \mathbb{R}$ (so-called Hamiltonian function) will be assumed smooth (at least of class $C^2$) and convex. In order to build smooth solutions (say, of class $C^2$) we will strengthen both the regularity and the convexity of $H$. In the simplest setting, this will require $H$ to be a $C^3$ function satisfying the following conditions (hereafter $H_p, H_{pp}, H_{ppp}$ denote the derivatives of $H(p)$):

\begin{equation}
\exists \alpha_H, \beta_H > 0 : \quad \alpha_H I_d \leq H_{pp}(p) \leq \beta_H I_d \quad \forall p \in \mathbb{R}^d
\end{equation}

(2.3)
where $I_d$ is the identity matrix in $\mathbb{R}^d$, and
\[
H_{pp}(p) \text{ is uniformly bounded for } p \in \mathbb{R}^d.
\]
(2.4)

Conditions (2.3)–(2.4) describe functions $H$ which are uniformly convex and with quadratic growth, whose model case is obviously $H = \frac{H(p)}{p^2}$. This set of assumptions will be required in the displacement convexity estimates, Section 5, and in the statements which rely on them.

However, in most of our results the two above conditions can be generalized into the following form, modeled on the case of Hamiltonian with superlinear growth of power $q$: \[
\exists \alpha_H, \beta_H > 0 \text{ and } \bar{\omega} \geq 0:\quad \alpha_H (|p| + \bar{\omega})^{q-2} I_d \leq H_{pp}(p) \leq \beta_H (|p| + \bar{\omega})^{q-2} I_d \quad \forall p \in \mathbb{R}^d \setminus \{0\} \tag{2.5}
\]
\[
|H_{pp}(p)| \leq \gamma (1 + |p|)^{2(q-2)} \quad \forall p \in \mathbb{R}^d. \tag{2.6}
\]

Let us notice that if $H \in C^2$ is strictly convex and satisfies $\alpha |p|^{q-2} \leq H_{pp}(p) \leq \beta |p|^{q-2}$ only for $|p|$ large, then it satisfies (2.5) (with $\bar{\omega} = 1$) for every $p \in \mathbb{R}^d$.

Assumptions (2.5)–(2.6) describe functions $H$ with $q$-growth in the gradient (thus generalizing (2.3)–(2.4) which correspond to $q = 2$). Of course, the case when $\bar{\omega} = 0$ is special, and corresponds to singular, or degenerate, $H_{pp}$. In that case, we will be able to show only existence of weak, rather than classical, solutions.

We now rephrase system (2.4) into a single elliptic equation in $(t, x)$ variables. To this purpose, we denote
\[
f^\varepsilon(r) := f(r) + \varepsilon \log r, \quad \varphi(r) := (f^\varepsilon)^{-1}(r) = (f + \varepsilon \log)^{-1}(r).
\]
Hence, the first equation in (2.1) implies
\[
m = \varphi(-u_t + H(Du) - V(x)). \tag{2.7}
\]
Computing formally, we have, using the first equation,
\[
(f'(m) + \varepsilon/m)m_t = [-u_{tt} + H_p(Du) \cdot Du] \\
(f'(m) + \varepsilon/m) \text{div}(m H_p(Du)) = [-Du_t \cdot H_p(Du) + D^2 u H_p(Du) \cdot H_p(Du) - DV(x) \cdot H_p(Du)] \\
+ (f'(m)m + \varepsilon) \text{tr} H_{pp}(Du) D^2 u \\
\]
Therefore, subtracting the previous two terms and using the second equation in (2.1), we get
\[
\begin{cases}
m = (f + \varepsilon \log)^{-1}(-u_t + H(Du) - V(x)) \\m - 2H_p(Du) \cdot Du_t + D^2 u H_p(Du) \cdot H_p(Du) - (m f'(m) + \varepsilon) \text{tr} H_{pp}(Du) D^2 u + DV(x) \cdot H_p(Du) = 0
\end{cases}
\]
Eventually, the second equation of the system can be shortly written as a quasilinear equation in the space-time variables:
\[
-\text{tr} (A D^2 u) + DV(x) \cdot H_p(Du) = 0
\]
where
\[
A := \begin{pmatrix}
1 & -H_p(Du) \\
-H_p(Du) & H_p(Du) \otimes H_p(Du) + (m f'(m) + \varepsilon) H_{pp}(Du)
\end{pmatrix}
\]
and we recall that $m$ is a function of $Du$ (see (2.7)). Let us observe that $A$ is the contribution of two terms, namely
\[
A := A_1 + A_2 = \begin{pmatrix}
1 & -H_p(Du) \\
-H_p(Du) & H_p(Du) \otimes H_p(Du)
\end{pmatrix} + \begin{pmatrix}
0 & 0 \\
0 & (m f'(m) + \varepsilon) H_{pp}(Du)
\end{pmatrix} \tag{2.8}
\]
and it can be checked that $A$ is elliptic provided $H_{pp} > 0$:
\[
\text{if } \eta = (s, \xi) \Rightarrow A \eta \cdot \xi = (s - H_p(Du) \cdot \xi)^2 + (m f'(m) + \varepsilon) H_{pp}(Du) \xi \cdot \xi > 0 \quad \text{if } \eta \neq 0. \tag{2.9}
\]
Let us notice that, when $\varepsilon = 0$, the operator becomes degenerate elliptic; in particular, the matrix $A$ would be elliptic in the set $\{m > 0\}$ and would possibly degenerate only in the set where $m$ vanishes, assuming $f$ increasing. Otherwise, even if $\varepsilon > 0$, $A$ can degenerate if $H_{pp}$ vanishes at some point.

We point out that the space-time elliptic problem also enjoys a divergence form structure, which is clearly induced by the continuity equation in (2.1). This is very important for considering possibly weak formulations. Anyway, in the following we will first approach the elliptic problem in its non-divergence form, through the use of maximum principle.
Finally, we observe that the planning conditions which prescribe the marginals of \( m \) turn into a nonlinear Neumann condition for \( u \):

\[-u_t(0, \cdot) + H(Du(0, \cdot)) = f^c(m_0) + V; \quad -u_t(T, \cdot) + H(Du(T, \cdot)) = f^c(m_1) + V.\]

To conclude, problem (2.1) can be rephrased as the following quasilinear elliptic problem:

\[
\begin{cases}
- \text{tr} \left( A(x, Du) D^2 u \right) + D V(x) \cdot H_p(Du) = 0 & \text{in } Q, \\
- u_t + H(Du) = f^c(m_1) + V(x) & \text{at } t = T, x \in \Omega, \\
- u_t + H(Du) = f^c(m_0) + V(x) & \text{at } t = 0, x \in \Omega, \\
H_p(Du) \cdot \vec{\nu} = 0 & \text{on } (0, T) \times \partial \Omega \\
\end{cases}
\]

(2.10)

where \( f^c(r) = f(r) + \varepsilon \log r \). Unfortunately, in (2.10) the boundary condition takes the form of a nonlinear Neumann condition defined piecewisely on the time-space boundary. This makes it harder to construct smooth solutions; it is known that, in general quasilinear problems with nonlinear boundary conditions, Lipschitz solutions may even lack of \( C^1 \)-regularity if the domain is not sufficiently smooth. This is the only reason why we will simplify our setting requiring the Hamiltonian \( H \) to be a radial function, namely that

\[ H(p) = h(|p|) \quad \text{for some } h \in C^1([0, \infty)), \quad h'(0) = 0. \]

(2.11)

This structure assumption on \( H \) reduces the boundary condition on \( \partial \Omega \) to a standard Neumann condition. The only point where this is needed appears in the Appendix, where we prove the \( C^{1,\alpha} \) regularity (up to the boundary) for the elliptic equation. Of course, all conditions given above on \( H \) can be rephrased in terms of the real function \( h(r) \), but we decided to keep the whole exposition for a general function \( H \), since all results would stand in the general case up to a modification of the \( C^{1,\alpha} \) result in the Appendix. Moreover, all results would hold for general \( H \), not necessarily radial, in case of compact state space, e.g. if the state variable \( x \) belongs to the flat \( d \)-dimensional torus.

### 3 Quasilinear elliptic equations

In this Section we directly study the quasilinear elliptic equation

\[- \text{tr} \left( A(y, Du) D^2 u \right) = f, \]

(3.1)

where \( u = u(y), \ y \in \mathcal{O} \subset \mathbb{R}^n \).

We start with a computational lemma which underlines the typical structure used for gradient bounds of (possibly degenerate) quasilinear equations.

**Lemma 3.1** Let \( u \in C^3(\mathcal{O}) \) be a solution to (3.1), with \( f \in C^1(\mathcal{O}) \) and \( A(y, \eta) \in C^1(\mathcal{O} \times \mathbb{R}^n) \). Then we have:

(i) for any \( M \in \mathbb{R} \), the function \((u + M)^2\) satisfies

\[- \text{tr} \left( A(y, Du) D^2 (u + M)^2 \right) = 2f(u + M) - 2A(y, Du)Du \cdot Du \]

(3.2)

(ii) for any \( C^2 \) function \( \psi : \mathbb{R}^n \to \mathbb{R} \), we have that the function

\[ w := \psi(Du) \]

satisfies

\[- \text{tr} \left( A D^2 w \right) = Df \cdot D \psi + \text{tr} \left( [D_y A \cdot Dw + D_w A \cdot D \psi] D^2 u \right) - \sum_{k,l=1}^n (D^2 \psi)_{kl} (A D u_k \cdot Du_l) \]

(3.3)

where \( A \) and \( \psi \) are computed on \( Du \), and where \( D_y A \cdot Dw \) and \( D_w A \cdot D \psi \) are the matrices with \( ij \)-th component given, respectively, by \( \sum_{t=1}^n \partial_{y_t} (a_{ij}) w_t \) and \( \sum_{k=1}^n \partial_{u_k} (a_{ij}) \psi_k \).

**Proof.** To obtain (3.2), it is enough to observe that

\[ D^2 (u + M)^2 = 2(u + M) D^2 u + 2Du \otimes Du. \]

Applying the matrix \( A(y, Du) \) and taking traces, (3.2) follows from (3.1).

8
In the case of (ii), we compute (denoting $z_k = \partial_k z$ the partial derivatives):

$$w_i = \psi_k(Du)u_{ki}, \quad w_{ij} = \psi_k(Du)u_{ki,j} + \psi_{kk}u_{ki}u_{kj}$$

so that

$$\text{tr} \left( A(y, Du) D^2 u \right) = \psi_k(Du)\partial_k [a_{ij}u_{ij}] - \psi_k(Du)\partial_{ij}(a_{ij})u_{ki}u_{ij} - \partial_{y_k}(a_{ij})\psi_k(Du)u_{ij} + \psi_{kk}a_{ij}u_{ki}u_{kj}$$

$$= -\psi_k(Du)\partial_k f - (\partial_{ij}(a_{ij})w_l) u_{ij} - \partial_{y_k}(a_{ij})\psi_k(Du)u_{ij} + (D^2\psi)_{kl} (A Du_k \cdot Du_l)$$

Now we consider our specific case, where the matrix $A$ is induced by the optimal transport problem and given by $(2.8)$. In particular, we consider the solution $u$ to the elliptic problem

$$-\text{tr} \left( A(x, Du) D^2 u \right) + pu + DV(x) \cdot H_p(Du) = 0 \quad \text{in } Q,$$

$$-u_t + H(Du) = f^\ast(m_1) + V(x) \quad \text{at } t = T, x \in \Omega,$$

$$-u_t + H(Du) + \delta u = f^\ast(m_0) + V(x) \quad \text{at } t = 0, x \in \Omega,$$

and satisfies conditions $(2.5)-(2.6)$ for some $q > 1$. Let us explicitly note that assumption $(2.5)$, together with the regularity of $H$, also implies the following:

$$H_p(p) \cdot p - H(p) \geq \gamma_0 |p|^q - \gamma_1 \quad \forall p \in \mathbb{R}^d$$

for some $\gamma_0, \gamma_1 > 0$.

We stress that the additional terms $\rho u$ and $\delta u$ in $(3.4)$, added in the interior and, respectively, on the boundary, reinforce the coercivity of the elliptic operator and guarantee the uniqueness of solutions and the validity of the maximum principle. We start with a bound which is uniform on $\rho$.

**Lemma 3.2** Let $u \in C^2(Q) \cap C^1(\overline{Q})$ be a solution of $(3.4)$. Then we have

$$\delta \|u\|_\infty \leq (|f^\ast(m_0)| + V\|_\infty + |f^\ast(m_1)| + V\|_\infty).$$

**Proof.** Apply the maximum principle (respectively, the minimum principle) to $u - C$ (respectively, $u + C$), with $C = \frac{1}{\delta}(|f^\ast(m_0)| + V\|_\infty + |f^\ast(m_1)| + V\|_\infty)$. □

The next step is the main tool in this approach and consists in a gradient bound for solutions to $(3.4)$. The idea was introduced by P.-L. Lions in [27, Lessons 27/11 and 4/12 2009], following the lines of the classical Bernstein method. It was recently refined in [30] to handle the general case with $x$-dependent nonlinearities. In the aforementioned results, $\Omega$ was assumed to be the flat torus. We extend here this kind of estimates for the case of bounded domains with Neumann conditions.

We start with a consequence of Lemma 3.1

**Lemma 3.3** Let $u \in C^2(Q) \cap C^1(\overline{Q})$ be a solution of $(3.4)$. Then we have

$$u_t(t,x) \leq \sup_{\Sigma_0 \cup \Sigma_T} (u_+) \quad \text{and} \quad |u_t(t,x)| \leq \sup_{\Sigma_0 \cup \Sigma_T} |u_t| \quad \forall (t,x) \in \overline{Q}. \quad (3.6)$$

**Proof.** Since $A$ is independent of $t$, from $(3.3)$ we deduce that $w = u_+$ (and $w = -u_-$) solve

$$-\text{tr} \left( A D^2 w \right) + H_{pp}(Du)DV(x) \cdot Dw - \text{tr} \left( |D_qA \cdot Dw| D^2 u \right) + pw = 0$$

so $w$ does not have positive maximum inside. On $\partial \Omega$ we have (taking time derivative of the Neumann condition)

$$Dw \cdot H_{pp}(Du)\bar{v} = 0. \quad (3.7)$$

If $w$ has a maximum on $\partial \Omega$, the tangential derivative would vanish so that

$$Dw \cdot (H_{pp}(Du)\bar{v}) = Dw \cdot \bar{v}(H_{pp}(Du)\bar{v} \cdot \bar{v}).$$
which implies $Dw \cdot \hat{v} = 0$ by (3.7). With a typical perturbation argument (replacing $w$ with $w + \theta d(x)$ and letting $\theta \to 0$), we conclude that the maximum of $w$ cannot be attained on $\partial \Omega$. Hence

$$w \leq \sup_{\Sigma_0 \cup \Sigma_T} w_+$$

for $w = \pm u_t$, which yields (3.6).

Now we estimate the derivatives of $u$ in terms of the sup-norm of $u$.

**Theorem 3.4** Let $u \in C^3(\bar{Q})$ be a solution of (3.4). Assume that $H$ is a $C^3$ radial function which satisfies conditions (2.5) and (2.6), that $f \in C^2(0, \infty)$ and there exist $\beta, s_0 > 0$ such that

$$|f''(s)|s^2 \leq \beta \sqrt{(1 + f'(s)s)^3}$$

for all $s \geq s_0$. Then, there exists a constant $C$, independent of $\rho$ and $\delta$, such that

$$\|Du\|_\infty \leq C(1 + \|u\|_\infty).$$

Here the constant $C$ depends on $\varepsilon, \beta, \|f'(m_0)\|_{W^{1, \infty}}, \|f'(m_1)\|_{W^{1, \infty}}, \|V\|_{W^{2, \infty}}$ and on the functions $H, f$.

**Remark 3.5** We stress that, in the proof below, the radial form of $H$ is only used to simplify the treatment of the boundary condition $H_p(Du) \cdot \hat{v} = 0$. However, the same result of Theorem 3.4 holds true for general $H$ (not radial) if we assume that $\Omega$ is a strictly convex domain, in the sense that there exists $\gamma > 0$ such that, in a $\delta_0$-neighborhood of the boundary, we have

$$D^2d(x) \leq -\gamma I_d \quad \forall x \in \Omega : d(x) \leq \delta_0$$

(3.9)

where $d(x) = \text{dist}(x, \partial \Omega)$. This reinforced convexity would be enough to handle the nonlinear condition at $\partial \Omega$.

**Remark 3.6** Assumption (3.8) is a technical condition, which however is satisfied by most natural examples, such as powers and exponentials. Functions which do not satisfy (3.8) can be highly oscillating such as, for instance, if $f'(s) = s(1 + \sin(s))$, with $\gamma > 1$.

Let us stress, however, that the growth of $f$ becomes irrelevant once we obtain an $L^\infty$ bound on $m$ prior to the gradient estimate of $u$. This is what we will exploit later, with different arguments, in Lemma 3.7 (using the convexity of the potential $V$) or in Proposition 5.2 (for quadratic growth Hamiltonians).

At this stage, we decided to keep Theorem 3.4 independent of any a priori estimate of $m$, at the expense of requiring the additional condition (3.8).

**Proof.**  Step 1 (tuning the $u$): we replace $u$ with

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

On account of (3.3), the function $v$ is therefore a solution to the following problem

$$
\begin{align*}
-\text{tr} \left( A(Du) D^2v \right) + DV(x) \cdot H_p(Dv) + \rho v &= \rho(M - C_0\frac{(T - t)}{T}) \quad \text{in } Q, \\
-v_t + H(Dv) - V(x) &= \delta u + f'(m_1) - C_0/T \quad \text{at } t = T, \ x \in \Omega, \\
-v_t + H(Dv) - V(x) + \delta u &= f'(m_0) - C_0/T \quad \text{at } t = 0, \ x \in \Omega, \\
H_p(Dv) \cdot \hat{v} &= 0 \quad \text{on } (0, T) \times \partial \Omega
\end{align*}
$$

(3.10)

where $f'(s) = f(s) + \varepsilon \log s$. We observe that

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

where, here and below, we denote by $C$ any (possibly different) constant independent of $\rho$ and $\delta$. We also notice that $v$ satisfies

$$v(T) \geq 1 + \|u\|_\infty \geq 1, \quad v(0) \leq -(1 + \|u\|_\infty) \leq -1.$$  

(3.11)

Now we define the function

$$z := H(Dv) + \frac{\lambda}{2}v^2$$

where

$$\lambda = \frac{\sigma}{1 + \|u\|_\infty^2}$$

(3.12)
for some (small) constant \( \sigma \) to be chosen later. The goal now is to estimate the function \( z \) through the maximum principle.

Step 2 \textit{(the boundary behavior)}. Suppose that \( z \) attains a maximum at the boundary \( t = T \), in which case we have \( Dz = 0 \) and \( z_t \geq 0 \) on the maximum point. By using the boundary condition for \( v \), we compute

\[
z_t = H_p(Dv) \cdot Dv + \lambda v v_t = H_p(Dv) \cdot D(Dv) - H_p(Dv) \cdot DV(x) - H_p(Dv) \cdot Df^z(m_1) - \delta H_p(Dv) \cdot Dv + \lambda v (C_0/T + H(Dv) - \delta u - f^z(m_1) - V) = H_p(Dv) \cdot Dz - \delta H_p(Dv) \cdot Dv - H_p(Dv) \cdot DV(x) - H_p(Dv) \cdot Df^z(m_1) - \lambda v (H_p(Dv) - H(Dv) - K)
\]

hence

\[
z_t = H_p(Dv) \cdot Dz + \delta H_p(Dv) \cdot Dv \leq (\|DV\|_\infty + \|Df^z(m_1)\|_\infty) |H_p(Dv)| - \lambda v (H_p(Dv) - H(Dv) - K)
\]

where \( K = K(\|f^z(m_1)\|_\infty, \|v\|_\infty, \|\mu\|_\infty, T) \).

Due to (3.5), we can suppose that \( H_p(Dv) - H(Dv) > 2K \), otherwise we have \( \gamma_0|Dv|^2 \leq \gamma_1 + 2K \), hence \( \max z \leq C K + \lambda |u|^2 \) and we are done. Therefore, using \( v(T) \geq 1 + \|u\|_\infty \), (3.12) and (3.5), and since \( |H_p(Dv)| \leq |Dv|^q-1 \) by (3.3), we estimate:

\[
z_t = H_p(Dv) \cdot Dz \leq (\|DV\|_\infty + \|Df^z(m_1)\|_\infty) |H_p(Dv)| - C \lambda (1 + \|u\|_\infty) [H_p(Dv) - H(Dv)] \leq C (\|DV\|_\infty + |Df^z(m_1)|_\infty) |Dv|^q-1 - \sigma \frac{C}{(1 + \|u\|_\infty)} |Dv|^q
\]

and we conclude that

\[
z_t = H_p(Dv) \cdot Dz < 0 \quad \text{if} \quad |Dv| > \frac{\sigma}{C} \left( \|DV\|_\infty + |Df^z(m_1)|_\infty \right) (1 + \|u\|_\infty).
\]

This implies that \( z \) cannot have a maximum at \( t = T \) unless \( |Dv| \leq \frac{K}{C} (1 + \|u\|_\infty) \) for some constant \( K > 0 \) independent of \( \rho, \delta \). Notice that the above argument only needs \( H_p(Dv) \cdot Dz = 0 \) on the maximum points, which holds even on \( \partial \Omega \) because \( H_p(Dv) \) is tangential.

Similarly we reason for \( t = 0 \) using that \( v(0) \leq -1 - \|u\|_\infty \leq 1 \). Thus, we conclude that \( z \) cannot have maximum at \( t = 0, T \) unless \( |Dv| \) is uniformly bounded.

Finally, we look at \((0, T) \times \partial \Omega\), where we use the Neumann condition for \( v \). Here we have

\[
Dz \cdot Dd(x) = H_p(Dv) \cdot [D^2vvDd(x)] + \lambda v Dv \cdot Dd(x) = H_p(Dv) \cdot D(Dv \cdot Dd(x)) - D^2d(x) H_p(Dv) \cdot Dv
\]

where we used that \( H_p(Dv) \cdot \vec{v} = 0 \) implies \( Dv \cdot Dd(x) = 0 \) on \( \partial \Omega \) (because \( H \) is radial). This also implies that \( D(Dv \cdot Dd(x)) \) is parallel to the normal direction, so it is orthogonal to \( H_p(Dv) \) using again the Neumann condition. Hence \( H_p(Dv) \cdot D(Dv \cdot Dd(x)) = 0 \). Finally, using that \( H_p(Dv) \) is parallel to \( Dv \) and \( \Omega \) is convex, we conclude that

\[
Dz \cdot Dd(x) = -D^2d(x) H_p(Dv) \cdot Dv \geq 0.
\]

This implies (up to the usual perturbation argument which consists in replacing \( u \) with \( u + \theta d(x) \) and letting \( \theta \to 0 \)) that no maximum point of \( z \) can occur on \( \partial \Omega \).

Step 3 \textit{(the interior estimate)}. We use (3.2) and we get

\[-\text{tr} \left( A(x, Du) D^2 \left[ \frac{v^2}{2} \right] \right) + DV(x) \cdot H_p(Dv) v + \rho v^2 = \rho \left( M - C_0 \frac{(T - t)}{T} \right) v - A(Du) Du \cdot Du.\]

Then we use (3.3) with \( \psi(v, Du) = H(Dv) \) and we get

\[-\text{tr} \left( A D^2 [H(Dv)] \right) + H_{pp}(Du)DV \cdot D^2v H_p(Dv) + D^2V(x) H_p(Dv) H_p(Dv) + \rho H_p(Dv) \cdot Dv = \text{tr} \left( [D_x A \cdot D[H(Dv)] + [D_x A \cdot H_p(Dv)]] D^2v \right) - \sum_{k,l=1}^d H_{pkpl}(Du) (A Du^k \cdot Du^l) \]

(3.13)
where $\mathcal{A}$ is computed on $(x, Du)$. Summing up the previous equalities and noticing that $D^2vH_p(Dv) = Dz - \lambda v Dv$, we deduce

$$- \operatorname{tr}(A D^2z) + H_{pp}DV \cdot Dz + \rho z + \lambda v (DV \cdot H_p - H_{pp}DV \cdot Dv) + D^2V(x)H_p \cdot H_p$$

$$+ \rho \left( H_p \cdot Dv - H + \lambda \frac{\epsilon^2}{2} \right) = \lambda \rho \left( M - C_0 \frac{(T - t)}{T} \right) v - \lambda ADv \cdot Dv$$

$$+ \operatorname{tr} \left( \left[ [D_\eta A : Dz] + [D_z A : H_p] \right] D^2v \right) - \lambda \sum_{i,j,\ell} v \partial_{\eta_i(a_{ij})} \partial_k v v_{ij} - \sum_{k,\ell=1} d H_{pp}\left( A Dv_k \cdot Dv_\ell \right)$$

where $H_p, H_{pp}$ are computed on $Dv$. The latter term in the left-hand side of the equality can be dropped by positivity. Hence we deduce

$$- \operatorname{tr}(A D^2z) + H_{pp}DV \cdot Dz + \rho z + \lambda v (DV \cdot H_p - H_{pp}DV \cdot Dv) + D^2V(x)H_p \cdot H_p$$

$$+ \lambda ADv \cdot Dv + \sum_{k,\ell=1} d H_{pp}\left( A Dv_k \cdot Dv_\ell \right) \leq |v| \rho M$$

$$+ \operatorname{tr} \left( \left[ [D_\eta A : Dz] + [D_z A : H_p] \right] D^2v \right) - \lambda \sum_{i,j,\ell} v \partial_{\eta_i(a_{ij})} \partial_k v v_{ij}. \quad (3.14)$$

We use the precise definition of $A$ and the coercivity of $H_{pp}$ from (2.5). Thus there exists $p_0, \gamma_H > 0$ such that, if $|Dv| > p_0$, we have

$$ADv \cdot Dv \geq | - v_t + H_p(Dv) \cdot Dv|^2 + \gamma_H (\epsilon + \chi(m)) |Dv|^q$$

$$\sum_{k,\ell=1} d H_{pp}\left( A Dv_k \cdot Dv_\ell \right) \geq \gamma_H |Dv|^q - 2(|Dv_t - D^2vH_p(Dv)|^2 + \gamma_H |Dv|^q - 2(\epsilon + \chi(m)) |D^2v|^2)$$

where we denoted $\chi(m) = f'(m)m$. Using these inequalities in (3.14) we get

$$- \operatorname{tr}(A D^2z) + H_{pp}DV \cdot Dz + \rho z + \lambda v (DV \cdot H_p - H_{pp}DV \cdot Dv) + D^2VH_p \cdot H_p$$

$$+ \lambda \left( | - v_t + H_p(Dv) \cdot Dv|^2 + \gamma_H (\epsilon + \chi(m)) |Dv|^q \right)$$

$$+ \gamma_H |Dv|^q - 2(Dv_t - D^2vH_p(Dv))^2 + \gamma_H |Dv|^q - 2(\epsilon + \chi(m)) |D^2v|^2 \quad (3.15)$$

$$\leq \lambda |v| \rho M + \operatorname{tr} \left( \left[ [D_\eta A : Dz] D^2v \right] \right) + \operatorname{tr} \left( \left[ [D_z A : H_p] D^2v \right] \right) - \lambda \sum_{i,j,\ell} v \partial_{\eta_i(a_{ij})} \partial_k v v_{ij}.$$

We now estimate last two terms. By definition of $A$, using (2.4) and that $(f(m) + \epsilon \log m) = (\chi(m) + \epsilon)m^{-1}$, we have

$$\sum_{i,j=1}^n \partial_{\eta_i(a_{ij})} \partial_k v v_{ij} = 2H_{pp}Dv[D^2vH_p(Dv) - Dv_t] + \sum_{i,j=1}^n \partial_{\eta_i((\epsilon + \chi(m))H_{pp})} \partial_k v v_{ij}$$

$$= 2H_{pp}Dv[D^2vH_p(Dv) - Dv_t] + \sum_{i,j=1}^n \partial_{\eta_i((\epsilon + \chi(m))H_{pp})} \partial_k v v_{ij}$$

$$+ \sum_{i,j=1}^n \partial_{\eta_i((\epsilon + \chi(m))H_{pp})} \partial_k v v_{ij}. \quad (3.16)$$

where $H_{pp}, H_{ppp}$ are computed on $Dv$. We observe that assumption (3.5) implies (and, for fixed $\epsilon$, is actually equivalent to) $m\chi'(m) \leq \epsilon(\epsilon + \chi(m)) \frac{\epsilon}{2}$ for some constant $c > 0$, and for $m > s_0 > 0$. Let us suppose by now that $m > s_0$ holds true, so that we can use this condition. Then, using also conditions (2.5)-(2.6) to handle
$H_{pp}$ and $H_{ppp}$, we can estimate the quantity in (3.16) multiplied by $\lambda v$ as follows:

$$\lambda \left| \sum_{ij} v \partial_{ij} (a_{ij}) \partial_{ij} v \right| \leq \frac{\gamma H}{2} (\varepsilon + \chi(m)) |DV|^q + \frac{\lambda}{2} - v_t + H_p(Dv) \cdot Dv|^2$$

$$+ C \lambda v^2 \frac{|DV|^q - |DV|^q}{\varepsilon + \chi(m)} |D^2 v H_p(Dv) - Dv|^2 + \lambda v^2 C (\varepsilon + \chi(m)) |DV|^{2(q-2)} |D^2 v|^2$$

$$\leq \frac{\gamma H}{2} (\varepsilon + \chi(m)) |DV|^q + \frac{\lambda}{2} - v_t + H_p(Dv) \cdot Dv|^2$$

$$+ \sigma \frac{C}{\varepsilon} |DV|^{q - 2} |D^2 v H_p(Dv) - Dv|^2 + \sigma C (\varepsilon + \chi(m)) |DV|^{2(q-2)} |D^2 v|^2$$

where we used the choice of $\lambda$ (see (3.12)). Let us stress that, if $\sigma$ is sufficiently small, last two terms are absorbed in the left-hand side of (3.15). Similarly we estimate, using again (3.8),

$$\text{tr} \left( [D_v A \cdot H_p] D^2 v \right) = - \frac{m \chi(m)}{\chi(m) + \varepsilon} (DV \cdot H_p) \text{tr} \left( H_{pp} D^2 v \right)$$

$$\leq \frac{\gamma H}{2} (\varepsilon + \chi(m)) |DV|^{2(q-2)} |D^2 v|^2 + C \|DV\|_\infty^2 |DV|^{2(q-1)},$$

and

$$\lambda v (DV \cdot H_p - H_{pp} DV \cdot Dv) + D^2 V H_p \cdot H_p \geq -\sigma - C (\|DV\|_\infty^2 + \|D^2 V\|_\infty) |DV|^{2(q-1)}.$$ 

Putting all together, we choose $\sigma$ sufficiently small and we deduce from (3.15)

$$- \text{tr} \left( A D^2 z \right) + H_{pp} DV \cdot Dz + \rho z + \frac{\lambda}{2} \left\{ | - v_t + H_p(Dv) \cdot Dv|^2 + \frac{\gamma H}{2} (\varepsilon + \chi(m)) |DV|^q \right\}$$

$$+ \frac{\gamma H}{2} |DV|^{q - 2} |DV| - D^2 v H_p(Dv)|^2 + \frac{\gamma H}{4} (\varepsilon + \chi(m)) |DV|^{2(q-2)} |D^2 v|^2$$

$$\leq \lambda |v| \rho M + \text{tr} \left( [D\eta A \cdot Dz] D^2 v \right) + C \left( \|DV\|_\infty^2 + \|D^2 V\|_\infty \right) |DV|^{2(q-1)}.$$ 

Dropping some positive terms, and noticing that $\lambda |v| \rho M \leq \rho C$ (by choice of $\lambda, M$), we get

$$- \text{tr} \left( A D^2 z \right) + H_{pp} DV \cdot Dz + \rho z + \frac{\lambda}{2} \left\{ | - v_t + H_p(Dv) \cdot Dv|^2 + \frac{\gamma H}{2} (\varepsilon + \chi(m)) |DV|^q \right\}$$

$$\leq C \rho + \text{tr} \left( [D\eta A \cdot Dz] D^2 v \right) + C \left( \|DV\|_\infty^2 + \|D^2 V\|_\infty \right) |DV|^{2(q-1)}.$$ 

(3.17)

It would be finished if not for the term containing the drift $V(x)$. To handle this part, we need to further exploit the quantity $| - v_t + H_p(Dv) \cdot Dv|^2$. In fact, by Lemma 3.2 we know that $(u_t)_+$ attains its maximum at $\Sigma_0 \cup \Sigma_T$. Suppose it holds at $t = 0$. Then we have, for some point $x_0 \in \Omega$,

$$u_t(t, x) \leq u_t(0, x_0) \leq H(Du(0, x_0)) - f(m_0(x_0)) - V(x_0) + \delta \|u\|_\infty$$

$$\leq \max z + C (\|f^\varepsilon(m_0)\|_\infty + \|V\|_\infty)$$

where we used Lemma 3.2. Similarly we reason if the maximum occurs at $t = T$. Hence, at any interior maximum point of $z$, we have

$$H(Dv) - v_t = \max z - u_t - \frac{\lambda}{2} v^2 - \frac{C_0}{T} \geq -K$$

for some $K$ depending on $\|f^\varepsilon(m_0)\|_\infty, \|f^\varepsilon(m_1)\|_\infty, \|V\|_\infty$. Notice that this also yields $H(Du) - u_t - V(x) \geq -K - \|V\|_\infty$ at any interior maximum point, so that $m > \kappa > 0$ for some $\kappa > 0$ only depending on $\|f^\varepsilon(m_0)\|_\infty, \|f^\varepsilon(m_1)\|_\infty, \|V\|_\infty$. This justifies that, at the interior maximum point, we used condition (3.8) (where we can assume $s_0 < \kappa$ with no loss of generality). On account of (3.5) we get

$$H_p(Dv) \cdot Dv - v_t \geq \gamma_0 |DV|^q - \gamma_1 - K.$$ 

Therefore, we deduce from (3.17)

$$- \text{tr} \left( A D^2 z \right) + H_{pp} DV \cdot Dz - \text{tr} \left( [D\eta A \cdot Dz] D^2 v \right) + \rho z + \epsilon_0 |DV|^q$$

$$\leq \tilde{K} + C (\|DV\|_\infty^2 + \|D^2 V\|_\infty) |DV|^{2(q-1)},$$

13
for some $c_0 > 0$ and some $\tilde{K}$ depending on $\|f^\varepsilon(m_0)\|_\infty, \|f^\varepsilon(m_1)\|_\infty, \|V\|_\infty$. Recalling the choice of $\lambda$, we conclude that, on any interior maximum point of $z$, we have

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

for some $C$ depending on $\varepsilon, \|V\|_{W^{2,\infty}}, \|f^\varepsilon(m_0)\|_{W^{1,\infty}}, \|f^\varepsilon(m_1)\|_{W^{1,\infty}}, \|u\|_\infty$. Let $K$ be a constant and we postpone it to the Appendix.

Step 4 (Conclusion). On account of the previous steps, we estimate the maximum of $z$ and in turn this yields

$$|Dv|_{\infty} \leq C,$$

where $C = C(\varepsilon, \|V\|_{W^{2,\infty}}, \|f^\varepsilon(m_0)\|_{W^{1,\infty}}, \|f^\varepsilon(m_1)\|_{W^{1,\infty}}, \|u\|_\infty)$. This estimate, together with Lemma 3.2 imply a similar bound for $u_t$ at $t = 0, t = T$. Using Lemma 3.3, we conclude with a full estimate for $u_t$. Hence, the desired estimate follows for $Du$.

We now give a further application of Lemma 3.1 to estimate the density $m$ in the optimal transport problem. Indeed, the next lemma gives sufficient conditions under which $m$, defined from (2.7) as $\varphi(-u_t + H(Du) - V(x))$, is bounded above.

**Lemma 3.7** Let $u \in C^2(Q) \cap C^1(\overline{Q})$ be a solution of (3.3). Assume that $V \in W^{2,\infty}(\Omega)$ is convex, and that $f'(m)m$ is nondecreasing. Then, there exists a constant $C$ such that

$$-u_t + H(Du) \leq C,$$

where $C = C(\|m_0\|_\infty, \|m_1\|_\infty, \|V\|_\infty, \|u\|_\infty, H, f)$. 

**Proof.** We follow the computations of (3.3) with the convex function $\psi(s, \xi) = -s + H(\xi)$, but we specialize to the matrix $A$ given by (2.8). This means that $w := -u_t + H(Du)$ satisfies

$$-\text{tr}(AD^2w) + \rho(-u_t + H_p(Du)Du) + H_{pp}DV \cdot Dw \leq -D^2VH_p(Du) \cdot H_p(Du) + \text{tr}((A_{pp} \cdot Dw)D^2u) + \varepsilon + m f'(m)H_{pp}w_t u_{ij}.$$  

where $H_{pp}, H_{ppp}$ are always computed on $Du$. Since $m_t - H_p(Du) \cdot Dm = m \text{tr}(H_{pp}D^2u)$, using the convexity of $V$ and the assumption that $mf'(m)$ is nondecreasing we have

$$D^2VH_p(Du) \cdot H_p(Du) + (mf'(m))'(m_t - H_p(Du) \cdot Dm)\text{tr}(H_{pp}D^2u) \geq 0.$$  

Hence we deduce that $w$ satisfies

$$-\text{tr}(AD^2w) + \rho(-u_t + H_p(Du)Du) + H_{pp}DV \cdot Dw \leq \text{tr}((A_{pp} \cdot Dw)D^2u) + \varepsilon + m f'(m)H_{pp}w_t u_{ij}.$$  

Since (3.3) implies $-u_t + H_p(Du)Du \geq w - \gamma_1$, we deduce that

$$-\text{tr}(AD^2w) + \rho(w - \gamma_1) + H_{pp}DV \cdot Dw \leq \text{tr}((A_{pp} \cdot Dw)D^2u) + \varepsilon + m f'(m)H_{pp}w_t u_{ij}.$$  

This implies that $w \leq \gamma_1$ on any maximum point inside the domain. Using the bound at $t = 0$, $t = T$ we have

$$w \leq C(1 + \|m_0\|_\infty + \|m_1\|_\infty + \|V\|_\infty) + \|u\|_\infty.$$  

On the boundary $\partial \Omega$, we have (with same arguments as in the previous propositions) that $Dw \cdot \vec{v} \leq 0$ and no maximum point can occur. Finally, we conclude that $w$ is bounded above by a constant only depending on $\gamma_1, \|m_0\|_\infty, \|m_1\|_\infty, \|V\|_\infty, \|u\|_\infty$. 

Finally, the estimates of the present Section, and specifically the gradient bound of Theorem 3.4 lead to the existence of solutions to (3.3). The proof of the following result, based on a continuity method, is quite technical and we postpone it to the Appendix.

**Theorem 3.8** Let $\rho, \delta > 0$. Assume that $f \in C^2(0, \infty)$ satisfies (3.3), and that $H \in C^3(\mathbb{R}^d)$ satisfies (2.11) and the growth conditions (2.26) – (2.27) (for some $\xi > 1$ and $\omega > 0$). Let $V \in W^{2,\infty}(\Omega)$ and $m_0, m_1 \in W^{1,\infty}(\Omega)$ with $m_0, m_1 > 0$ in $\Omega$. Then there exists a solution $u \in C^{2,\alpha}(Q) \cap C^{1,\alpha}(\overline{Q})$ to problem (3.3) and we have

$$\|u\|_{C^{1,\alpha}(\overline{Q})} \leq K$$

for some $K = K(\varepsilon, \|V\|_{W^{2,\infty}}, \|f^\varepsilon(m_0)\|_{W^{1,\infty}}, \|f^\varepsilon(m_1)\|_{W^{1,\infty}}, \|u\|_\infty)$ (independent of $\rho, \delta$).
4 A penalized optimal transport problem

At this stage we wish to let the parameters \( \rho, \delta \) in (3.4) tend to zero: they were only needed in the existence result to control the sup norm of \( u \). First of all, by letting \( \rho \to 0 \), we find a smooth solution of a control/transport problem with penalized initial-terminal conditions, which may have an interest in its own.

**Theorem 4.1** Let \( f, H, V, m_0, m_1 \) satisfy the same conditions of Theorem 3.8. Then, for any \( \delta > 0 \) there exists \( u \in C^{2,\alpha}(Q) \cap C^{1,\alpha}(\overline{Q}) \) which is a solution to the problem

\[
\begin{aligned}
- \text{tr} (A(Du) D^2 u) + DV(x) \cdot H_p(Du) &= 0 \quad \text{in } Q, \\
-u_t + H(Du) &= \delta u + f^\varepsilon(m_1) + V(x) \quad \text{at } t = T, x \in \Omega, \\
-u_t + H(Du) + \delta u &= f^\varepsilon(m_0) + V(x) \quad \text{at } t = 0, x \in \Omega, \\
H_p(Du) \cdot \nu &= 0 \quad \text{on } (0, T) \times \partial \Omega,
\end{aligned}
\]

(4.1)

where \( A \) is given by (2.8).

In particular, if we set \( m := (f^\varepsilon)^{-1}(-u_t + H(Du) - V(x)) \), then \( u \in C^{2,\alpha}(Q) \cap C^{1,\alpha}(\overline{Q}) \) is a solution of (4.1) if and only if the couple \((u, m)\) is a solution of the penalized problem

\[
\begin{aligned}
- u_t + H(Du) &= f(m) + \varepsilon \log m + V(x) \quad \text{in } Q, \\
\text{div}(m H_p(Du)) &= 0 \quad \text{in } Q, \\
H_p(Du) \cdot \nu &= 0 \quad \text{on } (0, T) \times \partial \Omega, \\
u(0) &= \frac{f^\varepsilon(m_0) - f^\varepsilon(m(0))}{\delta} \quad \text{in } \Omega, \\
u(T) &= \frac{f^\varepsilon(m(T)) - f^\varepsilon(m_1)}{\delta} \quad \text{in } \Omega,
\end{aligned}
\]

(4.2)

and \( m \in C^{1,\alpha}(Q) \cap C^{0,\alpha}(\overline{Q}) \), \( m > 0 \) in \( \overline{Q} \).

**Proof.** Let \( u_{\rho, \delta} \) be the solution to (3.4) given by Theorem 3.8. As a consequence of Lemma 3.2 and Theorem 3.3 \( u_{\rho, \delta} \) is bounded in \( W^{1,\infty}(Q) \) uniformly with respect to \( \rho \). We deduce from Theorem 3.8 that \( u_{\rho, \delta} \) is also bounded in \( C^{1,\alpha}(Q) \) for some \( \alpha > 0 \). This also yields that \( u_{\rho, \delta} \) is bounded in \( C^{2,\alpha}(Q) \) by classical Schauder’s elliptic regularity. As \( \rho \to 0 \), we deduce the existence of a smooth solution \( u \) to (4.1). Defining \( m \) as in (2.10), we have that \( m \in C^{0,\alpha}(Q) \cap C^{1,\alpha}(Q) \), and in addition \( m > 0 \) because \(-u_t + H(Du)\) is bounded and \( \varphi \) maps compact sets into compact sets of \((0, \infty)\). The equivalence between the elliptic equation (4.1) and the system (4.2) is straightforward for smooth solutions and the nonlinear Neumann condition for \( u \) at \( t = 0, T \) turns into the initial-terminal conditions of system (4.2).

The next step will consist in the limit as \( \delta \to 0 \). Here the main issue is to normalize \( u \) in order to stabilize the sup norm. In the next result we show that a control on the \( L^\infty \)-norm of the density \( m \) yields a control on a natural normalization of \( u \).

**Lemma 4.2** Assume that \( m_0, m_1 \in L^\infty(\Omega) \), and that \((u, m)\) is a (smooth) solution of (2.1) or, alternatively, a solution of (4.2) with \( \delta > 0 \). If we set

\[
\hat{u} := u - \int_\Omega u(T) m_1 \, dx
\]

(4.3)

then \( \hat{u} \) satisfies the global estimate

\[
\|\hat{u}\|_\infty \leq C(\|m\|_\infty, \|\varepsilon \log(m_1)\|_\infty, \|\varepsilon \log(m_0)\|_\infty, \|V\|_{W^{1,\infty}}, T, H, \Omega)
\]

(4.4)

and the local estimate

\[
-\frac{C_0}{(T-t)^{\frac{1}{2}}} \leq \hat{u}(t, x) \leq \frac{C_1}{(T-t)^{\frac{1}{2}}} \quad \forall (t, x) \in Q
\]

(4.5)

where \( C_0, C_1 \) only depend on \( \sup(f^\varepsilon(m) + V) \).

**Proof.** We consider the product between \( u(t) \) and \( m(t) \); using (2.1) (or alternatively (4.2)) we have

\[
\int_\Omega u(T)m(T) \, dx - \int_\Omega u(0)m(0) \, dx = - \int_0^T \int_\Omega m[H_p(Du) \cdot Du - H(Du)] \, dx \, dt - \int_0^T \int_\Omega (f^\varepsilon(m) + V)m \, dx \, dt,
\]

which implies, by convexity of \( H \) and since \( f^\varepsilon(s) \) is bounded below,

\[
\int_\Omega u(T)m(T) \, dx - \int_\Omega u(0)m(0) \, dx \leq C.
\]
Here and below all constants are independent of $\delta$ and $\varepsilon$ ($\varepsilon \leq 1$). This implies
\[
\int_{\Omega} u(T)(m(T) - m_1)dx - \int_{\Omega} u(0)(m(0) - m_0)dx \leq C + \int_{\Omega} u(0)m_0dx - \int_{\Omega} u(T)m_1dx
\]
\[
= C + \int_{\Omega} \tilde{u}(0)m_0dx
\]
where we used the definition of $\tilde{u}$ and the fact that $m_0$ has unit mass. The left-hand side vanishes if $(u, m)$ is a solution of (2.1). Alternatively, if $(u, m)$ is a solution of (1.2), we have
\[
\frac{1}{\delta} \int_{\Omega} f^{\varepsilon}(m(T)) - f^{\varepsilon}(m_1))(m(T) - m_1)dx + \frac{1}{\delta} \int_{\Omega} f^{\varepsilon}(m(0)) - f^{\varepsilon}(m_0))(m(0) - m_0)dx \leq C + \int_{\Omega} \tilde{u}(0)m_0dx,
\]
and the left-hand side is nonnegative by monotonicity of $f^{\varepsilon}$. Therefore, for both problems (2.1) and (4.2) we get
\[
\int_{\Omega} \tilde{u}(0)m_0dx \geq -C.
\]
For any given measure $\tilde{m}$, let us now consider the $p$-Wasserstein geodesic connecting $m_1$ and $\tilde{m}$ in $[s, T]$, specifically we consider the solution $\mu$ to the continuity equation
\[
\begin{cases}
\mu_t - \text{div}(\mu v) = 0 & \text{in } (s, T) \times \Omega, \\
v \cdot \vec{n} = 0 & \text{in } (s, T) \times \partial \Omega, \\
\mu(T) = m_1, \mu(s) = \tilde{m} & \text{in } \Omega,
\end{cases}
\]
which satisfies
\[
\int_{s}^{T} \int_{\Omega} |v|^p d\mu = \frac{c_p}{(T-s)^{p-1}} W_p(m_1, \tilde{m})^p.
\]
Here we use this geodesic with $p = q/(q - 1)$, where $q$ is given by the growth of the Hamiltonian (2.5). Let us set $M := \sup_{Q}[f^{\varepsilon}(m) + V]$. Since $\tilde{u}$ satisfies the inequality $-\tilde{u}_t + H(D\tilde{u}) \leq M$, multiplying by $\mu$ and integrating we get
\[
\int_{\Omega} \tilde{u}(s)\tilde{m} dx \leq \int_{s}^{T} \int_{\Omega} v \cdot D\tilde{u} \mu dxdt - \int_{s}^{T} \int_{\Omega} H(D\tilde{u}) \mu dxdt + M(T-s)
\]
where we used that $\int_{\Omega} \tilde{u} m_1 dx = 0$. Using the coercivity of $H$ we deduce
\[
\int_{\Omega} \tilde{u}(s)\tilde{m} dx \leq C(M(T-s) + \frac{1}{(T-s)^{\frac{p}{p-1}}} W_p(m_1, \tilde{m})^{\frac{q}{p}}).
\]
Since this holds for any measure $\tilde{m}$, we deduce that there exists a constant $C_1$ (only depending on $M, T$), such that
\[
\tilde{u}(s, x) \leq \frac{C_1}{(T-s)^{\frac{p}{p-1}}}.
\]
Reasoning in a similar way, namely using the geodesic between $m_0$ and any measure $\tilde{m}$, and the bound (4.7), we conclude that there exists a constant $C_0$ such that
\[
\tilde{u}(s, x) \geq \frac{C_0}{s^{\frac{p}{p-1}}}.
\]
This concludes the proof of (4.3). Now we transform the previous two bounds into a global $L^\infty$-bound for $\tilde{u}$, by using the strict positivity of the marginals. To this purpose, we consider $v := \tilde{u} - A(t - T)^2 + \tilde{\varepsilon}d(x)$, for small $\tilde{\varepsilon}$. Then one can see from (4.1) that $v$ cannot have positive maximum inside $Q$, nor on $\partial \Omega$, up to choosing $\tilde{\varepsilon} = o(\delta)$. We claim that, if $m_1 > 0$ in $\Omega$, then $v$ cannot have maximum at $t = T$ either, provided $A$ is sufficiently large. Indeed, if $v$ has a maximum at $t = T$, then we must have
\[
0 \leq \tilde{u}_t(t, x) - A = H(D\tilde{u}) - \delta u - V(x) - f^{\varepsilon}(m_1) - A \leq \delta \|u\|_\infty + \|V\|_\infty - f^{\varepsilon}(\min m_1) - A
\]
and we get a contradiction if $A$ is large enough (we use Lemma 3.2 if $\tilde{\varepsilon} > 0$). Then we deduce that $v$ attains its maximum at $t = 0$; using (4.8) and letting $\tilde{\varepsilon}, \theta \rightarrow 0$, we deduce that
\[
\tilde{u} \leq K_0 \quad \forall(t, x) \in \bar{Q}
\]
for some $K_0$, independent of $\delta$. Similarly we reason with the estimate from below provided $m_0$ is strictly positive (so that $f^{\varepsilon}(m_0)$ is bounded below). This concludes with the $L^\infty$ bound (4.4). \qed

Thanks to Lemma 3.2, the solution $u_\delta$ of (1.1) will be controlled in sup-norm after the normalization (4.5) provided $m_\delta$ is bounded in $L^\infty(Q)$. Sufficient conditions for this latter bound were already given in Lemma 3.7. In the next Section, we see an alternative way of controlling the density through displacement convexity arguments. This is the reason why we postpone to Section 6 the final convergence $\delta \rightarrow 0$ in (1.2).
5 Displacement convexity and density estimates

One of the main advantages in the construction of smooth solutions is the possibility to use the so-called Eulerian approach to displacement convexity estimates (see [35, Chapter 5.4]). In the context of mean-field game systems, general displacement convexity inequalities were first addressed in [10] in the periodic case. In this Section we extend their result (and later developments in [2]) to the case of Neumann boundary conditions and to general Lipschitz potentials $V$, and then we use it to get local and global bounds for the density. We start with an adaptation of [16, Thm 1.1]. Let us stress that in this Section we don’t use the radial structure of the Hamiltonian.

**Proposition 5.1** Let $\Omega$ be a smooth convex bounded domain in $\mathbb{R}^d$. Let $u \in C^2(\overline{\Omega}), m \in C^1(\overline{\Omega})$ be classical solutions to the system

$$
\begin{align*}
-ut &+ H(Du) = f(m) + V(x) \quad \text{in } Q, \\
m_t - \text{div}(m H_p(Du)) = 0 \quad \text{in } Q, \\
H_p(Du) \cdot \nu & = 0 \quad \text{on } (0,T) \times \partial \Omega
\end{align*}
$$

where $f \in C^1(0,\infty)$, $V \in W^{1,\infty}(\Omega)$, $H \in C^3$.

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

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

Then we have

$$
\frac{d^2}{dt^2} \int_{\Omega} U(m(t)) \geq \int_{\Omega} \left( P'(m)m - P(m) + \frac{1}{d} P'(m) \right) \left[ \text{div}(H_p(Du)) \right]^2 \\
+ \int_{\Omega} P'(m)f'(m)(H_{pp}(Du)Dm \cdot Dm) + \int_{\Omega} P'(m)(H_{pp}(Du)Dm \cdot DV(x))
$$

**Proof.** We borrow most of the computations from [16, Thm 1.1], while we take care of two new ingredients, i.e. the term with $V(x)$ and the boundary conditions. For the reader’s convenience, we write all the required steps. First of all, using the continuity equation we have

$$
\frac{d}{dt} \int_{\Omega} U(m(t)) = \int_{\Omega} U'(m)m_t = \int_{\Omega} U'(m) \text{div}(m H_p(Du))
$$

where we used the Neumann condition (in the last step) and the definition of $P(r) = U'(r)r - U(r)$. Hence, using again the equation of $m$ and integrating by parts,

$$
\frac{d^2}{dt^2} \int_{\Omega} U(m(t)) = \int_{\Omega} P'(m) [m(\text{div}(H_p(Du))) + Dm \cdot H_p(Du)] \text{div}(H_p(Du)) + \int_{\Omega} P(m) \text{div}(H_{pp}Du_t)
$$

$$
= \int_{\Omega} P'(m)m(\text{div}(H_p(Du)))^2 - \int_{\Omega} P(m) \text{div}(H_p(Du)) \text{div}(H_p(Du)) + \int_{\Omega} P(m) \text{div}(H_{pp}Du_t)
$$

$$
= \int_{\Omega} (P'(m)m - P(m)) \text{div}(H_p(Du))^2 - \int_{\Omega} P(m) H_p(Du) \cdot D(\text{div}(H_p(Du))) + \int_{\Omega} P(m) \text{div}(H_{pp}Du_t)
$$

where $H_{pp}$ is computed on $Du$. Now we use the equation of $u$ in the last term:

$$
\int_{\Omega} P(m) \text{div}(H_{pp}Du_t) = \int_{\Omega} P(m) \text{div}(H_{pp}D[u_t - H(Du)]) + \int_{\Omega} P(m) \text{div}(H_{pp}DH(Du))
$$

$$
= \int_{\partial \Omega} P(m)D[u_t - H(Du)] \cdot H_{pp} \nu - \int_{\Omega} P'(m)H_{pp}Dm \cdot D[u_t - H(Du)]
$$

$$
+ \int_{\Omega} P(m) \text{div}(H_{pp}DH(Du))
$$

$$
= \int_{\partial \Omega} P(m)D[u_t - H(Du)] \cdot H_{pp} \nu + \int_{\Omega} P'(m)H_{pp}Dm \cdot D[f(m) + V(x)]
$$

$$
+ \int_{\Omega} P(m) \text{div}(H_{pp}DH(Du)).
$$
We treat the boundary term writing \( \tilde{\nu} = -Dd(x) \), and using the convexity of \( \Omega \) (which implies that \( d(x) \) is concave). We have

\[
DH(Du) \cdot H_{pp}(Du) Dd = H_p(Du) \cdot D^2 u H_{pp}(Du) Dd
\]

\[
= H_p(Du) \cdot D(H_p(Du) \cdot Dd) - D^2 d H_p(Du) H_p(Du)
\]

\[
= -D^2 d H_p(Du) H_p(Du) \geq 0
\]

where we used that \( H_p(Du) \) is tangential and \( H_p(Du) \cdot Dd \) has zero tangential gradient. Since we also have

\[
Du_t \cdot H_{pp}(Du) Dd = \partial_t (H_p(Du) \cdot Dd) = 0
\]

we conclude that

\[
\int_{\partial \Omega} P(m) D[u_t - H(Du)] \cdot H_{pp} \tilde{\nu} \geq 0.
\]

Using this information in (5.4) and inserting that in (5.3) we get

\[
\frac{d^2}{dt^2} \int_{\Omega} U(m(t)) \geq \int_{\Omega} (P'(m)m - P(m))(\text{div}(H_p(Du))^2) + \int_{\Omega} P'(m)H_{pp} Dm \cdot D[f(m) + V(x)]
\]

\[
+ \int_{\Omega} P(m) \text{div}(H_{pp} DH(Du)) - \int_{\Omega} P(m)H_p(Du) \cdot D[\text{div}(H_p(Du))].
\]

We develop jointly last two terms observing

\[
\text{div}(H_{pp} DH(Du)) - H_p(Du) \cdot D[\text{div}(H_p(Du))] = H_{pp,p,pp,u,\ell} = \text{Tr}([H_{pp} D^2 u]^2)
\]

Using that \( \text{Tr}([AB]^2) \geq \frac{1}{4}\text{Tr}(AB)^2 \) for \( A, B \) symmetric with \( A \geq 0 \) (see e.g. [16, Lemma A.1]) we conclude that

\[
\text{div}(H_{pp} DH(Du)) - H_p(Du) \cdot D[\text{div}(H_p(Du))] \geq \frac{1}{d}(\text{div}(H_p(Du))^2)
\]

and then from (5.5) we get

\[
\frac{d^2}{dt^2} \int_{\Omega} U(m(t)) \geq \int_{\Omega} \left( P'(m)m - \left( 1 - \frac{1}{d} \right) P(m) \right) (\text{div}(H_p(Du))^2) + \int_{\Omega} P'(m)H_{pp} Dm \cdot D[f(m) + V(x)]
\]

which is (5.2).

We now deduce a priori estimates from the displacement convexity inequality (5.2). To this purpose, we require the Hamiltonian \( H \) to be of quadratic type, namely satisfying condition (2.3). While the global (in time) estimates are the typical application of displacement convexity (they are obtained e.g. in [16, 2]), the local estimates are more interesting, since they show a regularizing effect due to the congestion term in the cost functional of optimal transport. Similar local estimates were obtained in [22] for quadratic Hamiltonian with a very tricky use of flow-interchange techniques together with the variational interpretation of the functional as geodesic in Wasserstein space. Our next Proposition provides an alternative, different proof of the results in [22] in a slightly broader setting.

**Proposition 5.2** Assume that \( V \in W^{1,\infty}(\Omega) \) and \( f \in C^1(0, \infty) \) satisfies

\[
\exists c_0, r_0 > 0 : \quad f'(r) \geq c_0 r^{-1} \quad \forall r \geq r_0.
\]

Let \((u,m)\) satisfy (5.1), where \( H \) is a \( C^2 \) function satisfying (2.3). Then, for every \( 1 \leq p \leq \infty \) we have the global estimate

\[
\|m(t)\|_p \leq K_0 (\|m_0\|_p + \|m_1\|_p + 1) \quad \forall t \in [0, T]
\]

and the local estimate

\[
\|m(t)\|_p \leq K_1 (t^{-q} + (T-t)^{-q}) \quad \forall t \in (0, T)
\]

for some constants \( K_0, K_1 \) depending on \( c_0, r_0, p, d, T, \Omega, \|DV\|_\infty, \alpha_H, \beta_H, \) and for some \( q > 0 \).

**Proof.** The proof is done in two steps. We first establish global and local estimates in \( L^p \)-norm, for \( p < \infty \), and then we upgrade the estimates to the sup-norm.

**Step 1.** We use (5.2) with \( U(r) = (r - r_0)^p \), \( p \geq 2 \), where \( r_0 \) is given by (5.6). We notice that \( U \in C^1 \cap W^{2,\infty} \) and we have \( P(r) \geq 0 \) and \( P'(r) = U''(r) = p(p-1)(r-r_0)^{p-2} r^2 \mathbb{1}_{(r > r_0)} \). Moreover, we have \( P'(r) r - P(r) \geq 0 \). We deduce that

\[
\frac{d^2}{dt^2} \int_{\Omega} (m(t) - r_0)_+^p \geq p(p-1) \left\{ \int_{\Omega} (m - r_0)_+^{p-2} \mathbb{1}_{(m > r_0)} \left( f'(m) m (H_{pp} Dm \cdot Dm) + m (H_{pp} Dm \cdot DV(x)) \right) \right\}
\]

\[
\geq p(p-1) \left\{ c_0 \frac{\alpha_H}{2} \int_{\Omega} (m - r_0)_+^{p-2} \mathbb{1}_{(m > r_0)} \|Dm\|^2 - c_H \|DV\|_\infty \int_{\Omega} (m - r_0)_+^{p} + r_0^p \mathbb{1}_{\Omega} \right\}
\]
where we used assumption (5.6) and that $\alpha_H Id \leq H_{pp} \leq \beta_H Id$. Setting $\mu := (m - r_0)_+$, we rephrase the above inequality as

$$
\frac{d^2}{dt^2} \int_{\Omega} \mu(t)^p \geq p(p-1) \left\{ c_1 \int_{\Omega} \mu^{p-2} |D\mu|^2 - c_2 \|DV\|_\infty^2 \left( \int_{\Omega} \mu^p + 1 \right) \right\}.
$$

(5.9)

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

$$
\frac{d^2}{dt^2} \int_{\Omega} \mu^{p-2} |D\mu|^2 = \int_{\Omega} |D\mu|^2 \geq C_S \left( \int_{\Omega} \mu^p - (|\Omega|^{-1} \int_{\Omega} \mu)^2 \right)^{\frac{p}{2}}
$$

$$
\geq c_3 \left( \int_{\Omega} \mu^{\frac{p}{2}} \right)^{\frac{2}{p-1}} - c_4 \left( |\Omega|^{-1} \int_{\Omega} \mu \right)^2
$$

$$
\geq c_3 \left( \int_{\Omega} \mu^{\frac{p}{2}} \right)^{\frac{2}{p-1}} - c_4 |\Omega|^{-1} \int_{\Omega} \mu^p
$$

while using the interpolation inequality for $1 < p < \frac{2^*}{2}$ and $\|m(t)\|_{L^1(\Omega)} = 1$, we have

$$
\int_{\Omega} \mu^p \leq \left( \int_{\Omega} \mu^{\frac{p}{2}} \right)^{\frac{2}{p-1}}.
$$

Using the exact value of $2^*$ we conclude that

$$
\frac{d^2}{dt^2} \int_{\Omega} \mu^{p-2} |D\mu|^2 \geq c_3 \left( \int_{\Omega} \mu^p \right)^{\frac{1 + \frac{2}{p-1}}{2}} - c_4 |\Omega|^{-1} \int_{\Omega} \mu^p.
$$

From (5.9) we deduce (using e.g. $p(p-1) \geq \frac{p^2}{2}$ for $p \geq 2$)

$$
\frac{d^2}{dt^2} \int_{\Omega} \mu(t)^p \geq c_5 \left( \int_{\Omega} \mu(t)^p \right)^{\frac{1 + \frac{2}{p-1}}{2}} - c_6 (1 + p^2 \|DV\|_\infty^2) \int_{\Omega} \mu(t)^p - c_7 p^2 \|DV\|_\infty^2
$$

which implies

$$
- \varphi'' + c_5 \varphi^{1 + \frac{2}{p-1}} - c_8 \varphi \leq c_7 p^2 \|DV\|_\infty^2
$$

(5.10)

for $c_8 = c_6 (1 + p^2 \|DV\|_\infty^2)$ and $\varphi(t) = \int_{\Omega} \mu(t)^p$. Now we apply the maximum principle to $\varphi$ and we get

$$
\varphi(t) \leq \max \{ \varphi(0), \varphi(T), C_p \}
$$

for some constant $C_p$, depending on $c_0, r_0, p, d, \|DV\|_\infty, \alpha_H, \beta_H$. This yields the global estimate in $L^p$-norm

$$
\|m(t)\|_p \leq \|m_0\|_p + \|m_1\|_p + C(c_p \|DV\|_\infty + 1).
$$

(5.11)

But we also deduce a local estimate from (5.10), due to the superlinear term. In fact, the function

$$
\tilde{\varphi}(t) := L (t^{-\alpha} + (T - t)^{-\alpha}), \quad \alpha := d(p-1)
$$

is a supersolution of (5.10) for $L$ sufficiently large (possibly depending on $p$). By comparison between $\varphi$ and $\tilde{\varphi}$ we deduce the local estimate:

$$
\|m(t)\|_p \leq K \left( t^{-\frac{\alpha}{p}} + (T - t)^{-\frac{\alpha}{p}} \right)
$$

(5.12)

for some $K$ depending on $p, d, \|DV\|_\infty, T, r_0, \alpha_H, \beta_H$.

Step 2. We point out that, if not for the presence of the field $V(x)$, we could have deduced the $L^\infty$ bound by letting $p \to \infty$ in the estimates of the $L^p$ norm from the precised form (5.11). Due to the term containing $V(x)$ (the linear perturbation term in (5.10)), the Moser-type iteration would be more involved and we exploit a different argument, which might have an interest in its own. Namely, we obtain $L^\infty$ bounds with the classical level set iteration type argument which dates back to E. De Giorgi and G. Stampacchia, see e.g. [34].

To this purpose, we use (5.2) with $U(r) = \frac{(r-k)^2}{2}$, where now $k > 0$ is a positive parameter. As before, using (5.6) (we can suppose that $k \geq r_0$) and (2.3), we obtain from (5.2)

$$
\frac{d^2}{dt^2} \int_{\Omega} (m(t) - k)_+^2 \geq \int_{\Omega} m' \int_{\Omega} m H_{pp} Dm \cdot Dm \mathbb{I}_{(m > k)} + \int_{\Omega} m \mathbb{I}_{(m > k)} H_{pp} Dm \cdot DV(x)
$$

$$
\geq c_0 \frac{\alpha_H}{2} \int_{\Omega} |D(m - k)|^2 - c_H \int_{\Omega} m^2 \|DV\|^2 \mathbb{I}_{(m > k)}.
$$

(5.13)
We choose $k \geq \max(\|m_0\|_\infty, \|m_1\|_\infty)$; then we observe that
\[
\begin{align*}
\frac{d}{dt} \left( \int_\Omega (m(t) - k)^2 \right)_{t=T} &\leq 0, \\
\frac{d}{dt} \left( \int_\Omega (m(t) - k)^2 \right)_{t=0} &\geq 0,
\end{align*}
\forall k \geq \max(\|m_0\|_\infty, \|m_1\|_\infty)
\]
In particular, if we integrate in time we get from (5.13)
\[
c_H \frac{\alpha H}{2} \int_0^T \int_\Omega |D(m-k)^2| \leq c_H \int_0^T \int_\Omega m^2 \text{div} V |2 \mathds{1}_{\{m>k\}}.
\]
(5.14)
Similarly, if we first integrate (5.13) in $(t,T)$, we have
\[
\frac{d}{dt} \int_\Omega (m(t) - k)^2 \leq c_H \int_\Omega m^2 \text{div} V |2 \mathds{1}_{\{m>k\}} \leq c_H \|Dv\|_\infty^2 \int_Q m^2 \mathds{1}_{\{m>k\}}
\]
which yields, since $k \geq \|m_0\|_\infty$
\[
\forall t > 0, \quad \int_\Omega (m(t) - k)^2 = \int_0^t \int_\Omega |m(s) - k|^2 ds \leq T c_H \|Dv\|_\infty^2 \int_Q m^2 \mathds{1}_{\{m>k\}}
\]
(5.15)
Using (5.14) and (5.15), and the fact that $m^2$ is estimated in any $L^p$-space (by Step 1), we readily get the $L^\infty$-bound; this is well-known (see e.g. [18]) but we detail the steps for the reader’s convenience. We define the level set
\[
A_k := \{(t,x) \in Q : m(t,x) > k\}
\]
and we obtain from (5.14) and (5.15)
\[
\int_Q |D(m-k)|^2 + \sup_{t \in (0,T)} \left( \int_\Omega (m(t) - k)^2 \right) \leq C (1 + T) \|Dv\|_\infty^2 \|m_L^2(Q)|A_k|^{1-\frac{2}{d}}.
\]
(5.16)
We recall the interpolation inequality (see e.g. [11] Proposition 3.1, Chapter 1)
\[
\|v\|_{L^p(Q)} \leq c \|v\|_{L^\infty((0,T);L^2(\Omega))}^\beta \|Dv\|_{L^2(Q)}^{2(\frac{2}{d} - 1)} \quad \forall v \in L^2((0,T);W^{1,2}(\Omega)) \cap L^\infty((0,T);L^2(\Omega^d))
\]
where $\beta = 2 \left( \frac{d-2}{d} \right)$
\[
(5.17)
\]
for any $v \in L^2((0,T);W^{1,2}(\Omega)) \cap L^\infty((0,T);L^2(\Omega^d))$ such that $\int_\Omega v(t) dx = 0$ a.e. in $(0,T)$. Applying this inequality to $v = (m-k)_+ - |\Omega|^{-1} \int_\Omega (m(t) - k)_+ dx$ and using (5.16) we get
\[
\int_0^T \int_\Omega \left( \int_\Omega (m(t) - k)^2 \right) \leq \sup_{t \in (0,T)} \left( \int_\Omega (m(t) - k)^2 \right)^{\frac{1}{2}} + c \int_0^T \left( \int_\Omega (m(t) - k)^2 \right)^{\frac{2(d-2)}{d+2}}
\]
\[
\leq c |A_k|^{\frac{1}{d}}(1+\frac{\beta}{d}) + c \int_0^T |A_k(t)|^{1+\frac{2}{d}} \int_\Omega (m(t) - k)^2
\]
where $A_k(t) = \{x \in \Omega : m(t,x) > k\}$ is the time-section of $A_k$. Using (5.11), we know that $\sup_t |A_k(t)|$ is small for sufficiently large $k$. Hence the last term can be absorbed into the left-hand side and we deduce that there exists $k_0 > 0$ such that
\[
\int_0^T \int_\Omega (m-k)^{\frac{2(d-2)}{d}} dxdt \leq C |A_k|^{(1-\frac{2}{d})(1+\frac{\beta}{d})}
\]
\forall k \geq k_0 > 0
\]
We deduce from this inequality that
\[
\forall h > k \geq k_0 \quad |A_k|(h-k)^{\frac{2(d-2)}{d}} \leq C |A_k|^{\beta} \quad \beta := (1-\frac{2}{p})(1+\frac{2}{d})
\]
Choosing $p$ sufficiently large, we have $\beta = (1-\frac{2}{p})(1+\frac{2}{d}) > 1$; in that case, by a classical iteration lemma (see [34] Lemma 4.1, or similar arguments in [18] Chapter 2) we have $|A_k| = 0$ for some $k$ only depending on $k_0$ and $C$. This means that
\[
\|m\|_\infty \leq K
\]
for some $K$ depending on $\|m_0\|_\infty$, $\|m_1\|_\infty$, $\|Dv\|_\infty$, $T$, $c_0$, $r_0$, $H$. 

20
In a similar way we localize the previous estimate. To this goal, we fix $t_0 \in (0, T)$, and $R < R_0 := \min(t_0, T - t_0)$; then, for $\tau \in (0, R)$ let $\xi(t)$ be a smooth cut-off function such that

\[
\begin{align*}
\xi(t) &= 1 & \text{if } t \in (t_0 - \tau, t_0 + \tau) \\
\xi(t) &= 0 & \text{if } |t - t_0| \geq R \\
|\xi'(t)|^2 + |\xi''(t)| &\leq \frac{C}{(R - \tau)^2}
\end{align*}
\]

We denote

\[ A_{k, \tau} := \{(t, x) \in (t_0 - \tau, t_0 + \tau) \times \Omega : m(t, x) > k\}. \]

Then we have from (5.13)

\[
\frac{d^2}{dt^2} \left( \xi^2 \int_\Omega (m(t) - k_+^2) \right) \geq C_0 \frac{\alpha_H}{2} \int_\Omega \xi^2 |D(m - k_+)|^2 - c_H \|DV\|^2 \int_\Omega \xi^2 \mathbb{1}_{\{m > k\}} \\
- \frac{c}{(R - \tau)^2} \xi \int_\Omega (m(t) - k_+^2) + 4 \xi' \int_\Omega (m(t) - k_+^2).
\]  

(5.18)

Integrating in $(t, T)$ we get

\[
\frac{d}{dt} \left( \xi^2 \int_\Omega (m(t) - k_+^2) \right) \leq c_1 \int_{A_{k, T}} m^2 + \frac{c_2}{(R - \tau)^2} \int_{A_{k, T}} (m - k_+^2) + 4 \xi' \int_\Omega (m(t) - k_+^2),
\]

and one more integration yields

\[
\xi^2 \int_\Omega (m(t) - k_+^2) \leq c_1 T \int_{A_{k, T}} m^2 + \frac{c_2 T}{(R - \tau)^2} \int_{A_{k, T}} (m - k_+^2) + \int_Q 4 \xi' (m - k_+^2)
\]

\[
\leq c_1 T \int_{A_{k, T}} m^2 + \frac{c_3 T}{(R - \tau)^2} \int_{A_{k, T}} (m - k_+^2) + \int_Q 4 \xi' (m - k_+^2)
\]

\[
\leq \left( c_1 T + \frac{c_3 T}{(R - \tau)^2} \right) \|m\|^2_{L^p(A_{k, T})} |A_{k, R}|^{1 - \frac{2}{p}}.
\]

Similarly, integrating (5.18) in $(0, T)$, we estimate

\[
\int_Q \xi^2 |D(m - k_+)|^2 \leq c \int_{A_{k, T}} m^2 + \frac{cT}{(R - \tau)^2} \int_{A_{k, T}} (m - k_+^2)
\]

\[
\leq \left( c_1 T + \frac{c_3 T}{(R - \tau)^2} \right) \|m\|^2_{L^p(A_{k, T})} |A_{k, R}|^{1 - \frac{2}{p}}
\]

for $T \geq 1$. Hence, using as before the Gagliardo-Nirenberg inequality we get

\[
\int_Q (\xi(m - k_+)^{2(d+2)} \leq \left[ \sup_{t \in (0, T)} \left( \int_Q \xi^2 (m(t) - k_+^2) \right)^{\frac{d}{2}} \right] \left( \int_Q |D(m - k_+)|^2 \right)^{\frac{d+2}{2}}
\]

\[
\leq \left( c_1 T + \frac{c_3 T}{(R - \tau)^2} \right) \|m\|^2_{L^p(A_{k, T})} |A_{k, R}|^{1 - \frac{2}{p}}(1 + \frac{d}{2})
\]

\[
+ c \int_0^T \|A_{k}(t)\|^{1 + \frac{d}{2}} \int_\Omega (\xi(m(t) - k_+)^{2(d+2)}.
\]  

(5.19)

Notice that $\|m(t)\|^p$ can be estimated from (5.12), for any $p > 1$; indeed we have

\[
\text{for } t \in (t_0 - R, t_0 + R), \quad |A_{k}(t)| \leq k^{-p} \int_\Omega m(t)^p \leq C k^{-p} \left( (t_0 - R)^{-d(p-1)} + (T - t_0 - R)^{-d(p-1)} \right)
\]

hence, if $k$ is large, $|A_{k}(t)|$ is uniformly small for $t \in (t_0 - R, t_0 + R)$. Absorbing last term in (5.19) we deduce that

\[
\int_Q (\xi(m - k_+)^{2(d+2)} \leq C_R \left( c_1 T + \frac{c_3 T}{(R - \tau)^2} \right)^{1 + \frac{d}{2}} |A_{k, R}|^{(1 - \frac{2}{p})(1 + \frac{d}{2})}
\]

which implies

\[
|A_{k, \tau}|(h - k)^{2(d+2)} \leq C_R \left( \frac{c(R^2 + 1)T}{(R - \tau)^2} \right)^{1 + \frac{d}{2}} |A_{k, R}|^{(1 - \frac{2}{p})(1 + \frac{d}{2})}
\]

21
for every $h > k \geq k_0$ ($k_0$ possibly depending on $R$), and for every $\tau \in (0, R)$. As before, we can choose $p$ large so that $\beta := (1 - \frac{p}{2})(1 + \frac{2}{p}) > 1$ and we conclude with (a localized version of) the iteration lemma (see [34, Lemma 5.1]) that, for any $\sigma \in (0, 1)$, $m(t)$ is bounded for $t \in (t_0 - (1 - \sigma)R, t_0 + (1 - \sigma)R)$. In particular, we have
\[
\|m(t_0)\|_\infty \leq C(\max[t_0^{-1}, (T - t_0)^{-1}], \|DV\|_\infty, T).
\]

\[\square\]

Remark 5.3 Let us point out that the estimates of Proposition 5.2 apply to the solutions of the penalized problem (5.2) as well. Moreover, since the estimates only depend on $\|V\|_{W^{1,\infty}(\Omega)}$, on the uniform convexity (and upper bound) of the Hamiltonian and on the quite general condition (5.6), those bounds are inherited by weak solutions of general systems (2.11) which can be obtained in the limit as $\varepsilon \to 0$. See Definition 6.7 and Theorem 6.8 in the next Section.

6 Existence results

In this Section we collect all previous estimates and ingredients to deduce our main existence results.

6.1 Smooth solutions

Here we prove the existence of smooth solutions for the elliptic problem (2.10); equivalently, this yields the existence of smooth solutions to the general mean-field planning problem (2.1).

Theorem 6.1 Let $\Omega$ be a $C^3$, bounded, convex domain in $\mathbb{R}^d$ and let $V \in W^{2,\infty}(\Omega), m_0, m_1 \in W^{1,\infty}(\Omega)$ such that $m_0, m_1 > 0$ in $\Omega$. Let $H$ be a $C^3$ function satisfying (2.11) and $f \in C^2(0, \infty)$ be a nondecreasing function. Assume that at least one of the two following conditions is satisfied:

(i) $H$ satisfies conditions (2.6) - (2.14).
(ii) $H$ satisfies conditions (2.5) - (2.6) for some $q > 1$ and $\varpi > 0$, $r \mapsto f'(r)r$ is nondecreasing, and $V(x)$ is convex.

Then there exists $u \in C^{2,\alpha}(\overline{\Omega}) \cap C^{1,\alpha}(\Omega)$, $m \in C^{1,\alpha}(\overline{\Omega}) \cap C^{0,\alpha}(\Omega)$ such that $(u, m)$ is a smooth solution to the problem (2.1), and in addition $m > 0$ in $\Omega$. We also have that $u$ is a solution to the elliptic problem (2.10) (unique solution up to addition of a constant) and $m$ is the (unique) minimizer of the problem

\[
\min \int_0^T \int_\Omega L(v)dm + \int_0^T \int_\Omega m(\varepsilon|\text{log}(m)| - 1) + Vdxdt + \int_0^T \int_\Omega F(m)dxdt,
\]

where $F'(r) = f'(r)$ and $L$ is the Fenchel conjugate of $H$.

Proof. In a first step, we consider a suitable approximation $f_n$ of $f$ in a way that condition (3.8) is satisfied. If we are under condition (i), we simply take $f_n(r) = f(T_n(r))$, where $T_n(.)$ is a $C^2$ function such that $T_n(r) = r$ for $0 \leq r \leq n$, $T_n(r) \leq 2n$ for every $r$, and $|T_n'(r)r + T_n''(r)r^2| \leq n$. Then $|f_n''(r)r^2| \leq C_n$, and condition (3.8) is satisfied (for some $\beta$ depending on $n$). If we are in condition (ii), we wish to preserve the nondecreasing character of $f'(r)r$. To this purpose, given that $(f'(r)r)' \geq 0$, we may take

\[
f_n(r) := f(1) + f'(1) \log(r) + \int_1^r \frac{1}{s} \int_1^s [(f'(\tau)r)' \wedge n]d\tau.
\]

We observe that
\[
f_n'(r)r = f'(1) + \int_1^r [(f'(\tau)r)' \wedge n]d\tau
\]
and
\[
f_n''(r)r^2 = -f'(1) - \int_1^r [(f'(\tau)r)' \wedge n]d\tau + r[(f'(r)r)' \wedge n].
\]

A priori the minimizer can be considered in the class of $m, v \in L^\infty(Q)$, or even in the broader class of absolutely continuous curves $m \in C^0([0, T]; \mathcal{P}(\Omega))$, interpreting $v \in L^2(dt)$ as the metric derivative of $m$ (see [1]) and extending the functional to general measures in a classical way, see e.g. [22]. However, the a priori setting is a minor point here, because the minimizer $(m, H_p(Du))$ turns out to be smooth.
There is no loss of generality in assuming that \((f'(\tau))' \geq \delta_n > 0\) (otherwise replace \(f\) with \(f(\varphi + \frac{1}{\pi}r)\); then we have, for every \(r > 2\),

\[
r[(f'(\tau))' \wedge n] \leq n r \leq \frac{2n}{\delta_n} \int_1^T [(f'(\tau))' \wedge n] d\tau \leq \frac{2n}{\delta_n} f_n'(r) r.
\]

Hence (6.4) implies

\[
|f_n''(r)| r^2 \leq \left(1 + \frac{2n}{\delta_n}\right) f_n'(r) r \leq \left(1 + \frac{2n}{\delta_n}\right) (1 + f_n'(r)) r^2
\]

which means that \(f_n\) satisfies condition (6.5). At the same time, we have here that \(f_n'(r)\) is nondecreasing. Notice also that \(f_n\) is nondecreasing (from (6.3)) and satisfies

\[
f(1) + f'(1) \ln(r) \leq f_n(r) \leq f(r) \quad \forall r \in (0, \infty).
\]

Now we take the solution \((u_3, m_3)\) of problem (4.2) corresponding to \(f = f_n\), which is guaranteed by Theorem 4.1. We define \(\bar{u}_3\) as in Lemma 4.2. Now we show that there exists a constant \(M > 0\), independent of \(n\) and \(\delta\), such that \(\frac{1}{M} \leq m_3 \leq M\). First of all, we observe that under either conditions (i) or (ii), we have that \(m_3\) is bounded above (independently of \(n\)). Indeed, if (ii) holds true, this follows directly from Lemma 3.7 and the definition of \(m_3\); and since \(f_n \leq f\), the bound is independent of \(n\). If rather (i) holds true, we first observe that

\[
\|m_3(0)\|_{\infty} \leq (f_n^{-1}(\delta)\|u_3\|_{\infty} + f_n'_{\infty}(\|m_0\|_{\infty}))
\]

and then, due to Lemma 5.2 we have that \(\|m_3(0)\|_{\infty}\) is bounded. Similarly we have for \(\|m_3(T)\|_{\infty}\). Therefore, under condition (i), we deduce that \(m_3\) is uniformly bounded by Proposition 5.2. Notice again that, due to (6.5), the bound is independent of \(n\).

Thanks to the bound of \(m_3\) and since \(\log(m_0), \log(m_1)\) are Lipschitz continuous, we deduce from Lemma 4.2 that \(\bar{u}_3\) is uniformly bounded and solves

\[
\begin{cases}
-\operatorname{tr}(A(D\bar{u}_3)D^2\bar{u}_3) + DV(x) \cdot H(D\bar{u}_3) = 0 & \text{in } Q, \\
-(\bar{u}_3)_t + H(D\bar{u}_3) = \delta \bar{u}_3 + e_3 + f_n'(m_1) + V(x) & \text{at } t = T, x \in \Omega, \\
-(\bar{u}_3)_t + H(D\bar{u}_3) + \delta \bar{u}_3 + e_3 = f_n'(m_0) + V(x) & \text{at } t = 0, x \in \Omega,
\end{cases}
\]

where \(e_3 = \delta \int u_3(\tau)m_1 d\tau\) is a bounded sequence of real numbers. Now we apply Theorem 5.1 and we get that \(\bar{u}_3\) is bounded in Lipschitz norm by a constant which is independent of \(n\). In fact, if \(M\) is such that \(m_3 \leq M\), then condition (6.5) is only needed for \(r \leq M\); and clearly \(f_n\) satisfies (6.5) for \(1 \leq r \leq M\), with some constant \(\beta\) independent of \(n\). Thus, the Lipschitz bound of \(\bar{u}_3\) is independent of \(n, \delta\); by definition of \(m_3\), this means that there exists a constant \(M\), independent of \(n\), such that

\[
\frac{1}{M} \leq m_3 \leq M \quad \forall \delta > 0.
\]

Now, if \(f_n\) is defined by (6.2), since \((f'(\tau))'\) is continuous for \(\tau \in [\frac{1}{M}, M]\), for \(n\) sufficiently large we have \(f_n(m_3) = f(m_3)\). The same obviously holds if \(f_n = f(T_n(r))\). This means that, in both cases (i) or (ii), \((\bar{u}_3, m_3)\) are actually solutions of (4.2) with \(f\).

As a next step, from Theorem 5.3 we have that \(\bar{u}_3\) is actually bounded in \(C^{1,\alpha}(Q)\) for some \(\alpha > 0\). In particular, we have that \(m_3 = (f^c)^{-1}(\partial_x \bar{u}_3 + H(D\bar{u}_3) - V(x))\) is bounded in \(C^{0,\alpha}(Q)\) and, up to subsequences, \((\bar{u}_3, m_3)\) converge to some \((u, m)\) and this convergence is uniform up to \(t = 0, T\). In particular, we deduce from (1.6) and from the increasing character of \(f^c\) that \(m(0) = m_0\) and \(m(T) = m_1\). This also implies that \(e_3 \to 0\). Finally, the ellipticity implies that \(u\) is actually \(C^{2,\alpha}(Q)\), \(m \in C^{1,\alpha}(Q)\) and the system (2.1) is satisfied in a classical sense. Equivalently, \(u\) satisfies the quasilinear elliptic problem (2.10) in a classical sense. Notice that \(m > 0\) in \(Q\) because of the gradient bound on \(u\).

Finally, \(m\) is a minimizer of the variational problem (2.1); this is standard whenever \(m > 0\) and \((u, m)\) is a classical solution of (2.1). Indeed, system (2.1) represents the optimality conditions for the state-adjoint state of the optimization problem. In addition, setting \(w = mw\), it is well known (see [2, 3]) that (6.1) can be rephrased as a convex optimization problem in terms of the couple \((m, w)\); hence \(m\) is a minimizer. Due to the log term, here the convexity is strict and \(m\) is unique.

The case when the Hamiltonian may be degenerate, or singular, at \(p = 0\), e.g. in the model case \(H(p) = |p|^3\), corresponds to having \(\varpi = 0\) in assumption (2.5). In that situation, while the main gradient estimates remain true (see Theorem 5.1), the problem (2.10) lacks uniform ellipticity and \(u\) cannot be proved to be smooth. The natural framework in that context is to use the divergence structure of the operator, namely of the continuity equation for the density. This yields the following result.
Theorem 6.2 Let $\Omega$ be a $C^3$, bounded, convex domain in $\mathbb{R}^d$ and let $m_0, m_1 \in W^{1,\infty}(\Omega)$ be such that $m_0, m_1 > 0$ in $\Omega$. Assume that $H \in C^3(\mathbb{R}^d \setminus \{0\})$ satisfies (2.4), and that $f \in C^2(0, \infty)$ is such that $r \mapsto f'(r)r$ is nondecreasing in $[0, \infty)$. Let $V \in W^{2,\infty}(\Omega)$ be a convex function.

Then there exists $u \in W^{1,\infty}(Q)$, $m \in L^\infty(Q)$ such that $(u, m)$ is a solution to (2.1), where the continuity equation is taken in weak sense. Moreover, $u$ solves problem (2.11) in weak sense, i.e.

\[
\int_\Omega m_1 \xi_1(T) \, dx - \int_\Omega m_0 \xi(0) \, dx - \int_0^T \int_\Omega \varphi(-u_t + H(Du) - V(x)) [\xi_t - H_\nu(Du) \cdot D\xi] \, dx \, dt = 0 \quad \forall \xi \in C^1(\overline{Q}),
\]

where $\varphi(\cdot) = (f + \varepsilon \log)^{-1}(\cdot)$. We also have that $m > 0$ in $\overline{Q}$ and is the unique minimizer of (6.1), and $u$ is the unique (up to addition of a constant) weak solution of (6.1).

Proof. We approximate $H = h(|p|)$ with $H^\varepsilon = h(\sqrt{\frac{n}{\varepsilon} + |p|^{2\varepsilon}})$, so we can build a solution $(u^n, m^n)$ according to Theorem 6.1. Our assumptions allow us to apply Lemma 3.7 so $m^n$ is bounded in $Q$. From Lemma 4.2 we deduce that $\hat{u}^n$ (the normalization of $u^n$ as in (4.3)) is uniformly bounded. Then we use Theorem 3.4 (since $m^n$ is bounded, the assumption (3.8) is only needed in a compact set $[s_0, M]$, where it holds because $f$ is $C^2$) and we get that $\hat{u}^n$ is bounded in Lipschitz norm, and then it is relatively compact in $C(Q)$. We denote by $u \in W^{1,\infty}(Q)$ the uniform limit of (a subsequence of) $\hat{u}^n$. We observe that $\hat{u}^n$ is a sequence of solutions to divergence form problems

\[
-\partial_t (A_0(x, Du^n)) - \sum_{k=1}^d \frac{\partial}{\partial x_k} (A_k(x, Du^n)) = 0
\]

where $A_0 = -m^n = - (f^{\varepsilon})^{-1}(-\hat{u}^n + H^n(D\hat{u}^n) - V(x))$, $A_k = (f^{\varepsilon})^{-1}(-\hat{u}^n + H^n(D\hat{u}^n) - V(x))H^{\varepsilon}_{pk}(D\hat{u}^n)$. We notice that the uniform gradient bound on $u^n$ implies that $m^n$ is uniformly bounded from below and from above.

Now we show the convergence of $Du^n$ in $L^p(Q)$ for every $p < \infty$. To this purpose, we multiply the above equation by $(\hat{u}^n - u)$ and we get, using the divergence form structure,

\[
\int_0^T \int_\Omega \sum_{i=0}^d A_i(x, Du^n)D_i(\hat{u}^n - u) \, dx \, dt = \int_\Omega m_0(\hat{u}^n - u)(0) \, dx - \int_\Omega m_1(\hat{u}^n - u)(T) \, dx.
\]

The right-hand side goes to zero as $n \to \infty$ due to the uniform convergence of $\hat{u}^n$. In the left-hand side we use the definition of $m^n$ and we get

\[
-\int_0^T \int_\Omega m^n(\hat{u}^n - u)_t \, dx \, dt + \int_0^T \int_\Omega m^n H^{\varepsilon}_{pk}(D\hat{u}^n)D(\hat{u}^n - u) \, dx \, dt \overset{n \to \infty}{\to} 0.
\]

Recalling that $-\hat{u}^n = f^{\varepsilon}(m^n) + V - H^n(D\hat{u}^n)$, we obtain

\[
\int_0^T \int_\Omega m^n f^{\varepsilon}(m^n) \, dx \, dt = \int_0^T \int_\Omega m^n(-u_t + H(Du) - V) \, dx \, dt
\]

\[
+ \int_0^T \int_\Omega m^n \{H(D\hat{u}^n) - H(Du) - H_\nu(D\hat{u}^n)D(\hat{u}^n - u)\} \, dx \, dt + o(1)_n
\]

where $o(1)_n$ denotes some quantity which vanishes as $n \to \infty$. Now we observe that the last integral is nonpositive, due to the convexity of $H$. If we denote by $m$ the weak* limit of $m^n$ in $L^\infty(Q)$, we get

\[
\limsup_{n \to \infty} \int_0^T \int_\Omega m^n f^{\varepsilon}(m^n) \, dx \, dt \leq \int_0^T \int_\Omega m(-u_t + H(Du) - V) \, dx \, dt.
\]

But using again the convexity of $H$, the weak convergence of $D\hat{u}^n$ and denoting by $\bar{f}$ the weak* limit of $f^{\varepsilon}(m^n)$ in $L^\infty(Q)$, we have

\[
-\hat{u}^n_t + H^n(D\hat{u}^n) = f^{\varepsilon}(m^n) + V \Rightarrow -u_t + H(Du) \leq \bar{f} + V.
\]

Therefore (6.9) implies

\[
\limsup_{n \to \infty} \int_0^T \int_\Omega m^n f^{\varepsilon}(m^n) \, dx \, dt \leq \int_0^T \int_\Omega m \bar{f} \, dx \, dt
\]

which yields

\[
\limsup_{n \to \infty} \int_0^T \int_\Omega (m^n - m) (f^{\varepsilon}(m^n) - f^{\varepsilon}(m)) \, dx \, dt \leq 0.
\]
The strict monotonicity of \( f^\epsilon \) implies that \( m^n \to m \) in \( L^1(Q) \), and then in \( L^p(Q) \) for every \( p < \infty \). Hence we deduce from (6.8)

\[
\int_0^T \int_\Omega m^n(H_p(D\hat{u}^n) - H_p(Du))D(\hat{u}^n - u) \, dx \, dt = -\int_0^T \int_\Omega m^n H_p(Du) D(\hat{u}^n - u) \, dx \, dt
\]

\[
+ \int_0^T \int_\Omega m^n(\hat{u}^n - u) \, dx \, dt + o(1)_n .
\]

All terms in the right-hand side converge to zero. Then, using also the bound from below on \( m^n \), we obtain

\[
\int_0^T \int_\Omega (H_p(D\hat{u}^n) - H_p(Du))D(\hat{u}^n - u) \, dx \, dt \to 0 .
\]

The strict monotonicity of \( H_p(\cdot) \) implies that \( D\hat{u}^n \to Du \) in \( L^p(Q) \) for every \( p < \infty \). The identity \(-\hat{u}^n = f^\epsilon(m^n) + V - H^n(D\hat{u}^n)\) allows us to conclude that \( \hat{u}^n \to u \), strongly in \( L^p(Q) \) as well. Thus, we proved that \( D\hat{u}^n \to Du \) in \( L^p(Q) \) for every \( p < \infty \). Passing to the limit in the equation, we conclude that \( u \) is a weak solution of the limit problem in the sense of (6.6). Accordingly, \((u, m)\) satisfy (2.1), where \( u \) satisfies the first equation almost everywhere in \( Q \), and \( m \) is a bounded weak solution of the continuity equation. We notice that \( u \) is the unique Lipschitz solution (up to addition of a constant) of (6.6); indeed, (6.6) can be rephrased as

\[
\int_0^T \int_\Omega \sum_{i=0}^d A_i(x, Du) D_i \xi \, dx \, dt = \int_\Omega m_0 \xi(0) \, dx - \int_\Omega m_1 \xi(T) \, dx \quad \forall \xi \in C^1(\overline{Q})
\]

where \( A = (A_i)_{i=0, \ldots, d} \) is a strictly monotone vector field in \( \mathbb{R}^{d+1} \). Moreover, by a standard density argument, the weak formulation holds for all test functions \( \xi \in W^{1,\infty}(Q) \). Hence, if \( u_1, u_2 \) are two different solutions, using \( \xi = u_1 - u_2 \) in the weak formulation of both equations, we readily conclude that \( Du_1 = Du_2 \). So \( u_1 - u_2 \) differ by a constant. Finally, the uniqueness of \( m \) as a minimizer can be proved as in Theorem 6.1 by using the strict convexity of the associated functional.

\section{Transport with entropy in \( \mathbb{R}^d \)}

In this section we restrict to the case of optimal transportation without any further congestion term (or mean field interaction). This means that we set \( f = 0 \) and we focus on the very model problem

\[
\begin{align*}
- u_t + H(Du) &= \epsilon \log m + V(x) \quad \text{in } Q := (0, T) \times \Omega, \\
m_1 - \text{div}(m H_p(Du)) &= 0 \quad \text{in } Q, \\
H_p(Du) \cdot \nu &= 0 \quad \text{on } (0, T) \times \partial \Omega, \\
m(0, \cdot) &= m_0, \quad m(T, \cdot) = m_1 \quad \text{in } \Omega ,
\end{align*}
\]

(6.10)

For the reader’s convenience, we first collect the main result of the previous section in the specific case of (6.10).

\begin{theorem}
Let \( \Omega \) be a \( C^3 \), bounded, convex domain in \( \mathbb{R}^d \) and let \( V \in W^{2,\infty}(\Omega) \), \( m_0, m_1 \in W^{1,\infty}(\Omega) \) such that \( m_0, m_1 > 0 \) in \( \overline{\Omega} \). Let \( H \) be a \( C^3 \) function satisfying (2.11) and assume that at least one of the two following conditions is satisfied:

(i) \( H \) satisfies conditions (2.5) - (2.4)

(ii) \( H \) satisfies conditions (2.5) - (2.4) for some \( q > 1 \) and \( 0 < \varpi < \infty \), and \( V(x) \) is convex.

Then problem (6.10) admits a solution \((u, m)\) such that \( u \in C^{2,\alpha}(Q) \cap C^{1,\alpha}(\overline{Q}) \), \( m \in C^{1,\alpha}(Q) \cap C^{0,\alpha}(\overline{Q}) \), and this solution is unique (up to addition of a constant to \( u \) ). Moreover we have \( m(t) > 0 \) in \( \overline{\Omega} \) for every \( t \in [0, T] \). Finally, \( m \) is the unique minimizer of the problem

\[
\min \left\{ \int_0^T \int_\Omega L(v) \, dm + \epsilon \int_0^T \int_\Omega \log \left( \frac{dm}{d\varrho} \right) \, dm, \quad (m, v) : \begin{cases} m_t - \text{div}(mv) = 0, & \text{in } (0, T) \times \Omega, \\
v \cdot \nu = 0, & \text{on } (0, T) \times \partial \Omega, \\
m(0) = m_0, m(T) = m_1 \end{cases} \right\}
\]

where \( \varrho = e^{-\frac{V(x)}{\epsilon}} \) \( dx \) and \( L \) is the Fenchel conjugate of \( H \).
\end{theorem}
We now devote the remaining part of this subsection to a model result in the noncompact setting where $\Omega$ is replaced by the whole space $\mathbb{R}^d$. In this context, it is natural to consider the entropy of $m$ with respect to Gaussian-type measures.

Let us fix a reference measure $\nu := e^{-V}dx$, where $V$ is twice differentiable and satisfies

$$D^2V \xi \cdot \xi \geq \gamma |\xi|^2 \quad \gamma > 0.$$  \hfill (6.11)

In particular, this implies that $V$ is a convex, coercive function at infinity, whose model case is given by $V = \frac{|x|^2}{2}$. We consider the system

$$\begin{align*}
- u_t + H(Du) &= \varepsilon(\log m + V(x)) \quad \text{in } (0, T) \times \mathbb{R}^d, \\
m_t - \text{div}(m H_p(Du)) &= 0 \quad \text{in } (0, T) \times \mathbb{R}^d, \\
m(0, \cdot) &= m_0, m(T, \cdot) &= m_1 \quad \text{in } \mathbb{R}^d,
\end{align*}$$

where we assume that $m_0, m_1$ satisfy

$$m_0 e^{V}, m_1 e^{V} \in W^{1, \infty}(\mathbb{R}^d) \quad \text{and} \quad c_0 e^{-V(x)} \leq m_0 \leq C_0 e^{-V(x)}, \quad c_1 e^{-V(x)} \leq m_1 \leq C_1 e^{-V(x)} \quad (6.13)$$

for some positive constants $c_i, C_i > 0, i = 0, 1$. We can prove a similar result as above for the non compact case; for simplicity we restrict to Hamiltonians with quadratic growth.

**Theorem 6.4** Assume that $V \in W^{2, \infty}(\mathbb{R}^d)$ satisfies (6.11), and that $m_0, m_1$ satisfy (6.13). Let $H$ be a $C^3$ function satisfying (2.11) and (2.23)-(2.24).

Then problem (6.12) admits a solution $(u, m)$ such that $u \in C^2(\Omega), m \in C^1(\Omega)$, and this solution is unique (up to addition of a constant to $u$). Moreover we have that $m(t)e^{V(x)}$ is positive and bounded uniformly in $\mathbb{R}^d$ for every $t \in [0, T]$. Finally, $m$ is the unique minimizer of the problem

$$\begin{align*}
\min \int_0^T \int_{\Omega} m L(v) \, dx \, dt &+ \varepsilon \int_0^T \int_{\Omega} \log \left( \frac{dm}{d\nu} \right) \, dm, \\
\quad (m, v) : \quad &\begin{cases}
- m_t + \text{div}(vm) = 0 \quad \text{in } (0, T) \times \mathbb{R}^d, \\
m(0) = m_0, m(T) = m_1
\end{cases}
\end{align*}$$

where $\nu = e^{-V(x)}dx$.

**Proof.** We define $B_n$ the $d$-dimensional ball of radius $n$, and we let $(u_n, m_n)$ be the solution of the problem

$$\begin{align*}
- u_t + H(Du) &= \varepsilon(\log m + V(x)) \quad \text{in } (0, T) \times B_n, \\
m_t - \text{div}(m H_p(Du)) &= 0 \quad \text{in } (0, T) \times B_n, \\
H_p(Du) \cdot \nu &= 0 \quad \text{on } (0, T) \times \partial B_n, \\
m(0, \cdot) &= m_0, m(T, \cdot) &= m_1 \quad \text{in } B_n,
\end{align*}$$

which exists by Theorem 6.3 with $u_n, m_n$ smooth since $m_0, m_1$ are Lipschitz and positive in $B_n$. Since $V$ is convex, we can estimate $-u_t + H(Du)$ in terms of its maximum at $t = 0, T$ (see Lemma 3.7). We obtain that

$$m_n e^{V} \leq \sup_{i=0,1} \|m_i e^{V}\|_{\infty} \quad (6.14)$$

hence $m_n$ is bounded and uniformly decaying at infinity.

Now we estimate $Du$ with a variant of Theorem 5.4; in fact, we estimate $Du$ independently from the $L^\infty$-bound of $u$. Hereafter, we have $u = u_n$ (i.e. we avoid to write the index $n$). We consider $w = H(Du)$ and we recall (see (6.13)) that $w$ satisfies

$$- \text{tr}(A D^2 w) + H_{pp}DV \cdot Dw + D^2V(x)H_p(Du)H_p(Du) = \text{tr} ( [D_\eta A \cdot Dw ] D^2 u ) - \sum_{k,l=1}^d H_{p_k p_l} (A Du_k \cdot Du_l)$$

where, we recall, $H_{pp}$ is computed on $Du$. This implies, due to (6.11),

$$- \text{tr}(A D^2 w) + H_{pp} DV \cdot Dw - \text{tr} ( [D_\eta A \cdot Dw ] D^2 u ) + \gamma |H_p(Du)|^2 \leq 0.$$  \hfill (6.12)

Now we multiply $w$ by an auxiliary function to take care of boundary conditions. We set

$$z := e^{\sigma(t - \frac{\sigma^2}{2})} w.$$

...
Computing we find that \( z \) solves

\[
-\text{tr} \left( A(Du) D^2 z \right) - e^{\sigma(t - \frac{\gamma}{T})} \sigma w \left( \frac{2}{T} + (1 - 2 \frac{t}{T})^2 \sigma \right) + 2 \sigma (1 - 2 \frac{t}{T}) (z_t - H_p(Du) \cdot Dz) + H_p(DV \cdot Dz) - \text{tr} \left( [D_u A \cdot Dz] D^2 u \right) + e^{\sigma(t - \frac{\gamma}{T})} \gamma |H_p(Du)|^2 \leq 0
\]

where we used that \( A \) is independent of \( u_t \), so \( e^{\sigma(t - \frac{\gamma}{T})} D_u A \cdot Dz = D_u A \cdot Dz \). Using (3.30) (with \( q = 2 \)) we have \( |H_p|^2 \geq \gamma_0 H - \gamma_1 \) for some constants \( \gamma_0, \gamma_1 \). Hence, we deduce

\[
-\text{tr} \left( A(Du) D^2 z \right) + z \left[ \gamma \gamma_0 - \sigma \left( \frac{2}{T} + (1 - 2 \frac{t}{T})^2 \sigma \right) \right] + B \cdot Dz \leq \gamma \gamma_1 e^{\sigma(t - \frac{\gamma}{T})}
\]

for some vector field \( B \). In particular, for \( \sigma \) sufficiently small (only depending on \( T, \gamma_0, \gamma \)), we deduce that \( z \) is uniformly bounded on any internal maximum point. We can also exclude that \( z \) has a maximum on \( \partial B_n \), due to the Neumann condition and the convexity of \( B_n \). On the time-boundary, we have, using that 

\[
-z_t + H(Du) = \varepsilon \log(m_e e^V),
\]

\[
z_t = e^{\sigma(t - \frac{\gamma}{T})} \left( \sigma(1 - \frac{t}{T}) w + H_p \cdot Du_t \right) = e^{\sigma(t - \frac{\gamma}{T})} \left( \sigma(1 - \frac{t}{T}) w - \varepsilon H_p \cdot D \log(m_e e^V) \right) + H_p \cdot Dz \tag{6.15}
\]

for \( i = 0, 1 \). By assumption (6.13), we have that \( \log(m_e e^V) \in W^{1, \infty}(\mathbb{R}^d) \). Then it follows from (6.15) that \( z \) cannot have a maximum at \( t = 0 \) or \( t = T \) such that \( |Du| > K \), for some \( K \) depending on \( \sigma, H, \varepsilon \max_i \|D \log(m_e e^V)\|_{\infty} \). We deduce from the above steps the a priori estimate

\[
\|Du\|_{\infty} \leq C (1 + \varepsilon \max \left( \|D \log(m_0 e^V)\|_{\infty}, \|D \log(m_1 e^V)\|_{\infty} \right)),
\]

for some \( C \) only depending on \( \gamma, H, T \). This estimate, together with assumption (6.13), imply that \( |u_t| \) is bounded at \( t = 0, t = T \). Then, using Lemma 4.2 we also deduce a uniform bound for \( u_t \). Therefore we proved so far that

\[
\|Du_n\|_{\infty} \leq C, \quad \alpha_0 \leq m_n e^{V(x)} \leq \alpha_1 \quad \forall x \in B_n
\]

for some \( \alpha_0, \alpha_1 > 0 \). At this stage we can normalize \( u \) as in Lemma 4.2 by setting

\[
\tilde{u}_n := u_n - \int_{B_n} u_n(T)m_1 \, dx.
\]

We stress that the proof of Lemma 4.2 can be adapted here to give a bound on \( \tilde{u}_n \) which is uniform in \( B_n \). To this purpose one first needs to use \( 3.13 \) to obtain the estimate \( 3.7 \). Secondly, when using the Wasserstein geodesic in the proof of Lemma 4.2 it is enough to notice that the two marginals \( m_0, m_1 \) have finite moments (of any order), and one can still obtain estimate (4.3) using the geodesic between \( m_1 \) and measures \( m \) with finite second moments. Thus, following Lemma 4.2 we obtain that \( \tilde{u}_n \) is bounded in \( L^{\infty}(B_n) \).

We are only left with letting \( n \to \infty \). To this goal, we can use the local \( C^{1, \alpha} \) estimates (up to \( t = 0, t = T \), as in Proposition 7.2) to get compactness of \( Du_n \) and \( m_n \). With a diagonal process on a sequence of balls invading \( \mathbb{R}^d \), we build a solution \( u \) (and \( m \) in turn) of the problem in the whole space. \( \square \)

### 6.3 Weak solutions, extensions and further comments

We conclude with a few further remarks.

**A further estimate for optimal transport.** It is natural to ask what happens if we drop the strict positivity condition on the marginals \( m_0, m_1 \), together with letting the entropy term vanish, or even independently. Alternatively, one can wonder whether the positivity estimates for \( m \), or the gradient estimates for \( u \), can be localized in time. This seems to be an interesting question. In this direction, we only give the following estimate for the case of pure transport with entropy. It suggests, roughly speaking, that \( u \) should remain smooth at least on the support of \( m \), giving a precise quantitative estimate. Let us stress that this result only requires boundedness of the marginals \( m_0, m_1 \), and the estimate is independent of \( \varepsilon \).

**Theorem 6.5** Under the assumptions of Theorem 6.3 let \( (u, m) \) be the (smooth) solution of (6.10). Then there exists a constant \( \theta > 0 \) such that

\[
\theta H(Du(t)) + \varepsilon \log(m(t)) \leq K_t \quad \forall t \in (0, T) \tag{6.16}
\]

where \( K_t \) is a constant only depending on \( \|u\|_{\infty}, \|V\|_{\infty}, T \) and \( \min(t, T - t)^{-1} \).
Proof. By assumption \([\mathbb{P}_2]\), there exists \(\theta > 0\) such that
\[
H_{p}(p) \cdot p \geq (1 + 2\theta) H(p) - c_0 \quad \forall p \in \mathbb{R}^d.
\] (6.17)
We come back to the gradient estimate of Theorem 3.4, which we aim at localizing in time. To this purpose we set
\[
z := (1 + \theta) H(Du) - u_t + \frac{\nu^2}{2}, \quad \lambda = \frac{\sigma}{1 + \|u\|_{\infty}}
\]
where \(\sigma\) will be chosen (sufficiently small) as we did in Theorem 3.4. By using Lemma 3.1 we obtain (recall that \(H_{p}, H_{pp}\) are computed on \(Du\))
\[
- \text{tr} (A D^2 z) + H_{pp}DV \cdot Dz + \rho z + \lambda u (DV \cdot H_p - H_{pp}DV \cdot Du) + (1 + \theta) D^2 V(x) H_p \cdot H_p
\]
\[
+ \lambda ADu \cdot Du - \text{tr} ([D_{\eta}A \cdot Dz] D^2 u) = -\lambda \sum_{i,j,k} u \partial_{n}(a_{ij}) \partial_{k} u_{ij} - (1 + \theta) \sum_{k,l} H_{p_{k,l}} (ADu_{k} \cdot Du_{l})
\]
where we took advantage that, here, the matrix \(A\) does not depend on \((t,x)\) (and not either on \(u_{t}\), which will be used below). Using the coercivity of \(A\) and \(H_{pp}\), as we did in the proof of Theorem 3.4 we get that for \(|Du|\) sufficiently large
\[
- \text{tr} (A D^2 z) + H_{pp}DV \cdot Dz + \rho z - \text{tr} ([D_{\eta}A \cdot Dz] D^2 u)
\]
\[
+ \lambda u (DV \cdot H_p - H_{pp}DV \cdot Du) + (1 + \theta) D^2 V(x) H_p \cdot H_p + \lambda \left( |H_{p}(Du) \cdot Du - u_{t}|^2 + \varepsilon \gamma_H |Du|^q \right)
\]
\[
+ (1 + \theta) \gamma_H |Du|^q - 2 \left( |Du_t - D^2 u H_{p}(Du)|^2 + \varepsilon \gamma_H |Du|^q - 2 |D^2 u|^2 \right) \leq -\lambda u \sum_{i,j,k} \partial_{n}(a_{ij}) \partial_{k} u_{ij}.
\] (6.18)
We estimate last term in a similar way as in Theorem 3.3
\[
\lambda u \sum_{i,j,k} \partial_{n}(a_{ij}) \partial_{k} u_{ij} = \lambda u \left\{ 2 H_{p} Du[D^2 u H_{p}(Du) - Du_t] + \varepsilon \sum_{i,j,k} H_{p_{k,l}} \partial_{k} u_{ij} \right\}
\]
\[
\leq \lambda (\beta_H + \varepsilon) |Du|^q + \lambda u^2 \beta_H |Du|^{q-2} |D^2 u H_{p}(Du) - Du_t|^2 + C_H \lambda u^2 \varepsilon |Du|^{2(q-2)} |D^2 u|^2
\]
\[
\leq \lambda (\beta_H + \varepsilon) |Du|^q + \beta_H |Du|^{q-2} |D^2 u H_{p}(Du) - Du_t|^2 + \sigma C_{\varepsilon} |Du|^{2(q-2)} |D^2 u|^2
\]
where we used that \(\lambda u^2 \leq \sigma\) due to the choice of \(\lambda\). Notice that, for a sufficiently small \(\sigma\) (independent of \(\varepsilon\!\!\!\!\!\!\!\!\!\!\!\!\!\!), last two terms can be absorbed by the left-hand-side in (6.18). Similarly we estimate, still using \(\lambda u^2 \leq \sigma\),
\[
\lambda u (DV \cdot H_p - H_{pp}DV \cdot Du) + (1 + \theta) D^2 V(x) H_p \cdot H_p
\]
\[
\geq -C(\lambda + 1)|Du|^{2(q-1)} - \sigma C_{\varepsilon}.H.
\]
Once we insert the above inequalities in (6.18), choosing \(\sigma\) suitably small and dropping positive terms we get
\[
- \text{tr} (A D^2 z) + H_{pp}DV \cdot Dz + \rho z - \text{tr} ([D_{\eta}A \cdot Dz] D^2 u) + \lambda |H_{p}(Du) \cdot Du - u_{t}|^2
\]
\[
\leq C_0 + \lambda (\beta_H + \varepsilon) |Du|^q + K(1 + \lambda)|Du|^{2(q-1)}.\] (6.19)
We point out that the choice of \(\sigma\) is fixed, by now. We compare now the function \(z\) with
\[
\varphi(t) := L \left( \frac{1}{t^2} + \frac{1}{(T-t)^2} \right).
\]
Since \(\varphi\) blows-up at \(t = 0, T\), we have that \(z - \varphi\) admits a maximum point in \((0,T) \times \mathbb{R}\). As in Step 2 in Theorem 3.4 we can show that \(Dz \cdot \vec{\nu} \leq 0\) on \(\partial \Omega\), and no maximum could occur on the boundary. We now analyze maximum points inside the domain. In this case, we have \(Dz = 0\) and \([D_{\eta}A \cdot Dz] = 0\) (because \(A\) does not depend on \(u_t\)), and we also have
\[
\text{tr} (A D^2 z) \leq \text{tr} (A D^2 \varphi) = 6L \left( \frac{1}{t^2} + \frac{1}{(T-t)^2} \right).
\]
Moreover, if \(\max(z - \varphi) \geq L_0\), then using (6.17) we have
\[
H_p \cdot Du - u_t = H_p \cdot Du - (1 + \theta) H(Du) - \frac{\nu^2}{2} + z \geq \theta H(Du) + \varphi + L_0 - \sigma - c_0
\]

28
which implies, for $L_0 \geq \sigma + c_0$,  
$$|H_p \cdot Du - u_t| \geq \theta^2 H(Du)^2 + \varphi^2 \geq c_H |Du|^{2q} + L^2 \left( \frac{1}{t^4} + \frac{1}{(T-t)^4} \right).$$

Therefore (6.19) implies, on the maximum point,  
$$\lambda c_H |Du|^{2q} + (\lambda L^2 - 6L) \left( \frac{1}{t^4} + \frac{1}{(T-t)^4} \right) \leq C_0 + K(1 + \lambda)|Du|^{2(q-1)} + (\beta_H + \varepsilon)|Du|^q$$

which cannot hold if $L$ is sufficiently large. The conclusion is that, with a suitable choice of $L$, we have  
$$z \leq L \left( \frac{1}{t^4} + \frac{1}{(T-t)^4} \right) + L_0$$

which implies, using the definition of $z$ and $m$:  
$$\theta H(Du) + \varepsilon \log m + V(x) \leq L \left( \frac{1}{t^4} + \frac{1}{(T-t)^4} \right) + L_0.$$

This yields (6.16). \qed

**Remark 6.6** As we mentioned, the above estimate says that, once we drop the positivity condition on the marginals, $u$ should remain smooth on the support of $m$. But unfortunately, we do not have precise informations on the behavior of the support of the solution in that case (at least in dimension $d > 1$). This issue is possibly related to the regularity of solutions in critical sets for degenerate quasilinear elliptic problems.

**Convergence to weak solutions.** Since [3, 4, 5], a quite general theory of weak solutions is available for mean-field game systems as  
\[
\begin{align*}
-u_t + H(Du) &= f(m) + V(x) \quad \text{in } Q := (0, T) \times \Omega, \\
-m_t - \text{div}(m H_p(Du)) &= 0 \quad \text{in } Q, \\
H_p(Du) \cdot \nu &= 0 \quad \text{on } (0, T) \times \partial \Omega, \\
m(0, \cdot) &= m_0, \quad m(T, \cdot) = m_1 
\end{align*}
\]  
(6.20)

whenever $H$ is convex and $f$ is nondecreasing. The theory was actually developed when a terminal condition is prescribed on $u$ (a final pay-off in the cost functional), rather than for the transport problem in which both marginals are imposed on $m$ (at $t = 0, t = T$); but the transport case was also addressed (see [11, 17]) relying on the variational interpretation of those solutions as relaxed minima of the corresponding functionals. Up to minor differences (related to possibly different growth conditions on $H$ and $f$), weak solutions are defined as follows. For simplicity, we consider here the case that the Hamiltonian has quadratic-like growth. Let us recall that $P(\Omega)$ denotes the space of probability measures, endowed with the Kantorovich-Rubinstein distance  
$$d_1(m, m') = \sup_\phi \int_\Omega \phi d(m - m') \quad \forall m, m' \in P(\Omega),$$

where the supremum is taken over all 1-Lipschitz continuous maps $\phi : \Omega \to \mathbb{R}$.

**Definition 6.7** A pair $(u, m)$ is a weak solution of (6.20) if $m \in C^0([0, T]; P(\Omega)) \cap L^1(Q)$ with $m(0) = m_0$, $m(T) = m_1$, $u \in L_{loc}^2((0, T); H^1(\Omega))$ and in addition $m|Du|^2 \in L^1(Q)$, $f(m) \in L^1(Q)$ and $(u, m)$ satisfy:  

(i) $u$ is a weak sub-solution satisfying, in the sense of distributions,  
$$-u_t + H(Du) \leq f(m) + V(x) \quad \text{in } Q$$

(ii) $m$ is a weak solution satisfying, in the sense of distributions, the continuity equation  
$$m_t - \text{div}(m H_p(Du)) = 0 \quad \text{in } Q$$

(iii) $(u, m)$ satisfy the identity  
$$\int_{\Omega} m_0 u(0) \, dx - \int_{\Omega} u(T) \, m_1 \, dx = \int_0^T \int_{\Omega} f(m)m \, dx dt + \int_0^T \int_{\Omega} \frac{\partial}{\partial t} (m H_p(Du) \cdot Du - H(Du)) \, dx dt$$

(6.21)

where $u(0), u(T)$ are the one-sided traces of $u$ (which are well defined by properties of sub-solutions of Hamilton-Jacobi equations, see [37, Section 5]).
One motivation in the construction of smooth solutions (Theorem 6.1) is that it provides a regularization for weak solutions of \((\ref{6.20})\), by adding a small entropy term in the coupling function \(f\). In particular, this regularization allows one to justify the estimates proved in Section 5, which eventually holds for weak solutions in the limit as \(\varepsilon \to 0\). We give below a sample statement of this kind. This is to be compared with the results in \([22]\), where similar estimates are justified for weak solutions by using time-discretization and the geodesic interpretation of \(m(t)\) from optimal transport theory, whereas our approach is different and entirely relies on the PDE Eulerian approach developed in Section 5. Let us stress that weak solutions, as defined above, coincides with minima of the corresponding functional \((\ref{6.1})\) (with \(\varepsilon = 0\)).

**Theorem 6.8** Let \(Q\) be a \(C^3\), bounded, convex domain in \(\mathbb{R}^d\) and let \(V \in W^{1,\infty}(Q)\), \(m_0, m_1 \in L^\infty(Q) \cap P(Q)\). Let \(H\) be a \(C^2\) function satisfying \((\ref{2.11})\) and \((\ref{2.3})-\(\ref{2.4}\)). Assume that \(f \in C^1([0, \infty))\) is an increasing function satisfying condition \((\ref{5.4})\). Then there is a unique \(m\) and a unique \(u\) (up to \(m\)-negligible sets) such that \((u, m)\) is a weak solution of problem \((\ref{6.20})\), with \(\int_Q u(T)m_1 = 0\).

Moreover we have that \(m \in L^\infty(Q)\), it is the unique minimum of the functional \((\ref{6.1})\) (with \(\varepsilon = 0\)) and satisfies estimates \((\ref{5.7})-\(\ref{5.8}\)) Finally \(m\) is the limit as \(\varepsilon \to 0\) of the minima \(m^\varepsilon\) of regularized problems, obtained in Theorem 6.7.

**Proof.** We only sketch the argument. For \(\varepsilon > 0\), we consider the solution \((u^\varepsilon, m^\varepsilon)\) of \((\ref{2.1})\), where \(H, f, V\) are suitably regularized in order that Theorem 6.1 can be applied. We take for \(u^\varepsilon\) the normalized version given by Lemma 4.2 i.e. \(\int_Q u^\varepsilon(T)m_1dx = 0\). Applying Proposition 5.2 we have that \(\|m^\varepsilon\|_\infty\) is uniformly bounded. Moreover, by estimate (4.5) in Lemma 4.2 we also have that \(u^\varepsilon\) is locally uniformly bounded. We notice that \(u^\varepsilon\) satisfies, for some constant \(C\),

\[-u^\varepsilon_t + H(Du^\varepsilon) \leq f(m^\varepsilon) + V + \varepsilon C,\]

\((\ref{6.22})\)

where the right-hand side is bounded in \(L^\infty(Q)\). Now we use some arguments from \([31]\) and \([9\), Section 1.3.7]. Up to subsequences, we may assume that \(u^\varepsilon\) converges to some function \(u\) (at least weakly in \(L^2((a,b); H^1(\Omega))\) for all \(0 < a < b < T\), and that, for a.e. \(t \in (0, T)\), \(u(t)\) is the limit of \(u^\varepsilon(t)\) in the weak-* topology of \(L^\infty(\Omega)\). By stability of sub-solutions (see e.g. \([31\), Thm 5.11]), we have that \(u\) satisfies

\[-u_t + H(Du) \leq \bar{f} + V\]

\((\ref{6.23})\)

where \(\bar{f}\) is the weak-* limit of \(f(m^\varepsilon)\) in \(L^\infty(Q)\). By property of sub-solutions, \(u\) has one-sided limits as \(t \to 0^+, t \to T^-\); since from \((\ref{6.22})\) \(u^\varepsilon\) satisfies

\[\int_Q u^\varepsilon(t)m_1 \leq C(T-t)\]

we also deduce, letting first \(\varepsilon \to 0\) and then \(t \to T\), that \(\int_Q u(T)m_1 \leq 0\). As for \(m\), if \(w\) denotes the weak limit (e.g. in \(L^2(\Omega)\)) of \(m^\varepsilon H_p(Du^\varepsilon)\), then \(m\) is a bounded weak solution of the continuity equation

\[\left\{ \begin{array}{l} m_t - \text{div}(w) = 0 \\ m \in C([0,T],P(\Omega)) \text{, } m(0) = m_0 , m(T) = m_1. \end{array} \right.\]

\((\ref{6.24})\)

The duality between \((\ref{6.22})\) and \((\ref{6.24})\) was exploited in \([31\), Section 5.3]; roughly speaking, it is possible to cross multiply \((\ref{6.22})\) and \((\ref{6.24})\) Using that \(\int_Q u(T)m_1 \leq 0\), and reasoning as in \([9\), Thm 1.15], then one shows

\[\limsup_{\varepsilon \to 0} \int_Q f(m^\varepsilon)m^\varepsilon \leq \int_Q \bar{f} m dx dt.\]

A standard Minty-type argument implies that \(\bar{f} = f(m)\), and since \(f\) is increasing, one also deduces that \(m^\varepsilon \to m\) a.e. in \(Q\), and therefore in \(L^p(Q)\) for every \(p < \infty\). Moreover, the identification of \(\bar{f}\) goes together with the identification of \(w = m H_p(Du)\), using the convexity of \((m, w) \mapsto L(w/m)m\). Hence one follows arguments in \([31\) or \([9\): \((u,m)\) satisfy the equality

\[\int_Q m(H_p(Du)Du - H(Du)) + (f(m) + V)m = \int_Q m_0 u(0) dx\]

and \(\int_Q u(T)m_1 dx\) is proved to be zero. It follows from the above equality that \(m\) is a minimum of the limit functional and that \((u,m)\) is a weak solution of \((\ref{6.20})\) in the sense of Definition 6.7. The uniqueness of \(m\) and of \(u\) up to \(m\)-negligible sets can be proved as in \([9\), Thm 1.16].

\[\square\]
The compact case. All results proved here holds in more generality if the state space is the flat torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$. In fact, requiring the Hamiltonian to be radial was only needed to handle the regularity near the boundary $\partial \Omega$. Therefore, all results stated remain true in the torus removing the assumption (2.11). In the same spirit, the results could be extended to a compact Riemannian manifold without boundary, although this requires to use Bochner’s formula to handle the gradient estimates in the Riemannian setting. We will exploit this case in future work.

7 Appendix: existence of solutions to the elliptic problem

Here we show the existence of solutions to problem (3.4) where, for simplicity, we fix $\varepsilon = 1$, $T = 1$. In order to construct a solution, we follow the classical approach and we build a family of one-parameter problems. Let

\[ m^\tau := \varphi^\tau(-u_t + H(Du) - \tau V(x)), \quad \varphi^\tau := (\tau f(\cdot) + \log(\cdot))^{-1} \]  

(7.1)

and we consider the problem

\[
\begin{aligned}
& - \operatorname{tr}(A^\tau(t, x, Du)D^2 u) + \tau DV(x) \cdot H_p(Du) + \rho u = 0 & \text{in } (0, 1) \times \Omega, \\
& -u_t + H(Du) - \tau f(m_1) + \tilde{V}(x) - \delta u = \psi_1(x) & \text{at } t = 1, x \in \Omega, \\
& -u_t + H(Du) - \tau f(m_0) + \tilde{V}(x) + \delta u = \psi_0(x) & \text{at } t = 0, x \in \Omega, \\
& H_p(Du) \cdot \vec{n} = 0 & \text{on } (0, 1) \times \partial \Omega,
\end{aligned}
\]

(7.2)

where

\[ A^\tau = (a^\tau_{ij}((t, x), Du)) := \begin{pmatrix}
1 & -H_p(Du) \\
-H_p(Du) & H_p(Du) \otimes H_p(Du)
\end{pmatrix} + (1 + \tau m^\tau f'(m^\tau)) \begin{pmatrix}
0 & 0 \\
0 & H_{pp}
\end{pmatrix}, \]

and where $\tilde{V}, \psi_0, \psi_1 \in C^{1,\alpha}(\overline{\Omega})$. Problem (3.4) actually corresponds to $\tau = 1$, $\tilde{V} = V$, $\psi_0 = \log(m_0)$, $\psi_1 = \log(m_1)$.

To this problem we are going to apply Theorem 1.2 (Chapter X, page 462) in [19]. Unfortunately, we cannot apply directly this result because our domain does not look regular in the time-space environment; otherwise said, our nonlinear Neumann boundary condition is only piecewisely defined at the $(t, x)$ boundary. To overcome this technical issue, we will use the radial structure of the Hamiltonian $H$ (see (2.11)). This assumption simplifies the Neumann condition on $\partial \Omega$ and allows us to follow the classical steps up to using an extra reflection argument, in order to infer the second order estimates. A similar result for general Hamiltonians (without requiring (2.11)) would need to develop different technical tools to handle the $C^1$ estimates for piecewisely defined nonlinear boundary value problems; this is beyond our present scopes.

We recall that this kind of quasilinear problems (with nonlinear boundary conditions) can be solved provided one shows (uniformly in $\tau$):

- A global Lipschitz bound for (smooth) solutions of (2.5)
- A global $C^{1,\alpha}$ estimate for (smooth) solutions of (7.2)
- Solvability and $C^{2,\alpha}$ estimates for the linearized version of (7.2)

We split those three tasks in the next propositions.

**Proposition 7.1** Assume that $V \in C^{2,\alpha}(\overline{\Omega})$, that $H \in C^{3,1}(\mathbb{R}^d)$ satisfies (2.11) and (2.5)-(2.6) for some $q > 1$ and $\varpi > 0$, and that $f \in C^{2,1}(0, \infty)$ satisfies

\[ \exists \beta, s_0 > 0 : |f''(s)| \leq \beta (1 + f'(s)s) \quad \text{for all } s \geq s_0. \]

(7.3)

Let $m_0, m_1, \tilde{V}, \psi_0, \psi_1 \in C^{1,\alpha}(\overline{\Omega})$. Then there exists $M$, independent of $\tau \in [0, 1]$, such that any $u \in C^{2,\alpha}(\overline{Q})$ which is a solution of (7.2) satisfies

\[ \|u\|_{\infty} + \|Du\|_{\infty} \leq M \]

(7.4)

where $M$ depends on $\delta, \|m_0\|_{W^{1,\infty}(\Omega)}, \|m_1\|_{W^{1,\infty}(\Omega)}, \|\psi_0\|_{W^{1,\infty}(\Omega)}, \|\psi_1\|_{W^{1,\infty}(\Omega)}, \|	ilde{V}\|_{W^{1,\infty}(\Omega)}, \|V\|_{W^{2,\infty}(\Omega)}$ and on $H, f$.

**Proof.** This is the result of Theorem 3.3 (together with Lemma 5.2), up to minor remarks. The vector field $V$ is here replaced by $\tau V(x)$ in the interior and by $\tau \tilde{V}$ on the boundary, but this makes no difference in the proof. The condition (7.3) is slightly more restrictive than (3.3) in Theorem 3.3: this way we can assert that $\tau f(\cdot)$ satisfies the same condition as $f$, with the same constant. Hence, the gradient estimates are independent of $\tau$. We point out that the regularity required on $V, H, f$ guarantee that any $C^2$ solution $u$ actually belongs to $C^3(Q)$, so that the gradient bounds can be classically derived. Finally, we notice that the bound on $u$ depends on $\delta$ from maximum principle, as in Lemma 3.2. $\square$
Proposition 7.2 Under the same assumptions of Proposition 7.1 there exists \( C \), independent of \( \tau \in [0,1] \), such that any \( u \in C^{2,\alpha}(\overline{Q}) \) which is a solution of \( (7.2) \) satisfies
\[
\|u\|_{C^{1,\alpha}(\overline{Q})} \leq C
\]
where \( C \) depends on the constant \( M \) in \( (7.3) \) (as well as on the bounds of \( H_p, H_{pp}, H_{ppp} \) for \( |p| \leq M \) and again on \( \|m_0\|_{W^{1,\infty}(\Omega)}, \|m_1\|_{W^{1,\infty}(\Omega)}, \|\psi_0\|_{L^\infty(\Omega)}, \|\psi_1\|_{W^{1,\infty}(\Omega)}, \|\bar{\nabla}\|_{W^{1,\infty}(\Omega)} \).

Proof. The elliptic problem in \((7.2)\) can be rephrased as
\[
\begin{align*}
L_{\tau}(u) := & -a_{ij}^\tau(\eta, Du)u_{x_i x_j} + \tau D V(x) \cdot H_p(Du) + \rho u = 0 \quad \text{in } Q, \\
N^\tau(u) = & (\psi_1, \psi_0) \quad \text{on } \Sigma, \\
H_p(Du) \cdot \nu = & 0 \quad \text{on } (0,1) \times \partial \Omega,
\end{align*}
\]
where \( \eta = (t,x), \Sigma = (\{1\} \times \Omega) \cup (\{0\} \times \Omega) \) and the boundary operator \( N^\tau \) is defined as
\[
N_\tau(u) := \begin{cases} -u_t + H(Du) - \tau(f(m_1) + \bar{V}(x)) - \delta u & \text{at } t = 1, x \in \Omega, \\ -u_t + H(Du) - \tau(f(m_0) + \bar{V}(x)) + \delta u & \text{at } t = 0, x \in \Omega. \end{cases}
\]
Due to the estimate \((7.4)\) in Proposition 7.1, we have that the coefficients \( a_{ij}^\tau \) need only be considered in the compact set \( \mathcal{M} := \{ \eta \in Q, |u| \leq M, |Du| \leq M \} \). Assuming \( V \) of class \( C^{2,\alpha}_x \) and \( f \) of class \( C^{2,1}_x \) and \( H \) of class \( C^{3,1} \), we get that the coefficients \( a_{ij}^\tau \), as well as the first order coefficients, are \( C^{1,\alpha}_x \) in their arguments, with \( \frac{\partial a_{ij}^\tau(y,p)}{\partial y} \) being Lipschitz in \( p \). In particular, for the interior regularity we can apply the classical results (see e.g. [19, Chapter 6, Thm 1.1]) and we deduce the \( C^{1,\alpha} \) interior estimate for \( u \).

We are only concerned now with the boundary regularity. Without loss of generality (because \( H_{ppp} \) is bounded below), we assume here that
\[
H(p) = h(|p|^2),
\]
for some \( C^2 \) function \( h \) such that \( h' > 0 \). With this notation we have \( H_p = 2h'p \) and \( H_{ppp} = 2h'ppp \).

Now we use a (local) reflection argument in the \( x \) variable. To this purpose, recall that for \( x \in \Omega \), we denote by \( d(x) \) the distance of \( x \) to the boundary. It is well known (see e.g. [18]) that if \( \Omega \) is of class \( C^k \), \( d(x) \) is a \( C^k \) function in a neighborhood of the boundary. More precisely, there exists a positive number \( d_0 > 0 \) such that, if we denote \( \Omega_{d_0} := \{ x \in \Omega : d(x) < d_0 \} \), then any point \( x \in \Omega_{d_0} \) can be represented as
\[
x = -\delta \bar{n}(\bar{x}) + \bar{x}
\]
where \( \delta = d(x) \) and \( \bar{x} \) is the projection of \( x \) onto \( \partial \Omega \) (\( \bar{n}(\bar{x}) \) is the outward unit vector at \( \bar{x} \)). To ease notations, we denote afterhere by \( N \) the dimension of the euclidean space (i.e. \( \Omega \subset \mathbb{R}^N \)). Therefore, if we represent \( \bar{x} \in \partial \Omega \) in a system of coordinates of some local chart, we may assume that \( \bar{x} = (y', g(y')) \) for some \( C^k \) function \( g \) defined on some open subset \( U \subset \mathbb{R}^{N-1} \). Hence \((7.4)\) yields a local diffeomorphism between \( \Omega_{d_0} \) and \( U \times (0, d_0) \) defined as
\[
T(x) := (y', \delta) \iff x = -\delta \bar{n}(\bar{x}) + \bar{x}, \quad \bar{x} = (y', g(y')) \in \partial \Omega.
\]
Let us now set \( y_N := \delta \) so that \( y = (y', y_N) \) and \( y = T(x) \) will shortly denote the above change of variables. We notice that assuming \( \Omega \) of class \( C^3 \) guarantees that the mapping \( T \) is of class \( C^2 \). The advantage of this change of coordinates lies in the property that
\[
\nabla T_k(x) \cdot \nabla T_N(x) = \nabla T_k(x) \cdot \nabla d(x) = 0 \quad \forall k = 1, \ldots, N - 1
\]
because \( T_N(x) = d(x) \) (the distance function) and the projection coordinates remain constant if we move along the normal direction (i.e. \( \nabla T_k(x) \cdot \nabla d(x) = 0 \) for all \( N-1 \) coordinates \( y_k = T_k(x) \)).

Now we set
\[
\tilde{u}(t,y) := u(t, T^{-1}y) \quad \Rightarrow \quad u(t,x) = \tilde{u}(t,T(x))
\]
and we compute:
\[
u_{x_i} = \sum_{k=1}^N \tilde{u}_{y_k} \frac{\partial T_k}{\partial x_i} ; \quad u_{x_i x_j} = \sum_{k,l=1}^N \tilde{u}_{y_k y_l} \frac{\partial T_k}{\partial x_i} \frac{\partial T_l}{\partial x_j} + \sum_{k=1}^N \tilde{u}_{y_k} \frac{\partial^2 T_k}{\partial x_i \partial x_j}
\]
so the equation \((7.6)\) is transformed into
\[
\tilde{u}_{tt} - 2 \sum_k (DT_k \cdot H_p(Du)) \tilde{u}_{ty_k} + \sum_{k,t=1}^N \tilde{a}_{kt} \tilde{u}_{y_k y_t} + \sum_{k=1}^N \tilde{b}_k \tilde{u}_{y_k} = \rho \tilde{u}
\]
with \( \tilde{a}_{ij} = a_{ij}^\tau, \tilde{b}_i = b_i^\tau, \). (7.12)
where
\[ a_{kl}^\tau := \sum_{i,j} a_{ij}^\tau \frac{\partial T_{k}}{\partial x_i} \frac{\partial T_{\ell}}{\partial x_j}, \quad \mu_k^\tau := \sum_{i,j} a_{ij}^\tau \frac{\partial^2 T_{k}}{\partial x_i \partial x_j} - \tau 2 h'_D V \cdot DT_k \]
and where the coefficients \( a_{ij}^\tau \) are computed on \( Du \); recall that \( Du = M \tilde{u} \) for \( M = (m_{ij}) = \left( \frac{\partial T_{\ell}}{\partial x_i} \right) \).

We extend now \( \tilde{u} \) for \( y_N < 0 \) by setting
\[ u^*(t, y) := \begin{cases} \tilde{u}(t, y', y_N) \quad & \text{if } y' \in U, y_N \in (0, d_0) \\ \tilde{u}(t, y, -y_N) \quad & \text{if } y' \in U, y_N \in (-d_0, 0) \end{cases} \]
Notice that \( 7.10 \) implies so the Neumann condition \( H \) condition \( 7.10 \),

where \( m_{ij} \) for some smooth function \( \kappa \) for \( \text{some smooth function } \kappa \text{ with respect to reflection} \); hence, using the above notation,

Then we compute from \( 7.13 \), using the definition of \( \tilde{u}_{y_N} \) and \( \text{the radial structure of } H, \text{ and due to condition } 7.10 \),

we have
\[ (1 + \tau \eta' f'((m^\tau)) = \kappa(V^*(y), Du^*) \] (7.14)

for some smooth function \( \kappa^* \), where \( V^*(y) \) is the even reflection of \( V(T^{-1}(y)) \).

Then we compute from \( 7.13 \), using the definition of \( a_{ij}^\tau \) and the radial structure of \( H, \) and due to condition \( 7.10 \),

\[ \forall k \neq N, \quad \tilde{a}_{kN} := \sum_{i,j} \left[ H_{p_i} \otimes H_{p_j} + H_{p,p_j}(1 + \tau m^\tau f'((m^\tau))) \right] \frac{\partial T_{k}}{\partial x_i} \frac{\partial T_{N}}{\partial x_j} \]
\[ = \sum_{i,j} \left( 4(h')^2 + 4h'' m^\tau V^*(y), Du^* \right) u_{x_i} u_{x_j} \frac{\partial T_{k}}{\partial x_i} \frac{\partial T_{N}}{\partial x_j} \]
\[ = \left( 4(h')^2 + 4h'' m^\tau V^*(y), Du^* \right) \sum_{p=1}^{N-1} DT_p \cdot DT_k \tilde{u}_{y_N} \tilde{u}_{y_N} \]
where we used \( 7.11 \) (and once more \( 7.10 \)) in the last step. In particular, we notice that \( \tilde{a}_{kN}^\tau \) is an odd coefficient with respect to reflection; hence, using the above notation,

\[ \forall k \neq N, \quad \tilde{a}_{kN}^\tau \tilde{u}_{y_N} = \left( 4(h')^2 + 4h'' m^\tau V^*(y), Du^* \right) (DT_k \cdot M D u^*) u_{y_N}^* u_{y_N}^* \]

With similar computations we find
\[ \tilde{a}_{N N}^\tau = \left( 4(h')^2 + 4h'' m^\tau V^*(y), Du^* \right) (u_{y_N}^*)^2 + 2h' m^\tau V^*(y), Du^* \].

Still using \( 7.8, 7.11 \) and \( 7.10 \), we observe that
\[ \sum_{k} (DT_k \cdot H_p(Du)) \tilde{u}_{y_N} = 2h'(\tilde{u}_{y_N} \tilde{u}_{y_M} + \sum_{k,\ell=1}^{N-1} (DT_k \cdot DT_\ell) \tilde{u}_{y_N} \tilde{u}_{y_k}) \]
\[ = 2h'(u_{y_N}^* u_{y_N}^* + \sum_{k,\ell=1}^{N-1} (DT_k \cdot DT_\ell) u_{y_N}^* u_{y_N}^*) \]

33
Therefore, $u^*(t,y)$ solves the following elliptic equation for $t \in (0,1), y \in B$, where $B \subset \mathbb{R}^N$ is any open set which is contained in $U \times (-d_0,d_0)$:

$$u_{tt}^* - 4h'(u_{yy}^*)u_{ty}^* + \sum_{k=1}^{N-1} (DT_k \cdot DT_t) u_{y_k}^* u_{y_t}^* + \sum_{k=1}^{N} a^*_{kk} u_{y_k}^* + b^*(t,y) = \rho u^*$$

(7.15)

where

$$a^*_{ii} = (4(h')^2 + 4h'' \kappa^2 (V^*(y), \mathbf{Du}^*)) (DT_k \cdot M'D'u^*) (DT_k \cdot M'D'u^*) + 2h' \kappa^2 (V^*(y), \mathbf{Du}^*) (DT_k \cdot DT_t) \quad \text{if} \ k, \ell \neq N,$$

$$a^*_{NN} = (4(h')^2 + 4h'' \kappa^2 (V^*(y), \mathbf{Du}^*)) (DT_k \cdot M'D'u^*) u_{y_N}^* \quad \text{if} \ k \neq N,$$

$$a^*_{NN} = (4(h')^2 + 4h'' \kappa^2 (V^*(y), \mathbf{Du}^*)) |u_{y_N}^*|^2 + 2h' \kappa^2 (V^*(y), \mathbf{Du}^*)$$

and where $b^*(t,y)$ is the even extension (through $y_N = 0$) of the term $\bar{b}^* \cdot \bar{D} \bar{u}$.

Now we apply to equation (7.15) the $C^{1,0}$-estimates (up to the boundary) for elliptic equations with nonlinear first order condition at the boundary, see e.g. [24, Lemma 2.3]2. Let us point out that the boundary here is represented by the hypersurfaces $(\Sigma_0 \cap B) \cup (\Sigma_1 \cap B)$; the solution $u^*$ satisfies a uniform bound $|u^*| + |\mathbf{Du}^*| \leq K$ (by Proposition 7.1) and the coefficients $a_{ij}^*((t,y), \mathbf{Du}^*)$ are $C^1$ with respect to $\mathbf{Du}^*$ and $\frac{\partial a_{ij}^*}{\partial y^j}$ as well as $\frac{\partial a_{ij}^*}{\partial y^j}$ are uniformly bounded (because $H$ is $C^{2,1}$ in $p$ and $V$ is Lipschitz). Note that $a_{ij}^*$ depends on $y$ only through the even reflections of $V(y)$ and $T^{-1}(y)$. As for the coefficient $b^*(t,y)$, it is actually continuous (thanks to the fact that $u_{y_N}^* = 0$) and what only matters, is uniformly bounded by some constant depending on $K, \|DV\|_\infty$ and $\|D^2T\|_\infty$. Finally, the boundary operator $N^*$ is Lipschitz with respect to $y$ and $C^2$ with respect to the gradient variable. According to [24, Lemma 2.3], $\mathbf{Du}^*$ satisfies a bound in the Hölder norm, which is uniform with respect to $\tau$. With a standard localization argument, based on a partition of unity, we conclude the bound (7.5).

**Remark 7.3** We stress that a $C^{1,0}$ estimate (up to the boundary) can also be obtained if we look at the elliptic equation in the divergence form, which is clearly inherited from the continuity equation in (2.1). In the notations used in the previous proof, the function $u^*$, obtained after local change of coordinates and reflection, satisfies the equation

$$\partial_t m^* - \sum_{k=1}^{N} \frac{\partial}{\partial y_k} (m^* 2h(|MDu^*|^2)DT_k \cdot (MDu^*)) + \rho u^* \frac{m^*}{1 + \tau f(m^*) m^*} = 0$$

where

$$m^* = \varphi^* (-u_t + h(|MDu^*|^2) - \tau V),$$

where we recall that $\varphi^*$ is defined in (2.1) and $M = (m_{ij}) = (\frac{\partial T}{\partial x^j}), |MDu^*|^2 = |u_{y_N}^*|^2 + \sum_{k=1}^{N} \left( \sum_{\ell=1}^{N} \frac{\partial T}{\partial x^\ell} u_{y_N}^* \right)^2.$

Hence the equation can be rewritten as a nonlinear divergence form equation on $u^*$:

$$\partial_t (A_0(y, \mathbf{Du}^*)) + \sum_{k=1}^{N} \frac{\partial}{\partial y_k} (A_k(y, \mathbf{Du}^*)) = \beta(y, u^*, \mathbf{Du}^*) \quad (t,y) \in (0,1) \times B$$

complemented with a co-normal boundary condition at $t = 0, t = 1$:

$$A \cdot n = \begin{cases} A_0(y, \mathbf{Du}^*) & \text{for } t = 1, \\ -A_0(y, \mathbf{Du}^*) & \text{for } t = 0 \end{cases}$$

$$= \left( \varphi^*(\psi^*_t + \tau(f(m_1)^* + \bar{V} - V^*) + \delta u^*), -\varphi^*(\psi^*_0 + \tau(f(m_0)^* + \bar{V} - V^*) - \delta u^*) \right)$$

where $(\cdot)^*$ denotes the even extension of the various functions. Notice that when $\bar{V} = V$, $\delta = 0$ and $\psi_i = \log(m_i)$, then we have $A \cdot n = (m_1, -m_0)^*.$

Since $|\beta(y, u^*, \mathbf{Du}^*)| \leq C$ (because $\mathbf{Du}^*$ is bounded), and due to the regularity of the boundary terms, if $H_{pp}$ is nondegenerate the $C^{1,0}$ estimate for $u^*$ can then be deduced by [25, Thm 2].

We mention this alternative approach because it could be exploited if one aims at generalizing the $C^{1,0}$ estimate to possibly degenerate Hamiltonians, e.g. satisfying (2.5) with $\omega = 0.$

---

2We warn the reader that, literally, the result in [24, Lemma 2.3] assumes the coefficients $a_{ij}(y, \mathbf{Du})$ to be $C^1$ with respect to $y$; however this is not necessary, as can be readily checked by inspecting the proof as well as by noticing that the estimate only requires $\frac{\partial a_{ij}}{\partial y^j}$ to be bounded. In fact, the Lipschitz character of $a_{ij}$ with respect to $y$ is enough, which is here preserved by the reflection argument.
Last ingredient is the well-posedness of the linearized problem around one solution $u$ of (7.2) and, correspondingly, a compactness property for sequences of solutions of (7.2). In order to use the $C^{2,\alpha}$ regularity, here we need to use a compatibility condition at the boundary $\partial \Omega$. To this purpose, we denote $C^{1,\alpha}_{N} (\Omega)$ the space of functions $\phi \in C^{1,\alpha}(\Omega)$ such that $D\phi \cdot \nu = 0$ on $\partial \Omega$. We also denote $a_{ij}^{\alpha} = a_{ij}^\alpha (\eta_t, \eta_q)$.

**Proposition 7.4** In addition to the conditions of Proposition 7.1, assume that $\Omega$ is of class $C^4$, and that the functions $f(m_0)$, $f(m_1)$, $\psi_0$, $\psi_1$ belong to $C^{1,\alpha}_{N} (\Omega)$. Then the solution $u$ of (7.2) belongs to $C^{2,\alpha} (\Omega)$, and for every $\vartheta \in C^{0,\alpha} (\Omega), \zeta = (\zeta_1, \zeta_0) \in C^{1,\alpha}_{N} (\Omega)$ there exists a unique $\phi \in C^{2,\alpha} (\Omega)$ which is a solution of the linear problem

$$
\begin{aligned}
-a_{ij}^\alpha ((t,x), D\nu) \phi_{ij} - \partial_{\eta_{k}}^\alpha ((t,x), D\nu) u_{ij} \phi_k + \tau H_{pp} (D\nu) D\phi \cdot \rho \phi + \vartheta &= \text{in } Q, \\
-\phi_{t} + H_{p}(D\nu) \cdot D\phi &= \delta \phi + \zeta_1 \quad \text{on } \Sigma_T, \\
-\phi_{t} + H_{p}(D\nu) \cdot D\phi &= \delta \phi + \zeta_0 \quad \text{on } \Sigma_0, \\
D\phi \cdot \nu &= 0 \quad \text{on } (0,T) \times \partial \Omega.
\end{aligned}
$$

(7.16)

Moreover, if $\psi_{1m}, \psi_{0m}$ are sequences which converge in $C^{1,\alpha}_{N} (\Omega)$, then the corresponding solutions $u_m$ of (7.2) are relatively compact in $C^{2,\alpha} (\Omega)$.

**Proof.** The first assertion follows from Proposition 7.2 if we come back to the equation (7.15) complemented with the boundary condition satisfied by $u^{*}$, which reads as

$$
\begin{aligned}
-u^{*}_{t} + h(|M D u^{*}|^2) - \delta u^{*} &= \tau (f(m_0)^{*} + \tilde{V}^{*}) + \psi^{*}_1 \quad \text{at } t = 1, \\
-u^{*}_{t} + h(|M D u^{*}|^2) - \delta u^{*} &= \tau (f(m_0)^{*} + \tilde{V}^{*}) + \psi_0^{*} \quad \text{at } t = 0.
\end{aligned}
$$

Since the boundary data are assumed to belong to $C^{1,\alpha}_{N} (\Omega)$, they are reflected into $C^{1,\alpha}$ functions. This implies that $u^{*}_t - h(|M D u^{*}|^2) \in C^{1,\alpha} (U)$. One can therefore apply the results of [23] to invoke that $u^{*} \in C^{2,\alpha}$, which leads to the $C^{2,\alpha}$ regularity of $u$ up to the boundary.

Let us now consider the linearized problem (7.10). Since $u \in C^{2,\alpha} (\Omega), (7.10)$ is a linear problem with Hölder coefficients, and Schauder’s theory applies to get, at least, the interior estimates of $\phi$. Let us only check the boundary regularity. With the notations of Proposition 7.2, we localize near the boundary with the change of variable $y = T(x)$, and then we extend the problem by reflection, setting $\phi^{*}(t,y) := \phi(t,T^{-1}(y))$ through the hyperplane $\{y_N = 0\}$. From the computations of Proposition 7.2 we know that the matrix $A^{*}$ is transformed into a new matrix $A^{*}$ which depends on $(y, D\nu^{*})$ as follows:

$$
A^{*} = (a_{ij}^{*}) = \begin{pmatrix} 1 & -2h' \Lambda_u \\
-2h' \Lambda_u & [4h''^2] \Lambda_u \otimes \Lambda_u \end{pmatrix} + \kappa^7 (V^{*}(y), D\nu^{*}) \begin{pmatrix} 0 \\
0 \\
0 \\
2h''(DT_{k} \cdot DT_{j}) + (4h'') \Lambda_u \otimes \Lambda_u \end{pmatrix}
$$

with $\Lambda_u := ((DT_{k} \cdot M'Du^{*})_{k=1,...,N-1}, u_{y_{N}}^{*})$ and $\kappa^7 (V^{*}(y), D\nu^{*})$ defined as in (7.14).

We look now at the first order terms. By definition of $a_{ij}^{*}$, we have

$$
\frac{\partial a_{ij}^{*}}{\partial \eta_{k}}((t,x), D\nu) u_{ij} \phi_k = -2H_{pp} D\nu u_{ij} \cdot D\phi + 2D^2 u_{pp} H_{pp} D\phi + \tau g'(m^7)(\varphi^{*})' [-\phi_{t} + H_{p} D\phi] \text{tr}(H_{pp} D^2 u) + (1 + \tau g'(m^7)) \text{tr}(H_{pp} D\phi D^2 u)
$$

(7.17)

where $g(s) = sf'(s)$. Using the radial structure of $H$ (7.8) and the chosen reference frame where (7.10) holds true, one can check (with long but routine computations, similarly as in Proposition 7.2) that all terms in (7.17) can be transformed into functions of $\delta \phi$ and $u^{*}$ except for a few terms which are only first order in $u$ and terms which involve second derivatives of the map $T$. Then, after the change of coordinates and the reflection, $\phi^{*}$ satisfies a linear elliptic problem of the following form in the open set $(0,T) \times B$:

$$
a_{ij}^{*} \phi^{*}_{ij} + B^{*} \phi^{*} + c^{*} + \rho \phi^{*} = \vartheta^{*} \quad (t,y) \in (0,T) \times B.
$$

(7.18)

Now we wish to apply to $\phi^{*}$ the Hölder estimates up to the boundary in order to conclude that $\phi^{*} \in C^{2,\alpha}([0,T] \times U \times (-d_0, d_0))$ for some open set $U \subset \mathbb{R}^{N-1}$. The regularity of $\phi^{*}$ will yield the $C^{2,\alpha}$ regularity of $\phi$ in $[0,T] \times (B \cap \Omega)$ for any ball $B$ such that $\partial \Omega \cap B$ is the portion of a smooth graph.

In (7.18), the $(d+1)$-dimensional vector field $B^{*}$ depends on $D^2 u^{*}$ (which is Hölder continuous) and on the (even reflection) of $DT, DV, D\tilde{u}$, which are Lipschitz continuous. Therefore, overall $B^{*}$ belongs to $C^{0,\alpha}$ for some $\alpha \in (0,1)$. Conversely, the term $c^{*}$ in (7.18) is the even reflection of a function depending on $D\phi, D\tilde{u}$ and $D^2 T$; in this case we cannot say that $c^{*}$ is reflected into a Hölder continuous function until we establish the Hölder regularity for $D\phi.$
Therefore, since the coefficient \( c^* \) does not belong to \( C^{0,\alpha} \) a priori, we cannot apply the Schauder’s estimates in one shot to infer the \( C^{2,\alpha} \) regularity of \( \phi^* \). However, the coefficients \( a_{ij}^* \) belong to \( C^{0,1} \) (because \( u^* \) is \( C^{2,0} \), while \( V, DT \) are \( C^1 \) and so they are reflected into a Lipschitz function); hence, by rewriting the equation in divergence form, \( \phi^* \) is, in particular, a bounded weak solution to the equation

\[
(a_{ij}^* \phi_i^*)_t + H(y, \phi^*, \phi_t^*) = 0 \quad (t, y) \in (0, T) \times U \times (-d_0, d_0)
\]

for some function \( H(y, \phi^*, \phi_t^*) \) which satisfies \( |H(y, \phi^*, \phi_t^*)| \leq C_0(1 + |D\phi^*|) \). In addition, \( \phi^* \) satisfies a co-normal boundary condition at \( \Sigma_0, \Sigma_T \). By regularity for divergence form equations (which is true for general nonlinear problems, see e.g. [25, Thm 2]), we deduce that \( \phi^* \in C^{1,\alpha} \) and

\[
\|\phi^*\|_{C^{1,\alpha}} \leq K \quad \text{for some } K = K(\|\phi\|_\infty, \|u\|_{C^{2,\alpha}}, \|T\|_{L^2}, |D\|_{L^\infty}, \|\phi\|_{L^\infty}, \|\zeta\|_{C^{1,\alpha}}).
\]

We can now use this information, which implies that the term \( c^* \) in (7.18) belongs to \( C^{0,\alpha} \). Since \( a_{ij}^*, B^* \), \( \theta^* \) also belong to \( C^{0,\alpha} \), and since the boundary data \( \phi^* \in C^{1,\alpha} \) (because \( \zeta \) satisfies a Neumann condition), the classical Schauder’s estimates (see e.g. [23, Lemma 1]) imply that \( \phi^* \in C^{2,\alpha} \) and

\[
\|\phi^*\|_{C^{2,\alpha}} \leq K \quad \text{for some } K = K(\|\phi\|_\infty, \|u\|_{C^{2,\alpha}}, \|T\|_{C^{2,\alpha}}, \|V\|_{C^{1,\alpha}}, \|\theta\|_\alpha, \|\zeta\|_{C^{1,\alpha}}).
\]

We notice that \( \|\phi\|_\infty \) is estimated uniformly by maximum principle. We also recall that, requiring \( \partial \Omega \) of class \( C^4 \), we have that \( T \) is of class \( C^3 \). Finally, we proved that all solutions to the linear problem (7.16) satisfy the estimate

\[
\|\phi\|_{C^{2,\alpha}} \leq C(\|\phi\|_\alpha + \|\zeta\|_{C^{1,\alpha}})
\]

for some \( C = C(\|u\|_{C^{2,\alpha}}, \|\theta\|_{C^{1,\alpha}}) \). By linear theory, this implies that problem (7.16) is uniquely solvable for every \( \tau \in [0, 1] \).

Now we prove the last assertion of the Proposition. To this goal, we first observe that, since \( \psi_m = (\psi_{0m}, \psi_{1m}) \) is convergent in \( C^{1,\alpha}(\Omega) \), by Proposition 7.2 the sequence \( u_m \) is bounded in \( C^{1,\alpha}(\Omega) \), and therefore it is relatively compact in \( C^{1,\alpha}(\Omega) \) by Ascoli-Arzelà theorem. We observe that \( w^{m,n} := u_m - u_n \) solves

\[
\begin{align*}
- \alpha_{ij}^*(x, Du_m)w_{ij}^{m,n} + B_{ij}^{m,n}w_{k}^{m,n} + \tau \Gamma_{ij}^{n,m}DV(x) \cdot D^2w^{m,n} + \rho w^{m,n} &= 0 \quad \text{in } (0, 1) \times \Omega, \\
- \partial_t w_{ij}^{m,n} + \beta_{ij}^{m,n} \cdot Dw_{ij}^{m,n} - \delta w_{ij}^{m,n} &= \psi_{1n} - \psi_{0n}, \quad \text{at } t = 1, x \in \Omega, \\
- \partial_t w_{ij}^{m,n} + \beta_{ij}^{m,n} \cdot Dw_{ij}^{m,n} + \delta w_{ij}^{m,n} &= \psi_{0n} - \psi_{1n}, \quad \text{at } t = 0, x \in \Omega, \\
Dw_{ij}^{m,n} \cdot \nu &= 0 \quad \text{on } (0, 1) \times \partial \Omega,
\end{align*}
\]

where

\[
B_{ij}^{m,n} = \sum_{ij} (u_n)_{ij} \int_0^1 \frac{\alpha_{kl}^*}{\partial_k} (x, \lambda Du_m + (1 - \lambda) Du_n) d\lambda
\]

and

\[
\beta_{ij}^{m,n} = \int_0^1 H_p(\lambda Du_m + (1 - \lambda) Du_n) d\lambda, \quad \Gamma_{ij}^{n,m} = \int_0^1 H_p(\lambda Du_m + (1 - \lambda) Du_n) d\lambda.
\]

This is a linear problem for \( w^{m,n} \), which is of the same kind as (7.18). Therefore, we proceed as we did before, by localizing near the boundary and reflecting the solution through (the straightened part of) \( \delta \Omega \). Then we apply to the reflected function \( (w^{m,n})^* \) the \( C^{2,\alpha} \) estimates as before (23, Lemma 1). We only stress that, differently than it was for the function \( \phi^* \) above, now we already know that \( w^{m,n} \) is bounded in \( C^{1,\alpha}(\Omega) \); this allows us to apply directly the Schauder estimates because (in the notation used before) \( a_{ij}^*(w^{m,n})_{ij} \) belongs to and (can be estimated in) \( C^{0,\alpha}(\Omega) \). Applying the precise form of (23, Lemma 1), we conclude that

\[
\|w^{m,n}\|_{C^{2,\alpha}} \leq C_0\|w^{m,n}\|_{C^1} (1 + \|u_n\|_{C^{2,\alpha}} + \|u_m\|_{C^{2,\alpha}}) + C_1\|w^{m,n}\|_{C^{1,\alpha}} (\|D^2u_n\|_\infty + \|D^2u_m\|_\infty) \quad \text{.}
\]

From this estimate we conclude as in (23, Lemma 2); one shows first that \( \|u_n\|_{C^{2,\alpha}} \) is bounded (using the smallness of \( \|w^{m,n}\|_{C^1} \) in (7.20)) and this implies the compactness of \( u_n \) in \( C^{1,\alpha} \). Hence (7.20) yields the convergence of \( u_n \) in \( C^{2,\alpha}(\Omega) \).

We are ready to prove the existence of solutions to the elliptic problem (3.3).

**Proof of Theorem** 3.3 We start by assuming that the data are more regular, namely that \( \partial \Omega \) is of class \( C^4 \), \( V \in C^{2,\alpha}(\Omega) \), \( f \in C^{2,1}(0, \infty) \) satisfies (7.4) and \( H \in C^{3,1}(\mathbb{R}^d) \) satisfies (7.8). We also assume that the boundary data \( f(m_0), f(m_1), \tilde{V}(x), \psi_0, \psi_1 \) belong to \( C^{1,\alpha}(\Omega) \). Under those conditions, we can use
Propositions 7.1–7.4 and we establish the existence of a solution with a continuity method, following the same proof as in [19, Chapter X, Thm 1.1 & Thm 1.2]. Let us sketch some detail: we set

$$X := \{ v \in C^{2,\alpha}(\mathcal{Q}) : Dv \cdot \nu = 0 \text{ on } (0, T) \times \partial \Omega \}$$

and $X' := C^{0,\alpha}(\mathcal{Q}) \times C^{1,\alpha}(\mathcal{Q})$, and we define a family of mappings from $X$ into $X'$ by setting $\Phi(u, \tau) := (L^\tau(u), N_1^\tau(u), N_2^\tau(u))$, where $N_1^\tau, N_2^\tau$ are the restrictions to $t = 0, t = 1$ respectively, of the boundary operator $N^\tau$ in (7.4). We also set $X_0'$ the closed subset of $X'$ made of elements $(0, \psi_0, \psi_1)$, with $(\psi_0, \psi_1) \in C^{1,\alpha}(\mathcal{Q})^2$. Thanks to Proposition 7.2, the mapping $\Phi$ is locally invertible in a neighborhood of points $u \in X$ such that $\Phi(u, \tau) \in X_0'$ and in addition a sequence $u_k$ is compact in $X$ if $\Phi(u_k, \tau)$ is converging in $X_0'$. Finally, for $\tau = 0$ and $\psi_0 = \psi_1 = 0$, the problem (7.3) has the unique solution $u = 0$. Therefore, as in [19, Chapter X, Thm 1.1] one concludes that the problem with $\tau = 1$ is uniquely solvable.

In a second step, we approximate the data with smoother sequences. In particular, we approximate the domain $\Omega$ with a sequence of $C^4$ domains $\Omega_n$, and the functions $m_0, m_1$ with sequences $m_{0n}, m_{1n} \in C^{1,\alpha}(\mathcal{Q})$ such that $\|m_0\|_{W^{1,\infty}(\Omega)}, \|m_1\|_{W^{1,\infty}(\Omega)}$, are bounded, as well as $\log(m_0)\|_{W^{1,\infty}(\Omega)}$ and $\log(m_1)\|_{W^{1,\infty}(\Omega)}$. It is also possible to build $m_{0n}, m_{1n}$ in a way that those functions be constant along the normal in a small neighborhood of $\partial \Omega$, so that they belong to $C_N^{1,\alpha}(\mathcal{Q})$. Similarly we take a sequence $\tilde{V}_n \in C_N^{1,\alpha}(\mathcal{Q})$ which converges uniformly to $V$ and such that $\|\tilde{V}_n\|_{W^{1,\infty}(\Omega)}$ is bounded. Finally, we take another sequence $V_n \in C_N^{2,\alpha}(\mathcal{Q})$ converging to $V$ and such that $\|V_n\|_{W^{2,\infty}(\Omega)}$ is bounded. As for the function $f$, we approximate it with $f_n(m) = f\left(\frac{m}{1 + \varepsilon}\right)$ which satisfies the stronger condition (7.3) (for some constant $\beta$ possibly dependent on $n$) but also satisfies the weaker condition (7.3) for some constant which is uniform with respect to $n$. As for the function $H$, by assumption it is $C^{2,\alpha}$ and satisfies (7.8) and (2.5)–(2.6) for some $\tau > 0$; hence, up to replacing $h$ with $h(\sqrt{n^{-1} + 1})$, and up to a further regularization if needed, we can assume that $H$ is approximated by some $H^n$ which is $C^{4,1}(\mathbb{R}^d)$ and satisfies (7.8), and in a way that (2.5)–(2.6) hold uniformly for $H^n$. We build so far an approximating problem

$$\begin{align*}
- \text{tr} \left( A_n(x, Du_n) D^2 u_n + \rho u_n + D V_n(x) \cdot H^n_p(Du_n) = 0 \right) & \quad \text{in } Q, \\
- D_t u_n + H^n(Du_n) = \delta u_n + f^\varepsilon(m_{1n}) + \tilde{V}_n(x) & \quad \text{at } t = 1, x \in \Omega, \\
- D_t u_n + H^n(Du_n) + \delta u_n = f^\varepsilon(m_{0n}) + \tilde{V}_n(x) & \quad \text{at } t = 0, x \in \Omega, \\
H^n_p(Du_n) \cdot \nu = 0 & \quad \text{on } (0, 1) \times \partial \Omega.
\end{align*}$$

which admits a solution $u_n \in C^{2,\alpha}(\mathcal{Q})$ by what we proved in the first part. This solution $u_n$ also belongs to $C^{3,\alpha}(Q)$ and we can apply Lemma 3.2 and Theorem 3.3 to infer a uniform estimate for $\|u_n\|_{W^{1,\infty}(Q)}$. By Proposition 7.2, we also deduce that $u_n$ is uniformly bounded in $C^{1,\alpha}(Q)$, because the $C^{1,\alpha}$ estimate only depends on the bound for $\|u_n\|_{W^{1,\infty}(Q)}$ and on the Lipschitz bounds of the boundary terms $f^\varepsilon(m_{1n}), \tilde{V}_n$. Hence $u_n$ is relatively compact in $C^{1,\alpha}(Q)$. By interior regularity, we also have that $u_n$ is bounded in $C^{2,\alpha}(Q)$, so it is relatively compact in $C^2$ if restricted to compact subsets in the interior. This is enough to pass to the limit and conclude that, up to subsequences, $u_n$ converges towards some $u \in C^{2,\alpha}(Q) \cap C^{1,\alpha}(Q)$ which solves problem (7.4) and satisfies estimate (7.8). □

Acknowledgement. I warmly thank Giuseppe Savaré for stimulating my interest in this problem and for sharing with me several opinions and hints on the topics of the paper. I also thank Filippo Santambrogio for pointing me out reference 33.

References

[1] L. Ambrosio, N. Gigli, G. Savaré, Gradient flows in metric spaces and in the space of probability measures. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2nd ed. (2008).

[2] T. Bakaryan, R. Ferreira, D. Gomes, Some estimates for the planning problem with potential, Nonlinear Differ. Equ. Appl. NoDEA (2021).

[3] J.-D. Benamou, Y. Brenier, A computational fluid mechanics solution to the Monge- Kantrovich mass transfer problem, Numer. Math., 84 (2000), 375–393.

[4] J.-D. Benamou, G. Carlier, F. Santambrogio, Variational mean field games. Active Particles, Volume 1. Birkhäuser, Cham, (2017), 141-171.

[5] P. Cardaliaguet. Weak solutions for first order mean field games with local coupling, in Analysis and geometry in control theory and its applications, Springer 2015, pp. 111–158.
[6] P. Cardaliaguet, P. J. Graber, *Mean field games systems of first order*, ESAIM Control Optim. Calc. Var. 21 (2015), 690-722.

[7] P. Cardaliaguet, P. J. Graber, A. Porretta, D. Tonon *Second order mean field games with degenerate diffusion and local coupling*, NoDEA Nonlinear Differential Equations Appl., 22 (2015), 1287-1317.

[8] P. Cardaliaguet, A. R. Mészáros, F. Santambrogio, *First order mean field games with density constraints: pressure equals price*, SIAM J. Control Optim., 54 (2016), 2672-2709.

[9] P. Cardaliaguet, A. Porretta, *An introduction to Mean Field Game theory*, in: *Mean field games*, 1-158. Lecture Notes in Math., 2281 (CIME Found. Subser.), Springer, Cham, (2020).

[10] C. Clason, D.A. Lorenz, H. Mahler, B. Wirth, *Entropic regularization of continuous optimal transport problems*, arXiv:1906.01333 (2020).

[11] E. DiBenedetto: *Degenerate parabolic equations*, Springer-Verlag, 1993.

[12] L. Chizat, G. Peyre, B. Schmitzer, F.-X. Vialard, *Unbalanced optimal transport: dynamic and Kantorovich formulations*, J. Funct. Anal. 274 (2018), 3090-3123.

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

[14] N. Gigli, L. Tamanini, *Benamou-Brenier and duality formulas for the entropic cost on RCD*$(K,N)$* spaces*, Probability Theory and Related Fields 176 (2020), 1-34.

[15] D. Gilbarg, N. Trudinger. *Elliptic Partial Differential Equations of Second Order*, 2nd ed. Springer, Berlin (1983).

[16] D. Gomes, T. Seneci, *Displacement convexity for first-order mean-field games*, Minimax Theory and Appl. 3 (2018), 261-284.

[17] P. J. Graber, A. R. Mészáros, F. Silva, D. Tonon, *The planning problem in mean field games as regularized mass transport*, Calc. Var. Partial Differential Equations 58 (2019).

[18] O. A. Ladyženskaja, V.A. Solonnikov, N.N. Ural’tseva: *Linear and quasilinear equations of parabolic type*. Translations of Mathematical Monographs, Vol. 23 American Mathematical Society, Providence, R.I. 1967.

[19] O. A. Ladyzhenskaya, N. N. Ural’tseva: *Linear and quasilinear elliptic equations*. Academic Press, New York (1968).

[20] J.-M. Lasry and P.-L. Lions, *Jeux à champ moyen. II-Horizon fini et contrôle optimal*. Comptes Rendus Mathématique, 343 (2006), 679-684.

[21] J.-M. Lasry and P.-L. Lions, *Mean field games*, Japanese journal of mathematics 2 (2007), 229-260.

[22] H. Lavenant, F. Santambrogio, *Optimal density evolution with congestion: $L^\infty$ bounds via flow interchange techniques and applications to variational Mean Field Games*, Comm. P.D.E. 43 (2018), 1761-1802.

[23] G. Lieberman, *Solvability of quasilinear elliptic equations with nonlinear boundary conditions*, Trans. Amer. Math. Soc. 273 (1982), 753-765.

[24] G. Lieberman, *The nonlinear oblique derivative problem for quasilinear elliptic equations*, Nonlinear Anal. T.M.A. 8 (1984), 49-65.

[25] G. Lieberman, *Boundary regularity for solutions of degenerate elliptic equations*, Nonlinear Anal. 12 (1988), 1203-1219.

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

[27] P.-L. Lions, *Cours at Collège de France (2009-2010)*, https://www.college-de-france.fr/site/pierre-louis-lions/course-2009-2010.htm.

[28] R. J. McCann, *A convexity principle for interacting gases*, Adv. Math. 128 (1997), 153-179.
[29] N. Mimikos-Stamatopoulos, S. Munoz, *Regularity of one-dimensional first-order mean field games and the planning problem*, preprint arXiv:2204.06474 (2022).

[30] S. Munoz, *Classical and weak solutions to local first order mean field games through elliptic regularity*, Ann. I. H. Poincaré Anal. Nonlinéaire 39 (2022), 1-39.

[31] C. Orrieri, A. Porretta, G. Savaré, A variational approach to the mean field planning problem, J. Funct. Anal. 277 (2019), 1868-1957.

[32] F. Santambrogio: *Optimal transport for applied mathematicians*. Progress in Nonlinear Differential Equations and their applications 87, Birkhäuser, 2015.

[33] F. Santambrogio, X.-J. Wang, *Convexity of the support of the displacement interpolation: Counterexamples*, Applied Math. Letters 58 (2015), 152-158.

[34] G. Stampacchia, *Le problème de Dirichlet pour les équations elliptiques du second ordre à coefficients discontinus* Ann. Inst. Fourier (Grenoble) 15 (1965), 189–258.

[35] C. Villani: *Topics in Optimal transportation*. Number 58 in Graduate Studies in Mathematics. American Mathematical Soc., 2003.